For Online Publication

A Proofs to Propositions 1 and 2

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

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

$$V(a,h) = q_2(a)h + v_W(a), \qquad \text{for } a \in [6+S, R),$$
$$V(a,h) = q_1(a) \cdot \frac{h^{1-\alpha_1}}{1-\alpha_1} + e^{-r(6+S-a)} \cdot v(S, h_S), \qquad \text{for } a \in [6, 6+S), \text{if } S > 0,$$

where

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

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

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

$$q_2(a) = \frac{w}{r} \cdot q(a) \tag{17}$$
$$m(a) = \alpha_2 \left[\kappa q(a) z\right]^{\frac{1}{1-\alpha}} \tag{18}$$
$$h(a) = h_S + \frac{r}{w} \cdot \left[\int_{6+S}^{a} q(x)^{\frac{\alpha}{1-\alpha}} dx\right] \cdot (\kappa z)^{\frac{1}{1-\alpha}} \tag{19}$$

and

$$\frac{wh(a)n(a)}{\alpha_1} = \frac{m(a)}{\alpha_2}, \tag{20}$$

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 (3a) holds at equality, dividing by (3b) leads to equation (20), so once we know the optimal path of $h(a)$ and $m(a)$, $n(a)$ can be expressed explicitly. Plugging (3a) and the guess for the value function into equation (4), we obtain the linear, non-homogeneous first order differential equation

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

to which (17) is the solution. Using this result in (3a)-(3b) yields the solution for m, (18). Substituting (17), (18) and (20) into equation (1c) trivially leads to (19). $\square$

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

$$q_1(a) = e^{r(a-6)} q_0 \tag{21}$$

$$m(a)^{1-\alpha_2} = \alpha_2 e^{r(a-6)} \cdot q_0 z \tag{22}$$

$$h(a)^{1-\alpha_1} = h_0^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left[e^{\frac{\alpha_2 r(a-6)}{1-\alpha_2}} - 1 \right] \cdot (\alpha_2 q_0)^{\frac{\alpha_2}{1-\alpha_2}} z^{\frac{1}{1-\alpha_2}}. \tag{23}$$

Proof. Using the guess for the value function in (4) we have

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

to which solution is (21). Then equation (22) follows directly from (3b), and using this in (1c) yields the first order ordinary differential equation

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

to which (23) 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*

$$q_0 = \frac{e^{-rS}}{\alpha_2^{\alpha_2}} \cdot \left([\kappa q(6+S)]^{1-\alpha_2} z^{\alpha_1} \right)^{\frac{1}{1-\alpha}} \tag{24}$$

$$h_S = \frac{\alpha_1}{w} \cdot [\kappa q(6+S) z]^{\frac{1}{1-\alpha}} . \tag{25}$$

Proof. The value matching for this problem boils down to setting $n(6+S) = 1$ in the working phase, which yields (25).[53] 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 (25) we obtain (24). □

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

$$\begin{aligned} wh(a)(1-n(a)) &= wh(a) - wh(a)n(a) \\ &= (\kappa z)^{\frac{1}{1-\alpha}} \left\{ \alpha_1 \left[q(6+S)^{\frac{1}{1-\alpha}} - q(a)^{\frac{1}{1-\alpha}} \right] + r \cdot \int_{6+S}^{a} q(x)^{\frac{\alpha}{1-\alpha}} dx \right\} \\ &= (\kappa z)^{\frac{1}{1-\alpha}} \left\{ -\frac{\alpha_1}{1-\alpha} \cdot \int_{6+S}^{a} q(x)^{\frac{\alpha}{1-\alpha}} q'(x) dx + r \cdot \int_{6+S}^{a} q(x)^{\frac{\alpha}{1-\alpha}} dx \right\} \\ &= (\kappa z)^{\frac{1}{1-\alpha}} \left\{ \int_{6+S}^{a} q(x)^{\frac{\alpha}{1-\alpha}} \left[r - \frac{\alpha_1}{1-\alpha} q'(x) \right] dx \right\} \end{aligned}$$

after some simplification and inserting $q'(x) = -re^{-r(R-x)}$. We must still show Proposition 1.

