

Early Childhood Development, Earnings Inequality, and Social Mobility in an Education Signaling Model*

Lakshmi K. Raut¹

¹Visiting Fellow, University of Chicago

¹lakshmiraut@gmail.com

2026-10-11

Abstract

Much of earnings inequality results from the inequality of skill formation: children of low socioeconomic status acquire less schooling than equally talented children of better-off parents. In modern technology-intensive economies, channeling high-quality education to talented individuals and matching them to highly productive jobs is crucial for economic growth, social mobility, and earnings inequality. This paper studies a signaling equilibrium in which education is a costly signal of unobserved cognitive ability and the cost of schooling falls with both ability and parental schooling. The equilibrium wage schedule solves a nonlinear first-order ODE written in terms of a general cost index. The ODE characterizes equilibrium exactly, including the concavity of the wage schedule, and its local well-posedness holds under a monotone-posterior condition (implied by log-concavity when the cost index is separable). For log-normal economies, equilibrium distributions, Gini coefficients, intergenerational dynamics, and the stationary distribution are derived in closed form by Gaussian conditioning. Relative to full information, signaling compresses the dispersion of log schooling and log wages by exactly the variance of ability that schooling fails to reveal, but raises the intergenerational elasticity of schooling. For economies with a finite number of ability and schooling types, schooling depends on agents only through a cost index; the pure-strategy equilibria of a two-type economy are fully characterized, mobility and talent utilization differ across Pareto-ranked equilibria, and Perron–Frobenius theory characterizes the long-run distribution.

Short summary: The paper studies earnings inequality, social mobility, and economic growth in a signaling equilibrium framework, with policy implications for the labor market and education system.

JEL Classifications: J62, O15, D82, D86.

Keywords: Early Childhood Development, Social Mobility, Education Signaling, Earnings Inequality, Gaussian Conditioning.

*An earlier draft was presented at the 2017 ASSA Meetings, Chicago, January 6–8. I am grateful to Han Altae-Tran, Rinki Roy and Basit Zafar for many discussions.

1 Introduction

The growing income inequality has been a major concern for economists and policy makers around the world. Many factors are responsible for the observed burgeoning income inequality, such as capital outflow, relocation of jobs, declining labor union power, poor regulation of financial institutions, corruption, and all-encompassing globalization (Piketty 2014; Stiglitz 2015; Bourguignon 2015). In most economies, incomes of the bottom 99 percent come mainly from earnings, and much of the inequality in earnings results from the inequality of skill formation (Acemoglu and Autor 2011; Autor 2014).

A growing consensus has emerged among educators, researchers in sociology, psychology, and education (see, e.g., Schweinhart et al. 1993; Barnett 1995; Entwisle 1995; Reynolds et al. 2001, 2018) and economists (see, e.g., Keane and Wolpin 1997; Heckman 2000; Currie 2001, 2011; Raut 2003, 2018; Cunha and Heckman 2007; Maluccio et al. 2009; Duncan et al. 2010; Heckman et al. 2010; Currie and Almond 2011; Heckman and Raut 2013, 2016; García et al. 2016) that children of poor socioeconomic status (SES) are not prepared for college because they were not prepared for school to begin with. The most effective intervention for children of poor SES is therefore introduced at the preschool stage.

Much of the research in the last century focused on cognitive skills as the main determinant of socioeconomic behaviors, school performance, and labor market outcomes. An influential but controversial line of research argues that poor parents have poor cognitive abilities and that is why they are poor; children of poor SES inherit poor cognitive abilities from their parents; thus very little can be done to improve the cognitive skills of disadvantaged children, and hence their school performance and labor market outcomes—see Herrnstein and Murray (1994) and other references in Plomin and Deary (2015). This view has been challenged on statistical and biological grounds (Heckman 2006; Cunha et al. 2010; Noble et al. 2015), and the evidence shows that non-cognitive skills matter as much as cognitive skills for labor market outcomes (Heckman et al. 2006; Heckman and Raut 2016; Raut 2018).

It is the interplay of personality, emotion, and cognition that determines most socioeconomic behaviors. A branch of the psychology literature argues and empirically validates that emotional intelligence is an important factor in socioeconomic decisions and behaviors—not cognitive intelligence alone. The concept most relevant to our context is quoted from Mayer et al. (2004): “[Emotional Intelligence is the] capacity to reason about emotions, and of emotions to enhance thinking.” Measures based on various definitions of emotional intelligence are highly correlated with each other and each explains many socioeconomic behaviors independent of cognitive skills (see Heckman and Raut 2016; Chakrabarti and Chatterjea 2017; Raut 2018).

Where are cognitive and non-cognitive skills produced? The work of Bowlby (1982) argues that affect dysregulation beginning immediately after birth can have long-lasting effects on emotional development, and NICHD Early Child Care Research Network (2004) found longitudinal evidence for such mechanisms. A rapidly growing biological literature focuses on genetic and epigenetic mechanisms of personality, emotion, and cognitive development. For

instance, stress of various kinds can affect epigenetic reprogramming of the plasticity of various parts of the brain (see [Champagne et al. 2008](#); [Gluckman et al. 2008](#); [McEwen and Gianaros 2011](#)). A large-scale neurological study by Noble et al. (2015) found that family income significantly affects children’s brain size, particularly the surface area of the cerebral cortex that does most cognitive processing.

In modern technology-intensive economies, channeling high-quality education to talented individuals and matching them to highly productive technical jobs is crucial for economic growth, social mobility, and earnings inequality. Individuals know their own abilities but employers do not observe them, so education serves as a costly signal of ability ([Spence 1973](#); [Arrow 1973](#); [Stiglitz 1975](#); [Raut 2005](#)). Because children of poor SES face higher costs of acquiring education, the signal is contaminated by family background: schooling reflects not only talent but also the cost disadvantage associated with parental schooling. This paper studies the consequences of this mechanism for earnings inequality, social mobility, and growth in a signaling-equilibrium model that adapts the asymmetric information frameworks of Spence (1974) and Stiglitz (1975).

Relation to the literature. The model belongs to the class of Spence-type signaling models with a continuum of types, whose separating equilibria are characterized by a differential equation for the wage schedule ([Riley 1979, 2001](#); [Mailath 1987](#); [Mailath and Thadden 2013](#)). It differs in a respect that matters for inequality: agents differ in two dimensions, ability and parental schooling (which shifts the cost of schooling), so that schooling is only a noisy signal of ability (compare [Quinzii and Rochet 1985](#)). The intergenerational structure links the model to the literature on family background and mobility ([Becker and Tomes 1979](#); [Raut 1990, 2005](#); [Solon 1992](#); [Corak 2013](#); [Chetty et al. 2014](#)), and the self-fulfilling role of employers’ beliefs is related to models of statistical discrimination ([Coate and Loury 1993](#); [Lang and Manove 2011](#)). I do not take a stand in the debate between human-capital and signaling interpretations of schooling ([Weiss 1995](#)); the point is that, whatever its share, the signaling component transmits inequality in parental schooling (family background) into the wage structure.

Contributions. (i) I show that the equilibrium wage schedule solves an explicit first-order ODE written in terms of a general cost index; the ODE characterizes equilibrium exactly, including the concavity of the wage schedule, and is locally well posed under a monotone-posterior condition. When the cost index is separable in ability and parental schooling, the condition follows from log-concavity of the background component, via Efron’s monotonicity theorem ([Efron 1965](#)) (Section 3). (ii) For log-normal economies, I obtain the equilibrium in closed form, together with equilibrium schooling and wage distributions, Gini coefficients, intergenerational dynamics, and the stationary distribution, using Gaussian conditioning (Section 4). Relative to full information, signaling compresses the dispersion of log schooling and log wages by exactly $\sigma_*^2/(1-\rho)^2$ but raises the intergenerational elasticity of schooling by the factor $1+\beta$. (iii) For economies with finitely many types, I show that schooling depends on agents only through the cost index, characterize background-pooling and separating equilibria, fully characterize the pure-strategy equilibria of a two-type economy (which has Pareto-ranked equilibria with

different mobility and growth), and use Markov-chain theory for long-run mobility and talent utilization (Section 5). (iv) I show that a modest reduction in the cost disadvantage faced by children of low-schooling parents—plausibly achieved through early childhood development programs—converts the temporary mobility of the mobile equilibrium into a permanent one, raising both social mobility and economic growth (Proposition 5.4).

Roadmap. Section 2 specifies the economy and defines equilibrium, so that every later result is a statement about a precisely described model. Section 3 reduces the equilibrium to an ODE, to learn what an equilibrium looks like in general and when it exists. Section 4 specializes to log-normal distributions to obtain explicit answers about inequality and mobility. Section 5 considers finitely many types, to study discrete phenomena (pooling, multiple equilibria, temporary versus permanent mobility) that the smooth model cannot display. Section 6 concludes.

2 The Basic Model

Time is discrete. The economy consists of successive generations of agents and a continuum of risk-neutral, competitive producers. Each agent lives for one adult period, and a child is born to each adult; the child becomes an adult in the next period. Index by t the period in which an agent is an adult. Denote by $\tau_t \in \mathcal{T} = (0, \infty)$ the agent's cognitive ability, which affects both workplace productivity and learning in school (a higher number denotes greater ability), and by $s_t \in \mathcal{S} = (0, \infty)$ the agent's schooling level, which is observed by employers. The schooling level of the agent's parent is s_{t-1} . Throughout, *family background* or simply *background* means the schooling s_{t-1} of the agent's parent. The notation is summarized in Table 1.

Table 1: Notation.

Symbol	Meaning
τ_t, s_t, s_{t-1}	ability, schooling, and parental schooling of a period- t adult
$e(s, \tau)$	productivity of an agent with schooling s and ability τ
$\theta(s, \tau, s')$	cost of schooling s for ability τ and parental schooling s'
$\theta_1(s), K(\tau, s')$	schooling component and cost index of the cost function (Eq. (4))
$\kappa = K(\tau, s')$	value of the cost index; $u = \ln \kappa$
$m(s, u), H(s, w)$	expected productivity given the log cost index; right-hand side of the ODE (Eq. (5), Eq. (6))
b	wage elasticity of schooling, $(\rho + \beta)/(1 + \beta)$ (Section 3.3)
$w_t(s), q_t(e s)$	wage schedule and employers' belief in period t
$\varsigma_t(\tau, s')$	optimal schooling rule in period t
$g(\tau), \pi_t$	density of ability; distribution of schooling in period t
P_t	transition kernel (matrix in the finite case)
μ_τ, σ_τ^2	log-mean and log-variance of ability
$\mu_{s,t}, \sigma_{s,t}^2; \mu_{w,t}, \sigma_{w,t}^2$	log-mean and log-variance of schooling; of wages
α, γ, ρ	cost elasticities w.r.t. ability and parental schooling; productivity elasticity of schooling

Symbol	Meaning
β, σ_*^2, D	Gaussian-conditioning coefficient, residual variance, and scale; functions of the parental distribution (Eq. (12))
G	Gini coefficient
$g_i, \pi_{k,t}, c_j, \kappa_{ik}$	finite-type analogues (Section 5)

Given schooling s and ability τ , the agent's productivity is $e(s, \tau)$, and the cost of obtaining schooling s is $\theta(s, \tau, s')$, where s' is parental schooling.¹ The ability of a child born to a parent of ability τ_{t-1} and schooling s_{t-1} is drawn from a distribution with density $g(\tau)$ that does not depend on τ_{t-1} or s_{t-1} .²

2.1 Agent Optimization

All individuals have the linear utility function $u(c_t) = c_t$, where c_t is the consumption of an adult in period t .³ An adult of period t with ability τ and parental schooling s' takes the wage schedule $w_t(\cdot)$ as given and chooses schooling by solving

$$\varsigma_t(\tau, s') \in \arg \max_{s \in \mathcal{S}} [w_t(s) - \theta(s, \tau, s')]. \quad (1)$$

2.2 Producers and Beliefs

Employers cannot observe ability, so they pay each schooling level the expected productivity of the workers found there. This is where information is asymmetric, and it is why beliefs matter for equilibrium.

The production sector is competitive. The producer observes the schooling level s of a worker but not his ability τ . The employer holds a belief about productivity $e(s, \tau)$ given s , represented by the conditional distribution $q_t(e | s)$. Perfect competition and expected-profit maximization imply

$$w_t(s) = \int e q_t(e | s) de. \quad (2)$$

There is no affirmative action in hiring: workers with the same schooling level are treated identically regardless of family background.

¹The assumption that θ varies with τ is necessary for education to act as a signal for talent; see Stiglitz (1975), Spence (1973), or Kreps (1990).

²The assumption of independence is a simplification. Estimates of the intergenerational correlation of cognitive ability are positive and substantial (Plomin and Deary 2015); a positive correlation would strengthen the persistence results below because it adds a second channel of transmission.

³This abstracts away from risk-sharing between employers and workers.

2.3 Transition Dynamics

The economy begins in period $t = 1$ with parental schooling distributed as π_0 . The optimal choices $s_t = \varsigma_t(\tau_t, s_{t-1})$ determine the transition probabilities

$$P_t(A | s') = \int 1_A(\varsigma_t(\tau, s')) g(\tau) d\tau,$$

and the distribution $\pi_t(A) = \Pr(s_t \in A)$ of schooling satisfies

$$\pi_t(A) = \int P_t(A | s') \pi_{t-1}(ds'). \quad (3)$$

2.4 Signaling Equilibrium

Definition 2.1 (Signaling Equilibrium). Given an initial distribution π_0 , a **signaling equilibrium** is a sequence of beliefs $\{q_t(\cdot | s)\}_{t \geq 1}$ and schooling rules $\{\varsigma_t\}_{t \geq 1}$ such that, in every period $t \geq 1$:

1. (*Wages*) the wage schedule w_t is given by Eq. (2);
2. (*Optimality*) $\varsigma_t(\tau, s')$ solves Eq. (1) for every (τ, s') ; and
3. (*Bayes consistency*) for every schooling level s in the support of π_t , $q_t(\cdot | s)$ is the conditional distribution of $e(s, \tau_t)$ given $s_t = \varsigma_t(\tau_t, s_{t-1}) = s$, where $(\tau_t, s_{t-1}) \sim g \otimes \pi_{t-1}$ and π_t is given by Eq. (3). Off the support, beliefs are unrestricted.

