Essays on the Development of Inequalities over the Lifecycle

Helena Uta Bolt

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Economics

University College London

June 3, 2022

Declaration of Authorship

I, Helena Uta Bolt, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

[redacted]

Statement of Conjoint Work

Note on the joint work in Helena Uta Bolt's thesis "Essays on the Development of Inequalities over the Lifecycle":

The chapters, "The Intergenerational Elasticity of Earnings: Exploring the Mechanisms" and " Parental Altruism and Transfers of Time and Money - A Lifecycle Perspective" are co-authored work between Helena Uta Bolt, Eric French, Jamie Hentall Maccuish and Cormac O'Dea. Each author contributed equally. The chapter, "What is the Source of the Health Gradient? The Case of Obesity" is single authored by Helena Uta Bolt. "The Intergenerational Elasticity of Earnings: Exploring the Mechanisms" is currently an IFS working paper (21/07). An earlier version of "Parental Altruism and Transfers of Time and Money - A Lifecycle Perspective" was published as a Michigan Retirement Research Center Working Paper 2018-379.

[redacted]

Professor Eric French, Primary Supervisor

Abstract

Why do individuals who grew up in low income families tend to have lower income themselves? Why do they also tend to be unhealthier than those who grew up in richer families? This thesis studies how inequalities in income and health develop from childhood to adulthood. The first chapter quantifies the effect of various channels - education, cognitive skills, parental investments and family background - on intergenerational earnings persistence in Great Britain. We find that earnings persistence is largely driven by differences in parental investments during childhood, which encourage greater cognitive development and lifetime earnings. The second chapter jointly studies different channels through which parents invest in their children - by spending time with them to foster cognitive skills, by paying for their education, and by making monetary transfers. 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. The final chapter studies health inequality. I find that 34% of the differences in obesity rates between those coming from rich versus poor families is driven by differences in investments over the lifecycle. Differences in health investments received from the parents during childhood matter more than differences in investments in adulthood. A policy counterfactual in which the government directly invests in low income children's health reduces obesity rates and has substantial welfare effects.

Impact Statement

The topic of this thesis is the development of inequalities over the lifecycle. In particular, I study how parents invest in their children, how these investments shape a child's human capital and ultimately define their outcomes over the lifecycle. Early child development and its links to later life inequality has been the focus of recent policy debate (e.g. House of Commons debate, 9th June 2021). Especially in the light of the recent Covid 19 pandemic, concerns have been raised that differential investments by parents may lead to a long-run increase in inequality - both in income and health (Anders et al. (2020), Haelermans et al. (2022), Blundell et al. (2020)). This thesis thus is relevant not only for researchers, but also for policymakers wishing to reduce inequality. The three chapers each make a contribution to our understanding of how inequality evolves over the lifecycles, and what impact policy can have.

The second chapter answers the question of why rich parents tend to have rich children. There are many factors which affect social mobility, but few papers so far jointly quantify them. Knowing which factors matter the most is important for future research and policy makers alike. We find that differences in investments that parents make are a key driver of the intergenerational elasticity of earnings.

The next chapter studies investment decisions by parents in more detail. One key finding of chapter three is that there is a strong complementarity between educational investments and time investments. Those that end up going to university benefit the most from higher cognitive skills. This also shines light on one of the key drivers of differences in time investments. If parents know that they will never be able to afford to send their child to university, there is also less reason to ensure that the child will have high cognitive skills. This finding has direct implications for policy. It essentially says that to reduce income inequality, we have to jointly reduce barriers to higher education, as well as increase parental time investments.

The final chapter considers a different dimension of inequality - health inequality. Health inequality in the UK is on the rise and the gap in obesity rates between poor and affulent areas is a major policy concern (Holmes (2021), Theis and White (2021), Public Health England (2021), Office for Health Improvement and Disparities (2021)). However, to design effective policies we need to know how obesity actually evolves over the lifecycle. I find that differences in health investments over the lifecycle contribute to up to 34% of the differences in obesity rates between those coming from rich versus poor families. Moreover, I find that investments made by parents during childhood matter more than investments made during adulthood. Again, this has direct implications for policy, especially because a subsequent simulation that I run indicates that individuals would highly value a policy which increases investments in children's health.

Acknowledgements

This thesis is dedicated to Taha Saei (1992-2018),

who would have written a much more brilliant thesis than this one.

Doing this PhD was much harder than I ever anticipated and I would not have made it without the support of a number of amazing people. First and foremost, I would like to thank my advisor Eric French for supporting me right from the start, opening opportunities for me, and encouraging me to "keep at it" throughout. A big thank you goes out to Cormac O'Dea for the incredible amount of support he has provided over the past years - you have truly shown me that one can be a great academic *and* human at the same time. I would like to thank Gabriella Conti, Mariacristina De Nardi, and Milena Almagro for being extremely generous with time and advice, and for being fantastic role models to me. I am grateful to the ESRC for funding and to the Institute for Fiscal Studies for hosting me over the past years and providing me with a stimulating intellectual environment.

Thank you to both my parents who beyond doubt understood the importance of human capital investments. Finishing this PhD is my proudest achievement to date, and it will be forever painful to me that you didn't live to witness it. A special thank you also goes out to my friends - econ and non-econ - who went through all the highs and lows with me over the past years and without who I think I would have given up at some point. Thank you to my sisters Toki and Mika - you probably still don't know if I do micro or macro, but you care about everything else in my life and remind me that I am more than my work. Last, but not least, I would like to thank Eduardo for believing in me from day one and ensuring I make it over the finish line with a big smile, hopeful for the future.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Why do those coming from low income families end up with lower income themselves? Why do they tend to be less healthy? How do these gaps evolve over the lifecycle? These are the key questions that this dissertation tries to answer. In the following chapters I use reduced form and structural methods, combined with two unique cohort data sets from the UK to provide evidence as to how inequality evolves over the lifecycle.

The second chapter studies the intergenerational persistence in earnings and quantifies the drivers behind it. Individuals who come from richer families tend to attain more years of schooling, they have higher cognitive skills, during childhood they attend better quality schools and get to spend more time with their parents, and lastly, they tend to have more educated parents and fewer siblings.[1] Jointly quantifying how much each of these channels matter and how they interact with each other is important for policy makers and researchers alike.

Using a unique longitudinal data set covering a British cohort's first 55 years of life, we show that 54% of the IGE for males and 62% for females can be accounted for by differences in years of schooling, cognitive skills, parental investments in time and school quality, and family background. Of these channels, we show that

[1] See Solon (1992), Dearden et al. (1997), Mazumder (2005), Chetty et al. (2014) for evidence on the intergenerational correlation of earnings. The importance of time investments in children for their later life earnings has been extensively studied (Cunha and Heckman (2008) , Daruich (2018), Lee and Seshadri (2019)) and it has been well documented that richer parents invest more time in their children (Guryan et al. (2008)). Aizer and Cunha (2012) and Del Boca et al. (2014) show that richer parents have fewer children, allowing them to invest more time in each child.

parental investments are the main driver of the link between parents' income and children's earnings via higher cognitive development.

The key contributions of this paper are twofold. First, whilst other papers have considered the channels above individually, or jointly by matching moments from different data sources, to the best of our knowledge, this is the first paper to evaluate these major drivers of the IGE using a single sample of people.[2] Second, the mediation analysis technique that we develop in this paper provides a systematic way in which researchers can explore how potentially relevant channels operate. This is a step which can complement the explicit modeling of decision-making and can thus be useful for further research.

The third chapter then studies parental investment decisions in more detail. Parents have multiple ways of investing in their children. During childhood, they can spend time with their children, thus fostering their cognitive skills (Cunha et al. (2006), Heckman and Mosso (2014)). They can also pay for their child's education (Belley and Lochner (2007), Abbott et al. (2019)). Lastly, they can give cash transfers (Castaneda et al. (2003), De Nardi (2004)). What drives differences in parental investments and how much do early life circumstances contribute to differences in later life wages? Using the same data set as in chapter two, we estimate a dynastic model of parental decision-making. The following key findings emerge: First,we find only modest dynamic complementarity between early time investments in children and later time investments. However, similar to Delaney (2019) and Daruich (2018), we find substantial complementarities between final childhood ability at age 16 and education in wages. Among men with college education, a one standard deviation increase in cognitive ability leads to an additional 19 % in wages. Among those with the lowest level of education, this premium, at 9%, is much smaller. As a result, high ability individuals are more likely to select into education than

[2]See e.g. Keane and Wolpin (1997), Carneiro and Heckman (2002),Lochner and Monge-Naranjo (2012), Lee and Seshadri (2019), Caucutt and Lochner (2020) on education; Milligan and Stabile (2011), Dahl and Lochner (2012) and Agostinelli and Sorrenti (2018) on cognitive skills; Cunha and Heckman (2008), Cunha et al. (2010), Attanasio et al. (2020) on time investments; Altonji and Dunn (1996) and Dearden et al. (2002) on school quality, Black et al. (2005), Angrist et al. (2010), Bhalotra and Clarke (2020),Meghir and Palme (2005), Nybom and Stuhler (2014) on family background; Gayle et al. (2018) and Blanden et al. (2007) study some of these channels jointly.

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, we find that more than a quarter (28%) of the variance of lifetime wages can already explained by characteristics of the family, before an individual is even born. By the time individuals are 23, their characteristics can explain up to 62% of the variance in lifetime wages. Thus, more than half of the lifetime variability in wages is realized by age 23.

The fourth chapter studies a different dimension of inequality - health inequality. Those who grew up in lower income families not only tend to have lower income themselves as adults, but also tend to be less healthy (Almond and Currie (2011), Almond et al. (2018), Currie and Stabile (2003), Case et al. (2002), Case et al. (2005)). One aspect of health in which this is particularly prevalent is obesity (El-Sayed et al. (2012b), El-Sayed et al. (2012a)). Why do differences in obesity rates by childhood socioeconomic status arise and what - if anything - can policy achieve? Much research has shown that early life circumstances can have long-run effects on individual's health.[3] However, few paper to date explicitly model how childhood circumstances are linked to adulthood health.[4] My paper helps fill this gap by developing a framework that coherently models health and health investments over the whole lifecycle, starting from the in-utero period and ending with death. In particular, I model both parental health investments during childhood as well as continued own health investments during adulthood. This allows me to evaluate, for the first time, how much parental investments matter versus own investments during adulthood in the development of the obesity gradient. This is particularly important when considering potential policy interventions and deciding where government resources could be used most effectively to reduce health inequality. I estimate the model using data from a UK cohort that was born in 1970. I estimate health for-

[3]See references above. Almond et al. (2018) provides an overview of this literature.

[4]Two exceptions here are Kulikova (2015) and Dalgaard et al. (2021).

mation processes in childhood and adulthood allowing for heterogeneity which I identify using recent econometric techniques. I then incorporate these processes into a rich dynastic model. I find that differences in investments during childhood explain a larger share of the obesity gradient than differences in investments during adulthood. Investments during childhood are particularly productive as they can affect an individual's permanent health, as well as labour market productivity, which allows for higher self-investments in adulthood. A policy experiment in which the government invests directly in the health of low income children has large welfare effects, with individuals' willingness to pay for the program far exceeding the cost.

Chapter 2

The Intergenerational Elasticity of Earnings: Exploring the Mechanisms

2.1 Introduction

Income across generations is persistent. A developing literature investigates the mechanisms behind this persistence, such as differences in parental time, years of schooling, and family background.[1] However, there is still scarce evidence on the quantitative importance of each of these mechanisms for the intergenerational elasticity of earnings (IGE) and on the way they interact with each other. In this paper we provide such evidence.

Using a unique longitudinal data set covering a British cohort's first 55 years of life, we show that 54% of the IGE for males, and 62% for females, can be accounted for by differences in years of schooling, cognitive skills, parental investments in time and school quality, and family background. Of these channels, we show that parental investments are the main driver of the link between parents' income and children's earnings via higher cognitive development.

Our approach is to decompose the IGE at different levels of mediation, allowing us to quantify not only the impact of different factors (e.g., parental time and

[1] See Solon (1992), Dearden et al. (1997), Mazumder (2005), Chetty et al. (2014) for evidence on the intergenerational correlation of earnings. The importance of time investments in children for their later life earnings has been extensively studied (Cunha and Heckman (2008) , Daruich (2018), Lee and Seshadri (2019)) and it has been well documented that richer parents invest more time in their children (Guryan et al. (2008)). Aizer and Cunha (2012) and Del Boca et al. (2014) show that richer parents have fewer children, allowing them to invest more time in each child.

school quality investments) on lifetime earnings, but also to understand the mechanisms by which these factors impact earnings (e.g., through cognitive skills and years of education).[2] Specifically, we estimate how parental income affects each of the factors that we consider, and how these, directly or indirectly through the other factors, affect an individual's lifetime earnings. Similar to Heckman et al. (2018), we do not explicitly model agents' preferences, but instead we approximate their decision rules, and exploit restrictions arising from the lifecycle timing of human capital development.

We allow for four distinct mechanisms that can generate persistence in earnings across generations. The first mechanism is years of schooling. We find that the higher levels of schooling of children from richer families matter for the IGE.

However, once we consider the second mechanism, cognitive skills, we find that the effect of years of schooling on the IGE can be entirely explained by differences in age 16 cognition. Cognition, therefore, is crucial - it directly affects earnings, and it is the main driver of schooling decisions.

The third mechanism is investments. Richer families invest more time in their children and also send their children to better quality schools. These investments explain 47% (46%) of the IGE for males (females) once we allow investments to subsequently affect the development of cognition and years of schooling. Both types of investment are important, with time investments playing a more important role early in childhood, and school quality investments becoming more important later in childhood. Investments are critical for understanding later life earnings, but their importance arises through their effect on cognitive skills and years of schooling. We find little evidence that these investments are important for earnings over and above their impact via cognitive skills. This is an interesting finding in its own right, as it is typically *assumed* that investments impact later life outcomes only through their impact on observable skills. However, this assumption is rarely tested (see Heckman et al. (2013) for an example that does test this assumption). We fail

[2] See Klein and Goldberger (1955) and Theil (1958) for early applications of mediation analysis in economics, and Blanden et al. (2007), Heckman et al. (2013), Conti et al. (2016), Mogstad et al. (2021) for recent relevant applications.

to reject this key assumption.

The fourth is family background. Parental education and family size, through their impact on investments, explain 19% (34%) of the IGE for males (females). However, we also find that differences in investments between high and low income parents are not entirely driven by family background differences. The direct link between parental income and investments explains 33% (25%) of the IGE. These results are consistent with the view that if financial constraints are important for understanding the persistence in income across generations, it is because they constrain early-in-life investments, not years of schooling.

Whilst other papers have considered the channels above individually, or jointly by matching moments from different data sources, to the best of our knowledge, this is the first paper to evaluate these major drivers of the IGE using a single sample of people. Using the same sample throughout our mediation analysis enables us to measure inputs and outputs of one group of people in a single setting, and also to test whether early life investments (e.g. parental time) have independent effects on late life outcomes, over and above their effects on intermediate outcomes (e.g. cognition). Our mediation analysis also builds on recent advances in latent factor methods and carefully takes into account measurement error. This turns out to be important - ignoring measurement error attenuates the fraction of the IGE explained by parental investment by 45%.

The rest of this paper proceeds as follows. Section 2 relates our work to previous literature. Section 3 describes the data and documents descriptive statistics on schooling, cognition, investments, and family background. Section 4 outlines our mediation approach and Section 5 provides results and robustness checks. Section 6 concludes.

2.2 Literature

This paper focusses on understanding the importance of candidate mechanisms for the intergenerational elasticity in earnings. The setting for our empirical work is the UK, where the level of intergenerational persistence is similar to that observed in the US, and substantially higher than in most developed countries (Corak (2013)).[3]

A growing literature examines what drives these intergenerational associations (see Black et al. (2011)). Here, we summarize the literature on the intergenerational elasticity of earnings and its interplay with the mechanisms we explore: years of schooling, cognition, parental time investments, school quality and family background.

Consider the role of schooling first. Children of richer parents spend on average more years in education which leads to higher earnings. Two distinct channels have been emphasized in interpreting this. One emphasizes that children of richer parents develop greater cognitive skills, which keeps them in school longer (Keane and Wolpin (1997), Carneiro and Heckman (2002)). Another emphasizes the role of borrowing constraints (Lochner and Monge-Naranjo (2012), Lee and Seshadri (2019), Caucutt and Lochner (2020)). Our mediation analysis allows us to investigate the extent to which the effect of schooling on the IGE is mediated by cognition, and to what extent it can be directly accounted for by parental income.

The role of differences in the skills between children of low and high income parents has also been previously highlighted. Exploiting plausible variation in income, a number of papers (e.g Milligan and Stabile (2011), Dahl and Lochner (2012) and Agostinelli and Sorrenti (2018)) find that increases in parental income have a positive effect on children's test scores. Falk et al. (2021) also show that children of high income parents have higher cognition.[4] Our framework allows us to quantify how much of the differences in cognition can be explained by differential investments or differences in parental education, and to what extent they can be explained directly by parental income.

Papers in a large recent literature (e.g. Cunha and Heckman (2008), Cunha et al. (2010), Attanasio et al. (2020)) estimate production functions of human capital and have shown the formative role of parental time investments in the development

[3]Several papers estimate this elasticity using the same NCDS data we use, and find it to be between 0.21 and 0.43, depending on the measures and procedures used (Dearden et al. (1997), Gregg et al. (2016), Belfield et al. (2017))

[4]Non-cognitive skills have also been widely studied (Papageorge et al. (2019), Heckman et al. (2006)), though their effects on earnings are usually estimated to be smaller than for cognition.

of human capital. This evidence, combined with the fact that high income people spend more time with their children (Guryan et al. (2008)), means that parental differences in time investments are likely to be important in propagating correlations in earnings across generations. We estimate how parental investments affect cognitive skills at age 16 using latent factor methods and incorporate this into our mediation analysis. We find that differences in parental time investments are the dominant channel driving the IGE, at least for men. Applications of the production functions estimated in this literature often assume that investments affect earnings *only* through their impact on human capital. We can test whether this assumption holds, or whether parental investments can affect children's lifetime earnings above and beyond their effects on cognition. We fail to reject this assumption.

The literature on the role of school quality is smaller than that on time investments. Altonji and Dunn (1996) and Dearden et al. (2002) find modest effects of school quality on children's outcomes. More recently, Chetty et al. (2014) show that where a household lives has important impacts on income mobility, which may be related to school quality. Our data contains several measures of school quality, and allows us to look separately at the role of both parental time and school quality investments. We find that differences in school quality investments are the dominant channel driving the IGE, at least for women.

Turning finally to the role of family background, studies have investigated the impact of growing up in larger families (Black et al. (2005), Angrist et al. (2010), Bhalotra and Clarke (2020)), as well as effects of having more educated parents (Meghir and Palme (2005), Nybom and Stuhler (2014)). Whilst these papers are able to estimate the causal effect of family background on child outcomes by using plausibly exogenous variation, our aim is to trace mechanisms through which any family background effect operates. In particular, we can estimate whether these family characteristics affect lifetime earnings through improved cognition and schooling or through parental investments.

The two papers most similar to ours are Gayle et al. (2018) and Blanden et al. (2007). Gayle et al. (2018) estimate a rich intergenerational lifecycle model of parental decision-making that includes several of the channels that we consider. Our contribution relative to that paper, and other recent work which models parental decisions over certain investments[5], is that we can use our rich data set to quantify and test how different mechanisms affect the IGE, with very limited pre-imposition as to how these channels operate. Blanden et al. (2007) estimate the extent to which cognitive skills, non-cognitive skills, educational attainment, and labor force attachment have accounted for the rise of intergenerational earnings persistence between the NCDS 1958 cohort, and a cohort born in 1970. We build on their work in three ways. First, instead of taking cognition or schooling as given, we estimate the extent to which they are affected directly by parental income, but also indirectly through differences in investments and family background. Second, we use latent factor methods as recently developed in Heckman et al. (2006), Heckman et al. (2013), and Heckman et al. (2018), in order to account for the fact that ability, investment and school quality are measured with error. Finally, we study the lifetime earnings of both males and females rather than using male income at a particular point in time as an outcome.

To summarize, the aim of this paper is to quantify the effect of various channels on the IGE with minimal restrictions beyond those imposed by the lifecycle timing of human capital development. That is, within a framework informed by that lifecycle, we use the data to speak to the importance of each channel's direct and indirect effect on the IGE. Apart from being of interest in their own right, the results of this paper can be valuable for papers attempting to explicitly model parental behaviour, by being informative about what channels quantitatively matter, as well as how these should be included in a model.

2.3 Data

Our data comes from the UK's National Child Development Study (NCDS).[6] The dataset covers the full population of children born in one particular week of March

[5] For example Gayle et al. (2015), Lee and Seshadri (2019), Caucutt and Lochner (2020) and Daruich and Kozlowski (2020).

[6] The NCDS is provided by the Centre for Longitudinal Studies (2017) at the Institute of Education, University College London.

1958 and continues to follow them to this day. The NCDS is a globally-unrivalled resource for social scientists in its combination of information about investments in early childhood with information on ability, educational outcomes and later-life earnings. To the best of our knowledge, it is the only dataset in the world that provides detailed information on early-life investments and earnings outcomes over the entire working life. The data include multiple measures of children's ability, parental time investments and school quality in each wave. This allows us to posit the existence of underlying latent factors and to use a latent factor model to extract signals from noisy measures rather than assuming that cognition, investments and school quality can be adequately captured by one particular measure.

The initial survey at birth has been followed by subsequent surveys at ages: 7, 11, 16, 23, 33, 42, 46, 50 and 55.[7] The data from childhood includes information on several measures of cognition and parental investments, number of siblings, parental education, and parental income. Later waves of the study record educational outcomes, demographic characteristics, earnings, and hours of work. Table 2.2 gives a list of all the different cognition, time investment, and school quality investment measures we use. Details about the sample we use are given in Appendix 2.A.2.

In the rest of this section, we first describe how we construct the measures of parental income and child earnings that we use. We then document inequalities in family background, investments, and outcomes by parental income over the lifecycle in the NCDS data.

2.3.1 Parental Income

When NCDS cohort members were aged 16, comprehensive data on parental income was collected. Our measure of parental income sums across: father's earnings, mother's earnings, and other income, all net of taxes. Further details can be found in Appendix 2.A.1.1. In the descriptive table which follows, we split households by parental income tertile. Average annual income (at the time the child is 16

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

years old) is reported in the first row of Table 2.1 and ranges from £13,600 (at the 2014 price level) in the first tertile to £29,300 in the top tertile.

2.3.2 Children's Earnings

We observe cohort members' gross earnings at ages 23, 33, 42, 50, and 55. For further details on these earnings measures, see Appendix 2.A.1.2. Average annual earnings (setting earnings of non-workers to 0) over all ages 23-55 are reported in the second row of Table 2.1. The positive gradient with respect to parents' income is immediately apparent – average gross annual earnings of NCDS cohort members rises from £17,300 for those with parents in the bottom income tertile, to £20,400 for those whose parents are in the top income tertile.

2.3.3 Years of Schooling

The age at which an individual left school is constructed using the highest educational qualification recorded by the time cohort members reach age 33. The second panel of Table 2.1 shows that children with parents in the bottom income tertile spend 0.21 years less in school than children with parents in the highest income tertile.

2.3.4 Cognition

As part of the survey, cohort members took part in tests to measure their cognitive skills.[8] The third panel of Table 2.1 summarises standardized math and reading test scores at age 16 by parental income tertile. As one might expect, children from richer households develop greater cognitive skills; at age 16, reading scores were 21% of a standard deviation higher on average for children with parents in the highest income tertile compared to children with parents in the lowest tertile. For math scores, the difference between children from high versus low income families is 18% of a standard deviation.

Similar patterns can be found for teacher-rated English language and math ability at age 16.

[8] We use the terms cognitive skills and cognition interchangeably.

Table 2.1: Descriptive Statistics and Means by Parental Income Tertile

			Parental Income Tertile			
Variable	Mean	SD	Bottom	Middle	Top	P-val
Income variables						
Parents' household income at 16	21,134	7,523	13,598	20,496	29,308	0.00
Children's average annual earnings	18,899	16,913	17,293	19,019	20,386	0.00
Education						
Age left school	18.0	1.8	17.9	17.9	18.1	0.02
Cognition						
Reading at age 16	0.00	1.00	-0.11	0.01	0.10	0.00
Math at age 16	0.00	1.00	-0.08	-0.02	0.10	0.00
Teacher assessed English ability at age 16	0.00	1.00	-0.08	-0.04	0.12	0.00
Teacher assessed Math ability at age 16	0.00	1.00	-0.09	-0.01	0.10	0.00
Time investment						
% of fathers very interested at age 7	21.5	41.1	18.6	21.5	24.4	0.01
% of mothers very interested at age 7	34.4	47.5	31.5	34.1	37.4	0.03
% of fathers very interested at age 11	23.5	42.4	20.1	23.8	26.4	0.01
% of mothers very Interested at age 11	33.4	47.2	29.8	34.3	36.1	0.02
% of fathers very interested at age 16	29.9	45.8	28.2	28.6	32.9	0.08
% of mothers very interested at age 16	33.3	47.1	31.5	32.8	35.6	0.19
School quality						
% whose PTA holds meetings at age 7	57.7	49.4	56.8	57.6	58.7	0.71
Student-teacher ratio at age 11	24.8	9.35	25.4	24.7	24.3	0.06
% from child's class suitable for GCEs at age 11	23.6	14.0	23.4	23.1	24.3	0.20
Student-teacher ratio at age 16	17.2	1.93	17.2	17.3	17.1	0.05
% from child's class studying for GCEs at age 16	46.3	31.2	44.0	44.4	50.5	0.00
% from child's school who complete school	55.7	25.8	53.5	54.2	59.6	0.00
Family Background						
Number of siblings	2.04	1.46	2.13	1.93	2.05	0.01
Father's age left school	15.0	1.17	14.9	14.8	15.2	0.00
Mother's age left school	15.1	1.06	15.0	15.1	15.3	0.00

Note: This table presents descriptive statistics and means by parental income tertile for parental income, child earnings, and several measures of the mediators in our analysis (years of schooling, cognition, time investments, school quality, family background). The relevant sample consists of 2,637 individuals. Income and earnings variables are reported as annual values, deflated to 2014 prices. The final column reports *p-values* from the *F-tests* testing the null hypothesis of equality of means across income tertiles.

2.3.5 Parental Investments

The NCDS has detailed measures of the parental investments which cohort members received during childhood. The full set of parental time and school quality measures as well as cognitive skill measures are listed in Table 2.2. These measures come from different sources – some are from surveys of parents, others from surveys of school teachers and headteachers. The fourth and fifth panels of Table 2.1 document, and the next subsections discuss, the income gradients for a subset of these measures; these gradients are similar across all measures (see Tables 2.A.1 and 2.A.2 in Appendix 2.A for descriptive statistics on the full set of measures).

2.3.5.1 Parental Time Investments

Table 2.1 shows teachers' responses to a question about how interested parents are in their children's education (very interested, a little interested, not interested at all), asked when the children are 7, 11 and 16. We report the fraction of mothers and fathers who are reported to be very interested. The differences by parental income are evident at all ages for both parents. For example, 32% of mothers in the bottom tertile are adjudicated to be 'very interested' in their 7-year-old's education, compared to 37% of mothers at the top. The relevant figures for fathers are 19% and 24% respectively. Tables 2.A.1 and 2.A.2 in the Appendix shows that a qualitatively similar gradient between parental income and time investments is evident for our other measures of time investment.

2.3.5.2 School Quality

The fifth panel of Table 2.1 presents some measures of school quality at different ages. Children of high income parents are more likely to go to schools where: parents attend educational meetings at age 7, student-teacher ratios are low at age 11, and a high fraction of students are doing GCEs at age 16 (an optional exam for progressing to further education) and complete secondary education. However, some measures of school quality, such as the student-teacher ratio at age 16, hardly differ by parental income tertile. This highlights the advantage of our factor analytic approach, which does not focus on any individual measure of school quality.

2.3.6 Family Background

In our main analysis we use three family background variables: the number of siblings and mothers' and fathers' years of education (though in Section 2.5.5 we will show the robustness of our results to including several more). The sixth and final panel in Table 2.1 shows the well-known fact that poorer families tend to be larger. NCDS cohort members from households in the bottom income tertile grew up with more siblings on average than those in the other income tertiles. Mothers' education also differs significantly between parental income tertiles - mothers in the highest tertile spent on average four more months in school than those in the lowest tertile.

The next section shows how we use this rich data set to understand how the parental income gradients described here map into the gradient of lifetime earnings.

2.4 Approach

The goal of this paper is to decompose the IGE into fractions explained by schooling, cognition, investments, and family background using mediation analysis. The structure of this mediation framework follows logically from the lifecycle timing of human capital development. We allow factors that are determined later in life (such as years of schooling and cognition) to be affected by factors determined earlier in life (such as investments and family background).

Throughout this analysis, we correct for potential biases arising from measurement error in parental income, cognition, time investments, and school quality and use data on multiple measures of each of these last three factors. Our approach builds on recent work in economics that uses mediation analysis such as Heckman et al. (2013) and Heckman et al. (2018), and adds several more layers of mediation.[9]

The first part of this section describes our approach for estimating the IGE. The second part describes the multi-level mediation approach in more detail and illustrates the estimation of the equations for lifetime earnings, years of schooling, cognitive skill, and investments. The final part describes our latent factor approach and how we account for measurement error.

[9] For applications of mediation in economics see Blanden et al. (2007), Conti et al. (2016), Gelbach (2016), Bandiera et al. (2020) and Mogstad et al. (2021). Mediation analysis is also very common in the psychology literature, see MacKinnon et al. (2007) for an overview.

2.4.1 Measurement of the Intergenerational Elasticity of Earnings

To estimate the IGE, we would like to regress (de-meaned) log lifetime earnings of the NCDS cohort member $\ln Y_i$ on (de-meaned) log lifetime income of their parents $\ln Y_{Parent,i}$:

$$\ln Y_i = \rho \ln Y_{Parent,i} + u_i \tag{2.1}$$

where ρ represents the intergenerational elasticity of earnings. Since we consider de-meaned variables here, we omit the intercept.

We do not have a measure of parents' lifetime income and so cannot estimate ρ directly. We do, however, observe parents' income when the cohort member was 16. Thus, we estimate a regression of (de-meaned) log lifetime earnings of the NCDS cohort member $\ln Y_i$ on (de-meaned) log income of their parents measured when NCDS cohort members are 16 years old, $\ln Y_{Parent,i,16}$:

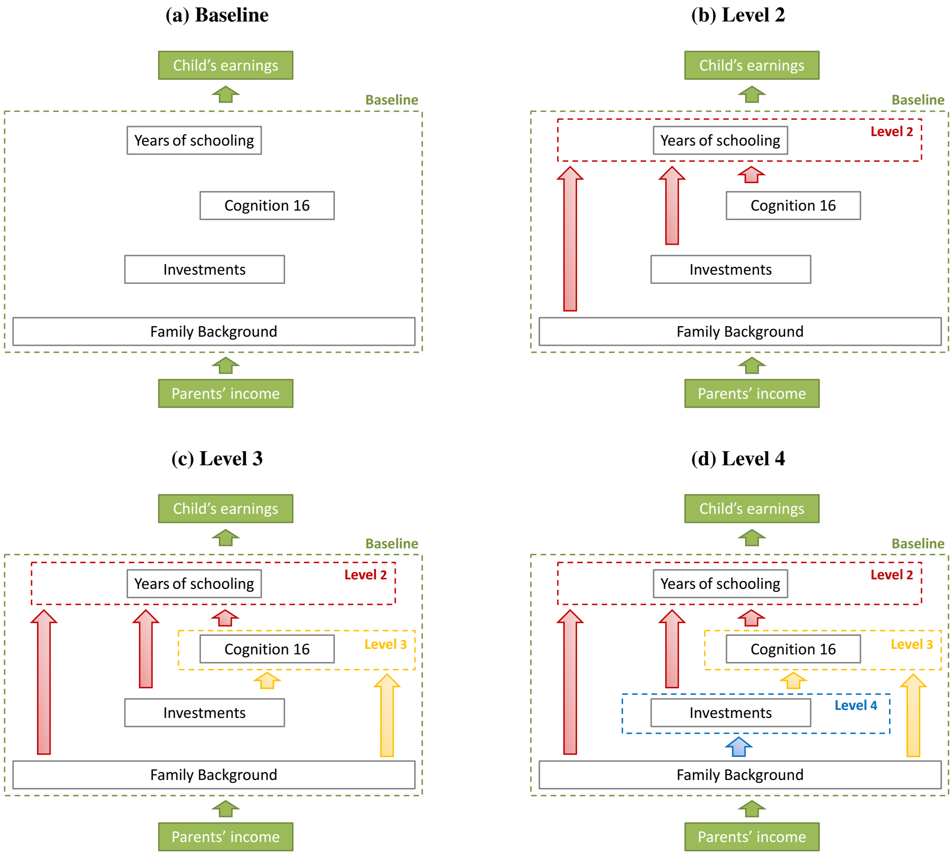
$$\ln Y_i = \rho^* \ln Y_{Parent,i,16} + u_i. \tag{2.2}$$

Previous research has shown that using parents' income at a point in time as a proxy for parents' lifetime income can lead to significant attenuation bias. For example, Haider and Solon (2006) find that using point in time income can bias the estimate of the IGE downwards by between 35-80% if parents' income is measured when the parents are very young or very old, although they also show that the bias is modest when the parents are in their 40s. Fortunately for our analysis, parents are on average age 43 when their income is measured.

Nevertheless, we also address the issue of measurement error explicitly. We use an errors-in-variables framework that is similar to the framework adopted in Haider and Solon (2006). We estimate the attenuation bias using information on children's lifetime and point in time income at age 42. As parents are on average age 43 when we observe their income, we assume that the reliability ratio is the same across generations and use it to correct the IGE. See Appendix 2.B for further details.

2.4.2 Decomposition and Multilevel Mediation Analysis

Figure 2.1: Overview of Mediation Approach



Note: In all levels, parents' income directly impacts all factors, which in turn affect the child's earnings. Additionally, in levels 2, 3, and 4, factors within the box impact each other as shown by the arrows.

Figure 2.1 illustrates our decomposition and mediation approach. We begin with a baseline level (illustrated in panel a) in which we estimate the impact of parental income on the mediating factors: years of schooling, cognition, investments and family background (which are depicted in the box in panel a). Next, we estimate the impact of each of these mediating factors on the cohort member's lifetime earnings. We then decompose the IGE into the fractions explained by the different mediating factors. In this decomposition, each element explains a fraction of the IGE, holding constant the effect of all other mediators. The decomposition in level 2 (panel b) allows for family background, investments and cognition to affect the IGE directly, but also indirectly via an effect on years of schooling. The decomposition in level 3 (panel c) allows for family background and investments to affect the IGE directly and indirectly through both cognition and schooling. Finally, the decomposition in level 4 (panel d) allows for family background to affect the IGE directly and through investments, cognition, and years of schooling. The decomposition procedure extends the one in Gelbach (2016) to nest multiple layers of mediation and also to account for measurement error. An attractive feature of the decomposition is that the shares explained in the decomposition are invariant to the level of mediation. We now give more detail about each of these stages in turn.

Baseline - Direct effects on lifetime earnings To implement this multi-level decomposition, we start by estimating the parameters of the following model:

$$\ln Y_i = \alpha_S S_i + \alpha_C C_i + \boldsymbol{\alpha}_I \boldsymbol{I}_i + \boldsymbol{\alpha}_F \boldsymbol{F}_i + \alpha_{Y_P} \ln Y_{Parent,i} + u_i^Y \tag{2.3}$$

where S is individual i's years of schooling and C is the child's cognition measured at age 16. $\boldsymbol{I} = [inv_7, inv_{11}, inv_{16}, sq_7, sq_{11}, sq_{16}]$ is a vector of all investments (time investments and school quality) at ages 7, 11, and 16. $\boldsymbol{F} = [ed_m, ed_f, sib]$ is a vector containing the family background variables (i.e. mother's and father's years of schooling, and the number of siblings of child i).

Our goal is to estimate the shares of the IGE explained by different channels. To do this, we first totally differentiate equation (2.3) with respect to $\ln Y_{Parent,i}$, which yields:

$$\frac{d\ln Y_i}{d\ln Y_{Parent,i}} = \alpha_S \frac{dS_i}{d\ln Y_{Parent,i}} + \alpha_C \frac{dC_i}{d\ln Y_{Parent,i}} + \boldsymbol{\alpha}_I \frac{d\boldsymbol{I}_i}{d\ln Y_{Parent,i}} + \boldsymbol{\alpha}_F \frac{d\boldsymbol{F}_i}{d\ln Y_{Parent,i}} + \alpha_{Y_P} \tag{2.4}$$

We then estimate how each of the covariates is individually related to parental income. For example, in the case of cognition, to calculate $\frac{dC}{d\ln Y_{Parent}}$ we estimate:

$$C_i = \kappa_C \ln Y_{Parent,i} + v_i^C \tag{2.5}$$

Using equations (2.4) and (2.5) we compute the total impact of parental income on earnings via cognition to be: $\frac{\partial \ln Y}{\partial C} \frac{dC}{d\ln Y_{Parent}} = \alpha_C \cdot \kappa_C$. Further, from equation (2.1) we know that the IGE is $\frac{d\ln Y}{d\ln Y_{Parent}} = \rho$. With these two quantities we can estimate the share of the IGE explained by cognition to be $\frac{\alpha_C \cdot \kappa_C}{\rho}$. We perform similar decompositions for all of the other variables – schooling, investments and family background.

As equation (2.4) shows, α_{Y_P} is the part of the IGE not explained by any of the other covariates, and thus $\frac{\alpha_{Y_P}}{\rho}$ is the total fraction of the IGE not explained by them.

Estimation of equation (2.3) allows us to test some restrictions in the lifetime earnings equation that are frequently made in the literature. Parental investments might impact future earnings only through their impact on cognition and subsequent school attainment. If this is the case, then they should have zero predictive power for child lifetime earnings, conditional on cognition and schooling. That is, in the language of psychology, there would be *complete mediation* (MacKinnon et al. (2007)). In fact, this is the standard assumption in the literature of dynamic skill investment (Caucutt and Lochner (2020), Lee and Seshadri (2019), Daruich (2018)). However, it is possible that, for example, school quality may have an effect on lifetime earnings above and beyond years of schooling, for instance by allowing individuals to have access to a more opportune social network and hence get better jobs. Whilst the coefficients on the covariates tell us individually whether this is the case, we also test whether groups of covariates jointly have a significant effect on lifetime earnings using *F-tests*. That is, we test for the joint significance of time investments at all ages, $\alpha_{inv_7} = \alpha_{inv_{11}} = \alpha_{inv_{16}} = 0$, school quality at all ages, $\alpha_{SQ_7} = \alpha_{SQ_{11}} = \alpha_{SQ_{16}} = 0$, and all family background variables $\alpha_{ed_m} = \alpha_{ed_f} = \alpha_{sib} = 0$.

Level 2 - Allowing for Indirect Effects via Years of Schooling Years of schooling can be affected by skills acquired by age 16, investments and family circumstances.

To estimate how these mediators affect the IGE both directly and indirectly via schooling, we first estimate the parameters of an equation which relates schooling to the income of the parents and each of the mediators:

$$S_i \;=\; \beta_C C_i + \boldsymbol{\beta}_I \boldsymbol{I}_i + \boldsymbol{\beta}_F \boldsymbol{F}_i + \beta_{Y_P} \ln Y_{Parent,i} + u_i^S \tag{2.6}$$

Now allowing S_i to depend on the other covariates, total differentiation of equation (2.3) with respect to $\ln Y_{Parent,i}$ yields:

$$\begin{aligned}\frac{d\ln Y_i}{d\ln Y_{Parent,i}} &= \alpha_S\left(\frac{\partial S_i}{\partial C_i}\frac{dC_i}{d\ln Y_{Parent,i}} + \frac{\partial S_i}{\partial \boldsymbol{I}_i}\frac{d\boldsymbol{I}_i}{d\ln Y_{Parent,i}} + \frac{\partial S_i}{\partial \boldsymbol{F}_i}\frac{d\boldsymbol{F}_i}{d\ln Y_{Parent,i}} + \beta_{Y_P}\right) \\ &\quad + \alpha_C\frac{dC_i}{d\ln Y_{Parent,i}} + \boldsymbol{\alpha}_I\frac{d\boldsymbol{I}_i}{d\ln Y_{Parent,i}} + \boldsymbol{\alpha}_F\frac{d\boldsymbol{F}_i}{d\ln Y_{Parent,i}} + \alpha_{Y_P}\end{aligned} \tag{2.7}$$

Noting from equation (2.6) that, for example, the effect of cognition on schooling is $\frac{\partial S_i}{\partial C_i} = \beta_C$, we can now compute the fraction of the IGE explained by cognition to be:

$$(\underbrace{\alpha_C \cdot \kappa_C}_{\text{Direct Effect}} + \underbrace{\alpha_S \cdot \beta_C \cdot \kappa_C}_{\text{Indirect Effect via schooling}})/\rho. \tag{2.8}$$

The "direct effect" of age 16 cognition on earnings is the same effect shown in level 1: $\alpha_C \cdot \kappa_C$. The new, indirect effect comes from parental income affecting schooling, through cognition: $\frac{\partial S_i}{\partial C_i} \cdot \frac{dC_i}{d\ln Y_{Parent,i}} = \beta_C \cdot \kappa_C$. As before, we know that $\frac{\partial \ln Y_i}{\partial S_i} = \alpha_S$. Together these imply that the share of the IGE explained by cognition, both directly as well as indirectly through schooling, is the expression in (2.8). Similar calculations are done to find the fractions explained by investments and family background.

