Intergenerational Altruism and Transfers of Time and Money: A Lifecycle Perspective*

Uta Bolt,† Eric French,† Jamie Hentall Maccuish† and Cormac O'Dea‡

January 21, 2020

Abstract

Parental investments in children can take one of three broad forms: (1) Time investments during childhood and adolescence that aid child development (2) Educational investments (3) Cash transfers in the form of inter-vivos gifts and bequests. Using panel data that covers a cohort of individuals from birth to retirement, we estimate a dynastic model of household decision-making with intergenerational altruism that nests a multi-period child production function, incorporates all three of these types of investments, and allows us to quantify the relative importance of those investments in explaining intergenerational persistence in outcomes. Preliminary estimates show that the model matches the intergenerational persistence in earnings observed in our data. The key factors driving this intergenerational persistence in income are: (i) we find strong complementarity between ability and education in the wage equation, meaning that the returns to education are only high for high ability individuals, (ii) because of borrowing constraints, high income households are more likely to send their kids to college, (iii) because of points (i) and (ii) high income households have higher incentive to invest in childrens' ability, generating the observed positive correlation between parents' education and time spent with children. Time investments contribute most of the three channels we consider to intergenerational propagation of inequality.

*Preliminary, please do not quote. We are very grateful to Peter Arcidiacono, Judith Delaney, Andrew Hood, George Levi-Gayle, Limor Golan, Ananth Seshadri, and multiple seminar participants for helpful discussions. We are very grateful to Jack Light, Hoon Pyo Jeon and Alan Liu for excellent research assistance. Funding from the Social Security Administration through the Michigan Retirement Research Center (MRRC grant UM17-15) and the Economic and Social Research Council (Centre for Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies (RES-544-28-50001) and Grant Inequality and the insurance value of transfers across the life cycle (ES/P001831/1)) for this work is gratefully acknowledged. O'Dea additionally acknowledges funding from the ESRC Secondary Data Analysis grant ref: ES/N011872/. The opinions and conclusions expressed are solely those of the authors and do not represent the opinions or policy of SSA or any agency of the Federal Government. Any errors are the authors' own.

†University College London and Institute for Fiscal Studies

‡Yale University and Institute for Fiscal Studies

1 Introduction

Intergenerational links are a key determinant of levels of inequality. Previous work looking at a range of developed economies finds very significant intergenerational correlations in education, incomes and wealth (e.g. **?**, **?**, **?**, **?**). The literature on understanding the mechanisms behind this persistence is much newer. This paper estimates a dynastic model of the decision-making of altruistic households to investigate the quantitative importance of three distinct mechanisms in generating intergenerational persistence in outcomes. Those mechanisms, each known to be important in linking outcomes across generations, are i) parental time investments during childhood and adolescence that aid child development (**?**, **?**) ii) parental aid for education (**?**, **?**) and iii) cash gifts in the form of inter-vivos transfers and bequests (**?**, **?**).

We use data from the National Child Development Survey, which is an ongoing panel containing the entire population of Britain born in a particular week in 1958. These data allow us to measure parental inputs and child ability throughout childhood and contains information on educational outcomes and earnings over the lifecycle. We use these data to estimate a child ability production function, applying the methods developed by **?** and **?** and leveraging the fact that we have multiple measures of parental inputs and child outcomes. We embed our estimated ability production function into a dynastic model in which ability and education generate productivity in the labor market and in which altruistic parents can give combinations of time, educational investments, and cash transfers to their children, while also making their own consumption and labor supply decisions.

Our model contains five distinct mechanisms which can generate persistence in outcomes across generations. The first three mechanisms generate a positive correlation between the earnings of an individual and the earnings of their parents. The first is the borrowing constraint, which limits families from accessing credit to provide higher education to their children. The second is that we allow parental productivity in investing in children to vary with productivity in the labor market. The estimated relationship is positive, which implies that the time investments more educated parents make in their children are more productive than those made by parents with less education. Third, while we find only modest complementarity between early childhood (0 to 7 years) and mid to late childhood (7 to 11 and 11 to 16 years) time investments, the complementarity between ability and education is much larger. [FACT]. This channel amplifies the effect of the first two channels.

The fourth channel – positive assortative matching – generates persistence in household earnings over and above that observed between parents and their children. The final mechanism – cash transfers from parents to children – allows for a persistence in income and consumption over and above that seen for earnings. To the best of our knowledge, this is the first paper to include all of the above channels.

The estimated model implies an intergenerational elasticity of earnings of 0.34, close to estimates for our cohort of interest in **?**. The model also replicates the fact (documented by **?** and observed in our data) that parents with more education spend more time with their children.

We have three principle findings. First, as noted above, we find modest dynamic complementarity between early time investments in children and later time investments. However, we find substantial complementarities between terminal childhood ability (measured at age 16) and education in wages. Among men those with college education, an increase in the standard deviation in this measure of ability leads to an additional 19% in wages. Among those with the lowest level of education – this premium, at 9%, is much smaller. As a result, high ability individuals are more likely to select into education than their low ability counterparts. This dynamic complementarity, in combination with self selection into education, is a key mechanism that perpetuates income inequality across generations. High income households, who have more resources to send their child to college, have higher returns to investing in their child's ability than their low income counterparts; thus they invest more in their children. Second, while we find that each of the three types of transfer contributes to persistence in outcomes, the quantitatively most important is parental time with children. [FACT]. Finally, using our model to evaluate the responsiveness of parental behavior to the policy environment we find that...

This paper relates to a number of different strands of the existing literature, including work measuring the drivers of inequality and intergenerational correlations in economic outcomes, the large literature seeking to understand child production functions and work on parental altruism and bequest motives. The most closely related papers, however, are those focused on the costs of and returns to parental investments in children. The three papers closest to ours are **?**, **?** and **?**. Each of those papers, like ours, contains a dynastic model in which parents can give time, education and money to their children. All three papers find that early life investments are key for understanding the intergenerational correlation of income. We build on the contributions of these papers in three ways. The first is that those papers lack data that links investments at young ages to earnings at older ages. As a result, they have to calibrate key parts of the model, while we are able to estimate the human capital production technology using recently-developed methods, and show how early life investments and the resulting human capital impacts later life earnings. The second is that we model explicitly the behavior of both men and women. This allows us to show the quantitatively important role that assortative matching plays in amplifying the role of parental transfers in generating persistence in outcomes at the household level. Finally, the focus of our paper is different. **?** focus on identifying the role of market imperfections in rationalizing observed levels of parental investments. The aim of **?** is to simultaneously rationalize intergenerational persistence in outcomes and cross-sectional inequality in outcomes. **?** focuses on the macroeconomic effects of large-scale policy interventions. Our primary focus, facilitated by our data on of each of the three parental inputs for our cohort of interest, is to quantitatively evaluate the role played by each.

Other closely related papers include **?** and **?**, both of which develop models in which parents choose how much time to allocate to the labor market, leisure and investment in children. Neither paper, however, incorporates household savings decisions, and hence the trade-off between time investments in children now and cash investments later in life. **?** focuses on the interaction between parental investments, state subsidies and education decisions, but abstract from the role of parents in influencing ability prior to the age of 16. **?** and **?** build overlapping-generations models of wealth inequality that includes both intergenerational correlation in human capital and bequests, but neither attempts to model the processes underpinning the correlation in earnings across generations.

The rest of this paper proceeds as follows. Section 2 describes the data, and documents descriptive statistics on ability, education and parental investments. Section 3 lays out the dynastic model used in the paper. Section 4 outlines our estimation approach while, Section 5 then provides some reduced-form evidence on the impact of parental investments, before Section 7 provides some results on the relative importance of different channels in explaining intergenerational correlations in education, earnings and welfare. Section 8 concludes, and draws out some implications for policy.

2 Data and Descriptive Statistics

The key data source for this paper is the National Child Development Study (NCDS). The NCDS follows the lives of all people born in Britain in one particular week of March 1958. The initial survey at birth has been followed by subsequent follow-up surveys at the ages of 7, 11, 16, 23, 33, 42, 46, 50 and 55.[1] During childhood, the data includes information on a number of ability measures, measures of parental time investments (discussed in more detail below) and parental income. Later waves of the study record educational outcomes, demographic characteristics, earnings and hours of work. For the descriptive analysis in this section, we focus on those individuals for whom we observe both their father's educational attainment (age left school) and their own educational qualifications by the age of 33. This leaves us with a sample of 9,436 individuals.

The main limitation of the NCDS data currently available for our purposes is that we do not have data on the inheritances received or expected by members of the cohort of interest. We therefore supplement the NCDS data using the English Longitudinal Study of Ageing (ELSA). This is a biennial survey of a representative sample of the 50-plus population in England, similar in form and purpose to the Health and Retirement Study (HRS) in the US. The 2012-13 wave of ELSA recorded lifetime histories of inheritance

[1] The age-46 survey is not used in any of the subsequent analysis as it was a telephone interview only, and the data are known to be of lower quality.

receipt which we can use to augment our description of the divergence in lifetime economic outcomes by parental background. We focus on individuals in ELSA born in the 1950s, leaving us with a sample of 3,001.[2]

Lastly, while the NCDS contains a rich set of qualitative measures on the quantity and quality of time spent with children, it does not contain information on the precise hours spent with children, we use the UK Time Use Survey (UKTUS) to map our NCDS investment measure into actual time. We describe these datasets in greater detail in Appendix C.

In the rest of this section, we document the evolution of inequalities over the lifecycle, and in particular how they relate to parental background and parental investments over time.

2.1 Ability and Time Investments

Table 1 shows the multiple measures of both ability and investments available in our data at ages 0, 7, 11, and 16. Having multiple measures is a key advantage of our data as each of the measures likely has measurement error, and we can a dynamic latent factor model (developed **?**), described below, to extract the separate signal from noise. Before doing that, however, we highlight some of the key features available in the raw data. We show how several measures of both investment and ability correlate with father's education to give an initial description of the intergenerational persistence in outcomes.

Figure 1 shows the cumulative distribution of normalized reading and math ability at each age, splitting the sample according to father's education. Throughout the paper we use low, medium and high to describe education groups – these correspond to having only compulsory levels of education, having some post-compulsory education and having some college respectively.[3] In the US context this would correspond roughly to high school dropout, high school graduate, and some college. Figure 1 shows that, as one might expect, children whose father has a higher level of education have higher ability; at the age of 7, 20% (18%) of the children of low-education fathers had a reading (math) score around one standard deviation or more below the mean, compared to just 4% (5%) of the children of high-education fathers. Similarly, 39% of the children of high-education fathers had a maths score around one standard deviation or more above the mean, compared to 18% of the children of low-education fathers.[4]

The second key thing to note from Figure 1 is that ability gaps by father's education widen through childhood. At the age of 7, 61% of the children of low-education fathers have above-average reading scores compared to 89% of the children of high-education parents - a gap of 28 percentage points. By age 11,

[2]The next wave of the NCDS, which will be in the field next year, is currently planned to collect information on lifetime inheritance receipt. We hope to use these new data in later versions of this work

[3]For this age group of fathers, compulsory education roughly corresponds to leaving school at age 14, post-compulsory means leaving school between ages 15 and 18, and some college means staying at school until at least age 19.

[4]For reading, a large number of students obtained a perfect score.

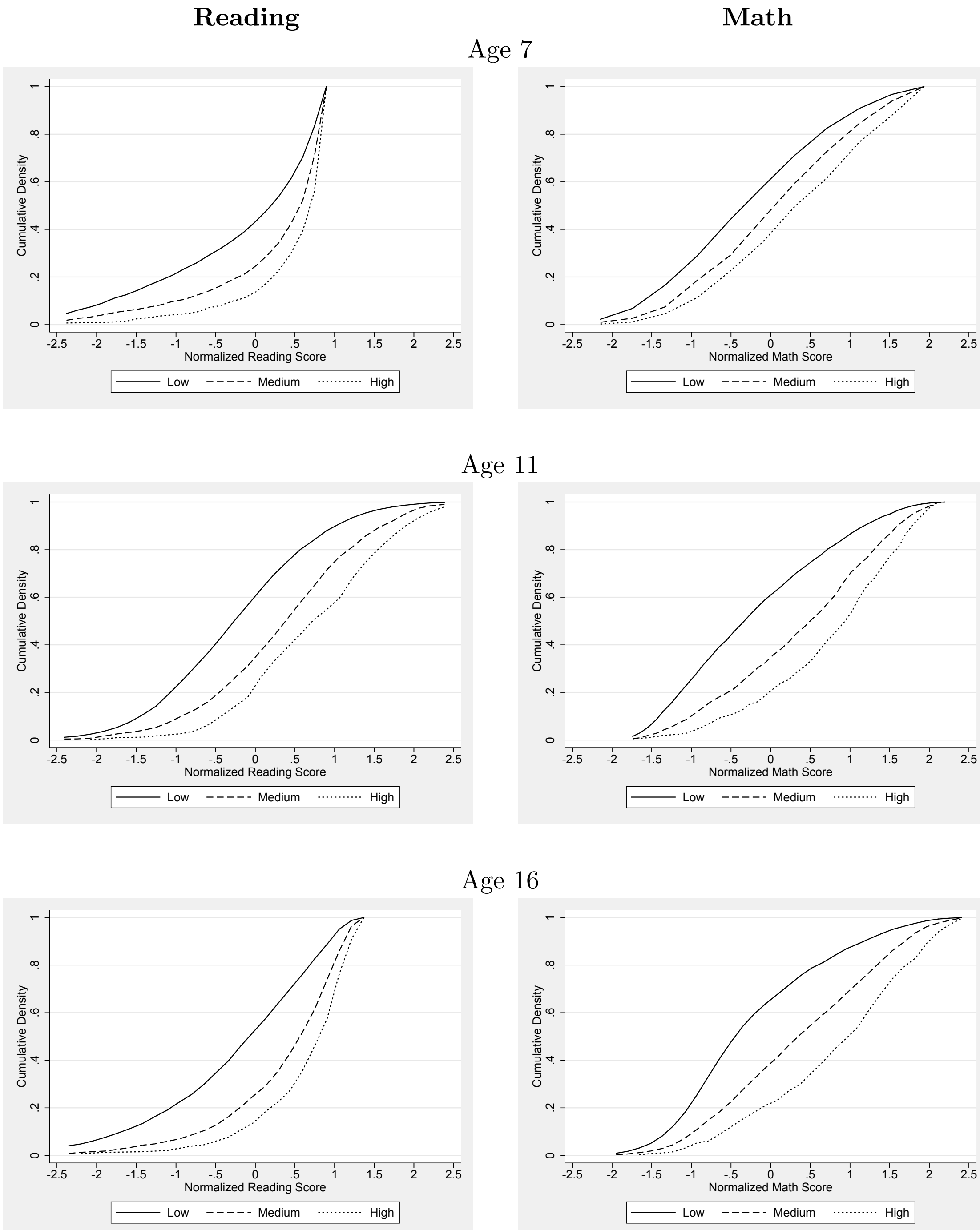
Table 1: List of all measures used

Ability measures	*Investment measures*
Age 0:	
Birthweight	Teacher's assessment of parents' interest in education (**mother** and father)
Gestation	Outings with child (mother and father)
	Read to child (mother and father)
	Father's involvement in upbringing
	Parental involvement in child's schooling
Age 7:	
Reading score	Teacher's assessment of parents' interest in education (**mother** and father)
Math score	Outings with child (mother and father)
Drawing score	Father's involvement in upbringing
Copying design score	Parents' ambitions regarding child's educational attainment (further educ & university)
	Parental involvement in child's schooling
	Library membership of parents
Age 11:	
Reading score	Teacher's assessment of parents' interest in education (**mother** and father)
Math score	Involvement of parents in child's schooling
Copying design score	Parents' ambitions regarding child's educational attainment
Age 16:	
Reading score	
Math score	

All investment measures are retrospective, so age 0 investments are measured at age 7, age 7 investments are measured at age 11, age 11 investments are measured at age 16.

that gap has widened to 37 percentage points, and by age 16 it stands at 38 percentage points. These gaps widen for math also.

Figure 1: Reading and maths skill, by parental education



Tables 2, 3 and 4 provide some descriptive evidence that at least some of the widening in ability gaps by parental characteristics between ages 7 and 16 age can be explained by differential parental investments (a hypothesis we test more formally using our model). Table 2 documents parental responses to a question about reading with their child, asked when the child is 7. It shows differences in the frequency with which both mothers and fathers read to their children, splitting families according to the education of the father.

