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

August 2015

Abstract

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

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

1 Introduction

Parents have a large influence on their children's outcomes. Does this merely reflect selection—correlation in unobserved heterogeneity across generations? Or does it also partly reflect human capital spillovers from parent to child? In the former case, government subsidies aimed at improving education would only impact one generation at best. But 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 affect intergenerational mobility.

Furthermore, intergenerational human capital spillovers (human capital of a parent being a direct input in producing human capital for a child) form an integral part of models of human capital accumulation such as Becker and Tomes (1986) and Benabou (1996). 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. To what extent do exogenous interventions or investments matter for the next generation's economic status? How do forced increases in the schooling of parents affect the schooling and earnings of children?

The model we build to address these questions is motivated by three empirical observations. First, parents' and children's schooling are highly correlated. Does this reflect correlation in innate abilities or a direct effect of being attached to a more educated mom? Given the difficulty in making causal statements from an OLS regression, researchers moved onto IV estimates. Second, many empirical studies find that the IV estimate of the effect of parents' schooling on children's schooling is zero (Behrman and Rosenzweig, 2002; Black et al., 2005). Does this imply that parental human capital spillovers are unimportant? Third, our own empirical analysis reveals that conditional on children's schooling, parents' schooling has a significant effect on children's earnings. Does this piece of evidence help identify the size of parental human capital spillovers?

We begin with a simple model of life-cycle human capital accumulation along the lines of Ben-Porath (1967), which encompasses both schooling and learning on-the-job. An individual's length of schooling and earnings profile is determined by his initial level of human capital as well as his learning ability—the speed with which he accumulates human capital. 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 a function of his parent's human capital, which is the source of parental spillovers in our model.[1] Second, we assume that the initial level of human capital is also affected by the child's learning ability, i.e., the speed at which the child accumulates human capital after age 6. Note that learning abilities are unobserved but correlated across generations and persist through life, unlike parents' human capital which only affects a 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 sep-

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

arating parental spillovers from selection on abilities can be solved by jointly using information on children's schooling and earnings. The key for identification is that parents' human capital affect children's schooling and earnings outcomes differently. In the model, all else equal, children of high human capital parents spend less time in school because there is less need to accumulate additional human capital before entering the labor market,[2] a feature of diminishing returns to human capital accumulation). While these children ultimately attain higher levels of human capital and consequently higher earnings, it takes them less time to reach that level, resulting in shorter schooling periods. Children of parents with more schooling in fact do spend more time in school. The model and data are reconciled if children of parents with more schooling are also better learners, since they would be induced to stay longer in school, overcoming the negative level effect. Such a mechanism could be understood as high human capital parents tending to have high learning abilities, which are passed on to their children.

The differentiation of years of schooling and quality of education is important: 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 years of schooling but different levels of earnings. Further suppose that their parents had 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 learning abilities is identified by observing how children's schooling levels vary across parents with different schooling levels.

This identification scheme relies on parental spillovers being constant over a child's life-cycle, conditional on the child's own schooling level. 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 (controlling for the child's own schooling). 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 (HRS) 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 about 20% of parents' education spills over to children's earnings.

We enrich the basic model with tastes for schooling and estimate it using HRS data on individual schooling and life-cycle earnings, as well as parents' schooling. In particular, in addition to parents' schooling levels, which are observed, and learning abilities, which are unobserved, we

[2]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).

account for a third source of heterogeneity—an unobserved taste for schooling. By doing so, not only are we able to replicate key moments in the data, but also ensure that we do not overestimate the effect of parents' education on children's schooling and earnings outcomes.

Our estimates suggest that the main determinant of the positive OLS correlation between mom and child's schooling is not learning ability or parental spillovers but the unobserved correlation between moms' education and children's tastes for schooling. That is, schooling outcome differences are mostly explained by non-pecuniary benefits being heterogeneous across children of parents with different levels of education. Hence *ceteris paribus*, pecuniary motivation has little causal effect on children's schooling outcomes.

At the same time, the estimates are also consistent with previous literature that finds that when using compulsory schooling as an instrument for parents' schooling, the estimated IV coefficient on children's schooling is zero. The intuition is straightforward. On the one hand, all else equal, increasing a parent's schooling and therefore the child's early human capital decreases the child's schooling, since children cut back on their length of schooling when human capital accumulation displays decreasing returns. On the other hand, parents who have been forced to stay in school longer may induce their children to develop a higher taste for schooling, leading to children increasing their schooling. When these two effects countervail each other, the IV estimate from model simulated data is close to zero.

Finally, a compulsory schooling law that results in mom's schooling rising by one year is found to have a 1.2% causal boost on the lifetime earnings of the child. This effect is, of course, heterogeneous across individuals and also 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, the earnings increase is spread more evenly over the child's 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 taking into consideration the potentially large impact of parents during early childhood (Cunha and Heckman, 2007).

Since Solon (1999), a broad literature has studied the causal effect of parents on children's earnings and/or education.[3] 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 relationship is close to zero or even negative, especially for mothers.[4] There is also evidence that

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

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

points to strong inherited genetic effects on children's schooling outcomes.[5]

Studies that look at children's outcomes other than schooling, most commonly earnings, tend to find a larger role for the environment. 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, they seem to have a large effect on other social behavior, such as college selectivity and drinking.

The fact that genetic effects are larger in some cases while environmental effects are larger in others is sometimes viewed as an inconsistency, especially in terms of children'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 likely motivated by the very strong relationship between an individual's education and earnings (Card, 1999). In contrast, we argue that a parent's education can have qualitatively different effects on the child's education and earnings. 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 returns to schooling literature (Heckman et al., 2006), as of yet no attempts have been made to link non-pecuniary motives to the parental effect on children's education and earnings.

Our first departure from this literature is to posit a non-linear model in which the individual takes his parent's variable(s) as a state to solve a lifetime optimization problem. 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.[6] 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 intergenerational spillovers. Since the resulting causal effect on earnings remains positive, our model is at once consistent with empirical findings that parents may have a negative causal effect on children's schooling, and the limited evidence that 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. Several studies have shown that

[5] Behrman and Taubman (1989), using data on twin parents and extended family relationships, decompose the variance of observed years of schooling into the variance of genetic and environmental variables and 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 non-genetic effects, but conclude that inherited genetics is still the dominant effect.

[6] For example, 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.

pecuniary motives alone fall short of explaining education choices (Heckman et al., 2006) and non-pecuniary motives are estimated to be quite large in life-cycle models with schooling choice (Heckman et al., 1998). We allow both the child's pecuniary and non-pecuniary returns to schooling be be correlated with his parent's schooling, through unobserved learning abilities and tastes for schooling, respectively. Non-pecuniary returns may capture psychological or non-cognitive factors that induce a child to attain more or less education (Oreopoulos et al., 2008; Rege et al., 2011), and/or the fact that children from less advantaged families are more likely to be misinformed about education returns (Betts, 1996; Avery and Turner, 2012; Hoxby and Avery, 2013).[7]

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 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), 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 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 the main result of increasing mother's schooling by 1 year, and also the counterfactual result of a hypothetical compulsory schooling reform. Section 6 concludes.

2 Schooling and Earnings Model with Parental Spillovers

We begin with a simple variant of a Ben-Porath (1967) life-cycle model of human capital accumulation, abstracting away from non-pecuniary tastes for schooling. An individual begins life at age

[7] In our context, we simply lump them together into heterogeneous tastes for schooling.

6, retires at age $R > 6$ and dies at age $T \geq R$, all exogenous to the individual. His state at age 6 is (h_0, z), which denotes his initial stock of human capital and (learning) ability, respectively.

Let $V(a, h)$ denote the present discounted value net income at age a given a human capital level of h. 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:[8]

$$V(6, h_0) = \max_{\{n(a), m(a)\}} \left\{ \int_6^R e^{-r(a-6)} \left[wh(a) \left[1 - n(a)\right] - m(a) \right] da \right\} \tag{1a}$$

$$\dot{h}(a) = z \left[n(a)h(a)\right]^{\alpha_1} m(a)^{\alpha_2}, \tag{1b}$$

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

given $h(6) = h_0$.

The exponents α_1 and α_2 are the returns to time and good investments, respectively. The wage w and discount rate r are assumed to be constant and taken as given by the individual. At any age, w multiplies human capital $h(a)$ to generate individual wages. 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 (h_0, z) captures slope and level effects. Assuming decreasing returns to scale, i.e. $\alpha_1 + \alpha_2 < 1$, if h_0 and 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 spent 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.

2.1 Initial Conditions

Now suppose there is a mass of children facing the same problem. 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}). We assume a statistical relationship between the parents and children. Learning abilities, which remain constant through life, are transmitted exogenously through generations according to

$$\log z_i = \mu_z + \rho_z \frac{\sigma_z}{\sigma_{z_P}} (\log z_{Pi} - \mu_{z_P}) + \epsilon_{zi}, \tag{2}$$

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

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

ϵ_{zi} a mean zero, i.i.d. shock with standard deviation σ_ϵ.

But we are not interested in the intergenerational persistence of learning abilities *per se*. So we also assume a statistical relationship between h_P and z_P, in which case z becomes multiplicatively independent of z_P conditional on h_P, i.e.

$$z_{Pi} = f(h_{Pi})\epsilon_{Pi} \quad \Rightarrow \quad z_i = g(h_{Pi})\epsilon_i, \tag{3}$$

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 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.

We further assume that z_P, which cannot be affected by the environment, does not directly affect h_{0i}. Then the causal effect from parents' education on children's earnings comes entirely from h_{Pi} affecting h_{0i}. This is to be understood as parental investments prior to age 6, which depend on the parent's economic status summarized 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 one's own learning ability z_i as well. So we write h_{0i} as a function of (z_i, h_{Pi}); specifically we assume

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

Then the elasticity of h_0 with respect to h_P, evaluated at individual i's state, is:

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

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 represents how much the child's ability matters for early human capital formation independently of h_P.

