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

Sang Yoon (Tim) Lee — QMUL, TSE & CEPR

Nicolas Roys — Royal Holloway

Ananth Seshadri — UW-Madison & NBER

February 2024

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 length of schooling, whose parents have different lengths of schooling. The model also features heterogeneous preferences 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 preferences 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.

*We thank Antonio Ciccone, Mariacristina De Nardi, Steven Durlauf, Eric French and Chris Taber for very helpful conversations. The paper also benefited from conference and seminar participants at the 2015 ASSA and EEA meetings, University of Cyprus, Essex, and TSE. Junjie Guo and Wei Song provided outstanding research assistance.

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 posit a standard life-cycle model of human capital accumulation à la Ben-Porath (1967, henceforth BP), which encompasses both schooling and learning on-the-job. An individual's length of schooling and earnings profile is endogenously determined by his initial level of human capital at age 6, and his learning ability, which governs the amount of human capital he can accumulate in a unit of time and remains constant throughout life.[1] We extend this canonical model by assuming that an individual's initial level of human capital is itself a function of his learning ability and his parent's human capital. In addition, we allow for correlation between individuals' learning abilities and their parents' human capital, to capture the unobserved persistence of abilities over generations. Thus the effect of parents' human capital on children's outcomes from altering their initial level of human capital is causal, while any effect through its correlation with learning abilities reflects selection.

The key result from our model is that an individual's schooling is a function of his learning ability and his parent's human capital, while earnings is a function of his learning ability and schooling. That is, parents' human capital will affect how long their children stay in school, but conditional on the length of schooling, have no direct affect on their children's earnings. This allows us to separately identify parental spillovers from selection by jointly using information on children's schooling and earnings, in addition to parental background.[2]

We argue that our model can rationalize findings from the previous literature that find a positive OLS coefficient when regressing children's years of schooling on their parents', but no evidence of a causal effect (e.g. Behrman and Rosenzweig, 2002; Black et al., 2005). High human capital parents tend to have higher learning abilities, which are passed on to their children. Such children would then stay longer in school to reap the benefits. This generates a positive selection effect. But in BP, all else equal, individuals with higher levels of initial human capital spend less time in school since they can reach higher levels of human capital more quickly. So in our model, as long as higher human capital parents have children with higher levels of initial human capital, such children will spend less time in school—a somewhat surprising result. We refer to this as the "level effect": better initial conditions substitute the need for longer investments.

[1] For the purposes of this paper, "schooling" refers to "years of education." We will also refer to the parent as "she" and child as "he."

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

Our model suggests that the observed positive intergenerational schooling relationships indicates that selection effects are stronger than the level effect. And, it explains why IV estimates of intergenerational schooling find no effect. The negative level effect is key, a feature we do not artificially assume, but inherent in most BP life-cycle models of human capital accumulation (e.g. Heckman et al., 1998; Huggett et al., 2011).

Importantly, this does not mean that the causal effect of parents' human capital on children's *earnings* (i.e., the spillover effect) is negative. According to BP, earnings are only a function of schooling and own abilities, which is unobserved. And our model implies that schooling is determined by abilities and parents' human capital. So among children who attain the same years of schooling, abilities have a one-to-one relationship with parents' human capital. This relationship allows us to identify the magnitude of the spillover from how parents' human capital affects the earnings of *children who attain the same level of schooling*.

Reduced-form evidence from the Health and Retirement Survey (HRS) data reveals that children with the same years of schooling have parallel earning profiles with constant gaps, which are increasing in their moms' years of schooling. The gaps indicate that moms affect how much human capital a child accumulates before labor market entry, compared to another child with the same years of schooling. The fact that these earnings profiles are parallel suggests that spillovers no longer operate once children enter the labor market. And, the size of the gaps are similar whether we look at children who do or don't finish high school, or children who do or don't enroll in college. All such evidence is consistent with our model assumptions.

We then estimate an extended model by indirect inference. We use moms' schooling as a proxy for parents' human capital. We allow children's human capital production technology to differ depending on whether he is in school or is working, and also allow for heterogeneous, nonpecuniary preferences for schooling. Including preference heterogeneity not only helps replicate key moments in the data,[3] but it also prevents the overestimation of the effects of parents' human capital and children's learning abilities on children's schooling and earnings outcomes.

Preferences are identified by an exclusion restriction: abilities affect both schooling and earnings outcomes, whereas preferences exclusively affect schooling choices. Empirically, the reduced-form effect of moms' schooling on children's earnings is too small for the positive intergenerational schooling relationship to be driven solely by selection on abilities. Instead, our estimates suggest that a positive correlation between moms' schooling and children's preferences for schooling (selection on preferences) is the main determinant of the positive intergenerational schooling relationship. And it plays a much more significant role than selection on abilities. Importantly, the causal effect is still negative, due to the level effect.

Finally, we find that forcefully increasing all moms' schooling by a year has a 1.2% causal

[3] In the canonical BP model, if the human capital production technology were the same in the schooling and working phases, initial earnings would be zero—which is clearly at odds with the data.

boost on the lifetime earnings of children, even though the causal effect on schooling is negative. This effect is heterogeneous across individuals and also over the life-cycle, and smaller than our reduced-form estimate of 1.7%. We argue that this constitutes a lower-bound for parental spillovers, as the model only allows parents' human capital to affect children's initial level of human capital and suppresses other channels that may also be operational, such as parents' human capital having a causal impact on children's learning abilities.

Related literature A broad literature has studied the causal effect of parents on children's outcomes.[4] The common challenge for all these studies is to separately identify the unobserved correlation between parents' and their children's endowments (abilities, preferences) from the unobserved, causal impact of parental spillovers. Unfortunately, different approaches have led to a wide range of estimates (Black and Devereux, 2011). We contribute to this literature by incorporating insights from a human capital model of education and earnings.

But if we focus on the intergenerational transmission of education only, there is some consensus that parents' schooling has little causal impact on children's schooling, with the estimated effect being small or insignificant, or even negative (Holmlund et al., 2011). Some studies focus on teasing out the selection effect, such as Behrman and Taubman (1989) and Plug and Vijverberg (2003), who find strong, inherited genetic effects on children's schooling outcomes, and only limited room for causal effects. A larger literature has focused on measuring causal effects directly via means of an instrumental variable. Some solutions have been to employ special data on twin parents (Behrman and Rosenzweig, 2002), adopted vs. biological children (Plug, 2004) or to take advantage of compulsory schooling reforms during the parents' generation as a natural experiment (Black et al., 2005). Our structural model is explicit about the unobserved factors of learning abilities and preferences for schooling, and can rationalize the wide range of estimates found in the empirical literature.

Most of the existing literature attempts to control for selection on abilities. But parents can affect children's outcomes through non-pecuniary channels as well. 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). In the context of our model, non-pecuniary motives are captured by the heterogeneous preferences for schooling. While some preferences can be purely genetic, others can be affected by the environment. So the intergenerational transmission of preferences presents difficulties when estimating causal effects, but also provides some discipline on how to interpret results from special data sets.

Behrman and Rosenzweig (2002) find a quantitatively large, negative causal effect of moms' schooling on children's schooling, and conclude that the observed positive intergenerational school-

[4]See Black and Devereux (2011); Holmlund et al. (2011); Sacerdote (2011) for extensive surveys of this literature.

ing relationship is entirely driven by correlation between moms' schooling and unobserved factors. Black et al. (2005) find that the causal schooling relationship is not significantly different from zero. Through the lenses of our model, twin moms who have different levels of schooling (as in Behrman and Rosenzweig, 2002) likely differ less in abilities and preferences than adjacent cohorts before and after a schooling reform (as in Black et al., 2005). In particular, even though a reform is unlikely to cause a jump in abilities, it could cause a change in preferences. So in the case of twin moms, the child of the mom with more schooling should himself attain less schooling, due to the level effect. But in the case of a schooling reform, latter cohorts may develop a stronger preference for schooling that is passed on to their children, which can countervail the level effect.[5]

More generally, while our model posits that the relationship between parents' human capital and children's preferences should reflect selection, how much we can control for this endogeneity might be context-dependent. Specifically, an instrumental variable itself may directly affect preferences, or there could be an omitted variable that affects both the instrument and preferences. Thus, our findings can rationalize different findings in the literature to the extent to which any exogenous variation they exploit is independent of preferences.

These preferences 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). 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).[6] In our context, we simply lump such factors together into heterogeneous preferences for schooling. Of course, non-cognitive skills and traits can also be inherited through causal and non-causal transmission channels, but for now we take the stance that they are non-causal, and posit that some instruments used in the empirical literature are more likely to be correlated with such traits than others.

As another example, while adoptee studies find a causal effect of parents' schooling that is lower than the observed level of intergenerational schooling persistence, they still remain slightly positive (Sacerdote, 2002) or are close to zero (Plug, 2004). Children adopted by different families likely do develop the non-cognitive skills or acquire information that affect their preference for schooling. If these unobservables are positively correlated with parents' education, to some ex-

[5]The working paper version of Black et al. (2005) alludes to the possibility of confounding effects. Chevalier et al. (2013) present evidence that daughters of higher educated moms stay in school longer, controlling for moms' education by month of birth. In our interpretation then, younger moms who start school at an earlier age than their classmates may not only attain fewer years of schooling, but also pass on weaker preferences for schooling to their children.

[6]Blanden et al. (2007) find evidence that children's non-cognitive skills can explain some of the intergenerational persistence of income in Britain, while Groves (2005) attributes a chunk to the persistence of personalities.

tent it should be expected that the effect of parents' schooling on children's schooling should be larger in adoptee studies than in twin studies, although the latter in particular may be sensitive to sampling bias. Similarly, the schooling of moms and dads may also have different effects on the schooling of daughters and sons due to differences in how preferences are transmitted.[7]

Like schooling, earnings also exhibits strong intergenerational persistence (Solon, 1999; Chetty et al., 2014). But in contrast to the empirical literature on schooling, most studies find that non-genetic factors do have a positive causal effect on children's earnings (Björklund et al., 2006; Sacerdote, 2007). The fact that parental causal effects are small or non-existent for children's schooling but not for their earnings is sometimes viewed as an inconsistency (Black and Devereux, 2011). This seems to stem from an implicit understanding that parents should have a qualitatively similar effect on children's education and earnings outcomes, which is likely motivated by the strong, positive relationship between an individual's own education and earnings (Card, 1999).

