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

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

November 1, 2022

PRELIMINARY AND INCOMPLETE

Abstract

Parental investments in children can take one of three broad forms: (1) Time investments during childhood that affect child ability (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 and incorporates all three of these types of investments. We find that 28% of the variance of lifetime wages can already be explained by characteristics of the parents before individuals are born and 62% of the variance can be explained by age 23 characteristics of the individual. In terms of investments, we find evidence of dynamic complementarity between time and educational investments – the returns to education are higher for high ability individuals. This is a potentially important mechanism in perpetuating intergenerational outcomes, as borrowing constraints prevent low-income families from investing in education, thus simultaneously reducing the incentive to invest in time.

*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 Sree Ayyar, 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 NORFACE for Grant 462- 16-120: Trends in Inequality: Sources and Policy", 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 of Bristol and Institute for Fiscal Studies

‡University of Cambridge and Institute for Fiscal Studies

§University of 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. Solon (1992), Dearden et al. (1997), Mazumder (2005), Chetty et al. (2014), Gallipoli et al. (2020)). 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 economic 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 (Cunha et al. (2006), Heckman and Mosso (2014)) ii) parental aid for education (Belley and Lochner (2007), Abbott et al. (2019)) and iii) cash gifts in the form of inter-vivos transfers and bequests (Castaneda et al. (2003), De Nardi (2004)).

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 Cunha et al. (2010) and Agostinelli and Wiswall (2022) 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. The model is used to a) show how particular intergenerational transfers affect the outcomes of household members, b) compare the relative importance of these types of transfers and c) evaluate counterfactual policies.

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. These are, first, the borrowing constraint, which limits low income low wealth families from accessing credit to provide higher education to their children. The second mechanism is that we allow parental productivity in investing in children to be correlated 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, we allow for a dynamic complementarity between early and late time investments and between time investments and educational invesments. 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. 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 wages of 0.24, close to estimates for our cohort of interest in Dearden et al. (1997). The model also replicates the fact (documented by Guryan et al. (2008) and observed in our data) that parents with more education spend more time with their children. We have two key 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 low education men this premium is only 9%. 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, who one happens to be born to and who one happens to marry are central for explaining life's outcomes. We find that 30% of the variance of men's and 13% of the variance of women's lifetime wages can already explained by characteristics of their parents, before the individual is even born. By the time individuals are 23, the shares rise to 65% and 45% for men and women, respectively. By modeling marriage and the behavior of both members of couples, we can assess not only the variability of individual but also household income. The characteristics of one's spouse is an important source of uncertainty in lifetime income prior to marriage, especially for women who on average earn less than their spouses. Resolution of this uncertainty explains almost half of the variability in *household* lifetime income for women.

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 Caucutt and Lochner (2020), Lee and Seshadri (2019) and Daruich (2018). 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. Caucutt and Lochner (2020) focus on identifying the role of market imperfections in rationalizing observed levels of parental investments. The aim of Lee and Seshadri (2019) is to simultaneously rationalize intergenerational persistence in outcomes and cross-sectional inequality in outcomes. Daruich (2018) focuses on the macroeconomic effects of large-scale policy interventions. Our primary focus, facilitated by our data on 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 Del Boca et al. (2014) and Gayle et al. (2018), 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. Abbott et al. (2019) 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. Castaneda et al. (2003) and De Nardi (2004) 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. Bolt et al. (2022) use the same data as in this paper and mediation analysis to show that the mechanisms we consider in this paper are the key ones for explaining the persistence of income across generations. But that paper does not allow for behavioral responses, and so cannot be used to consider counterfactuals.

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 two step estimation approach. Section 5 then presents results from the first step estimation, whereas Section 6 presents identification arguments and results from the second step estimation. Section 7 presents results from counterfactuals and Section 8 concludes.

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.

As the NCDS currently does not have data on the inheritances received or expected, we supplement it using data on individuals drawn from similar birth cohorts in the English Longitudinal Study of Ageing (ELSA). ELSA 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 gift and inheritance receipt which we can use to augment our description of the divergence in lifetime economic outcomes by parental background. We use data on ELSA members who are born in the 1950s, which gives us a sample of 3,001.[2]

Lastly, to convert the investment measures observed into units of time, we use the UK Time Use Survey (UKTUS), which has detailed measures of time spent in educational investments in the child. We describe these measures in the notes of Table 2 and in greater detail in Appendix C.

The rest of this section documents inequalities in the three types of parental transfers we are interested it (time investments, educational investments, and cash transfers), as well as subsequent outcomes (ability, lifetime income). 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.

[1]The age-46 survey is not used in any of the subsequent analysis as it was a more limited telephone-only interview.

[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.

2.1 Transfer Type 1: Parental Time Investments

The NCDS has detailed measures of parental time investments received during childhood. The full set of measures we use to estimate the impact of parental time on cognition are listed in Table 1.[4] These measures come from different sources – some are from surveys of parents, others from surveys of teachers. Here we highlight some of the key features in the data.

The first panel of Table 2 documents paternal education gradients for some of the investment measures we use. Whilst 52% of high educated fathers read to their age 7 child each week, only 33% of low educated fathers do so. The gradient is even more pronounced for the teacher's assessment of the parents' interest in the child's education: when the child is 7, 66% of high educated fathers are judged by the child's teacher to be 'very interested' in their child's education but only 20% of low education fathers are. While mothers are assessed as having greater interest in their child's education than fathers, there are large differences according to education group (75% of the highest education group are very interested, compared to 33% in the lowest education group).

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	

Notes: 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.

[4] While some of these measures are potentially costly in terms of money as well as time, we focus on the time cost which the previous literature has found to be the key determinant of child cognition (e.g., Del Boca et al. (2014)).

Table 2: Transfers and outcomes by father's education

	Avg	SD	Father's education Low	Medium	High	p-val*
Transfer 1: Parental Investments						
Mother reads each week 7	0.49	0.50	0.46	0.56	0.67	0.00
Father reads each week 7	0.36	0.48	0.33	0.44	0.52	0.00
Mother outings most weeks 11	0.54	0.50	0.53	0.61	0.59	0.00
Father outings most weeks 11	0.51	0.50	0.50	0.58	0.56	0.00
Father very interested in education 7	0.26	0.44	0.20	0.43	0.66	0.00
Mother very interested in education 7	0.39	0.49	0.33	0.58	0.75	0.00
Father very interested in education 11	0.31	0.46	0.23	0.52	0.73	0.00
Mother very interested in education 11	0.39	0.49	0.33	0.59	0.76	0.00
Father very interested in education 16	0.36	0.48	0.28	0.57	0.80	0.00
Mother very interested in education 16	0.38	0.49	0.32	0.59	0.78	0.00
Hours spent with child per week [UKTUS]**	9.06	10.05	8.35	8.91	9.87	.52
Transfer 2: Child Education						
Fraction low education	0.25	0.43	0.30	0.10	0.02	0.00
Fraction high education	0.16	0.37	0.13	0.31	0.46	0.00
Transfer 3: Cash Transfers						
Inter-vivos transfers (>£1000)	0.07	0.26	0.06	0.10	0.20	0.06
Gift value (among recipients only)	39,400	104,600	30,600	77,900	49,100	0.72
Fraction receiving inheritance	0.39	0.49	0.36	0.58	0.54	0.00
Inheritance value (among recipients)	88,200	114,700	75,600	122,400	174,300	0.00
Outcome 1: Child Ability						
Reading 7	0.00	1.00	-0.09	0.33	0.58	0.00
Reading 11	0.00	1.00	-0.13	0.46	0.90	0.00
Reading 16	0.00	1.00	-0.11	0.47	0.77	0.00
Maths 7	0.00	1.00	-0.08	0.26	0.54	0.00
Maths 11	0.00	1.00	-0.13	0.48	0.91	0.00
Maths 16	0.00	1.00	-0.14	0.48	0.99	0.00
Outcome 2: Real Lifetime Earnings in £1,000						
Men	1,347	352	1,289	1,533	1,740	0.00
Women	925	239	879	1,048	1,197	0.00

Notes: For different types of transfers and outcomes, Table 2 shows: Mean, standard deviation, mean conditional on each paternal education group (low, medium, high) *P-values for an F-test of the difference in the mean between the low and high father's education group. ** Sum of father's and mother's time spent on the following activities spent with the child in UKTUS data: teaching the child, reading/playing/talking with child, travel escorting to/from education.

2.2 Transfer Type 2: Educational Investments

Panel 2 of Table 2 shows that there is a substantial intergenerational correlation in educational attainment between fathers and their children. Having a high-educated father makes it much more likely that a child will end up with high education. 46% of the children of high educated fathers also end up with high education, compared to only 13% of those whose fathers have low education.

2.3 Transfer 3: Inter-vivos Transfers and Bequests

The third panel of Table 2 documents the receipt of inter-vivos transfers and bequests as reported in ELSA by father's education. The table shows significant differences in the receipt of inter-vivos transfers depending on parental education. Only 6% of individuals from low education families report having received a transfer worth more than £1,000, compared to 20% from high educated families. Moreover, conditional on receipt of a gift, the average value for the two groups differs by about £18,400.

Differences in inheritance receipt by parental background are also significant. 54% of those with high educated fathers have received an inheritance, compared to 36% of those with low-educated fathers, and among those who have received an inheritance, those with high educated fathers have received more than twice as much on average (£174,300 compared to £75,600). The net result is that those with high educated fathers inherit £66,000 more than those with low-educated fathers.

2.4 Outcome 1: Ability

The fourth panel of Table 2 shows the average of reading and math ability of children at ages 7, 11, and 16, by father's education. As one might expect, children whose father has a higher level of education have higher ability; at the age of 7, reading ability of children of low educated fathers is 0.09 standard deviations below average, whereas it is 0.58 above average for children of high educated fathers. This ability gap widens with age: by the time the children are 16, reading ability of children of low educated fathers is 0.11 standard deviations below average, whereas it is 0.77 above average for children of high educated fathers. Similar patterns are found for math scores.

2.5 Outcome 2: Lifetime Earnings

Finally, we can see that children of more educated fathers have higher lifetime earnings. The gap in lifetime earnings between men with high educated fathers versus those with low educated fathers is £451k. For women, the difference is £318k.

To summarize, we find that children from more highly educated fathers tend to receive more of each of the three kinds of transfers, and they end up with higher ability, as well as lifetime income. In the following, we present a model bringing together these different types of transfers, to explain how these operate in generating the intergenerational persistence in outcomes that we observe.

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. Figure 1 illustrates the model's timeline. During childhood, parental time investments in children and educational choices affect the evolution of the child's ability and their educational attainment. Upon reaching age 23, they are matched in couples, possibly receive transfers of cash from their parents and begin adult life. They then have their own children, choose consumption, labor supply, and how much to invest in their own children, with implications for their children's future outcomes.

The NCDS interviews respondents 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 length). 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 but does not make decisions.

2. Independence consists of one period at $t = 5$ corresponding to ages 23-25. The individual receives a parental cash transfer (which is potentially 0), is matched into a couple and begins making labor supply and savings decisions.

3. Parenthood has five periods $t = 6, 7, 8, 9,$ 10, corresponding to ages 26-32, 33-36, 37-41,42-48, 49-54. The couple have identical twin children at the start of the 'Parenthood' phase. In addition to making labor supply and savings decisions, the couple 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 their children who in turn are matched into couples.

4. Late adult phase consists of 10 regularly-spaced periods corresponding to ages 55-59, ..., 100-104. The household separates from their children and makes their own saving and consumption decisions.

In outlining the dynastic model we describe below a lifecycle decision problem of a single generation. All generations are, of course, linked; each couple has children. These children, in turn, will form couples have children, too. To index generations we use t to denote the age (in model periods) of the generation

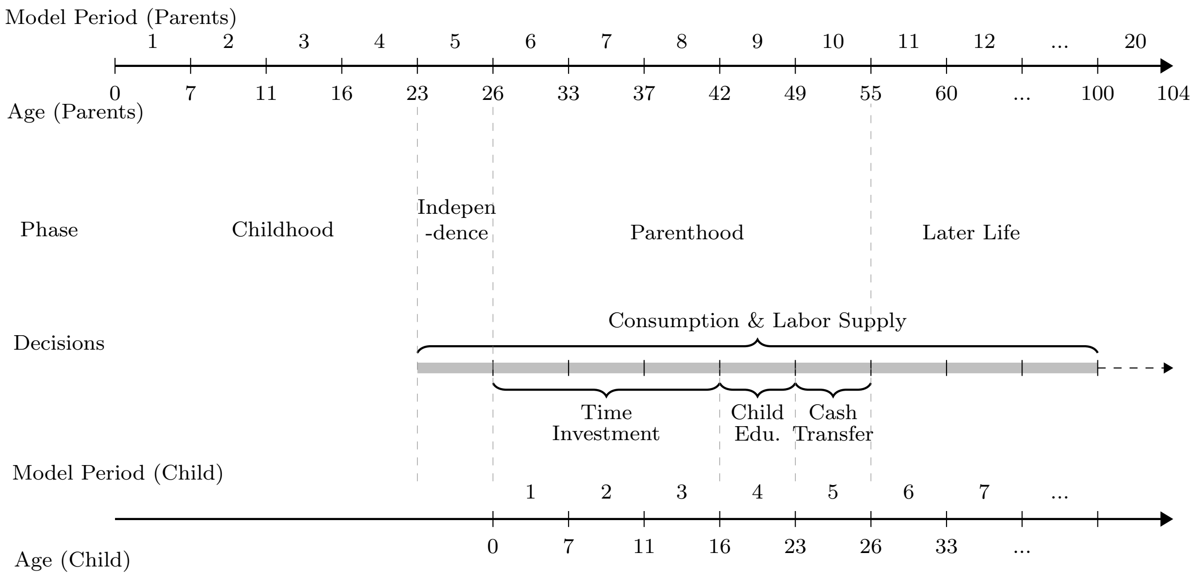


Figure 1: The Life Cycle of an Individual

we consider and 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 ($l_{g,t}$):

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

We allow preferences for consumption and leisure to vary with gender. Households equally weight the sum of male and female utility. The household utility function is 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 quantity of leisure is:

$$l_{g,t} = T - (\theta ti_{g,t} + h_{g,t}) \tag{1}$$

[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 $t = 6$.

where T is a time endowment, $ti_{g,t}$ is time investment hours in children, $h_{g,t}$ is work hours, and $l_{g,t}$ is leisure time. $1-\theta$ is the share of time with the child that represents leisure to the parent: if θ =0 then time with children is pure leisure for the parent, whereas if θ =1 then time with children generates no leisure value.

