The Causal Effect of Parents' Education
on Children's Earnings [*]

Sang Yoon (Tim) Lee
TSE and CEPR

Nicolas Roys
Royal Holloway

Ananth Seshadri
UW-Madison

March 2018

Abstract

We present a model of endogenous schooling and earnings to isolate the causal effect of parents' education on children's education and earnings outcomes. The model suggests that parents' education is positively related to children's earnings, but its relationship with children's education is ambiguous. Identification is achieved by comparing the earnings of children with the same education, whose parents have different levels of education. The model also features heterogeneous tastes for schooling, and is estimated using HRS data. The empirically observed positive OLS coefficient obtained by regressing children's schooling on parents' schooling is mainly accounted for by the correlation between parents' schooling and children's unobserved tastes for schooling. This is countered by a negative, structural relationship between parents' and children's schooling choices, resulting in an IV coefficient close to zero when exogenously increasing parents' schooling. Nonetheless, an exogenous one-year increase in parents' schooling increases children's lifetime earnings by 1.2 percent on average.

1

1   Introduction

Parents have a large influence on their children's outcomes. Does this merely reflect selection—correlation in unobserved heterogeneity across generations? Or does it also partly reflect human capital spillovers from parent to child? In the former case, government subsidies aimed at improving education would only impact one generation. But in the presence of intergenerational spillovers, the returns from such public investments are reaped by all succeeding members of a dynasty, resulting in long-lasting effects.

To address these questions, we begin with a simple model of life-cycle human capital accumulation along the lines of Ben-Porath (1967), which encompasses both schooling and learning on-the-job. An individual's length of schooling and earnings profile is determined by his initial level of human capital at age 6, and his learning ability—the amount of human capital he can accumulate in a unit of time. We assume that an individual's initial level of human capital is a function of his learning ability and his parent's human capital, the latter of which is the source of parental spillovers in our model.[1] Learning abilities are unobserved but correlated across generations and persist through life, unlike parents' human capital which only affect children early on in life.

We separately identify parental spillovers from selection on abilities by using information on children's schooling and earnings, in addition to parental background. The key assumption is that children's earnings is a function of their learning abilities and schooling, while schooling is a function of their learning abilities and their parents' human capital. That is, parents directly affect how long their children stay in school, but only indirectly affect their children's earnings through children's schooling. In our model, if higher human capital parents have children with higher initial levels of human capital, all else equal, such children spend less time in school since they can reach higher levels of human capital more quickly. We refer to this as the "level effect": higher initial conditions substitute the need for later investments.

Thus, our model can rationalize findings from the previous literature that find a negative causal relationship between mothers' and children's schooling (Behrman and Rosenzweig, 2002). Of course, this does not mean that the observed relationship between parents' and children's schooling should be negative. If high human capital parents also have high learning abilities, and such abilities are passed on to their children, the children will want to stay in school longer since the returns are higher. This can countervail and dominate the level effect, generating a positive OLS relationship between parents' and children's schooling. This is consistent with the empirical literature which finds that the positive intergenerational OLS coefficient is not causal, but rather driven by selection: the positive correlation between parents' schooling and unobserved factors (Behrman and Rosenzweig, 2002; Black et al., 2005).

Differentiating the length of schooling from human capital—the quality of education—is im-

---

[1]Henceforth we refer to the causal effect of parents' human capital on children's earnings as the "parental spillover."

portant for identification: children who attain the same years of schooling can have different levels of human capital, which is manifested in the data as differential earnings. In the model, such earnings differences can be fully accounted for by the schooling differences of their parents, from which we can recover the size of parental spillovers. Conversely, the contribution of learning abilities can be identified from the relationship between the earnings and schooling of children whose *parents* attained the same level of schooling.

Reduced form evidence from the Health and Retirement Survey (HRS) data suggests that children who attain the same years of schooling, but whose mothers have different years of schooling, have parallel earning profiles with a constant gap. The parallel gap points toward the existence of a parental spillover that only affects how much human capital the child accumulates before entering the labor market (schooling), and then remaining constant (controlling for the child's own schooling). Moreover, this gap is similar across different child schooling levels. Both pieces of evidence are consistent with our Ben-Porath model in which parents' human capital only affect children's initial level of human capital.

We then structurally estimate an extended model with non-pecuniary tastes for schooling by indirect inference. By including taste heterogeneity, not only are we able to replicate key moments in the data, but also ensure that we do not overestimate the spillover effect and children's learning abilities on their schooling and earnings outcomes. It turns out that tastes for schooling are important: our estimates suggest that the positive correlation between parents' schooling and children's tastes for schooling is the main determinant of the positive OLS correlation between mom and child's schooling, more important than the positive correlation between parents' schooling and children's abilities.

Our model can also rationalize the insignificant relationship between parents' and children's schooling when using compulsory schooling reforms as an instrument for parents' schooling (Black et al., 2005). Forcing parents to stay in school longer may induce their children to develop a higher taste for schooling, which can countervail the level effect. Our counterfactual results suggest that for our model to be consistent with such IV estimates that find a zero effect of schooling reforms, some of the correlation between mothers' schooling and children's tastes for schooling must be causal.

Finally, we find that forcing all moms to increase their schooling by one year can have a 1.2% causal boost on the lifetime earnings of the child. This effect is heterogeneous across individuals and also over the life-cycle.

Related literature By no means are we the first to estimate the causal effect of parents on children's outcomes. We contribute to this literature by incorporating insights from a human capital model of earnings and education.

Since Solon (1999), a broad literature has studied the causal effect of parents on children's earn-

ings and/or education.[2] The common challenge for all these studies is to separately identify the unobserved correlation between parents' and the children's endowments (abilities, tastes) from the unobserved causal impact of parental spillovers. The typical approach has been to posit a linear relationship between parents' and children's education, and employ special data on twin parents (Behrman and Rosenzweig, 2002), adopted vs. biological children (Plug, 2004) or compulsory schooling reforms during the parents' generation as a natural experiment (Black et al., 2005). In particular, Behrman and Rosenzweig (2002) find a quantitatively large, negative causal effect of mothers' schooling on children's schooling, and conclude that the observed positive intergenerational schooling relationship is entirely driven by correlation between mother's schooling and unobserved factors. Our structural model rationalizes the negative causal effect through the level effect, and we explicitly model unobserved factors as learning abilities and tastes for schooling. Black et al. (2005) find that the causal intergenerational schooling relationship is not significantly different from zero. Through the lenses of our model, if the positive relationship between parents' schooling and children's tastes for schooling is partially causal, forcefully increasing parents' schooling will countervail the level effect, resulting in a IV estimate close to zero.

There is also evidence that points to strong inherited genetic effects on children's schooling outcomes (Behrman and Taubman, 1989; Plug and Vijverberg, 2003). In contrast, other studies find that non-genetic factors have a positive causal effect on children's earnings (Björklund et al., 2006; Sacerdote, 2007). The fact that genetic effects are large for children's schooling but not for their earnings is sometimes viewed as an inconsistency (Black and Devereux, 2011). This stems at least partially from an implicit understanding that parents should have a qualitatively similar effect on the children's education and earnings outcomes, which is likely motivated by the relationship between an individual's education and earnings (Card, 1999).

Instead, we present a non-linear model in which parent's education can have a negative effect on children's schooling but positive effect on earnings, as the empirical evidence suggests. The model allows us to consider children's schooling and earnings outcomes jointly, rather than considering single outcomes separately.[3] By considering two outcome variables, we are able to identify the two unobservables: the cross-sectional correlation between children's abilities and their parents' schooling (selection) and the size of intergenerational spillovers (causal).

The level effect, which we exploit to rationalize the negative causal effect of parents' schooling on children's schooling, is inherent in most Ben-Porath life-cycle models of human capital accumulation (e.g., Heckman et al., 1998; Huggett et al., 2011). Despite inducing less schooling, higher initial human capital leads to higher earnings, so our model is at once consistent with empirical

---

[2]Black and Devereux (2011); Sacerdote (2011) are recent surveys of this literature.

[3]In this sense, our study is related to Bowles and Gintis (2002), which estimates how much of the intergenerational persistence in earnings can be explained by the correlation between fathers' earnings and other variables, such as children's education. However, it has nothing to say about whether the effect of fathers' earnings on children's education and earnings are causal or not.

findings that parents have negative or zero causal effect on children's schooling and a positive effect on their earnings.

Another confounding factor when interpreting how parents affect children's schooling is to what extent the education choice for (or by) the child was economically motivated. Several studies have shown that pecuniary motives alone fall short of explaining education choices (Heckman et al., 2006) and non-pecuniary motives are estimated to be quite large in life-cycle models with schooling choice (Heckman et al., 1998). But as of yet, no attempts have been made to link non-pecuniary motives to the parental effect on children's education and earnings.

In our estimation, we explicitly separate children's pecuniary and non-pecuniary motives for schooling and allow both to be correlated with his parent's schooling, through unobserved learning abilities and tastes for schooling, respectively. Non-pecuniary returns may capture psychological or non-cognitive factors that induce a child to attain more or less education (Oreopoulos et al., 2008; Rege et al., 2011), and/or the fact that children from less advantaged families are more likely to be misinformed about education returns (Betts, 1996; Avery and Turner, 2012; Hoxby and Avery, 2013).[4]

Recent research differentiates how cognitive and non-cognitive skills formed early in life can explain various measures of well-being in adulthood (Cunha et al., 2010), and the childhood environment has long been suspected as what may explain the large estimates for non-pecuniary motives found in structural models of earnings (Bowles et al., 2001; Heckman et al., 2006). The spillover in our model can be understood as the parental effect on cognitive skills that increases the child's earnings ability, while the correlation of a parent's education with her child's taste for schooling can be understood as the parental effect on non-cognitive abilities that do not directly relate to earnings and only schooling. We find that this effect must be at least partially causal in order to reconcile our model with the data.

The rest of the paper is organized as follows. Section 2 posits a model of human capital accumulation with spillovers, and we analytically solve a simple version of the model from which we can make empirical predictions. In section 3 we describe the HRS data that we use and interpret the reduced form evidence through the lenses of our model. Section 4 presents how we structurally estimated the model to the HRS. We also show, quantitatively, that the estimated model inherits properties of the simpler version. Section 5 examines the main result of increasing mother's schooling by 1 year, and also the counterfactual result of a hypothetical compulsory

---

[4]In our context, we simply lump such factors together into heterogeneous tastes for schooling. The existence of non-pecuniary motives presents difficulties when estimating causal effects, but also provides some discipline on how to interpret results from special data sets. For example, the schooling difference between twin parents or compulsory increases in years of schooling are less likely to be related to other unobserved family characteristics that affect children's schooling decisions, while children adopted to different families very likely do develop the non-cognitive skills or acquire information that conform to such unobservables. If these unobservables are positively correlated with parents' education, to some extent it should be expected that the effect of a parents' education on children's education should be smaller in twins or IV studies (although the former is also subject to sampling bias) than adoptee studies.

schooling reform. Section 6 concludes.

2   Schooling and Earnings Model with Parental Spillovers

Our model is a variant of the Ben-Porath (1967) life-cycle model of human capital accumulation. An individual begins life at age 6 and retires at an exogenous age R. His initial state contains his parent's human capital h_{Pi}, his learning ability exogenously transmitted across generations z_i, and his tastes for schooling $\tilde{\xi}_i(s)$. The latter is an individual-specific function of years of schooling s that represents individual i's non-pecuniary benefit for obtaining s years of schooling in present value terms. Credit markets are complete and we can thus focus on the income-maximisation problem.[5] An individual at age 6 solves[6]

$$W(h_{Pi}, z_i, \tilde{\xi}_i(s)) = \max_s \left\{ \tilde{W}\left(h_{Pi}, z_i; s\right) + \tilde{\xi}_i(s) \right\} = \tilde{W}(h_{Pi}, z_i; S_i) + \tilde{\xi}_i(S_i), \tag{1a}$$

where S_i denotes individual i's optimal choice for schooling, and $\tilde{W}(\cdot)$ is the present discounted value of net income from attaining s years of schooling :

$$\tilde{W}(h_{Pi}, z_i; s) = \max_{\{n(a), m(a)\}} \left\{ - \int_6^{6+s} e^{-r(a-6)} m(a) da + \int_{6+s}^R e^{-r(a-6)} w h(a) \left[1 - n(a)\right] da \right\} \tag{1b}$$

subject to

$$\dot{h}(a) = \begin{cases} z_i h(a)^{\alpha_1} m(a)^{\alpha_2}, & \text{for } a \in [6, s), \\ z_i [n(a) h(a)]^{\alpha_{1W}} m(a)^{\alpha_{2W}}, & \text{for } a \in [s, R), \end{cases} \tag{1c}$$

$$n(a) \in [0, 1], \quad m(a) > 0, \tag{1d}$$

$$h(6) = h_{i0} = b z_i^\lambda h_{Pi}^\nu. \tag{1e}$$

The exponents (α_1, α_{1W}) and (α_2, α_{2W}) are the returns to time and good investments, $[n(a), m(a)]$, into human capital at age a, $h(a)$. We allow these returns to potentially vary between the schooling and working phases (before and after s). Tastes for schooling only affect an individual's desire to remain (or not) in school while having no direct effect on earnings.

The causal effect from parents' education on children's earnings comes from (1e): the child's initial level of human capital at age 6 is a function of his parent's human capital, h_{Pi}.[7] This can

---

[5]This is reasonable approximation given that our data is from the 1931-1941 cohort (Belley and Lochner, 2007).

[6]A young child making decisions can easily be recast as an altruistic parents making decisions for him.

[7]This assumes that parents' human capital is measured when all parents are of the same age. One could think of an overlapping generations setup in which parents make their own human capital investment decisions while simultaneously making decisions for a young child. Since we take parents' education directly from the data, we choose a parsimonious representation in which it enters only as a state in the child's optimization problem. We are abstracting

be understood as investment in children prior to age 6 depending on the parent's economic status through the parameter ν. Clearly, initial human capital h_{i0} may also be affected by one's own learning ability z_i as well, the extent of which is measured by λ.

2.1 Initial Conditions

In what follows we drop the individual subscript i unless necessary. The population distribution of $(\log h_P, \log z)$ is joint lognormal,

$$\begin{bmatrix} \log h_P \\ \log z \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu_{h_P} \\ \mu_z \end{bmatrix} , \begin{bmatrix} \sigma_{h_P}^2 & \rho_{zh_P}\sigma_h\sigma_z \\ \rho_{zh_P}\sigma_h\sigma_z & \sigma_z^2 \end{bmatrix} \right). \tag{2}$$

where (μ_{h_P}, μ_z) are the population means and (σ_{h_P}, σ_z) the corresponding standard deviations of (h_P, z). The correlation between the two variables may not equal zero $\rho_{zh_P} \neq 0$ because of the exogenous transmission of learning abilities and because the parent's ability is a input in the production of her human capital h_P. Consequently, z and h_P are both functions of the parent's learning ability.

Now suppose we try to measure the relationship between parents and children's initial level of human capital from data on (h_0, h_P, z):

$$\frac{d \log h_0}{d \log h_P} = \underbrace{\frac{\partial \log h_0}{\partial \log h_P}}_{=\nu} + \underbrace{\frac{d \log z}{d \log h_P}}_{\equiv \Xi(\rho_{zh_P})} \cdot \underbrace{\frac{\partial \log h_0}{\partial \log z}}_{=\lambda}.$$

The causal effect from parents' to children's human capital is captured by the parameter ν, and we define the spillover as *the increase in earnings caused by this transmission*. The term $\Xi(\rho_{zh_P})$ is non-zero if and only if ρ_{zh_P} is non-zero, which we already argued above captures only selection. Lastly, λ is also selection, as it captures how much the child's ability, which is exogenously transmitted from the parent's ability, directly affects early human capital formation.[8]

In the estimation, we proxy h_P by mothers' education, so it is observed. But since abilities are unobserved, we face the problem of separately identifying ν, λ and ρ_{zh_P}, i.e., how much a child's schooling or earnings are correlated with the parent's because of h_P directly affecting h_0,

---

away from the fact that children may consider how their human capital investment decisions affect their future children. Such a setup would requires micro-data for an additional generation and further complicate identification (how to separate altruism from spillovers, for example). We chose a simpler model that we identify with HRS data only. Moreover, as long as individuals maximize income that can be passed on to future children, without explicitly taking into account the spillover effect, life-cycle decisions would remain the same under perfect credit markets.

[8]We are assuming away a causal effect that parents' human capital can have on their children's learning abilities, but if this were the case, λ would pick up a third channel: How much of z is directly affected by h_P before age 6. We are also assuming away a causal effect that h_P can have on z after age 6, but such an effect is not separately identified from ρ_{zh_P}. In both cases, however, our estimate would deliver a lower bound of the spillover, which still turns out to be significantly positive. We discuss this more in Sections 2.3 and 4.3.

or indirectly through h_P's correlation with z. Simlilarly, we need to identify the impact of h_P on the children's tastes for schooling $\tilde{\xi}(s)$. For now, we abstract from tastes for schooling and only include them later in Section 4. In the remainder of this section, we show how the solution to the model without tastes for schooling gives us insights into the precise mechanisms at work, and how $(\nu, \lambda, \rho_{zh_P})$ can be recovered from a panel of individuals' schooling and earnings, and their parents' schooling.

