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

Sang Yoon (Tim) Lee
Department of Economics
University of Mannheim

Nicolas Roys
Department of Economics
University of Wisconsin-Madison

Ananth Seshadri
Department of Economics
University of Wisconsin-Madison

March 2015

Preliminary; comments welcome.

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 delivers a positive relationship between parents' education and children's earnings and an ambiguous relationship with children's schooling. Identification is achieved by comparing the earnings levels of children with the same schooling level but of parents with different schooling levels. A generalized version of the model with heterogeneous tastes for schooling is estimated using the HRS data with parents' schooling. The empirically observed, positive OLS coefficient from regressing children's schooling on parents' schooling is mainly accounted for by the correlation between parents' schooling and children's unobserved tastes for schooling. But this is countered by the negative structural relationship between parents' and children's schooling choices, resulting in a negative or close to zero IV coefficient when exogenously increasing parents' schooling. Nonetheless, an exogenous one-year increase in parents' schooling increases a child's lifetime earnings by 1.2 percent on average.

*We thank Steven Durlauf and Chris Taber for very helpful conversations which motivated this paper. Junjie Guo and Wei Song provided outstanding research assistance. We also thank Mariacristina De Nardi, Eric French and seminar participants at the IFS.

1 Introduction

Parents have a large influence on their children's outcomes. The correlation between parents' and children's schooling is as high as 0.5, and even after controlling for children's schooling, parents' schooling or earning levels have a positively significant effect on children's earnings. Does this merely represent correlation in unobserved heterogeneity across generations? Or does it also partly reflect human capital spillovers from parent to child? These are important questions from the perspective of education policy. After all, if intergenerational correlations merely reflect selection, government subsidies aimed at improving education would only impact one generation. On the other hand, in the presence of intergenerational externalities, the returns from such public investments are reaped by all succeeding members of a dynasty, resulting in long-lasting effects. Redistributive education policies may also go beyond reducing inequality within a single generation and have positive impacts on intergenerational mobility.

Furthermore, intergenerational spillovers form an integral part of models of human capital accumulation such as Becker and Tomes (1986). Being able to identify and estimate the magnitude of these spillovers is consequently important not just for better understanding the effects of socioeconomic policy, but also to gain a better understanding of the mechanics through which inequality is formed and persists over generations. How important is nature (individual abilities that cannot be affected by exogenous intervention) as opposed to nurture (investments into human capital) in determining the next generation's economic status? How do forced increases in the schooling of parents affect the schooling and earnings of children? These are the questions we seek to address in this paper.

We begin with a very simple life-cycle model of human capital accumulation owing to Ben-Porath (1967), which encompasses both schooling and learning on-the-job, and in which human capital represents an individual's economic earnings ability. In this model, individual earning profiles are determined by their initial level of human capital, and the speed, or innate learning ability, at which the individual accumulates human capital.[1] We augment this model in two aspects. First, we posit that an individual's initial level of human capital when he begins schooling at age 6 is also a function of his parent's human capital, which is what represents the parental spillover in our model.[2] Second, we assume that the initial level of human capital is also affected by learning ability, i.e., the speed at which human capital is accumulated after age 6. Thus, the initial level and speed of human capital accumulation are affected by his own ability, which is unobserved but correlated across generations, while the parent's human capital, which can be proxied by the schooling or earnings levels of the parent and thus observed, only affects the child early on in life.

The simple model delivers analytical expressions for the schooling and earnings of a child as a function of the parent's human capital, which closely resemble those that have been estimated in the empirical literature. In particular, we demonstrate that the identification problem of separating

[1] The Mincer regression is a special case of this set-up when post-schooling time allocation declines linearly until retirement.

[2] For the purposes of this paper, we refer to the causal effect of parents' education on children's earnings as the "parental spillover."

parental spillovers from selection on abilities can be rectified by using information on children's earnings and schooling jointly. The key for identification is that the parents' human capital affects children's education and earnings outcomes differently, which is true in our model because the spillover occurs early on in life while other factors persist through life. In our model, all else equal, children with high human capital parents spend less time in school because there is less need to accumulate additional human capital before entering the labor market.[3] It is still the case that these children ultimately attain higher levels of human capital and hence higher earnings, but it takes them less time to reach that level, resulting in shorter schooling periods. But in the data, these children in fact spend more time in school. So it must be that high human capital parents, who also tend to have high innate learning abilities, also pass on these abilities which overcomes the negative level effect, by inducing high learning children to stay longer in school.

What allows identification in our model is this differentiation of the quantity and quality of years of schooling: children who attain the same years of schooling can still have different levels of human capital, which is manifested in the data as differential earnings. Suppose that we observe two individuals with the same schooling level but different levels of earnings, and parents with different levels of schooling. Through the lens of our model, this reveals the relative magnitudes of their learning abilities as a function of the relative magnitudes of their parents' schooling. Then, the earnings differences between these two individuals can be completely accounted for by the schooling differences of their parents, from which we can recover the size of parental spillovers. Once this is done, the contribution of nature is identified by observing how children's schooling levels vary across parents with different schooling levels.

Our identification scheme relies on parental spillovers having a constant effect over the children's life-cycle earnings conditional on the children's own schooling levels. Is such an assumption empirically reasonable? Reduced form evidence suggests that children who attain the same years of schooling, but whose mothers have different years of schooling, have parallel earning profiles with a constant gap. The parallel gap points toward the existence of a parental spillover that only affects how much human capital the child accumulates before entering the labor market (schooling), and then remaining constant once controlling for the child's schooling level. Moreover, this gap is indeed similar across different child schooling levels. We show analytically that the spillover is precisely picking up this gap. If our model were true, reduced form estimates from the Health and Retirement Survey data indicate that educating a mom for an extra year is equivalent to having a mom with an extra year of education—i.e., the treatment and selection effects are similar. Furthermore, five extra years of mom's education has the same reduced form effect as one additional year of own schooling, suggesting that as much as 20% of parents' education spills over to child's earnings.

Of course, because the child's schooling choice is also endogenous, the reduced form evidence does not tell us what the exact model implied magnitude of parental spillovers on children's earnings is. To bring our model closer to the data, we extend the basic model and estimate it to HRS

[3]This is in fact what many empirical studies find, as we soon discuss. The idea that less schooling may indicate higher earnings prospects goes as far back as Willis and Rosen (1979).

data on individual schooling and earnings, and parents' schooling. In addition to parents' schooling levels, which is observed, and learning abilities, which is unobserved, we account for a third source of heterogeneity—an unobserved taste for schooling.

We let both the child's ability and taste for schooling be correlated with his parent's schooling level, in line with evidence that other factors not related to economic or financial returns (psychological factors, non-cognitive skills, misinformation) may induce a child to attain more or less education (Betts, 1996; Oreopoulos et al., 2008; Rege et al., 2011). Moreover, several studies that estimate the returns to education have shown that pecuniary motives alone fall short of explaining education choices (Heckman et al., 2006, 2008). In our context, less educated parents may discourage children's academic achievement due to lack of information, or more educated parents may motivate the children to emulate their parents regardless of future economic outcomes. Since we do not explicitly model expectations formulations or non-cognitive skills, our heterogeneous tastes for schooling effectively lumps all such factors together.

By letting the child's taste for schooling be correlated with his parent's schooling, the relationship between schooling and parents freely varies from the relationship between earnings and parents. Our estimates suggest that most of schooling differences can be explained by this heterogeneity in tastes across parents with different levels of education, evidence that *ceteris paribus*, pecuniary motivation has little causal effect on children's schooling outcomes. Nonetheless, a forceful increase in mom's schooling is found to have a 1.2% causal boost on lifetime earnings. However, the effect is heterogeneous both across individuals and over the life-cycle. On average, the earnings increase mainly comes from reducing children's need for schooling so that they enter the labor market early (thereby decreasing foregone earnings), but there is almost no difference in their earnings after age 25. For children of less educated parents, however, the earnings boost sustains over the child's entire lifetime.

Related literature By no means are we the first to estimate the causal effect of parents on children's outcomes. We contribute to this literature by incorporating insights from a human capital model of earnings and education, and the recent literature that emphasizes the potentially large impact of early childhood interventions. In particular, Cunha and Heckman (2007) note that early childhood investments are difficult to separate from genetic transmissions. Ours is a first attempt to fill this gap.

Since Solon (1999), a broad literature has studied the causal effect of parents on children's earnings and/or education.[4] The common challenge for all these studies is to separately identify the unobserved correlation between parents' and the children's abilities from the unobserved causal impact of parental spillovers. The typical approach has been to posit a linear relationship between parents' and children's education, and employ special data on twins, adoptees or compulsory schooling reforms as natural experiments for identification (Behrman and Rosenzweig, 2002; Plug, 2004; Black et al., 2005). All studies find that the causal intergenerational schooling re-

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

lationship is close to zero or even negative, especially for mothers.[5] Because conventional wisdom tells us that the causal intergenerational effect of schooling should be positive, they contemplate whether the inability of being able to control gender differences in parenting, unobserved effects from mating, etc., in the data, leads to such surprising results.

Conversely, there is also evidence that points to strong inherited genetic effects on children's schooling outcomes. Early work by Behrman and Taubman (1989), using data on twin parents and extended family relationships to decompose the variance of observed years of schooling into the variance of genetic and environmental variables, conclude that as high as 80% is accounted for by genetics alone. More recently, Plug and Vijverberg (2003), using data on biological and adopted children to separately identify how much of inherited IQ can explain children's schooling outcomes, find more evidence for environment effects, but conclude that the inherited genetic effect is still at least larger than 50%.

Studies that look at children's earnings and/or outcomes other than schooling tend to find a larger role for the environment. Bowles and Gintis (2002) estimates how much of the intergenerational persistence in earnings can be explained by the correlation between the child's education and IQ, and his father's earnings, but has little to say about causal effects as the intermediate variables themselves are both subject to ability selection and spillovers. In a rare study where both information on adopted and biological parents were used, Björklund et al. (2006) find that the biological mother's years of schooling has a larger effect on the child's years of schooling than the adopted mother's, but adoptive father's earnings or income has a larger effect on the child's earnings or income than the biological father's. Sacerdote (2007), using a large sample of Korean adoptees, finds that while adopted families tend to matter less for both education and earnings, the seem to have a large effect on other social behavior, such as college selectivity and drinking.

Taken altogether, the fact that genetic effects are larger in some cases while environmental effects are larger in others is viewed as an inconsistency, especially in terms of the child's education and earnings outcomes (Black and Devereux, 2011). This stems at least partially from an implicit understanding that parents should have a qualitatively similar effect on the child's education and earnings outcomes, which is most likely motivated by the very strong relationship between an individual's education and earnings (Card, 1999). However, just as it may not be natural to expect the parental environment to have the same effect on the child's cognitive traits as non-cognitive traits (Cunha et al., 2010; Heckman et al., 2013), we argue that a parent's education can have qualitatively different effects on the child's education and earnings as well. Another confounding factor when interpreting the parental effect on children's schooling is to what extent the education choice for or by the child was economically motivated. While this is well recognized in the literature on returns to schooling (Heckman et al., 2006, 2008), as of yet no attempts have been made to link non-pecuniary motives with a potential discrepancy between the parental effect on children's education as opposed to children's earnings.

Our first departure from the literature is to posit a non-linear model in which the individual

[5] With the exception of Black et al. (2005), who find weak evidence of a positive mother-son relationship.

takes his parent's variable(s) as a state to solve a lifetime optimization problem, as opposed to a linear explanatory variable. The negative causal effect of parents on children's schooling that the solution admits may sound counterintuitive at first. But such a level effect (that high initial conditions substitute later investments) have been found to be important in previous life-cycle models of human capital accumulation (Heckman et al., 1998; Huggett et al., 2011). Moreover, the solution explains the child's schooling and earnings outcomes jointly, rather than single outcomes one by one. In a mechanical sense, this is how we are able to identify the two unobservables, namely, the cross-sectional correlation between children's abilities and their parents' schooling, and the size of the intergenerational spillovers. Since the resulting causal effect on earnings is still positive, our model is at once consistent with the empirical findings that parents may have a negative causal effect on children's schooling, and the limited evidence that fathers' or family incomes have a positive causal effect on children's earnings.

The second aspect that differentiates our approach is that in our estimation, we explicitly separate pecuniary and non-pecuniary motives for schooling. Non-pecuniary motives are estimated to be quite large in life-cycle models with schooling and/or occupation choice.[6] There is also substantial evidence 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), which would also be captured as part of the taste for schooling in our setting. The existence of non-pecuniary motives presents difficulties when estimating causal effects from the data, but also provides some discipline on how to interpret the results from special data sets. For example, the schooling difference between twin parents or compulsory increases in years of schooling are less likely to to be related to other unobserved family characteristics that affect children's schooling decisions, while children adopted to different families very likely do develop the non-cognitive skills or acquire information that conform to such unobservables. If these unobservables are positively correlated with parents' education, to some extent it should be expected that the effect of a parents' education on children's education should be smaller in twins or IV studies (although the former is also subject to sampling bias) than adoptee studies.

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; Heckman et al., 2013), and the childhood environment has long been suspected as what may explain the large estimates for non-pecuniary motives found in structural models of earnings (Bowles et al., 2001; Heckman et al., 2006). The spillover in our model can be understood as the parental effect on cognitive skills that increases the child's earnings ability, while the correlation of a parent's education with her child's taste for school can be understood as the parental effect on non-cognitive abilities that do not directly relate to earnings and only schooling.

The rest of the paper is organized as follows. Section 2 posits a simple model of human capital accumulation and adds to it a parental spillover. The solution to this model is derived analytically from which we can make empirical predictions. In section 3 we describe the HRS data that we use

[6] Keane and Wolpin (1997); Heckman et al. (1998) and almost the entire literature that followed.

and interpret the reduced form evidence through the lenses of our model. Section 4 presents the more comprehensive model which is estimated to the HRS. We also show, quantitatively, that the estimated model inherits properties of the simpler model. Section 5 examines counterfactual predictions of the model including a hypothetical compulsory schooling reform. Section 7 concludes.

2 A Simple Model with Parental Spillovers

In this section, we present a simple variant of a Ben-Porath (1967) life-cycle model of human capital accumulation.

2.1 Preliminaries

The Ben-Porath Model Consider first the basic Ben-Porath model without parental spillovers or any tastes for schooling. An individual begins life at age 6, retires at age $R > 6$ and dies at age $T \geq R$, all exogenous to the individual. At age 6, an individual is described by the state vector (h_0, z), which denotes his initial stock of human capital and (learning) ability, respectively. An individual chooses time and good investments into human capital accumulation, $[n(a), m(a)]$, $a \in [6, R)$, to maximize the present discounted value of net income. His problem is

$$\max_{n(a),m(a)} \left\{ \int_6^R e^{-r(a-6)} \left[wh(a)(1-n(a)) - m(a) \right] da \right\}$$

subject to

$$\begin{aligned}
&\dot{h}(a) = z[n(a)h(a)]^{\alpha_1} m(a)^{\alpha_2}, \qquad a \in [6, R), \\
&n(a) \in [0,1], \qquad m(a) \geq 0, \\
&h(6) = h_0
\end{aligned}$$

where α_1 and α_2 are the returns to time and good investments, respectively. The market wage w and discount rate r are constant and taken as given by the individual. Since the market wage multiplies human capital to generate earnings, human capital is understood as an individual's "earning ability" as opposed to z, "ability to learn" (Heckman et al., 1998). Starting from h_0, human capital $h(a)$ is accumulated and evolves over the life-cycle, while z is constant through life.

The above decision problem is a finite horizon problem. When the individual retires, his stock of human capital depreciates completely. Thus, the time path of $n(a)$ weakly decreases with age. The individual state vector captures slope and level effects. Assuming decreasing returns to scale, i.e. $\alpha_1 + \alpha_2 < 1$, if the initial stock of human capital h_0 and ability z are low and high enough, respectively, the time allocation decision is constrained at the upper bound of 1 for some time and then strictly declines. Since all time is spend either working or accumulating human capital, the length of time that $n(a) = 1$ can be understood as the schooling period. All else equal, individuals with higher ability levels (high z) make more human capital investments and stay longer in school, while individuals with a higher initial stock of human capital (h_0) stay less time in school.

Although the model is posed as if a child, at age 6, chooses a lifetime plan for human capital investment decisions, it is clear that parents would have a dominant role in making those decisions at least during earlier stages of life. This could be factored in easily by assuming a parent who has altruistic preferences toward the lifetime utility or human capital of her children. But with no uncertainty and short-run credit constraints, an optimizing altruistic parent would make the same decisions for her child as the child would have chosen on his own.

Assumptions for Identification Now suppose a mass of individuals face the same problem over multiple generations. We think of intergenerational transmission as parents influencing the initial state vector (h_0, z). To set ideas, denote the initial level of human capital and learning ability of individual i as (h_{0i}, z_i), and his parent's human capital and learning ability as (h_{Pi}, z_{Pi}). Our approach is to assume a statistical relationship for the parents to focus on the children's initial conditions. Learning abilities, which remain constant through life, are transmitted exogenously through generations. We assume that $\log z_i$ is log-linear in $\log z_{Pi}$, specifically

$$\log z_i = \mu_z + \rho_z \frac{\sigma_z}{\sigma_{z_P}} (\log z_{Pi} - \mu_{z_P}) + \epsilon_{zi}, \tag{1}$$

where ρ_z represents the intergenerational correlation of abilities, (μ_z, μ_{z_P}) are the population means of the child and parent generations' abilities, (σ_z, σ_{z_P}) the corresponding standard deviations, and ϵ_{zi} a mean zero, i.i.d. shock with standard deviation σ_ϵ.