For example, 34% of fathers with low education read to their 7-year-old children each week, compared to 53% of fathers with high education. The relevant figures for mothers are 47% and 67% respectively.

Tables 3 and 4 present the child's teachers assessment of parental interest in the child's education, at the ages of 7 and 11 respectively. The differences by father's educational attainment are perhaps even more striking than those in reading patterns. When the child is 7, high educated fathers are three times more likely to be judged by the teacher to be 'very interested' in their child's education as low education fathers (65% compared to 22%). While mothers are assessed as having greater interest in their child's education in all groups, there are large difference according to education group (76% in the highest education group are very interested, compared to 35% in the lowest education group). At the age of 11, the gap in paternal interest is very similar. The tables also show that having a higher-educated father dramatically reduces the risk of a child having parents with little interest in their education. Among those with a high educated father, only around 10% have a mother or father who is judged to show 'little interest' in their education at the age of 11. On the other hand, among those whose father is low educated, that figure rises to around a quarter of mothers and nearly half of fathers.

Table 2: Frequency with which parents read to age-7 children

		Father		
		Never	Sometimes	Each week
Father's education	Low	30%	36%	34%
	Middle	20%	35%	45%
	High	18%	29%	53%
		Mother		
		Never	Sometimes	Each week
Father's education	Low	16%	37%	47%
	Middle	12%	31%	57%
	High	10%	23%	67%

Table 3: Teacher assessment of parental interest in education of age-7 child

		Father		
		Little interest	Some interest	Very interested
Father's education	low	55%	24%	22%
	middle	34%	22%	44%
	high	20%	15%	65%
		Mother		
		Little interest	Some interest	Very interested
Father's education	low	23%	43%	35%
	middle	10%	30%	60%
	high	6%	18%	76%

Table 4: Teacher assessment of parental interest in education of age-11 child

		Father		
		Little interest	Some interest	Very interested
Father's education	low	46%	29%	25%
	middle	21%	25%	54%
	high	12%	16%	72%
		Mother		
		Little interest	Some interest	Very interested
Father's education	low	26%	38%	35%
	middle	12%	27%	61%
	high	8%	16%	76%

2.2 Educational Attainment

Table 5 shows that there is a substantial inter-generational correlation in educational attainment between fathers and their children. First, having a high-educated father makes it much less likely that a child will end up with low education. 30% of the children of fathers with low education end up with low education, compared to only 10% of those with middle educated fathers, and just 2% of high educated fathers. Second, having a high-educated father makes it much more likely that a child will end up with high education. Fully 66% of the children of high educated fathers also end up with high education, compared to only 20% of those whose fathers have low education.

Educational investments differ from inter-vivos transfers and bequests in terms of timing, but also more importantly in that they directly impact on children's earnings. In addition, educational attainment affects who the individual partners with. It is these mechanisms by which high education parents transfer resources to their children that we estimate and model below.

Table 5: Intergenerational correlation in education

	Child's education		
Father's education	low	middle	high
low	30%	50%	20%
middle	10%	47%	43%
high	2%	32%	66%

2.3 Inter-vivos Transfers and Bequests

Table 6 documents the receipt of inter-vivos transfers and bequests of the NCDS cohort so far, again splitting by father's education. As explained at the start of this section, the top panel draws on the NCDS data itself, while the bottom panel uses ELSA data instead, as information on inheritance receipt is not yet available in the NCDS.

The table shows that inter-vivos transfers are a significant source of economic resources for young adults, and that as one would expect are much more significant for those with higher-educated parents. By the age of 33, 55% of those with a high education father had received an inter-vivos transfer, of an average of around £50,000. While this is the mean of a highly right-skewed distribution, these figures indicate an important role for inter-vivos transfers relieving borrowing constraints in this part of the lifecycle. At the same age, 24% of those with low-educated fathers had received an inter-vivos transfer, of an average size of just under £25,000.

Evidence from ELSA data suggests that differences in inheritance receipt by parental background are also significant. 46% of those with high educated fathers have received an inheritance, compared to 26% of those with low-educated fathers, and among those who have received an inheritance, those with high educated fathers have received around twice as much on average (£120,843 compared to £66,545) . The net result is that those with high educated fathers have inherited around £40,000 more than those with low-educated fathers. This is likely to understate the true difference in mean lifetime inheritance receipt between these groups; some of those born in the 1950s will still have living parents, and differential mortality means it is in fact likely that this applies to a larger share of those with high-educated fathers.

3 Model

This section describes a dynastic model of consumption and labor supply in which parents can make different types of transfers to their children. The model can be used to a) evaluate how particular intergenerational transfers affect the outcomes of household members, b) compare the relative insurance

Table 6: Receipt of inter-vivos transfers and bequests by father's education

		Inter-vivos transfers		
		Mean (\pounds)	Received	Mean exc. zeros (\pounds)
	Low	1921.763	6%	30,639
Father's education	Medium	7949.932	10%	77,909
	High	9575.253	20%	49,073
		Inheritances (1950s birth-cohort)		
		Mean (\pounds)	Received	Mean exc. zeros (\pounds)
	Low	27,394	36%	75,648
Father's education	Medium	71,189	58%	122,395
	High	93,553	54%	174,349

Notes: ELSA life-time inter-vivos transfers and inheritances in £2014. Trimmed at the top 1%. Only includes individuals whose parents have both died by the time of the interview.

value of these types of transfers and c) simulate household behavior and welfare under counterfactual policies (for example, under reforms to estate taxation).

Figure 2 provides an overview of the dynastic model. During childhood, parental time investments in children and educational choices affect the evolution of the child's ability and their educational attainment. Children are then matched in couples, possibly receive transfers of cash from their parents and begin adult life. They then have their own children, and alongside the standard choices of consumption and labour supply they choose how much to invest in their own children, with implications for their children's future outcomes.

Our NCDS data respondents were surveyed every four to seven years from the age of 0 and 55. To be consistent with the data, each of our model periods will cover the time between interviews (and each period will be of different lengths). A model timeline illustrating this is shown in Figure 2 (while a more comprehensive Appendix Table 17 summarizes the rest of the detail that will be laid out in this section).

Each individual has a lifecycle of 20 model periods which can be broken into four phases.

1. Childhood has periods $t = 1, 2, 3, 4$ which corresponds to ages 0-6, 7-10, 11-15, 16-22. During childhood the individual accumulates human capital and education, remains attached to their parents and does not make active decisions.

2. Independence consists of one period at $t = 5$ at ages 23-25. The individual (potentially) receives a parental cash transfer, is matched into a couple and begins working life.

3. Parenthood has five periods at ages $t = 6, 7, 8, 9, 10$, corresponding to ages 26-32, 33-36, 37-41,42-48, 49-54. Identical twin children are born at the start of the 'Parenthood' phase and now additionally

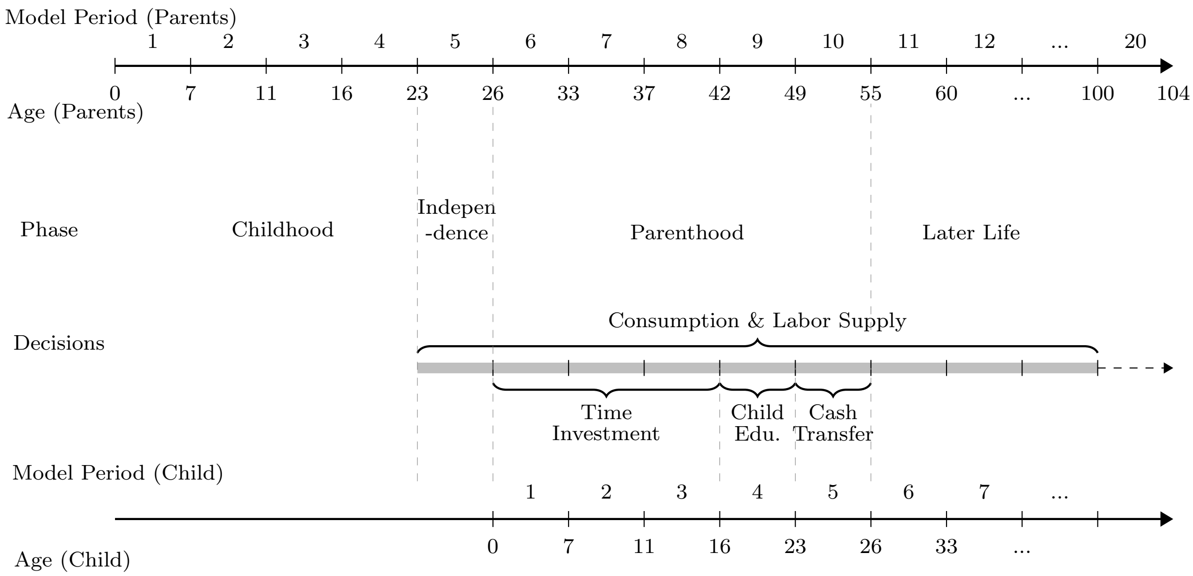


Figure 2: The Life Cycle of an Individual

to making their own participation saving decision the parents decide how much to invest in their childrens' human capital and education. At the end of this period they have an opportunity to transfer wealth to the children.

4. Late adult phase consists of 10 regularly space periods from 55-59, ..., 100-104. The household is separated from their children and makes decision about their own saving and consumption.

In outlining the dynastic model we describe below a lifecycle decision problem of a single generation. All generations are, of course, linked; each member of the couple whose decision problem we specify has parents, and they, in turn, will have children. The index t will denote the age (in model periods) of the generation whose problem we are laying out, we use a prime to denote their childrens' variables: for example, in the model period when adults are aged t, their children are aged t'.[5]

We now provide formal details of the model.

3.1 Preferences

The utility of each member of the couple $g \in \{m, f\}$ (male and female respectively) depends on their consumption ($c_{g,t}$) and leisure ($l_{g,t}$):

$$u_g(c_{g,t}, l_{g,t}) = \frac{(c_{g,t}^{\nu_g} l_{g,t}^{(1-\nu_g)})^{1-\gamma}}{1-\gamma}$$

[5]Children are born five model periods after their parents, therefore they are aged $t = 1$ in model periods when the parent is model-aged $t = 6$.

We allow the relative preferences for consumption and leisure to vary with gender. Household preferences are given by the equally-weighted sum of male and female utility, multiplied by a factor n_t which represents the number of equivalized adults in a household in time t (scaled so that for a childless couple $n_t = 1$).

$$u(c_{m,t}, c_{f,t}, l_{m,t}, l_{f,t}, n_t) = n_t \Big(u_m(c_{m,t}, l_{m,t}) + u_f(c_{f,t}, l_{f,t}) \Big)$$

Total household consumption is split between children, who receive a fraction $\frac{n_t - 1}{n_t}$, and adults who get a share $\frac{1}{n_t}$. The latter quantity is efficiently allocated between spouses.

In discounting their future utility each generation applies an annual discount factor (β). We assume that β represents an annual discount factor. Model period length aligns with the differences in time between interviews and so the discount factor between model period varies. Thus β_{t+1} is the discount rate between t and $t+1$ and is equal to β^{τ_t} where τ is length of model period t.[6]

Each generation is altruistic regarding the utility of their offspring (and future generations). In addition to the time discounting of their children's future utility (which they discount at the same rate they discount their own future utility), they additionally discount it with an intergenerational altruism parameter (λ).

3.2 Initial Conditions and Parental Cash Transfers Received

Individuals begin their independent life at age 23, and are matched into couples at this age. Individuals differ at the start of life in their ability, their level of education and their initial wealth (which comes from a parental cash transfer). We describe the determinants of ability, education, and wealth in the sections below.

3.3 Demographics

There is probabilistic matching between men and women which is based on education, where education takes one of three values: low, middle and high. The probability that a man of education ed_m gets married to a women with education ed_f is given by $Q_m(ed_m, ed_f)$. The (symmetric) matching probabilities for females are $Q_f(ed_f, ed_m)$. Everyone is matched into couples – there are no singles in the model. The draw of spousal ability and initial wealth is therefore drawn from a distribution that depends on one's own education.

At age 26, a pair of identical twins is born to the couple. In order to match the average fertility for this sample, which is close to two, yet still maintain computational tractability, we follow **?** and assume that the twins face with identical sequences of shocks.

[6] In addition, utility in each period is weighted by the number of years between periods.

Mortality is stochastic - the probability of survival of a couple (we assume that both members of a couple die in the same year) to age $t+1$ conditional on survival to age t is given by s_{t+1}. We assume that death is not possible until the household enters the late adult phase of life at the age of 50 and that death occurs by the age of 110 at the latest.

3.4 Constraints and Income Sources

Constraints Parents face three constraints – an intertemporal budget constraint, a liquidity constraint, and an intratemporal time constraint. The first is at the household level:

$$a_{t+1} = (1+r_t)(a_t + y_t - (c_{m,t} + c_{f,t}) - x_t) \tag{1}$$

where a_t is parental wealth, y_t is household income and x_t is a cash transfer to children that can only be made when the parent is 49 and the child is 23 (and $x_t = 0$ otherwise). The gross interest rate $(1+r_t)$ is equal to $(1+r)^{\tau_t}$ where r is an annual interest rate and τ_t is the length in years of model period t.

The second constraint is a liquidity constraint. In our empirical work we match explicitly total inter-vivos transfers received over the life cycle. In the model we assume that all transfers received are at age 23. In practice, however, many transfers are after this age. To account for this, we assume that some of the transfers from parents to children are placed in a trust, and only a share of these assets π_t are liquid in period t. However, all earned income in period t is fully liquid. Thus we have the borrowing constraint

$$(c_{m,t} + c_{f,t}) + x_t \leq \pi_t a_t + y_t \iff a_{t+1} \geq (1+r_t)(1-\pi_t)a_t \tag{2}$$

π_t is estimated outside the model to match the share of total wealth represented by received transfer at that age.

The third constraint is a per-parent ($g \in \{m, f\}$) intratemporal time budget constraint:

$$T = l_{g,t} + ti_{g,t} + h_{g,t} \tag{3}$$

where T is a time endowment, $ti_{g,t}$ is time investments in children, and $h_{g,t}$ is work hours, and $l_{g,t}$ is leisure time.

Earnings and household income Household income is given by

$$y_t = \begin{cases} \tau(e_{m,t}, e_{f,t}, e'_{t'}, t) & \text{if the children are age 16} \\ \tau(e_{m,t}, e_{f,t}, t) & \text{otherwise} \end{cases} . \tag{4}$$

where $\tau(.)$ is a function which returns net-of-tax income, and $e_{m,t}$ and $e_{f,t}$ are male and female earnings respectively. In their last period before the 'Independence' phase of life (age 16), children can participate in the labor market if they are no longer in education (discussed below). Their parents are still the decision-maker in this period and any income they earn ($e'_{t'}$) is part of household resources in that period.

Earnings are equal to hours worked (h) multiplied by the wage rate, for example: $e_{f,t} = h_{f,t}w_{f,t}$. That wage rate evolves according to a process that has a deterministic component which varies with age, a term which depends on ability and a stochastic (AR(1)) component.

$$\begin{aligned}
\ln\ w_t &= \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 \ln ab + \delta_5 PT_t + v_t \\
v_t &= \rho v_{t-1} + \eta_t \\
\eta_t &\sim N(0, \sigma^2)
\end{aligned}$$

where PT_t is a dummy for working part time. While the associated subscripts are suppressed here, each of $\{\delta_0, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \rho, \sigma^2\}$ varies by gender (g) and education (ed).

3.5 Ability and Education of Children

This section describes the production function for ability and education from birth to age 23. Over this part of the life cycle, parental investments of time received by their children do not directly impact the contemporaneous utility of the child, but leads (in expectation) to the children having higher wages, more able spouses and more able childrens' children later in life, all of which matters to the altruistic parent.

3.5.1 Child Ability Production Function

A child's initial ability at birth $ab'_{1'}$ is given by:

$$ab'_{1'} = f_{ab_1}(ed_m, ed_f, u'_{ab,1'}) \tag{5}$$

which depends on his/her parents' level of education, ed_m and ed_f, and $u'_{ab',1}$, which is a stochastic variable that generates heterogeneity in initial ability, conditional on parental education.