Consequently, we are solving a simple life-cycle decision problem in which a child takes his ability and parent's human capital as exogenous state variables. In the estimation, we proxy h_P by a parent's education and/or earnings, so it is observed. But since abilities are unobserved, we face the problem of separately identifying (ν, λ) from ρ_{zh_P}, i.e., how much a child's schooling or earnings are correlated with the parent's because of h_P directly affecting h_0, or indirectly affecting z (through z_P). The goal of this section is to show how the solution to the model allows this, which is why we abstract from tastes for schooling for now and only include them later in section 4. This gives us insights into the precise mechanisms at work, and how some model parameters can be recovered from a panel of individuals' schooling and earnings.

2.2 Solution

In what follows we drop the individual subscript i unless necessary. We solve program (1) given (4). Program (1) 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 (HJB) equation is

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

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 the marginal return. Equations (5), (6) and (7) along with the law of motion (1b), initial condition (4) and terminal condition $V_h = 0$—the appropriate transversality condition for a fixed terminal time problem—characterize the complete solution. 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

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

All proofs are contained in Appendix A.

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

$$F'(S) > 0, \qquad F(S) \geq z^{1-\lambda(1-\alpha)} h_P^{-\nu(1-\alpha)} \tag{8}$$

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 individuals benefit more from schooling. But 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 because children of high human capital parents start off with a higher level of human capital and have less of a need to stay in school, or put differently, children of higher human capital parents can learn more during the same length of schooling (a substitution effect between the length and quality of schooling).

For the parent generation, we assume a Mincerian representation between schooling and earnings (the parent's human capital):[9]

$$h_P = \exp(\beta S_P) \quad \Leftrightarrow \quad \log h_P = \beta S_P \tag{9}$$

Then the observed dependence of children's schooling on parents' schooling depends on $(\rho_{zh_P}, \nu, \lambda)$. To see this, suppose (3) takes the form

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

where $(\mu_{h_P}, \sigma_{h_P})$ denote the population mean and standard deviation of (log) parents' human capital, and $\tilde{\rho}_{zh_P}$ the elasticity of z with respect to S_P. Plugging this into (8) at equality yields

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

The correlation in schooling across generations is then positively related to the parameter $\tilde{\rho}_{zh_P}$ as long as $\lambda(1-\alpha) < 1$, but negatively related to the spillover parameter ν. Estimating the magnitudes of of these different components is the key objective of this paper.[10]

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

[9] This assumption is further discussed in Section 4.2.

[10] Later we also assume that parents' education S_P is correlated with children's tastes for schooling.

schooling, h_S, satisfies

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

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.

Using Proposition 2, we can now derive an expression for individual earnings profiles. 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 individuals pay for job training costs in the form of lower wages (i.e., 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 $a \in [6+S, R)$ and any $(\pi_n, \pi_m) \in [0,1]^2$,*

$$e(a) = wh(a)\,[1 - \pi_n n(a)] - \pi_m m(a) = [C_1(S) + C_2(a;S)] \cdot z^{\frac{1}{1-\alpha}},$$
$$C_2(a;s) = \kappa^{\frac{1}{1-\alpha}} \cdot \left\{ r \cdot \int_{6+s}^{a} q(x)^{\frac{\alpha}{1-\alpha}} dx - (\alpha_1 \pi_n + \alpha_2 \pi_m) q(a) \right\}$$

So we can write earnings as a function of potential experience $x = a - 6 - S$ and schooling S,

$$e(x, S) = [C_1(S) + C_2(x;S)] \cdot z^{\frac{1}{1-\alpha}}.$$

Proof. See Appendix A. □

Corollary 1 implies that the fraction of job training costs deducted from earnings only depends on age, as long as the fraction deducted remains constant. More importantly, conditional on ability z, (C_1, C_2) govern the intercept and slope of individuals' (potential) experience-earnings profiles by schooling, respectively. According to the model, these slopes must be close to parallel across different z's, which is standard in models that use a Ben-Porath approach.

The earnings equation in Corollary 1 can also be interpreted as a Mincer equation that relates earnings to schooling. The functions C_1 and C_2 determines the returns to schooling and potential experience, respectively. So given a balanced sample of individual earnings profiles, their intercept and slopes identify the exponents for the human capital production function, (α_1, α_2). Of course for identification, we first need to control for the unobserved heterogeneity in z.

2.3 Identification of Key Parameters

A robust finding in empirical studies is that even after controlling for observables, mothers' education has a statistically significant relationship with children's schooling and earnings. The parameters $(\tilde{\rho}_{zh_P}, \lambda)$ are a structural representation of the selection effect, and ν of the causal effect. We now demonstrate that all three parameters are separately identified given a data sample of children's schooling and earnings outcomes, and the schooling 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 schooling levels of their parents and themselves, (S_{Pi}, S_i), and experience-earnings profiles $e_{i,x}$ (but not z_i).*

1. *If we select only those individuals whose parents have the same $S_{Pi} = \hat{S}_P$, then for all $x > 0$, earnings depend only on S through λ and nothing else:*

$$e_{i,x}(S_{Pi} = \hat{S}_P) \propto [C_1(S_i) + C_2(x, S_i)] \cdot F(S_i)^{\frac{1}{(1-\alpha)[[1-\lambda(1-\alpha)]}}. \tag{12}$$

If we regress

$$\log e_{i,x}(S_{Pi} = \hat{S}_P) = a_0 + a_1 \log [C_1(S_i) + C_2(x; S_i)] + a_2 \log F(S_i) + \epsilon_i, \tag{13}$$

we recover

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

2. *Suppose that (z, h_P, S_P) follow (9)-(10). If we regress*

$$\log e_{i,x} = b_0 + b_1 \log [C_1(S_i) + C_2(x; S_i)] + b_2 S_{Pi} + \eta_i, \tag{14}$$

we recover

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

So if we know the functions (C_1, C_2, F)—i.e. (α_1, α_2)—are known, λ is recovered from (13). If β is also known, $\tilde{\rho}_{zh_P}$ is identified from (14).

Proof. Suppose we observe two individuals with different levels of schooling and earnings but whose parents have the same schooling, denoted by (S_1, S_2), $(e_{1,x}, e_{2,x})$, and (S_{P1}, S_{P2}), respectively. Let (z_1, z_2) denote their *unobserved* learning abilities. Then by Corollary 1,

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

but since $S_{P_1} = S_{P_2}$, by Proposition 1

$$\frac{F(S_1)}{F(S_2)} = \left(\frac{z_1}{z_2}\right)^{1-\lambda(1-\alpha)} \quad \Rightarrow \quad \frac{e_{1,x}}{e_{2,x}} = \frac{[C_1(S_1) + C_2(x;S_1)]}{[C_1(S_2) + C_2(x;S_2)]} \cdot \left[\frac{F(S_1)}{F(S_2)}\right]^{\frac{1}{(1-\alpha)[1-\lambda(1-\alpha)]}}$$

which is (12). Part 2 follows trivially from Corollary 1 after plugging in the assumed relationship between z and h_P from (10) and h_P and S_P from (9). □

Put simply, the magnitude of λ is identified by the Mincer coefficient of own schooling on earnings, *conditional on parents' schooling*. Clearly for children with identical S_P, the spillover has no role in explaining earnings differences. Also, earnings are not directly 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 is common to all individuals, the only way that schooling can have heterogeneous effects on earnings is through z's influence on $F(S)$, which reveals λ.

Even if λ is known, the schooling regression in (11) does not identify ν separately from $\tilde{\rho}_{zh_P}$. But a Mincer regression that controls for an individual's schooling and potential experience (which corresponds to (C_1, C_2)) reveals that the coefficient on parents' schooling S_P captures only $\tilde{\rho}_{zh_P}$ and completely misses ν (part 2 of the corollary). This also implies that once we control for own schooling and parents' schooling, average experience-earnings profiles should be close to parallel, since C_2 does not vary across individuals. Furthermore, the gaps between the profiles are determined by C_1, so the parameters (α_1, α_2) are identified.[11]

That a simple regression of log earnings on a function of schooling and potential experience, and parents' schooling S_P completely reveals ability selection (given knowledge of C_1, C_2, F) may be somewhat surprising. We will see in Corollary 3, however, that if run separate regressions on groups of children with the exact same years of schooling, the coefficient on parents' schooling would identify ν instead. According to our model then, the difficulty in separating the two effects stems from how we control for own schooling S.

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 $x > 0$, earnings depend only on S_P through (ν, λ) and nothing else:*

$$e_{i,x}(S_i = \hat{S}) \propto h_{Pi}^{\frac{\nu}{1-\lambda(1-\alpha)}}. \tag{15}$$

So if we regress

$$\log e_{i,x}(S_i = \hat{S}) = b_0 + b_2 S_{Pi} + \epsilon_i, \tag{16}$$

[11] Because $C_1(s)$ is exponential in s, $\log C_1(s)$ is close to linear, meaning these gaps should be even across different schooling levels. We show in the next section that this is indeed true in the data.

we recover

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

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 individuals with the same level of schooling but different levels of earnings and parents' schooling, denoted by (S_1, S_2), $(e_{1,x}, e_{2,x})$, and (S_{P_1}, S_{P_2}), respectively. Let (z_1, z_2) denote their unobserved learning abilities. Then by Corollary 1, since $S_1 = S_2$,

$$\frac{e_{1,x}}{e_{2,x}} = \left(\frac{z_1}{z_2}\right)^{\frac{1}{1-\alpha}} \quad \Rightarrow \quad \left(\frac{z_1}{z_2}\right)^{1-\lambda(1-\alpha)} = \left(\frac{h_{P_1}}{h_{P_2}}\right)^{\nu(1-\alpha)} \quad \Rightarrow \quad \frac{e_{1,x}}{e_{2,x}} = \left(\frac{S_{P_1}}{S_{P_2}}\right)^{\frac{\nu}{1-\lambda(1-\alpha)}}$$

where the second equation follows from Proposition 1 and (9). □

Equation (14) in Corollary 2 implied that among children whose parents attained the same schooling, experience-earnings profiles are parallel with the gaps explained by own schooling. Corollary 3 now shows that *conditional on children's years of schooling*, it is the difference in their parents' schooling that manifests itself as constant gaps across parallel profiles. Then if we know λ, these gaps identify the magnitude of ν.