2.2 Model Solution Without Tastes for Schooling

For now we ignore tastes for schooling by assuming $\tilde{\xi}(s) = 0$, and solve the income maximization problem (1b) subject to the constraints while also imposing $\alpha_{1W} = \alpha_1$ and $\alpha_{2W} = \alpha_2$. This continuous time deterministic control problem is a direct application of Ben-Porath (1967). The individual maximizes the present discounted value of net income at age 6, where the state is h and controls are (n, m). The terminal time is fixed at R but the terminal state $h(R)$ must be chosen, as well as an optimal stopping time S. Since the objective function is linear, the constraint set strictly convex, and the law of motion strictly positive and concave (as long as $\alpha_1 + \alpha_2 < 1$), the optimization problem is well-defined and the solution is unique (Léonard and Van Long, 1992). The Hamilton-Jacobi-Bellman (HJB) equation is

$$rV(a,h) - \frac{\partial V(a,h)}{\partial a} = \max_{n,m} \left\{ wh(1-n) - m + \frac{\partial V(a,h)}{\partial h} \cdot z(nh)^{\alpha_1} m^{\alpha_2} \right\}.$$

where $V(a,h)$ denotes the present discounted value of net income at age a. As usual, the HJB equation can be interpreted as a no-arbitrage condition. The left-hand side is the instantaneous cost of holding a human capital level of h at age a, while the the right-hand side is the instantaneous return. The first order conditions for the controls are

$$whn \leq \alpha_1 z(nh)^{\alpha_1} m^{\alpha_2} \cdot V_h, \qquad \text{with equality if } n < 1 \tag{3}$$

$$m = \alpha_2 z(nh)^{\alpha_1} m^{\alpha_2} \cdot V_h, \tag{4}$$

where V_h is the partial of $V(a,h)$ with respect to h. These conditions simply state that the marginal cost of investment is equal to the marginal return. The envelope condition gives (at the optimum)

$$r \cdot V_h - V_{ah} = w(1-n) + \frac{\alpha_1 z(nh)^{\alpha_1} m^{\alpha_2}}{h} \cdot V_h + z(nh)^{\alpha_1} m^{\alpha_2} \cdot V_{hh}, \tag{5}$$

where V_{xh} is the partial of V_h with respect to $x \in \{a, h\}$. This "Euler equation" states that at the optimum, the marginal cost of increasing human capital must be equal the marginal return. Equations (3)-(5) along with the law of motion (1c), initial condition (1e) and terminal condition $V_h = 0$—the appropriate transversality condition for a fixed terminal time problem—characterize

8

the complete solution. Here we only present the important results; all proofs are relegated to Appendix A. To save on notation, it is useful to define $\alpha \equiv \alpha_1 + \alpha_2$ and

$$q(a) \equiv \left[1 - e^{-r(R-a)}\right], \qquad \kappa \equiv \frac{\alpha_1^{\alpha_1} \alpha_2^{\alpha_2} w^{1-\alpha_1}}{r}.$$

PROPOSITION 1: OPTIMAL SCHOOLING CHOICE *Define $\alpha \equiv \alpha_1 + \alpha_2$ and*

$$F(s)^{-1} \equiv \kappa \left(\frac{\alpha_1}{w}\right)^{1-\alpha} \cdot \left[1 - \frac{(1-\alpha_1)(1-\alpha_2)}{\alpha_1 \alpha_2} \cdot \frac{1 - e^{-\frac{\alpha_2 rs}{1-\alpha_2}}}{q(6+s)}\right]^{\frac{1-\alpha}{1-\alpha_1}} \cdot q(6+s),$$

which is only a function of prices and the parameters (α_1, α_2). The optimal choice of schooling S is uniquely determined by

$$F'(S) > 0, \qquad F(S) \geq z/h_0^{1-\alpha} = z^{1-\lambda(1-\alpha)} h_P^{-\nu(1-\alpha)} \quad \text{with equality if } S > 0. \tag{6}$$

Proof. See Appendix A. □

The higher the learning ability z of an individual, the higher the optimal choice of his schooling (as long as $\lambda(1-\alpha) < 1$). Intuitively, conditional on an initial level of human capital, higher z individuals benefit more from schooling. But the causal effect from the parent's human capital h_P on schooling is *negative* (as long as $\nu(1-\alpha) > 0$), in line with the empirical evidence in Behrman and Rosenzweig (2002). In our model, this level effect—substitution between the initial level of human capital and length of schooling—arises as long as lifetimes are finite and the returns to human capital investments displays decreasing returns ($\alpha < 1$): individuals with high initial values face low returns to schooling earlier, reducing their incentives to stay in school and instead enter the labor market earlier.

Using Proposition 1, we can relate children's schooling to the parents'. For the parent generation, we assume a Mincerian representation between their schooling S_P and earnings (the parent's human capital h_P):

$$h_P = \exp(\beta S_P) \quad \Leftrightarrow \quad \log h_P = \beta S_P. \tag{7}$$

where β is the returns to schooling of the parent's generation. Since (h_{Pi}, z_i) are assumed to be distributed joint lognormal (Equation 2), we can write the statistical relationship

$$\log z_i = \mu_z + \epsilon_{zi}, \quad \text{where} \quad \epsilon_{zi} \sim \mathcal{N}(0, \sigma_z) \quad \text{and} \tag{8a}$$

$$\epsilon_{zi} | \log h_{Pi} = \rho_{zh_P} \cdot \frac{\sigma_z}{\sigma_{h_P}} \cdot (\log h_{Pi} - \mu_{h_P}) + \epsilon_{Pi} \equiv \tilde{\mu}_z + \tilde{\rho}_{zh_P} S_{Pi} + \epsilon_{Pi}, \tag{8b}$$

where $\tilde{\rho}_{zh_P} = \beta \cdot \rho_{zh_P}\sigma_z/\sigma_{h_P}$ measures selection and ϵ_{Pi} is distributed normal with mean zero. That is, ϵ_{zi} is correlated with h_{Pi} but with no causal relationship. Plugging this into (6) at equality yields[9]

$$\log F(S_i) = [1 - \lambda(1 - \alpha)] \log z_i - \nu(1 - \alpha)\beta S_{Pi}$$
$$= [\tilde{\rho}_{zh_P} - (\beta\nu + \lambda\tilde{\rho}_{zh_P})(1 - \alpha)] \cdot S_{Pi} + [1 - \lambda(1 - \alpha)](\tilde{\mu}_z + \epsilon_{Pi}). \tag{9}$$

This shows that the correlation in schooling across generations is affected both by selection and spillovers. In particular, while the correlation is positively related to the selection parameter $\tilde{\rho}_{zh_P}$ (as long as $\lambda(1 - \alpha) < 1$), it is negatively related to the spillover parameter ν.

PROPOSITION 2: POST-SCHOOLING HUMAN CAPITAL AND EARNING PROFILES *For an individual who attains S years of schooling, human capital at the end of schooling is*

$$h_S = C_1(S) \cdot z^{\frac{1}{1-\alpha}}/w$$

where $C_1(s) \equiv \alpha_1 \cdot [\kappa q(6 + s)]^{\frac{1}{1-\alpha}}$. *Earnings at age* $a \in [6 + S, R)$ *is*

$$e(a) = wh(a)[1 - n(a)] = [C_1(S) + C_2(a; S)] \cdot z^{\frac{1}{1-\alpha}}, \tag{10}$$

where $C_2(a; s) = \kappa^{\frac{1}{1-\alpha}} \cdot \left\{ r \cdot \int_{6+s}^{a} q(x)^{\frac{\alpha}{1-\alpha}} dx - \alpha_1 q(a) \right\}$.

Proof. See Appendix A. □

The above proposition tells us that, once the length of schooling is known, the human capital level of a child is affected only by his own learning ability z. His initial stock of human capital, h_0, has no effect on the *amount* of human capital accumulated (quality) except through the *length* of schooling (quantity), S. So both the causal effect of ν and early childhood learning λ are subsumed in the length of schooling.

Equation (10) implies that conditional on ability z, (C_1, C_2) govern the intercept and slope of individuals' (log-)age-earnings profiles by schooling, respectively. According to the model, these slopes must be close to parallel across different z's, which is standard in models that use a Ben-Porath approach. This equation can also be interpreted as a Mincer equation that relates earnings to schooling. The functions C_1 and C_2—both of which are close to exponential—determine the returns to schooling and potential experience, respectively, and their shapes only depend on exogenous prices and the exponents of the human capital production function (α_1, α_2). So from a balanced sample of individual earnings profiles, the intercept and slopes of the profiles would

---

[9]In what follows, a capital S denotes the optimal schooling choice, which is assumed to be what we observe in the data.

identify (α_1, α_2) if we could control for unobserved heterogeneity in z.

2.3 Identification of Key Parameters

A robust finding in empirical studies is that even after controlling for observables, mothers' education has a statistically significant relationship with children's schooling and earnings. The parameters $(\tilde{\rho}_{zh_P}, \lambda)$ are structural representations of the selection effects and ν the causal effect. Here we demonstrate how to identify $\tilde{\rho}_{zh_P}$ and ν given a data sample of children's schooling and earnings outcomes, and the schooling levels of their parents. Identification of λ is relegated to Appendix C as the selection effect coming from λ is only secondary and its point estimate is close to zero.

For what follows, suppose we observe a random sample of individuals whose schooling levels of their parents and themselves, (S_{Pi}, S_i), and potential experience $x_i \equiv a_i - 6 - S_i$, are observed without error. Their observed log earnings, $\log e_{i,x}^*$, is the sum of their log earnings $\log e_{i,x}$ implied by the model and an i.i.d. measurement error term $u_{i,x}$.

COROLLARY 1: IDENTIFYING ρ_{zh_P}**.** *The observations* $(e_{i,x}^*, S_i, S_{Pi})$ *satisfy the regression equation*

$$\log e_{i,x}^* = b_0 + b_1 \log \left[C_1(S_i) + \tilde{C}_2(x; S_i) \right] + b_2 S_{Pi} + v_{i,x}, \tag{11}$$

where the regression error $v_{i,x}$ *is independent of the regressors,* $\tilde{C}_2(x, s) \equiv C_2(x + 6 + s, s)$*, and*

$$b_2 = \tilde{\rho}_{zh_P} / (1 - \alpha).$$

Proof. For a random sample, the cross-sectional relationship between (z_i, S_{Pi}) satisfies

$$\log z_i = (\mu_z + \tilde{\mu}_z) + \tilde{\rho}_{zh_P} S_{Pi} + \epsilon_{Pi}$$

where ϵ_{Pi} is normally distributed with mean zero. Applying this to (10) in Proposition 2 yields

$$\log e_{i,x} = \log \left[C_1(S_i) + \tilde{C}_2(x; S_i) \right] + \frac{1}{1 - \alpha} \left[(\mu_z + \tilde{\mu}_z) + \tilde{\rho}_{zh_P} S_{Pi} + \epsilon_{Pi} \right]$$

It follows the regression error $v_{i,x} = u_{i,x} + \epsilon_{Pi} / (1 - \alpha)$ and $b_2 = \tilde{\rho}_{zh_P} / (1 - \alpha)$. □

A Mincer regression that controls for an individual's schooling and potential experience (which correspond to $(C_1, \tilde{C}_2)$, respectively) reveals that the coefficient on parents' schooling S_P captures only $\tilde{\rho}_{zh_P}$ and completely misses ν. This also implies that once we control for own schooling and parents' schooling, average (log-)experience-earnings profiles should be close to parallel, since $\tilde{C}_2$, which controls the slope of the profiles, does not vary across individuals. Furthermore, the gaps between the profiles of individuals with different levels of schooling are determined by C_1, so

(α_1, α_2) are identified.[10]

That a simple regression of log earnings on a function of schooling and potential experience, and parents' schooling S_P completely reveals ability selection may be somewhat surprising. The intuition is the same as in Proposition 2: once we control for own schooling S, there is no longer any role for the causal spillover effect, which stems from the child's initial level of human capital.

We will see in Corollary 2, however, that if run separate regressions on subsamples of children with the exact same years of schooling, the coefficient on parents' schooling would identify ν instead. According to our model then, the difficulty in separating the two effects stems from how we treat own schooling S.

COROLLARY 2: IDENTIFYING SPILLOVERS. *If we select a subsample of individuals with the same level of schooling, $S_i = \hat{S}$ and positive experience level $x > 0$, the observations $(e_{i,x}^*, S_{Pi})$ satisfy the regression equation*

$$\log e_{i,x}^*(S_i = \hat{S}) = \bar{b}_{0,x}(S_i = \hat{S}) + \bar{b}_2 S_{Pi} + u_{i,x}, \tag{12}$$

with

$$\bar{b}_2 = \beta \cdot \nu / \left[1 - \lambda(1 - \alpha)\right].$$

Proof. Proposition 1 implies that, for a subsample of children such that $S_i = \hat{S}$, it must hold across the subsample that

$$z_i = F(\hat{S})^{\frac{1}{1-\lambda(1-\alpha)}} \cdot h_{Pi}^{\frac{\nu(1-\alpha)}{1-\lambda(1-\alpha)}} = F(\hat{S})^{\frac{1}{1-\lambda(1-\alpha)}} \cdot \exp(\beta S_{Pi})^{\frac{\nu(1-\alpha)}{1-\lambda(1-\alpha)}}$$

where the equality follows from (7). Applying this and $S_i = \hat{S}$ in (10) yields

$$\log e_{i,x} = \underbrace{\log\left[C_1(\hat{S}) + \tilde{C}_2(x; \hat{S})\right] + \frac{\log F(\hat{S})}{(1-\alpha)[1-\lambda(1-\alpha)]}}_{\bar{b}_{0,x}(S_i = \hat{S})} + \frac{\beta \cdot \nu}{1 - \lambda(1-\alpha)} S_{P,i},$$

from which (12) follows. □

Corollary 2 shows that *for a subsample of children* with the same years of schooling, the difference in their parents' schooling manifests itself as constant gaps across parallel (log-)earnings profiles. Corollary 3 in Appendix C shows how λ can be identified; Then since $\alpha = \alpha_1 + \alpha_2$ is known,

---

[10]Recall that the shapes of $C_1(s), C_2(a;s)$ only depend on prices and (α_1, α_2). Because $C_1(s)$ is exponential in s, $\log C_1(s)$ is close to linear, meaning these gaps should be even across different schooling levels. So controlling for age or potential experience interacted with schooling levels can capture the shape of the function $\log[C_1(s) + C_2(a;s)]$, because $C_2(a;s)$ is also close to log-linear. We show in the next section that this is approximately true in the data.

and β is easily recovered from a Mincer regression on the parents' generation (Equation 7), the magnitude of these gaps identifies ν.

Why does the coefficient on parents' schooling in (11) reveal selection while in (12) it reveals spillovers? In the former, we are measuring the effect of parents on children's residual earnings, and the effect of own schooling on earnings only comes through the functions $(C_1, \tilde{C}_2)$. Importantly, it does *not* control for the effect that parents have on schooling. Since we are looking across the entire population, and since parents have no direct role on earnings once schooling is controlled for, the regression only reveals selection (correlation between parental human capital and abilities). The effect of schooling in this case is understood as the average effect of schooling on earnings across all children, regardless of parental background.

In the latter, we are selecting a subsample of children with the exactly same level of schooling. So obviously there is no role for own schooling to explain earnings differences, but more importantly, the distribution of abilities differ from the population distribution. Regressing earnings on parents' schooling does not reveal the population (h_P, z) correlation, but the structural relationship between the two *within* each schooling subgroup. And by means of (6)-(7), we know exactly what this relationship is in terms of (ν, λ), i.e., we know how ability varies as a function of parents' schooling. So once λ is known, ν is revealed. By looking at the *within-schooling group variation*, earnings differences reveal the effect that parents have on schooling.[11]

This intuition for why (11) reveals selection while (12) reveals spillovers is visualized in Appendix F, Figure 5. The parental spillover is understood as a child of a more educated parent getting more out of the same level of schooling (i.e., the quality of schooling). Then given a sample of children with the same level of schooling, we recover the underlying relationship between parents schooling and children's abilities.

We believe our identification scheme is more general than our model. The main assumption is that children's schooling is a function of their abilities and parents' schooling, and that earnings is a function of their abilities and their *own* schooling. That is, while parents directly affect children's schooling, they do not directly influence children's earnings. The effect is only indirect through how children's earnings are affected by children's schooling.[12]

In the next section, we run the proposed regressions in Corollaries 1-2 using HRS data, in which earnings profiles are indeed parallel with gaps determined by own and parents' schooling, and present some raw evidence on the relative magnitudes of $\tilde{\rho}_{zh_P}$ and ν. Then in Section 4, we estimate a model that includes unobserved heterogeneity in tastes for schooling.

---

[11]The model also implies that the within-subgroup relationships should be the same across groups, which we empirically verify in the next section.

[12]If the true data-generating process *did* include a direct effect, our estimated effect would be a lower-bound for the spillover.

3 Data Analysis

The Health and Retirement Study (HRS) is sponsored by the National Institute of Aging and conducted by the University of Michigan with supplemental support from the Social Security Administration. It is a national panel study with an initial sample (in 1992) of 12,652 persons in 7,702 households that over-samples blacks, Hispanics, and residents of Florida. The sample is nationally representative of the American population 50 years old and above. The baseline 1992 sample that we use for our study consisted of in-home, face-to-face interviews of the 1931-41 birth cohort and their spouses, if they were married. Follow up interviews have continued every two years after 1992. As the HRS has matured, new cohorts have been added.

The HRS is usually used to study elderly Americans close to or in retirement, but there are several features that make it suitable for us. First, these older individuals and their parents were less affected by compulsory schooling regulations and other government interventions, and more than half never advanced to college. This makes the sample suitable for a model such as ours in which a large schooling variation is important for identifying causal effects.[13] Second, the education premium was quite stable prior to the 1980s, so it is unlikely for these cohorts to have been surprised by an unexpected rise in education returns. It is also less likely that the effect of education on earnings outcomes or the reverse effect of expected earnings on education choices changes much from cohort to cohort. Third, the HRS contains information on own schooling, schooling of both parents, and can be augmented with restricted Social Security earnings data through which we observe entire life-cycle earnings histories.[14]

3.1 Descriptive Statistics

For the purposes of our study, we keep 5,760 male respondents born between 1924 and 1941 from the 1992 sample.[15] We further drop 646 individuals with missing information on their own education or mother's years of schooling. This leaves us with 5,114 individuals. Table 1 describes this sample by level of education. Children and mom's schooling are about 12.3 and 9.2 years, respectively, both with a standard deviation of approximately 3.5 years.

A large fraction of HRS respondents allowed researchers restricted access to their Social Secu-

---

[13]Indeed, the HRS displays much more schooling variation than found in other datasets. Many papers studying intergenerational schooling relationships, such as those cited in the introduction, also focus on earlier periods, but lack life-cycle earnings information (except for Black et al. (2005), which uses Norwegian administrative data, although they do not use that information in their analysis).

[14]One limitation of the HRS is that information on parents is limited to their education. But in most recent datasets with richer information on parents and family background, we only observe the beginning of children's age-earnings profiles, which is also very noisy because the average age for labor market entry is increasing. Few datasets do have long earnings histories as well as richer parental background information, such as NLSY79, but these individuals face compulsory schooling regulations and a rising college premium.

[15]Most women from this sample have only very short earnings histories. Although the initial HRS sample selected individuals from the 1931-1941 cohort, of course many of their spouses were born in different years.