Between birth and age 16, child ability updates each period according to the restricted translog production function:

$$\ln ab'_{t'+1} = \gamma_{1,t'} \ln ab'_{t'} + \gamma_{2,t'} \ln ti_{t'} + \gamma_{3,t'} \ln ti_t \cdot \ln ab_{t'} + \gamma_{4,t'} ed^m + \gamma_{5,t'} ed^f + u'_{ab,t'} \tag{6}$$

The rate of growth of a child's ability depends on his/her parents' level of education and the sum of the time investments ($ti_t = ti_{m,t} + ti_{f,t}$) those parents make. There is also a stochastic component to the ability transition equation ($u'_{ab,t+1}$). Ability evolves until the period 4 (age of 16), after which it does not change. ab without a subscript denotes final ability.

We include parental education to impact ability to capture the idea that high skill individuals who are productive in the labor market may also be productive at producing skills in their children. This is a mechanism that features prominently in several recent studies of the labor market (e.g., **?**).

3.5.2 Education

When the child is age 16 the parent chooses the educational level of the child. There was compulsory education to age 16 for our sample members and additional years of education were free of tuition at this point in time. Thus we model the decision to send the child to school until age 18 or 22. Because tuition was 0 for this cohort, the key cost of education at this time was the cost of forgone labor income from the child. We model the potential income if the child works, and thus the loss of household income if the child receives additional years of education as in equation (4).

3.6 Decision Problem

3.6.1 Decision Problem in the Parenthood Phase

Choices Households make decisions on behalf of both the adults and children within the household.

These are (with the time periods in which those decisions are taken given in parentheses):

1. Consumption of each parent– $c_{m,t}$ and $c_{f,t}$ where m and f index consumption by the male and female respectively (each period).

2. Hours of work of each parent – $h_{m,t}$, $h_{f,t}$ We allow each parent to work full-time, part-time or not at all (each period).

3. Time investments in children of each parent – $ti_{m,t}$ and $ti_{f,t}$ (up to and including the age at which their child turns 11)

4. Childrens' education ed' (in the period the children turn 16)

Uncertainty In this phase couples face uncertainty over the innovation to each of their wages and the innovations to the child ability production function. The joint distribution of these stochastic variables ($\mathbf{q}_t \equiv \{\eta_{m,t}, \eta_{f,t}, u_{ab,t}\}$) is given by $F_t(\mathbf{q}_t)$.

State variables The vector of state variables ($\mathbf{X}_t$) during parenthood (where we suppress time subscripts) is $\mathbf{X}_t = \{t, a, w_m, w_f, ed_m, ed_f, ab_m, ab_f, g', ab'\}$ where t is age, w_m, w_f are the wages of each of both parents, ab_m, ab_f are their abilities, g' is the child's gender and ab' is the child's ability.

Value function The value function for the Parenthood phase is given below in expression (7):

$$
\begin{aligned}
V_t(\mathbf{X}_t) \quad = \quad & \max_{c_{m,t}, c_{f,t}, h_{m,t}, h_{f,t}, ti_{m,t}, ti_{f,t}, ed'_g} \left\{ u(c_{m,t}, c_{f,t}, l_{m,t}, l_{f,t}, n_t) + \beta_{t+1} \int V_{t+1}(\mathbf{X}_{t+1}) dF_{t+1}(\mathbf{q}_{t+1}) \right\} \\
s.t. \quad & \text{i) the intertemporal budget constraint in equation (1)} \\
& \text{ii) the borrowing constraint in equation (2)} \\
& \text{iii) and the time budget constraints in equation (3)} \\
& \text{and education of the child is a choice when the child is 16, but not otherwise.}
\end{aligned}
$$

3.6.2 Decision Problem in the Independence Phase

The final period in which a couple is making decisions on behalf of their dependent child is when they are 49 (and their child is 23).

Choices During this phase couples make three sets of choices:

1. Household consumption – $c_{mt}, c_{f,t}$
2. Hours of work for each parent – $h_{m,t}$, $h_{f,t}$ where m and f index hours of work by the male and female respectively
3. A cash gift (x_t) to their children made when the parent is age 49 and the child is 23.

Uncertainty Couples face two distinct types of uncertainty. The first is uncertainty over the characteristics of their children as they start adulthood. The dimensions of uncertainty here are the childrens' initial wage draw and the attributes of their future spouse (his/her ability, education level, assets, and initial wage draw). The stochastic variables are collected in a vector $\mathbf{p}'_{t'}$, and their joint distribution is given by $H()$. These are realized after the parent makes their decision.

The second dimension of uncertainty is with respect to their own circumstances next year – that is their next period wage draws ($\mathbf{q}_{t+1} \equiv \{\eta_{m,t+1}, \eta_{f,t+1}\}$ with distribution given by $F_{t+1}(\mathbf{q}_{t+1})$).

State variables The set of state variables in this phase is that same as in the early phase of adulthood plus childrens' education (ed').

Value function The decision problem in the Independence phase where age of parent is 49 and age of child is 23 ($t = 10$ and $t' = 5$) is:

$$V_t(\mathbf{X}_t) = \max_{c_{m,t}, c_{f,t}, h_{m,t}, h_{f,t}, x_t} \Big\{ u(c_{m,t}, c_{f,t}, l_{m,t}, l_{f,t}, n_t) \quad (7)$$

$$+\lambda \int V'_{t'}(\mathbf{X}'_{t'}) dH(\mathbf{p}'_{t'}) + \beta_{t+1} \int V_{t+1}(\mathbf{X}_{t+1}) dF_{t+1}(\mathbf{q}_{t+1}) \Big\} \quad (8)$$

$s.t.$ i) the intertemporal budget constraint in equation (1)

ii) the borrowing constraint in equation (2)

iii) and the time budget constraint in equation (3)

Note that there are two continuation value functions here. The first is the expected value of the couple to which the (soon to be independent) child of the parent will belong to. The (altruistic) parents take this into account in making their decisions. This continuation utility is discounted by the altruism parameter (λ) and the integration is with respect to the children's intial wage draw and the characteristics of their spouse (these shocks are realized after the couple makes their decisions). The second continuation value function is the future expected utility that the parents will enjoy in the next period (when they will enter the late adult phase). The value function (given in equation (9)) must be integrated with respect to next period's wage draws, which are stochastic, and discounted by β_{t+1}, the time discount factor.

3.6.3 Decision Problem in the Late Adult phase

At this stage the children of generation 1 have entered their own early adult phase and the generation 1 couple enters a 'late adult phase',

Choices During this phase households make labor supply and consumption/saving decisions only.

Uncertainty There is uncertainty over their next period wage draws ($\mathbf{q}_{t+1} \equiv \{\eta_{t+1,f}, \eta_{t+1,m}\}$ with distribution given by $F_{t+1}(\mathbf{q}_{t+1})$) and there is now stochastic mortality (where assume that both members of the couple die in the same period).

State variables The vector state variables ($\mathbf{X}_t$) during the late adult phase of life is $\mathbf{X}_t = \{t, a, w_m, w_f, ed_m, ed_f, ab_m, ab_f\}$ The ability and education of the (now-grown-up) child are no longer state variables.

Value function The decision problem in the 'late adult' phase of life can be expressed as:

$$
\begin{aligned}
V_t(\mathbf{X}_t) \quad &= \max_{c_{m,t},c_{f,t},h_{m,t},h_{f,t}} \left\{ u(c_{m,t}, c_{f,t}, l_{m,t}, l_{f,t}) + \beta_{t+1} s_{t+1} \int V_{t+1}(\mathbf{X}_{t+1}) dF_{t+1}(\mathbf{q}_{t+1}) \right\} \qquad (9) \\
s.t. \quad & \text{the intertemporal budget constraint in equation (1)} \\
& \text{and the time budget constraint in equation (3)}
\end{aligned}
$$

where s_{t+1} is the probability of surviving to period $t+1$, conditional on having survived to period t. All cash transfers *received* by the couple are now liquid ($\pi_t = 1$), and so the couple of not subject to the liquidity constraint in the late adult phase of life.

4 Estimation

We adopt a two-step strategy to estimate the model. In the first step, we estimate or calibrate those parameters that, given our assumptions, can be cleanly identified outside our model. In particular, we estimate the distribution of ability at age 0, the human capital production function, the wage process, marital sorting process, mortality rates, and the process for the share of assets that are liquid π_t from the data. In addition, we also estimate the initial conditions (of the joint distribution of education, ability, gender, and parental transfers received at age 23) directly from the data. We calibrate the interest rate, parameters of the tax code (taken from IFS TAXBEN), and household equivalence scale parameter.

In the second step, we estimate the rest of the model's parameters (discount factor, consumption weight for both husband and wife, risk aversion, and altruism parameter)

$$
\Delta = (\beta, \nu_f, \nu_m, \gamma, \lambda)
$$

with the method of simulated moments (MSM), taking as given the parameters that were estimated in the first step. In particular, we find the parameter values that allow simulated life-cycle decision profiles to "best match" (as measured by a GMM criterion function) the profiles from the data.

Because our underlying motivations are to explain sources of income and parental investments in children, we match employment choices for both husbands and wives and also household time spent with children, by parents' age and education. Because we wish to understand study money as well as time transfers to children, we also match educational decisions (and thus the forgone income from not working), as well as cash transfers to children when the children are older. Finally, to understand how households value their own utility in the present versus future, we match wealth data, which should be informative of the discount factor.

In particular, the moment conditions that comprise our estimator are given by

1. Employment rates, by age, gender, and education, from the NCDS data

2. Fraction in full time work conditional on being employed, by age, gender, and education, from the NCDS data

3. Mean annual time spent with children, by child's age and parent's gender and education, from the UKTUS data

4. The shares in each education group, by gender of child and parental education level, from the NCDS data

5. Mean lifetime receipt of inter-vivos transfers, by education and gender of recipient, from ELSA

6. Tertiles of wealth at 33 and 50, by education of the husband, where wealth at age 33 is from the NCDS and wealth at 50 is from ELSA

We observe hours and investment choices of individuals in the NCDS, and thus match data for these individuals for the following years: 1981, 1991, 2000, 2008, and 2013: when they were 23, 33, 42, 50 and 55. See table

1) For transfers, what shall we do? [Eric: I think we calculate the PDV of inter-vivos transfers received, then we will assume bequests are thrown into the sea (or the government takes the transfers or potentially we randomly or equally distribute the bequests), given that we never figured out what to do with the bequests.]

The mechanics of our MSM approach are as follows. We compute life-cycle histories for a large number of artificial households, each composed of a man and woman. Each member of these households is endowed with a value of the age-23 ability and wages drawn from the empirical distribution from the NCDS data, and wealth which is drawn from ELSA.

We discretize the asset and also the ability and wage grids for both spouses and, using value function iteration, we solve the model numerically. This yields a set of decision rules, which, in combination with the simulated endowments and shocks, allows us to simulate each individual's assets, work hours and home investment hours, child's educational choices, and inter-vivos transfers. We use the resulting profiles to construct moment conditions, and evaluate the match using our GMM criterion. We search over the parameter space for the values that minimize the criterion. Appendix L contains a detailed description of our moment conditions, the weighting matrix in our GMM criterion function, and the asymptotic distribution of our parameter estimates.

4.1 Estimating the Model of Parental Investments and Human Capital Accumulation

In our data we cannot directly observe children's skills (ab_t) or parents investment (I). However we have multiple noisy measures of each. In our analysis we explicitly account for this measurement error, using an approach where we an efficient implementation of the methods in **?** to account for these multiple measures. We show how to use multiple measures (as in **?**, **?**) but using a simpler system GMM approach rather than maximum likelihood and filtering methods.

Following AW, we use a restricted translog production function (6) which we repeat here for convenience:

$$\ln ab'_{t'+1} = \gamma_{1,t'} \ln ab'_{t'} + \gamma_{2,t'} \ln ti_{t'} + \gamma_{3,t'} \ln ti_t \cdot \ln ab_{t'} + \gamma_{4,t'} ed^m + \gamma_{5,t'} ed^f + u'_{ab,t'}$$

We assume independence of measurement errors and use the noisy measures to instrument for one another. An important extension, relative to AW, is that we use many possible combinations of input measures to instrument for one another.

We use the same methodology to estimate the function

$$\ln ti_t = \alpha_{1,t} \ln ab'_{t'} + \alpha_{2,t} ed^f + \alpha_{3,t} ed^m + \alpha_{4,t} \ln y_t + u_{ti,t} \tag{10}$$

This equation is not derived from the structural model, but is infomative of how changes in ability, parental education, and income impact time investments, addressing measurement error. Thus we have our model structural parameters match the estimated α parameters in equation 10 using an indirect inference procedure. See the appendix for details on the estimation and identification of the parameters in this section.

The NCDS contains multiple measures of investments at different ages, but does not measure the actual time spent. As parents choose investments in terms of time in the model, we convert the latent time investment index to time using UKTUS data. For this, we assign the percentiles of latent investments the amount of time related to the same percentile of time investments in the UKTUS. See the appendix for further details.

4.2 Estimating the Wage Equation, Accounting for Measurement Error in Ability and Wages and Selection

We estimate the wage equation laid out in Section 3:

$$\ln w_t^* = \ln w_t + u_t = \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 lnab + \delta_5 PT_t + v_t + u_t \text{ where} \tag{11}$$

$$v_t = \rho v_{t-1} + \eta_t,$$

$$u_t \text{ is IID measurement error in wages}$$

for each gender and education group. We estimate the wage equation parameters in two steps.

In the first step we estimate the δ parameters, accounting for measurement error in $lnab$ using the measurement error framework of AW. In the second step we estimate the parameters of the wage shocks $\rho, Var(\eta)$, and the variance of wages upon entry into the labor market$Var(v_5)$ using a standard error components model, accounting for measurement errror in both $lnab$ and u_t.

The above procedure addresses problems of measurement error in ability and wages but not selection. We control for selection bias by finding the wage profile that, when fed into our model, generates the same estimated profile (i.e., the same δ parameters from equation (I.1)) that we estimated in the data. Because the simulated profiles are computed using only the wages of those simulated agents that work, the profiles should be biased for the same reasons they are in the data. We find this bias-adjusted wage profile using the iterative procedure described in French (2005).

See appendix H for details.

4.3 Imputing Time Spent with Children

The NCDS data includes high quality information on measures of parental investments in children, but does not include the model consistent measure of actual hours spent with children. In order to construct a measure of actual hours spent with children, we first use our latent factor model to construct an index of time spent with children. To convert the index into hours, we use time data from the UKTUS survey for children aged 5-9, summing up investment hours of husband and wife. We calculate rank percentiles of our time index in the NCDS and and the rank percentiles of hours in the time use data. For example, a household is predicted to be at the 75th percentile of the time investments in children distribution in the NCDS data, then to that household we assign the hours with children observed at the 75th percentile of hours in the time use data. We use this method to convert model generated time investment in children into the latent time investment measure in equation (6) to generate ability realizations for the children in the model.

5 First Step Estimation Results

In this section we describe results from our first-step estimation, that we use as inputs for our structural model, and the outputs that we require our model to match. These first step inputs describe the determinants of investments in children, how those investments affect childrens' ability, and how ability impacts success in the labor market. In particular, we present estimates of how children's ability, as well as parental resources, affect investments in children. We also present estimates of the effect of parental time investments on children's ability, and how that ability in turn affects subsequent education and adult earnings. This exploits a key advantage of our data - that we measure for the same individuals their parents' investments, their ability and the value of that ability in the labour market.

5.1 The Determinants of Parental Investments in Children

In Section 2, we documented that higher-educated parents spend more time time reading to their children and show more interest in their educational progress. Here we exploit this variation to create a measure of the time investments of parents in children using the methods described briefly in Section 4.1 and in more detail in the appendix. Throughout in estimation we use a GMM estimator with a diagonal weighting matrix.

We estimate equation (10) for the periods when the children are between ages 0-6, 7-10, 11-16 with parents of the corresponding ages 26-32,33-36,37-41. Time investment in the data is measured at the end of the period (e.g., age 7 for when the child was age 0-6). Our time index is measured in logs, and thus we can interpret our coefficients as being measured as elasticities.

For children between ages 0-6, we find that a 1% increase in parents' income increases the index of time spent with the age 0 child by .11%, and a one year increase in mother's and father's education increases time spent with the child by .039% and .034%, respectively.