Why does the coefficient on parents' schooling in (14) reveal selection while in (16) it reveals spillovers? In the former, we are measuring the effect of parents on children's residual earnings, accounting for the effect that schooling has on earnings through the functions (C_1, C_2). But note that we are still looking across the entire population, and since parents have no direct role on earnings once schooling is controlled for, the regression only reveals selection (correlation between parental human capital and abilities). The effect of schooling in this case is understood as the average effect of schooling on earnings across all children, regardless of parental background.

In the latter, we are selecting a subgroup of children with the same level of schooling. So obviously there is no role for schooling to explain earnings differences, but more importantly, the distribution of abilities differ from the population distribution. Hence regressing earnings on parents' schooling does not reveal the population (z, h_P) correlation, but only the correlation within each schooling subgroup. However, by means of (8), we know exactly what this correlation is in terms of (λ, ν), i.e., we know how ability varies as a function of parents' human capital. Since parents' human capital is (assumed to be) observed, once λ is known, ν is revealed.

The next natural question is whether the posited structure of the model is empirically reasonable. In the next section, we run the proposed regressions in Corollaries 2-3 using HRS data, in which earnings profiles are indeed parallel with gaps determined by own and parents' schooling, and present some raw evidence on the relative magnitudes of $\tilde{\rho}_{zh_P}$ and ν. We then estimate a richer, extended model in section 4, which includes unobserved heterogeneity in tastes for schooling. This is so that we do not overestimate the role of abilities, the only source of unobserved heterogeneity in the simple, stylized model presented thus far. We use corollaries 2 and 3 to discipline our choice of moments, and show through simulations that the basic intuition for identification from carries over.

The main takeaway from these corollaries does not depend on whether our model is a true representation of the data; our identification scheme does not depend on the particular model *per se*. The essential assumption is that schooling is a function of both own abilities and parents' schooling, and that earnings is a function of own abilities and schooling. That is, parents only influence children's earnings through children's schooling. Otherwise, even if we had an ideal instrument for S_P, we would not be able to identify the causal effect on earnings since an exogenous increase in S_P would affect earnings not only directly but also indirectly through S.Under these assumptions, the parental spillover is understood as a child of a more educated parent getting more out of the same level of schooling (i.e., the quality of schooling). Then given a sample of children with the same level of schooling, we can always recover the underlying relationship between parents schooling and children's abilities, regardless of the underlying model. The intuition for this is in Figure 5 in the Appendix.

3 Data Analysis

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

The HRS is usually used to study elderly Americans close to or in retirement, but there are several features that make it suitable for us. First, these older individuals and their parents were less affected by compulsory schooling regulations and other government interventions, and more than half never advanced to college. This makes the sample suitable for a model such as ours in which a large schooling variation is important for identifying causal effects.[12] Second, the education premium was quite stable prior to the 1980s, so it is unlikely for these cohorts to have been surprised by an unexpected rise in education returns. It is also less likely that the effect of education on earnings outcomes or the reverse effect of expected earnings on education choices changes much from birth year to birth year. Third, the HRS contains information on own schooling, schooling of both parents, and can be augmented with restricted Social Security earnings data through which we observe entire life-cycle earnings histories.[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

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

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

Table 1: Summary Statistics by Education

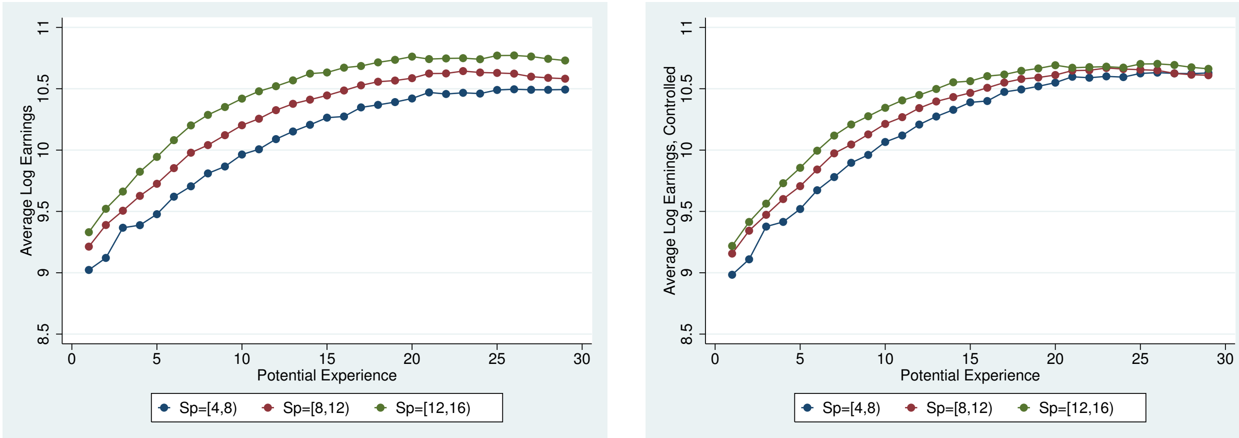
3.1 Descriptive Statistics

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 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

with richer information on parents and family background, we only observe the beginning of children's age-earnings profiles, which is also very noisy because the average age for labor market entry is increasing. Few datasets do have long earnings histories as well as richer parental background information, such as NLSY79, but these individuals face compulsory schooling regulations and a rising college premium.

[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.



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

(b) Children of mother's with [4,8), [8,12), and [12,16) years of schooling, controlling for own schooling.

Figure 1: Identifying Selection

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

records assuming the following individual log-earnings process[15]

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

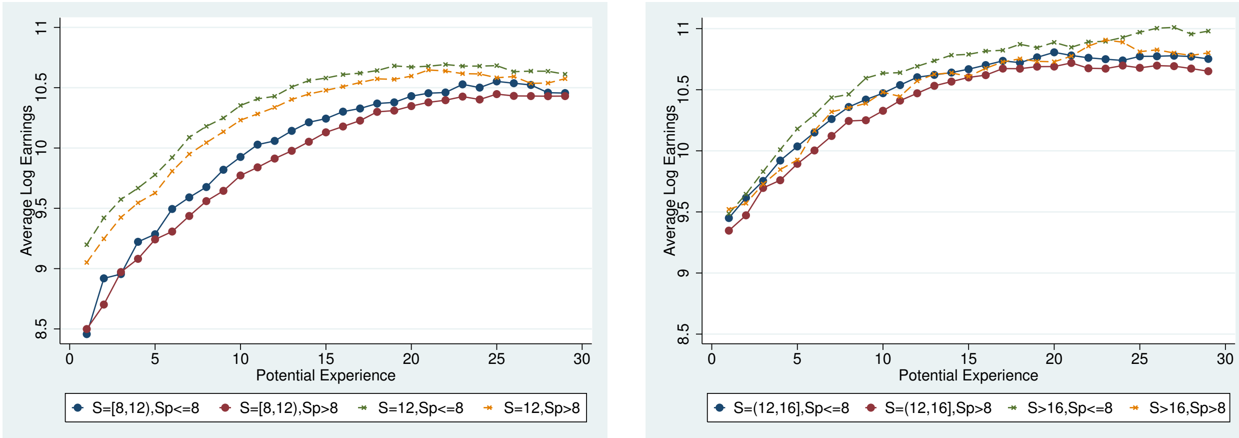
where $e^*_{i,t}$ is the latent earnings of individual i at time t in 2008 dollars, $X_{i,t}$ is the vector of characteristics at time t, and the error term $\varepsilon_{i,t}$ includes an individual specific component α_i, which is constant over time, and an unanticipated white noise component $u_{i,t}$. 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 Spillovers and Ability Selection

According to our model, the spillover is subsumed in the choice of schooling (Proposition 1). Selection on abilities is revealed as the Mincer coefficient on parents in a regression over the population (Corollary 2) and spillovers by the same coefficient but in a regression over subgroups with identical levels of schooling (Corollary 3). Moreover, the model predicts that for both regressions, earnings profiles should be parallel with a constant gap.

In Figure 1(a), we divide individuals by 3 subsamples, depending on whether their mothers

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



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

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

Figure 2: Identifying Spillovers

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

attain $S_P \in [4,8)$, $[8,12)$, or $[12,16)$ years of schooling. These categories correspond approximately to primary, secondary, and tertiary education. Notice that the average earnings profiles are parallel by mother's schooling, as predicted by the model (the function C_2).[16] In Figure 1(b), we take out a mean effect of own schooling on own earnings for each level of own schooling S. In terms of the model, it is akin to controlling for the effect that comes through $C_1(S)$ for all levels of S. Since schooling is correlated across generations, the gaps narrow. The remaining gap is the selection effect: the fact that higher human capital parents have higher human capital children. Also notice that these gaps are of similar magnitude across different S_P, justifying our parametric assumption of log-normally distributed abilities.