	HSD	HSG	SMC	CLG	Total
Schooling	7.97	12.00	13.88	16.54	12.33
	(2.67)		(0.68)	(0.50)	(3.41)
Mom's Schooling	7.00	9.26	10.11	11.07	9.24
	(3.70)	(2.98)	(3.11)	(3.27)	(3.60)
Dad's Schooling	6.41	8.72	9.88	11.02	8.91
	(3.70)	(3.36)	(3.59)	(3.77)	(3.96)
% White	73.31	84.70	85.00	89.30	82.81
% Black	21.94	13.24	12.23	6.62	13.82
% Hispanic	20.36	4.98	5.86	2.72	8.67
Earnings 23-27	16.69	23.39	22.20	18.34	20.24
	(12.23)	(13.23)	(12.77)	(12.37)	(13.00)
Earnings 28-32	24.96	34.78	34.13	33.42	31.76
	(16.36)	(17.92)	(18.60)	(19.56)	(18.50)
Earnings 33-37	30.90	41.46	41.68	44.15	39.33
	(18.78)	(21.04)	(21.72)	(23.64)	(21.85)
Earnings 37-42	34.38	44.27	45.92	52.17	43.79
	(21.21)	(24.16)	(28.16)	(31.88)	(26.96)
# Obs	1349	1647	940	1178	5114

*HSD<12, HSG=12, 12<SMC<16, CLG=16+ years of schooling
**Years of schooling top-coded at 17.
** Standard deviations in parentheses.
***Earnings inflated to 2008, measured in $1000.

Table 1: Summary Statistics by Education

rity earnings records. Combined with self-reported earnings in the HRS, these earnings records provide almost the entire history of earnings for most of the HRS respondents. Some records were top-coded, which we impute assuming the following individual log-earnings process[16]

$$\log e_{i,0}^* = X_{i,t}'\beta_0 + \varepsilon_{i,0}$$
$$\log e_{i,t}^* = \rho \log e_{i,t-1}^* + X_{i,t}'\beta_x + \varepsilon_{i,t}, \qquad t \in \{1, 2, ..., T\}$$
$$\varepsilon_{i,t} = \alpha_i + u_{i,t}$$

where $e_{i,t}^*$ is the latent earnings of individual i at time t in 2008 dollars, $X_{i,t}$ is the vector of characteristics at time t, and the error term $\varepsilon_{i,t}$ includes an individual specific component α_i, which is constant over time, and an unanticipated white noise component $u_{i,t}$. We employed random-effect assumptions with homoskedastic errors to estimate the above model separately for men with and without a college degree. Scholz et al. (2006) gives details of the above earnings model, the procedure used to impute top-coded earnings, and the resulting coefficient estimates.

---

[16] Social security earnings records exceeding the maximum level subject to social security taxes were top-coded in the years 1951 through 1977.
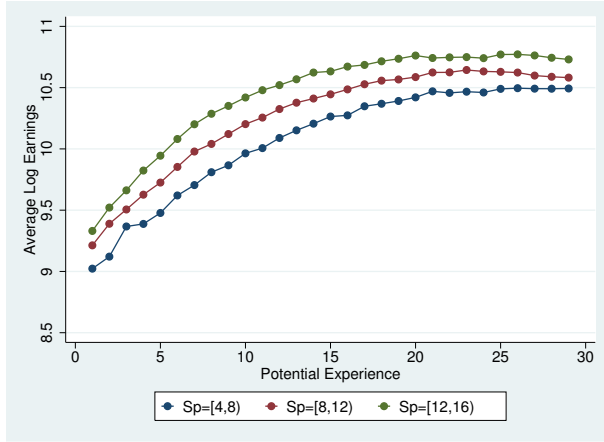
(a) Children of mother's with [4,8), [8,12), and [12,16) years of schooling.
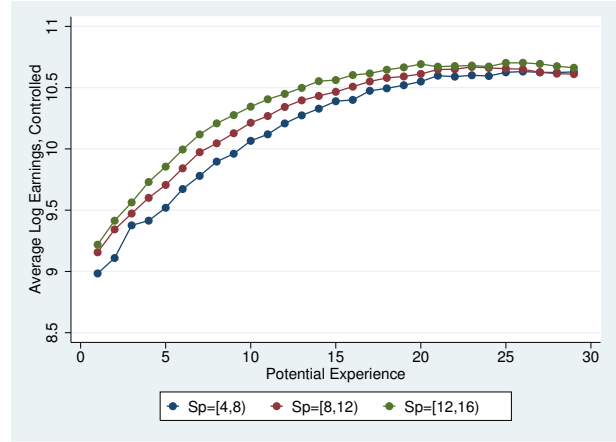
(b) Children of mother's with [4,8), [8,12), and [12,16) years of schooling, controlling for own schooling.

Figure 1: Identifying Selection
Earnings profiles of children of different schooling levels by mother's schooling level: 1924-1941 birth cohort. The y-axis is average log annual earnings in 2008 USD. Mothers' schooling levels are divided by 8 years or below, and more than 8 years.

3.2 Spillovers and Selection

According to our model, the spillover is subsumed in the choice of schooling (Proposition 1). Selection on abilities is revealed as the Mincer coefficient on parents in a regression over the population (Corollary 1) and spillovers by the same coefficient but in a regression over subsamples of individuals with identical levels of schooling (Corollary 2). Moreover, the model predicts that for both regressions, earnings profiles should be parallel with a constant gap.

In Figure 1(a), we divide individuals into 3 subsamples, depending on whether their mothers attain $S_{Pi} \in [4,8)$, $[8,12)$, or $[12,16)$ years of schooling. These correspond approximately to primary, secondary, and tertiary education. Average earnings profiles are parallel by mother's schooling, as predicted by the model (the function $\tilde{C}_2$).[17] In Figure 1(b), we take out a mean effect of own schooling on own earnings for each level of own schooling S. In terms of the model, it is akin to controlling for the effect that comes through $C_1(S)$ for all levels of S. Since schooling is correlated across generations, the gaps narrow. The remaining gap is the selection effect: the fact that higher human capital parents have higher ability children. These gaps are of similar magnitude across different S_P, justifying our parametric assumption of log-normally distributed abilities.

To visualize spillovers, we need to divide individuals into different levels of own schooling, rather than control for an average schooling effect that applies equally to everyone. So we split individuals into 4 subsamples depending on whether they have $S_i \in [8,12)$, exactly 12, $(12,16]$, or

---

[17]For robustness, we have tried dividing children and their mothers according to different levels of education. Experience-earnings are parallel except when we split mother's education categories into very fine levels with few observations. We also confirmed that this evidence is present in available data from the NLSY79 and PSID.

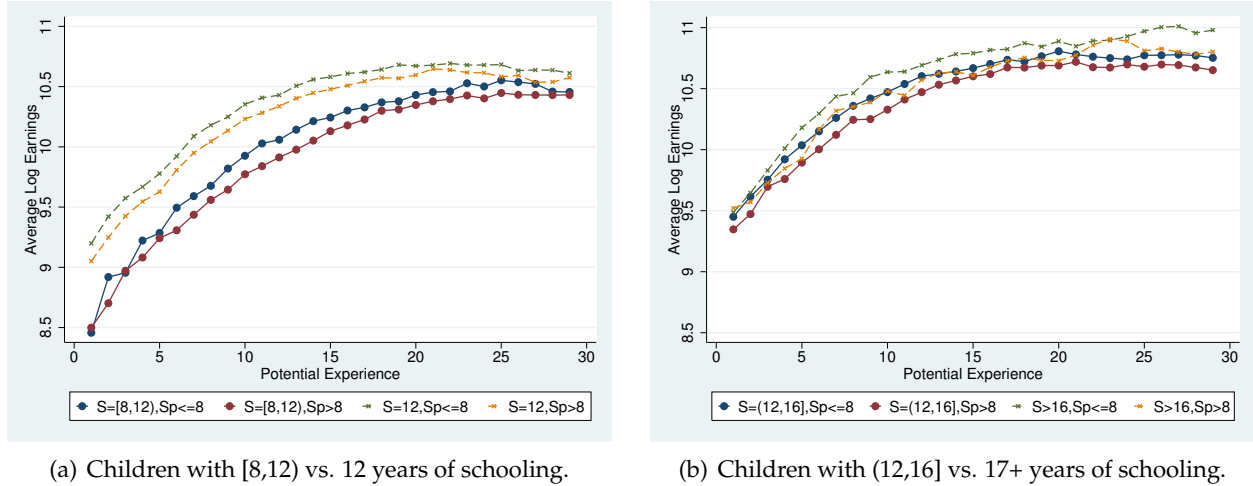(a) Children with [8,12) vs. 12 years of schooling. (b) Children with (12,16] vs. 17+ years of schooling.

Figure 2: Identifying Spillovers

Earnings profiles of children of different schooling levels by mother's schooling level: 1924-1941 birth cohort. The y-axis is average log annual earnings in 2008 USD. Mothers' schooling levels are divided by 8 years or below, and more than 8 years.

more than 16 years of schooling. Each group is further divided according to whether the mother has more than 8 years of schooling, which is about half of the mothers in our sample.[18] Figure 2 depicts the average log-earnings profiles for each subsample. The left panel compares individuals with $[8, 12)$ to those with exactly 12 years of schooling, and the right panel individuals with $(12, 16]$ to those with 17 or more years of schooling, respectively.

For all 4 education levels, the average log-earnings profiles of children with the same schooling but different mother's schooling are nearly parallel with a constant gap. This points to a permanent level effect that persists throughout an individual's career, which is precisely what the spillover parameter ν is intended to capture.[19] Moreover, the gaps are similar across the three categories of children's education, as implied by our model . Next, we compute the reduced-form magnitudes of the selection and spillover effects presented in Figures 1-2.

3.3 Mincer Regressions

The figures suggest the following simple Mincer regression:

$$\log e_{i,x} = \beta_0 + \beta_1 S_i + \beta_2 S_{Pi} + f(x_i) + \epsilon_{i,x} \tag{13}$$

where $e_{i,x}$ is the earnings of individual i with potential experience $x_i = \text{age} - 6 - S_i$, $f(\cdot)$ a function of x which we specify in various different ways below, and $\epsilon_{i,x}$ an error term. We estimate different

---

[18]And corresponds to the end of junior high in most states in 1900 U.S..

[19]To be precise, the gap should capture $\hat{b}_2 = \beta \cdot \nu / [1 - \lambda(1 - \alpha)]$ in (12).

17

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
S	0.090	0.082	0.083	0.080	0.080	0.076	0.097	0.095	
	(99.73)	(84.21)	(82.29)	(78.14)	(78.14)	(81.02)	(35.24)	(33.10)	
Mom S_P		0.017			0.015	0.017	0.018		0.016
		(22.71)			(15.60)	(22.44)	(22.92)		(21.38)
Dad S_P			0.011		0.003				
			(15.39)		(3.45)				
Mom + Dad				0.009					
				(20.64)					
EXP $\times$ S							-0.001	-0.001	
							(-6.09)	(-4.85)	
EXP $\times$ S_P								0.000	
								(-2.99)	
R^2	0.190	0.194	0.193	0.195	0.195	0.190	0.195	0.195	0.196

Table 2: Mincer Regressions

OLS regressions of (log) earnings on own and parents' years of schooling. HRS initial cohort, 5,114 individuals males born 1924-1941, ages 23-42. All columns include a full set of dummies for each level of potential experience (age-6-S), except for (6), which includes only a linear and quadratic term instead. t-stats shown in parentheses.

versions of (13) for earnings data from ages 23 to 42, and tabulate the results in Table 2.[20,21]

We consider four measures for S_{Pi}: the mother's and father's years of schooling, respectively, their sum, and also including both as separate controls. Our theory does not speak to which of these measures is more appropriate. However, many studies suggest that mothers have a larger influence on children (Behrman and Rosenzweig, 2002; Del Boca et al., 2012), which is true in our sample as shown in Table 2.

The first specification (1) is a standard Mincer regression with dummies for each potential experience level observed in the data. The return to schooling is estimated to be 9%. This is in the lower range of the estimated returns to schooling for more recent cohorts, which is in line with the increased return to education over the last century (Goldin and Katz, 2007). The returns slightly decrease to 8.2% when we include mothers' years of schooling in (2). The coefficient on mothers' education is 1.7% and statistically significant. This suggests that being born to a mother with five additional years of schooling has about the same effect on earnings as would an additional year of own schooling.

The coefficient on own schooling remains similar when we measure parents' human capital

---

[20]Although the estimates barely change even if we use all the available earnings records, we restrict ourselves to this age interval because this is what we estimate the model to in the next section. We begin at age 23 because many earnings records are missing prior to this age, especially for individuals choosing higher levels of schooling. We end at age 42 because it is close to the peak of the earnings profiles for most individuals, and our model has nothing to say about the irregular labor supply or retirement behavior at older ages.

[21]Including race or cohort dummies barely affect our estimates. In our structural estimation to follow, such effects should be absorbed in the unobserved heterogeneity in tastes for schooling, so controlling for them here would make the reduced form estimates incomparable with the structural estimates.

by fathers' years of schooling in (3), but the parent's coefficient is attenuated to 1.1% but still statistically significant. The parent's coefficient drops further when we measure parents' human capital as the sum of the schooling of both parents in (4). If we include both separately as in (5), mothers' education is very slightly reduced from 1.7% to 1.5% while the coefficient on fathers drop significantly from 1.1% to 0.3%. This implies that moms' schooling has dominant explanatory power, which we take as the proxy for parent's human capital in what follows.

In column (6) we replaced the experience dummies with a linear and quadratic term in experience. The results are similar to column (2), where we had a full set of dummies for experience. In columns (7) and (8) we added two interactions terms to experience dummies: one between own education and experience and another between mother's years of schooling and experience. Although the Mincer coefficient on S is as much as 2 percentage points higher than other specifications, the coefficients on the interactions are small and/or insignificant.

Comparing column (2) with columns (6)-(8), we find that using a full set of experience dummies and interactions improves little over controlling for experience with only a linear and quadratic term, in particular on the coefficient of interest on S_P. So in column (9), we use linear and quadratic while instead controlling for individual schooling by including a full set of dummies for all observed years of schooling in the data (0 to 17) instead of linearly. The coefficient on mother's years of schooling is 1.6% and highly significant.

We have rerun these regressions controlling for race and cohort dummies, and also for white males separately. The coefficient on mom's schooling is quite stable across all specifications. The striking feature is that no matter how we control for experience, the coefficients on S and mom's S_P are similar in magnitude and highly significant. This means that the function C_1 in the model can be controlled for linearly, while $\tilde{C}_2$ can be controlled for using a simple linear and quadratic term for potential experience.[22]

To summarize, mothers' schooling has a stronger relationship with sons' earnings than fathers', and the estimated effect of mother's schooling, or $\hat{\beta}_2$, is about 1.7%, which is a measure of the gaps between the three lines in Figure 1. This is in the range of previous empirical work (Card, 1999), and according to our model is not a causal effect but interpreted as selection on abilities. To get the causal spillover, we need to slice the sample into subgroups of children with similar levels of schooling, whose distribution of abilities will differ from the population.

We do exactly this in Table 3. For each of the 4 subsamples divided by children's own level of schooling, we regress

$$\log e_{i,x} = \beta_0 + \beta_2 S_{P,i} + \beta_3 x + \beta_4 x^2 + \epsilon_{i,x},$$

---

[22]In fact, since $C_1(s)$ is close to exponential, its logarithm is close to linear, and a comparison of columns (6) and (9) suggest that the log earnings-schooling relationship is also linear in the data. Indeed, in Figure 8 in Appendix F, we plot the values of each schooling dummy against the linear return of 7.6%, and find that the dummies increase almost linearly. This gives us further confidence that the model can replicate important features of the data.

	$S \in [8, 12)$	$S = 12$	$S \in (12, 16]$	$S > 16$	W.Avg
Mom S_P	0.022	0.022	0.020	0.016	0.020
	(12.32)	(16.38)	(13.93)	(7.09)	
Cons	7.861	8.655	8.929	9.152	8.649
	(53.31)	(170.85)	(313.90)	(391.67)	
R^2	0.121	0.151	0.182	0.256	
Sample	944	1647	1478	640	4709

Table 3: Mincer Regressions by Own Schooling

OLS regressions of (log) earnings on mothers' years of schooling. HRS initial cohort, males born 1924-1941, ages 23-42. All columns include a linear and quadratic for potential experience (age-6-S). t-stats shown in parentheses. The last column is the population weighted average of the coefficients.

i.e., equation (13) without controlling for S_i and using a linear and quadratic term for potential experience x. Again, the Mincer coefficient on mother's schooling is significant for all subgroups at approximately 2%. This coefficient measures the average gaps between the pairs of earnings profiles for children with the same level of schooling but different levels of mothers' schooling in Figure 2. Moreover, the magnitudes of the coefficients are similar across all groups, although slightly lower for children with very high education (more than college).[23] This justifies our assumption of a constant ν that applies equally for all education groups (Corollary 2).

4   Estimation

Since schooling and earnings are only functions of abilities (z) and parents' schooling (S_P), the previous regressions may overestimate the role of both selection and spillovers in the presence of other dimensions of unobserved heterogeneity. Furthermore, the predictions are too strong to explain other important features, such as why the spillover seems to dampen at higher levels of education, or relatedly, why some children attain very high levels of schooling without any apparent gain in earnings. In this section we make the following assumptions on Program (1) to bring it closer to the data:

1. The choice of schooling is constrained to be discrete, consistently with the data.

2. Tastes for schooling, or non-pecuniary benefits, are modelled using a nested logit.

3. OJT only involves time inputs, and the return to human capital investments differ during the schooling and working phases: $(\alpha_{1W}, \alpha_{2W}) = (\alpha_W \neq \alpha_1, 0)$.

Although we cannot derive closed form solutions for schooling and earnings as in Section 2, in Appendix B we characterize the solution under these assumptions, and and describe how a so-

---

[23]As shown in the table, the sample size of this group smaller. Moreover, the years of schooling of both mother and child are top-coded at 17 years in the data.

lution is found numerically in Appendix D. We still want to be able to use Corollaries 3 and 2 to discipline our choice of moments, and will show through simulations that the basic intuition for identification carries over.

4.1 Population Distribution Assumptions

Our dataset contains information on individuals' schooling and their complete earnings profiles, and the schooling of parents. Since we only have information on the schooling of the parent, we approximate the parent's human capital, or earnings, of the parent by the standard Mincerian equation (7). The only role of β is to transform parents' schooling into human capital units before applying them to the children's initial condition (1e).[24]

We assume that $(\log h_P, \log z)$ are joint normal as in (2), and note that h_P is known given S_P and β. The distribution of S_P is taken from the empirical distriubtion function of mother's schooling in the HRS. It is observed in discrete years ranging from 0 to 16. Then for each mother's schooling level and corresponding h_P, the distributional assumption implies

$$\log z_i | \log h_P \sim \mathcal{N}\left(\mu_z + \rho_{zh_P}\frac{\sigma_z}{\sigma_{h_P}}\left(\log h_P - \mu_{h_P}\right), \sigma_z^2\left(1 - \rho_{zh_P}^2\right)\right). \tag{14}$$

For each combination of $\{h_P, z, S \in \{8, 10, 12, 14, 16, 18\}\}$, we solve the model numerically as described in Appendix D. This induces the optimal choice of S and resulting life-cycle earnings for any given initial condition (h_P, z).