Proof of Proposition 1. The length of the schooling period can be determined by plugging equations (24)-(25) into (23) 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 $S > 0$. This implies that human capital accumulation must be positive in school-

[53] 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.

ing, which is guaranteed by the law of motion for human capital. Rearranging terms,

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

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

$$z^{1-\lambda(1-\alpha)} h_P^{-\nu(1-\alpha)} \leq F(S), \tag{26}$$

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

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

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

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

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

and S is determined by (26) at equality. The full solution is given by Lemmas 1-3 and we obtain Proposition 2. Otherwise $S = 0$ and the solution is given by Lemma 1. □

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

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

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

$$V(a,h) = \frac{w}{r} \cdot q(a)h + D_W(a) \tag{27}$$

and the solution is characterized by

$$n(a)h(a) = \left[\frac{\alpha_W}{r} \cdot q(a)z\right]^{\frac{1}{1-\alpha_W}} \tag{28}$$

$$h(a) = h_J + \left(\frac{\alpha_W}{r}\right)^{\frac{\alpha_W}{1-\alpha_W}} \cdot \left[\int_{6+S+J}^{a} q(x)^{\frac{\alpha_W}{1-\alpha_W}} dx\right] \cdot z^{\frac{1}{1-\alpha_W}}, \tag{29}$$

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

$$V(a,h) = e^{r(a-6-S)} q_S \cdot \frac{h^{1-\alpha_W}}{1-\alpha_W} + e^{-r(6+S+J-a)} D(J, h_J) \tag{30}$$

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

$$h(a)^{1-\alpha_W} = h_S^{1-\alpha_W} + (1-\alpha_W)(a-6-S)z. \tag{31}$$

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

$$q_S = we^{-rJ} \cdot \left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J)z^{\alpha_W}\right]^{\frac{1}{1-\alpha_W}} \tag{32}$$

$$h_J = \left[\frac{\alpha_W}{r} \cdot q(6+S+J)z\right]^{\frac{1}{1-\alpha_W}}. \tag{33}$$

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,

$$\frac{z}{h_S^{1-\alpha_W}} \leq G(J) \equiv \left[\frac{\alpha_W}{r} \cdot q(6+S+J) - (1-\alpha_W)J\right]^{-1}. \tag{34}$$

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

$$G(0) < \frac{z}{h_S^{1-\alpha_W}} \qquad \Leftrightarrow \qquad \frac{r}{\alpha_W q(6+S)} < \frac{z}{h_S^{1-\alpha_W}}, \tag{35}$$

and J is determined by (34) 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 $J = 0$,*

$$\epsilon + (1-\alpha_2)\left[\frac{\alpha_2^{\alpha_2} w}{r} \cdot q(6+S) z h_S^{\alpha_1}\right]^{\frac{1}{1-\alpha_2}} = w \cdot \left(h_S + (1-\alpha_W)\left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S) z\right]^{\frac{1}{1-\alpha_W}}\right) \tag{36}$$

$$h_S^{1-\alpha_1} \leq h_0^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left(1 - e^{-\frac{\alpha_2 r S}{1-\alpha_2}}\right) \cdot \left[\frac{\alpha_2 w}{r} \cdot q(6+S) h_S^{\alpha_1}\right]^{\frac{\alpha_2}{1-\alpha_2}} z^{\frac{1}{1-\alpha_2}} \tag{37}$$

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

$$q_0 = \frac{we^{-rS}}{r} \cdot q(6+S) h_S^{\alpha_1}. \tag{38}$$

2. *if $J > 0$,*

$$\begin{aligned} \epsilon + (1-\alpha_2)\left(\alpha_2^{\alpha_2} w e^{-rJ}\left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J) z\right]^{\frac{1}{1-\alpha_W}} \cdot h_S^{\alpha_1-\alpha_W}\right)^{\frac{1}{1-\alpha_2}} \\ = we^{-rJ}\left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J) z\right]^{\frac{1}{1-\alpha_W}} \end{aligned} \tag{39}$$

$$\begin{aligned} h_S^{1-\alpha_1} \leq h_0^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left(1 - e^{-\frac{\alpha_2 r S}{1-\alpha_2}}\right) \\ \cdot \left(\alpha_2 w e^{-rJ}\left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J) z^{\alpha_W}\right]^{\frac{1}{1-\alpha_W}} h_S^{\alpha_1-\alpha_W}\right)^{\frac{\alpha_2}{1-\alpha_2}} \cdot z^{\frac{1}{1-\alpha_2}} \end{aligned} \tag{40}$$