The annual discount factor is β. The model period length aligns with the differences in time between interviews and so the discount factor between model period varies. Thus the discount rate between t and $t+1$ is $\beta_{t+1} = \beta^{\tau_t}$, where τ_t 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 Demographics

All individuals are matched probabilistically into couples, conditional on education. 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 matching probabilities for females are $Q_f(ed_f, ed_m)$. 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 Abbott et al. (2019) and assume that the twins are faced with identical sequences of shocks.

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 households face mortality risk after the age of 50 and that death occurs by the age of 105 at the latest.

3.3 Human Capital

This section describes the production function for ability and education from birth to age 23. During this part of the life cycle, parental time investments 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, all of which matters to the altruistic parent.

3.3.1 Child Ability Production Function

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

[6] In addition, to account for varying period length and within period discounting we weight each period's utility by $\sum_{q=0}^{\tau_t} \beta^q = \frac{1-\beta^{\tau_t+1}}{1-\beta}$.

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

where $ab^{'}_{t'}$ represents child's ability when the child is age t'. Child's ability depends on his/her parents' level of education, the sum of the time investments ($ti_{t'} = ti_{m,t'} + ti_{f,t'}$) those parents make, past ability, and a shock ($u'_{ab,t+1}$). Ability evolves until period 4 (age of 16), after which it does not change.

We allow education of the mother and father, ed_m and ed_f, 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., Lee and Seshadri (2019)).

A child's initial ability at birth $ab'_{1'}$ is a function of his/her parents' level of education and a shock:

$$ab'_{1'} = \gamma_{4,0'}ed_m + \gamma_{5,0'}ed_f + u'_{ab,0'}. \tag{3}$$

3.3.2 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. Thus we model the decision to send the child to school until age 16, age 18 (completing secondary education) or 21 (completing undergraduate education). Because there were no tuition fees for the cohort we study, we model the cost of education as forgone labor income when at school.

3.3.3 Wages

The wage rate evolves according to a process that has a deterministic component which varies with age and whether the individual works part-time or fulltime, and a stochastic component:

$$\ln\ w_t \quad = \quad \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + v_t \tag{4}$$

where PT_t is a dummy for working part time. To capture the impact of ability on lifetime wages we model the initial wage draw in period 5 (age 23) as a function of final ability (ab) and a shock: subsequent values follow a random walk

$$v_t = \begin{cases} \delta_5 ab_t + \eta_t, & \eta_t \sim N(0, \sigma^2_{\eta 4}) \quad \text{if } t = 5 \\ v_{t-1} + \eta_t, & \eta_t \sim N(0, \sigma^2_{\eta}) \quad \text{if } t > 5 \end{cases} . \tag{5}$$

Ability impacts the age 23 wage shock v_5 and thus impacts wages at all ages because v_t is modeled as having a unit root. Thus we do not need to keep track of ability after turning age 23, but instead we keep track of wages as a state variable, which includes v_t and thus final ability. While the associated subscripts are suppressed above, each of $\{\delta_0, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \sigma_\eta, \sigma_{\eta 4}\}$ varies by gender (g) and education (ed). This flexibility means that we allow ability to impact wages through its relationship with education δ_5. As we show below this flexibility is important as the returns to ability are higher for the highly educated.

3.4 Budget Constraints

Constraints Households an intertemporal budget constraint and a borrowing constraint:

$$a_{t+1} = (1 + r_t)(a_t + y_t - (c_{m,t} + c_{f,t}) - x_t) \tag{6}$$

$$a_{t+1} \geq 0 \tag{7}$$

where a_t is the household wealth, y_t is household income and x_t is a cash transfer to children that can only be made when the members of the couple are 49 and their children are 23 (and so $x_t = 0$ in all other periods). 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.

Earnings and household income Earnings are equal to hours worked (h) multiplied by the wage rate, for example: $e_{f,t} = h_{f,t} w_{f,t}$. Household net-of-tax income is

$$y_t = \tau(e_{m,t}, e_{f,t}, e'_t, t) \tag{8}$$

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. Prior to age 16 childrens' earnings (e'_t) are 0. Upon turning age 16 (period t =4) children work full time at the median wage given their age and gender for the years in the model period they are no longer in education. Their parents are still the decision-maker in this period and any income the children earn is part of household income.

3.5 Decision Problem

3.5.1 Decision Problem in the Young Adult

An individual becomes an active decision maker at age 23 when they are already formed into a household as part of a childless couples. As such $t = 5$ is the first model period with a decision problem to solve.

Choices Each period during this phase couples choose consumption $(c_{m,t}, c_{f,t})$ and hours of work of each parent $(h_{m,t} h_{f,t})$ where $h_{g,t} \in \{0, 20, 40, 50\}$ hours per week. The resulting vector of decision variables is $\mathbf{d_t} = (c_{mt}, c_{f,t}, h_{m,t}, h_{f,t})$.

Uncertainty Couples face uncertainty over the innovation to each of their wages next period $\{\eta_{m,t}, \eta_{f,t}\}$ and the initial ability of their future children $u'_{ab,0'}$.

State variables The vector of state variables $(\mathbf{X}_t)$ during young adulthood is (suppressing time subscripts) $\mathbf{X}_t = \{t, a_t, w_{m,t}, w_{f,t}, ed_m, ed_f\}$ where t is age, a_t is assets, $w_{m,t}, w_{f,t}$ are the wages of each parent, and ed_m, ed_f the education of each spouse.

Value function The value function for the independent adult phase is given below in expression (10):

$$V_t(\mathbf{X}_t) \quad = \quad \max_{\mathbf{d_t}} \left\{ u(c_{mt}, c_{f,t}, h_{m,t}, h_{f,t}, n_t) + \beta_{t+1} \mathbb{E}_t \big[V_{t+1}(\mathbf{X}_{t+1}) \big] \right\} \tag{9}$$

subject to the intertemporal budget constraint in equation (6) and the borrowing constraint in equation (7) where the expectation operator is over the innovation to the wage of each of spouse $(\eta_{m,t}, \eta_{f,t})$ and the initial ability of the child $(u'_{ab,0'})$.

3.5.2 Decision Problem in the Parenthood Phase

Choices Households make decisions on behalf of both the adults and children within the household each period. They choose consumption and hours of work of each parent $(c_{m,t},\ c_{f,t}, h_{m,t},\ h_{f,t})$, time investments in children of each parent ($ti_{m,t}$ and $ti_{f,t}$) until their child turns 16, and childrens' education ed' in the period the children turn 16. The resulting vector of decision variables is $\mathbf{d_t} = (c_{m,t}, c_{f,t}, h_{m,t}, h_{f,t}, ti_{m,t}, ti_{f,t}, ed')$.

Uncertainty Couples face uncertainty over the innovation to each of their wages $\{\eta_{m,t}, \eta_{f,t}\}$ and the innovations to the childrens' ability $(u'_{ab,t})$.

State variables The set of state variables in this phase is $\mathbf{X}_t = \{t, a_t, w_{m,t}, w_{f,t}, ed_m, ed_f, g', ab'_t\}$, which is the same as in the independent adult phase plus the childrens' gender (g') and their ability (ab'_t).

Value function The household's value function and the constraints are the same as in equation (10), except adapted to have the sets of choices, uncertainty, states described immediately above.

3.5.3 Decision Problem in the Independence of Children Phase

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

Choices During this phase couples choose consumption $(c_{m,t}, c_{f,t})$, hours of work for each parent $(h_{m,t}, h_{f,t})$, and a cash gift (x_t) which is split equally between their two children. The resulting vector of decision variables is $\mathbf{d_t} = (c_{m,t}, c_{f,t}, h_{m,t}, h_{f,t}, x_t)$.

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 second dimension of uncertainty is with respect to their own circumstances next year – that is their next period wage draws.

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

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

$$V_t(\mathbf{X}_t) \quad = \quad \max_{\mathbf{d_t}} \left\{ u(c_{m,t}, c_{f,t}, l_{m,t}, l_{f,t}, n_t) + 2\lambda \mathbb{E}_t[V'_{t'}(\mathbf{X'}_{t'})] + \beta_{t+1} \mathbb{E}_t[V_{t+1}(\mathbf{X}_{t+1})] \right\}$$

subject to equations (6) and (7). 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) children of the parent will belong to, and the expectation operator is over the children's initial wage draw and their future spouse's attributes. 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 initial wage draw and the characteristics of their spouse (these shocks are realized after the parents makes their decisions). We have assumed that parents have two identical children and therefore we multiply this continuation value by 2. 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). This expectation operator is with respect to next period's wage draws, which are stochastic, and discounted by β_{t+1}, the time discount factor.

3.5.4 Decision Problem in the Late Adult phase

At this stage the children have entered their own young adult phase and the parent couple enters a late adult phase.

Choices Households make labor supply and consumption/saving decisions only ($\mathbf{d_t} = (c_{mt}, c_{f,t}, h_{m,t}, h_{f,t})$).

Uncertainty There is uncertainty over next period's wage draws and survival s_t (we assume both members of the couple die in the same period).

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

Value function Given the definitions of choices, states, and uncertainty for the late life phase the value function and the constraints take the same form as for the young adult phase (expression (10)).

$$V_t(\mathbf{X}_t) \quad = \quad \max_{\mathbf{d_t}} \left\{ u(c_{mt}, c_{f,t}, h_{m,t}, h_{f,t}, n_t) + \beta_{t+1} s_{t+1} \mathbb{E}\big[V_{t+1}(\mathbf{X}_{t+1})\big] \right\} \tag{10}$$

subject to equations (6) and (7).

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 human capital production function, the wage process, marital sorting process, and mortality rates. 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 remaining parameters using the method of simulated moments and correct for selection bias in the wage equation.

4.1 Estimating the Human Capital Production Function

Estimating the latent factor production function We have multiple noisy measures of children's latent ability ($ab'_{t'}$) and parental investment ($inv_{t'}$) in our NCDS data. Following the recent literature (Agostinelli and Wiswall (2022)), we estimate a human capital production function where latent ability is a function of previous period's (latent) ability and investments, parental education, and a shock:

$$ab'_{t'+1} = \alpha_{1,t'}ab'_{t'} + \alpha_{2,t'}inv_{t'} + \alpha_{3,t'}inv_{t'} \cdot ab'_{t'} + \alpha_{4,t'}ed^m + \alpha_{5,t'}ed^f + u'_{ab,t'} \tag{11}$$

We explicitly account for measurement error in the latent factors using a GMM implementation of the methods in (Agostinelli and Wiswall (2022)). Following the literature (Cunha and Heckman (2008), Cunha et al. (2010)), we assume independence of measurement errors, allowing us to use all possible combinations of (noisy) input measures to instrument for one another using a system GMM approach described in Appendix E.

Converting latent investments to time Equation (11) gives us the coefficient of a unit of latent investment on a unit of latent ability. However, latent ability and latent investments do not have a natural scale. We normalize the scale of the ability measure via the wage equations (4 and 5), which we discuss in Appendix H below.

We anchor latent parental investments to hours of investment time, as this is the relevant object in the model. To anchor the latent investments estimated using the NCDS to time, we use another data set that contains information on hours of time spent with children – the UK Time Use Survey (UKTUS). We assume time investments with children impact latent investments according to:

$$inv_{t'} = \kappa_{0,t'} + \kappa_{1,t'}(ti_{m,t'} + ti_{f,t'}) \tag{12}$$

where $\kappa_{1,t'}$ is the hours-to-latent investments conversion parameter which determines the productivity of time investments and $\kappa_{0,t'}$ is a constant that ensures we match mean time investments. We allow the κ parameters to vary by age, to reflect that parental time investments, and the productivity of those investments, varies by age.

The parameters $\kappa_{0,t'}$ and $\kappa_{1,t'}$ are estimated using MSM by matching age 16 ability by father's education in the NCDS data and time investments by parental education in the UKTUS data. We discuss the estimation and identification of $\kappa_{0,t'}$ and $\kappa_{1,t'}$ in Section 6.

With the parameters $\kappa_{0,t'}$ and $\kappa_{1,t'}$ in hand, we substitute equation (12) into equation (11) as follows, where the second line is (2) which is the production function we use in our dynamic programming model:

$$\begin{aligned} ab'_{t'+1} &= \alpha_{1,t'}ab'_{t'} + \alpha_{2,t'}(\kappa_{0,t'} + \kappa_{1,t'}ti_{t'}) + \alpha_{3,t'}(\kappa_{0,t'} + \kappa_{1,t'}ti_{t'}) \cdot ab'_{t'} + \alpha_{4,t'}ed^m + \alpha_{5,t'}ed^f + u'_{ab,t'} \\ &= \gamma_{0,t'} + \gamma_{1,t'}ab'_{t'} + \gamma_{2,t'}ti_{t'} + \gamma_{3,t'}ti_t \cdot ab'_{t'} + \gamma_{4,t'}ed^m + \gamma_{5,t'}ed^f + u'_{ab,t'} \end{aligned}$$

where $\gamma_{0,t'} = \alpha_{2,t'}\kappa_{0,t'}$, $\gamma_{1,t'} = (\alpha_{3,t'}\kappa_{0,t'}+\alpha_{1,t'})$, $\gamma_{2,t'} = \alpha_{2,t'}\kappa_{1,t'}$, $\gamma_{3,t'} = \alpha_{3,t'}\kappa_{1,t'}$, $\gamma_{4,t'} = \alpha_{4,t'}$, $\gamma_{5,t'} =$

$\alpha_{5,t'}$.