The parents' human capital h_P is likely correlated with their abilities z_P, which transmits to z. However, we are not interested in the intergenerational persistence of learning abilities. By assuming a statistical relationship between h_P and z_P, z becomes multiplicatively independent of z_P conditional on h_P, i.e.

$$z_{Pi} = f(h_{Pi})\epsilon_{Pi} \tag{2}$$

$$\Rightarrow \qquad z_i = g(h_{Pi})\epsilon_i, \tag{3}$$

where $(\epsilon_{Pi}, \epsilon_i)$ are i.i.d. across the population. We can now remain completely silent about z_P and its correlation with the child's (h_0, z), as long as we know the correlation between $\log z$ and $\log h_P$ induced by (3), which we denote by ρ_{zh_P}. This correlation is positive if children of high human capital parents, who likely have high abilities themselves, also have higher ability children.

If each subsequent cohort began life with the same initial stock of human capital $h_{0i} \equiv h_0$, parents' education would not have any causal influence on the earnings of children. For there to be any causal (treatment) effect there must exist a channel through which h_{0i} can be affected by h_{Pi}. To the extent that $h_i(6) \equiv h_{0i}$ represents the amount of learning that happens before school entry, clearly its formation is also affected by both one's own learning ability z_i and parental investments prior to age 6. Such parental investments will depend on the parent's economic status, which is summarized by h_{Pi}. So we can write h_{0i} as a function of (z_i, h_{Pi}); specifically we assume

$$h_i(6) \equiv h_{0i} = f(z_i, h_{Pi}) = z_i^{\lambda} h_{Pi}^{\nu}. \tag{4}$$

We want to identify the elasticity of h_0 with respect to h_P, evaluated at individual i's state:

$$\nu \equiv \left. \frac{\partial \log f(z, h_P)}{\partial \log h_P} \right|_{(z,h_P)=(z_i,h_{Pi})}$$

which is assumed to be constant for all i. The parameter ν captures the degree to which a higher human capital parent transmits more human capital to her child during the first 6 years of life. The spillover effect is defined as the increase in earnings induced by this transmission. Such a channel is rather standard in the literature on intergenerational transmissions (Becker and Tomes, 1986), and our particular formulation can be considered a reduced form representation of the importance of early childhood. The parameter λ can be understood analogously: it captures the population correlation between z and h_0 conditional on h_P, which comes from two channels: i) how much the child's ability matters for early human capital formation, and ii) how much of childrens' abilities are *not* explained by h_P.[7]

In the data, h_P can be proxied by a parent's education and/or earnings, and is thus observed. But since abilities are unobserved, we may face the problem of separately identifying (ν, λ) from ρ_{zh_P}, i.e., whether a child's schooling or earnings are correlated with the parent because of h_P or z. The goal of this section is to show how the solution to the model allows this. To do so, we abstract from tastes for schooling for now, which are only included later in section 4. We do this primarily to gain insight into the precise mechanisms at work, and understand how the model parameters can be recovered from a panel individuals assuming (1)-(4). Taste for schooling are included to account for any remaining unobserved heterogeneity.

Consequently, we are simply solving a life-cycle decision problem in which a child takes his ability and his parent's human capital as an exogenous state variable. This means the parental spillover is essentially an intergenerational externality that is not internalized by previous generations.[8]

2.2 The Individual's Problem

For what follows we drop the individual subscript i unless necessary. Let $V(a, h)$ denote the value function for an individual of age a and human capital level h. The problem faced by an individual

[7] We are assuming away that parents' human capital can directly affect their children's learning abilities, but if this were the case, λ would pick up a third channel, namely, how much of z is directly affected by h_P before age 6. In doing so, we would still be losing any direct effect that h_P can have on z after age 6. However, not only do parents have the largest effect on children in their early years (Del Boca et al., 2013), such an effect is not separately identified from ρ_{zh_P} in our simple model. We discuss this more in Section 5.

[8] In other words, individuals do not invest more in their own human capital in anticipation of that investment spilling over to subsequent generations. If they do, it would only have a larger effect on adult earnings.

at age 6 given $h(6) = h_0$ can be written

$$V(6, h_0) = \max_{\{n(a), m(a)\}} \left\{ \int_6^R e^{-r(a-6)} g\left(h(a); n(a), m(a)\right) da \right\}$$
$$\dot{h}(a) = f\left(h(a); n(a), m(a)\right), \qquad n(a) \in [0, 1], \qquad m(a) > 0.$$

where the objective function and law of motion are given as

$$g(h; n, m) = -whn - m$$
$$f(h; n, m) = z(nh)^{\alpha_1} m^{\alpha_2}.$$

This is a continuous time deterministic control problem with state h and controls (n, m). The terminal time is fixed at R but the terminal state $h(R)$ must be chosen. 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 equation is

$$rV(a, h) - \frac{\partial V(a, h)}{\partial a} = \max_{n,m} \left\{ g(h; n, m) + \frac{\partial V(a, h)}{\partial h} f(h; n, m) \right\}.$$

As usual, the HJB equation can be interpreted as a no-arbitrage condition. The left-hand side is the instantaneous cost of holding a human capital level of h at age a, while the the right-hand side is the instantaneous return. The first order conditions for the controls are

$$whn \leq \alpha_1 z(nh)^{\alpha_1} m^{\alpha_2} \cdot V_h, \qquad \text{with equality if } n < 1 \tag{5}$$
$$m = \alpha_2 z(nh)^{\alpha_1} m^{\alpha_2} \cdot V_h, \tag{6}$$

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

$$r \cdot V_h - V_{ah} = w(1 - n) + \frac{\alpha_1 z(nh)^{\alpha_1} m^{\alpha_2}}{h} \cdot V_h + z(nh)^{\alpha_1} m^{\alpha_2} \cdot V_{hh}, \tag{7}$$

where V_{xh} is the partial of V_h with respect to $x \in \{a, h\}$. This "Euler equation" states that at the optimum, the marginal cost of increasing human capital must be equal to the marginal return. Equations (5), (6) and (7) along with the law of motion

$$\dot{h} = z(nh)^{\alpha_1} m^{\alpha_2} \tag{8}$$

characterize the complete solution, given the initial state $h(6) = h_0$ and terminal condition $V_h = 0$, the appropriate transversality condition for a fixed terminal time problem. We solve this problem in Appendix A and here only present the important results.

To save on notation, it is useful to define $\alpha \equiv \alpha_1 + \alpha_2$ and

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

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

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

The optimal choice of schooling S is uniquely determined by

$$F'(S) > 0, \qquad F(S) \geq z^{1-\lambda(1-\alpha)} h_P^{-\nu(1-\alpha)} \tag{9}$$

with equality if $S > 0$.

Proof. See Appendix A. □

The higher the learning ability z of an individual, the higher the optimal choice of his schooling (as long as $\lambda(1-\alpha) < 1$). Intuitively, conditional on an initial level of human capital, higher z individual benefit more from schooling. Importantly, the causal, spillover effect from the parent's human capital h_P on schooling is *negative* (as long as $\nu(1-\alpha) > 0$). While this may sound counterintuitive, it is in line with the empirical evidence in Behrman and Rosenzweig (2002) and Black et al. (2005) who find that increasing mothers' years of schooling by 1 year has a negative effect on their children's schooling outcomes. In our model, this relationship holds true because children with high human capital parents have less of a need to stay in school because they start off with a higher level of human capital (a quantity-quality substitution effect).

In the estimation, we proxy for h_P by parents' schooling.[9] Then according to our model, the observed dependence of children's schooling on parents' schooling depends on both ρ_{zh_P}, the correlation between (z, h_P), the spillover effect ν, and early childhood learning λ. To see this, suppose (3) takes the form

$$\begin{aligned} \log z_i &= \mu_z + \rho_{zh_P}\frac{\sigma_z}{\sigma_{h_P}}(\log h_{Pi} - \mu_{h_P}) + \epsilon_i \\ &= \tilde{\mu}_z + \tilde{\rho}_{zh_P}\log h_{Pi} + \epsilon_i, \end{aligned} \tag{10}$$

where $(\mu_{h_P}, \sigma_{h_P})$ denote the mean and standard deviation of the parents' human capital.[10] Now $\tilde{\rho}_{zh_P}$ denotes the elasticity of z with respect to h_P, rather than the correlation. At equality, (9) can

[9] The model is estimated to the original HRS cohort. In the HRS data, we observe parents' years of schooling, but not earnings. To proxy the human capital distribution of the parents by earnings instead of schooling, we assume that the parents' schooling-earnings relationship is identical to a separate Mincer regression run on the HRS AHEAD cohort, who are approximately one generation older than the original HRS cohort.

[10] This is in fact what we assume for the estimation later in Section 4, equation (19).

be written as

$$\log F(S_i) = [1-\lambda(1-\alpha)]\log z_i - \nu(1-\alpha)\log h_{Pi}$$
$$= [\tilde{\rho}_{zh_P} - (\nu + \lambda\tilde{\rho}_{zh_P})(1-\alpha)]\log h_{Pi} + [1-\lambda(1-\alpha)]\,(\tilde{\mu}_z + \epsilon_i). \tag{11}$$

Hence depending on other parameter values, we may observe a positive OLS effect between children and parent's schooling even if ν is large. Moreover, even if we were to observe that increasing a parent's human capital exogenously had a small effect on a child's schooling, it may not imply that there are no spillovers but a large α. But $(\nu, \lambda, \rho_{zh_P})$ are not separately identified in such a simple schooling regression, even if (α_1, α_2) are known.

Proposition 2: Post-Schooling Human Capital *For any S, human capital at the end of schooling, h_S, satisfies*

$$wh_S = C_1(S)\cdot z^{\frac{1}{1-\alpha}}, \qquad where \;\; C_1(s) = \alpha_1 \cdot [\kappa q(6+s)]^{\frac{1}{1-\alpha}}$$

Proof. See Appendix A. □

The above proposition tells us that, once the length of schooling is known, the human capital level of a child is affected only by his own learning ability z. His initial stock of human capital, h_0, has no effect on the *amount* of human capital accumulated (quality) except through the *length* of schooling (quantity), S. So both the parental effects of ν and early childhood learning λ are subsumed in the length of schooling. Although ν has no effect on a given individual's earnings once his schooling is determined, we will see in Corollary 3 that earnings differences across individuals with the same level of schooling identifies ν from the data.

Using Proposition 2, we can also describe the dependence of children's earnings profiles on the human capital of their parents. Assume that a fraction π_n and π_m of time and goods investments (n, m), respectively, are subtracted from the value of the human capital to obtain measured earnings. This simply amounts to assuming that the individual pays for the job training costs in the form of lower contemporaneous wages (i.e., that employers deduct this fraction before paying employee wages) .

Corollary 1: Evolution of Earnings Profiles *For an individual who attains S years of schooling, for all $a \in [6+S, R)$,*

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

for any $(\pi_n, \pi_m) \in [0,1]^2$, where

$$C_2(a;s) = \kappa^{\frac{1}{1-\alpha}} \cdot \left\{ r \cdot \int_{6+s}^{a} q(x)^{\frac{\alpha}{1-\alpha}} dx - (\alpha_1\pi_n + \alpha_2\pi_m) q(a) \right\}$$

for $a \geq s-6$.

Proof. See Appendix A. ☐

By virtue of Corollary 1, what fraction of job training costs are paid for by the firm only depends on age, as long as it is assumed to be constant. Moreover, the exponents for the human capital production function, (α_1, α_2), are identified by the slopes of the log age-earnings profiles of individuals with the same level of schooling, since age only affects earnings through the function C_2. This is standard in models that use this approach.

This expression for earnings in Corollary 1 can also be interpreted as a Mincer-type equation that relates earnings to schooling. The functions C_1 and C_2 are common to all individuals and dictates the returns to schooling, while C_2 would further tell us the shape of the average age-earnings profile. Of course, we would need to know the exact parameters for these functions and also be able to control for the unobserved z for testing.

2.3 Identification of Key Parameters

A robust finding in empirical studies is that even after controlling for observables, mothers' education has a statistically significant relationship with children's schooling and earnings. The parameters $(\nu, \lambda, \rho_{zh_P})$ are a structural representation of this. In this section, we demonstrate that Proposition 1 and Corollary 1 imply that all three parameters are separately identified once we have data on children's schooling and earnings outcomes, and the human capital levels of their parents.

COROLLARY 2: IDENTIFYING λ AND ρ_{zh_P} *Suppose we observe a large, representative sample of individuals for whom we know the human capital level of their parents h_{Pi}, schooling outcomes S_i, and age-earnings profiles $e_i(a)$ (but not z_i).*

1. *If we select only those individuals whose parents have the same $h_{Pi} = \hat{h}_P$, then for any $a \in [6+S, R)$, earnings depend only on S through λ and nothing else:*

$$e_i(a|h_{Pi} = \hat{h}_P) \propto [C_1(S_i) + C_2(a; S_i)] \cdot F(S_i)^{\frac{1}{(1-\alpha)[[1-\lambda(1-\alpha)]}}. \tag{12}$$

   So if we regress

$$\log e_i(a|h_{Pi} = \hat{h}_P) = a_0 + a_1 \log [C_1(S_i) + C_2(a; S_i)] + a_2 \log F(S_i) + \epsilon_i,$$

   we recover

$$\hat{a}_2 = 1/\left\{(1-\alpha)\left[1-\lambda(1-\alpha)\right]\right\}.$$

2. *Suppose that (z, h_P) follows (10). If we regress*

$$\log e_i(a) = b_0 + b_1 \log [C_1(S_i) + C_2(a; S_i)] + b_2 \log h_{Pi} + \eta_i, \tag{13}$$

we recover

$$\hat{b}_2 = \hat{\tilde{\rho}}_{zh_P}/(1-\alpha).$$

So if (α_1, α_2) *is known, as well as the functions* (C_1, C_2, F)*,* λ *is recovered from a Mincer regression that includes a complete set of dummies for all human capital levels of parents. Likewise,* (ρ_{zh_P}, σ_z) *are jointly identified from a Mincer regression linearly controlling for* $\log h_P$*.*

Proof. Suppose we observe two age $\hat{a}$ individuals with different levels of schooling and age $\hat{a}$ earnings but whose parents have the same level of human capital, denoted by (S_1, S_2), (e_1, e_2), and (h_{P1}, h_{P2}), respectively. Let (z_1, z_2) denote their *unobserved* learning abilities. Then by Corollary 1,

$$\frac{e_1}{e_2} = \frac{[C_1(S_1) + C_2(\hat{a}; S_1)]}{[C_1(S_2) + C_2(\hat{a}; S_2)]} \cdot \left(\frac{z_1}{z_2}\right)^{\frac{1}{1-\alpha}},$$

but since $h_{P1} = h_{P2}$, by Proposition 1

$$\frac{F(S_1)}{F(S_2)} = \left(\frac{z_1}{z_2}\right)^{1-\lambda(1-\alpha)} \quad \Rightarrow \quad \frac{e_1}{e_2} = \frac{[C_1(S_1) + C_2(\hat{a}; S_1)]}{[C_1(S_2) + C_2(\hat{a}; S_2)]} \cdot \left[\frac{F(S_1)}{F(S_2)}\right]^{\frac{1}{(1-\alpha)[1-\lambda(1-\alpha)]}}$$

which is (12). Part 2 follows trivially from Corollary 1 after plugging in the assumed relationship between z and h_P from (10). □

Put simply, the magnitude of λ is identified by the Mincer coefficient of children's schooling on earnings, *conditional on parental human capital*. Clearly for children with identical levels of h_P, the spillover has no role in explaining earnings differences. Also, earnings are not affected by a child's initial level of human capital directly, once controlling for schooling (Proposition 2 and Corollary 1). Since the human capital technology parameters (α_1, α_2) are common to all individuals, the only way that schooling can have heterogeneous effects on earnings is through z's influence on the determination of S, which reveals λ.

Even when λ is known, however, the schooling regression in (11) still does not identify ν separately from ρ_{zh_P}. On the other hand, if we know exactly the underlying functions (C_1, C_2, F), a Mincer regression that controls for an individual's age and schooling would reveal that the coefficient on the human capital of the individual's parent captures only ρ_{zh_P} and completely misses ν.[11] Since earnings are entirely explained by abilities after controlling for schooling, in which all parental effects are subsumed, a linear regression only reveals the relationship between parents and abilities.