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

$$q_0 = we^{-r(S+J)} \cdot \left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J) z^{\alpha_W} \right]^{\frac{1}{1-\alpha_W}} \cdot h_S^{\alpha_1 - \alpha_W}. \tag{41}$$

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

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

Hence (38) follows from the smooth pasting condition. Likewise, (36) follow from plugging $n(6+S)$, $m(6+S)$ from Lemmas 2 and 4 and q_0 from (38) in the value matching condition. Lastly, (37) merely states that the optimal h_S must be consistent with optimal accumulation in the schooling phase, $h(6+S)$.

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

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

Hence (41) follows from plugging q_S and h_S from (32)-(33) in the smooth pasting condition. Likewise, (39) follow from plugging $n(6+S) = 1$, $m(6+S)$ from Lemma 2, and q_0 from (41) in the value matching condition. Again, (40) 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 (35) does not or does hold.

C Identifying b_{select} and λ

Applying (9) to (6) in Proposition 2 yields

$$\log e_{i,x} = \log C(x; S_i) + \frac{\tilde{\mu}_z}{1-\alpha} + b_{\text{select}} \cdot S_{Pi} + \frac{\epsilon_i}{1-\alpha}. \tag{42}$$

where $e_{i,x}$ is individual i's earnings at x years of experience, and the selection effect of parents' schooling on children's earnings, b_{select}, was defined in (10b). If we knew the function $C(\cdot)$, we could regress log-earnings on parents' schooling, controlling for experience and own schooling, to identify b_{select}.

A byproduct of Corollary 1 is that the schooling-specific experience effect $\varphi_{s,x}$ should capture the functions $C(x;s)$ and $F(s)$ (Equation 12). If we were able to separate the two, either (10b) or (42) could be used to infer b_{select}. However, this is generally infeasible for both theoretical and empirical reasons. Theoretically, $\log C(0;s)$ diverges to $-\infty$ in the model, but $x=0$ is required for identification. And in the data, not only do we not observe zero earnings but neither do we observe early-experience earnings for individuals with low levels of schooling.

To gain some traction independently of the structural estimation in Section 4, take an nth-order approximation around (12):

$$\log C(x;s) + \frac{\log F(s)}{(1-\alpha)[1-\lambda(1-\alpha)]} = A_0 + \sum_{j=1}^{n} A_j^s (s-\hat{s})^j + \sum_{m=1}^{n}\sum_{j=0}^{m} A_{mj}^x (s-\hat{s})^j (x-\hat{x})^{m-j} + \dots \tag{43}$$

where $(\hat{x},\hat{s})$ are some reference levels of experience and schooling, and (A_j^s, A_{mj}^s) are some constants. Then instead of including a full set of schooling-specific experience effects as in (11), regress

$$\log e_{i,x}^* = a_0 + \sum_{j=1}^{n} a_j^s S_i^j + \sum_{m=1}^{n}\sum_{j=0}^{m} a_{mj}^x S_i^j x^{m-j} + bS_{P_i} + \tilde{u}_{i,x} \tag{44}$$

where (a_j^s, a_{mj}^x) are coefficients to be estimated. If the approximation is accurate, the estimate $\hat{b}$ would still capture b_{spill}. More importantly, the first sum captures all the variation coming from $F(s)$, which is only a function of s, and the second sum the variation from $C(\cdot)$. Thus

COROLLARY 2: IDENTIFYING SELECTION. *Suppose the assumptions in Corollary 1 hold, the approximation (43) is accurate, and we estimate (44). Then*

1. *Regressing $\sum_{j=1}^{n} \hat{a}_j^s S_i^j$ on S_{Pi} admits an unbiased estimate for $b_{select} - b_{spill}$.*
2. *Regressing $(\log e_{i,x}^* - \sum_{m=1}^{n}\sum_{j=0}^{m} a_{mj}^x S_i^j x^{m-j})$ on S_{Pi} admits an unbiased estimate for b_{select}.*