This is likely also a reason the vast majority of the literature exclusively focuses on only the schooling-to-schooling, or earnings-to-earnings (or income) relationship. Some do discuss whether the effect of parents' earnings on children's schooling is similar to its effect on children's earnings (e.g. Chang et al., 2023), but little is known on the effect of *parents' schooling on children's earnings*.[8]

Our data does not allow us to differentiate between parents' education and earnings, so we proxy their earnings by their years of schooling. Instead, we focus on differentiating their effect on *their children's education and earnings* in a structural model in which both are jointly determined, rather than focusing on single outcomes one-by-one.[9] By considering two sets of outcome variables, we are able to isolate the size of intergenerational spillovers (causal) from selection. In our model, parents can have opposing effects on their children's schooling and earnings outcomes. Even though the model-implied causal effect is negative, we find that the overall relationship between parents' and children's schooling is positive due to strong selection on preferences. But higher initial human capital still implies higher earnings, leading us to conclude that parents have a positive causal effect on children's earnings despite the negative level effect on schooling.

The rest of the paper is organized as follows. Section 2 posits a model of human capital accumulation with spillovers, of which we solve a simpler version to derive analytical results and make empirical predictions. In Section 3 we describe the HRS data that we use and compare reduced-

[7]Our model is based only on the mom-son relationship. Fernández et al. (2004) and subsequent studies find evidence that supports a causal transmission of attitudes across generations, but none focus on exogenous changes in parents' education.

[8]The only paper we know that investigates the effect of *parents' education on children's earnings or incomes* is Connolly et al. (2021). They find a positively significant effect, but do not control for endogeneity or selection and do not make any claims of causality. Consistent with their findings, we also find that moms' education has a larger effect on children's earnings at lower levels of moms' education.

[9]In this sense, our study is also related to Bowles and Gintis (2002), which estimates how much of the intergenerational persistence in earnings can be explained by the correlation between dads' earnings and other variables, such as children's education.

form evidence against the simple model's predictions. Section 4 estimates the full structural model to the HRS. We show, quantitatively, that the estimated model inherits properties of the simpler version. Section 5 examines the main result of increasing moms' schooling by 1 year, and also the counterfactual result of a hypothetical compulsory schooling reform. Section 6 concludes.

2 Schooling and Earnings Model with Parental Spillovers

An individual begins life at age 6 as the child of a parent and retires at an exogenous age R.[10] His initial state contains his parent's human capital h_{Pi}, his learning ability z_i, and his preference for schooling $\xi_i(s)$. The latter is an individual-specific function of years of schooling s that represents individual i's non-pecuniary benefit from obtaining s years of schooling in present value terms. Credit markets are complete and we can thus focus on the income-maximization problem.[11]

An individual at age 6 chooses how long to stay in school s to maximize the present discounted sum of net income $W(h_{Pi}, z_i; s)$ plus the non-pecuniary benefits of schooling $\xi_i(s)$[12]

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

where S_i denotes individual i's optimal choice for schooling, and $W(\cdot)$ is defined by

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

subject to

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

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

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

where r is the interest rate, w the wage per unit of human capital, and ϕ a productivity parameter that is constant across all individuals.

The exponents (α_1, α_{1W}) and (α_2, α_{2W}) are the returns to time and goods 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). The coefficients on time n and human capital h are the same in line with the empirical evidence surveyed in Browning et al. (1999).[13] Tastes for school-

[10] Our empirical analysis and estimation will assume that the parent is the mom, for reasons explained below.

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

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

[13] For instance, Heckman et al. (1998) estimates this technology using NLSY79 and they are not able to reject this

ing only affect an individual's desire to remain (or not) in school while having no direct effect on earnings.

The causal effect of 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}.[14] This can be understood as parents' economic status affecting how much they invest in their children, the effect of which is governed by the parameter ν. Clearly, initial human capital h_{i0} may also be affected by one's own learning ability z_i, the extent of which is measured by λ.

2.1 Initial Conditions

Our model makes three departures from the canonical BP model. First, we allow for non-pecuniary preferences for schooling (equation 1a). Second, the returns to human capital investments are allowed to vary between the schooling and working phases (equation 1c). Third and most important, we assume that individuals' initial human capital depend on their own ability and their parents' human capital (equation 1e). We further assume that the population distribution of $(\log h_{Pi}, \log z_i)$ is joint normal:

$$\begin{bmatrix} \log h_{Pi} \\ \log z_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{h_P} \\ \mu_z \end{bmatrix}, \begin{bmatrix} \sigma^2_{h_P} & \rho_{zh_P}\sigma_h\sigma_z \\ \rho_{zh_P}\sigma_h\sigma_z & \sigma^2_z \end{bmatrix} \right). \tag{2}$$

where (μ_{h_P}, μ_z) are the population means and (σ_{h_P}, σ_z) the standard deviations of (h_P, z). In what follows we drop the individual subscript i unless necessary.

Given (1e) and (2), the theoretical OLS coefficient of projecting h_0 on h_P must satisfy:

$$\frac{\text{Cov}(\log h_0, \log h_P)}{\text{Var}(\log h_P)} = \underbrace{\nu}_{\text{causal}} + \underbrace{\rho_{zh_P} \cdot \frac{\sigma_z}{\sigma_{h_P}} \cdot \lambda}_{\text{selection}}.$$

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 selection term ρ_{zh_P} is non-zero when a parent's human capital (h_P) is a function of the parent's learning ability, which we assume is exogenously transmitted to her child (as z). In that case, z and h_P are both functions of the parent's learning ability and would thus be correlated, reflecting selection. The parameter λ

assumption for persons of different ability and education groups.

[14]One could also consider the parents' choice of when to have children, or investing in their own human capital to raise their own human capital when their children are age 6. But 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. If parents are the same age and maximize income that can be passed on to future children, *without* internalizing the spillover effect, life-cycle decisions would remain the same under perfect credit markets. Similarly, we are also abstracting away from the fact that children may consider how their human capital investment decisions affect their future children. Such a setup would require 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 can identify solely from the HRS data.

also reflects 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.[15]

In the estimation, we proxy h_P by moms' schooling S_P, so it is observed.[16] But since abilities are unobserved, we face the problem of separately identifying ν from ρ_{zh_P} and λ, i.e., how much a child's schooling or earnings are correlated with the parent's because of h_P directly affecting h_0, or indirectly through h_P's correlation with z. Simlilarly, we need to identify the relationship between h_P and children's preferences for schooling $\xi(s)$.

For now, we abstract from preferences for schooling and only include them later in Section 4, noting only that preferences would have a similar selection effect as abilities on children's schooling but the opposite effect on their earnings. In the remainder of this section, we show that the model without preferences for schooling gives us insights into the precise mechanisms at work, and how the spillover from ν can be separately identified from (λ, ρ_{zh_P}) using a panel of individuals' schooling and earnings, and their parents' schooling.

2.2 Model Solution Without Tastes for Schooling

Suppose that $\xi(s) = 0$, and solve the income maximization problem (1) subject to the constraints while also imposing $\alpha_{1W} = \alpha_1$ and $\alpha_{2W} = \alpha_2$. Then the model is identical to BP, with the only addition being that initial human capital is a function of z and h_P. An individual maximizes the present discounted sum 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 sum 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 having a human capital level of h at age a, while the right-hand side is the instantaneous

[15] We are assuming away a potential effect that parents' human capital can have on their children's learning abilities. But if this were the case, λ would pick up this 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.

[16] In other words, we assume that h_P and S_P have a one-to-one relationship. We have verified that allowing h_P to vary conditional on S_P has only a limited effect on our estimates. Results are available upon request.

return. The first order conditions for the controls are

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

$$m = \alpha_2 z(nh)^{\alpha_1} m^{\alpha_2} \cdot V_h, \tag{3b}$$

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)

$$r \cdot V_h - V_{ah} = w(1-n) + \frac{\alpha_1 z(nh)^{\alpha_1} m^{\alpha_2}}{h} \cdot V_h + z(nh)^{\alpha_1} m^{\alpha_2} \cdot V_{hh}, \tag{4}$$

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 equal the marginal return. Equations (3)-(4) 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 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

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

PROPOSITION 1: OPTIMAL SCHOOLING CHOICE *Define the function $F(s)$ as*

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

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

$$F'(S) > 0, \qquad F(S) \geq z^{1-\lambda(1-\alpha)} h_P^{-\nu(1-\alpha)} \quad \textit{with equality if } S > 0, \tag{5}$$

so F implicitly defines S as a function of (h_P, z).

Proof. See Appendix A. □

The higher the learning ability z of an individual, the higher his optimal choice of 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 are decreasing ($\alpha < 1$): Individuals with high initial values face low returns to schooling earlier, reducing their incentive to stay in school and instead enter the labor market earlier.

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*

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

Earnings at experience $x \in [0, R-6-S)$ is

$$e(x;S,z) = wh(x)\,[1-n(x)] = C(x;S) \cdot z^{\frac{1}{1-\alpha}}, \tag{6}$$

where

$$C(x;s) = r\kappa^{\frac{1}{1-\alpha}} \left\{ \int_0^x q(6+s+\hat{x})^{\frac{\alpha}{1-\alpha}} \left[1 + \frac{\alpha_1 e^{-r(R-6-s-\hat{x})}}{1-\alpha} \right] d\hat{x} \right\}. \tag{7}$$

Proof. See Appendix A. □

The above proposition tells us that, conditional on a child's length of schooling, the human capital level of a child during his working years is determined only by his 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.[17]

Equation (6) can be interpreted as a type of Mincer equation that relates earnings to schooling. Conditional on ability z, the slope of individuals' log-earnings experience profiles vary by length of schooling S according to the function C. Conversely, for individuals with the same S, experience profiles must be parallel across different z's, which is standard in models that use BP. But while Proposition 2 does rely on the fact that S subsumes the effect of initial human capital, it does not exploit the structural relationship that S is a function of (h_P, z) in Proposition 1. In the next subsection we show that this is key: Both propositions must be utilized for us to isolate the spillover effect.