Definition 2.1 describes a signalling equilibrium as a fixed point in each time period t : A competitive producer holds subjective belief $q(e_t | s_t)$ and announces the wage schedule $w_t(s_t)$ as in Eq. (2) (cf. *wages* condition); Given $w_t(s_t)$, each worker (τ_t, s_{t-1}) decides his optimal education level $\varsigma_t(\tau_t, s_{t-1})$ solving Eq. (1) (cf. *Optimality* condition). Given the distributions $g(\tau_t)$ and $\pi_{t-1}(s_{t-1})$, the optimal schooling decision $s_t \equiv \varsigma_t(\tau_t, s_{t-1})$ induces a joint distribution of (s_t, τ_t) ; which induces a conditional distribution of τ_t given s_t . This conditional distribution together with the productivity function $e_t = e(s_t, \tau_t)$ produces the observed distribution of productivity levels $\hat{q}_t(e | s_t)$ for each level of s_t chosen by some agent. Denoting it as $\hat{q}_t(e | s_t) = \Phi(q_t(e | s_t))$. A signaling equilibrium is the fixed point of the map Φ , i.e., the anticipated distribution coincides with the observed distribution, $q_t(e_t | s_t) = \hat{q}_t(e_t | s_t)$ (cf. *Bayes consistency* condition).

Note that the equilibrium of period t depends on the past only through π_{t-1} .

2.5 The Cost Index and Sorting

I impose a structural restriction to keep the model tractable: the cost of schooling is a product of a schooling component and a *cost index* summarizing all agent-specific factors influencing cost. The restriction implies that agents sort by the cost index, so that employers can infer only the cost index from schooling, and ability and family background are confounded. It also defines the class of *regular* equilibria studied in the rest of the paper.

Assumption A1 (Technology and costs). (i) The cost of schooling factors into a schooling component and a **cost index**,

$$\theta(s, \tau, s') = \theta_1(s) K(\tau, s'), \quad (4)$$

where $\theta_1 \in C^2(\mathcal{S})$ with $\theta_1' > 0$ and $\theta_1(0^+) = 0$, and $K : \mathcal{T} \times \mathcal{S} \rightarrow (0, \infty)$ is C^1 , strictly decreasing in τ , and nonincreasing in s' . (ii) The productivity function $e \in C^1(\mathcal{S} \times \mathcal{T})$ is positive, strictly increasing in τ , and nondecreasing in s .

Write $\kappa \equiv K(\tau, s')$ for the agent's cost index and $u \equiv \ln \kappa$. A higher ability or higher parental schooling lowers κ ; the background disadvantage of a child is the amount by which her index exceeds that of an equally able child of high-schooling parents. Under A1 the marginal cost of schooling is $\theta_1'(s)\kappa$, so the single-crossing (Spence–Mirrlees) condition holds automatically, $\partial[\theta_1'(s)\kappa]/\partial\tau < 0$ (see [Milgrom and Shannon 1994](#) for the role of this condition in monotone comparative statics). The cost index is a general function of ability and parental schooling; I impose the product form $K = \theta_2(\tau)\theta_3(s')$ only where it is needed (Lemma 3.1 and the examples of [Section 4](#) and [Section 5](#)). I restrict attention to equilibria in which the wage schedule gives every agent a unique interior optimum.

Definition 2.2 (Regular Equilibrium). A signaling equilibrium is **regular** in period t if $w_t \in C^1(\mathcal{S})$, $w_t' > 0$, and the function

$$u_t(s) \equiv \ln \frac{w_t'(s)}{\theta_1'(s)}$$

is strictly decreasing with range $\mathbb{R}$.

From here on I fix a period and suppress the index t on w_t , u_t , and ς_t . When θ_1 is convex, strict concavity of w implies that u is strictly decreasing, because $(w'/\theta_1')' = (w''\theta_1' - w'\theta_1'')/\theta_1'^2 < 0$; when $\theta_1(s) = s$, regularity means that w is strictly concave with $w'(0^+) = \infty$ and $w'(\infty) = 0$.

Lemma 2.1 (Optimal Schooling and Sorting by the Cost Index). *Under A1, let w be regular. Then every agent has a unique optimal schooling level,*

$$\varsigma(\tau, s') = u^{-1}(\ln K(\tau, s')),$$

the solution of the first-order condition $w'(s)/\theta_1'(s) = \kappa$. The rule ς depends on (τ, s') only through κ and is strictly decreasing in κ ; hence it is strictly increasing in τ (and nondecreasing in s'). Agents with the same cost index make the same choice.

Proof. The derivative of the objective $w(s) - \theta_1(s)\kappa$ is $\theta_1'(s)[e^{u(s)} - \kappa]$. Since u is strictly decreasing with range $\mathbb{R}$, this derivative is positive for $s < u^{-1}(\ln \kappa)$ and negative for $s > u^{-1}(\ln \kappa)$, so $u^{-1}(\ln \kappa)$ is the unique global maximizer. It is strictly decreasing in κ because u^{-1} is strictly decreasing. ■

Figure 1 illustrates the iso-schooling curves $\{(s', \tau) : K(\tau, s') = \bar{\kappa}\}$ in the (s_{t-1}, τ_t) plane. A higher schooling level corresponds to a lower cost index, hence to a curve farther from the origin. Along a curve, a child of low ability and rich parents makes the same choice as a child of high ability and poor parents; the market therefore cannot tell them apart.

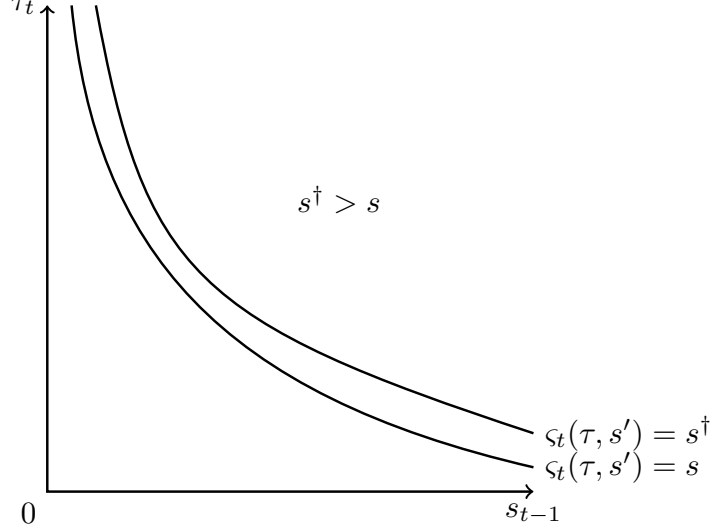


Figure 1: Iso-schooling curves $\{(s', \tau) : \varsigma_t(\tau, s') = s\}$ in the (s_{t-1}, τ_t) plane. For $s' > s$ the iso-schooling curve for s' lies northeast of the curve for s : higher ability or better family background raises equilibrium schooling. Agents on the same curve are pooled by the market.

Remark 2.1 (Endogenous Parental Investment). In the model, parental schooling s_{t-1} shifts costs exogenously. Parents who care about their children may instead invest in pre-school human capital and thereby choose the cost parameter; this makes the background effect (the parameter γ of Section 4) endogenous and would permit a welfare evaluation of public pre-school subsidies. A framework for a stationary Markov perfect signaling equilibrium with parental investment is given in Raut (2005, sec. 6). This could be further extended in the spirit of Heckman and Raut (2016), which is left for future work. A related extension would make the joint distribution of (τ_t, s_{t-1}) the outcome of the epigenetic and neurodevelopmental mechanisms surveyed in Section 1.

3 Equilibrium Characterization: An ODE for the Wage Schedule

A fixed point of the kind in Definition 2.1 and its elaboration in the text below the definition is hard to analyze directly. This section shows that, under A1, the fixed-point problem collapses to a single first-order ordinary differential equation (ODE) for the wage schedule, and it states exactly when a solution of that ODE is an equilibrium and a regularity condition (Definition 2.2) selects an equilibrium among the one-parameter family of solutions of the ODE. A period is fixed throughout the section and the index t is suppressed.

3.1 The Posterior of Ability Given the Cost Index

Employers do not observe ability; at schooling level s , they see a pool of workers. By Lemma 2.1, the schooling choice depends on a worker only through the cost index $\kappa = K(\tau, s')$, so all that the market can learn from s is the value of κ . What matters for wages is therefore the *posterior distribution of ability given the cost index*. This subsection defines the expected productivity implied by the posterior and states its core property—monotonicity—upon which the rest of the section rests.

Let $U \equiv \ln K(\tau_t, s_{t-1})$ be the log cost index of a randomly drawn adult. For $s \in \mathcal{S}$ and $u \in \mathbb{R}$, let

$$m(s, u) \equiv \mathbb{E}[e(s, \tau_t) \mid U = u] \quad (5)$$

be the expected productivity of a worker with schooling s whose log cost index is u : it is the wage that competitive employers would pay at schooling level s if they knew that the workers choosing s have log cost index u . For fixed s , write $m(s, \mathbb{R}) \equiv \{m(s, u) : u \in \mathbb{R}\}$ for the set of values taken by $m(s, \cdot)$; that is, the set of all wages at s that are consistent with *some* value of the cost index. If $m(s, \cdot)$ is continuous and strictly decreasing (as assumed below), $m(s, \mathbb{R})$ is the open interval whose endpoints are the limits of $m(s, u)$ as $u \rightarrow \pm\infty$.

Assumption A2 (Posterior mean). U has a strictly positive density, the conditional distribution of τ_t given $U = u$ has a density $h(\cdot \mid u)$, and $m(s, u)$ is finite and C^1 on $\mathcal{S} \times \mathbb{R}$ with $\partial_u m < 0$.

The sign condition is natural: a higher cost index signals lower ability, hence lower expected productivity. The next lemma shows that, when the cost index is separable in ability and parental schooling, the sign condition follows from log-concavity of the background component; this is the only place where separability of K is used in the general theory.

Lemma 3.1 (Separable Cost Index). *Let $K(\tau, s') = \theta_2(\tau)\theta_3(s')$ with θ_2 strictly decreasing, let τ_t (density g) and s_{t-1} be independent, and suppose $\eta \equiv \ln \theta_3(s_{t-1})$ has a C^1 density f_η that is strictly positive and strictly log-concave on $\mathbb{R}$. Then $U = \ln \theta_2(\tau_t) + \eta$ has the positive density $f_U(u) = \int g(\tau)f_\eta(u - \ln \theta_2(\tau)) d\tau$, the posterior density is*

$$h(\tau \mid u) = \frac{g(\tau)f_\eta(u - \ln \theta_2(\tau))}{f_U(u)},$$

the likelihood ratio $h(\tau \mid u')/h(\tau \mid u)$ is strictly decreasing in τ whenever $u < u'$, and $u \mapsto m(s, u)$ is strictly decreasing for every s .

Proof. Fix $u < u'$ and write $a(\tau) = \ln \theta_2(\tau)$, strictly decreasing in τ . The derivative of $\ln[f_\eta(u' - a)/f_\eta(u - a)]$ with respect to a is $(\ln f_\eta)'(u - a) - (\ln f_\eta)'(u' - a) > 0$ by strict log-concavity, so $h(\tau \mid u')/h(\tau \mid u)$ is strictly decreasing in τ . Hence $h(\cdot \mid u') - h(\cdot \mid u)$, which

integrates to zero, is positive to the left and negative to the right of some τ° , and

$$m(s, u') - m(s, u) = \int [e(s, \tau) - e(s, \tau^\circ)][h(\tau | u') - h(\tau | u)] d\tau < 0,$$

because the integrand is nonpositive and negative on a set of positive measure (e is strictly increasing in τ). ■

The heart of the proof is a *monotone likelihood ratio* (MLR) property of the posterior: when the cost index rises (u to u'), it doesn't just lower average ability—it shifts the entire probability distribution toward lower abilities. High ability levels become strictly less likely relative to lower ones. The property comes from the log-concavity of the density of the family background component η . This shape assumption prevents the distribution from having multiple “bumps,” ensuring a higher cost index cannot be explained by two completely different extremes (e.g., very high vs. very low ability). A classical result of Efron (1965) (the conditional expectation of an increasing function of one of two independent log-concave variables, given their sum, is increasing in the sum) is the best-known statement of this fact; here only log-concavity of f_η is needed because the kernel $f_\eta(u - a)$ is then totally positive of order two (see also Karlin and Rubin 1956).

MLR is what makes $m(s, \cdot)$ strictly decreasing, so that a higher cost index means lower expected productivity at every schooling level. Hence, for a given (s, w) there is at most one cost index u with $m(s, u) = w$; the function $\hat{u}$ defined below is well defined, and the equilibrium condition becomes an ODE with an explicit right-hand side (Theorem 3.1). Without monotonicity, the same wage at s could be rationalized by several different pools of workers, and the equilibrium would not be characterized by a single ODE.

3.2 The Main Characterization Theorem

This subsection states exactly when a candidate wage schedule is an equilibrium wage schedule. For (s, w) with $w \in m(s, \mathbb{R})$, let $\hat{u}(s, w)$ be the unique solution u of $m(s, u) = w$, and define

$$\begin{aligned} \mathcal{D} &\equiv \{(s, w) : s \in \mathcal{S}, w \in m(s, \mathbb{R})\}, \\ H(s, w) &\equiv \theta'_1(s) \exp\{\hat{u}(s, w)\}. \end{aligned} \tag{6}$$

The set $\mathcal{D}$ collects the pairs (schooling level, wage) that employers could *rationally* announce: $(s, w) \in \mathcal{D}$ if there is a cost index at which the expected productivity of the workers choosing s is exactly w . $\hat{u}(s, w)$ is that cost index (in logs): the characteristic of the workers who must be found at s if the wage paid at s is w . $H(s, w)$ is the slope that the wage schedule must have at (s, w) : by the first-order condition of the workers' problem, the slope of the wage schedule equals their marginal cost of schooling, $\theta'_1(s)\kappa = \theta'_1(s)e^{\hat{u}(s, w)}$. So $H(s, w)$ answers the question “if the wage at s is w , how steeply must wages rise at s so that the workers who are there find it optimal to be there?”