So although they are not directly related to parents, knowledge of the functions (C_1, C_2, F) are essential for identifying λ and ρ_{zh_P}. If we could control for age and its interaction with schooling properly (the functions (C_1, C_2)), Corollary 2 would imply that we could simply regress log earn-

[11] Technically, we can also recover an estimate for the variance of abilities σ_z^2 from (13), so that we can infer not just the elasticity $\tilde{\rho}_{zh_P}$ but also the correlation ρ_{zh_P}.

ings on parents' human capital h_P, and a function of age and schooling, and the coefficient on the first variable would completely reveal ability selection. This is somewhat surprising, since it does not pose any identification problems in terms of separating selection from the causal effect ν. We will see in Corollary 3, however, that if we instead control for all levels of children's schooling with a full set of dummies, instead of relying on linearity or specific functional forms, the coefficient on parents' human capital would identify ν instead.

Corollary 3: Identification of Spillovers *Suppose we have the same sample as in Corollary 2. If we select only those children with the same level of schooling, $S_i = \hat{S} \neq 0$, then for any $a \in [6+S, R)$, earnings depend only on h_P through (ν, λ) and nothing else:*

$$e_i(a|S_i = \hat{S}) \propto h_{Pi}^{\frac{\nu}{1-\lambda(1-\alpha)}}. \tag{14}$$

So if we regress

$$\log e_i(a|S_i = \hat{S}) = b_0 + b_1 C_2(a; S_i) + b_2 \log h_{Pi} + \epsilon_i,$$

we recover

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

So if α is known, and since λ is identified from Corollary 2, ν is recovered from a Mincer regression that includes a complete set of dummies for all levels of schooling.

Proof. Suppose we observe two age $\hat{a}$ individuals with the same level of schooling but different levels of age $\hat{a}$ earnings and parents with different levels of human capital, denoted by (S_1, S_2), (e_1, e_2), and (h_{P1}, h_{P2}), respectively. Let (z_1, z_2) denote their unobserved learning abilities. Then by Corollary 1, since $S_1 = S_2$,

$$\frac{e_1}{e_2} = \left(\frac{z_1}{z_2}\right)^{\frac{1}{1-\alpha}},$$

and by Proposition 1,

$$\left(\frac{z_1}{z_2}\right)^{1-\lambda(1-\alpha)} = \left(\frac{h_{P1}}{h_{P2}}\right)^{\nu(1-\alpha)} \qquad \Rightarrow \qquad \frac{e_1}{e_2} = \left(\frac{h_{P_1}}{h_{P_2}}\right)^{\frac{\nu}{1-\lambda(1-\alpha)}}$$

which is (14). □

Corollary 3 shows that our model has clear implications for the level and steepness of the age-earnings profiles of two individuals with the same years of schooling but parents with different levels of human capital: the steepness of the profiles should be identical while differences in parents' human capital manifest as level differences. In other words, the profiles should be parallel with constant gaps, and given λ, this gap identifies ν.

The next natural question is whether the posited structure of the model is empirically reasonable, and why we would not just stop here and run the proposed regressions in Corollaries 2-3. We address these questions in the next section and also present some raw evidence on the relative magnitudes of ρ_{zh_P} and ν in the next section. In section 4, we present a generalized model with tastes for schooling so that we do not overestimate the structural effects from the stylized model that ignores all sources of unobserved heterogeneity other than abilities.

2.4 Discussion

This section only describes a simple, stylized model. Although we estimate a more complicated model to the HRS data in section 4, we use Corollaries 2 and 3 to discipline our choice of GMM moments, and show through simulations that the basic intuition carries over. However, the main takeaway from these corollaries is more than whether our model is a true representation of the data: it shows that in order to identify the causal effect of parents on children's earnings, we first need to correctly specify the relationship between the child's own schooling and earnings, and also between early childhood and later human capital accumulation. Otherwise, even if we had an ideal instrument for h_P, we would not be able to identify the causal effect on earnings. Since an exogenous increase in h_P would affect earnings not only directly but also indirectly through h_0 and S, we would not be able to identify the causal effect without knowledge of the latter indirect channels.

In the simple model, learning ability and parents' human capital alone accounts for the observed correlation of children's schooling and earnings with parents' schooling. Later we add tastes for schooling to account for additional sources of unobserved heterogeneity. Notably missing from our model is a consideration for life-cycle uncertainty and short-run borrowing constraints, which we abstract from for a variety of reasons.

Since we focus on how intergenerational linkages affect childhood, schooling outcomes and lifetime earnings, life-cycle uncertainty during the working stage is not of foremost interest. Also, individual age-earnings profiles are noisy in the very beginning and end of the life-cycle, but quite stable in between ages 23 and 42. Most individuals (in the HRS cohort) display stable earnings profiles from age 23 onward, peaking around age 42 after which it flattens out (until retirement behavior begins causing irregularities). Given this smoothness of individual profiles during these ages, which we use as our GMM moments, it is unlikely our estimates are affected by working-age borrowing constraints.

There is also substantial evidence that short-run constraints were not a major factor for higher education outcomes in the U.S., at least prior to the 1990s (Carneiro and Heckman, 2002; Belley and Lochner, 2007). While we do not have direct evidence for the HRS cohort that we estimate our model to, these individuals would have gone to college in the 1950s, a period when foregone earnings would have been more relevant for their education decision than direct costs. Moreover, the major dividing factor in this period was whether an individual finished high school, not college-entry (Taber, 2001), and it is unlikely that secondary schooling costs were binding for schooling choices in this period. There is also evidence that schooling outcomes not rationalized by expected earnings are better explained by non-pecuniary costs for schooling rather than borrowing constraints (Cameron and Taber, 2004; Heckman et al., 2006), which we do incorporate into our estimated model in section 4. In line with such studies, we also find that such taste heterogeneity is indeed important to account for schooling outcomes.

Borrowing constraints for the parent generation may affect, however, how the parent's human capital affects early human capital formation of the child, so some caution is required when interpreting the parameter ν. In addition to a purely exogenous spillover effect, ν would also capture how investment opportunities differ by parents with different levels of education. Hence ν should be interpreted as a reduced form representation capturing both an exogenous spillover and possibly borrowing constraints during the parent's young adulthood, not a structural representation of an early childhood human capital production function. Nonetheless, since we add taste heterogeneity into the estimated model and allow it to be correlated with parents' education, all other unobserved heterogeneity pertaining to parental background would be absorbed there, and the effect captured by ν would be restricted only to pecuniary ones.

3 Data Analysis

While the Health and Retirement Study (HRS) is usually used to study elderly Americans close to or in retirement, there are several features that make it suitable for studying intergenerational effects in the context of our model. First, these older individuals and their parents were less affected by compulsory schooling regulations and other government interventions, and more than half never advanced to college. This makes the sample suitable for a model such as ours in which schooling is a continuous choice.[12] Second, education premia 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 outcomes changes much from birth year to birth year. Third, the HRS contains information on the schooling level of both parents, and augmented with restricted Social Security earnings data, we can observe individuals' entire life-cycle earnings histories.[13]

[12]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).

[13]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 can 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. More importantly though, even datasets that have both richer parental background information and life-cycle earnings histories, such as NLSY79, are usually based on individuals facing compulsory schooling regulations.

Table 1: Summary Statistics by Education

	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.

3.1 Descriptive Statistics

The 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 oversamples blacks, Hispanics, and residents of Florida. The sample is nationally representative of the American population 50 years old and above. The baseline 1992 sample that we use for our study consisted of in-home, face-to-face interviews of the 1931-41 birth cohort and their spouses, if they were married. Follow up interviews have continued every two years after 1992. As the HRS has matured, new cohorts have been added.

For the purposes of our study, we keep 5,760 male respondents born between 1924 and 1941 from the 1992 sample.[14] We further drop 646 individuals with missing information on their own education or mother's years of schooling. This leaves us with 5,114 individuals. Table 1 describes this sample by level of education. Children and mom's schooling are about 12.3 and 9.2 years, respectively, both with a standard deviation of approximately 3.5 years.

A large fraction of HRS respondents gave permission for researchers to gain access, under tightly restricted conditions, to their Social Security earnings records. Combined with self reported

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

earnings in the HRS, these earnings records, although top-coded in some cases, provide almost the entire history of earnings for most of the HRS respondents. We imputed top-coded earnings records assuming the following individual log-earnings process[15]

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

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, and an unanticipated white noise component $u_{i,t}$. We employed random-effect assumptions with homoskedastic errors to estimate 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.

3.2 Evidence of Spillovers

According to our model, the spillover is subsumed in the choice of schooling (Proposition 1) and manifests as a gap in log earnings that remains constant through life (Corollary 3). So we split individuals into three subsamples depending on their own education levels, according to whether they have less than, exactly, or more than 12 years of schooling. Each group is further divided according to whether the mother has more than 8 years of schooling, corresponding to the end of junior high in most states in 1900 U.S.. Figure 1 depicts the average age-(log)earnings profiles for each subsample. The left and right panels compare children with 12 years against children with less and more than 12 years of schooling, respectively.

The profiles support the importance of parents' human capital: individuals with more educated mothers have higher earnings throughout their life-cycles. For all 3 education levels, the average log earnings profiles of children with the same education level but different mother's schooling levels are nearly parallel with a constant gap. This points to a permanent level effect that persists throughout an individual's career with no evidence of increasing steepness in the log earnings profile. It is precisely this gap that is captured by the spillover parameter ν in our model[16]. Moreover, the gaps are similar across all three categories of the children's educational attainment.[17]

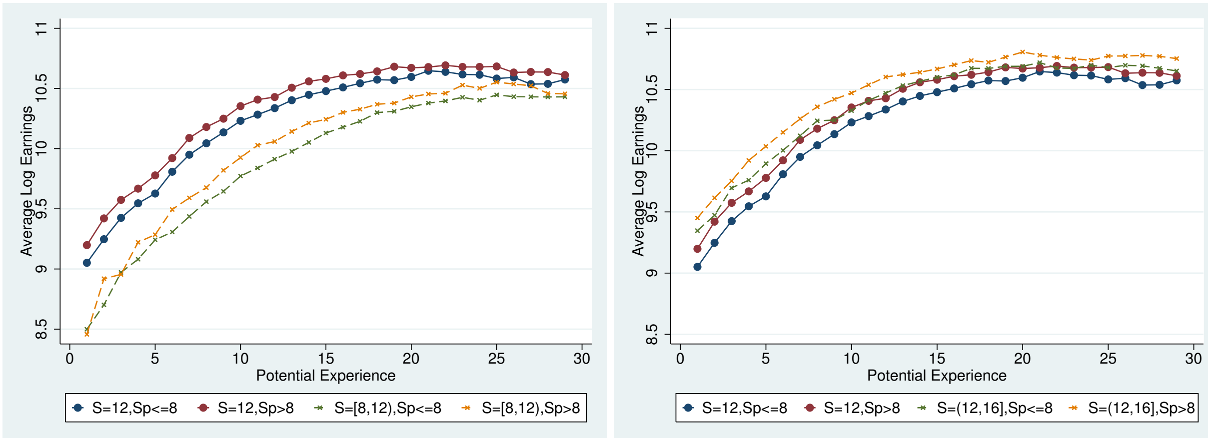
In contrast, note that the profiles of children with the same mother's schooling levels become steeper for higher education levels. This is consistent with equation (12) in Corollary 2, where

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

[16] To be precise, the gap should capture $\beta\nu / [1 - \lambda(1 - \alpha)]$, where β is defined in (18).

[17] For robustness, we have tried dividing children and their mothers according to different levels of education. The parallel gaps remain, although when we split children's education categories into very fine levels with few observations, the gaps between different levels of mother's schooling vary slightly. We also confirmed that this evidence is present in recent cohorts, through similar exercises using the NLSY79 and PSID.

Figure 1: Identifying Spillovers



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

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

earning differences across different education levels would increase in age through the function C_2. If we had perfect knowledge of the relationship between age, schooling, and earnings, and can control for this non-linearity, the differences in the controlled gaps between the profiles of children with different schooling but same mother's schooling level would identify λ (Corollary 2). Also remember that λ is also needed to quantify the size of ν. But these relationships are non-linear and difficult to control for, as we will see in the next subsection.

3.3 Mincer Regressions

Consider a standard Mincer regression augmented with parental schooling:

$$\log e_{i,a} = \beta_0 + \beta_1 S_i + \beta_2 S_{P,i} + f\left(\text{EXP}_{i,a}\right) + \epsilon_{i,a} \tag{15}$$

where $e_{i,a}$ is the earnings of individual i at age a, and $f(\cdot)$ is a flexible function of $\text{EXP}_{i,a}$, potential experience (age-6-S), which we specify in various different ways below. The variables S_i and $S_{P,i}$ denote years of schooling of individual i and individual i's parents, respectively, and $\epsilon_{i,a}$ is an error term. We estimate different specifications of (15) for earnings data from ages 23 to 42, and tabulate the results in Table 2.[18,19]

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

[19] Including race or cohort dummies did not affect our estimates much, so we did not include them in the baseline regression. Moreover, such effects would be absorbed in the unobserved heterogeneity in tastes for schooling, which is included in the estimated model, so controlling for them would make the reduced form estimates incomparable with our GMM estimates.

Table 2: Mincer Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
S	0.090	0.082	0.083	0.080	0.080	0.076	0.097	0.095	
	(99.73)	(84.21)	(82.29)	(78.14)	(78.14)	(81.02)	(35.24)	(33.10)	
Mom S_P		0.017			0.015	0.017	0.018		0.017
		(22.71)			(15.60)	(22.44)	(22.92)		(21.63)
Dad S_P			0.011		0.003				
			(15.39)		(3.45)				
Mom + Dad				0.009					
				(20.64)					
EXP$\times S$							-0.001	-0.001	
							(-6.09)	(-4.85)	
EXP$\times S_P$								-0.000	
								(-2.99)	
R^2	0.190	0.194	0.193	0.195	0.195	0.190	0.195	0.195	0.200

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

We consider four measures for $S_{P,i}$: the mother's and father's years of schooling, respectively, their sum, and also including both as separate controls. Our theory is silent on which of these measures is more appropriate. However, many studies have found that parental inputs have its strongest impact on children's human capital early on (the ν effect), e.g. Del Boca et al. (2012), who also find that mothers spend more time with their children at an early age. This leads us to suspect mothers should play the dominant role, which is confirmed in our results in Table 2.

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

The results are quite similar when we measure parents' human capital with the father's year of schooling in (3) but with an attenuated coefficient. The rate of return of paternal education is 1.1% and is statistically significant. The coefficient drops further when we measure parents' human capital as the sum of the schooling of both parents in (4). If we include both separately as in (5), mother's education is very slightly reduced from 1.7% to 1.5% while the coefficient on fathers drop significantly from 1.1% to 0.3%. This implies the lack of perfectly assortative mating, and that mom's schooling has dominant explanatory power. The relative magnitudes of the parental coefficients are consistent with previous studies that have run similar regressions. So for what follows, we take mother's schooling as the proxy for parent's human capital.

In column (6) we replaced the experience dummies with a linear and quadratic in experience, and in columns (7) and (8) we added two interactions terms to experience dummies: one between own education and experience and another between mother's years of schooling and experience. What is noticeable here is not the coefficients on the interaction terms themselves, but that the Mincer coefficient on S is as much as 2 percentage points higher than other specifications. Lastly in column (9), we controlled for individual schooling by including a full set of dummies for all observed years of schooling in the data (0 to 17) instead of linearly. Again, the coefficient on mother's years of schooling is 1.7% and highly significant.

We have also rerun these sets of regressions controlling for race and cohort dummies, and also for white males separately. The results are quite stable across all specifications. The striking feature is that no matter how we control for experience, the coefficients on S and mom's S_P are similar in magnitude and highly significant. To summarize, mothers' schooling have a stronger relationship with sons' earnings than fathers', and the estimated effect of mother's schooling, or β_2, is about 1.7%. This is in the range of previous empirical work on this topic (Card, 1999). The question is, does this reflect intergenerational persistence of abilities or causal spillovers?

Insofar as S is controlled for linearly in (2), our model would indicate that β_2 captures only ability selection (Corollary 2). On the other hand, (9) would capture spillover effects (Corollary 3). But without knowledge of the true schooling-earnings relationship, depending on what the different specifications for experience are capturing in (6)-(8), it is unclear whether the coefficients in these columns can be interpreted as selection or spillovers. The small but statistically significant interaction terms and the larger coefficients on S in (7) and (8) point toward a non-linear relationship between education and earnings. We also saw in Figure 1 that the main difference in the profiles between children with different levels of schooling but same mother's schooling was the steepness of their profiles. This may take the interpretation of (7) and (8) closer to (9), where we control for all levels of schooling, than (2). Hence, the fact that β_2 is invariant across different specifications should rather be interpreted as the overall magnitudes of selection and spillovers being similar, with each having approximately one-fifth an effect on earnings as own schooling.

The takeaway is that the interpretation of a parental effect is sensitive to the econometrician controls for age and schooling. But since schooling (S) itself is a function of abilities (z) and parents' schooling (S_P), no matter how one controls for age and schooling the estimates do not reveal the isolated causal effect of abilities nor spillovers. In terms of our model, this means that without knowing the magnitudes of the rest of the parameters in the model, in particular α, which governs the returns to human capital accumulation, and λ, ability's role in early human capital formation, we cannot say much about how much of β_2 is the selection and/or spillover effect, even in cases (2) and (9). As we emphasized in the introduction, we need to be able to exploit information on schooling and earnings outcomes jointly. The Ben-Porath model does this for us, by admitting an age-schooling-earnings relationship that has been used ubiquitously in the literature. To estimate the exact values of the spillover and its causal effect, we generalize the simple model of section 2 to bring it closer to the data.

