Supplemental Appendices

Chau Pham

A Technical details

A.1 Functional form of h

Let h be a concave function of $\hat{h}$ and I, i.e. $h = f(\hat{h}, I), f_i \geq 0, f_{ii} \leq 0$, where $i \in \{\hat{h}, I\}$. In case $I = 0, h = f(\hat{h}, 0) = f(\hat{h})$. Absent any uncertainty in the "schooling" process, the level of human capital of an agent is defined by f. On the contrary, when uncertainty arises, to be consistent with our setting, we assume h is uniformly distributed over the interval $[f(\hat{h}), f(\hat{h}, I)]$, that is $h \sim U\big(f(\hat{h}), f(\hat{h}, I)\big)$. We now prove that given f, our results remain robust under an appropriate set of conditions.

Proof. The expected utility of an agent is,

$$\begin{aligned}
U &= u(L-I) + p(h)\Big(u(H-I) - u(L-I)\Big) \\
&= -e^{\lambda(L-I)} + \Big[e^{-\lambda(L-I)} - e^{-\lambda(H-I)}\Big] \int_{f(\hat{h})}^{f(\hat{h},I)} \gamma h \big(f(\hat{h}, I) - f(\hat{h})\big)^{-1} \, \mathrm{d}h \qquad \text{(A.1)} \\
&= -e^{\lambda(L-I)} + \frac{\gamma}{2}\big(f(\hat{h}, I) + f(\hat{h})\big)\Big[e^{-\lambda(L-I)} - e^{-\lambda(H-I)}\Big]
\end{aligned}$$

The corresponding first and second derivative with respect to I are,

$$U' = e^{\lambda I}\Big\{ -\lambda e^{-\lambda L} + \frac{\gamma}{2}\big(e^{-\lambda L} - e^{-\lambda H}\big) \underbrace{\big[\lambda f(\hat{h}, I) + \lambda f(\hat{h}) + f_I\big]}_{A} \Big\} \qquad \text{(A.2)}$$

$$U'' = \lambda U' + \frac{\gamma}{2} e^{\lambda I}\big(e^{-\lambda L} - e^{-\lambda H}\big) \underbrace{(\lambda f_I + f_{II})}_{A'} \qquad \text{(A.3)}$$

For our standard results to hold, we require that $U'' > 0$ if $U' > 0$, which is equivalent to $A' \geq 0$ or A is non-decreasing with I and $A > \frac{2\lambda e^{-\lambda L}}{\gamma(e^{-\lambda L} - e^{-\lambda H})}$. Insofar that these two conditions are satisfied within the feasible range of I given λ, the expected utility is U-shaped in I and the discontinuity of investment occurs. In the basic model, the functional form used implies $f_{II} = 0$, which effectively translates into U'' being positive when U' is positive. □

A.2 Concavity of the expected utility function in equation 2.5

$$U = pE\Big[u\big(f(h) - I\big)\Big] + (1-p)u(L-I)$$

To prove that U is concave in I, we prove that $E\Big[u\big(f(h) - I\big)\Big]$ is concave in I.

Proof. Following the concavity of u, by definition,

$$u\big(f(h) - tI_1 - (1-t)I_2\big) = u\big(t(f(h) - I_1) + (t-1)(f(h) - I_2)\big) \tag{A.4}$$

$$\geq tu\big(f(h) - I_1\big) + (t-1)u\big(f(h) - I_2\big) \tag{A.5}$$

$$\Rightarrow E\Big[u\big(f(h) - tI_1 - (1-t)I_2\big)\Big] \geq E\Big[tu\big(f(h) - I_1\big) + (t-1)u\big(f(h) - I_2\big)\Big] \tag{A.6}$$

$$\geq tE\Big[u\big(f(h) - I_1\big)\Big] + (t-1)E\Big[u\big(f(h) - I_2\big)\Big] \tag{A.7}$$

where $t \in [0, 1]$ and the last line appeals to Jensen's inequality. As $E\Big[u\big(f(h) - I\big)\Big]$ is concave and u is also concave, their convex combination is concave, i.e. U is concave. □

A.3 Convergence of a non-linear recurrence relation

The sequence $\{h_{1,k}\}$ is of the general form

$$x_{t+1} = x_t^a + b \tag{A.8}$$

where $a, b \in (0, 1)$ and $x_0 > 1$. This is known as a non-linear recurrence relation, which is notorious for its lack of a closed form solution. We therefore do not attempt to find a closed form solution but instead find the limit towards which the sequence converges. To prove the existence of the limit of $\{x_t\}$, we need to prove that the sequence is monotonic and bounded.