Example 3.1 (Reading the Admissible Set and the Required Slope: An Example). Let $\theta_1(s) = s$, $e(s, \tau) = s^\rho \tau$, and $K = \tau^{-\alpha}$, with $\rho = \frac{1}{2}$ and $\alpha = 1$. Then $\tau = e^{-u}$ and $m(s, u) = s^{1/2} e^{-u}$, which takes every positive value as u ranges over $\mathbb{R}$: $m(s, \mathbb{R}) = (0, \infty)$, so $\mathcal{D} = \mathcal{S} \times (0, \infty)$, $\hat{u}(s, w) = -\ln(w/s^{1/2})$, and $H(s, w) = s^{1/2}/w$.

Theorem 3.1 (Equilibrium Wage Schedule as the Solution of an ODE). *Assume A1 and A2. Let $w \in C^1(\mathcal{S})$ with $w' > 0$ be such that $u(s) = \ln[w'(s)/\theta_1'(s)]$ is strictly decreasing with range $\mathbb{R}$. Then:*

- (i) *Characterization. w is the wage schedule of a regular signaling equilibrium if and only if it solves*

$$w'(s) = H(s, w(s)), \quad (s, w(s)) \in \mathcal{D}, \quad s \in \mathcal{S}, \quad (7)$$

that is, if and only if $w(s) = m(s, u(s))$: the wage at s equals the expected productivity of the workers who choose s .

- (ii) *Second-order condition and concavity. Every solution is C^2 , and $w''(s) \leq \theta_1''(s) w'(s) / \theta_1'(s)$ for all s , with strict inequality except on a set with empty interior; this is the second-order condition of every agent's schooling problem. In particular, w is strictly concave whenever $\theta_1'' \leq 0$, for example when $\theta_1(s) = s$.*
- (iii) *Local uniqueness. The domain $\mathcal{D}$ is open and $H \in C^1(\mathcal{D})$. Through every point of $\mathcal{D}$ passes exactly one maximal solution of Eq. (7), and distinct maximal solutions never intersect. Hence a regular equilibrium wage schedule is determined by any single one of its points.*

Proof. Step 1: conditioning on schooling is conditioning on the cost index. Since u is a strictly decreasing homeomorphism of $\mathcal{S}$ onto $\mathbb{R}$, Lemma 2.1 gives the unique optimal choice $\varsigma = u^{-1}(\ln K(\tau, s'))$. Hence, for every s , the event $\{s_t = s\}$ coincides with $\{U = u(s)\}$. Because U has full support, every s is chosen by a positive mass of agents, so Bayes consistency pins down the belief at every s : $q(\cdot | s)$ is the law of $e(s, \tau_t)$ given $U = u(s)$.

Step 2: (i). By Eq. (2), the wage consistent with these beliefs is $m(s, u(s))$. Hence w is the wage schedule of a regular equilibrium if and only if $w(s) = m(s, u(s))$ for all s (for the “if” direction, define the beliefs of Step 1; they are Bayes-consistent and reproduce w). By A2, $m(s, \cdot)$ is a continuous strictly decreasing bijection of $\mathbb{R}$ onto $m(s, \mathbb{R})$, so $w(s) = m(s, u(s))$ holds if and only if $(s, w(s)) \in \mathcal{D}$ and $u(s) = \hat{u}(s, w(s))$, that is, $w'(s) = \theta_1'(s) e^{\hat{u}(s, w(s))} = H(s, w(s))$.

Step 3: (ii). Once the ODE holds, $w'(s) = H(s, w(s))$ with $H \in C^1$ (Step 4) shows $w \in C^2$. Differentiating $w' = \theta_1' e^u$ gives $w'' = e^u(\theta_1'' + \theta_1' u')$, so $w'' - \theta_1'' w' / \theta_1' = \theta_1' e^u u'$. Since u is strictly decreasing, $u' \leq 0$ and u' vanishes on no interval, which gives the inequality. The left side equals $w'' - \theta_1'' \kappa$ at $\kappa = e^u$, the second derivative of the agent's objective $w(s) - \theta_1(s) \kappa$ at the optimum. If $\theta_1'' \leq 0$, then $w'' = e^u(\theta_1'' + \theta_1' u') \leq 0$, with strict inequality except on a set with empty interior, so w is strictly concave.

Step 4: (iii). By A2, $\Gamma(s, u) = (s, m(s, u))$ is an injective C^1 map with Jacobian determinant

$\partial_u m \neq 0$. By the inverse function theorem, Γ is a C^1 diffeomorphism onto the open set $\mathcal{D}$, with inverse $(s, w) \mapsto (s, \hat{u}(s, w))$. Hence $\hat{u} \in C^1(\mathcal{D})$ and $H \in C^1(\mathcal{D})$ because $\theta'_1 \in C^1$. A C^1 right-hand side is locally Lipschitz, so the Picard–Lindelöf theorem (Coddington and Levinson 1955, chap. 1; Hartman 2002, II) gives a unique maximal solution through each point of $\mathcal{D}$, and two maximal solutions with a common point coincide. A regular wage schedule is defined on all of $\mathcal{S}$, hence is the maximal solution through any of its points. ■

An equilibrium wage schedule is exactly a solution of one first-order ODE that is defined for all schooling levels and satisfies the regularity requirement that marginal cost fall with schooling. Two further facts come with it: concavity of wages is *implied*, not assumed (when θ_1 is linear or concave), and the equilibrium is pinned down by a single point, so the family of candidate equilibria is one-dimensional. What the theorem does not say is that a regular solution always exists: it reduces that question to whether some member of the one-parameter family of solutions is defined on all of $\mathcal{S}$ and has the regularity property. Subsection 3.3 answers this in closed form for the economies used in Section 4. In implicit form, the equilibrium condition reads

$$w(s) = \Psi(s, w'(s)), \quad \Psi(s, p) \equiv m(s, \ln[p/\theta'_1(s)]); \quad (8)$$

the theorem says that this implicit equation can be solved for w' , which is what yields the explicit ODE Eq. (7).

Remark 3.1 (Boundary Conditions and Multiplicity). An ODE has a one-parameter family of solutions, so one asks what plays the role of the initial condition. The answer differs across signaling models.

In the classical Spence model with a bounded type space, the lowest type’s schooling choice is a free boundary condition, which generates a continuum of separating equilibria; the least-cost separating equilibrium is the one selected by standard refinements (Riley 1979; Mailath 1987; Cho and Kreps 1987; Cho and Sobel 1990). Here the type space $(0, \infty)$ has no lowest type to supply an initial condition, and every schooling level is chosen by a positive mass of agents. The requirement in Theorem 3.1 that u be strictly decreasing with range $\mathbb{R}$ takes its place: it rules out solutions that are convex near $s = 0$ (the first-order condition would then identify a minimum rather than a maximum) and solutions that vanish at a positive schooling level (no agent would choose lower levels). In the power class exactly one solution survives, the one with $w(0^+) = 0$, the continuum analogue of the lowest type doing no schooling and earning nothing (Proposition 3.1)

Remark 3.2 (Beyond the Gaussian Case). Theorem 3.1 does not use log-normality. For other families of distributions, $m(s, u)$ is available by one-dimensional numerical integration of Eq. (5), and Eq. (7) can be solved numerically. Closed-form posteriors exist for conjugate families, such as the Gaussian family of Section 4 and the natural exponential families with quadratic variance functions.

3.3 Closed-Form Solution in the Power Class

[Theorem 3.1](#) leaves open whether a regular solution exists. This subsection settles the question, in closed form, for the class of economies on which the rest of the paper is built: constant-elasticity productivity, linear schooling cost, and a posterior mean that is log-linear in the cost index. The log-normal economy of [Section 4](#) belongs to this class. Because the ODE is then separable, the result requires only elementary integration.

Suppose

- **(PC)** $\theta_1(s) = s$, $e(s, \tau) = s^\rho \tau$ with $0 < \rho < 1$, and the posterior mean of ability is log-linear in the cost index,

$$T(u) \equiv \mathbb{E}[\tau \mid U = u] = \chi e^{-\beta u}, \quad \chi, \beta > 0. \quad (9)$$

Then $m(s, u) = s^\rho T(u) = \chi s^\rho e^{-\beta u}$, which is C^1 with $\partial_u m = -\beta m < 0$, so that A2 holds, $\mathcal{D} = \mathcal{S} \times (0, \infty)$, $\hat{u}(s, w) = -\beta^{-1} \ln(w/(\chi s^\rho))$, and $H(s, w) = (\chi s^\rho/w)^{1/\beta}$. Let

$$b \equiv \frac{\rho + \beta}{1 + \beta} \in (\rho, 1).$$

Proposition 3.1 (Closed-Form Equilibrium in the Power Class). *Under (PC):*

- (a) *Every positive solution of [Eq. \(7\)](#) satisfies*

$$w(s)^{(1+\beta)/\beta} = C + \frac{\chi^{1/\beta}}{b} s^{(\rho+\beta)/\beta}, \quad C \in \mathbb{R}. \quad (10)$$

- (b) *Exactly one solution is the wage schedule of a regular equilibrium, namely $C = 0$:*

$$w(s) = A s^b, \quad A = \chi^{1/(1+\beta)} b^{-\beta/(1+\beta)}.$$

It is strictly concave, and its elasticity with respect to schooling is the constant $b < 1$.

- (c) *Solutions with $C > 0$ have $w(0^+) > 0$ and $w'(0^+) = 0$, and solutions with $C < 0$ are defined only for s above a positive threshold, where $w = 0$; neither is regular. All solutions with $C \neq 0$ nevertheless approach the equilibrium for large s : $d \ln w / d \ln s \rightarrow b$ as $s \rightarrow \infty$.*

Proof.

- (a) The ODE $w' = (\chi s^\rho/w)^{1/\beta}$ is separable: $w^{1/\beta} w' = (\chi s^\rho)^{1/\beta}$. Integrating, $\frac{\beta}{1+\beta} w^{(1+\beta)/\beta} = \chi^{1/\beta} \frac{\beta}{\rho+\beta} s^{(\rho+\beta)/\beta} + \text{const}$, and $(1+\beta)/(\rho+\beta) = 1/b$ gives [Eq. \(10\)](#).
 (b)–(c) A regular schedule must be positive and defined on all of $\mathcal{S}$, which excludes $C < 0$ (the right side of [Eq. \(10\)](#) is negative for small s). If $C > 0$, $w(0^+) = C^{\beta/(1+\beta)} > 0$, so $w' = \chi^{1/\beta} s^{\rho/\beta} w^{-1/\beta} \rightarrow 0$ as $s \downarrow 0$, and then $u = \ln w'$ cannot have range $\mathbb{R}$. For $C = 0$, $w = A s^b$ with A as stated, and $u(s) = \ln(bA) + (b-1) \ln s$ is strictly decreasing with range $\mathbb{R}$ because $b < 1$; $w'' = bA(b-1)s^{b-2} < 0$, and by [Theorem 3.1\(i\)](#) w

is the wage schedule of a regular equilibrium. For any $C \neq 0$, (Eq. (10)) gives $w^{(1+\beta)/\beta} s^{-(\rho+\beta)/\beta} \rightarrow \chi^{1/\beta}/b$ as $s \rightarrow \infty$, so $d \ln w / d \ln s \rightarrow \frac{\beta}{1+\beta} \cdot \frac{\rho+\beta}{\beta} = b$. Why $\rho < 1$. Part (a) does not use $\rho < 1$. If $\rho \geq 1$, then $b \geq 1$ and, for every C , $w \sim As^b$ and $w' = (\chi s^\rho / w)^{1/\beta} \sim s^{b-1}$ as $s \rightarrow \infty$, which does not decrease to zero (if $b > 1$ it diverges; if $b = 1$ it tends to a positive constant). A regular schedule has w' strictly decreasing, with $u = \ln w'$ tending to $-\infty$, so no solution is regular: there is no equilibrium. ■

Proposition 3.1 shows that in the power class, the equilibrium exists, is unique, is concave, and has constant elasticity $b < 1$. The wage rises with schooling for two reasons: schooling is productive (elasticity ρ), and more schooling signals higher ability (an additional elasticity $b - \rho = \beta(1 - \rho)/(1 + \beta) > 0$); the total is still below one, so wages are less dispersed than schooling. The closed form is the building block of all results in [Section 4](#).

Remark 3.3 (Scale Invariance and the Role of Log-Linearity). The equilibrium has a scale-invariance interpretation. The ODE satisfies $H(\lambda s, \lambda^b w) = \lambda^{b-1} H(s, w)$, so it is unchanged by the scaling $(s, w) \mapsto (\lambda s, \lambda^b w)$, and the equilibrium $w = As^b$ is the unique solution that is invariant under it. This is the viewpoint of Lie's symmetry methods for differential equations ([Olver 1993, chap. 2 and 3](#)). The converse also holds: by differentiating $T(\hat{u}(s, w)) = ws^{-\rho}$ one finds that the ODE is invariant under a scaling $(s, w) \mapsto (\lambda s, \lambda^b w)$ if and only if T'/T is constant, that is, if and only if the posterior mean is log-linear as in [Eq. \(9\)](#). The closed forms of [Section 4](#) are therefore tied to Gaussian-like posteriors; for other posteriors one uses the numerical method of [3.2](#). Since elementary integration already solves the power class, I do not develop the symmetry machinery further.

4 The Log-Normal Economy

The characterization of [Section 3](#) is general but abstract. To obtain explicit answers about inequality and mobility, this section specializes to log-normal ability and parental schooling. In that case the employers' inference problem has a closed form (Gaussian conditioning) and the ODE of [Section 3](#) is solved in closed form ([Proposition 3.1](#)). The results answer three questions: how unequal are schooling and wages in equilibrium (Gini coefficients); how persistent is schooling across generations (intergenerational dynamics and mobility); and what does the signaling friction itself do to inequality and mobility, relative to full information and relative to a world in which cost does not depend on family background?

