Teachers and the Evolution of Aggregate Inequality

Anson Zhou*

May 2024

Abstract

This paper develops an overlapping generations model to study the dynamic effects of teacher selection on aggregate income inequality. I propose that there is a two-way relationship between teacher quality and population-wide human capital distribution. On the one hand, declining teacher quality disproportionately hurts children from low-income households, elevating the dispersion of human capital in future generations. On the other hand, a greater dispersion of human capital amplifies (dampens) the decline in teacher quality if the return to human capital among teachers is lower (higher) than that in other occupations. I provide a constructive proof of identification of key model parameters and calibrate the model to match the U.S. data. The results indicate that static changes in teacher selection generate large dynamic effects on the level and dispersion of children's human capital that exceed one-generation estimates.

JEL classification: I24, J24, J31, J45
Keywords: teachers, human capital, inequality

*Anson Zhou: Faculty of Business and Economics, The University of Hong Kong. I am indebted to Juan Carlos Cordoba, Chao Fu, Minseon Park, Mike Ricks, Cailin Slattery, Uta Schöenberg, Michela Tincani, Joanna Venator, Alexander Willén, Alan Yang, Matt Zahn, Lichen Zhang, Xiaodong Zhu for helpful comments and suggestions.

1 Introduction

While teachers account for less than 5% of the workforce, they play a fundamental role in shaping students' achievements and path to upward mobility (Rivkin et al. 2005, Card et al. 2022). It is widely acknowledged that teacher selection has profound implications not only on the equilibrium of the teacher labor market but also on the next generation's outcomes (Hanushek 2011, Chetty et al. 2014c).

While recent research has made significant progress in measuring such implications empirically,[1] some important questions remain unanswered. In particular, how does teacher selection affect the dynamics of income inequality among non-teachers and how will the effects vary as the children being affected grow up, join the labor market, and (some of whom) become teachers themselves? The answers to these questions are challenging to obtain empirically using one-generation estimates because the human capital distribution of workers is time-varying and policy-dependent. Nevertheless, these questions are important for understanding the long-run general equilibrium impacts of teacher selection and the unique role of teachers in shaping the aggregate labor market.

In this paper, I study teacher selection in an overlapping-generations economy with heterogeneous agents and endogenous human capital formation in the spirit of Bénabou (2002) and Daruich (2018). In the environment, I introduce self-selection into teaching or non-teaching occupations depending on the base wages and returns to human capital. After choosing an occupation, adults become parents and make investments in their children's human capital formation. Parental efforts and teacher quality jointly contribute to the human capital of the next generation through a production function that allows for imperfect substitutability between these two inputs. Furthermore, motivated by Tamura (2001), I assume that the effectiveness of an educator, whether she is a teacher or a parent, depends on the relative position of her human capital to the average of the population. When children grow up and become adults, the economy moves on to the next generation.

The key mechanism of this model lies in a two-way relationship between teacher quality and population-wide human capital distribution. To be more specific, consider a scenario

[1] For example, see Jacob et al. 2018, Biasi et al. 2021, Lovenheim and Willén 2019, Lavy 2020, and Tincani 2021.

where decreasing rewards to human capital in the teaching profession pushes the most talented individuals into non-teaching occupations and thus reduces teacher quality. Deteriorating teacher quality harms human capital in the next generation, with the effects being disproportionately larger among low-income families when teacher quality and parental investments are substitutes in the production of human capital.[2] As a result, the economy has lower intergenerational mobility and a greater dispersion of human capital in the next generation. Importantly, a greater dispersion of human capital itself further reduces (raises) teacher quality if the return to human capital among teachers is lower (higher) than that in other occupations. Therefore, *static* changes in teacher quality could be amplified or dampened in the *dynamic* economy depending on the relative returns to human capital across occupations.

To quantify the magnitude of the mechanism, I provide a constructive proof of the identification of key model parameters. The labor market parameters, such as base wages, returns to human capital, and non-pecuniary benefits of teaching and non-teaching professions, are identified using income ratio across occupations, income dispersion within occupations, and teacher share of the labor force calculated using data from the Current Population Survey (CPS). The preference parameters, in particular the preference weight on children's human capital, are identified using data on parental investments from the Panel Study of Income Dynamics (PSID). Lastly, key parameters in the human capital production functions are identified using empirical evidence on teacher value-added (VA) by Chetty et al. (2014a) and the heterogeneous effects of duty-to-bargain (DTB) laws by Lovenheim and Willén (2019).

I use the calibrated model to simulate two counterfactuals. In the first counterfactual, I simulate the transition path after a small reduction of returns to human capital among teachers relative to non-teachers. I show that the outflow of talent from the teaching profession gets propagated over time – the short-run decline in teacher quality is only 60% of that in the long run. Likewise, the short-run effect on intergenerational mobility understates the long-run change by 35%. In addition, changing teacher selection spills over to non-teaching occupations, with the effects approaching the new steady state in three to four generations.

In the second counterfactual, I plot the combination of the mean and standard deviation

[2] See Kotera and Seshadri (2017), Yum (2023), and Agostinelli et al. (2022).

of the population-wide human capital distribution in the long run as I vary the relative skill bias across occupations to give a sense of optimal policy. I find that there is a local trade-off where greater human capital dispersion accompanies a higher population average due to the endogenous response of parental investments.

In Section 6, I explore an alternative measure of skill bias across occupations by leveraging individual-level data from the National Longitudinal Survey of Youth 1979 (NLSY79). I find that the correlation between the Armed Force Qualification Test (AFQT) percentile score with hourly wage is smaller among teachers. This result is consistent with the calibration results in the main analysis where the return to human capital is lower in the teaching profession.

Related Literature

This paper contributes to the large body of empirical literature in labor and education economics that studies the structure of the teacher labor market and its implications on students. The most related papers include Hoxby (1996) and Lovenheim (2009) on teacher unions, Bacolod (2007) on the importance of women's alternative employment opportunities, Lovenheim and Willén (2019), Lavy (2020), Biasi et al. (2021), and Tincani (2021) on collective bargaining and performance-based compensation, and Card et al. (2022) on school quality and minimum teacher salary laws. In various institutional settings, the literature finds that rewarding more effective teachers raises teacher quality and improves children's outcomes, especially for those with disadvantaged backgrounds. To the best of my knowledge, this is the first paper to incorporate these one-generation estimates into an overlapping-generations general equilibrium setting and uncover large dynamic spillover effects into non-teaching professions. The contribution is a dynamic perspective where I take the pool of potential teachers in the future as endogenous.

This paper also builds on the literature that studies the determinants of aggregate inequality and intergenerational mobility. The literature has traditionally focused on factors that affect the demand for education, such as credit constraints (Lee and Seshadri 2019, Caucutt and Lochner 2020), information frictions (Hoxby and Turner 2015), the role of differential fertility (De la Croix and Doepke 2004), and neighborhood effects (Chetty et al.

2014c, Durlauf and Seshadri 2018, Fogli and Guerrieri 2019, Chyn and Daruich 2022).[3] This paper contributes to the less studied literature regarding the supply side of education (e.g., Agostinelli et al. 2021, Fu et al. 2022, and Agostinelli et al. 2022) by showing that teacher market reforms could be powerful instruments that move the economy along the Great Gatsby curve toward lower inequalities and greater intergenerational mobility.

Last, this paper contributes to the literature that studies the aggregate impacts of reward structure in some special occupations, such as government officials (Murphy et al. 1991, Acemoglu 1995) and entrepreneurs (King and Levine 1993, Baumol 1996). This paper contributes to the literature by studying another pivotal occupation in the economy – teachers. Due to teachers' role in the production of human capital for the whole future generation, I show that the aggregate impacts of teacher selection are far-reaching. The framework developed here could be applied to other occupations if the same amplifying mechanism applies and the identifying moments are available in the data.

The rest of the paper is organized as follows. I present the model and discuss its key assumptions in Section 2 and explain the mechanism in Section 3. In Section 4, I present the identification proof and the calibration results. Section 5 contains the main results on policy counterfactuals. Section 6 discusses the alternative measure of occupation-specific returns to human capital. Section 7 concludes.

2 Model

I study an overlapping-generations economy populated by agents that live for two periods – children and adults. Children do not make any decisions. Their human capital is formed through a production function that takes teacher quality and parental investments as inputs. Adults with different levels of human capital supply labor inelastically and self-select into two occupations: teachers and non-teachers (also called workers). After occupation selection, adults become parents and choose child investments to maximize their utility from consumption and preferences on their children's human capital. To simplify notations, I omit the time/generation subscript t from variables.

[3] See Blanden et al. (2023) for a recent summary of the literature.