The effect of child's ability and parents' income on investments grows with age. We find that parents reinforce children's skills - a 1% increase in child ability at age 7 is associated with a 0.35% increase in age 7 investments. Parental education again has a positive effect on age 7 investments. The effect of income on investments becomes stronger such that a 1% increase in income leads to a 0.16% increase in investments.

These estimates grow further at age 11: a 1% increase in child ability at age 11 is associated with a 0.68% increase in age 11 investments. The effect of income on investments becomes stronger such that a 1% increase in income leads to a 0.26% increase in age 11 investments. Interestingly, father's education has a larger impact than mother's education on our age 11 time index.

These results are robust to also including a number of other covariates into the equation, such as parental age and number children in the household.

Table 7: Determinants of time investments.

	Coeff	90% CI
Investment function age 0-6		
log ability	0.000	[-0.015,0.019]
log parental income	0.078	[0.057,0.112]
mum: medium ed	0.086	[0.060,0.135]
mum: high ed	0.194	[0.119,0.281]
dad: medium ed	0.113	[0.081,0.154]
dad: high ed	0.154	[0.113,0.233]
Investment function age 7-10		
log ability	0.363	[0.314,0.401]
log parental income	0.125	[0.096,0.176]
mum: medium ed	0.260	[0.210,0.326]
mum: high ed	0.263	[0.203,0.322]
dad: medium ed	0.276	[0.181,0.377]
dad: high ed	0.365	[0.278,0.466]
Investment function age 11-15		
log ability	0.589	[0.536,0.635]
log parental income	0.174	[0.123,0.220]
mum: medium ed	0.223	[0.152,0.287]
mum: high ed	0.302	[0.145,0.429]
dad: medium ed	0.240	[0.163,0.315]
dad: high ed	0.333	[0.211,0.439]

Notes: GMM estimates. Confidence intervals are bootstrapped using 100 replications. Parent's income is log after tax annual labor income, measured at age 16. For the investment equation in period 1 (age 0-6), we use ability measured at age 0, and investments at 7. For the investment equation in period 2 (ages 7-10), we use ability measured at age 7, and investments at 11. For the investment equation in period 3 (ages 11-15), we use ability measured at age 11, and investments at 16.

5.2 The Determinants of Ability

In Section 2 we documented that children of high educated parents do better in cognitive tests, and that the ability gaps between children of high and low educated parents grow over time. Here we combine the multiple test scores to create a measure of skills, and estimate a human capital production function using the methods described briefly in Section 4.1 and in more detail in the appendix. Similar to our approach to estimating the determinants of time investment, we use a GMM estimator with a diagonal weighting matrix.

We allow for the fact that high education parents may have high ability children. We assume that parental education determines the initial draw of the child ability (i.e. at birth). See appendix F for more on this.

We estimate equation (6) for ability at ages 7, 11, and 16. The time investments entering the equation are those corresponding to ages 0-6, 7-10, and 11-16, respectively (when the parents are aged 26-32,33-36, and 37-41).

Our ability index is measured in logs, and thus we can interpret our coefficients as being measured as elasticities.

Table 8: Determinants of log ability.

	Coeff	90% CI
Production function age 7		
log ability	0.107	[0.037,0.206]
log investment	0.205	[0.194,0.296]
ability x inv	-0.039	[-0.129,0.003]
mum: medium ed	0.303	[0.251,0.374]
mum: high ed	0.403	[0.262,0.485]
dad: medium ed	0.314	[0.226,0.426]
dad: high ed	0.265	[0.204,0.333]
Production function age 11		
log ability	0.861	[0.819,0.986]
log investment	0.117	[0.091,0.152]
ability x inv	0.079	[0.059,0.113]
mum: medium ed	0.146	[0.057,0.173]
mum: high ed	0.325	[0.212,0.456]
dad: medium ed	0.202	[0.115,0.245]
dad: high ed	0.357	[0.255,0.429]
Production function age 16		
log ability	0.945	[0.908,0.980]
log investment	0.111	[0.082,0.140]
ability x inv	-0.041	[-0.071,-0.009]
mum: medium ed	0.021	[-0.014,0.054]
mum: high ed	-0.066	[-0.185,0.047]
dad: medium ed	0.042	[0.001,0.072]
dad: high ed	0.080	[-0.017,0.153]

Notes: GMM estimates. Confidence intervals are bootstrapped using 100 replications. For the production function at age 7, we use log ability measured at age 7 as a function of log ability at age 0, log time investments measured at age 7 (and referring to investments at age 0-6). For the production function at age 11, we use log ability measured at age 11 as a function of log ability at age 7, log time investments measured at age 11 (and referring to investments at age 7-10). For the production function at age 16, we use log ability measured at age 16 as a function of log ability at age 11, log time investments measured at age 16 (and referring to investments at age 11-15).

We estimate the relationship between age 7 ability as a function of age 0 ability, age 0 time investments, the interaction of ability and time investments, and mother's and father's education. Estimates are presented in Table 8. It shows that time investments have a significant effect on changes in ability over time, even after conditioning on background characteristics and initial ability. A one percent increase in time investments at age 0 raises age-7 ability by 0.225 percent, a one percent increase in time investments at age 7 raises age-11 ability by 0.09 percent, and a one percent increase in time investments at age 11 raises age-16 ability by 0.13 percent.

Ability is very persistent, especially at older ages.

Interestingly, the interaction between ability and investments is negative for age 7, but positive for age 11. This implies that whilst at young ages, investments are more productive for low-skilled children, at older ages, productivity is higher for the higher-skilled ones. The positive and statistically significant coefficients on the age 11 interactions terms indicates that the ability production function does in fact exhibit dynamic complementarity at this stage of childhood (as found by **?**).

We find that parental skill, as measured by parental education, is strongly impacts future ability, providing empirical support for a key mechanism for perpetuating inequality across generations. High skill parents are effective in producing human capital in their children (as also shown in some of the papers cited in **?** and is assumed in **?** and **?**) in addition to having more resources to afford college.

These results are robust to also including a number of other covariates into the equation, such as parental age and number children in the household.

5.3 The Effect of Ability and Education on Wages

In the dynastic model with intergenerational altruism laid out in Section 3 parents do not receive any direct return from their children having higher ability at the age of 23. Instead, they include their children's expected lifetime utility in their own value function, with a weight determined by the intergenerational altruism parameter λ. Hence parental investments in children's ability (both through time and money investments in education) will be driven by the return to ability in the labour market. Here we focus on the return to ability in the labour market, as measured by its impact on wages. We estimate the wage equation (I.1) laid out in Section 3 and H for each gender and education group.

Ability impacts wages through its relationship with education, but it also has a direct impact on wages conditional on education. This is shown by Table **??**, which plots the estimates of δ_4 for each gender and education group. The interpretation of these coefficients is that they are estimates of the log-point increase in wages associated with a log-point increase in age-16 ability, conditional on education. The top panel does not correct for selection, whereas the bottom panel does. The extent of complementary similar to that estimated in **?**, and is implicit in much of the literature on match quality in educational choice (e.g., **?**), although this finding has received less attention in the literature on estimation of ability production functions.

Table 9: Log-point change in wages for a 1 log-point increase in ability, by education level

	Male	Female
	No selection correction	
Low	0.113 (0.025)	0.137 (0.030)
Middle	0.194 (0.020)	0.134 (0.020)
High	0.259 (0.028)	0.179 (0.029)
	With selection correction	
Low	0.113X	0.137X
Middle	0.194X	0.134X
High	0.259X	0.179X

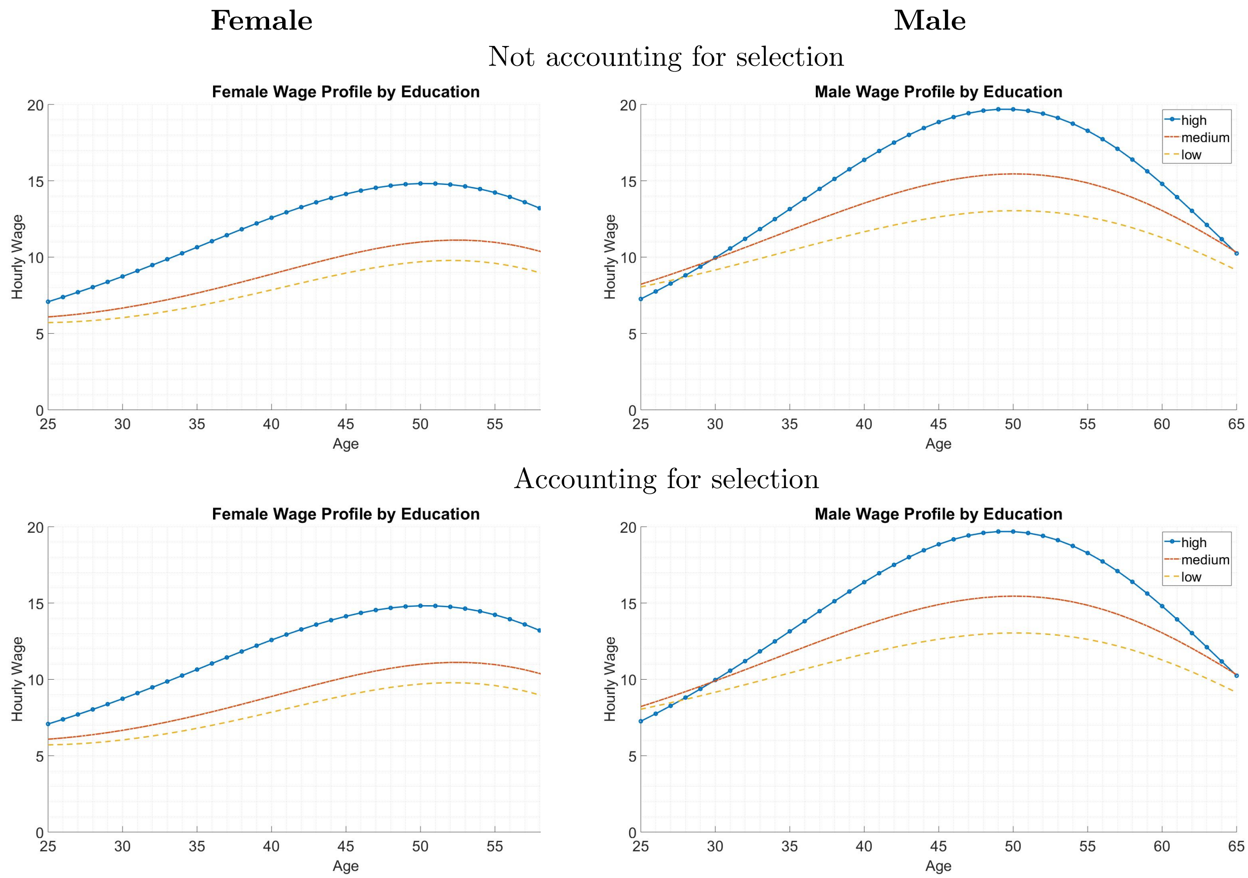
Note: Cluster bootstrapped standard errors in parentheses (500 repetitions). The variance of log ability is .57.

The table shows that, as one would expect, age-16 ability has a significant positive impact on wages conditional on education for all groups. Perhaps most interestingly, it shows evidence of complementarity between education and ability in the labour market, particularly for men. While low education men see only a 0.113 log-point increase in hourly wages for every additional log-point of ability, high education men (with some college education) see an average increase of 0.259 log-points in hourly wages for every additional log-point of ability. To give a sense of the magnitude of the estimated impact of ability on wages, the variance of log ability is 0.57. Thus increasing ability for a high education male by one standard deviation increases his wages by $\sqrt{.57} \times .259 = 20\%$.

As we show below this dynamic complementarity between ability and education gives rise to self selection. Those with high ability tend to select into high education. Furthermore, because of forward looking behavior, households more likely to invest in the education of their child also invest more time in producing high ability children.

Figure 3 shows wage profiles by age, education and gender for full time workers with average ability. The top panel shows estimates not accounting for selection. The bottom panel accounts for selection. Men and those with high education have higher wages and faster wage growth.

Figure 3: Wages, by age, education and gender



Note: Wages measured in 2014 £.

Finally, table 10 shows the persistence and variance of innovations to wages. The results indicate that $\rho = 0.937 - 0.999$ depending on the group; wages are almost a random walk. The estimate of σ^2_η =0.005-0.021; one standard deviation of an innovation in the wage is 7-14% of wages, depending on the group. These estimates are similar to other papers in the literature (e.g. **?**, **?**) and imply that long run forecast errors may be large. Furthermore, we find evidence that the variance of wage innovations is increasing with education.

Interestingly, however, we estimate $\sigma^2_{\nu_5}$ to be small for all groups. While there is significant cross sectional variation in wages, even early early in life, we estimate that most of that variation is explainable by our latent ability measure and measurement error in wages.

Table 10: Persistence and variance of innovations to wages, by education level

Parameter	Description	Low	Middle	High
			Men	
ρ	autocorrelation of wage innovation	0.964	0.984	0.978
		(0.015)	(0.007)	(0.010)
σ_η^2	innovation variance of wages	0.0065	0.0079	0.0103
		(0.0024)	(0.0017)	(0.0020)
			Women	
ρ	autocorrelation of wage innovation	0.960	0.942	0.943
		(0.0021)	(0.008)	(0.014)
σ_η^2	innovation variance of wages	0.0061	0.0142	0.0230
		(0.0036)	(0.0027)	(0.0062)

5.4 Marital Matching Probabilities

Table **??** shows the distribution of marriages, conditional on education, that we observe in the NCDS data. It also shows the share of men and women in each educational group. An important incentive for education is that it increases the probability of marrying another high education, high wage person. Table **??** shows evidence of assortive mating, as shown by the high share of all matches that are along the diagonal on the table: 12% of all marriages are between couples who are both low educated, 38% are between those who are middle educated and 4% among those who are highly educated.

Table 11: Marital matching probabilities, by education

	Low education male	Medium education male	High education male	Share of females in education group
Low education female	0.12	0.19	0.02	0.33
Medium education female	0.13	0.38	0.05	0.56
High education female	0.01	0.07	0.04	0.12
Share of men in education group	0.26	0.64	0.11	

Note: The numbers represent cell proportions, which are the percentage of all marriages involving a particular match, i.e. these frequencies sum to one. NCDS data, marriages at age 23

5.5 Other Calibrations

Other parameters set outside the model are the interest rate r, parameters of the tax system τ, the household equivalence scale (n_t), time endowment T, and survival probabilities s_t.

The interest rate is set to 2.2%, following **?**. To model taxes, we use IFS TAXBEN which is a microsimulation model which calculates both taxes and benefits of each family member as a function of their income and other detailed characteristics . We then calculated taxes and benefits (including state pensions) for our sample memebers at each point in their life, and estimated a three-parameter tax system which varies across three different phases of life: young without children (age 23), working adult (ages 26-60), pension age (age 65, onwards). This three parameter tax system has the following functional form: $y_t = d_{0,t} + d_{1,t}(e_{m,t} + e_{f,t})^{d_{2,t}}$. We set the time endowment to $T = 16$ available hours per day $\times$ 7 days per week $\times$ 52 weeks per year=5,824 hours per year. We use the modified OECD equivalence scale and set $n_t = 1.4$ for couples with children. Survival probabilities are calculated using National life tables from the Office for National Statistics.

5.6 The Moment Conditions We Match

6 Second Step Results, Identification, and Model Fit

We also report our second-step parameter estimates for our structural model, the model fit, and we discuss the model's identification.

6.1 Utility Function Estimates and Identification

Table 12: Estimated structural parameters.

	Benchmark	No wage selection	Set $\lambda = 0$
β: discount factor	0.943	0.999	0.97X
	(0.04)	(0.0X)X	(0.05)
ν_f: consumption weight, female	0.378	0.377	3.84
	(0.018)	(0.01X)X	(0.55)
ν_m: consumption weight, male	0.430	0.441	x
	(0.016)	(0.01X)X	NA
γ: risk aversion	6.362	4.552	2,360
	(0.021)	X	(8,122)X
λ: altruism parameter	0.416	.385	0
	(0.012)		
Coefficient of relative risk aversion, consumption*	3.2	2.5	
Overidentification test	xx	80.6	81.5
Degrees of freedom	xx	97	96
p-value	xx	88.5%	85.4%

Notes: Standard errors: in parentheses below estimated parameters. NA: parameters fixed for a given estimation.