We can now estimate whether there is any direct effect of parental income on schooling (which affects the IGE), or whether the effect is entirely mediated through the other covariates. For this, we estimate how much the coefficient on schooling differs between a regression of schooling on parental income with no other covariates (κ_S from an equation analogous to (2.5) but with schooling in place of cognition), versus one that includes all other covariates (β_{Y_P} from equation (2.6)). The fraction of the effect of parental income on schooling which is not mediated by the other covariates then becomes $\frac{\beta_{Y_P}}{\kappa_S}$. The fraction of the IGE that is explained by schooling, $\frac{\alpha_S \kappa_S}{\rho}$, can therefore be split into a part that is mediated by other covariates, $\frac{\kappa_S - \beta_{Y_P}}{\kappa_S} \cdot \frac{\alpha_S \kappa_S}{\rho}$, and the remaining part $\frac{\beta_{Y_P}}{\kappa_S} \cdot \frac{\alpha_S \kappa_S}{\rho}$.

We then test whether groups of covariates jointly have a significant effect on years of schooling. That is, in equation (2.6) we test for the joint significance of time investments, $\beta_{inv_7} = \beta_{inv_{11}} = \beta_{inv_{16}} = 0$, school quality, $\beta_{SQ_7} = \beta_{SQ_{11}} = \beta_{SQ_{16}} = 0$, and all family background variables $\beta_{ed_m} = \beta_{ed_f} = \beta_{sib} = 0$.

Level 3 - Allowing for Indirect Effects via Years of Schooling and Cognition In level 3 we allow investments and family background to affect cognition (C_i). We estimate the following equation:

$$C_i = \boldsymbol{\gamma_I I}_i + \boldsymbol{\gamma_F F}_i + \gamma_{Y_P} \ln Y_{Parent,i} + u_i^C \tag{2.9}$$

We model cognition as a function of all investments made up to age 16. This approach can be thought of as the reduced form to a set of structural cognition production functions of the sort estimated in Agostinelli and Wiswall (2016b), Cunha and Heckman (2008) and Bolt et al. (2021), if the cognition functions at different ages are linear in their arguments.[10]

[10]In particular, note that if the cognition production function is:

$$\begin{aligned} C_{i,t} = \; & \rho_t C_{i,t-1} + \Gamma_{inv,t} inv_{it} + \Gamma_{sq,t} sq_{it} + \Gamma_{ed_m,t} ed_{m,i} + \Gamma_{ed_f,t} ed_{f,i} + \Gamma_{siblings,t} siblings_i + \\ & \Gamma_{parinc,t} \ln Y_{Parent,i} + \Upsilon_{i,t}^C \end{aligned}$$

where $t-1$ is the previous period (so if $t = 16$, then $t-1$=11), and

$$C_{i,0} = \Gamma_{ed_m,0} ed_{m,i} + \Gamma_{ed_f,0} ed_{f,i} + \Gamma_{siblings,0} siblings_i + \Gamma_{parinc,0} \ln Y_{Parent,i} + \Upsilon_{i,0}^C$$

then recursive substitution on the above equations yields:

$$\begin{aligned} C_{i,16} &= \Gamma_{inv,16} inv_{i16} + \Gamma_{sq,16} sq_{i16} + \rho_{16} \left(\Gamma_{inv,11} inv_{i11} + \Gamma_{sq,11} sq_{i11} \right) + \rho_{16}\rho_{11} \left(\Gamma_{inv,7} inv_{i7} + \Gamma_{sq,7} sq_{i7} \right) \\ &+ \left(\Gamma_{ed_m,16} + \rho_{16}\Gamma_{ed_m,11} + \rho_{16}\rho_{11}\Gamma_{ed_m,7} + \rho_{16}\rho_{11}\rho_7\Gamma_{ed_m,0} \right) ed_{m,i} \\ &+ \left(\Gamma_{ed_f,16} + \rho_{16}\Gamma_{ed_f,11} + \rho_{16}\rho_{11}\Gamma_{ed_f,7} + \rho_{16}\rho_{11}\rho_7\Gamma_{ed_f,0} \right) ed_{f,i} \\ &+ \left(\Gamma_{siblings,16} + \rho_{16}\Gamma_{siblings,11} + \rho_{16}\rho_{11}\Gamma_{siblings,7} + \rho_{16}\rho_{11}\rho_7\Gamma_{siblings,0} \right) siblings_i \\ &+ \left(\Gamma_{parinc,16} + \rho_{16}\Gamma_{parinc,11} + \rho_{16}\rho_{11}\Gamma_{parinc,7} + \rho_{16}\rho_{11}\rho_7\Gamma_{parinc,0} \right) \ln Y_{Parent,i} + u_i^C \end{aligned} \tag{2.10}$$

where $C_i = C_{i,16}, u_i^C = \Upsilon_{i,16}^C + \rho_{16}\Upsilon_{i,11}^C + \rho_{16}\rho_{11}\Upsilon_{i,7}^C + \rho_{16}\rho_{11}\rho_7\Upsilon_{i,0}^C$. Note that equation (2.10) implies

Equation (2.9) parsimoniously captures how all of the different channels potentially impact cognition. Allowing for the effect of parental income on cognition to be mediated by parental investments and family background we can see that we can insert $\frac{dC_i}{d\ln Y_{Parent,i}} = \frac{\partial C_i}{\partial \boldsymbol{I}_i}\frac{d\boldsymbol{I}_i}{d\ln Y_{Parent,i}} + \frac{\partial C_i}{\partial \boldsymbol{F}_i}\frac{d\boldsymbol{F}_i}{d\ln Y_{Parent,i}}$ into equation (2.7) to infer the total impact of investments and family background and can use this to compute the fractions of the IGE explained when allowing covariates to affect cognition and schooling. For example, the fraction of the IGE explained by time investments at age 16 is:

$$[\underbrace{\alpha_{inv_{16}}}_{\substack{\text{Direct Effect}\\ \text{of } inv_{16} \text{ on Earnings}}} + \underbrace{\beta_{inv_{16}}\alpha_S}_{\substack{\text{Indirect Effect}\\ \text{of } inv_{16} \text{ via schooling}}} + \underbrace{(\underbrace{\alpha_C}_{\substack{\text{Direct Effect}\\ \text{of cognition on Earnings}}} + \underbrace{\beta_C\alpha_S}_{\substack{\text{Indirect Effect}\\ \text{of cognition via schooling}}})\gamma_{inv_{16}}}_{\text{Indirect Effect of } inv_{16} \text{ via cognition}}] \cdot \kappa_{inv_{16}}/\rho \tag{2.11}$$

where $\kappa_{inv_{16}}$ is a regression coefficient of inv_{16} on parental income and captures the fact that richer parents invest more in their kids. We can also estimate how much of the effect of parental income on cognition can be explained by investments and family background: $\frac{\kappa_C - \gamma_{Y_P}}{\kappa_C}$.

Analogously to the previous levels, we now test whether groups of covariates jointly have a significant effect on cognition. That is, in equation (2.9) we test for the joint significance of time investments at all ages, $\gamma_{inv_7} = \gamma_{inv_{11}} = \gamma_{inv_{16}} = 0$, school quality at all ages, $\gamma_{SQ_7} = \gamma_{SQ_{11}} = \gamma_{SQ_{16}} = 0$, and all family background variables $\gamma_{ed_m} = \gamma_{ed_f} = \gamma_{sib} = 0$.

Level 4 - Allowing for Indirect Effects via Years of Schooling, Cognition, and Investments Finally, we also allow for investments to be affected by family background. We estimate the parameters of equations that relate each element of the vector $\boldsymbol{I}$ (containing parental time and school quality investments at each age) to family background characteristics and parental income. For example, for age 16 time investments we estimate:

that reduced form equation (2.9).

$$inv_{16,i} \quad = \quad \boldsymbol{\delta}_F \boldsymbol{F}_i + \delta_{Y_P} \ln Y_{Parent,i} + u_i^{inv_{16}} \tag{2.12}$$

The share of the IGE explained by an element of $\boldsymbol{F}$, for example, maternal education (ed_m), then becomes:

$$\begin{aligned} \{ \quad & \underbrace{\alpha_{ed_m}}_{\substack{\text{Direct Effect} \\ \text{of mum ed on Earnings}}} + \underbrace{\alpha_S \beta_{ed_m}}_{\substack{\text{Indirect Effect} \\ \text{of mum ed via Schooling}}} + \underbrace{(\alpha_C + \beta_C \alpha_S)\gamma_{ed_m}}_{\substack{\text{Indirect Effect} \\ \text{of mum ed via Cognition}}} + \\ & \underbrace{[\underbrace{\alpha_{inv_{16}}}_{\substack{\text{Direct Effect of} \\ inv_{16} \text{ on Earnings}}} + \underbrace{\beta_{inv_{16}} \alpha_S}_{\substack{\text{Indirect Effect of} \\ inv_{16} \text{ via schooling}}} + \underbrace{(\underbrace{\alpha_C}_{\substack{\text{Direct Effect of} \\ \text{cognition on Earnings}}} + \underbrace{\beta_C \alpha_S}_{\substack{\text{Indirect Effect of} \\ \text{cognition via schooling}}})\gamma_{inv_{16}}]\delta_{ed_m,inv_{16}}}_{\text{Indirect Effect of } inv_{16} \text{ via cognition}}}_{\substack{\text{Indirect effect} \\ \text{via } inv_{16}}} \} \cdot \kappa_{ed_m}/\rho \end{aligned} \tag{2.13}$$

where κ_{ed_m} is the parental income gradient for maternal education. This captures how maternal education directly affects lifetime earnings as well as all indirect pathways running through each of years of schooling, cognition, and investments. Again, we can estimate the extent to which the relationship between investments at age 16 and parental income is mediated by family background: $\frac{\kappa_{inv_{16}} - \delta_{Y_P}}{\kappa_{inv_{16}}}$.

2.4.3 Latent Factors and Measurement Error Correction

Following recent developments in the literature on human capital development, we interpret the skill and investment measures in our dataset as noisy measures of unobserved, underlying factors. We allow for investments to be potentially multi-dimensional.[11] In this section we first address the question of how many underlying dimensions of investments to model. We then explain how we deal with the fact that we observe multiple, likely noisy, measures of skills and investment.

2.4.3.1 Choosing the Number of Latent Factors

Given the large number of measures of investments available, it is difficult to determine a priori how many underlying latent factors there are. For example, the

[11] In the robustness check in Section 2.5.5, we also allow for skills to be multidimensional.

number of outings with parents, parental interest in the child's education, and the student-teacher ratio at the child's school may all represent measures of investment in the child's human capital. However, they may represent three distinct types of investments (e.g. leisure time with child, academic interest of parents, school environment). Our approach is to carefully select the number of underlying factors via exploratory factor analysis (Gorsuch et al. (2003), Thompson (2004), Heckman et al. (2013)), which leads us to retain only two factors which we label time investments and school quality. Further description of these procedures, tables including the eigenvalues and loadings, as well as further information on our measures can be found in Appendix 2.D.1.

2.4.3.2 Using Latent Factors in Our Analysis

We now describe how our measures of investments and cognition relate to these underlying latent factors. Following previous literature (Cunha and Heckman (2008), Agostinelli and Wiswall (2016b)), we assume a linear relationship between measures (Z) and underlying latent factors $\Omega \in \{C, inv, sq\}$:

$$Z_{\Omega,i,t,j} = \mu_{\Omega,t,j} + \lambda_{\Omega,t,j}\Omega_{i,t} + \varepsilon_{\Omega,i,t,j} \tag{2.14}$$

Here, $Z_{\Omega,i,t,j}$ denotes measure j of latent factor Ω (e.g. a math score as a measure of latent cognitive skills) for individual i at time t. $\mu_{\Omega,t,j}$ and $\lambda_{\Omega,t,j}$, respectively, are the location and scale of this measure and are constant across individuals. $\varepsilon_{\Omega,i,t,j}$ denotes an idiosyncratic measurement error, assumed to be independent across individuals, measures, and time. The measurement errors are also assumed to be independent of the latent variables and all other controls and shocks. As the latent factors do not have a natural scale or location, we normalize their means to be zero in every period, and their variances to be one. This allows us to estimate the location and scale parameter for each measure (see Appendix 2.D.2 for details). We then predict the latent factors for each individual, using the Bartlett score method (Heckman et al. (2013)). Bartlett scores are a linear combination of all retained measures, inversely weighted by their noise. This means that measures with little measurement error get more weight than those with a lot of measurement error. The measurement error in each cognitive skill and investment measure is minimized, but not eliminated, by combining them via the Barlett scoring method. As a result, we cannot simply estimate the above equations via OLS, but have to correct for the remaining measurement error. We do this by using the errors-in-variables correction as suggested by Heckman et al. (2013). Appendix 2.D.2.4 gives more details.

2.5 Results

2.5.1 Estimates of the IGE

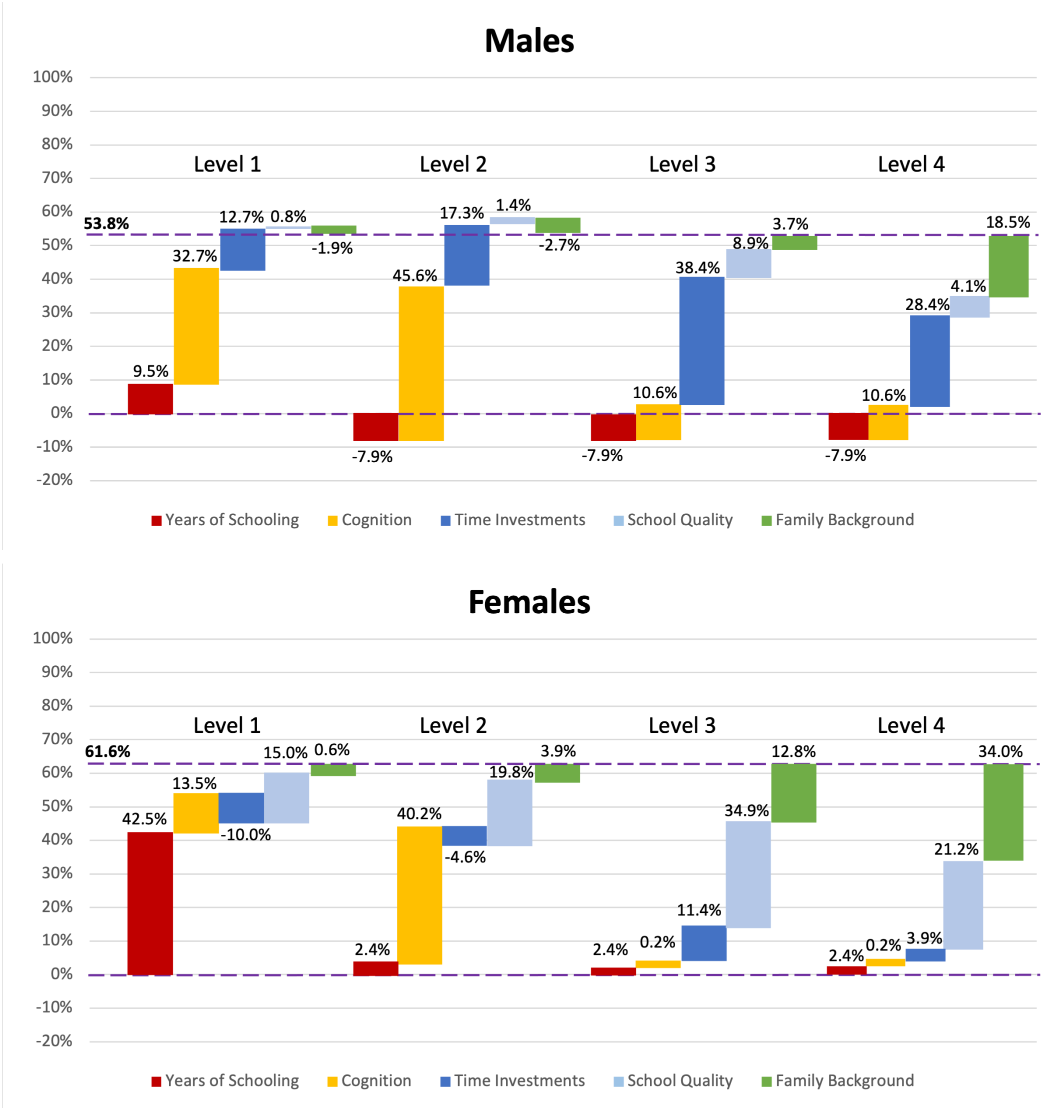
Applying the approach described in Section 2.4.1, we first present our estimates of the the IGE. Table 2.3 shows estimates both with and without corrections for measurement error. Without correcting for measurement error, the estimated IGE is 0.16 and 0.12 for males and females respectively. These estimates are similar to estimates in other studies using the same data (Belfield et al. (2017), Gregg et al. (2016)), and are robust to alternative measures of earnings (see Appendix 2.C). This increases to 0.32 and 0.24 once we account for measurement error that comes from using parental income at age 16 as a proxy for parental lifetime income.

2.5.2 Mediation Analysis

Tables 2.4 and 2.5 give the full results of our mediation analysis for males and females, respectively. Overall, variation in the factors we consider – schooling, cognition, investments and family background – can explain 54% (62%) of the intergenerational elasticity in earnings of males (females). The next four subsections discuss, in turn, the results in each of the four columns of Tables 2.4 and 2.5, which represent the fractions of the IGE that can explained by each covariate at each of our four levels of mediation.

Baseline - Direct effects on lifetime earnings The baseline column in Tables 2.4 and 2.5 decompose the IGE by only allowing for direct effects of each factor on the IGE. We find that differences in cognition and years of schooling explain the largest fraction of the IGE. The fraction explained by these two factors is high for both males and females, but their relative importance differs across genders. For

Figure 2.2: Graphical Summary of Mediation Analysis: Share of IGE Explained



males, 33% of the IGE is explained by cognition and 10% by years of schooling. For females, 13% are explained by cognition and 43% by years of schooling.

We find that parental investments have only a small impact, and family background has virtually no impact on lifetime earnings once we control for cognition, school attainment, and parental income, as we do at this stage of mediation. This is an attractive aspect of both our data and our framework. Whereas most of the literature *assumes* that time and school quality investments impact lifetime earnings only through their role in promoting higher cognition and school attainment, we formally test to see if that is the case. We find that neither parental education nor number of siblings play a role, conditional on cognition and years of schooling. Hence, conditional on parents' income, we find no evidence that parents transmit an advantage to their children's earnings beyond that which comes from improved cognition and schooling outcomes.

These estimated shares are obtained using estimates of the relationship between lifetime earnings and our covariates in equation (2.3) and the relationship between parental income and those covariates (equation (2.5)). Table 2.6 shows determinants of lifetime earnings, schooling and cognition for both males and females. Columns 1 and 2 – the relevant ones for understanding the baseline mediation results – show that cognition and years of schooling are important determinants of lifetime earnings. For men, one more year of schooling increases lifetime earnings by 5.7%, whereas a one standard deviation increase in cognition increases lifetime earnings by 14%. For women, we find a larger return to education and a smaller return to cognition compared to males. The bottom of Table 2.6 reports p-values for the tests of the joint significance of time investments at different ages, school quality at different ages, and family background. These are modified F-tests that account for measurement error in all the variables we use – see Appendix 2.E for details. We find that the null hypothesis of no effect of these categories on lifetime earnings cannot be rejected. This confirms the usual view that parental time investments and family background affect lifetime earnings primarily through the channel of cognition and schooling. Estimates for the relationship between mediators and parental income (equation (2.5)) can be found in Appendix Table 2.F.1.

Level 2 - Allowing for Indirect Effects via Years of Schooling The second columns in Tables 2.4 and 2.5 show results which allow the fraction of the IGE explained by years of schooling to be mediated by the other covariates – cognition, investments and family background. Once we allow for this mediation, the fraction of the IGE explained by years of schooling is no longer statistically different from zero for women and becomes negative for men. For both males and females, the higher educational attainment of individuals from high income households is completely mediated by these three factors, as opposed to being a direct effect of parental income on years of schooling. The most important mediator is cognition, which is consistent with the literature. Multiple previous studies have shown that those with greater cognitive skills are more likely to complete more years of education, in part due to the greater success they appear to have in higher education (Keane and Wolpin (2001), Carneiro and Heckman (2002), Arcidiacono (2005)). For males, we find that cognition, in addition to explaining 33% of the IGE through a direct effect, also accounts for 13% of the IGE through an indirect effect on schooling, thus explaining a total share of 46% of the IGE. The total effect for women is similar - at 40%. Neither the education of parents nor the number of siblings has a significant impact on years of school attainment, once cognition and investments are controlled for.

The estimates in columns 3 and 4 of Table 2.6, which show the determinants of schooling, help to understand the drivers of this result. Cognition is the major determinant of educational attainment - a one standard deviation increase in cognition at age 16 leads to almost a whole additional year of schooling. This estimate is remarkably similar for both males and females.

The F-tests reported at the bottom of the table show that we cannot reject the null hypothesis that time investments, school quality, and family background have no effect on years of schooling, once cognition is controlled for. This highlights the importance of age 16 cognition as a key explanatory variable for lifetime earnings, as it has both an important direct effect on earnings, and also a significant indirect effect through schooling.

Level 3 - Allowing for Indirect Effects via Years of Schooling and Cognition
In Level 3, the results of which are reported in the third columns of Tables 2.4 and 2.5, we allow the effect of parental income on age 16 cognition to be mediated by parental investments and family background. We find that once we consider indirect effects of investments on cognition and schooling, investments become the key driver of the IGE. Over 40% of the IGE can be explained by the increased time and school quality investments of high income parents, mostly through their impact

on cognition. The relative importance of different types of investment varies by gender – time investments are quantitatively more important than school quality for men, whereas for women this is reversed. Both types of investment are important, with time investments more important than school quality early in childhood, and school quality investments becoming relatively more important later in childhood. For example, for men, parental time investments at age 7 contribute much more than school quality investments at age 7 to the IGE (14% versus 0%). However, parental time and school quality investments at age 16 contribute similar amounts to the IGE (17% and 13%, respectively). All investments appear important at age 16, which is of particular interest given the recent interest in early life investments. Interestingly, parental education and number of siblings have only a modest impact on cognition once we control for investments, and thus the shares of the IGE explained by these factors only change slightly between level 2 and level 3.

Table 2.6 presents the drivers of these results. The third panel, which shows the relationship between cognition and each of investments and background, indicates that time investments and school quality have strong effects on cognition at age 16. This result is in keeping with the results from the large recent literature on the estimation of production functions for child ability (Cunha and Heckman (2008), Cunha et al. (2010), Attanasio et al. (2020)).

The education of the mother and the father have a small positive effect on cognition even after controlling for investments. This could be due to direct ability transmission which we cannot capture, or due to differential productivity of higher education parents in producing child ability. We also find a negative effect of the presence of siblings.

Level 4 - Allowing for Indirect Effects via Years of Schooling, Cognition, and Investments In our final mediation stage we additionally allow the effect of parental income on investments to be mediated by measures of family background – parental education and number of siblings. Of these, parental education plays an important role, especially for females, in mediating the effect of investments. Although in our level 3 analysis we find that parental education has little effect once we control for investments, in level 4 we find that parental education is important because it impacts investments. For females, mother's education explains 1% of the IGE when we control for investments, but explains 10% of the IGE in level 4 when we allow investments to be a function of mother's education.

However, the investments of high income parents cannot be fully explained by parental education or by the number of siblings. In Table 2.4 (2.5), the fraction of the IGE explained by investments at level 3 is 0.473 (0.463), yet it remains at 0.325 (0.251) for men (women) once we allow for family background as a mediator in level 4. This suggests a potential direct role for parents' income in determining investments. Turning to Table 2.7, which provides estimates of our investment equation (2.12), we can see strong parental income gradients for time investments at younger ages and for school quality at older ages after controlling for family background. Together with the large impacts of investments on cognition and years of schooling (Table 2.6), we conclude that parental income gradients in investments are a powerful source for perpetuating inequality.

Summary Figure 2.2 summarizes the main results of our mediation analysis. The most striking aspect of the graph for men is the substantial role that cognition plays in explaining the IGE in the baseline, with a comparatively modest impact of family background. This role for cognition is substantially attenuated by level 3 – that is, the seeds for the cognitive advantage enjoyed by the children of richer parents are found in the greater investments that those parents make. For women, we see that cognition and schooling are the most important in the baseline, but by level 4, when these are allowed to be mediated by investments and family background, we find that those latter factors play an important role.

In the next section we provide a summary and compare our results to the literature.

2.5.3 Key Findings and Comparisons to the Literature

In this section, we summarize our main findings and relate them to some important results in the literature.

1) Lifetime earnings are mainly driven by cognition and years of schooling.

Parental investments and family background have no significant effect on the IGE once we control for cognition and years of schooling.

Columns 1 and 2 of Table 2.6 shows that cognition and schooling have sizeable and significant effects on lifetime earnings for men and women, respectively. The return to a year of schooling on lifetime earnings is 5.7% (13.4%) for men (women), which falls within the range of commonly reported returns to schooling, as summarized in Card (1999). For males, the effect of a standard deviation of cognition on lifetime earnings is 13.9%, similar to the estimates in Heckman et al. (2006). Whilst many papers *assume* that earnings is mainly determined by schooling and cognition at the beginning of adulthood (Huggett et al. (2011a), Lee and Seshadri (2019), Daruich (2018)), our analysis confirms that once schooling and cognition are controlled for, other early life factors have no significant effect.

2) The role of years of schooling in explaining the relationship between parents' income and children's earnings is mediated entirely by differences in cognition. This means that direct effects of parental income on schooling are insignificant in our setting.

Our paper contributes to the important question as to whether borrowing constraints limit educational attainment of those from low income families and thus propagate inequality. Consistent with Arcidiacono (2005), Stinebrickner and Stinebrickner (2008), Carneiro and Heckman (2002) and Carneiro et al. (2011), we find a large effect of cognitive skills on schooling decisions. Our finding that parental income has no significant impact on educational attainment, conditional on cognition, is consistent with Belley and Lochner (2007), who find that conditional on ability, parental income did not matter for the schooling decisions of the NLSY79 cohort (although they find that it does for more recent cohorts). Part of the explanation for this may lie in the fact that there were no university tuition fees in the UK for our cohort. We do not find a direct effect of parental education on years of schooling, once we control for cognition. This can be interpreted as evidence for the intergenerational persistence in education being transmitted mostly through cognition (Lee and Seshadri (2019)).

3) Parental investments in time and school quality explain most of the differences in cognition in individuals from low and high income families.

Consistent with recent literature, we find that parental time investments are crucial for child human capital development (see Heckman and Mosso (2014) for a summary). Whilst we cannot rule out potential endogeneity of our investments, several papers have found parents compensate for shocks during childhood (Attanasio et al. (2017), Attanasio et al. (2020)) – that is, parents invest more in children who have lower ability. If this dynamic is at work in our cohort, our estimates of the effect of time investments on cognition will be downward biased and our results regarding the impact of time investments on the IGE would be a lower bound.

Several papers estimate the effect of school quality on lifetime outcomes; the majority use a single measure such as classroom size (Akerhielm (1995), Altonji and Dunn (1996), Dearden et al. (2002), Chetty et al. (2011)). Our approach has the benefit of combining multiple measures of school quality and correcting for measurement error. This leads us to find larger effects of school quality on cognition than, for example, Akerhielm (1995). In Appendix 2.H, we provide further comparisons regarding the effect of school quality on lifetime earnings.

4) We find that income-related differences in child investments (and their consequences) explain 32.5% of the IGE, whereas family background-related differences in child investments account for 18.5% of the IGE. This means that parental income directly affects investments, controlling for family background.

The importance of family circumstances during childhood on human capital investments has recently been documented in the literature. Carneiro et al. (2013) estimate the causal effect of an increase in maternal education on reading to the child at ages 7-8 to be approximately 6% of a standard deviation. This is similar in magnitude to the effect of maternal education on investments which we find at those ages (8-10%). Bhalotra and Clarke (2020) find that an exogenous increase in the number of children in the household due to twin birth decreases the amount that parents read to their children. Our results point in the same direction.

Our finding that it is not only family background, but parental income itself which matters for parental investments is consistent with the growing literature which studies the effect of financial transfers on human capital development (Duncan et al. (2011), Milligan and Stabile (2011), Dahl and Lochner (2012)), and in particular with Caucutt and Lochner (2020), who provide evidence that the mechanism behind this link is that parental income impacts investments in children.

See Appendix 2.H for further comparisons regarding the effect of family size and parental education on lifetime earnings.

2.5.4 The Importance of Correcting for Measurement Error

Throughout this analysis, we have confronted the fact that parental income, cognition, and time and school quality investments are all measured with error. In this section we discuss the importance of allowing for measurement error in these variables. Table 2.8 gives results of our mediation analysis when *not* accounting for measurement error. To run this analysis, instead of using the Bartlett scores for each of factors, we choose the single measure with the highest signal-to-noise ratio for cognition, school quality, and cognition. We then conduct the same mediation analysis as outlined in Section 2.4, but without using the errors-in-variables approach to correct for measurement error.

Although the overall share of the IGE explained is not substantially affected by controlling for measurement error, there are two key differences between the results in Table 2.8 and those we reported in Tables 2.4 and 2.5. First, consistent with the usual attenuating role of measurement error, the estimated coefficients on the potentially error-ridden variables are smaller when not accounting for measurement error, leading to a smaller share explained by our variables of interest. Second, the objects not measured with error explain a larger share of the IGE when not accounting for measurement error. This is as expected. Since all the variables considered are correlated with one another, erroneously understating the impact of one variable will lead us to erroneously overstate the role of the other variables.

2.5.5 Robustness Checks

In this section, we summarize the results of the following robustness checks: including more family background variables, including non-cognitive skills, and allowing for interactions between lifetime earnings and cognition.

Additional Family Background Variables We have focused on key drivers of lifetime earnings that are highlighted in the literature. We find that our results are robust to the inclusion of other variables that are commonly thought to be important for child development, such as marital stability, and parental age. Results are presented in Appendix Table 2.G.2. The coefficients on variables included in the original specification change only very modestly. Moreover, none of the additional family background variables are found to explain a significant fraction of the IGE.

Non-cognitive skills Our approach focuses on the importance of cognition and years of schooling in explaining lifetime earnings. A number of recent papers have highlighted the importance of non-cognitive skills (e.g., Heckman et al. (2006), Papageorge et al. (2019), Todd and Zhang (2020)). To assess the importance of non-cognitive skills in our data, we use the same latent variables approach to measure non-cognitive skills that we use to measure cognition. Following Heckman et al. (2013), we use exploratory factor analysis to determine the appropriate measures to use. Previous papers using the NCDS often exploit teacher-reported behaviors at age 11 as measures of non-cognitive skills (e.g., Papageorge et al. (2019), Blanden et al. (2007)). However, these are not available at age 16. In order to allow for non-cognitive skills to enter into the mediation analysis in a manner analogous to how cognition enters, we thus focus on parent-reported behaviors at age 16 as measures of non-cognitive skills. Using exploratory factor analysis, we retain the following seven out of 14 possible measures for our non-cognitive skills factor: difficulty concentrating, destructive behavior, tearfulness, restlessness, belligerence, irritability, and disobedience. Appendix 2.D.2.4 contains a full list of measures and results of the exploratory factor analysis.

Table 2.G.3 presents the results of the mediation analysis when non-cognitive skills are included. Non-cognitive skills explain a small and insignificant fraction of

the IGE. This holds at the baseline level, but also when we allow for non-cognitive skills to potentially affect schooling. The fractions explained by years of schooling and cognition only marginally decrease when non-cognitive skills are included.

The robustness of our mediation results to the inclusion of non-cognitive skills can be explained by Appendix Table 2.G.4, which shows the determinants of lifetime earnings and years of schooling. The coefficients on non-cognitive skills are statistically insignificant, except for in the females' years of schooling regression, and everywhere are about five times smaller than the coefficients on cognition. For example, a one standard deviation increase in cognition increases men's lifetime earnings by 13.5%, while a standard deviation increase in non-cognitive skills increases their lifetime earnings by 3.8%. Further, comparing the lifetime earnings regressions in Table 2.G.4, which control for non-cognitive ability, with our main results in Table 2.6, we find the effect of cognition on lifetime earnings and years of schooling is essentially unchanged, for both males and females.

Finally, and crucially for this paper, we find that the correlation of parental income with non-cognitive skills is small, especially when compared to the relationship between parental income and cognitive skills. Thus, even if non-cognitive skills were important for lifetime earnings, they would not help explain the IGE in our data. For example, Appendix Table 2.G.5 shows that a 1% increase in parental income is associated with a 0.5% of a standard deviation increase in cognition for females at age 16, while it is associated with only a 0.001% of a standard deviation increase in non-cognitive skills at the same age.

In short, our approach is intentionally parsimonious. We focus on what we have found to be the two most robust measures for predicting lifetime earnings (cognition and years of schooling), and we use measures of investment and family background that are important for determining these.

Complementarity between cognition and years of schooling Throughout we have used equations that are linear in factors. The linear specification facilitates easily interpretable decompositions, especially when using multiple levels. However, Bolt et al. (2021) find evidence of complementarity between schooling and cognition in the wage equation using NCDS data. Therefore we conduct a robustness check in which we allow for an interaction between cognition and schooling in the lifetime earnings equation. We find no major change in the decomposition when allowing for this complementarity. Moreover, the interaction term explains an insignificant fraction of the IGE. See Appendix 2.G.3 for further details.

2.6 Conclusion

In this paper, we quantify the direct and indirect effects of a large number of potential drivers of the IGE by performing a systematic multilevel mediation analysis. Overall, we are able to explain 54% of the IGE for males and 62% for females.

The four mechanisms we model as generating a link between income of parents and children are: years of schooling, cognition of the children, investments the parents make in the children, and family background. Of these, cognition and years of schooling play the most important role when all mechanisms are considered to have only a direct effect on transmission of income across generations. Once we allow for investments at earlier ages to affect cognition and years of schooling at later ages, the perspective changes. The role of years of schooling in explaining the relationship between parents' income and children's earnings is completely mediated by differences in cognition, and the role of cognition is substantially mediated by differences in time investments. Finally, around one third of the differences in investments by rich and poor parents can be attributed to differences in family background. Once we account for the effect of family background and early investments on cognition and years of schooling, we find that family background and investments are by far the most important mechanisms explaining the IGE. Failure to account for measurement error leads to an underestimated role of investments.

Apart from these results being interesting in their own right, we believe that the methodology used in this paper provides a systematic way in which researchers can explore how potentially relevant channels operate, a step which can complement the explicit modeling of decision-making.

Our paper focuses on the link between early life circumstances and lifetime earnings through their impact on human capital. However, we do not address the channels by which human capital impacts earnings in adulthood, whether it be through higher wages or higher employment rates. We also do not consider other important adult outcomes, such as marital choices. Building on the current mediation framework for further investigations of these issues would be a fruitful direction for future research.

Table 2.2: List of ability and investment measures

Cognition	
Age 16	Reading score, Math score, Teacher-assessed Math and English ability
Time Investments	
Age 0-7	Teacher's assessment of parents' interest in child's education, Parent outings with child, Whether parents read to the child
Age 7-11	Teacher's assessment of parents' interest in child's education, Parent outings with child, Parents' ambitions regarding child's educational attainment
Age 11-16	Teacher's assessment of parents' interest in child's education, Parents' ambitions regarding child's educational attainment
School Quality Investments	
Age 0-7	Whether school has a parent-teacher association (PTA), whether PTA conducts educational meetings, Social class of fathers in the child's class, Type of school
Age 7-11	Class size, Student-teacher ratio, Type of school, Proportion of children in class suitable for GCEs
Age 11-16	Student-teacher ratio, Type of school, Proportion of children in child's school who: complete school / move on to full-time degrees / pass 2+ A-levels / are studying for GCEs

Note: Data collected at ages 7,11 and 16 for children includes measures of ability, non-cognitive skills, parental time investments, family circumstances, and parental income. Note, all investment measures are retrospective, so age 0-7 investments are measured at age 7, age 7-11 investments are measured at age 11, age 11-16 investments are measured at age 16.

Table 2.3: IGE estimates

	Male	Female	Male uncorrected	Female uncorrected
IGE	0.317	0.236	0.155	0.115
	(0.097)	(0.105)	(0.045)	(0.050)
N	1350	1347	1350	1347

Note: The first two columns report the IGE estimates for male and female children, after parental income has been corrected for measurement error. The third and fourth column report estimates without this correction.

Table 2.4: Multilevel Mediation Analysis: Share of IGE Explained - Males

	Males			
	Level 1	Level 2	Level 3	
Years of Schooling	**0.095**	**-0.079**	**-0.079**	**-0.079**
	[0.020, 0.191]	[-0.189, -0.005]	[-0.189, -0.005]	[-0.189, -0.005]
Cognition	**0.327**	**0.456**	0.106	0.106
	[0.164, 0.652]	[0.258, 0.931]	[-0.069, 0.329]	[-0.069, 0.329]
Investments	0.135	0.187	**0.473**	**0.325**
	[-0.138, 0.393]	[-0.087, 0.480]	[0.175, 0.930]	[0.084, 0.643]
Time Investments	0.127	0.173	**0.384**	**0.284**
	[-0.090, 0.331]	[-0.029, 0.431]	[0.149, 0.791]	[0.085, 0.589]
... Age 7	0.126	0.147	0.143	0.111
	[-0.080, 0.441]	[-0.059, 0.450]	[-0.075, 0.441]	[-0.067, 0.325]
... Age 11	-0.054	-0.027	0.076	0.066
	[-0.392, 0.183]	[-0.341, 0.223]	[-0.162, 0.386]	[-0.147, 0.326]
... Age 16	0.056	0.053	**0.166**	**0.108**
	[-0.008, 0.146]	[-0.010, 0.147]	[0.080, 0.304]	[0.030, 0.205]
School Quality	0.008	0.014	0.089	0.041
	[-0.124, 0.135]	[-0.121, 0.142]	[-0.076, 0.234]	[-0.106, 0.126]
... Age 7	-0.001	-0.001	0.000	-0.000
	[-0.028, 0.021]	[-0.029, 0.019]	[-0.020, 0.020]	[-0.029, 0.022]
... Age 11	-0.024	-0.023	-0.044	-0.030
	[-0.137, 0.075]	[-0.139, 0.080]	[-0.179, 0.051]	[-0.135, 0.026]
... Age 16	0.033	0.038	**0.133**	0.072
	[-0.017, 0.085]	[-0.016, 0.094]	[0.048, 0.231]	[-0.006, 0.140]
Family Background	-0.019	-0.027	0.037	**0.185**
	[-0.131, 0.089]	[-0.146, 0.078]	[-0.089, 0.166]	[0.063, 0.453]
Mother's education	-0.051	-0.049	-0.029	0.020
	[-0.137, 0.012]	[-0.139, 0.015]	[-0.103, 0.038]	[-0.039, 0.095]
Father's education	0.016	0.008	0.035	0.084
	[-0.077, 0.096]	[-0.084, 0.088]	[-0.063, 0.114]	[-0.008, 0.184]
Number of Siblings	0.016	0.014	0.031	**0.081**
	[-0.012, 0.053]	[-0.015, 0.047]	[-0.001, 0.085]	[0.008, 0.191]
Total	0.538	0.538	0.538	0.538
N	1350	1350	1350	1350

Notes: 95% Confidence intervals constructed using 250 clustered bootstrap replications. Coefficients that are significant at the 5% level are **bold**. The row labelled 'Investments' shows the fraction explained by the sum of all investments, i.e. time investments and school quality. The row labelled 'Time Investments' shows the fraction explained by the sum of all time investments, i.e. for ages 7, 11, 16. Similarly for 'School Quality' and 'Family Background'.

Table 2.5: Multilevel Mediation Analysis: Share of IGE Explained - Females

	Females			
	Level 1	Level 2	Level 3	Level 4
Years of Schooling	**0.425**	0.024	0.024	0.024
	[0.165, 1.361]	[-0.249, 0.340]	[-0.249, 0.340]	[-0.249, 0.340]
Cognition	**0.135**	**0.402**	0.002	0.002
	[0.007, 0.508]	[0.185, 1.237]	[-0.286, 0.204]	[-0.286, 0.204]
Investments	0.050	0.151	**0.463**	0.251
	[-0.271, 0.302]	[-0.142, 0.560]	[0.126, 1.527]	[-0.049, 0.947]
Time Investments	-0.100	-0.046	0.114	0.039
	[-0.428, 0.083]	[-0.372, 0.157]	[-0.106, 0.655]	[-0.223, 0.325]
... Age 7	0.143	0.157	**0.176**	0.105
	[-0.010, 0.690]	[-0.005, 0.669]	[0.002, 0.824]	[-0.006, 0.458]
... Age 11	**-0.180**	**-0.175**	-0.133	-0.083
	[-0.705, -0.017]	[-0.702, -0.019]	[-0.544, 0.037]	[-0.385, 0.034]
... Age 16	-0.062	-0.029	**0.070**	0.016
	[-0.285, 0.012]	[-0.156, 0.054]	[0.008, 0.289]	[-0.055, 0.137]
School Quality	**0.150**	**0.198**	**0.349**	**0.212**
	[0.008, 0.483]	[0.038, 0.698]	[0.139, 1.187]	[0.014, 0.619]
... Age 7	0.047	0.044	0.046	0.032
	[-0.016, 0.219]	[-0.021, 0.213]	[-0.020, 0.228]	[-0.044, 0.187]
... Age 11	0.019	0.022	0.016	-0.010
	[-0.061, 0.109]	[-0.064, 0.098]	[-0.074, 0.080]	[-0.129, 0.030]
... Age 16	0.084	**0.132**	**0.287**	**0.191**
	[-0.020, 0.300]	[0.021, 0.408]	[0.102, 0.851]	[0.043, 0.591]
Family Background	0.006	0.039	0.128	**0.340**
	[-0.288, 0.221]	[-0.220, 0.323]	[-0.089, 0.554]	[0.136, 1.174]
Mother's education	-0.043	-0.024	0.010	0.104
	[-0.304, 0.152]	[-0.300, 0.173]	[-0.230, 0.262]	[-0.047, 0.567]
Father's education	0.068	0.081	0.126	**0.227**
	[-0.088, 0.277]	[-0.079, 0.348]	[-0.018, 0.418]	[0.076, 0.679]
Number of Siblings	-0.019	-0.019	-0.008	0.009
	[-0.140, 0.027]	[-0.146, 0.028]	[-0.099, 0.023]	[-0.033, 0.063]
Total	0.616	0.616	0.616	0.616
N	1347	1347	1347	1347

Notes: 95% Confidence intervals constructed using 250 clustered bootstrap replications. Coefficients that are significant at the 5% level are **bold**. The row labelled 'Investments' shows the fraction explained by the sum of all investments, i.e. time investments and school quality. The row labelled 'Time Investments' shows the fraction explained by the sum of all time investments, i.e. for ages 7, 11, 16. Similarly for 'School Quality' and 'Family Background'.