Once we know b_{spill} from Corollary 1, Part 1 can be used to back out b_{select}. Part 2 implies that once we filter out an individual's schooling-specific experience profile, the coefficient on parents' schooling S_P captures only b_{select} . This may be somewhat surprising, but a direct consequence of Proposition 2: (ν,λ) only affect the child's initial level of human capital, of which effect is absorbed into the length of schooling, which in turn only affects the shape of experience profiles.[54]

According to our model then, the difficulty in separating the two effects stems from how we control for own schooling S. In Corollary 1, the coefficient on parents' schooling identifies b_{spill}

[54]Of course, the result is an outcome of our model. The advantage of this model implication is that we can establish a lower-bound as opposed to an upper-bound for causal effects. Note that our identification scheme cannot speak to the general returns to an individual's own schooling, but only his parent's schooling.

	1.	2.
3rd order	0.114	0.106
	(37.72)	(117.40)
R^2	0.218	0.133
4th order	0.115	0.109
	(38.00)	(119.11)
R^2	0.218	0.136
Sample	5,114	89,893

Table C.11: Implied Selection Coefficient

Selection coefficient defined in (10b), implied by Corollary 2. Columns (1) and (2) correspond to parts 1 and 2 of the corollary, respectively. The 1st (2nd) panel uses coefficients from a regression of (log) earnings on moms' years of schooling and a 3rd (th) order polynomial of potential experience (age-6-S) and own 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.

because when holding own schooling S fixed, the regression reveals the relationship that z must have with S_P to be consistent with the observed level of S. But if we could directly control for how schooling affects earnings as in Corollary 2, the regression reveals the sample correlation that z must have with S_P. This intuition is visualized in Appendix Figure F.5. The parental spillover is understood as a child of a more educated parent getting more out of the same level of schooling (i.e., the quality of schooling). And within a sample of children with the same level of schooling, we can recover the combinations of parents' schooling and children's earnings that are consistent with the observed level of schooling, assuming that the spillover effect is subsumed in the choice of schooling.

In Table 3, we showed that the estimates of b_{spill} is robust to alternative specifications. In particular, columns (7)-(8) included 3rd and 4th-order polynomials of (x, S_i). We use these estimates to compute the implied values of b_{select} via the two methods suggested by Corollary 2.

A formal test rejects equality of parts 1 and 2, but the magnitudes are strikingly similar. Similarly, the estimates barely change from a 3rd to 4th order polynomial, suggesting that higher order terms are quantitatively irrelevant. All in all, the results imply that the selection effect is almost 6 times larger than the spillover effect. However, as we find in Section 4, most of selection is in fact driven by preferences rather than abilities.

Similarly, we can use Corollary 1 to estimate λ:

COROLLARY 3 *Suppose the assumptions in Corollary 1 hold, the approximation (43) is accurate, and we obtain estimates from (44). If we regress*

$$\log e^*_{i,x} - \sum_{m=1}^{n}\sum_{j=0}^{m} \hat{a}^x_{mj} S_i^j x^{m-j} = \left[\vartheta_{S_{Pi}} + b_{S_{Pi}} \sum_{j=1}^{n} \hat{a}^s_j S_i^j\right] + \tilde{u}_{ix} \tag{45}$$

where $\{\vartheta_{s_P}\}$ *is a set of parents' schooling fixed effects, and* $\tilde{u}_{i,x}$ *regression error, the estimates*

$$\hat{b}_{S_P} = b_{S_P} = 1/\left\{(1-\alpha)\left[1-\lambda(1-\alpha)\right]\right\}.$$

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

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

since we assume (8). Applying this and $S_{Pi} = \hat{S}_P$ in (6) yields

$$\log e_{i,x} - \log C(x; S_i) = K_{\hat{S}_P} + \frac{\log F(S_i)}{(1-\alpha)\left[1-\lambda(1-\alpha)\right]},$$

among individuals with $S_{Pi} = \hat{S}_P$, where $K_{\hat{S}_P}$ is some constant. Since the term in brackets in (45) apply only to individuals whose parents have the same level of schooling and the 1st and 2nd sums in (45) capture the functions $(\log C, \log F)$, respectively, b_{S_P} is estimated as claimed. □

Corollary 2 shows that among children whose *parents* attained the same years of schooling, the magnitude of λ is identified by the within-group variation in a function of log-earnings and a function of own schooling. So we can use (C, F) recovered from (44) to perform the suggested regression. The intuition is similar as Corollary 1, except that we focus on the within-group variation among different groups.