I now examine how inequality in earnings and schooling evolves when ability and parental schooling are log-normal. Write $X \sim \text{LN}(\mu, \sigma^2)$ if $\ln X \sim N(\mu, \sigma^2)$. Assume throughout this section, for every $t \geq 1$,

$$\begin{aligned} \tau_t &\sim \text{LN}(\mu_\tau, \sigma_\tau^2), \quad s_{t-1} \sim \text{LN}(\mu_{s,t-1}, \sigma_{s,t-1}^2) \text{ independent of } \tau_t, \\ e(s, \tau) &= s^\rho \tau, \quad 0 < \rho < 1, \quad \theta_1(s) = s, \quad K(\tau, s') = \tau^{-\alpha} s'^{-\gamma}, \end{aligned} \tag{11}$$

with $\alpha > 0$ and $\gamma \geq 0$. Here, and only here, the cost index takes the product form $K = \theta_2(\tau)\theta_3(s')$ with $\theta_2(\tau) = \tau^{-\alpha}$ and $\theta_3(s') = s'^{-\gamma}$. The restriction $\rho < 1$ (diminishing returns to schooling) is needed for the existence of an interior optimum. I consider two cases: (i) the cost depends only on ability ($\gamma = 0$), so that schooling is a perfect signal; and (ii) the cost also depends on parental schooling ($\gamma > 0$), so that the signal is imperfect. In both cases the posterior mean of ability given the cost index is log-linear, so that the power-class results of [Subsection 3.3](#) apply: for $\gamma > 0$, A1 and A2 hold (see the proof of [Proposition 4.3](#)), and for $\gamma = 0$ the posterior is degenerate, $\tau = e^{-u/\alpha}$, and [Theorem 3.1](#) holds verbatim with $m(s, u) = s^\rho e^{-u/\alpha}$.

The following lemma calculates the posterior ability based on the signal using the *orthogonal projection*. This is the only probabilistic calculation required to achieve the closed-form results. Regard random variables with finite variance as vectors, with the covariance as inner product. The orthogonal projection of $Y = \ln \tau$ on the observed signal $V = \alpha Y + \gamma Z$ is the part of Y that is linear in V and leaves a remainder uncorrelated with V : it equals $\mathbb{E}Y + \beta(V - \mathbb{E}V)$ with $\beta = \text{Cov}(Y, V) / \text{Var}(V)$. Geometrically it is the foot of the perpendicular dropped from Y onto the line spanned by V ; statistically it is the best linear predictor of Y given V , the one that minimizes the mean squared error. For jointly normal variables, “uncorrelated” means “independent”, so the remainder carries no information about V . The projection is then not just the best *linear* predictor but the full conditional expectation, and the conditional variance is the variance of the remainder, which does not depend on the realized value of the signal. Two consequences drive the results below: the posterior mean of $\ln \tau$ is linear in the cost index (so the posterior mean of τ is log-linear, as required by (PC) in [Subsection 3.3](#)), and the posterior variance is a constant. Standard references are Anderson (2003, sec. 2.5) for the conditional normal distribution and Brockwell and Davis (1991, chap. 2) for the Hilbert-space projection theorem. The lemma is a static computation and requires neither continuous-time filtering nor a Riccati equation.

Lemma 4.1 (Gaussian Conditioning). *Let $Y = \ln \tau \sim N(\mu_\tau, \sigma_\tau^2)$ and $Z = \ln s' \sim N(\mu_z, \sigma_z^2)$ be independent, $\alpha, \gamma > 0$, and $D^2 \equiv \alpha^2 \sigma_\tau^2 + \gamma^2 \sigma_z^2$. Then, conditional on $\alpha Y + \gamma Z = v$, Y is normal with*

$$\begin{aligned} \mathbb{E}[Y \mid \alpha Y + \gamma Z = v] &= \mu_\tau + \beta(v - \alpha \mu_\tau - \gamma \mu_z), \\ \text{Var}(Y \mid \alpha Y + \gamma Z) &= \sigma_*^2, \text{ where} \\ \beta &= \frac{\alpha \sigma_\tau^2}{D^2}, \quad \sigma_*^2 = \frac{\gamma^2 \sigma_\tau^2 \sigma_z^2}{D^2}. \end{aligned} \tag{12}$$

The signal quality $\mathcal{R}^2 \equiv 1 - \sigma_*^2 / \sigma_\tau^2 = \alpha \beta = \alpha^2 \sigma_\tau^2 / D^2 \in (0, 1)$ is the fraction of the variance of $\ln \tau$ explained by the signal; it is increasing in α and σ_τ and decreasing in γ and σ_z .

Proof. $(Y, \alpha Y + \gamma Z)$ is bivariate normal with $\text{Cov}(Y, \alpha Y + \gamma Z) = \alpha \sigma_\tau^2$ and $\text{Var}(\alpha Y + \gamma Z) = D^2$. The conditional mean and variance are the standard regression formulas for the multivariate normal distribution; the residual $Y - \mathbb{E}[Y \mid \alpha Y + \gamma Z]$ is independent of $\alpha Y + \gamma Z$. The monotonicity

statements follow from $\mathcal{R}^2 = 1/(1 + \gamma^2 \sigma_z^2/(\alpha^2 \sigma_\tau^2))$. ■

4.1 Perfect Signal: Cost Independent of Family Background

This subsection treats the benchmark in which schooling reveals ability perfectly. It shows the basic mechanism in its simplest form and supplies the formulas to which the imperfect-signal case reduces when the background effect on cost vanishes.

Let $\gamma = 0$, so that the log cost index $U = -\alpha \ln \tau$ reveals ability, $\tau = e^{-U/\alpha}$, and $T(u) = e^{-u/\alpha}$. This is Eq. (9) with $\chi = 1$ and $\beta = 1/\alpha$, so $b = (\alpha \rho + 1)/(\alpha + 1)$ and Eq. (7) becomes

$$w'(s) = \left[\frac{s^\rho}{w(s)} \right]^\alpha. \quad (13)$$

Proposition 4.1 (Equilibrium under a Perfect Signal). *Under Eq. (11) with $\gamma = 0$, the signaling equilibrium is unique among regular equilibria and is the same in every period:*

$$\begin{aligned} w(s) &= \left[\frac{\alpha + 1}{\alpha \rho + 1} \right]^{1/(1+\alpha)} s^{(\alpha \rho + 1)/(\alpha + 1)}, \\ \varsigma(\tau) &= \left[\frac{\alpha \rho + 1}{\alpha + 1} \right]^{1/(1-\rho)} \tau^{(\alpha + 1)/(1-\rho)}. \end{aligned} \quad (14)$$

Equilibrium schooling and wages are log-normal, $s_t \sim \text{LN}(\mu_s^{\text{pf}}, (\sigma_s^{\text{pf}})^2)$ and $w_t = w(s_t) \sim \text{LN}(\mu_w^{\text{pf}}, (\sigma_w^{\text{pf}})^2)$, with

$$\mu_s^{\text{pf}} = \frac{1}{1-\rho} \left[\ln \frac{\alpha \rho + 1}{\alpha + 1} + (\alpha + 1) \mu_\tau \right], \quad \sigma_s^{\text{pf}} = \frac{(\alpha + 1) \sigma_\tau}{1-\rho}, \quad (15)$$

$$\mu_w^{\text{pf}} = \frac{1}{1-\rho} \left[\rho \ln \frac{\alpha \rho + 1}{\alpha + 1} + (\alpha \rho + 1) \mu_\tau \right], \quad \sigma_w^{\text{pf}} = \frac{(\alpha \rho + 1) \sigma_\tau}{1-\rho}, \quad (16)$$

and the Gini coefficients are, with Φ the standard normal CDF,

$$G_s^{\text{pf}} = 2\Phi\left(\frac{(\alpha + 1) \sigma_\tau}{\sqrt{2} (1-\rho)}\right) - 1, \quad G_w^{\text{pf}} = 2\Phi\left(\frac{(\alpha \rho + 1) \sigma_\tau}{\sqrt{2} (1-\rho)}\right) - 1. \quad (17)$$

Proof. By Theorem 3.1, a regular equilibrium solves Eq. (13). By Proposition 3.1 with $\chi = 1$ and $\beta = 1/\alpha$, the unique regular solution is $w = As^b$ with $b = (\alpha \rho + 1)/(\alpha + 1)$ and $A = b^{-1/(1+\alpha)} = [(\alpha + 1)/(\alpha \rho + 1)]^{1/(1+\alpha)}$, which is Eq. (14) for w . The first-order condition $w'(s) = \tau^{-\alpha}$ gives $\ln s = [\ln(bA) + \alpha \ln \tau]/(1-b)$, which simplifies to $\ln s = \frac{1}{1-\rho} \ln \frac{\alpha \rho + 1}{\alpha + 1} + \frac{\alpha + 1}{1-\rho} \ln \tau$, that is, Eq. (14) for ς and Eq. (15). Then $\ln w_t = b \ln s_t + \ln A$ gives Eq. (16). The Gini coefficient of $\text{LN}(\mu, \sigma^2)$ is $2\Phi(\sigma/\sqrt{2}) - 1$ (Aitchison and Brown 1957). ■

With a perfect signal, schooling reveals ability one-for-one: each schooling level is chosen by a single ability level, wages equal productivity, and log-normality of ability carries over to

schooling and wages. Wages are less dispersed than schooling because the wage rises less than proportionally with schooling (elasticity $b = (\alpha\rho + 1)/(\alpha + 1) < 1$).

Proposition 4.2 (Wage Inequality vs. Schooling Inequality). *Under a perfect signal, $G_w^{\text{pf}} < G_s^{\text{pf}}$.*

Proof. Since Φ is strictly increasing, it suffices to compare the arguments in Eq. (17). The argument of G_w^{pf} equals $\frac{\alpha\rho+1}{\alpha+1}$ times that of G_s^{pf} , and $\frac{\alpha\rho+1}{\alpha+1} < 1$ if and only if $\rho < 1$. ■

4.2 Imperfect Signal: Cost Depends on Ability and Family Background

This subsection treats the case of interest: when parental schooling also affects the cost, schooling confounds ability and background, and employers can infer only the cost index. The aim is to find how the equilibrium wage schedule, and the distributions of schooling and wages, change when the signal becomes noisy.

Let $\gamma > 0$, so that $w'(s_t) = \tau_t^{-\alpha} s_{t-1}^{-\gamma}$ no longer reveals τ_t . Fix a period t . The following quantities depend on t only through the parental log-moments $\mu_{s,t-1}$ and $\sigma_{s,t-1}^2$, and I suppress this dependence:

$$D^2 = \alpha^2 \sigma_\tau^2 + \gamma^2 \sigma_{s,t-1}^2, \quad \beta = \frac{\alpha \sigma_\tau^2}{D^2}, \quad \sigma_*^2 = \frac{\gamma^2 \sigma_\tau^2 \sigma_{s,t-1}^2}{D^2},$$

(Eq. (12) with $\sigma_z = \sigma_{s,t-1}$), and

$$b = \frac{\rho + \beta}{1 + \beta} \in (\rho, 1), \quad \chi = \exp\left\{(1 - \alpha\beta)\mu_\tau - \gamma\beta\mu_{s,t-1} + \frac{1}{2}\sigma_*^2\right\}.$$

Proposition 4.3 (Equilibrium under an Imperfect Signal). *Under Eq. (11) with $\gamma > 0$ and $\sigma_{s,t-1} > 0$, the period- t signaling equilibrium is unique among regular equilibria and is given by*

$$w(s) = A s^b, \quad A = \chi^{1/(1+\beta)} b^{-\beta/(1+\beta)}, \quad (18)$$

$$\varsigma(\tau, s') = \left[b A \tau^\alpha s'^\gamma \right]^{(1+\beta)/(1-\rho)}.$$

Schooling and wages are log-normal, $s_t \sim \text{LN}(\mu_{s,t}, \sigma_{s,t}^2)$ and $w_t = w(s_t) \sim \text{LN}(\mu_{w,t}, \sigma_{w,t}^2)$, with

$$\begin{aligned} \mu_{s,t} &= \frac{1}{1-\rho} \left[\ln b + \frac{1}{2}\sigma_*^2 + (1+\alpha)\mu_\tau + \gamma\mu_{s,t-1} \right], \\ \sigma_{s,t} &= \frac{1+\beta}{1-\rho} D = \frac{D^2 + \alpha\sigma_\tau^2}{(1-\rho)D}, \end{aligned} \quad (19)$$

$$\begin{aligned}\mu_{w,t} &= \frac{1}{1-\rho} \left[\rho \ln b + \frac{1}{2} \sigma_*^2 + (1 + \alpha \rho) \mu_\tau + \rho \gamma \mu_{s,t-1} \right], \\ \sigma_{w,t} &= \frac{\rho + \beta}{1-\rho} D = \frac{\rho D^2 + \alpha \sigma_\tau^2}{(1-\rho)D}.\end{aligned}\tag{20}$$

The Gini coefficients are

$$G_{s,t}^{\text{im}} = 2\Phi\left(\frac{\sigma_{s,t}}{\sqrt{2}}\right) - 1, \quad G_{w,t}^{\text{im}} = 2\Phi\left(\frac{\sigma_{w,t}}{\sqrt{2}}\right) - 1,$$

and $G_{w,t}^{\text{im}} < G_{s,t}^{\text{im}}$, because $\sigma_{w,t}/\sigma_{s,t} = b < 1$.

Proof. Posterior mean. Put $Y = \ln \tau_t$ and $Z = \ln s_{t-1}$. The log cost index $U = -(\alpha Y + \gamma Z)$ is a nondegenerate normal variable, so it has a positive density, and conditioning on $U = u$ is conditioning on $\alpha Y + \gamma Z = -u$. By Lemma 4.1,

$$T(u) = \mathbb{E}[e^Y | U = u] = \exp\{\mu_\tau - \beta(u + \alpha \mu_\tau + \gamma \mu_{s,t-1}) + \frac{1}{2} \sigma_*^2\} = \chi e^{-\beta u},$$

so $m(s, u) = s^\rho T(u)$ is C^1 with $\partial_u m = -\beta m < 0$: A1 and A2 hold, and (PC) and Eq. (9) hold with β as in Eq. (12).