Table 2.6: Determinants of Lifetime Earnings, Schooling, Cognition

	Lifetime Earnings		**Years of Schooling**		**Cognition**	
	Males	Females	Males	Females	Males	Females
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Schooling	**0.057**	**0.134**				
	(0.013)	(0.014)				
Cognition	**0.139**	0.064	**0.973**	**0.944**		
	(0.030)	(0.036)	(0.054)	(0.062)		
Time Investments						
Age 7	0.041	0.062	0.124	0.047	-0.006	0.043
	(0.040)	(0.043)	(0.090)	(0.087)	(0.066)	(0.069)
Age 11	-0.015	-0.076	0.128	0.016	0.142	0.094
	(0.045)	(0.041)	(0.101)	(0.078)	(0.091)	(0.088)
Age 16	0.032	-0.042	-0.024	**0.169**	**0.328**	**0.349**
	(0.023)	(0.027)	(0.053)	(0.052)	(0.036)	(0.032)
School Quality Investments						
Age 7	-0.012	0.047	0.006	-0.019	0.078	0.008
	(0.060)	(0.025)	(0.169)	(0.047)	(0.056)	(0.029)
Age 11	-0.013	0.025	0.011	0.024	-0.057	-0.037
	(0.034)	(0.030)	(0.053)	(0.136)	(0.038)	(0.068)
Age 16	0.024	0.035	0.071	**0.151**	**0.362**	**0.342**
	(0.022)	(0.023)	(0.050)	(0.049)	(0.029)	(0.032)
Family Background						
Mother's education	-0.025	-0.010	0.020	0.031	**0.051**	0.041
	(0.020)	(0.023)	(0.041)	(0.055)	(0.024)	(0.025)
Father's education	0.007	0.015	-0.062	0.022	**0.057**	**0.050**
	(0.020)	(0.017)	(0.038)	(0.036)	(0.025)	(0.022)
Number of siblings	-0.013	0.025	0.030	-0.001	**-0.070**	**-0.072**
	(0.013)	(0.016)	(0.029)	(0.032)	(0.021)	(0.020)
Log parental income	0.146	0.091	-0.440	0.042	0.173	0.002
	(0.110)	(0.111)	(0.252)	(0.236)	(0.165)	(0.124)
P-values for joint significance:						
Time Investments	0.708	0.842	0.490	0.315	0.096	0.031
School Quality	0.501	0.285	0.424	0.183	0.017	0.009
Family Background	0.291	0.276	0.218	0.408	0.012	0.020
N	1350	1347	1350	1347	1350	1347

Note: Lower panel shows p-values of F-tests for joint significance of time investments at all ages, school quality at all ages, and all family background variables. The null hypothesis is that the relevant group of variables jointly have zero coefficient in the corresponding regression. Importantly, the F-tests test restrictions while accounting for measurement error in the data. See Appendix (2.E) for more details.

Table 2.7: Determinants of Parental Investments

Time Investments

	Males			Females		
	Age 7	Age 11	Age 16	Age 7	Age 11	Age 16
Mother's Education	**0.080**	0.064	**0.112**	**0.095**	**0.110**	**0.182**
	(0.037)	(0.039)	(0.026)	(0.030)	(0.039)	(0.028)
Father's Education	0.058	0.024	**0.090**	**0.068**	0.055	0.045
	(0.031)	(0.038)	(0.025)	(0.027)	(0.035)	(0.025)
Number of Siblings	**-0.310**	**-0.237**	**-0.136**	**-0.261**	**-0.205**	**-0.175**
	(0.022)	(0.024)	(0.018)	(0.025)	(0.028)	(0.020)
Log parental income	**0.746**	**1.021**	**0.362**	0.323	0.350	0.081
	(0.210)	(0.205)	(0.148)	(0.178)	(0.188)	(0.161)
N	1350	1350	1350	1347	1347	1347

School Quality Investments

	Males			Females		
	Age 7	Age 11	Age 16	Age 7	Age 11	Age 16
Mother's Education	0.052	0.058	**0.104**	0.023	**0.169**	0.056
	(0.039)	(0.049)	(0.033)	(0.086)	(0.067)	(0.034)
Father's Education	0.004	**0.171**	**0.124**	0.035	**0.103**	**0.116**
	(0.039)	(0.056)	(0.036)	(0.047)	(0.051)	(0.027)
Number of Siblings	-0.034	**-0.052**	**-0.084**	-0.056	-0.026	-0.023
	(0.026)	(0.019)	(0.017)	(0.051)	(0.016)	(0.021)
Log parental income	-0.032	0.410	0.231	0.165	-0.113	**0.374**
	(0.178)	(0.217)	(0.141)	(0.314)	(0.237)	(0.172)
N	1350	1350	1350	1347	1347	1347

Note: These tables show the regression of investments on family background variables. This provides some explanation of the difference between levels 3 and 4 of our mediation analysis. All investment measures here are corrected for measurement error.

Table 2.8: Main Decomposition without Measurement Error Correction

	Males			
	Level 1	Level 2	Level 3	Level 4
Years of School	**0.177**	0.007	0.007	0.007
	[0.059, 0.439]	[-0.149, 0.109]	[-0.149, 0.109]	[-0.149, 0.109]
Cognition	**0.175**	**0.294**	0.123	0.123
	[0.087, 0.468]	[0.157, 0.745]	[-0.014, 0.351]	[-0.014, 0.351]
Investments	**0.132**	**0.175**	**0.287**	**0.178**
	[0.004, 0.382]	[0.042, 0.483]	[0.111, 0.732]	[0.044, 0.493]
Family Background	0.018	0.027	0.085	**0.194**
	[-0.124, 0.168]	[-0.115, 0.191]	[-0.049, 0.315]	[0.029, 0.563]
Total	0.502	0.502	0.502	0.502
N	1092	1092	1092	1092

	Females			
	Level 1	Level 2	Level 3	Level 4
Years of School	**0.530**	0.104	0.104	0.104
	[0.234, 1.980]	[-0.234, 0.584]	[-0.234, 0.584]	[-0.234, 0.584]
Cognition	0.042	**0.282**	**0.132**	**0.132**
	[-0.079, 0.198]	[0.110, 1.007]	[0.010, 0.616]	[0.010, 0.616]
Investments	0.044	**0.153**	**0.243**	**0.136**
	[-0.121, 0.441]	[0.010, 0.791]	[0.083, 1.094]	[0.005, 0.702]
Family Background	0.009	0.087	0.146	**0.254**
	[-0.268, 0.299]	[-0.136, 0.515]	[-0.057, 0.686]	[0.053, 1.016]
Total	0.626	0.626	0.626	0.626
N	1127	1127	1127	1127

Note: This replicates the mediation analysis in Table 4, except it uses an individual measure rather than latent factors for each of cognition, time investments and school quality; and we do not correct for measurement error.

Appendix

2.A Data

We use only one dataset - the National Child Development Study (NCDS) - to estimate the IGE and the equations for the mediation analysis. The NCDS is a panel tracking a cohort of individuals born in the UK in a particular week in March 1958. In this section, we describe how the parental income and children's lifetime earnings variables are constructed from the NCDS data, as well as how we select our analytical sample.

2.A.1 Variable Construction

2.A.1.1 Parental Income

Parental income variables are only observed in the third wave of the NCDS, when cohort members are aged 16. There are three relevant variables: fathers' net earnings, mothers' net earnings and other net income. We drop individuals whose parents did not answer any of the earnings questions. Adopting Blanden et al. (2013)'s procedure, we also drop households where at least one parent is working but that parents' earnings is unobserved. Finally, we restrict age of father at the time of the survey at birth to be between 20 and 45, to ensure that the point-in-time income measure is a reasonable proxy for lifetime income (a detailed breakdown of how we arrive to the final sample can be found in 2.A.2).

Fathers' and mothers' earnings and other net income are recorded as interval, or "banded" data. We use the continuous parental income variables which were imputed by the Institute for Fiscal Studies as part of an effort to harmonize income variables in different cohort studies (Belfield et al. (2017)). This procedure can be summarized by three steps. First, income bands identical to the NCDS are created for the 1973, 1974 and 1975 Family Expenditure Survey (FES) data. Second, within each income band, net male, net female and net other income data are estimated using variables that are thought to be correlated with income.[12]. Third, using these prediction equations, each of the three income components are separately predicted within each income band in the NCDS, using the same key covariates. The sum of the three net, predicted earnings variables in the NCDS is the measure of parental income that we use.

2.A.1.2 Children's Earnings

NCDS cohort members' gross wages are observed at ages 23, 33, 42, 50, and 55. Gross earnings are calculated using usual wages and pay period on the respondent's main job. Current or last wage variables are used where usual wages are unavailable.

Based on NCDS employment spells data from 1974-2008, we apply three criteria for sample selection as in Gregg et al. (2016). Firstly, cohort members are excluded if less than 60 months of working history is observed. Secondly, cohort members with no earnings observations across the five survey waves are excluded. Finally, cohort members who do not have parental income observation from childhood are not considered in our analysis.

The crucial step in dealing with this sample is in how we interpret missing and zero earnings figures for cohort members, which could be due to attrition, item non-response or non-employment. To address this, we adapt Gregg et al. (2016)'s method of imputing missing earnings at the time of a survey. For this imputation, earnings are modeled as a function of an individual-specific fixed effect and a categorical education variable interacted with age of cohort member:

$$\ln Y_{it} = f_i + \chi \cdot S_i^* \cdot t + \zeta_{it}$$

[12]The following variables are used to estimate income data within each band: year of interview, age left school of mother and father, employment status of mother and father, age of mother and father, occupation of mother and father, number of siblings of the cohort member, housing tenure, region, number of rooms in household, number of cars, marital status, whether benefits are received by household, type of school cohort member attends, interactions between age left school and occupation of mother and father, interactions between age left school and employment status of mother and father, and interactions between year of interview and occupation of mother and father

where $S_i^* = 1$ if less than O level, 2 if O level qualification, 3 if A level qualification, 4 if some college or above.[13] Once earnings at ages 23, 33, 42, 50, and 55 have been predicted as $\widehat{\ln Y_{it}} = \hat{f}_i + \hat{\chi} \cdot S_i^* \cdot t$, the earnings value for individuals recorded as being out of work at the time of a survey is replaced with zero. This accounts for the fact that missing earnings in those cases represented non-employment, as opposed to attrition. Monthly earnings are calculated from earnings at each survey date by imputing a linear trajectory between each earnings data point. Aggregating monthly earnings provides an estimate of children's lifetime earnings.

2.A.2 Sample Selection

The initial wave of the NCDS consists of 18,558 individuals. We impose the following requirements for individuals to be in the mediation sample:

- Parental income available (see section 2.A.1.1 for details) - drop 12,170 observations

- Father's age is between 20 and 45 at the time child was born - drop 682 observations

- Children's years of school available - drop 1,572 observations

- Children's lifetime earnings available (see section 2.A.1.2 for details) - drop 117 observations

- Mother and father's years of schooling available - drop 151 observations

- Number of siblings is available - drop 1 observation

- At least one measure of school quality available at each age 7,11,16 - drop 1,048 observations

- At least one measure of time investment available at each age 7,11,16 - drop 106 observations

[13] O levels and A levels were national exams taken at approximately age 16 and 18 respectively.

- At least one measure of ability available at each age 0,7,11,16 - drop 14 observations

The final sample consists of 1350 males and 1347 females.

2.A.3 Further Descriptives

Tables 2.A.1 and 2.A.2 include descriptive statistics on parental investment measures which we use for our analysis. We also report the gradients by parental income tertile, and F-tests on mean-equivalence across the groups.

Table 2.A.1: Time Investments: Descriptive Statistics and Means by Parental Income Tertile

			Parental Income Tertile			
Variable	Mean	SD	Bottom	Middle	Top	P-values
Time Investments						
Age 7						
% of fathers very interested in child's education	21.5	41.1	18.6	21.5	24.4	0.01
% of mothers very interested in education	34.4	47.5	31.5	34.1	37.4	0.03
% of fathers who go on weekly outings with child	69.8	45.8	65.2	72.5	71.5	0.00
% of mothers who go on weekly outings with child	85.0	35.7	82.8	85.5	86.8	0.07
% of fathers who read to child every week	35.5	47.9	35.9	37.6	32.9	0.12
% of mothers who read to child every week	47.3	49.9	46.0	50.5	45.2	0.06
Age 11						
% of fathers very interested in child's education	23.5	42.4	20.1	23.8	26.4	0.01
% of mothers very interested in child's education	33.4	47.2	29.8	34.3	36.1	0.02
% of fathers who go on weekly outings with child	50.7	50.0	45.0	53.6	53.1	0.00
% of mothers who go on weekly outings with child	53.5	49.9	49.4	56.0	55.1	0.01
% of parents expecting child to finish school	74.5	43.6	70.0	73.6	80.0	0.00
% of parents expecting child to attend university	83.1	37.5	81.2	82.8	85.2	0.08
Age 16						
% of fathers very interested in child's education	29.9	45.8	28.2	28.6	32.9	0.08
% of mothers very interested in child's education	33.3	47.1	31.5	32.8	35.6	0.19
% of parents expecting child to attend university	26.6	44.2	24.1	25.1	30.6	0.01

Note: This table presents descriptive statistics and means by parental income tertile for all of the investment measures we use. The relevant sample consists of 2,637 individuals. The final column reports *p-values* from the *F-tests* testing the null hypothesis of equality of means across income tertiles.

Table 2.A.2: School Quality Investments: Descriptive Statistics and Means by Parental Income Tertile

			Parental Income Tertile			
Variable	Mean	SD	Bottom	Middle	Top	P-values
School Quality Investments						
Age 7						
Private school (in %)	1.45	11.9	1.34	0.89	2.11	0.10
% whose school has a PTA	15.5	36.2	16.9	15.9	13.5	0.11
% whose PTA holds educational meetings	57.7	49.4	56.8	57.6	58.7	0.71
Social class of fathers in child's class	3.04	0.61	3.09	3.07	2.98	0.00
% of children discussed in PTA meetings	48.7	31.1	45.8	48.9	51.4	0.00
Age 11						
Private school (in %)	1.00	10.1	0.92	0.58	1.66	0.08
Class size	35.0	6.70	34.5	35.1	35.5	0.01
Student-teacher ratio	24.8	9.35	25.4	24.7	24.3	0.06
% from child's class suitable for GCEs	23.6	14.0	23.4	23.1	24.3	0.20
Age 16						
Private school (in %)	1.93	13.5	1.22	0.98	3.29	0.00
% from child's school who complete school	55.7	25.8	53.5	54.2	59.6	0.00
% from child's school now in full-time degrees	1.02	1.71	0.91	0.95	1.19	0.00
% from child's school who passed 2+ A-levels	1.84	2.76	1.68	1.72	2.12	0.00
% from child's class studying for GCEs	46.3	31.2	44.0	44.4	50.5	0.00
Student-teacher ratio	17.2	1.93	17.2	17.3	17.1	0.05

Note: This table presents descriptive statistics and means by parental income tertile for all of the investment measures we use. The relevant sample consists of 2,637 individuals. The final column reports *p-values* from the *F-tests* testing the null hypothesis of equality of means across income tertiles.

2.B Measurement Error in Parental Income

Our goal is to estimate the IGE, which is the slope coefficient in the OLS regression of child's log lifetime earnings on parent's log lifetime income, where both variables are de-meaned. It is ρ in the following regression:

$$\ln Y_i = \rho \ln Y_{Parent,i} + u_i$$

However, parent's lifetime income is unobserved in the NCDS. It is proxied using income observed in the year the children turn 16 (and median age of their parents is close to 42). Previous research has shown that attenuation bias from such proxying can be significant (Haider and Solon (2006)). To address this problem, we use the errors-in-variables framework described below.

First, consider the log-linear projection of age 42 income on lifetime income for parents:

$$\ln Y_{Parent,i,42} = \varphi_{Parent} \ln Y_{Parent,i} + \varepsilon_{Parent,i} \tag{2.15}$$

Allowing φ_{Parent} to be different than one makes this framework more general than the textbook errors-in-variables framework. We assume the following:

- $Cov(\ln Y_{Parent,i}, u_i) = 0$
- $Cov(\ln Y_{Parent,i}, \varepsilon_{Parent,i}) = 0$
- $Cov(u_i, \varepsilon_{Parent,i}) = 0$

Using this projection and these assumptions, the slope coefficient that we can estimate using observables is ρ^*, the slope coefficient from the regression of child's lifetime earnings on parents' age 42 income:

$$\ln Y_i = \rho^* \ln Y_{Parent,i,42} + u_i$$

Note that by substituting for $\ln Y_i$ using the IGE regression and then substituting for $\ln Y_{Parent,i,42}$ using the log-linear projection we can rewrite ρ^* to be:

$$\rho^* = \frac{\mathbb{E}[\ln Y_i \ln Y_{Parent,42} - \mathbb{E}[\ln Y_{Child}]\mathbb{E}[\ln Y_{Parent,42}]]}{\mathbb{E}[\ln Y_{Parent,42}{}^2 - \mathbb{E}^2[\ln Y_{Parent,42}]]} = \rho \frac{\varphi_{Parent} Var(Y_{Parent})}{\varphi^2_{Parent} Var(Y_{Parent}) + Var(\varepsilon_{Parent})} \tag{2.16}$$

Thus, ρ^* is equal to the true IGE multiplied by the reliability ratio, defined:

$$RR_{Parent} = \frac{\varphi_{Parent} Var(Y_{Parent})}{\varphi^2_{Parent} Var(Y_{Parent}) + Var(\varepsilon_{Parent})}$$

We can estimate ρ^* using available NCDS data. However, this does not identify ρ because the reliability ratio for the parent generation, RR_{Parent}, is unobserved. Our approach to identify RR_{Parent} is to assume that the relationship between point-in-time income and lifetime income is the same across both generations, that is, $\varphi_{Parent} = \varphi_{Child}$. We also assume that the variance of measurement error in lifetime income when using point in time income as a proxy is the same across the two generations, i.e. $Var(\varepsilon_{Parent}) = Var(\varepsilon_{Child})$. Here, φ_{Child} and $Var(\varepsilon_{Child})$ are identified using equation (2.15) for the children. We then calculate $Var(Y_{Parent})$ from equation (2.15) and arrive at an estimate of RR_{Parent}. The final step is to divide our estimate of ρ^* in equation (2.17) by RR_{Parent} to arrive at ρ. Adjusting our main estimates for measurement error pushes the IGE up from 0.155 to 0.317 for males and from 0.115 to 0.236 for females.

2.C Different measures of intergenerational elasticity of earnings and income

In our estimates of the intergenerational elasticity of earnings we use children's gross (pre-tax) lifetime earnings and net (post-tax) parental income (father's and mother's net earnings, and other net income) when children are 16. This choice is made because the NCDS only reports net income for parents, while gross earnings for children are found to haver significantly less attrition than net earnings for children. Moreover, the components included in the net income measures are different for parents and children.

In this appendix, we report five alternative estimates of the IGE, and compare them to our preferred measures. None of the measures here are corrected for measurement error.

Table 2.C.1: Different measures of the IGE

	Measure of children's earnings	Measure of parent's earnings	IGE for Males	IGE for Females
	Full Sample			
(1)	Gross lifetime earnings	Within-band	0.155	0.115
		imputed net income	(0.045)	(0.050)
(2)	Gross mean annual earnings	Within-band	0.120	0.141
		imputed net income	(0.053)	(0.050)
(3)	Gross lifetime earnings	Within-band	0.139	0.106
		median net income	(0.050)	(0.053)
	N		1350	1347
	Restricted Sample			
(4)	Gross lifetime earnings	Within-band	0.211	0.047
		imputed net income	(0.063)	(0.089)
(5)	Net lifetime earnings	Within-band	0.179	0.090
		imputed net income	(0.043)	(0.059)
(6)	Net family lifetime earnings	Within-band	0.096	0.122
		imputed net income	(0.045)	(0.044)
N		404	377	

Note: Row (1) presents our main baseline estimates of the IGE using children's gross lifetime earnings and net parental income when children are 16. Each subsequent row changes one aspect of the estimation, holding all other aspects constant. Gross mean annual earnings is the simple average of children's gross earnings at 23, 33, 42, 50 and 55 multiplied by length of child's working life. Within-band median income uses medians for each band from the FES when imputing for net parental income at 16. Net lifetime earnings is constructed by using take-home pay at 23, 33, 42 and 50 for children and our fixed-effects imputation. Net lifetime family earnings sums across the imputed net lifetime earnings of the child and their partner. Note, row (4) re-estimates the same IGE measure as in row (1) but on a restricted sample, for comparison with estimates in rows (5) and (6).

Table 2.C.1 presents our different versions of the IGE. Our main estimates without measurement error corrections for parental income are 0.155 and 0.115 for males and females, respectively. As we discuss in the main text, corrected for measurement error, the estimates become 0.317 and 0.236.

Using a simple mean of point-in-time earnings over the time we see them to construct children's lifetime earnings instead of our fixed-effects imputation pushes our estimate of the IGE down by 3.5 percentage points for males and pushes it up by 3 percentage points for females. This is reported in row (2) of Table 2.C.1.

In our main estimate, we predict the three components of parent's income within each income band, using characteristics that are correlated with income. An alternative procedure is to directly calculate median income within each income band from the FES, and replace interval data in the NCDS with the corresponding medians. Row (3) of Table 2.C.1 presents the IGE using this alternative measure of parents' income. This reduces our estimate of the IGE by no more than 2 percentage points. Gregg et al. (2016) find an estimate of 0.18 for males when they use this construction of parents' income, although their sample selection criteria differ from ours. Using their criteria, we obtain an IGE estimate of 0.178 when median income is used to construct parental income.

Belfield et. al (2017) make a distinction between net and gross earnings for children. Our baseline uses gross lifetime earnings, while the estimate in row (5) of Table 2.C.1 uses the child's net lifetime earnings. To construct net lifetime earnings, we simply use take-home pay at ages 23, 33, 42 and 50 instead of gross wage. The NCDS measures take-home pay as "pay after deductions for tax and National Insurance, including any overtime, bonus, commission or tips". Since data on take-home pay is only available for some observations in our main sample, we re-estimate our baseline IGE using the restricted sample for comparison (this is presented in row (4)). This gives us 0.211 and 0.047 for males and females, respectively. Using net instead of gross lifetime earnings for children pushes our IGE for males down to 0.179 , and pushes the IGE for females up to 0.090, both relative to our baseline calculated on the restricted sample.

A final way to estimate the IGE is to use family income for the children as well as for the parents. Belfield et. al (2017) find an IGE of 0.203 when they use son's gross family income at 42 and parent's net income from the NCDS. We find estimates of 0.096 for males and 0.122 for females but we use net lifetime family income for children, which is the sum of net lifetime earnings of the NCDS cohort member and their partner.

Overall, we find that the uncorrected IGE is relatively robust to the measure of child's earnings used. We prefer using our fixed effect imputed child gross lifetime earnings measure, as it minimizes problems from attrition, and is the most accurate reflection of lifetime earnings of the child.

2.D Estimation

2.D.1 Choice of Measures and Number of Factors - Exploratory Factor Analysis

As described in Section 2.4.3.1, we conduct an exploratory factor analysis (EFA) to select measures of investments and human capital. EFA is a statistical technique for data reduction. It reduces the number of variables in an analysis by describing linear combinations of the measures using a smaller number of "factors" that contain most of the information. By taking the eigenvalues of the data matrix of measures it allows us to understand how the different measures are related to one another. Based on the Kaiser rule, which suggests to only retain only factors with an eigenvalue greater than one, we retain only two factors which we label time investments and school quality.

Table 2.D.1 presents the result for investments. As can be seen from the tables, for ages 11 and 16, the measures clearly load onto two different dimensions: one that is related to parental behaviours, and another that is related to the school environment. It is also noticeable that both dimensions of investments have eigenvalues greater than one. One variable at age 16 – whether parents wish for their child to continue to further education – loads similarly on both time investments and school quality. As we are using a dedicated measurement system where each measure can only load onto one factor, we drop this measure. At age 7, parental interest in education does not load onto the same factor as outings and reading with the child. However, in order to ensure that measures capture the same underlying factor at different ages, we decide to assign it to the time investment category (as in the other periods) rather than the school quality investment category in the final analysis. Classroom size at age 7 loads similarly on both time and school quality, hence we drop it from the final analysis. Finally, following Heckman et al. (2013), we exclude measures with low loadings from the final analysis.

Turning to human capital at age 16, Table 2.D.2 present the results of the EFA. We see the measures load onto what is commonly considered a cognitive and a non-cognitive dimension of human capital. It should be noted that only the first two eigenvalues are greater than one, thus indicating that there is no third dimension (such as a second non-cognitive dimension) for this sample. Again, measures with low loadings are excluded from the final analysis.

2.D.2 Estimating the Cognition, Parental Investment, Schooling Choice, and Lifetime Earnings Equations

Here we discuss how to estimate the production function for: cognition at 16, investment decision rules, schooling choices, and lifetime earnings equations. Estimating these equations is not straightforward because both both the left- and the right-hand-side variables in the equations are latent factors that we only observe through multiple error ridden measures. We closely follow Heckman et al. (2013) and use an errors-in-variables approach to address measurement error.

There are three main steps when using our noisy measures of skills, school quality, investments, and parental income. First, we establish how each individual measure relates to the underlying unobserved latent factor. This is what we lay out in Sections 2.D.2.1, 2.D.2.2 and 2.D.2.3, where we describe the measurement system, the statistical assumptions on the measurement system, and estimation of the measurement system. Second, we combine these different measures into one comprehensive index that can be used for estimation. We do this by estimating Bartlett scores, described in Section 2.D.2.4. Third, in Section 2.D.2.5 we correct for the remaining measurement error in the Bartlett scores using an errors in variables formula.

2.D.2.1 Measurement System

We do not observe parental income (Y_{Parent}), children's skills (C), parental time investments (inv), or school quality (sq) directly, and thus they are latent factors. Instead, we observe multiple error-ridden measurements of each.[14] These measures

[14] We do not have multiple measures of parents' income, but we show that we can use multiple measures of children's earnings in our errors in variables framework in appendix 2.B.

have arbitrary scale and location. In particular:

$$Z_{\Omega,i,t,j} = \mu_{\Omega,t,j} + \lambda_{\Omega,t,j}\Omega_{i,t} + \varepsilon_{\Omega,i,t,j} \tag{2.17}$$

for $\Omega \in \{Y_{Parent}, C, inv, sq\}$ and each $j = \{1, ..., J_{\Omega,t}\}$ error-ridden measurements of each latent factor Ω.

2.D.2.2 Assumptions on Measurement Errors and Normalizations

Measurement Errors Measurement errors are assumed to be independent across individuals, measures, and time. Measurement errors are also assumed to be independent of the latent variables, and all other controls and shocks. In particular, we make the following assumptions on our measurement model:

1. $\varepsilon_{\Omega,t,j} \perp \varepsilon_{\Omega,t,j'}$ for all t, Ω and $j \neq j'$
2. $\varepsilon_{\Omega,t,j} \perp \varepsilon_{\Omega,t',j'}$ for all Ω and $t \neq t'$ and j, j'
3. $\varepsilon_{\Omega,t,j} \perp \Omega'_{t'}$ for all Ω, Ω', t, t', and j
4. $\varepsilon_{\Omega,t,j} \perp X_{t'}$ for all Ω, X, t, t'
5. $\varepsilon_{\Omega,t,j} \perp u_{t'}$ for all Ω, t, t' where $u_{t'}$ represents a structural shock (for investment, cognition, schooling)

Although we drop the i subscripts for notational convenience, all of the above independence assumptions hold for each individual i. Furthermore, all measurement errors are assumed independent across individuals.

Assumption 1 is that measurement errors are independent contemporaneously across measures. Assumption 2 is that measurement errors are independent over time. Assumption 3-5 are that measurement errors in any period are independent of the latent factors (Assumption 3), covariates (Assumption 4) and structural shocks (Assumption 5) in any period. While these assumptions are strong, they are common in the literature.

Normalizations Our latent factors do not have a natural scale or location. We thus normalize the latent factors to have mean zero and variance one in every period:

$$Var(\Omega_t) = 1 \tag{2.18}$$

$$E[\Omega_t] = 0 \tag{2.19}$$

2.D.2.3 Estimation of Measurement Parameters

Using the measurement system in appendix 2.D.2.1 and the statistical assumptions made in appendix 2.D.2.2, here we describe the procedure to estimate the measurement parameters.

1. **Location parameters (μs) in measurement equations**

   In all periods, we normalize the mean of the latent factors C, inv, and sq to zero. Furthermore, we de-mean all of our measures, $Z_{\Omega,t,j}$. Therefore, using equation (3.17), we note that:

$$\mu_{\Omega,t,j} = \mathbb{E}[Z_{\Omega,t,j}] = 0, \tag{2.20}$$

2. **Variance of latent factors**

   Using equation (3.17) and our normalization of the variance of the latent factor, we can derive the scale parameters of each of the latent factors from the covariances of the observed measures:

$$Cov(Z_{\Omega,t,j}, Z_{\Omega,t,j^*}) = \lambda_{\Omega,t,j}\lambda_{\Omega,t,j^*}\, Var(\Omega_t) \tag{2.21}$$

   Note that as long as we have at least three measures, we can identify the scaling parameters. For example, if we have three measures, we have three covariances available, and three λs to estimate.

3. **Variance of the measurement error**

   Finally, we can estimate the measurement error variance for each measure using the observed variance of our measures, our estimated scaling parameters, and the normalization of the latent factor:

$$Var(Z_{\Omega,t,j}) = \lambda^2_{\Omega,t,j} Var(\Omega_t) + Var(\varepsilon_{\Omega,t,j}) \tag{2.22}$$

2.D.2.4 Predicting Children's Skills, Parental Time Investments and School Quality using Bartlett Scores

We have multiple measures of cognition, time investments, and school quality. In this section, which borrows heavily from Heckman et al. (2013), we show how to use the Bartlett score method to take a weighted average of these measurements. In the Bartlett score method, the weights are constructed so that noise from measurement error is minimized. In particular, the procedure is as follows:

Step 1 Estimate measurement system as described above.

Step 2 For each individual, predict Bartlett score (after demeaning): $\Omega_{S,i,t} = (\lambda'_{\Omega,\mathbf{t}}\boldsymbol{\Omega}^{-\mathbf{1}}\lambda_{\Omega,\mathbf{t}})^{-\mathbf{1}}\lambda'_{\Omega,\mathbf{t}}\boldsymbol{\Omega}^{-\mathbf{1}}\mathbf{Z}_{\Omega,\mathbf{i},\mathbf{t}}$ - where all the objects are replaced by their estimated counterparts. Here $\lambda_{\Omega,\mathbf{t}}$ is a $J_{\Omega,t} \times 1$ vector of scaling parameters $\lambda_{\Omega,t,j}$ of all measures j for factor Ω, and Ω is a $J_{\Omega,t} \times J_{\Omega,t}$ diagonal matrix with the variances of the measurement errors on the diagonal. Hence this step is equivalent to estimating a weighted regression of the measurement equation for each individual where the coefficient of interest is $\Omega_{i,t}$. The weights ensure that noisier measures receive a lower weight.

Note, however, that the Bartlett scores are still contaminated with measurement error. Taking a weighted average reduces but does not eliminate measurement error so the Bartlett scores themselves can thus be seen as a measure of the true latent factor: $\Omega_{S,i,t} = \Omega_{i,t} + \xi_{\Omega,i,t}$. Hence, we follow Heckman et al. (2013), who use the following measurement error correction approach:

2.D.2.5 Measurement Error Correction

Say we are interested in the following outcome equation, where Y_{it} can be an arbitrary object of interest (e.g., years of schooling, skills, investments) :

$$Y_{i,t} = \alpha \Omega_{i,t} + \gamma X_{i,t} + u_{i,t}$$

Now we have the predicted score:

$$\Omega_{S,i,t} = \Omega_{i,t} + \xi_{\Omega,i,t}$$

Because $\xi_{\Omega,i,t}$ is a weighted average of $\varepsilon_{\Omega,i,t,j}$, the assumptions made in appendix 2.D.2.2 implies:

$$(\boldsymbol{X}_i, \boldsymbol{\Omega}_i) \perp \boldsymbol{\xi}_{\Omega,i}, \quad E\left(\boldsymbol{\xi}_{\Omega,i}\right) = 0, \quad \mathrm{Cov}(\boldsymbol{\xi}_\Omega, \boldsymbol{\xi}_\Omega) = \boldsymbol{\Sigma}_{\boldsymbol{\xi}_\Omega \boldsymbol{\xi}_\Omega}$$

Now OLS estimates are inconsistent because

$$Y_{i,t} = \alpha \Omega_{S,i,t} + \gamma X_{i,t} + u_{i,t} - \alpha \xi_{\Omega,i,t}$$

.

$$\operatorname{plim} \begin{pmatrix} \hat{\alpha} \\ \hat{\gamma} \end{pmatrix} =$$
$$\begin{pmatrix} \mathrm{Cov}\left(\boldsymbol{\Omega}_S, \boldsymbol{\Omega}_S\right) & \mathrm{Cov}\left(\boldsymbol{\Omega}_S, \boldsymbol{X}\right) \\ \mathrm{Cov}\left(\boldsymbol{X}, \boldsymbol{\Omega}_S\right) & \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}) \end{pmatrix}^{-1} \begin{pmatrix} \mathrm{Cov}(\boldsymbol{\Omega}, \boldsymbol{\Omega}) & \mathrm{Cov}(\boldsymbol{\Omega}, \boldsymbol{X}) \\ \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{\Omega}) & \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}) \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \gamma \end{pmatrix}$$

Note to see more clearly:

$$\operatorname{plim} \begin{pmatrix} \hat{\alpha} \\ \gamma \end{pmatrix} = \underbrace{\begin{pmatrix} \Sigma_{\Omega,\Omega} + \Sigma_{\xi_\Omega,\xi_\Omega} & \Sigma_{\Omega,X} \\ \Sigma_{X,\Omega} & \Sigma_{X,X} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_{\Omega,\Omega} & \Sigma_{\Omega,X} \\ \Sigma_{X,\Omega} & \Sigma_{X,X} \end{pmatrix}}_{A} \begin{pmatrix} \alpha \\ \gamma \end{pmatrix}$$

As we know the variance of the latent factors, we thus know $\xi_{\Omega,i,t}$ and can correct for it. We can then pre-multiply the estimated coefficients by the inverse of the attenuation factor (A^{-1}) to get the consistent coefficients.

We use the approach described above to estimate the schooling equation, the cognition equations, and also the equations for parental investments in time and school quality. We also use it to correct for measurement error in the regression of mediators on parental income.

2.D.3 Signal-to-Noise Ratios

Table 2.D.3 presents the signal-to-noise ratios for the variables included in our final analysis. The signal-to-noise ratio is defined as $SN_{\Omega,t,j} = \frac{\lambda^2_{\Omega,t,j} Var(\Omega_t)}{Var(Z_{\Omega,t,j})}$. Intuitively, this is the appropriately scaled variance of the latent factor (signal) to the variance of the measure (signal+noise) and thus describes the information content of each measure. Note that we have normalized the variance of the latent factors as well as of the measures to one. Hence, the factor loadings $\lambda_{\Omega,t,j}$ are the square-root of the signal-to-noise ratio.

2.E Testing Restrictions in the Presence of Measurement Error

To test for the joint significance of groups of coefficients in Section 2.4.2, we use a measurement-error-corrected *F-test* procedure. We illustrate this procedure here using a simple example. Suppose the model we are interested in is:

$$y = \alpha_1 x_1^* + \alpha_2 x_2^* + u \tag{2.23}$$

Under the null hypothesis $H_0 : \alpha_2 = 0$:

$$y = \alpha_1^R x_1^* + u_R$$

Under the alternative hypothesis $H_1 : \alpha_2 \neq 0$:

$$y = \alpha_1^{UR} x_1^* + \alpha_2^{UR} x_2^* + u_{UR}$$

The standard *F-statistic* is then constructed as:

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(n-k-1)}, \tag{2.24}$$

where $SSR_R = \sum_{i=1}^{n}(y_i - \hat{\alpha}_1^R x_{1,i}^*)^2$, $SSR_{UR} = \sum_{i=1}^{n}(y_i - \hat{\alpha}_1^{UR} x_{1,i}^* - \hat{\alpha}_2^{UR} x_{2,i}^*)^2$, q = 1 is the number of restrictions, n is the sample size and k = 1 is the number of independent variables in the unrestricted model. However, in our model, we do not observe the true values of independent variables (i.e. x_1^* and x_2^*). Instead our observed variables, x_1 and x_2, are measured with an additive error:

$$\begin{aligned} x_1 &= x_1^* + \varepsilon_1 \\ x_2 &= x_2^* + \varepsilon_2 \end{aligned} \tag{2.25}$$

Using this framework, the aim is to now express the *F-statistic* in terms of observables only. First, we re-write both the sum of squared residual terms in terms of variances:

$$SSR_R = \textstyle\sum_{i=1}^{n}(y_i - \hat{\alpha}_1^R x_{1,i}^*)^2 = n \cdot Var(y - \hat{\alpha}_1^R x_1^*) \tag{2.26}$$

$$SSR_{UR} = \textstyle\sum_{i=1}^{n}(y_i - \hat{\alpha}_1^{UR} x_{1,i}^* - \hat{\alpha}_2^{UR} x_{2,i}^*)^2 = n \cdot Var(y - \hat{\alpha}_1^{UR} x_1^* - \hat{\alpha}_2^{UR} x_2^*) \tag{2.27}$$

Substituting the expressions in (2.25) into the RHS of (2.26) and (2.27):

$$
\begin{aligned}
n \cdot Var(y - \hat{\alpha}_1^R x_1^*) &= n \cdot Var(y - \hat{\alpha}_1^R (x_1 - \varepsilon_1)) \\
&= n[Var(y - \hat{\alpha}_1^R x_1) + (\hat{\alpha}_1^R)^2 Var(\varepsilon_1) + 2Cov(y - \hat{\alpha}_1^R x_1, \varepsilon_1)] \\
&= n[Var(y - \hat{\alpha}_1^R x_1) + (\hat{\alpha}_1^R)^2 Var(\varepsilon_1) - 2(\hat{\alpha}_1^R)^2 Var(\varepsilon_1)] \\
&= n[Var(y - \hat{\alpha}_1^R x_1) - (\hat{\alpha}_1^R)^2 Var(\varepsilon_1)], \qquad (2.28)
\end{aligned}
$$

$$
\begin{aligned}
& n \cdot Var(y - \hat{\alpha}_1^{UR} x_1^* - \hat{\alpha}_2^{UR} x_2^*) \\
&= n[Var(y - \hat{\alpha}_1^{UR}(x_1 - \varepsilon_1) - \hat{\alpha}_2^{UR}(x_2 - \varepsilon_2))] \\
&= n[Var((y - \hat{\alpha}_1^{UR} x_1 - \hat{\alpha}_2^{UR} x_2) + (\hat{\alpha}_1^{UR}\varepsilon_1 + \hat{\alpha}_2^{UR}\varepsilon_2))] \\
&= n[Var(y - \hat{\alpha}_1^{UR} x_1 - \hat{\alpha}_2^{UR} x_2) + Var(\hat{\alpha}_1^{UR}\varepsilon_1 + \hat{\alpha}_2^{UR}\varepsilon_2) \\
&\quad + 2Cov(y - \hat{\alpha}_1^{UR} x_1 - \hat{\alpha}_2^{UR} x_2, \hat{\alpha}_1^{UR}\varepsilon_1 + \hat{\alpha}_2^{UR}\varepsilon_2)] \\
&= n[Var(y - \hat{\alpha}_1^{UR} x_1 - \hat{\alpha}_2^{UR} x_2) + (\hat{\alpha}_1^{UR})^2 Var(\varepsilon_1) + (\hat{\alpha}_2^{UR})^2 Var(\varepsilon_2) \\
&\quad - 2(\hat{\alpha}_1^{UR})^2 Var(\varepsilon_1) - 2(\hat{\alpha}_2^{UR})^2 Var(\varepsilon_2)] \\
&= n[Var(y - \hat{\alpha}_1^{UR} x_1 - \hat{\alpha}_2^{UR} x_2) - (\hat{\alpha}_1^{UR})^2 Var(\varepsilon_1) - (\hat{\alpha}_2^{UR})^2 Var(\varepsilon_2)], \quad (2.29)
\end{aligned}
$$

where we assume the two measurement errors to be uncorrelated, $Cov(\varepsilon_1, \varepsilon_2) = 0$, and the unobserved factors to be uncorrelated with the measurement errors $Cov(\varepsilon_j, x_k^*) = 0$ for all $j = \{1,2\}$ and $k = \{1,2\}$.

Expressions (2.28) and (2.29) can now be calculated using observed independent variables and estimated measurement error variances. Using (2.26), (2.27), (2.28) and (2.29), the $F - statistic$ can be expressed in terms of observables.

Within this framework, our procedure to test restrictions in the presence of measurement error is:

1. Estimate α_1 and α_2. Since these are the coefficients from the true model, we estimate them using an errors-in-variables correction. Call these estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$.

2. Obtain residuals $y - \hat{\alpha}_1 x_1$ and $y - \hat{\alpha}_1 x_1 - \hat{\alpha}_2 x_2$ and estimate their variances.