In order to visualize spillovers, we need to divide individuals into different levels of own schooling, rather than control for an average schooling effect that applies equally to everyone. So we split individuals into 4 subsamples depending on whether they have $S \in [8,12)$, exactly 12, $(12,16]$, or more than 16 years of schooling. Each group is further divided according to whether the mother has more than 8 years of schooling, which is about half of the mothers in our sample.[17] Figure 2 depicts the average log-earnings profiles for each subsample. The left panel compares individuals with $[8,12)$ to those with exactly 12 years of schooling, and the right panel individuals with $(12,16]$ to those with 17 or more years of schooling, respectively.

Again, for all 4 education levels, the average log-earnings profiles of children with the same schooling but different mother's schooling are nearly parallel with a constant gap. This points to a permanent level effect that persists throughout an individual's career. It is precisely this gap that

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

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

is captured by the spillover parameter ν in our model.[18] Moreover, the gaps are similar across all three categories of the children's educational attainment. We now present reduced form evidence on the magnitudes of the selection and spillover effects presented in Figures 1-2.

3.3 Mincer Regressions

The figures suggest the following simple Mincer regression:

$$\log e_{i,x} = \beta_0 + \beta_1 S_i + \beta_2 S_{P_i} + f(x) + \epsilon_{i,x} \tag{17}$$

where $e_{i,x}$ is the earnings of individual i with potential experience x=age-6-S, $f(\cdot)$ a function of x which we specify in various different ways below, and $\epsilon_{i,x}$ an error term. We estimate different versions of (17) for earnings data from ages 23 to 42, and tabulate the results in Table 2.[19,20]

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 does not speak to which of these measures is more appropriate. However, many studies suggest that mother's have a larger influence on children (Behrman and Rosenzweig, 2002; Del Boca et al., 2012), which is true in our sample as shown in Table 2.

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

The coefficient on own schooling remains similar when we measure parents' human capital with the father's year of schooling in (3), but the parent's coefficient is attenuated. The rate of return on paternal education is 1.1% and statistically significant. The parent's coefficient drops further when we measure parents' human capital as the sum of the schooling of both parents in (4). If we include both separately as in (5), 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 that mom's schooling has dominant explanatory power. So in what follows, we take mother's schooling as the proxy for parent's human capital.

[18] To be precise, the gap should capture $\hat{b}_2 = \beta \cdot \nu / [1 - \lambda(1-\alpha)]$ in (16).

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

[20] Including race or cohort dummies do not affect our estimates much. Moreover in our estimated model, such effects should be absorbed in the unobserved heterogeneity in tastes for schooling, so controlling for them here would make the reduced form estimates incomparable with our GMM estimates.

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

Table 2: Mincer Regressions

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

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

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

We have rerun these regressions controlling for race and cohort dummies, and also for white males separately. The coefficient on mom's schooling is quite stable across all specifications. The striking feature is that no matter how we control for experience, the coefficients on S and mom's S_P are similar in magnitude and highly significant. This means that the function C_2 in the model can be controlled for using a simple linear and quadratic term for potential experience, while C_1 can be controlled for linearly. In fact, since C_1 is close to exponential, its logarithm is close to linear, and a comparison of columns (6) and (9) suggest that the log earnings-schooling relationship is also linear in the data. Indeed, in Figure 6 in the Appendix, we plot the values of each schooling dummy against the linear return of 7.6%, and find that the dummies increase almost linearly. This gives us further confidence that the model can replicate important features of the data.

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

Table 3: Mincer Regressions by Own Schooling

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

To summarize, mothers' schooling has a stronger relationship with sons' earnings than fathers', and the estimated effect of mother's schooling, or β_2, is about 1.7%, which is a measure of the gaps between the three lines in Figure 1. This is in the range of previous empirical work on this topic (Card, 1999), and according to our model is not a causal effect but is interpreted as selection on abilities. To get the (causal) parental spillover, we need to slice the sample into subgroups of children with similar levels of schooling, whose distribution of abilities will differ from the population.

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

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

i.e., equation (17) without controlling for S_i and using a linear and quadratic term for potential experience x. Again, the Mincer coefficient on mother's schooling is significant for all subgroups at approximately 2%. This coefficient measures the average gaps between the pairs of earnings profiles for children with the same level of schooling but different levels of mother's schooling in Figure 2. Moreover, the magnitudes of the coefficients are similar across all groups, although slightly lower for children with very high education (more than college).[21] This justifies our assumption of a constant ν that applies equally for all education groups (Corollary 3).

4 Estimation

Since schooling and earnings are only functions of abilities (z) and parents' schooling (S_P), the previous regressions may overestimate the role of both selection and spillovers in the presence of other dimensions of unobserved heterogeneity. Furthermore, the predictions are too strong to explain other important features, such as why the spillover seems to dampen at higher levels of education, or relatedly, why some children attain very high levels of schooling without any

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

apparent gain in earnings. In this section we extend the simple model to bring it closer to the data by including heterogeneous tastes for schooling, among other smaller modifications. The goal of the extended model we estimate is to retain the simplicity of the framework presented in section 2 and empirical intuition from section 3, while being able to explain the selection and spillover effects at finer levels of own and mother's schooling.

4.1 An Extended Model of Parental Spillovers

We make the following modifications to the model of section 2. First, we constrain the choice of schooling to be discrete, consistently with the data. Second, we introduce tastes, or non-pecuniary benefits, from schooling, using a nested logit. Lastly, we allow the human capital production technology during school to differ from on-the-job training (OJT). The problem faced by an individual at age 6 is then

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

where $\tilde{\xi} \equiv [\tilde{\xi}_S]$ is a vector that represents a non-pecuniary benefit for each level of schooling expressed in present value terms. This only affects an individual's desire to remain (or not) in school while having no direct effect on earnings. The age 6 pecuniary value from attaining S years of schooling is

$$\tilde{V}(S; 6, h_0) = \max_{\{n(a), m(a)\}} \left\{ -\int_6^{6+S} e^{-r(a-6)} m(a) da + \int_{6+S}^{R} e^{-r(a-6)} h(a) \left[1 - n(a)\right] da \right\} \tag{18}$$

subject to

$$\dot{h}(a) = \begin{cases} z h(a)^{\alpha_1} m(a)^{\alpha_2}, & \text{for } a \in [6, S), \\ z[n(a)h(a)]^{\alpha_W}, & \text{for } a \in [S, R), \end{cases}$$
$$h(6) = h_0 = b z^{\lambda} h_P^{\nu}. \tag{19}$$

Schooling only involves goods inputs and OJT only involves time inputs. The technology in the schooling phase is identical to the simple model while the working phase is identical to the the simple model with $\alpha_2 = 0$, but note that we let the return to human capital investments differ during the school and working phases,. The parameter b that multiplies initial human capital in (19) captures the level of human capital common to the population, so we have dropped the wage rate w as it is not separately identified from b in our partial equilibrium setup.

At age 6, an individual is characterized by $\{h_P, z, \tilde{\xi}\}$. Although we cannot derive closed form solutions for schooling and earnings as in the simple model, in Appendix B we characterize the solution to (18) and and describe how a solution is found numerically in Appendix C.

4.2 Population Distribution Assumptions

Our dataset contains information on individuals' schooling and their complete earnings profiles, and the schooling of parents. Since we only have information on the schooling of the parent, we approximate the parent's human capital, or earnings, of the parent by the standard Mincerian equation $h_P = \exp(\beta S_P)$ in (9). 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 (19). 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.[22]

We assume that $(\log h_P, \log z)$ are joint normal, specifically

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

Note that h_P is known given 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

$$\log z | \log h_P \sim \mathcal{N}\left(\mu_z + \rho_{zh_P}\frac{\sigma_z}{\sigma_{h_P}}\left(\log h_P - \mu_{h_P}\right), \sigma^2_z\left(1-\rho^2_{zh_P}\right)\right). \tag{20}$$

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 choice of S and resulting life-cycle earnings for any given initial condition (h_P, z).

Tastes for schooling $\tilde{\xi}$ vary with schooling S, parent's human capital h_P and ability z:

$$\tilde{\xi}_S \equiv \delta_S\left(\gamma_{h_P} h_P + \gamma_z z\right) + \xi_S. \tag{21}$$

where $\xi \equiv [\xi_S]$ is 6-dimensional logit. The constants γ_{h_p} and γ_z capture the correlations between ξ and h_P and z, respectively. We normalize $\delta_8 = 0$ and $\delta_{10} = 1$ and estimate δ_S, $S \in \{12, 14, 16, 18\}$.