Tastes for schooling vary across the population with parent's human capital and ability:

$$\tilde{\xi}_i(S_i) \equiv \delta_S\left(1 + \gamma_{h_p}h_{Pi} + \gamma_z z_i\right) + \xi_{S_i}. \tag{15}$$

where $\xi_i \equiv [\xi_{S_i}]_{S_i \in \{10,12,14,16,18\}}$ is 6-dimensional logit. The constants γ_{h_p} and γ_z capture the correlations between ξ_i and (h_{Pi}, z_i), respectively.[25]

Schooling decisions are nested depending on college-entry, i.e. the distributions of tastes for $S_i \in \{8, 10, 12\}$ and $S_i \in \{14, 16, 18\}$ are modeled as nested logit. The vector ξ_i is drawn from a 6-dimensional, generalized extreme value distribution with c.d.f. G and scale parameter σ_ξ:

$$G(\xi_i) = \exp\left\{-\left[\exp\left(-\xi_{8_i}/\sigma_\xi\zeta_h\right) + \exp\left(-\xi_{10_i}/\sigma_\xi\zeta_h\right) + \exp\left(-\xi_{12_i}/\sigma_\xi\zeta_h\right)\right]^{\zeta_h}\right.$$
$$\left.-\left[\exp\left(-\xi_{14_i}/\sigma_\xi\zeta_c\right) + \exp\left(-\xi_{16_i}/\sigma_\xi\zeta_c\right) + \exp\left(-\xi_{18_i}/\sigma_\xi\zeta_c\right)\right]^{\zeta_c}\right\}$$

---

[24] If β were not included and parent's human capital is $\exp(S_P)$, the estimated spillover would be $\beta\nu$.

[25] We have no formal theory of whether the correlation between parents' schooling and tastes is causal or not. However, our counterfactual results suggest that for our model to be consistent with IV studies that find a close-to-zero causal effect of mother's schooling on children, at least part of the correlation must be causal.

where $(1 - \zeta_h, 1 - \zeta_c) \in [0, 1]$ proxies the correlation within each nest. Now let

$$\tilde{u}_{i,S_i} \equiv \tilde{W}(h_{Pi}, z_i; S_i) + \delta_S \left(1 + \gamma_{h_P} h_{Pi} + \gamma_z z_i\right), \quad S_i \in \{8, 10, 12, 14, 16, 18\}.$$

Then for individuals with the same (h_{Pi}, z_i),

$$\Pr(S_i = 8) = \Pr(S_i = 8 | S_i \in \{8, 10, 12\}) \cdot \Pr(S_i \in \{8, 10, 12\}),$$

where

$$\Pr(S_i = 8 | S_i \in \{8, 10, 12\}) = \frac{\exp\left(\tilde{u}_{i,8_i}/\sigma_\xi \zeta_h\right)}{\exp\left(\tilde{u}_{i,8_i}/\sigma_\xi \zeta_h\right) + \exp\left(\tilde{u}_{i,10_i}/\sigma_\xi \zeta_h\right) + \exp\left(\tilde{u}_{i,12_i}/\sigma_\xi \zeta_h\right)} \tag{16}$$

$$\Pr\left(S_i \in \{8, 10, 12\}\right) = \tag{17}$$

$$\frac{\left[\exp\left\{\frac{\tilde{u}_{i,8_i}}{\sigma_\xi \zeta_h}\right\} + \exp\left\{\frac{\tilde{u}_{i,10_i}}{\sigma_\xi \zeta_h}\right\} + \exp\left\{\frac{\tilde{u}_{i,12_i}}{\sigma_\xi \zeta_h}\right\}\right]^{\zeta_h}}{\left[\exp\left\{\frac{\tilde{u}_{i,8_i}}{\sigma_\xi \zeta_h}\right\} + \exp\left\{\frac{\tilde{u}_{i,10_i}}{\sigma_\xi \zeta_h}\right\} + \exp\left\{\frac{\tilde{u}_{i,12_i}}{\sigma_\xi \zeta_h}\right\}\right]^{\zeta_h} + \left[\exp\left\{\frac{\tilde{u}_{i,14_i}}{\sigma_\xi \zeta_c}\right\} + \exp\left\{\frac{\tilde{u}_{i,16_i}}{\sigma_\xi \zeta_c}\right\} + \exp\left\{\frac{\tilde{u}_{i,18_i}}{\sigma_\xi \zeta_c}\right\}\right]^{\zeta_c}}.$$

4.2 Generalized Method of Moments

There are 26 parameters in the model, which we partition into

$$\theta_0 = [w, r, R, \beta, \mu_{h_P}, \sigma_{h_P}, \delta_8, \delta_{10}]$$
$$\theta_1 = \left[(\alpha_1, \alpha_2, \alpha_W, \nu, \lambda, b, \rho_{zh_P}, \mu_z, \sigma_z); \ (\sigma_\xi, \gamma_{h_P}, \gamma_z, \zeta_h, \zeta_c, \delta_{12}, \delta_{14}, \delta_{16}, \delta_{18})\right].$$

The first partition, θ_0, are parameters that are set *a priori*. The rest of the parameters in θ_1 are from the simple model and the taste structure in the extended model, respectively. These are estimated by GMM to fit schooling and earnings moments from the extended model, which can be computed exactly subject only to numerical approximation error (see Appendix D for details).[26]

Parameters Set a Priori We fix the interest rate at 5%, which is in the range of the after-tax rate of return on capital reported in Poterba (1998) and used in Heckman et al. (1998).[27] Retirement age is fixed at 65.

The coefficient β is recovered from a standard Mincer regression applied to the HRS AHEAD cohorts (without including mother's schooling). The purpose is to induce a statistical distribution of parents' earnings from their schooling, including all endogenous effects.[28] The resulting coefficient is quite stable across cohorts, ranging from approximately 0.04 to 0.06 for men and 0.05

---

[26]While we could derive the likelihood of the model, we choose a method of moments because it allows us to derive identification from key moments of the data that are important for our purposes. A likelihood estimator would attempt to fit individual behavior which our parsimonious model is not designed to match.

[27]They find that a time-varying interest rate does not lead to significant differences using a similar model.

[28]This regression also admits a constant term, which we ignore since it is not separately identified from b.

Parameter	Value	Description
w	1	Normalization; not separately identified from b
r	5%	interest rate, after-tax rate of return on capital in Poterba (1998)
R	65	Retirement age, exogenous
β	6%	Mincer return to schooling, HRS AHEAD cohorts
μ_{h_P}	0.55	Mean human capital of mothers, from data (see text)
σ_{h_P}	0.22	Standard deviation of mothers' human capital, from data (see text)
δ_8	0	Taste for 8 years of schooling, normalization

* See text for more details.

Table 4: Parameters Set a Priori

to 0.09 for women; we fix $\beta = 0.06$. This is not very different from the coefficients we recover from the (later-born) HRS cohorts in our sample in Table 2 which includes more controls, and our estimates are not sensitive to different values of a β within this range.

Since $\log h_P = \beta S_P$, we have $\mu_{h_P} = \beta \mu_{S_P}$ and $\sigma_{h_P} = \beta \sigma_{S_P}$. We take the mean and variance of mother's schooling, μ_{S_P} and σ_{S_P}, directly from their sample analogs in the data. Hence the only parametric assumption we are imposing by assuming that S_P is Gaussian is its correlation structure with z. Since $(\mu_{S_P}, \sigma_{S_P}) = (9.24, 3.60)$, we obtain $(\mu_{h_P}, \sigma_{h_P}) = (0.55, 0.21)$.

The tastes for 8 years of schooling, δ_8, is not separately identified from the other taste parameters (namely, γ_{h_P}, γ_z and σ_ξ) and normalized to 0. The entire list of exogenously fixed parameters are summarized in Table 4.

Estimated Parameters   The remaining 18 parameters in θ_1, are estimated by GMM to empirical moments of interest. These moments are schooling and earnings outcomes by level of mother's schooling, tabulated in the last 5 columns of Table 5. Since we constrain schooling choices to lie on 6 grid points, rather than targeting average years of schooling we target the probability of attaining high or low levels of schooling by 6 levels of mother's schooling. For each of these 12 groups, we construct average earnings for ages 25, 30, 35 and 40, which are in turn computed by simply averaging an individual's earnings from ages 23-27, 28-32, and so forth.

For each level of mom's schooling, 1 of the 2 educational attainment shares of the children are dropped (since they add up to 1). All average earnings are normalized by the lowest level of average earnings, i.e. the age 25 average earnings of children with less than 12 years of schooling and 5 or less years of mom's schooling, which is also dropped. Four additional moments are included: the correlation between S and S_P, the OLS coefficient from regressing S on S_P, and the Mincer regression coefficients on S and S_P from specification (2) of Table 2. These are included to capture the earnings and schooling gradients in the data we may miss by targeting aggregated moments. In sum, we have 57 moments to match with 18 model parameters.

Denote these target moments $\hat{\Psi}$. For an arbitrary value of θ_1, we compute the implied model

Mom's S_P	Fraction (%)	Child's S	Average $\bar{S}$	Fraction (%)	Average Earnings at age 23-27	28-32	33-37	38-42
≤5	12.75	≤11	6.35	60.04	1.00	1.41	1.71	1.85
		≥12	13.12	39.96	1.26	1.91	2.32	2.51
6-7	12.30	≤11	8.02	42.48	1.10	1.66	1.98	2.11
		≥12	13.53	57.52	1.27	1.91	2.37	2.65
8	21.76	≤12	10.72	66.23	1.35	1.99	2.42	2.62
		≥13	15.21	33.77	1.31	2.12	2.71	2.99
9-11	13.77	≤12	11.00	63.73	1.37	1.98	2.38	2.57
		≥13	15.24	36.27	1.30	2.15	2.65	2.96
12	30.00	≤12	11.21	45.41	1.52	2.24	2.62	2.77
		≥13	15.38	54.59	1.36	2.29	2.87	3.34
≥13	9.43	≤12	11.41	19.56	1.39	1.96	2.33	2.49
		≥13	15.83	80.44	1.30	2.23	2.82	3.27
(S, S_P) correlation and OLS:			0.48	0.46				
Mincer coefficients (β_1, β_2):			0.08	0.02				

Table 5: Targeted Empirical Moments

Note that for mom's with low S_P (the first four rows), we divide whether the child's educational attainment was low or high by whether or not he graduated from high school, while for the rest by whether he advanced beyond high school. In the third column, $\bar{S}$ denotes the average years of schooling attained in each category. All average earnings are normalized by the average earnings from 23-27 of the group with less than 12 years of schooling whose moms attained 5 years or less of schooling. (S, S_P) OLS denotes the coefficient from regressing S on S_P, and the Mincer coefficients are from specification (2) in Table 2.

moments $\Psi(\theta_1)$ as described in Appendix D. The parameter estimate $\hat{\theta}_1$ is found by

$$\hat{\theta}_1 = \arg \min_{\theta_1 \in \Theta_1} \left(\hat{\Psi} - \Psi(\theta_1) \right)' W \left(\hat{\Psi} - \Psi(\theta_1) \right)$$

where W is a weighting matrix. This procedure generates a consistent estimate of θ_1. We use a diagonal weighting matrix $W = diag(V^{-1})$, where V is the variance-covariance matrix of $\hat{\Psi}$. This weighting scheme allows for heteroskedasticity and can have better finite sample properties than the optimal weighting matrix (Altonji and Segal, 1996) in practice.[29] Minimization is performed using a Nelder-Mead simplex algorithm, and since this method does not guarantee global optima we tried several thousand different starting values to numerically search over a wide range of the parameter space (most of which have naturally defined boundaries). Asymptotic standard errors for the parameter estimates are obtained from

$$\sqrt{N}(\hat{\theta}_1 - \theta_1^*) \to \left(G'WG \right)^{-1} G'W\hat{V}WG \left(G'WG \right)^{-1}$$

---

[29]Using the optimal weighting matrix, we are not able to fit earnings moments, as shown in Appendix Figure 9. This is likely due to earnings moments being correlated across ages, which the diagonal matrix ignores.

as N approaches ∞, where $\hat{\theta}_1$ is the estimate, θ_1^* is the unknown, true parameter, and N the sample size. The matrix G is the $M \times P$ Jacobian of $\Psi(\theta_1)$ with respect to θ_1, where $(M = 57, P = 18)$ are the number of moments and parameters, respectively, and is computed numerically. The estimate of V, $\hat{V}$, is estimated via 2000 bootstrap draws.

Identification   As is usual with this class of models, identification is hard to prove formally. Our choice of moments is guided by intuition from the simpler model of how certain moments should have more influence on certain parameters, and are summarized in Table 11 in the appendix.

Corollaries 1-2 showed that in the simple model, the parameters of interest (ρ_{zh_P}, ν) can be identified by the coefficient on mother's schooling in a Mincer regression over the population (Table 2), and the coefficient on mothers' schooling in a regression over subsamples of individuals with the same level of schooling (Table 3). We show that this intuition carries over to our extended model through in Section 4.4.

Identification of the human capital production technologly parameters $(\alpha_1, \alpha_2, \alpha_W)$ are similar to other studies using Ben-Porath-type technologies. Since α_W governs the speed of human capital growth in the working phase, it is identified by the slope of average experience-earnings profiles. Conversely, the schooling parameters (α_1, α_2) determine can be identified by slopes of profiles by own and mother's schooling levels. at what level of earnings an individual begins his working phase. For high values of $\alpha \equiv\equiv \alpha_1 + \alpha_2$, individuals with high S will have flatter earnings profiles since they will have higher early age earnings, and for high values of α_1, individuals with high S_P will have flatter earnings profiles since they will benefit more from higher age 6 human capital and thus have higher early age earnings.

The level parameter b controls the overall amount of age 6 human capital, since with higher values schooling becomes less important for all individuals uniformly. Given an average level of schooling, the 5 taste parameters δ_S for $S \in \{10, 12, 14, 16, 18\}$ (δ_8 is normalized) should perfectly account for the shares of individuals choosing the 6 schooling levels, $S \in \{8, 10, 12, 14, 16, 18\}$, while (ζ_h, ζ_c) would capture the overall variation in wages by high school and college. Given an overall variation in schooling across all groups controlled by σ_{ξ}, the parameters (γ_{h_P}, γ_z) are identified by how educational attainment varies across mother's schooling and individual earnings.

4.3   Interpreting the Parameters

Table 6 reports the 18 parameter estimates and their asymptotic standard errors. The model generated educational attainment shares (empirical counterparts in fourth column of Table 5) are matched nearly exactly, as well as the 4 additional gradient moments (in the lower panel of Table 5 and first panel in Table 8). The earnings moments are compared with the data visually in Figure 7 in Appendix F.

	HC prod		Spillovers		"Ability"
α_1	0.258 (0.001)	ν	0.778 (0.010)	ρ_{zh_P}	0.229 (0.004)
α_2	0.348 (0.005)	λ	0.060 (0.026)	μ_z	-1.198 (0.012)
α_W	0.426 (0.004)	b	0.881 (0.026)	σ_z	0.117 (0.001)
	Tastes		Taste levels		
σ_ξ	0.612 (0.017)	ζ_h	0.266 (0.007)	δ_{14}	1.894 (0.008)
γ_{h_P}	1.551 (0.036)	ζ_c	0.846 (0.010)	δ_{16}	2.699 (0.021)
γ_z	0.549 (0.027)	δ_{12}	1.646 (0.004)	δ_{18}	4.198 (0.021)

*Standard errors in parentheses.

Table 6: Parameter Estimates

Human Capital Production   The human capital production parameters $(\alpha_1, \alpha_2, \alpha_W)$ are in the lower range of estimates found in the literature that use comparable Ben-Porath technologies. Estimates surveyed by Browning et al. (1999) lie in the range 0.5 to 0.9. This may have to do with the fact that the HRS cohort lived in a period in which observed education returns were much lower, for example, the college premium was about 40% prior to the 1980s rising to above 100% in 2000. The returns to human capital investment are slightly larger in school ($\alpha_1 + \alpha_2 = 0.606$) than on-the-job ($\alpha_W = 0.426$). This means that for purposes of human capital accumulation, an individual would prefer to stay in school rather than work.

Parental Spillover and Early Childhood   The magnitude of ν seemingly implies large spillovers—a mom with 10 percent higher human capital has a child with 8 percent higher initial human capital, controlling for selection on abilities and tastes. Increasing mom's schooling by 1 year increases her child's initial human capital by 12.3%. On the other hand, the estimated λ is both small and almost insignificant. This indicates that parents are much more important than the child's learning abilities for early human capital formation. Yet, we will show in Section 5 that the effect of increasing mom's schooling on lifetime earnings is an order of magnitude lower at about 1.2%. This is because the higher human capital early in life leads to less human capital accumulation and thus reduces schooling duration.

The astute reader might wonder whether we recover a large estimate because we assume that parents only have a level effect, i.e., that h_P only has a causal effect on h_6 but not on z, nor on the speed of human capital accumulation. Unfortunately, such a slope effect is not separately identified from ρ_{zh_P} in the simple model of Section 2. We have run several numerical simulations with

the extended model to verify that adding a slope spillover has small influence on all other parameter estimates except ρ_{zh_P}. This implies that a slope spillover would only crowd out selection effects, so our results can be viewed as a conservative lower-bound estimate for parental spillovers.

Selection on Learning Abilities   The estimate for ρ_{zh_P} implies that on average, mothers with 1 standard deviation of schooling above the population mean have children with learning abilities 0.23 standard deviations higher. Given the empirical estimate of σ_{S_P} and the model estimated σ_z, mothers with 1 more year of schooling have children with 0.7% higher abilities.

Unlike the spillover, this is a permanent difference that sustains through life. The impact on earnings at all ages, according to Lemma 4 in Appendix B, is similar to what we found in the simple model in Proposition 2, and can be approximated by $\Delta \log z / (1 - \alpha_W) = 1.3\%$ for a 1 year difference in mom's schooling. This is more or less similar to the selection effect on lifetime earnings, as we soon show in Section 5. Following similar calculations, a standard deviation of 0.117 for z translates into $\sigma_z / (1 - \alpha_W) = 0.204$ or a 20.4% standard deviation in earnings, once we control for schooling.