2.1 Occupation Choice

There exist two occupations $j \in \{1, 2\}$ in the economy where $j = 1$ denotes teachers and $j = 2$ denotes workers. Let $F(h)$ stand for the human capital distribution of the adult population – an endogenous object in the competitive equilibrium.

At the beginning of each period, individual adults with human capital h receive idiosyncratic preference shocks ν_j that follow a Gumbel distribution with scale parameter θ. Then, she solves the following occupation choice problem to maximize utility:

$$\max_{j=1,2} \qquad \alpha_j + \psi_j \log(h) + \mathbb{1}_{j=1}\kappa + \nu_j$$

where α_j is occupation-specific base wage; ψ_j is returns to human capital across occupations; and κ is non-pecuniary benefits of teachers relative to workers.

Individual's occupation choices generate labor supply to teachers and workers. Define $p(h)$ as the share of workers who have human capital h and choose to become teachers. The aggregate labor supply to the teaching profession is thus given by:

$$\pi^s = \int p(h)\, dF(h). \tag{1}$$

2.2 Labor Demand

I assume that workers' baseline wage α_2 and returns to skill ψ_2 are governed by exogenous technologies. Employers of workers also have perfectly elastic labor demand that can accommodate as much labor as supplied.

Teachers' baseline wage α_2 and returns to skill ψ_2, on the other hand, are posted by the government and their salaries are financed by income taxes. Given that the paper focuses on teacher quality, I abstract from changes in teacher quantity by assuming that the labor demand of teachers is fixed at $\pi^d = \overline{\pi}$. The non-pecuniary benefits κ adjust to clear the labor supply and demand for the teaching occupation.

2.3 Teaching Resources

First, following the classic work on teacher quality by Tamura (2001), I assume that the human capital of individuals is transformed into teaching resources $\tilde{h}$ through technology:

$$\tilde{h} = \frac{h}{\overline{h}} \tag{2}$$

where $\overline{h}$ denotes the average human capital of parents in the population. The core idea behind this assumption, which dates back to the seminal work by Galor and Weil (2000), is that as human capital accumulates over time, the qualifications of educators, whether they are teachers or parents, also need to rise in the economy to sustain their value added. Empirically, recent work by Biasi and Ma (2022) on the education-innovation gap confirms the importance of teachers' relative human capital position in the economy for their effectiveness.

Summing up across individuals, I use $\mathcal{Q}$ to denote the aggregate teaching resources, i.e.,

$$\mathcal{Q} = \int p(h) \cdot \tilde{h} \, dF(h). \tag{3}$$

Second and more importantly, I assume that the aggregate teaching resources $\mathcal{Q}$ is distributed uniformly across households, both in terms of teacher-student ratio and teacher quality. Therefore, from an individual household point of view, the amount of teaching resources received is given by

$$q(h) = q = \frac{1}{\overline{\pi}} \cdot \mathcal{Q}$$

where $1/\overline{\pi}$ is the ratio between teachers to parents.

This assumption has two parts. Regarding teacher-student ratio, existing empirical research, such as Hoxby (2000) and Adusumilli et al. (2024) on the Tennessee STAR experiment, Cho et al. (2012) using data from Minnesota, and Angrist et al. (2019) using the Maimonides' rule in Israel, finds little evidence of class size effects on student achievements. Thus, despite the potential correlation between family income and class size, abstracting away from the heterogeneous number of teachers across households is unlikely to lead to significant biases in the results.

Regarding teacher quality, while existing research finds large effects of teacher value added (VA) on student outcomes (e.g., Chetty et al. (2014a)), the sorting between the socioeconomic status of parents and teacher VA is extremely small. Most notably, Chetty et al. (2014b) find that "a $10,000 increase in parent income is associated with less than a 0.0001 SD improvement in teacher VA (measured in student test score SDs)." They argue that sorting is so limited because 85% of the variation in teacher VA is within rather than between schools, and most sorting occurs through the choice of schools – a margin that the model here captures via endogenous parental effort in the child's human capital formation.

2.4 Human Capital Formation

After occupation selection, individuals become parents and have one child. An individual adult with human capital h and occupation j pays linear income taxes, takes teacher quality as given, and solves the following optimization problem:

$$\max_{e \in (0,1)} \quad \log(c) + \beta \mathbb{E}_{\epsilon} \log(h') \tag{4}$$

subject to budget constraint

$$c = w_j(h)(1-e)(1-\tau) \quad \text{where} \quad \log(w_j(h)) = \alpha_j + \psi_j \log(h)$$

and child human capital production function

$$\log(h') = A + \log(\epsilon) + \lambda_1 \log(e\tilde{h}) + \lambda_2 \log(q) + \lambda_3 \log(e\tilde{h}) \log(q) + \rho \log(\tilde{h}) \tag{5}$$

where ϵ is an idiosyncratic ability shock that follows a lognormal distribution

$$\log(\epsilon) \sim \mathcal{N}(-\sigma_{\epsilon}^2/2, \sigma_{\epsilon}^2).$$

As discussed by Agostinelli and Wiswall (2016), the translog production function in Equation (5) has two advantages over the commonly used CES production function (with Cobb-Douglas being a special case). First, it does not restrict the elasticity of substitution between inputs to be constant. Second, it allows the output elasticity with respect to individual inputs to be higher than unity. These properties are useful for the model to match the level and disparities in the impacts of changing teacher quality established in the existing empirical evidence to be discussed in Section 4. Lastly, an additional advantage of the translog production function, as will be shown shortly, is that it allows for an analytical solution of the model.

2.5 Government and Labor Markets

The government collects linear income tax at the rate τ to finance teachers' salaries. Thus, the government budget balance is given by

$$\underbrace{(1-\tau)\cdot\int \underbrace{p(h)(1-e_1(h))}_{\text{hours of teachers}}\underbrace{w_1(h)}_{\text{wage}}\,dF(h)}_{\text{net payments to teachers}} = \underbrace{\tau\cdot\int \underbrace{(1-p(h))(1-e_2(h))}_{\text{hours of workers}}\underbrace{w_2(h)}_{\text{wage}}\,dF(h)}_{\text{net payments by workers}}. \tag{6}$$

2.6 Competitive Equilibrium

Definition: A competitive equilibrium of this economy is a sequence of wages $w_{j,t}(h)$, tax rates τ_t, occupation choices $p_t(h)$, consumption $c_{j,t}(h)$, child investments $e_{j,t}(h)$, and human capital distributions $F_t(h)$ such that households optimize, the government balances budget, and the human capital distribution follows the law of motion governed by individual choices and idiosyncratic shocks.

Definition: A stationary equilibrium is a competitive equilibrium where prices, allocations, and human capital distributions stay invariant over time.

3 Solution and Model Mechanism

This section presents the solution to the stationary equilibrium of the economy and discusses the model mechanism. To save space, the detailed steps of derivation are presented in the Appendix A.

3.1 Model Solution

To generate a closed-form solution to the model, I first assume that the human capital distribution in the stationary equilibrium is lognormal:

$$\log(h) \sim \mathcal{N}(\mu, \sigma^2) \tag{7}$$

As will be shown below, this assumption is self-fulfilling because the lognormality is preserved over time under the optimal behavior of agents in the economy.

Labor Market

Define relative base wage α and relative skill bias ψ as

$$\alpha = \alpha_1 - \alpha_2, \qquad \psi = \psi_1 - \psi_2.$$

The equilibrium conditions of the labor market can be summarized in the following set of equations. First, the conditional probability of becoming teachers $p(h)$ is given by

$$p(h) = \big(\exp(\alpha + \kappa) \cdot h^{\psi}\big)^{\theta} \tag{8}$$

which is increasing in the relative base wage of teachers α and the relative non-pecuniary benefits of teachers κ. The probability is also increasing in the human capital of adults h if and only if the teaching profession has higher returns to human capital, i.e., $\psi > 0$. Parameter θ governs the elasticity of choice probability to payoff differentials, reflecting the role of idiosyncratic taste shocks ϵ in occupation choices.

Integrating the occupation choice probability $p(h)$ over the population human capital distribution $F(h)$ gives the aggregate share of teachers in the labor force:

$$\overline{\pi} = \exp(\theta(\alpha + \kappa)) \cdot \exp(\theta\psi\mu + (\theta\psi\sigma)^2/2). \tag{9}$$

We can also derive the wage ratio between teachers and non-teachers in the economy:

$$\frac{\mathbb{E}(w|j=1)}{\mathbb{E}(w|j=2)} = \exp(\alpha) \cdot \exp(\psi\mu + (\sigma\psi)^2(1+2\theta)/2) \tag{10}$$

which provides a measure of inequality across occupations.