Lemma A.1. *The sequence $\{x_t\}$ is monotonic. Namely, for all $t \in \mathbb{N}$, the sequence $\{x_t\}$ is either non-increasing or non-decreasing.*

Proof. Without loss of generality, assume $x_{k+1} \geq x_k$, we now prove that $x_{k+2} \geq x_{k+1}$.

$$x_{k+1} \geq x_k \Rightarrow\ x_{k+1}^a + b \geq x_k^a + b \Rightarrow\ x_{k+2} \geq x_{k+1}$$

Through a similar procedure, it can be shown that the opposite is also true, that is if $x_{k+1} \leq x_k$, $x_{k+2} \leq x_{k+1}$. □

Lemma A.2. *Sequence $\{x_t\}$ is bounded below if $\{x_t\}$ is non-increasing and bounded above if $\{x_t\}$ is non-decreasing.*

Proof. As the values of a and b are arbitrary, we consider two cases that arise, the non-increasing case and the non-decreasing case. In the former case, $x_{t+1} \leq x_t$ for all t while the opposite is true for the latter case.

Case 1: Sequence $\{x_t\}$ is non-increasing
Since $\{x_t\}$ is non-increasing by supposition, $x_{t+1} \leq x_t$. However, by definition, $x_t > b$ for all t. Consequently, $\{x_t\}$ is bounded below when $\{x_t\}$ is non-increasing.

Case 2: Sequence $\{x_t\}$ is non-decreasing
Let $F(y) = y^a + b - y$, defined for $y > 1$. $F(y)$ is increasing for $y < a^{\frac{1}{1-a}}$ and decreasing for $y > a^{\frac{1}{1-a}}$. By definition, $y > 1 > a^{\frac{1}{1-a}}$ meaning $F(y)$ is decreasing over its entire range. As $F(1) = b$ is positive and $F(.)$ is decreasing, $F(y)$ eventually becomes negative as $y > y^*$, where y^* is the point at which $F(y^*) = 0$.
Now notice that $F(x_t) = x_t^a + b - x_t = x_{t+1} - x_t$ while $F(x_t) \geq 0$ for all t by supposition. As $F(x_t)$ is always non-negative, $x_t < y^*\ \forall t$, i.e. $\{x_t\}$ is bounded above. □

Theorem A.3. *The sequence $\{x_t\}$ defined by a recurrence relation $x_{t+1} = x_t^a + b$ converges to a unique limit M if $a \in (0,1)$ and $x_0 > 1$.*

Proof. Following from Lemma A.1 & A.2, $\{x_t\}$ converges as it is monotonic and bounded. Let M be the limit of $\{x_t\}$,

$$\lim_{t\to\infty} x_t = M = \lim_{t\to\infty} x_{t+1} \Rightarrow M = M^a + b \tag{A.9}$$

The solution of (A.9) is then the limit to which $\{x_t\}$ converges. Due to the inverted U-shaped of F and $F(1) = b > 0$, $F = 0$ only has a unique positive root. In other words, M is unique. □

B Calibration exercise

B.1 Summary of sub-samples

The variables to be used for calibration include the highest level of education an individual achieved, the highest level of education either parent achieved, the amount of annual educational investment, and the annual wage an individual received. The first two variables recorded the highest qualifications obtained, divided into 8 categories from 1 - no qualification, to 8 - Professional Degree. Table B.1 and Table B.2 summarise data for the sub-samples used in the calibration.

From Table B.1, we notice a couple of features worth mentioning. First, for individuals in the low wage group, they are slightly less educated than their parents. On the contrary, individuals in the high wage group are, *on average*, more educated than their parents. In aggregate, the children's generation is more educated than the parent's generation but this trend is driven by the high-wage group. Second, an average person

Table B.1: Summary statistics (top and bottom 25% wage)

	Observations	Mean	SD	Min	Max
Wage at the 25th percentile or lower					
h	1227	3.08	1.23	1	8
$\hat{h}$	1227	3.27	1.44	1	8
I	1227	619.74	1,767.07	0	24450
L	1227	15,020.81	7,555.46	48	25000
Wage at the 75th percentile or higher					
h	1215	4.66	1.51	1	8
$\hat{h}$	1215	4.19	1.58	1	8
I	1215	2,562.51	5,271.92	0	68600
H	1215	100,736.88	47,612.31	62000	235884
Aggregate					
h	2442	3.86	1.58	1	8
$\hat{h}$	2442	3.73	1.58	1	8
I	2442	1,586.35	4,041.64	0	68600
Wage[a]	2442	57,689.32	54,742.40	48	235884
p	2442	0.50	0.50	0	1

[a] Total income wage in 2017, to be classified as either L or H.