Schooling decisions 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 modeled as nested logit. 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

$$\tilde{u}_S \equiv \tilde{V}(S; 6, h_0) + \delta_s\left(\gamma_{h_p} h_P + \gamma_z z\right), \quad S \in \{8, 10, 12, 14, 16, 18\}.$$

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

Nesting yields the following conditional choice probabilities (CCP), given (h_P, z):

$$\Pr(S=8) = \Pr(S=8|S \in \{8,10,12\}) \cdot \Pr(S \in \{8,10,12\}) \tag{22a}$$

$$\Pr(S=8|s \in \{8,10,12\}) = \frac{\exp\left(\tilde{u}_8/\sigma_\xi \zeta_h\right)}{\exp\left(\tilde{u}_8/\sigma_\xi \zeta_h\right) + \exp\left(\tilde{u}_{10}/\sigma_\xi \zeta_h\right) + \exp\left(\tilde{u}_{12}/\sigma_\xi \zeta_h\right)} \tag{22b}$$

$$\Pr(S=14|S \in \{14,16,18\}) = \frac{\exp\left(\tilde{u}_{14}/\sigma_\xi \zeta_c\right)}{\exp\left(\tilde{u}_{14}/\sigma_\xi \zeta_c\right) + \exp\left(\tilde{u}_{16}/\sigma_\xi \zeta_c\right) + \exp\left(\tilde{u}_{18}/\sigma_\xi \zeta_c\right)} \tag{22c}$$

and

$$\Pr\left(s \in \{8,10,12\}\right) = \tag{23}$$

$$\frac{\left[\exp\left\{\frac{\tilde{u}_8}{\sigma_\xi \zeta_h}\right\} + \exp\left\{\frac{\tilde{u}_{10}}{\sigma_\xi \zeta_h}\right\} + \exp\left\{\frac{\tilde{u}_{12}}{\sigma_\xi \zeta_h}\right\}\right]^{\zeta_h}}{\left[\exp\left\{\frac{\tilde{u}_8}{\sigma_\xi \zeta_h}\right\} + \exp\left\{\frac{\tilde{u}_{10}}{\sigma_\xi \zeta_h}\right\} + \exp\left\{\frac{\tilde{u}_{12}}{\sigma_\xi \zeta_h}\right\}\right]^{\zeta_h} + \left[\exp\left\{\frac{\tilde{u}_{14}}{\sigma_\xi \zeta_c}\right\} + \exp\left\{\frac{\tilde{u}_{16}}{\sigma_\xi \zeta_c}\right\} + \exp\left\{\frac{\tilde{u}_{18}}{\sigma_\xi \zeta_c}\right\}\right]^{\zeta_c}}.$$

4.3 Generalized Method of Moments

There are 25 parameters in the model, which we partition into

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

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 extended model, respectively. These are estimated by GMM to fit schooling and earnings moments from the extended model, which can be computed exactly subject only to numerical approximation error (see Appendix C for details).[23]

Parameters Set a Priori The interest rate is fixed at 5%, which is in the range of the after-tax rate of return on capital reported in Poterba (1998) and used in Heckman et al. (1998).[24] We follow their procedure and fix the interest rate at 5%, As noted earlier, the wage rate is normalized to 1 since it is not separately identified from b. The retirement age is fixed at 65.

The coefficient β is recovered from a standard Mincer regression applied to the HRS AHEAD cohorts (without including mother's schooling). The purpose is to induce a statistical distribution of parents' earnings from their schooling, including all endogenous effects.[25] The resulting coefficient is quite stable across cohorts, ranging from approximately 0.04 to 0.06 for men and 0.05 to 0.09 for women; we fix $\beta = 0.06$. This is not very different from the coefficients we recover from

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

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

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

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

* See text for more details.

Table 4: Parameters Set a Priori

the (later-born) HRS cohorts in our sample in Table 2 which includes more controls; our estimates are not sensitive to different values of a β within this range.

Since $\log h_P = \beta S_P$, we have $\mu_{h_P} = \beta \mu_{S_P}$ and $\sigma_{h_P} = \beta \sigma_{S_P}$. We take the mean and variance of mother's schooling, μ_{S_P} and σ_{S_P}, directly from their sample analogs in the data. Hence the only parametric assumption we are imposing by assuming that S_P is Gaussian is its correlation structure with z. Since $\mu_{S_P} = 9.24$ and $\sigma_{S_P} = 3.60$, we 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 not separately identified from the other taste parameters (namely, γ_{h_P}, γ_z and σ_ξ) and normalized to 0 and 1, respectively. The entire list of exogenously fixed parameters are summarized in Table 4.

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

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

Denote these target moments $\hat{\Psi}$. For an arbitrary value of θ_1, we compute the implied model moments $\Psi(\theta_1)$ as described in Appendix C. The parameter estimate $\hat{\theta}_1$ is found by

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

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

Table 5: Targeted Empirical Moments

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

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

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

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

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

Corollaries 2-3 showed that in the simple model, the parameters of interest $(\nu, \lambda, \rho_{zh_P})$ are identified from earnings by mother's schooling, earnings by own schooling, 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 in Section 4.5.

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

The level parameter b controls the overall amount of age 6 human capital, since with higher values schooling becomes less important for all individuals uniformly. Given an average level of schooling, the 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 σ_ξ, the parameters (γ_{h_P}, γ_z) are identified by how educational attainment varies across mother's schooling and individual earnings.

4.4 Interpreting the Parameters

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

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

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

*Standard errors in parentheses.

Table 6: Parameter Estimates

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

The astute reader might wonder whether we recover a large estimate because we assume that parents only have a level effect, i.e., that h_P only has a causal effect on h_6 but not z. Unfortunately, such a slope effect is not separately identified from ρ_{zh_P} in the simple model of Section 2. We have run several numerical simulations with the extended model to verify that adding a slope spillover has small influence on all other parameter estimates except ρ_{zh_P}. This implies that a slope spillover would only crowd out selection effects, so our results can be viewed as a conservative lower bound estimate for parental spillovers.

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

Unlike the spillover, this is a permanent difference that sustains through life. The impact on earnings at all ages, according to Lemma 4 in Appendix B, is similar to what we found in the

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

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

simple model in Corollary 1, and can be approximated by $\Delta \log z/(1 - \alpha_W) = 1.3\%$ for a 1 year difference in mom's schooling. This is more or less similar to the selection effect on lifetime earnings, as we soon show in Section 5. Following similar calculations, a standard deviation of 0.117 for z translates into $\sigma_z/(1 - \alpha_W) = 0.204$ or a 20.4% standard deviation in earnings, once we control for schooling.

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

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

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

4.5 Fit Analysis

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

$$\text{selection}: \; b_2 = \beta\rho_{zh_P}\sigma_z/(1-\alpha)\sigma_{h_P} = \begin{cases} 0.019 & \text{if } \alpha = \alpha_1 + \alpha_2 \\ 0.013 & \text{if } \alpha = \alpha_W \end{cases} \tag{24}$$

$$\text{spillover}: \; \bar{b}_2 = \beta\nu / \left[1 - \lambda(1-\alpha)\right] \approx 0.048 \quad \text{for both cases.} \tag{25}$$

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

All else equal, tastes for schooling induce individuals to stay in school longer at the the detriment of lifetime earnings. To make up for this and explain a 2% observed, reduced form spillover, the estimated value of ν must be larger than what would be implied by the simple model. Moreover, since selection on tastes generates a large intergenerational schooling correlation, it needs to be moderated by a negative effect on schooling coming from a large ν. For both these reasons, if we were to ignore the taste heterogeneity as in (25), the implied spillover effect on earnings will be counterfactually high.

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

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

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

out, the reduced form effect is similar to when we ignore taste heterogeneity as in (24).

Shutting down parameters To interpret the effect of spillovers and ability selection in light of our empirical analysis from Section 3, it is useful to see how the gradient moments are affected when shutting down the key parameters (ν, ρ_{zh_P}).[26] These exercises also show that the intuition from the simple model in the previous sections carries over to the estimated model.

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

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

The third row is when there is no selection on either (h_P, z).

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

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

[26] We have tried similar exercises with λ, which had small effects because of its smaller estimated magnitude.

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 become negative. This implies that for schooling choices, selection on tastes dominates selection on abilities: even though children with high z would spend more time in school, they do not if 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 shut down, the numbers are more or less identical to when only γ_{h_P} is shut down.

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

5 Counterfactual Experiments

We conduct two main experiments using the model estimates. First, we decompose the effects of a one-year increase in mother's schooling. Second, we implement a counterfactual compulsory schooling reform. Although the spillover coefficient ν has only a small effect on the reduced form coefficient on mother's schooling in a Mincer regression, a uniform one-year increase in mother's schooling leads to an average 1.2% increase in children's earnings controlling for selection, while further allowing for selection leads to an additional 1.3% increase. And, this happens without increasing children's schooling. This is explained by parents with higher human capital having a negative effect on children's schooling as we saw in Proposition 1, which is countervailed by children of higher human capital parents having higher tastes for schooling.

5.1 Decomposing Spillovers from Selection

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

We perform several experiments to control for spillovers, selection effects and tastes for schooling. First, we hold constant the individual's (z, ξ) and also the schooling choice S, which isolates the pure quality effect coming from higher S_P. Next we still hold (z, ξ) constant, but let the individuals re-optimize their choice of S—since the higher initial human capital substitutes for the need to stay in school longer, earnings further increases while schooling decreases. The combined effect of the pure quality increase and schooling choice adjustment is the spillover effect. Then, we let either z or ξ vary with h_P as dictated by the distributional assumptions in Section 4.2; 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

	(1) fixed S	(2) spillover	(3) ability	(4) tastes	(5) RF
Schooling diff (years)	-	-0.425	-0.384	0.451	0.479
Avg earnings diff (%)	0.007	0.012	0.025	0.002	0.016

Table 9: Aggregate Effect of 1 Year Increase in Mom's Schooling.

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

S_P years of mother's schooling, to the average outcome of those with $S_P + 1$ years of mother's schooling. Refer to Appendix D for a more formal description.