[17] All else equal, for individuals wanting to enter the working phase with a certain amount of human capital, raising h_0 (quality) would reduce S (quantity), diluting its effect on earnings (the level effect). The reason why h_0 has *exactly* zero effect on earnings, though, is because we are imposing $(\alpha_1, \alpha_2) = (\alpha_{1W}, \alpha_{2W})$. Then the path of earnings must be continuous and differentiable before and after S. Continuity implies that the level of human capital at age S can be solely determined by the optimality conditions in (3) when $n = 1$ (Lemma 3 in the Appendix). Note that the benchmark BP model indeed assumes that the human capital production function remains constant throughout life. If the production function changes before and after S, there will be a dependence of earnings on h_0, but we verify in the estimated model that this does not affect the implications we derive from Proposition 2.

2.3 Identification of Spillovers

We will now demonstrate how to identify spillovers given a data sample of children's schooling and earnings outcomes, and the schooling levels of their parents. In addition, we will also argue that a positive intergenerational schooling relationship, a robust finding in empirical studies, implies that the spillover effect is dominated by selection. However, quantifying the exact magnitude of selection is less straightforward, which motivates the structural model in Section 4. There, we also find that selection on preferences rather than abilities is the dominating force, while the identification scheme for spillovers carries over. Nonetheless, we also provide a reduced-form approximation of ability selection in Appendix C. The appendix also proposes a reduced-form identification scheme for λ.

To isolate the spillover effect, we use Proposition 1 to 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):

$$h_P = \exp(\beta_P S_P) \quad \Leftrightarrow \quad \log h_P = \beta_P S_P. \tag{8}$$

where β_P is the statistical returns to schooling of the parent generation. This is only a proxy that transforms parents' schooling into human capital units, including all endogenous effects.[18] Then since we assumed (h_{Pi}, z_i) to be distributed joint lognormal in (2), we can write the statistical relationship

$$\log z_i = \mu_z + \rho_{zh_P} \cdot \frac{\sigma_z}{\sigma_{h_P}} \cdot (\log h_{Pi} - \mu_{h_P}) + \epsilon_i \equiv \tilde{\mu}_z + \tilde{\rho}_{zh_P} S_{Pi} + \epsilon_i, \tag{9}$$

where $\tilde{\rho}_{zh_P} \equiv \beta_P \cdot \rho_{zh_P} \sigma_z / \sigma_{h_P}$ measures the non-causal relationship between (z, h_{Pi}), and ϵ_i is distributed normal with mean zero, and independent of S_{Pi}. Applying this to (5) at equality,[19]

$$\begin{aligned} \log F(S_i) &= [1 - \lambda(1-\alpha)] \log z_i - (1-\alpha)\beta_P \nu S_{Pi} && \text{(10a)} \\ &= [\tilde{\rho}_{zh_P} - (1-\alpha)(\beta_P \nu + \lambda \tilde{\rho}_{zh_P})] \cdot S_{Pi} + [1 - \lambda(1-\alpha)] (\tilde{\mu}_z + \epsilon_i) \\ \frac{\log F(S_i)}{(1-\alpha)[1-\lambda(1-\alpha)]} &= \frac{\tilde{\mu}_z}{1-\alpha} + \Big[\underbrace{\frac{\tilde{\rho}_{zh_P}}{1-\alpha}}_{\equiv b_{\text{select}}} - \underbrace{\frac{\beta_P \nu}{1 - \lambda(1-\alpha)}}_{\equiv b_{\text{spill}}} \Big] \cdot S_{Pi} + \frac{\epsilon_i}{1-\alpha}. && \text{(10b)} \end{aligned}$$

where the function F is an increasing function of S (Proposition 3). This shows that the intergenerational schooling coefficient is positively related to the selection parameter $\tilde{\rho}_{zh_P}$ but negatively

[18] That is, we do not claim that β_P is a causal effect. This is a shortcut we take as we do not have data on parents' earnings, nor grandparents. As explained in Section 3, our estimation uses HRS AHEAD cohorts born approximately one generation before our HRS sample to indirectly infer parents' earnings, and we benchmark our results only against parents' schooling and not earnings. None of our results are results are sensitive to the value of β_P.

[19] Recall that a capital S denotes the optimal schooling choice, which is assumed to be what we observe in the data.

related to the spillover parameter ν under reasonable parameter values of (λ, α). Therefore a positive coefficient would indicate that the selection effect outweighs the spillover effect. In the richer model that we incorporate preferences for schooling in the next section, a positive correlation between preferences and h_P will also contribute to a positive coefficient.

Next, we jointly exploit the structural relationships in (5) and (6) to separately identify spillover effects. In what follows, suppose we have 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. But the earnings of an individual i with x years of experience, $e_{i,x}$, is only observed with error, so that $\log e^*_{i,x} = \log e_{i,x} + u_{i,x}$ where $u_{i,x}$ is i.i.d. across all (i, x).

Corollary 1: Identifying Spillovers. *Suppose that $(S_i, x) > 0$ for all i, and assume (8)-(9) hold. If we regress*

$$\log e^*_{i,x} = [\varphi_{S_i,x} + b_{S_i} S_{Pi}] + \tilde{u}_{i,x}. \tag{11}$$

where $\{\varphi_{s,x}\}$ is a full set of schooling$\times$experience effects, and $\tilde{u}_{i,x}$ regression error, the estimates

$$\hat{b}_s = \bar{b} = b_{spill} = \beta_P \nu / \left[1 - \lambda(1-\alpha)\right].$$

Proof. Proposition 1 implies that, among children with $S_i = \hat{S} > 0$, it must hold that

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

where the equality follows from (8). Applying this and $S_i = \hat{S}$ in (6) yields

$$\log e_{i,x}(\hat{S}) = \underbrace{\log C(x;\hat{S}) + \frac{\log F(\hat{S})}{(1-\alpha)[1-\lambda(1-\alpha)]}}_{\hat{\varphi}_{\hat{S},x}} + \frac{\beta_P \nu}{1-\lambda(1-\alpha)} S_{Pi} \tag{12}$$

among individuals with $S_i = \hat{S}$. Since the term in brackets in (11) apply only to individuals with the same level of schooling, the coefficients b_s are estimated as claimed, and the regression error only captures measurement error. □

Corollary 1 shows that holding own schooling levels fixed, the spillover is identified by the within-group variation in log-earnings and parents' schooling. Moreover, it implies that among children with the same length of schooling, differences in parents' schooling manifests itself as constant gaps across log-earnings profiles, which reveals the spillover.[20]

Why does the coefficient on parents' schooling in (11) reveal spillovers? According to Propo-

[20] Corollaries 2-3 in Appendix C show how b_{select} and λ can be identified in reduced-form.

sition 2, conditional on S and potential experience, children's earnings are only a function of z. While S itself is endogenous to z (equation 10a), holding S fixed controls for the selection effect, which is exactly what is being done in (11). Then the regression explicitly reveals the structural relationship between (h_P, z) *within* each schooling subgroup given in (10a), rather than the population (h_P, z) relationship in (9). Our model also implies that within-subgroup relationships should be the same across groups, which we empirically verify in the next section.

We stress that while the propositions and corollaries of course depend on our model, the identification scheme is more general than that. The two identifying assumptions are that $S = S(S_P, z)$: Children's schooling is a function of their abilities and parents' schooling; and $e = e(x; S, z) = e[x; S(S_P, z), z]$: earnings is a function of their abilities and only their *own* schooling. That is, parents directly influence children's schooling but not their earnings. The earnings effect is only indirect through how children's earnings are affected by children's schooling.[21]

In the next section, we run the proposed regressions in Corollary 1 using HRS data. We will show that earnings profiles are close to parallel with gaps determined by own and parents' schooling, and present some raw evidence on the magnitude of b_{spill}. Then in Section 4, we estimate a generalized model that includes unobserved heterogeneity in preferences for schooling.

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 were conducted every two years since 1992. As the HRS 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 our purposes. 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 our model in which large schooling variations are important for identifying causal effects.[22] In particular, we can directly exploit information on the length of schooling as implied by our model, rather than

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

[22] 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).

imputing educational attainment based on whether individuals gain certain qualifications (such as a college degree). 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.[23]

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.[24] We further drop 646 individuals with missing information on their own education or mom's years of schooling. This leaves us with 5,114 individuals. Table 1 describes this sample by level of education. Average schooling 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 Security 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[25]

$$
\begin{aligned}
\log e^*_{i,0} &= X'_{i,t}\beta_0 + \varepsilon_{i,0} \\
\log e^*_{i,t} &= \rho \log e^*_{i,t-1} + X'_{i,t}\beta_x + \varepsilon_{i,t}, \qquad t \in \{1, 2, ..., T\} \\
\varepsilon_{i,t} &= \alpha_i + u_{i,t}
\end{aligned}
$$

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, as well as 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.

[23] 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 have longer earnings histories as well as richer parental background information, such as NLSY79, but these individuals face compulsory schooling regulations and a rising college premium.

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

[25] Social security earnings records exceeding the maximum level subject to social security taxes were top-coded in the years 1951 through 1977.