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

Intuitively, we can identify the three parameters $(\nu, \lambda, \rho_{zh_P})$ because we have 3 sets of observations: earnings e and own schooling S, and parents' schooling S_P. Ignoring experience, the model imposes two non-linear, monotonic relationships $e = e(S, z)$ and $S = S(S_P, z)$, and we additionally parameterized the distribution of (S_P, z). Therefore, all three parameters can be separately identified.

D Numerical Algorithm

For the purposes of our estimated model in which S is fixed, the solution method in Appendix B is straightforward. We need not worry about value-matching conditions and only need to solve the smooth-pasting conditions given S, which are equations (37) and (40), to obtain h_S. Note that there is always a solution to (37) or (40)—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)

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

$$\begin{aligned} 1 = \left(\frac{h_0}{\mathbf{h_S}}\right)^{1-\alpha_1} + \frac{(1-\alpha_1)(1-\alpha_2)}{r\alpha_2} \cdot \left(1 - e^{-\frac{\alpha_2 r S}{1-\alpha_2}}\right) \\ \cdot \left(\alpha_2 w e^{-rJ}\left[\frac{\alpha_W^{\alpha_W}}{r} \cdot q(6+S+J) z^{\alpha_W}\right]^{\frac{1}{1-\alpha_W}}\right)^{\frac{\alpha_2}{1-\alpha_2}} \cdot z^{\frac{1}{1-\alpha_2}} \cdot \mathbf{h_S}^{-\frac{1-\alpha+\alpha_2\alpha_W}{1-\alpha_2}}, \end{aligned} \tag{47}$$

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 (46) and (47) 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 (35) holds, we may have two solutions:

1. If only one solution satisfies (35), it is the solution.

2. If both satisfy (35), 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 (27) 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 mom's schooling, learning abilities and preferences 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 $\beta_P = 0.06$, construct the h_P-grid which is just a transformation of the S_P-grid according to (8).

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

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

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

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

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

and similarly for other possible choices of S.

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

E Formal Description of Experiments in Section 5

Formally, for any initial condition $x = (S_P, z, \tilde{\xi})$, the model implied schooling and age-a earnings outcomes can be written as functions of x, $S = S(x)$, $E(a) = \tilde{E}(x; S; a)$.[55] Then schooling following a j-year increase in S_P, holding $(z, \tilde{\xi})$ constant, is

$$S_v^j(x) \equiv S(S_P + j, z, \tilde{\xi}). \tag{48}$$

Age a earnings following a j-year increase in S_P, holding $(z, \tilde{\xi})$ and S constant, is

$$E_0^j(x; a) \equiv \tilde{E}(S_P + j, z, \tilde{\xi}; S; a)|_{S = S(S_P, z, \tilde{\xi})} \tag{49}$$

i.e., the schooling choice is fixed as if mom's education is S_P, but the earnings outcome, or amount of human capital accumulated, is computed assuming that mom's education is $S_P + j$. This cap-

[55]Since although the parent variable in the initial condition is h_P, it is defined as $\log h_P = \beta_P S_P$ in (8).

tures the spillover effect that is independent of quantity (schooling) adjustment. Now if we define

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

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

$$E_v^j(x;a) \equiv \tilde{E}(S_P + j, z, \tilde{\xi}; a), \tag{50}$$

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

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

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

$$S_z^j(x) \equiv S(S_P + j, z \cdot \Delta_z^j, \tilde{\xi}), \qquad E_z^j(x;a) \equiv E(S_P + j, z \cdot \Delta_z^j, \tilde{\xi}; a), \tag{51}$$

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

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

$$S_{rf}^j(x) \equiv S(S_P + j, z \cdot \Delta_z^j, \tilde{\xi} + \Delta_\xi^j(S_P) + \Delta_{z\xi}^j(z)) \tag{53a}$$

$$E_{rf}^j(x;a) \equiv E(S_P + j, z \cdot \Delta_z^j, \tilde{\xi} + \Delta_\xi^j(S_P) + \Delta_{z\xi}^j(z); a), \tag{53b}$$