3. Adjust $Var(y - \hat{\alpha}_1 x_1)$ and $Var(y - \hat{\alpha}_1 x_1 - \hat{\alpha}_2 x_2)$ for (estimated) measurement error variances $\hat{Var}(\varepsilon_1)$ and $\hat{Var}(\varepsilon_2)$ as in (2.28) and (2.29). This gives us SSR_R and SSR_{UR}.

4. Calculate the *F-statistic* as in (2.24) and using estimates of SSR_R and SSR_{UR}.

5. Bootstrap for the distribution of the *F-statistic*. We bootstrap the distribution of the statistic because measurement error make the distribution of the statistic complex.

Table 2.D.1: Exploratory Factor Analysis of Investment Measures

Age 7			Age 11			Age 16		
	Factor 1	Factor 2		Factor 1	Factor 2		Factor 1	Factor 2
	SQ	TI		SQ	TI		SQ	TI
Eigenvalues:	1.88	0.93		2.01	1.35		4.27	1.66
Loadings:								
F:interest in ed	0.65	0.01	**F:interest in ed**	0.11	0.62	**F:interest in ed**	0.01	0.87
M:interest in ed	0.68	0.04	**M:interest in ed**	0.09	0.61	**M:interest in ed**	-0.03	0.92
F: outings	0.01	0.63	**F: outings**	-0.15	0.55	**P:support schooling**	0.05	0.60
M:outings	-0.01	0.58	**M:outings**	-0.14	0.53	P:wish ct ed	0.30	0.41
F:reads	-0.02	0.61	**P: wish ct ed**	0.02	0.44	**school type**	0.45	0.04
M:reads	0.03	0.54	**P: wish uni**	0.00	0.35	**%Ct School**	0.63	0.12
school type	0.35	-0.14	**class size**	-0.40	0.10	**%FT degree**	0.89	-0.04
avg soc class of school	-0.47	0.05	**school type**	0.86	-0.01	**%GCE (past)**	0.94	-0.05
PT meetings available	0.19	-0.04	**% GCE**	0.49	0.19	**%GCE (present)**	0.70	0.04
Parental engagement	0.34	-0.06	**teacher-student ratio**	0.61	-0.02	**teacher-student ratio**	0.50	-0.07
PTA	0.17	-0.04	teacher experience	0.14	-0.01	avg soc class of school	0.62	0.13
class size	-0.17	0.12	teacher initiative	0.00	0.25	class size	-0.12	0.25
						career guidance	0.11	-0.03
						hours of english	-0.20	0.04
						hours of maths	-0.10	0.10

Note: The top panel presents eigenvalues of the exploratory factor analysis. The bottom panel presents loadings of the exploratory factor analysis. Measures that are subsequently included in the main analysis are in **bold**. Factor loadings rotated. ‘F’ denotes father, ‘M’ denotes mother, ‘P’ denotes parents.

Table 2.D.2: Exploratory Factor Analysis of Skills at 16

	Factor 1 Cognitive	Factor 2 Non-Cognitive
Eigenvalue:	3.215	1.648
Loadings:		
Reading score	0.78	0.01
Math score	0.81	0.01
Teacher-rated Math ability	0.88	0.00
Teacher-rated English ability	0.87	-0.01
Unconcentrated	0.09	0.46
Destructive	0.03	0.30
Miserable	-0.05	0.52
Fidgety	0.00	0.44
Worried	-0.14	0.44
Irritable	0.01	0.54
Fights	0.06	0.44
Disobedient	0.10	0.45
Solitary	-0.11	0.25
Bullied	-0.02	0.29
Sucks on thumb	-0.02	0.14
Bites nails	0.04	0.21
Upset	-0.05	0.33
Has mannerisms	-0.02	0.21

Table presents loadings for exploratory factor analysis. Measures that are subsequently included in the main analysis are in **bold**.

Table 2.D.3: Signal-to-Noise Ratio for Measures Used

Cognition at 16					
Reading	0.56				
Math	0.62				
Teacher assessed math	0.80				
Teacher assessed english	0.72				
Parental Investments					
Age 7		*Age 11*		*Age 16*	
M:Outings	0.29	M:Outings	0.53	F: Interest in ed	0.75
F:Outings	0.37	F:Outings	0.57	M:Interest in ed	0.90
M:Reads	0.31	F:Interest in Ed	0.12	P:Support schooling	0.32
F:Reads	0.29	M: Interest in Ed	0.14		
M:Interest in Ed	0.15	P:Wish cnt ed	0.09		
F:Interest in Ed	0.11	P:Wish uni	0.06		
School Quality					
Age 7		*Age 11*		*Age 16*	
Avg soc class of school	0.06	Class Size	0.10	School Type	0.08
PT meetings available	0.42	School Type	0.42	%Cnt School	0.35
PTA	0.28	% GCE	0.20	%FT degree	0.82
School type	0.01	Teacher student ratio	0.21	% passed A-levels	0.93
Parental engagement	0.10			% studying towards A-levels	0.45
				Teacher student Ratio	0.20

Note: 'F' denotes father, 'M' denotes mother, 'P' denotes parents.

2.F Effect of Parental Income on Mediating Variables

Table 2.F.1 shows the results of a regression of each mediating variable on log parental income, as illustrated in equation 2.5. We do this separately for males and females and correct for measurement error in parental income using the correction approach outlined above.

Table 2.F.1: Coefficient on Log Parental Income

Dependent Variable	Males	Females
Years of Schooling	0.530	**0.748**
	(0.301)	(0.242)
Cognition	**0.743**	**0.499**
	(0.175)	(0.142)
Time Investments		
Age 7	**0.966**	**0.542**
	(0.185)	(0.204)
Age 11	**1.175**	**0.562**
	(0.202)	(0.188)
Age 16	**0.557**	**0.351**
	(0.165)	(0.155)
School Quality Investments		
Age 7	0.017	0.238
	(0.172)	(0.204)
Age 11	**0.600**	**0.180**
	(0.231)	(0.030)
Age 16	**0.427**	**0.563**
	(0.150)	(0.162)
Family Background		
Mother's education	**0.636**	**1.049**
	(0.159)	(0.190)
Father's education	**0.776**	**1.086**
	(0.218)	(0.238)
Number of siblings	-0.400	-0.176
	(0.235)	(0.245)
N	1350	1347

Standard errors bootstrapped with 250 repetitions.

Note: This table presents estimates of Equation (2.5) for each determinant we use for lifetime earnings. Each row contains the coefficient (and corresponding standard error) on log parental income from a measurement error corrected regression of the variable in the first column on log parental income, for males and females.

2.G Results for Robustness Checks

2.G.1 Robustness Check 1: More Covariates

Table 2.G.2 shows results for the mediation analysis when we additionally control for marital stability and parental age. Marital stability is a dummy that takes on value 1 when the child lives with both natural parents from birth to age 16. We find

that none of the additional covariates explain a significant fraction of the IGE.

Note, despite the effect of birth order on lifetime outcomes being widely studied, we do not include it here as a mediator for the IGE. The reason for this is that conditional on the number of children, birth order is uncorrelated with parental income. Hence it cannot mediate the effect of parental income on child earnings.

2.G.2 Robustness Check 2: Including Non-Cognitive Skills

Table 2.G.3 is analogous to the mediation analysis Tables 2.4 and 2.5 in the main text, but also includes non-cognitive skills. Non-cognitive skills enter in the same way as cognition; they can affect years of schooling, and are affected by investments and family background. We can see that non-cognitive skills explain a small and insignificant fraction of the IGE. Table 2.G.4 shows estimated determinants of lifetime earnings and years of schooling with non-cognitive skills. For both lifetime earnings and years of schooling, cognitive skills matter much more than non-cognitive skills although both are positively related to parental income. Finally, Table 2.G.5 shows the relationship between parental income and non-cognitive skills at age 16. We find a much stronger relationship between parental income and child cognition, than between parental income and non-cognitive skills. In fact, for men, non-cognitive skills are negatively correlated with parental income.

2.G.3 Robustness Check 3: Including Interactions

The errors-in-variables correction we are using is suitable only for linear models. In order to conduct a robustness check which allows for an interaction between cognition and schooling in the lifetime earnings equation, we estimate the lifetime earnings equation using a system GMM approach where we use the different measures as instruments for each other to correct for measurement error following Bolt et al. (2021). This approach extends the one in Agostinelli and Wiswall (2016b). The downside of this approach is that we cannot correct for measurement error in parental income anymore (as we only have one measure of parental income). Thus, we first estimate a version of the level 1 mediation where we do not correct for measurement error in parental income. Other than that, the approach is the same as in the main text. Column 1 of Table 2.G.6 shows that Level 1 mediation results remain largely unchanged compared to the baseline model, even if we do not correct for measurement error in parental income. Column 2 of Table 2.G.6 then provides a comparison between the GMM approach and errors-in-variables, when we assume only linear effects. We can see that the results of using GMM to correct for measurement error or using errors-in-variables to correct for measurement error yields very similar results. Finally, in column 3, we add the interaction between years of schooling and cognition. The interaction term explains a small and insignificant fraction of the IGE, both for males and females. Although we find evidence of an interaction between years of schooling and cognition in the earnings equation, this interaction is not highly correlated with parental income.

Table 2.G.1: Main Mediation Analysis Including Additional Covariates - Males

	Males			
	Level 1	Level 2	Level 3	Level 4
Years of Schooling	**0.095**	-0.096	-0.096	-0.096
	[0.029, 0.203]	[-0.260, -0.001]	[-0.260, -0.001]	[-0.260, -0.001]
Cognition	**0.323**	**0.454**	0.149	0.149
	[0.168, 0.631]	[0.255, 0.837]	[-0.086, 0.406]	[-0.086, 0.406]
Investments	0.134	0.187	**0.469**	0.306
	[-0.104, 0.433]	[-0.043, 0.518]	[0.178, 0.962]	[-0.007, 0.799]
Time Investments	0.132	0.178	**0.388**	**0.281**
	[-0.075, 0.445]	[-0.040, 0.519]	[0.156, 0.826]	[0.055, 0.753]
Age 7	0.135	0.156	0.152	0.084
	[-0.055, 0.420]	[-0.046, 0.463]	[-0.054, 0.412]	[-0.038, 0.283]
Age 11	-0.057	-0.030	0.075	0.066
	[-0.331, 0.238]	[-0.293, 0.267]	[-0.163, 0.421]	[-0.145, 0.379]
Age 16	0.054	0.052	**0.162**	**0.131**
	[-0.008, 0.148]	[-0.009, 0.144]	[0.071, 0.365]	[0.032, 0.331]
School Quality	0.002	0.010	0.081	0.024
	[-0.124, 0.135]	[-0.121, 0.142]	[-0.076, 0.234]	[-0.106, 0.126]
Age 7	-0.001	-0.001	0.000	0.001
	[-0.022, 0.032]	[-0.021, 0.028]	[-0.025, 0.029]	[-0.035, 0.046]
Age 11	-0.030	-0.028	-0.051	-0.056
	[-0.158, 0.050]	[-0.161, 0.053]	[-0.205, 0.029]	[-0.252, 0.031]
Age 16	0.033	0.038	**0.132**	0.080
	[-0.010, 0.106]	[-0.006, 0.118]	[0.046, 0.265]	[-0.018, 0.215]
Family Background	-0.205	-0.197	-0.174	-0.011
	[-0.131, 0.089]	[-0.146, 0.078]	[-0.089, 0.166]	[0.063, 0.453]
Mother's education	-0.045	-0.044	-0.021	0.031
	[-0.154, 0.010]	[-0.162, 0.016]	[-0.126, 0.038]	[-0.054, 0.103]
Father's education	0.012	0.004	0.032	0.084
	[-0.073, 0.101]	[-0.092, 0.086]	[-0.053, 0.131]	[0.009, 0.194]
Number of Siblings	0.013	0.012	0.028	0.077
	[-0.418, 0.096]	[-0.381, 0.114]	[-0.426, 0.110]	[-0.390, 0.182]
Stable	-0.145	-0.133	-0.150	-0.103
	[-0.418, 0.096]	[-0.381, 0.114]	[-0.426, 0.110]	[-0.390, 0.182]
Mum's age	-0.032	-0.028	-0.038	-0.037
	[-0.179, 0.089]	[-0.174, 0.090]	[-0.171, 0.092]	[-0.186, 0.100]
Dad's age	-0.008	-0.007	-0.025	-0.063
	[-0.188, 0.149]	[-0.191, 0.147]	[-0.212, 0.121]	[-0.263, 0.076]
N	1350	1350	1350	1350

Note: This replicates the mediation analysis in Tables 2.4 and 2.5, except it includes additional family background variables.

Table 2.G.2: Main Mediation Analysis Including Additional Covariates - Females

	Females			
	Level 1	Level 2	Level 3	Level 4
Years of Schooling	**0.423**	0.093	0.093	0.093
	[-0.260, -0.001]	[-0.280, 0.733]	[-0.280, 0.733]	[-0.280, 0.733]
Cognition	0.129	**0.396**	-0.021	-0.021
	[-0.030, 0.435]	[0.152, 1.142]	[-0.649, 0.319]	[-0.649, 0.319]
Investments	0.049	0.150	**0.449**	0.277
	[-0.398, 0.458]	[-0.224, 0.699]	[0.106, 1.579]	[-0.341, 1.202]
Time Investments	-0.093	-0.038	0.116	-0.070
	[-0.427, 0.122]	[-0.321, 0.203]	[-0.088, 0.454]	[-0.467, 0.239]
Age 7	0.149	0.167	0.181	-0.018
	[-0.140, 0.619]	[-0.053, 0.722]	[-0.062, 0.769]	[-0.366, 0.247]
Age 11	-0.180	-0.176	-0.133	-0.053
	[-0.637, -0.037]	[-0.634, 0.028]	[-0.518, 0.123]	[-0.371, 0.194]
Age 16	-0.062	-0.028	0.067	0.002
	[-0.276, 0.032]	[-0.184, 0.086]	[-0.016, 0.231]	[-0.106, 0.116]
School Quality	**0.142**	**0.188**	**0.333**	**0.347**
	[0.008, 0.483]	[0.038, 0.698]	[0.139, 1.187]	[0.014, 0.619]
Age 7	0.047	0.044	0.047	0.062
	[-0.060, 0.198]	[-0.050, 0.184]	[-0.054, 0.186]	[-0.092, 0.300]
Age 11	0.017	0.019	0.012	0.005
	[-0.069, 0.129]	[-0.101, 0.113]	[-0.182, 0.086]	[-0.143, 0.089]
Age 16	0.078	0.125	**0.274**	**0.280**
	[-0.052, 0.349]	[-0.009, 0.456]	[0.099, 0.884]	[0.064, 0.919]
Family Background	-0.264	-0.302	-0.183	-0.012
	[-0.288, 0.221]	[-0.220, 0.323]	[-0.089, 0.554]	[0.136, 1.174]
Mother's education	-0.027	-0.009	0.043	0.147
	[-0.341, 0.267]	[-0.258, 0.284]	[-0.228, 0.368]	[-0.072, 0.631]
Father's education	0.055	0.066	0.113	0.209
	[-0.133, 0.297]	[-0.124, 0.293]	[-0.052, 0.442]	[0.041, 0.763]
Number of Siblings	-0.020	-0.021	-0.011	0.004
	[-0.118, 0.038]	[-0.114, 0.040]	[-0.070, 0.029]	[-0.037, 0.056]
Stable	-0.189	-0.243	-0.178	-0.155
	[-0.923, 0.598]	[-0.893, 0.468]	[-0.780, 0.609]	[-0.834, 0.527]
Mum's age	-0.099	-0.087	-0.155	-0.213
	[-0.474, 0.105]	[-0.417, 0.100]	[-0.601, 0.039]	[-0.778, 0.025]
Dad's age	0.017	-0.008	0.004	-0.004
	[-0.264, 0.342]	[-0.294, 0.222]	[-0.285, 0.281]	[-0.283, 0.232]
N	1347	1347	1347	1347

Note: This replicates the mediation analysis in Tables 2.4 and 2.5, except it includes additional family background variables.

Table 2.G.3: Main Decomposition with Non-Cognitive Skills

	Males			
	Level 1	Level 2	Level 3	Level 4
Years of Schooling	**0.104**	**-0.078**	**-0.078**	**-0.078**
	[0.031, 0.266]	[-0.274, -0.012]	[-0.274, -0.012]	[-0.274, -0.012]
Cognition	**0.338**	**0.474**	0.107	0.107
	[0.181, 0.759]	[0.296, 1.007]	[-0.096, 0.378]	[-0.096, 0.378]
Non-cognitive skills	-0.004	-0.005	- 0.046	- 0.046
	[-0.079, 0.042]	[-0.082, 0.043]	[-0.169, 0.007]	[-0.169, 0.007]
Investments	0.123	0.178	**0.517**	**0.354**
	[-0.133, 0.454]	[-0.063, 0.623]	[0.212, 1.346]	[0.112, 0.974]
Family Background	-0.008	-0.018	0.051	0.214
	[-0.173, 0.109]	[-0.194, 0.103]	[-0.093, 0.188]	[0.092, 0.558]
N	1339	1339	1339	1339
	Females			
	Level 1	Level 2	Level 3	Level 4
Years of Schooling	**0.420**	0.039	0.039	0.039
	[0.194, 1.127]	[-0.171, 0.329]	[-0.171, 0.329]	[-0.171, 0.329]
Cognition	0.135	**0.394**	0.012	0.012
	[-0.016, 0.400]	[0.161, 1.071]	[-0.297, 0.212]	[-0.297, 0.212]
Non-cognitive skills	0.000	0.000	- 0.022	-0.022
	[-0.047, 0.039]	[-0.073, 0.061]	[-0.151, 0.022]	[-0.151, 0.022]
Investments	0.033	0.128	**0.444**	0.239
	[-0.306, 0.351]	[-0.142, 0.525]	[0.158, 1.278]	[-0.038, 0.745]
Family Background	-0.006	0.020	0.108	0.314
	[-0.238, 0.215]	[-0.189, 0.303]	[-0.066, 0.504]	[0.089, 0.997]
N	1336	1336	1336	1336

Note: This replicates the mediation analysis in Tables 2.4 and 2.4, except it includes Non-cognitive Skills.

Table 2.G.4: Determinants of lifetime earnings and years of schooling with non-cognitive measures

	Lifetime Earnings		**Years of Schooling**	
	Males	Females	Males	Females
Years of Schooling	0.055	0.133		
	(0.014)	(0.015)		
Cognition	0.135	0.064	0.947	0.927
	(0.028)	(0.039)	(0.057)	(0.066)
Non-Cognitive Skills	0.038	0.013	0.076	0.099
	(0.028)	(0.024)	(0.051)	(0.049)
Time Investments				
Age 7	0.036	0.066	0.110	0.038
	(0.036)	(0.044)	(0.085)	(0.081)
Age 11	-0.010	-0.078	0.111	0.013
	(0.035)	(0.043)	(0.101)	(0.080)
Age 16	0.033	-0.045	-0.018	0.164
	(0.024)	(0.026)	(0.056)	(0.050)
School Quality Investments				
Age 7	-0.007	0.048	0.013	-0.016
	(0.023)	(0.026)	(0.051)	(0.049)
Age 11	-0.013	0.023	0.009	0.021
	(0.033)	(0.032)	(0.058)	(0.255)
Age 16	0.024	0.037	0.070	0.158
	(0.020)	(0.023)	(0.048)	(0.066)
Family Background				
Mother's age left school	-0.024	-0.008	0.016	0.037
	(0.018)	(0.023)	(0.040)	(0.055)
Father's age left school	0.011	0.016	-0.067	0.013
	(0.019)	(0.019)	(0.037)	(0.036)
Number of Siblings	-0.013	0.027	0.030	0.001
	(0.013)	(0.016)	(0.031)	(0.032)
Log parental income	0.059	0.047	-0.183	0.033
	(0.048)	(0.048)	(0.103)	(0.107)
N	1339	1336	1339	1336

Note: This table reports estimates from the same model presented in Table 2.6, except here variables of non-cognitive skills, internalising behaviour and externalising behaviour are included. These are shown to have negligible effect on log lifetime earnings once cognition at 16 and years of schooling are controlled for.

Table 2.G.5: Effect of log parental income on cognition and non-cognitive skills

	Males		Females	
	Cognition	Non-Cognitive Skills	Cognition	Non-Cognitive Skills
$\ln Y_{Parent}$	0.743	-0.032	0.499	0.001
	(0.168)	(0.181)	(0.159)	(0.191)
N	1339	1339	1336	1336

Note: Estimates of the direct effects of log parental income on cognition and non-cognitive skills for males and females at age 16.

Table 2.G.6: Robustness including interaction terms

	Males		
	EIV	GMM	GMM
Years of Schooling	**0.093**	**0.165**	**0.162**
	[0.019, 0.228]	[0.073, 0.325]	[0.066, 0.310]
Cognition	**0.333**	**0.368**	**0.365**
	[0.193, 0.729]	[0.173, 0.646]	[0.184, 0.625]
Years of Schooling × Cognition			-0.016
			[-0.066, 0.017]
Investments	0.163	0.137	0.122
	[-0.060, 0.456]	[-0.112, 0.428]	[-0.119, 0.392]
Family Background	-0.012	-0.055	-0.053
	[-0.150, 0.112]	[-0.232, 0.074]	[-0.215, 0.077]
	Females		
	EIV	GMM	GMM
Years of Schooling	**0.425**	**0.452**	**0.487**
	[0.158, 1.337]	[0.231, 1.083]	[0.265, 1.206]
Cognition	0.135	0.094	0.078
	[-0.008, 0.502]	[-0.058, 0.268]	[-0.081, 0.229]
Years of Schooling × Cognition			0.003
			[-0.054, 0.070]
Investments	0.057	0.149	0.122
	[-0.266, 0.437]	[-0.140, 0.554]	[-0.124, 0.513]
Family Background	0.022	0.055	0.102
	[-0.233, 0.302]	[-0.164, 0.297]	[-0.136, 0.374]

2.H Further Comparisons with the Literature

Here we present some further comparisons with estimates in the literature.

School Quality: Altonji and Dunn (1996) use within-family variation in where siblings attended schools, instrumented using family residence, to estimate the direct return to school quality. Interestingly, they find that if anything, OLS estimates understate the return to school quality. In their IV estimates they find that a one standard deviation improvement in the quality of the child's school increases earnings by approximately 10.5%. [15]This is bigger than, but not statistically different from, their OLS estimate. Using the NCDS data, we estimate an identical model to the OLS regression used by Altonji and Dunn (1996) (see Column 1 of Table 2.H.1). We find that one standard deviation increase in school quality at all ages increases lifetime earnings by 8.9%. Thus, our estimated impact of school quality is in a similar range as Altonji and Dunn (1996)'s preferred IV estimates.

Family Size: Several papers that use twins as a source of plausibly exogenous variation tend to find small effects of family size on earnings, and suggest that OLS estimates overstate the effect of family size. For example, Black et al. (2005) use Norwegian administrative data on twins and find that an increase in the number of siblings leads to a change in earnings of -2% using OLS, but of 0% for males and 3% for females when using twins as an instrument. [16] These IV estimates are not statistically significant. In Table 2.H.1, we use the same controls as Black et al. (2005) and find a larger effect of family size on earnings of -3.1% (see Column 2 of Table 2.H.1), implying that we may potentially be over-stating the effect of family size. This observation is also supported by Angrist et al. (2010), who show that the effects of family size on earnings are in the range -2% to +2% when using twins and/or sex composition as instruments, and are -3% when using OLS [17]. However, Daruich and Kozlowski (2020) find that an equalisation of fertility across the in-

[15] See Table 1 in Altonji and Dunn (1996). To construct this estimate, we use the level coefficient from Column 1a of their paper - 0.501 (1.17). Since their coefficients are scaled by 100, we divide by 100, and then multiply by 2.1 (which is SD of the school-quality index).

[16] See Table 9 in the appendix of Black et al. (2005). These estimates are taken from the log(earnings). Row 1 contains relevant OLS estimates, and row 3 contains IV estimates.

[17] OLS estimates are taken from Tables 5 and 6 of Angrist et al. (2010). Pooled IV estimates are from Table 7. The relevant dependent variable in Angrist et al. (2010) is log earnings.

come gradient would increase income persistence by 6% - an estimate very close to ours.

Table 2.H.1: Estimated effects of school quality on lifetime earnings

	Children's Lifetime Earnings	
	(1)	(2)
School Quality at 7	0.023	
	(0.011)	
School Quality at 11	0.013	
	(0.011)	
School Quality at 16	0.053	
	(0.015)	
Number of siblings		-0.031
		(0.011)
Constant	0.000	12.999
	(0.014)	(0.258)

Column (1) presents OLS estimates from a regression of children's lifetime earnings on school quality at all ages. It includes the following controls: a cubic in years of schooling, years of schooling interacted with parental education, gender, ethnicity and region of residence during childhood. This replicates the controls used in Altonji and Dunn (1996) as relevant for our data. Column (2) has estimates from OLS regression of children's lifetime earnings on number of siblings, controlling for: parents' age at birth, parental education, gender and birth-order. This replicates the controls used in Black et al. (2005) as relevant for our data.

Chapter 3

Parental Altruism and Transfers of Time and Money - A Lifecycle Perspective

3.1 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)). The literature on understanding the mechanisms behind this persistence is much newer. This paper estimates a dynastic model of the decision-making of altruistic households to investigate the quantitative importance of three distinct mechanisms in generating intergenerational persistence in outcomes. Those mechanisms, each known to be important in linking outcomes across generations, are i) parental time investments during childhood and adolescence that aid child development (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 (2016a) and leveraging the fact that we have multiple measures of parental inputs and child outcomes. We embed our estimated ability production function into a dynastic model in which ability and education generate productivity in the labor market and in which altruistic parents can give combinations of time, educational investments, and cash transfers to their children, while also making their own consumption and labor supply decisions.

Our model contains five distinct mechanisms which can generate persistence in outcomes across generations. The first three mechanisms generate a positive correlation between the earnings of an individual and the earnings of their parents. The first is the borrowing constraint, which limits low income low wealth families from accessing credit to provide higher education to their children. The second 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, 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 0.32, 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 with college education an increase in the standard deviation in ability at age 16 leads to an additional 19% in wages. Among those with the lowest level of education – this premium, at 9%, is much smaller. As a result, high ability individuals are more likely to select into education than their low ability counterparts. This dynamic complementarity, in combination with self selection into education, is a key mechanism that perpetuates income inequality across generations. High income households, who have more resources to send their child to college, have higher returns to investing in their child's ability than their low income counterparts; thus they invest more in their children. Second, we find that more than a quarter (28%) of the variance of lifetime wages can already explained by characteristics of the family, before an individual is even born. By the time individuals are 23, their characteristics can explain up to 62% of the variance in lifetime wages. Thus, more than half of the lifetime variability in wages is realized by age 23.

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.

The rest of this paper proceeds as follows. Section 3.2 describes the data, and documents descriptive statistics on ability, education and parental investments. Section 3.3 lays out the dynamic 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 3.8 concludes.

3.2 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 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 focus on individuals in ELSA born in the 1950s, leaving us with a sample of 3,001.[2]

Lastly, to convert the investment measures observed into actual units of time, we use the UK Time Use Survey (UKTUS), which has detailed information on time spent in educational investments in the child. We describe the measures of time

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

[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

investments we use in the notes of Table 3.2.2 and in greater detail in Appendix 3.C.

In the rest of this section, we document 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.

3.2.1 Transfer 1: Parental Time Investments

The NCDS has detailed measures of parental time investments which cohort members received during childhood. The full set of parental time investment as well as cognitive skill measures we use are listed in Table 3.2.1.[4] These measures come from different sources – some are from surveys of parents, others from surveys of school teachers. Whilst in later sections of the paper, we exploit the full range of measures using latent factor methods, here we highlight some of the key features directly observable in the raw data.

The first panel of Table 3.2.2 documents the gradients for some of the investment measures in father's education at different ages. Whilst 52% of high educated fathers read to their child each week, only 33% of low educated fathers do so. The gradient is somewhat less stark for taking the child on outings, but much more pronounced for the parents' interest in the child's education. For example, when the child is 7, high educated fathers are three times more likely to be judged by the child's teacher to be 'very interested' in their child's education as low education fathers (66% compared to 20%). While mothers are assessed as having greater in-

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

[4]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)).

terest in their child's education in all groups, there are large difference according to education group (75% in the highest education group are very interested, compared to 33% in the lowest education group). The gaps increase slightly with age.

Table 3.2.1: List of all measures used

Ability measures	*Investment measures*
Age 0:	
Birthweight	Teacher's assessment of
Gestation	parents' interest in education (**mother** and father)
	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
Math score	in education (**mother** and father)
Drawing score	Outings with child (mother and father)
Copying design score	Father's involvement in upbringing
	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'
Math score	interest in education (**mother** and father)
Copying design score	Involvement of parents in child's schooling
	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.

3.2.2 Transfer 2: Educational Investments

Panel 2 of Table 3.2.2 shows that there is a substantial intergenerational correlation in educational attainment between fathers and their children. First, having a high-educated father makes it much less likely that a child will end up with low education. 30% of the children of fathers with low education end up with low education, compared to only 10% of those with middle educated fathers, and just 2% of high educated fathers. Second, having a high-educated father makes it much more

Table 3.2.2: Transfers and outcomes by father's education

			Father's education			
	Avg	SD	low	med	high	F-test
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 educ 7	0.26	0.44	0.20	0.43	0.66	0.00
Mother very interested in educ 7	0.39	0.49	0.33	0.58	0.75	0.00
Father very interested in educ 11	0.31	0.46	0.23	0.52	0.73	0.00
Mother very interested in educ 11	0.39	0.49	0.33	0.59	0.76	0.00
Father very interested in educ 16	0.36	0.48	0.28	0.57	0.80	0.00
Mother very interested in educ 16	0.38	0.49	0.32	0.59	0.78	0.00
Time spent with child [UKTUS]*	9.06	10.05	8.35	8.91	9.87	.52
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
Cash Transfers						
Fraction receiving gift (>£1000)	0.07	0.26	0.06	0.10	0.20	0.06
Gift value in £1000 (excl zeros)	39.4	104.6	30.6	77.9	49.1	0.72
Fraction receiving inheritance	0.39	0.49	0.36	0.58	0.54	0.00
Inheritance value in £1000 (excl zeros)	88.2	114.7	75.6	122.4	174.3	0.00
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
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 3.2.2 shows: Mean, standard deviation, mean conditional on each paternal education group (low, medium, high), and P-values for an F-test of the difference in the mean between the low and high father's education group. *We include all of the following activities as time spent with the child when constructing the investment measure in UKTUS: teaching the child, reading/playing/talking with child, travel escorting to/from education.

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.

3.2.3 Transfer 3: Inter-vivos Transfers and Bequests

The third panel of Table 3.2.2 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 have inherited around £66,000 more than those with low-educated fathers. This is likely to understate the true difference in mean lifetime inheritance receipt between these groups; some of those born in the 1950s will still have living parents, and differential mortality means it is in fact likely that this applies to a larger share of those with high-educated fathers.

3.2.4 Outcome 1: Ability

The first panel of Table 3.2.2 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, the difference in reading (math) ability between children of low and high educated fathers is 0.67 (0.62) of a standard deviation. These ability gaps widen with age, such that by the time the children are 16, they amount to a 0.88 standard deviation difference in reading score and 1.03 standard deviation difference in math.

3.2.5 Outcome 2: Lifetime Earnings

Finally, we can see that children of more educated fathers have a 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.

3.3 Model

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

Figure 3.3.1 provides an overview of the dynastic model. During childhood, parental time investments in children and educational choices affect the evolution of the child's ability and their educational attainment. Children are then matched in couples, possibly receive transfers of cash from their parents and begin adult life. They then have their own children, and alongside the standard choices of consumption and labour supply they choose how much to invest in their own children, with implications for their children's future outcomes.

Our NCDS data respondents were surveyed every four to seven years from the age of 0 and 55. To be consistent with the data, each of our model periods will cover the time between interviews (and each period will be of different lengths). A model timeline illustrating this is shown in Figure 3.3.1 (while a more comprehensive Appendix Table 3.B.1 summarizes the rest of the detail that will be laid out in this section).

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

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

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

3. Parenthood has five periods at ages $t = 6,7,8,9$, 10, corresponding to ages 26-32, 33-36, 37-41,42-48, 49-54. Identical twin children are born at the start of the 'Parenthood' phase and now additionally to making labour supply and savings decisions, the parents decide how much to invest in their childrens' human capital and education. At the end of this period they have an opportunity to transfer wealth to the children.

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

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

We now provide formal details of the model.

[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 $t = 6$.

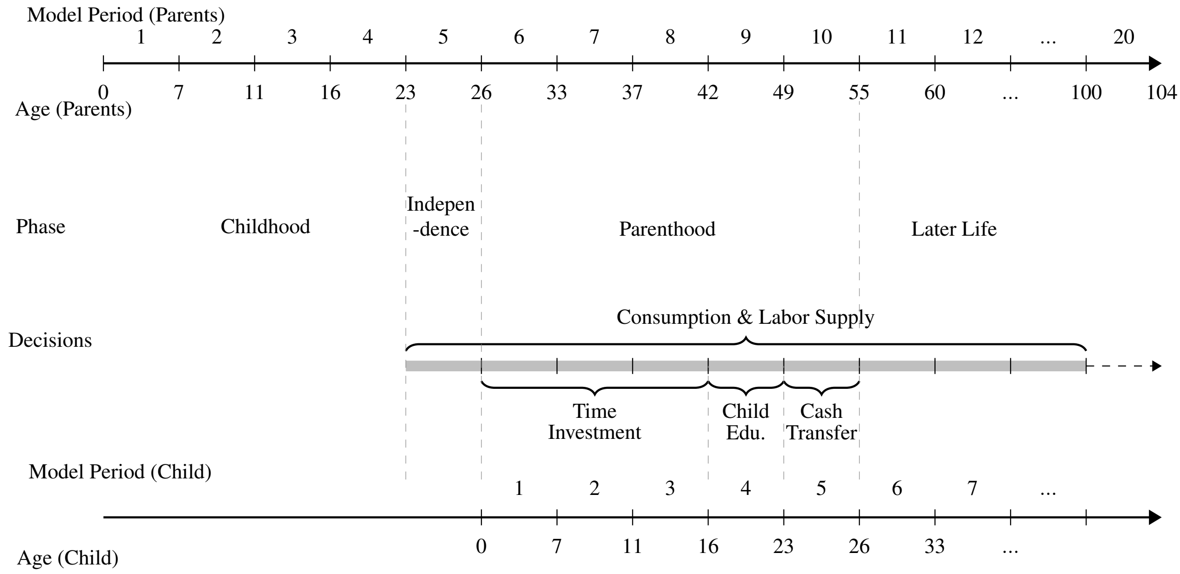


Figure 3.3.1: The Life Cycle of an Individual

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

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

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

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

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

3.3.2 Initial Conditions and Parental Cash Transfers Received

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

3.3.3 Demographics

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

At age 26, a pair of identical twins is born to the couple. In order to match the average fertility for this sample, which is close to two, yet still maintain computational tractability, we follow 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 that death is not possible until the household enters the late adult phase of life at the age of 50 and that death occurs by the age of 110 at the latest.

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

3.3.4 Constraints and Income Sources

Constraints Parents face three constraints – an intertemporal budget constraint, a borrowing constraint, and an intratemporal time constraint. The budget constraint and borrowing constraint are:

$$a_{t+1} = (1+r_t)(a_t + y_t - (c_{m,t} + c_{f,t}) - x_t) \tag{3.1}$$

$$a_{t+1} \geq 0 \tag{3.2}$$

where a_t is parental wealth, y_t is household income and x_t is a cash transfer to children that can only be made when the parents are 50 and the children are 23 (and $x_t = 0$ otherwise). The gross interest rate $(1+r_t)$ is equal to $(1+r)^{\tau_t}$ where r is an annual interest rate and τ_t is the length in years of model period t.

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

$$l_{g,t} = T - (\theta ti_{g,t} + h_{g,t}) \tag{3.3}$$

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.

Earnings and household income Household income is given by

$$y_t = \begin{cases} \tau(e_{m,t}, e_{f,t}, e'_{t'}, t) & \text{if the children are age 16} \\ \tau(e_{m,t}, e_{f,t}, t) & \text{otherwise} \end{cases}. \tag{3.4}$$

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

Earnings are equal to hours worked (h) multiplied by the wage rate, for example: $e_{f,t} = h_{f,t}w_{f,t}$. That wage rate evolves according to a process that has a deterministic component which varies with age and whether the individual works part-time or fulltime, and a stochastic (AR(1)) component:

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

where PT_t is a dummy for working part time. To capture the impact of ability on lifetime wages we model the initial wage draw in period 4 (age 16) as a function of final ability and a shock:

$$v_4 = \delta_5 ab + \eta_{4.}$$

Ability impacts the age 16 wage shock v_4 and thus impacts wages at all ages because v_t is persistent; if $\rho = 1$ then ability has a permanent effect on wages. Thus we do not need to keep track of ability after age 16, 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 here, each of $\{\delta_0, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \rho, \sigma^2\}$ varies by gender (g) and education (ed). This flexibility means that we allow ability to impact wages through its relationship with education δ_5. As we show below this flexibility is critical because the returns to ability are higher for the highly educated.

3.3.5 Ability and Education of Children

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

3.3.5.1 Child Ability Production Function

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

$$ab'_{t'+1} = \gamma_{1,t'} ab'_{t'} + \gamma_{2,t'} ti_{t'} + \gamma_{3,t'} ti_t \cdot ab_{t'} + \gamma_{4,t'} ed_m + \gamma_{5,t'} ed_f + u'_{ab,t'} \tag{3.5}$$

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

We include 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 given by his/her parents' level of education and a shock:

$$ab'_{1'} = \gamma_{4,0'} ed_m, + \gamma_{5,0'} ed_f + u'_{ab,0'}. \tag{3.6}$$

3.3.5.2 Education

When the child is age 16 the parent chooses the educational level of the child. There was compulsory education to age 16 for our sample members and additional years of education were free of tuition at this point in time. Thus we model the decision to send the child to school until age 18 or 22. Because there were no tuition fees for this cohort, the key cost of education at this time was the cost of forgone labor income from the child. We model the potential income if the child works, and thus the loss of household income if the child receives additional years of education as in equation (3.4).

3.3.6 Decision Problem

3.3.6.1 Decision Problem in the Parenthood Phase

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

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

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

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

3. Time investments in children of each parent – $ti_{m,t}$ and $ti_{f,t}$ (up to the age at which their child turns 16.)

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

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

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

Value function The value function for the Parenthood phase is given below in expression (3.7):

$$V_t(\mathbf{X}_t) = \max_{\substack{c_{m,t}, c_{f,t}, h_{m,t}, \\ h_{f,t}, ti_{m,t}, ti_{f,t}, ed'_g}} \left\{ u(c_{m,t}, c_{f,t}, l_{m,t}, l_{f,t}, n_t) + \beta_{t+1} \int V_{t+1}(\mathbf{X}_{t+1}) dF_{t+1}(\mathbf{q}_{t+1}) \right\} \quad (3.7)$$

$s.t.$ i) the intertemporal budget constraint in equation (3.1)

ii) the borrowing constraint in equation (3.2)

iii) and the time budget constraints in equation (3.11)

and education of the child is a choice when the child is 16, but not otherwise.

3.3.6.2 Decision Problem in the Independence Phase

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

Choices During this phase couples make three sets of choices:

1. Household consumption – $c_{mt}, c_{f,t}$

2. Hours of work for each parent – $h_{m,t}$, $h_{f,t}$ where m and f index hours of work by the male and female respectively

3. A cash gift (x_t) which is split equally between their children. The gift is made when the parents are age 49 and the children are 23.

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

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

State variables The set of state variables in this phase is the same as in the early 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:

$$\begin{aligned} V_t(\mathbf{X}_t) \quad &= \max_{c_{m,t},c_{f,t},h_{m,t},h_{f,t},x_t} \Big\{ u(c_{m,t}, c_{f,t}, l_{m,t}, l_{f,t}, n_t) \\ &+2\lambda \int V'_{t'}(\mathbf{X'}_{t'}) dH(\mathbf{p}'_{t'}) + \beta_{t+1} \int V_{t+1}(\mathbf{X}_{t+1}) dF_{t+1}(\mathbf{q}_{t+1}) \Big\} \\ s.t. \quad & \text{i) the intertemporal budget constraint in equation (3.1)} \\ & \text{ii) the borrowing constraint in equation (3.2)} \\ & \text{iii) and the time budget constraint in equation (3.11)} \end{aligned} \tag{3.8}$$

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. The (altruistic) parents take this into account in making their decisions. This continuation utility is discounted by the altruism parameter (λ) and the integration is with respect to the children's intial wage draw and the characteristics of their spouse (these shocks are realized after the couple makes their decisions). 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). The value function (given in equation (3.9)) must be integrated with respect to next period's wage draws, which are stochastic, and discounted by β_{t+1}, the time discount factor.

3.3.6.3 Decision Problem in the Late Adult phase

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

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

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

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

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

$$
\begin{aligned}
V_t(\mathbf{X}_t) \quad &= \quad \max_{c_{m,t}, c_{f,t}, h_{m,t}, h_{f,t}} \Big\{ u(c_{m,t}, c_{f,t}, l_{m,t}, l_{f,t}) + \\
& \qquad \beta_{t+1} s_{t+1} \int V_{t+1}(\mathbf{X}_{t+1}) dF_{t+1}(\mathbf{q}_{t+1}) \Big\} \\
s.t. & \quad \text{the intertemporal budget constraint in equation (3.1)} \\
& \quad \text{and the time budget constraint in equation (3.11)}
\end{aligned} \tag{3.9}
$$

where s_{t+1} is the probability of surviving to period $t+1$, conditional on having survived to period t.

3.4 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.