	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

3.2 Visualizing Spillovers

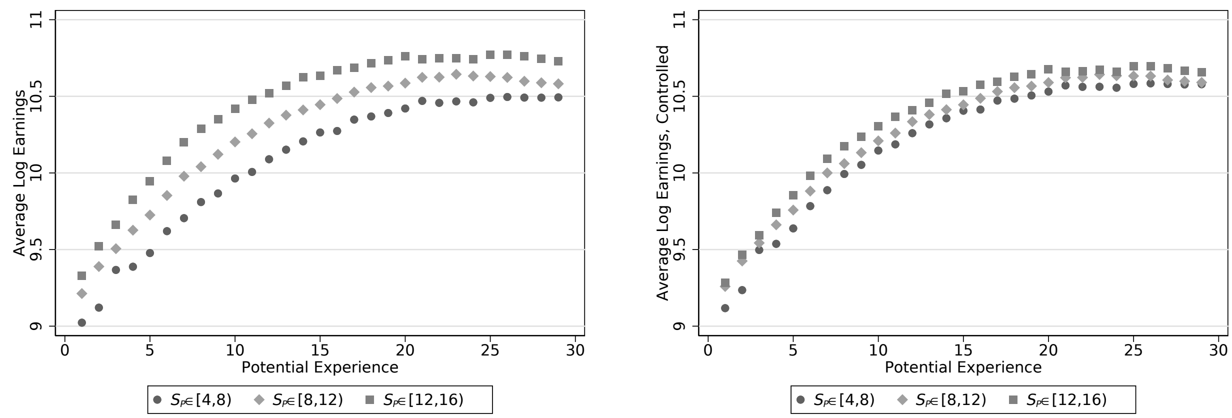
According to our model, the spillover is subsumed in the choice of schooling (Proposition 1). For individuals with the same level of schooling, earnings profiles should be parallel with a constant gap. Spillovers are identified from the within group variation in parents' schooling among individuals with identical levels of schooling (Corollary 1).

In Figure 1(a), we divide individuals into 3 subsamples, depending on whether their moms attained $S_P \in [4,8)$, $[8,12)$, or $[12,16)$ years of schooling. These correspond approximately to primary, secondary, and tertiary education. Average earnings profiles are close to parallel by mom's schooling, suggesting that potential experience effects do not vary much across schooling levels.[26,27] In Figure 1(b), we control for own schooling-specific experience effects, corresponding to the estimates for $\hat{\varphi}_{s,x}$ in Corollary 1. Since schooling is correlated across generations, the gaps narrow but remain close to parallel.[28] The corollary also tells us that the gap between these controlled

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

[27]It could also mean that own schooling levels are evenly distributed across moms' schooling levels, but this is not the case in the data as shown in Table 5.

[28]In the model, the slope of earnings profiles is determined by $\log C(x; S)$, so controlling for S, the slopes should be



(a) Children of mom's with [4,8), [8,12), and [12,16) years of schooling.

(b) Children of mom's with [4,8), [8,12), and [12,16) years of schooling, controlling for own schooling-specific experience effects.

Figure 1: Experience-Earnings Profiles by Mothers' Schooling

Earnings profiles of children by different levels of moms' schooling: 1924-1941 birth cohort. The y-axis is average log annual earnings in 2008 USD. Mothers' schooling levels are divided into 4 to 7 years, 8 to 11 years, and 12 or more years.

profiles is the spillover effect *if* it is constant across subsamples of children with the same levels of schooling.

So in Figure 2, we compare earnings profiles across moms' schooling levels by groups with similar levels of own schooling. Individuals are categorized into 4 subsamples depending on whether they have $S \in [8, 12)$, exactly 12, $(12, 16]$, or more than 16 years of schooling. Each group is further divided according to whether the mom has more than 8 years of schooling, which is about half of the moms in our sample.[29] Figure 2(a) depicts the average log-earnings profiles of individuals with $[8, 12)$ and those with exactly 12 years of schooling, and Figure 2(b) individuals with $(12, 16]$ and those with 17 or more years of schooling, respectively.

For all four education levels, the average log-earnings profiles of children with the same schooling but different mom's schooling are nearly parallel with a constant gap, as predicted by Corollary 1.[30] This points to a permanent parental effect that persists throughout an individual's career, which is precisely what the spillover intends to capture.[31] Moreover, the magnitude of the gaps are similar across the four categories of children's education, as implied by the corollary.

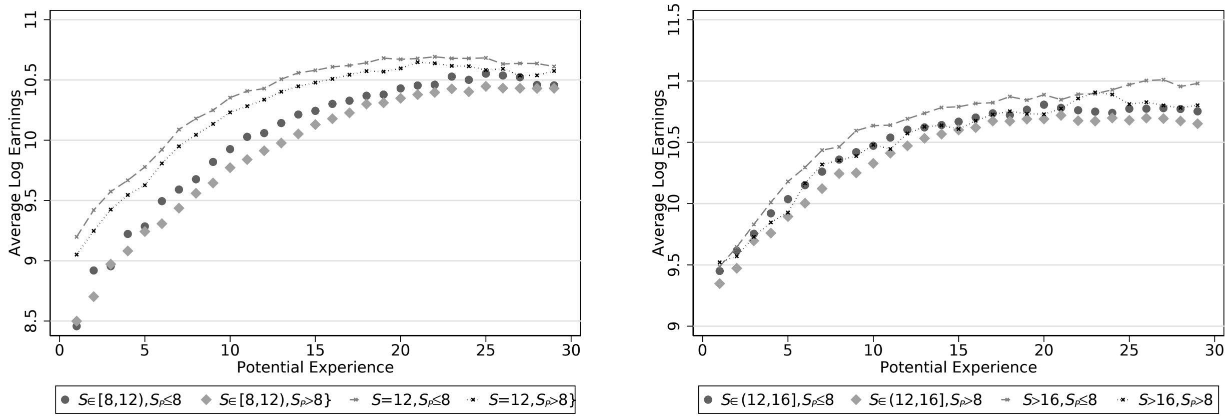
Given this visual evidence, we next obtain the reduced-form magnitudes of the spillover effect b_{spill} presented in Figures 1-2. To capture the spillover, we run different versions of a Mincerian

identical. See Appendix Figure F.6 that plots the gaps between the profiles in Figure 1(b). Parallelism fails at low and high experiences, especially for low S_P's, but holds from 5 to 25 years of potential experience.

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

[30] Appendix Figure F.7 plots the gaps between the profiles in Figure 2. Just like the average profiles, parallelism fails at low and high experiences for schooling-specific profiles, but holds from 5 to 25 years of potential experience.

[31] To be precise, the gaps should capture $b_s = \bar{b} = b_{\text{spill}} = \beta_P \cdot \nu / [1 - \lambda(1 - \alpha)]$ in Corollary 1.



(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 moms' schooling: 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.

regression:

$$\log e_{i,x} = f(S_i, x) + bS_{Pi} + \epsilon_{i,x} \tag{13}$$

where $f(\cdot)$ is a function of both schooling and potential experience which we specify in various different ways below, and $\epsilon_{i,x}$ an error term. We estimate different versions of (13) for earnings data from ages 23 to 42, and tabulate the results in Table 2.[32,33] In Table 3, we also allow the estimates of b to vary by own-schooling, S_i.

We consider four measures for S_{Pi}: the mom's and dad'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 moms have a larger influence on children (Behrman and Rosenzweig, 2002; Del Boca et al., 2012), which is also true in our sample as shown in Table 2.

Column (1) is a standard Mincer regression with a linear coefficient for own schooling, and a full set of potential experience effects (but not interacted). 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 in the last century (Goldin and Katz, 2007).

[32] Although the estimates barely change if we use all 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, and 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.

[33] 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 preferences for schooling, so controlling for them here would make the reduced-form estimates incomparable with the structural estimates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)*	(8)*
Mom S_P		0.017	0.017			0.015	0.017	0.017
		(22.71)	(21.82)			(14.98)	(21.87)	(21.89)
Dad S_P				0.011		0.003		
				(15.03)		(3.53)		
Mom + Dad					0.008			
					(20.05)			
S	0.090	0.082					0.190	0.228
	(99.73)	(84.21)					(17.19)	(6.29)
x effects	o	o	o	o	o	o	x	x
S effects	x	x	o	o	o	o	x	x
(S,x) effects	x	x	o	o	o	o	x	x
3rd-order	x	x	x	x	x	x	o	x
4th-order	x	x	x	x	x	x	x	o
R^2	0.190	0.194	0.204	0.203	0.205	0.205	0.201	0.203

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. Columns (1)-(6) include a full set of potential experience (age-6-S) effects, and columns (3)-(6) also include a full set of own schooling$\times$experience effects. Columns (7)-(8) instead include 3rd and 4th order polynomials of (x, S). t-stats shown in parentheses.
$*$The coefficients shown for S is conditional on $S = 12$ years of schooling.

The returns slightly decrease to 8.2% when we include moms' years of schooling in column (2). The coefficient on moms' education is 1.7% and statistically significant. This suggests that being born to a mom with five additional years of schooling has about the same effect on earnings as would an additional year of own schooling.

Column (3) corresponds exactly to the regression in Corollary 1, which includes a full set of own schooling-specific experience effects. When we replace moms' schooling with dads' schooling in (4), the parent's coefficient is attenuated to 1.1% but still statistically significant. The parents' coefficient drops further when we let S_{Pi} equal the sum of the schooling of both parents in (5). If we include both separately as in (6), the coefficient on moms is slightly reduced from 1.7% to 1.5% while the coefficient on dads drops significantly from 1.1% to 0.3%. This implies that moms' schooling has dominant explanatory power, which we take as the proxy for parents' human capital in what follows. In columns (7)-(8), we drop all schooling and experience effects and instead control for schooling and experience using 3rd- and 4th-order polynomials.[34] The coefficient on S_P remains virtually the same regardless of how we control for experience.

The striking feature is that no matter how we control for own schooling and experience, the coefficients on mom's S_P are virtually identical and highly significant across all specifications. The coefficient is stable even when we re-ran all these regressions controlling for race and cohort dummies, and also for white males separately.

[34] The exact specification can be found in Appendix C, equation (44).

	$S < 8$	$S \in [8,12)$	$S = 12$	$S \in (12,16]$	$S > 16$	W.Avg
Mom S_P	0.017	0.021	0.022	0.020	0.016	0.020
	(6.40)	(12.21)	(16.40)	(13.93)	(7.09)	
x effects	o	o	o	o	o	
R^2	0.076	0.121	0.151	0.182	0.256	
Sample	6,116	16,961	29,727	26,256	10,833	

Table 3: Mincer Regressions by Own Schooling

OLS regressions of (log) earnings on moms' years of schooling. HRS initial cohort, males born 1924-1941, ages 23-42. All columns include a full set of potential experience (age-6-S) effects. t-stats shown in parentheses. The sample size is the number of earnings observations. The last column is the population weighted average of the coefficients.