* Average coefficient of relative risk aversion, consumption, averaged over men and women. Calculated as $-(1/2)[(\nu_m(1-\gamma)-1)+(\nu_f(1-\gamma)-1)]$.

The parameter γ is the coefficient of relative risk aversion γ, (or the inverse of the intertemporal elasticity) for the consumption-leisure aggregate. It is the key parameter for understanding both the coefficient of relative risk aversion for consumption and for understanding the willingness to intertemporally substitute labor supply. Identification of this parameter comes from both consumption and labor supply decisions.

The coefficient of relative risk aversion for the consumption is on average XX to XX (depending on the specification) averaging over men and women,[7] which is similar to previous estimates that rely on different methodologies (see Attanasio and Weber (1995) and Browning et al for reviews of the estimates). Identification of the coefficient of relative risk aversion for consumption is similar to Cagetti (2003) and French (2005) who estimate models of buffer stock consumption over the life cycle using asset data. Within this framework, a small estimate of the coefficient of relative risk aversion means that individuals save little given their level of assets and their level of uncertainty. If they were more risk averse, they would save more in order to buffer themselves against the risk of bad income shocks in the future. These precautionary motives can explain high employment rates when young, despite the low wages of the young: more risk averse individuals work more hours when young in order to accumulate a buffer stock of assets.

We also obtain identification from labor supply, since γ is the inverse of the intertemporal elasticity of substitution for utility, and is thus key for determining the intertemporal elasticity of labor supply.[8] Wage changes cause both substitution from work both into leisure and into time spent with children.

Interestingly, we also obtain identification from the relative time versus money transfers to children. Time investments in children tend to pay a high return, but are risky since the child's human capital is a risky. The parent can also transfer a riskless financial asset to the children. A low coefficient of relative risk aversion implies counterfactually high hours levels spent with children when young, and counterfactually low transfers to the children.

Our estimate of the time discount factor, β, is larger than most estimates for three reasons. The first two reasons are clear upon inspection of the Euler Equation: $\frac{\partial u_t}{\partial c_{g,t}} \geq \beta s_{t+1}(1+r))E_t \frac{\partial u_{t+1}}{\partial c_{g,t+1}}$.[9] This equation identifies $\beta s_{t+1}(1+r)$, although not the elements of this equation separately. Therefore, a lower value of s_{t+1} or $(1+r)$ results in a higher value of β. The first reason for our high estimate of β is that most studies omit mortality risk. In our model, individuals discount the future not by the discount rate β, but by the discount factor multiplied by the survivor function s_{t+1}. Since the survivor function is necessarily

[7]We measure the individual's coefficient of relative risk aversion using the formula $-\frac{(\partial^2 u_t/\partial c_{g,t}^2)c_{g,t}}{(\partial u_t/\partial c_{g,t})} = -(\nu_g(1-\gamma)-1)$, and so the average is $-(1/2)[(\nu_m(1-\gamma)-1)+(\nu_f(1-\gamma)-1)]$. Note that this variable is measured holding labor supply fixed. The coefficient of relative risk aversion for consumption is poorly defined when labor supply is flexible.

[8] Assuming certainty, linear budget sets, and interior conditions, the Frisch elasticity of leisure is $\frac{\nu_g(1-\gamma)-1}{\gamma}$ and the Frisch elasticity of labor supply is $-\frac{l_{g,t}}{h_{g,t}} \times \frac{\nu_g(1-\gamma)-1}{\gamma}$. However, one of the advantages of the dynamic programming approach is that it is not necessary to assume certainty, linear budget sets, or interior conditions.

[9]Note that the Euler Equation holds with equality when assets are positive.

less than one, omitting mortality risk will bias β downwards. Third, the life cycle profile of hours shows that young individuals work many hours even though their wage, on average, is low. This is equivalent to stating that young people buy relatively little leisure, even though the price of leisure (their wage) is low. Between ages 35 and 60, people buy more leisure (i.e., work fewer hours) as they age even though their price of leisure (or wage) increases. Therefore, life cycle labor supply profiles provide evidence that individuals are patient. Heckman and MaCurdy (1980), and French (2005) also find that $\beta(1+r) > 1$ when using life cycle labor supply data.

The parameters ν_m and ν_f are identified by the level of hours worked, both in the market and at home investing in children. To see this note that, the marginal rate of substitution between consumption and leisure is approximately

$$\begin{aligned} w_{g,t}(1-\tau'_{g,t}) &\leq -\frac{\partial u_t}{\partial h_{g,t}} / \frac{\partial u}{\partial c_g} \\ &\leq -\frac{1-\nu_{g,t}}{\nu_{g,t}} / \frac{c_{g,t}}{l_{g,t}} \end{aligned} \tag{12}$$

which holds with equality when work hours are positive, where $\tau'_{g,t}$ is individual g's marginal tax rate at time t.[10] Inserting the time endownment equation (3) into equation (12) and making the approximation $c_{g,t} \approx w_{g,t} h_{g,t}(1-\tau'_{g,t})$ yields

$$\nu_g \approx \frac{h_{g,t}}{T - ti_{g,t}}. \tag{13}$$

Thus ν_g is approximately equal to the share of non-childcare hours that is spent at work. We find that this share is somewhat less than .5, and thus our estimate of ν_g is modestly less than .5 for both men and women.

The parameter λ is identified from three sources.

First, households make time investments in their children when young. The opportunity cost of this time is considerable.

Second, households invest in the formal education of their children. The foregone household income from children going to school represents a direct loss of resources to the household. We estimate that the returns to both time investments in education are high. For example, we find that the return to education is XX%, assuming average, which is well above the market interest rate of XX%.

Third, households make cash transfers to their children. We find that cash transfers to children are modest. However, they are the most direct source of altruism. To see this, note that from equation (7) that in the transition phase ($t = 9$, when the parent is 49 and the child is 23), parents have the opportunity to transfer resources, and the following optimality condition holds

[10]This relationship is not exact, for three reasons. First, we allow for a part time penalty to work hours.

$$\frac{\partial u_t}{\partial c_{g,t}} \geq \frac{\partial \lambda \int V'_{t'}(\mathbf{X'}_{t'})dH(\mathbf{p}'_{t'})}{\partial A'_{t'}} = \frac{\lambda \int \partial u'_t dH(\mathbf{p}'_{t'})}{\partial c'_{g,t'}}$$

that holds with equality if transfers are positive. The term on the right is the child's expected marginal utility of consumption value of assets, which the parent can transfer to the child when the child is age 23. Children are borrowing constrained and thus likely have a higher marginal utility of consumption than their parents, for multiple reasons. First, because wages rise over the life cycle, their wages are lower than their parents. Second, the children will have their own children, which is both expensive. Third, those children of children also involve time investments, reducing time available for work.

We estimated that inter-vivos cash transfers are modest relative to lifetime income. Given the higher marginal utility of consumption of children than adults when the children and the small size of transfers, this suggests that $\lambda < 1$. Nevertheless, the fact that these transfers are made is perhaps the strongest evidence that $\lambda > 0$ and households are altruistic.

6.2 Model Fit

7 Results

7.1 The Returns to Education

In our model wages depend on ability, education, and the interaction of ability and education, among other variables. We measure the return to education by exogenously changing education levels of agents in the model, then calculate their lifetime income, from age 23-65, and calculate the percent difference in earnings given these different education levels. In column (1) of table **??** we do this by assuming that the education was unanticipated: the household's decision rules are thus calculated assuming that the household can choose the child's education level. Thus changing education holds constant the decision to invest in child's ability. If everyone received low education, lifetime earnings would be xxx. If everyone selected high education, earnings would be xxx, a difference of xx%. Given that the average difference in years of schooling between low and high educated individuals is approximately 5, this translates to a xx% increase in lifetime earnings per year of education.

Section 5.3 shows that the returns to ability are higher for the highly educated, which we interpret as a form of dynamic complementarity. This implies the return to education is higher for high ability individuals. High ability individuals are thus more likely to select into education.

We focus on the return to education for two groups: those who in the baseline case selected high (college) education, and for those who selected low (compulsory) education.

For those who select low education, we calculate their lifetime income in the baseline case, then force them to have high (college) education, again calculating their lifetime income. Table **??** shows that for those who would have selected low education, lifetime income is virtually unchanged by sending them to high education. The reason for this is that the wage gains from education are modest for these low ability individuals. In contrast, the returns to education are high for those who select high education. Complementarity between ability and education, in combination with self selection in the model, is crucial.

Table 13: Returns to education.

	Unanticipated (1)	Anticipated (2)
	All	
Lifetime income, if low education		
Lifetime income, if high education		
Percent change		
Percent change, annualized*		
	Those who selected high education	
Lifetime income, if low education		
Lifetime income, if high education		
Percent change		
Percent change, annualized*		
	Those who selected low education	
Lifetime income, if low education		
Lifetime income, if high education		
Percent change		
Percent change, annualized*		

Notes: * Percent change, annualized= percent change/5.

As with column (1), column (2) solves the model both for the case where the everyone is low education then everyone is high education. However, in column (2) everyone is aware of their future education level. This means households can change their time investment and other decisions in response to the education change. When households are forced to provide more education to their children, they respond by increasing time investments, since the return to child investment is now higher. Thus, the return to education is so some extent, a choice. Column (2) shows that when allowing for these anticipation effects, the return to education is larger.

7.2 A Statistical Decomposition of the Persistence in Inequality

In this analysis, we quantify the difference in expected lifetime income (as defined below) across children with fathers of the three different education levels defined and discussed in Section 2: compulsory, some post-compulsory and some college.

7.2.1 Methods

In this analysis, we focus on male members of the NCDS cohort, and define lifetime income as the sum of gross earnings during prime working age (between the ages of 25 and 55), plus any cash transfers and bequests from parents.

Differences in cash transfers and bequests from parents can be directly observed in the NCDS and ELSA data respectively, as reported in Table 6. To calculate differences in prime-age earnings we proceed in two steps. First, we estimate the earnings equation given in Sections 3 and **??**. Second, we calculate in the NCDS data the distribution across education and ability levels of individuals with each level of father's education. By combining these two things, we can calculate expected lifetime earnings for each paternal education group.

Having calculated expected earnings for each paternal education group given the actual distributions of ability and education within each group, we then do the same calculation for three counterfactual distributions of ability and education across each paternal education group.

See appendix I for details.

Table 14: Statistical Decomposition of lifetime income gap

	Education of dad:		
	Low	Medium	High
Average lifetime income of children (£1,000):	1,129	1,443	1,720
Reduction in income gap if...			
... time investments are equalized		12%	12%
... everyone goes to college		-5%	10%
... equalize time investments and college		11%	25%
... equalizing cash transfers		11%	8%

7.3 Intergenerational Elasticity of Income

Prior work (reviewed e.g., in Black and Devereux 2011) has used many different statistics to summarize the joint distribution of parents and childrens income: (1) log-log intergenerational elasticities (IGE) of child income with respect to parent income; (2) the correlation between parent and child percentile ranks; and (3) quintile transition matrices. We focus on approaches (1) and (2).

We focus in estimating the regression

$$\ln y' = a_0 + a_1 \ln y + u \tag{14}$$

where $'$ denotes a children's variable, and thus it is the regression of child's log income on father's log income. Assuming that the model is stationary (and so $Var(\ln y') = Var(\ln y)$), the parameter a_1 can also be interpreted as the intergenerational correlation of income.

7.3.1 Measurement

Studies estimating the intergenerational correlation of income often focus issues of measurement error, which can lead to attenuation bias. Some papers have argued that there is less measurement error in earnings at ages 35-50 than at other ages and so earnings should be measured in this age range. For this reason, and to be close to much of the empirical work, we choose father's income at age 43 and child's income at age 33, which is what

For these reasons, and following much of the literature, we estimate equation (14) using four different approaches. First, we add measurement error to our simulated wage data, multiply this by model hours, and use child's income at 33 and father's income at 43 and estimate equation (14) using OLS (which is approximately the ages of fathers and sons that Dearden et al. (1997) use). Second, we use average father's income between ages 40-45, which follows much of the literature (e.g., Solon, Mazumder). Third, we instrument for father's income using father's education. This is a common procedure to address measurement error in parental income. We should not think that this instrumental variables procedure should deliver the true correlation, but merely as a point of reference to the literature. Finally, we calculate the entire lifetime labor income of fathers and children, assuming no measurement error in wages or any other variable. This would be the approach we would use if we had ideal data.

7.3.2 Intergenerational Correlation of other Variables

Education, wealth, earnings versus household income.

Table 15: Model Predicted Intergnerational Elasticity of Income

	IGE
Estimates (from Dearden et al. (1997))	0.45
Model prediction	0.42
Model prediction, equalize education	0.04
Model prediction, allow heterogeneity in education, shut down assortive mating	0.28

7.3.3 Decomposing the IGE

Our model has several channels to explain the intergenerational correlation of income. First, we allow for the fact that highly educated parents may be more productive with their time. Second,we allow for the fact that high income households spend more time with their children, increasing their age 16 ability. Third, they are more likely likely to send their children to college. To investigate the relative importance of these channels, we use both a purely statistical model, and also our structural economic model.

8 Conclusions and Policy Implications

At the most general level, the paper shows that policymakers interested in tackling the intergenerational transmission of inequalities need to consider policies designed to counter the inequality-increasing effects of each of the three forms of parental investment, since each proves to be quantitatively important in driving inequalities in income. Moreover, policymakers should bear in mind the substitutability of these different forms of investment - any attempt to shut down one channel of parental investments is likely to provoke a shift towards investment in other forms.

A Parameter definitions

Table 16 summarises the parameters that enter the model and which are introduced in the body of the paper (excluding the appendices).

Table 16: Parameter definitions

Preference Parameters		**State variables**	
β	Discount factor, annual	$g \in \{m, f\}$	Gender
β_{t+1}	Discount factor, between model periods	t	Model period
ν_g	Consumption weight in utility function	ed_g	Educational Attainment
λ	Intergenerational altruism parameter	a_t	Wealth
		$w_{g,t}$	Wage
Labour market		**Household choices**	
y_t	Household income	$c_{g,t}$	Consumption
$\tau(.)$	Net-of-tax income function	$l_{g,t}$	Leisure
$e_{g,t}$	Earnings	$h_{g,t}$	Work hours
η_t	wage innovation	$ti_{g,t}$	Time investment in children
σ_η^2	Variance of wage innovation	x_t	Cash transfer ($t = 10$)
ρ	Autocorrelation of wage innovation		
Ability		**Utility function and arguments**	
$ab'_{t'}$	Child's ability at t'	$u()$	Single period utility function
$f_{ab_t}()$	Childhood ability production function	$V_t(\mathbf{X_t})$	Value function (parenthood phase)
μ_{ab}	Stochastic ability component	$\mathbf{X_t}$	Vector of all state variables
$F_t()$	Joint distrib. of parental uncertainty	n_t	Number of equiv. adults in household
$\mathbf{q_t}$	Parental uncertainty vector	T	Time endowment
$\mathbf{p'_{t'}}$	Independence uncertainty vector		
Assets		**Other**	
$(1 + r_t)$	Gross interest rate, between model periods	τ	Length (years) of period t
r	Annual interest rate	$Q_g()$	Marriage probability function
π_t	Transfer liquidity parameter	s_{t+1}	Survival rate across period t
Measurement Systems			
ω	Vector of child ability and time investment		

B Time Periods, States, Choices and Uncertainty

Table 17 lists all model time periods, parents' and chilrens' age in those time periods, the state variables, choice variables, and sources of uncertainty during those time periods.

Table 17: Model time periods, and states, choices and sources of uncertainty during those time periods

Time Periods																				
Model period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Parent generation's age	0	7	11	16	23	26	33	37	42	49	55	60	65	70	75	80	85	90	95	100
Child generation's age						0	7	11	16	23										
Parent generation's datasets																				
NCDS					x		x		x	x	x									
Time use survey						x	x	x												
ELSA					x		x			x										
Child generation's datasets																				
NCDS						x	x	x												
Parent generation's states																				
Assets					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Wage of male and female					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Education of male and female					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Ability of male and female					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Children's gender						x	x	x	x	x										
Children's ability						x	x	x	x	x										
Children's education										x										
Parent generation's choices																				
Work hours of male and female					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Time spent with children, male and female						x	x	x												
Consumption, male and female					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Cash transfer to children										x										
Education of children									x											
Parent generation's uncertainty																				
Wage shock of male and female					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Initial ability of children						x														
Ability shock to children							x	x	x											
Children's partner										x										
Children's initial wage										x										
Mortality										x	x	x	x	x	x	x	x	x	x	x

Table 18: Note: Between periods 1 and 4 the parent generation makes no choices, and in this sense has no state variables or uncertainty.