Table B.2: Summary statistics (top and bottom 15% wage)

	Observations	Mean	SD	Min	Max
Wage at the 15^{th} percentile or lower					
h	729	3.02	1.25	1	8
$\hat{h}$	729	3.30	1.45	1	8
I	729	611.23	1,774.95	0	20905.55
L	729	9,885.97	5,268.66	48	18000
Wage at the 85^{th} percentile or higher					
h	720	4.80	1.55	1	8
$\hat{h}$	720	4.34	1.65	1	8
I	720	2,939.22	6,157.33	0	68600
H	720	122,318.36	51,726.74	80000	235884
Aggregate					
h	1449	3.91	1.66	1	8
$\hat{h}$	1449	3.81	1.64	1	8
I	1460	1,767.99	4,665.32	0	68600
Wage	1449	65,752.99	67,118.30	48	235884
p	1449	0.50	0.50	0	1

from the high wage group invests more in education than does the low wage group counterpart. Third, in both groups, there are individuals who invested an amount several times higher than the mean of the group they are in. For instance, there is an individual in the low-wage group who invested an amount in surplus of $\$24,000$, an amount almost 10 times the average of the high-wage group and close to 40 times the average of her own group. These features are similarly present in the subset of data for the top and bottom 15% wage.

B.2 Estimates of α

We run the following regression equation for the two sub-samples to get a measure of α,

$$\log(h_{i,j}) = \alpha \log(\hat{h}_i) + \beta + \theta FE_j + \epsilon_i$$

As previously mentioned, there are several assumptions implicit in this regression equation. First, as the variables h and $\hat{h}$ denote the highest qualification obtained by children and parents respectively, we are treating qualification synonymous with years of schooling. Additionally, the incremental effect between, having a Master's degree and a PhD degree for instance, is thus assumed to be similar to the incremental effect between having no qualification and having a GED. Second, we are ignoring other factors other than "human capital" that might have an effect on the educational level of an agent. These factors may encompass genetics, neighborhood effects, rearing effects, ambitions, among many other things. Lastly, we do not account for assortative mating in parents.

In our defence, we argue that isolating the pure causal effect of parents' schooling on that of a child is only secondary in our case. Instead, we are most interested in the aggregate effect of parents' schooling, which may manifest through a number of channels including assortative mating, nurture, school choice, and to a

Table B.3: OLS estimates of the intergenerational linkage of human capital

	25% top and bottom wage			15% top and bottom wage		
	(1)	(2)	(3)	(4)	(5)	(6)
$\log(\hat{h})$	0.427***	0.427***	0.409***	0.456***	0.455***	0.433***
	(0.0189)	(0.0189)	(0.0197)	(0.0245)	(0.0243)	(0.0257)
Birth year FE	No	Yes	Yes	No	Yes	Yes
Ethnicity FE	No	No	Yes	No	No	Yes
Observations	2442	2442	2442	1449	1449	1449
R^2	0.216	0.218	0.239	0.230	0.234	0.257
Adjusted R^2	0.215	0.217	0.236	0.230	0.231	0.253

lesser extent, neighborhood effects. While the problem of omitted variable bias remains, to the extent that the omitted variables are positively correlated with educational attainment of a child, we expect a positive bias. Estimates of α in Table B.3 are therefore the upper bound of the true effect of parental educational attainment on that of a child. Nonetheless, these estimates are not unreasonable and closely match those of studies that make use of years of schooling to estimate the causal effect of parent's schooling.[1]

C Cases of multiple equilibria

The four possible scenarios: (i) increasing $E[h_{i,k}]$ and $M > 2$; (ii) increasing $E[h_{i,k}]$ and $\widetilde{h} < M < 2$; (iii) decreasing $E[h_{i,k}]$ and $\widetilde{h} < M < 2$; and (iv) decreasing $E[h_{i,k}]$ and $M < \widetilde{h}$.