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 correlation effect on preferences that comes from $(z, \tilde{\xi})$ being correlated, even conditional on h_P. We coin this the "reduced-form" effect since by construction,

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

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 (48) and (51)-(53)

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

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

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

F Tables and Figures not in text

Mom's S_P	Fraction (%)	Child's S	Average $\bar{S}$
≤5	12.75	≤11	6.35
		≥12	13.12
6-7	12.30	≤11	8.02
		≥12	13.53
8	21.76	≤12	10.72
		≥13	15.21
9-11	13.77	≤12	11.00
		≥13	15.24
12	30.00	≤12	11.21
		≥13	15.38
≥13	9.43	≤12	11.41
		≥13	15.83

Table F.1: Summary Moments

For moms with low S_P (the first four rows), we divide low/high educational attainment of children by whether or not he graduated from high school, while for the rest we divide by whether or not he advanced beyond high school. In the third column, $\bar{S}$ denotes the average years of schooling attained in each category. These numbers can be compared with the target moments in Table 5.

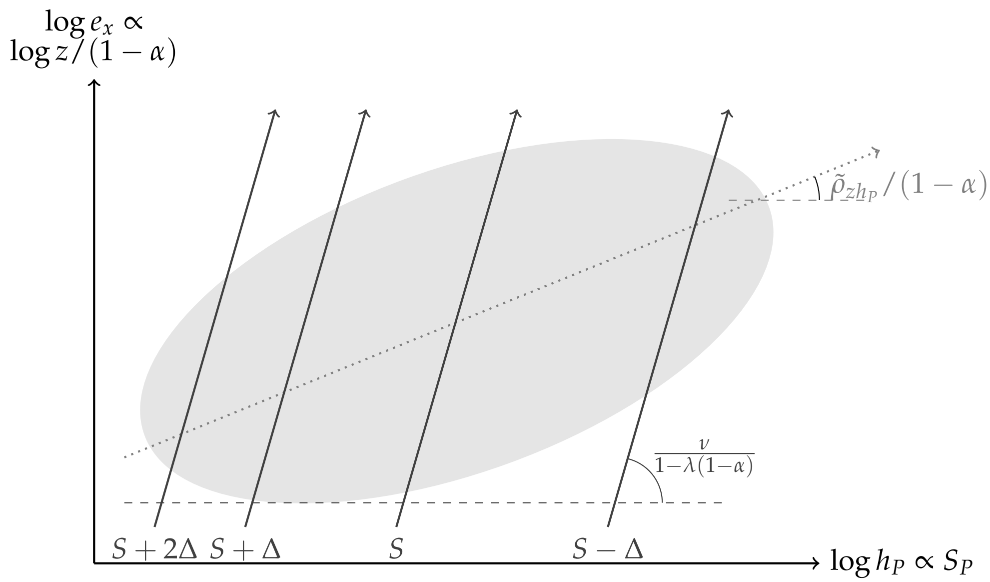


Figure F.5: Intuition for Identifying Ability Selection and Spillovers.
x-axis: parents' human capital, or their schooling levels, y-axis: children's log abilities, or log-earnings controlling for own schooling. The light gray area represents the distribution of abilities and parents' schooling, (h_P, z). According to our model, selection is captured by the population correlation between parent's schooling and earnings, controlling for schooling (given knowledge of the earnings function C, defined in Proposition 2). This is captured by the slope of the gray dotted 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 dark gray lines. Note that conditional on abilities, schooling is decreasing in parents' schooling.