C Data

We use data from the NCDS, ELSA, and UKTUS in our analysis, and use sample selection rules which are consistent across the three data sets. The sample selection rules are described in more detail below.

C.1 NCDS

Our main data set is the National Child Development Survey (NCDS) which started with 18,558 individuals born in one week in March 1958. We use the NCDS Data in three different ways: First, for estimating the ability production functions. Second, for estimating the income process. And third, to derive moment conditions on marital matching, education shares, employment rates, the fraction of full-time work, and wealth at age 33. We explain the samples used for these three purposes in more detail below.

Production function estimation: For the production function estimation, we require individuals to have a full set of observations on ability all measures, investment measures between the ages of 0-16, parental education, and parental income (see table 1 for a full list of measures). This reduces the original sample of 48,644 observations to 11,596 observations across the four waves considered.

Income process: For the estimation of the income process, we consider the waves collected at ages 23, 33, 42, 50, and 55, leaving out age 46 due to low-quality data. This leads to a total of 54,352 observations in adulthood. Of these, we drop all self-employed people (5,932 excluded), those who are unmarried after age 23 (7,602 excluded), those for who we only have one wage observation (9,909 excluded) leaving us with 30,909 observations. We trim wages at the top and bottom 1% for each sex and education group.

Moments: For the moments, we exclude all self-employed people (5,932 excluded), and those who are unmarried after age 23 (7,602 excluded), leaving us with a total sample of 40,818.

C.2 ELSA

We use the ELSA data both for asset data at age 50 which we use in our moment conditions and also for the gift and inheritance data which we use both in our moment conditions and also to construct the liquidity share π_t. ELSA is a biannual survey of those 50 and older, starting in 2002. We use data up through 2016.

Although NCDS sample members are asked about assets at age 50, these data are considered to be of low quality because the data omit housing wealth; thus we use ELSA instead. For our wealth measure, we use the sum of housing wealth including second homes, savings, investments including stocks and bonds, trusts, business wealth, and physical wealth such as land, after financial debt and mortgage debt has been subtracted.

For the asset moment condition at age 50 we begin with 924 respondents who are age 50 at the time of the survey. We drop members of cohorts not born between 1950-1959 (which excludes 255 observations), unmarried people (which excludes 88 observations), and the self-employed (which excludes

54 observations). Finally, we have 14 households where both members were exactly age 50. In order to not double count these households, we exclude one observation from these two person households, resulting in 513 individuals remaining.

ELSA has high quality data on gifts and inheritances in wave 6 (collected in 2012-2013). In this wave respondents were asked to recall receipt of inheritances and substantial gifts (defined as those worth over £1,000 at 2013 prices) over their entire lifetimes. Respondent are asked age of receipt and value for three largest gifts and three largest inheritances.[11] From our original sample of 10,601 in 2012, we drop members of cohorts not born between 1950-1959 (which excludes 7,223 observations), singles (921 excluded), and self-employed (328 excluded), resulting in 2,129 individuals remaining. Of those 2,129 individuals, 1,884 had at least one parent has died and 1,094 had both parents died by the time of the survey. Thus 51% of our sample already had both parents die by this point and thus have likely received all transfers they will ever receive.

Table 19: Sample comparison: NCDS and ELSA

	Education shares			
	Male		Female	
	NCDS	ELSA	NCDS	ELSA
Low	16%	20%	22%	26%
Medium	49%	38%	49%	40%
High	35%	43%	29%	34%
	Median net weekly earnings in £			
	Male		Female	
	NCDS	ELSA	NCDS	ELSA
Low	399	315	223	171
Medium	479	383	266	221
High	665	519	399	358

Notes: In NCDS, low education includes no educational qualification or CSE2-5, Medium education includes O-level or A-level, High education includes higher qualifications or a degree. In ELSA, low education includes no educational qualification or CSE, Medium education includes O-level or A-level, High education includes higher qualifications below a degree or a degree. Earnings are median net weekly earnings in £2013.

C.3 UKTUS

Using the NCDS we can construct a latent time investment index, but not investment time itself. For measuring investment time we use UKTUS data from 2000-2001. Respondents use a time diary to record activities of their day in 144 x 10-minute time slots for one weekday and one weekend day. In each of these slots the respondent records their main ("main activity for each ten minute slot") and secondary

[11]Only 3.6% of all individuals have three or more large inheritances or bequests (Crawford 2014), so the restriction is unlikely to signficantly affect our results.

activities ("most important activity you were doing at the same time"), as well as who it was carried out with. We have diaries for both parents and the children, but use only the parent diaries.

We construct our measure of time spent with children by summing up across both parents the ten minute time slots during which an investment activity with a child takes place either as a main or a secondary activity. Although we know the number of children and the age of each child within the household, we do not know the precise age of the child that received the investment, we assume this to be the youngest child. We include all of the following activities as time spent with the child when constructing the investment measure: unspecified childcare, physical care and supervision, feeding the child, other specified physical care and supervision, teaching the child, reading/playing/talking with child, accompanying child, other specified childcare, travel escorting a child (other than education), travel escorting to/from education.

Our original sample includes 11,053 diary entries. We keep only married individuals with a child ≤ 15 yrs (which excludes 6,694 observations), drop households with more than 2 adults (797 excluded), keep those for whom we have diary information on both parents for both a weekend day and a weekday (506 excluded), and keep only 2 kid families (1,660 excluded), leaving us with 1,396 remaining observations: (349 households with 4 entries (weekend, weekday for mum, dad)).

D Appendix Table

Table 20: Proportion of children in each father's education group

	Low (compulsory)	75%
Father's education	Middle (post-compulsory)	20%
	High (some college)	5%

E Our Implementation of the Agostinelli-Wiswall Approach

E.1 Production Function

The production function for skills takes a restricted translog specification, equation (6) from the main text.

$$\ln ab'_{t'+1} = \gamma_{1,t'} \ln ab'_{t'} + \gamma_{2,t'} \ln ti_{t'} + \gamma_{3,t'} \ln ti_t \cdot \ln ab_{t'} + \gamma_{4,t'} ed_m + \gamma_{5,t'} ed_f + u'_{ab,t'}$$

E.2 Parental Investment Function

We estimate a reduced form parental investment equation to gain intuition about the effect of parental characteristics on investments. Parental investments depend on a child's current skills, parents' education and parental income.

$$\ln ti_t = \alpha_{1,t} \ln ab'_{t'} + \alpha_{2,t} ed_f + \alpha_{3,t} ed_m + \alpha_{4,t} \ln y_t + u_{ti,t} \tag{15}$$

E.3 Measurement

We do not observe children's skills ($ab'_{t'}$), or investments (ti_t) directly. However we observe $j = \{1, ..., J_{\omega,t}\}$ error-ridden measurements of each. These measurements have arbitrary scale and location. That is for each $\omega \in \{ab, ti\}$ we observe:

$$Z_{\omega,t,j} = \mu_{\omega,t,j} + \lambda_{\omega,t,j} \ln \omega_t + \epsilon_{\omega,t,j} \tag{16}$$

All other variables are assumed to be measured without error.

E.4 Assumptions on Measurement Errors and Shocks

Measurement errors are assumed to be independent across measures and across time. Measurement errors are also assumed to be independent of the latent variables, household income and the structural shocks $(u_{ti,t}, u_{ab,t}, ed_f, ed_m, \ln y_t)$.

E.5 Normalizations

As mentioned above, skills and investments do not have a fixed location or scale which is why we need to normalize them. In the first period, we normalize the mean of the log-latent factor to be zero which fixes the location. In all other periods, the mean of the log-latent factor is allowed to be different from zero. Moreover, for each period, we set the scale $\lambda_{\omega,t,1} = 1$ for one normalizing measure $Z_{\omega,t,1}$.

AW have shown that renormalization of the scale can lead to biases in the estimation of coefficients in the case of overidentification of the production function coefficients. This is not the case in our estimation as we do not assume constant returns to scale which means our estimation does not suffer from overidentification. For more details, see **?**.

E.6 Intial Conditions Assumptions

Period 1 for the child and period 6 for the parent is the time of the child's birth.

$$\mathbf{\Omega} = (\ln ab_1^{'}, \ln ti_6, ed_f, ed_m, \ln y_t) \sim N(\mu_\Omega, \mathbf{\Sigma}_\Omega) \tag{17}$$

$\mu_\Sigma = [0,0,0,0,0]$. The mean of $\ln ab_1^{'}$, $\ln ed_f$, $\ln ed_m$ and $\ln ti_6$ are 0 by normalization and without loss of generality.

E.7 Estimation

1. **Variance of latent factors** (the elements of $\mathbf{\Sigma}_\Omega$).

   Using equation (16) we can derive the variance of each of the latent factors:

   $$Cov(Z_{\omega,t,j}, Z_{\omega,t,j^*}) = \lambda_{\omega,t,j}\lambda_{\omega,t,j^*}\, Var(\ln \omega_t) \tag{18}$$

   Note that this is overidentified as there are many different combinations of j and j^* that can be used here ($j^* \neq j$). Whilst AW select one of the combinations, we use a bootstrap to estimate the variances of the objects in equation (18), and run a diagonal GMM in order to construct a unique $Var(\ln \omega_0)$. Because y, ed_m, ed_f are observable, it is straightforward to estimate the covariance of these with each other, as well as their covariance with ability and time investments.

2. **Scale parameters (λs) in measurement equations**. Here we estimate the parameters for the measurement equations for the child skill and investment latent variables. We have normalised $\lambda_{ab,t,1} = \lambda_{ti,t,1} = 1$ to set the scale of ab_t and of ti_t. For each other measure $j \neq 1$, and for $\omega \in \{ab, ti\}$, using equation (18) we can show that:

   $$\lambda_{\theta_\omega,t,m} = \frac{Cov(Z_{\omega,t,j}, Z_{\omega,t,j^*})}{Cov(Z_{\omega,t,1}, Z_{\omega,t,j^*})} \tag{19}$$

   Note that this is overidentified as there are many different combinations of j and j^* that can be used here.

3. **Location parameters (μs) in measurement equations** At the child's birth, we normalize the mean of $\ln \theta_1$ and $\ln ti_6$ to zero. Therefore:

   $$\mu_{\omega,t,j} = \mathbb{E}[Z_{\omega,t,j}] \tag{20}$$

4. **Calculation for next step** For each measure we need to calculate a residualized measure of each Z for $\omega_t \in \{ab_t, ti\}$:

$$\tilde{Z}_{\omega,t,j} = \frac{Z_{\omega,t,j} - \mu_{\omega,t,j}}{\lambda_{\omega,t,j}} \tag{21}$$

This will be used below in Step 1. Note that:

$$\ln \omega_t = \tilde{Z}_{\omega,t,j} - \underbrace{\frac{\epsilon_{\omega,t,j}}{\lambda_{\omega,t,j}}}_{\equiv \tilde{\epsilon}_{\omega,t,j}} \tag{22}$$

It gives log skills (or time input) plus an error rescaled to match scale of the skills (which is also the scale of skill measure 1).

5. **Estimate latent skill production technology**

   We will only describe the estimation of the production technology, as the estimation of the investment equation is analogous. Recall the production function:

$$\ln ab^{'}_{t'+1} = \gamma_{1,t'} \ln ab^{'}_{t'} + \gamma_{2,t'} \ln ti_{t'} + \gamma_{3,t'} \ln ti_t \cdot \ln ab_{t'} + \gamma_{4,t} ed_m + \gamma_{5,t} ed_f + u^{'}_{ab,t'}$$

   and using equation (22) note that we can rewrite the above equation as:

$$\begin{aligned} \frac{Z_{ab',t'+1,j} - \mu_{ab',t'+1,j} - \epsilon_{ab',t'+1,j}}{\lambda_{ab',t'+1,j}} \quad = \quad & \gamma_{1,t'}(\tilde{Z}_{ab',t',j} - \tilde{\epsilon}_{ab',t',j}) + \\ & \gamma_{2,t'}(\tilde{Z}_{ti,t,j} - \tilde{\epsilon}_{ti,t,j}) + \\ & \gamma_{3,t'}(\tilde{Z}_{ti,t,j} - \tilde{\epsilon}_{ti,t,j}) \cdot (\tilde{Z}_{ab',t',j} - \tilde{\epsilon}_{ab',t',j}) + \\ & \gamma_{4,t'} ed_m + \gamma_{5,t'} ed_f + \\ & u_{ab',t'} \end{aligned} \tag{23}$$

   or

$$\begin{aligned} \frac{Z_{ab',t'+1,j} - \mu_{ab',t'+1,j}}{\lambda_{ab',t'+1,j}} \quad = \quad & \gamma_{1,t'} \tilde{Z}_{ab',t',j} + \\ & \gamma_{2,t'} \tilde{Z}_{ti,t,j} + \\ & \gamma_{3,t'} \tilde{Z}_{ti,t,j} \cdot \tilde{Z}_{ab',t',j} + \\ & \gamma_{4,t'} ed_m + \gamma_{5,t'} ed_f + \\ & (u_{ab',t'} - \tilde{\epsilon}_{ab',t',j} - \tilde{\epsilon}_{ti,t,j} - \tilde{\epsilon}_{ti,t,j} \cdot \tilde{\epsilon}_{ab',t',j} + \frac{\epsilon_{ab',t'+1,j}}{\lambda_{ab',t'+1,j}}). \end{aligned} \tag{24}$$

OLS is inconsistent here, as $\tilde{Z}_{ab',t',j}$ and $\tilde{\epsilon}_{ab',t',j}$ are correlated. We resolve this issue by instrumenting for $\tilde{Z}_{ab',t',j}$ using the other measures of ability $\tilde{Z}_{ab',t',j^*}$ in that period.

Recall that we only normalized the location of factors in the first period, but have not done so for the subsequent periods (in this case $\mu_{ab',t'+1,j}$). We estimate the location parameter for each measure j by estimating equation (24) using only output measure j on the left hand side. The intercept then identifies $\mu_{ab',t'+1,j}$.

Once we have estimated all location parameters, we allow for the whole set of relevant input and output measures, and estimate equation (24) by using a system GMM with diagonal weights. By using the system GMM we make efficient use of all available measures.

6. **Estimate the variance of the production function shocks**

The variance of the structural skills shock can be obtained using residuals from equation (24), where $\pi_{\theta,t,j} \equiv (u_{ab,t} - \tilde{\epsilon}_{ab,t,j} - \tilde{\epsilon}_{ti,t,j} - \tilde{\epsilon}_{ti,t,j} \cdot \tilde{\epsilon}_{ab,t,j} + \frac{\epsilon_{ab,t+1,j}}{\lambda_{ab,t+1,j}})$:

$$Cov\left(\frac{\pi_{\theta,t,j}}{\lambda_{\theta,t,j}}, \tilde{Z}_{\theta,t,j^*}\right) = \sigma^2_{\theta,t,j}$$

As again, these covariances are overidentified, we use a bootstrap and diagonal GMM to estimate the shock variances efficiently. Again, the variance of the time investment shocks is estimated similarly.

F Initial Ability

We allow for the fact that high education parents may have high ability children. We assume that child ability at birth is log-normally distributed conditional on parental education. The mean of this distribution is estimated by adjusting the means of the gestation and birthweight measure by their respective scaling parameters (λ defined above) and combining them using a minimum distance estimator. The variance of the distribution is estimated by dividing the covariance of the two measures by the product of the scaling parameters, as described in step 1 of section E.7. This is done separately for each combination of maternal and paternal education groups. Table **??** shows the mean and initial variance of ability for each parental education group.