4 Estimation

The theory and evidence up to now suggest a novel way of estimating the causal effect of parent's education on children's earnings. By conditioning on the schooling of the child, we are able to get around some of the selection issues commonly encountered in empirical work. The goal of the generalized model we estimate in this section is to retain the simplicity of the framework presented in section 2 and empirical intuition from section 3, and yet be able to obtain realistic predictions on the transmission of human capital and schooling across generations.

The underlying environment is one in which learning ability z is transmitted across generations. However, our framework does not require us to make specific assumptions on this transmission process per se. The magnitude of the spillover is determined only the state of the child, which consists of his or her learning ability z and the human capital level of the parent h_P. Hence, for our purposes all that is required is the estimation of ρ_{zh_P}, the correlation between z and h_P. Since an overriding objective is to fit the distribution of schooling, we will also include a taste for schooling in the objective function. As we already argued, many previous studies find that non-pecuniary benefits, in addition to learning ability (selection), play an important role in rationalizing schooling decisions.

In line with the previous sections, the counterfactual experiments in this section show that the coefficient on mother's schooling in a Mincer regression captures selection, in the sense that when ρ_{zh_P} is set to zero, the counterfactual coefficient is also close to zero. We also show that the coefficient changes little even if we set $\nu = 0$. Nonetheless, we show that a counterfactual increase in mother's schooling leads to a 1.2% increase in children's earnings controlling for selection on average, while further allowing for selection leads to an additional 1.3% increase. Furthermore, this happens without increasing children's schooling. This is explained by parents with higher human capital having a negative effect on children's schooling, as we saw in Proposition 1, while children with higher human capital parents having higher tastes for schooling is what drives the observed positive intergenerational relationship in schooling. Given this, we conduct a counterfactual schooling reform which shows that a large OLS coefficient when regressing child's schooling outcomes on parent's schooling is consistent with a negative or zero IV coefficient, but increases children's lifetime earnings by 2.8%.

4.1 A Generalized Model of Parental Spillovers

The biggest change we make in comparison to the model of section 2 is to constrain the optimal choice of schooling to be discrete (but with many nodes) to be consistent with the data. Another benefit of doing so is that we can use a nested logit model to capture unobserved, non-pecuniary benefits from schooling. The generalized model also assumes different laws of motion for human capital during schooling and on-the-job (OJT). The problem faced by an individual at age 6 can be written as

$$V(6,h_0) = \max_{S\in\{8,10,12,14,16,18\}} \left\{\tilde{V}\left(S;6,h_0\right) + \tilde{\xi}_S\right\}$$

where $\tilde{\xi} \equiv [\tilde{\xi}_S]$ is a vector that represents a non-pecuniary benefit for each level of schooling (but measured in pecuniary units) that varies across individuals and schooling levels, but stays constant throughout the life-cycle. The age 6 pecuniary value from attaining S years of schooling is

$$\tilde{V}(S;6,h_0) = \max_{\{n(a),m(a)\}} \left\{-\int_6^{6+S} e^{-r(a-6)} m(a)da + \int_{6+S}^{R} e^{-r(a-6)} h(a)\left[1-n(a)\right]da\right\} \tag{16}$$

subject to

$$\begin{aligned} \dot{h}(a) &= \begin{cases} zh(a)^{\alpha_1} m(a)^{\alpha_2}, & \text{for } a \in [6,S), \\ z[n(a)h(a)]^{\alpha_W}, & \text{for } a \in [S,R), \end{cases} \\ h(6) &= h_0 = bz^{\lambda} h_P^{\nu}. \end{aligned} \tag{17}$$

Since $\tilde{\xi}$ only affects an individual's desire to remain (or not) in school while having no direct effect on earnings, the inclusion of tastes for schooling allows the model to flexibly account for unobserved heterogeneity in schooling-earnings relationships that do not solely rely on economic factors (human capital and learning ability). This not only helps account for the data but also ties our hands to not label everything as ability selection or parental spillovers.

The only changes we have made in addition to the tastes for schooling is to explicitly split the schooling and working phase, during which the human capital accumulation technology differs. Schooling only involves goods inputs while OJT only involves time inputs. The technology in the schooling phase is identical to the simple model with $n(a) = 1$, while the working phase is identical to the the simple model with $\alpha_2 = 0$. We also allow α_W, the returns to human capital investments during the working phase, differ from the schooling phase. The parameter b that multiplies initial human capital captures the overall level of human capital in the model, while we have dropped the wage rate w since it is not separately identified from b in our partial equilibrium setup (i.e., it is not separately identified from units of human capital without modeling the demand for labor).

At age 6, an individual is completely characterized by $\left\{h_P, z, \tilde{\xi}\right\}$. If the schooling choice were continuous, we cannot derive closed form solutions for schooling and earnings as in the simple model,[20] but when S is given exogenously the resulting profile of earnings can be characterized using similar methods. In Appendix B, we characterize to solve (16) and the equations governing earnings given $\{h_P, z, S\}$, and describe how a solution is found numerically in Appendix C.

[20]This is primarily because in general, the supply of labor, $1-n(a)$, jumps from 0 to a strictly positive amount once an individual begins to work, unlike in the simple model where it increases continuously over time.

4.2 Population Distribution Assumptions

Our dataset contains information on parental schooling, children's schooling as well as complete earnings profiles of children. Since we only have information on the schooling of he parent, we approximate the parent's human capital, or earnings, of the parent by a standard Mincerian equation relating parental schooling to earnings,

$$h_P = a\exp(\beta S_P). \tag{18}$$

The only usage of β is to normalize parents' schooling and transform it into human capital units before applying them to the children's initial condition (17). Since the effect we are interested in is the causal effect of increasing mother's schooling by one year on children's earnings, this normalization is innocuous.[21] The coefficient a is normalized to 1 since it is not separately identified from b in the child's initial human capital (17).

The population distribution of $\{h_P, z, \tilde{\xi}\}$ are parametrized as follows. First, we assume that $(\log h_P, \log z)$ are joint normal, specifically

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

Given the relationship $h_P = \exp(\beta S_P)$, we know h_P once we know S_P and β. The distribution of S_P is taken from the empirical p.m.f. of mother's schooling in the HRS, so it is observed in discrete years ranging from 0 to 16. Then for each mother's schooling level and corresponding h_P, our distributional assumptions imply

$$\log z | \log h_P \sim \mathcal{N}\left(\mu_z + \rho_{zh_P}\frac{\sigma_z}{\sigma_{h_P}}\left(\log h_P - \mu_{h_P}\right), \sigma_z^2\left(1-\rho_{zh_P}^2\right) \right). \tag{19}$$

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

Schooling choices are determined as follows. We incorporate a very rich structure for taste heterogeneity to account for a large range of unobserved heterogeneity. First, the tastes for schooling $\tilde{\xi}$ vary not only with the schooling levels but also parent's human capital h_P and ability z:

$$\tilde{\xi}_S \equiv \delta_S\left(\gamma_{h_P} h_P + \gamma_z z\right) + \xi_S. \tag{20}$$

where $\xi \equiv [\xi_S]$ is 6-dimensional logit. The constants γ_{h_p} and γ_z captures the correlation between preferences for schooling, and parents' human capital h_P and ability z, respectively. We normalize $\delta_8 = 0$ and $\delta_{10} = 1$, so that only δ_S, $S \in \{12, 14, 16, 18\}$ need to be estimated.[22]

A nested logit model is used to model the unobserved taste heterogeneity ξ. Schooling de-

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

[22] Other normalizations would be colinear.

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

$$\begin{aligned} G(\xi) = \exp\Big\{ & - \left[\exp\left(-\xi_8/\sigma_\xi\zeta_h\right) + \exp\left(-\xi_{10}/\sigma_\xi\zeta_h\right) + \exp\left(-\xi_{12}/\sigma_\xi\zeta_h\right)\right]^{\zeta_h} \\ & - \left[\exp\left(-\xi_{14}/\sigma_\xi\zeta_c\right) + \exp\left(-\xi_{16}/\sigma_\xi\zeta_c\right) + \exp\left(-\xi_{18}/\sigma_\xi\zeta_c\right)\right]^{\zeta_c} \Big\} \end{aligned}$$

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

$$\tilde{u}_S \equiv \tilde{V}(S;6,h_0) + \delta_s\left(\gamma_{h_P} h_P + \gamma_z z\right)$$

for $S \in \{8,10,12,14,16,18\}$. Nesting yields the following conditional choice probabilities (CCP), given (h_P, z):

$$\Pr(S=8) = \Pr(S=8|S \in \{8,10,12\}) \cdot \Pr(S \in \{8,10,12\}) \tag{21a}$$

$$\Pr(S=8|s \in \{8,10,12\}) = \frac{\exp\left(\tilde{u}_8/\sigma_\xi\zeta_h\right)}{\exp\left(\tilde{u}_8/\sigma_\xi\zeta_h\right) + \exp\left(\tilde{u}_{10}/\sigma_\xi\zeta_h\right) + \exp\left(\tilde{u}_{12}/\sigma_\xi\zeta_h\right)} \tag{21b}$$

$$\Pr(S=14|S \in \{14,16,18\}) = \frac{\exp\left(\tilde{u}_{14}/\sigma_\xi\zeta_c\right)}{\exp\left(\tilde{u}_{14}/\sigma_\xi\zeta_c\right) + \exp\left(\tilde{u}_{16}/\sigma_\xi\zeta_c\right) + \exp\left(\tilde{u}_{18}/\sigma_\xi\zeta_c\right)} \tag{21c}$$

and

$$\begin{aligned} & \Pr\left(s \in \{8,10,12\}\right) = \qquad (22) \\ & \frac{\left[\exp\left\{\frac{\tilde{u}_8}{\sigma_\xi\zeta_h}\right\} + \exp\left\{\frac{\tilde{u}_{10}}{\sigma_\xi\zeta_h}\right\} + \exp\left\{\frac{\tilde{u}_{12}}{\sigma_\xi\zeta_h}\right\}\right]^{\zeta_h}}{\left[\exp\left\{\frac{\tilde{u}_8}{\sigma_\xi\zeta_h}\right\} + \exp\left\{\frac{\tilde{u}_{10}}{\sigma_\xi\zeta_h}\right\} + \exp\left\{\frac{\tilde{u}_{12}}{\sigma_\xi\zeta_h}\right\}\right]^{\zeta_h} + \left[\exp\left\{\frac{\tilde{u}_{14}}{\sigma_\xi\zeta_c}\right\} + \exp\left\{\frac{\tilde{u}_{16}}{\sigma_\xi\zeta_c}\right\} + \exp\left\{\frac{\tilde{u}_{18}}{\sigma_\xi\zeta_c}\right\}\right]^{\zeta_c}}. \end{aligned}$$

4.3 Generalized Method of Moments

This sub-section proposes a simple estimation of our model using the method of moments. We take advantage of uor closed-form solution, which make calculating the model moments computationally easy. The schooling and earnings moments of the extended model can be computed exactly subject only to numerical approximation error (See the Appendix C for details). We minimize the distance between sample moments and theoretical moments with respect to the model's parameters. While we could derive the likelihood of the model, we choose to use the method of moments because it allows us to derive identification from key features of the data we believe our ought to match precisely. A likelihood estimator would attempt to fit other dimensions of the data which our stylised model is not designed to match.

There are 27 parameters in total, which we denote by a vector θ. We partition θ into three

Table 3: Parameters Set a Priori

Parameter	Value	Description
r	5%	interest rate, after-tax rate of return on capital in Poterba (1998)
w	1	Wage rate per human capital unit, normalization
R	65	Retirement age, exogenous
a	1	Normalization (see text)
β	6%	Mincer return to schooling, HRS AHEAD cohorts
μ_{h_P}	0.55	Mean human capital of mothers, from data (see text)
σ_{h_P}	0.22	Standard deviation of mothers' human capital, from data (see text)
δ_8	0	Taste for 8 years of schooling, normalization
δ_{10}	1	Taste for 10 years of schooling, normalization

* See text for more details.

vectors, i.e. $\theta \equiv [\theta_0, \theta_{10}, \theta_{11}]$, where

$$\begin{aligned}
\theta_0 &= [r, w, R, \beta, \mu_{h_P}, \sigma_{h_P}, a, \delta_8, \delta_{10}] \\
\theta_{10} &= [\alpha_1, \alpha_2, \alpha_W, \nu, \lambda, b, \rho_{zh_P}, \mu_z, \sigma_z] \\
\theta_{11} &= \left[\sigma_\xi, \gamma_{h_P}, \gamma_z, \zeta_h, \zeta_c, \delta_{12}, \delta_{14}, \delta_{16}, \delta_{18}\right] .
\end{aligned}$$

The first partition, θ_0, are parameters that are set a priori. The rest of the parameters, $\theta_1 \equiv [\theta_{10}\, \theta_{11}]$, are from the simple model and the taste structure in the generalized model, respectively, and are estimated by GMM.

Parameters Set a Priori The interest rate and wage per human capital unit are fixed at constant levels, but Heckman et al. (1998) find that this misspecification barely affects a similar human capital production technology that they estimate for the NLSY79. We follow their procedure and fix the interest rate at 0.05, which is in the range of the after-tax rate of return on capital reported in Poterba (1998).[23] The wage w is normalized to 1, because it cannot be separately identified from b, the constant multiplying the initial human capital of children. Since human capital in our model is essentially efficiency wage units, without a demand side for human capital we cannot separate the average level of human capital from the wage. The retirement age, R, is fixed at the statutory retirement age of 65.

The coefficient β is recovered from running a standard Mincer regression similar to (15) for the HRS AHEAD cohorts, only without including mother's schooling. This regression is kept simple since the purpose is to induce a statistical distribution of earnings from the schooling distribution, including all endogenous effects. Technically, the parameter a would be the constant when running this regression, but as is the case with w, is not separately identified from b and hence normalized to 1. The β coefficient is quite stable across cohorts, ranging from approximately 0.04

[23]On the other hand, we did not explicitly model taxes on earnings. This is innocuous if earnings were taxed at a flat rate. Since we abstract from uncertainty, a more complicated tax structure would also not affect our results much as long as relative lifetime earnings remain similar.

to 0.06 for men and 0.05 to 0.09 for women; we fix $\beta = 0.06$. This value is not very different from the coefficients we recover from the the original HRS cohorts in Table 2 which includes more controls; our estimates are not sensitive to different values of a β within this range.

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

The tastes for 8 and 10 years of schooling, δ_8 and δ_{10}, are normalized to 0 and 1, respectively. These two parameters are not separately identified from the other taste parameters (namely, γ_{h_P}, γ_z and σ_ξ). The entire list of exogenously fixed parameters are summarized in Table 3.

Estimated Parameters We are left with 18 parameters to be estimated, the model and taste parameters $\theta_1 = [\theta_{10}\, \theta_{11}]$. These parameters are chosen by GMM so that the model reproduces a set of empirical moments of interest. The moments used in the estimation are tabulated in the last 5 columns of Table 4. The moments of interest for us are schooling and earnings outcomes by level of mother's schooling. Since we constrain schooling choices to lie on 6 grid points, rather than targeting average years of schooling we target the probability of attaining high or low levels of schooling by 6 levels of mother's schooling. For each of these 12 groups, we construct average earnings for ages 25,30,35 and 40, which are in turn computed by simply averaging an individual's earnings from ages 23-27, 28-32, and so forth.

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

Denote these target moments by $\hat{\Psi}$. We first compute the variance-covariance matrix of $\hat{\Psi}$ using 2000 bootstrap repetitions. For an arbitrary value of θ_1, we numerically compute the implied model moments, $\Psi(\theta_1)$, as described in Appendix C. The parameter estimate $\hat{\theta}_1$ is found by searching over the parameter space Θ_1 to find the parameter vector which minimizes the criterion function:

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

where W is a weighting matrix. This procedure generates a consistent estimate of θ_1. Following (**?**), we use a diagonal weighting matrix $W = diag(V^{-1})$, where V is the variance-covariance matrix of the sample auxiliary parameters. This weighting scheme allows for heteroskedasticity and it has

Table 4: Targeted Empirical Moments

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

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

better finite sample properties than the optimal weighting matrix (see (Altonji and Segal, 1996)). The minimization is performed using a Nelder-Mead simplex algorithm. While this method does not guarantee global optima, we used several thousand different starting values to numerically search over a wide range of parameter values (most of which have naturally defined boundaries). We use asymptotic standard errors for the parameter estimates, obtained from

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

as N approaches ∞, where $\hat{\theta}_1$ is the estimate, θ_1^* is the unknown, true parameter, and $N = 5114$ is the size of our sample. Let $P = 18$, $M = 57$ denote the number of parameters and moments, respectively. Then G is the $M \times P$ Jacobian matrix of $\Psi(\theta_1)$ with respect to θ_1, which we compute numerically, and $\hat{V}$ is the variance-covariance matrix of $\hat{\Psi}$.