3.4.1 Estimating the Human Capital Production Function

Estimating the production function of human capital 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 (2016a)), we estimate a production function where latent ability is a function of previous period's (latent) ability and investments, parental education, and a shock:

$$ab'_{t'+1} = \alpha_{1,t'} ab'_{t'} + \alpha_{2,t'} inv_{t'} + \alpha_{3,t'} inv_{t'} \cdot ab_{t'} + \alpha_{4,t'} ed^m + \alpha_{5,t'} ed^f + u'_{ab,t'} \quad (3.10)$$

We explicitly account for measurement error in the latent factors, by developing an efficient GMM implementation of the methods in Agostinelli and Wiswall (2016a, AW). In the appendix, we show how to exploit multiple measures (as in Cunha and Heckman (2008), Cunha et al. (2010)) but using a simpler system GMM approach rather than maximum likelihood and filtering methods. We assume independence of measurement errors and use the noisy measures to instrument for one another. An important extension, relative to AW, is that we use many possible combinations of input measures to instrument for one another. See Appendix 3.D for more details.

Converting latent investments to time Equation (3.10) 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 equation, which we discuss in Section 3.G 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:

$$inv_{t'} = \kappa_{0,t'} + \kappa_{1,t'}(ti_{m,t'} + ti_{f,t'}) \quad (3.11)$$

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.

Next, we estimate the parameters $\kappa_{0,t'}$ and $\kappa_{1,t'}$ 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. See Section 3.4.3 for details on the estimation and identification of $\kappa_{0,t'}$ and $\kappa_{1,t'}$.

With the parameters κ_0 and κ_1 at hand, we substitute equation (3.11) into equation (3.10), giving us equation (3.5) which is the production function we use in our dynamic programming model:

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

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

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

We estimate the wage equation laid out in Section 3.3:

$$\ln w_t^* = \ln w_t + u_t = \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 PT_t + v_t + u_t \text{ where} \tag{3.12}$$

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

$$v_4 = \delta_5 ab + \eta_4$$

u_t is IID measurement error in wages

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

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

The above procedure addresses problems of measurement error in ability and wages but not selection. We control for selection bias by finding the wage profile that, when fed into our model, generates the same estimated profile (i.e., the same δ parameters from equation (3.12)) 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 3.G for details.

3.4.3 Method of Simulated Moments

In the second step, we estimate the rest of the model's parameters (discount factor, consumption weight for both husband and wife, risk aversion, altruism weight, share of time with the child that represents leisure to the parent, the hours-to-latent investments conversions):

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

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

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

We discretize the asset and also the ability and wage grids for both spouses and, using value function iteration, we solve the model numerically. This yields a set of decision rules, which, in combination with the simulated endowments and shocks, allows us to simulate each individual's assets, work hours and home investment hours, child's educational choices, and inter-vivos transfers. We use the resulting profiles to construct moment conditions, and evaluate the match using our GMM criterion. We search over the parameter space for the values that minimize the criterion. Appendix 3.I contains a detailed description of our moment conditions, the weighting matrix in our GMM criterion function, and the asymptotic distribution of our parameter estimates.

3.5 First Step Estimation Results

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

3.5.1 The Determinants of Ability

In Section 3.2 we documented that children of high educated parents do better in cognitive tests, and that the ability gaps between children of high and low educated parents grow over time. Here we combine the multiple test scores to create a measure of skills, and estimate a human capital production function using the methods described briefly in Section 3.4.1 and in more detail in the appendix. Similar to our approach to estimating the determinants of time investment, we use a GMM estimator with a diagonal weighting matrix.

We allow for the fact that high education parents may have high ability children. We assume that parental education determines the initial draw of the child ability (i.e. at birth). See appendix 3.E for more on this.

We estimate equation (3.5) for ability at ages 7, 11, and 16. The time investments entering the equation are those corresponding to ages 0-6, 7-10, and 11-16, respectively (when the parents are aged 26-32,33-36, and 37-41).

To ease interpretation, we normalize our ability and time measure to have variance one in every period.

We estimate the relationship between age 7 ability as a function of age 0 ability, age 0 time investments, the interaction of ability and time investments, and mother's and father's education. Estimates are presented in Table 3.5.1. It shows that time investments have a significant effect on changes in ability over time, even after conditioning on background characteristics and initial ability. A one SD increase in time investments at age 0 raises age-7 ability by 0.21 SD, a one SD increase in time investments at age 7 raises age-11 ability by 0.12 SD, and a one percent increase in time investments at age 11 raises age-16 ability by 0.11 SD.

Ability is very persistent, especially at older ages.

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 does in fact exhibit dynamic complementarity at this stage of childhood (as found by Cunha et al. (2010)).

We find that parental skill, as measured by parental education, strongly impacts future ability, providing empirical support for a key mechanism for perpetuating inequality across generations. High skill parents are effective in producing human capital in their children (as also shown in some of the papers cited in Heckman and

Table 3.5.1: Determinants of log ability.

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

Notes: GMM estimates. Confidence intervals are bootstrapped using 100 replications. For the production function at age 7, we use 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).

Mosso (2014) and is assumed in Becker et al. (2018) and Lee and Seshadri (2019)) in addition to having more resources to afford college.

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

3.5.2 The Effect of Ability and Education on Wages

In the dynastic model with intergenerational altruism laid out in Section 3.3 parents do not receive any direct return from their children having higher ability at the age of 23. Instead, they include their children's expected lifetime utility in their own value function, with a weight determined by the intergenerational altruism parameter λ. Hence parental investments in children's ability (both through time and money investments in education) will be driven by the return to ability in the labour market. Here we focus on the return to ability in the labour market, as measured by its impact on wages. We estimate the wage equation (3.12) laid out in Section 3.3 and Appendix 3.G for each gender and education group.

We first turn to Table 3.5.2 which shows the persistence and variance of innovations to wages. The results indicate that $\rho = 0.942 - 0.984$ depending on the group; in several groups the coefficient is not significantly different from 1. We thus set $\rho = 1$ in the model – this also means that ability has a permanent effect on wages. The estimate of σ_η^2 =0.005-0.021; one standard deviation of an innovation in the wage is 7-14% of wages, depending on the group. These estimates are similar to other papers in the literature (e.g. French (2005), Blundell et al. (2016)) and imply that long run forecast errors may be large. Furthermore, we find evidence that the variance of wage innovations is increasing with education.

Table 3.5.2: Persistence and variance of innovations to wages, by education level

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

Ability impacts wages through its relationship with education, but it also has a direct impact on wages conditional on education via the initial wage shock. This is shown by Table 3.5.3, which shows the estimates of δ_5 for each gender and education group. The interpretation of these coefficients is that they are estimates of the log-point increase in wages associated with a one SD increase in age-16 ability, conditional on education. The extent of complementary is similar to that estimated in Delaney (2019), 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)) although this finding has received less attention in the literature on estimation of ability production functions.

Table 3.5.3: Log-point change in wages for a 1 SD increase in ability, by education level

	Male	Female
Low	0.085 (0.020)	0.103 (0.025)
Middle	0.146 (0.016)	0.101 (0.016)
High	0.194 (0.022)	0.134 (0.023)

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

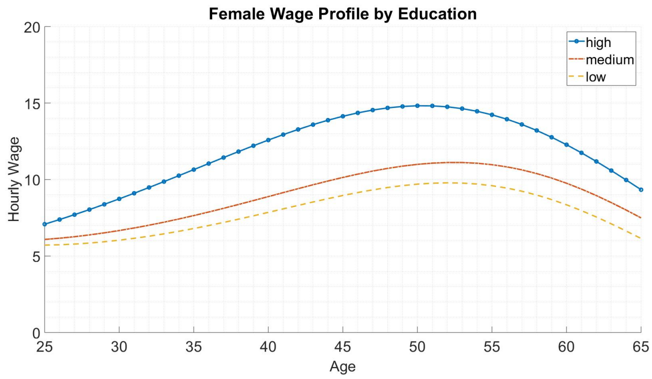
The table shows that, as one would expect, age-16 ability has a significant positive impact on wages conditional on education for all groups. Perhaps most interestingly, it shows evidence of complementarity between education and ability in the labour market, particularly for men. While low education men see only a 0.085 log-point increase in hourly wages for every additional SD of ability, high education men (with some college education) see an average increase of 0.19 log-points in hourly wages for every additional SD of ability.

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

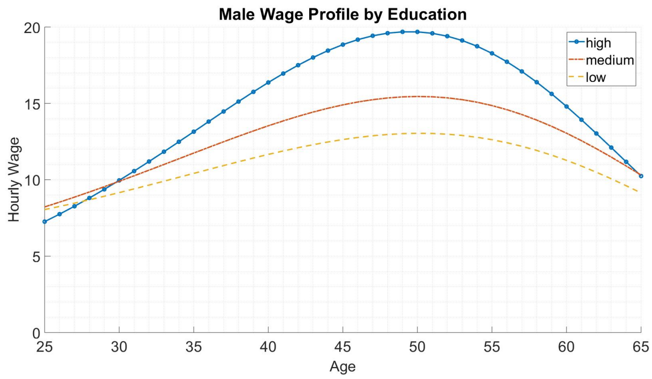
Figure 3.5.1 shows wage profiles by age, education and gender for full time workers with average ability. Men and those with high education have higher wages and faster wage growth.

Figure 3.5.1: Wages, by age, education and gender

Females



Males



Note: Wages measured in £2014.

3.5.3 Marital Matching Probabilities

Table 3.5.4 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 3.5.4 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 3.5.4: Marital matching probabilities, by education

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

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

3.5.4 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 (age 23), working adult (ages 26-60), pension age (age 65, onwards). This three parameter tax system has the following functional form: $y_t = d_{0,t} + d_{1,t}(e_{m,t} + e_{f,t})^{d_{2,t}}$. We set the time endowment to $T = 16$ available hours per day $\times$ 7 days per week $\times$ 52 weeks per year=5,824 hours per year. We use the modified OECD equivalence scale and set $n_t = 1.4$ for couples with children. Survival probabilities are calculated using national life tables from the Office for National Statistics.

3.6 Second Step Results, Identification, and Model Fit

We now turn to presenting our second-step parameter estimates for our structural model, the model fit, and we discuss the model's identification.

3.6.1 Utility Function Estimates and Identification

Table 3.6.1: Estimated structural parameters.

Parameter	Estimate
β: discount factor	0.98
ν_f: consumption weight, female	0.43
ν_m: consumption weight, male	0.40
γ: risk aversion	2.17
λ: altruism parameter	0.40
θ: time cost of investment	0.05
$\kappa_{1,1}$: latent investments per hour, ages 0-6	0.12
$\kappa_{1,2}$: latent investments per hour, ages 7-10	0.18
$\kappa_{1,3}$: latent investments per hour, ages 11-15	0.25
Coefficient of relative risk aversion, consumption*	1.49

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

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

The coefficient of relative risk aversion for consumption is 1.49 averaging over men and women,[7] which is similar to previous estimates that rely on different

[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.

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

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

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

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

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

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

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

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

$$\begin{aligned} w_{g,t}(1-\tau'_{g,t}) &\leq -\frac{\partial u_t}{\partial h_{g,t}} / \frac{\partial u}{\partial c_g} \\ &\leq -\frac{1-v_{g,t}}{v_{g,t}} / \frac{c_{g,t}}{l_{g,t}} \end{aligned} \tag{3.13}$$

which holds with equality when work hours are positive, where $\tau'_{g,t}$ is individual g's marginal tax rate at time t.[10] Inserting the time endownment equation (3.11) into equation (3.13) and making the approximation $c_{g,t} \approx w_{g,t}h_{g,t}(1-\tau'_{g,t})$ yields

$$v_g \approx \frac{h_{g,t}}{T - ti_{g,t}}. \tag{3.14}$$

Thus v_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 v_g is modestly less than .5 for both men and women.

The parameter λ is identified from three sources.

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

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

Second, households invest in the formal education of their children. The foregone household income from children going to school represents a direct loss of resources to the household.

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

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

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

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

Our estimate of the altruism parameter λ is 0.4 and in line with estimates by Daruich who estimates it to be 0.48 and Lee & Seshadri who estimate it to be 0.32. Some papers in the literature, however, also present higher estimates such as Gayle, Golan, Soytas who's estimate is 0.80 and Caucutt & Lochner who's estimate is 0.86.

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 θ =0 then time with children is pure leisure for the parent, whereas if $\theta = 1$ then time with children generates no leisure value. Thus, if $\theta = 1$ then the opportunity cost of one additional hour with the child should be (approximately) equal to the resulting increase in the expected present discounted value of higher lifetime income of child from that time investment. Likewise, if θ =0 and time with children is pure leisure, then parents would spend time with their children even if it did not affect the child's future wages. See appendix 3.K for a more formal derivation. We estimate θ to be 0.05, meaning that 95% of the time that parents spend with their children actually 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.

3.6.2 Model Fit

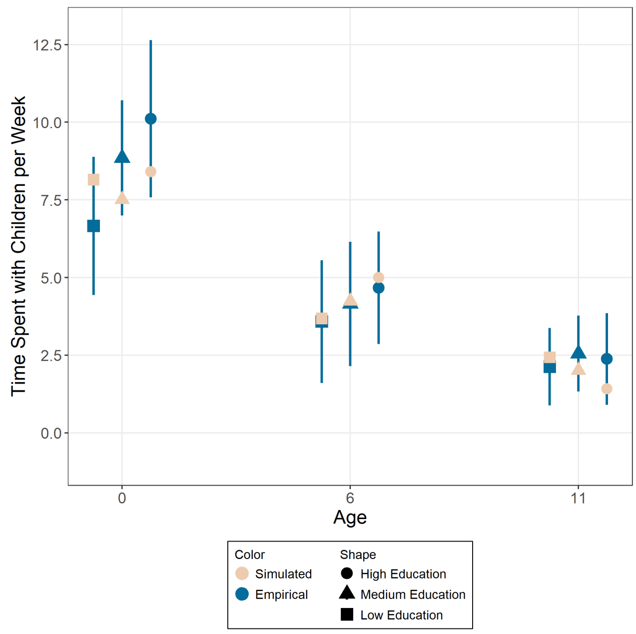
In this section we focus on the moments that are critical for understanding intergenerational altruism: transfers of time, educational investments, and money.

Figure 3.6.1 shows transfers of time from mothers and fathers in the left and right panels, respectively. The model fits well three key patterns in the data. First, time investments decline with age. Second, mothers invest more in their children than fathers. Third, high education parents invest more time in their children than low education parents.

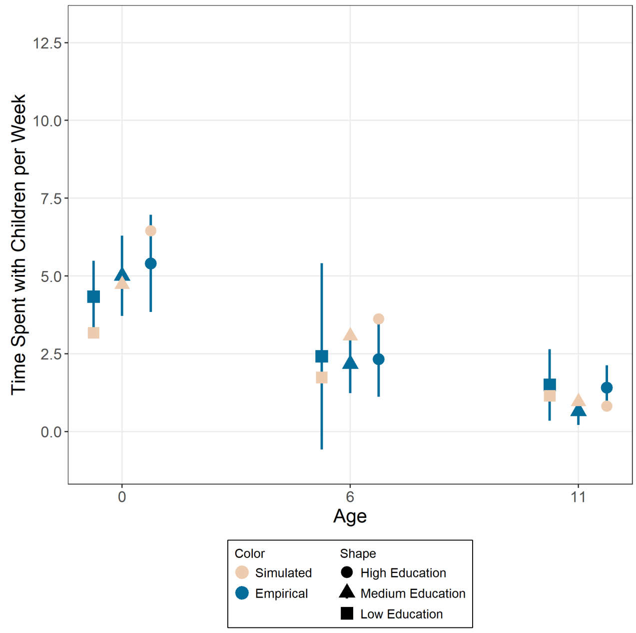
This higher level of time investments of educated parents, in combination with their greater productivity of these investments, leads to higher ability of their children as can be seen in panel (a) of Figure 3.6.2. Our model captures well how higher time investments of the educated lead to higher ability 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 3.6.2 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.

Figure 3.6.1: Model fit: parental time with children



(a) Mother's Time with Kids



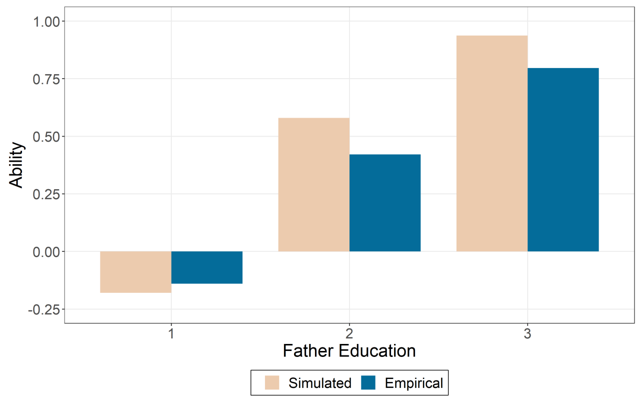
(b) Father's Time with Kids

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

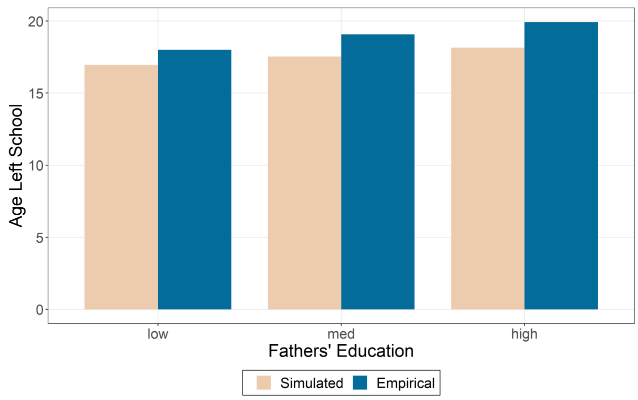
Table 3.6.2 shows that we match well the mean level of financial transfers received and the median level of assets at age 60. These financial transfers include inter-vivos transfers when younger and bequests received when older. These amounts are discounted to age 23: when undiscounted, the amounts are considerably larger.

Finally, our model can reproduce key labour supply moments of men and women with different education levels as shown in Appendix 3.J. 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 3.6.2: 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 3.6.2: 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 3.C for more details.

3.7 Results

3.7.1 How is Income Risk Resolved with Age?

How much of the cross-sectional variance in lifetime income can we predict 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. In this section, we show how much of the cross-sectional variance in lifetime income can be predicted at each point in the lifecycle. We also disentangle the relative roles of choices, luck and family circumstances in determining lifetime income.

We simulate data for two generations. These data contain information, at each age, on parental assets, wages, and education, as well as the child's ability, gender, education, wages, and their spouse's education, and wages. Using these data we first calculate lifetime income for each child's household; next, we calculate the share of the variability of lifetime income that is predictable given information known at different points in time. 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. (2011b) and Lee and Seshadri (2019), who focus on the amount of variability known at a particular age. We extend their approach to show the evolution of the amount of variability known as the child ages.

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 variation 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.

Table 3.7.1: Explained outcome variance: Evolution over childhood

Parents' age	23	26	33	37	42	49
Child's age	NA	0	7	11	16	23
Male Children						
Individual's wage	28%	32%	37%	41%	45%	62%
Household's income	16%	19%	23%	26%	28%	41%
Female Children						
Individual's wage	6%	7%	8%	10%	11%	36%
Household's income	4%	6%	8%	10%	12%	20%

Notes: Table shows variance of individual's wage and household income that can be explained by state variables known at different ages over an individual's lifecycle.

Table 3.7.1 reports these decompositions for both males and females. It shows decompositions for individual lifetime wages and for total household lifetime income. Individual lifetime wages are calculated as after-tax income from wages between ages 23 and 50 if the individual worked full time in every period and the spouse did not work. In reality, many females work part-time rather than full-time; hence for females, these numbers reflect the difference in earnings *potential* rather than actual earnings differences. Wages are discounted using the model interest rate. The table shows that, before birth, when the parent is 23, 28% and 6% of lifetime wages are known for males and females, respectively. This is the share explained by parents' education (which affects initial ability and the productivity of parental investments) and also household financial resources at that age (which affects the amount of investments the child receives over their life). As the child ages, new information is realized, both about their own ability and their parents' financial resources. Immediately after birth, the share of lifetime wages explained is 32% and 7% for males and females, respectively and this share rises to 45% and 11% at age 16. At age 23, after educational choices are made, and the initial wage draw is received, this rises to 62% and 36% for males and females respectively. 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. In our model, wages are explained by ability, education, and subsequent wage shocks. The higher return to ability means that the share of lifetime wage variability that is attributable to ability is higher for men.

Household lifetime income more fully captures well-being: it comprises the labor income of both self and spouse, and also parental transfers received. This measure differs from individual wages because of the extent of assortive mating, labor supply choices, and transfers received. At age 23 (before any characteristics of the spouse are realized) the share of lifetime household income explained is 41% and 20% for males and females, respectively. The main reason for the large difference in shares explained by gender is that men contribute more to household income, hence women's characteristics are less predictive of household income.

Upon reaching adulthood, the process of household formation explains more of the remaining variability in outcomes. Whereas Table 3.7.1 only conditions on characteristics of the individual, Table 3.7.2 also conditions on characteristics of the spouse at the time of marriage at age 23, and thus includes all information at the household level. The first column brings in information from the spouse's wage shock. This shows that the share of household lifetime wages (the sum of lifetime wages for both husband and wife) explained rises to 55%. In the second column we also include the spouse's education, which increases the share explained to 65%. Household wages are more explainable than individual wages because lifetime wages are a function of known age 23 characteristics and future wage shocks. These shocks average out over the two spouses, giving a measure of insurance within the household. The third column includes parental transfers. Given all the other variables known, these transfers are highly predictable and thus the added contribution from these transfers is small.

This exercise shows us how uncertainty is resolved over an individual's lifecycle, which gives a measure of the importance of luck but does not disentangle the

Table 3.7.2: Outcome variance explained after revelation of: Spouse's wage, spouse's education, cash transfers

	Spouse's wage	Spouse's education	Transfers
Wage of household	55%	65%	65%
Household's income	38%	44%	46%

Notes: First two columns show variance of wage of household and household income, that can be explained once a spouse's wage and education is known. Last column shows fraction of variance explained once cash transfers received by the household are known (at age 23). Household wage is the sum of lifetime wages for both husband and wife.

relative importance of differences in investments versus the productivity of those investments. To address the relative importance of choices relative to other variables, we run counterfactual experiments where we equalize parental time investments, education, and money transfers, and assess how this impacts lifetime wages and income. To do this we constrain choices when simulating the model, thus forcing an unanticipated change on the parents. For example, in the case of education, all parents must now send their children to university. However, parental labor supply and savings are unconstrained, as in the baseline specification. Table 3.7.3 shows the results for males. Differences in time investments matter substantially for wages and account for 28% of the variance. Moreover, we can see that differences in time investments during childhood matter more in affecting an individual's wages than differences in educational investments. However, once we look at the household wage, we can see that time investments and education explain the same fraction. This is because assortative matching in the model depends on education. Thus, when considering the variance of the combined wage of a household, differences in time investments received matter as much as differences in educational investments.

3.7.2 Intergenerational Elasticities

We now turn to the patterns of intergenerational persistence that the model generates. We focus on log-log intergenerational elasticities (IGE) of child outcomes with respect to parental outcomes. We do this for the following outcomes: lifetime earnings, lifetime wages, lifetime hours, lifetime consumption, and average wealth

Table 3.7.3: Fraction of outcome variance for males explained by time investments and education

Equalize:	Time Investments	Education
Individual's wage	28%	16%
Wage of household	24%	24%

Notes: In the first column, all time investments are equalized to mean time investments in the simulated sample. In the second column, education is set to the highest level of education for everyone. Household wage is the sum of male and female wage in the household.

over the lifecycle. We estimate the following regression on our simulated data:

$$\ln y' = a_0 + a_1 \ln y + u \tag{3.15}$$

where $\ln y'$ denotes the log of the child's outcome (e.g. child's lifetime income) and $\ln y$ the log of parents' outcome(e.g. parents' lifetime income).

3.7.2.1 Intergenerational Elasticities

We begin by focussing on the intergenerational elasticity of lifetime household earnings. The model predicts a value of 0.16, which is similar to the estimated values reported in Belfield et al. (2017) and Bolt et al. (2021). In contrast, the intergenerational elasticity of wages is much higher at 0.32. These two numbers can be reconciled when considering that earnings are a function of wages and labor supply. When we look at the intergenerational elasticity of labour supply, we actually find it to be negative, as individuals from richer families expect to receive more transfers from their parents, lowering their incentive to work. Thus, even though children from rich families receive higher wages, they supply less labour, which leads to a lower intergenerational elasticity of earnings. The intergenerational elasticity of consumption is 0.36 which is in line with Charles et al. (2014), who find that the intergenerational correlation in total consumption is 0.28 using PSID data. Finally, we calculate the intergenerational elasticity of average wealth over the lifecycle and find it to be 0.33. This is similar to estimates by Charles and Hurst (2003) and Gregg et al. (2021). From this exercise, we conclude that out model is very capable of reproducing patterns of intergenerational persistence, thus making it suitable for evaluating policy counterfactuals.

Table 3.7.4: Model Predicted Intergenerational Elasticity, Selected Variables

Earnings	0.22
Wages	0.32
Labour Supply	-0.19
Consumption	0.36
Wealth	0.33

Notes: Intergenerational elasticities calculated from model simulated data. Earnings, wages, labour supply, consumption calculated as average over ages 23-65. Wealth is calculated as the average wealth between ages 23-65.

3.8 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.

Appendix

3.A Parameter definitions

Table 3.A.1 summarises the parameters that enter the model and which are introduced in the body of the paper (excluding the appendices).

Table 3.A.1: Parameter definitions

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

3.B Time Periods, States, Choices and Uncertainty

Table 3.B.1 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.

3.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.

3.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 3.2.1 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.

3.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 conditions. ELSA is a biannual survey of those 50 and older, starting in 2002. We use data up through 2016.

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

For the asset moment condition at age 50 we begin with 924 respondents who are age 50 at the time of the survey. We drop members of cohorts not born between 1950-1959 (which excludes 255 observations), unmarried people (which excludes 88 observations), and the self-employed (which excludes 54 observations). Finally, we have 14 households where both members were exactly age 50. In order to not double count these households, we exclude one observation from these two person households, resulting in 513 individuals remaining.

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

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

3.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 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 $\leq$ 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)).

3.D Estimation of the Ability Production Function and Wage Function

3.D.1 Production Function

The production function for skills that we estimate is as specified in equation (3.10) in the main text.

$$ab'_{t'+1} = \alpha_{1,t'}ab'_{t'} + \alpha_{2,t'}inv_{t'} + \alpha_{3,t'}inv_{t'} \cdot ab_{t'} + \alpha_{4,t'}ed^m + \alpha_{5,t'}ed^f + u'_{ab,t'} \quad (3.16)$$

3.D.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:

$$Z_{\omega,t,j} = \mu_{\omega,t,j} + \lambda_{\omega,t,j}\omega_t + \varepsilon_{\omega,t,j} \quad (3.17)$$

All other variables are assumed to be measured without error.

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

3.D.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 (3.16). This is not the case in our estimation as we do not assume $\alpha_{1,t'} + \alpha_{2,t'} + \alpha_{3,t'} = 1$ when estimating equation (3.16). For more details, see Agostinelli and Wiswall (2016b).

3.D.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.

3.D.6 Estimation

1. **Variance of latent factors**.

   Using equation (3.17) we can derive the variance of each of the latent factors:

   $$Cov(Z_{\omega,t,j}, Z_{\omega,t,j^*}) = \lambda_{\omega,t,j}\lambda_{\omega,t,j^*}\, Var(\omega_t) \tag{3.18}$$

   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 (3.18), 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 (3.18) we can show that:

   $$\lambda_{\omega,t,m} = \frac{Cov(Z_{\omega,t,j}, Z_{\omega,t,j^*})}{Cov(Z_{\omega,t,1}, Z_{\omega,t,j^*})} \tag{3.19}$$

   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:

$$\mu_{ab',1,j} = \mathbb{E}[Z_{ab',1,j}], \;\; \mu_{inv,6,j} = \mathbb{E}[Z_{inv,6,j}] \tag{3.20}$$

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

$$\tilde{Z}_{\omega,t,j} = \frac{Z_{\omega,t,j} - \mu_{\omega,t,j}}{\lambda_{\omega,t,j}} \tag{3.21}$$

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

$$\omega_t = \tilde{Z}_{\omega,t,j} - \underbrace{\frac{\varepsilon_{\omega,t,j}}{\lambda_{\omega,t,j}}}_{\equiv\tilde{\varepsilon}_{\omega,t,j}} \tag{3.22}$$

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 (3.22) note that we can rewrite the above equation as:

$$\begin{aligned} &\frac{Z_{ab',t'+1,j} - \mu_{ab',t'+1,j} - \varepsilon_{ab',t'+1,j}}{\lambda_{ab',t'+1,j}} = \\ &\alpha_{1,t'}(\tilde{Z}_{ab',t',j} - \tilde{\varepsilon}_{ab',t',j}) + \alpha_{2,t'}(\tilde{Z}_{inv,t',j} - \tilde{\varepsilon}_{inv,t',j}) + \\ &\alpha_{3,t'}(\tilde{Z}_{inv,t',j} - \tilde{\varepsilon}_{inv,t',j}) \cdot (\tilde{Z}_{ab',t',j} - \tilde{\varepsilon}_{ab',t',j}) + \\ &\alpha_{4,t'} ed_m + \alpha_{5,t'} ed_f + u_{ab',t'} \end{aligned} \tag{3.23}$$

or

$$\begin{aligned}
&\frac{Z_{ab',t'+1,j} - \mu_{ab',t'+1,j}}{\lambda_{ab',t'+1,j}} = \\
&\alpha_{1,t'}\tilde{Z}_{ab',t',j} + \alpha_{2,t'}\tilde{Z}_{inv,t',j} + \\
&\alpha_{3,t'}\tilde{Z}_{inv,t',j} \cdot \tilde{Z}_{ab',t',j} + \\
&\alpha_{4,t'}ed_m + \alpha_{5,t'}ed_f + \\
&\Bigg(u_{ab',t'} - \tilde{\varepsilon}_{ab',t',j} - \tilde{\varepsilon}_{inv,t',j} - \tilde{\varepsilon}_{inv,t',j} \cdot \tilde{\varepsilon}_{ab',t',j} + \frac{\varepsilon_{ab',t'+1,j}}{\lambda_{ab',t'+1,j}} \\
&- \alpha_{3,t'}(\tilde{Z}_{inv,t',j}\tilde{\varepsilon}_{ab',t',j} + \tilde{Z}_{ab',t',j}\tilde{\varepsilon}_{inv,t',j}) \Bigg).
\end{aligned} \tag{3.24}$$

OLS is inconsistent here, as $\tilde{Z}_{ab',t',j}$ and $\tilde{\varepsilon}_{ab',t',j}$ 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 (3.24) 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 (3.24) 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 (3.24), where $\pi_{ab,t,j} \equiv \Bigg(u_{ab',t'} - \tilde{\varepsilon}_{ab',t',j} - \tilde{\varepsilon}_{inv,t',j} - \tilde{\varepsilon}_{inv,t',j} \cdot \tilde{\varepsilon}_{ab',t',j} + \frac{\varepsilon_{ab',t'+1,j}}{\lambda_{ab',t'+1,j}} - \alpha_{3,t'}(\tilde{Z}_{inv,t',j}\tilde{\varepsilon}_{ab',t',j} + \tilde{Z}_{ab',t',j}\tilde{\varepsilon}_{inv,t,j}) \Bigg)$:

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

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.

3.E Initial Ability

We allow for the fact that high education parents may have high ability children. We assume that child ability at birth is log-normally distributed conditional on parental education. The mean of this distribution is estimated by adjusting the means of the gestation and birthweight measure by their respective scaling parameters (λ defined above) and combining them using a minimum distance estimator. The variance of the distribution is estimated by dividing the covariance of the two measures by the product of the scaling parameters, as described in step 1 of section 3.D.6. This is done separately for each combination of maternal and paternal education groups. Table 3.E.1 shows the mean and initial variance of ability for each parental education group.

3.F Signal to Noise Ratios

Note that using equation (3.17) the variance of measure $Z_{\omega,t,j} = (\lambda^2_{\omega,t,j})Var(\theta_{\omega,t}) + Var(\varepsilon_{\omega,t,j})$, where $(\lambda^2_{\omega,t,j})Var(\theta_{\omega,t})$ comes from the variability in the signal in the measure and $Var(\varepsilon_{\omega,t,j})$ represents measurement error, or "noise". The signal to noise ratios for measure $Z_{\omega,t,j}$ is calculated in the following way:

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

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 3.F.1 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 3.F.2 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 3.F.2: 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.

3.G Estimation of the Wage Equation

We estimate the wage equation laid out in Section 3.3. We present the estimation equation in (3.12) for convenience:

$$\ln w_t^* = \ln w_t + u_t = \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 ab_4 + \delta_5 PT_t + v_t + u_t \text{ where} \tag{3.25}$$

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

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

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

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

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

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

$$\begin{aligned} \ln w_t^* &= \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_4 ab_4 + \delta_5 PT_t + u_t + v_5 + \sum_{k=6}^{t} \eta_k \\ &= \delta_1 t + \delta_2 t^2 + \delta_3 t^3 + \delta_5 PT_t + FE + \xi_t \end{aligned} \tag{3.26}$$

where FE is a person specific fixed effect capturing the time invariant factors $\delta_4 ab_4 + v_5$ and ξ_t is a residual.

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

Step 2: Predict the fixed effect:

$$\begin{aligned} \widehat{FE} &\equiv \bar{\ln w_t^*} - \hat{\delta}_1 \bar{t} - \hat{\delta}_2 \bar{t}^2 - \hat{\delta}_3 \bar{t}^3 - \hat{\delta}_5 \bar{PT}_t \\ &= \delta_0 + \delta_4 ab_4 + v_5 \\ &= \delta_0 + \delta_4 \tilde{Z}_{ab,4,j} + v_5 - \delta_4 \tilde{\varepsilon}_{ab,4,j} \end{aligned} \tag{3.27}$$

where $ab_4 = \tilde{Z}_{ab,4,j} - \tilde{\varepsilon}_{ab,4,j}$ is defined in equation (3.22). Although the estimated fixed effect, $\widehat{FE}$, is affected by variability in the sequence of wage shocks $\{\eta_t\}_{t=6}^{12}$ and measurement errors $\{u_t\}_{t=5}^{12}$, this merely adds in measurement error on the left hand side variable in equation (3.27) . 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 $\tilde{\varepsilon}_{ab,4,j}$ by construction. We address this problem in the next step.

Step 3: Using GMM, project the predicted fixed effect ($\widehat{FE}$) on each measure of ability, $\tilde{Z}_{ab,4,j}$, and instrument by using the respective other measures, $\tilde{Z}_{ab,4,j'}$, to get $\hat{\delta}_0$ and $\hat{\delta}_4$. Since we have two measures of ability (reading and math), we have two equations and two instruments. When reading is the ability measure, we instrument for this using math, and vice versa. The GMM efficiently combines different measures of ability and simultaneously correct for measurement error.

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

Redefining equation (3.25) using $ab_4 = \tilde{Z}_{ab,4,j} - \tilde{\varepsilon}_{ab,4,j}$ as defined in equation (3.22).:

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

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

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

Note that from the measurement equation 3.17, $V(\tilde{Z}_{ab,4,j}) = V(ab_4) + V(\tilde{\varepsilon}_{ab,4,j})$, where we have previously estimated $V(ab_4)$ using equation 3.18 and $V(\tilde{Z}_{ab,4,j})$ is the variance of the renormalized measures in the data. We can then back out the variance of the measurement error and plug it into the following equation to estimate the parameters of the wage shocks:

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

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

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

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).

3.H Computational Details

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

1. To find optimal decision rules, we solve the model backwards using value function iteration. The state variables of the model are model period, assets, wage rates, education levels, own ability, childrens' gender, childrens' ability, and childrens' education. At each model period, we solve the model for 25 grid points for assets, 15 points for wage rates (for each spouse), 3 education levels for each spouse, 5 points for own ability for each spouse, childrens' gender, childrens' ability (5 points), and childrens' education. Because we assume that the two children are identical, receive identical shocks, and that parent make identical decisions towards the two children, we only need to keep track of the state variables for one child. Our approach for discretizing wage shocks follows 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 3 levels of working hours (non-employed, part-time, full-time) and in model period $t = 6, 7$ and 8 they can choose between four levels of time spent with children. In all model periods except $t = 10$ we solve for the optimal level of next period assets using golden search. In period $t = 10$ as parents may also transfer assets to children: we solve this two-dimensional optimization problem using Nelder-Mead. We back out household consumption from the budget constraint and then solve for individual level consumption from the intra-temporal first order condition, which delivers the share of household consumption going to the male in the household. As this first order condition is a non-linear function we approximate the solution using the first step of a third order Householder algorithm. This allows us to use the information contained in the first three derivatives of the first order condition. We found this method to give fast and accurate solutions to the intra-temporal problem. Details of this are available from the authors.

Next we describe our simulation procedure.

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

2. Given the optimal decision rules, the initial conditions of the state variables, and the histories of shocks faced by both parents and children, we calculate life histories for savings, consumption, labor supply, time and education

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

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

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

3.I Moment Conditions and Asymptotic Distribution of Parameter Estimates

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

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

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

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

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

Suppose we have a dataset of I independent individuals that are each observed at up to T separate calendar years.

Let $\varphi(\Delta;\chi_0)$ denote the J-element vector of moment conditions described immediately above, and let $\hat{\varphi}_I(.)$ denote its sample analog. Letting $\widehat{\mathbf{W}}_I$ denote a $J \times J$ weighting matrix, the MSM estimator $\hat{\Delta}$ is given by

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

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

In practice, we estimate χ_0 as well, using the approach described in the main text. Computational concerns, however, compel us to treat χ_0 as known in the analysis that follows. Under regularity conditions stated in 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;

$$\mathbf{D} = \frac{\partial\varphi(\Delta;\chi_0)}{\partial\Delta'}\Big|_{\Delta=\Delta_0} \tag{3.28}$$

is the $J \times M$ gradient matrix of the population moment vector; and $\mathbf{W} = \text{plim}_{I\to\infty}\{\widehat{\mathbf{W}}_I\}$. Moreover, Newey (1985) shows that if the model is properly specified,

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

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

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

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

But even though the optimal weighting matrix is asymptotically efficient, it can be biased in small samples. (See, for example, Altonji and Segal (1996).) We thus use a "diagonal" weighting matrix, as suggested by Pischke (1995). This diagonal weighting scheme uses the inverse of the matrix that is the same as $\mathbf{S}$ along the diagonal and has zeros off the diagonal of the matrix.

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

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). The data variance for moment m corresponding to mean hours at age t then becomes:

$$\begin{aligned}\hat{S}_{m,m} = \quad & \tfrac{1}{I}\textstyle\sum_{i=1}^{I}[hours_{i,t} - \sum_{i=1}^{I_t} hours_{i,t}]^2 \times I_{hours_{i,t} \neq missing} = \\ & \tfrac{I_t}{I}\tfrac{1}{I_t}\textstyle\sum_{i=1}^{I_t}[hours_{i,t} - \sum_{i=1}^{I_t} hours_{i,t}]^2\end{aligned} \tag{3.29}$$

Table 3.B.1: Model time periods, and states, choices and sources of uncertainty during those time periods

Time Periods																				
Model period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Parent generation's age	0	7	11	16	23	26	33	37	42	49	55	60	65	70	75	80	85	90	95	100
Child generation's age						0	7	11	16	23										
Parent generation's datasets																				
NCDS					x		x		x	x	x									
Time use survey						x	x	x												
ELSA					x		x			x										
Child generation's datasets																				
NCDS						x	x	x												
Parent generation's states																				
Assets					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Wage of male & female					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Education of male & female					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Ability of male & 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 & female					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Time spent with children, male & female						x	x	x												
Consumption, male & 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 & 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.

Table 3.C.1: 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.