Table 21: Means and variances of initial ability conditional on parental education group

Means:				
		Mother's education		
		Low	Medium	High
	Low	0.845	-0.097	0.021
Father's education	Medium	-0.341	0.367	-0.916
	High	-1.345	0.389	0.915
Variances:				
		Mother's education		
		Low	Medium	High
	Low	2.606	3.487	4.878
Father's education	Medium	2.969	1.787	6.526
	High	4.319	2.194	0.840

G Signal to Noise Ratios

Signal to noise ratios for all measures are calculated in the following way:

$$s_{\omega,t,j} = \frac{(\lambda^2_{\omega,t,j})Var(\ln\theta_{\omega,t})}{(\lambda^2_{\omega,t,j})Var(\ln\theta_{\omega,t}) + Var(\epsilon_{\omega,t,j})}$$

Intuitively, this is the variance of the latent factor (signal) to the variance of the measure (signal+noise) and thus describes the information content of each measure.

Table 22 presents signal to noise ratios for ability. At birth, birthweight is the most informative measure. At age 7, reading, maths, coping, and drawing scores are all roughly equally informative. At ages 11 and 16, maths scores become the most informative.

Table 22: Signal to noise ratios: Ability measures

Age 0		*Age 7*		*Age 11*		*Age 16*	
birthweight	0.862	read	0.385	read	0.555	read	0.57
gestation	0.14	maths	0.335	maths	0.942	maths	0.713
		copy	0.259	copy	0.104		
		draw	0.281				

Table 23 presents signal to noise ratios for investment. Here we have many measures of investment. The most informative measures when young are the freqency of father's outings with the child, and both mother's and father's frequency of reading to the child. At older ages, the most informative variable is the teacher's assessment of each parent's interest in the child's education.

Table 23: Signal to noise ratios: Investment measures

Age 0		*Age 7*		*Age 11*	
mum: interest	0.164	mum: interest	0.356	mum: interest	0.796
mum: outing	0.27	mum: outings	0.235	dad: interest	0.765
mum: read	0.456	dad: outings	0.166	other index	0.344
dad: outing	0.773	dad:interest	0.386	parental ambition	0.221
dad: interest	0.082	dad:role	0.033		
dad:read	0.539	parents initiative	0.206		
dad: large role	0.069	parents ambition uni	0.093		
other index	0.136	parents ambition school	0.249		
		library member	0.253		

All investment measures are retrospective, so age 0 investments are measured at age 7, age 7 investments are measured at age 11, age 11 investments are measured at age 16.

H Estimating the Wage Equation, Accounting for Measurement Error in Ability and Wages

We estimate the wage equation laid out in Section 3. We present the estimation equation in (I.1) for convenience:

$$\ln w_t^* = \ln w_t + u_t = \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 lnab_4 + \delta_5 PT_t + v_t + u_t \text{ where} \tag{25}$$

$$v_t = \rho v_{t-1} + \eta_t,$$

$$u_t \text{ is IID measurement error in wages}$$

and PT_t relates to part time status, for each gender and education group.

In our procedure we must address three issues. First, wages are measured with error u_t. Second, ability ab_4 is measured with error. Third, we only observe the wage for those who work, which is a selected sample.

However, we also wish to address issues of selectivity relying on our panel data as much as is possible. To address the issue of composition bias (the issue of whether lifetime high or low wage individuals drop out of the labor market first), we use a fixed effects estimator. Assuming that $\rho = 1$, which we estimate to be close to the truth, we can allow v_5 (the first period of working life, age 23) to be correlated with other observables, and estimate the model using fixed effects. In particular, the procedure is:

Step 0: Note that if $\rho = 1$, then: $v_t = v_5 + \sum_{k=6}^{t} \eta_k$. Hence:

$$\begin{aligned} \ln w_t^* &= \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 lnab_4 + \delta_5 PT_t + u_t + v_5 + \sum_{k=6}^{t} \eta_k \\ &= \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_5 PT_t + FE + \xi_t \end{aligned} \tag{26}$$

where FE is a person specific fixed effect capturing the time invariant factors $\delta_4 lnab_4 + v_5$ and ξ_t is a residual.

Step 1: Estimate δ_1, δ_2, δ_3, δ_5 using fixed effects (FE) regression.

Step 2: Predict the fixed effect:

$$\begin{aligned} \widehat{FE} &\equiv \bar{\ln} w_t^* - \hat{\delta}_1 \bar{t} - \hat{\delta}_2 \bar{t}^2 - \hat{\delta}_3 \bar{t}^3 - \hat{\delta}_5 \bar{PT}_t \\ &= \delta_0 + \delta_4 lnab_4 + v_5 \\ &= \delta_0 + \delta_4 \tilde{Z}_{ab,4,j} + v_5 - \delta_4 \tilde{\epsilon}_{ab,4,j} \end{aligned} \tag{27}$$

where $lnab_4 = \tilde{Z}_{ab,4,j} - \tilde{\epsilon}_{ab,4,j}$ is defined in equation (22). Although the estimated fixed effect, $\widehat{FE}$, is affected by variability in the sequence of wage shocks $\{\eta_t\}_{t=6}^{12}$ and measurement errors $\{u_t\}_{t=5}^{12}$, this merely adds in measurement error on the left hand side variable in equation (27) . However, measurement error on the right hand side $\ln ab_4$ is more serious: we only have the noisy proxies $\tilde{Z}_{ab,4,j}$ which are correlated with $\tilde{\epsilon}_{ab,4,j}$ by construction. We address this problem in the next step.

Step 3: Using GMM, project the predicted fixed effect ($\widehat{FE}$) on each measure of ability, $\tilde{Z}_{ab,4,j}$, and instrument by using the respective other measures, $\tilde{Z}_{ab,4,j'}$, to get $\hat{\delta_0}$ and $\hat{\delta_4}$. Since we have two measures of ability (reading and math), we have two equations and two instruments. When reading is the ability measure, we instrument for this using math, and vice versa. The GMM efficiently combines different measures of ability and simultaneously correct for measurement error.

Step 4: Then use covariances and variances of residuals to calculate shock variances.

Redefining equation (25) using $lnab_4 = \tilde{Z}_{ab,4,j} - \tilde{\epsilon}_{ab,4,j}$ as defined in equation (22).:

$$\ln w_t^* = \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 \tilde{Z}_{ab,4,j} + \delta_5 PT_t + v_t + u_t - \delta_4 \tilde{\epsilon}_{ab,4,j}.$$

We estimate the parameters of the wage shocks $\rho, Var(\eta)$, and $Var(v_5)$.

$$\tilde{\ln w_t} \equiv \ln w_t^* - (\delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 \tilde{Z}_{ab,4,j} + \delta_5 PT_t) = v_t + u_t - \delta_4 \tilde{\epsilon}_{ab,4,j}$$

Note that from the measurement equation 16, $V(\tilde{Z}_{ab,4,j}) = V(lnab_4) + V(\tilde{\epsilon}_{ab,4,j})$, where we have previously estimated $V(lnab_4)$ using equation 18 and $V(\tilde{Z}_{ab,4,j})$ is the variance of the renormalized measures in the data. We can then back out the variance of the measurement error and plug it into the following equation to estimate the parameters of the wage shocks:

$$C(\widetilde{\ln w_t}, \widetilde{\ln w_{t+k}}) = \rho^k V(v_t) + V(\tilde{\epsilon}_{ab,4,j})$$

$$V(\widetilde{\ln w_t}) = V(v_t) + V(u_t) + V(\tilde{\epsilon}_{ab,4,j})$$

$$V(\widetilde{\ln w_{t+k}}) = \rho^k V(v_t) + \sum_{j=1}^{k} \rho^j V(\eta_{t+k}) + V(u_{t+k}) + V(\tilde{\epsilon}_{ab,4,j}).$$

Step 5: correct the δ parameters for selection. The fixed-effects estimator is identified using wage growth for workers. If wage growth rates for workers and non-workers are the same, composition bias problems—the question of whether high wage individuals drop out of the labor market later than low wage individuals—are not a problem. However, if individuals leave the market because of a wage drop, such as from job loss, then wage growth rates for workers will be greater than wage growth for non-workers. This selection problem will bias estimated wage growth upward.

We control for selection bias by finding the wage profile that, when fed into our model, generates the same fixed effects profile as in the estimates using the NCDS data. Because the simulated fixed effect profiles are computed using only the wages of those simulated agents that work, the profiles should be biased upwards for the same reasons they are in the data. We find this bias-adjusted wage profile using the iterative procedure described in French (2005).

I Decomposing the Difference in Lifetime Income by Parental Education

In this analysis, we quantify the difference in expected lifetime income (as defined below) across children with fathers of the three different education levels defined and discussed in Section 2: low (which corresponds to compulsory education in the UK), middle (completion of secondary schooling), and high (some college). Importantly, we explicitly account for the fact that ability and parental investments are measured with error and are thus latent variables.

I.1 Methods

We define lifetime income as the sum of gross earnings during prime working age (between the ages of 23 and 65) which are taken from the NCDS, plus any cash transfers and bequests from parents which are taken from ELSA.

To calculate differences in prime-age earnings we first predict the evolution of ability from age 0 to age 16. We then use predicted ability at age 16, and education as observed in the data to predict earnings using the earnings equation described in section H. We assume that everyone works full time. The steps are described in more detail below.

Step 1 We combine multiple measures of investments measured at ages 7, 11, and 16, as well as multiple measures of ability at birth to produce latent factor scores using the weighted GLS procedure in Appendix F.2 of Heckman et al. (2013). Recall that investments measured at age 7 represent investments for the period from age 0-6, those measured at age 11 represent ages 7-10, and those measured at 16 represent ages 11-15. We begin with the prediction of age 7 ability, by plugging in the latent ability score at birth, the age 7 latent investment score, maternal and paternal education as observed in the data into the production function described in section 4.1. Using this predicted age 7 ability, the age 11 latent investment score, and parental education, we predict age 11 ability, again using our estimated production functions. Finally, we repeat this process for age 16 ability.

Step 2 We then use predicted ability at 16, and education as observed in the data to predict wages for each age using the earnings equation I.1. We then assume that individuals work for 40 hours per week, and 52 weeks per year, thus giving us annual earnings. We then sum the earnings between ages 23 to 65, and calculate means for each paternal education group.

We then add mean transfers for each paternal education group as observed in ELSA to the predicted earnings of each group. The differences between paternal education groups serve as a baseline for comparison with earnings differentials derived from the following counterfactuals.

1. Equalizing ability at age 7: In Step 1, we set ability at age 7 equal to the mean predicted value of the latent ability score at age 7 for all individuals, and predict ability at 11, 16, and earnings.

2. Equalizing ability at age 16: In Step 2 we set ability at age 16 equal to the mean predicted value of the latent ability score at age 16 for all individuals, and predict earnings.

3. Equalizing investments at age 7 (11, or 16): In step 1, we set investments at age 7 (11, or 16) for all individuals equal to the mean of the latent investment score at that age. We then predict ability at all subsequent ages as well as earnings.

4. Equalizing investments at all ages: In step 1, we set investments in each period equal to the mean of the latent investment score for the respective age. We then predict ability at 7, 11, 16, as well as earnings.

5. Equalizing education: In step 3, we set education equal to the highest level for all individuals, and predict earnings.

6. Equalizing time investments & college: In step 1, we set investments in each period equal to the mean of the latent investment score for all individuals. We then predict ability at 7, 11, 16. We additionally set education in step 3 equal to college for everyone, and predict earnings.

7. Equalizing cash transfers. We set cash transfers equal to zero for everyone.

We then add the respective mean transfers from ELSA to each group, and calculate the percentage change in earnings across the different groups.

To compare our predicted lifetime income to the data, we calculate a model-free benchmark of lifetime income using a weighted sum of gross earnings plus transfers in the raw data. Recall that we only have information on wages at ages 23, 33, 42, 50, 55. Moreover, for some individuals, earnings data are missing for some periods (we restrict the sample to individuals with least two observations). We first multiply observed wages in each period by 40 hours and 52 weeks. We then calculate a weighted average of earnings, allowing for missing earnings observations:

$$Average\ yearly\ earnings_i = \frac{1}{\sum_{j=1}^{J} \mathbb{1}_{obs_{i,j}}(\#years\ in\ period\ j)} \sum_{j=1}^{J} \mathbb{1}_{obs_{i,j}} earnings_{i,j}(\#years\ in\ period\ j)$$

where $j = \{23, 33, 42, 50, 55\}$ and the respective years in each period are $\#years\ in\ period\ j = \{10, 9, 8, 5, 10\}$. We then multiply the average earnings by a total of 42 years, and add mean lifetime transfers for each paternal education group from ELSA.

I.2 Results

Overall differences in expected lifetime income as defined above are shown in the first row of Table **??**. Males (females) with low education fathers have expected earnings that are £282,000 (£233,000) lower than those with middle education fathers and have expected earnings that are £539,000 (£439,000) lower than those with high education fathers. For reference, the lifetime income of males with low-educated fathers is £1,152,000, whereas for females it is £848,000. We assume for these predictions that everyone works 40 hours. In reality, many females work part-time rather than full-time; hence for females, these numbers reflect the difference in earnings *potential* rather than actual earnings differences.

Table 24: Lifetime income, by fathers' education

	Males			Females		
	Education of dad:			Education of dad:		
	Low	Medium	High	Low	Medium	High
Predicted lifetime income in 1000 pounds	1,208	1,517	1,815	873	1,148	1,330
Predicted earnings in 1000 pounds	1,181	1,459	1,717	842	1,058	1,221
Cash transfers	27	58	99	31	90	109
Change in income gap when...						
Ability						
... if all children had same ability at age 7		36%	28%		20%	16%
... if all children had same ability at age 16		72%	69%		41%	43%
... no effect of parental education on ability		61%	59%		34%	35%
Time Investments						
... Equalizing early investments		2%	2%		1%	1%
... Equalizing middle investments		6%	6%		4%	4%
... Equalizing late investments		5%	4%		3%	3%
... if all children received the same time investments		12%	12%		8%	8%
Education						
... if all children attained the highest level of education		-3%	9%		23%	31%
... if all children had medium education		23%	32%		45%	52%
... if all children had low education		62%	65%		57%	63%
... equalize investments and college		13%	23%		34%	42%
Cash transfers						
...if all children received the same cash transfers		10%	12%		21%	17%
Benchmark Earnings in 1000 pounds (data)	1,326	1,708	1,987	1,326	1,708	1,987
Fraction explained by model:	91%	89%	91%	66%	67%	67%

* Average inter-vivos transfers and bequests received over life-cycle

For both males and females, there is a large effect of equalizing ability at age 7, because ability differences by parents' education are already large by age 7. Furthermore, ability is persistent, and thus ability at age 16 is strongly influenced by age 7. Higher age 16 ability leads to higher lifetime income, especially for those who select high education (as shown in table **??**). As a result, 69% (45%) of the difference in lifetime income between males (females) born to high versus low education fathers can be explained by age 16 ability.

These differences in ability arise both because of differences in investments and also the direct effects of different parental education levels on children's ability. For example, equalizing parental time investments leads to a sizable reduction of 12% in the earnings gap for males, but only 9% for females. The direct effect of parents' education is also important; for example, equalizing the direct effect of maternal education in the production function for everyone reduces the earnings gap by XX% and XX%, for males and females respectively.

Next we test for dynamic complementarity in the production of human capital. Complementarity implies that increasing investments at one age increases the productivity of investments at other ages. Thus, increasing investments at all ages should result in larger impacts than the sum of the impacts of equalizing investments one age at a time. Interestingly, we do not evidence for complementarity using this approach. For example, whereas equalizing all equalizing parental time investments leads to a reduction of 12% in the earnings gap for males, equalizing investments at ages 0-6, 7-10, and 11-15 reduce the gap 2%, 6%, and 4%, respectively. Note that these individual contributions sum to 12%, which is the same as the increase from increasing investments at all ages.[HUB: essentially the sign of the interaction changes in the different periods, which seems to cancel out some of the dynamic complementarity. I did some comparisons (seeing whether equalizing 0-6 and 7-10 sums to equalizing at age 11 etc.), but didn't include it as table was getting too messy. We should possibly include a new table that only looks at dynamic complementarity in investments when we get back after summer.]