Identification Identification is achieved by a combination of functional forms and parametric assumption. It is hard to formally prove as is usual with this class of models. However, our choice of moments is guided by intuition of how certain moments should have more influence on certain parameters, as summarized in Table 10 in the appendix.

Although we showed identification of $(\nu, \lambda, \rho_{zh_P})$ in Corollaries 2-3 with the simple model, we cannot get as tight predictions in the extended model. This is because even though incorporating taste heterogeneity and different human capital accumulation technologies in the school and working phases may seem like minor extensions, closed form solutions do not exist for a general choice of parameters.

Nonetheless, the choice of target moments in Table 4 is guided by intuition gained from the simple model. According to the corollaries, the most important parameters of interest, $(\nu, \lambda, \rho_{zh_P})$, are identified by earnings levels by mother's schooling levels, earnings levels by children's schooling levels, and the Mincer coefficient on mother's schooling in regression (2) of Table 2. We show that this intuition carries over to our extended model through counterfactuals in Table 7.

The identification argument for the exponents on the human capital production functions, $(\alpha_1, \alpha_2, \alpha_W)$, are similar to other studies using Ben-Porath type functions. Since α_W governs the speed of human capital growth in the working phase, it is identified by the slopes of the average age-earnings profile. Conversely, (α_1, α_2) determines at what level of earnings an individual begins his working phase. When early age earnings are lower, the slope of the age-earnings profile must be higher given a level of lifetime earnings. For high values of $\alpha \equiv= \alpha_1 + \alpha_2$, individuals with high S will have flatter earnings profiles, since they will have higher early age earnings. For high values of α_1, individuals with high S_P will have flatter earnings profiles, since they will benefit more from higher age 6 human capital levels in the schooling phase and thus have higher early age earnings.

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

5 Results

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

5.1 Interpreting the Parameters

Human Capital Production The parameters that govern human capital production, $(\alpha_1, \alpha_2, \alpha_W)$ are in the lower range of estimates found in the literature that use comparable Ben-Porath human capital models, e.g. Heckman et al. (1998) estimate $\alpha_W = 0.9$ in the NLSY79 for post-schooling

Table 5: Parameter Estimates

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

*Standard errors in parentheses.

human capital production. This may have to do with the fact that the HRS cohort lived in a period in which observed education returns were much lower, for example, the college premium was about 40% prior to the 1980s rising to above 100% in 2000. The returns to human capital investment are slightly larger in school ($\alpha_1 + \alpha_2 = 0.606$) than on-the-job ($\alpha_W = 0.426$). This means that for purposes of human capital accumulation, an individual would prefer to stay in school rather than work.

Parental Spillover and Early Childhood The magnitude of ν seemingly implies large spillovers—a mom with 10 percent higher human capital has a child with 8 percent higher initial human capital, controlling for ability and taste selection. More simply put, increasing mom's schooling by 1 year increases her child's initial human capital by 12.3%. On the other hand, the estimated λ is both small and almost insignificant. This indicates that parents are much more important than the child's learning abilities for early human capital formation.[24]

Although the estimated ν is large, note that by construction of the model, it must encompass any causal input that influences the child's human capital that can be explained by mother's schooling.[25] For example, if the child's human capital is sensitive to early childhood investments as in Cunha and Heckman (2007), it would imply a large ν that summarizes the compounded effects of dynamic complementarity in their model. But also note that the impact from increasing mom's h_P is attenuated later in the life, because children with higher h_0 (age 6 human capital) accumulate less human capital in school and onward.[26] This is for two reasons: i) the human cap-

[24]The only boundaries we imposed on in the estimation were that $\alpha, \alpha_W \in (0,1)$, to guarantee an interior solution for both schooling and working times, and $\nu, \lambda > 0$.

[25]The effects coming from its correlations with abilities and tastes are selection, not causal.

[26]This is somewhat reminiscent of the Perry early intervention program initially boosting children's cognitive skills,

ital accumulation technology displays decreasing returns, so they accumulate less human capital within any given time period, and moreover ii) schooling time decreases, i.e., they use the high returns technology ($\alpha > \alpha_W$) for a shorter amount of time (Proposition 1). As we show soon, the causal effect of an additional year of mom's schooling is 1.2%, much smaller than 12.3%, and most of this comes from early labor market entry rather than an increase in life-cycle earnings.[27]

One may also question that we recover a large estimate because we have assumed that parents only have a level effect, i.e., that h_P only has a causal effect on age 6 human capital but not learning abilities z. As we discussed in footnote 7, the problem with including such a slope effect is that it is not separately identified from ρ_{zh_P} in the simple model of Section 2. We have run several numerical simulations with the extended model to verify that assuming a slope spillover has small influence on all other parameter estimates except ρ_{zh_P}.[28] This implies that a slope spillover would only crowd out selection effects, so our results can be viewed as a conservative lower bound estimate for parental spillovers.

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

Unlike the spillover, this is a permanent difference that sustains through life. The impact on earnings at all ages, according to Lemma 4 in Appendix B, is similar to what we found in the simple model in Corollary 1, and can be approximated by $\Delta \log z/(1-\alpha_W)$, approximately 1.3%. However, because high z children also go to school longer, the impact on lifetime earnings is a bit less at 1% on average, as we soon show in our counterfactuals. Following similar calculations, a standard deviation of 0.117 for z translates into $\sigma_z/(1-\alpha_W) = 0.204$ or a 20.4% standard deviation in earnings, once controlling for schooling.

The estimates for (ρ_{zh_P}, σ_z) can also be used to put a lower bound on the intergenerational transmission of abilities. To this end, suppose that (2) takes the form

$$\log z_{Pi} = \mu_{z_P} + \rho_P \cdot \frac{\sigma_{z_P}}{\sigma_{h_P}}(\log h_{Pi} - \mu_{h_P}) + \epsilon_{Pi}$$

where ρ_P is the correlation between the human capital and ability of the parent. If we further assume that the cross-sectional distribution of abilities and correlation structure between human capital and abilities remain stable over generations, we can impose $\rho_P = \rho$, where ρ is the cor-

but the effect quickly dissipating during school years.

[27] A large ν may also indicate a large degree of intergenerational altruism. Although we have abstracted from individuals internalizing the spillover, if the underlying environment is one in which it is internalized, parents of all generations would invest more in their children. Then, ν would be capturing a composite of spillovers (the causal effect of parents on children's earnings) and altruism (how much parents care about children). For our purposes, however, ν would still capture the causal effect as long as altruism does not vary much across the mother's education.

[28] Analytical results from the simple model and simulated results from the extended model are available upon request.

Table 6: Schooling Benefits, Average

	$S=8$	$S=10$	$S=12$	$S=14$	$S=16$	$S=18$
Present value earnings by ability quartile						
Q(0)	224,000	281,236	352,036	378,771	379,583	337,408
Q(1)	211,681	247,796	267,454	246,399	242,531	242,651
Q(2)	255,637	292,894	314,487	297,382	295,694	300,278
Q(3)	286,640	326,009	356,178	342,649	342,001	347,917
Q(4)	322,800	362,776	407,678	453,431	459,233	447,386
Nonpecuniary benefits correlated with h_P by ability quartile						
Q(0)	-	5,097	9,560	10,343	15,653	28,297
Q(1)	-	5,297	10,127	10,220	15,468	28,159
Q(2)	-	5,079	9,950	10,304	15,605	28,368
Q(3)	-	4,782	9,659	10,296	15,614	28,407
Q(4)	-	4,369	8,945	10,406	15,734	28,285
Nonpecuniary benefits correlated with z by ability quartile						
Q(0)	-	324	598	717	1,029	1,524
Q(1)	-	301	513	576	824	1,296
Q(2)	-	332	562	638	915	1,450
Q(3)	-	354	603	687	986	1,562
Q(4)	-	377	650	790	1,139	1,766

PDV pecuniary value at age 6, in 2000 USD. Non-pecuniary benefits are all relative to $S=8$.

relation between (h, z) of the children in our estimation, and $\sigma_{z_P} = \sigma_z$. Using lifetime income as a proxy for h, our model implies a correlation of $\rho = 0.794$ between the children's h and z. The implied correlation of abilities across generations in our model is then

$$\rho_z = \frac{\rho_{zh_P}}{\rho_P} \cdot \frac{\sigma_z}{\sigma_{z_P}} \approx \frac{\rho_{zh_P}}{\rho} = 0.232.$$

Since it is unlikely that the independence assumptions (1)-(3) hold exactly in reality, i.e., $\log z | \log h_P$ is likely still correlated with $\log z_P$, this number should be viewed as a lower bound.

Tastes for Schooling The relative magnitudes of the constants δ_S measure the overall non-pecuniary benefits to each schooling level $S \in \{12, 14, 16, 18\}$. For example, the relative non-pecuniary benefit from 12 as opposed to 10 years of schooling is more than 4 times as large as getting some college education (14 years) rather than starting work after high school. But a college degree (16 years) comes with a substantial benefit, while continuing further (18 years) comes with a benefit as large as graduating from high school. These values imply that the large number of high school graduates and college graduates in the data are hard to justify in our model with continuous heterogeneity in learning abilities based only on pecuniary gains, so to some extent the tastes for schooling are capturing individuals conforming to the exogenous schooling conventions.

Recall that $(1-\zeta_h, 1-\zeta_c)$ are a measure of correlation of tastes for schooling levels for high school and below, and some college and above. Hence, tastes for staying in high school are much more correlated than in college. To some degree, this must explain the small jump to δ_{14} from δ_{12}, since the two nests are uncorrelated and the possibility of choosing across nests partially compensates for smaller non-pecuniary benefit. But it also implies that some college is associated with a large enough pecuniary return so that the non-pecuniary benefits need not be large, and conversely, the large jumps at δ_{12} and δ_{16} must also reflect that the pecuniary returns are not large enough to justify stopping at 12 or 16 years of schooling. This is in contrast to recent trends in which some non-pecuniary *costs* and even larger costs are needed to rationalize some college education and 4-year college graduation (Taber, 2001), but consistent with the low college graduate premium prior to the 1980s.

The discrete schooling choices in our model are bounded between 8 and 18 years. Although years of schooling in the HRS are top-coded at 17 years, if this were a problem top-coded individuals would have had much higher earnings than lower years, and δ_{18} would not need to jump far beyond δ_{16}. But the large value of δ_{18} implies that in fact the pecuniary gains from graduate degrees or staying in college for more than 4 years are so low that a very large non-pecuniary gain is needed to justify such a choice. In contrast, the large value of $\delta_{10} = 1$ may simply imply that the pecuniary returns to schooling are large at years below 8 in the data, but that we are missing that feature by imposing a minimum of 8 years of schooling.

The most interesting parameters for our purposes are (γ_{h_P}, γ_z) that govern correlation of the tastes for schooling with the population heterogeneity in (h_P, z). Note that γ_{h_P} is much larger than γ_z, implying that non-pecuniary or non-cognitive motives for staying in school are much more influenced by parents than by a child's learning ability. Fast-learning children tend to like school more, but the major determinant is the mother, or more broadly the family background. This conforms to the notion that highly educated mothers are more likely to provide a family environment conducive for longer schooling, and also inculcate in their children a higher motivation to advance further in education.

Notice that such a desire for schooling necessarily decreases lifetime earnings, since the pecuniary gains would be higher in the absence of the non-pecuniary gain by definition of lifetime income maximization. But schooling cannot be a bad thing for future earnings; since more human capital is accumulated with a better technology than when working, later wages will always be higher. Hence, the lost earnings differential that is made up for by the non-pecuniary gain from staying in school longer must come from later labor market entry. So the tastes for schooling can also be thought of as putting a larger weight on future rather than the present discounted value of earnings, and the large value of γ_{h_P} may indicate that mother's with more schooling push their children to stay in school to be better off later in life, rather than earn a small amounts as a teenager.

Table 7: Counterfactual Effect of Spillover and Correlation Parameters.

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

A barred variable $\bar{x}$ implies that we set the value of the x, which is heterogeneous across the population, to is mean value. Corr_S and OLS_S denote, respectively, the correlation between S and S_P, and the OLS coefficient from regression S on S_P. (β_1, β_2) are the coefficients on (S, S_P) from the Mincer regression in column (2) of Table 2.

6 Counterfactuals

We conduct three counterfactuals:

1. Shutting down key parameters to verify the sources of intergenerational transmissions,
2. The main counterfactual of decomposing a one-year increase in mother's education, and
3. A counterfactual compulsory schooling reform.

Counterfactual effects on moments To interpret the effect of spillovers and ability selection in light of our empirical analysis from Section 3, it is useful to see how the gradient moments are affected when shutting down the key parameters (ν, ρ_{zh_P}).[29] These results are shown in the second panel of Table 7. The third panel tabulates the counterfactual changes in the same moments when controlling for selection on tastes for schooling. The first and second rows labeled $\gamma_{h_P}\bar{h}_P$ and $\gamma_z\bar{z}$ denote the cases where we keep all else equal and set

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

instead of (20), and the third row is when there is no selection on either (h_P, z).

Both ν and ρ_{zh_P} do little to affect β_1, the Mincer schooling coefficient, and shutting down ν has only a small effect on the Mincer parental coefficient β_2 as well. As expected, β_2 is mostly affected by ρ_{zh_P} than anything else. This is in line with our intuition from Section 2 that β_2 captures selection rather than spillovers, which pervades also into our general model. But while ability selection explains much of the linear parent's schooling-earnings relationship, it does little to affect

[29]We have tried similar exercises with λ, which had small effects because the estimate it small and less imprecise.

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 (11) in the simple model.

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

One interesting result that is not predicted by our theoretical or reduced form analysis is that in all cases where we shut down taste selection, β_1 is unaffected while β_2 increases by a third. Remember that the role of tastes for schooling is essentially neutral for earnings outcomes (the effect of own education on earnings is unchanged), but since without taste selection schooling is explained by (h_P, z) only, the parental effect on earnings (β_2) becomes larger. This is because now individuals diverge less from the schooling levels that would maximize their lifetime incomes.

6.1 Decomposing Spillover from Selection

Reduced form prediction Having confirmed that our intuition from Sections 2-3 carries over, we can apply the GMM estimates to Corollaries 2-3 and compare the model-predicted values of b_2 to the regression coefficients on S_P in columns (2) and (9), Table 2. This gives us the model and data predicted ability selection and spillover effects from having a 1-year more educated mother, according to the simple model. Since $\log h_P = \beta S_P$, the b_2's in both corollaries must be multiplied by $\beta = 0.06$ to be comparable with the regression coefficients, and since we have different α's in the extended model, we can get a range by computing

$$\text{selection}: \ \beta b_2 = \beta\rho_{zh_P}/(1-\alpha) \qquad = \begin{cases} 0.035 & \text{if } \alpha = \alpha_1 + \alpha_2 \\ 0.024 & \text{if } \alpha = \alpha_W \end{cases} \tag{23}$$

$$\text{spillover}: \ \beta b_2 = \beta\nu/\left[1-\lambda(1-\alpha)\right] \approx 0.048 \quad \text{for both cases.} \tag{24}$$

Both the implied ability selection and spillover coefficients, in particular the latter, are larger than its reduced form estimate of 0.017. All else equal, tastes for schooling induce individuals to stay in school at the the detriment of lifetime earnings. To make up for this and explain a 1.7% observed, reduced form return, the value of ρ_{zh_P} must be larger than what would be implied by the simple model. Conversely, if we ignore taste heterogeneity as in (23), the implied effect of ability selection on earnings will be counterfactually high. On the other hand, as we saw above, selection on tastes is what generates the large intergenerational schooling correlation, which needed to be moderated by a negative effect on schooling coming from a large ν. So again, if we were to ignore the taste heterogeneity as in (25), the implied spillover effect on earnings will be counterfactually high. The estimated ν has to be larger to moderate the direct effect of tastes on schooling, while the effect of tastes on earnings is only indirect, requiring only a modestly larger estimate for ρ_{zh_P}.

The relative effects of (ρ_{zh_P}, ν) can also be understood intuitively from (11). We cannot make a direct comparison without explicitly computing the function $F(\cdot)$, but given that tastes for schooling generate a very high persistence in schooling as evidenced in Table 7, the estimate for ρ_{zh_P} must remain small; otherwise the correlation and OLS coefficient would become even larger. On the other hand, a large value of ν is needed to moderate the large, positive impact that tastes have on schooling.

Counterfactual prediction The simple model predictions, while helpful for identification, does not tell us the causal effect of increasing mother's education on children's earnings (and schooling). But we can simulate the effect through counterfactuals using the estimated model. Given the individual state (h_P, z, ξ) at age 6, we compute the change in the child's schooling and earnings outcomes in response to a 1-year increase in S_P by 1 year, which translates into an increase of β units of h_P in logarithms. We do this in several ways to control for spillovers and selection effects.

First, we hold constant the individual's (z, ξ), which isolates the pure effect coming only from the spillover. Next, we let either z or ξ vary with h_P as dictated by the distributional assumptions in Section 4.2; this separately captures the selection effects from each. Finally, we let both z and ξ vary together, which we label a "reduced-form" effect—i.e., this is just comparing the average outcomes of children with S_P years of mother's schooling, to the average outcome of those with $S_P + 1$ years of mother's schooling.