4.2 Identification and Estimation of the Wage Equation

We estimate the wage equation laid out in equations (4) and (5), but allow for i.i.d. measurement error in wages u_t. Using those equations and noting that $v_t = \delta_5 ab_4 + \sum_{k=5}^{t} \eta_k$ yields:

$$\ln w_t^* = \ln w_t + u_t = \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + \delta_5 ab_4 + \sum_{k=5}^{t} \eta_k + u_t \tag{13}$$

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.

We can address some problems of selectivity using our panel data. To address the issue of composition bias (the issue differential labor force entry and exit by lifetime wages), we use a fixed effects estimator. Given our assumption of a unit root in $v_t = \delta_5 ab_4 + \sum_{k=5}^{t} \eta_k$, which we estimate to be close to the truth, we can allow v_4 (the first period of working life, age 16) to be correlated with other observables, and estimate the model using fixed effects. In particular, we estimate δ_1, δ_2, δ_3, δ_4 and an individual fixed effects use a fixed effects estimator:

$$\ln w_t^* \quad = \quad \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + FE + \xi_t$$

where $FE = \delta_0 + \delta_5 ab_4 + \eta_5$ is a person specific fixed effect capturing the time invariant factors and $\xi_t = \sum_{k=6}^{t} \eta_k + u_t$ is a residual. We then use a methodology similar to that described in section 4.1 to estimate δ_5 where we use multiple noisy measures of ability to instrument for each other. We then estimate the variances of the wage shocks $(\sigma^2_{\eta_5}, \sigma^2_{\eta})$ and the variance of the measurement error (σ^2_u) using an error components procedure.

The above procedure addresses problems of measurement error in ability as well as selection based on permanent differences in productivity but not selection based on wage shocks. We control for this last aspect of 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 (**??**)) 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 Method of Simulated Moments

We estimate the rest of the model's parameters (discount factor, consumption weight for both spouses, risk aversion, altruism weight, share of time with the child that represents leisure to the parent, the hours-to-latent investments conversions):

$$\Delta = (\beta, \nu_f, \nu_m, \gamma, \lambda, \theta, \{\kappa_{0,t'}, \kappa_{1,t'}\}_{\{t'=1,2,3\}})$$

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 we wish to understand the drivers of parental labor supply and time investments, we match employment choices for both spouses and also household time spent with children, by parents' age and education. Because we wish to understand the drivers of education and money transfers, we also match educational decisions, as well as cash transfers to children when the children are older. Because we wish to understand how households discount the future, we match wealth data. Finally, to understand the relationship between time and latent investments, we match observed hours spent with children and their observed ability. In particular, the moment conditions that comprise our estimator are given by

1. Employment rates, by age, gender, and education, from the NCDS data (30 moments)
2. Fraction in full time work conditional on being employed, by age, gender, and education, from the NCDS data (30 moments)
3. Mean annual time spent with children, by child's age and parent's gender and education, from the UKTUS data (18 moments)
4. Mean age at which individuals left fulltime education by fathers' education level from the NCDS data (3 moments)
5. Mean lifetime receipt of inter-vivos transfers, from ELSA (1 moment)
6. Median wealth at 60 from ELSA (1 moment)
7. Mean ability at age 16 by father's education (3 moments)

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.

The mechanics of our MSM approach are as follows. We simulate life-cycle histories of shocks to ability, wages, partnering and childrens' gender and ability for a large number of artificial individuals over multiple generations. Each individual is endowed with a gender and a value of the age-23 education, wealth, and partner characteristics drawn from the empirical distribution from the NCDS data. The initial stochastic component of wages v_5 is drawn from a parametric distribution estimated on the NCDS data (see, section 4.2).

Next, using value function iteration, we solve the model numerically. We solve backwards through time, embedding a backwards recursion over each lifecycle of multiple generations. Our solution concept involves finding a fixed point in decisions rules over generations. Using these decision rules, in combination with simulated endowments and the trajectories of shocks, we simulate the profiles of behavior for a large number of artificial households, each composed of a man and woman. The behaviors that we can simulate are those that our modelled agents decide: assets, work hours and time investments, 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 function. We search over the parameter space for the values that minimize this criterion. Appendix J contains a detailed description of our moment conditions, the weighting matrix in our GMM criterion function, and the asymptotic distribution of our parameter estimates.

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. We 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 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. Combining multiple test scores to create a measure of skills, we estimate a human capital production function using the methods described briefly in Section 4.1 and in more detail in the appendix.

We estimate equation (2) 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. Estimates are presented in Table 3 (Appendix F gives estimates of the initial ability draw). To ease interpretation, we normalize our ability and time measures to have unit variance in every period.

Table 3: Determinants of ability.

	Age 7	Age 11	Age 16
Lagged Ability	0.154	0.739	0.939
	[0.057, 0.251]	[0.696, 0.834]	[0.918, 0.993]
Investment	0.146	0.097	0.131
	[0.113, 0.171]	[0.079, 0.116]	[0.093, 0.161]
Lagged Ability $\times$ Investment	-0.021	0.040	-0.038
	[-0.067, 0.010]	[0.027, 0.068]	[-0.066, -0.009]
Mum: Medium Education	0.448	0.181	0.027
	[0.347, 0.552]	[0.109, 0.235]	[-0.026, 0.075]
Mum: High Eduction	0.593	0.414	-0.088
	[0.388, 0.776]	[0.292, 0.571]	[-0.242, 0.055]
Dad: Medium Education	0.472	0.262	0.056
	[0.252, 0.611]	[0.179, 0.321]	[0.002, 0.115]
Dad: High Education	0.401	0.460	0.107
	[0.313, 0.495]	[0.290, 0.548]	[0.010, 0.218]
Ability shock: $Var(u_{ab',t'})$	0.031	0.067	0.026

Notes: GMM estimates. Confidence intervals are bootstrapped using 100 replications. For the production function at age 7, we use ability measured at age 7 as a function of ability at age 0, time investments measured at age 7 (and referring to investments at age 0-6). For the production function at age 11, we use ability measured at age 11 as a function of ability at age 7, time investments measured at age 11 (and referring to investments at age 7-10). For the production function at age 16, we use ability measured at age 16 as a function of ability at age 11, 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. It shows that time investments have a significant effect on changes in ability over time, even after conditioning on background characteristics and initial ability. Evaluated at mean ability, a one standard deviation increase in time investments at age 0-6 raises age-7 ability by 0.15 standard deviations, a one standard deviation increase in time investments at age 7-10 raises age-11 ability by 0.10 standard deviations, and a one standard deviation increase in time investments at age 11 raises age-16 ability by 0.13 standard deviations. Ability is very persistent, especially at older ages, implying a high level of self-productivity.

Interestingly, the interaction between ability and investments is negative for age 7 and 16, 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 exhibits dynamic complementarity at this stage of childhood (as found by Cunha et al. (2010)). However, at all ages the extent of complementarity or substitutability is modest. For example, for those with age 7 ability one standard deviation below (above) average, a one standard deviation in investment delivers a 0.097-0.040=0.057 (0.097+0.040=0.137) increase in age 11 ability.

While the richness and breadth of our data allows us to account for measurement error in ability and investments, we do not believe our setting allows for credible exclusion restrictions that would allow us to account for the potential endogeneity of investments. The literature has not yet come to a consensus as to whether potential endogeneity would lead us to over- or understate the returns to investments. Attanasio et al. (2020) find that failure to account for endogeneity leads to an understatement of the returns to investments, whereas find that it leads to an overstatement of the returns.

We find that parental education strongly impacts future ability, providing empirical support for a key mechanism for perpetuating inequality across generations. High education parents are effective in producing human capital in their children (as also shown in some of the papers cited in Heckman and Mosso (2014) and is assumed in Becker et al. (2018) and Lee and Seshadri (2019)) in addition to having more resources to afford college. The high productivity of high education parents means that all else equal, their children will be of high ability. As we will show below, ability and years of education are highly complementary in the production of wages. The combination of these features of human capital production gives high education parents yet another incentive to send their children to higher education.

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

5.2 The Effect of Ability and Education on Wages

Our approach allows us to better understand whether differences in wages across individuals represents differences in ability versus shocks. In this section we give our estimates of the wage process shown in Section H for each gender and education group.

We allow the impact of ability on wages to depend on education to capture the possibility that returns to ability are greater for the more educated. Table 4 shows estimates of this impact (δ_5) for each gender and education group. These estimates show the log-point increase in wages associated with a one standard deviation increase in age-16 ability for each education and gender group. The extent of complementarity is similar to that estimated in Delaney (2019) and **?**, and is implicit in much of the literature on match quality (e.g., Arcidiacono (2005)) and college preparedness in educational choice (e.g., Blandin and Herrington (2018)).

Table 4: Log-point change in wages for a 1 SD increase in ability, by education level

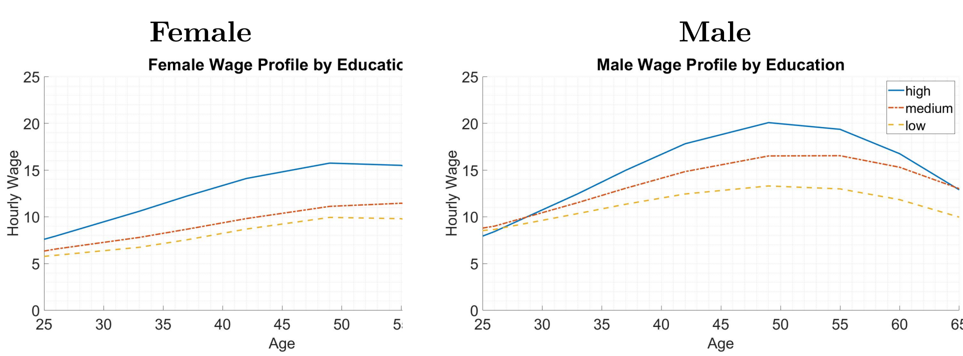
	Male	Female
Low	0.084 (0.025)	0.078 (0.024)
Middle	0.167 (0.019)	0.103(0.018)
High	0.205 (0.027)	0.127(0.027)

Notes: Cluster bootstrapped standard errors in parentheses (500 repetitions).

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 0.08 log-point increase in hourly wages for every additional SD of ability, high education men (with some college education) see an average increase of 0.21 log-points in hourly wages for every additional SD of ability. High educated women also receive greater returns to ability than low or middle educated women, although the gradient is more modest relative to that of men.

Figure 2 shows wage profiles by age, education and gender for full time workers with average ability and also ability that is one standard deviation above average. Men and those with high education have higher wages and faster wage growth.

Figure 2: Wages, by age, education and gender



Note: Wages measured in 2014 pounds. Wage profiles have been corrected for selection and are evaluated at mean ability.

As we show below this dynamic complementarity between ability and education has implications both for optimal time investments in children, and also for optimal educational decisions. Because of forward looking behavior, households who are more likely to invest in the education of their child have a stronger incentive to invest time in producing high ability children. Furthermore, those with high ability have an incentive to select into high education.

Turning to the variance of innovations to wages (σ_η^2), Table 5 shows that the estimated variance ranges from 0.0024 to 0.0048 implying that a one standard deviation of an innovation in the wage is 5-7% of wages, depending on the group. These estimates are similar to other papers in the literature (e.g. French (2005), Blundell et al. (2016)). Furthermore, we find evidence that the variance of wage innovations is increasing with education, implying that education is a risky investment.

Interestingly, we estimate the variance of the initial wage shock $\sigma^2_{\eta_5}$ to be small for all groups. While there is significant cross sectional variation in wages, even early in life, we estimate that most of that variation is explainable by our latent ability measure and measurement error in wages.

Table 5: Variance of innovations to wages, by education level

		Men	
	Low	Middle	High
σ^2_η	0.0024	0.0038	0.0045
	(0.0006)	(0.0006)	(0.001)
		Women	
	Low	Middle	High
σ^2_η	0.0020	0.0034	0.0048
	(0.0003)	(0.0004)	(0.0006)

Note: σ^2_η is the variance of the annual innovation to wages. Bootstrapped standard errors in parentheses.

In our formulation wage shocks have an autocovariance of one: wages are a random walk with drift. This implies ability has a permanent effect on wages. To test this restriction we also estimated versions of the wage process where we allowed the autocovariance to be less than one. However, we found little evidence against this restriction and thus use the more parsimonious formulation.

5.3 Marital Matching Probabilities

Table 6 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 6 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 6: 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 males in education group	0.26	0.64	0.11	

Notes: 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

5.4 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 4.69%, following Jordà et al. (2019). 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 members at each point in their life, and estimated a three-parameter tax system which varies across three different phases of life: young without children (ages 23-25), working adult (ages 26-64), pension age (age 65, onwards). This three parameter tax system has the following functional form: $y_t = d_{0,t} + d_{1,t}(e_{m,t} + e_{f,t} + e_{f,t} + e'_t)^{d_{2,t}}$. 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.

6 Second Step Results, Identification, and Model Fit

We now present the estimated structural parameters, how they are identified and the model's fit. Table 7 presents estimates from the structural model.

6.1 Utility Function Estimates and Identification

Table 7: Estimated structural parameters.