C.1 Increasing or decreasing $E[h_{i,k}]$ and $\widetilde{h} < M < 2$

As case (ii) and (iii) have similar results in that members of household i still invest the maximum amount in education at the infinite horizon, their results can be obtained from 3.12 as follows,

$$\begin{aligned}\lim_{k\to\infty} E[b_{i,k}] &= \lim_{k\to\infty}(\beta R)^k b_{i,0} + \left[\frac{\beta\gamma}{2}\bar{I}(H-L) + \beta L - \beta R\bar{I}\right]\lim_{k\to\infty}\sum_{j=0}^{k-1}(\beta R)^j \\ &\quad + \beta\gamma(H-L)\lim_{k\to\infty}\sum_{j=0}^{k-1}(\beta R)^{k-j}E[h_{i,j}]^\alpha \\ &= \frac{1}{1-\beta R}\left[\frac{\beta\gamma}{2}\bar{I}(H-L) + \beta L - \beta R\bar{I}\right] + \beta\gamma(H-L)\lim_{k\to\infty}\sum_{j=0}^{k-1}(\beta R)^{k-j}E[h_{i,j}]^\alpha\end{aligned} \tag{C.1}$$

C.2 Increasing $E[h_{i,k}]$ and $M > 2$

From period $\bar{k}+1$ onward, $E[h_{i,\bar{k}+1}]^\alpha + \bar{I} > 2$ implying members living in period $\bar{k}+2$ and thereafter invest an amount $2 - E[h_{i,\bar{k}+1}]^\alpha < \bar{I}$ rather than the maximum amount $\bar{I}$. Clearly, as the amount of investment now falls below $\bar{I}$, we shall wonder whether it is still in the agent's best interest to invest in education. In

[1]See for instance Behrman and Rosenzweig (2002); Plug (2004) and Antonovics and Goldberger (2005).

this case, an agent invests if and only if $U(0) < U(2 - E[h_{i,\bar{k}+1+n}]^{\alpha})$ for $n \in \mathbb{N}^{+}$, which is equivalent to

$$e^{-\lambda L}\Big(1 - e^{\lambda(2-h^{\alpha}_{i,\bar{k}+1+n})R}\Big) + \frac{\gamma}{2}(e^{-\lambda L} - e^{-\lambda H})\Big(e^{\lambda(2-h^{\alpha}_{i,\bar{k}+1+n})R}(2 + h^{\alpha}_{i,\bar{k}+1+n}) - 2h^{\alpha}_{i,\bar{k}+1+n}\Big) > 0 \tag{C.2}$$

Suppose (C.2) is satisfied, at time $\bar{k}+1+n$, the level of human capital of a member of household i is

$$E[h_{i,\bar{k}+1+n}] = \frac{1}{2}(2 + E[h_{i,\bar{k}+n}]^{\alpha}) = \frac{1}{2}E[h_{i,\bar{k}+n}]^{\alpha} + 1 \tag{C.3}$$

The expected bequest at time $\bar{k}+1+n$ is,

$$\begin{aligned} E[b_{i,\bar{k}+1+n}] &= \beta\Big[\big(E[b_{i,\bar{k}+n}] - 2 + E[h_{i,\bar{k}+n}]^{\alpha}\big)R + \frac{\gamma}{2}(H-L)(2 + E[h_{i,\bar{k}+n}]^{\alpha}) + L\Big] \\ &= (\beta R)^{n} E[b_{i,\bar{k}+1}] + \beta L\sum_{j=0}^{n-1}(\beta R)^{j} + 2\beta\Big[\frac{\gamma}{2}(H-L) - R\Big]\sum_{j=0}^{n-1}(\beta R)^{j} \\ &\quad + \beta\Big[\frac{\gamma}{2}(H-L) + R\Big]\sum_{j=0}^{n-1}(\beta R)^{n-1-j}E[h_{i,\bar{k}+1+j}]^{\alpha} \end{aligned} \tag{C.4}$$

$$\begin{aligned} \Rightarrow \lim_{n\to\infty} E[b_{1,\bar{k}+1+n}] &= \frac{\beta}{1-\beta R}L + \beta\Big[\frac{\gamma}{2}(H-L) + R\Big]\lim_{n\to\infty}\sum_{j=0}^{n-1}(\beta R)^{n-1-j}E[h_{1,\bar{k}+1+j}]^{\alpha} \\ &\quad + \frac{\beta}{1-\beta R}\Big[\gamma(H-L) - 2R\Big] \end{aligned} \tag{C.5}$$