In terms of inequality within occupations measured using the coefficient of variations (CV), the model generates

$$\mathbb{CV}(w|j=1) = \sigma\psi_1 \qquad \text{and} \qquad \mathbb{CV}(w|j=2) = \sigma\psi_2. \tag{11}$$

One important thing to note here is that within-occupation inequality is not only proportional to the returns to human capital ψ_j but also scales with the dispersion of the underlying human capital in the population σ. Because teacher quality shapes σ through human capital production, this condition provides an important channel where teacher selection affects inequalities within non-teaching occupations over time.

Teaching Resources

Using the definition of aggregate teaching resource $\mathcal{Q}$ and the assumption on how it is distributed, there is a closed-form expression of the amount of teacher quality received at the household level:

$$q = \exp(\theta\psi\sigma^2) \tag{TS}$$

The relationship above illustrates the channel from human capital dispersion in the population σ to teacher quality q through selection, and hence named the teacher selection (TS) equation. In particular, teacher quality q increases in human capital dispersion σ if and only if the return to human capital is higher in the teaching profession, i.e., $\psi > 0$, so that teachers are selected from the upper spectrum of the human capital distribution.

Another interpretation of Equation (TS) is that teacher quality q is more sensitive to teacher selection ψ in an economy with greater dispersion of human capital σ because (TS) implies that

$$\frac{d\log(q)}{d\psi} = \theta\sigma^2.$$

That is to say, selection is a more powerful instrument when there is more heterogeneity in the underlying population.

Lastly, while the aggregate teaching resource $\mathcal{Q}$ is distributed uniformly, teachers are nevertheless heterogeneous. The standard deviation of teachers' effectiveness is given by

$$\text{std}(\tilde{h}|j=1) = \sigma q. \tag{12}$$

In Section 4, I show that this equation is helpful in the identification of the human capital production function.

Child Investments

Using the first-order condition of the household, it can be shown that parental investment in children's human capital formation is

$$e(h) = \beta(\lambda_1 + \lambda_3 \log(q)) \qquad \text{for all } h. \tag{13}$$

Intuitively, the amount of investment $e(h)$ is increasing in the preference weight on children's human capital β and the direct productivity of the investment λ_1. Whether $e(h)$ increases or decreases with teacher quality q depends on whether these two inputs are complements or substitutes in the production function governed by λ_3.

Plugging the optimal investment (13) back into the human capital production function (5), we can obtain the relationship between parents and their children's human capital:

$$\begin{aligned} \log(h') = A + \log(\epsilon) &+ (\rho + \lambda_1 + \lambda_3 \log(q)) \log(\tilde{h}) \\ &+ \lambda_1 \log(e) + \lambda_2 \log(q) + \lambda_3 \log(e) \log(q) \end{aligned} \tag{14}$$

Using Equation (14), we can also derive a closed-form expression for the intergenerational elasticity of human capital (IGE):

$$\text{IGE} = \frac{\partial \log(h')}{\partial \log(h)} = \rho + \lambda_1 + \lambda_3 \log(q) \tag{15}$$

As can be seen, when an improvement in teacher quality q has smaller positive impacts on high-income households, i.e., $\lambda_3 < 0$, such improvement reduces IGE and raises intergenerational mobility.

Taking the derivative of Equation (14) and using the Envelope theorem, we have a closed-form expression of the average impact of a marginal improvement of teacher quality q on children's human capital in the population:

$$\frac{\partial \log(h')}{\partial \log(q)} = \lambda_3 \log(\tilde{h}) + \lambda_2 + \lambda_3 \log(e) \Longrightarrow \mathbb{E}\left[\frac{\partial \log(h')}{\partial \log(q)}\right] = \lambda_2 + \lambda_3 \log(e) \tag{16}$$

which is synonymous with teachers' value added (VA) in the literature.

Lastly, due to the interaction between parental input and teacher effectiveness, the impact of teacher quality on children's human capital depends on household characteristics. Taking an additional derivative with respect to parental human capital,

$$\frac{\partial^2 \log(h'))}{\partial \log(q) \partial \log(h)} = \lambda_3 \tag{17}$$

Equations (15), (16), and (17), provided that we can measure their left-hand-sides, are extremely useful in helping identify the key parameters in the human capital production function $\{\lambda_1, \lambda_2, \lambda_3\}$.

Dynamics of Human Capital Distribution

Equation (14) indicates that a child's (log) human capital $\log(h)$ is a sum of an idiosyncratic shock ϵ which follows a normal distribution and a linear transformation of the parent's (log) human capital $\log(h)$ which is also normal by assumption. Because the sum of independent normal distributions is also normal, Equation (14) provides an analytical characterization of the human capital distribution dynamics in the aggregate.

In particular, taking expectation on both sides of (14) yields:

$$\mu' = A - \frac{\sigma_\epsilon^2}{2} + \lambda_1 \log(e) + \lambda_2 \log(q) + \lambda_3 \log(e) \log(q). \tag{18}$$

Furthermore, the evolution of the human capital dispersion follows

$$(\sigma')^2 = \sigma_\epsilon^2 + (\underbrace{\rho + \lambda_1 + \lambda_3 \log(q)}_{\text{IGE}})^2 \cdot \sigma^2. \tag{19}$$

In the stationary competitive equilibrium of the economy, Equation (19) implies that the relationship between human capital dispersion and teacher quality follows a dispersion formation (DF) equation:

$$\sigma^2 = \frac{\sigma_\epsilon^2}{1 - (\rho + \lambda_1 + \lambda_3 \log(q))^2} \tag{DF}$$

When IGE is decreasing in teacher quality q, Equation (DF) indicates that human capital dispersion σ^2 in the stationary equilibrium is also decreasing in q.

3.2 Model Mechanism

The main mechanism of the model can be summarized in a phase diagram on the relationship between human capital dispersion σ^2 and teacher quality q. As discussed above, these two variables affect each other through the teacher selection (TS) equation and the dispersion formation (DF) equation. While (DF) is downward sloping, whether (TS) is upward or downward sloping depends on the relative returns to human capital across occupations governed by parameter ψ.

Figure 1 plots the case where $\psi > 0$. Point A is the original steady state (σ_A^2, q_A). The dashed line plots the teacher selection (TS) equation in the case where ψ is slightly reduced. As can be seen, when the economy wakes up in the original steady state and faces a shock where ψ falls, teacher quality q deteriorates from q_A to q_C in the short run. When the economy reaches the new steady-state at point B, however, teacher quality recovers to q_B but the economy faces a greater dispersion in human capital σ_B^2.

Figure 2 plots the scenario where $\psi < 0$. When ψ is slightly reduced in this case, the short-run impacts on teacher quality (from q_A to q_C) understates the long-run impacts (from q_A to q_B). Teacher selection and dispersion formation reinforce each other and generate a vicious loop. Importantly, Equations (10) and (11) indicate that the increase in σ^2 puts pressure on inequalities both within and across occupations.

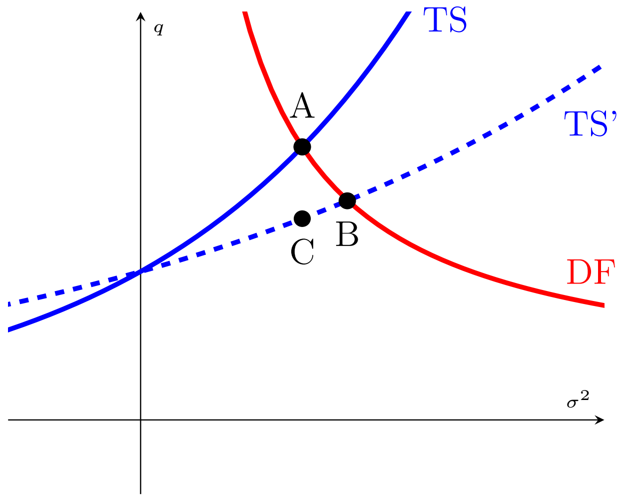


Figure 1: Phase diagram when $\psi > 0$

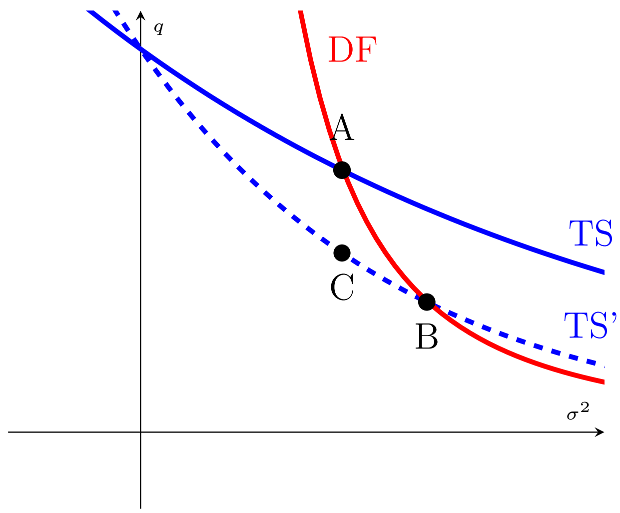


Figure 2: Phase diagram when $\psi < 0$

In both cases, inspecting points A, B, and C shows the difference between short- and long-run impacts of changing teacher selection. The full transition path, on the other hand, can be analytically characterized by repeatedly applying Equations (TS) and (DF) under the new ψ.