Formally, for any initial condition $x = (S_P, \log z, \xi)$, the model implied schooling and age-a earnings outcomes can be written as functions of x, $S = S(x)$, $E(a) = E(x; a)$.[30] Then schooling and age a earnings following a j-year increase in S_P, holding (z, ξ) constant, are

$$S_\nu^j(x) \equiv S(S_P + j, z, \xi), \qquad E_\nu^j(x; a) \equiv E(S_P + j, z, \xi; a). \tag{25}$$

Selection on abilities and tastes associated with a j-year increase in S_P can be written as

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

respectively, where $\Delta_\xi^j(S_P)$ is a 6-dimensional vector for each level of schooling $S \in \{8, \dots, 18\}$. The first expression follows since $(S_P, \log z)$ are joint-normal, and the second from the definition

[30] Since although the parent variable in the initial condition is h_P, it is defined as $\log h_P = \beta S_P$ in (18).

Table 8: Aggregate Effect of 1 Year Increase in Mom's Schooling.

	fixed S	spillover	ability	tastes	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

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

of tastes in (20). Then schooling and age a earnings following a j-year increase in S_P, including partial selection effects on z or ξ, are

$$S_z^j(x) \equiv S(S_P + j, z \cdot \Delta_z^j, \xi), \qquad E_z^j(x; a) \equiv E(S_P + j, z \cdot \Delta_z^j, \xi; a), \tag{26}$$

$$S_\xi^j(x) \equiv S(S_P + j, z, \xi + \Delta_\xi^j(S_P)), \qquad E_\xi^j(x; a) \equiv E(S_P + j, z, \xi + \Delta_\xi^j(S_P); a). \tag{27}$$

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

$$S_{rf}^j(x) \equiv S(S_P + j, z \cdot \Delta_z^j, \xi + \Delta_\xi^j(S_P) + \Delta_{z\xi}^j(z)) \tag{28a}$$

$$E_{rf}^j(x; a) \equiv E(S_P + j, z \cdot \Delta_z^j, \xi + \Delta_\xi^j(S_P) + \Delta_{z\xi}^j(z); a), \tag{28b}$$

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

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

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

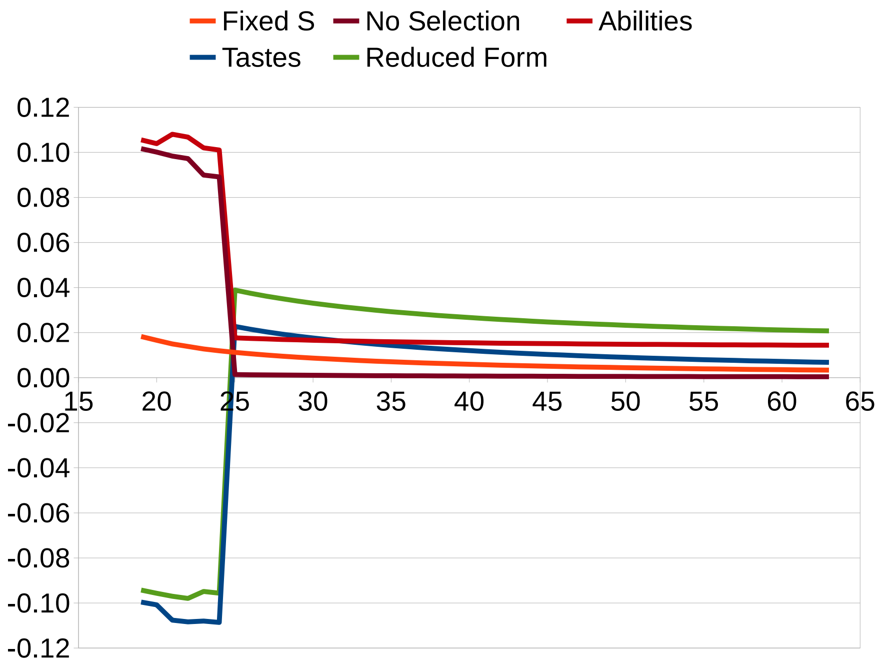
The first row of Table 8 lists the average spillover and selection effects following a 1 year increase in S_P, obtained by integrating the change from $S(x)$ to (25)-(28) over the population distribution Φ. The second row lists the average spillover and selection effects, in log-points (to approximate percentage changes) on the present-discounted value of lifetime earnings:

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

for $k \in \{v, z, \xi, rf\}$.

As expected, the spillover has a negative effect on schooling. What may be slightly surprising is that allowing for selection on abilities only moderates this by 0.039 years (comparing the 1st and 2nd columns) or 0.028 years (3rd vs. 4th columns). As we saw in Table 7 earlier, selection on abilities did not have much of an effect on intergenerational schooling relationships once selection on tastes were taken into account. Since tastes already induce individuals to stay in school longer

Figure 2: Lifecycle Effect of 1 Year Increase in Mom's Schooling.



than what maximizes lifetime earnings, the fact that abilities become higher does not further increase schooling much. In addition, 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), although this effect is likely small given the small value of $\lambda = 0.06$.

As was implied from the previous counterfactual of shutting down the taste correlations with other variables, only when we allow for selection on tastes do we see a positive relationship between mother's schooling on child's schooling. This relationship is large (0.876 years), but is moderated by the negative spillover effect. Even ignoring selection on abilities, the parental spillover and selection on tastes generate a level of intergenerational schooling that is close to its empirically observed OLS relationship.

In contrast, the spillover has an approximately 1.2 percent positive effect on lifetime earnings, while selection on abilities have a 1.3 percent effect. We conclude that independently of selection on tastes, the causal effect of mother's education on earnings is more or less similar to ability selection, i.e., highly ability mothers having high ability children. Selection on tastes have a negative average impact: the increase in lifetime earnings drops by 1 percentage point once we allow for selection (1st vs. 3rd columns). The fact that the effect is negative is expected, since tastes for schooling make individuals deviate from lifetime earnings maximization. What is surprising is that this effect is so large that it almost dominates the spillover.

However, it is important to remember that this is only an average effect over the entire lifecycle; the earnings effects differ substantially by age, and also across children of mother's with different levels of schooling.[31] Such life-cycle effects are depicted in Figure 2. Each line plots

$$\log\left[\int E_k^1(x;a)d\Phi(x)\right]-\log\left[\int E(x;a)d\Phi(x)\right]$$

for $k \in \{\nu, z, \xi, rf\}$. Longer schooling induced by high tastes increase lifetime earnings later in life (through more human capital accumulated in school), but this is dominated by the foregone earnings earlier in life. Conversely, the spillover effect comes almost entirely from early labor market entry, while there is no change in earnings after age 24 (the latest age we allow labor market entry in the model). This means that on average, children of mothers with less schooling catch up with those of mothers with more schooling by staying in school longer. In Figure 7 in the appendix we show the median life-cycle effects, which basically tell the same story.

In Table 4, we saw that although schooling and average earnings increase with mothers' schooling level over all, children of mothers with very high schooling ($\geq$ 13 years) earn less than children of mothers with 12 years of schooling. In fact, their earnings levels are similar to those whose mothers' schooling is between 9 and 11 years. The implications of this is born out in Figure 3, which depicts the cross-sectional impact of spillovers, abilities and tastes, in response to a one year increase in mom's schooling. The top panel is for schooling, where each bar plots

$$\log\left[\int S_k^1(x)d\Phi(\hat{S}_P \in M, \hat{z}, \hat{\xi})\right]-\log\left[\int S(x)d\Phi(\hat{S}_P \in M, \hat{z}, \hat{\xi})\right]$$

and the bottom panel for lifetime earnings, which plots

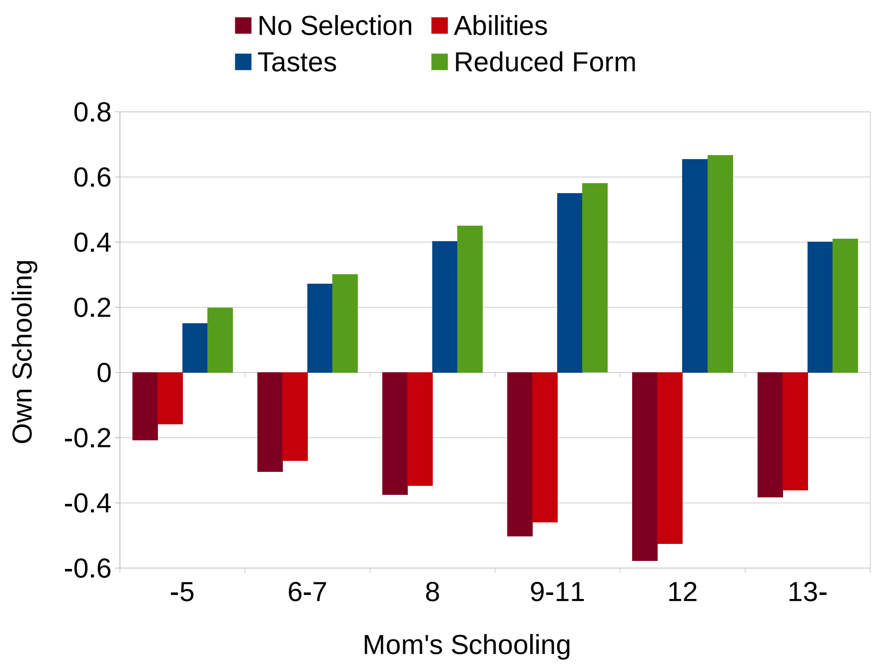
$$\begin{aligned}&\log\left[\sum_{a=14}^{R}\left(\frac{1}{1+r}\right)^{a-14}\int E_k^1(x;a)d\Phi(\hat{S}_P \in M_s, \hat{z}, \hat{\xi})\right]\\ &-\log\left[\sum_{a=14}^{R}\left(\frac{1}{1+r}\right)^{a-14}\int E(x;a)d\Phi(\hat{S}_P \in M_s, \hat{z}, \hat{\xi})\right],\end{aligned}$$

for $k \in \{\nu, z, \xi, rf\}$ and the 6 mothers' schooling group M_s, in the first column of Table 4. The structural impact of spillovers is always negative on schooling and positive on earnings; ability selection has a positive impact on both; and taste selection always has a positive impact on schooling at the expense of a negative impact on earnings.

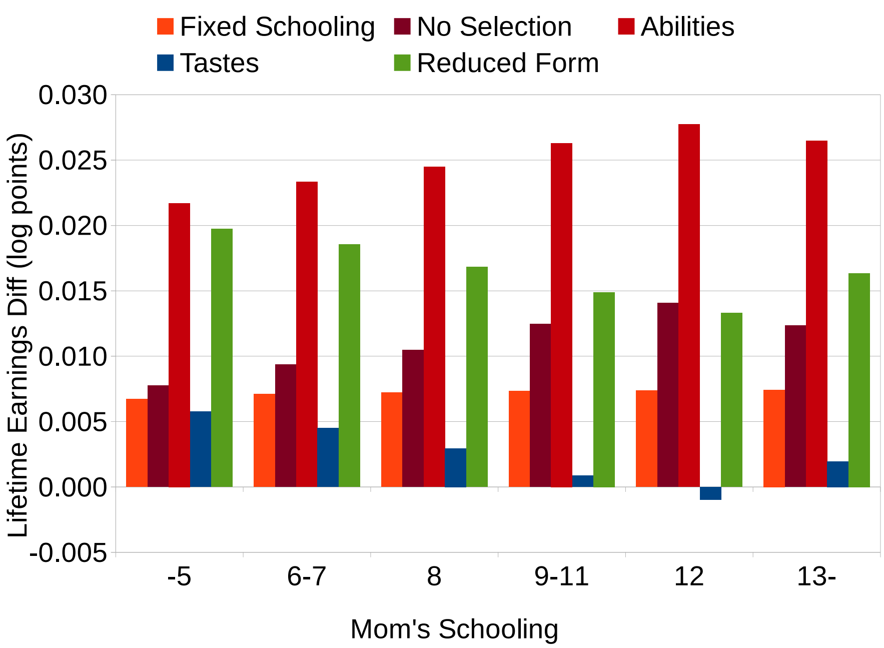
Note that ability selection has a similar impact on schooling and earnings for all levels of S_P (the gaps between the first 2 bars, and the gaps between the next 2 bars). But both spillovers and taste selection has an increasing impact up to 12 years of S_P. In particular, the parental spillover on earnings is increasing by inducing less schooling, despite the human capital technology displaying decreasing returns. This means that preferences for longer schooling are increasing faster than the spillover in the population. Since high S_P children deviate farther from their lifetime earnings

[31] This may also have to do with the fact that the HRS cohort faced low education returns compared to recent cohorts. In the data, the most highly educated individuals do not earn more than their slightly less-educated counterparts. As discussed in Section 5.1, large, positive tastes for schooling are needed to justify why these individuals obtained higher education despite the low returns.

Figure 3: Cross-section Effects of 1 Year Increase in Mom's Schooling



(a) Children's Schooling



(b) (log) Lifetime Earnings

maximizing choice, an increase in S_P controlling for tastes has a more negative causal impact on schooling and larger, positive impact on lifetime earnings.

But in the last cell, this trend is reversed. Children in this category are those with the highest tastes for schooling, and are closer to the upper bound of 18 years of schooling. Their observed choices deviate less from what maximizes lifetime earnings, and also decreasing returns to human capital accumulation sets in more strongly. As a result, both the selection and spillover effects become smaller. Conversely, for the same reasons, although children of less educated mothers experience the smallest impact from a 1-year increase in their mothers' schooling, the reduced form impact on their earnings is the largest since they are the least affected by preferences for longer schooling.

Counterfactual Compulsory Schooling Reform We saw above that the fact that even when the causal effect of mother's education on children's earnings is positive, the effect on children's schooling can be negative. The observed, positive intergenerational schooling correlation is instead explained by unobserved heterogeneity in tastes for schooling. Having understood the marginal effects of increasing a parental schooling by 1 year, we now conduct a counterfactual experiment by imposing a minimum schooling requirement, which is intended to mimic compulsory schooling reforms that took place in many countries throughout the 20th century. A minimum schooling requirement has heterogeneous affects across the population since it only affects those parents who would otherwise not attain the required level of schooling, and even within this group, the additional number of years that is attained will vary. The goal of the exercise is to show that a large schooling OLS coefficient is consistent with a small or negative IV coefficient, and the evidence for spillovers would be found in children's earnings, not schooling.

We simply impose a minimum 8 years of schooling for all parents. The initial level of human capital in the economy is set to

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

We choose 8 years as the hypothetical schooling requirement because 8 years already was the compulsory schooling requirements in many U.S. states at the time, or soon after.[32] Such a reform would affect 25% of the individuals in our data, increasing the schooling of their mothers by an average of 3.5 years.

For the schooling OLS and IV regressions, we combine samples of two regimes: one without the minimum requirement (our benchmark model) and one imposing the requirement. These represent mother-child pairs pre- and post-reform, respectively. Then we run both an OLS and IV regression on the merged data, using a dummy variable for the different regimes as an instrument.

[32] In contrast, Black et al. (2005) study the case of Norway, whose compulsory schooling requirement went up from 7 to 9 years in the 1960s. Of course, while the location and timing differs (many of the parents in our data would have been in school at turn of the 20th century), because the U.S. was a forerunner of public schooling (Goldin and Katz, 2007), mothers' average years of schooling is only 1 year less (10.5 years vs 9.3 years). However, the percentage of the population that would be affected is almost two-fold (12.4% vs 25%).

Table 9: Counterfactual Schooling Reform

		fixed S	spillover	ability	tastes	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

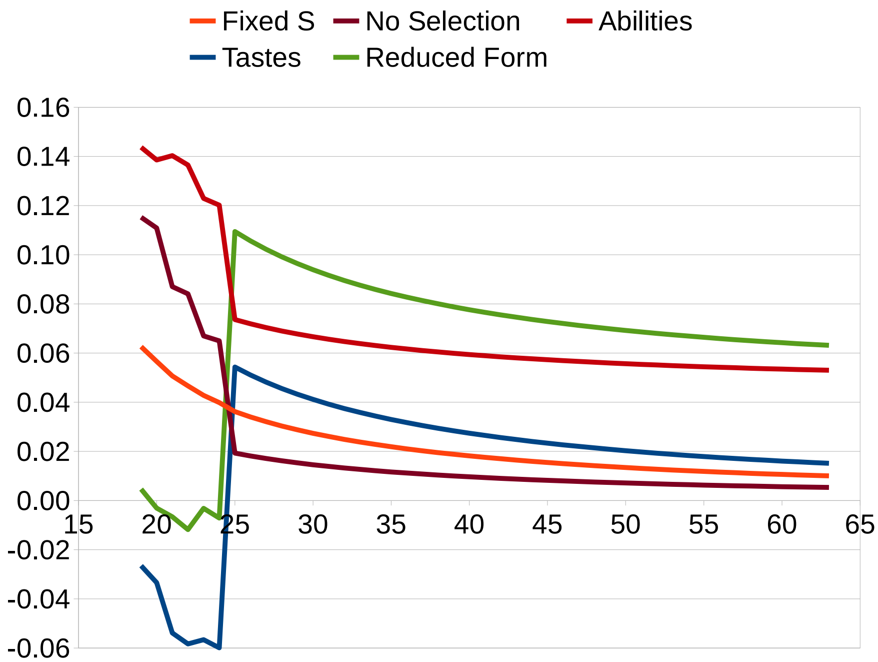
The second row denotes the change in the cross section average of the present discounted value of lifetime earnings, in logarithms, for only those individuals affected by the reform. The first column holds abilities and tastes constant, while the next two columns let ability or taste also vary according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

As above, we repeat this exercise for four cases controlling for selection on z and ξ, and including the selection effects from z and/or ξ. The top panel in 9 shows the regression results for all cases. The bottom panel shows the change in the present discounted value of lifetime earnings, in log-points (to approximate percentages).