To summarize, moms' schooling has a stronger relationship with sons' earnings than dads', and the estimated effect $\hat{b}$ 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 (Corollaries 1) captures a causal spillover effect to the extent they are identical across subgroups of children with similar levels of schooling.

In Table 3, we group individuals into schooling intervals rather than exact years (S_i less than 8, $\in [8, 12)$, exactly 12, $\in (12, 16]$, or more than 16), and for each group, regress earnings only on moms' schooling, controlling for experience effects.[35] The coefficient is significant across all groups at approximately 2%, although slightly lower for children with either very high or low education.[36] We also formally test for equality of the coefficients using an F-test, resulting in a test statistic equal to 1.53 with an associated p-value of 0.19, so we cannot reject that the coefficients are equal. This justifies our assumption of a constant ν that applies equally for all education groups (Corollary 1).

Motivation for a Structural Model In our sample, the intergenerational schooling regression coefficient is 0.46. Thus, (10b) implies that the selection effect must be larger than the spillover effect. In fact, in Appendix C we find that $b_{\text{select}} \approx 0.111$, almost 6 times larger than b_{spill}. However, this ignores that children's schooling and earnings are jointly determined endogenous outcomes, and simulating the model leads to a counterfactually high relationship between parents' schooling and children's earnings. Instead, in the next section we will find that it is positive selection on preferences rather than abilities that predominantly explains intergenerational schooling persistence and the ability selection effect is muted.

Moreover, to compare our model against the empirical literature on intergenerational causal effects, we must perform counterfactuals which necessitates consistent estimates of all the parameters of the model (α, φ, etc.) This is what we turn to next.

[35] The estimated S_P coefficient barely changes when we use a polynomial for experience rather than dummies.

[36] As shown in the table, the sample size of these two groups are smaller. Moreover, the years of schooling of both mom and child are top-coded at 17 years in the data.

4 Structural Estimation

Since schooling and earnings are only functions of abilities (z) and parents' schooling (S_P), the previous regressions may overestimate the role of ability selection and spillovers in the presence of other dimensions of unobserved heterogeneity, in particular non-pecuniary preferences for schooling. On the one hand, it could be that children of higher educated moms have stronger preferences for schooling. Then to explain a positive intergenerational schooling relationship in the context of (10b), it is necessary that the effect of ability selection is smaller, the spillover effect larger or a combination of both. On the other hand, since we also observe a positive coefficient of moms' schooling on children's earnings, one or both effects must be larger if children stay in school longer because of preferences rather than any pecuniary motive.

To quantitatively assess the relative importance of each, we now turn to estimating our generalized model. Then in Section 5, we will use these estimates to perform counterfactuals and to analyze the empirical results from the instrumental variables literature seeking to estimate the causal effect of parents' education on children's earnings. We make the following assumptions in program (1) to bring it closer to the data and incorporate preferences for schooling:

1. The choice of schooling is constrained to be discrete, consistent with the data.
2. Preferences for schooling, or non-pecuniary benefits, are modeled as nested logit.
3. On-the-job training 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, we can still characterize the solution (Appendix B) which can be used to solve the model numerically (details in Appendix D). And since Corollary 1 disciplines our choice of moments, we will also demonstrate that the underlying intuition for identification carries over to the generalized model.

4.1 Population Distribution Assumptions

Our HRS sample only reports parents' schooling, but does not contain any information on parents' human capital—namely earnings. So we first run a separate Mincer regression as in (8) on HRS AHEAD cohorts born between 1890-1923, approximately one generation older than the original HRS cohort, to obtain β_P. Then assuming that the parents' schooling-earnings relationship is identical to that of the AHEAD cohorts, we can use it to transform parents' schooling S_P into human capital units h_P before applying them to childrens' initial condition (1e).[37] But since β_P is a reduced-form parameter (it is not causal), we interpret all our results against S_P rather than h_P.

[37] If β_P were not included and parent's human capital is $\exp(S_P)$, the estimated parameter would be $\beta_P \nu$.

The empirical distribution of S_P in our HRS sample ranges in discrete years from 0 to 16, which we take as given. Then as in (2), we assume that the distribution of z conditional on h_P is

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

For all possible combinations 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).

Preferences for schooling depend on parent's human capital and ability according to

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

where $\tilde{\xi}_i \equiv \left[\tilde{\xi}_{S_i}\right]_{S_i \in \{10,12,14,16,18\}}$ is randomly distributed across the population. The constants γ_{h_p} and γ_z capture the correlations between ξ_i and (h_{Pi}, z_i), respectively.[38] We normalize $\delta_8 = 0$ and $\delta_{10} = 1$ and estimate $\delta_S, S \in \{12, 14, 16, 18\}$. We further assume that schooling decisions are nested depending on college-entry, i.e. preferences for $S_i \in \{8, 10, 12\}$ and $S_i \in \{14, 16, 18\}$ are modeled as nested logit. Thus the vector $\tilde{\xi}_i$ is drawn from a 6-dimensional, generalized extreme value distribution with c.d.f. G and scale parameter σ_ξ:

$$\begin{aligned} G(\tilde{\xi}_i) = \exp\Big\{ & -\left[\exp\left(-\tilde{\xi}_{8_i}/\sigma_\xi\zeta_h\right) + \exp\left(-\tilde{\xi}_{10_i}/\sigma_\xi\zeta_h\right) + \exp\left(-\tilde{\xi}_{12_i}/\sigma_\xi\zeta_h\right)\right]^{\zeta_h} \\ & -\left[\exp\left(-\tilde{\xi}_{14_i}/\sigma_\xi\zeta_c\right) + \exp\left(-\tilde{\xi}_{16_i}/\sigma_\xi\zeta_c\right) + \exp\left(-\tilde{\xi}_{18_i}/\sigma_\xi\zeta_c\right)\right]^{\zeta_c} \Big\} \end{aligned}$$

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

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

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

$$\Pr(S_i = 8) = \Pr(S_i = 8 | S_i \in \{8, 10, 12\}) \cdot \Pr(S_i \in \{8, 10, 12\}), \tag{16}$$

for which we can easily compute the conditional choice probabilities given $G(\cdot)$, and similarly for other possible choices of S. All details can be found in Appendix D.

[38]We have no formal theory of whether the correlation between parents' schooling and preferences is causal or not. But as we alluded to in the introduction, we take the default stance that they only reflect selection.

4.2 Generalized Method of Moments

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

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

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 preference 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.[39]

Parameters Set a Priori The wage w is normalized to 1 and 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).[40] Retirement age is fixed at 65.

The value of β_P is quite stable across HRS AHEAD cohorts, ranging from approximately 0.04 to 0.06 for men and 0.05 to 0.09 for women; we fix $\beta_P = 0.06$. 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 results are not sensitive to different values of β_P within this range.[41] Since $\log h_P = \beta_P S_P$, it follows that $\mu_{h_P} = \beta_P \mu_{S_P}$ and $\sigma_{h_P} = \beta_P \sigma_{S_P}$. We take the mean and variance of mom's schooling directly from their sample analogs in the data, which are $(\mu_{S_P}, \sigma_{S_P}) = (9.24, 3.60)$ and thus $(\mu_{h_P}, \sigma_{h_P}) = (0.55, 0.21)$.

Preferences for 8 and 10 years of schooling, δ_8 and δ_{10}, are not separately identified from ϕ and normalized to 0 and 1, respectively. 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 moms' 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 share of individuals attaining high or low levels of schooling by 6 levels of mom's schooling.[42] 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.

[39] 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.

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

[41] This regression also admits a constant term, which we ignore since it is not separately identified from ϕ in (1e).

[42] The empirical share of each of the 6 mom's schooling groups, as well as the average years of schooling for each subgroup with higher or lower education, are tabulated in Appendix Table F.1.

Parameter	Value	Description
w	1	Normalization; not separately identified from μ_z
r	5%	Interest rate; after-tax rate of return on capital in Poterba (1998)
R	65	Retirement age, exogenous
β_P	6%	Mincer return to schooling, HRS AHEAD cohorts
μ_{h_P}	0.55	Mean human capital of moms, from data (see text)
σ_{h_P}	0.22	Standard deviation of moms' human capital, from data (see text)
δ_8	0	Preference for 8 years of schooling, normalization
δ_{10}	1	Preference for 10 years of schooling, normalization

* See text for more details.

Table 4: Parameters Set a Priori

For each level of moms' schooling, 1 of the 2 educational attainment shares of the children are redundant (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 whose moms had 5 or less years of 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 earnings and schooling gradients in the data we may miss by targeting aggregated moments. In sum, we target 57 empirical moments with 18 model parameters.

Denote these target moments $\hat{\Psi}$. For an arbitrary value of θ_1, we compute the implied model 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.[43] 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

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

[43]The diagonal matrix is unable to fit the curvature of lifecycle earnings, as shown in Appendix Figure F.8. This is because the diagonal matrix ignores the age correlations of earnings. However we found that while the optimal weighting matrix does get the curvature right, it does not match relative earning gaps across schooling groups of moms and children, which is our main focus.

Mom's S_P	Child's S	Fraction (%)	Average Earnings at age 23-27	28-32	33-37	38-42
≤ 5	≤ 11	60.04	1.00	1.41	1.71	1.85
	≥ 12	39.96	1.26	1.91	2.32	2.51
6-7	≤ 11	42.48	1.10	1.66	1.98	2.11
	≥ 12	57.52	1.27	1.91	2.37	2.65
8	≤ 12	66.23	1.35	1.99	2.42	2.62
	≥ 13	33.77	1.31	2.12	2.71	2.99
9-11	≤ 12	63.73	1.37	1.98	2.38	2.57
	≥ 13	36.27	1.30	2.15	2.65	2.96
12	≤ 12	45.41	1.52	2.24	2.62	2.77
	≥ 13	54.59	1.36	2.29	2.87	3.34
≥ 13	≤ 12	19.56	1.39	1.96	2.33	2.49
	≥ 13	80.44	1.30	2.23	2.82	3.27
(S, S_P) correlation and OLS:			0.48	0.46		
Mincer coefficients on (S, S_P):			0.08	0.02		

Table 5: Targeted Empirical Moments

For moms with low S_P (the first four rows), we divide low/high educational attainment of children by whether or not he graduated from high school, while for the rest we divide by whether or not he advanced beyond high school. The share of each S_P-group, and the average years of schooling for each S-group are summarized in Appendix Table F.1. 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.