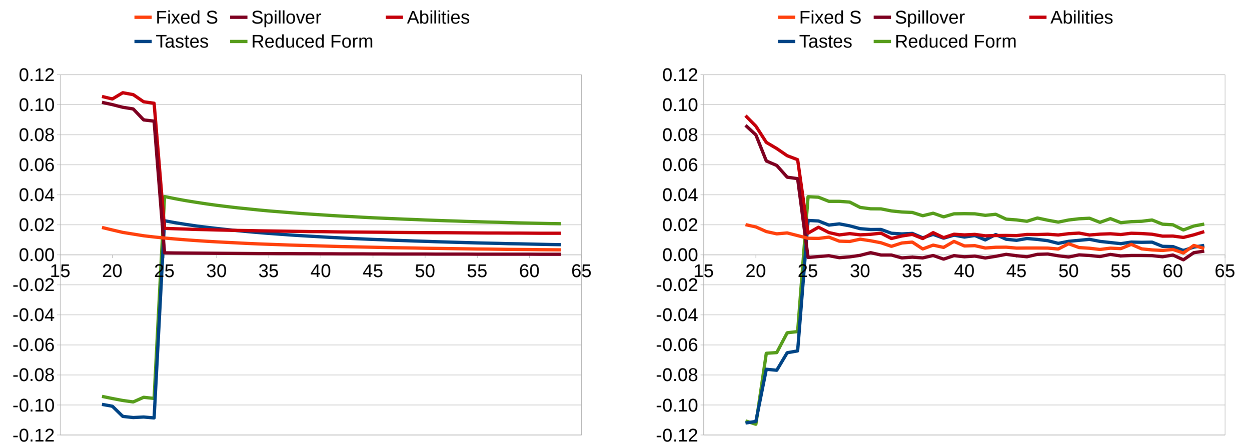
The first and second rows of Table 9 lists the average effects of a 1 year increase in S_P on schooling S and the present-discounted value of lifetime earnings, respectively. The change in S is in years and the changing in earnings in log-point differences, to approximate percentage changes. Column (1) shows the change in the child's earnings when his schooling is held constant. Columns (2) to (5) depict the change in schooling or earnings when holding each individual's (z, ξ) constant, when allowing for selection on abilities or tastes, and lastly the reduced form effect.

As expected, the spillover effect on schooling is negative, namely, the model predicts that increasing all mothers' schooling by a year would lead to an average 0.425 year decline in the schooling of the child generation. Surprisingly, allowing for selection on abilities only moderates this by 0.039 years (comparing columns (2) and (3)) or 0.028 years (columns (4) and (5)). As we saw in Table 8, selection on abilities does not have much of an effect on intergenerational schooling relationships once selection on tastes are taken into account. Since tastes already induce individuals to stay in school longer than what maximizes lifetime earnings, letting abilities become higher does not further increase schooling much.[27]

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

The spillover increases lifetime earnings by about 0.7 percent holding schooling constant, which increases to 1.2 percent when allowing schooling to adjust. Selection on abilities has a 1.3 percent effect. We conclude that independent of selection on tastes, the causal effect of mother's education on earnings is more or less similar to the selection effect on abilities, i.e., high ability mothers having high ability children. Selection on tastes has a negative average impact: the increase in lifetime earnings drops by 1 percentage point once we allow for selection (2nd vs. 4th

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



(a) change in average

(b) change in median

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

columns). The effect is negative, since tastes for schooling make individuals deviate from lifetime earnings maximization, but not quite enough to dominate the positive effects from both the spillover and ability selection.

Table 9 shows the average change in the sum of net present discounted value of lifetime earnings, but the earnings effects differ substantially over the life-cycle, and also across children of mother's with different levels of schooling. Life-cycle effects are depicted in Figure 3, where each line plots the change in average (on the left panel) or median (on the right) earnings following a 1-year increase in mother's schooling by age, compared to the benchmark estimates. Comparing "Fixed S" and "Spillover," we see that the total spillover effect comes almost entirely from early labor market entry followed by almost no change in earnings after age 24 (the latest age we allow labor market entry in the model). This also means that on average, children of mothers with less schooling catch up with those of mothers with more schooling by staying in school longer, so that their earnings do not differ much later in life. Allowing for selection on abilities has a fixed positive effect throughout the life-cycle, as expected from Section 2. Conversely, longer schooling induced by selection on 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.

5.2 Counterfactual Compulsory Schooling Reform

We next impose a minimum schooling requirement which is intended to mimic compulsory schooling reforms that took place in many countries throughout the 20th century. Such a reform only affects those parents who would otherwise not attain the required level of schooling. It shows it is possible to simultaneously estimate a large schooling OLS coefficient and a small or negative IV coefficient. The evidence for spillovers would be found in children's earnings, not schooling.

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

		(1) fixed S	(2) spillover	(3) ability	(4) tastes	(5) RF
Schooling	OLS	-	0.492	0.487	0.458	0.455
	IV	-	-0.204	-0.148	0.227	0.261
Avg Earnings diff (%)		0.025	0.028	0.075	0.017	0.066

Table 10: Counterfactual Schooling Reform

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

capital of all children to

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

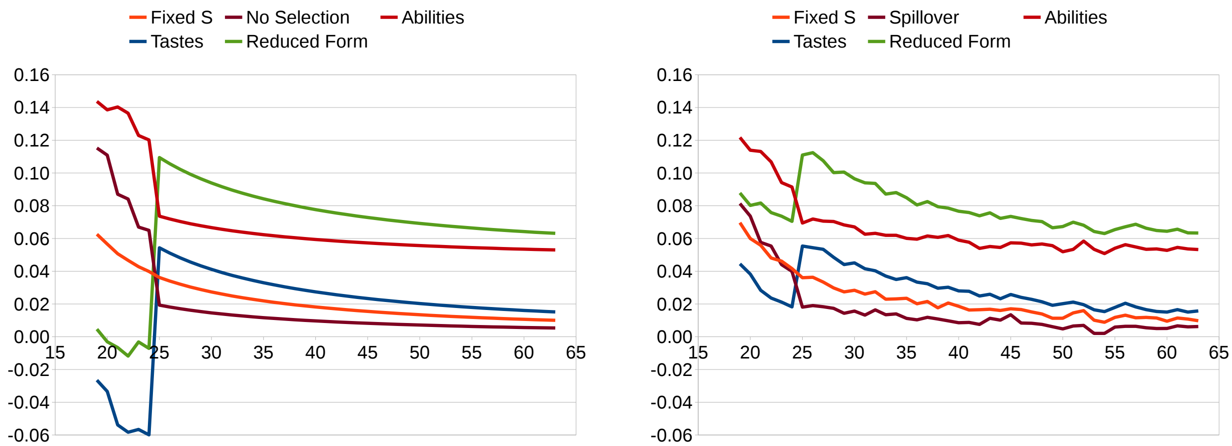
We choose 8 years as the hypothetical requirement because it was the compulsory schooling requirements in many U.S. states at the time, or soon after.[28] Such a reform would affect 25% of the individuals in our data, increasing the schooling of their mothers by an average of 3.5 years.

For the schooling OLS and IV regressions, we combine samples of two regimes: one without the minimum requirement (our benchmark model) and one imposing the requirement. These represent mother-child pairs pre- and post-reform, respectively. Then we run OLS and IV regressions on the merged data, using a dummy variable for the different regimes as an instrument. As above, we repeat this exercise for four cases controlling for selection on z and ξ, and including the selection effects from z and/or ξ. The top panel in 10 shows the regression results for all cases. The bottom panel shows the change in the present discounted value of lifetime earnings, in log-points.

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 9 qualitatively. The controlled effect is negative, but of smaller magnitude. The IV coefficient increases when including ability selection (columns 2 vs. 3) but much more when including taste selection (columns 2 vs. 4). The reduced form effect is only half of the population reduced form effect following a 1-year increase, in Table 9. As discussed there, this is because both the spillover and 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

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



(a) change in average

(b) change in median

Figure 4: Lifecycle Impact of Reform

columns 2 and 4, 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 of their children anywhere from -0.2 to 0.2 years.[29]

The spillover effect is positive through life, as can be seen in Figure 4(a), unlike in Figure 3 where it became virtually zero at ages 24 and above. Again, this is because for children with low S_P, the spillover effect is dominated by a pure quality effect, as can be seen by the much smaller difference between the "Fixed S" and "No Selection" lines. 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 be close to zero. But the negative effect is smaller in magnitude (although the lifetime earnings effect is -1.1 percentage points for both Tables 9 and 10, mother's schooling increases by 3.5 years in the latter compared to 1 year in the former). This is because these children have lower tastes for schooling to begin with, and have much to lose by earlier labor market entry because of decreasing returns

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

to human capital accumulation. This is even more obvious when we look at median earnings in Figure 4(b). There, at nowhere during the life-cycle does the reform have a negative effect, not even at early ages, when including selection on tastes.

6 Conclusion

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

Our model is consistent with several features of the joint distribution of parent schooling, child schooling and child earnings over the life-cycle. It is also consistent with a positive OLS correlation between parent and child schooling. The model feature that plays an important role in generating this feature is the correlation in tastes for schooling across generations. Thus, our model is consistent with the previous literature that finds when using compulsory schooling as an instrument for parents' schooling, the estimated IV coefficient on children's schooling is zero. The intuition is straightforward. On the one hand, all else equal, increasing a parent's human capital decreases a child's schooling, a feature of diminishing returns to human capital accumulation). On the other hand, a parent with higher human capital tends to have higher learning ability which transmits to the child, leading to increased schooling. Because these two effects countervail each other, the IV estimate from model simulated data is zero.

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

References

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

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

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.