Tastes for Schooling   As expected, the constant δ_S rises with schooling attainment S. Recall that $(1 - \zeta_h, 1 - \zeta_c)$ is a measure of the correlation in tastes for schooling levels for high school and below, and some college and above. Hence, unobserved tastes for staying in high school are much more correlated than in college.[30] The idiosyncratic component of non-pecuniary benefits has a standard deviation of 3,846 dollars.

To facilitate the interpretation of the taste parameters, Table 7 reports the pecuniary and non-pecuniary benefits of schooling for different groups of individuals. The top panel reports the pecuniary benefits of schooling by ability quartiles. Individuals in the highest quartile have the highest pecuniary benefits when graduating from college while all other quartiles have the highest pecuniary benefits when graduating from high school.

That γ_{h_P} is much larger than γ_z implies that non-pecuniary or non-cognitive motives for staying in school are much more influenced by parents than by children's learning abilities. Consequently, non-pecuniary benefits are much higher for children of mother's with high S_P, although it is also higher for children with higher z, as reported in the middle and bottom panels of Table 7, respectively. Fast-learning children tend to like school more, but the major determinant is the mother, or more broadly the family background. This conforms to the notion that highly educated mothers are more likely to provide a family environment conducive for longer schooling, and also inculcate in their children a higher motivation to advance further in education. The correlation between children's tastes for schooling (non-pecuniary benefits) and parents' schooling plays an

---

[30]We thank an anonymous referee who had the insight that higher education levels may be less correlated also because of higher attrition, due to dropout risks or occupational requirements.

	$S = 8$	$S = 10$	$S = 12$	$S = 14$	$S = 16$	$S = 18$
Pecuniary benefits by ability quartile						
All	280,007	344,059	419,099	440,588	439,346	394,687
Q1	266,522	308,293	328,030	295,265	288,900	290,403
Q2	314,621	356,932	379,236	351,205	347,202	353,819
Q3	348,637	391,599	424,096	400,711	397,892	406,226
Q4	388,525	429,301	477,997	522,664	526,884	515,749
Nonpecuniary benefits correlated with h_P by S_P group						
All	-	11,682	21,912	23,706	35,877	64,857
$S_P < 8$	-	9,850	16,955	19,038	27,679	44,800
$S_P \in [8, 12)$	-	12,694	21,188	24,222	34,756	55,276
$S_P = 12$	-	15,622	25,711	29,584	42,161	65,578
$S_P > 12$	-	17,555	29,584	33,794	48,726	78,404
Nonpecuniary benefits correlated with z by ability quartile						
All	-	742	1,370	1,643	2,359	3,492
Q1	-	689	1,176	1,321	1,889	2,971
Q2	-	762	1,289	1,462	2,098	3,323
Q3	-	812	1,382	1,574	2,260	3,580
Q4	-	865	1,490	1,810	2,611	4,047

Table 7: Pecuniary and non pecuniary benefits of schooling
PDV pecuniary value at age 23, in 2008 USD. Non-pecuniary benefits are all relative to $S = 8$.

important role in the following analyses.

4.4 Fit Analysis

Comparison with reduced from prediction   Having confirmed that our intuition from Sections 2-3 carries over, we can apply the GMM estimates to Corollaries 1-2 and compare the model-predicted values of $(b_2, \bar{b}_2)$, the regression coefficients on S_P in a Mincer regression, to what we found in columns (2) and (9) of Table 2. This gives us the model and data predicted ability selection and spillover effects from having a 1-year more educated mother, according to the simple model. Since we have different α's in the extended model, we can get a range by computing

$$
\text{selection}: \ b_2 = \beta \rho_{zh_P} \sigma_z / (1 - \alpha) \sigma_{h_P} = \begin{cases} 0.019 & \text{if } \alpha = \alpha_1 + \alpha_2 \\ 0.013 & \text{if } \alpha = \alpha_W \end{cases} \tag{18}
$$

$$
\text{spillover}: \ \bar{b}_2 = \beta \nu / \left[1 - \lambda(1 - \alpha) \right] \ \approx 0.048 \quad \text{for both cases.} \tag{19}
$$

28

The implied ability selection coefficient is more or less in the range of its reduced form estimate of 1.7% that we obtained in column (6) of Table 2. However, the spillover coefficient is noticeably larger than the reduced form weighted average of 2% in Table 3.

All else equal, tastes for schooling induce individuals to stay in school longer at the the detriment of lifetime earnings. To make up for this and explain a 2% observed, reduced form spillover, the estimated value of ν must be larger than what would be implied by the simple model. Moreover, since tastes generate a large intergenerational schooling correlation, it needs to be moderated by a negative effect on schooling coming from a larger ν. For both these reasons, if we were to ignore taste heterogeneity as in (19), the implied spillover effect on earnings would become counterfactually high.

For the similar reasons, we may expect the estimated value of ρ_{zh_P} to be larger to make up for the negative effect of tastes on lifetime earnings and explain an observed 1.7% return. On the other hand, both children's schooling and earnings are increasing in z, unlike h_P which increases earnings but reduces schooling. So the estimate for ρ_{zh_P} must remain small; otherwise the observed intergenerational schooling correlation would become too large. Because these two forces cancel out, the reduced form effect is similar to when we ignore taste heterogeneity as in (18).

Shutting down parameters   To interpret the effect of spillovers and ability selection in light of our empirical analysis from Section 3, it is useful to see how the gradient moments are affected when shutting down the key parameters (ν, ρ_{zh_P}).[31] These exercises also show that the intuition from the simple model in the previous sections carries over to the estimated model.

The results are shown in the second panel of Table 8. The third panel tabulates the changes in the same moments when controlling for tastes for schooling. The first and second rows labeled $\gamma_{h_P}\mu_{h_P}$ and $\gamma_z\mu_z$ denote the cases where we keep all else equal and set

$$\tilde{\xi}(S) = \delta_S(1 + \gamma_{h_P}\mu_{h_P} + \gamma_z z) + \xi(S), \qquad \tilde{\xi}(S) = \delta_S(1 + \gamma_{h_P}h_P + \gamma_z\mu_z) + \xi(S)$$

The third row is when there is no variation in tastes across both (h_P, z).

Both the spillover ν and selection ρ_{zh_P} do little to affect the Mincer schooling coefficient β_1. And shutting down the spillover ν has only a small effect on the Mincer parental coefficient β_2. As expected from the theoretical results of Section 2, β_2 is mostly affected by selection ρ_{zh_P}. The addition of taste shocks does not alter this result.

While ability selection explains much of the linear parent's schooling-earnings relationship, it does little to affect the intergenerational schooling relationship. In fact, when both ρ_{zh_P} and ν are set to zero, schooling persistence becomes even higher, while it should be zero according to (9) in the simple model. This indicates that the observed intergenerational schooling persistence is

---

[31]Similar exercises with λ had minimal effects, since its point estimate is already close to zero.

	Corr_S	OLS_S	β_1	β_2
Data	0.469	0.444	0.076	0.017
Model	0.494	0.419	0.076	0.018
Structural				
$\nu = 0$	0.656	0.487	0.072	0.012
$\rho_{zh_P} = 0$	0.461	0.387	0.078	0.004
$\rho_{zh_P} = \nu = 0$	0.643	0.464	0.074	-0.003
Tastes				
$\gamma_{h_P}\mu_{h_P}$	-0.454	-0.381	0.076	0.025
$\gamma_z\mu_z$	0.492	0.415	0.075	0.018
$\gamma_z\mu_z, \gamma_{h_P}\mu_{h_P}$	-0.461	-0.385	0.074	0.025

Table 8: Effect of Spillover and Correlation Parameters.
(μ_{h_P}, μ_z) means that we shut down the correlation between tastes and (h_P, z). Corr_S and OLS_S denote, respectively, the correlation between S and S_P, and the OLS coefficient from regression S on S_P. (β_1, β_2) are the coefficients on (S, S_P) from the Mincer regression in column (2) of Table 2.

mainly a result of unobserved heterogeneity in tastes for schooling. Specifically, it is this and the countervailing force from ν that explains both the schooling correlation and OLS coefficient. When $\nu = 0$, the parental level affect disappears, inducing high ability individuals (whose parents tend to have higher levels of schooling) to increase their length of schooling, which in turn increases the correlation of schooling across generations. When tastes do not vary with mothers' schooling $(\gamma_{h_P}\mu_{h_P})$, children of high human capital parents (who tend to have higher levels of schooling) no longer have a desire to remain in school longer, and both Corr_S and OLS_S become negative. This implies that for schooling choices, tastes dominate selection on abilities: even though children with high z would spend more time in school, they do not if tastes are shut down. Letting tastes vary across different learning abilities has little effect on all moments as can be seen in the row $\gamma_z\mu_z$; this is somewhat expected since γ_{h_P} is much larger than γ_z. Indeed, when tastes do not vary along either dimension, the resulting numbers are more or less identical to when it varies only with abilities (last row).

Note that in all cases where we control for tastes, β_2 increases by a third. Without heterogeneity in tastes, individuals now diverge less from the schooling levels that would maximize their lifetime incomes. Consequently, the role of (h_P, z) becomes larger, which, in turn, results in a larger reduced form parental effect on earnings, β_2.

5 Counterfactual Experiments

We conduct two main experiments using the model estimates. First, we decompose the effects of a one-year increase in mother's schooling. Second, we implement a counterfactual compulsory schooling reform. Although the spillover coefficient ν has only a small effect on the reduced form

	(1) fixed S	(2) spillover	(3) ability	(4) tastes	(5) RF
Schooling diff (years)	-	-0.425	-0.384	0.451	0.479
Avg earnings diff (%)	0.007	0.012	0.025	0.002	0.016
Age 30 earnings diff (%)	0.009	0.001	0.017	0.018	0.033
Age 50 earnings diff (%)	0.004	0.001	0.015	0.009	0.023

Table 9: Aggregate Effect of 1 Year Increase in Mom's Schooling.
The second row denotes the change in the cross section average of the present discounted value of lifetime earnings, in logarithims. The first column holds abilities and tastes constant, while the next two columns let ability or taste also vary according to their estimated correlations with h_P. The column RF is when we allow for both selection on abilities and correlation with tastes for schooling. Schooling OLS in data and estimated model is 0.458 and 0.478, respectively.

coefficient on mother's schooling in a Mincer regression, a uniform one-year increase in mother's schooling leads to an average 1.2% increase in children's earnings controlling for selection on abilities and tastes for schooling, while further allowing for selection and changes in tastes for schooling leads to an additional 1.3% increase. And, this happens without increasing children's schooling. This is explained by parents with higher human capital having a negative effect on children's schooling as we saw in Proposition 1, which is countervailed by children of higher human capital parents having higher tastes for schooling.

5.1 Decomposing Spillovers from Selection

Given his state (h_P, z, ξ) at age 6, we compute the change in a child's schooling and earnings outcomes in response to a 1-year increase in S_P, which translates into an increase of β units of h_P in logarithms.

We perform several experiments to control for spillovers, selection effects and tastes for schooling. First, we hold constant the individual's (z, ξ) and also the schooling choice S, which isolates the pure quality effect coming from higher S_P. Next we still hold (z, ξ) constant, but let the individuals re-optimize their choice of S—since the higher initial human capital substitutes for the need to stay in school longer, earnings further increases while schooling decreases. The combined effect of the pure quality increase and schooling choice adjustment is the spillover effect. Then, we let either z or ξ vary with h_P as dictated by the distributional assumptions in Section 4.1; this separately captures the effects of selection on abilities and the correlation of parents' schooling with children's tastes for schooling. Finally, we let both z and ξ vary together, which we label a "reduced-form" effect—i.e., this is just comparing the average outcomes of children with S_P years of mother's schooling, to the average outcome of those with $S_P + 1$ years of mother's schooling. Refer to Appendix E for a more formal description.

The first and second rows of Table 9 lists the average effects of a 1 year increase in S_P on schooling S and the present-discounted value of lifetime earnings, respectively. The third and

fourth rows show the average change in earnings at ages 30 and 50. The change in S is in years and the changing in earnings in log-point differences, to approximate percentage changes. Column (1) shows the change in the child's earnings when his schooling is held constant. Columns (2) to (5) depict the change in schooling or earnings when holding each individual's (z, ξ) constant, when allowing them to vary according to their correlations with h_P, and lastly the combined, reduced form effect.

As expected, the spillover effect on schooling is negative, namely, the model predicts that increasing all mothers' schooling by a year would lead to an average 0.425 year decline in the schooling of the child generation. Surprisingly, allowing for selection on abilities only moderates this by 0.039 years (comparing columns (2) and (3)) or 0.028 years (columns (4) and (5)). As we saw in Table 8, selection on abilities does not have much of an effect on intergenerational schooling relationships once tastes are taken into account. Since tastes already induce individuals to stay in school longer than what maximizes lifetime earnings, letting abilities become higher does not further increase schooling much.[32]

Only when we allow allow tastes to vary with mothers' schooling do we see a positive effect of a 0.451 year increase in children's schooling following the 1 year increase in mothers' schooling. The effect is large, but is moderated by the negative spillover effect: comparing columns (2) and (4), the sole effect of tastes is an increase of 0.876 years. Even ignoring selection on abilities, spillovers and tastes alone generate an intergenerational schooling relationship that is close to its empirical counterpart.

The spillover increases lifetime earnings by about 0.7 percent holding schooling constant, which increases to 1.2 percent when allowing schooling to adjust. Selection on abilities has a 1.3 percent effect. We conclude that independent of children's tastes for schooling, the causal effect of mother's education on earnings is more or less similar to the selection effect on abilities, i.e., high ability mothers having high ability children. But if the correlation between mother's schooling and children's tastes for schooling are entirely casual, the average impact is negative: the increase in lifetime earnings drops by 1 percentage point (2nd vs. 4th columns). The effect is negative, since tastes for schooling make individuals deviate from lifetime earnings maximization, but not quite enough to dominate the positive effects from the spillover.

Table 9 shows the average change in the sum of net present discounted value of lifetime earnings, but the earnings effects differ substantially over the life-cycle, and also across children of mother's with different levels of schooling. Life-cycle effects are depicted in Figure 3, where each line plots the change in average (on the left panel) or median (on the right) earnings following a 1-year increase in mother's schooling by age, compared to the benchmark estimates. For clarity,

_____

[32]Since $h_0 = z^\lambda h_P^\nu$, some of the desire to increase schooling (since children can learn more in a fixed amount of time) is countervailed by a higher h_0 (since there is less need to learn when human capital is already high). Yet, this effect is not large given the low estimated value of $\lambda = 0.06$.

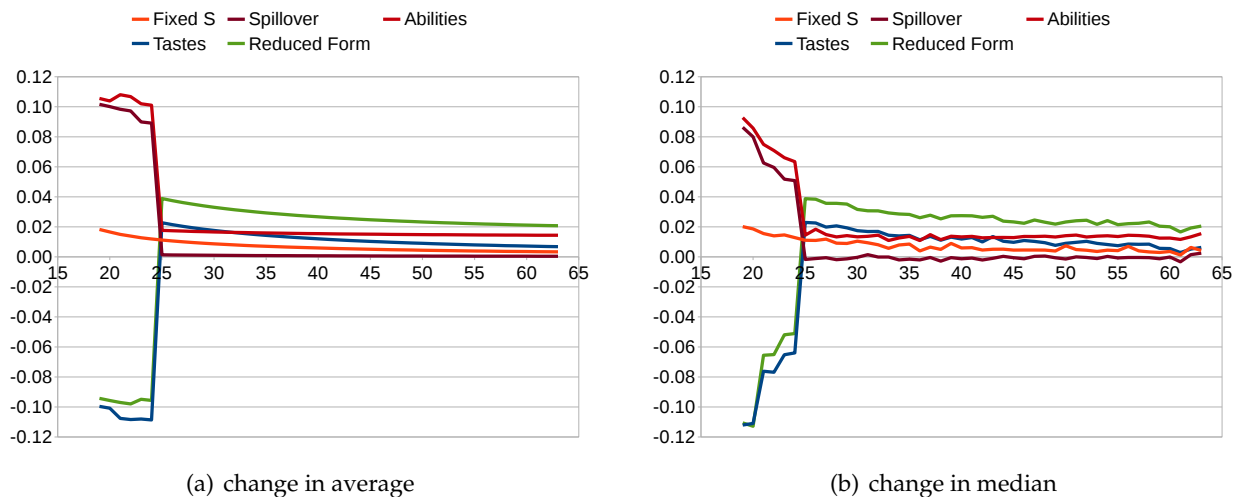(a) change in average   (b) change in median

Figure 3: Lifecycle Effect of 1 Year Increase in Mom's Schooling.

we also plot earnings differences at ages 30 and 50 in Table 9 as well.

Comparing "Fixed S" and "Spillover," we see that the positive lifetime earnings spillover effect comes almost entirely from early labor market entry followed by almost no change in earnings after age 24 (the latest age we allow labor market entry in the model). This also means that on average, children of mothers with less schooling catch up with those of mothers with more schooling by staying in school longer, so that their earnings do not differ much later in life. Allowing for selection on abilities has a fixed positive effect throughout the life-cycle, as expected from Section 2. Conversely, longer schooling induced by tastes for schooling increase lifetime earnings later in life (through more human capital accumulated in school), but this is dominated by the foregone earnings earlier in life.

5.2 Counterfactual Compulsory Schooling Reform

We next impose a minimum schooling requirement which is intended to mimic compulsory schooling reforms that took place in many countries throughout the 20th century. Such a reform only affects those parents who would otherwise not attain the required level of schooling. It shows it is possible to simultaneously estimate a large schooling OLS coefficient and a small or negative IV coefficient, indicating that the reform in the parents' generation has little causal effect on children's schooling. Nonetheless, it still has a positive causal effect on children's earnings.

We impose a minimum 8 years of schooling for all parents and set the initial level of human capital of all children to

$$h_0 = bz^\lambda h_P^\nu = bz^\lambda \exp\left[\nu\beta \max\{S_P, 8\}\right].$$

33

		(1) fixed S	(2) spillover	(3) ability	(4) tastes	(5) RF
Schooling	OLS	-	0.492	0.487	0.458	0.455
	IV	-	-0.204	-0.148	0.227	0.261
Avg earnings diff (%)		0.025	0.028	0.075	0.017	0.066
Age 30 earnings diff (%)		0.027	0.015	0.067	0.041	0.094
Age 50 earnings diff (%)		0.013	0.007	0.056	0.020	0.069

Table 10: Counterfactual Schooling Reform

The second row denotes the change in the cross section average of the present discounted value of lifetime earnings, in logarithms, for only those individuals affected by the reform. The first column holds abilities and tastes constant, while the next two columns let ability or taste also vary according to their estimated correlations with h_P. The column RF is when we let both vary. Schooling OLS in data and estimated model is 0.444 and 0.412, respectively.