Parameter	Estimate
β: discount factor	0.985
	(0.0001)
ν_f: consumption weight, female	0.454
	(0.0002)
ν_m: consumption weight, male	0.433
	(0.0003)
γ: risk aversion	3.462
	(0.0078)
λ: altruism parameter	0.313
	(0.0010)
θ: time cost of investment	0.041
	(0.0002)
$\kappa_{1,1}$: latent investments per hour, ages 0-6	0.175
	(0.0006)
$\kappa_{1,2}$: latent investments per hour, ages 7-10	0.153
	(0.0010)
$\kappa_{1,3}$: latent investments per hour, ages 11-15	0.224
	(0.0009)
Coefficient of relative risk aversion, consumption*	2.092

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 $-(1/2)[(\nu_m(1-\gamma)-1)+(\nu_f(1-\gamma)-1)]$. β is an annual value.

The parameter γ is the coefficient of relative risk aversion (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 consumption and labor supply. The coefficient of relative risk aversion for consumption is 2.09 averaging over men and women,[7] which is similar to previous estimates that rely on different methodologies (see Browning et al. (1999) 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 savings over the life cycle using asset data as we do. 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

[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.

averse, they would save more in order to buffer themselves against the risk of bad income shocks in the future. We also obtain identification from labor supply since 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. Furthermore, 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.

Our estimate of the time discount factor β is equal to 0.985, and is also identified using our wealth data and our data on labor supply over the lifecycle, both of which suggest households are relatively patient. First, wealth holdings at age 60 are relatively high given pension benefits and high consumption demands up to this age. Second, 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. 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 share of total non-childcare hours devoted to time worked in the market. To see this note that, the marginal rate of substitution between consumption and leisure is approximately

$$\begin{aligned} w_{g,t}(1-\tau'_{g,t}) &\leq -\frac{\partial u_t}{\partial h_{g,t}} / \frac{\partial u}{\partial c_g} \\ &\leq -\frac{1-\nu_{g,t}}{\nu_{g,t}} / \frac{c_{g,t}}{l_{g,t}} \end{aligned} \tag{14}$$

which holds with equality when work hours are positive, where $\tau'_{g,t}$ is individual g's marginal tax rate at time t.[9] Inserting the time endowment equation (1) into equation (14) and making the approximation $c_{g,t} \approx w_{g,t} h_{g,t}(1-\tau'_{g,t})$ yields

$$\nu_g \approx \frac{h_{g,t}}{T - ti_{g,t}}. \tag{15}$$

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.

[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, an advantage of the dynamic programming approach is that it is not necessary to assume certainty, linear budget sets, or interior conditions.

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

Our estimate of the weight that the altruistic parents place on the utility of both their children (2λ) is 0.63 which is the middle of the range of estimates reported in the literature. This is higher than estimates by Daruich who estimated it to be 0.48 and Lee & Seshadri who estimate it to be 0.32, and lower than Gayle, Golan, and Soytas whose estimate is 0.80 and Caucutt & Lochner whose estimate is 0.86. These papers model a parent with only one child, whereas in our framework a parent has two children. Thus we multiply by 2 the continuation values of the children.

The parameter λ is identified from two sources. First, 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. Second, 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

$$\frac{\partial u_t}{\partial c_{g,t}} \geq \frac{2\lambda \partial \mathbb{E}_t V'_{t'}(\mathbf{X'}_{t'})}{\partial A'_{t'}} = \frac{2\lambda \mathbb{E}_t \partial u'_t}{\partial c'_{g,t'}}$$

that holds with equality if transfers are positive. The term on the right is the sum (over both children) of the childrens' expected marginal utility of consumption value of assets, which the parent can transfer to the child when the child is age 23. At the time of the transfer the children will be at a low earning time during their life cycles, and will soon have their own children and the time and money expenses of those children. This, and the fact that they are likely to be borrowing constrained, will mean they will a higher marginal utility of consumption than their parents. In order to rationalize relatively modest, transfers to children λ must be less than 1. Nevertheless, the fact that these transfers are made is perhaps the strongest evidence that $\lambda > 0$ and households are altruistic. Furthermore, in Section 7.1 we show that the returns to education average 7.4% per year of education, which is well above the market interest rate of 4.7%. Recall that the returns to education accrue to the child, whereas the return to cash accrues to the parents. The fact that many parents do not invest in their childrens' education, but some do, again provides evidence that λ is less than 1 but is greater than 0.

The parameter θ is identified by the relative productivity of time investments with children. Recall that $1-\theta$ is the share of time with the child that represents leisure to the parent: if $\theta = 1$ then time with children has the same utility cost as work, whereas if $\theta = 0$ then time with children has the same utility benefit as leisure. Thus, if $\theta = 1$ optimal behavior implies that the economic benefit of an additional hour of investment in the child (i.e., the increase in the expected present value of the childrens' lifetime income) will be (approximately) equal the economic benefit of an additional hour of work (the parent's wage). Conversely, if $\theta = 0$ then parents will spend time with their children even if it does not affect the childrens' future wages. Appendix L provides a more formal discussion of identification of θ. Because we find that the impact of parents' time on childrens' ability is positive but modest, we estimate θ to be 0.04, meaning that 96% of the time that parents spend with their children is leisure for them. There is little evidence on the magnitude of this parameter. The closest study to ours is Daruich (2018) who uses a specification slightly different than ours, but also finds that time spent with children is largely leisure.

The $\kappa_{1,t'}$ parameters govern the relationship between units of latent investments and units of time. Identification of these parameters comes from that we observe gradients in time investments (from the UKTUS data) by parental education and that we also observe corresponding ability gradients from the NCDS. That is, we observe more educated parents spending more time with their children, as well as a gradient in final ability by parental education. This, together with the production function in first stage pins down how a unit of time maps into a unit of investments. Appendix L contains a more formal derivation.

6.2 Model Fit

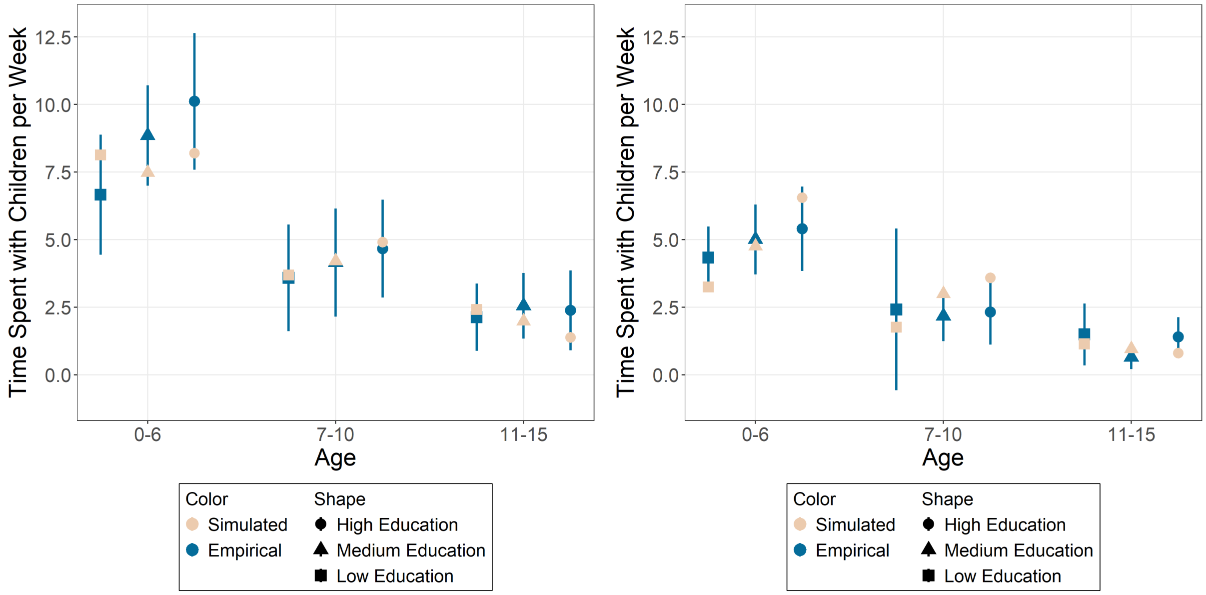
In this section we focus on the moments that are critical for understanding intergenerational altruism: transfers of time, educational investments, and money.

Figure 3 shows transfers of time from mothers and fathers in the left and right panels, respectively. The model fits three key patterns in the data well. First, time investments decline with age. Second, mothers invest more in their children than fathers. This higher rate of investment reflects the lower wage, and thus the lower opportunity cost of time for women. Third, high education parents invest more time in their children than low education parents. This pattern is driven by a combination of the higher education levels of their children and the complementarity between ability and education in wages that we have estimated.

This higher level of time investments of educated parents, in combination with their greater productivity of these investments, leads to higher ability of their children as can be seen in panel (a) of Figure 4. Our model captures well how higher time investments of the educated lead to higher ability of their children. Children of low education fathers have ability that is 0.14 standard deviations below average, whereas children born to high education fathers have ability that is 0.80 standard deviations above average. Our model matches these patterns well, although we slightly overstate the gradient.

Next, panel (b) of Figure 4 shows children's education, by father's education. Although the model slightly underpredicts educational attainment of children, it captures the gradient of children's education by parent's education. The difference between the average age left school of the children with high educated fathers and those with low educated fathers that our model predicts is 1.12 years, close to the

Figure 3: Model fit: parental time with children



(a) Mother's Time with Kids

(b) Father's Time with Kids

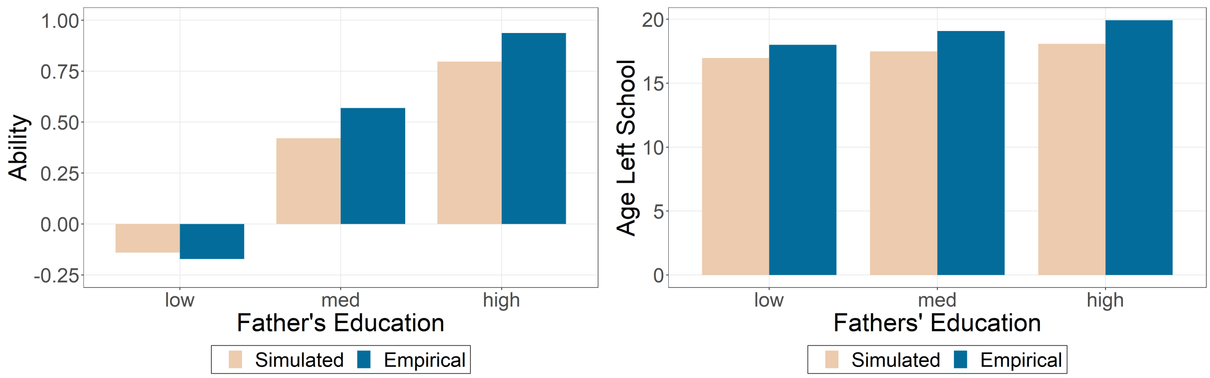
Notes: Measures of educational time investments. Source: UKTUS. See Appendix C.3 for details.

difference of 1.92 years found in the data.

Table 9 shows that we match well the mean level of financial transfers received and the median level of assets at age 60. These financial transfers include inter-vivos transfers when younger and bequests received when older. These amounts are discounted to age 23: when undiscounted, the amounts are considerably larger. In the data, as in the model, median transfers are 0. Thus we match mean transfers.

Finally, our model can reproduce key labor supply moments of men and women with different education levels as shown in Appendix K. Both female labour force participation and fulltime work conditional on employment are slightly overpredicted in the model. However, the model does well in generating a dip in female participation and fulltime work between ages 33 and 48 (when children are in the household). Moreover, as in the data, the model predicts higher participation rates for more educated women at older ages. For men, the model does well in generating a level of labour supply that is consistent with the data both on the intensive and the extensive margin.

Figure 4: Model fit: education and ability



(a) Ability by father's education

(b) Education choices by father's education

Notes: Empirical education and ability from NCDS data.

Table 8: Model fit: transfers and assets

	Empirical	Simulated
Mean transfers	£12,900	£12,800
Median Assets	£306,400	£291,700

Notes: Values in 2014 GBP. Mean transfers and median assets calculated using ELSA data. Transfers include inter-vivos transfers and bequests and are discounted to age 23 at the real rate of return. See Appendix C for more details.

Table 9: Model fit: transfers and assets

	Mean transfers		**Median assets**	
Education	Empirical	Simulated	Empirical	Simulated
Low	£10,000	XXX	£144,100	XXX
Medium	£24,000	XXX	£316,000	XXX
High	£48,400	XXX	£468,200	XXX

Notes: Values in 2014 GBP. Mean transfers calculated by *father's* education. Median assets calculated by individual's *own* education. Empirical mean transfers and median assets calculated using ELSA data. Transfers include inter-vivos transfers and bequests and are discounted to age 23 at the real rate of return. See Appendix C for more details.

6.3 Intergenerational Persistence

Although we do not target them directly, our model fits well the intergenerational persistence in economic outcomes that are commonly estimated in other studies. This include the intergenerational correlation of education and the intergenerational elasticity (IGE) of lifetime earnings and consumption. We estimate the following regression on our simulated data: $y' = a_0 + a_1 y + u$ where y' denotes the child's outcome (e.g., the number of years of schooling or the log of childrens' household earnings) and y the parents' corresponding outcome (e.g. parents' years of schooling or lifetime household earning).

The model predicted correlation of childrens' and parent's education is 0.23, with a correlation with father's education of 0.19 and mothers of 0.18, which is similar to the estimates presented in Hertz et al. (2007) who report 0.31 for Great Britain. The model predicted intergenerational elasticity of lifetime household earnings is 0.24, which is similar to the estimated values reported in Belfield et al. (2017) and Bolt et al. (2021). A more complete measure of lifetime resources is consumption. The model predicted intergenerational elasticity of consumption is 0.51 which is in line with the findings in Gallipoli et al. (2020) who find an average consumption IGE in the PSID of 0.46, a value that is substantially above their estimates of the elasticity of earnings. Wealth transfers across generations cause consumption to be more persistent than earnings. That our model reproduces key patterns of intergenerational persistence gives us additional confidence in its use for evaluating the drivers of this persistence and policy counterfactuals.