To sum up, with human capital formation, static changes in teacher quality can be either amplified or dampened depending on the relative returns to human capital across occupations. The magnitude of the channel depends critically on parameters characterizing the labor market $\{\psi, \theta\}$ and the human capital production $\{\rho, \lambda_1, \lambda_3\}$. The main task of the quantitative section is therefore to identify these parameters from the data.

4 Calibration

This section discusses the calibration strategy and results.

4.1 Calibration Strategy

The set of parameters to be pinned down includes:

$$\underbrace{\alpha_1, \alpha_2, \psi_1, \psi_2, \kappa,}_{\text{labor market}} \underbrace{\theta, \beta}_{\text{preference}}, \underbrace{\lambda_1, \lambda_2, \lambda_3, A, \rho, \sigma_\epsilon}_{\text{human capital production}}, \underbrace{\tau}_{\text{taxes}}$$

I divide these parameters into three groups:

1. **Parameters to be normalized.**

   I set $A = 1$ because human capital in the model does not have an inherent scale. In addition, because only the relative base wage $\alpha \equiv \alpha_1 - \alpha_2$ and the relative return to human capital $\psi \equiv \psi_1 - \psi_2$ matters for individual decisions and aggregate outcome, I normalize $\alpha_2 = 0$ and $\psi_2 = 1$. Therefore, identifying α_1 and ψ_1 is equivalent to identifying α and ψ.

2. **Parameters to be calibrated exogenously.**

   I follow Hsieh et al. (2019) and set $\theta = 2$. This parameter governs the extensive margin elasticity of labor supply with respect to a wage change across occupations.

   Following Lefgren et al. (2012), I choose ρ to be 60% of the observed intergenerational elasticity of income (IGE). This puts an upper bound on how much the channels included in the model, in particular parental investment and its interaction with teacher quality, account for intergenerational human capital persistence.

3. **Parameters to be identified using data.**

   The remaining parameters are $\alpha_1, \psi_1, \kappa, \beta, \lambda_1, \lambda_2, \lambda_3, \sigma_\epsilon, \tau$. Here, I provide a constructive proof of identification with the following steps:

   i As ψ_2 is normalized to be 1, we can recover $\sigma = \mathbb{CV}(w|j = 2)$ from the second part of Equation (11).

ii Once we know σ, we can back out ψ_1 using the first part of Equation (11). We can also back out σ_ϵ^2 from Equation (DF) provided that we can measure IGE.

iii Given that we have derived σ and $\psi \equiv \psi_1 - \psi_2$, we can calculate teacher quality q using Equation (TS).

iv Equation (17) provides a direct way to identify λ_3 from the differential impacts of changing teacher quality across households.

v Once λ_3 is known, Equation (16) allows us to back out λ_2 from the average teacher value-added provided that we can measure parental input e.

vi Given that we know λ_3 and q, we can compute λ_1 from Equation (15).

vii Because we know λ_1, λ_3, and q, we can back out β from Equation (13).

viii Equation (18) allows us to calculate μ.

ix Once μ is known, Equation (10) provides a direct way to back out α provided that we can measure wage differentials across occupations in the data.

x Knowing α, we can back out the relative non-pecuniary benefits κ from Equation (9) provided that we can measure the share of teachers in the labor force $\overline{\pi}$.

xi Tax parameter τ is backed out from the government budget balance (6).

4.2 Data Moments and Calibration Results

In this section, I discuss the data moments used in the identification and the calibration results.

Table 1 contains the set of moments targeted in the calibration. These moments come from several sources, and if possible, reflect the underlying data generating process in the U.S. around the early 2000s.

Among these moments, the ones related to the labor market of teachers versus non-teachers are collected from the Current Population Survey Annual Social and Economic Supplement (CPS-ASEC). Further details regarding the data and sample selection procedures are relegated to Appendix B.

I target the amount of child investment as a share of total resources using moments from the Panel Study of Income Dynamics Child Development Supplement (PSID-CDS) and the Consumer Expenditure Survey (CEX). In particular, Daruich (2018) reports that the average weekly quality time spent with children is about 17.8 hours in the PSID-CDS sample. Dividing the estimate by 18 hours a day net of sleep time, 7 days a week, and the presence of two parents gives the moment condition $e = 0.07$. Regarding monetary investment, Daruich (2018) reports that the average expenditures on two children are \$3,848 in the 1996-2000 sample. Dividing the estimate by the median household income (\$42,148 in 2000) results in the moment condition $e = 0.09$. Taking the average of the two conditions, I target $e = 0.08$ in the calibration.

The moment on the correlation of income across generations, which is the same as the correlation between their human capital in this model, comes from Chetty et al. (2014c) where the authors link parental and children's income using administrative data. The value of 0.34 is consistent with other papers in the literature, such as Lee and Seshadri (2019).

I target the average effect of teacher quality using estimates by Chetty et al. (2014a) where the authors report that a one standard deviation increase in teacher quality in a single grade raises annual earnings by 1.3 percent. I make two adjustments to make this estimate comparable to the model object $\mathbb{E}(\partial \log(h')\partial \log(q))$. First, I multiply the estimate by 12 to reflect the number of years of compulsory schooling. Second, I divide it by σq because it represents the standard deviation of teachers' effectiveness in the model (see Equation (12)).[4]

Lastly, I target the differential effect of teacher quality using the evidence from the staggered adoption of duty-to-bargain (DTB) laws, i.e., mandates that school districts had to negotiate in good faith with a teacher union. Lovenheim and Willén (2019) construct exposures to DTB laws at the individual level and link the exposure measures to students' lifetime outcomes such as education, labor supply, and earnings. They find that DTB laws reduce the earnings of affected students. Relative to Black and Hispanic students, White and Asian students have a 60% higher level of parental income on average (more details in

[4]Note that the values of σ and q have already been identified in previous steps when we use the information on teacher value added to identify λ_2.

Appendix B) and witnessed an 80% smaller effect of DTB laws on earnings. As shown in Section 3.1, the ratio of these numbers informs the elasticity of substitution between parental and teacher inputs in human capital formation.[5]

Applying the constructive identification steps provided above, the calibration results are presented in Table 2. The fact that $\psi_1 < \psi_2$ in the calibration results implies that the propagation mechanism is represented by Figure 2. Parameter $\lambda_3 < 0$ indicates that parental input and teacher quality are substitutes, consistent with prior research in the literature such as Kotera and Seshadri (2017) and Yum (2023). Also, note that $\lambda_1, \lambda_2 < 0$ does not mean that the returns to parental input or teacher quality are negative because one also needs to consider the interaction term to fully account for the total effects. For example, $\partial \log h' / \partial e = \lambda_1 + \lambda_3 \log(q)$ with the second part $\lambda_3 \log(q)$ being positive in the baseline calibration.

5 Counterfactuals

I use the model to analyze two counterfactuals.

5.1 Dynamic Effects of Teacher Selection

In the first counterfactual, I decrease the relative skill bias ψ by 0.01 from the benchmark value. In reality, this change could come from technological changes such as an increase in ψ_2 due to the convexification of the labor market (Katz and Murphy 1992) or from institutional changes such as a reduction in ψ_1 due to the enactment of collective bargaining agreements (Hoxby 1996). While the immediate impacts of such change on the teacher labor market, or labor market in general, have been studied before, e.g., Autor et al. (2020) and Biasi et al. (2021), the model offers a unique perspective in extrapolating its dynamic consequences, and

[5] One challenging question is about how to interpret the differential effects of unionization across households. Within the context of the model, such differences inform the degree of imperfect substitution of private and public inputs, but one could also imagine a scenario where such differences reflect elements like incomplete markets (e.g., heterogeneous degrees of borrowing constraints). This concern is important if one wants to understand the best policy responses to such disparities and deserves to be explored in a separate paper. The goal of this paper, however, is different given that I am focusing on the dynamic effects of teacher selection. For that purpose, a mechanical model where I assume heterogeneous treatment effects without specifying the micro-foundation will generate the same dynamic predictions as the approach currently used.