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

Benabou, R. (1996, April). Equity and efficiency in human capital investment: The local connection. *Review of Economic Studies 63*(2), 237–64.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

For Online Publication

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 $a \in [6+S, R)$,*

$$q_2(a) = \frac{w}{r} \cdot q(a) \tag{26}$$

$$m(a) = \alpha_2 \left[\kappa q(a) z\right]^{\frac{1}{1-\alpha}} \tag{27}$$

$$h(a) = h_S + \frac{r}{w} \cdot \left[\int_{6+S}^{a} q(x)^{\frac{\alpha}{1-\alpha}} dx\right] \cdot (\kappa z)^{\frac{1}{1-\alpha}} \tag{28}$$

and

$$\frac{wh(a)n(a)}{\alpha_1} = \frac{m(a)}{\alpha_2}, \tag{29}$$

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 (29), 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 (26) is the solution. Using this result in (5)-(6) yields the solution for m, (27). Substituting (26), (27) and (29) into equation (1b) trivially leads to (28). $\square$

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 $a \in [6, 6+S)$,*

$$q_1(a) = e^{r(a-6)} q_0 \tag{30}$$

$$m(a)^{1-\alpha_2} = \alpha_2 e^{r(a-6)} \cdot q_0 z \tag{31}$$

$$h(a)^{1-\alpha_1} = h_0^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left[e^{\frac{\alpha_2 r(a-6)}{1-\alpha_2}} - 1 \right] \cdot (\alpha_2 q_0)^{\frac{\alpha_2}{1-\alpha_2}} z^{\frac{1}{1-\alpha_2}}. \tag{32}$$

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 (30). Then equation (31) follows directly from (6), and using this in (1b) 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 (32) is the solution. $\square$

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*

$$q_0 = \frac{e^{-rS}}{\alpha_2^{\alpha_2}} \cdot \left([\kappa q(6+S)]^{1-\alpha_2} z^{\alpha_1} \right)^{\frac{1}{1-\alpha}} \tag{33}$$

$$h_S = \frac{\alpha_1}{w} \cdot [\kappa q(6+S) z]^{\frac{1}{1-\alpha}}. \tag{34}$$

Proof. The value matching for this problem boils down to setting $n(6+S) = 1$ in the working phase, which yields (34).[30] 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 (34) we obtain (33). □

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 (33)-(34) into (32) evaluated at age $6+S$:

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

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,

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

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

$$z^{1-\lambda(1-\alpha)} h_P^{-\nu(1-\alpha)} \leq F(S), \tag{35}$$

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

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

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

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

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

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

and S is determined by (35) 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 Extended 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, Extended *Assume that the solution to the income maximization problem is such that $n(a) = 1$ for $a \in [6+S, 6+S+J)$ for some $J \in [0, R-6-S)$. Then given $h_S \equiv h(6+S)$, the value function for $a \in [6+S+J, R)$ can be written as*

$$V(a,h) = \frac{w}{r} \cdot q(a)h + D_W(a) \tag{36}$$

and the solution is characterized by

$$n(a)h(a) = \left[\frac{\alpha_W}{r} \cdot q(a)z\right]^{\frac{1}{1-\alpha_W}} \tag{37}$$

$$h(a) = h_J + \left(\frac{\alpha_W}{r}\right)^{\frac{\alpha_W}{1-\alpha_W}} \cdot \left[\int_{6+S+J}^{a} q(x)^{\frac{\alpha_W}{1-\alpha_W}} dx\right] \cdot z^{\frac{1}{1-\alpha_W}}, \tag{38}$$

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

$$V(a,h) = e^{r(a-6-S)} q_S \cdot \frac{h^{1-\alpha_W}}{1-\alpha_W} + e^{-r(6+S+J-a)} D(J, h_J) \tag{39}$$

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

$$h(a)^{1-\alpha_W} = h_S^{1-\alpha_W} + (1-\alpha_W)(a-6-S)z. \tag{40}$$

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

$$q_S = we^{-rJ} \cdot \left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J)z^{\alpha_W}\right]^{\frac{1}{1-\alpha_W}} \tag{41}$$

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

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

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

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

$$\frac{z}{h_S^{1-\alpha_W}} \leq G(J) \equiv \left[\frac{\alpha_W}{r} \cdot q(6+S+J) - (1-\alpha_W)J\right]^{-1}. \tag{43}$$

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

$$G(0) < \frac{z}{h_S^{1-\alpha_W}} \qquad \Leftrightarrow \qquad \frac{r}{\alpha_W q(6+S)} < \frac{z}{h_S^{1-\alpha_W}}, \tag{44}$$

and J is determined by (43) at equality. Otherwise $J = 0$.

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

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

1. *if $J = 0$,*

$$\epsilon + (1-\alpha_2)\left[\frac{\alpha_2^{\alpha_2} w}{r}\cdot q(6+S) z h_S^{\alpha_1}\right]^{\frac{1}{1-\alpha_2}} = w\cdot\left(h_S + (1-\alpha_W)\left[\frac{\alpha_W^{\alpha_W}}{r}\cdot q(6+S)z\right]^{\frac{1}{1-\alpha_W}}\right) \tag{45}$$

$$h_S^{1-\alpha_1} \le h_0^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2}\cdot\left(1-e^{-\frac{\alpha_2 rS}{1-\alpha_2}}\right)\cdot\left[\frac{\alpha_2 w}{r}\cdot q(6+S)h_S^{\alpha_1}\right]^{\frac{\alpha_2}{1-\alpha_2}} z^{\frac{1}{1-\alpha_2}} \tag{46}$$

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

$$q_0 = \frac{we^{-rS}}{r}\cdot q(6+S)h_S^{\alpha_1}. \tag{47}$$

2. *if $J > 0$,*

$$\begin{aligned}&\epsilon + (1-\alpha_2)\left(\alpha_2^{\alpha_2} we^{-rJ}\left[\frac{\alpha_W^{\alpha_W}}{r}\cdot q(6+S+J)z\right]^{\frac{1}{1-\alpha_W}}\cdot h_S^{\alpha_1-\alpha_W}\right)^{\frac{1}{1-\alpha_2}}\\ &\quad = we^{-rJ}\left[\frac{\alpha_W^{\alpha_W}}{r}\cdot q(6+S+J)z\right]^{\frac{1}{1-\alpha_W}}\end{aligned} \tag{48}$$

$$\begin{aligned}h_S^{1-\alpha_1} \le h_0^{1-\alpha_1} &+ \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2}\cdot\left(1-e^{-\frac{\alpha_2 rS}{1-\alpha_2}}\right)\\ &\cdot\left(\alpha_2 we^{-rJ}\left[\frac{\alpha_W^{\alpha_W}}{r}\cdot q(6+S+J)z^{\alpha_W}\right]^{\frac{1}{1-\alpha_W}} h_S^{\alpha_1-\alpha_W}\right)^{\frac{\alpha_2}{1-\alpha_2}}\cdot z^{\frac{1}{1-\alpha_2}}\end{aligned} \tag{49}$$

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

$$q_0 = we^{-r(S+J)}\cdot\left[\frac{\alpha_W^{\alpha_W}}{r}\cdot q(6+S+J)z^{\alpha_W}\right]^{\frac{1}{1-\alpha_W}}\cdot h_S^{\alpha_1-\alpha_W}. \tag{50}$$

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

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

Hence (47) follows from the smooth pasting condition. Likewise, (45) follow from plugging $n(6+S)$, $m(6+S)$ from Lemmas 2 and 4 and q_0 from (47) in the value matching condition. Lastly, (46) 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 (50) follows from plugging q_S and h_S from (41)-(42) in the smooth pasting condition. Likewise, (48) follow from plugging $n(6+S) = 1$, $m(6+S)$ from Lemma 2, and q_0 from (50) in the value matching condition. Again, (49) requires consistency between h_S and $h(6+S)$. □

For each case where we assume $J = 0$ or $J > 0$, it must also be the case that condition (44) 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 (46) and (49), to obtain h_S. Note that there is always a solution to (46) or (49)—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)

$$1 = \left(\frac{h_0}{\mathbf{h_S}}\right)^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left(1 - e^{-\frac{\alpha_2 rS}{1-\alpha_2}}\right) \cdot \left[\frac{\alpha_2 w}{r} \cdot q(6+S)\right]^{\frac{\alpha_2}{1-\alpha_2}} z^{\frac{1}{1-\alpha_2}} \cdot \mathbf{h_S}^{-\frac{1-\alpha}{1-\alpha_2}} \tag{51}$$

$$\begin{aligned}
1 = \left(\frac{h_0}{\mathbf{h_S}}\right)^{1-\alpha_1} &+ \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left(1 - e^{-\frac{\alpha_2 rS}{1-\alpha_2}}\right) \\
&\cdot \left(\alpha_2 w e^{-rJ} \left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J) z^{\alpha_W}\right]^{\frac{1}{1-\alpha_W}}\right)^{\frac{\alpha_2}{1-\alpha_2}} \cdot z^{\frac{1}{1-\alpha_2}} \cdot \mathbf{h_S}^{-\frac{1-\alpha+\alpha_2\alpha_W}{1-\alpha_2}},
\end{aligned} \tag{52}$$

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 (51) and (52) 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 (44) holds, we may have two solutions:

1. If only one solution satisfies (44), it is the solution.
2. If both satisfy (44), 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 (36) 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 (9).

3. For each node on the h_P-grid, construct z-grids according to (20), 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 (22)-(23).

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 Formal Description of Experiments in Section 5

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, $S = S(x)$, $E(a) = \tilde{E}(x; S; a)$.[31] Then schooling following a j-year increase in S_P, holding (z, ξ) constant, is

$$
S_v^j(x) \equiv S(S_P + j, z, \xi). \tag{53}
$$

[31] Since although the parent variable in the initial condition is h_P, it is defined as $\log h_P = \beta S_P$ in (9).