Equilibrium. By Proposition 3.1, the unique regular equilibrium is $w = As^b$ with $b = (\rho + \beta)/(1 + \beta)$ and $A = \chi^{1/(1+\beta)} b^{-\beta/(1+\beta)}$, which is Eq. (18).

Distributions. From $w'(s) = bAs^{b-1}$, the first-order condition $\ln w'(s_t) = -\alpha Y - \gamma Z$ and $1 - b = (1 - \rho)/(1 + \beta)$,

$$\ln s_t = \frac{1 + \beta}{1 - \rho} [\ln(bA) + \alpha Y + \gamma Z],$$

which is normal with variance $(\frac{1+\beta}{1-\rho})^2 D^2$. Using $\ln(bA) = \frac{1}{1+\beta} [\ln b + (1 - \alpha \beta) \mu_\tau - \gamma \beta \mu_{s,t-1} + \frac{1}{2} \sigma_*^2]$ and collecting terms gives Eq. (19). Finally $\ln w_t = b \ln s_t + \ln A$, and algebra (the coefficient of $\ln b$ is $\frac{\rho}{1-\rho}$, that of μ_τ is $\frac{1+\alpha\rho}{1-\rho}$, that of $\mu_{s,t-1}$ is $\frac{\rho\gamma}{1-\rho}$) gives Eq. (20). The identities $(1 + \beta)D = (D^2 + \alpha \sigma_\tau^2)/D$ and $(\rho + \beta)D = (\rho D^2 + \alpha \sigma_\tau^2)/D$ follow from $\beta D^2 = \alpha \sigma_\tau^2$. ■

Each schooling level is now chosen by a mixture of abilities and backgrounds. Employers pay the expected productivity of the pool, which depends on how informative the signal is: β is the weight that the posterior mean of $\ln \tau$ puts on the signal, and σ_*^2 is the ability variance left unexplained. The wage elasticity b lies between ρ and 1, and the dispersion of schooling is driven by D , the dispersion of the cost index, scaled up by the factor $(1 + \beta)/(1 - \rho)$.

As $\gamma \downarrow 0$, $\beta \rightarrow 1/\alpha$, $\sigma_*^2 \rightarrow 0$, and the formulas for w and the distributions converge to those of the perfect-signal case.

4.3 Intergenerational Dynamics, Mobility, and the Stationary Distribution

This subsection links generations: today's distribution of schooling is tomorrow's distribution of family backgrounds. The aim is to measure social mobility by the intergenerational elasticity (IGE) and to find when inequality settles down to a stationary distribution rather than growing without bound.

Because τ_t is independent of s_{t-1} , the proof of [Proposition 4.3](#) shows that

$$\ln s_t = \text{const} + \phi \ln s_{t-1} + \frac{\alpha(1+\beta)}{1-\rho} \ln \tau_t, \quad \phi \equiv \frac{\gamma(1+\beta)}{1-\rho}, \quad (21)$$

so ϕ is the intergenerational elasticity (IGE) of schooling ([Solon 1992](#); [Corak 2013](#)), the standard measure of (im)mobility: a larger ϕ means lower social mobility. The coefficient depends on t because β depends on the cross-sectional dispersion of parental schooling.

Proposition 4.4 (Dynamics and Stationary Distribution). *Under [Eq. \(11\)](#) with $\gamma > 0$, suppose s_0 is log-normal. Then:*

- (i) Recursion. s_t is log-normal for all t , with $\sigma_{s,t}^2 = F(\sigma_{s,t-1}^2)$,

$$F(n) \equiv \frac{(\alpha^2 \sigma_\tau^2 + \gamma^2 n + \alpha \sigma_\tau^2)^2}{(1-\rho)^2(\alpha^2 \sigma_\tau^2 + \gamma^2 n)},$$

and $\mu_{s,t}$ given by [Eq. \(19\)](#).

- (ii) Stationarity. A stationary variance $n^* > 0$ with $F(n^*) = n^*$ exists if and only if $\gamma < 1 - \rho$, and then it is unique: $n^* = (x^* - \alpha^2 \sigma_\tau^2) / \gamma^2$, where $x^* = [b_0 + \sqrt{b_0^2 + 4a_0c_0}] / (2a_0)$ with $a_0 = (1-\rho)^2 - \gamma^2$, $b_0 = [(1-\rho)^2 \alpha^2 + 2\gamma^2 \alpha] \sigma_\tau^2$, and $c_0 = \gamma^2 \alpha^2 \sigma_\tau^4$. If $\gamma \geq 1 - \rho$, then $F(n) > n$ for all n , and $\sigma_{s,t}^2 \rightarrow \infty$.
- (iii) Stationary IGE. At the stationary variance, the IGE $\phi^* = \gamma(1+\beta)/(1-\rho)$, with $\beta = \alpha \sigma_\tau^2 / x^*$, lies in $(0, 1)$.
- (iv) Convergence. If $\alpha \geq 1$ and $\gamma < 1 - \rho$, then $\sigma_{s,t}^2 \rightarrow n^*$ monotonically from any initial variance, and $\mu_{s,t}$ converges to the unique solution of the linear fixed-point equation implied by [Eq. \(19\)](#), because its feedback coefficient $\gamma/(1-\rho) < 1$.

Proof.

- (i) is [Eq. \(19\)](#) rewritten with $\sigma_{s,t} = (D^2 + \alpha \sigma_\tau^2) / ((1-\rho)D)$.
- (ii) Write $x = \alpha^2 \sigma_\tau^2 + \gamma^2 n$, so $n = (x - \alpha^2 \sigma_\tau^2) / \gamma^2$. Then $F(n) > n$ if and only if $\gamma^2(x + \alpha \sigma_\tau^2)^2 > (1-\rho)^2 x(x - \alpha^2 \sigma_\tau^2)$, that is, $Q(x) \equiv a_0 x^2 - b_0 x - c_0 < 0$. Since $c_0 > 0$, $Q(0) < 0$. If $a_0 > 0$, Q has exactly one positive root x^* , $Q < 0$ on $(0, x^*)$ and $Q > 0$ beyond. Moreover $Q(\alpha^2 \sigma_\tau^2) = -\gamma^2 \alpha^2 \sigma_\tau^4 (1+\alpha)^2 < 0$, so $x^* > \alpha^2 \sigma_\tau^2$ and $n^* > 0$. Thus $F(n) > n$ for $n < n^*$ and $F(n) < n$ for $n > n^*$. If $a_0 \leq 0$, then $Q < 0$ for all $x > 0$, so $F(n) > n$ for all n ; an increasing sequence with no fixed point is unbounded.
- (iii) At the stationary state [Eq. \(21\)](#) gives $n^* = \phi^{*2} n^* + [\alpha(1+\beta)/(1-\rho)]^2 \sigma_\tau^2$ with $n^* > 0$,

hence $\phi^{*2} < 1$.

- (iv) For $\alpha \geq 1$, F is nondecreasing: with $x = D^2 \geq \alpha^2 \sigma_\tau^2 \geq \alpha \sigma_\tau^2$, $F = [x + 2\alpha \sigma_\tau^2 + \alpha^2 \sigma_\tau^4/x]/(1 - \rho)^2$ has derivative in x proportional to $1 - \alpha^2 \sigma_\tau^4/x^2 \geq 0$. Since F is nondecreasing and has the sign pattern from (ii), the sequence $\sigma_{s,t}^2 = F(\sigma_{s,t-1}^2)$ is monotone and bounded, and converges to n^* . The mean converges because it solves a linear recursion with coefficient $\gamma/(1 - \rho) < 1$ and a convergent forcing term. ■

The dispersion of schooling of one generation is a function of that of the previous generation. Inequality settles at a stable level if and only if $\gamma < 1 - \rho$: the elasticity of cost with respect to parental schooling must be smaller than $1 - \rho$, which measures how strongly diminishing returns damp the transmission of parental advantage. If $\gamma \geq 1 - \rho$, inequality in schooling grows without bound across generations. When it settles, the IGE is below one: children of better-schooled parents are better schooled, but regress partly toward the mean.

For $\alpha < 1$, F is not monotone for small n , and I claim only existence and uniqueness of the stationary variance. A comparative-statics remark: holding the parental distribution fixed, ϕ is increasing in γ whenever $\alpha \geq 1/8$ (in particular if $\alpha \geq 1$), since $\partial[(1 - \rho)\phi]/\partial\gamma = 1 + (a/\alpha)(a - x)/(a + x)^2 \geq 1 - 1/(8\alpha)$ with $a = \alpha^2 \sigma_\tau^2$ and $x = \gamma^2 \sigma_{s,t-1}^2$; and the signal quality $\mathcal{R}^2$ is strictly decreasing in γ .

4.4 Signaling versus Full Information

This subsection isolates the effect of the signaling friction itself. It compares the equilibrium with a benchmark in which employers see ability directly, keeping everything else (costs, distributions) the same, to find what the friction does to the dispersion of schooling and wages and to mobility.

To isolate the effect of the signaling friction, compare the equilibrium with the **full-information benchmark** in which employers observe τ_t and pay $w = e(s, \tau)$. Agents then maximize $s^\rho \tau - s \tau^{-\alpha} s'^{-\gamma}$, so that

$$\ln s^{\text{FI}} = \frac{1}{1 - \rho} [\ln \rho + (1 + \alpha) \ln \tau + \gamma \ln s'], \quad \ln w^{\text{FI}} = \rho \ln s^{\text{FI}} + \ln \tau.$$

Proposition 4.5 (Signaling Compresses Dispersion but Ties Schooling to Background). *Fix period t and the parental distribution $\text{LN}(\mu_{s,t-1}, \sigma_{s,t-1}^2)$ with $\gamma > 0$. Then:*

- (i) Compression.

$$\text{Var}(\ln s^{\text{FI}}) - \text{Var}(\ln s_t) = \text{Var}(\ln w^{\text{FI}}) - \text{Var}(\ln w_t) = \frac{\sigma_*^2}{(1 - \rho)^2} > 0,$$

so the Gini coefficients of schooling and wages are strictly lower under signaling than under full information.

- (ii) Sorting on ability versus background. *The elasticity of schooling with respect to ability is $\alpha(1 + \beta)/(1 - \rho)$ under signaling and $(1 + \alpha)/(1 - \rho)$ under full information, a shortfall of $(1 - \mathcal{R}^2)/(1 - \rho) > 0$; the elasticity with respect to parental schooling, that is, the IGE, is $\gamma(1 + \beta)/(1 - \rho)$ under signaling and $\gamma/(1 - \rho)$ under full information, an excess of $\gamma\beta/(1 - \rho) > 0$.*
- (iii) Perfect signal. *If $\gamma = 0$, dispersion coincides with full information but schooling is excessive: the mean of $\ln s$ exceeds its full-information value by $\frac{1}{1-\rho} \ln \frac{\alpha\rho+1}{\rho(\alpha+1)} > 0$.*

Proof.

- (i) With $x = \gamma^2 \sigma_{s,t-1}^2$ and $D^2 = \alpha^2 \sigma_\tau^2 + x$, Eq. (19) gives $(1 - \rho)^2 \text{Var}(\ln s_t) = D^2 + 2\alpha\sigma_\tau^2 + \alpha^2\sigma_\tau^4/D^2$, while $(1 - \rho)^2 \text{Var}(\ln s^{\text{FI}}) = (1 + \alpha)^2 \sigma_\tau^2 + x = D^2 + 2\alpha\sigma_\tau^2 + \sigma_\tau^2$. The difference is $\sigma_\tau^2 - \alpha^2\sigma_\tau^4/D^2 = \sigma_\tau^2 x/D^2 = \sigma_*^2$. For wages, $(1 - \rho)^2 \text{Var}(\ln w^{\text{FI}}) = (1 + \alpha\rho)^2 \sigma_\tau^2 + \rho^2 x$ and, from Eq. (20), $(1 - \rho)^2 \text{Var}(\ln w_t) = \rho^2 D^2 + 2\rho\alpha\sigma_\tau^2 + \alpha^2\sigma_\tau^4/D^2$; again the difference is σ_*^2 . (ii) Read off the coefficients of $\ln \tau$ and $\ln s'$ in Eq. (21) and in $\ln s^{\text{FI}}$, using $\alpha\beta = \mathcal{R}^2$. (iii) Compare Eq. (15) with $\ln s^{\text{FI}}$ for $\gamma = 0$ and use $\frac{\alpha\rho+1}{\alpha+1} > \rho$, which is equivalent to $\rho < 1$. ■

The friction compresses the cross-sectional distribution, because employers cannot tell high-ability from low-ability workers within a schooling level and so pay them alike. The size of the compression equals the part of ability variance that the market cannot infer, σ_*^2 . But the friction also reduces how much schooling responds to ability and raises how much it responds to parental schooling: this is the mobility cost of signaling.

The next proposition asks a different question: relative to an economy in which the cost depends only on ability, does adding a background effect on cost raise or lower inequality?

Proposition 4.6 (Adding Family Background to the Cost: Ambiguous Effect on Inequality). *Compare the period- t imperfect-signal equilibrium ($\gamma > 0$) with the perfect-signal equilibrium (same $\alpha, \rho, \sigma_\tau$). Then*

$$(1 - \rho)(\sigma_{s,t} - \sigma_s^{\text{pf}}) = \frac{(D - \alpha\sigma_\tau)(D - \sigma_\tau)}{D},$$

$$(1 - \rho)(\sigma_{w,t} - \sigma_w^{\text{pf}}) = \frac{(D - \alpha\sigma_\tau)(\rho D - \sigma_\tau)}{D}.$$

Since $D > \alpha\sigma_\tau$, schooling inequality is higher with background effects if and only if $D > \sigma_\tau$, that is, $\gamma^2 \sigma_{s,t-1}^2 > (1 - \alpha^2) \sigma_\tau^2$ (always true if $\alpha \geq 1$), and wage inequality is higher if and only if $\rho D > \sigma_\tau$. For $\sigma_\tau < D < \sigma_\tau/\rho$, schooling is more unequal while wages are less unequal. The Gini comparisons have the same signs.