Table 10: Intergenerational Persistence

Outcome	Model-Implied	Literature
Intergenerational Correlation, Education	0.23	Hertz (2007) $\approx$ 0.3
Intergenerational Elasticity, Earnings	0.24	Dearden et al. (2007), Bolt et al. (2021) $\approx$0.3
Intergenerational Elasticity, Consumption	0.51	Gallipoli et al. (2022) $\approx$0.5 (in the US)

Notes: Intergenerational correlations and elasticities calculated from model simulated data. Earnings and consumption calculated as average over ages 23-65.

7 Results

7.1 The Returns to Education

In this section we present model predicted returns to education. In our model the return to education is heterogenous across the population since wages depend on ability, education, and their interaction, among other variables. We measure the return to education for different groups by exogenously changing the education levels of agents in the model, then calculating the resulting annualized percent change in lifetime wages from age 23-65. We take as given the age-23 joint distribution of the state variables of the NCDS sample members, draw histories of shocks, and calculate optimal decisions for both NCDS sample members and their children. In order to measure the childrens' return to education, we simulate their childrens' lifetime wages twice: first if they receive high education then if they receive low education. In column (1) of Table 11 we assume that the education level is unanticipated: the household's decision rules are thus calculated assuming that the household (erroneously) believes they can choose the child's education level. Thus, in this first experiment, changing education holds constant the decision to invest in child's ability. If everyone received low education, average lifetime wages (i.e., their pre-tax earning if they worked full time) would be £382,000. Conversely, if everyone selected high education, wages would be £563,000, a difference of 47.3%. Given the 5 year difference in schooling between low and high educated individuals, annualizing this translates to a 8.1% increase in lifetime wages per year of education: education is a lucrative investment. To put the size of education transfers in context: the increase in lifetime earnings from moving from low to high education (£181,000) is significantly higher than the average cash transfer to children reported in Table 9 (£13,000).

Section 5.2 presents evidence of dynamic complementarity: the returns to education are higher for those with high ability. This has two implications for economic behavior that are evident in Table 11.

First, it provides an incentive for high ability individuals to self-select into education. To measure the amount of the resulting self-selection, we calculate the return to education for two groups: those who in the baseline case select high (college) education, and for those who in the baseline case select low (compulsory) education. The bottom panel of Table 11 shows that the return to education is higher for those who would have selected high education (8.5%, which is the treatment effect on the treated) than the return for those who select low education (7.9%, which is the treatment effect on the untreated). Complementarity between ability and education, in combination with self selection in the model, explains this result.

Table 11: Returns to education.

	Unanticipated (1)	Anticipated (2)
	A: Full Sample	
Lifetime earnings, if low education	382,000	362,000
Lifetime earnings, if high education	563,000	595,000
Return	47.3	64.4
Annualized return	8.1	10.5
	B: By Baseline Education Choice	
Annualized return		
... among those who selected high education	8.5	11.9
... among those who selected low education	7.9	9.6

Notes: The return R is calculated as percent change in discounted pre-tax lifetime wage earnings between having high and low education. Annualized returns equal $(1+R)^{1/5}-1$. Anticipated means that the household is certain that it will be forced to either have high or low education starting from birth. Unanticipated means that the household believes it will make the optimal educational choice at age 16 and makes time investments given the belief of optimal educational choice, then is forced to either have high or low education.

Second, if parents are forward-looking their investment decisions will depend on the probability their children continue education. As with column (1), column (2) solves the model both for the case where children attain high education and for the case where children attain low education, and reports the resulting return to education. However, in column (2) households are certain of their childrens' future education level. This means households can change their time investment and other decisions in response to the education change. When households are certain their children will attain high education, they respond by increasing time investments, since the return to child investment is now higher. Column (2) shows that when allowing for these anticipation effects, the return to education rises from 8.1% to 10.5% once households anticipate this higher level of education. This highlights the importance of pre-announced policies that can deliver higher returns than policies that are not pre-announced. Pre-announcing the policy allows parents to adjust their time investments accordingly.

7.2 How is Income Risk Resolved over the Life Cycle?

How much of the cross-sectional variance in lifetime income can we predict using information known at different ages? Already before birth, information on the parents can help us predict an individual's lifetime income via predicted future investments that parents will make as well as through the productivity of those investments. As the child is born and grows older, more and more decisions are made and shocks are realized, thus increasing the extent to which lifetime income can be predicted.

As in previous sections, we take as given the age-23 joint distribution of the state variables of the NCDS sample members, and solve and simulate their choices and their children, including the childrens' lifetime income when they become adults. Next, we calculate the share of the variance in their lifetime income that can be predicted by the following variables that are known at each age: parental assets, wages, and education; the child's ability, gender, education, and wages; and the education and wages of the child's spouse. This approach allows us to decompose the relative importance of (predictable) circumstances and choices; the remainder being explained by shocks. This builds upon the approach in Huggett et al. (2011) and Lee and Seshadri (2019), who calculate the share of lifetime income known to the individual at age 23 and age 0, respectively. By showing the amount of lifetime income variability known at multiple ages, we illustrate how this uncertainty is resolved with age.

Our decomposition makes use of the law of total variance: a random variable can be written as the sum of its conditional mean plus the deviation from its conditional mean. As these two components are orthogonal, the total variance equals the sum of the variance in the conditional mean plus the variance around the conditional mean. We divide the variance in the conditional mean of lifetime income by the total variance of lifetime income, and report this for three measures of household resources in Table 12.

The first of these is individual lifetime wages which are calculated as the discounted pre-tax earnings

Table 12: Explained Outcome Variance: Evolution Over Childhood

Parents' age	23	26	33	37	42	49	+ Spouse	+
Child's age	NA	0	7	11	16	23	Ed, Wage	Transfer
Male Children								
Individual's wage	30%	37%	42%	46%	48%	65%	65%	
Household's wage	26%	32%	38%	41%	44%	57%	64%	
Household's income	23%	30%	35%	38%	41%	53%	61%	62%
Female Children								
Individual's wage	13%	15%	20%	23%	25%	45%	45%	
Household's wage	4%	5%	7%	8%	9%	17%	64%	
Household's income	5%	7%	10%	11%	13%	22%	61%	62%

Notes: Share of variance of given variables explained by all variables of both parent and child known at a given age. Wages, earnings, and income are discounted pre-tax values received between ages 23 and 54. Wages here are measured as the potential earnings if working full time. Household income is the sum of earnings plus parental transfers received by both the individual and their spouse.

between ages 23 and 64 that an individual would earn if they worked full time in every period. These numbers reflect the difference in *potential* rather than realized earnings, which depend on labor supply choices. The table shows that 30% and 13% of lifetime wages are known for males and females, respectively, even before they are born. These shares are explained by parents' education (which affects initial ability and the productivity of parental investments) and also household financial resources (which affects the quantity of investments the child receives). As the child ages, new information is realized, both about their own ability and their parents' financial resources. Immediately after birth, initial ability and parental wage shocks are revealed, which leads to a rise in the shares explained to 37% and 15% for males and females, respectively. By the time the children are aged 23, educational choices have been made, and their initial wage draw has been realized, and these shares are 65% and 45%. Thus close to half of lifetime wage variability is realized by age 23. The higher share of lifetime wages that is explainable for men reflects the higher return to ability for men, especially those who obtain a high level of education.

Our second measure of resources is household wages, the sum of lifetime wages for both spouses. At age 23 individuals marry, resolving uncertainty about age 23-characteristics of the spouse (their wage, education, and parental transfer). Before matching occurs, household wages (the sum of lifetime wages for both self and spouse) are less explainable than individual wages since matching is not perfectly assortative. Just before marriage, the share of lifetime household wages explained is 57% and 17% for males and females, respectively. The share explained is much lower for women than for men because wages are both lower and less variable for women than men. The final column shows that marriage explains much of the remaining variability in outcomes, especially for women: the share of the variability in lifetime household income explained jumps to 64% for both men and women after marriage. Household wages are less explainable than individual wages before marriage, but are more explainable afterwards. That is, before marriage, the characteristics of one's future spouse is an important risk; after marriage one's spouse becomes an important form of insurance.

Our third measure of resources is household lifetime income, which is the sum of realized earnings and parental transfers received by both self and spouse. Household income is about as explainable as the household wage. The final column includes parental transfers. Transfers explain little of lifetime income, both because transfers are small relative to lifetime earnings and also because the transfers made are highly explainable given all the other variables known.

7.3 What Explains Income Inequality?

The previous section showed how uncertainty is resolved over childhood as shocks are realized and choices are made. However, it does not show the relative importance of the different parental choices which lead to intergenerational persistence in outcomes. This section shows the relative roles of different types of parental transfers in contributing to variability in lifetime income. To address the importance of choices relative to other variables, we perform counterfactual experiments where we hold all choices, of both the parents' and childrens' cohorts, constant except that we equalize in turn parental time investments, education, and money transfers. We evaluate individual lifetime wages, household lifetime wages, and household lifetime income for the childrens' cohort and report the proportionate fall in variance that these equalizations would induce.

Table 13: Fraction of outcome variance for males explained by time investments, education, and ability

Equalise:	Final Ability	Education	Time Investments	Transfers
Individual's wage	18%	8%	13%	-
Wage of household	19%	16%	16%	-
Household's income	16%	14%	11%	4%

Notes: Percentage reduction in variance of variables when equalizing a channel to its model median for education and means otherwise. Wages, earnings, and income are discounted pre-tax values received between ages 23 and 64. Wages here are measured as the potential earnings if working full time. Household income is the sum of earnings plus parental transfers received by both the individual and their spouse.

Table 13 shows that equalizing time investments received would reduce the variance of individual lifetime wages by 13%. Equalizing educational investments would have a smaller impact, reducing this variance by 8%. However, the proportion of *household* wages explained by time investments and education is greater. This is because equalizing education not only removes the variation in household wages coming from the education of the spouses, but also removes the additional variation across households due to assortative matching. This highlights the interplay between education, the family, and lifetime risks:

because highly-educated individuals are more likely to marry other highly-educated individuals, inequality in education contributes more to variability in household wages than it does to individual wages. Assortive matching amplifies inequality.

The final row of Table 13, considers household income, which includes transfers as well as labor earnings. Equalizing transfers across households would reduce the variability in income less than equalizing either of time investments or education. If all households received mean transfers, the variance of household income would fall by 4%; equalizing education and time investments would reduce this variance by more. Altonji and Villaneuva (2008) and Black et al. (2022) also emphasize the modest role that transfers play in lifetime inequality.

8 Conclusion

This paper estimates a dynastic model of parental altruism where parents can invest in their children through time, educational expenditures, and transfers of cash. We estimate human capital production functions and the effect of ability on wages using data from a cohort of children born in 1958, thus presenting the first results of an estimated model that links early life investments to late life earnings by estimating (rather than calibrating) ability and wage functions. Our model is able to replicate realistic patterns of intergenerational persistence in wages, earnings, wealth and consumption.

We find that 28% of the variance of lifetime wages can already be explained by characteristics of the parents before individuals are born. This is due to both - direct effects of parental characteristics on individual's ability, and also due to increased investments of higher educated parents. In terms of investments, we find evidence of dynamic complementarity between time and educational investments – the returns to education are higher for high ability individuals. We find that this is a potentially important mechanism in perpetuating intergenerational outcomes, as borrowing constraints prevent low-income families from investing in education, thus simultaneously reducing the incentive to invest in time.

References

Abbott, B., G. Gallipoli, C. Meghir, and G. Violante (2019). Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy*.

Agostinelli, F. and M. Wiswall (2016, July). Identification of dynamic latent factor models: The implications of re-normalization in a model of child development. Working Paper 22441, National Bureau of Economic Research.

Agostinelli, F. and M. Wiswall (2022, July). Estimating the technology of children's skill formation. Working Paper 22442, National Bureau of Economic Research.

Altonji, J. G. and L. M. Segal (1996). Small sample bias in gmm estimation of covariance structures. *Journal of Business and Economic Statistics 14*(3), 353–366.

Arcidiacono, P. (2005). Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica 73*(5), 1477–1524.

Becker, G. S., S. D. Kominers, K. M. Murphy, and J. L. Spenkuch (2018). A theory of intergenerational mobility. *Journal of Political Economy 126*(S1), S7–S25.

Belfield, C., C. Crawford, E. Greaves, P. Gregg, L. Macmillan, et al. (2017). Intergenerational income persistence within families. Technical report.

Belley, P. and L. Lochner (2007). The changing role of family income and ability in determining educational achievement. *Journal of Human capital 1*(1), 37–89.

Blandin, A. and C. Herrington (2018). Family structure, human capital investment, and aggregate college attainment.

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw (2016). Female labor supply, human capital, and welfare reform. *Econometrica 84*(5), 1705–1753.

Bolt, U., E. French, J. H. Maccuish, and C. O'Dea (2021). The intergenerational elasticity of earnings: Exploring the mechanisms.

Browning, M., L. P. Hansen, and J. J. Heckman (1999). Micro data and general equilibrium models. *Handbook of macroeconomics 1*, 543–633.

Cagetti, M. (2003). Wealth accumulation over the life cycle and precautionary savings. *Journal of Business & Economic Statistics 21*(3), 339–353.

Castaneda, A., J. Diaz-Gimenez, and J.-V. Rios-Rull (2003). Accounting for the u.s. earnings and wealth inequality. *Journal of Political Economy 111*(4), 818–857.

Caucutt, E. M. and L. Lochner (2020). Early and late human capital investments, borrowing constraints, and the family. *Journal of Political Economy 128*(3), 1065–1147.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity: The geography of intergenerational mobility in the united states. *Quarterly Journal of Economics 129*(4), 1553–1623.

Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education 1*, 697–812.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Daruich, D. (2018). The macroeconomic consequences of early childhood development policies. *FRB St. Louis Working Paper* (2018-29).