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. While all moments influence all parameters, our choice of moments is guided by intuition from the simpler model of how certain moments should have more influence on certain parameters. Here we provide a brief description with a more detailed explanation on key parameters in the following subsection.

According to Corollary 1, ν can be identified by the coefficient on mom's schooling in a earnings regression among subsamples of individuals with the same level of schooling. While this is distorted by the inclusion of preferences for schooling, the intuition should still hold conditional on such preferences. Thus ν can still be identified from earnings differences among children with similar levels of own schooling but different levels of moms' schooling, as tabulated in Table 5. Similarly, λ is identified from how earnings vary by own schooling levels among children whose moms have similar levels of schooling (Corollary 3). Conditional on preferences and given (ν, λ), ρ_{zh_P} determines the coefficient on mom's schooling in a earnings regression over the population (Corollary 2). As we explained in Section 2.3, the intuition for identification carries over from the simple model since our identifying assumption is that schooling S is determined by (h_P, z) and earnings only by (S, z), not the specifics of the model.

The human capital production technology parameters are identified by life-cycle earnings moments. Since α_W governs the speed of human capital growth in the working phase, it is identified by the average slope of experience-earnings profiles. On the other hand, the schooling parameters (α_1, α_2) are identified by how initial earnings and the subsequent slope of earnings profiles vary by own and moms' schooling levels. For high values of $\alpha \equiv \alpha_1 + \alpha_2$, individuals with high S will have higher early age earnings and thus flatter earnings profiles. Likewise for high values of α_1, individuals with high S_P will benefit more from higher age 6 human capital and thus have higher initial earnings, leading to flatter earnings profiles. The mean and standard deviation of abilities (μ_z, σ_z) is identified by the overall mean and variation of earnings.[44]

Since δ_8 and δ_{10} is normalized to 0 and 1, respectively, and schooling becomes less important at higher values of the level parameter ϕ, it determines the share of children who attain 8 versus 10 years of schooling. The remaining 4 preference parameters δ_S, $S \in \{12, 14, 16, 18\}$ account for the shares of the remaining 4 schooling levels. While we do not target schooling shares exactly, Table 5 includes 12 own schooling moments which are sufficient to identify 5 parameters, as well as the scale parameter σ_ξ and (γ_{h_P}, γ_z), which determines how educational attainment varies by moms' schooling and individual earnings. The inclusion of γ_{h_P} is particularly important: Without it, moms' effect on how long children stay in school would be purely pecuniary.

The parameters (ζ_h, ζ_c) captures variation in earnings by high school and by college, conditional on the overall variation implied by σ_z.

4.3 Interpreting the Parameters

Table 6 reports the 18 parameter estimates and their asymptotic standard errors. The model generated educational attainment shares are nearly identical to the data in the fourth column of Table 5, as expected. More importantly, the 4 additional gradient moments (lower panel of Table 5 and first panel in Table 8) are almost identical to the data as well. Earnings moments are compared with the data visually in Appendix Figure F.8.

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 BP technologies. Estimates surveyed by Browning et al. (1999) lie in the range 0.5 to 0.9. This may have to do with the

[44] Since we normalized the earnings of the first cell in Table 5 to 1, μ_z is in normalized units.

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)	ϕ	0.881 (0.026)	σ_z	0.117 (0.001)
Preferences					
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

fact that the HRS cohort lived in a period in which observed education returns were 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 preferences. Increasing mom's schooling by 1 year increases her child's initial human capital by approximately $\beta_P \nu \approx 5\%$. However, as we will show in Section 5, the causal effect of increasing mom's schooling on lifetime earnings is more modest, at 1.2%, which is lower than the 1.7% we found in Section 3.2, Table 2. This is because higher human capital early in life leads to less schooling and less human capital accumulation.

The astute reader might wonder if our large estimate arises from the assumption that parents only have a level effect, i.e., that h_P only has a causal effect on h_0 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 model of Section 2. To address this, we run several numerical simulations with the extended model finding that incorporating a slope spillover has little influence on all other parameter estimates except for ρ_{zh_P}. This suggests that adding a slope spillover would mainly reduce the estimated selection effects, so contrary to the estimate being large it can in fact be viewed as a conservative lower-bound of the spillover effect.

Selection on Learning Abilities The estimate for ρ_{zh_P} implies that on average, moms 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, this means that moms with 1 more year of schooling have children with $\rho_{zh_P}\sigma_z/\sigma_{S_P} \approx 0.7\%$ higher abilities. Combined with $\lambda = 0.06$, this implies that selection on abilities has almost no affect on initial human capital.

Instead, higher abilities sustain through life and raise both schooling and earnings. According to Proposition 2, the impact on earnings at all ages is $\Delta \log z/(1-\alpha_W) \approx 1.3\%$ for a 1 year difference in moms' schooling.[45] This is more or less similar to the selection effect on lifetime earnings, as we soon show in Section 5, and in the ballpark of a 2% Mincer coefficient on moms' schooling.

This is also why the estimated λ is small and almost insignificant, indicating that parents are much more important than children's learning abilities for early human capital formation. Given the estimated value of ρ_{zh_P}, a higher λ would lead to a larger overall effect of moms' schooling on lifetime earnings than observed in the data.

Following similar calculations, a standard deviation of 0.117 for z translates into a $\sigma_z/(1-\alpha_W) \approx 20.4\%$ standard deviation in log-earnings, once we control for schooling.

Preferences 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 preferences for schooling levels of high school and below, and some college and above, respectively. Hence, preferences for staying in high school are much more correlated than those for college, possible due to higher attrition.[46] The idiosyncratic component of non-pecuniary benefits has a standard deviation of 3,846 dollars.

To facilitate the interpretation of the preference 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 more influenced by moms' schooling than by children's learning abilities. Consequently, non-pecuniary benefits increase much more with moms' schooling S_P than with children's z, as reported in the middle and bottom panels of Table 7, respectively. This conforms to the notion that highly educated moms are also those that are more likely to provide a family environment that encourages their own children to advance further in education. This correlation between children's preferences for schooling (non-pecuniary benefits) and parents' schooling

[45] Although the proposition is based on the simple model, a first-order approximation of the earnings equation implied by the general model in Appendix B, Lemma 4 leads to the same expression.

[46] For example, college students face dropout risks and graduates may face stricter occupational requirements. We thank an anonymous referee for this insight.

	$S=8$	$S=10$	$S=12$	$S=14$	$S=16$	$S=18$
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 $S = 8$.

plays an important role in the following analyses.

The estimate for γ_z is small because abilities play a limited role explaining the intergenerational schooling relationship. Just like the point estimate for λ was small given the estimate of ρ_{zh_P}, γ_z must be small given σ_z, since otherwise the effect of own schooling on observed earnings would be too large.

4.4 Fit Analysis

Shutting down parameters To interpret the effect of spillovers and selection in light of our empirical analysis from Section 3, it is useful to see how gradient moments are affected when shutting down the key parameters. The second panel of Table 8 tabulates the changes in moments when we set the key parameters (ν, ρ_{zh_P}) to zero.[47] The third panel tabulates the changes in the same moments when we don't allow preferences for schooling to vary by h_P or z. Specifically, the rows

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

	Corr_S	OLS_S	S coeff.	S_P coeff.
Data	0.469	0.444	0.076	0.017
Model	0.494	0.419	0.076	0.018
Pecuniary				
$\nu = 0$	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
Preferences				
$\gamma_{h_P}\mu_{h_P}$	-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 preferences and (h_P, z). Corr_S and OLS_S denote, respectively, the correlation between S and S_P, and the OLS coefficient from regressing S on S_P. (S, S_P) coefficients correspond to the coefficients from the Mincer regression in column (2) of Table 2.

"$\gamma_{h_P}\mu_{h_P}$" and "$\gamma_z\mu_z$" denote the cases where we keep all else equal and set

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

The third row is when there is no variation in preferences across both (h_P, z). These exercises also show that the intuition for identifying spillovers based on the simple model in Section 2 carries over to the estimated model, as well as how selection on abilities and preferences are separately identified.

Both ν and ρ_{zh_P} have limited impact on the own schooling coefficient. The intuition for this can be found in the propositions in Section 2.2. According to Proposition 1, the initial human capital effect of both are absorbed into the choice for schooling, while the relationship between earnings and schooling in Proposition 2 remains unchanged. In similar vein, shutting down ν has a much smaller effect than ρ_{zh_P} on the moms' schooling coefficient. This is because ν affects earnings only through the choice of schooling, while ρ_{zh_P} has a direct effect. The addition of preference heterogeneity do not alter these results.

When $\nu = 0$, the level effect 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. But ability selection barely affects 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 (10b) in the simple model. This indicates that the persistence of schooling across generations is attributed to the heterogeneity in preferences for schooling, along with the opposing influence of ν.

The third panel confirms the role of preferences in explaining the persistence of schooling across generations. When preferences do not vary with moms' schooling (row $\gamma_{h_P}\mu_{h_P}$), children of high human capital parents (who tend to have higher levels of schooling) no longer prefer to remain in school longer, and both Corr_S and OLS_S become negative. In the absence of preference heterogeneity by moms' schooling, selection on abilities alone is not enough to counter the negative level effect on children's schooling.

At the same time, the Mincer coefficient on moms' schooling increases by a third: When differences in moms' schooling do not drive any differences in non-pecuniary benefits, its effect on pecuniary earnings becomes larger. Thus if we were to force the model to explain the intergenerational schooling relationship only through selection on abilities, the Mincer coefficient on moms' schooling would be too high in comparison to the data. Conversely, the observed coefficient is too low for the intergenerational persistence of schooling to be explained solely by selection on abilities.