Proof. Expand: $(D - \alpha\sigma)(D - \sigma)/D = D + \alpha\sigma^2/D - (1 + \alpha)\sigma$, which equals $(1 - \rho)\sigma_{s,t} - (1 + \alpha)\sigma_\tau$ by Eq. (19) and Eq. (15); similarly $(D - \alpha\sigma)(\rho D - \sigma)/D = \rho D + \alpha\sigma^2/D - (1 + \alpha\rho)\sigma$.

The Gini coefficient is increasing in the log-standard deviation. ■

Proposition 4.6 shows that two forces act in opposite directions: the cost index becomes more dispersed (it now varies with background as well as ability), which raises inequality, while the signal becomes less informative, which compresses it. Which force dominates depends on whether the dispersion of the cost index exceeds the dispersion of ability.

Remark 4.1 (Interpretation). **Proposition 4.5** shows that the cross-sectional compression of inequality produced by an imperfect signal is not benign. The compression equals the variance σ_*^2 of ability that the market cannot infer from schooling; it comes with a loss of ability-based sorting and a *higher* weight on family background, hence lower social mobility. Relative to the full-information allocation, signaling both misallocates schooling and strengthens its dependence on parental schooling. The model treats wages as deterministic functions of schooling; match-specific productivity shocks, revealed after hiring (compare [Rothschild and Stiglitz 1976](#) for the adverse-selection logic), would add within-group dispersion without changing these between-group results.

5 Finite Ability and Schooling Types

The log-normal economy gives smooth, closed-form results but cannot display discrete features of the problem: the coexistence of equilibria, the pooling of talented poor children with others, and the difference between temporary and permanent mobility. With finitely many ability and schooling types these features can be analyzed completely. The section shows how the same mechanism, the confounding of ability and background in the cost index, generates (i) pooling by background, (ii) multiple Pareto-ranked equilibria, and (iii) consequences for mobility and growth. It proceeds from general properties (sorting, separating and pooling equilibria) to a fully solved two-type economy and then to long-run implications.

To gain further analytical insight, I consider an economy with finitely many ability types and schooling levels,

$$\mathcal{T} = \{\tau_1 < \tau_2 < \dots < \tau_N\}, \quad \mathcal{S} = \{s_1 < s_2 < \dots < s_M\},$$

with $g_i = \Pr(\tau_t = \tau_i) > 0$ and $\pi_{k,t} = \Pr(s_t = s_k)$; parental schooling takes values in $\mathcal{S}$ as well. An agent is indexed by (i, k) : ability τ_i and parental schooling s_k . Costs have the form [Eq. \(4\)](#),

$$\theta(s_j, \tau_i, s_k) = c_j \kappa_{ik}, \quad c_j \equiv \theta_1(s_j), \quad \kappa_{ik} \equiv K(\tau_i, s_k), \quad (22)$$

with $0 = c_1 < c_2 < \dots < c_M$ (the lowest schooling level is costless) and κ_{ik} strictly decreasing in i and nonincreasing in k . The cost index need not be separable in τ_i and s_k . Productivity $e(s_j, \tau_i)$ is nondecreasing in i and in j . I write $\theta_{jik} \equiv \theta(s_j, \tau_i, s_k)$.

A **pure-strategy signaling equilibrium** of period t is a rule $\varsigma : \{1, \dots, N\} \times \{1, \dots, M\} \rightarrow \{1, \dots, M\}$ (the index j of the chosen level) and wages $(W_1, \dots, W_M)$ such that: (a) *optimality*:

$\varsigma(i, k) \in \arg \max_j [W_j - c_j \kappa_{ik}]$; (b) *Bayes consistency*: for every level j chosen with positive probability,

$$W_j = \frac{\sum_{(i,k): \varsigma(i,k)=j} g_i \pi_{k,t-1} e(s_j, \tau_i)}{\sum_{(i,k): \varsigma(i,k)=j} g_i \pi_{k,t-1}};$$

and (c) *off-path beliefs*: for a level j that is never chosen, $W_j \in [e(s_j, \tau_1), e(s_j, \tau_N)]$. The next state is $\pi_{j,t} = \sum_{(i,k): \varsigma(i,k)=j} g_i \pi_{k,t-1}$, and the **transition matrix** (rows indexed by parental schooling, rows summing to one) is

$$(P_t)_{kj} = \Pr(s_t = s_j \mid s_{t-1} = s_k) = \sum_{i=1}^N g_i 1\{\varsigma(i, k) = j\}, \quad \pi_t = \pi_{t-1} P_t. \quad (23)$$

5.1 Sorting by the Cost Index

This subsection establishes the finite-type analogue of Lemma 2.1: schooling reveals the cost index and nothing else. This is the basic structural fact behind all results of the section.

Lemma 5.1 (Monotone Sorting). *In any pure-strategy equilibrium, if $\kappa_{ik} < \kappa_{i'k'}$, then $\varsigma(i, k) \geq \varsigma(i', k')$. Agents with equal cost indices ($\kappa_{ik} = \kappa_{i'k'}$) face the same maximization problem and, up to tie-breaking, choose the same level.*

Proof. Let $j = \varsigma(i, k)$, $j' = \varsigma(i', k')$, $\kappa = \kappa_{ik} < \kappa' = \kappa_{i'k'}$. Optimality of both choices gives $W_j - c_j \kappa \geq W_{j'} - c_{j'} \kappa$ and $W_{j'} - c_{j'} \kappa' \geq W_j - c_j \kappa'$. Adding, $(c_j - c_{j'}) (\kappa' - \kappa) \geq 0$, so $c_j \geq c_{j'}$ and $j \geq j'$. ■

Thus schooling reveals the cost index and nothing else. If family background matters (θ_3 not constant), a talented child of poor parents and a less talented child of richer parents can have the same or interleaved cost indices, and then schooling cannot rank them by ability. This is the finite-type counterpart of the pooling in Figure 1.

Definition 5.1 (Equilibrium Types). A pure-strategy equilibrium ς is called

- (i) **fully separating** if $\varsigma(i, k) = \varsigma_0(i)$ for a strictly increasing ς_0 (schooling reveals ability and is independent of background);
- (ii) **background-pooling** if $\varsigma(i, k) = \varsigma_0(k)$ (ability plays no role in schooling given the family background); and
- (iii) **hybrid** (semi-separating) otherwise: some, but not all, ability differences are revealed.

5.2 Separating and Pooling Equilibria

This subsection characterizes the two polar kinds of equilibrium: separating equilibria, in which schooling reveals ability, and background-pooling equilibria, in which schooling simply reproduces the parent's level. The aim is to know exactly when each can arise.

Proposition 5.1 (Separating Equilibria when Background is Irrelevant). *Let $\theta_3 \equiv \text{const}$, so $\kappa_i \equiv \theta_2(\tau_i)\theta_3$ is strictly decreasing in i . Let $j_1 < \dots < j_N$ be schooling indices, and set $C_i = c_{j_i}$ and $W_i = e(s_{j_i}, \tau_i)$ (full revelation). Then no type wants to switch to another used level if and only if the adjacent constraints hold:*

$$\kappa_i (C_i - C_{i-1}) \leq W_i - W_{i-1} \leq \kappa_{i-1} (C_i - C_{i-1}), \quad i = 2, \dots, N. \quad (24)$$

The profile is an equilibrium (with the most pessimistic off-path wage $e(s_j, \tau_1)$ at unused levels j) if and only if, in addition, $e(s_j, \tau_1) - c_j \kappa_i \leq W_i - C_i \kappa_i$ for all i and all unused j .

Proof. Necessity of Eq. (24): type i does not mimic $i-1$ gives $W_i - \kappa_i C_i \geq W_{i-1} - \kappa_i C_{i-1}$, and type $i-1$ does not mimic i gives $W_{i-1} - \kappa_{i-1} C_{i-1} \geq W_i - \kappa_{i-1} C_i$. Sufficiency: let $r_l = (W_l - W_{l-1})/(C_l - C_{l-1}) \in [\kappa_l, \kappa_{l-1}]$. For type i , the payoff $V(l) = W_l - \kappa_i C_l$ satisfies $V(l) - V(l-1) = (C_l - C_{l-1})(r_l - \kappa_i) \geq 0$ for $l \leq i$ (since $r_l \geq \kappa_l \geq \kappa_i$) and ≤ 0 for $l > i$ (since $r_l \leq \kappa_{l-1} \leq \kappa_i$). Hence $l = i$ maximizes V over used levels. The off-path statement is the definition of equilibrium with the lowest admissible off-path wage, which makes deviations least attractive. ■

The constraints Eq. (24) are the discrete analogue of the first-order condition $w'(s) = \theta'_1(s)\kappa$ of Section 3 (the slope of the wage between adjacent signals is pinned between the marginal costs of the two types). Existence of a fully separating equilibrium requires a sufficiently rich grid of schooling levels; with a coarse grid only hybrid equilibria may exist, as in the two-type economy below.

Proposition 5.2 (Background-Pooling Equilibria). *Let $\varsigma_0 : \{1, \dots, M\} \rightarrow \{1, \dots, M\}$ and set $\bar{e}_j = \sum_i g_i e(s_j, \tau_i)$. Define $W_j = \bar{e}_j$ for levels j in the image of ς_0 on the support of π_{t-1} and $W_j = e(s_j, \tau_1)$ otherwise. The rule $\varsigma(i, k) = \varsigma_0(k)$ is a signaling equilibrium (with these wages) if and only if, for all i, k, j ,*

$$\bar{e}_{\varsigma_0(k)} - c_{\varsigma_0(k)} \kappa_{ik} \geq W_j - c_j \kappa_{ik}. \quad (25)$$

In such an equilibrium $s_t = s_{\varsigma_0(k)}$ for a child of a parent with schooling s_k ; in particular, if $\varsigma_0(k) = k$, then $P_t = I$ (no intergenerational mobility).

Proof. Because τ_t is independent of s_{t-1} and the choice depends only on k , the ability distribution among workers at any used level is the prior (g_i), so Bayes consistency yields $W_j = \bar{e}_j$. Condition Eq. (25) is optimality of $\varsigma_0(k)$ for every (i, k) . If an equilibrium of this form exists with some other off-path wages, replacing them by the lowest admissible ones weakly reduces deviation payoffs, so Eq. (25) holds; the converse is immediate. ■

5.3 Two-Type Illustration

This subsection solves a minimal economy completely, to see how pooling and mobility coexist as distinct equilibria and what determines which one is available over time.

Specialize to $N = M = 2$. Productivity is

$$e(s_1, \tau_i) = e_1 \ (i = 1, 2), \quad e(s_2, \tau_1) = e_2, \quad e(s_2, \tau_2) = e_3, \quad e_1 < e_2 < e_3, \quad (26)$$

so unskilled workers' productivity is independent of talent, while educated talented workers are more productive. The proportion of high-ability workers is $g_2 = p \in (0, 1)$. Let

$$\Delta(x) \equiv (e_2 - e_1) + x(e_3 - e_2), \quad x \in [0, 1],$$

be the wage premium of schooling s_2 over s_1 when a fraction x of the workers at s_2 is talented. Assume

$$\theta_{222} < \theta_{212} < \Delta(p) < \theta_{221} < \theta_{211}, \quad \theta_{222} < \Delta(0). \quad (27)$$

(The chain imposes no restriction on the cost index beyond monotonicity; if K is separable, $K = \theta_2 \theta_3$, it is consistent with Eq. (22) whenever $\theta_{222} \theta_{211} = \theta_{212} \theta_{221}$. For example, $p = \frac{1}{2}$, $(e_1, e_2, e_3) = (1, 2, 3)$, and $(\theta_{222}, \theta_{212}, \theta_{221}, \theta_{211}) = (0.6, 1.2, 1.8, 3.6)$ give $\Delta(p) = 1.5$.) Write $\pi_{1,t-1}$ for the fraction of parents with low schooling, assumed to lie in $(0, 1)$, and define

$$v_t \equiv \frac{p}{1 - (1 - p) \pi_{1,t-1}} \in (p, 1), \quad (28)$$

the fraction of talented workers among those with schooling s_2 when all agents except the least-able poor ones choose s_2 .

Proposition 5.3 (Equilibria of the Two-Type Economy). *Under Eq. (26) and Eq. (27), up to knife-edge indifferences, the pure-strategy equilibria of period t are exactly:*

- (E1) Background-pooling (immobility): $\varsigma(i, k) = k$. Wages are $W_1 = e_1$ and $W_2 = (1 - p)e_2 + pe_3$; the transition matrix is $P = I$; and the mass of talented agents with high schooling is $R_t = p \pi_{2,t-1}$. It exists in every period.
- (E2) Hybrid (mobility): all agents choose s_2 except type $(i, k) = (1, 1)$. Wages are $W_1 = e_1$ and $W_2 = (1 - v_t)e_2 + v_t e_3$; the transition matrix is $P_t = \begin{pmatrix} 1 - p & p \\ 0 & 1 \end{pmatrix}$; and $R_t = p$. It exists if and only if

$$\theta_{221} < \Delta(v_t) < \theta_{211}. \quad (29)$$

No other pure-strategy equilibrium exists. When both exist, E2 Pareto-dominates E1: agents $(2, 1)$ and all agents with $k = 2$ are strictly better off, and agent $(1, 1)$ is indifferent.