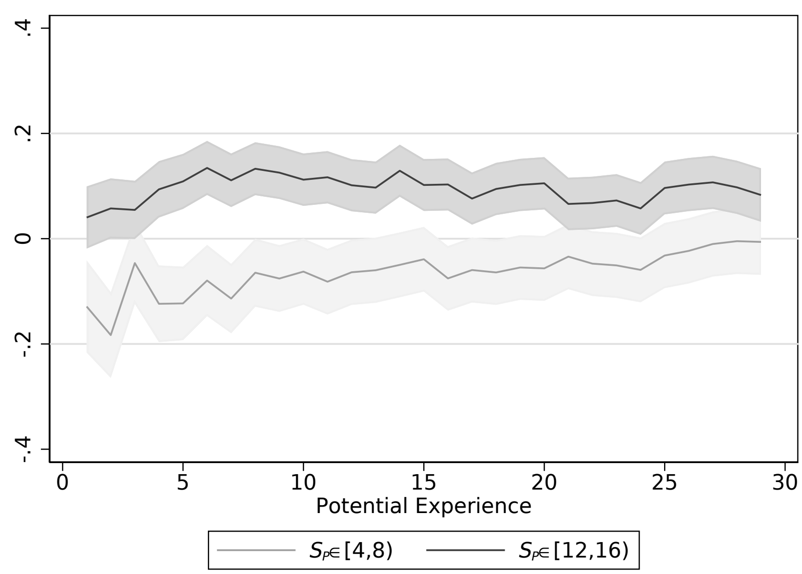


Figure F.6: Gaps between Average Earnings Profiles, All
Gaps between the earnings profiles of children by different levels of moms' schooling: 1924-1941 birth cohort. The y-axis is average log annual earnings in 2008 USD. Mothers' schooling levels are divided into 4 to 7 years and 12 or more years, and the light and dark lines correspond to the difference in the profiles compared to those whose moms attained 8 to 11 years of schooling.

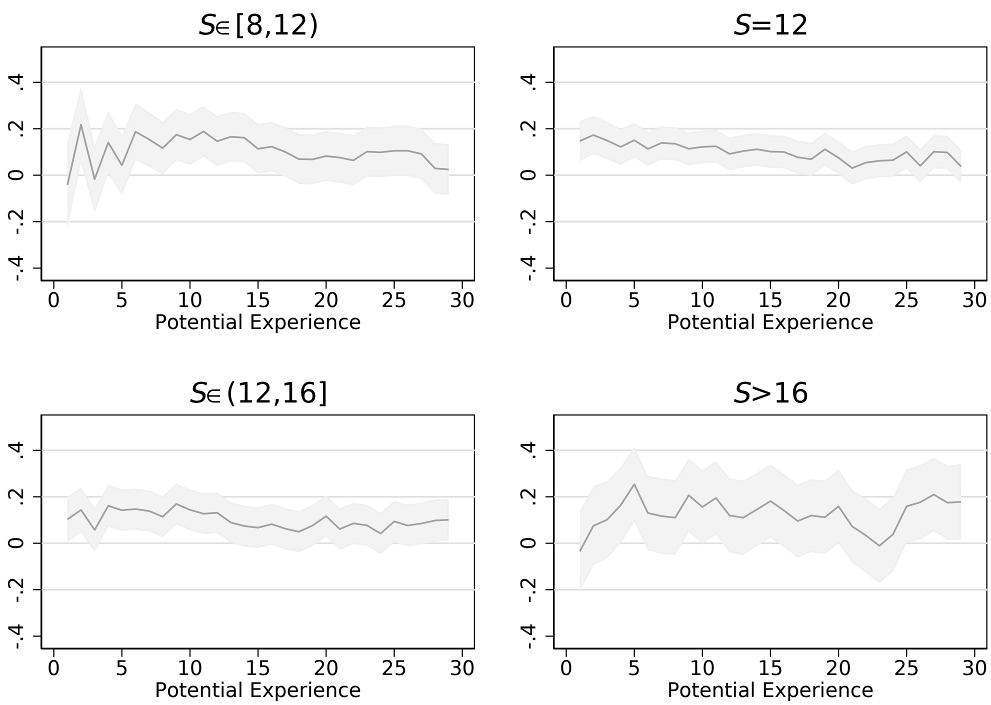
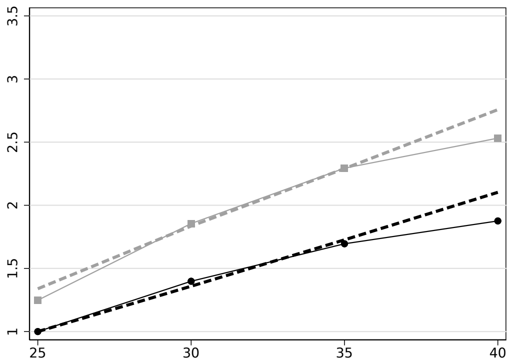
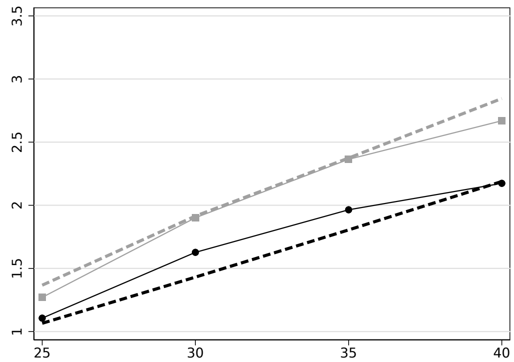