We choose 8 years as the hypothetical requirement because it was the compulsory schooling requirements in many U.S. states at the time, or soon after.[33] Such a reform would affect 25% of the individuals in our data, increasing the schooling of their mothers by an average of 3.5 years.

For the schooling OLS and IV regressions, we combine samples of two regimes: one without the minimum requirement (our benchmark model) and one imposing the requirement. These represent mother-child pairs pre- and post-reform, respectively. Then we run OLS and IV regressions on the merged data, using a dummy variable for the different regimes as an instrument. As above, we repeat this exercise for four cases controlling for (z, ξ), and then allowing z and/or ξ to vary with h_P according to their correlations. The first row in 10 shows the regression results for all cases. The second row shows the change in the present discounted value of lifetime earnings, in log-points, and the following two rows the average change in earnings at ages 30 and 50.

The OLS coefficients are somewhat difficult to interpret, since the constants in the regressions are also changing. But with low-educated parents no longer in the sample post-reform, the OLS increases in all 4 cases to a value slightly higher than in the benchmark (0.412). The IV regressions, which measures the average 1-year effect among children whose mother's became more educated due to the reform. The controlled effect is negative, but of smaller magnitude. The IV coefficient increases when including ability selection (columns 2 vs. 3) but much more when including correlation with tastes (columns 2 vs. 4). The reduced form effect is only half of the population reduced form effect following a 1-year increase, in Table 9. As discussed there, this is because both the spillover and correlation with tastes has a smaller effect for lower levels of S_P, who are the only ones affected by the reform.

We conjecture that this partially explains the puzzling fact that many studies using special data sets on twins, adoptees, or compulsory schooling reforms find a zero or negative effect of parents'

---

[33]Black et al. (2005) study the case of Norway, whose compulsory schooling requirement went up from 7 to 9 years in the 1960s. While the location and timing differs, our moms' average years of schooling is only 1 year less (10.5 vs 9.3 years). However, the percentage of the moms that would be affected is almost two-fold (12.4% vs 25%).

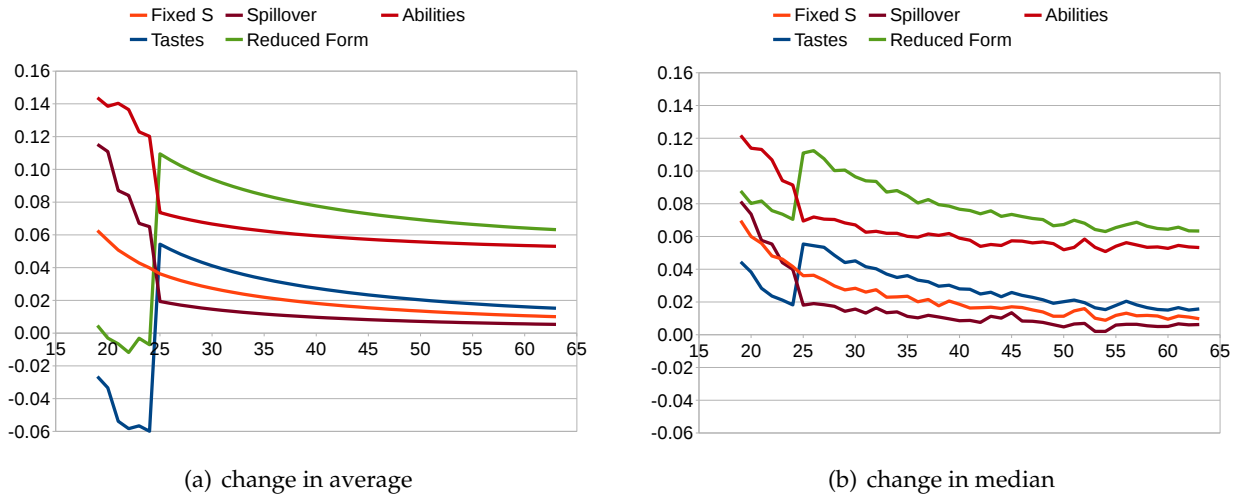|	
(a) change in average	(b) change in median

Figure 4: Lifecycle Impact of Reform

schooling on children's schooling. First, the IV coefficients are small in absolute magnitude in columns 2 and 4, because children of less educated mothers enjoy less spillovers, but also lose less in terms of lifetime earnings due to tastes. Second, the structural effect may in fact be negative, as we have argued throughout this paper. The fact that in some studies the effect is found to be close to zero can be due to the fact that forcing mothers to obtain more schooling at least partially induces their children to develop higher tastes for schooling. We can interpret this as mothers who are forced to attend school longer, but would not have attained higher levels of schooling otherwise, having some impact on children's non-cognitive abilities or perception of schooling that help them stay in school, although perhaps not to the extent that mothers who *choose* to become highly educated transmit high tastes for schooling. In our experiment, that would mean that education for mothers who would otherwise attain very low levels of schooling could increase the schooling of their children anywhere from -0.2 to 0.2 years.[34]

The spillover effect is positive through life, as can be seen in Figure 4(a), unlike in Figure 3 where it became virtually zero at ages 24 and above. Again, this is because for children with low S_P, the spillover effect is dominated by a pure quality effect, as can be seen by the much smaller difference between the "Fixed S" and "Spillover" lines. The negative effect that comes from tastes for schooling is still there, meaning that, if mothers with low education, following a reform, raise their children's tastes for schooling, we may expect children's schooling *and* lifetime earnings to be close to zero. But the negative effect is smaller in magnitude (although the lifetime earnings effect is -1.1 percentage points for both Tables 9 and 10, mother's schooling increases by 3.5 years in the latter compared to 1 year in the former). This is because these children have lower

---

[34]This explanation could reconcile Black et al. (2005)'s finding of a zero IV with Oreopoulos and Page (2006)'s finding that compulsory schooling reforms in the U.S. reduced grade repetitions among children of the reform cohort.

tastes for schooling to begin with, and have much to lose by earlier labor market entry because of decreasing returns to human capital accumulation. This is even more obvious when we look at median earnings in Figure 4(b). There, at nowhere during the life-cycle does the reform have a negative effect, not even at early ages, when including the correlation between mothers' schooling and children's tastes for schooling. And as is evident there and also in the 3rd and 4th rows of Table 10, the positive effect at later stages in life can be quite large.

6   Conclusion

In this paper we present a model of human capital which features endogenous schooling and earnings to isolate the causal effect of parents' education on children's education and earnings outcomes. The model has several important ingredients—tastes for schooling are correlated across generations, ability is also persistent across generations, and finally, the human capital of a parent is an input in producing human capital for the child. We label the last feature a parental spillover. Despite the positive relationship between the child's own schooling and earnings, the causal effect of parent's schooling on children's schooling can be negative, even when the causal effect is positive for children's earnings. Children of higher human capital parents begin life with higher human capital themselves, and when schooling is endogenous, they can spend less time in school but still attain the same or higher level of earnings. A simple version of the model is solved in closed form and its implications compared to empirical evidence in the HRS data.

Our model is consistent with several features of the joint distribution of parent schooling, child schooling and child earnings over the life-cycle. It is also consistent with a positive OLS correlation between parent and child schooling. The model feature that plays an important role in generating this feature is the correlation in tastes for schooling across generations. Our model is consistent with the previous literature that finds when using compulsory schooling as an instrument for parents' schooling, the estimated IV coefficient on children's schooling is zero. On the one hand, all else equal, increasing a parent's human capital decreases a child's schooling, a feature of diminishing returns to human capital accumulation. On the other hand, if a parent with higher human capital partially increases a child's taste for schooling, the negative effect is countervailed, resulting in an IV estimate that is close to zero.

Another important finding is that the unobserved correlation between mothers' education and children's tastes for schooling is the main determinant of children's schooling, not selection on abilities or parental spillovers. Finally, even though the causal effect of parent schooling on child schooling is negative, the estimated causal effect of mother's education on children's lifetime earnings can be as large as 1.2%. Although not directly comparable, our result that the causal effect on earnings is similar to the selection effect is in line with the "nature-nurture" literature, which finds that nurture effects are at most similar or less than nature effects.

References

Altonji, J. G. and L. M. Segal (1996, July). Small-Sample Bias in GMM Estimation of Covariance Structures. *Journal of Business & Economic Statistics 14*(3), 353–66.

Avery, C. and S. Turner (2012, Winter). Student loans: Do college students borrow too much–or not enough? *Journal of Economic Perspectives 26*(1), 165–92.

Behrman, J. R. and M. R. Rosenzweig (2002, March). Does increasing women's schooling raise the schooling of the next generation? *American Economic Review 92*(1), 323–334.

Behrman, J. R. and P. Taubman (1989, December). Is Schooling "Mostly in the Genes"? Nature-Nurture Decomposition Using Data on Relatives. *Journal of Political Economy 97*(6), 1425–46.

Belley, P. and L. Lochner (2007). The Changing Role of Family Income and Ability in Determining Educational Achievement. *Journal of Human Capital 1*(1), 37–89.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy 75*, 352.

Betts, J. R. (1996). What do students know about wages? evidence from a survey of undergraduates. *The Journal of Human Resources 31*(1), pp. 27–56.

Björklund, A., M. Lindahl, and E. Plug (2006). The origins of intergenerational associations: Lessons from swedish adoption data. *The Quarterly Journal of Economics 121*(3), 999–1028.

Black, S. E. and P. J. Devereux (2011). *Recent Developments in Intergenerational Mobility*, Volume 4 of *Handbook of Labor Economics*, Chapter 16, pp. 1487–1541. Elsevier.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. *The American Economic Review 95*(1), pp. 437–449.

Bowles, S. and H. Gintis (2002, Summer). The Inheritance of Inequality. *Journal of Economic Perspectives 16*(3), 3–30.

Bowles, S., H. Gintis, and M. Osborne (2001). The determinants of earnings: A behavioral approach. *Journal of Economic Literature 39*(4), 1137–1176.

Browning, M., L. P. Hansen, and J. J. Heckman (1999). Micro data and general equilibrium models. In J. B. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, Volume 1 of *Handbook of Macroeconomics*, Chapter 8, pp. 543–633. Elsevier.

Card, D. (1999). The causal effect of education on earnings. Volume 3A of *Handbook of Labor Economics*. Amsterdam: Elsevier.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010, 05). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Del Boca, D., C. Monfardini, and C. Nicoletti (2012, March). Children's and parents' time-use choices and cognitive development during adolescence. Working Papers 2012-006, Human Capital and Economic Opportunity Working Group.

Goldin, C. and L. F. Katz (2007, March). The race between education and technology: The evolution of u.s. educational wage differentials, 1890 to 2005. NBER Working Papers 12984, National Bureau of Economic Research, Inc.

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

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

Heckman, J. J., J. Stixrud, and S. Urzua (2006, July). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics 24*(3), 411–482.

Hoxby, C. and C. Avery (2013). The Missing "One-Offs": The Hidden Supply of High-Achieving, Low-Income Students. *Brookings Papers on Economic Activity 46*(1 (Spring), 1–65.

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

Kennan, J. (2006). A note on discrete approximations of continuous distributions.

Léonard, D. and N. Van Long (1992). *Optimal Control Theory and Static Optimization in Economics*. Cambridge University Press.

Oreopoulos, P., M. Page, and A. H. Stevens (2008). The intergenerational effects of worker displacement. *Journal of Labor Economics 26*(3), pp. 455–000.

Oreopoulos, P. and M. E. Page (2006, October). The Intergenerational Effects of Compulsory Schooling. *Journal of Labor Economics 24*(4), 729–760.

Plug, E. (2004). Estimating the effect of mother's schooling on children's schooling using a sample of adoptees. *The American Economic Review 94*(1), pp. 358–368.

Plug, E. and W. Vijverberg (2003, June). Schooling, family background, and adoption: Is it nature or is it nurture? *Journal of Political Economy 111*(3), 611–641.

Poterba, J. M. (1998). The rate of return to corporate capital and factor shares: new estimates using revised national income accounts and capital stock data. *Carnegie-Rochester Conference Series on Public Policy 48*(0), 211 – 246.

Rege, M., K. Telle, and M. Votruba (2011). Parental job loss and children's school performance. *The Review of Economic Studies 78*(4), pp. 1462–1489.

Sacerdote, B. (2007, February). How Large Are the Effects from Changes in Family Environment? A Study of Korean American Adoptees. *The Quarterly Journal of Economics 122*(1), 119–157.

Sacerdote, B. (2011). Nature and nurture effects on children's outcomes: What have we learned from studies of twins and adoptees? Volume 1 of *Handbook of Social Economics*, pp. 1 – 30. North-Holland.

Scholz, J. K., A. Seshadri, and S. Khitatrakun (2006). Are americans saving "optimally" for retirement? *Journal of Political Economy 114*, 607–643.

Solon, G. (1999, October). Intergenerational mobility in the labor market. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 29, pp. 1761–1800. Elsevier.

For Online Publication

A   Proofs to Propositions 1 and 2

The proof requires a complete characterization of the income maximization problem. While we can use standard methods to derive the solution, here we simply guess and verify the value function. For notational convenience, we drop the age argument a unless necessary. Although we normalize $w = 1$ in the estimation, we keep it here for analytical completeness.

We separately characterize the solutions during the schooling and working phases in Lemmas 1 and 2. Schooling choice S is characterized as the solution to an optimal stopping time problem in Lemma 3. To this end, we guess that

$$V(a, h) = q_2(a)h + C_W(a), \qquad\qquad\qquad \text{for } a \in [6 + S, R),$$

$$V(a, h) = q_1(a) \cdot \frac{h^{1-\alpha_1}}{1 - \alpha_1} + e^{-r(6+S-a)} \cdot C(S, h_S), \qquad \text{for } a \in [6, 6+S), \text{if } S > 0,$$

where

$$C(S, h_S) = q_2(6 + S)h_S + C_W(6 + S) - q_1(6 + S) \cdot \frac{h_S^{1-\alpha_1}}{1 - \alpha_1},$$

for which the length of schooling S and level of human capital at age $6 + S$, h_S, are given, and C_W is some redundant function of age. These are the appropriate guesses for the solution, and the transversality condition becomes $q(R) = 0$. Now we characterize the working phase:

LEMMA 1: WORKING PHASE *Suppose optimal schooling satisfies $S \in [0, R - 6)$. Then given $h(6 + S) \equiv h_S$ and $q(R) = 0$, the solution satisfies, for $a \in [6 + S, R)$,*

$$q_2(a) = \frac{w}{r} \cdot q(a) \tag{20}$$

$$m(a) = \alpha_2 \left[\kappa q(a) z \right]^{\frac{1}{1-\alpha}} \tag{21}$$

$$h(a) = h_S + \frac{r}{w} \cdot \left[\int_{6+S}^{a} q(x)^{\frac{\alpha}{1-\alpha}} dx \right] \cdot (\kappa z)^{\frac{1}{1-\alpha}} \tag{22}$$

and

$$\frac{wh(a)n(a)}{\alpha_1} = \frac{m(a)}{\alpha_2}, \tag{23}$$

where

$$q(a) \equiv \left[1 - e^{-r(R-a)} \right], \qquad \kappa \equiv \frac{\alpha_1^{\alpha_1} \alpha_2^{\alpha_2} w^{1-\alpha_1}}{r}.$$

Proof. Given that equation (3) holds at equality, dividing by (4) leads to equation (23), so once we know the optimal path of $h(a)$ and $m(a)$, $n(a)$ can be expressed explicitly. Plugging (3) and the guess for the value function into equation (5), we obtain the linear, non-homogeneous first order differential equation

$$\dot{q}_2(a) = rq_2(a) - w,$$

to which (20) is the solution. Using this result in (3)-(4) yields the solution for m, (21). Substituting (20), (21) and (23) into equation (1c) trivially leads to (22). □

If $S = 0$ (which must be determined), the previous lemma gives the unique solution to the income maximization problem. If $S > 0$, what follows solves the rest of the problem, beginning with the next lemma describing the solution during the schooling period.

LEMMA 2: SCHOOLING PHASE *Suppose optimal schooling satifies $S \in (0, R - 6)$. Then given $h(6) = h_0$ and $q_1(6) = q_0$, the solution satisfies, for $a \in [6, 6 + S)$,*

$$q_1(a) = e^{r(a-6)}q_0 \tag{24}$$

$$m(a)^{1-\alpha_2} = \alpha_2 e^{r(a-6)} \cdot q_0 z \tag{25}$$

$$h(a)^{1-\alpha_1} = h_0^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left[e^{\frac{\alpha_2 r(a-6)}{1-\alpha_2}} - 1 \right] \cdot (\alpha_2 q_0)^{\frac{\alpha_2}{1-\alpha_2}} z^{\frac{1}{1-\alpha_2}}. \tag{26}$$

Proof. Using the guess for the value function in (5) we have

$$\dot{q}_1(a) = rq_1(a),$$

to which solution is (24). Then equation (25) follows directly from (4), and using this in (1c) yields the first order ordinary differential equation

$$\dot{h}(a) = h(a)^{\alpha_1} \left[\alpha_2 q_1(a) \right]^{\frac{\alpha_2}{1-\alpha_2}} z^{\frac{1}{1-\alpha_2}},$$

to which (26) is the solution. □

The only two remaining unknowns in the problem are the age-dependent component of the value function at age 6, q_0, and human capital level at age $6 + S$, h_S. This naturally pins down the length of the schooling phase, S. The solution is solved for as a standard stopping time problem.

LEMMA 3: VALUE MATCHING AND SMOOTH PASTING *Assume $S > 0$ is optimal. Then (q_0, h_S),*

are given by

$$q_0 = \frac{e^{-rS}}{\alpha_2^{\alpha_2}} \cdot \left([\kappa q(6+S)]^{1-\alpha_2} z^{\alpha_1} \right)^{\frac{1}{1-\alpha}} \tag{27}$$

$$h_S = \frac{\alpha_1}{w} \cdot [\kappa q(6+S)z]^{\frac{1}{1-\alpha}} . \tag{28}$$

Proof. The value matching for this problem boils down to setting $n(6+S) = 1$ in the working phase, which yields (28).[35] The smooth pasting condition for this problem is

$$\lim_{a\uparrow 6+S} \frac{\partial V(a,h)}{\partial h} = \lim_{a\downarrow 6+S} \frac{\partial V(a,h)}{\partial h}.$$

Using the guesses for the value functions, we have

$$q_1(6+S)h_S^{-\alpha_1} = q_2(6+S) \qquad \Leftrightarrow \qquad h_S^{\alpha_1} = \frac{r}{w} \cdot \frac{e^{rS}}{q(6+S)} \cdot q_0,$$

and by replacing h_S with (28) we obtain (27). □

This, and the solutions for $[n(a)h(a), m(a)]$ during the working phase in Lemma 1 proves Proposition 2. We must still show Proposition 1.