C.3 Decreasing $E[h_{i,k}]$ and $M < \widetilde{h}$

This scenario yields the same results as the case of no investment in education. Recall that $\bar{k}$ is the last period where $E[h_{i,\bar{k}}] > \widetilde{h}$, which implicates $E[h_{i,\bar{k}+1}] < \widetilde{h}$. As the expected level of human capital falls below the threshold level, members living in period $\bar{k}+2$ and onward stop investing in education due to the U-shaped utility and henceforth traverse the same evolution path as the case where the initial human capital is below the threshold. From equation (3.9) and (3.10), the initial bequest and human capital are inconsequential to the steady-state bequest and human capital. By the same token, the amount of bequests and level of human capital of a member of household i before $\bar{k}+2$ plays no role in the determination of equilibrium bequest and human capital.

C.4 A comparison

Having shown all the possible scenarios, it is now of interest to compare two households 1 and 2 that differ only in their initial human capital. Without loss of generality, let $h_{1,0} > h_{2,0}$ and $b_{1,0} = b_{2,0}$. It is readily evident that if both households invest at the infinite horizon, members of household 1 always inherit a higher level of human capital and bequest than do members of household 2 when they belong to different categories. By contrast, if both households do not invest at the infinite horizon, including the case of decreasing $E[h_{i,k}]$, their human capital and bequest are identical in the long run. In addition, if the two households belong to the same category, they converge to the same limit yielding a pooling equilibrium.

Consider cases where household 1 invests in education while household 2 does not at the infinite horizon. Specifically, we are interested in comparing household 1 with decreasing $E[h_{i,k}]$ and $\widetilde{h} < M < 2$ with household 2 that never invests in education as their results can be generalized to other cases. Subtracting

(C.1) by equation (3.10) and cancelling similar terms yield,

$$\frac{\beta \bar{I}}{2(1-\beta R)}[\gamma(H-L)-2R]+\beta\gamma(H-L)\lim_{k\to\infty}\sum_{j=0}^{k}(\beta R)^{k-j}(E[h_{1,j}]^{\alpha}-h_{2,0}^{\alpha}) \tag{C.6}$$

The second term of (C.6) is always positive whereas the sign of the first term hinges on the difference between the returns to education and the returns to capital. Consequently, for sufficiently high γ or $H-L$, (C.6) is positive leading to a separating equilibrium. By contrast, another separating equilibrium emerges when γ or $H-L$ are not high enough to offset the foregone income from capital investment. The former reflects the story the paper is trying to narrate. To elaborate, suppose two households are originally below the poverty line. In addition, if an agent from either household gets an offer to work the good job, she can escape poverty but this does not necessarily mean future members of the same household are also able to escape poverty. A member of household 1 invests in education and increases her chance of getting the offer. While the expected human capital of the next member falls, which effectively reduces the probability that the agent gets an offer, members of household 1 still find it in their best interest to invest in education to maintain the probability. They are, thus, non-poor *on average*. A member of household 2, on the other hand, chooses to invest in the capital market as this option yields a higher expected utility for her. However, as periodic optimization, in this case, lowers the probability of getting the offer in the future, future members of household 2 suffer as they remain below the poverty line *on average*. Clearly, a bad draw for household 1 and good draw for household 2 would reverse their position in a single period but as long as there is no change in the investment decision, household 1 is expected to cross the poverty line while household 2 remains below it.

C.5 Discussion

We have shown that below a certain threshold of human capital inheritance, an agent ceases investment in education in favour of investment in the capital market. This threshold is not fixed but depends on other parameters of the model. Formally, it is the level of human capital equating the expected utility from investing in education and the utility obtained in case of no such investment,

$$\widetilde{h} \equiv \left(\frac{e^{-\lambda L}}{\gamma(e^{-\lambda L}-e^{-\lambda H})}-\frac{\bar{I}e^{\lambda \bar{I}R}}{2(e^{\lambda \bar{I}R}-1)}\right)^{1/\alpha} \tag{C.7}$$