De Nardi, M. (2004). Wealth inequality and intergenerational links. *The Review of Economic Studies 71*(3), pp. 743–768.

Dearden, L., S. Machin, and H. Reed (1997). Intergenerational mobility in britain. *The Economic Journal*, 47–66.

Del Boca, D., C. Flinn, and M. Wiswall (2014). Household choices and child development. *The Review of Economic Studies 81*(1), 137.

Delaney, J. (2019). Born unequal? understanding the components of life time inequality. Technical report, ESRI working paper.

Duffie, D. and K. J. Singleton (1993). Simulated moments estimation of markov models of asset prices. *Econometrica 61*(4), 929–952.

French, E. (2005, 04). The effects of health, wealth, and wages on labour supply and retirement behaviour. *Review of Economic Studies 72*(2), 395–427.

Gallipoli, G., H. Low, and A. Mitra (2020). Consumption and income inequality across generations.

García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2020). Quantifying the life-cycle benefits of an influential early-childhood program. *Journal of Political Economy 128*(7), 2502–2541.

Gayle, G.-L., L. Golan, and M. Soytas (2018). What is the source of the intergenerational correlation in earnings. Technical report, Mimeo.

Guryan, J., E. Hurst, and M. Kearney (2008). Parental education and parental time with children. *Journal of Economic perspectives 22*(3), 23–46.

Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. *Annu. Rev. Econ. 6*(1), 689–733.

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

Jordà, Ò., K. Knoll, D. Kuvshinov, M. Schularick, and A. M. Taylor (2019). The rate of return on everything, 1870–2015. *The Quarterly Journal of Economics 134*(3), 1225–1298.

Lee, S. Y. and A. Seshadri (2019). On the intergenerational transmission of economic status. *Journal of Political Economy 127*(2), 000–000.

Mazumder, B. (2005). Fortunate sons: New estimates of intergenerational mobility in the united states using social security earnings data. *The Review of Economics and Statistics 87*(2), 235–255.

Newey, W. K. and D. L. McFadden (1994). Large sample estimation and hypothesis testing. In R. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics, Volume 4*. Elsevier, Amsterdam.

Pakes, A. and D. Pollard (1989). Simulation and the aysmptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

Powell, J. (1994). Estimation of semiparametric models. In R. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics, Volume 4*. Elsevier, Amsterdam.

Powell, M. J. (2009). The bobyqa algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge 26.*

Solon, G. (1992). Intergenerational income mobility in the united states. *The American Economic Review*, 393–408.

Tauchen, G. (1986). Finite state markov-chain approximations to univariate and vector autoregressions. *Economics Letters 20*(2), 177–181.

A Parameter definitions

Table 14 summarises the parameters that enter the model and which are introduced in the body of the paper (excluding the appendices).

Table 14: 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	Educational Attainment
λ	Intergenerational altruism parameter	a_t	Wealth
$1-\theta$	Share of investment time perceived as leisure	$w_{g,t}$	Wage
$\kappa_{0,t}, \kappa_{1,t}$	Time to investment conversion parameters		
	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 ($t = 10$)
δ_j	Wage profile parameters		
	Ability		**Utility function and arguments**
$ab'_{t'}$	Child's ability at t'	$u()$	Single period utility function
γ_j	Ability production parameters	$V_t(\mathbf{X_t})$	Value function (parenthood phase)
u_{ab}	Stochastic ability component	$\mathbf{X_t}$	Vector of all state variables
		n_t	Number of equiv. adults in household
		T	Time endowment
		$\mathbf{d_t}$	Vector of decision variables
	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
		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 15 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 15: 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										
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
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

Notes: 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 in our moment condition. 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.[10] 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 16: 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 activities ("most important activity you were doing at the same time"), as well as who it was carried out

[10] 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.

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: teaching the child, reading/playing/talking with child, 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 17: 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 Approach to Estimation of the Ability Production Function, Parental Investment Function, and Wage Function

E.1 Production Function

The production function for skills that we estimate is as specified in equation (11) in the main text.

$$ab'_{t'+1} = \alpha_{1,t'} ab'_{t'} + \alpha_{2,t'} inv_{t'} + \alpha_{3,t'} inv_{t'} \cdot ab_{t'} + \alpha_{4,t'} ed^m + \alpha_{5,t'} ed^f + u'_{ab,t'} \tag{16}$$

E.2 Measurement

We do not observe children's skills ($ab'_{t'}$), or investments (inv_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 $\omega \in \{ab, inv\}$ we observe:

$$Z_{\omega,t,j} = \mu_{\omega,t,j} + \lambda_{\omega,t,j}\omega_t + \epsilon_{\omega,t,j} \tag{17}$$

All other variables are assumed to be measured without error.

E.3 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 $(u_{inv,t}, u_{ab,t}, ed_f, ed_m)$.

E.4 Normalizations

As mentioned above, ability 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 latent factors to be zero which fixes the location of the latent factors. In all other periods, the mean of the latent factor for ability ab_t is allowed to be different from zero. Moreover, for each period, we set the scale parameter $\lambda_{\omega,t,1} = 1$ for one normalizing measure $Z_{\omega,t,1}$.

AW have shown that renormalization of the scale parameter $\lambda_{\omega,t,1} = 1$ can lead to biases in the estimation of coefficients in the case of overidentification of the production function coefficients when assuming that $\alpha_{1,t'} + \alpha_{2,t'} + \alpha_{3,t'} = 1$ in equation (16). This is not the case in our estimation as we do not assume $\alpha_{1,t'} + \alpha_{2,t'} + \alpha_{3,t'} = 1$ in when estimating equation (16). For more details, see Agostinelli and Wiswall (2016).

E.5 Intial Conditions Assumptions

Period 1 for the child and period 6 for the parent is the time of the child's birth. The mean of ab'_1, ed_f, ed_m and inv_6 are 0 by normalization and without loss of generality. ab'_1 depends on parents' education and is normally distributed conditional on parents' education.

E.6 Estimation

1. **Variance of latent factors**.

   Using equation (17) we can derive the variance of each of the latent factors:

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

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 $Var(\omega_0)$. Because 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 parental 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 $\lambda_{ab,t,1} = \lambda_{inv,t,1} = 1$ to set the scale of ab_t and of inv_t. For each other measure $j \neq 1$, and for $\omega \in \{ab, inv\}$, using equation (18) we can show that:

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

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 ab_1' and inv_6 to zero. Therefore:

$$\mu_{ab',1,j} = \mathbb{E}[Z_{ab',1,j}], \quad \mu_{inv,6,j} = \mathbb{E}[Z_{inv,6,j}] \tag{20}$$

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

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

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

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

It gives ability (or investment) plus an error rescaled to match scale of the ability (which is also the scale of ability 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:

$$ab'_{t'+1} = \alpha_{1,t'}ab'_{t'} + \alpha_{2,t'}inv_{t'} + \alpha_{3,t'}inv_{t'} \cdot ab_{t'} + \alpha_{4,t}ed_m + \alpha_{5,t}ed_f + u'_{ab,t'}$$

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

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

or

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

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 $\pi_{ab,t,j} \equiv \left(u_{ab',t'} - \tilde{\epsilon}_{ab',t',j} - \tilde{\epsilon}_{inv,t',j} - \tilde{\epsilon}_{inv,t',j} \cdot \tilde{\epsilon}_{ab',t',j} + \frac{\epsilon_{ab',t'+1,j}}{\lambda_{ab',t'+1,j}} - \alpha_{3,t'}(\tilde{Z}_{inv,t',j}\tilde{\epsilon}_{ab',t',j} + \tilde{Z}_{ab',t',j}\tilde{\epsilon}_{inv,t,j}) \right)$:

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

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

Initial ability at birth is a function of mother's education level, father's education level, and a shock. We estimate this function by running a minimum distance estimation of our initial ability measures (after adjusting for their different scales) on parental education dummies. We then estimate the variance of the shock analogously to Step 6 in the previous section. Table 18 shows the results of the minimum distance, and the variance of the initial ability shock.

Table 18: Initial ability regression

	Coefficient	SE
Mother's education		
Medium	0.092	(0.041)
High	0.079	(0.103)
Father's education		
Medium	0.066	(0.044)
High	-0.007	(0.081)
Constant	-0.038	(0.022)
Variance of shock	0.880	

G Signal to Noise Ratios

Note that using equation (17) the variance of measure $Z_{\omega,t,j} = (\lambda^2_{\omega,t,j})Var(\theta_{\omega,t}) + Var(\epsilon_{\omega,t,j})$, where $(\lambda^2_{\omega,t,j})Var(\theta_{\omega,t})$ comes from the variability in the signal in the measure and $Var(\epsilon_{\omega,t,j})$ represents measurement error, or "noise". The signal to noise ratios for measure $Z_{\omega,t,j}$ is calculated in the following way:

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

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 19 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 19: Signal to noise ratios: Ability measures

Age 0		*Age 7*		*Age 11*		*Age 16*	
birthweight	0.862	read	0.385	read	0.555	read	0.570
gestation	0.140	maths	0.335	maths	0.942	maths	0.713
		copy	0.259	copy	0.104		
		draw	0.281				

Table 20 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 20: 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	0.270	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		

Notes: 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 equations (4) and (5), but allow for i.i.d. measurement error in wages u_t. Using those equations and noting that $v_t = \delta_5 ab_4 + \sum_{k=5}^{t} \eta_k$ yields:

$$\ln w_t^* = \ln w_t + u_t = \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + \delta_5 ab_4 + \sum_{k=5}^{t} \eta_k + u_t \tag{25}$$

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. Given our assumption of a unit root in $v_t = \delta_5 ab_4 + \sum_{k=5}^{t} \eta_k$, 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: From equation (25) note that:

$$\ln w_t^* = \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + FE + \xi_t$$

where FE is a person specific fixed effect capturing the time invariant factors $\delta_5 ab_4 + \eta_5$ 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:

$$\begin{aligned} \widehat{FE} &\equiv \bar{\ln w_t^*} - \hat{\delta}_1 \bar{t} - \hat{\delta}_2 \bar{t^2} - \hat{\delta}_3 \bar{t^3} - \hat{\delta}_4 \bar{PT}_t \\ &= \delta_0 + \delta_5 ab_4 + \eta_5 \\ &= \delta_0 + \delta_5 \tilde{Z}_{ab,4,j} + \eta_5 - \delta_5 \tilde{\epsilon}_{ab,4,j} \end{aligned} \tag{26}$$

where $ab_4 = \tilde{Z}_{ab,4,j} - \tilde{\epsilon}_{ab,4,j}$ and where $\tilde{Z}_{ab,4,j}$ and $\tilde{\epsilon}_{ab,4,j}$ have been defined in equation (22). The above equation holds for all measures j. Although the estimated fixed effect, $\widehat{FE}$, is affected by variability in the sequence of wage shocks $\{\eta_t\}_{t=5}^{12}$ and measurement errors $\{u_t\}_{t=5}^{12}$, this merely adds in measurement error on the left hand side variable in equation (26). However, measurement error on the right hand side 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, we 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 obtain $\hat{\delta}_0$ and $\hat{\delta}_5$. 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. Our GMM procedure efficiently combines different measures of ability and yields consistent estimates of $\hat{\delta_0}$ and $\hat{\delta_5}$ even in the presence of measurement error in the ability measures.

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

Substituting a noisy measure of ability into the wage equation (**??**) yields

$$\ln w_t^* = \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + \delta_5 \tilde{Z}_{ab,4,j} + \sum_{k=5}^{t} \eta_k + u_t - \delta_5 \tilde{\epsilon}_{ab,4,j}$$

where we use the fact that $ab_4 = \tilde{Z}_{ab,4,j} - \tilde{\epsilon}_{ab,4,j}$ as defined in equation (22). Next we define a wage residual that will exist for each ability measure:

$$\widetilde{\ln w_{t,j}} \equiv \ln w_t^* - (\delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + \delta_5 \tilde{Z}_{ab,4,j}) = \sum_{k=5}^{t} \eta_k + u_t - \delta_5 \tilde{\epsilon}_{ab,4,j}$$

Note that from the measurement equation 17, $Var(\tilde{Z}_{ab,4,j}) = Var(ab_4) + Var(\tilde{\epsilon}_{ab,4,j})$, where we have previously estimated $Var(ab_4)$ using equation 18 and $Var(\tilde{Z}_{ab,4,j})$ 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:

$$Cov(\widetilde{\ln w_{t,j}}, \widetilde{\ln w_{t+k,j}}) = Var(\sum_{k=5}^{t} \eta_k) + \delta_5^2 Var(\tilde{\epsilon}_{ab,4,j})$$

$$Var(\widetilde{\ln w_{t,j}}) = Var(\sum_{k=5}^{t} \eta_k) + Var(u_t) + \delta_5^2 Var(\tilde{\epsilon}_{ab,4,j})$$

$$Var(\widetilde{\ln w_{t+k,j}}) = Var(\sum_{k=5}^{t+k} \eta_k) + Var(u_{t+k}) + \delta_5^2 Var(\tilde{\epsilon}_{ab,4,j}).$$

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).

H.1 Wage shock process estimates without imposing random walk

In section 5.2, we impose that wage shocks have an autocovariance of 1. Table 21 shows the coefficients and standard errors when we relax this assumption and allow the persistence parameter to be different from one. We also report results from an overidentification test statistic. To do this we initially regress log wages on age, education, ability and part time status as before, then estimate the process for the residuals using an error components model where we match the variance covariance matrix of wage residuals. When estimating, we allow for an AR(1) process with homoskedastic (i.e., with age-invariant variances) innovations and a transitory shocks in which we allow for heteroskedasticity. We have 5 periods of data, and thus 15 unique elements of the variance covariance matrix which we treat as moment conditions for each gender/education group. We estimate the variances of the transitory shocks (5 parameters), the initial variance of the AR(1) component, the variance of the AR(1) shocks, and ρ, meaning that we have 8 parameters to estimate and thus 15-8=7 degrees of freedom, meaning that under the null of correct model specification our test statistic should be distributed $\chi^2(7)$. Overall, the model fits the data well and we cannot reject the hypothesis of correct model specification for many groups. Perhaps more importantly, we can see that for all groups except low educated females, we cannot reject the hypothesis that the persistence parameter is 1. Even for this group, the value of $\rho = 0.94$. Thus throughout we assume $\rho = 1$ for all groups.