Table 3.E.1: Means and variances of initial ability conditional on parental education group

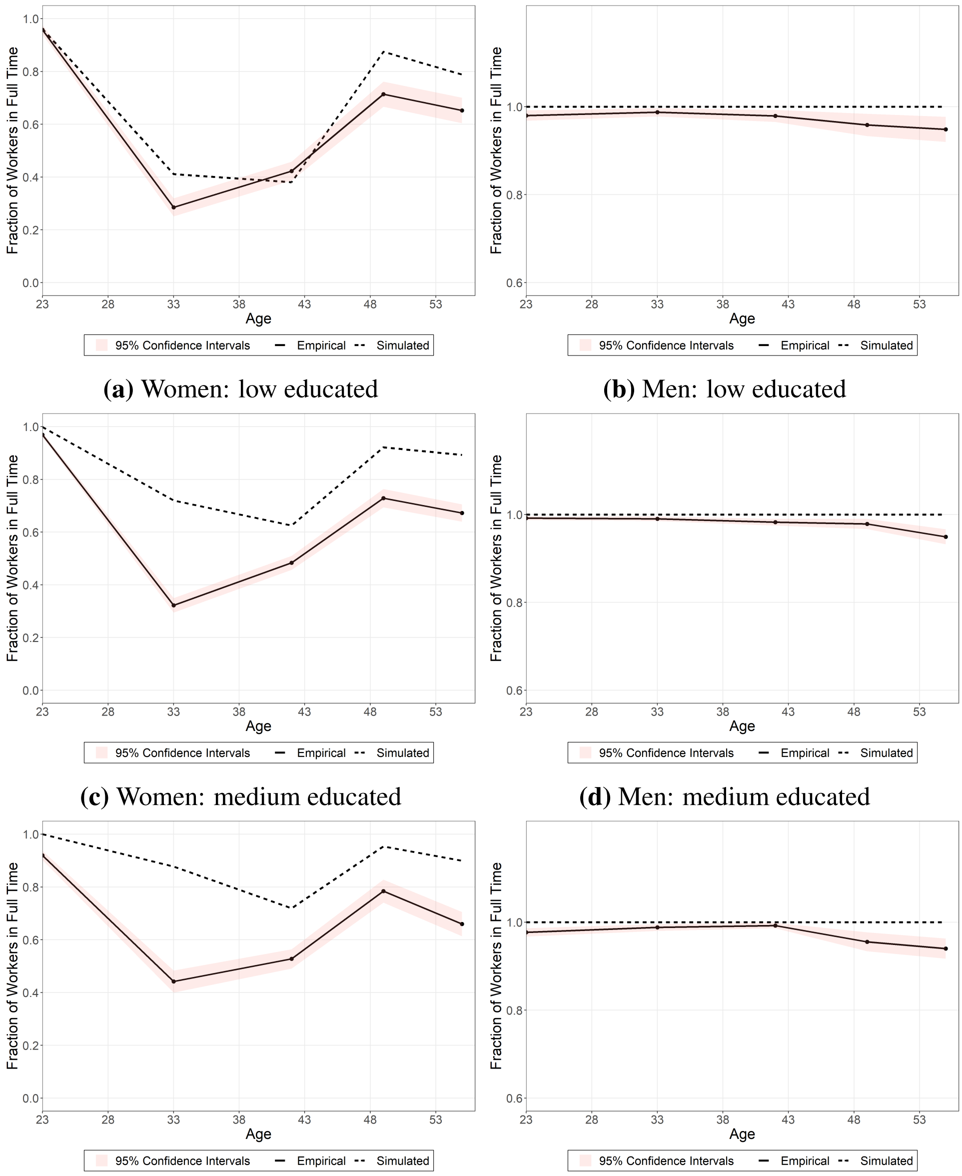
Means:				
		Mother's education		
		Low	Medium	High
	Low	0.845	-0.097	0.021
Father's education	Medium	-0.341	0.367	-0.916
	High	-1.345	0.389	0.915
Variances:				
		Mother's education		
		Low	Medium	High
	Low	2.606	3.487	4.878
Father's education	Medium	2.969	1.787	6.526
	High	4.319	2.194	0.840

Table 3.F.1: 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				

3.J Further Details on Model Fit

Figure 3.J.1: Model fit: full-time work conditional on employment



(a) Women: low educated

(b) Men: low educated

(c) Women: medium educated

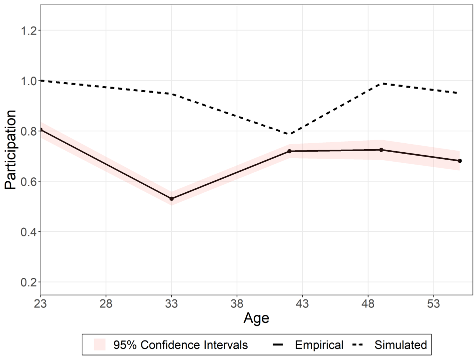
(d) Men: medium educated

(e) Women: high educated

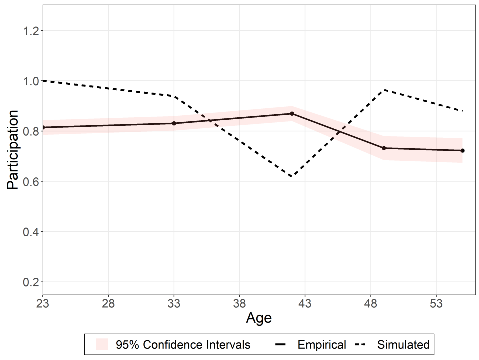
(f) Men: high educated

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

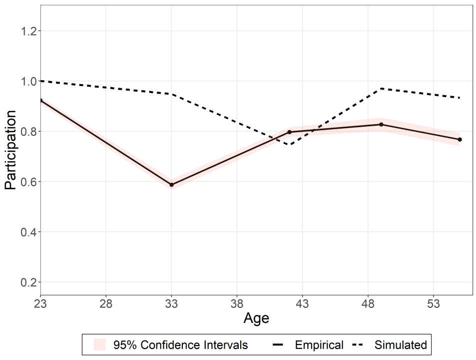
Figure 3.J.2: Model fit: participation



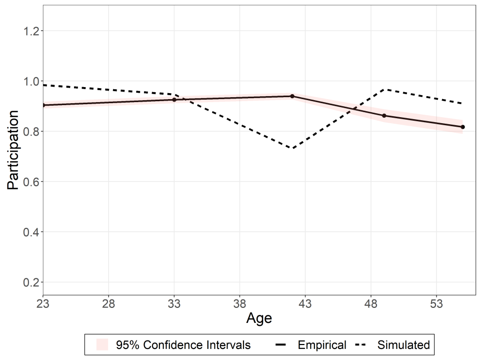
(a) Women: low educated



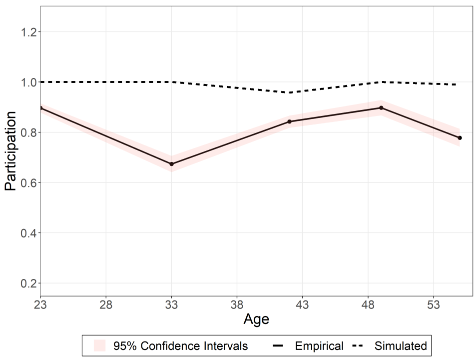
(b) Men: low educated



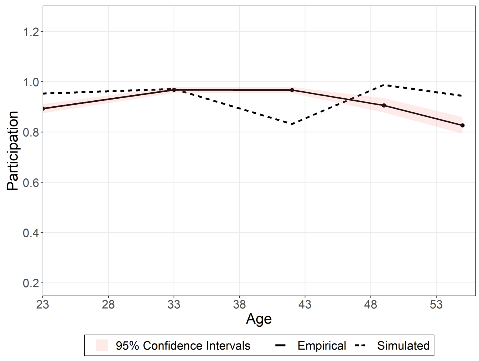
(c) Women: medium educated



(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.

3.K 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 uncertainty. The agent's states are: education *ed*, ability *ab*, and their initial assets x_1. Agents are altruistic towards their child and incorporate their child's value function into their problem, but discount it by factor λ. Agents 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. The child initially has no other assets than the monetary transfer from the parent. We first describe the discrete education choice decision problem of the parent:

$$V(ed, ab, x_1) = \max_{ed'=\{D,HS,C\}} \{V_{ed'=D}, V_{ed'=HS}, V_{ed'=C}\} \tag{3.30}$$

The above nests the following decision problem over consumption, leisure, time investments and asset transfers:

$$V_{ed'=k}(S) = \max\nolimits_{c_1,l_1,ti_1,x_1',c_2,l_2,ti_2} u(c_1,l_1) + \beta u(c_2,l_2) + \lambda V'(ab', ed', x_1') \tag{3.31}$$

subject to:

$$(1+r)^2 x_1 + (1+r)h_1 w_1(ab, ed) + h_2 w_2(ab, ed) = (1+r)c_1 + c_2 + x_1' + p_k \mathbf{1}_{[ed'=k]} \tag{3.32}$$

$$ab' = \alpha_0 + \alpha_1 ti_1 + \alpha_2 ti_2 \tag{3.33}$$

$$T = \theta ti_1 + h_1 + l_1 \tag{3.34}$$

$$T = \theta ti_2 + h_2 + l_2 \tag{3.35}$$

where (3.32) describes the monetary budget constraint over 2 periods, (3.33) shows the human capital production function over two periods where α_1, α_2 are the productivity of time investments for final ability. (3.34) is the time constraint in period 1 and (3.35) is the time constraint in period 2.

We can now rewrite this and derive optimality conditions:

$$V_{ed'=k}(S) = \max_{c_1,l_1,ti_1,x'_1,c_2,l_2,ti_2} u(c_1,l_1) + \beta u(c_2,l_2) + \lambda V'(ab',ed',x'_1)$$
$$+\mu[(1+r)^2 x_1 + (1+r)h_1 w_1(ab,ed) + h_2 w_2(ab,ed) - (1+r)c_1 - c_2 - x'_1 - p_k \mathbf{1}_{[ed'=k]}]$$
$$+\kappa(\alpha_0 + \alpha_1 ti_1 + \alpha_2 ti_2 - ab')$$
$$+\zeta_1(\theta ti_1 + h_1 + l_1 - T)$$
$$+\zeta_2(\theta ti_2 + h_2 + l_2 - T)$$

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

$$\kappa + \lambda \mu'[(1+r)\frac{\partial w'_1(ed'=k)}{\partial ab'} h'_1 + \frac{\partial w'_2(ed'=k)}{\partial ab'} h'_2] = 0$$

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:$-\mu + \lambda \frac{\partial V'}{\partial x'_1} = 0$

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

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

$$w_1(ab,ed)\theta = \alpha_1 [\frac{1}{(1+r)^2} \frac{\partial w'_{1'}(ed')}{\partial ab'} h'_{1'} + \frac{1}{(1+r)^3} \frac{\partial w_{2'}(ed')}{\partial ab'} h'_{2'}]$$

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 which is the productivity of an hour of time, multiplied by the resulting marginal increase in income over the lifecycle to the child. For completeness, we can describe the optimal education decision rule as follows:

$$ed' = \begin{cases} D & \text{if } V_{ed'=D} > \max\{V_{ed'=HS}, V_{ed'=C}\} \\ HS & \text{if } V_{ed'=HS} > \max\{V_{ed'=D}, V_{ed'=C}\} \\ C & \text{otherwise} \end{cases}$$

Because these educational choices are those that maximize the value function, we take these educational choices as given when solving for the first order conditions with respect to the other variables.

Chapter 4

What is the Source of the Health Gradient? The Case of Obesity

4.1 Introduction

It is a well-established fact that health and childhood socio-economic status are closely related (Almond and Currie (2011), Almond et al. (2018), Currie and Stabile (2003), Case et al. (2002), Case et al. (2005)). Yet the mechanisms which give rise to this relationship over the lifecycle are not well-understood. Children from richer families tend to have parents that are healthier and more productive in the labour market. Although these traits could be inherited directly, differences in behaviours may also matter. Children from richer families receive more health investments during childhood, leading them to start adulthood in better health and with higher labour market productivity. As adults, they also invest more in their own health. This paper disentangles and quantifies these channels to evaluate the degree to which policy can affect health inequality. The main novelty and contribution is that I model health and health investments during both childhood *and* adulthood: to the best of my knowledge this is the first estimated model of lifecycle health that includes health investments from conception to death. This turns out to be important – differences in parental investments matter more for the health gradient than differences in individual's own investments.

The model I develop and estimate is a dynastic model of endogenous health formation from the in-utero period until death. During childhood, parents make nutritional investments that can affect both health and later labour market productivity of their child and are costly in terms of money. During adulthood, individuals continue to make nutritional investments to increase their labour market productivity and to reduce their mortality risk and sick time. The model is estimated on data from a cohort born in Great Britain in 1970. This data set is uniquely suitable to study health development over the lifecycle, as it contains 1) repeated measures of BMI, which is the health measure I focus on, and 2) measures of health investments during childhood and throughout adulthood. I focus on BMI as the main measure of health for several reasons. First, obesity has become one of the most pressing health issues of recent decades. It is the leading cause of preventable death in the UK (ahead of smoking), and most recently received a lot of attention as one of the major comorbidities affecting Covid 19 survival (Ho et al. (2021)). Second, whilst there is a strong genetic basis to BMI, it can be affected by individuals' health behaviours, making it an interesting case to study endogenous health formation over the lifecycle (Biroli 2015). Third, it is a measure of health for which the gradient in childhood socioeconomic status is particularly prevalent.

Of those who grew up in the top quartile of the income distribution, 25% are obese by age 47, whereas of those who grew up in the bottom quartile, 35% end up obese. I find that 34% of this *obesity gradient* arises due to differences in health investments over the lifecycle, whilst the rest can be explained by differences in inherited health and labour market productivity. This suggests that there may be considerable scope for policy intervention. Differentiating between parental and own investments, I find that differences in health investments made by one's parents during childhood contribute more to the obesity gradient than differences made in adulthood by the individual herself. This is mainly because parental investments lead to better initial conditions, and higher earnings of children, which in turn lead to higher investments by the adult individual. However, parents strongly discount the benefits of investments in their children and face borrowing constraints. This means that for some children, government investments in their health might be welfare improving. I thus evaluate a counterfactual policy experiment in which the government directly invests in the health of low income children. The policy reduces obesity rates of those eligible by 2.1 percentage points by age 47. I find that individuals have a willingness to pay for this intervention which exceeds its cost by a factor of 1.84.

The rest of the paper proceeds as follows. The next section discusses related literature and the contribution of the paper. Section 3 presents the data, followed by descriptive facts in Section 4. Section 5 explains the model and Section 6 provides details on the estimation of the model. Section 7 presents model results, Section 8 presents the decomposition of the health gradient. Section 9 shows a policy counterfactual and the final Section concludes.

4.2 Related Literature

This paper relates to several different strands of literature. First, it builds on a large empirical literature which studies the long-run effect of childhood circumstances on health. Papers exploiting quasi-random variation have shown that investments and shocks during the pregnancy period can affect both health and labour market productivity in the long-run (e.g. Almond and Mazumder 2011, Scholte et al. 2015, see Almond and Currie 2011 for a summary). Beyond pregnancy, several papers have considered the childhood period, showing that e.g. access to the social safety net can improve outcomes for children in the long-run (see e.g. Almond et al. 2018 for a detailed summary of this literature). Whilst most of this literature carefully establishes causality, it does not model explicitly how early life shocks or investments generate the observed outcomes, and to what extent subsequent behaviours – either by the parent or the individual herself – matter.

Second, there is a growing literature which uses dynamic lifecycle models to study the causes and consequences of health. Whilst some of these papers evaluate the effect of exogenous health on labour supply, savings, and retirement (French 2005, De Nardi et al. 2010, De Nardi et al. 2017), others consider health to be endogenous by modelling health investments (Grossman 1972, Galama and Van Kippersluis 2019, Ozkan 2014, Scholz and Seshadri 2013, Margaris and Wallenius 2020). Most of this literature starts modelling health only from adulthood. Hence, the importance of the early part of the lifecycle is abstracted from, and the effect of early childhood policies on the health gradient cannot be evaluated. Two exceptions here are Kulikova (2015) and Dalgaard et al. (2021). Kulikova presents a model of parental investments in health and cognition during childhood, but she abstracts from health investment decisions in adulthood. Dalgaard et al. present a theoretical model of a health process which can capture long-run effects of early life shocks, but do not apply it to data.

Third, this paper adds to a growing literature that incorporates human capital formation during childhood into lifecycle models of parental decision-making (Caucutt and Lochner 2020, Daruich 2018, Gayle et al. 2018, Bolt et al. 2021). Most of these papers consider cognitive and non-cognitive skill formation, and assume that this process is completed at the end of childhood, rather than allowing for further investments during adulthood. Individuals' own actions during adulthood, however, might be crucial in amplifying or undoing parental choices, which is why modelling investments over the whole lifecycle may lead to different policy conclusions. A notable exception here is Lee and Seshadri (2019) who combine a model of parental investments in cognition with a Ben-Porath human capital investment model in adulthood. My model also allows for continued investments during adulthood, but focuses on health. Given that the development of health over the lifecycle may follow a very different path compared to skill formation, studying health formation on its own, rather than extrapolating from skill formation models is important.

Lastly, this paper relates to a growing literature in economics on BMI, nutrition, and obesity (for a summary, see e.g. Cawley 2015). In terms of modelling individual choices, there has been a focus on studying consumer behaviour over healthy and unhealthy foods (Dubois et al. 2018, Barahona et al. 2020), but so far the direct link of those choices to body mass and obesity has rarely been made (an exception is Margaris 2017). However, with obesity becoming an increasingly pressing issue for governments around the world, it is crucial to understand to what extent individual decision-making can affect BMI and whether policies can be effective in reducing obesity.

Given the previous literature, my contribution is twofold. First, I develop a framework that coherently models health and health investments over the whole lifecycle. The key novelties here are that 1) I allow for direct long-run effects of in-utero investments and 2) I allow for both parental health investments during childhood as well as continued own health investments during adulthood. Doing so means I can capture long-run effects of early life circumstances via various channels - directly through permanent changes to an individual's health, and indirectly through accumulation of a health stock and continued self-investments during adulthood. This also allows me to evaluate the relative contribution of parental investments versus own investments during adulthood.

Secondly, I estimate the model on longitudinal cohort data. To the best of my knowledge, no paper has used micro-data to estimate endogenous health development from the in-utero period up until middle-age. The use of data from a single cohort allows me to carefully take into account heterogeneity in the health process, which is particularly important for evaluating what policy can achieve to reduce the obesity gradient.

I will now turn to describing the data set used in this study.

4.3 Data

The main data set used for this paper is the 1970 British Cohort Study (BCS70) which started with all children born in one particular week in April 1970 in Great Britain.[1] Cohort members were initially surveyed shortly after birth, and subsequently at ages 22 months, (5), 10, 16, 26, 30, 34, 38, 42, and 46/47.[2] Whilst several other surveys contain information on health and socio-economics status over the lifecycle, this data set is unique in that it also contains information on health

[1]The BCS70 is provided by the Centre for Longitudinal Studies (2018) at the Institute of Education, University College London.

[2]Age 5 contains no information on BMI and is only used for retrospective information on breastfeeding behaviour.

investments from the antenatal period up until middle-age.

I use information on cohort members' BMI which was measured regularly from 22 months onwards, as well as cohort members' mother's BMI which was reported in the age 10 wave. Parents' income was reported twice, when the cohort member was age 10 and 16. Individual's adult earnings and hours are also reported in every wave from age 26 onwards. Information on assets of the cohort member are available at age 42. In terms of health investments, mothers were asked several questions on antenatal care during the first wave of the survey. In subsequent waves, various measures of nutritional intake, such as food frequency questionnaires (age 10, 30, 42) or diet diaries (age 16, 47), were taken.

As I focus on mothers as the main decision-makers in the model, for the sample I select only female cohort members. I drop non-white individuals and those with congenital disabilities, as well as those who died before age 42.[3] My main sample consists of 6,187 individuals.

In addition to the BCS70, I use two more data sets for supplementary purposes: the English Longitudinal Study of Ageing (ELSA) and the Wealth and Assets Survey (WAS). More details on these data sets can be found in the appendix 4.E.

Health: BMI The measure of health that I use throughout this paper is the body-mass index (BMI) which is calculated as (weight in kg)/(height in m^2). For adults, a BMI of above 25 is considered overweight, and a BMI above 30 is considered obese. For children, obesity is defined as being above the 95th percentile of the WHO BMI for age distribution. Whilst the BMI is not a direct measure of body fat, and hence may lead to misclassification of some individuals (e.g. athletes), it is generally considered an important measure of health and correlates strongly with co-morbidities such as cardiovascular diseases, diabetes and respiratory problems (Apovian 2016, Cawley 2015). Given the lack of underweight individuals in my data set, in this paper an increase in BMI means worse health, whereas a lower BMI means better health.

[3]Non-white individuals are dropped as previous literature has found that their BMI processes can differ substantially from Caucasians, and the non-white sample is too small to conduct the analysis by race.

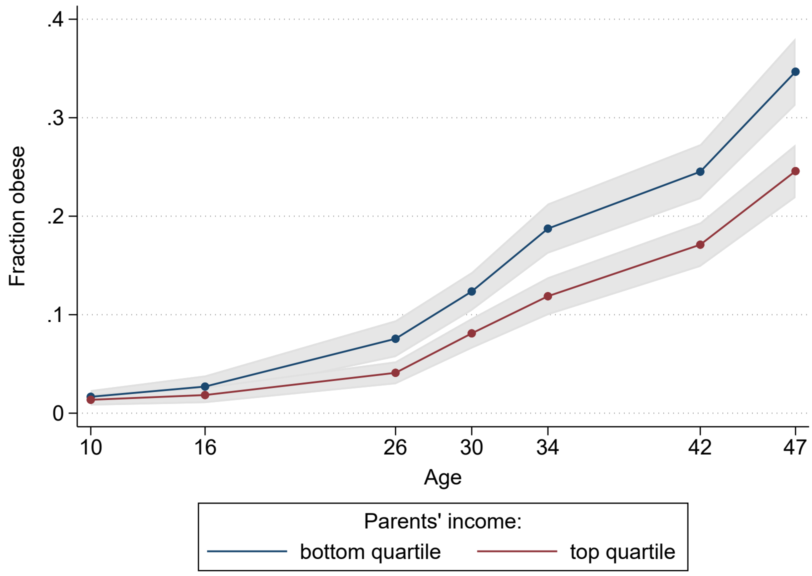
Health Investments: Modified Healthy Eating Index At ages 10, 16, 30, 42, and 47, the BCS70 includes information on individual's diets, either in the form of a food frequency questionnaire (FFQ, ages 10,16,30,42), or a diet diary (16, 47). The FFQ asks individuals how often they eat certain types of food, e.g. "How often do you eat fruit? Rarely, once a month, once a week, two to three times a week, four to five times a week, every day". The diet diary asks individuals to record everything they ate over a period of two (four) days at age 47 (16).

The Healthy Eating Index (HEI) reflects the quality of an individual's diet by measuring how much the diet adheres to Dietary Guidelines as set out by the USDA.[4] Depending on the food group, points are given for adequacy (e.g. eating enough fruit), or for moderation (e.g. not eating too much fat). The original Healthy Eating Index requires a high level of detail to be constructed, e.g. information on cups of fruit per 1,000 kcal. Whilst this level of detail is available at age 47, the information in older waves of the survey is less granular. For the older waves, I thus develop a modified version of the HEI similar to Masip et al. (2019), in which I follow the scoring of the HEI as closely as possible given the ordinal responses of the FFQs. See Appendix 4.C.1 for further details on the original HEI and the construction of the modified index. The final HEI scores for each wave are normalized to have mean zero and unit variance.

Health Investments: Perinatal investments For the period between pregnancy and age 2, no food frequency questionnaires or diet diaries are available. Instead, the cohort member's mothers were asked several questions about prenatal care and breastfeeding. I use measures on the attendance of mothercraft and labour preparation classes, week of initiation of prenatal care, alcohol drinking during pregnancy, and breastfeeding duration to construct a score of perinatal care. The construction of this score follows Heckman et al. (2013) and gives measures with little measurement error more weight than those measures with a lot of measurement error. See Appendix 4.C.3 for more details.

[4] Whilst there are many alternative ways of measuring diets, I choose this method as it 1) is most applicable to the type of data that I have, 2) has been found to have less measurement error than e.g. total calories consumed, 3) is a scientific measure of diet quality as opposed to diet types (which is what factor methods produce. For more details, see appendix 4.C.1).

Figure 4.4.1: Obesity over the lifecycle by parents' income quartile



Note: Graph shows the fraction of individuals obese at different ages conditional on the quartile of family income at age 10. Data from BCS70, females only. 90% Confidence Intervals are in gray.

4.4 The Obesity Gradient and Related Facts

In this section, I first show evidence on the relationship between family income in childhood and lifecycle obesity. I then present four related facts that help us understand what might be driving these differences in obesity rates, and that my model builds on. For the descriptives in this section, the family income measure I use is net family income measured when the cohort member was age 10.

Obesity Gradient Cohort members who grew up in poorer families are more likely to be obese. Figure 4.4.1 plots the share of obese individuals at each age for the top and bottom family income quartile. During childhood there are small, but insignificant differences by parental income group, which become large and significant during adulthood. By age 47, the difference in obesity rates for those coming from rich versus poor families is 10 percentage points. This difference in obesity rates at age 47 by parental income during childhood is what I call the *obesity gradient*. The benefit of considering the relationship between an individual's health and their parents' income rather than their own income is that it mitigates issues of simultaneity.

While a cohort member's BMI might affect her own income, it is less likely that it affected her parents' income during childhood.

Fact 1: Poorer parents are more likely to be obese. The first row of Table 4.4.1 shows the fraction of mothers that are obese by family income quartile during childhood. Mothers at the bottom quartile are almost twice as likely to be obese than those from the top quartile. One interpretation of this is that poor parents might directly transmit their poor health to their children, thus leading to the obesity gradient we observe.

Table 4.4.1: Facts related to the obesity gradient

			Parents' Income Quartile:			
	Mean	SD	Bottom	Middle	Top	F-test
Parental Health						
Obese mum	0.06	0.24	0.07	0.06	0.04	0.00
Childhood Investments						
Perinatal Investments	0.00	1.00	-0.21	-0.01	0.20	0.00
Healthy Eating Index 10	0.00	1.00	-0.13	-0.07	0.24	0.00
Healthy Eating Index 16	-0.00	1.00	-0.24	-0.02	0.22	0.00
Wages and Cognition						
Maths 16	0.00	1.00	-0.22	-0.01	0.31	0.00
Reading 16	0.00	1.00	-0.19	0.00	0.14	0.00
Wage 26	7.76	2.94	7.09	7.59	8.58	0.00
Adulthood Investments						
Healthy Eating Index 30	0.00	1.00	-0.14	-0.02	0.14	0.00
Healthy Eating Index 42	0.00	1.00	-0.11	-0.04	0.17	0.00
Healthy Eating Index 47	0.00	1.00	-0.12	-0.02	0.16	0.00

Note: This table presents descriptive statistics and means by parental income quartile for variables related to the four facts in the main text. Bottom refers to quartile 1 , middle to quartiles 2& 3, and top to quartile 4 of parental income at age 10. The final column reports p-values from the F-tests testing the null hypothesis of equality of means across the bottom and top quartile. The perinatal investment score, healthy eating indices, and reading and maths scores at 16 are normalized to have mean zero, variance 1. Hourly wages at 26 are in 2021 prices.

Fact 2: Poorer parents invest less in child health. Poorer parents invest less in child health - both in the perinatal period and in later childhood periods.

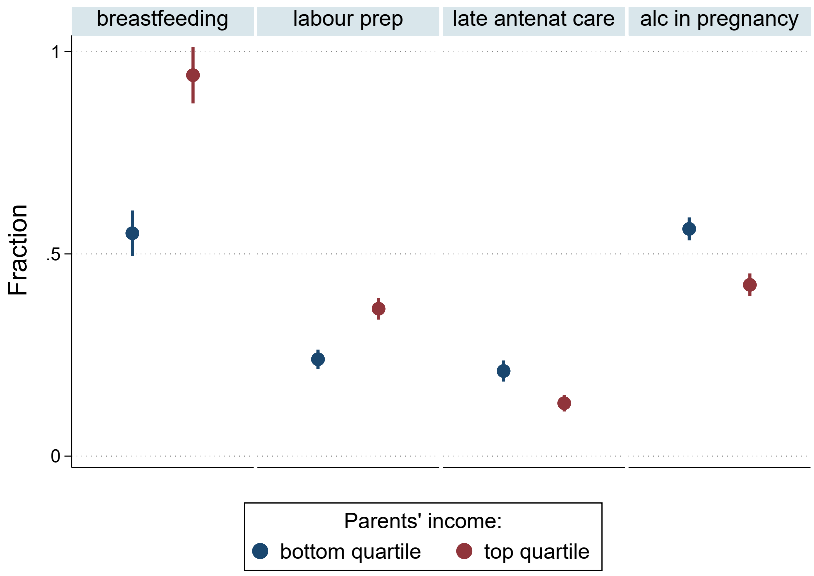
Figure 4.4.2 shows the gradients for perinatal investments. Panel a) shows differences in some of the measures of perinatal investments: poor mothers are less likely to breastfeed or to attend labour preparation classes, and are more likely to initiate antenatal care late and to drink during pregnancy. In panel b), I plot the distribution of the perinatal care index, described in section 4.3, which reflects the same pattern. Children from low income families are much more likely to have perinatal investments below the mean. Perinatal investments might be a particularly important factor in the development of the obesity gradient, as the first 1000 days of life (the period between conception to age 2) have been found to affect long-run health and skill development. For example, nutritional deprivation in-utero and under/overfeeding immediately after birth can alter an individual's metabolism (see e.g. Patel and Srinivasan 2011, Zhu et al. 2019). Moreover, nutrition in the perinatal period can affect the formation of neural connections, thus causing long-run changes in cognitive development (Cusick and Georgieff 2016).

Figure 4.4.3 shows the gradients in diet quality by parental income at age 16. Poor children eat salad, fruit, and wholegrains fewer times a week than their richer peers; they eat ice-cream equally often and chips much more frequently. The bottom panel shows the distribution of the Healthy Eating Index once I combine and score all measures. On average, the difference in diet quality between the top and bottom income quartile is 0.46 of a standard deviation at 16. The second panel of Table 4.4.1 also includes summary statistics and income gradients for the Healthy Eating Index at age 10.

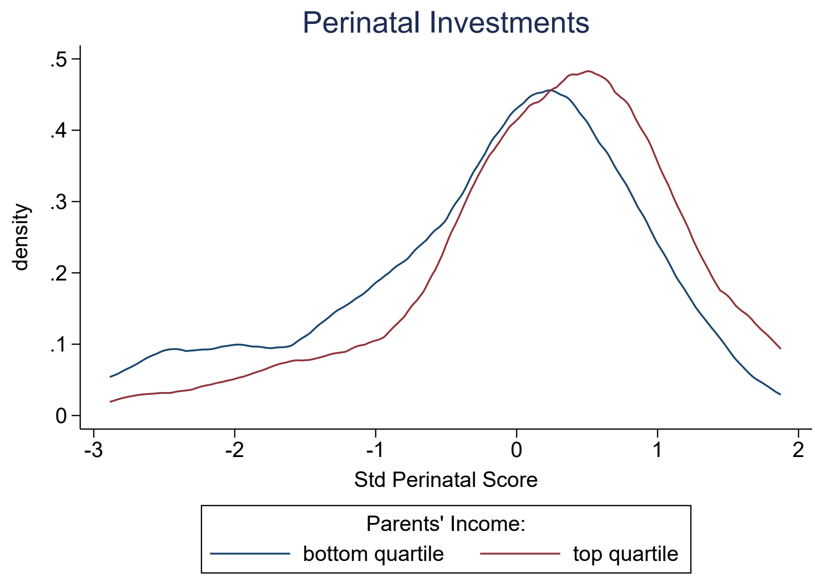
Nutritional investments after the perinatal period may be relevant for the obesity gradient for two reasons. First, they can affect an individual's childhood BMI which could then persist into adulthood. Second, they may affect an individual's brain development (Bellisle 2004, Bryan et al. 2004), and thus affect the resources available to the individual later in life, and hence potential later self-investments.

Fact 3: Individuals from poorer families have worse cognition and lower earnings at the beginning of adulthood. The third panel of Table 4.4.1 shows differences in test scores at age 16, as well as income at the beginning of adulthood. We can see that individuals from poor families have math (reading) scores that are 0.53 (0.33) of a standard deviation lower than those from rich families. We also observe significant wage differences at age 26. Wages for those coming from the top income quartile are on average £1.49 higher than wages for those coming from the bottom

Figure 4.4.2: Differences in perinatal investments



(a) Perinatal Investments: Raw Data



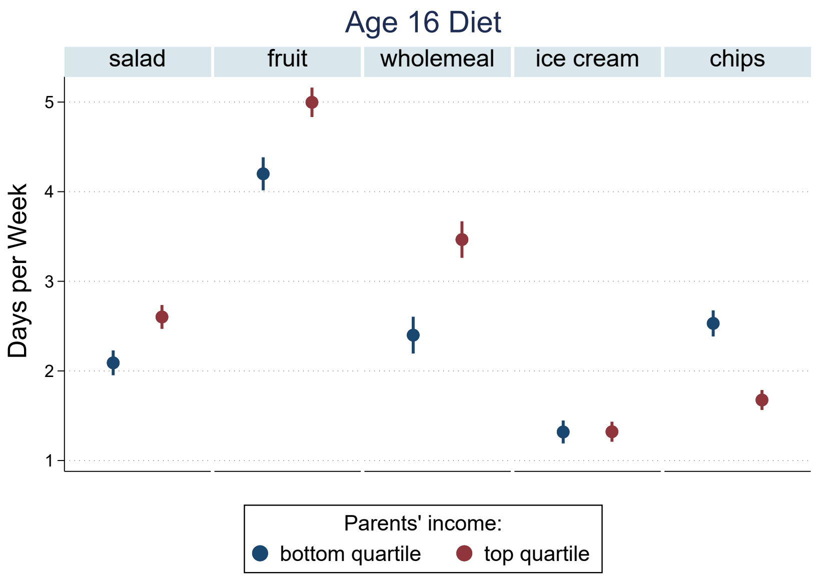
(b) Perinatal Investment Index

Note: Top panel shows fraction of mothers who breastfed, took part in labour preparation classes, had a late first antenatal care visit, and drank alcohol during pregnancy. Bars indicate 95% confidence intervals. Bottom panel shows kernel density plots of the predicted factor scores for perinatal investments.

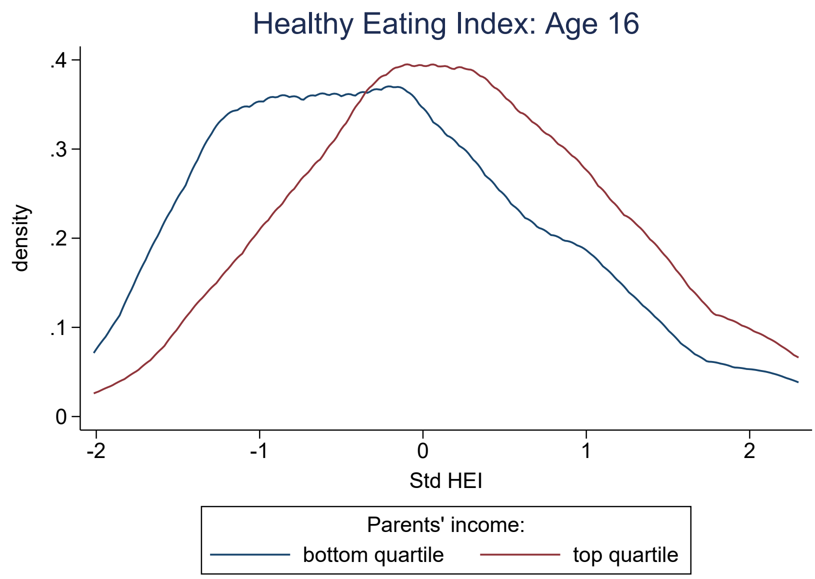
income quartile.

This fact is relevant for the obesity gradient as a healthy diet is more costly than an unhealthy diet. For example, Morris et al. (2014) show that the healthiest type

Figure 4.4.3: Diet by parental income at age 16



(a) Diet at age 16: Raw Data

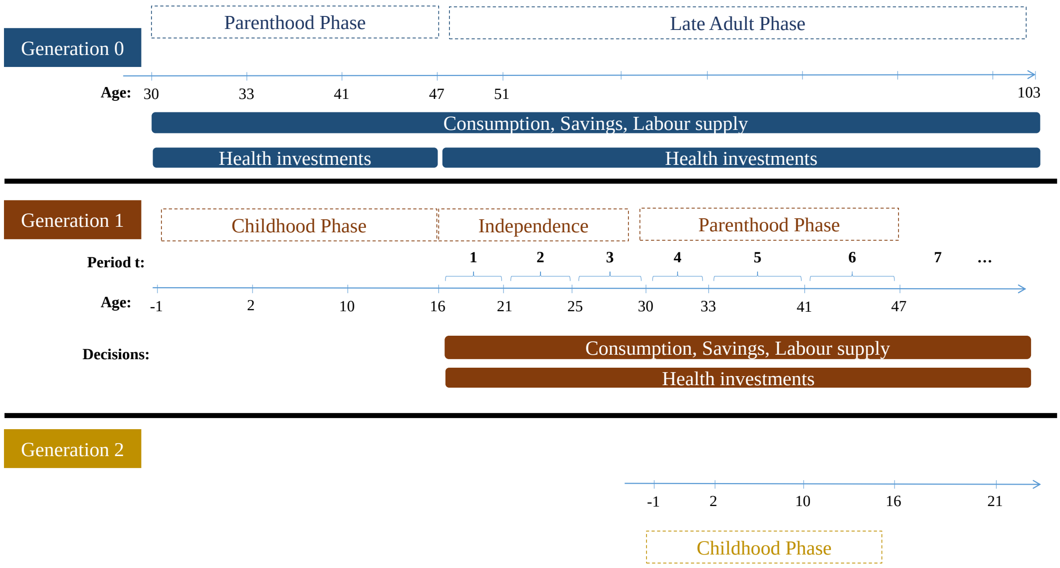


(b) Diet at age 16: Healthy Eating Index

Note: Top panel shows average days per week at age 16 that the cohort member reports to eat: salad, fruit, wholemeal foods, ice cream, and chips. Bars indicate 95% confidence intervals. Bottom panel shows kernel density plots of the Healthy Eating Index at age 16.

of diet on average costs £3.31 more per day than the unhealthiest diet (in £2021). Thus, another mechanism that potentially gives rise to the obesity gradient is that individuals from poor families continue to have less resources available to them as adults, thus preventing them from investing in their health.

Figure 4.5.1: Illustration of stages of the model



Fact 4: Individuals from poorer families invest less as adults. Finally, individuals from poorer families continue to have worse diets as adults. The difference in diet quality is 0.28 of a standard deviation and is constant across ages 30, 42, and 47, as can be seen in the bottom panel of Table 4.4.1.

Summary To summarize, we observe that individuals from poor families are disadvantaged on multiple dimensions. Their parents are unhealthier, they receive less investments during childhood, they have lower cognition and earnings, and continue to invest less in themselves as adults. We will now turn to modelling in detail how these factors might operate individually and jointly in generating the obesity gradient.

4.5 The Model

This is a dynastic model where each generation lives through four stages of the lifecycle: childhood, independence, parenthood, and late adulthood. There is endogenous BMI formation throughout the lifecycle and endogenous cognitive development during childhood. The lifecycle of an agent starts in-utero (age -1) and lasts at most until age 103. Model periods are irregular and correspond to the time interval between interviews of the BCS70. Figure 4.5.1 shows a model timeline illustrating the different period lengths and stages of the lifecycle. Before explaining the stages in more detail, I describe some elements that make the rest of the model clearer.

Preferences Agents in the model derive utility from consumption (c_t), leisure (l_t), BMI (B_t), and being alive.

$$u(c_t, l_t, B_t) = \frac{[(c_t/N_t)^{\nu} l_t^{1-\nu}]^{1-\sigma}}{1-\sigma} - \delta(B_t) + \zeta \tag{4.1}$$

where

$$\delta(B_t) = \delta_1 (B_t - \underline{B})^{\delta_2} \tag{4.2}$$

The utility function is CRRA with consumption weight ν, and curvature parameter σ. N_t is the consumption equivalence scale which accounts for household composition changing over time. Agents face a disutility of having a higher than healthy BMI, where δ_1 determines the level of disutility and δ_2 the curvature. As death is normalized to have value zero, a positive ζ means that agents value being alive (Hall and Jones, 2007).

In discounting their future utility, agents apply an annual discount factor (β). Model period length aligns with the differences in time between interviews and so the discount factor between model period varies. Thus β_t is the discount factor between t and t+1 and is equal to β^{τ} where τ is length of model period t.[5]

Agents are altruistic towards the next generation, with altruism parameter λ.

Demographics Agents in the model are female and exogenously become pregnant at age 30 with one female child. Agents have husbands that exogenously provide income and consume a share of consumption according to the consumption equivalence scale.

Credit Constraints Agents in this model cannot borrow. They can save in an asset with a save annual rate of return r. The return to saving assets a_t in period t will then be $(1+r)^{\tau} a_t$ in $t+1$, accounting for the years between periods, τ.

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

Heterogeneity Individuals are of a discrete type, $g = \{1,..,G\}$. The BMI, cognition, and wage process that individuals face differ by type g.

4.5.1 Independence

The Independence phase consists of periods $t = \{1,2,3\}$ which correspond to ages 16 to 21, 21 to 25, and 25 to 30. The agent's states are one's wage w_t, BMI (B_t), type (g), and cash on hand (m_t) - which consists of the stochastic income from a husband (y_t) and assets ($(1+r)^{\tau} a_t$).

The agent chooses consumption (c_t), savings (a_{t+1}), hours worked (n_t), and the quality of diet they consume (inv_t) at a price (p_t). Agents face two budget constraints: First, there is a monetary budget constraint. Second, there is a time budget constraint. Total time (T_t) can be spent on leisure (l_t) or labour supply (n_t); additionally, some time gets lost if the agent is in bad health, $\varphi(B_t)$.[6] BMI next period is a function of age, current period BMI, diet quality, the agent's type, and a shock. Wages are a function of age, BMI, labour supply, the agent's type, and a shock. Lastly, uncertainty is over BMI shocks, wage shocks, and income from the husband.

The agent solves the following decision problem:

$$V(\mathbb{S}_t) = \max_{c_t, inv_t, n_t} u(c_t, l_t, B_t) + \beta_t E[V(\mathbb{S}_{t+1})] \tag{4.3}$$

s.t.

Monetary Budget Constraint:	$a_{t+1} = m_t + w_t n_t - c_t - inv_t p_t$
Cash on Hand:	$m_t = (1+r)^{\tau} a_t + y_t$
Time Budget Constraint:	$T_t = l_t + n_t - \varphi(B_t)$
Adult Health Process:	$B_{t+1} = \mathscr{B}^A(t, inv_t, B_t, g, \varepsilon_{t,B})$
Adult Wage Process:	$w_t = \mathscr{W}(t, B_t, n_t, g, \varepsilon_{t,w})$

4.5.2 Parenthood

The parenthood phase consists of periods $t = \{4,5,6\}$ which correspond to ages 30 to 33, 33 to 41, 41 to 47. The corresponding child ages are -1 to 2 (perinatal period),

[6]The latter may capture, for example, reduced mobility or time spent attending to co-morbidities such as diabetes.

2 to 10 (early childhood), and 10 to 16 (late childhood).

Perinatal period, t=4: At the beginning of period $t = 4$, the agent exogenously becomes pregnant with one child. The agent's states are age, cash on hand, wage, BMI, type, $\mathbb{S}_4 = \{t, m_t, w_t, B_t, g\}$. The agent chooses consumption, savings, hours worked and investments. This choice is subject to the BMI and wage process which are the same as in the independence phase, as well as the two budget constraints.

Investments continue to affect the agent's BMI, but now also affect her child. In particular, during the perinatal period, the type of the child (g^C) is determined by BMI and wage of the agent, investments, and a shock. The type of the child depends on the agent's BMI and wage to allow for potential direct transmission of health and labour market productivity from parent to child. The type formation process depends on parental investments to model potentially permanent effects of perinatal investments on health and cognition.