Table 1: Data Moments

Object	Interpretation	Value	Source
$\overline{\pi}$	Share of teachers in the labor force	0.045	CPS-ASEC
$\mathbb{CV}(w\|j=1)$	Coefficient of variation of income among teachers	0.52	CPS-ASEC
$\mathbb{CV}(w\|j=2)$	Coefficient of variation of income among non-teachers	0.75	CPS-ASEC
$\mathbb{E}(w\|j=1)/\mathbb{E}(w\|j=2)$	Income ratio between teachers and non-teachers	1.03	CPS-ASEC
e	Child investments as a share of total resources	0.07	Daruich (2018)
$d\log(h')/d\log(h)$	Intergenerational elasticity of income	0.344	Chetty et al. (2014c)
$\mathbb{E}(\partial\log(h')\partial\log(q))$	Average effect of teacher quality	0.013	Chetty et al. (2014a)
$\partial^2\log(h'))/(\partial\log(q)\partial\log(h))$	Differential effect of teacher quality	misc.	Lovenheim and Willén (2019)

Notes: This table displays the list of moments used in the identification.

Table 2: Calibration Results

Parameter	Interpretation	Value
$\{\alpha_1, \psi_1\}$	base wage and return to human capital among teachers	$\{0.35, 0.69\}$
$\{\alpha_2, \psi_2\}$	base wage and return to human capital among non-teachers	$\{0, 1\}$
κ	relative non-pecuniary benefits	-1.5
θ	taste shock dispersion	2
β	preference weight on child's human capital	0.71
$\{\lambda_1, \lambda_2, \lambda_3\}$	human capital production parameter	$\{-0.34, -2.91, -1.31\}$
A	human capital scale	1
ρ	exogenous human capital persistence	0.23
σ_ϵ	ability shock dispersion	0.71
τ	budget-clearing tax rate	0.05

Notes: This table displays the calibration results.

in particular, through the teacher selection channel.

Figure 3 collects the transition path of several key variables in the model after the decrease in ψ takes place at $t = 0$.

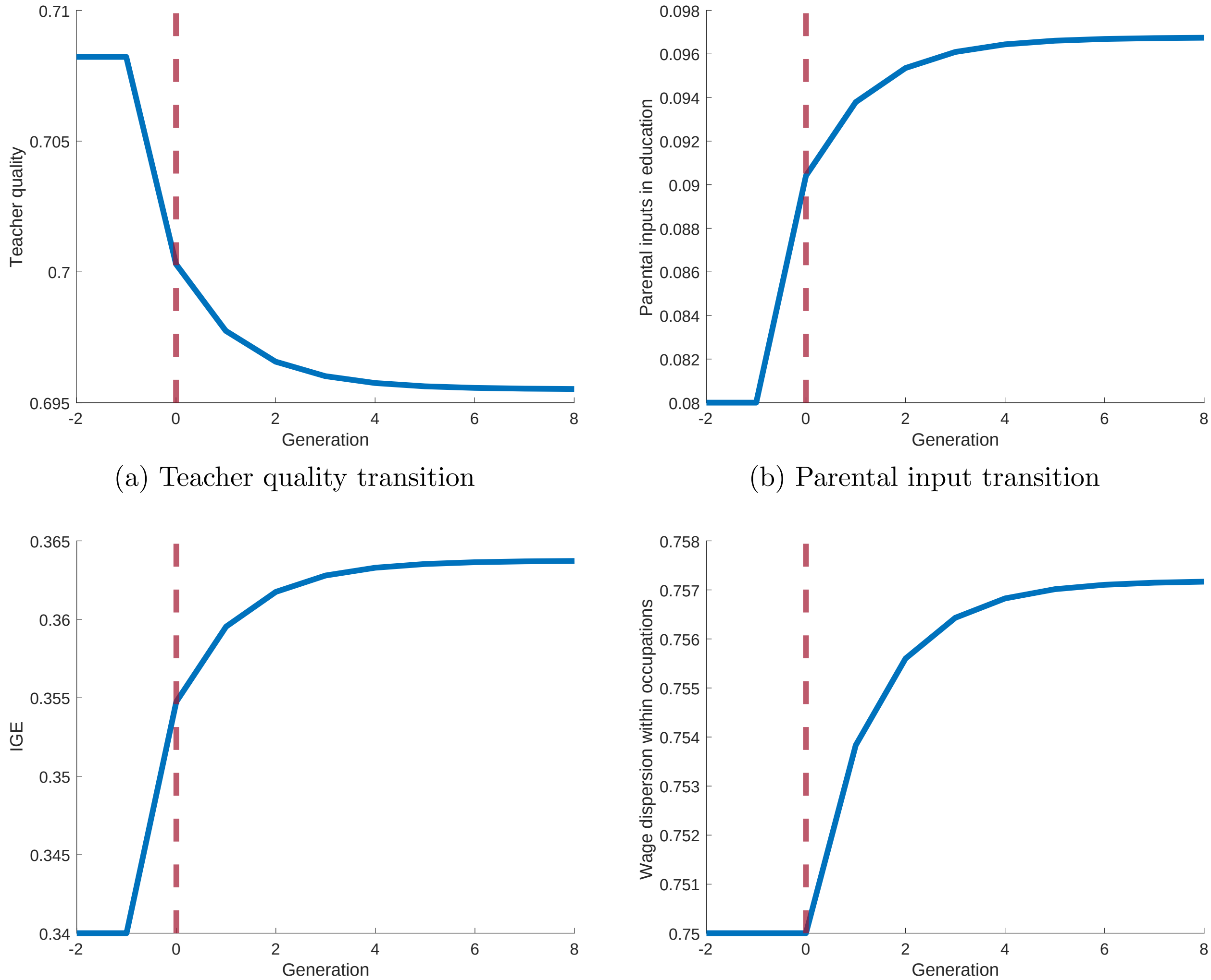
Regarding short-run impacts, Figure 3a indicates that teacher quality q deteriorates immediately when ψ falls as individuals with high human capital choose to leave the teaching profession and seek employment elsewhere. As teacher quality declines, Figure 3b indicates that parents optimally step up in their efforts to make up for the losses in education resources from teachers. Because parents have different levels of human capital and hence heterogeneous ability to compensate, the changing composition of educational resources has distributive consequences, elevating educational inequalities and pushing the IGE from 0.34 up to 0.355 at $t = 0$ (see Figure 3c). The greater degree of intergenerational persistence, in turn, sustains a larger dispersion of human capital in the cross-section (see Figure 3d) and higher inequalities within teaching and non-teaching occupations.[6]

In addition to these short-run impacts which are partly targeted in the calibration, the main pay-off of the model structure comes in regarding the dynamic impacts along the transition path. In particular, due to the feedback effects of the Equations (TS) and (DF) illustrated in Figure 2, the short-run reduction in teacher quality q is only 60% of the changes in the long run (see Figure 3a). Likewise, the short-run effect on IGE understates the long-run level by 35%. The effects on within-occupation inequalities also unfold gradually, occurring after $t = 1$ and approaching the new steady state in three to four generations. These propagation effects will be neglected by policymakers unless they take into account the changing human capital distribution in the population and how it affects future teacher selection.

5.2 Mean-Variance Trade-off

In the second counterfactual, I plot the tuple of the mean and variance of the population human capital distribution in the long run across a range of ψ.

[6]The model also generates a transition path of income inequality across occupations according to Equation (10), but the magnitude of such changes is not economically significant. This is because while a decrease in ψ reduces the relative wage of teachers to non-teachers for individuals with high human capital, it simultaneously raises the relative wage for individuals with low human capital. The overall impact is rather small in the equilibrium.



(a) Teacher quality transition

(b) Parental input transition

(c) IGE transition

(d) Wage dispersion transition

Figure 3: Transition path after a small decrease in ψ

In particular, Figure 4 traces the frontier of μ and σ when I perturb the value of σ from the original level. When ψ falls, human capital dispersion σ increases, but interestingly, average human capital μ also rises. On the other hand, when ψ increases, the effects go in the other direction.