Table 21: Estimates for AR(1) process without random walk restriction

		Male	
Education:	Low	Medium	High
ρ	1.034	0.968	1.027
	(0.022)	(0.018)	(0.019)
Test stat:	10.74	12.96	36.60
		Female	
Education:	Low	Medium	High
ρ	0.940	0.985	0.971
	(0.029)	(0.023)	(0.023)
Test stat:	35.38	23.56	19.54

This table shows the persistency parameter for an AR(1) wage shock process when we relax the assumption that the process is a random walk. Bootstrapped standard errors are in parentheses. The rows entitled "Test stat" show the overidentification test statistic.

I 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 50 grid points for assets, 10 points for wage rates (for each spouse), 3 education levels 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 Tauchen (1986). 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 4 levels of working hours (non-employed, part-time, full-time, over-time) and in model period $t = 6, 7$ and 8 they can choose between six 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 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 J. Our method of simulated moments procedure delivers the model parameters that minimize a GMM criterion function, which we also describe in Appendix J.

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

J 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 $\Delta =(\beta, \nu_f, \nu_m, \gamma, \lambda, \theta, \{\kappa_{1,t'}\}_{\{t'=1,2,3\}})$. 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 periods: $t \in \{5, 7, 9, 10, 11\}$ (which corresponds to ages 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 ($6T$). In addition, from the NCDS we match age 16 ability and the mean education leaving age, conditional on father's education level (6 moment conditions).

From ELSA we match mean lifetime inter-vivos transfers received (1 moment) and also household median wealth at age 50 (1 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).

In the end, we have a total of $J = 86$ 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 in up to J separate moment conditions. 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 Pakes and Pollard (1989) and Duffie and Singleton (1993), 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;

$$\mathbf{D} = \frac{\partial \varphi(\Delta; \chi_0)}{\partial \Delta'}\Big|_{\Delta=\Delta_0} \tag{27}$$

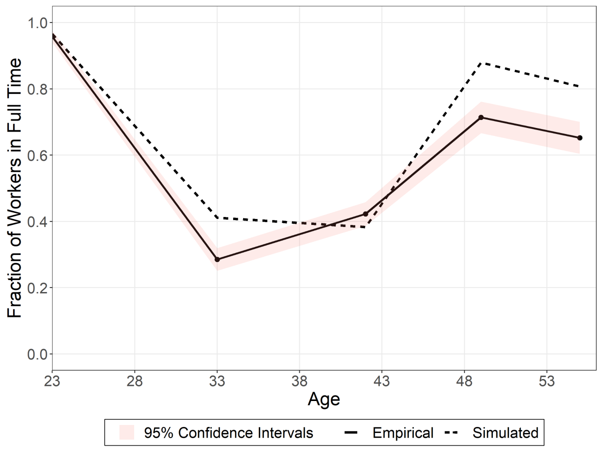
is the $J \times M$ gradient matrix of the population moment vector; and $\mathbf{W} = \text{plim}_{I \to \infty}\{\widehat{\mathbf{W}}_I\}$. The asymptotically efficient weighting matrix arises when $\widehat{\mathbf{W}}_I$ converges to $\mathbf{S}^{-1}$, the inverse of the variance-covariance matrix of the data. When $\mathbf{W} = \mathbf{S}^{-1}$, $\mathbf{V}$ simplifies to $(1+\tau)(\mathbf{D}'\mathbf{S}^{-1}\mathbf{D})^{-1}$.

But even though the optimal weighting matrix is asymptotically efficient, it can be biased in small samples. (See, for example, Altonji and Segal (1996).) We thus use a "diagonal" weighting matrix. This diagonal weighting scheme consists of the inverse of the moments along the diagonal, which will weight more heavily moments with low means so that they too will contribute significantly to the GMM criterion function, regardless of how precisely estimated they are.

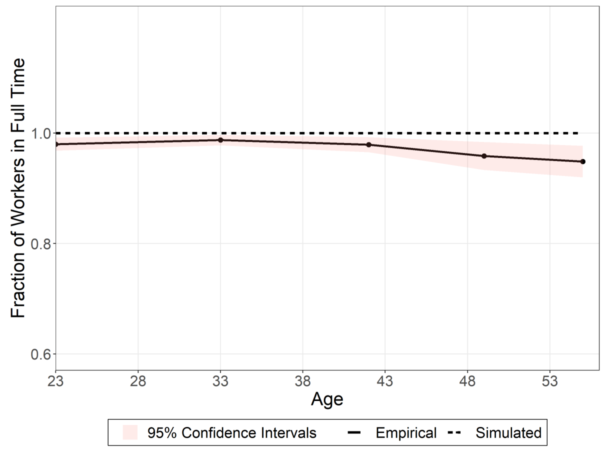
We estimate $\mathbf{D}$, $\mathbf{S}$, and $\mathbf{W}$ with their sample analogs. For example, our estimate of $\mathbf{S}$ is the $J \times J$ estimated variance-covariance matrix of the sample data. One complication in estimating the gradient matrix $\mathbf{D}$ is that the functions inside the moment condition $\varphi(\Delta; \chi)$ are non-differentiable at certain data points (e.g., for employment). 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 Pakes and Pollard (1989), Newey and McFadden (1994), and Powell (1994). When calculating gradients we vary step-sizes, then take the average gradient over the different step-sizes.

K Further Details on Model Fit

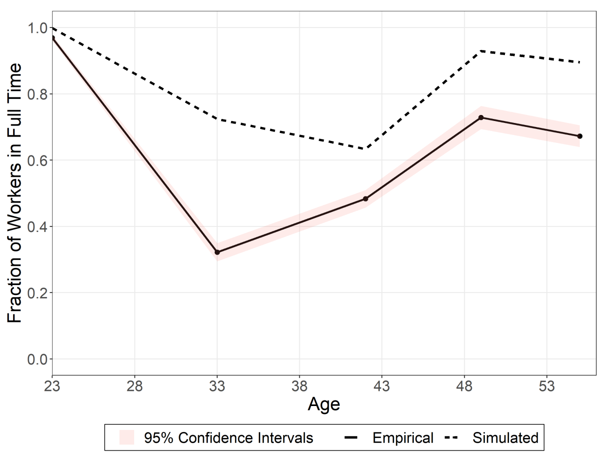
Figure 5: Model fit: full-time work conditional on employment



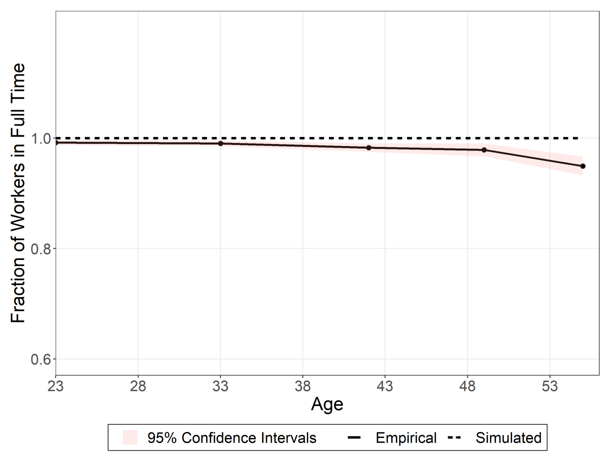
(a) Women: low educated



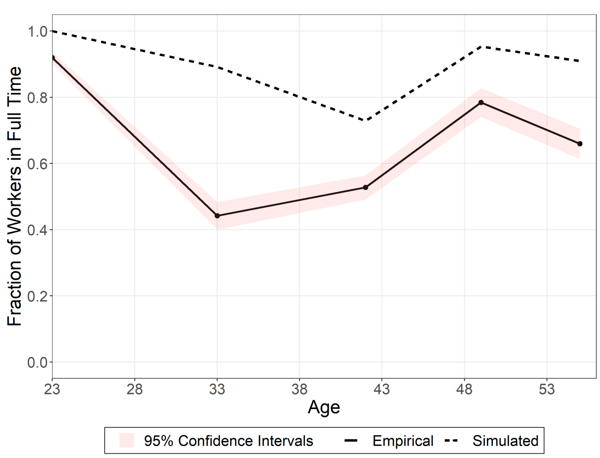
(b) Men: low educated



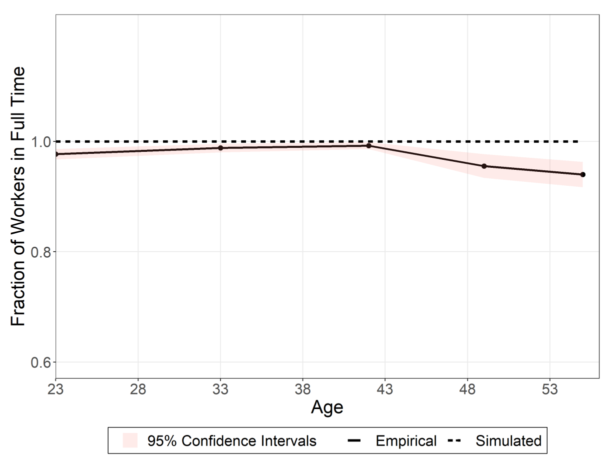
(c) Women: medium educated



(d) Men: medium educated



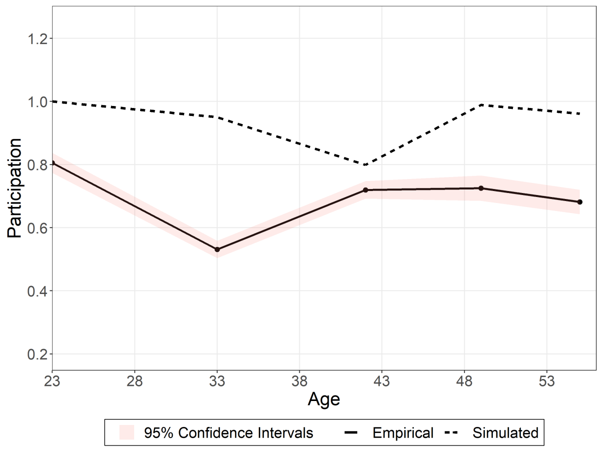
(e) Women: high educated



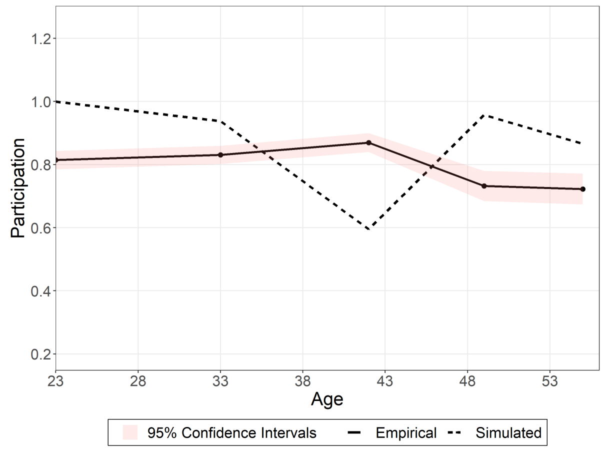
(f) Men: high educated

Notes: Figures show fraction in fulltime work at different ages conditional on being employed for women and men. Empirical data come from NCDS.

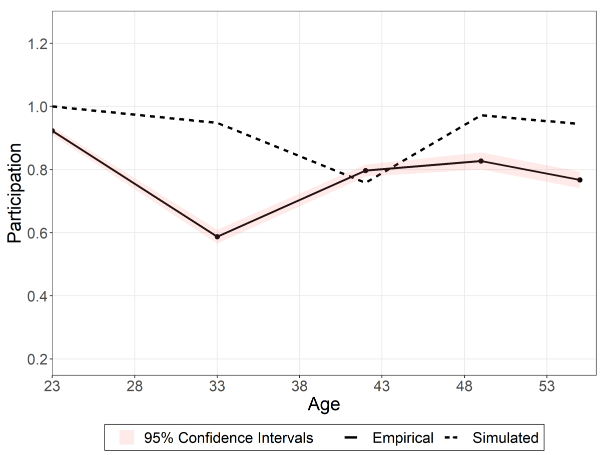
Figure 6: Model fit: participation



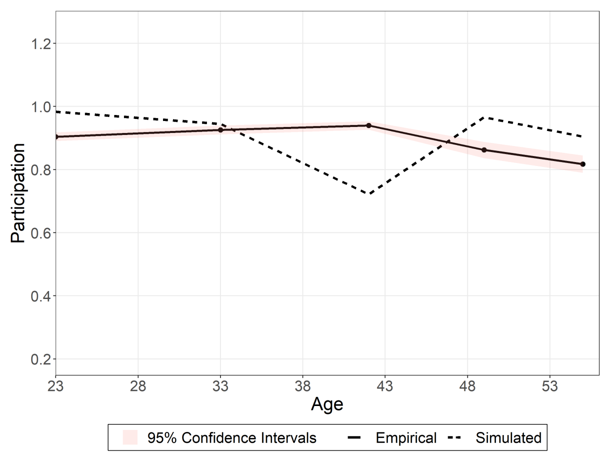
(a) Women: low educated



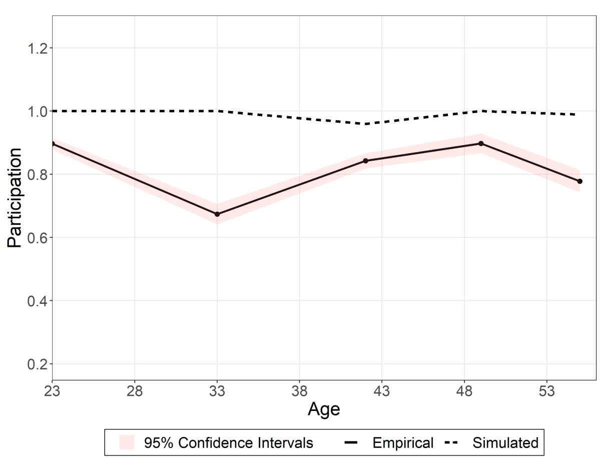
(b) Men: low educated



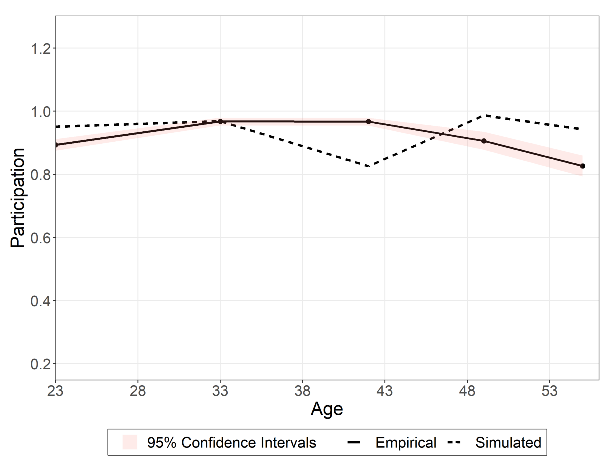
(c) Women: medium educated



(d) Men: medium educated



(e) Women: high educated



(f) Men: high educated

Notes: Figures show fraction of individuals employed at different ages for women and men. Empirical data come from NCDS.