Letting preference vary across abilities has little effect on all moments (row $\gamma_z\mu_z$). This is somewhat expected since γ_{h_P} is much larger than γ_z. Indeed, when preferences do not vary along either dimension, the resulting numbers are more or less identical to when it varies only with abilities (last row).

Comparison with reduced-form prediction Given the large point estimate of γ_{h_P}, children of higher educated moms like to stay in school longer. We can now evaluate the degree to which overlooking preferences in the basic model results in biased estimations of selection on abilities and spillover effects. Compare the model-predicted values of $(b_{\text{spill}}, b_{\text{select}})$ in equation (10b) to their reduced-form estimates, which were 0.017 and 0.111 in Table 2 and Appendix Table C.11, respectively. Their model-implied values are

$$
\begin{aligned}
b_{\text{select}} &= \beta_P \rho_{zh_P}\sigma_z/(1-\alpha)\sigma_{h_P} = \begin{cases} 0.019 & \text{if } \alpha = \alpha_1 + \alpha_2 \\ 0.013 & \text{if } \alpha = \alpha_W \end{cases} \\
b_{\text{spill}} &= \beta_P \nu / \left[1 - \lambda(1-\alpha)\right] \approx 0.048 \quad \text{for both cases.}
\end{aligned}
$$

Indeed, the implied ability selection coefficient is much smaller than its reduced-form estimate, while the implied spillover coefficient is larger. This occurs because the extended model allows for non-pecuniary benefits while the simple model (and thus Corollary 1) does not. And since choices based on non-pecuniary benefits come at the detriment of lifetime earnings,[48] to make up for this and explain a reduced-form spillover in the range of 2%, the values of ν and ρ_{zh_P} must be larger.

And if positive selection on preferences for schooling contribute to the intergenerational school-

[48] The effect of preferences and ν are the opposite for schooling and earnings, since a higher value for ν leads to less time in school but higher lifetime earnings.

ing relationship, the contribution of selection on abilities must be smaller. And thus the spillover effect on earnings must rise relative to the ability selection effect. If we run the same regression as Corollary 1 on simulated data from the model, we recover a coefficient of 1.4%, which is similar to the structural effect of 1.2% we find in Section 5 and less than the 1.7% we found in Section 3.2.

5 Counterfactual Experiments

We conduct two main experiments using the model estimates. First, we decompose the effects of a one-year increase in moms' schooling. Second, we implement a counterfactual compulsory schooling reform. Although the spillover parameter ν has only a small effect on the reduced-form coefficient on S_P in a Mincer regression, we find that a uniform one-year increase in moms' schooling leads to an average 1.2% increase in children's lifetime earnings controlling for selection on abilities and preferences for schooling. Further allowing for selection leads to an additional 1.3 percentage point increase. At the same time, the effect on children's years of schooling is small. This is explained by parents with higher human capital having a negative effect on children's schooling (Proposition 1) which is offset by the stronger preferences for schooling of children from higher human capital parents.

5.1 Decomposing Spillovers from Selection

Given their state $(h_P, z, \tilde{\xi})$ at age 6, we compute the change in children's schooling and earnings outcomes in response to a 1-year increase in their moms' schooling, which translates into an increase of β_P units of $\log h_P$.

We perform several experiments to control for spillovers, ability selection and preferences for schooling. In Table 9, column (1), we hold constant the individual's $(z, \tilde{\xi})$ and also the schooling choice S, which isolates the pure quality effect coming from higher S_P. In column (2), we still hold $(z, \tilde{\xi})$ constant, but let the individuals re-optimize their choice of S. Since the higher initial human capital substitutes for schooling, earnings further increases while schooling decreases. The combined effect of the pure quality increase and schooling choice adjustment is the spillover effect. In columns (3) and (4), we respectively allow z then $\tilde{\xi}$ to rise 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 moms' schooling with children's preferences for schooling, in addition to the spillover effect. Finally in column (5), we let both z and $\tilde{\xi}$ rise together, which we label a "reduced-form" effect—i.e., this is just comparing the average outcomes of children with S_P years of moms' schooling, to the average outcome of those with $S_P + 1$ years of moms' schooling. See Appendix E for a formal description of how the counterfactual outcomes are computed.

The first and second rows of Table 9 list the average effects of a 1 year increase in S_P on school-

	(1) fixed S	(2) spillover	(3) ability	(4) pref.	(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 Moms' Schooling.
Units in % are approximated by log-point differences. The second row denotes the change in the cross section average of the present discounted sum of lifetime earnings. Column (1) holds children's schooling constant, while column (2) holds abilities and preferences constant but let's schooling vary according to the model. Columns (3)-(4) let abilities or preferences also vary according to their estimated correlations with h_P. Column (5) is when we allow for both selection on abilities and correlation with preferences for schooling. The coefficients from an intergenerational schooling OLS in data and model are 0.444 and 0.419, respectively.

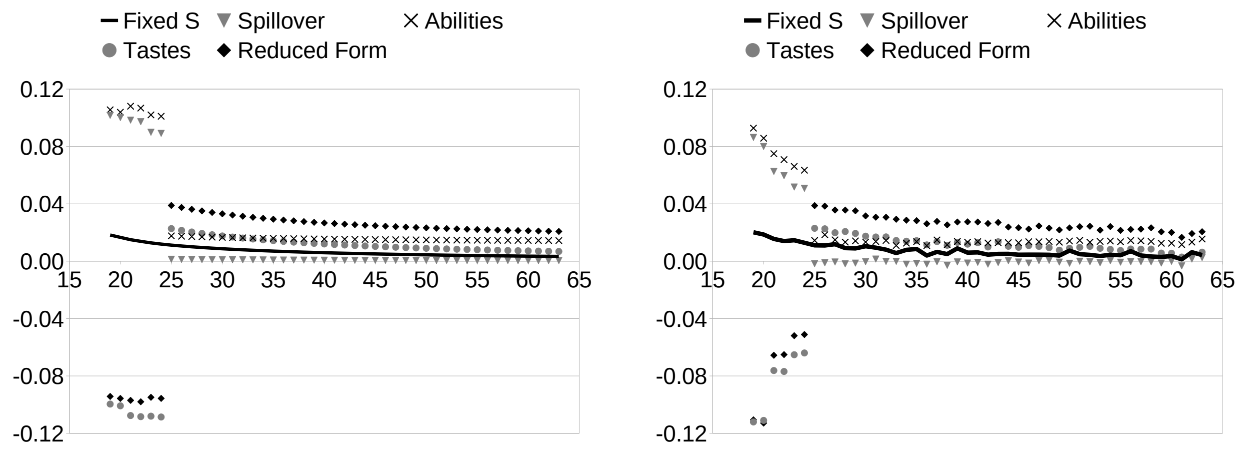
ing S and the present-discounted value of lifetime earnings, respectively. The third and fourth rows show the average change in earnings at ages 30 and 50. The change in S is in years and the change in earnings in log-point differences.

As expected, the spillover effect on schooling is negative. The model predicts that increasing all moms' schooling by a year would lead to an average 0.425 year decline in the next generation's schooling. Surprisingly, allowing for a rise in abilities only moderates this by 0.039 years (columns 2 vs 3) or 0.028 years (columns 4 vs 5). Higher abilities do not have much of an effect on intergenerational schooling relationships once preferences are taken into account, as we saw in Table 8. Preferences induce individuals to stay in school longer than what maximizes lifetime earnings, and since human capital accumulation faces decreasing returns, higher abilities do not further increase schooling by much.[49]

Only when we allow preferences to rise with moms' schooling do we see an increase in children's schooling of 0.451 years. The effect is large, but moderated by the negative spillover effect: Comparing columns (2) and (4), the sole effect of preferences is an even larger increase of 0.876 years. In sum, spillovers and preferences alone generate an intergenerational schooling relationship that is close to its empirical counterpart, with selection on abilities playing only a minimal role.

The spillover increases lifetime earnings by about 0.7 percent holding children's schooling constant, which increases to 1.2 percent when allowing schooling to adjust. Selection on abilities has an additional 1.3 percentage point effect. We conclude that independently of children's preferences for schooling, the causal effect of moms' education on earnings is more or less similar to the selection effect on abilities, i.e., higher ability moms having higher ability children. But once we also account for the correlation between moms' schooling and children's preferences for school-

[49] 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 minuscule 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.

ing, the average impact is negative: The increase in lifetime earnings drops by 1 percentage point (columns 2 vs. 4). This is because preferences for schooling induce individuals to deviate even further from lifetime earnings maximization by staying in schooling even longer, but the negative effect falls short of canceling out the positive spillover effects.

From which stage of the life cycle do the gains in lifetime earnings arise? In Figure 3 we plot earnings effects over an individual's entire life-cycle. Each line plots the change in average (left panel) or median (right panel) earnings following a 1-year increase in mom's schooling, for each age, compared to the benchmark estimates.

Comparing "Fixed *S*" and "Spillover," we see that the lifetime earnings spillover effect comes almost entirely from early labor market entry followed by almost no change after age 24 (the latest age we allow for labor market entry in the model). This also means that on average, children of moms with less schooling catch up with those of moms with more schooling by staying in school longer, so that their earnings do not differ by much later in life. Allowing for higher abilities has a fixed positive effect throughout the life-cycle, as expected from Section 2. Conversely, longer schooling induced by preferences increases 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 that mimics compulsory schooling reforms. Such reforms took place in many countries throughout the 20th century, and only affect parents who would otherwise not have attained the required level of schooling. Our exercise will show that it is possible to simultaneously estimate a large schooling OLS coefficient and a small or negative IV coefficient, indicating that a reform in the parents' generation may have little causal

		(1) fixed S	(2) spillover	(3) ability	(4) pref.	(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

Units in % are approximated by log-point differences. The second row denotes the change in the cross section average of the present discounted sum of lifetime earnings, in logarithms, for only those individuals affected by the reform. Column (1) holds children's schooling constant, while column (2) holds abilities and preferences constant but let's schooling vary according to the model. Columns (3)-(4) let abilities or preferences also vary according to their estimated correlations with h_P. Column (5) is when we allow for both selection on abilities and preferences for schooling. The coefficients from an intergenerational schooling OLS in data and model are 0.444 and 0.419, respectively.

effect on children's schooling. Nonetheless, it can still have 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

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

We choose 8 years as it was in fact the compulsory schooling requirement in many U.S. states at the time, or soon after.[50] In our data, 25% of moms have less than 8 years of schooling, and the reform would raise the schooling of these moms by an average of 3.5 years.