Children of the highly educated are more likely to continue their studies. Thus education is an additional force in propagating intergenerational inequality. To assess the importance of education, we fix age 16 ability at its baseline value, then change education from its baseline value to giving everyone the highest level of education. This leads to a small reduction in the earnings gap for men (10%), but leads to a sizable reduction for women (27%). The reduction in income inequality is smaller for men than women, for two reasons. First, there is more wage dispersion for men, but the return to education at average ability is slightly larger for women than men. Thus equalizing education at average ability has a smaller relative reduction in the amount of wage dispersion. Second, for men with high education, the effect of ability on wages is strong. Thus education amplifies the effect of differences in ability. Giving everyone a high level of education does not lead to a strong reduction in the earnings gap for men, unless ability is also equalized. For high-educated women, however, the amplification effect is smaller, which is why equalizing ability leads to a larger reduction in income inequality.

Equalizing both time investments and education jointly leads to a reduction of the lifetime income gap by 25% for men and 39% for women. Interestingly, these impacts are larger than the sum of the two in isolation. The reason for this is that ability and education are complementary in producing higher wages: those with high education have the highest returns to investments in ability. Due to this complementarity, the effects of equalizing ability and the subsequent equalization of education reinforce each other in the earnings predictions, thus leading to ability playing a key role in lifetime earnings.

Cash transfers are less important than earnings for explaining lifetime differences in income, and also matter less in the lifetime income differential for males compared to females. Eliminating differences in cash transfers reduces the lifetime income gap by 9% for men, and by 18% for women. The smaller effects for men can be explained by higher labour income for men over the life cycle, which reduces the relative importance of cash transfers. Moreover, the difference in transfers received by children of low versus medium or high educated parents is larger for women (£56,900) than for men (£34,100).

To summarise, the decomposition analysis suggests that for men (women) 28% (18%) of the difference in lifetime income across paternal education groups is attributable to differences in ability at age 7, 69% (45%) by ability at age 16, and 8% (18%) by inter-vivos transfers and bequests received.

Thus, while inter-vivos transfers are important, most of the differences in lifetime income between children of low versus higher education fathers are realized by the age of 16.

Lastly we should note that all of the earnings amounts were predicted given ability and education and estimated model coefficients. The bottom row, however, shows that these factors can explain most of the differences in earnings of earnings between children of high versus low education parents. Whereas the raw data show that sons (daughters) of high education fathers have lifetime income that is £631,000 (£439,000) greater than that of sons (daughters) of low education fathers, our model predicts a difference of £539,000 (£451,000). These modest difference between model prediction and data suggest that we have captured the key drivers of the intergenerational persistence of lifetime income.

J Estimation of the Liquidity Parameter π_t

In the model we assume that all transfers received are at age 23 but in practice many transfers are received after this age. To account for this, we assume that only a share of period t assets are liquid, which leads to the liquidity constraint equation (2), which we repeat here:

$$(c_{m,t} + c_{f,t}) + x_t \leq \pi_t a_t + y_t \iff a_{t+1} \geq (1 + r_t)(1 - \pi_t)a_t. \tag{28}$$

We use two data sources to estimate π_t: asset data from the NCDS at different 23, 33, and 50 which we refer to as $Assets_{NCDS,t}$. Some of these assets are the result of savings of earned income, whereas some of these assets come from parental transfers. We use information on size and timing of parental cash transfers using retrospective data to construct $Intervivos_{ELSA,t}$ and $Bequest_{ELSA,t}$, which is total inter-vivos and bequest transfers received up to time t using ELSA data.

We wish to create a measure of current assets plus the discounted value of all future transfers received, which is our concept of assets in the model. To do this we create a measure of all assets that does not derive in any way from transfers $Assets_{own,t}$, i.e. the assets excluding parental cash transfers, then we add to this the sum of all lifetime transfers:

$$Assets_{own,t} = Assets_{NCDS,t} - Intervivos_{ELSA,t} - Bequest_{ELSA,t}$$

We then estimate total life time parental transfers in the form of inter-vivos trans $Intervivos_{ELSA,LT}$ and bequests $Bequest_{ELSA,LT}$, that is transfers received up until the date of the ELSA survey. We assume that these are equal to total lifetime transfers received, which appears reasonable since the average age of ELSA sample members is XX. By this age most of their parents have already died. We then use this to construct the model analog of total assets from all sources, which is the sum of $Assets_{own,t} + Intervivos_{ELSA,LT} + Bequest_{ELSA,LT}$

? π_t as the fraction of liquid assets (as observed in the NCDS) over

$$\pi_t = \frac{Assets_{NCDS,t}}{Assets_{own,t} + Intervivos_{ELSA,LT} + Bequest_{ELSA,LT}}$$

As we do not observe the same individuals in the NCDS as in ELSA, we use means of the assets and transfers for the estimations above.

K Computational Details

This Appendix details how we solve for optimal decision rules as well as our simulation procedure. We describe solving for optimal decision rules first.

1. To find optimal decision rules, we solve the model backwards using value function iteration. The

state variables of the model are model period, assets, wage rates, education levels, own ability, childrens' gender, childrens' ability, and childrens' education. At each model period, we solve the model for 25 grid points for assets, 15 points for wage rates (for each spouse), 3 education levels for each spouse, 5 points for own ability for each spouse, childrens' gender, childrens' ability (5 points), and childrens' education. Because we assume that the two children are identical, receive identical shocks, and that parent make identical decisions towards the two children, we only need to keep track of the state variables for one child. Our approach for discretizing wage shocks follows **?**. The bounds for the discretisation of the wage process is $\pm$ 3 standard deviations. For ability we use Gauss-Hermite procedures to integrate. We use linear interpolation between grid points when on the grid, and use linear extrapolation outside of the grid.

2. Parents can each choose between between 3 levels of working hours (non-employed, part-time, full-time) and in model period $t = 6, 7$ and 8 they can choose between four levels of time spent with children. In all model periods except $t = 10$ we solve for the optimal level of next period assets using golden search. In period $t = 10$ as parents may also transfer assets to children: we solve this two-dimensional optimization problem using Nelder-Mead. We back out household consumption from the budget constraint and then solve for individual level consumption from the intra-temporal first order condition, which delivers the share of household consumption going to the male in the household. As this first order condition is a non-linear function we approximate the solution using the first step of a third order Householder algorithm. This allows us to use the information contained in the first three derivatives of the first order condition. We found this method to give fast and accurate solutions to the intra-temporal problem. Details of this are available from the authors.

Next we describe our simulation procedure.

1. Our initial sample of simulated individuals is large, consisting of 50,000 random draws of individuals in the first wave of our data at age 23. Given that we randomly simulate a sample of individuals that is larger than the number of individuals observed in the data, most observations will be drawn multiple times. We take random Monte Carlo draws of education and assets, which are the state variables that we believe are measured accurately and are observed for everyone in the data. For the variables with a large amount measurement error, or are not observed for all sample members (i.e., initial ability of each parent and child, and wages of each parent), we exploit the model implied joint distribution of these state variables. We assume child's gender is randomly distributed across the population.

2. Given the optimal decision rules, the initial conditions of the state variables, and the histories of

shocks faced by both parents and children, we calculate life histories for savings, consumption, labor supply, time and education investments in children, which then implies histories for childrens' ability, educational attainment. For discreet choice variables (e.g. participation), we evaluate whether the choice is the same at all surrounding grid points. If not, we resolve the households problem given each of the households' choices (e.g., work and not work), and choose the value that delivers the highest value. If so, we take the implied discreet variable, and if any of the continuous state variables (e.g. assets) is between grid-points we interpolate to find the implied decision rule.

3. We aggregate the simulated data in the same way we aggregate the observed data, and construct moment conditions. We describe these moments in greater detail in Appendix L. Our method of simulated moments procedure delivers the model parameters that minimize a GMM criterion function, which we also describe in Appendix L.

4. To search for the parameters that minimize the GMM criterion function, we use the BOBYQUA algorithm developed by Powel (2009). This is a derivative free algorithm that use a trust region approach to build quadratic models of the objective function on sub-regions.

L Moment Conditions and Asymptotic Distribution of Parameter Estimates

We estimate the parameters of our model in two steps. In the first step, we estimate the vector χ, the set of parameters than can be estimated without explicitly using our model. In the second step, we use the method of simulated moments (MSM) to estimate the remaining parameters, which are contained in the $M \times 1$ parameter vector $\Delta =(\beta, \nu_f, \nu_m, \gamma, \lambda)$. Our estimate, $\hat{\Delta}$, of the "true" parameter vector Δ_0 is the value of Δ that minimizes the (weighted) distance between the life-cycle profiles found in the data and the simulated profiles generated by the model.

We match data from three different sources. For most of our moments we use data from the NCDS. However, the NCDS currently lacks high quality asset and transfer data after age 23, and does not have detailed time use information with children. For the asset and transfer data we also match data from ELSA, and for the information on time with children we also use UKTUS.

From the NCDS we match, for three education groups ed, two genders (male and female) g, $T = 5$ different ages: $t \in \{23, 33, 42, 50, 55\}$ the following moment conditions: $3 \times 2 \times T = 6T$ moment conditions: employment rates (forming 6T moment conditions), mean annual work hours of workers ($6T$), the shares in each education group, by gender of child and and father's education level (18 moment conditions), from the NCDS data. In addition, we also match wealth tertiles at age 33, by education of the husband, using the NCDS (9 moments).

From ELSA we match mean lifetime inter-vivos transfers received, by education and gender of recipient (6 moments) and also household wealth tertiles at age 50, by education of the husband (9 moments).

From UKTUS we match mean annual time spent with children, by age of child (ages 0-7, 8-11, 12-16) and gender and education of parent (18 moments).

To properly match the entire asset distribution, we match asset tertiles. Our procedure for matching these tertiles is described below. Suppose that individual i is observed at time t and lives in a household where the husband's education group is ed_m. Let $a_{\pi_j ed_m t}(\Delta, \chi)$ denote the model-predicted π_j asset quantile for individuals in individual i's group at time t, where χ includes all parameters estimated in the first stage. Assuming that observed assets have a continuous conditional density, a_{it} will satisfy

$$\Pr\Big(a_{it} \leq a_{\pi_j ed_m t}(\Delta_0, \chi_0) \,|\, ed_m, t, \text{individual } i \text{ observed at } t\Big) = \pi_j.$$

The preceding equation can be rewritten as a moment condition (**??**). In particular, applying the indicator function produces

$$E\big(1\{a_{it} \leq a_{\pi_j ed_m t}(\Delta_0, \chi_0)\} - \pi_j \,|\, ed_m, t, \text{individual } i \text{ observed at } t\big) = 0. \tag{29}$$

We can convert this conditional moment equation into an unconditional one (e.g., **?**):

$$E\Big([1\{a_{it} \leq a_{\pi_j ed_m t}(\Delta_0, \chi_0)\} - \pi_j] \times 1\{ed_{m_i} = ed_m\} \times 1\{\text{individual } i \text{ observed at } t\} \;\Big|\; t\Big) = 0, \tag{30}$$

for $ed_m \in \{\text{low (compulsory)}, \text{middle (post-compulsory)}, \text{high (some college)}\}$. In the end, we have a total of $J = 120$ moment conditions.

Our approach accounts explicitly for the fact that the data are unbalanced: some individuals leave the sample, and we use multiple datasets, so an individual who belongs in one sample (e.g., NCDS) likely does not belong to another sample (e.g., ELSA or UKTUS). Suppose we have a dataset of I independent individuals that are each observed at up to T separate calendar years. Let $\varphi(\Delta; \chi_0)$ denote the J-element vector of moment conditions described immediately above, and let $\hat{\varphi}_I(.)$ denote its sample analog. Letting $\widehat{\mathbf{W}}_I$ denote a $J \times J$ weighting matrix, the MSM estimator $\hat{\Delta}$ is given by

$$\underset{\Delta}{\operatorname{argmin}} \; \frac{I}{1+\tau} \, \hat{\varphi}_I(\Delta; \chi_0)' \widehat{\mathbf{W}}_I \hat{\varphi}_I(\Delta; \chi_0),$$

where τ is the ratio of the number of observations to the number of simulated observations.

In practice, we estimate χ_0 as well, using the approach described in the main text. Computational concerns, however, compel us to treat χ_0 as known in the analysis that follows. Under regularity conditions stated in **?** and **?**, the MSM estimator $\hat{\Delta}$ is both consistent and asymptotically normally distributed:

$$\sqrt{I}\left(\hat{\Delta} - \Delta_0\right) \rightsquigarrow N(0, \mathbf{V}),$$

with the variance-covariance matrix $\mathbf{V}$ given by

$$\mathbf{V} = (1+\tau)(\mathbf{D}'\mathbf{W}\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}\mathbf{S}\mathbf{W}\mathbf{D}(\mathbf{D}'\mathbf{W}\mathbf{D})^{-1},$$

where $\mathbf{S}$ is the variance-covariance matrix of the data;

$$\mathbf{D} = \frac{\partial\varphi(\Delta;\chi_0)}{\partial\Delta'}\Big|_{\Delta=\Delta_0} \tag{31}$$

is the $J \times M$ gradient matrix of the population moment vector; and $\mathbf{W} = \text{plim}_{I\to\infty}\{\widehat{\mathbf{W}}_I\}$. Moreover, **?** shows that if the model is properly specified,

$$\frac{I}{1+\tau}\hat{\varphi}_I(\hat{\Delta};\chi_0)'\mathbf{R}^{-1}\hat{\varphi}_I(\hat{\Delta};\chi_0) \rightsquigarrow \chi^2_{J-M},$$

where $\mathbf{R}^{-1}$ is the generalized inverse of

$$\begin{aligned}\mathbf{R} &= \mathbf{P}\mathbf{S}\mathbf{P},\\ \mathbf{P} &= \mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{W}\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}.\end{aligned}$$

The asymptotically efficient weighting matrix arises when $\widehat{\mathbf{W}}_I$ converges to $\mathbf{S}^{-1}$, the inverse of the variance-covariance matrix of the data. When $\mathbf{W} = \mathbf{S}^{-1}$, $\mathbf{V}$ simplifies to $(1+\tau)(\mathbf{D}'\mathbf{S}^{-1}\mathbf{D})^{-1}$, and $\mathbf{R}$ is replaced with $\mathbf{S}$.

But even though the optimal weighting matrix is asymptotically efficient, it can be biased in small samples. (See, for example, **?**.) We thus use a "diagonal" weighting matrix, as suggested by **?**. This diagonal weighting scheme uses the inverse of the matrix that is the same as $\mathbf{S}$ along the diagonal and has zeros off the diagonal of the matrix.

We estimate $\mathbf{D}$, $\mathbf{S}$, and $\mathbf{W}$ with their sample analogs. For example, our estimate of $\mathbf{S}$ is the $J \times J$ estimated variance-covariance matrix of the sample data. When estimating this matrix, we use sample statistics, so that, for example, $a_{\pi_j ed_m t}(\Delta, \chi)$ is replaced with the sample quantile for group $ed_m t$.

One complication in estimating the gradient matrix $\mathbf{D}$ is that the functions inside the moment condition $\varphi(\Delta;\chi)$ are non-differentiable at certain data points; see equation (30). This means that we cannot consistently estimate $\mathbf{D}$ as the numerical derivative of $\hat{\varphi}_I(.)$. Our asymptotic results therefore do not follow from the standard GMM approach, but rather the approach for non-smooth functions described in **?**, **?** (section 7), and **?**.

To find $\mathbf{D}$, it is helpful to rewrite equation (30) as

$$\Pr\Big(ed_{m_i} = ed_m \ \& \ \text{ individual } i \text{ observed at } t\Big) \times \left[\int_{a_{\pi_{j-1}ed_mt}(\Delta_0,\chi_0)}^{a_{\pi_j ed_m t}(\Delta_0,\chi_0)} f\Big(a_{it} \mid ed_m, t\Big)\, da_{it} \ - \ \pi_j\right] = 0. \tag{32}$$

It follows that the rows of $\mathbf{D}$ are given by

$$\Pr\Big(ed_{m_i} = ed_m \ \& \ \text{individual } i \text{ observed at } t\Big) \times f\Big(a_{\pi_j ed_m t} \;\Big|\; ed_m, t\Big) \times \frac{\partial a_{\pi_j ed_m t}(\Delta_0,\chi_0)}{\partial \Delta'}. \tag{33}$$

In practice, we find $f\Big(a_{\pi_j ed_m t} \;\Big|\; ed_m, t\Big)$, the conditional p.d.f. of assets evaluated at the quantile $a_{\pi_j ed_m t}$, with a kernel density estimator written by **?**. The gradients for equations (**??**) and (**??**) are found in a similar fashion.