(C.7) is derived under the assumption that $\widetilde{h}^{\alpha}+\bar{I}\leq 2$. If this assumption does not hold, the one derived above is the *shadow threshold*, as agent cannot invest an amount $\bar{I}$ at $\widetilde{h}$, whereas the actual threshold is higher on account of a smaller amount of investment made. It is the level of $\hat{h}$ satisfying

$$-e^{-\lambda L}\left(e^{\lambda(2-\hat{h}^{\alpha})R}-1\right)+\frac{\gamma}{2}\left(e^{-\lambda L}-e^{-\lambda H}\right)\left(e^{\lambda(2-\hat{h}^{\alpha})R}(2+\hat{h}^{\alpha})-2\hat{h}^{\alpha}\right)=0 \tag{C.8}$$

As $\widetilde{h}$ is pivotal in the current discussion of underinvestment as well as persistent poverty, it is useful to mention how the threshold level of human capital reacts to changes in other existing parameters. With the exception of λ and $\bar{I}$, whose changes bring about ambiguous effects on $\widetilde{h}$, the direction of changes of $\widetilde{h}$ in response to a change in other parameters is self-evident in the corresponding first-derivatives. *Ceteris paribus*, the threshold level increases with R and L while decreases with γ and H. This pattern is in alignment with our expectation that an agent will require a higher threshold level as the returns to non-educational investment increase, whereas higher returns to education clearly lower $\widetilde{h}$.

This additional text has shown that the underinvestment in education by itself does not, *a priori*, constitute a poverty trap. Rather, certain conditions apply for the divergence in the evolution path to emerge. We arrive at this conclusion in an environment absent any kind of frictions prevalent in the social mobility literature including fixed cost with credit constraints (Galor and Zeira, 1993; Maoz and Moav, 1999), non-homotheticity (Moav, 2002; Banerjee and Mullainathan, 2010; Ghatak, 2015), among many others. In our model, a sufficiently high return to education and the indivisibility of outcome are necessary for the existence of a separating equilibrium. Whether the return is high enough to the point of being "disproportionate" is more of an empirical issue.

D Generalisation to CRRA utility

This section is dedicated to the case when an agent has a CRRA utility function, which arguably provides a better description of how one invests. Unlike the CARA case, expected utility is concave in the amount of investment, and thus, the optimal investment needs not occur at either corners. However, for a given level of bequest, children of poorly educated parents do not invest while children of sufficiently educated parents invest the maximum amount. Due to the nature of CRRA utility, however, the discussion cannot be held in the absence of wealth although it needs not be. Nonetheless, we do not shift the focus away from inherited human capital as a major, albeit non-exclusive, source of investment decision. Instead, we examine whether it is feasible to make up for a lower initial bequest with a higher inherited human capital for a given investment, and if it is, to what extent. To facilitate the study of bequest-human capital trade-off, we provide a comparison between households with different combinations of bequests and human capital inheritance in a multi-period setting.

D.1 Educational investment in a single period

Assume an agent has the following CRRA utility,

$$u(x) = \begin{cases} \dfrac{x^{1-\sigma}-1}{1-\sigma} & \sigma \neq 1 \\ \log(x) & \sigma = 1 \end{cases} \tag{D.1}$$

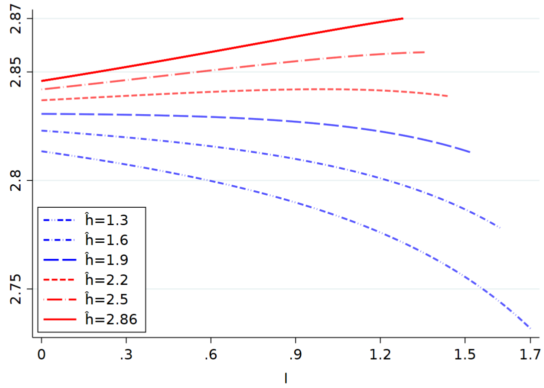
The expected utility that corresponds to equation (3.2) is,

$$U_{i,t} = p(h)\frac{\big(H+(b_{i,t-1}-I_{i,t})R\big)^{1-\sigma}-1}{1-\sigma} + (1-p(h))\frac{\big(L+(b_{i,t-1}-I_{i,t})R\big)^{1-\sigma}-1}{1-\sigma} \tag{D.2}$$

where $p(h) = \frac{\gamma}{2}(2h_{i,t-1}^{\alpha} + I_{i,t})$. Figure D.1 draws the expected utility and Figure D.2 draws the expected bequest and investment against the amount of parental human capital.