Proof of Proposition 1. The length of the schooling period can be determined by plugging equations (27)-(28) into (26) evaluated at age $6+S$:

$$\left(\frac{\alpha_1}{w} \cdot [\kappa q(6+S)z]^{\frac{1}{1-\alpha}} \right)^{1-\alpha_1}$$
$$\leq h_0^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2^{1-\alpha_2}} \cdot \left(1 - e^{-\frac{\alpha_2 rS}{1-\alpha_2}} \right) \cdot \left([\kappa q(6+S)]^{\alpha_2} z^{1-\alpha_1} \right)^{\frac{1}{1-\alpha}},$$

with equality if $S > 0$. This implies that human capital accumulation must be positive in schooling, which is guaranteed by the law of motion for human capital. Rearranging terms,

$$h_0 \geq \frac{\alpha_1}{w} \cdot \left[1 - \frac{(1-\alpha_1)(1-\alpha_2)}{\alpha_1\alpha_2} \cdot \frac{1 - e^{-\frac{\alpha_2 rS}{1-\alpha_2}}}{q(6+S)} \right]^{\frac{1}{1-\alpha_1}} \cdot [\kappa q(6+S)z]^{\frac{1}{1-\alpha}},$$

or now replacing $h_0 \equiv z^\lambda h_P^\nu$,

$$z^{1-\lambda(1-\alpha)} h_P^{-\nu(1-\alpha)} \leq F(S), \tag{29}$$

---

[35]This means that there are no jumps in the controls. When the controls may jump at age $6 + S$, we need the entire value matching condition.

$$F(S)^{-1} \equiv \kappa \left(\frac{\alpha_1}{w} \right)^{1-\alpha} \cdot \left[1 - \frac{(1-\alpha_1)(1-\alpha_2)}{\alpha_1 \alpha_2} \cdot \frac{1 - e^{-\frac{\alpha_2 rS}{1-\alpha_2}}}{q(6+S)} \right]^{\frac{1-\alpha}{1-\alpha_1}} \cdot q(6+S)$$

which is the equation in the proposition. Define $\bar{S}$ as the solution to

$$\alpha_1 \alpha_2 q(6+\bar{S}) = (1-\alpha_1)(1-\alpha_2) \left(1 - e^{-\frac{\alpha_2 r\bar{S}}{1-\alpha_2}} \right),$$

i.e. the zero of the term in the square brackets. Clearly, $\bar{S} < R - 6$, $F'(S) > 0$ on $S \in [0, \bar{S})$, and $\lim_{S \to \bar{S}} F(S) = \infty$. An interior solution ($S > 0$) requires that

$$F(0) < z^{1-\lambda(1-\alpha)} h_P^{-\nu(1-\alpha)} \qquad \Leftrightarrow \qquad z^{1-\lambda(1-\alpha)} h_P^{-\nu(1-\alpha)} > \frac{r}{\alpha_1^{1-\alpha_2} (\alpha_2 w)^{\alpha_2} \cdot q(6)},$$

and S is determined by (29) at equality. The full solution is given by Lemmas 1-3 and we obtain Proposition 2. Otherwise $S = 0$ and the solution is given by Lemma 1. $\qquad \square$

B   Analytical Characterization when $(\alpha_{1W}, \alpha_{2W}) = (0, \alpha_W)$

It is instructive to first characterize the solution to the model when the schooling choice, S, is still continuous. In this case, the solution to the schooling phase is identical to Lemma 2. In the working phase, there can potentially be a region where $n(a) = 1$ for $a \in 6 + [S, S + J)$, and $n(a) < 1$ for $a \in [6 + S + J, R)$, so we can characterize the "full-time OJT" duration, J, following Appendix A.

LEMMA 4: WORKING PHASE, EXTENDED *Assume that the solution to the income maximization problem is such that $n(a) = 1$ for $a \in [6 + S, 6 + S + J)$ for some $J \in [0, R - 6 - S)$. Then given $h_S \equiv h(6 + S)$, the value function for $a \in [6 + S + J, R)$ can be written as*

$$V(a, h) = \frac{w}{r} \cdot q(a)h + D_W(a) \tag{30}$$

and the solution is characterized by

$$n(a)h(a) = \left[\frac{\alpha_W}{r} \cdot q(a)z \right]^{\frac{1}{1-\alpha_W}} \tag{31}$$

$$h(a) = h_J + \left(\frac{\alpha_W}{r} \right)^{\frac{\alpha_W}{1-\alpha_W}} \cdot \left[\int_{6+S+J}^{a} q(x)^{\frac{\alpha_W}{1-\alpha_W}} dx \right] \cdot z^{\frac{1}{1-\alpha_W}}, \tag{32}$$

where $h_J \equiv h(6 + S + J)$ is the level of human capital upon ending full-time OJT. If $J = 0$, there is nothing further to consider. If $J > 0$, the value function in the full-time OJT phase, i.e. $a \in [6 + S, 6 + S + J)$ can

43

be written as

$$V(a,h) = e^{r(a-6-S)}q_S \cdot \frac{h^{1-\alpha_W}}{1-\alpha_W} + e^{-r(6+S+J-a)}D(J,h_J) \tag{33}$$

where

$$D(J,h_J) = \frac{w}{r} \cdot q(6+S+J)h_J + D_W(6+S+J) - e^{rJ}q_S \cdot \frac{h_J^{1-\alpha_W}}{1-\alpha_W}$$

while human capital evolves as

$$h(a)^{1-\alpha_W} = h_S^{1-\alpha_W} + (1-\alpha_W)(a-6-S)z. \tag{34}$$

If $J > 0$, the age-dependent component of value function at age $6+S$, q_S, and age $6+S+J$ level of human capital, h_J, are determined by

$$q_S = we^{-rJ} \cdot \left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J)z^{\alpha_W} \right]^{\frac{1}{1-\alpha_W}} \tag{35}$$

$$h_J = \left[\frac{\alpha_W}{r} \cdot q(6+S+J)z \right]^{\frac{1}{1-\alpha_W}}. \tag{36}$$

The previous Lemma follows from applying the proof in Appendix A. The solution for J is also obtained in a similar way we obtained S. Since human capital accumulation must be positive during the full-time OJT phase,

$$\frac{\alpha_W}{r} \cdot q(6+S+J)z \leq h_S^{1-\alpha_W} + (1-\alpha_W)Jz,$$

with equality if $J > 0$. Rearranging terms,

$$\frac{z}{h_S^{1-\alpha_W}} \leq G(J) \equiv \left[\frac{\alpha_W}{r} \cdot q(6+S+J) - (1-\alpha_W)J \right]^{-1}. \tag{37}$$

Define $\bar{J}$ as the zero to the term in the square brackets, then clearly $\bar{J} < R - S - 6$, $G'(J) > 0$ on $J \in [0, \bar{J})$, and $\lim_{J \to \bar{J}} G(J) = \infty$. Hence an interior solution $J > 0$ requires that

$$G(0) < \frac{z}{h_S^{1-\alpha_W}} \qquad \Leftrightarrow \qquad \frac{r}{\alpha_W q(6+S)} < \frac{z}{h_S^{1-\alpha_W}}, \tag{38}$$

and J is determined by (37) at equality. Otherwise $J = 0$.

Now if S were discrete, as in the model we estimate, we only need to solve for h_S, the level of human capital at age $6 + S$. Then we can solve for $V(h_0, z; s)$ for all 6 possible values of s, using Lemmas 2 and 4 for the schooling and working phases, respectively. But it is also possible to

characterize the unconstrained continuous choice of S, even though a closed form solution does not exist in general. We only need consider new value matching and smooth pasting conditions.

LEMMA 5: SCHOOLING PHASE, EXTENDED *The length of schooling, S, and level of human capital at age $6 + S$, h_S, are determined by*

1. *if $J = 0$,*

$$\epsilon + (1 - \alpha_2)\left[\frac{\alpha_2^{\alpha_2}w}{r} \cdot q(6+S)zh_S^{\alpha_1}\right]^{\frac{1}{1-\alpha_2}} = w \cdot \left(h_S + (1 - \alpha_W)\left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S)z\right]^{\frac{1}{1-\alpha_W}}\right)$$
(39)

$$h_S^{1-\alpha_1} \leq h_0^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left(1 - e^{-\frac{\alpha_2 rS}{1-\alpha_2}}\right) \cdot \left[\frac{\alpha_2 w}{r} \cdot q(6+S)h_S^{\alpha_1}\right]^{\frac{\alpha_2}{1-\alpha_2}} z^{\frac{1}{1-\alpha_2}} \quad (40)$$

with equality if $S > 0$. In an interior solution $S \in (0, R - 6)$, the age-dependent component of the value function at age 6, q_0 is determined by

$$q_0 = \frac{we^{-rS}}{r} \cdot q(6+S)h_S^{\alpha_1}. \tag{41}$$

2. *if $J > 0$,*

$$\epsilon + (1 - \alpha_2)\left(\alpha_2^{\alpha_2}we^{-rJ}\left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J)z\right]^{\frac{1}{1-\alpha_W}} \cdot h_S^{\alpha_1 - \alpha_W}\right)^{\frac{1}{1-\alpha_2}}$$

$$= we^{-rJ}\left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J)z\right]^{\frac{1}{1-\alpha_W}} \tag{42}$$

$$h_S^{1-\alpha_1} \leq h_0^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left(1 - e^{-\frac{\alpha_2 rS}{1-\alpha_2}}\right) \tag{43}$$

$$\cdot \left(\alpha_2 we^{-rJ}\left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J)z^{\alpha_W}\right]^{\frac{1}{1-\alpha_W}} h_S^{\alpha_1 - \alpha_W}\right)^{\frac{\alpha_2}{1-\alpha_2}} \cdot z^{\frac{1}{1-\alpha_2}}$$

with equality if $S > 0$. In an interior solution $S \in (0, R - 6)$, the age-dependent component of the value function at age 6, q_0 is determined by

$$q_0 = we^{-r(S+J)} \cdot \left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J)z^{\alpha_W}\right]^{\frac{1}{1-\alpha_W}} \cdot h_S^{\alpha_1 - \alpha_W}. \tag{44}$$

Proof. Suppose $S \in (0, R - 6)$. The value matching and smooth pasting conditions when $J = 0$

45

are, respectively,

$$\epsilon - m(6+S) + e^{rS}q_0 z m(6+S)^{\alpha_2} = w h_S\left[1 - n(6+S)\right] + \frac{w}{r}\cdot q(6+S)z\left[n(6+S)h_S\right]^{\alpha_W}$$

$$e^{rS}q_0 h_S^{-\alpha_1} = \frac{w}{r}\cdot q(6+S).$$

Hence (41) follows from the smooth pasting condition. Likewise, (39) follow from plugging $n(6+S)$, $m(6+S)$ from Lemmas 2 and 4 and q_0 from (41) in the value matching condition. Lastly, (40) merely states that the optimal h_S must be consistent with optimal accumulation in the schooling phase, $h(6+S)$.

The LHS of the value matching and smooth pasting conditions when $J > 0$ are identical to when $J = 0$, and only the RHS changes:

$$\epsilon - m(6+S) + e^{rS}q_0 z m(6+S)^{\alpha_2} = q_S z$$

$$e^{rS}q_0 h_S^{-\alpha_1} = q_S h_S^{-\alpha_W}.$$

Hence (44) follows from plugging q_S and h_S from (35)-(36) in the smooth pasting condition. Likewise, (42) follow from plugging $n(6+S) = 1$, $m(6+S)$ from Lemma 2, and q_0 from (44) in the value matching condition. Again, (43) requires consistency between h_S and $h(6+S)$. □

For each case where we assume $J = 0$ or $J > 0$, it must also be the case that condition (38) does not or does hold.

C Identifying λ

COROLLARY 3 *If we select a subsample of individuals whose parents have the same $S_{Pi} = \hat{S}_P$, then for all positive experience levels $x > 0$, the relationship between earnings and S satisfies:*

$$e_{i,x}(S_{Pi} = \hat{S}_P) \propto \left[C_1(S_i) + \tilde{C}_2(x, S_i)\right] \cdot F(S_i)^{\frac{1}{(1-\alpha)\left[\left[1-\lambda(1-\alpha)\right]\right]}}, \tag{45}$$

where $\tilde{C}_2(x,s) \equiv C_2(x+6+s,s)$. So the observations $(e_{i,x}^*, S_i, S_{Pi})$ satisfy the regression equation

$$\log e_{i,x}(S_{Pi} = \hat{S}_P) = a_0 + a_1 \log\left[C_1(S_i) + \tilde{C}_2(x; S_i)\right] + a_2 \log F(S_i) + u_i \tag{46}$$

with

$$a_2 = 1/\left\{(1-\alpha)\left[1 - \lambda(1-\alpha)\right]\right\}.$$

Proof. Since (6) holds with equality when $S > 0$, for a subsample of children such that $S_{Pi} = \hat{S}_P$, it

must hold across the subsample that

$$z_i \propto F(S_i)^{\frac{1}{1-\lambda(1-\alpha)}}.$$

Plugging this in (10) yields (45). $\qquad\square$

Equation (46) implies that among children whose *parents* attained the same years of schooling, (log-)earnings profiles are parallel (since $\tilde{C}_2$ does not vary across individuals). More importantly, the gaps between the profiles are explained only by own schooling. But since (α_1, α_2) and consequently $(C_1, \tilde{C}_2)$ can be recovered from the argument following Corollary 1, the gaps can be used to identify the magnitude of λ.

For children with identical S_P, earnings differences cannot be attributed to causal effects from parents. So the only way that schooling can have heterogeneous effects on earnings of individuals whose parents have the same S_P is through z's influence on $F(S)$, which reveals λ.

Admittedly, such a regression is less satisfactory than Corollaries 1-2, as both regressors in (46) are functions of own schooling S_i. Since $(C_1, \tilde{C}_2, F)$ are all close to exponential in schooling S, we cannot separately recover λ from regressions as in Section 3 without knowledge of (α_1, α_2). This is another motivation for estimating the model structurally, in which we find that the estimated effect of λ is close to zero.

D   Numerical Algorithm

For the purposes of our estimated model in which S is fixed, the solution method in Appendix B is straightforward. We need not worry about value-matching conditions and only need to solve the smooth-pasting conditions given S, which are equations (40) and (43), to obtain h_S. Note that there is always a solution to (40) or (43)—i.e., we can always define a function $h_S(S)$ as a function of S. This is seen by you rearranging the equations as (bold-face for emphasis)

$$1 = \left(\frac{h_0}{\mathbf{h_S}}\right)^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left(1 - e^{-\frac{\alpha_2 rS}{1-\alpha_2}}\right) \cdot \left[\frac{\alpha_2 w}{r} \cdot q(6+S)\right]^{\frac{\alpha_2}{1-\alpha_2}} z^{\frac{1}{1-\alpha_2}} \cdot \mathbf{h_S}^{-\frac{1-\alpha}{1-\alpha_2}} \quad (47)$$

$$1 = \left(\frac{h_0}{\mathbf{h_S}}\right)^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left(1 - e^{-\frac{\alpha_2 rS}{1-\alpha_2}}\right) \qquad (48)$$

$$\cdot \left(\alpha_2 w e^{-rJ} \left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J)z^{\alpha_W}\right]^{\frac{1}{1-\alpha_W}}\right)^{\frac{\alpha_2}{1-\alpha_2}} \cdot z^{\frac{1}{1-\alpha_2}} \cdot \mathbf{h_S}^{-\frac{1-\alpha+\alpha_2 \alpha_W}{1-\alpha_2}},$$

respectively. Hence, for any given value of S, both RHS's begin at or above 1 at $h_S = h_0$, goes to 0 as $h_S \to \infty$, and is strictly decreasing in h_S. The solution $h_S(S)$ to both (47) and (48) are such that

1. $h_S = h_0$ when $S = 0$ or $S + J = R - 6$

2. $h_S(S)$ is hump-shaped in S (i.e., there $\exists S$ s.t. h_S reaches a maximum).

The rest of the model can be solved by Lemmas 2 and 4, and we can use Lemma 4 to determine J. Depending on whether condition (38) holds, we may have two solutions:

1. If only one solution satisfies (38), it is the solution.

2. If both satisfy (38), compare the two value functions at age 6 given S and candidate solutions $J_1 = 0$ and $J_2 > 0$ from Lemma 4 using the fact that the function D_W in (30) can be written

$$
D_W(6 + S + J)
$$
$$
= w\left(\frac{\alpha_W}{r}\right)^{\frac{\alpha_W}{1-\alpha_W}} \left\{ \int_{6+S+J}^{R} e^{-r(a-6-S-J)} \left[\int_{6+S+J}^{a} q(x)^{\frac{\alpha_W}{1-\alpha_W}} dx - \frac{\alpha_W}{r} \cdot q(a)^{\frac{1}{1-\alpha_W}} \right] da \right\} \cdot z^{\frac{1}{1-\alpha_W}}
$$

and

$$
V(S; 6, h_0) = \int_6^{6+S} e^{-r(a-6)} \left[\epsilon - m(a) \right] da + e^{-rS} V(6 + S, h_S)
$$
$$
= \frac{1 - e^{-rS}}{r} \cdot \epsilon - \frac{1 - \alpha_2}{r\alpha_2} \cdot (\alpha_2 z q_0)^{\frac{1}{1-\alpha_2}} \left(e^{\frac{r\alpha_2 S}{1-\alpha_2}} - 1 \right) + e^{-rS} V(6 + S, h_S).
$$

The candidate solution that yields the larger value is the solution.

Computing Model Moments   Given our distributional assumptions on mother's schooling, learning abilities and tastes for schooling, we can compute the exact model implied moments as follows. We set grids over h_P, z, and S, with $N_{h_P} = 17$, $N_z = 100$ and $N_S = 6$ nodes each.

1. Construct a grid over all observed levels of S_P in the data. This varies from 0 to 16 with mean 9.26 and standard deviation 3.52. Save the p.m.f. of S_P to use as sampling weights.

2. Assuming $\beta = 0.06$, construct the h_P-grid which is just a transformation of the S_P-grid according to (7).

3. For each node on the h_P-grid, construct z-grids according to (14), according to Kennan (2006). This results in a total of $N_{h_P} \times N_z$ nodes and probability weights, where for each h_P node we have a discretized normal distribution.

4. For each (h_P, z) compute the pecuniary of choosing $S \in \{8, 10, 12, 14, 16, 18\}$ (solve for $V(S; 6, h_0)$ according to the above) and compute the fraction of individuals choosing each schooling level using the CCP's in (16)-(17).