For the schooling OLS and IV regressions, we combine simulated samples from two regimes: one without the minimum requirement (our benchmark model) and one imposing the requirement. These represent mom-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. We repeat this exercise for the five cases we considered in the previous subsection: Keeping the next generation's schooling S fixed; letting S vary but controlling for $(z, \tilde{\xi})$; and then allowing z and/or $\tilde{\xi}$ to rise with h_P according to their correlations. The first row in Table 10 shows the regression results for all cases. The second row shows the change in the present discounted sum of lifetime earnings, and the following two rows the average change in earnings at ages 30 and 50.

We note that our experiment admits a range of possible *model* IV estimates. Which column we should compare against *empirical* IV estimates depends on the context of the instrument used in a particular study. As we noted in the introduction, most instruments are chosen with an emphasis

[50] 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 moms affected is almost two-fold (12.4% vs 25%).

to capture exogenous variation in parents' schooling independently of their *abilities*. But we suspect that some instruments may likely be correlated with their preferences, which are passed on to the next generation. In that case, we could expect a range of estimates between columns (2) and (4).

The OLS coefficients are somewhat difficult to interpret, since a quarter of the population from one regime receives varying levels of treatment (some moms are farther from 8 years than others). Nonetheless, they increase in all 4 cases to a value slightly higher than the model's benchmark value of 0.419. The IV regressions measure how children's schooling change when their moms go to school for an additional year, but only among those moms who are affected by the reform. The controlled effect is negative, but of smaller magnitude. The IV coefficient increases when allowing children's abilities to also rise (columns 2 vs 3) but much more when allowing children's preferences to rise (columns 2 vs 4). The reduced-form coefficient is only half of what we saw in Table 9, when we imposed a 1-year increase in the schooling of all moms. This is because both the spillover and preferences for schooling have a smaller effect for lower levels of S_P, who are the only ones affected by the reform.[51]

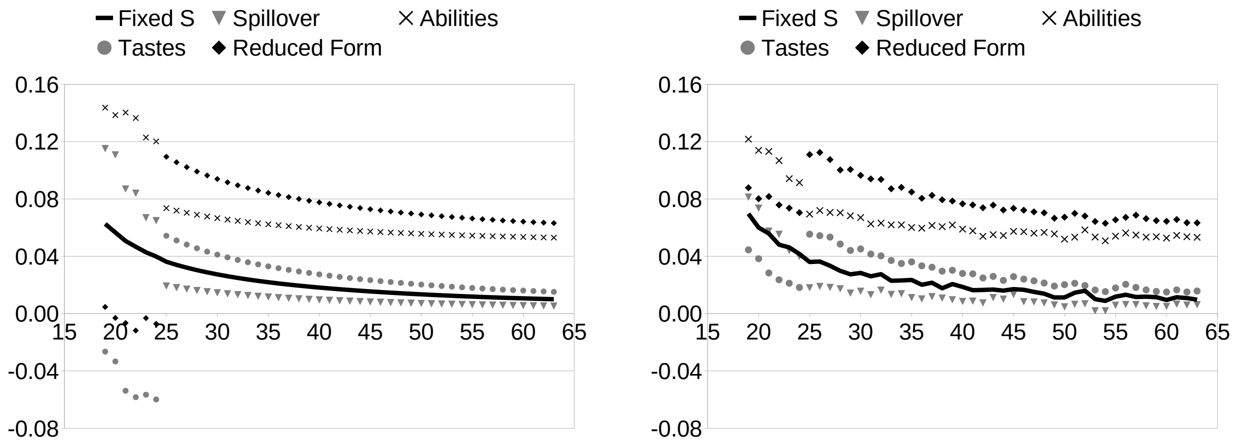
We conjecture that this partly explains the puzzling fact that many studies using special data sets on twins, adoptees, or compulsory schooling reforms find a zero or negative effect of parents' schooling on children's schooling. First, the structural effect may in fact be negative, as we have argued throughout this paper. More important, while the instruments may control for ability transmission, whether they also control for selection on preferences is unclear. Thus the reason that in some studies the IV coefficient is found to be close to zero can potentially be because a national-level reform, or being adopted into a new family, at least partly induces children to develop stronger preferences for schooling.

We can interpret this as moms *compelled* to attain more schooling having some impact on the family environment or their children's non-cognitive abilities, which in turn help them stay in school longer. However, such a transmission of non-pecuniary preferences from mom to child is likely weaker compared to moms who *choose* to become highly educated. In our experiment, that would mean that educating moms who would otherwise attain very low levels of schooling could increase their children's schooling by anywhere from -0.2 to 0.2 years.[52] Likewise, adoptees are likely to respond to the family environment of their adopted family, even if how much they adapt might be weaker than biological children.

Regardless, note that the effect on lifetime earnings is always positive, no matter the sign and

[51] Due to our structural assumptions on initial human capital in (1e) and on the dependence of preferences on h_P in (15).

[52] We could also understand 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 in the context of preferences. The reform could have different effects on families' preferences for longer schooling vis á vis grade repetition.



(a) change in average

(b) change in median

Figure 4: Lifecycle Impact of Reform

magnitude of the IV coefficients. Moreover, as can be seen in Figure 4, the positive effect persists throughout the life-cycle, unlike in Figure 3 where it became virtually zero from 24 onward. For children with low S_P, earnings rise significantly even without the endogenous adjustment in schooling (the "Fixed S" line). Consequently, the extra increase from early labor market entry is lower in comparison (the "Spillover" line).

The negative effect from preferences is still there, meaning that if a reform impacting moms with low education also raise their children's preferences for schooling, we may expect the effect on children's schooling *and* lifetime earnings to be close to zero. But the negative effect is smaller in magnitude for children of less educated moms (although the lifetime earnings effect is -1.1 percentage points in both Tables 9 and 10, moms' schooling increases by 1 year in the former, by design, but an average of 3.5 years in the latter). This is because these children have weak preferences for schooling to begin with and their returns to schooling are higher, so the increase in later-in-life earnings can dominate the foregone earnings from schooling. 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 that are affected by labor market entry. 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

We presented 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—preferences 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 on children's earnings is positive. Since children of higher human capital parents begin life with higher human capital themselves, 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 parents' schooling, children's schooling and children's earnings over the life-cycle. It is also consistent with a positive OLS correlation between parents' and children's schooling. In generating this last feature, our estimated results attribute an important role to the correlation in preferences for schooling across generations. The unobserved correlation between moms' education and children's preferences for schooling is the main determinant of children's schooling, not selection on abilities or parental spillovers.

Our results are also consistent with the previous literature that finds that, when using instruments for parents' schooling, the estimated IV coefficient on children's schooling can be small or even negative. All else equal, increasing a parent's human capital decreases a child's schooling, a feature of diminishing returns to human capital accumulation, even if parents with higher human capital have children with stronger preferences for schooling. But even though the causal effect of moms' schooling on children's schooling is negative, the estimated causal effect of mom's schooling 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 (Plug and Vijverberg, 2003; Björklund et al., 2006).

Our model was estimated to U.S. cohorts born in the 1930s (with parents were born in the 1900s to 1910s). While education structures have since changed dramatically in advanced countries, we believe our structural framework contributes to the literature that is also often based on data from older cohorts. In addition, the insights from our model-based approach is still relevant for more recent cohorts, for whom high-school and college graduation is a more important indicator of educational attainment than years of schooling, at least in two dimensions: i) When considering education, it is important to consider the heterogeneous gains across individuals who attain the same education levels (e.g. the *quality* of degrees), and thus ii) Causal, intergenerational effects on quantity-based indicators of education (e.g. whether or not an individual attains a college degree) cannot be found without controlling for such differences in quality.

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-N urture 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.

Blanden, J., P. Gregg, and L. Macmillan (2007). Accounting for intergenerational income persistence: Noncognitive skills, ability and education. *The Economic Journal 117*(519), C43–C60.

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.

Chang, Y., S. N. Durlauf, S. Lee, and J. Y. Park (2023, March). A trajectories-based approach to measuring intergenerational mobility. Working Paper 31020, National Bureau of Economic Research.

Chetty, R., N. Hendren, P. Kline, E. Saez, and N. Turner (2014, May). Is the united states still a land of opportunity? recent trends in intergenerational mobility. *American Economic Review 104*(5), 141–47.

Chevalier, A., C. Harmon, V. O'Sullivan, and I. Walker (2013, December). The impact of parental income and education on the schooling of their children. *IZA Journal of Labor Economics 2*(8).

Connolly, M., C. Haeck, and J.-W. P. LalibertÃ© (2021, June). Parental Education and the Rising Transmission of Income between Generations. In *Measuring Distribution and Mobility of Income and Wealth*, NBER Chapters, pp. 289–315. National Bureau of Economic Research, Inc.

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.

Fernández, R., A. Fogli, and C. Olivetti (2004). Mothers and Sons: Preference Formation and Female Labor Force Dynamics. *The Quarterly Journal of Economics 119*(4), 1249–1299.

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.

Groves, M. O. (2005). *Chapter Seven. Personality and the Intergenerational Transmission of Economic Status*, pp. 208–231. Princeton: Princeton University Press.

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.

Holmlund, H., M. Lindahl, and E. Plug (2011). The causal effect of parents' schooling on children's schooling: A comparison of estimation methods. *Journal of Economic Literature 49*(3), 615–651.

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. (2002). The nature and nurture of economic outcomes. *American Economic Review 92*(2), 344–348.

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.