Figure F.7: Gaps between Average Earnings Profiles, All

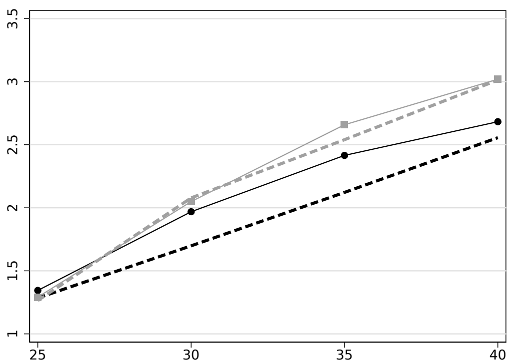
Gaps between the earnings profiles of children by different levels of moms' schooling: 1924-1941 birth cohort. The y-axis is average log annual earnings in 2008 USD. Mothers' schooling levels are divided into 4 to 7 years and 12 or more years, and the light and dark lines correspond to the difference in the profiles compared to those whose moms attained 8 to 11 years of schooling.



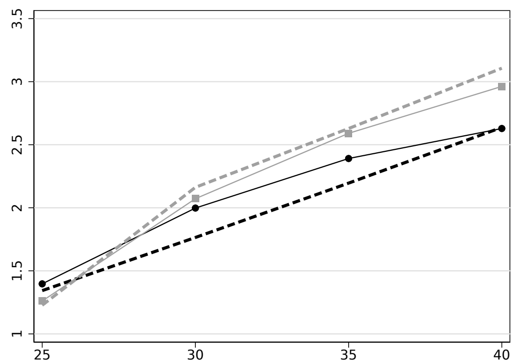
(a) $S_P < 5$



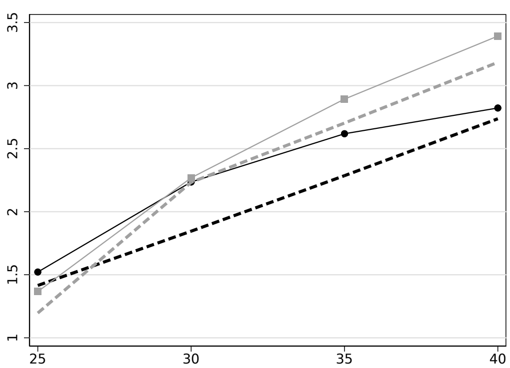
(b) $S_P \in \{6,7\}$



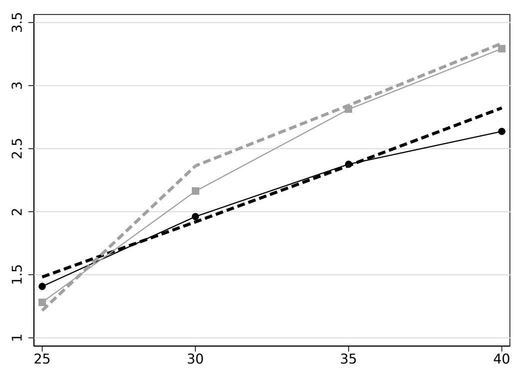
(c) $S_P = 8$



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



(e) $S_P = 12$



(f) $S_P > 12$

Figure F.8: 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 gray lines on top correspond to individual's with $S \geq 12$ for the first row of plots, and $S > 12$ for the rest. The dark gray lines correspond to the converse (less educated).