Proof. By Lemma 5.1 and Eq. (27), whose chain orders the cost indices as $\kappa_{22} < \kappa_{12} < \kappa_{21} < \kappa_{11}$, the set of agents choosing s_2 is an upper set in this order. Since $e(s_1, \cdot) \equiv e_1$, $W_1 = e_1$ always, and agent (i, k) chooses s_2 if and only if $W_2 - e_1 > \theta_{2ik}$. The five candidate sets: (a) *none*: then $W_2 \geq e_2$ off path, so agent $(2, 2)$ gains at least $\Delta(0) > \theta_{222}$ by deviating; (b) *only* $(2, 2)$: then $W_2 = e_3$ and agent $(1, 2)$ gains $\Delta(1) > \Delta(p) > \theta_{212}$ by choosing s_2 ; (c) $\{(2, 2), (1, 2)\}$: the ability distribution at s_2 is the prior, $W_2 - e_1 = \Delta(p)$, and the choices are optimal by Eq. (27); this is E1; (d) *all but* $(1, 1)$: the share of talent at s_2 is $[p\pi_{2,t-1} + p\pi_{1,t-1}]/[\pi_{2,t-1} + p\pi_{1,t-1}] = v_t$, and optimality requires $\theta_{221} < \Delta(v_t) < \theta_{211}$; (e) *all*: then $W_2 - e_1 = \Delta(p) < \theta_{221}$, so agent $(2, 1)$ deviates. For the Pareto statement: payoffs at $(1, 1)$ equal e_1 in both; at $(2, 1)$ they are e_1 in E1 and $e_1 + \Delta(v_t) - \theta_{221} > e_1$ in E2; for $k = 2$ they are $e_1 + \Delta(\cdot) - \theta_{212}$ with $\Delta(v_t) > \Delta(p)$. ■

Proposition 5.3 shows that in E1 every child ends up where her parents were: talented children of low-schooling parents do not acquire the signal and earn the low wage. In E2 all children except the least able poor ones acquire the signal, so schooling at the top is a mix of talents, the wage at the top is lower than under perfect revelation but higher than under E1, and talented poor children move up. Every agent except the least able poor (who are indifferent) is strictly better off in E2, so E2 Pareto-dominates E1 when it exists.

Proposition 5.4 (Dynamics and Growth in the Two-Type Economy). *Suppose E2 is played from $t = 1$ while it exists. Then $\pi_{1,t} = (1 - p)^t \pi_{1,0}$, $v_t = p/[1 - (1 - p)^t \pi_{1,0}]$ is strictly decreasing and converges to p , and E2 is an equilibrium in period t if and only if $t \leq t_0$, where t_0 is the largest t with $\Delta(v_t) > \theta_{221}$ (finite, since $\Delta(v_t) \rightarrow \Delta(p) < \theta_{221}$; $t_0 \geq 1$ iff $\Delta(v_1) > \theta_{221}$), provided $\Delta(v_1) < \theta_{211}$. After t_0 the economy plays E1 forever, so $\pi_{1,t} = (1 - p)^{t_0} \pi_{1,0}$ for $t \geq t_0$. Under the growth assumption below, the path with E2 until t_0 has strictly higher talent utilization than the permanent E1 path whenever $t_0 \geq 1$:*

$$R_t = p \quad (t \leq t_0), \quad R_t = p[1 - (1 - p)^{t_0} \pi_{1,0}] \quad (t > t_0), \quad \text{versus} \\ R_t = p(1 - \pi_{1,0}) \text{ under permanent E1.}$$

Proof. Under E2, $\pi_{1,t} = (1 - p)\pi_{1,t-1}$ (only $(1, 1)$ chooses s_1), which gives the closed form and monotonicity of v_t . The upper bound in Eq. (29) is monotone: if $\Delta(v_1) < \theta_{211}$ it holds for all t . The lower bound fails once $v_t \leq (\theta_{221} - e_2 + e_1)/(e_3 - e_2)$; since π_1 stops changing under E1, v does not rise again and E2 never becomes available again. The expressions for R_t follow from Proposition 5.3. ■

Mobility under E2 shrinks the pool of poor families, which dilutes the share of talent among educated workers; once that share falls below a threshold, the signal is no longer worth its cost for talented poor children and the economy falls back into immobility. The mobile equilibrium is therefore temporary unless costs for poor children fall.

Example 5.1 (Numerical Illustration). Let $p = \frac{1}{2}$, $(e_1, e_2, e_3) = (1, 2, 3)$, so that $\Delta(x) = 1 + x$, and $(\theta_{222}, \theta_{212}, \theta_{221}, \theta_{211}) = (0.6, 1.2, 1.8, 3.6)$, so that $\Delta(p) = 1.5$ and Eq. (27) holds. With

$\pi_{1,0} = 0.9$, Eq. (28) gives $v_1 = 0.909$ and $\Delta(v_1) = 1.909 \in (1.8, 3.6)$, so E2 exists in period 1. In period 2, $v_2 = 0.645$ and $\Delta(v_2) = 1.645 < 1.8$, so E2 no longer exists and $t_0 = 1$. Talent utilization is $R_1 = 0.5$ in period 1 and $R_t = 0.5 \times (1 - 0.5 \times 0.9) = 0.275$ for $t \geq 2$, against $R_t = 0.5 \times 0.1 = 0.05$ under permanent E1.

5.4 General Finite Economies: Mobility, Long-Run Distribution, and Growth

This subsection extends the analysis to any finite economy, to measure mobility and talent utilization by standard tools (Markov chains) and to link them to growth.

Growth. Let $R_t = \Pr(\tau_t = \tau_N, s_t = s_M)$ be the mass of agents who combine top ability with top schooling (knowledge workers). In the spirit of knowledge externalities (Romer 1990; Raut 2005), I assume the growth rate of the economy's knowledge stock in period t is $\varpi(R_t)$ for a strictly increasing function ϖ . By independence of τ_t and s_{t-1} ,

$$R_t = g_N \sum_k \pi_{k,t-1} 1\{\varsigma(N, k) = M\} \leq g_N, \quad (30)$$

with equality if and only if every top-ability agent attains s_M .

Proposition 5.5 (Mobility, Stationary Distribution, and Talent Utilization). *Let ς be time-invariant and P the matrix Eq. (23).*

- (i) *Polar cases. In a fully separating equilibrium all rows of P equal $\pi_j^{\text{sep}} = \sum_i g_i 1\{\varsigma_0(i) = j\}$; hence $\pi_t = \pi^{\text{sep}}$ for all $t \geq 1$, the second eigenvalue is $\lambda_2 = 0$, and the Shorrocks mobility index $(M - \text{tr } P)/(M - 1)$ equals one (Shorrocks 1978). In a background-pooling equilibrium with $\varsigma_0(k) = k$, $P = I$, every distribution is stationary, and the index equals zero.*
- (ii) *Stationary distribution. If P is primitive (irreducible and aperiodic), there is a unique $\pi^* > 0$ with $\pi^* = \pi^* P$, and for every $r \in (|\lambda_2|, 1)$ there is C with $\|\pi_t - \pi^*\|_1 \leq C r^t$ for every initial π_0 , where λ_2 is the eigenvalue of second-largest modulus; if P is diagonalizable, $r = |\lambda_2|$ is attainable. The spectral gap $1 - |\lambda_2|$ is a standard mobility index (Sommers and Conlisk 1979).*
- (iii) *Talent utilization. In a fully separating equilibrium with $\varsigma_0(N) = M$, $R_t = g_N$ for all $t \geq 1$, the maximum possible value; in a background-pooling equilibrium, $R_t = g_N \Pr(\varsigma_0(s_{t-1}) = M)$.*

Proof.

- (i) If $\varsigma(i, k) = \varsigma_0(i)$, the sum in Eq. (23) does not depend on k ; then $\text{tr } P = \sum_k \pi_k^{\text{sep}} = 1$. If $\varsigma(i, k) = k$, $(P)_{kj} = \sum_i g_i 1\{k = j\} = 1\{k = j\}$. (ii) is the Perron–Frobenius theorem for primitive stochastic matrices and the spectral decomposition of P^t (Seneta 1981; Stokey and Lucas 1989); without diagonalizability, Jordan blocks of size m at $|\lambda_2|$ produce a factor t^{m-1} , absorbed by taking $r > |\lambda_2|$. (iii) follows from Eq. (30).

The condition that P be *primitive* (not merely irreducible) matters: an irreducible but periodic matrix, such as a permutation, does not converge. The two-type hybrid equilibrium E2 is not time-invariant (it depends on v_t); its transition matrix has the absorbing state s_2 and eigenvalues 1 and $1 - p$, consistent with $\pi_{1,t} = (1 - p)^t \pi_{1,0}$.

Remark 5.1 (Multiplicity, Beliefs, and Refinements). The two-type economy has two coexisting equilibria, E1 and E2, which differ in mobility, talent utilization, and welfare, and which are supported by different employer beliefs about the ability composition at s_2 . Equilibrium selection therefore requires a theory of expectations. When family background is irrelevant, the least-cost separating equilibrium is the outcome that survives standard refinements under single crossing (Cho and Kreps 1987; Cho and Sobel 1990); with two-dimensional heterogeneity no such selection result is available. That self-fulfilling beliefs can sustain group differences is familiar from the statistical-discrimination literature (Coate and Loury 1993; Lang and Manove 2011).

Remark 5.2 (Policy Implications). Early childhood development programs reduce the cost disadvantage of children from poor families, that is, they lower the effective γ in the log-normal economy (and θ_{221} , θ_{211} in the two-type economy). In the log-normal economy, a lower γ raises the signal quality $\mathcal{R}^2$ and, when $\alpha \geq 1/8$, lowers the IGE (see the comparative statics after Proposition 4.4). In the two-type economy, suppose a program lowers θ_{221} below $\Delta(p)$. Then E1 is no longer an equilibrium: talented children of poor parents find schooling worthwhile even under pooled beliefs. If $\Delta(v_1) < \theta_{211}$, E2 is an equilibrium in every period, because $\Delta(v_t) \downarrow \Delta(p) > \theta_{221}$; then $R_t = p$ and $\pi_{1,t} = (1 - p)^t \pi_{1,0} \rightarrow 0$. A modest reduction of the cost disadvantage can thus convert the temporary mobility of E2 into a permanent one. Apart from the inequalities $\theta_{222} < \theta_{212} < \Delta(p)$ and $\theta_{222} < \Delta(0)$ of Eq. (27), which any cost reduction preserves, the existence conditions for E1 and E2 involve only θ_{221} and θ_{211} . A uniform proportional subsidy to schooling, which scales every θ_{2ik} by the same factor $\delta < 1$, therefore yields the same equilibria as a reduction of θ_{221} and θ_{211} by the same factor alone, but it also subsidizes children who would invest anyway. The long-run returns to early childhood programs documented by Heckman et al. (2010), Heckman and Raut (2016), García-Peña et al. (2016), Raut (2018), and Reynolds et al. (2018) are consistent with this channel, but the model does not by itself identify it.

6 Conclusion

This paper studied how education, used as a costly signal of unobserved ability, transmits family-background disadvantage into earnings inequality and low social mobility. By *background* I mean parental schooling: because the cost of schooling falls with it, a child of low-schooling parents faces a higher cost index than an equally able child of high-schooling parents. “Background inequality” is therefore the dispersion of parental schooling as it enters the cost index (in the log-normal economy, the term $\gamma \ln s_{t-1}$); it is not inequality of income or wealth, which

the model does not include. Because schooling choices depend on agents only through the cost index, employers can infer ability only imperfectly, and the equilibrium wage schedule solves a first-order ODE (Theorem 3.1). In log-normal economies, signaling compresses the dispersion of log schooling and log wages by exactly the variance of ability that schooling fails to reveal, yet it raises the intergenerational elasticity of schooling, and a stationary distribution exists if and only if $\gamma < 1 - \rho$ (Proposition 4.5, Proposition 4.4).

The signaling trap. With finitely many types, the dependence of the cost $\theta(s, \tau, s_{t-1})$ on parental schooling can sustain a pooling equilibrium in which talented children of low-schooling parents do not acquire the signal and are under-utilized, while a Pareto-superior equilibrium with upward mobility coexists. Which one prevails depends on beliefs, and without policy the mobile equilibrium is temporary (Proposition 5.3, Proposition 5.4, Proposition 5.5).

Early childhood development and policy. The model identifies the cost differential between children of low- and high-schooling parents as the lever that governs mobility and talent utilization. Early childhood development (ECD) programs plausibly act on this lever by improving cognitive and affective development before schooling begins. A sufficient reduction in the cost faced by talented poor children raises the probability that they obtain the signal, shifts the economy from pooling to the mobile equilibrium, raises the mass R_t of talented workers in technical jobs and hence growth, and raises the upward transition probability $\Pr(s_t = s_2 \mid s_{t-1} = s_1)$ (5.2). In the two-type economy only the costs of poor-background children enter the equilibrium conditions, so a uniform schooling subsidy yields the same equilibria as a targeted reduction of the same proportion but also subsidizes children who invest anyway. This favors capability investments aimed at the cost differential over credentialing subsidies, although the relative effectiveness of the two instruments is an empirical question that the model does not settle. Extensions, namely endogenous parental investment and non-Gaussian distributions, are discussed in 2.1 and 3.2.

References

- Acemoglu, Daron, and David Autor. 2011. “Skills, Tasks and Technologies: Implications for Employment and Earnings.” Human Capital, Technology. *Handbook of Labor Economics* 4: 1043–171. [https://doi.org/10.1016/S0169-7218\(11\)02410-5](https://doi.org/10.1016/S0169-7218(11)02410-5).
- Aitchison, John, and James A. C. Brown. 1957. *The Lognormal Distribution*. Cambridge University Press.
- Anderson, Theodore W. 2003. *An Introduction to Multivariate Statistical Analysis*. 3rd ed. Wiley.
- Arrow, Kenneth J. 1973. “Higher Education as a Filter.” *Journal of Public Economics* 2 (3): 193–216.