The reason why σ is decreasing in ψ is intuitive and has been discussed in the first counterfactual. To better understand why μ is also decreasing in ψ locally, I consider the

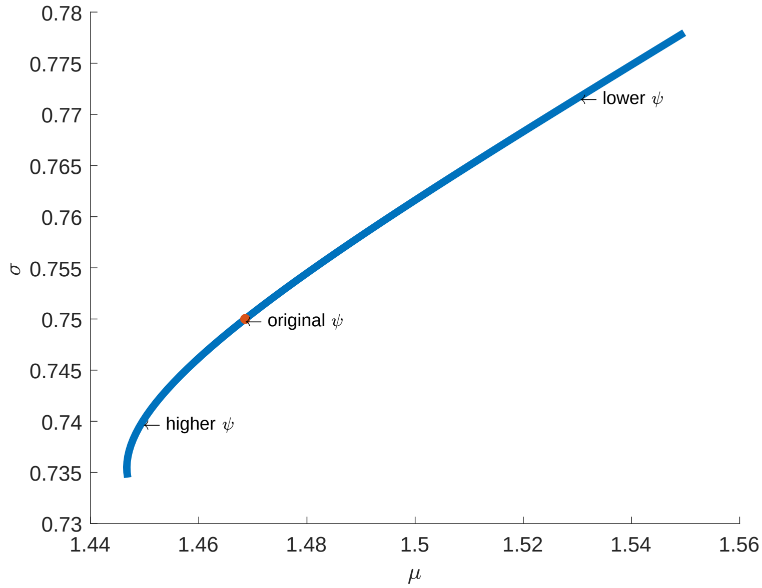


Figure 4: Mean-dispersion frontier

model-based decomposition

$$\begin{aligned}\mu &= A - \frac{\sigma_\epsilon^2}{2} + \lambda_1 \log(e) + \lambda_2 \log(q) + \lambda_3 \log(e)\log(q) \\ &= \underbrace{A - \frac{\sigma_\epsilon^2}{2}}_{\text{constant}} + \underbrace{\left(\lambda_1 \log(e) + \frac{1}{2}\lambda_3 \log(e)\log(q)\right)}_{\text{contribution by } e} + \underbrace{\left(\lambda_2 \log(q) + \frac{1}{2}\lambda_3 \log(e)\log(q)\right)}_{\text{contribution by } q}\end{aligned} \tag{20}$$

where I split the interaction term half-half into "contribution by e" and "contribution by q". Figure 5 plots the two parts and their sum as I by changes in ψ. As can be seen, when ψ falls, teacher quality deteriorates but parental investment rises to compensate. Quantitatively, the model predicts that the rise in e overcompensates the fall in q in the long-run steady state, and thus the economy has a higher μ accompanying the higher σ.

To the extent that the social planner may prefer higher mean μ and lower dispersion σ, i.e., the indifference curve is increasing in the south-east direction of the diagram, there is a local mean-variance trade-off. A full characterization of the optimal policy requires taking a stance on the social welfare function of agents within and across generations.

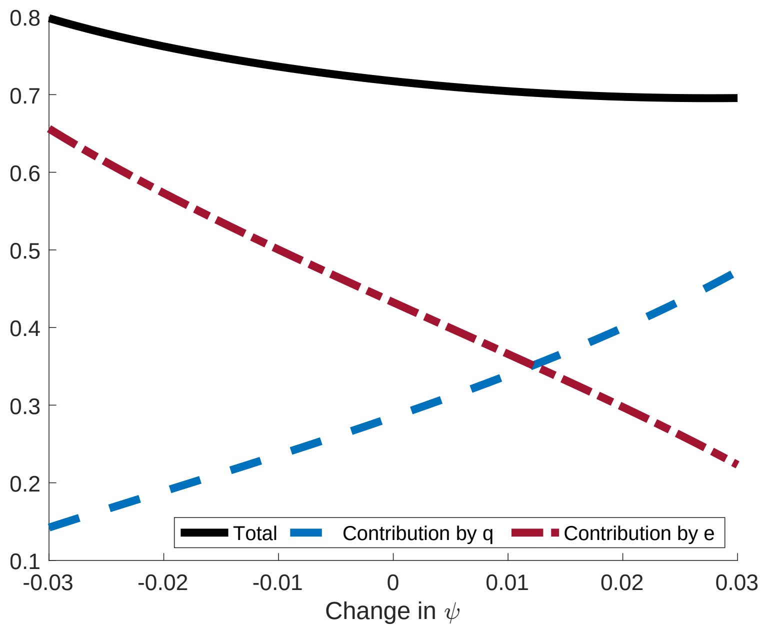


Figure 5: Source of human capital formation

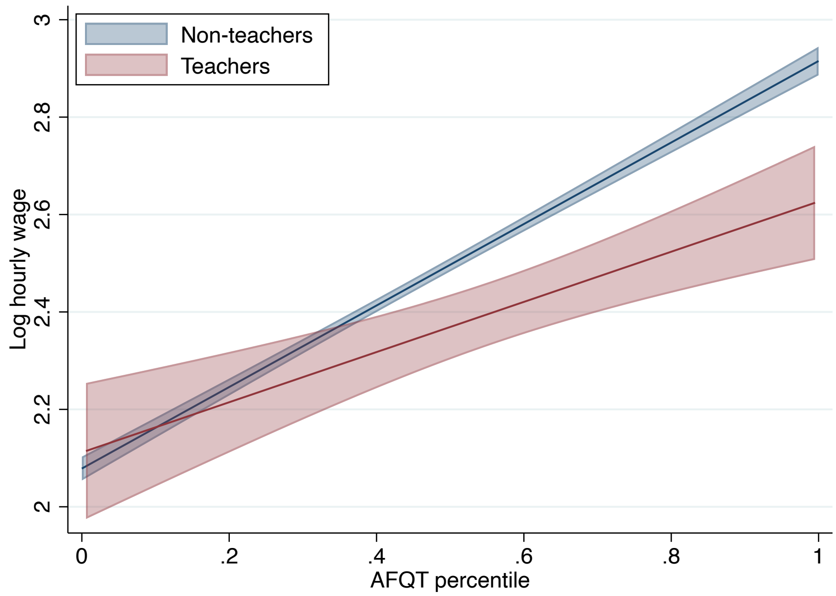
6 Alternative Measure of Returns to Human Capital

In Section 4, the two important parameters governing skill bias across occupations ψ_1 and ψ_2 are identified using aggregate moments on the income dispersion in teaching and non-teaching occupations respectively. In this section, I take a more micro-level approach by directly regressing wage on measures of cognitive ability using the Armed Forces Qualification Test (AFQT) percentile score from the National Longitudinal Study of Youth 1979 (NLSY79) data.

The NLSY79 survey tracks a cohort of individuals aged 14 to 22 when they were initially interviewed in 1979. The survey collects their labor market history, including weeks worked, occupation codes, and hourly wages. For each individual, their cognitive ability was assessed in 1980 through ten intelligence tests known as the Armed Services Vocational Aptitude Battery (ASVAB), and a summarizing measure known as the Armed Forces Qualification Test (AFQT) percentile score was computed. With some caveats, the AFQT score was a commonly used measure of cognitive skills in the literature (e.g., Neal and Johnson 1996). I restrict the sample to college-educated individuals who worked at least 30 weeks on the primary job last year with an hourly wage of at least one dollar.

Figure 6 plots the relationship between the AFQT score and the hourly wage of individuals in the data. As can be seen, for both teachers and non-teachers, the hourly wage is positively correlated with the AFQT score, but the correlation is stronger among non-teachers.

Figure 6: Relationship between the AFQT Score and Hourly Wage



Notes: This figure plots the relationship between the AFQT percentile score and log hourly wage in year 1996 across occupations for individuals in the NLSY79 sample. The line plots the best linear fitted value. The shaded area plots the 90% confidence interval around the fitted value.

To show this pattern more systematically, I run the following regressions:

$$Y_{i,t} = \alpha_{j,t} + \Psi_{j,t} \cdot \mathrm{AFQT}_i + \varepsilon_{i,t} \tag{21}$$

where i indexes individuals, $j \in \{1,2\}$ indexes teachers and non-teachers respectively, t represents survey year, and $Y_{i,t}$ is the log of hourly wage. I use the notation $\Psi_{j,t}$ because the independent variable AFQT_i denotes skill percentiles instead of skill levels, hence the interpretation of the coefficient is a little different from the occupation-specific skill bias $\psi_{j,t}$ in the model.

Table 3 reports the regression results. As can be seen, AFQT percentiles are strongly correlated with hourly wage. For example, a one percentile increase in the ranking of AFQT score is correlated with a 0.515% higher hourly wage. Importantly, the regression results suggest that the coefficient is larger in non-teaching occupations ($j = 2$) than that among

Table 3: Regression Results

	$t = 1996$		$t = 2006$	
	$j = 1$	$j = 2$	$j = 1$	$j = 2$
$\Psi_{j,t}$	0.515	0.837	0.827	0.927
	(0.113)	(0.024)	(0.122)	(0.029)
# Observations	240	2490	227	2193

Notes: This table displays the results of regression (21). Subscript $j \in \{1, 2\}$ indexes teachers and non-teachers respectively. Standard errors in parentheses.

teacher ($j = 1$). Across the two waves of data, the ratio $\Psi_{1,t}/\Psi_{2,t}$ is on average 0.754 which is quite close to the ratio of skill bias $\psi_1/\psi_2 = 0.69$ in the model.