Lastly, uncertainty is over agent's BMI, wage, husband's income, the type of the child (g^c), initial BMI (B_5^c) of the child, and initial cognition of the child (C_5^c).

Early childhood, late childhood, t=5, 6: The type of the child is now fixed. Agents can, however, continue to affect their child's outcomes by choosing diet quality, which affect the child's BMI and cognition. Diet quality also continues to affect the agent's own BMI.

The agent's states are cash on hand, agent's wage, agent's BMI, agent's type, as well as child's BMI, child's cognition, and child's type: $\mathbb{S}_{t=5,6} = \{t, m_t, w_t, B_t, g; B_t^C, C_t^C, g^C\}$. The agent chooses consumption, savings, hours worked and expenditures for diet. The choices are subject to the BMI and wage process of the agent, the BMI and cognition process of the child, and the two budget constraints. The BMI process of the child depends on previous BMI and previous cognition (thus allowing for potential cross-productivity between cognition and BMI), investments, and a shock. The cognition process of the child is analogous, and thus also depends on diet quality. In doing so, the model can capture the effects of diet on brain development in children documented by Bellisle (2004), Bryan et al. (2004).

Agents are imperfectly altruistic: In the final period of parenthood $t = 6$, they incorporate the continuation value of their child into their problem, but discount it by altruism parameter λ.

Uncertainty is over agent's own BMI, agent's wage, husband's income, child's BMI (B_5^c), and child's cognition (C_5^c). In period 6, the agent also needs to form expectations over the initial wage of the child which is a stochastic function of the child's final cognition.

The decision problem that the agent faces in the parenthood period is the following:

$$V(\mathbb{S}_t) = \max_{c_t, inv_t, n_t} u(c_t, l_t, B_t) + \beta_t E[V(\mathbb{S}_{t+1})] + \mathbb{1}_{t=6} \lambda \beta_t E[V^c(\mathbb{S}_{t+1}^c)] \tag{4.4}$$

s.t.

Monetary Budget Constraint:	$a_{t+1} = m_t + w_t n_t - c_t - inv_t p_t$
Cash on Hand:	$m_t = (1+r)^\tau a_t + y_t$
Time Budget Constraint:	$T_t = l_t + n_t - \varphi(B_t) - \kappa_t$
Adult Health Process:	$B_{t+1} = \mathscr{B}^A(t, inv_t, B_t, g, \varepsilon_{t,B})$
Adult Wage Process:	$w_t = \mathscr{W}(t, B_t, n_t, g, \varepsilon_{t,w})$
Type formation t=4:	$Pr(g^C = k) = \Gamma(B_t, w_t, inv_t)$
Child Health Process t=5,6:	$B_{t+1} = \mathscr{B}_t^C(inv_t, B_t, C_t, g, \varepsilon_{t,B})$
Child Cognition Process t=5,6:	$C_{t+1} = \mathscr{C}_t(inv_t, B_t, C_t, g, \varepsilon_{t,C})$

4.5.3 Late Adulthood

Late adulthood begins once the child has left the household, in period $t = 7$ which corresponds to age 47. Periods from now on are regular and 4 years long. Survival becomes stochastic and depends on BMI; the agent can live at most until the end of period 20 (age 103), then she dies with certainty. The agent chooses consumption, savings, labour supply and diet quality which now only affect the agent's own BMI. The agent retires exogenously at $t = 11$ (age 63). From then on the household receives a fixed pension.

The state variables are age, cash on hand (m_t), agent's BMI (B_t), agent's type

(g). There is uncertainty over wages and husband's income (before retirement), BMI, and survival (s_t).

$$V(\mathbb{S}_t) = \max_{c_t, inv_t, inv_t^N, n_t} u(c_t, l_t, B_t) + s_t(B_t)\beta_t E[V^c(\mathbb{S}_{t+1})]$$

$$\mathbb{S} = \{t, m_t, B_t, w_t, g\}$$

s.t.

Monetary Budget Constraint:	$a_{t+1} = m_t + w_t n_t - c_t - inv_t p_t$
Cash on Hand:	$m_t = (1+r)^\tau a_t + y_t$
Time Budget Constraint:	$T_t = l_t + n_t - \varphi(B_t)$
Adult Health Process:	$B_{t+1} = \mathscr{B}^A(t, inv_t, B_t, g, \varepsilon_{t,B})$
Adult Wage Process:	$w_t = \mathscr{W}(t, B_t, n_t, g, \varepsilon_{t,w})$

4.6 Estimation

The estimation procedure consists of three main steps. In the first step, I estimate agents' types using k-means clustering. In the second step, I estimate the BMI, wage, and cognition processes conditional on type, as well as the type formation process, and survival. In the third step, I estimate preference parameters using my dynastic model and the Method of Simulated Moments. I will now turn to explaining each step in more detail.

4.6.1 First Step: Fixed Heterogeneity Types

From age 2 onwards, individuals are of a latent type $g = 1, .., G$. I estimate these heterogeneity types using k-means clustering following Bonhomme et al. (2017) (BLM), and Jolivet and Postel-Vinay (2020). This approach has two main advantages: First, compared to using e.g. an Arellano-Bond type estimator, BLM allows for non-additive and time-varying effects of heterogeneity. This is important especially for this paper, where heterogeneity is relevant over the whole lifecycle, and effects of heterogeneity on BMI might differ between childhood and adulthood. Second, compared to estimating heterogeneity types inside the model, BLM has the benefit of a) being computationally less intensive and b) splitting the state space into clusters that are actually empirically relevant, rather than imposing an arbitrary grid.

To implement this, I choose a set of relevant individual-specific moments $\mathbf{m}$ (more on this below). The objective is then to split the state-space into clusters such that individuals within a cluster are as similar as possible in terms of their moments. More formally, individuals are clustered, such that:

$$\min_{\{g_i\}\in\{1,\cdots,K\}^N} \sum_{i=1}^{N} \|\mathbf{m}_i - \overline{\mathbf{m}}(g_i)\|^2$$

where $\overline{\mathbf{m}}(g_i)$ is the mean of vector $\mathbf{m}$ in group g_i.

The moments used for clustering need to satisfy the injectivity condition which says that individuals of the same type would converge to having the same moments as $T \to \infty$. In my model, the types affect the BMI process, cognition process, and wage process. I thus use the following individual specific moments that are informative about the latent type of an individual:

1. Average BMI over the lifecycle
2. Average wage over the lifecycle
3. Average cognition at ages 2, 10, and 16
4. Average investments over the lifecycle

I standardize moments to have mean zero and variance one to use them for the k-means algorithm. I set the number of types to three, following the inspection of the gap statistic for the optimal number of clusters. Details can be found in appendix 4.A.1.

Table 4.6.1 provides summary statistics for each group. Type 1 agents on average earn high wages and have a low BMI. Type 2 agents have a low BMI, but earn low wages. Type 3 agents on average are in have a high BMI and earn low wages. We can also see that the investment pattern closely follows wages. Finally cognition and wages capture similar content: Type 1 agents who earn the highest wages are also the ones who have the highest cognition.

Table 4.6.1: Summary Statistics for each of the three k-means clustering groups

Type	avg BMI	avg wage	avg cognition	avg inv
1	22.06	14.44	0.85	1.94
2	21.71	8.78	-0.58	1.42
3	27.93	9.47	-0.36	1.56

Note: This table shows summary statistics for each of the three clustering groups resulting from k-means procedure: Average BMI over the lifecycle, average wage over the lifecycle, average cognition during childhood, and average daily investment expenditure in 2021 GBP (amount individuals pay for diet beyond subsistence).

4.6.2 Second Step: Externally Estimated Processes

4.6.2.1 Adulthood Health Process

Equation 4.5 shows the health production function I estimate.

$$\ln(B_{t+1}) = \eta_0 + \eta_1 t + \eta_2 t^2 + \eta_3 inv_t + \eta_4 \ln(B_t) + \eta_{5,k} \mathbb{1}_{[g=k]} + \varepsilon_{B,t+1} \tag{4.5}$$

$$\text{where } \varepsilon_{B,t+1} \sim \mathcal{N}(0, \sigma_B^2)$$

Log BMI is a linear function of a quadratic in age, diet quality, previous log BMI, and the type g of the agent. Here, $\eta_{5,k}$ denotes the coefficient on a dummy which takes value one if the agent is of type k.

As the intervals at which I observe BMI in adulthood are irregular, I estimate the health process using non-linear regression following Rosner and Munoz (1988). Table 4.6.2 shows the results. Confidence intervals are calculated by bootstrapping around the whole estimation procedure, i.e. over the estimation of the types, as well as the health process estimation. BMI is highly persistent; the persistency parameter is estimated to be 0.96. Investments have a negative and significant effect on log BMI. Being of type 3 strongly increases BMI. I also test interactions between investments and BMI following recent literature on heterogeneity in the productivity of health investments (Margaris and Wallenius 2020). However, these turn out to be insignificant (see bottom 2 rows), and are thus excluded from the model.

Table 4.6.2: Adulthood health process

log BMI_{t+1}		
inv_t	-0.0014	[-0.002;-0.001]
log BMI_t	0.9612	[0.958;0.965]
Type 2	-0.0007	[-0.002;0]
Type 3	0.0139	[0.013;0.015]
Interactions:		
Type 2 x inv_t	0.0003	[-0.001;0.001]
Type 3 x inv_t	-0.0004	[-0.002;0.001]

Note: GMM estimation results for annual health process in adulthood. Coefficients for the age-polynomial are not reported. inv_t is in pounds spent on diet quality per day. The bottom panel shows coefficients and confidence intervals for the interaction between types and investments. Confidence intervals are cluster bootstrapped with 250 repetitions, and are reported at the 90% level.

4.6.2.2 Adulthood: Wage Process

I estimate the following wage process:

$$\ln(\tilde{w}_t) = \ln(w_t) + \xi_t = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \gamma_3 n_t + \gamma_4 B_t + \gamma_5 B_t^2 + \gamma_{6,k} \mathbb{1}_{[g=k]} + \varepsilon_{w,t} + \xi_t \tag{4.6}$$

$$\text{where } \varepsilon_{w,t} = \rho \varepsilon_{w,t-1} + u_t \tag{4.7}$$

$$u_t \sim \mathcal{N}(0, \sigma_u^2)$$

$$\varepsilon_{w,1} = \alpha \ln C_1 + u_1 \tag{4.8}$$

$$\xi_t \text{ is iid measurement error in wages}$$

Wages (w_t) depend on age, hours worked (n_t), BMI (B_t), the type (g) and an AR1 persistent shock $\varepsilon_{w,t}$ that initially depends on cognition $\ln C_1$.[7] Wages are mismeasured with idiosyncratic measurement error ξ_t. $\gamma_{6,k}$ denotes the coefficient on a dummy which takes value one if the agent is of type k.

I estimate the wage equation parameters in two steps. In the first step I estimate the γ parameters, in the second step I estimate the parameters of the wage shock

[7] Kline and Tobias (2008) find the effects of BMI on wages are non-linear over the BMI distribution. I thus allow for a square term to enter here.

process.

Step 1 Bias from endogenous selection into work is a concern when estimating the wage process outside the model. Thus, the deterministic part of the wage equation is estimated using a Heckman selection correction where I use marital status and the number of children as instruments (Heckman 1976). Moreover, to reduce concerns about simultaneity of wages and BMI, the process is estimated using the lag of BMI rather than concurrent BMI. Results are presented in Table 4.6.3. Both type 2 and type 3 earn significantly less wages than type 1. I also find that there is a significant and sizeable part-time penalty. My estimate of the BMI penalty is somewhat smaller than the estimates reported in Cawley (2004). All else equal, the BMI penalty I estimate corresponds to a predicted wage at age 42 of £9.99 for a type 2 woman with a healthy BMI of 20 compared to a predicted wage of £9.74 for an obese type 2 woman with BMI 30. [8]

Table 4.6.3: Adulthood wage process: Deterministic part

log Wages		
part-time	-0.073	[-0.092;-0.054]
BMI	-0.017	[-0.035;0]
BMI^2	0.0003	[0;0.001]
Type 2	-0.455	[-0.484;-0.427]
Type 3	-0.387	[-0.433;-0.342]

Note: Estimation results for health process in equation (4.6). Coefficients for the age-polynomial are not reported. Confidence intervals are cluster bootstrapped with 250 repetitions, and are reported at the 90% level.

Step 2 Wage shocks follow the AR1 process described in (4.7) and (4.8). To estimate $\rho, Var(u), Var(\xi)$ and α, I first predict wages $\ln \hat{w}_t$ using our coefficients estimated in step 1. I then calculate the residuals: $\ln \tilde{w}_t - \ln \hat{w}_t = \varepsilon_{w,t} + \xi_t$. Finally I exploit variances, and covariances of the residuals over time to estimate the parameters of the shock process and how the initial shock depends on cognition. See appendix 4.B.4 for further details on the estimation procedure.

The parameters for the stochastic part of the wage process are presented in

[8] For this part, I impute BMI at age 38 using a weighted average of age 34 and age 42. The coefficients are not sensitive to this - dropping age 38 data overall yields very similar results.

Table 4.6.4. The wage shock is highly persistent with an annualized persistency parameter of 0.97. The coefficient on cognition implies that a one standard deviation increase in cognition at age 16 increases initial wages by 2.7 percent. This is smaller than similar estimates found in e.g. Bolt et al. (2021). This discrepancy is due to the inclusion of the heterogeneous types. For comparison, the right hand panel of Table 4.6.4 shows the wage process estimates when not accounting for heterogeneity. As expected, wage shocks are much more persistent if we do not allow for group fixed effects. Moreover, excluding group fixed effects overestimates the effect of cognition.

Table 4.6.4: Parameters for stochastic wage process and comparison with version without GFE

		with GFE	without GFE
AR1 persistency	ρ	0.965	0.993
Effect of cognition	α	0.027	0.166
SD of persistent wage shock	SD(u)	0.078	0.067
SD of measurement error in wages	SD(ξ)	0.212	0.226

Note: Estimation results for the stochastic part of the wage process described in equations (4.7) and (4.8). Estimates are obtained using a minimum distance estimator. The persistency parameter is converted to an annual frequency.

4.6.2.3 Childhood: Health and Cognition Production Function

I now turn to the production functions of childhood health and cognition, described in equations (4.9) and (4.10), respectively.

$$\ln(B^C_{t+1}) = \eta_{0,t} + \eta_{1,t} inv_t + \eta_{2,t}\ln(B^C_t) + \eta_{3,t}\ln(C^C_t) + \eta_{4,k,t}\mathbb{1}_{[g^C=k]} + \varepsilon^C_{B,t+1} \quad (4.9)$$

$$\ln(C^C_{t+1}) = \gamma_{0,g,t} + \gamma_{1,t} inv_t + \gamma_{2,t}\ln(C^C_t) + \gamma_{3,t}\ln(B^C_t) + \gamma_{4,k,t}\mathbb{1}_{[g^C=k]} + \varepsilon^C_{C,t+1} \quad (4.10)$$

$$\text{where } \varepsilon^C_{B,t} \sim \mathcal{N}(0, \sigma^2_{B^c,t}) \text{ and } \varepsilon^C_{C,t} \sim \mathcal{N}(0, \sigma^2_{C^c,t})$$

I estimate equations 4.9 and 4.10 for each of the two stages of childhood: early (ages 2-10) and late (ages 10-16). Log BMI is a linear function of investments, lagged log

BMI, lagged log cognition, type dummies and a normally distributed i.i.d. shock.[9] The production function of cognition is analogous to the one for BMI.

An issue when estimating these equations is that cognition is not observed directly, but only indirectly via noisy measures such as test scores at each age.[10] To estimate the production functions using the noisy measures, I start by assuming that these measures relate to an individual's cognition according to a log-linear measurement system frequently used in the literature (Cunha and Heckman 2008, Agostinelli and Wiswall 2016a, Attanasio et al. 2020):

$$Z_{m,i,t} = \lambda_{m,t} \ln C_{i,t} + \varepsilon_{m,i,t} \tag{4.11}$$

Here, $Z_{m,i,t}$ is a measure m of cognition (e.g. a reading score at age 16), where $m = \{1,..M\}$ for individual i at time t. $\lambda_{m,t}$ is a scaling parameter and $\varepsilon_{m,i,t}$ is idiosyncratic measurement error. To reduce clutter, I drop the superscript C to denote this as the child's cognition in this exposition. I normalize the variance of $\ln C_t$ to one and assume that measurement errors are uncorrelated over time, individuals, and across measures. Measurement error is also orthogonal to $\ln C_{i,t}$.

The estimation of the production function then follows Agostinelli and Wiswall (2016a) and takes the following steps:

I first estimate the parameter $\lambda_{m,t}$ from ratios of covariances of different measures. I then divide each measure by its scaling parameter, so we can rewrite equation 4.11:

$$\tilde{Z}_{m,i,t} = \frac{Z_{m,i,t}}{\lambda_{m,t}} = \ln C_{i,t} + \frac{\varepsilon_{m,i,t}}{\lambda_{m,t}} = \ln C_{i,t} + \tilde{\varepsilon}_{m,i,t} \tag{4.12}$$

As we can see from (4.12), each rescaled measure of cognition $\tilde{Z}_{m,i,t}$ is equivalent to log cognition plus idiosyncratic measurement error. This measurement error can bias our production function estimates. To avoid this bias, I use the different measures as instruments for each other in the production function. For example measure

[9] At age 16, I also include controls for dieting behaviour as Crawley and Summerbell (1997) find, using the BCS70 age 16 data, that including these controls is important when analyzing the relationship between diet and BMI for teenagers.

[10] I assume that BMI is measured perfectly, as it is measured by the health visitors, thus reporting error etc is a minor concern.

$\tilde{Z}_{m',i,t}$ can serve as an instrument for $\tilde{Z}_{m,i,t}$. The estimation is done using a GMM estimator. Again, to get confidence intervals, I bootstrap around the entire procedure, i.e. the estimation of the measurement parameters, the types, and the production functions. Further details on the assumptions about the measurement errors, and estimates of measurement system parameters are in appendix 4.B.3.

Table 4.6.5 provides results for the estimated production functions. Investments, both at age 10 and 16 reduce BMI, though the coefficient is estimated imprecisely at age 16. Investments also have a positive effect on cognition. A one pound increase in daily nutritional expenditure increases log cognition by 0.08 of a standard deviation at age 10 and 0.13 of a standard deviation at age 16. BMI is less persistent than in adulthood, and also less persistent than cognition. The persistency parameter for cognition has an annualized value of 0.88 between ages 2 to 10, as well as between ages 10 to 16 which is similar to estimates found in other papers (Attanasio et al. 2020, Cunha et al. 2010). There is no evidence of cross-productivity between BMI and cognition. Similar to adulthood, being of type 3 increases an individual's BMI, whereas being of type 2 or 3 reduces cognition.

Table 4.6.5: Estimates of cognition and health production function

	Age 10		Age 16	
ln BMI				
inv_t	-0.009	[-0.015;-0.002]	-0.008	[-0.023;0.007]
$\ln BMI_{t-1}$	0.170	[0.111;0.228]	0.418	[0.362;0.475]
C_{t-1}	-0.003	[-0.008;0.003]	0.003	[-0.005;0.01]
Type 2	-0.025	[-0.04;-0.011]	0.001	[-0.013;0.016]
Type 3	0.128	[0.111;0.145]	0.114	[0.096;0.132]
Cognition				
inv_t	0.078	[0.035;0.121]	0.132	[0.068;0.196]
$\ln BMI_{t-1}$	0.127	[-0.212;0.466]	-0.089	[-0.476;0.297]
C_{t-1}	0.358	[0.322;0.394]	0.460	[0.391;0.529]
Type 2	-0.795	[-0.884;-0.707]	-0.801	[-0.943;-0.66]
Type 3	-0.676	[-0.777;-0.575]	-0.758	[-0.923;-0.592]

Note: Estimation results for child health and child cognition production functions as outlined in equations (4.9) and (4.10). Log cognition is standardized to have variance 1. inv_t is in pounds spent on diet quality per day. Confidence intervals are at the 90% level and are calculated using a bootstrap with 250 repetitions.

4.6.2.4 Type Formation

The probability of a child being of type $g^C = k$, where $k = 1, ...G$, is a function of perinatal investments (inv_4) and BMI and family income. I estimate this using the following multinomial logit specification:

$$Pr(g^C = k) = \frac{\exp(x'\beta_k)}{\sum_{j=1}^{G} \exp(x'\beta_j)} \tag{4.13}$$

and

$$x'\beta_k = \beta_{0,k} + \beta_{1,k} inv_4 + \beta_{2,k} \ln B_4 + \beta_{3,k} \ln y_4 \tag{4.14}$$

Results are presented in Table 4.6.6. I find that perinatal investments significantly increase the probability of being in the high wage group, thus decreasing the probability of being in any of the other groups. The same holds for parental income. Maternal BMI especially matters in affecting the probability of being of type 3, which is the group most likely to be obese. This indicates that genetic transmission of obesity may matter strongly here. [11]

Table 4.6.6: Multinomial logit - Base group is Type 1

Type 2			Type 3		
inv_4	-0.270	[-0.384;-0.156]	inv_4	-0.197	[-0.337;-0.057]
$\ln BMI_4$	0.155	[-0.571;0.882]	$\ln BMI_4$	3.412	[2.747;4.076]
$\ln wage_4$	-1.071	[-1.313;-0.83]	$\ln wage_4$	-1.140	[-1.444;-0.836]

Note: This table shows estimates of the multinomial type formation equation as outlined in equation 4.13. The baseline type is Type 1. $inv_4, BMI_4, wage_4$ refer to parental investments, BMI and wages in the perinatal period (child age -1 to 2). The Confidence intervals are at the 90% level and are calculated using a bootstrap with 250 repetitions.

4.6.2.5 Other Processes

Initial distribution of child BMI and cognition A child's initial BMI and cognition at age 2 (agent period 5) depends on the child's type and an i.i.d. normally

[11] Note that due to data limitations I do not directly observe $\ln B_4$ and $\ln y_4$, i.e. maternal BMI and income at age 30. Instead, I observe parental BMI and income when the child is age 10. In the above regression, I thus adjust parental BMI and income to have the same mean and variance as observed for cohort members at age 30 which is the age at which individuals become pregnant in the model. This ensures that the coefficients have a magnitude that is consistent with the model. In future versions I intend to estimate this process inside the model to circumvent this issue.

distributed shock. Further details are in appendix 4.B.2.

BMI and Mortality In late adulthood, survival depends on the agent's BMI and age. To estimate this, I turn to the English Longitudinal Study of Ageing. I use the first available BMI information in ELSA (wave 2), and estimate a logit regression of one year survival on an age-polynomial and health, including various controls. Following Bhaskaran et al. (2018), I drop the first 4 years after BMI is measured to minimize the effects of reverse causality and focus on the effect of BMI on mortality for the region of BMI ≥ 25. I find that BMI has a strong negative effect on survival, with the magnitudes consistent with those from studies in the medical literature. Details can be found in Appendix 4.B.5.

Husband's income Husband's income, $y_{t,h}(g)$, is a function of age, the agent's type and an iid shock. Details on the estimation of the husband's income function and results are in appendix 4.B.6.

Externally Set Parameters

I set the following parameters of the model externally: The price for better diet quality is taken Morris et al. (2014) who find that the healthiest diet costs £3.31 more than the unhealthiest diet per day. To make this applicable for my data, I assume that a diet at the 95th percentile of the Healthy Eating Index costs £3.03, whereas diets at or below the bottom 5th percentile of the Healthy Eating Index costs zero. I interpolate the cost linearly in between these two percentiles.

The annual interest rate r is set to 5%. This is relatively high, but consistent with the high returns to housing wealth that this cohort attained (Jordà et al. 2019).

The time cost of children is estimated directly from the UK Time Use Survey. Details can be found in appendix 4.E.2.

4.6.3 Third Step Estimation - Method of Simulated Moments

In the third step, I estimate ten parameters using the method of simulated moments. I will first describe the parameters estimated and moments used, followed by a discussion of identification. Further computational details can be found in appendix 4.D.

Parameters I estimate the following ten parameters using the method of simulated moments: the discount factor β, the consumption weight ν, the scale and curvature of disutility of a high BMI (δ_1, δ_2), the scale and curvature of sick time (φ_0, φ_1), the curvature parameter of the utility function σ, the altruism parameter λ, the cost of perinatal investments p_4, and the additive utility parameter ζ.

Moments I match the following 36 moments for the BCS70 cohort: Mean hours worked at ages 26, 30, 34, 42, 47 by BMI category (20 moments); asset medians at ages 34, 42, 47 (3 moments); mean BMI at ages 16, 26, 30, 34, 42, 47 (6 moments), the distribution of health types (3 moments), the correlation between wealth and BMI (1 moment), the obesity rate and variance of BMI at 47 (2 moments), and the value of statistical life (1 moment).[12]

To match employment profiles of the model and data, one has to take into consideration that women in the BCS70 do not necessarily all become pregnant at age 30. To ensure consistency between the model assumption and the data, I estimate the hours profiles in the following way: I pool observations across ages in the BCS70 and estimate a logit regression of hours on an age polynomial, BMI, BMI squared, and the number of children. I then predict the hours worked for those *without* children at ages 26 and 47, and for those *with* children at ages 30, 34, 42. This ensures that the moments are based on similar household composition as in the data.

Financial wealth and housing assets are reported at age 42. For ages 34 and 47, only financial wealth is reported. I thus turn to the Wealth and Assets Survey to estimate asset moments for ages 34 and 47 for individuals born in the 1970s, and check consistency with the BCS70 using the available information. More details can be found in appendix 4.E.1.

The value of statistical life that I target is £5m. This is the central VSL found across several metanalayses by Banzhaf (2021).

[12] Note that I cannot exactly align the timing of the model with the data during childhood *and* adulthood because of the overlapping generations structure of this model. I thus primarily follow the structure of the data during childhood, and assume that age 33 in the model corresponds to age 34 in the data, and age 41 in the model corresponds to age 42 to in the data.

Identification All parameters of the model can affect all simulated moments as they jointly generate the behaviour of agents. However, some moments are more informative than others about particular parameter estimates, and I will provide some intuition in the following.

The discount factor β is mainly pinned down by asset moments. The more patient individuals are, the more willing they are to save. ν shifts the weight of consumption versus leisure in the utility function, and can thus be identified from labour supply moments. σ, the curvature of the utility function, is also commonly pinned down by asset and labour supply moments. In this model, it is also identified from the correlation between wealth and BMI. Following Hall and Jones' argument, σ is crucial, as it determines how quickly the marginal utility of consumption decreases for individuals. For rich individuals, gaining more consumption within one period may yield little extra utility, so they prefer to invest in extending their life-span.

The parameters determining the level of the disutility of a high BMI, δ_1 is pinned down by the average BMI, especially after the child has left the household. The curvature of the disutility of a high BMI, δ_2, is pinned down by the variance of BMI at 47 and the obesity rate at 47. If δ_2 is high, it indicates that individuals want to avoid having a very high BMI which compresses the variance of BMI. Moreover, for the same average level of BMI, we expect to observe a lower share of individuals that are obese.

Sick time parameters φ_0 and φ_1 are pinned down by differences in hours worked for different BMI groups. The altruism parameter λ is pinned down by the level of BMI of generation 2 during childhood. If parents are more altruistic, then they will invest more in their children. The cost of perinatal investments, p_4, is pinned down by the type distribution of generation 2. This is because perinatal investments strongly affect the type of a child, and so changes in the price of these investments will affect the resulting distribution of types.

Finally, the additive parameter ζ is pinned down by the value of statistical life (VSL). The value of statistical life is the willingness of an individual to pay to reduce their mortality risk, i.e. $VSL_t = \frac{\frac{\partial V_t}{\partial s_t}}{\frac{\partial V_t}{\partial w_t}}$ (Bommier et al. 2020). The numerator of this term is a function of ζ. Hence, a shift in ζ will lead to a change in the simulated VSL. Finally, how much individuals value their lives also affects how much they invest, and thus the simulated BMI moments.

4.7 Results

4.7.1 Model Fit and Parameter Estimates

Targeted Moments The model fits the data well. Figure 4.7.1 shows the model fit for BMI over the lifecycle, the type distribution, hours worked, and assets. The fit for BMI over the lifecycle and the distribution of types in generation 2 is good. The model underpredicts hours at age 30, and slightly underpredicts assets. It also generates a correlation between BMI and wealth that is slightly too low in magnitude (-0.13 in the Model vs -0.17 in the data), and a variance of BMI of 33.4 which is close to the data variance of 34.6. The obesity rate in the model is slightly too low at 0.27 versus 0.3 in the data. Finally, the model matches the targeted value of statistical life well (£4.98m in model vs £5m targeted).

Table 4.7.1 presents the parameter estimates resulting from the method of simulated moments. We find that the discount factor, consumption weight, and risk aversion parameters are within a range consistent with previous literature. The altruism parameter at 0.22 is slightly lower compared to other dynastic models where the estimates are usually on a range between 0.29-0.4 (Daruich (2018), Yum (2020)) . The curvature parameter for the disutility of BMI, δ_2, interestingly is close to one. Hence along the distribution of BMI, a one unit increase in BMI leads to a similar decrease in utility. This also means that a percentage increase in BMI leads to a stronger decrease in utility for those with a low BMI compared to those with a high BMI. Finally, the price of perinatal investments is high. p_4 implies that a standard deviation of investments in the perinatal period costs a family 4.18 times as much as a standard deviation of investments in a later period.

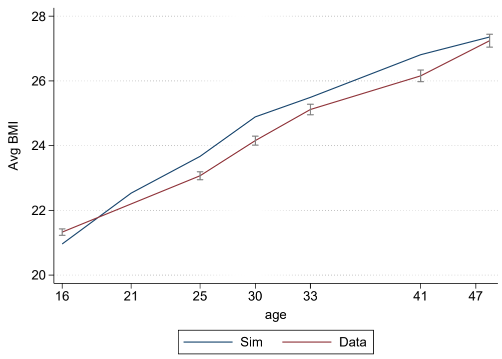
Untargeted Moments None of the moments in the model explicitly targets the relationship between income of the first generation and BMI or obesity in the second generation. We can see from Figure 4.7.2 that the model is able to generate this

Table 4.7.1: Results: Parameter Estimates

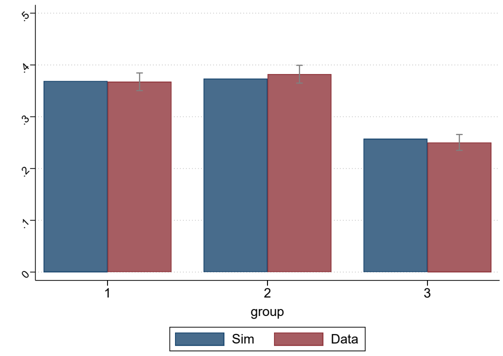
Parameter	Definition	Estimate	SE
β	Discount factor	0.984	(0.0005)
ν	Consumption weight	0.579	(0.0009)
δ_1,δ_2	Disutility of BMI	0.003,0.976	(0.0007),(0.1784)
φ_1,φ_2	Sick time parameters	0.100, 0.579	(0.0037), (0.0469)
σ	Risk aversion	4.372	(0.0815)
p_4	Price of perinatal inv	5.271	(1.1096)
λ	Altruism parameter	0.226	(0.0540)
ζ	Value of life parameter	0.120	(0.0580)

Note: Column 1 and 2 show parameters that are estimated and their respective definitions. Column 3 shows estimates resulting from the Method of Simulated Moments, Column 4 shows Standard Errors. Details on the estimation procedure and calculation of standard errors can be found in Appendix 4.D.

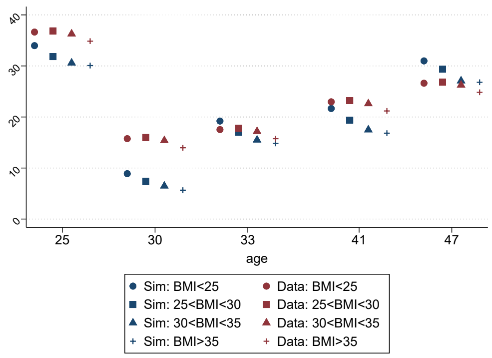
Figure 4.7.1: Model Fit: Targeted moments



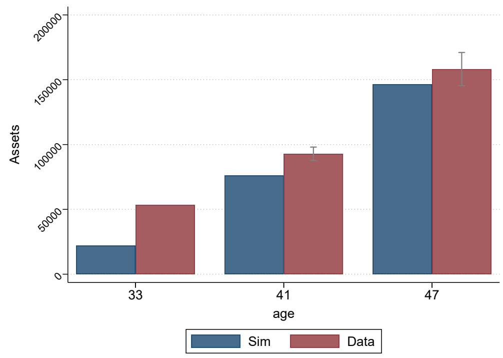
(a) BMI moments



(b) Type distribution



(c) Hours by BMI

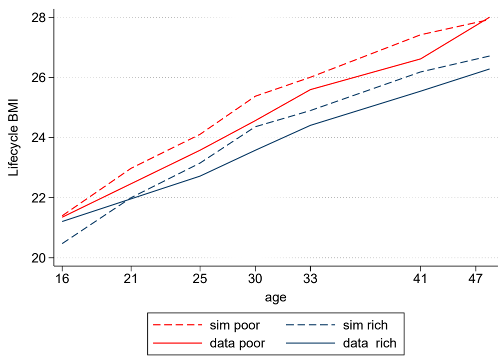


(d) Assets

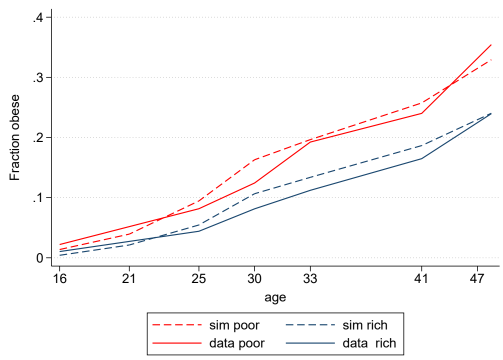
Note: Figure shows empirical moments and moments simulated using the model. The number of simulated individuals is 25,000. Grey bars show empirical 95% confidence intervals.

relationship over the lifecycle. This means that the model is able to endogenously generate the obesity gradient observed in the data.

Figure 4.7.2: Model Fit: Targeted moments



(a) BMI moments



(b) Fraction obese

Note: Figure shows empirical moments and moments simulated using the model. Solid line uses data from BCS70, dotted line uses data from simulation. Poor (rich) refers to the bottom (top) quartile of parents' income at age 10.

4.8 Decomposing the Obesity Gradient

The main focus here is the difference in obesity rates at age 47 between individuals coming from the bottom quartile (Q1) of the parental income distribution at age 10 versus the top (Q4), which is 8.4 percentage points in my simulation. I first investigate how much investments over the lifecycle matter for this gradient by evaluating how much it would shrink if all individuals exogenously received the same average investments as Q4 in every period, starting from pregnancy up until age 47. The remaining differences in Q1 and Q4 arise from differences in the health and productivity of the parents. Table 4.8.1 shows that the difference in obesity rates between individuals from rich and poor families shrinks by 34% once investments are equalized over the whole lifecycle.

I next compare how much of the gap is due to differences in parental behaviours versus differences in individuals' own behaviours. To look at the effect of differences in childhood investments alone, I re-simulate the model by giving individuals in the bottom quartile the same investments as the top but only from pregnancy up until age 16. This reduces the obesity gradient by 24%. I then consider what happens if investments were the same between those from the bottom vs top during adulthood. This reduces the obesity gap by 17%. It is notable that these two numbers add up to more than the 34% found when equalizing all behaviours. The reason for this is that increasing investments of Q1 during childhood leads to more self-investments during adulthood.

Given the importance of investments in the childhood period, I then consider *when* investments during childhood matter. Here, I equalize investments during the perinatal period, versus investments later in childhood. Results indicate a strong effect of the first investment period. Differences in the perinatal period contribute to 11% of the obesity gradient; differences in the later childhood periods contribute to 13% of the obesity gradient. This shows how formative the perinatal period is, despite it being short.

Finally, I investigate more closely *how* childhood investments decrease the obesity gap. Childhood investments can matter via several channels in this model. First, they can change the long-run type of an individual, thus leading to direct changes in the BMI process of an individual. Second, they can change the BMI of an individual at the start of their independent life. As we found that BMI is very persistent, differences in BMI at the beginning of adulthood can continue to lead to differences in BMI at age 47. Third, nutritional investments during childhood can also affect a child's cognition leading to a higher initial wage, which is also fairly persistent. These differences in wages can allow an individual to buy better nutrition in adulthood, thus leading to lower BMI.

To investigate the first channel, I give individuals in Q1 the same average investments as those in Q4 during childhood. However, when simulating their BMI, rather than using the health type that is predicted in this counterfactual, I use the health type predicted in the baseline case. Thus, the BMI production function they face is the same as in the baseline case, but all else is allowed to differ. This isolates the effect of parental investments on the obesity gradient that goes via direct long-run changes of the BMI production function. I find that this accounts for 7% of the obesity gradient.

To investigate the second case, I again simulate the model by giving all Q1 individuals the average investments of Q4 during all childhood periods. However, at age 16, rather than continuing with the resulting BMI, I give every individual their original BMI from the baseline case. This helps us isolate to what extent parental investments matter because they affect age 16 BMI. We find that 12% of the obesity gradient is a result of the persistency of differences in initial adulthood BMI.

Finally, we investigate to what extent the effect of nutrition on initial wages matters. For this, I again equalize investments during childhood between Q1 and Q4. However, at age 16, I now set cognition for each individual to be the same as in the baseline case. This helps us isolate the effect of investments via the initial wage shock. This only accounts for 0.7% of the obesity gradient. Thus, the effect of investments on BMI via initial wages is neglible.

Table 4.8.1: Decomposition of Obesity Gradient

Differences in investments	34%
Differences in own investments	17%
Differences in parental investments	24%
Perinatal	11%
Later childhood	13%
Long-run direct health of child (type)	7%
Initial health of child at 16 (health stock)	12%
Initial wage shock of child at 16	0.7%

Note: This table shows results from a decomposition of the obesity gradient. The first column describes the counterfactual scenario, the second column shows the associated decrease in the obesity gradient. More details can be found in the main text.

4.9 Policy Counterfactual: Improving the Diets of Low Income Children

In the previous section, I found that differences in childhood investments explain the largest share of the obesity gradient. Children, however, cannot invest in their own health, but instead rely on the choices their parents make which may not be optimal due to lifecycle and intergenerational borrowing constraints. Moreover, the low altruism parameter implies that parents strongly discount the benefits of such investments to the child. I thus implement a policy counterfactual in which the government increases investments for those from poor families to a level similar to those from rich families. I then evaluate to what extent this reduces the obesity gradient, whether this is welfare improving, and whether individuals' willingness to pay exceeds the cost of such a policy. Moreover, I analyze parental reactions to such a policy and compare its effects to a similarly sized monetary transfer to parents.

Implementation The government directly invests in the health of the children who are at the bottom quartile of the parental income distribution in each childhood stage. I calculate the 25th percentile of income, which I use as the eligibility cutoff, at each age from the baseline simulation. The amount that the government invests is equivalent to the difference in average investments at each age between those from the bottom and the top income quartile.

Parents know that this policy is in place and can react to it accordingly. In terms of the cost of the policy, I assume that the government faces the same prices as parents in each period, and that its investments are as effective as parental investments. The cost of the policy is on average £795 per year between the ages of -1 to 16. Note that the cost of this intervention is similar to some interventions already in place in the UK. For example, Gaheer and Paull (2016) find that the per child annual cost of providing pregnancy and baby health care services at children's centers costs £1,200 per year.

Welfare To fix ideas, recall that individuals start their own independent lifecycle at age 16. The maximized value of their problem in the baseline case is

$$V^{Base}(\mathbb{S}_1) = \max_{\boldsymbol{d}_{1..J}} u(c,l,B) + \sum_{t=2..J} \beta^t E[u(c,l,B)] \tag{4.15}$$

where for brevity, we denote S to be state variables in period 1 and $\boldsymbol{d}_{1..J}$ is the sequence of choices over the lifecycle. This is an additive version of the full recursive problem described in Section 4.5.

In the policy counterfactual, agents receive additional investments from the government if their parents are poor, but expect a fraction $(1-\Delta)$ of consumption to be subtracted in every period from age 33 onwards. I only introduce the tax from period 5 (age 33) onwards, as individuals in the model cannot borrow. Thus, at young ages, the willingness to pay is very low due to this constraint. As I am interested in the unconstrained willingness to pay, I calculate the WTP at ages when individuals are less constrained. [13]

$$V^{Policy}(\mathbb{S}_1) = \max_{\boldsymbol{d}_{1..J}} u(c(1-\Delta_t), l, B) + \sum_{t=2..J} \beta^t E[u(c(1-\Delta_t), l, B)] \qquad (4.16)$$

where

$$\Delta_t = \begin{cases} 0 & \text{if } t < 5 \\ \Delta & \text{if } \geq 5 \end{cases}$$

I then solve for the Δ_t such that $E[V^{Base}(\mathbb{S}_1)] = E[V^{Policy}(\mathbb{S}_1)]$. Note that I re-solve the model, which means that I allow for all kinds of behavioural effects that the reform might generate (e.g. crowding out of parental investments, changes in asset accumulation and labour supply due to the deduction of Δ etc.).

Results I find that the policy reduces obesity rates at age 47 by 2.1 percentage points for those eligible in all childhood periods. Table 4.9.1 shows that the reduction in obesity rates differs only slightly depending on the age at which an individual is eligible.

Table 4.9.1: Results: Policy counterfactual

	Effect		Selection		Crowding out:
	% point reduction in obesity at 47		% of families eligible		% eligible families reducing investments
	Anticip	Unanticip	Anticip	Unanticip	Anticip
All periods	2.1	2.3			
Perinatal	1.9	2.2	30%	25%	3%
Age 2-10	1.4	1.6	25%	25%	0.1%
Age 10-16	1.5	1.6	27%	25%	25%

Note: Anticip shows results for the case when individuals anticipate the policy change. Unanticip shows results for the case when individuals do NOT anticipate the policy change.