- Autor, D. H. 2014. “Skills, Education, and the Rise of Earnings Inequality Among the ‘Other 99 Percent’.” Inequality. *Science* 344 (6186): 843–51. <https://doi.org/10.1126/science.1251868>.
- Barnett, W. Steven. 1995. “Long-Term Effects of Early Childhood Programs on Cognitive and School Outcomes.” Preschool. *The Future of Children* 5 (3): 25–50. <https://doi.org/10.2307/1602366>.
- Becker, Gary S., and Nigel Toms. 1979. “An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility.” *Journal of Political Economy* 87 (6): 1153–89.
- Bourguignon, François. 2015. *The Globalization of Inequality*. Inequality. Princeton University Press.
- Bowlby, John. 1982. “Attachment and Loss: Retrospect and Prospect.” Child Development. *American Journal of Orthopsychiatry* 52 (4): 664–78. <https://doi.org/10.1111/j.1939-0025.1982.tb01456.x>.
- Brockwell, Peter J., and Richard A. Davis. 1991. *Time Series: Theory and Methods*. 2nd ed. Springer.
- Chakrabarti, Gagari, and Tapas Chatterjea. 2017. “Intelligence? ... Emotions? ... or, the Emotional Intelligence: Theories and Evidence in Global Context.” Psychology, Emotional Intelligence to *Employees’ Emotional Intelligence, Motivation & Productivity, and Organizational Excellence*. Springer Singapore. https://doi.org/10.1007/978-981-10-5759-5_2.
- Champagne, Danielle L., Rosemary C. Bagot, Felisa van Hasselt, et al. 2008. “Maternal Care and Hippocampal Plasticity: Evidence for Experience-Dependent Structural Plasticity, Altered Synaptic Functioning, and Differential Responsiveness to Glucocorticoids and Stress.” Neuroscience, Epigenetics. *Journal of Neuroscience* 28 (23): 6037–45. <https://doi.org/10.1523/JNEUROSCI.0526-08.2008>.
- Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. “Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States.” *The Quarterly Journal of Economics* 129 (4): 1553–623.
- Cho, In-Koo, and David M. Kreps. 1987. “Signaling Games and Stable Equilibria.” Signaling, Equilibrium Refinements. *The Quarterly Journal of Economics* 102 (2): 179–221. <https://doi.org/10.2307/1885060>.
- Cho, In-Koo, and Joel Sobel. 1990. “Strategic Stability and Uniqueness in Signaling Games.” *Journal of Economic Theory* 50 (2): 381–413.

- Coate, Stephen, and Glenn C. Loury. 1993. "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" *The American Economic Review* 83 (5): 1220–40.
- Coddington, Earl A., and Norman Levinson. 1955. *Theory of Ordinary Differential Equations*. ODE Theory. McGraw-Hill.
- Corak, Miles. 2013. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." *Journal of Economic Perspectives* 27 (3): 79–102.
- Cunha, Flavio, and James J. Heckman. 2007. "The Technology of Skill Formation." Human Capital, Skill Formation. *American Economic Review* 97 (2): 31–47. <https://doi.org/10.1257/aer.97.2.31>.
- Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883–931.
- Currie, Janet. 2001. "Early Childhood Education Programs." Preschool. *Journal of Economic Perspectives* 15 (2): 213–38. <https://www.jstor.org/stable/2696599>.
- Currie, Janet. 2011. "Inequality at Birth: Some Causes and Consequences." Inequality. *The American Economic Review* 101 (3): 1–22. <http://www.jstor.org/stable/29783707>.
- Currie, Janet, and Douglas Almond. 2011. "Human Capital Development Before Age Five." Human Capital, Child Development. *Handbook of Labor Economics*, 1315–486. [https://doi.org/10.1016/s0169-7218\(11\)02413-0](https://doi.org/10.1016/s0169-7218(11)02413-0).
- Duncan, Greg J., Kathleen M. Ziol-Guest, and Ariel Kalil. 2010. "Early-Childhood Poverty and Adult Attainment, Behavior, and Health." Inequality. *Child Development* 81 (1): 306–25. <https://doi.org/10.1111/j.1467-8624.2009.01396.x>.
- Efron, Bradley. 1965. "Increasing Properties of Pólya Frequency Functions." *The Annals of Mathematical Statistics* 36 (1): 272–79.
- Entwisle, Doris R. 1995. "The Role of Schools in Sustaining Early Childhood Program Benefits." Preschool. *The Future of Children* 5 (3): 133–44. <https://doi.org/10.2307/1602371>.
- García, Jorge Luis, James Heckman, Duncan Ermini Leaf, and María José Prados. 2016. *The Life-Cycle Benefits of an Influential Early Childhood Program*. Child Development No. 22993. NBER Working Paper Series. National Bureau of Economic Research. <https://doi.org/10.3386/w22993>.
- Gluckman, Peter D., Mark A. Hanson, Cyrus Cooper, and Kent L. Thornburg. 2008. "Effect

- of in Utero and Early-Life Conditions on Adult Health and Disease.” *Early Childhood Events, Health, Epigenetics. New England Journal of Medicine* 359 (1): 61–73. <https://doi.org/10.1056/nejmra0708473>.
- Hartman, Philip. 2002. *Ordinary Differential Equations*. ODE Theory. 2nd ed. Society for Industrial; Applied Mathematics. <https://doi.org/10.1137/1.9780898719222>.
- Heckman, James J. 2000. “Policies to Foster Human Capital.” *Preschool, Human Capital. Research in Economics* 54 (1): 3–56. <https://doi.org/10.1006/reec.1999.0225>.
- Heckman, James J. 2006. “Skill Formation and the Economics of Investing in Disadvantaged Children.” *Science* 312 (5782): 1900–1902. <https://doi.org/10.1126/science.1128898>.
- Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Q. Yavitz. 2010. “Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program.” *Preschool. Quantitative Economics* 1 (1): 1–46. <https://doi.org/10.3982/QE8>.
- Heckman, James J., and Lakshmi K. Raut. 2013. *Intergenerational Long Term Effects of Preschool: Structural Estimates from a Discrete Dynamic Programming Model*. Working Paper No. 19077. NBER Working Paper Series. National Bureau of Economic Research. <https://doi.org/10.3386/w19077>.
- Heckman, James J., and Lakshmi K. Raut. 2016. “Intergenerational Long-Term Effects of Preschool: Structural Estimates from a Discrete Dynamic Programming Model.” *Preschool, Dynamic Programming. Journal of Econometrics* 191 (1): 164–75. <https://doi.org/10.1016/j.jeconom.2015.10.001>.
- Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior.” *Journal of Labor Economics* 24 (3): 411–82.
- Herrnstein, Richard, and Charles Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. Behavioral Genetics. 1st Free Press pbk. Free Press.
- Karlin, Samuel, and Herman Rubin. 1956. “The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio.” *The Annals of Mathematical Statistics* 27 (2): 272–99. <https://doi.org/10.1214/aoms/1177728259>.
- Keane, Michael P., and Kenneth I. Wolpin. 1997. “The Career Decisions of Young Men.” *Human Capital, Labor Economics. Journal of Political Economy* 105 (3): 473–522. <https://doi.org/10.1086/262080>.

- Kreps, David M. 1990. *A Course in Microeconomic Theory*. Microeconomics, Game Theory. Princeton University Press.
- Lang, Kevin, and Michael Manove. 2011. “Education and Labor Market Discrimination.” *The American Economic Review* 101 (4): 1467–96.
- Mailath, George J. 1987. “Incentive Compatibility in Signaling Games with a Continuum of Types.” Signaling, Game Theory. *Econometrica* 55 (6): 1349–65. <https://doi.org/10.2307/1913577>.
- Mailath, George J., and Ernst-Ludwig von Thadden. 2013. “Incentive Compatibility and Differentiability: New Results and Classic Applications.” *Journal of Economic Theory* 148 (5): 1841–61.
- Maluccio, John A., John Hoddinott, Jere R. Behrman, Reynaldo Martorell, Agnes R. Quisumbing, and Aryeh D. Stein. 2009. “The Impact of Improving Nutrition During Early Childhood on Education Among Guatemalan Adults.” Child Development, Nutrition. *The Economic Journal* 119 (537): 734–63. <https://doi.org/10.1111/j.1468-0297.2009.02220.x>.
- Mayer, John D., Peter Salovey, and David R. Caruso. 2004. “Emotional Intelligence: Theory, Findings, and Implications.” Psychology, Emotional Intelligence. *Psychological Inquiry* 15 (3): 197–215. https://doi.org/10.1207/s15327965pli1503_02.
- McEwen, Bruce S., and Peter J. Gianaros. 2011. “Stress- and Allostasis-Induced Brain Plasticity.” Neuroscience, Stress. *Annual Review of Medicine* 62 (1): 431–45. <https://doi.org/10.1146/annurev-med-052209-100430>.
- Milgrom, Paul, and Chris Shannon. 1994. “Monotone Comparative Statics.” *Econometrica* 62 (1): 157–80. <https://doi.org/10.2307/2951479>.
- NICHD Early Child Care Research Network. 2004. “Affect Dysregulation in the Mother–Child Relationship in the Toddler Years: Antecedents and Consequences.” Child Development, Attachment. *Development and Psychopathology* 16 (1). <https://doi.org/10.1017/s0954579404044402>.
- Noble, Kimberly G., Suzanne M. Houston, Natalie H. Brito, et al. 2015. “Family Income, Parental Education and Brain Structure in Children and Adolescents.” Child Development, Neuroscience. *Nature Neuroscience* 18 (5): 773–78. <https://doi.org/10.1038/nn.3983>.
- Olver, Peter J. 1993. *Applications of Lie Groups to Differential Equations*. 2nd ed. Graduate Texts in Mathematics 107. Springer. <https://doi.org/10.1007/978-1-4612-4350-2>.

- Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Growth and Inequality. Belknap Press.
- Plomin, R., and I. J. Deary. 2015. “Genetics and Intelligence Differences: Five Special Findings.” Behavioral Genetics. *Molecular Psychiatry* 20 (1): 98–108. <https://doi.org/10.1038/mp.2014.105>.
- Quinzii, Martine, and Jean-Charles Rochet. 1985. “Multidimensional Signalling.” Signaling, PDE. *Journal of Mathematical Economics* 14 (3): 261–84. [https://doi.org/10.1016/0304-4068\(85\)90003-5](https://doi.org/10.1016/0304-4068(85)90003-5).
- Raut, Lakshmi K. 1990. “Capital Accumulation, Income Distribution and Endogenous Fertility in an Overlapping Generations General Equilibrium Model.” *Journal of Development Economics* 34 (1-2): 123–50. [https://doi.org/10.1016/0304-3878\(90\)90079-q](https://doi.org/10.1016/0304-3878(90)90079-q).
- Raut, Lakshmi K. 2003. *Long Term Effects of Preschool Investment on School Performance and Labor Market Outcome*. Working Paper No. 0307002. University Library of Munich, Germany. <https://ideas.repec.org/p/wpa/wuwpla/0307002.html>.
- Raut, Lakshmi K. 2005. *Signaling Equilibrium, Intergenerational Social Mobility and Long Run Growth*. <https://doi.org/10.2139/ssrn.852924>.
- Raut, Lakshmi K. 2018. “Long-Term Effects of Preschool on School Performance, Earnings and Social Mobility.” Preschool, Social Mobility. *Studies in Microeconomics* 6 (1-2): 24–49. <https://doi.org/10.1177/2321022218802023>.
- Reynolds, Arthur J., Suh-Ruu Ou, and Judy A. Temple. 2018. “A Multicomponent, Preschool to Third Grade Preventive Intervention and Educational Attainment at 35 Years of Age.” Preschool. *JAMA Pediatrics*, ahead of print, January. <https://doi.org/10.1001/jamapediatrics.2017.4673>.
- Reynolds, Arthur J., Judy A. Temple, Dylan L. Robertson, and Emily A. Mann. 2001. “Long-Term Effects of an Early Childhood Intervention on Educational Achievement and Juvenile Arrest.” Preschool. *JAMA* 285 (18): 2339–46. <https://doi.org/10.1001/jama.285.18.2339>.
- Riley, John G. 1979. “Informational Equilibrium.” Signaling, Information Economics. *Econometrica* 47 (2): 331–59. <https://doi.org/10.2307/1913366>.
- Riley, John G. 2001. “Silver Signals: Twenty-Five Years of Screening and Signaling.” Signaling, Survey. *Journal of Economic Literature* 39 (2): 432–78. <https://doi.org/10.1257/jes.39.2.432>.

- Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5, Part 2): S71–102.
- Rothschild, Michael, and Joseph E. Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." Signaling, Insurance. *The Quarterly Journal of Economics* 90 (4): 629–49. <https://doi.org/10.2307/1885326>.
- Schweinhart, L. J., Helen V. Barnes, and David Weikart. 1993. *Significant Benefits: The High-Scope Perry Preschool Study Through Age 27*. Preschool. High/Scope Press. <https://eric.ed.gov/?id=ED366433>.
- Seneta, Eugene. 1981. *Non-Negative Matrices and Markov Chains*. 2nd ed. Springer.
- Shorrocks, Anthony F. 1978. "Income Mobility and the Markov Assumption." Social Mobility, Income Distribution. *The Economic Journal* 88 (311): 566–78. <https://doi.org/10.2307/2231972>.
- Solon, Gary. 1992. "Intergenerational Income Mobility in the United States." *The American Economic Review* 82 (3): 393–408.
- Sommers, Paul M., and John Conlisk. 1979. "Eigenvalue Immobility Measures for Markov Chains." *The Journal of Mathematical Sociology* 6 (2): 253–76.
- Spence, A. Michael. 1974. *Market Signaling: Informational Transfer in Hiring and Related Screening Processes*. Signaling, Labor Economics. Harvard Economic Studies. Harvard University Press.
- Spence, Michael. 1973. "Job Market Signaling." Signaling. *The Quarterly Journal of Economics* 87 (3): 355–74. <http://www.jstor.org/stable/1882010>.
- Stiglitz, Joseph E. 1975. "The Theory of 'Screening,' Education, and the Distribution of Income." Signaling, Screening. *The American Economic Review* 65 (3): 283–300. <http://www.jstor.org/stable/1804834>.
- Stiglitz, Joseph E. 2015. *The Great Divide: Unequal Societies and What We Can Do about Them*. Inequality, Policy. W. W. Norton & Company.
- Stokey, Nancy L., and Robert E. Lucas. 1989. *Recursive Methods in Economic Dynamics*. Dynamic Programming, Macroeconomics. Harvard University Press.
- Weiss, Andrew. 1995. "Human Capital Vs. Signalling Explanations of Wages." *Journal of Economic Perspectives* 9 (4): 133–54.