To sum up, using more direct measures of ability, empirical estimates using micro-level wage data confirm that teachers have a more compressed wage distribution than non-teachers, translating to a smaller skill bias in the model.

7 Conclusion

Teacher selection affects who becomes teachers among the current population and shapes the distribution of human capital in the next generation. This, in turn, determines the occupational sorting among the next generation.

This paper develops an overlapping generations model to study this feedback mechanism. With assumptions on the human capital production function and how aggregate teaching resource is distributed, I show that the model is highly tractable with analytical characterizations. I then provide a constructive proof of identification of key model parameters and calibrate the model to match the U.S. data. I find that static changes in teacher selection generate large dynamic effects on the level and dispersion of children's human capital that exceed one-generation estimates. Furthermore, there is a local mean-variance trade-off regarding the long-run human capital distribution.

References

Daron Acemoglu. Reward structures and the allocation of talent. *European Economic Review*, 39(1):17–33, 1995.

Karun Adusumilli, Francesco Agostinelli, and Emilio Borghesan. Heterogeneity and endogenous compliance: Implications for scaling class size interventions. Technical report, National Bureau of Economic Research, 2024.

Francesco Agostinelli and Matthew Wiswall. Estimating the technology of children's skill formation. Technical report, National Bureau of Economic Research, 2016.

Francesco Agostinelli, Margaux Luflade, Paolo Martellini, et al. On the spatial determinants of educational access. Technical report, 2021.

Francesco Agostinelli, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti. When the great equalizer shuts down: Schools, peers, and parents in pandemic times. *Journal of public economics*, 206:104574, 2022.

Joshua D Angrist, Victor Lavy, Jetson Leder-Luis, and Adi Shany. Maimonides rule redux. *American Economic Review: Insights*, 1(3):309–324, 2019.

David Autor, Claudia Goldin, and Lawrence F Katz. Extending the race between education and technology. In *AEA Papers and Proceedings*, volume 110, pages 347–351. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2020.

Marigee P Bacolod. Do alternative opportunities matter? the role of female labor markets in the decline of teacher quality. *The Review of Economics and Statistics*, 89(4):737–751, 2007.

William J Baumol. Entrepreneurship: Productive, unproductive, and destructive. *Journal of Business Venturing*, 11(1):3–22, 1996.

Roland Bénabou. Tax and education policy in a heterogeneous-agent economy: What levels of redistribution maximize growth and efficiency? *Econometrica*, 70(2):481–517, 2002.

Barbara Biasi and Song Ma. The education-innovation gap. Technical report, National Bureau of Economic Research, 2022.

Barbara Biasi, Chao Fu, and John Stromme. Equilibrium in the market for public school teachers: District wage strategies and teacher comparative advantage. Technical report, National Bureau of Economic Research, 2021.

Jo Blanden, Matthias Doepke, and Jan Stuhler. Educational inequality. In *Handbook of the Economics of Education*, volume 6, pages 405–497. Elsevier, 2023.

David Card, Ciprian Domnisoru, and Lowell Taylor. The intergenerational transmission of human capital: Evidence from the golden age of upward mobility. *Journal of Labor Economics*, 40(S1):S39–S95, 2022.

Elizabeth M Caucutt and Lance Lochner. Early and late human capital investments, borrowing constraints, and the family. *Journal of Political Economy*, 128(3):1065–1147, 2020.

Raj Chetty, John N Friedman, and Jonah E Rockoff. Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–2679, 2014a.

Raj Chetty, John N Friedman, and Jonah E Rockoff. Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American economic review*, 104(9): 2593–2632, 2014b.

Raj Chetty, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, 129(4):1553–1623, 2014c.

Hyunkuk Cho, Paul Glewwe, and Melissa Whitler. Do reductions in class size raise students' test scores? evidence from population variation in minnesota's elementary schools. *Economics of Education Review*, 31(3):77–95, 2012.

Eric Chyn and Diego Daruich. An equilibrium analysis of the effects of neighborhood-based interventions on children. Technical report, National Bureau of Economic Research, 2022.

Diego Daruich. The macroeconomic consequences of early childhood development policies. *FRB St. Louis Working Paper*, (2018-29), 2018.

David De la Croix and Matthias Doepke. Public versus private education when differential fertility matters. *Journal of Development Economics*, 73(2):607–629, 2004.

Steven N Durlauf and Ananth Seshadri. Understanding the great gatsby curve. *NBER Macroeconomics Annual*, 32(1):333–393, 2018.

Alessandra Fogli and Veronica Guerrieri. The end of the american dream? inequality and segregation in us cities. Technical report, National Bureau of Economic Research, 2019.

Chao Fu, Junjie Guo, Adam J Smith, and Alan Sorensen. Students' heterogeneous preferences and the uneven spatial distribution of colleges. *Journal of Monetary Economics*, 129:49–64, 2022.

Oded Galor and David N Weil. Population, technology, and growth: From malthusian stagnation to the demographic transition and beyond. *American economic review*, 90(4): 806–828, 2000.

Eric A Hanushek. The economic value of higher teacher quality. *Economics of Education Review*, 30(3):466–479, 2011.

Caroline M Hoxby. The effects of class size on student achievement: New evidence from population variation. *The quarterly journal of economics*, 115(4):1239–1285, 2000.

Caroline M Hoxby and Sarah Turner. What high-achieving low-income students know about college. *American Economic Review*, 105(5):514–517, 2015.

Caroline Minter Hoxby. How teachers' unions affect education production. *The Quarterly Journal of Economics*, 111(3):671–718, 1996.

Chang-Tai Hsieh, Erik Hurst, Charles I Jones, and Peter J Klenow. The allocation of talent and us economic growth. *Econometrica*, 87(5):1439–1474, 2019.

Brian A Jacob, Jonah E Rockoff, Eric S Taylor, Benjamin Lindy, and Rachel Rosen. Teacher applicant hiring and teacher performance: Evidence from dc public schools. *Journal of Public Economics*, 166:81–97, 2018.

Lawrence F Katz and Kevin M Murphy. Changes in relative wages, 1963–1987: supply and demand factors. *The quarterly journal of economics*, 107(1):35–78, 1992.

Robert G King and Ross Levine. Finance, entrepreneurship and growth. *Journal of Monetary economics*, 32(3):513–542, 1993.

Tomoaki Kotera and Ananth Seshadri. Educational policy and intergenerational mobility. *Review of Economic Dynamics*, 25:187–207, 2017.

Victor Lavy. Teachers' pay for performance in the long-run: The dynamic pattern of treatment effects on students' educational and labour market outcomes in adulthood. *The Review of Economic Studies*, 87(5):2322–2355, 2020.

Sang Yoon Lee and Ananth Seshadri. On the intergenerational transmission of economic status. *Journal of Political Economy*, 127(2):855–921, 2019.

Lars Lefgren, David Sims, and Matthew J Lindquist. Rich dad, smart dad: Decomposing the intergenerational transmission of income. *Journal of Political Economy*, 120(2):268–303, 2012.

Michael F Lovenheim. The effect of teachers' unions on education production: Evidence from union election certifications in three midwestern states. *Journal of Labor Economics*, 27(4):525–587, 2009.

Michael F Lovenheim and Alexander Willén. The long-run effects of teacher collective bargaining. *American Economic Journal: Economic Policy*, 11(3):292–324, 2019.

Kevin M Murphy, Andrei Shleifer, and Robert W Vishny. The allocation of talent: Implications for growth. *The Quarterly Journal of Economics*, 106(2):503–530, 1991.

Derek A Neal and William R Johnson. The role of premarket factors in black-white wage differences. *Journal of political Economy*, 104(5):869–895, 1996.

Steven G Rivkin, Eric A Hanushek, and John F Kain. Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458, 2005.

Robert Tamura. Teachers, growth, and convergence. *Journal of Political Economy*, 109(5): 1021–1059, 2001.

Michela M Tincani. Teacher labor markets, school vouchers, and student cognitive achievement: Evidence from chile. *Quantitative Economics*, 12(1):173–216, 2021.

Minchul Yum. Parental time investment and intergenerational mobility. *International Economic Review*, 64(1):187–223, 2023.