The OLS coefficients are somewhat difficult to interpret, since the constants in the regressions are also changing. But with low-educated parents no longer in the sample post-reform, the OLS increases in all 4 cases to a value slightly higher than in the benchmark (0.412). The IV regressions, which measures the average 1-year effect among children whose mother's became more educated due to the reform, basically reflects the results from Table 8 qualitatively. The controlled effect is negative, but of smaller magnitude. The IV coefficient increases when including ability selection (column 1 vs. column 2) but much more when including taste selection (column 1 vs. column 3). The reduced form effect is only half of the population reduced form effect following a 1-year increase, in Table 8. As discussed there, this is because both the spillover and selection on tastes has a smaller effect for lower levels of S_P, who are the only ones affected by the reform.

We conjecture that this partially explains the puzzling fact that many studies using special data sets on twins, adoptees, or compulsory schooling reforms find a zero or negative effect of parents' schooling on children's schooling. First, the IV coefficients are small in absolute magnitude in columns 1 and 3, because children of less educated mothers enjoy less spillovers from their parents but also lose less in terms of lifetime earnings due to tastes. Second, the structural effect may in fact be negative, as we have argued throughout this paper. The fact that in some studies the effect is found to be close to zero but not negative, but nonetheless small, can be due to the fact that rather than the schooling tastes for children not being affected at all, as in the hypothetical schooling reform of column 3, forcing mothers to obtain more schooling does induce their children to develop higher tastes for schooling to a certain degree. We can interpret this as mothers who are forced to attend school longer, but would not have attained higher levels of schooling otherwise, having some impact on children's non-cognitive abilities or perception of schooling that help them stay in school, although perhaps not to the extent that mothers who choose to become highly educated transmit high tastes for schooling. In our experiment, that would mean that education for mothers who would otherwise attain very low levels of schooling could increase the schooling

Figure 4: Lifecycle Impact of Reform



of their children anywhere from -0.2 to 0.2 years.[33]

The negative structural effect that comes from a quantity-quality trade-off in the effect of mother's schooling, and the positive effect that comes from higher tastes for schooling, is a qualitatively different explanation to explain intergenerational schooling relationships. Previous studies implicitly assume that a causal effect should be positive, and have been sought for pecuniary or cognitive answers such as assortative mating. Some studies have posited that increased female labor market participation of more highly educated women may have reduced their time spent with children, thereby reducing their early education levels, but recent data tells us that more educated women in fact spend more, not less, time with children in their first formative years of life. Our findings indicate that in order to understand intergenerational schooling relationships, one should not assume that the causal effect should necessarily be positive, and that we need to investigate deeper into the non-pecuniary or information effects that a parent's education has on their children. For example, if schooling has non-cognitive benefits that are not captured by earnings, the total causal spillover would be larger than just the pecuniary benefit gained by boosting early human capital formation. But if our tastes for schooling parameters are capturing children being over-optimistic about the pecuniary gains from longer schooling, and such beliefs are correlated with their parents' education levels, more caution needs to be taken when thinking about intergenerational education spillovers.

That such an investigation would be important becomes more apparent when observing the life-cycle earnings effects. The isolated spillover effect on lifetime earnings is 2.8%, on average,

[33]This explanation would reconcile **?**'s finding of a zero IV with **?**'s finding that compulsory schooling reforms in the U.S. decreased the number of the reform cohort's children repeating grades.

although this is in response not to a 1-year increase in mom's schooling but an average of 3.5 years. It is also important to note that the spillover effect is positive through life, as can be seen in Figure 4, unlike in Figure 2 where it became virtually zero at ages 24 and above. The negative effect that comes from taste selection is still there, meaning that, if mothers with low education, following a reform, raise their children's tastes for schooling, we may expect children's schooling *and* lifetime earnings to drop. But the negative effect is smaller in magnitude (although the lifetime earnings effect is -1.1 percentage points for both Tables 8 and 9, mother's schooling increases by 3.5 years in the latter compared to 1 year in the former). Part of this is because the earnings they could have made through early labor market entry is smaller for these children, which is even more obvious when we look at median earnings in Figure 8 in the appendix. There, at nowhere during the life-cycle does the reform have a negative effect, not even at early ages when including selection on tastes.

This has policy implications for public education. On the one hand, if we are only concerned with the child's lifetime earnings, or if tastes for schooling represent misinformation, public education enforcement should be careful not to "over-educate" individuals by over-emphasizing the causal effect of schooling on earnings. On the other hand, if parents are less concerned about sending a teenager to work and more about a child having higher earnings past his mid-20s, passing on non-cognitive skills or beliefs that schooling is important is worthwhile even for children's earnings, as can be seen in Figure 4. It may be even more worthwhile given that the earnings they forgo as a teenager would be small.

7 Conclusion

In this paper we presented a Ben-Porath human capital model of endogenous schooling and earnings to isolate the causal effect of parent's education on children's education and earnings outcomes. Despite the positive relationship between the child's own schooling and earnings, the causal effect of parent's schooling on children's schooling can be negative, even when the causal effect is positive for children's earnings. Children of higher human capital parents begin life with higher human capital themselves, and when schooling is endogenous, they can spend less time in school but still attain the same or higher level of earnings. A simple version of the model was solved in closed form and its implications compared to empirical evidence in the HRS data.

We argue that without controlling for its quality, schooling outcomes may not be a good indicator of intergenerational effects since years of schooling is only a quantity measure. Conversely, although earnings is closer to a quality measure, years of experience must be taken into account when studying parental effects. Indeed in our estimation, the causal spillover effect of parents on children's earnings came mostly from expediting labor market entry. Another important finding from the estimated model is that unobserved correlation between mothers' education and children's tastes for schooling is the main determinant of children's schooling, not the causal effect and not selection on abilities.

The estimated causal effect of mother's education on children's lifetime earnings is found to be

1.2%, compared to 1.3% that can be explained by ability selection across generations. 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, who find that nurture effects are at most similar or less than nature effects.

Since this stems from our consideration of a quantity-quality trade-off in education and unobserved heterogeneity in tastes for schooling, more attention may need to be taken to construct qualitative measures for education attainment and how parents affect the non-cognitive abilities of children and their beliefs about individual-specific education returns. There already is a blossoming literature on estimating major-specific returns to education and how much students know about such returns. A separate literature studies the non-cognitive benefits to early childhood interventions, how mothers and fathers jointly influence children’s perceptions and changes of educational attainment, as well as the effect of borrowing constraints faced by parents with such younger children. But as of yet there has been few attempts to link these these findings to intergenerational transmissions. A deeper consideration for such factors will surely imply larger returns for parental spillovers.

Appendices

A Proofs to Propositions 1, 2, and Corollary 1.

The proof requires a complete characterization of the income maximization problem. While we can use standard methods to obtain the solution, we do this elsewhere and in what follows simply guess and verify the value function. For notational convenience, we drop the age argument a unless necessary. We separately characterize the solutions before and after the constraint $n \leq 1$ is binding in Lemmas 1 and 2. Then schooling time S is characterized as the solution to an optimal stopping time problem in Lemma 3. To this end, we further assume that

$$
\begin{aligned}
V(a,h) &= q_2(a)h + C_W(a), && \text{for } a \in [6+S, R), \\
V(a,h) &= q_1(a)\cdot\frac{h^{1-\alpha_1}}{1-\alpha_1} + e^{-r(6+S-a)}C(S,h_S), && \text{for } a \in [6, 6+S), \text{if } S > 0,
\end{aligned}
$$

where

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

for which the length of schooling S and level of human capital at age $6+S$, h_S, are given, and C_W is some redundant function of age. Given the forms of $g(\cdot)$ and $f(\cdot)$, these are the appropriate guesses for the solution, and the transversality condition becomes $q(R) = 0$. Given the structure of the problem, we first characterize the working phase.

Lemma 1: Working Phase *Assume that the solution to the income maximization problem is such that $n(a) = 1$ for $a \leq 6+S$ for some $S \in [0, R-6)$. Then given $h(6+S) \equiv h_S$ and $q(R) = 0$, the solution satisfies, for $a \in [6+S, R)$,*

$$
q_2(a) = \frac{w}{r}\cdot q(a) \tag{29}
$$

$$
m(a) = \alpha_2\left[\kappa q(a) z\right]^{\frac{1}{1-\alpha}} \tag{30}
$$

$$
h(a) = h_S + \frac{r}{w}\cdot\left[\int_{6+S}^{a} q(x)^{\frac{\alpha}{1-\alpha}}dx\right]\cdot(\kappa z)^{\frac{1}{1-\alpha}} \tag{31}
$$

and

$$
\frac{wh(a)n(a)}{\alpha_1} = \frac{m(a)}{\alpha_2}, \tag{32}
$$

where

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

Proof. Given that equation (5) holds at equality, dividing by (6) leads to equation (32), so once we know the optimal path of $h(a)$ and $m(a)$, $n(a)$ can be expressed explicitly. Plugging (5) and the guess for the value function into equation (7), we obtain the linear, non-homogeneous first order differential equation

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

to which (29) is the solution. Using this result in (5)-(6) yields the solution for m, (30). Substituting (29), (30) and (32) into equation (8) trivially leads to (31). □

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

Lemma 2: Schooling Phase *Assume that the solution to the income maximization problem is such that $n(a) = 1$ for $a \in [6, 6+S)$ for some $S \in (0, R-6)$. Then given $h(6) = h_0$ and $q_1(6) = q_0$, the solution satisfies, for $a \in [6, 6+S)$,*

$$q_1(a) = e^{r(a-6)}q_0 \tag{33}$$

$$m(a)^{1-\alpha_2} = \alpha_2 e^{r(a-6)} \cdot q_0 z \tag{34}$$

$$h(a)^{1-\alpha_1} = h_0^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left[e^{\frac{\alpha_2 r(a-6)}{1-\alpha_2}} - 1 \right] \cdot (\alpha_2 q_0)^{\frac{\alpha_2}{1-\alpha_2}} z^{\frac{1}{1-\alpha_2}}. \tag{35}$$

Proof. Since $n(a) = 1$ during the schooling phase, using the guess for the value function in (7) we have

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

to which solution is (33). Then equation (34) follows directly from (6), and using this in (8) yields the first order ordinary differential equation

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

to which (35) is the solution. □

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

Lemma 3: Value Matching and Smooth Pasting *Assume $S > 0$ is optimal. Then (q_0, h_S),* *are given by*

$$q_0 = \frac{e^{-rS}}{\alpha_2^{\alpha_2}} \cdot \left([\kappa q(6+S)]^{1-\alpha_2} z^{\alpha_1} \right)^{\frac{1}{1-\alpha}} \tag{36}$$

$$h_S = \frac{\alpha_1}{w} \cdot [\kappa q(6+S) z]^{\frac{1}{1-\alpha}}. \tag{37}$$

Proof. The value matching for this problem boils down to setting $n(6+S) = 1$ in the working phase, which yields (37).[34] The smooth pasting condition for this problem is

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

Using the guesses for the value functions, we have

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

and by replacing h_S with (37) we obtain (36). □

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

Proof of Proposition 1. The length of the schooling period can be determined by plugging equations (36)-(37) into (35) evaluated at age $6+S$:

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

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

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

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

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

$$z^{1-\lambda(1-\alpha)} h_P^{-\nu(1-\alpha)} \leq F(S), \tag{38}$$

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

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

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

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

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

and S is determined by (38) at equality. The full solution is given by Lemmas 1-3 and we obtain Proposition 2 and Corollary 1. Otherwise $S = 0$ and the solution is given by Lemma 1. □

B Analytical Characterization of the Generalized Model

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

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

$$V(a,h) = \frac{w}{r} \cdot q(a)h + D_W(a) \tag{39}$$

and the solution is characterized by

$$n(a)h(a) = \left[\frac{\alpha_W}{r} \cdot q(a)z\right]^{\frac{1}{1-\alpha_W}} \tag{40}$$

$$h(a) = h_J + \left(\frac{\alpha_W}{r}\right)^{\frac{\alpha_W}{1-\alpha_W}} \cdot \left[\int_{6+S+J}^{a} q(x)^{\frac{\alpha_W}{1-\alpha_W}} dx\right] \cdot z^{\frac{1}{1-\alpha_W}}, \tag{41}$$

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

$$V(a,h) = e^{r(a-6-S)} q_S \cdot \frac{h^{1-\alpha_W}}{1-\alpha_W} + e^{-r(6+S+J-a)} D(J, h_J) \tag{42}$$

where

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

while human capital evolves as

$$h(a)^{1-\alpha_W} = h_S^{1-\alpha_W} + (1-\alpha_W)(a-6-S)z. \tag{43}$$

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

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

$$h_J = \left[\frac{\alpha_W}{r} \cdot q(6+S+J) z \right]^{\frac{1}{1-\alpha_W}}. \tag{45}$$

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

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

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

$$\frac{z}{h_S^{1-\alpha_W}} \leq G(J) \equiv \left[\frac{\alpha_W}{r} \cdot q(6+S+J) - (1-\alpha_W) J \right]^{-1}. \tag{46}$$

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

$$G(0) < \frac{z}{h_S^{1-\alpha_W}} \qquad \Leftrightarrow \qquad \frac{r}{\alpha_W q(6+S)} < \frac{z}{h_S^{1-\alpha_W}}, \tag{47}$$

and J is determined by (46) at equality. Otherwise $J = 0$.

Now if S were discrete, as in the model we estimate, we only need to solve for h_S, the level of human capital at age $6+S$. Then we can solve for $V(h_0, z; s)$ for all 6 possible values of s, using Lemmas 2 and 4 for the schooling and working phases, respectively. But it is also possible to characterize the unconstrained continuous choice of S, even though a closed form solution does not exist in general. We only need consider new value matching and smooth pasting conditions.

Lemma 5: Schooling Phase, Generalized *The length of schooling, S, and level of human capital at age $6+S$, h_S, are determined by*

1. *if $J=0$,*

$$\epsilon + (1-\alpha_2)\left[\frac{\alpha_2^{\alpha_2} w}{r}\cdot q(6+S) z h_S^{\alpha_1}\right]^{\frac{1}{1-\alpha_2}} = w\cdot\left(h_S + (1-\alpha_W)\left[\frac{\alpha_W^{\alpha_W}}{r}\cdot q(6+S)z\right]^{\frac{1}{1-\alpha_W}}\right) \tag{48}$$

$$h_S^{1-\alpha_1} \leq h_0^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2}\cdot\left(1-e^{-\frac{\alpha_2 r S}{1-\alpha_2}}\right)\cdot\left[\frac{\alpha_2 w}{r}\cdot q(6+S)h_S^{\alpha_1}\right]^{\frac{\alpha_2}{1-\alpha_2}} z^{\frac{1}{1-\alpha_2}} \tag{49}$$

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

$$q_0 = \frac{we^{-rS}}{r}\cdot q(6+S)h_S^{\alpha_1}. \tag{50}$$

2. *if $J>0$,*

$$\epsilon + (1-\alpha_2)\left(\alpha_2^{\alpha_2} w e^{-rJ}\left[\frac{\alpha_W^{\alpha_W}}{r}\cdot q(6+S+J)z\right]^{\frac{1}{1-\alpha_W}}\cdot h_S^{\alpha_1-\alpha_W}\right)^{\frac{1}{1-\alpha_2}}$$
$$= we^{-rJ}\left[\frac{\alpha_W^{\alpha_W}}{r}\cdot q(6+S+J)z\right]^{\frac{1}{1-\alpha_W}} \tag{51}$$

$$h_S^{1-\alpha_1} \leq h_0^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2}\cdot\left(1-e^{-\frac{\alpha_2 r S}{1-\alpha_2}}\right) \tag{52}$$
$$\cdot\left(\alpha_2 w e^{-rJ}\left[\frac{\alpha_W^{\alpha_W}}{r}\cdot q(6+S+J)z^{\alpha_W}\right]^{\frac{1}{1-\alpha_W}} h_S^{\alpha_1-\alpha_W}\right)^{\frac{\alpha_2}{1-\alpha_2}}\cdot z^{\frac{1}{1-\alpha_2}}$$

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

$$q_0 = we^{-r(S+J)}\cdot\left[\frac{\alpha_W^{\alpha_W}}{r}\cdot q(6+S+J)z^{\alpha_W}\right]^{\frac{1}{1-\alpha_W}}\cdot h_S^{\alpha_1-\alpha_W}. \tag{53}$$

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

$$\epsilon - m(6+S) + e^{rS}q_0 z m(6+S)^{\alpha_2} = wh_S\left[1-n(6+S)\right] + \frac{w}{r}\cdot q(6+S)z\left[n(6+S)h_S\right]^{\alpha_W}$$
$$e^{rS}q_0 h_S^{-\alpha_1} = \frac{w}{r}\cdot q(6+S).$$

Hence (50) follows from the smooth pasting condition. Likewise, (48) follow from plugging $n(6+$

$S)$, $m(6+S)$ from Lemmas 2 and 4 and q_0 from (50) in the value matching condition. Lastly, (49) merely states that the optimal h_S must be consistent with optimal accumulation in the schooling phase, $h(6+S)$.