L Identification of the time cost of investments θ

To give some intuition regarding the identification of θ, we use a simplified two period version of our dynastic model, where we abstract from couples, uncertainty, and where we additionally assume a linear production function. The household's state variables are: education ed, ability ab, and their initial assets a_1. The parent is altruistic towards their child and incorporate their child's value function into their problem, but discount it by factor λ. Households choose consumption c_t, leisure l_t, time investments ti_t, monetary transfers to their child x_1' and the education of the child ed' which can be dropout (D), high school (HS) or college (C). Each education choice is associated with a price p_k, $k \in \{D, HS, C\}$, which can be interpreted as the price of foregone labor earnings of the child. The child initially has no other assets than the monetary transfer from the parent. We first describe the discrete decision problem of the parent who selects their children's education level. They maximize their value function which nests the child's value function:

$$V(ed, ab, a_1) = \max_{ed'=\{D,HS,C\}} \{V_{ed'=D}, V_{ed'=HS}, V_{ed'=C}\} \tag{28}$$

where $V_{ed'=k}$ denotes the value of the problem if the parents choose education level k for their child. The above nests the following decision problem over consumption, leisure, time investments and asset transfers:

$$V_{ed'=k}(ed, ab, a_1) = \quad \max_{c_1,l_1,ti_1,x_1',c_2,l_2,ti_2} u(c_1, l_1) + \beta u(c_2, l_2) + \lambda V'(ed' = k, ab', a_1') \tag{29}$$

subject to:

$$(1+r)^2 a_1 + (1+r)h_1 w_1(ab, ed) + h_2 w_2(ab, ed) = (1+r)c_1 + c_2 + x_1' + \sum_{ed'=\{D,HS,C\}} p_k \mathbf{1}_{[ed'=k]} \tag{30}$$

$$ab' = \alpha_0 + \alpha_1 ti_1 + \alpha_2 ti_2 \tag{31}$$

$$T = \theta ti_1 + h_1 + l_1 \tag{32}$$

$$T = \theta ti_2 + h_2 + l_2 \tag{33}$$

$$a_1' = x_1', x_1' \geq 0 \tag{34}$$

where (30) describes the monetary budget constraint over 2 periods, (31) shows the human capital production function over two periods where α_1, α_2 are the productivity of time investments for final ability. (32) and (33) are the time constraints in period 1 and 2, and (34) states that initial assets equal the initial parental cash transfer.

Assuming interior conditions for the choice variables $\{c_1, l_1, ti_1, c_2, l_2, ti_2\}$ but allowing the constraint x_1' to bind we can now rewrite this problem and derive optimality conditions:

$$\begin{aligned}
V_{ed'=k}(ed, ab, a_1) = \quad & \max_{c_1,l_1,ti_1,x_1',c_2,l_2,ti_2} u(c_1, l_1) + \beta u(c_2, l_2) + \lambda V'(ab', ed', x_1') \\
& +\mu[(1+r)^2 a_1 + (1+r)h_1 w_1(ab, ed) + h_2 w_2(ab, ed) - (1+r)c_1 - c_2 - x_1' - p_k \mathbf{1}_{[ed'=k]}] \\
& +\kappa(\alpha_0 + \alpha_1 ti_1 + \alpha_2 ti_2 - ab') + \phi(x_1') \\
& +\zeta_1(\theta ti_1 + h_1 + l_1 - T) \\
& +\zeta_2(\theta ti_2 + h_2 + l_2 - T)
\end{aligned}$$

Euler equation: $\frac{\partial u}{\partial c_1} = \beta \frac{\partial u}{\partial c_2}(1+r)$

FOC wrt ti_1: $\zeta_1 \theta - \kappa \alpha_1 = 0$

FOC wrt ab': $\kappa + \lambda \frac{\partial V'}{\partial ab'} = 0$

$$\kappa + \lambda \mu'[(1+r)\frac{\partial w_1'(ab', ed'=k)}{\partial ab'} h_1' + \frac{\partial w_2'(ab', ed'=k)}{\partial ab'} h_2'] = 0$$

FOC wrt l_1:$-\zeta_1 + \frac{\partial u}{\partial l_1} = 0$

FOC wrt l_2:$-\zeta_2 + \beta \frac{\partial u}{\partial l_2} = 0$

FOC wrt h_1:$-\zeta_1 + \mu(1+r)w_1(ab, ed) = 0$

FOC wrt h_2:$-\zeta_2 + \mu w_2(ab, ed) = 0$

FOC wrt x_1':$-\mu + \lambda \frac{\partial V'}{\partial x_1'} = 0$

$$\mu = \lambda \mu'(1+r)^2 + \phi \Rightarrow \mu \geq \lambda \mu'(1+r)^2$$

From this, we can derive the following optimality condition for investments in period 1:

$$w_1(ab, ed)\theta \leq \alpha_1[\frac{1}{(1+r)^2} \frac{\partial w_{1'}'(ab', ed')}{\partial ab'} h_{1'}' + \frac{1}{(1+r)^3} \frac{\partial w_{2'}(ab', ed')}{\partial ab'} h_{2'}'] \tag{35}$$

This equation is key to understanding the identification of θ. On the left hand side, we have the marginal cost of investments to the parent which is their wage times θ – the amount of leisure they lose per hour of time spent with the child. On the right hand side, we have the marginal benefit of an hour spent with the child; this is the increase in the present discounted value of the child's future income from the hour of investment. The increase equals the productivity of an hour of time α_1, multiplied by the resulting marginal increase in income over the lifecycle to the child. If cash transfers are positive equation (35) holds with equality, although if cash transfers are 0 then it is an inequality. Dividing both sides by $w_1(ab, ed)$ shows that we can place an upper bound on θ by calculating the present values of the gain in child's lifetime income from one hour of time investment relative to the wage. In the appendix below we perform exactly this calculation.

L.1 Approximating the PDV of time investments

We estimate θ to be 0.049, implying that 95% of time spent with the child is leisure time. This is surprising, given that some studies have estimated sizeable returns to early life investments such as Perry Pre-School (e.g. García et al. (2020)). Furthermore, many studies have found positive returns to parental investment. To gain some intuition, we conduct a back-of-the envelope calculation using equation (35) that takes into account the production function of ability, the opportunity cost of time to parents, and the returns to time investments in the form of higher lifetime earnings of the child.

In the following calculation, we consider a family with a low educated father. We assume that at baseline, the amount of time that the family invests in their child is at the mean in each period and calculate the resulting ability at age 16. We also calculate the child's expected lifetime earnings, assuming that the child is male, has low education, and works 40 hours per week up until age 65.

We then consider the impact of one additional hour per week spent with the child in the first period of life (0-6). The resulting ability increase translates to a wage increase of 0.5% at each age, causing lifetime earnings to increase by £5,079 when not discounted, or by £1,141 when discounting back to age 6 (using an interest rate of r=0.0469).

To calculate the lifetime costs to the parents of increasing investments by 1 hour per week, we assume that parents have an hourly wage of £9.3, which is the average expected wage between ages 26-32 for a low educated male. They thus forgo $\pounds 9.3 \times 52$ weeks $\times\, 6$ years $= \pounds 2,902$ when they increase their investment by 1 hour per week for during the first childhood stage 1. Thus, the ratio of the NPV of returns to cost is $\frac{1,141}{2,902} = 0.39$.

M Identification of κ

Our structural model maps hours of parental time into future ability. However, the NCDS has latent investments and future ability. Here we show more on the mapping between hours of time and latent investments.

As described in section 4.1, we assume hours of parental time spent with children $ti_{m,t'}, ti_{f,t'}$ are converted to latent investment units $inv_{t'}$ according to equation (12) which we reproduce here:

$$inv_{t'} = \kappa_{0,t'} + \kappa_{1,t'}(ti_{m,t'} + ti_{f,t'})$$

where $\kappa_{1,t'}$ is the hours-to-latent investments conversion parameter which determines the productivity of an hour of time and $\kappa_{0,t'}$ is a constant. We allow the κ parameters to vary by age. With three investment periods and two parameters in each period, this gives us six parameters to estimate.

To gain intuition regarding the identification of the κ parameters, recall equation (**??**) which shows the relationship between time investments and ability:

$$\begin{aligned} ab'_{t'+1} &= \alpha_{1,t'}ab'_{t'} + \alpha_{2,t'}inv_{t'} + \alpha_{3,t'}inv_{t'} \cdot ab_{t'} + \alpha_{4,t'}ed^m + \alpha_{5,t'}ed^f + u'_{ab,t'} \\ &= \alpha_{1,t'}ab'_{t'} + \alpha_{2,t'}(\kappa_{0,t'} + \kappa_{1,t'}ti_{t'}) + \alpha_{3,t'}(\kappa_{0,t'} + \kappa_{1,t'}ti_{t'}) \cdot ab'_{t'} + \alpha_{4,t'}ed^m + \alpha_{5,t'}ed^f + u'_{ab,t'} \end{aligned}$$

The top line is the estimating equation using the NCDS data: the α parameters are estimated using the latent investment, ability, and parental education measures in the NCDS data. The κ parameters are estimated within the dynamic programming model. Identification of the κ_1 parameters comes from the gradients in time investments $ti_{m,t'} + ti_{f,t'}$ (from the UKTUS data) by parental education and the corresponding ability gradients (from the NCDS). All the α parameters and parental education ed^m, ed^f are known. From UKTUS we know that at each age high education parents spend more time with their children than low education parents and from the NCDS we know that at each age the children of high education parents have higher ability levels, even after controlling for the direct effect of parental education on ability: $\alpha_{4,t'}ed^m + \alpha_{5,t'}ed^f$. κ_1 thus captures how differences in time investments by parental education translate into differences in ability, controlling for parental education. The κ_0 parameters allow us to match mean time investments at different ages, observed in the UKTUS. The means of hours of time is positive, whereas the mean of latent investment is 0.

N Updating the matching probabilities in counterfactuals

In Section 5.3, we show that marital matching probabilities depend on the education level of the male and the female. These probabilities reflect the prevailing distribution of education levels in the population for the cohort we study. When we turn to our policy analysis, we must account for the fact that, in counterfactual settings, the distribution of education levels in the population may change, which will lead to changes in the marital matching probabilities. We account for this in our counterfactuals by allowing matching probabilities to depend on population education shares.

We estimate these matching probabilities as a function of the distribution of education levels observed in the population using data from the Family Expenditure Survey (FES) and its successor surveys from 1978 to 2017. During this time, there were major changes in the distribution of education, both for men and women. For example, the share of women with high education increased from less than 10% in 1987 to more than 40% in 2017. We use these data to estimate the following ordered logit model where for each gender and education level, we estimate the probability of matching with someone of the other gender with a certain education level, conditional on the distribution of education in the population of both genders. For example, we estimate the probability that an individual of education level $ed = j$ (e.g, a low educated male) partnering with an individual of education level $ed = i \in \{$ low, medium, and high educated female $\}$ as:

$$\begin{aligned} p_{ij} = Pr\left(y_j = i\right) &= Pr\left(\kappa_{i-1,j} < \mathbf{x}\boldsymbol{\beta}_j + u \leq \kappa_{i,j}\right) \\ &= \frac{1}{1+\exp\left(-\kappa_{i,j} + \mathbf{x}\boldsymbol{\beta}_j\right)} - \frac{1}{1+\exp\left(-\kappa_{i-1,j} + \mathbf{x}\boldsymbol{\beta}_j\right)} \end{aligned} \tag{36}$$

where $\mathbf{x}\boldsymbol{\beta}_j = \beta_{1,j} S_{m,low} + \beta_{2,j} S_{m,medium} + \beta_{3,j} S_{f,low} + \beta_{4,j} S_{f,med}$, $S_{g,ed}$ denotes the share in the population who are in gender group g and education group ed and the $\kappa_{i,j}$ parameters are the estimated thresholds for each group. In our dynastic model, any given policy environment generates population shares $S_{g,ed}$. These can be used with the parameters estimated here to deliver the matching probabilities that characterize the marriage market under the new equilibrium.