All moments are computed by aggregating over the $N_{h_P} \times N_z \times N_S$ grids using the product of the empirical p.m.f. of h_P, the discretized normal p.d.f. of z, and CCP's of S as sampling weights.

E Formal Description of Experiments in Section 5

Formally, for any initial condition $x = (S_P, z, \xi)$, the model implied schooling and age-a earnings outcomes can be written as functions of x, $S = S(x)$, $E(a) = \tilde{E}(x; S; a)$.[36] Then schooling following a j-year increase in S_P, holding (z, ξ) constant, is

$$S_\nu^j(x) \equiv S(S_P + j, z, \xi). \tag{49}$$

Age a earnings following a j-year increase in S_P, holding (z, ξ) and S constant, is

$$E_0^j(x; a) \equiv \tilde{E}(S_P + j, z, \xi; S; a)|_{S = S(S_P, z, \xi)} \tag{50}$$

i.e., the schooling choice is fixed as if mother's education is S_P, but the earnings outcome, or amount of human capital accumulated, is computed assuming that mother's education is $S_P + j$. This captures the spillover effect that is independent of quantity (schooling) adjustment. Now if we define

$$E(x; a) \equiv \tilde{E}(S_P, z, \xi; S; a)|_{S = S(S_P, z, \xi)},$$

i.e. the earnings outcome when both the amount of human capital accumulation and the schooling choice are computed from the same level of S_P, we can write the total spillover effect as

$$E_\nu^j(x; a) \equiv \tilde{E}(S_P + j, z, \xi; a), \tag{51}$$

which also includes the substitution effect between the length and quality of schooling. Selection on abilities and tastes associated with a j-year increase in S_P can be written as

$$\Delta_z^j \equiv \exp\left[(\rho_{zh_P}\sigma_z/\sigma_{h_P}) \cdot \beta j\right]$$
$$\Delta_\xi^j(S_P) \equiv \left\{\Delta_\xi^j(S; S_P)\right\}_S \equiv \left\{\delta_S \gamma_{h_P} \exp(\beta S_P)\left[\exp(\beta j) - 1\right]\right\}_S,$$

respectively, where $\Delta_\xi^j(S_P)$ is a 6-dimensional vector for each level of schooling $S \in \{8, \ldots, 18\}$. The first expression follows since $(S_P, \log z)$ are joint-normal, and the second from the definition of tastes in (15). Then schooling and age a earnings following a j-year increase in S_P, including its correlation with z or ξ, are

$$S_z^j(x) \equiv S(S_P + j, z \cdot \Delta_z^j, \xi), \qquad E_z^j(x; a) \equiv E(S_P + j, z \cdot \Delta_z^j, \xi; a), \tag{52}$$
$$S_\xi^j(x) \equiv S(S_P + j, z, \xi + \Delta_\xi^j(S_P)), \qquad E_\xi^j(x; a) \equiv E(S_P + j, z, \xi + \Delta_\xi^j(S_P); a). \tag{53}$$

---

[36]Since although the parent variable in the initial condition is h_P, it is defined as $\log h_P = \beta S_P$ in (7).

Outcomes incorporating all spillover and correlation effects following a j-year increase are

$$S^j_{rf}(x) \equiv S(S_P + j, z \cdot \Delta^j_z, \xi + \Delta^j_\xi(S_P) + \Delta^j_{z\xi}(z)) \tag{54a}$$

$$E^j_{rf}(x;a) \equiv E(S_P + j, z \cdot \Delta^j_z, \xi + \Delta^j_\xi(S_P) + \Delta^j_{z\xi}(z);a), \tag{54b}$$

where $\Delta^j_{z\xi}(z) \equiv \left\{ \Delta^j_{z\xi}(S;z) \right\}_S = \left\{ \delta_S \gamma_{zz} \left[\Delta^j_z - 1 \right] \right\}_S$ is a compounded correlation effect on tastes that comes from (z,ξ) being correlated, even conditional on h_P. We coin this the "reduced form" effect since by construction,

$$\int S^j_{rf}(x)d\Phi(\hat{S}_P = S_P, \hat{z}, \hat{\xi}) = \int S(x)d\Phi(\hat{S}_P = S_P + j, \hat{z}, \hat{\xi}),$$

where Φ is the joint distribution over x, and $\hat{x}$ are dummies for integration.

The first row of Table 9 is obtained by integrating the change from $S(x)$ in (49) and (52)-(54) over the population distribution Φ, when $j = 1$. The second row is the outcome of

$$\log \left[\sum_{a=14}^{R} \left(\frac{1}{1+r} \right)^{a-14} \int E^1_k(x;a)d\Phi(x) \right] - \log \left[\sum_{a=14}^{R} \left(\frac{1}{1+r} \right)^{a-14} \int E(x;a)d\Phi(x) \right]$$

for $k \in \{0, v, z, \xi, rf\}$. Figure 3 is obtained by plotting

$$\log \left[\int E^1_k(x;a)d\Phi(x) \right] - \log \left[\int E(x;a)d\Phi(x) \right], \qquad \text{for } k \in \{0, v, z, \xi, rf\}.$$

The tables and figures in Section 5.2 are similarly computed by choosing the right change in j for each affected mother's cohort, and integrating over the affected population.

F   Tables and Figures not in text

	α	E slopes across S given E
HC Production	α_1	E slopes across S_P given E
	α_W	E slope controller
	ν	E levels across S_P
Spillovers	λ	E levels across S
	b	S level controller
	ρ_{zh_P}	Mincer coefficient β_2
Abilities	μ_z	E level controller
	σ_z	E level variation
	$\delta_S, \zeta_h, \zeta_c$	S levels
Tastes	γ_h, γ_z	S levels across S_P and E
	σ_{ξ}	S level variation

Table 11: Identification

(S_P, S, E) stand for mother's schooling, and the individuals' schooling and earnings levels, respectively. Taste heterogeneity picks up the residual unobserved heterogeneity not captured by the simple model.
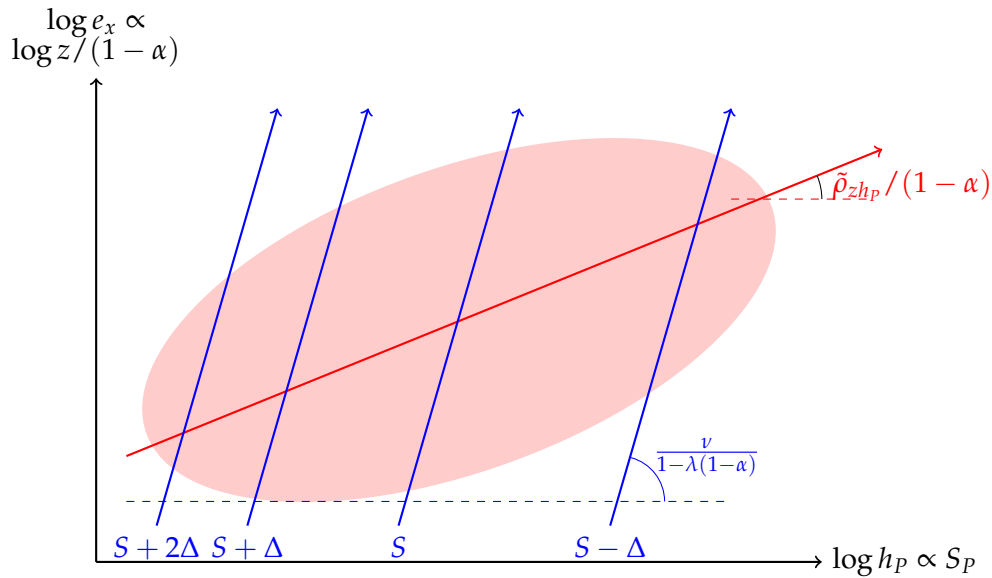
Figure 5: Intuition for Identifying Ability Selection and Spillovers.

x-axis: parents' human capital, or their schooling levels, y-axis: children's log abilities, or log earnings controlling for own schooling. The pink area represents the distribution of abilities and parents' schooling, (h_P, z). According to our model, selection is captured by the population correlation between parent's schooling and earnings, controlling for schooling. This is captured by the slope of the red line. Spillovers are captured by the relationship between earnings and abilities among children with the same level of schooling, which is captured by the slope of the blue lines. Note that conditional on abilities, schooling is decreasing in parents' schooling.
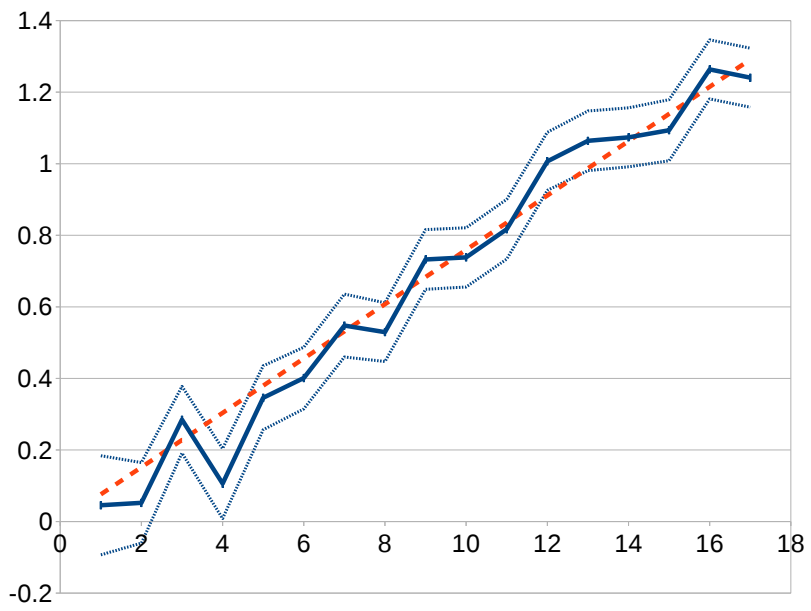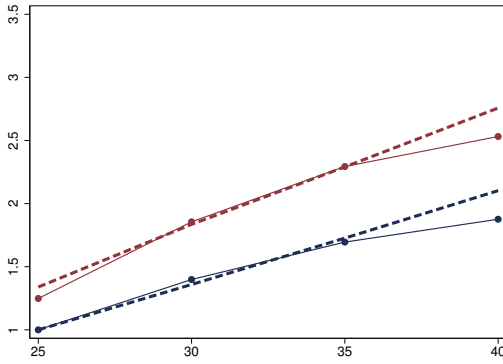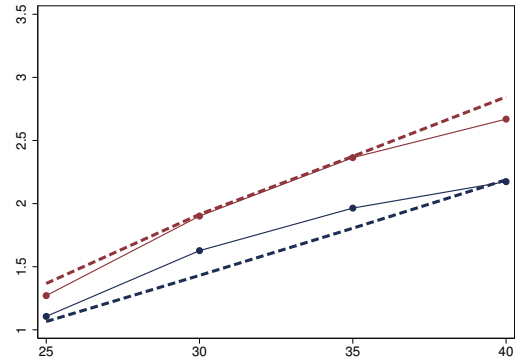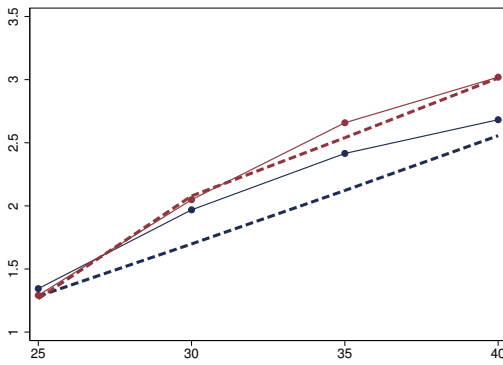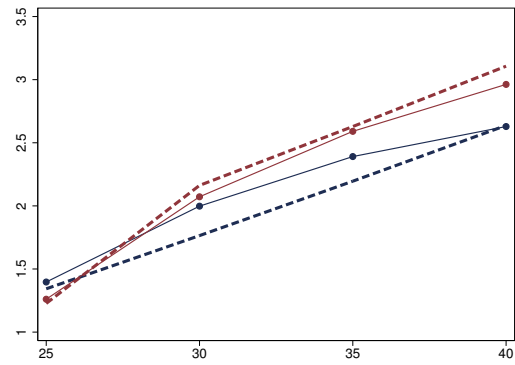
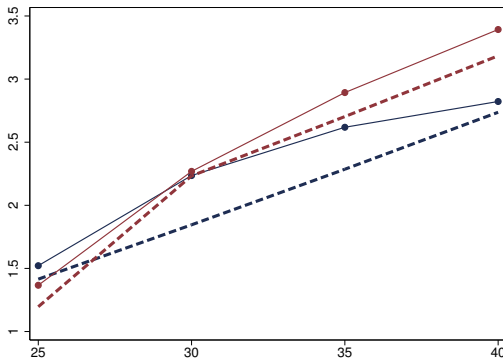Figure 6: Mincerian Return to Schooling, Linear vs. Dummies.

Figure 7: Model Fit

y-axis: normalized average earnings, *x*-axis: ages 25,30,35,40. Solid and dashed lines are, respectively, the data and model moments implied by the GMM parameter estimate values. The red lines on top correspond to individual's with $S < 12$ for the first row of plots, and $S \leq 12$ for the rest. The blue lines on the bottom correspond to the converse.

HC prod		Spillovers		"Ability"	
α_1	0.272 (0.001)	ν	0.739 (0.010)	ρ_{zh_P}	0.180 (0.004)
α_2	0.297 (0.005)	λ	0.001 (0.026)	μ_z	-1.225 (0.012)
α_W	0.391 (0.004)	b	0.930 (0.026)	σ_z	0.082 (0.001)
Tastes		Taste levels			
σ_ξ	0.612 (0.017)	ζ_h	0.266 (0.007)	δ_{14}	1.894 (0.008)
γ_{h_P}	1.551 (0.036)	ζ_c	0.846 (0.010)	δ_{16}	2.699 (0.021)
γ_z	0.549 (0.027)	δ_{12}	1.646 (0.004)	δ_{18}	4.198 (0.021)

*Standard errors in parentheses.

Table 12: Parameter Estimates

Figure 8: Mincerian Return to Schooling, Linear vs. Dummies.

(a) $S_P < 5$

(b) $S_P \in \{6,7\}$

(c) $S_P = 8$

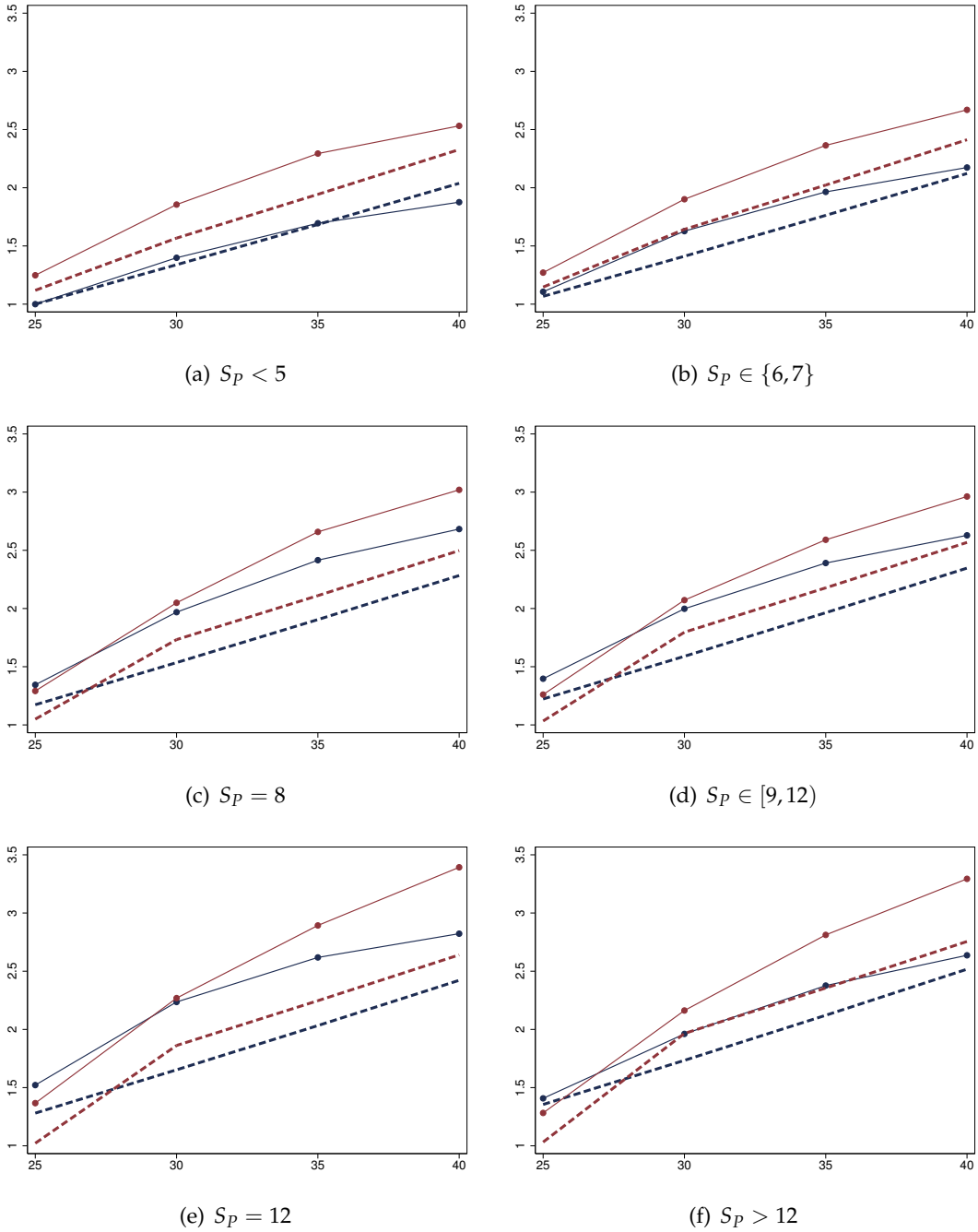(d) $S_P \in [9,12)$

(e) $S_P = 12$

(f) $S_P > 12$

Figure 9: Model Fit with Optimal Weighting Matrix

y-axis: normalized average earnings, x-axis: ages 25,30,35,40. Solid and dashed lines are, respectively, the data and model moments implied by the GMM parameter estimate values. The red lines on top correspond to individual's with $S < 12$ for the first row of plots, and $S \le 12$ for the rest. The blue lines on the bottom correspond to the converse.