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

$$\begin{aligned}
&\epsilon - m(6+S) + e^{rS} q_0 z m(6+S)^{\alpha_2} = q_S z \\
&e^{rS} q_0 h_S^{-\alpha_1} = q_S h_S^{-\alpha_W}.
\end{aligned}$$

Hence (53) follows from plugging q_S and h_S from (44)-(45) in the smooth pasting condition. Likewise, (51) follow from plugging $n(6+S) = 1$, $m(6+S)$ from Lemma 2, and q_0 from (53) in the value matching condition. Again, (52) requires consistency between h_S and $h(6+S)$. $\square$

For each case where we assume $J = 0$ or $J > 0$, it must also be the case that condition (47) does not or does hold.

C Numerical Algorithm

For the purposes of our estimated model in which S is fixed, the solution method in Appendix B is straightforward. We need not worry about value-matching conditions and only need to solve the smooth-pasting conditions given S, which are equations (49) and (52), to obtain h_S. Note that there is always a solution to (49) or (52)—i.e., we can always define a function $h_S(S)$ as a function of S. This is seen by you rearranging the equations as (bold-face for emphasis)

$$1 = \left(\frac{h_0}{\mathbf{h_S}}\right)^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left(1 - e^{-\frac{\alpha_2 rS}{1-\alpha_2}}\right) \cdot \left[\frac{\alpha_2 w}{r} \cdot q(6+S)\right]^{\frac{\alpha_2}{1-\alpha_2}} z^{\frac{1}{1-\alpha_2}} \cdot \mathbf{h_S}^{-\frac{1-\alpha}{1-\alpha_2}} \tag{54}$$

$$\begin{aligned}
1 = &\left(\frac{h_0}{\mathbf{h_S}}\right)^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left(1 - e^{-\frac{\alpha_2 rS}{1-\alpha_2}}\right) \\
&\cdot \left(\alpha_2 w e^{-rJ}\left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J) z^{\alpha_W}\right]^{\frac{1}{1-\alpha_W}}\right)^{\frac{\alpha_2}{1-\alpha_2}} \cdot z^{\frac{1}{1-\alpha_2}} \cdot \mathbf{h_S}^{-\frac{1-\alpha+\alpha_2\alpha_W}{1-\alpha_2}},
\end{aligned} \tag{55}$$

respectively. Hence, for any given value of S, both RHS's begin at or above 1 at $h_S = h_0$, goes to 0 as $h_S \to \infty$, and is strictly decreasing in h_S. The solution $h_S(S)$ to both (54) and (55) are such that

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

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

The rest of the model can be solved by Lemmas 2 and 4, and we can use Lemma 4 to determine J. Depending on whether condition (47) holds, we may have two solutions:

1. If only one solution satisfies (47), it is the solution.

2. If both satisfy (47), compare the two value functions at age 6 given S and candidate solutions $J_1 = 0$ and $J_2 > 0$ from Lemma 4 using the fact that the function D_W in (39) can be written

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

and

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

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

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

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

2. Assuming $\beta = 0.06$, construct the h_P-grid which is just a transformation of the S_P-grid according to (18).

3. For each node on the h_P-grid, construct z-grids according to (19), according to Kennan (2006). This results in a total of $N_{h_P} \times N_z$ nodes and probability weights, where for each h_P node we have a discretized normal distribution.

4. For each (h_P, z) compute the pecuniary of choosing $S \in \{8, 10, 12, 14, 16, 18\}$ (solve for $V(S;6,h_0)$ according to the above) and compute the fraction of individuals choosing each schooling level using the CCP's in (21)-(22).

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

D Tables and Figures not in text

Figure 5: Log earnings by Mother's Schooling

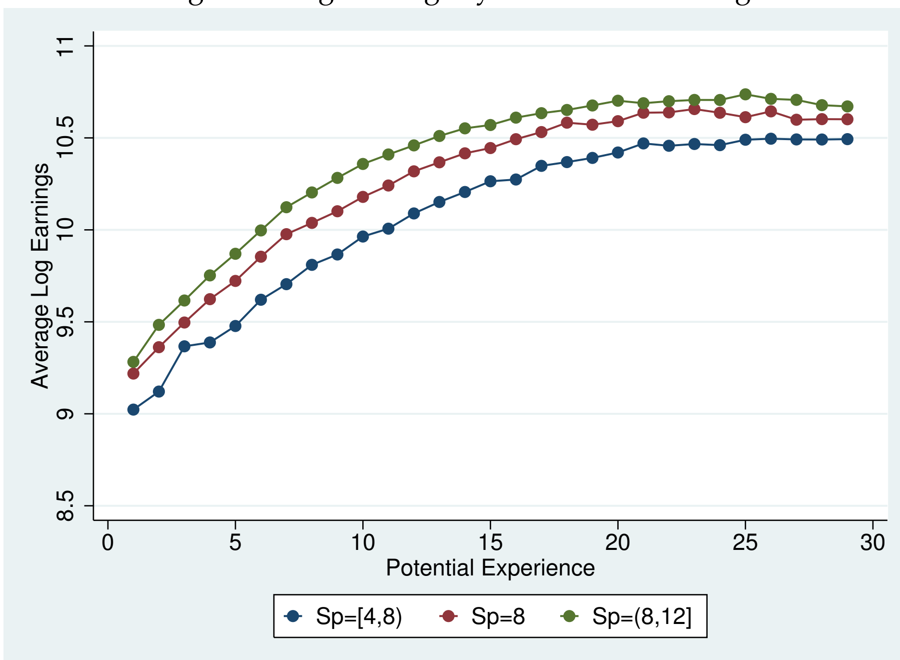


Table 10: Identification

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

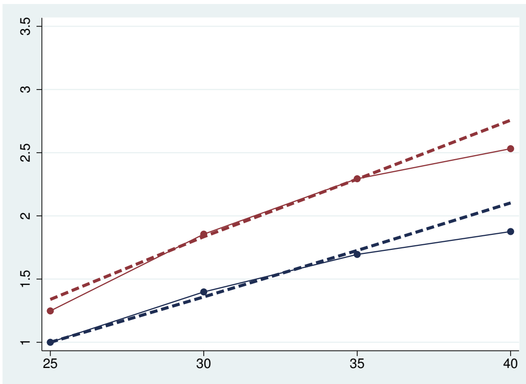
(S_P, S, E) stand for mother's schooling, and the individuals' schooling and earnings levels, respectively. Taste heterogeneity picks up the residual unobserved heterogeneity not captured by the simple model.

Table 11: Schooling Benefits, Median

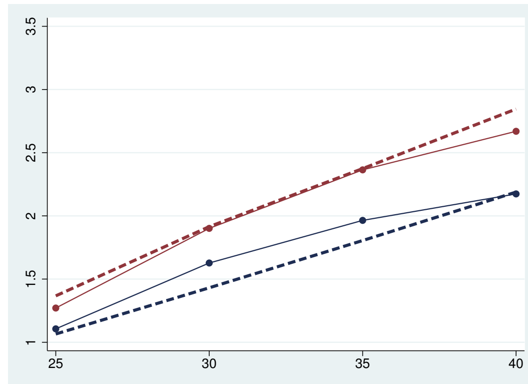
	$S=8$	$S=10$	$S=12$	$S=14$	$S=16$	$S=18$
Present value earnings by ability quartile						
Q(0)	223,874	280,365	353,766	366,731	366,986	325,786
Q(1)	209,734	252,868	273,449	247,519	245,504	245,175
Q(2)	265,120	293,683	319,773	299,677	292,973	309,249
Q(3)	282,891	322,135	357,059	342,983	345,109	349,229
Q(4)	340,101	359,287	409,854	456,748	472,201	443,605
Nonpecuniary benefits correlated with h_P by ability quartile						
Q(0)	-	5,344	9,381	10,424	15,368	28,627
Q(1)	-	5,313	10,573	10,257	14,448	28,492
Q(2)	-	5,206	9,922	10,087	14,480	28,882
Q(3)	-	4,780	9,356	10,047	14,438	28,987
Q(4)	-	4,205	8,793	10,229	15,371	28,772
Nonpecuniary benefits correlated with z by ability quartile						
Q(0)	-	324	600	712	1,023	1,510
Q(1)	-	304	523	588	837	1,302
Q(2)	-	334	575	641	920	1,486
Q(3)	-	354	606	682	984	1,566
Q(4)	-	376	648	794	1,143	1,771

PDV pecuniary value at age 6, in 2000 USD.

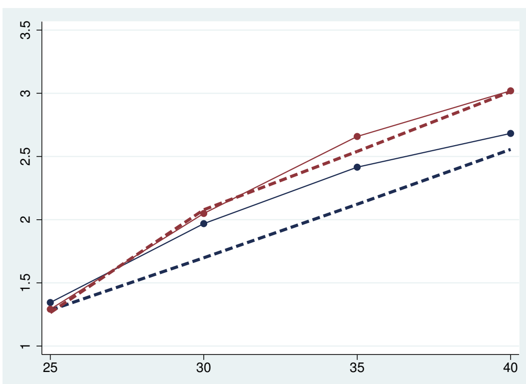
Figure 6: Model Fit



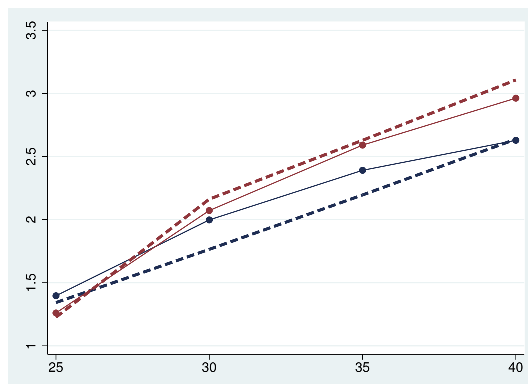
(a) $S_P < 5$



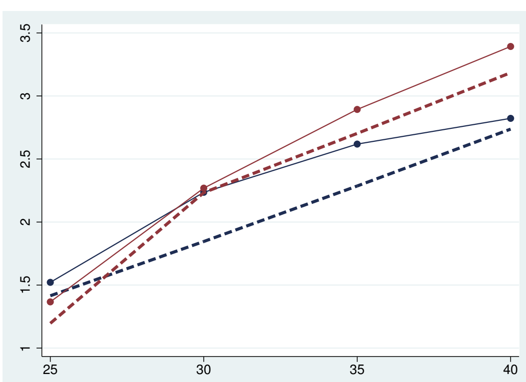
(b) $S_P \in \{6, 7\}$



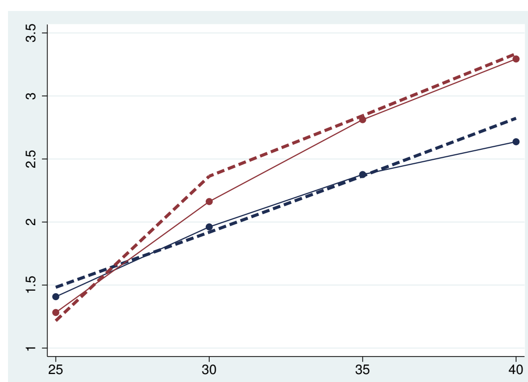
(c) $S_P = 8$



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



(e) $S_P = 12$



(f) $S_P > 12$

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

Figure 7: Median Lifecycle Effect of 1 Year Increase in Mom's Schooling.

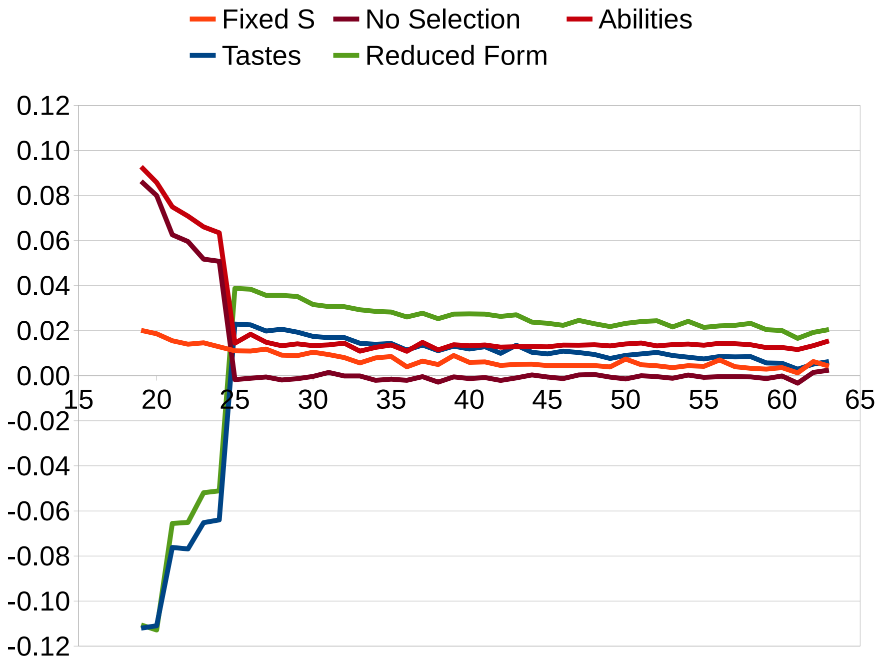
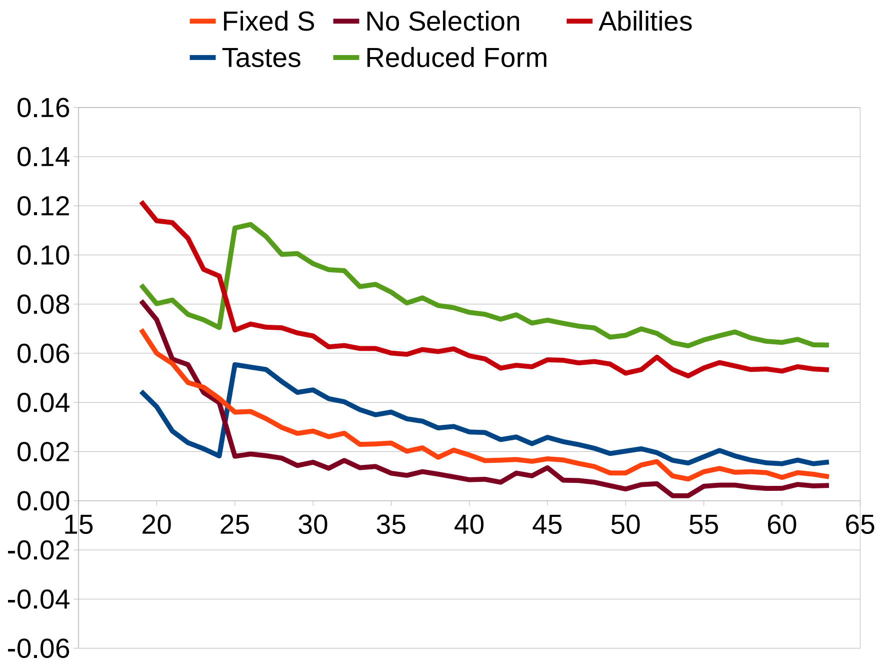


Figure 8: Median Lifecycle Effect of 9 Year Compulsory Schooling Reform.



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.

Becker, G. S. and N. Tomes (1986, July). Human capital and the rise and fall of families. *Journal of Labor Economics 4*(3), S1–39.

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.

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.

Cameron, S. V. and C. Taber (2004). Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy 112*(1), pp. 132–182.

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

Carneiro, P. and J. J. Heckman (2002). The evidence on credit constraints in post-secondary schooling*. *The Economic Journal 112*(482), 705–734.

Cunha, F. and J. J. Heckman (2007, May). The technology of skill formation. *American Economic Review 97*(2), 31–47.

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. Flinn, and M. Wiswall (2013, April). Transfers to households with children and child development. Working Papers 2013-001, Human Capital and Economic Opportunity Working Group.

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

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

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

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–86.

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., L. J. Lochner, and P. E. Todd (2008). Earnings functions and rates of return. *Journal of Human Capital 2*(1), pp. 1–31.

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

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

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

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

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.

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

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

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

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

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

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

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

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

Taber, C. R. (2001, July). The rising college premium in the eighties: Return to college or return to unobserved ability? *Review of Economic Studies 68*(3), 665–91.

Willis, R. J. and S. Rosen (1979, October). Education and self-selection. *Journal of Political Economy 87*(5), S7–36.