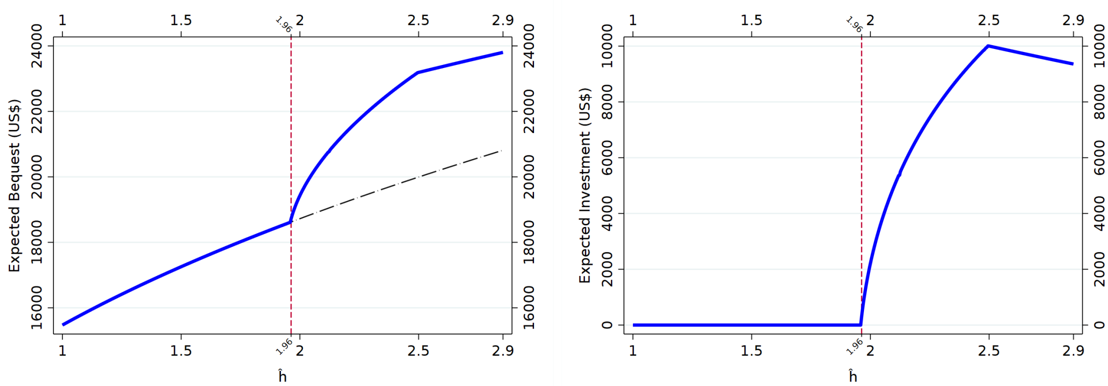
For sufficiently high $\hat{h}$ or $\hat{h} < \widetilde{h}$, the optimal investment still occurs at either corners. For $\hat{h}$ that is neither too low nor too high, e.g., the lowest red line in Figure D.1, the optimal investment is an interior solution. The ensuing concave utility means that bequest and investment are continuous in the amount of parental human capital. We do observe, however, a significant leap in the expected bequest from Figure D.2 as agents start investing in education. The thick blue line describes the expected bequest if agents invest in accordance with the investing rules, i.e. invest in education if $\hat{h} > \widetilde{h}$, whereas the black dash-dot line

Figure D.1: **Expected Utility for different level of $\hat{h}$**



Calibrated parameters: $H = 16.73$, $L = 1.35$, $\sigma = 1$, $\alpha = 0.43$, $\gamma = \underline{p} = 0.28$, $\bar{p} = 0.79$, $R = 1.05$, $\bar{I} = 1.86$, $h \in [1, 2.86]$, $b_0 = 1.4$. Threshold $\widetilde{h} \simeq 1.963$.

Figure D.2: **Expected Bequest and Investment for different level of $\hat{h}$**



Modified parameters: $H = \$122,318$; $L = \$9,886$; $b_0 = \$10,234$.

shows the expected bequest if agents only invest in the capital market. For this particular configuration, the capital market allows agents to lend their idle capital rather than borrowing for educational investment.

D.2 Educational investment in a multi-period setting

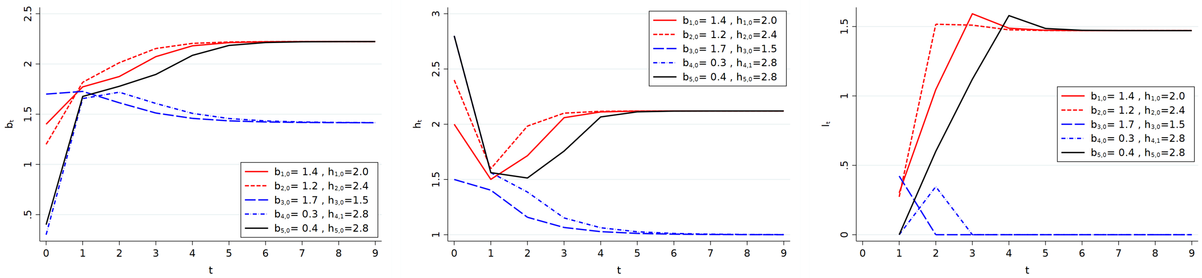
We start first by proving that human capital and bequest are substitutes. The first derivative with respect to $h_{i,t-1}$ and the cross partial-derivative $U_{h_{i,t-1}b_{i,t-1}}$ are,