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

$$E_0^j(x;a) \equiv \tilde{E}(S_P + j, z, \xi; S; a)|_{S=S(S_P,z,\xi)} \tag{54}$$

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

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

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

$$E_v^j(x;a) \equiv \tilde{E}(S_P + j, z, \xi; a), \tag{55}$$

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

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

respectively, where $\Delta_\xi^j(S_P)$ is a 6-dimensional vector for each level of schooling $S \in \{8, \ldots, 18\}$. The first expression follows since $(S_P, \log z)$ are joint-normal, and the second from the definition of tastes in (21). Then schooling and age a earnings following a j-year increase in S_P, including partial selection effects on z or ξ, are

$$S_z^j(x) \equiv S(S_P + j, z \cdot \Delta_z^j, \xi), \qquad E_z^j(x;a) \equiv E(S_P + j, z \cdot \Delta_z^j, \xi; a), \tag{56}$$

$$S_\xi^j(x) \equiv S(S_P + j, z, \xi + \Delta_\xi^j(S_P)), \qquad E_\xi^j(x;a) \equiv E(S_P + j, z, \xi + \Delta_\xi^j(S_P); a). \tag{57}$$

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

$$S_{rf}^j(x) \equiv S(S_P + j, z \cdot \Delta_z^j, \xi + \Delta_\xi^j(S_P) + \Delta_{z\xi}^j(z)) \tag{58a}$$

$$E_{rf}^j(x;a) \equiv E(S_P + j, z \cdot \Delta_z^j, \xi + \Delta_\xi^j(S_P) + \Delta_{z\xi}^j(z); a), \tag{58b}$$

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 9 is obtained by integrating the change from $S(x)$ in (53) and (56)-(58) over the population distribution Φ, when $j = 1$. The second row is the outcome of

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

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

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

and each bar in the left and right panels of Figure **??** plots, respectively,

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

for the distribution in change in schooling S, and

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

for the distribution in the change in average lifetime earnings, for $k \in \{0, \nu, z, \xi, rf\}$.

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

E Tables and Figures not in text

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

Table 11: Identification

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

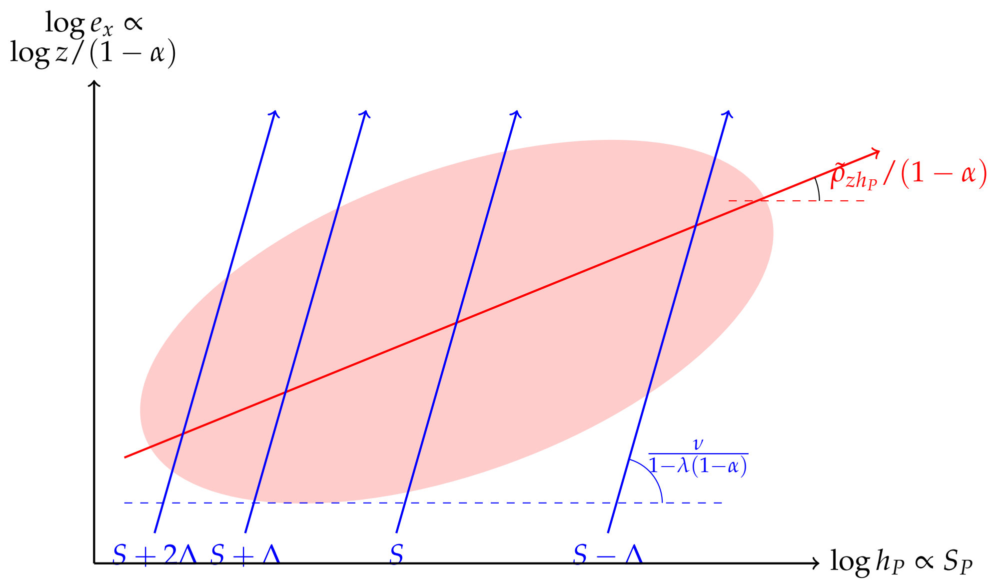


Figure 5: Intuition for Identifying Ability Selection and Spillovers.
x-axis: parents' human capital, or their schooling levels, y-axis: children's log abilities, or log earnings controlling for own schooling. The pink area represents the distribution of abilities and parents' schooling, (h_P, z). According to our model, selection is captured by the population correlation between parent's schooling and earnings, controlling for schooling. This is captured by the slope of the red line. Spillovers are captured by the relationship between earnings and abilities among children with the same level of schooling, which is captured by the slope of the blue lines. Note that conditional on abilities, schooling is decreasing in parents' schooling.

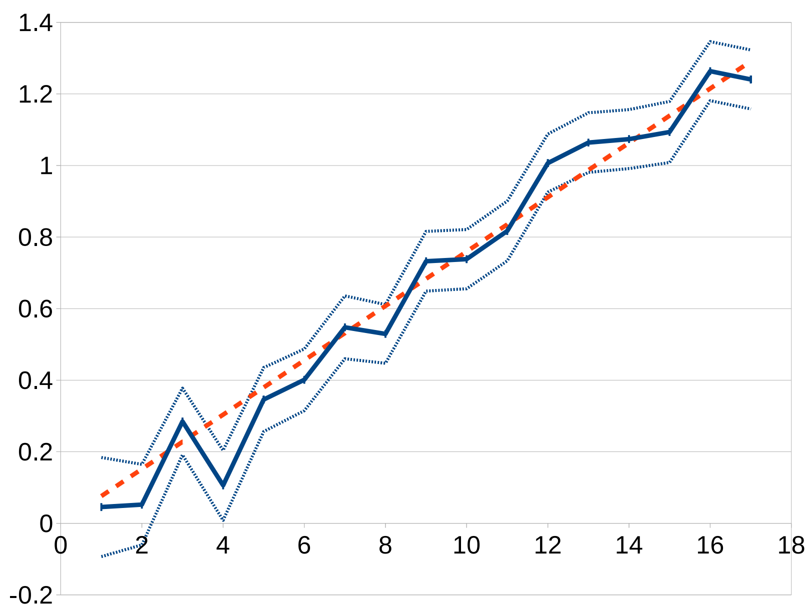
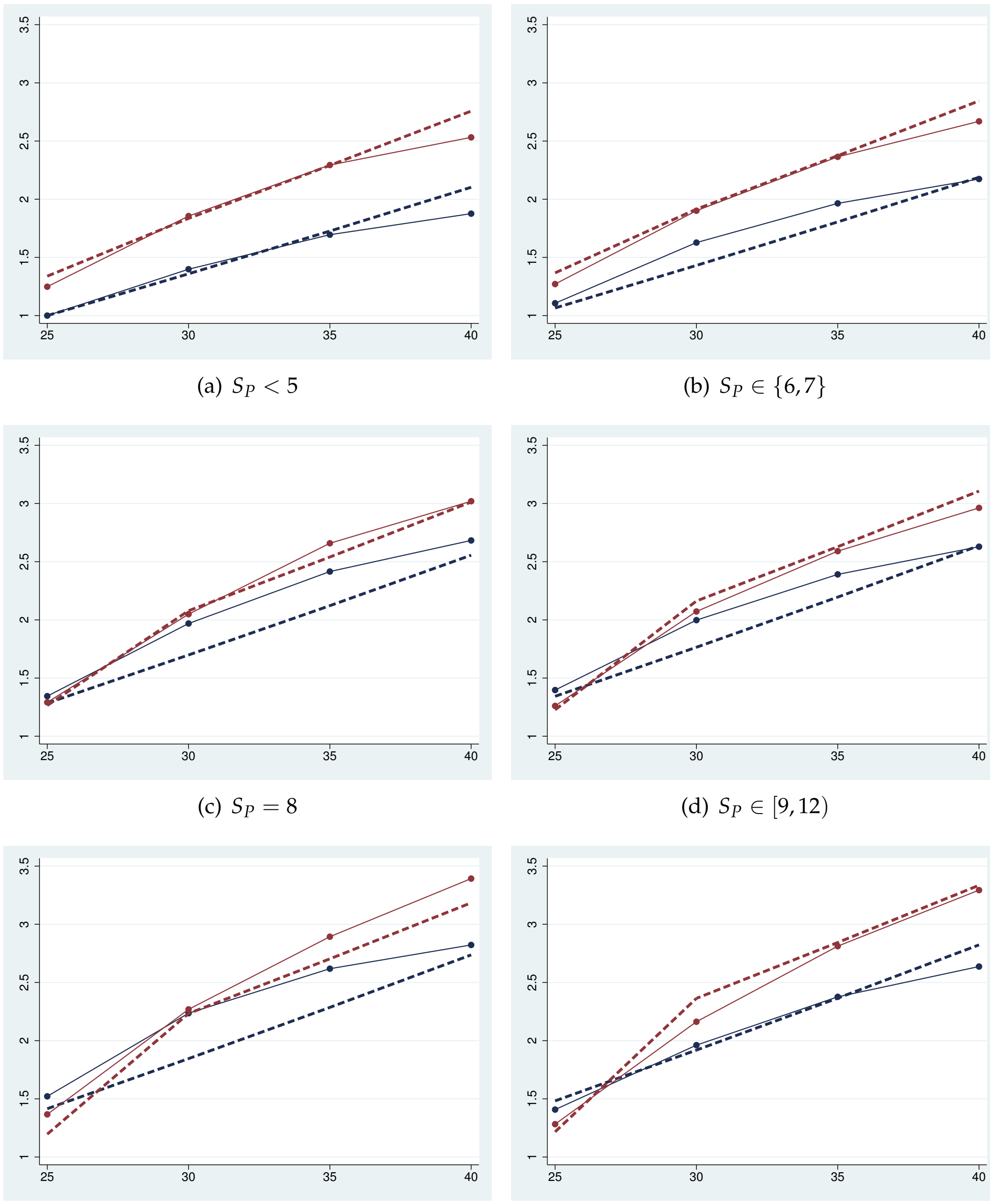


Figure 6: Mincerian Return to Schooling, Linear vs. Dummies.



(a) $S_P < 5$

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

(c) $S_P = 8$

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

(e) $S_P = 12$

(f) $S_P > 12$

Figure 7: Model Fit

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