Appendix

A Derivation of Model Solution

A.1 Labor Market

As the taste shock follows a Gumbel distribution, the share of workers choosing to become teachers conditional on their human capital h is given by:

$$p(h) = \frac{\left(\exp(\alpha_1 - \alpha_2 + \kappa) \cdot h^{\psi_1 - \psi_2}\right)^{\theta}}{1 + \left(\exp(\alpha_1 - \alpha_2 + \kappa) \cdot h^{\psi_1 - \psi_2}\right)^{\theta}} \tag{22}$$

Define $\alpha = \alpha_1 - \alpha_2$ and $\psi = \psi_1 - \psi_2$, I adopt a simplifying approximation given that the share of teachers is small in the labor force:

$$p(h) \approx \exp(\theta(\alpha + \kappa)) \cdot h^{\theta\psi} \tag{23}$$

I make an assumption that the human capital distribution in the stationary equilibrium is lognormal:

$$\log(h) \sim \mathcal{N}(\mu, \sigma^2) \tag{24}$$

As will be shown below, this assumption will be self-fulfilling in the model.

Log-normal distribution has a convenient property that will be used repeatedly in the model solution: if $\log(X) \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\mathbb{E}(X^k) = \exp(k\mu + (k\sigma)^2/2)$$

With the assumption on the human capital distribution, the aggregate share of teachers in the labor force is therefore:

$$\begin{aligned} \overline{\pi} &= \int p(h)\, dF(h) \\ &= \exp(\theta(\alpha + \kappa)) \cdot \mathbb{E}(h^{\theta\psi}) \\ &= \exp(\theta(\alpha + \kappa)) \cdot \exp(\theta\psi\mu + (\theta\psi\sigma)^2/2) \end{aligned} \tag{25}$$

The average wage among teachers is

$$\begin{aligned}\mathbb{E}(w|j=1) &= \frac{1}{\pi}\cdot\int p(h)w_1(h)\,dF(h)\\ &= \frac{1}{\pi}\cdot\int \exp(\theta(\alpha+\kappa)+\alpha_1)\cdot h^{\theta\psi+\psi_1}\,dF(h)\\ &= \exp(\alpha_1)\cdot\exp(\psi_1\mu+\sigma^2\cdot(2\theta\psi\psi_1+\psi_1^2)/2)\end{aligned}\tag{26}$$

In addition,

$$\begin{aligned}\mathbb{E}(w^2|j=1) &= \frac{1}{\pi}\cdot\int p(h)(w_1(h))^2\,dF(h)\\ &= \frac{1}{\pi}\cdot\int \exp(\theta(\alpha+\kappa)+2\alpha_1)\cdot h^{\theta\psi+2\psi_1}\,dF(h)\\ &= \exp(2\alpha_1)\cdot\exp(2\psi_1\mu+\sigma^2\cdot(4\theta\psi\psi_1+4\psi_1^2)/2)\end{aligned}\tag{27}$$

Therefore,

$$\frac{\mathbb{E}(w^2|j=1)}{(\mathbb{E}(w|j=1)^2)} = \exp(\sigma^2\psi_1^2) = 1+(\mathbb{CV}(w|j=1))^2\tag{28}$$

where $\mathbb{CV}(w|j=1)=\text{std}(w|j=1)/\mathbb{E}(w|j=1)$ represents the coefficient of variation of w within teachers.

Using Taylor expansion around zero,

$$\mathbb{CV}(w|j=1)=\sigma\psi_1\tag{29}$$

On the other hand, the average wage among workers is

$$\begin{aligned}\mathbb{E}(w|j=2) &= \frac{1}{1-\pi}\cdot\int(1-p(h))w_2(h)\,dF(h)\\ &\approx \int \exp(\alpha_2)\cdot h^{\psi_2}\,dF(h)\\ &= \exp(\alpha_2)\cdot\exp(\psi_2\mu+\sigma^2\psi_2^2/2)\end{aligned}\tag{30}$$

In addition,

$$\begin{aligned}\mathbb{E}(w^2|j=2) &= \frac{1}{1-\overline{\pi}} \cdot \int (1-p(h))(w_2(h))^2\, dF(h) \\ &\approx \int \exp(2\alpha_2) \cdot h^{2\psi_2}\, dF(h) \\ &= \exp(2\alpha_2) \cdot \exp(2\psi_2\mu + \sigma^2 \cdot 2\psi_2^2)\end{aligned} \tag{31}$$

Therefore,

$$\frac{\mathbb{E}(w^2|j=2)}{\mathbb{E}(w|j=2)^2} = \exp(\sigma^2\psi_2^2) = 1 + (\mathbb{CV}(w|j=2))^2$$

Again, using Taylor expansion around zero,

$$\mathbb{CV}(w|j=2) = \sigma\psi_2 \tag{32}$$

To compute the wage ratio across occupations:

$$\begin{aligned}\frac{\mathbb{E}(w|j=1)}{\mathbb{E}(w|j=2)} &\approx \mathbb{E}\left(\frac{w_1}{w_2}\right) \\ &= \frac{1}{\overline{\pi}} \int p(h) \cdot \exp(\alpha) \cdot h^{\psi}\, dF(h) \\ &= \exp(\alpha) \cdot \exp(\psi\mu + (\sigma\psi)^2(1+2\theta)/2)\end{aligned} \tag{33}$$

A.2 Teaching Resources

By definition, teaching resource is given by

$$\begin{aligned}\mathcal{Q} &= \int p(h) \cdot \frac{h}{\overline{h}}\, dF(h) \\ &= \frac{1}{\overline{h}} \int \exp(\theta(\alpha+\kappa)) \cdot h^{\theta\psi+1}\, dF(h) \\ &= \frac{1}{\overline{h}} \cdot \exp(\theta(\alpha+\kappa)) \cdot \exp((\theta\psi+1)\mu + \sigma^2(\theta\psi+1)^2/2) \\ &= \exp(\theta(\alpha+\kappa)) \cdot \exp(\theta\psi\mu + \sigma^2 \cdot (\theta^2\psi^2 + 2\theta\psi)/2)\end{aligned} \tag{34}$$

Then, the amount of teaching resources each household receives is

$$q = \frac{1}{\overline{\pi}} \cdot \mathcal{Q} = \exp(\theta\psi\sigma^2) \tag{35}$$

The average amount of teaching resources is

$$\mathbb{E}(\tilde{h}|j=1) = \frac{1}{\overline{\pi}} \cdot \int p(h) \cdot \tilde{h} \, dF(h) = q \tag{36}$$

$$\begin{aligned} \mathbb{E}(\tilde{h}^2|j=1) &= \frac{1}{\overline{\pi}} \cdot \int p(h) \cdot \tilde{h}^2 \, dF(h) \\ &= \frac{1}{\overline{\pi}\overline{h}^2} \cdot \int \exp(\theta(\alpha+\kappa)) \cdot h^{\theta\psi+2} \, dF(h) \\ &= \exp(\sigma^2(1+2\theta\psi)) \end{aligned} \tag{37}$$

Therefore,

$$\text{std}(\tilde{h}|j=1) = \sqrt{\mathbb{E}(\tilde{h}^2|j=1) - \mathbb{E}(\tilde{h}|j=1)^2} = \sqrt{\exp(\sigma^2)-1} \cdot q \approx \sigma q \tag{38}$$

A.3 Child Investments

The first-order condition of e is given by

$$\frac{1}{1-e} = \beta\left(\frac{\lambda_1}{e} + \frac{\lambda_3 \log(q)}{e}\right) \Longrightarrow e = \frac{\beta(\lambda_1 + \lambda_3 \log(q))}{1+\beta(\lambda_1+\lambda_3\log(q))} \approx \beta(\lambda_3 + \lambda_3 \log(q)) \tag{39}$$

B Moments

To generate the moments describing the labor market of teaching and non-teaching professions, I collect micro-level data from the Current Population Survey Annual Social and Economic Supplement (CPS-ASEC) from 1980. I keep the observations that satisfy the following criteria:

- weeks worked last year was greater than or equal to 30,
- usual weekly hours worked last year was greater than or equal to 20, and
- had income greater than or equal to 100 dollars.

Then, I label the worker as a teacher if her occupation code last year (1990 basis) falls between 113 and 163. Based on this sample, I compute the share of teachers in the labor force, wage ratio across occupations, and wage dispersion within occupations.

To compute the ratio of parental income across races, I further restrict the sample to those with at least one child below the age of 18 in the household. To be more consistent with the time frame in Lovenheim and Willén (2019), I experimented with dropping the sample after the 2000s and the results are largely unchanged.

For all calculations, I use the Annual Social and Economic Supplement weight.