$$\frac{\partial U_{i,t}}{\partial h_{i,t-1}} = \alpha\gamma h_{i,t-1}^{\alpha-1} \frac{\big(H + (b_{i,t-1} - I_{i,t})R\big)^{1-\sigma} - \big(L + (b_{i,t-1} - I_{i,t})R\big)^{1-\sigma}}{1-\sigma} \tag{D.3}$$

$$\frac{\partial^2 U_{i,t}}{\partial h_{i,t-1}\,\partial b_{i,t-1}} = \alpha\gamma h_{i,t-1}^{\alpha-1} R\Big[\big(H + (b_{i,t-1} - I_{i,t})R\big)^{-\sigma} - \big(L + (b_{i,t-1} - I_{i,t})R\big)^{-\sigma}\Big] \tag{D.4}$$

Since $\sigma > 0$ for risk-averse agents, the cross partial-derivative is negative as human capital and bequest are substitutes in determining the optimal investment. We now compare households with different combinations

Figure D.3: **The intergenerational evolution of $I_{i,t}$, $h_{i,t}$ and $b_{i,t}$**



Calibrated parameters: $H = 16.73$, $L = 1.35$, $\sigma = 1$, $\alpha = 0.43$, $h_{i,t} \in [1, 2.86]$, $R = 1.05$, $\beta = 0.2$.

of initial bequests and human capital. It should be kept in mind that a higher bequest lowers the threshold level $\widetilde{h}$, allowing wealthier yet less educated household to invest in education.[2] Figure D.3 illustrates the evolution of human capital, bequest and investment of 5 different households over 9 generations.

Overall, depending on whether a household invests at the end of the studied period, there are two equilibria, i.e. the high equilibrium for investing households, and the low equilibrium for non-investing households. We take household 1, which is average in both human and non-human capital as the baseline for comparison. Even though household 2, which has higher initial human capital yet a lower bequest, is similar to household 1 in that they both invest by the 9^{th} period, notice that household 2 converges to the high equilibrium faster than does household 1. Household 3, on the other hand, starts off relatively richer yet gradually converges to the low equilibrium because of the low initial human capital. By contrast, household 4, which is highly educated in the beginning, ends up uneducated as the low initial bequest bars members of this household from investing a sufficient amount in education. Lastly, the 5^{th} household is in a situation identical to household 4 at the beginning, albeit slightly wealthier. This meager increase in bequest, however, has a long-lasting impact as it allows future members to invest in education a sufficient amount to sustain the human capital inheritance above the threshold.

This simple calibration highlights the importance of both inputs in the determination of long-run wealth. It is evident that, *in our model*, a typical poor household, which has both a low bequest and low human capital inheritance, is doomed to poverty. Additionally, our model implies that a lump-sump transfer to poor households is not an effective way to reduce poverty if the amount is not sufficient to encourage future investment. On the contrary, subsidising education is also lacking if the poor cannot become sufficiently educated to sustain their educational investment. Consequently, the most effective approach is to do both, education susidisation and transfer. The higher level of education can the government subsidise, the lower is the transfer required. Additionally, our results suggest that the transfer needs not be conditional upon educational investment because at sufficiently high levels of parental education and household wealth, an agent will invest because it is in her best interest to do so. To conclude, it is of essence for government to monitor the elasticity of substitution between the two inputs to formulate a policy that is most effective in poverty alleviation.

[2] To see this, take the first derivative of $U_{i,t}$ with respect to $b_{i,t}$.

References

Antonovics, K. L. and Goldberger, A. S. (2005). Does increasing women's schooling raise the schooling of the next generation? comment. *American Economic Review*, 95(5):1738–1744.

Banerjee, A. and Mullainathan, S. (2010). The shape of temptation: Implications for the economic lives of the poor. Technical report, National Bureau of Economic Research.

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

Galor, O. and Zeira, J. (1993). Income distribution and macroeconomics. *The review of economic studies*, 60(1):35–52.

Ghatak, M. (2015). Theories of poverty traps and anti-poverty policies. *The World Bank Economic Review*, 29(1):S77–S105.

Maoz, Y. D. and Moav, O. (1999). Intergenerational mobility and the process of development. *The Economic Journal*, 109(458):677–697.

Moav, O. (2002). Income distribution and macroeconomics: the persistence of inequality in a convex technology framework. *Economics Letters*, 75(2):187–192.

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