Turning to welfare, individuals are willing to give up 3.99 percent of their consumption from age 33 onwards to live under this policy rather than in the baseline case. This means that the net present value of consumption agents are willing to give

[13] Note that I use the net present value to evaluate the willingness to pay versus the costs, thus taking into account that the tax only gets collected from period 5 onwards.

up exceeds the net present value of the cost by a factor of 1.84. Hence, I conclude that individuals strongly value this intervention.

In the case where the policy is unanticipated by the parents, it leads to a 2.3 percentage point reduction in obesity which is slightly higher than in the anticipated case. There are two opposing forces at play here. First, anticipation of the policy leads low income parents to reduce their labour supply especially in the perinatal period, such that they fall below the eligibility threshold. In the perinatal period, 30% of individuals are now eligible for the policy, rather than the intended 25%.[14] Second, parental investments are crowded out, that is parents respond by reducing investments in their children. However, as most parents who are eligible for this policy do not invest in their child at all in the baseline, this reaction is small. Together, these behavioural forces have a negligible effect on the reduction of the obesity gradient.

Lastly, I consider an alternative policy in which the government directly transfers the amount it invests to parents in the bottom quartile, rather than investing directly. Interestingly, this leads to hardly any increases in investments during childhood. Only 10% of parents who receive the transfer increase investments in their children. Instead, parents increase their consumption and their savings. Thus, in order to ensure that children receive higher investments in health, direct government investments or transfers tied to nutritional expenditures such as food vouchers might be more effective than monetary transfers.

4.10 Conclusion

This paper studies the origins of the obesity gradient and the extent to which it can be affected by policy. Individuals coming from poor families are 10 percentage points more likely to be obese at age 47 than those coming from rich families. I find that 34% of this gradient stems from differences in behaviour between those coming from rich versus poor families. Over the lifecycle, differences in early

[14] In the calculation of cost to WTP, I account for this increase in eligibility. However, I cannot evaluate forgone labour income tax revenue of the parents for the government, as I do not explicitly model this. I leave this for further research.

childhood investments matter more than investments during adulthood . However, children cannot invest in their own health, and parents strongly discount benefits to their children in making health investment choices. Individuals thus highly value an intervention in which the government directly invests in the health of low income children.

This paper is an important step in connecting the dots between early childhood health and later life health. Whilst the focus here is on obesity, the model can be applied and extended to other dimensions of health, such as physical health more generally or even mental health. In terms of obesity specific extensions, including physical activity or incorporating habit formation may be fruitful avenues for future research.

Appendix

Appendix

4.A Further Details: K-means

4.A.1 Optimal number of clusters

To estimate heterogeneity groups using the k-means clustering algorithm, the researcher needs choose the number of clustering groups. This is unknown a-priori, but data driven methods can be used to determine it. A commonly used criterion to evaluate the appropriate number of clusters is the Elbow statistic. Recall that for a given number of clusters K, k-means minimizes the total within cluster variance:

$$\min_{\{g_i\}\in\{1,\cdots,K\}^N} \sum_{i=1}^{N} \|\mathbf{m}_i - \overline{\mathbf{m}}\,(g_i)\|^2 = SSE_K$$

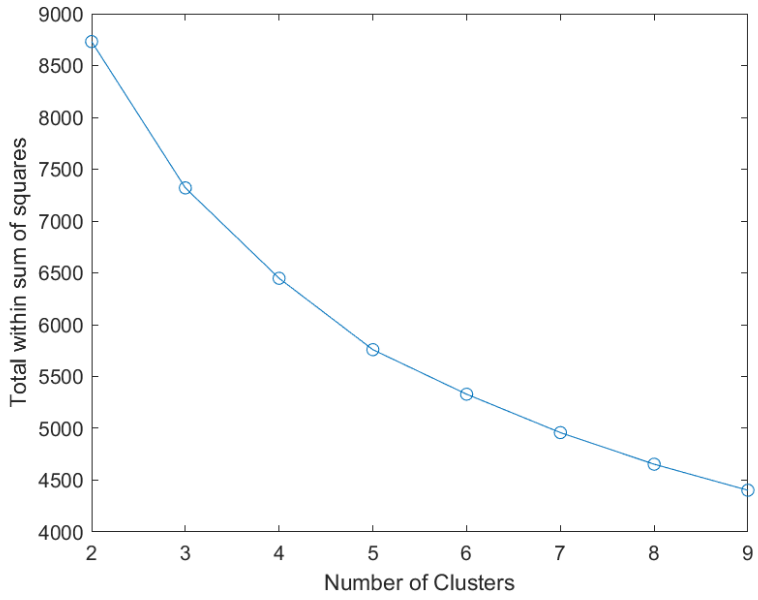
To generate the Elbow statistic, I first calculate the sum of squared errors, SSE_K for an increasing number of clusters, $K = 1, ..., K_{max}$. I then plot the SSE against the number of clusters. As the number of clusters increases, the total within cluster variance decreases (as observations within a cluster become more similar). The optimal number of clusters is at the "kink" in the plot, i.e. the point where the decrease in SSE changes the most. The intuition behind this is that at that point, adding one more cluster adds limited value in explaining the variation in the data.

As can be seen from Figure 4.A.1a, the Elbow plot does not result in a visually obvious kink point in my case. I thus use the Gap Statistic to formally assess where it is (Tibshirani et al., 2001). The Gap Statistic is calculated as the difference between the expected change in SSE versus the actual change in SSE:

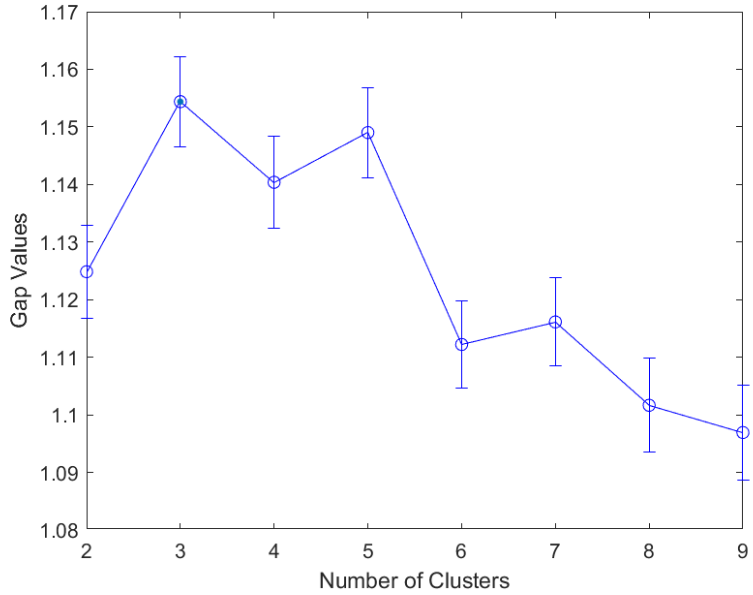
$$\mathrm{Gap}_n(K) = E_n^*\left[\log\left(SSE_K\right)\right] - \log\left(SSE_K\right) \tag{4.17}$$

The expected change E_n^* is determined using a Monte Carlo simulation based on a uniform distribution. The final value for the Gap Statistic takes into account simulation error (hence the error bars in the graph). More details can be found in Tibshirani et al. 2001.

Figure 4.A.1: Choosing the optimal number of clusters



(a) Elbow Plot



(b) Gap Statistics

4.B Further Details on Second Step Estimation

4.B.1 Data: Age 2 BMI and cognition

Data at 22 months was only collected for a random 10% subset of cohort members, all multiple births, those with very long gestation (> 42 weeks), and those with very low birthweight. I only keep data on those individuals that comprised the random subset. I then impute age 2 BMI and cognition for the rest of the sample members from information at age 5 (when all sample members were surveyed) in the following way: For cognition, I first regress age 2 cognition on cognition at age 5, birth weight, mother's education and father's education for the subsample. I correct for measurement error in cognition at age 5. I then predict age 2 cognition for the entire sample. For BMI, I regress age 2 BMI on log height, log head circumference, log BMI of the mother, and birth weight.[15] I then predict age 2 BMI for the entire sample.

4.B.2 Estimation: Child initial BMI and cognition

A child's initial health and cognition at age 2 (agent period 5) depends on their type and additional iid shocks as described in equation (4.18). Table 4.B.1 shows the estimates.

$$\ln C_5^c = \gamma_{0,5} I_{g=k} + \varepsilon_{C,5}^c \tag{4.18}$$

$$\ln B_5^c = \eta_{0,5} I_{g=k} + \varepsilon_{B,5}^c \tag{4.19}$$

$$\text{where } \varepsilon_{C,5}^c \sim N(0, \sigma_g^C) \text{ and } \varepsilon_{B,5}^c \sim N(0, \sigma_g^B)$$

Table 4.B.1: Initial cognition and BMI at age 2

	Cognition (Std)		BMI (Std)	
	Mean	SD	Mean	SD
Type 1	1.06	0.36	-0.34	0.62
Type 2	-0.72	0.39	-0.38	0.49
Type 3	-0.45	0.26	1.06	0.77

Note: Table shows results for mean and variance of initial cognition and BMI at age 2 depending on type. Cognition and BMI here are standardized to have mean zero, variance one.

4.B.3 Further Details: Childhood Cognition Process

I make the following assumptions about the measurement error in cognition $\varepsilon_{m,i,t}$:

- Measurement errors are uncorrelated over time t

[15]Unfortunately, no information on weight of cohort members at age 5 is available.

- Measurement errors are uncorrelated across measures m
- Measurement errors are uncorrelated with cognition $\ln C_{i,t}$
- Measurement errors are uncorrelated across individuals i

Table 4.B.2 shows the measures used at ages 10 and 16 and the corresponding signal to noise ratios, where the signal to noise ratios are defined as the ratio of the variance of the latent factor to the variance of the measure: $\frac{Var(\ln C_{i,t})}{Var(Z_{m,i,t})}$

Table 4.B.2: Cognition measures and their signal to noise ratios

Age 10		Age 16	
Variable	Signal/Noise	Variable	Signal/Noise
		Number of CSE grades	0.95
Reading Score	0.81	Number of O Levels	0.96
Maths Score	0.80	Maths Score	0.57
British Ability Scales	0.81	Reading Score	0.28

Note: Reading Scores, Maths Scores, and British Ability Scales from tests taken as part of survey. O-Levels and Certificate of Secondary Education (CSE) were qualifications attained at age 16 in Great Britain at the time.

4.B.4 Wage shock process

Stochastic part - Detailed version Log wages follow the AR1 process described in (4.7) and (4.8) and are measured with error:

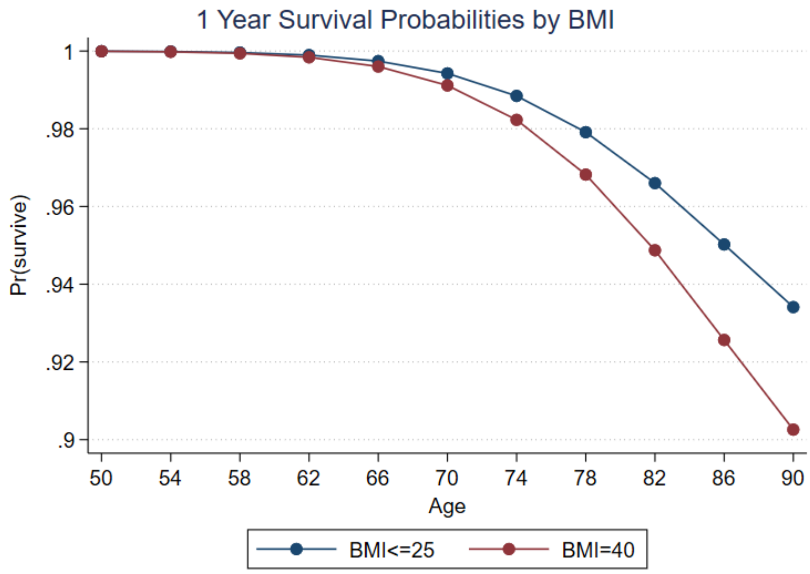
$$\ln(\tilde{W}_t) = \ln(W_t) + \xi_t \tag{4.20}$$

ξ_t is measurement error in period t.

We observe wages every 4 years from age 26 onwards and assume that persistent shocks as well as the measurement error have the same variance in every period: $Var(u_t) = Var(u)$, $Var(\varepsilon_t) = Var(\varepsilon)$ for $t > 16$.

As period lengths differ in the model, I estimate an annual wage process and then adjust it to the appropriate period length. The assumptions to do this are: 1) transitory shocks are assumed to be measurement error, 2) persistent shocks only occur in the periods observed (starting from age 26)

Figure 4.B.1: Predicted 1 year survival probabilities by BMI



Note: Females and non-smokers only. Predicted survival probabilities from a logistic regression using ELSA, controlling for quadratic polynomial in age, marital status, and education.

4.B.5 BMI and mortality

In late adulthood, survival depends on the agent's BMI and age. To estimate this, I use the first available BMI information in ELSA (wave 2), and estimate a logit regression of one year survival on an age-polynomial and health, controlling for education levels and marital status. Following Bhaskaran et al. (2018), I drop smokers, and I drop the first 4 years after BMI is measured first to minimize the effects of reverse causality (diseases causing changes in BMI). I focus on the effect of BMI on mortality for the region of BMI $\geq$ 25. I find that BMI has a strong negative effect on survival. Whilst omitted variable bias may be a concern, the estimates are close in magnitude to the effect of BMI on mortality estimated in the medical literature (Bhaskaran et al. (2018), more details on the comparison are available upon requested from the author). Figure 4.B.1 plots the predicted survival probability for an individual with a BMI of 25 versus an individual with a BMI of 30. The gap in survival probabilities between those with high and low BMI rises with age.

4.B.6 Husband's income

In the BCS70, I observe the cohort members' husband's net earnings at ages 30,34, and 42, but not at ages 26 or 46. To get the full age profile of husband's earnings, I thus first estimate log weekly earnings as a function of age and age squared for the male cohort members of the BCS70 (as opposed to the partner's of the female cohort members). Second, to estimate how husband's earnings depend on wife's type, I take reported husband's earnings and partial out the age profile I previously estimated. I then regress the residual on the wife's type. Finally I estimate the wage shock from the residuals after also partialling out the wife's type. Table 4.B.3 shows the results:

Table 4.B.3: Husband's earnings function

	Coeff	SE
age	0.283	0.036
age^2	-0.004	0.000
Wife's type:		
Type 2	-0.203	0.021
Type 3	-.260	0.023
$Var(\varepsilon_h^Y)$	0.395	

Note: This table shows estimates of the husband's earnings function. Log weekly earnings are a function of an age polynomial and the wife's type. The last row shows the variance of the shock in log weekly earnings of the husband.

4.C Further Details: Health Investments

4.C.1 Construction of the modified Healthy Eating Index

As described in the main text, I use a modified version of the Healthy Eating Index to score the diet qualities of BCS70 cohort members. The original healthy eating index requires a high level of detail to be constructed, e.g. information on cups of fruit per 1,000 kcal. Whilst this level of detail is available at age 47, the information in older waves of the survey is less granular. I thus develop a modified version of the HEI similar to Masip et al. (2019) in which I follow the scoring of the HEI as closely as possible given the ordinal responses of the FFQs. Two major modifications as opposed to the original HEI are made. First, specific nutrients (e.g. grams of protein) are not available in the FFQs. Instead, I assign specific food items to be representative of a particular nutrient. For example, I use the sum of an individual's fish, egg, and meat intake as a measure of protein intake. Second, the FFQ are based on frequency of consumption, rather than grams. I thus use frequency based cut-offs to assign points. For example, rather than giving individuals a maximum score for fruit consumption if they eat at least 0.8 cups of fruit per 1,000 kcal, they get a maximum score if they eat fresh fruit every day.

At age 16, information from diet diaries was coded in grams per day. To calculate the maximum scores, I calculate grams per day/total calories for each food group and then weigh the zscore by maximum points. I then standardize this index to have mean zero, variance 1.

A full list of the scoring I use is presented in Table 4.C.1.

4.C.2 Relating the Healthy Eating Index and latent investments to food expenditure

I now turn to the question of the price for a better quality diet. Here, I use estimates from a paper by Morris et al. (2014) who estimate the costs associated with a better diet for women in the UK. In particular, they report that the lowest quality diet on average cost £3.31 less per day than the highest quality diet after adjusting for covariates. To make this cost applicable to my data, I assume that the cost of the highest (lowest) quality diet applies to the average of the top (bottom) Healthy Eating Index quintile. [16] From this, I can estimate the price of a unit of HEI as the coefficient of a (demeaned) linear regression that has diet expenditure on the left hand side and the healthy eating index on the right hand side: $C_t = p_t HEI_t$.

4.C.3 Investments in the perinatal period

We do not observe an individual's diet quality in the perinatal period. Instead, we observe the following pre- and postnatal investments made by mothers: Week of first antenatal visit, attendance of mothercraft classes, attendance of labour preparation classes, alcohol consumption during pregnancy, breastfeeding duration. I

[16] The original number in Morris et al is £2.04 between the period of 1995-1998. I inflate this number using the CPI inflator.

assume that these measures are informative of the health investments received by the child in the perinatal period and apply a linear measurement system akin to the one used for cognition:

$$Z_{m,i,t} = \mu_{m,t} + \lambda_{m,t} inv_{i,t} + \varepsilon_{m,i,t} \tag{4.21}$$

Here, $Z_{m,i,t}$ is the measure of investment used, $\mu_{m,t}$ is a location parameter and $\lambda_{m,t}$ is a scaling parameter and $\varepsilon_{m,i,t}$ is idiosyncratic measurement error. I then predict a perinatal investment score for each individual using the Bartlett score method, which inversely weighs measures by their noise, thus minimizing bias from measurement error Heckman et al. (2013).

Table 4.C.1: Scoring for modified Healthy Eating Index

	Age 10	
Food Group	**Adequacy**	**Max Pts**
Whole Grains	Brown bread $\geq$ quite often	10
	Breakfast= every day	10
Dairy	Cheese $\geq$ quite often	5
	Milk $\geq$ 1 cup	5
Protein	Eggs+Meat+Fish $\geq$ quite often	5
Fish	Fish $\geq$ quite often	5
	Moderation	
Sugar	Decile of (cola+cocoa+chocolate)	10
Caffeine	Coffee=0	5
	Black tea = 0	5
	Age 16	
Food Group	**Adequacy**	**Max Pts**
Whole Grains	Wholemeal bread,granary bread,high fibre cereals	10
Fruit	Fruit juice, fruit, apple,citrus	10
Veg	Fresh/frozen, canned, tomatoes, green veg, carrots, peas, salad	10
Legumes	Beans, nuts	5
Protein	Eggs, chicken,meat,bacon,yoghurt,milk,cheese,sausage	5
Fish	Fish	5
Food Group	**Moderation**	
Fat	Savoury biscuits, pizza, butter, chips, burger, chips, savoury snacks	10
Sugar	Fizzy drinks, biscuits, chocolate, cake, ice cream, puddings, jam, hot chocolate chocolate, pies, sugar, squash, sugared cereal	10
	Age 30	
Food Group	**Adequacy**	**Max Pts**
Whole Grains	Wholemeal $\geq$ 3-6 days pw	10
Fruit	Fresh fruit $\geq$ every day	10
Vegetables	Sum of (salads, cookedveg)$\geq$ 7 days pw	10
Legumes	pulses $\geq$ 3-6pw	5
Protein	(red meat+poultry+fish+pulses) >=once per day	5
Fish	Fish $\geq$ 1-2pw	5
	Moderation	
Fat	Decile of (fat fried+oil fried+ chips)	10
Sugar	Decile of (sweets+ cakes)	10
	Age 47	
Food Group	**Adequacy**	**Max Pts**
Whole Grains	$\geq$ 1.5 ounce equivalent per 1,000 kcal	10
Total Fruits	$\geq$ 0.8 cup equivalent per 1,000 kcal	5
Whole Fruits	$\geq$ 0.4 cup equivalent per 1,000 kcal	5
Total Vegetables	$\geq$ 1.1 cup equivalent per 1,000 kcal	5
Greens and Beans	$\geq$ 0.2 cup equivalent per 1,000 kcal	5
Dairy	$\geq$ 1.3 cup equivalent per 1,000 kcal	10
Total Protein Foods	$\geq$ 2.5 ounce equivalent per 1,000 kcal	5
Seafood and Plant Proteins	$\geq$ 0.8 ounce equivalent per 1,000 kcal	5
Fatty Acids	(PUFAs + MUFAs) / SFAs $\geq$ 2.5	10
	Moderation	
Refined Grains	$\leq$ 1.8 ounce equivalent per 1,000 kcal	10
Added Sugars	$\leq$ 6.5% of energy	10
Saturated Fats	$\leq$ 8% of energy	10

4.D Further Details: Method of Simulated Moments

To estimate preference parameters, I use the method of simulated moments (MSM). This involves 1) solving the model numerically 2) simulating the model and estimating moments from the simulated data 3) Calculating the GMM criterion 4) iterating through steps 1-3 until the GMM criterion is minimized. Each step is described in more detail now.

4.D.1 Solving the model

The model described is a dynastic model. To solve it, we first need to solve the lifecycle problem via backwards induction, and then iterate until we get a fixed point across generations:

$$V(\mathbb{S}_1) = V^c(\mathbb{S}_1)$$

Discretization The model has two naturally discrete states which are an agent's type and their child's type; each has three grid points. The model has 5 continuous state variables that need to be discretised. These are cash on hand, wage, health, child health, child cognition. Wages are placed on a grid (that has 10 elements) using Tauchen discretization procedure. The cash on hand grid consists of 45 points which are distributed with higher density at lower values, taking the curvature of the value function into consideration. Health in adulthood is also on a growing grid, with 12 points. Child health and child cognition are on a linear grid with 5 points each. There are three control variables in the model. How much to work, how much to save and how much to invest in health. The labour supply choice is discretized into three points - no work, part-time work (20 hours) and full-time work (40 hours). The savings choice is discretized linearly into 20 points within the feasible set given an agent's assets. The investment choice is discretized into seven points between £0 and £3 per day (which is the maximum investment I observe in the data).

Integration Agents in the model need to form expectations over the realizations of shocks to their survival, their wages, their health, child health, child cognition, and the child's type. This involves integrating the value functions over those variables. For wages and survival, I form a weighted average of the value functions evaluated at different grid points, where the weights are the probability of reaching that point[17]. For all other variables, I use the Gauss-Hermite integration procedure as the value function can be "off-grid" given the realization of the shock.

Approximation I use linear interpolation to evaluate the value functions at points that are not on the grid.

Optimization I solve the model by grid search. At each given state, utility for all possible choices is calculated. The utility maximizing choice is the household's optimal decision at the given state.

Backwards Induction and Iterating to a Fixed Point The value of death is normalized to zero. I then iterate backwards through an agent's lifecycle, solving for optimal choices. Agents incorporate their child's problem into their problem as given in equation (4.4). The discounting by λ ensures that this is a contraction mapping. Thus, due to the contraction mapping theorem, we can find the fixed point by iterating across generations. We start by setting $V^c(\mathbb{S}_1) = 0$ and solve the first generation's lifecycle until we get their value at the beginning of the lifecycle $V(\mathbb{S}_1)$. We then calculate the difference between generation 1 and generation 2's value functions: $\max|V^c(\mathbb{S}_1) - V(\mathbb{S}_1)|$. If this is greater than the tolerance value, we update $V^c(\mathbb{S}_1) = V(\mathbb{S}_1)$ and follow the same procedure, until convergence.

4.D.2 Simulating the model

I use the policy functions resulting from the model to simulate the behaviour of 25,000 households (i.e. two generations each). I assume that the first generation starts off with an initial health, type, and wage distribution as estimated from the BCS70 data. Initial assets at age 16 are set to zero. The second generation starts off their independent lifecycle with the health, type and wages as developed during childhood. Again, their initial assets are set to zero.

4.D.3 Calculating the GMM Criterion

Once all individual lifecycles are simulated, the moment conditions are formed from the simulated data analogously to the empirical data. I then find the optimal set of

[17] Note that wages cannot be off-grid due to the Tauchen procedure applied previously

parameters Γ, by minimizing the GMM criterion function:

$$\min_{\Gamma} \mathbf{M}_{\Gamma} \mathbf{I} \mathbf{M}_{\Gamma} \tag{4.22}$$

where $\mathbf{M}$ is a vector of distances of moments in the data versus simulated moments. This depends on the vector of parameters to be estimated Γ. I currently weigh the moments using a diagonal matrix which adjusts for the scale of the moments and weighs them by the number of moments in each group, where the groups are: asset moments (3), employment moments (20), BMI moments (6), type moments (3), other moments (4). Whilst this is not efficient, it is asymptotically consistent. The minimization is conducted by using the BOBYQA algorithm developed by Powell (2009).

4.D.4 Calculating the Standard Errors

I calculate MSM standard errors following standard methods (see e.g. Bolt et al. (2021)), adjusting for simulation error and missing observations.

4.E Further Details on Moments

4.E.1 Wealth Moments

Financial wealth and housing assets are reported at age 42. For ages 34 and 46, only financial wealth is reported. I thus turn to the Wealth and Assets Survey (WAS) to estimate asset moments for ages 34 and 46 for individuals born in the 1970s. The WAS is a biannual representative longitudinal survey of household asset holdings in the United Kingdom starting from 2006. I use the earliest available wave (2006) to get age 34 wealth of the BCS70 cohort by restricting to individuals between 32-35 years of age. I use the wave conducted 8 years later to get at age 42 wealth, now restricting to individuals aged 40-44 (as there is no date of birth available after the first wave). Finally, I use the wave 4 years subsequent (wave 6) and restrict to individuals aged 35-54. Whilst these bins are coarse, I find that the WAS data are consistent with the available wealth variables in the BCS70 (See Table 4.E.1).

Table 4.E.1: Comparison Asset Data WAS and BCS70

	Age 34		Age 42		Age 46	
	Financial Wealth		Total Wealth		Financial Wealth	
	WAS	BCS70	WAS	BCS70	WAS	BCS70
Mean	15,609	13,569	185,763	172,772	61,781	42,581
Median	1,113	1,459	110,323	96,387	9,192	6,649

Note: Values in 2021 GBP

4.E.2 Time with Kids

During years in which children are in the household, a fraction κ_t for general child care is subtracted from a woman's time endowment. This fraction is estimated from the UK Time Use Survey (UKTUS) 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. I select non-single households with at least one child and focus on the age group of the youngest child. I then estimate for child age groups 0-5, 5-10, and 10-16 the fraction of their total time endowment (16 hours per day) that women spent in childcare activities.

Bibliography

Abbott, B., G. Gallipoli, C. Meghir, and G. L. Violante (2019). Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy 127*(6), 2569–2624.

Agostinelli, F. and G. Sorrenti (2018). Money vs. Time: Family income, maternal labor supply, and child development. *University of Zurich, Department of Economics, Working Paper* (273).

Agostinelli, F. and M. Wiswall (2016a, July). Estimating the technology of children's skill formation. *NBER Working Paper 22442*.

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

Aizer, A. and F. Cunha (2012, September). The production of human capital: Endowments, investments and fertility. *NBER Working Paper 18429* (18429).

Akerhielm, K. (1995). Does class size matter? *Economics of Education Review 14*(3), 229–241.

Almond, D. and J. Currie (2011). Killing me softly: The fetal origins hypothesis. *Journal of economic perspectives 25*(3), 153–72.

Almond, D., J. Currie, and V. Duque (2018). Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature 56*(4), 1360–1446.

Almond, D. and B. Mazumder (2011). Health capital and the prenatal environment: the effect of ramadan observance during pregnancy. *American Economic Journal: Applied Economics 3*(4), 56–85.

Altonji, J. G. and T. A. Dunn (1996). Using siblings to estimate the effect of school quality on wages. *The Review of Economics and Statistics 78*(4), 365–671.

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.

Anders, J., L. Macmillan, P. Sturgis, and G. Wyness (2020). Homeschooling during lockdown deepens inequality. *LSE Covid 19 Blog*.

Angrist, J., V. Lavy, and A. Schlosser (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics 28*(4), 773–824.

Apovian, C. M. (2016). Obesity: definition, comorbidities, causes, and burden. *Am J Manag Care 22*(7 Suppl), s176–85.

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

Attanasio, O., C. Meghir, and E. Nix (2020). Human capital development and parental investment in india. *The Review of Economic Studies 87*(6), 2511–2541.

Attanasio, O., C. Meghir, E. Nix, and F. Salvati (2017). Human capital growth and poverty: Evidence from Ethiopia and Peru. *Review of economic dynamics 25*, 234–259.

Bandiera, O., N. Buehren, R. Burgess, M. Goldstein, S. Gulesci, I. Rasul, and M. Sulaiman (2020, January). Women's empowerment in action: Evidence from a randomized control trial in Africa. *American Economic Journal: Applied Economics 12*(1), 210–59.

Banzhaf, H. S. (2021). The value of statistical life: A meta-analysis of meta-analyses. *NBER Working Paper 29185*.

Barahona, N., C. Otero, S. Otero, and J. Kim (2020). Equilibrium effects of food labeling policies. *SSRN Working Paper 3698473*.

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. *Institute of Fiscal Studies Working Paper W17/11*.

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.

Bellisle, F. (2004). Effects of diet on behaviour and cognition in children. *British Journal of nutrition 92*(S2), S227–S232.

Bhalotra, S. and D. Clarke (2020). The twin instrument: Fertility and human capital investment. *Journal of the European Economic Association 18*(6), 3090–3139.

Bhaskaran, K., I. dos Santos-Silva, D. A. Leon, I. J. Douglas, and L. Smeeth (2018). Association of bmi with overall and cause-specific mortality: a population-based cohort study of 3· 6 million adults in the uk. *The lancet Diabetes & endocrinology 6*(12), 944–953.

Biroli, P. (2015). Genetic and economic interaction in the formation of human capital: the case of obesity. *Mimeo. University of Zurich*.

Black, S. E., P. J. Devereux, et al. (2011). Recent developments in intergenerational mobility. *Handbook of Labor Economics 4*, 1487–1541.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). The more the merrier? The effect of family size and birth order on children's education. *The Quarterly Journal of Economics 120*(2), 669–700.

Blanden, J., P. Gregg, and L. Macmillan (2007). Accounting for intergenerational income persistence: Noncognitive skills, ability and education. *The Economic Journal 117*(519), C43–C60.

Blanden, J., P. Gregg, and L. Macmillan (2013). Intergenerational persistence in income and social class: The effect of within-group inequality. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 176*(2), 541–563.

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

Blundell, R., M. Costa Dias, R. Joyce, and X. Xu (2020). Covid-19 and inequalities. *Fiscal studies 41*(2), 291–319.

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). Intergenerational altruism and transfers of time and money: A life-cycle perspective. *Mimeo*.

Bommier, A., D. Harenberg, F. Le Grand, and C. O'Dea (2020). Recursive preferences, the value of life, and household finance. *Cowles Foundation Discussion Paper No. 2231*.

Bonhomme, S., T. Lamadon, and E. Manresa (2017). Discretizing unobserved heterogeneity. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2019-16).

Bryan, J., S. Osendarp, D. Hughes, E. Calvaresi, K. Baghurst, and J.-W. van Klinken (2004). Nutrients for cognitive development in school-aged children. *Nutrition reviews 62*(8), 295–306.

Card, D. (1999). Chapter 30 - The causal effect of education on earnings. *Handbook of Labor Economics 3*, 1801–1863.

Carneiro, P. and J. J. Heckman (2002). The evidence on credit constraints in post-secondary schooling. *The Economic Journal 112*(482), 705–734.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011). Estimating marginal returns to education. *American Economic Review 101*(6), 2754–81.

Carneiro, P., C. Meghir, and M. Parey (2013). Maternal education, home environments, and the development of children and adolescents. *Journal of the European Economic Association 11*(suppl_1), 123–160.

Case, A., A. Fertig, and C. Paxson (2005). The lasting impact of childhood health and circumstance. *Journal of health economics 24*(2), 365–389.

Case, A., D. Lubotsky, and C. Paxson (2002). Economic status and health in childhood: The origins of the gradient. *American Economic Review 92*(5), 1308–1334.

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.

Cawley, J. (2015). An economy of scales: A selective review of obesity's economic causes, consequences, and solutions. *Journal of health economics 43*, 244–268.

Centre for Longitudinal Studies (2017). National child development study [data collection]. http://doi.org/10.5255/UKDA-SN-7669-1.

Centre for Longitudinal Studies (2018). 1970 British Cohort Study [Data Collection]. https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=200001.

Charles, K. K., S. Danziger, G. Li, and R. Schoeni (2014). The intergenerational correlation of consumption expenditures. *American Economic Review 104*(5), 136–40.

Charles, K. K. and E. Hurst (2003). The correlation of wealth across generations. *Journal of political Economy 111*(6), 1155–1182.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly journal of economics 126*(4), 1593–1660.

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.

Conti, G., J. J. Heckman, and R. Pinto (2016). The effects of two influential early childhood interventions on health and healthy behaviour. *The Economic Journal 126*(596), F28–F65.

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives 27*(3), 79–102.

Crawley, H. and C. Summerbell (1997, February). Feeding frequency and BMI among teenagers aged 16-17 years. *International Journal of Obesity 21*(2), 159–161.

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.

Currie, J. and M. Stabile (2003). Socioeconomic status and child health: why is the relationship stronger for older children? *American Economic Review 93*(5), 1813–1823.

Cusick, S. E. and M. K. Georgieff (2016). The role of nutrition in brain development: the golden opportunity of the "first 1000 days". *The Journal of pediatrics 175*, 16–21.

Dahl, G. B. and L. Lochner (2012). The impact of family income on child achievement: Evidence from the Earned Income Tax Credit. *American Economic Review 102*(5), 1927–56.

Dalgaard, C. J., C. W. Hansen, and H. Strulik (2021). Fetal origins -a life cycle model of health and aging from conception to death. *Health Economics 30*(6), 1276–1290.

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

Daruich, D. and J. Kozlowski (2020). Explaining intergenerational mobility: The role of fertility and family transfers. *Review of Economic Dynamics 36*, 220–245.

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

De Nardi, M., E. B. French, and J. B. Jones (2010). Why do the elderly save? the role of medical expenses. *Journal of Political Economy 118*(1), 39–75.

De Nardi, M., S. Pashchenko, and P. Porapakkarm (2017, October). The lifetime costs of bad health. *NBER Working Paper 23963* (23963).

Dearden, L., J. Ferri, and C. Meghir (2002). The effect of school quality on educational attainment and wages. *The Review of Economics and Statistics 84*(1), 1–20.

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. *ESRI working paper*.

Dubois, P., R. Griffith, and M. O'Connell (2018). The effects of banning advertising in junk food markets. *The Review of Economic Studies 85*(1), 396–436.

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

Duncan, G. J., P. A. Morris, and C. Rodrigues (2011). Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments. *Developmental psychology 47*(5), 1263.

El-Sayed, A. M., P. Scarborough, and S. Galea (2012a). Socioeconomic inequalities in childhood obesity in the united kingdom: a systematic review of the literature. *Obesity facts 5*(5), 671–692.

El-Sayed, A. M., P. Scarborough, and S. Galea (2012b). Unevenly distributed: a systematic review of the health literature about socioeconomic inequalities in adult obesity in the united kingdom. *BMC public health 12*(1), 1–12.

Falk, A., F. Kosse, P. Pinger, H. Schildberg-Horisch, and T. Deckers (2021). Socioeconomic status and inequalities in children's iq and economic preferences. *Journal of Political Economy 129*(9), 2504–2545.

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.

Gaheer, S. and G. Paull (2016). The value for money of children's centre services: Evaluation of children's centres in england (ecce). *Department for Education, Research Report*.

Galama, T. J. and H. Van Kippersluis (2019). A theory of socio-economic disparities in health over the life cycle. *The Economic Journal 129*(617), 338–374.

Gayle, G.-L., L. Golan, and M. Soytas (2015). What accounts for the racial gap in time allocation and intergenerational transmission of human capital? *Mimeo*.

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

Gelbach, J. B. (2016). When do covariates matter? And which ones, and how much? *Journal of Labor Economics 34*(2), 509–543.

Gorsuch, R. L., J. Shinka, and W. Velicer (2003). Handbook of psychology: Vol 2. research methods in psychology. *ed. JA Schinka and WF Velicer, Chapter Factor Analysis*, 143–164.

Gregg, P., R. Kanabar, et al. (2021). Intergenerational wealth transmission in great britain. *Centre for Education Policy and Equalising Opportunities, Working Paper*, 21–06.

Gregg, P., L. Macmillan, and C. Vittori (2016). Moving towards estimating sons lifetime intergenerational economic mobility. *Oxford Bulletin of Economics and Statistics, forthcoming*.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy 80*(2), 223–255.

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

Haelermans, C., R. Korthals, M. Jacobs, S. de Leeuw, S. Vermeulen, L. van Vugt, B. Aarts, T. Prokic-Breuer, R. van der Velden, S. van Wetten, et al. (2022). Sharp increase in inequality in education in times of the covid-19-pandemic. *Plos one 17*(2), e0261114.

Haider, S. and G. Solon (2006). Life-cycle variation in the association between current and lifetime earnings. *American Economic Review 96*(4), 1308–1320.

Hall, R. E. and C. I. Jones (2007, 02). The value of life and the rise in health spending. *The Quarterly Journal of Economics 122*(1), 39–72.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–86.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement, volume 5, number 4*, pp. 475–492. NBER.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2018). Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy 126*(S1), S197–S246.

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

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics 24*(3), 411–482.

Ho, F. K., C. Celis-Morales, F. Petermann-Rocha, S. L. Parra-Soto, J. Lewsey, D. Mackay, and J. P. Pell (2021). Changes over 15 years in the contribution of adiposity and smoking to deaths in england and scotland. *BMC public health 21*(1), 1–8.

Holmes, J. (2021, Nov). King's fund briefing: Tackling obesity. https://www.kingsfund.org.uk/sites/default/files/2021-07/Tackling20obesity.pdf.

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

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

Jolivet, G. and F. Postel-Vinay (2020). A structural analysis of mental health and labor market trajectories. *IZA Discussion Paper 13518*.

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.

Keane, M. P. and K. I. Wolpin (1997). The career decisions of young men. *Journal of political Economy 105*(3), 473–522.

Keane, M. P. and K. I. Wolpin (2001). The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review 42*(4), 1051–1103.

Klein, L. and A. Goldberger (1955). An econometric model of the United States, 1929-1952. *Amsterdam: North-Holland Publishing Company*.

Kline, B. and J. L. Tobias (2008). The wages of bmi: Bayesian analysis of a skewed treatment–response model with nonparametric endogeneity. *Journal of Applied Econometrics 23*(6), 767–793.

Kulikova, Y. (2015). Health policies and intergenerational mobility. *Mimeo*.

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

Lochner, L. and A. Monge-Naranjo (2012). Credit constraints in education. *Annu. Rev. Econ. 4*(1), 225–256.

MacKinnon, D. P., A. J. Fairchild, and M. S. Fritz (2007). Mediation analysis. *Annu. Rev. Psychol. 58*, 593–614.

Margaris, P. (2017). The macroeconomic effects of obesity: An application to the uk. *Mimeo*.

Margaris, P. and J. Wallenius (2020). Can wealth buy health? a model of pecuniary and non-pecuniary investments in health. *CEPR Discussion Paper No. DP14796*.

Masip, G., A. Keski-Rahkonen, K. H. Pietiläinen, U. M. Kujala, M. Rottensteiner, K. Väisänen, J. Kaprio, and L. H. Bogl (2019). Development of a food-based diet quality score from a short ffq and associations with obesity measures, eating styles and nutrient intakes in finnish twins. *Nutrients 11*(11), 2561.

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.

Meghir, C. and M. Palme (2005). Educational reform, ability, and family background. *American Economic Review 95*(1), 414–424.

Milligan, K. and M. Stabile (2011). Do child tax benefits affect the well-being of children? Evidence from Canadian child benefit expansions. *American Economic Journal: Economic Policy 3*(3), 175–205.

Mogstad, M., A. Fagereng, and M. Ronning (2021). Why do wealthy parents have wealthy children? *Journal of Political Economy*.

Morris, M. A., C. Hulme, G. P. Clarke, K. L. Edwards, and J. E. Cade (2014). What is the cost of a healthy diet? using diet data from the uk women's cohort study. *J Epidemiol Community Health 68*(11), 1043–1049.

Newey, W. K. (1985). Generalized method of moments specification testing. *Journal of Econometrics 29*(3), 229–256.

Nybom, M. and J. Stuhler (2014). Interpreting trends in intergenerational mobility. *Swedish Institute for Social Research Working Paper Series* (3/2014).

Office for Health Improvement and Disparities (2021). New era of public health to tackle inequalities and level up the uk. https://www.gov.uk/government/news/new-era-of-public-health-to-tackle -inequalities-and-level-up-the-uk,.

Ozkan, S. (2014). Preventive vs. curative medicine: A macroeconomic analysis of health care over the life cycle. *Mimeo*.

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

Papageorge, N. W., V. Ronda, and Y. Zheng (2019). The economic value of breaking bad: Misbehavior, schooling and the labor market. *NBER Working Paper 25602*.

Patel, M. S. and M. Srinivasan (2011). Metabolic programming in the immediate postnatal life. *Annals of nutrition and metabolism 58*(Suppl. 2), 18–28.

Pischke, J.-S. (1995). Measurement error and earnings dynamics: Some estimates from the PSID validation study. *Journal of Business & Economics Statistics 13*(3), 305–314.

Powell, M. J. (2009). The bobyqa algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge*, 26–46.

Public Health England (2021, Nov). Health profile for england: 2021. https://www.gov.uk/government/publications/health-profile-for-england-2021.

Rosner, B. and A. Munoz (1988). Autoregressive modelling for the analysis of longitudinal data with unequally spaced examinations. *Statistics in Medicine 7*(1-2), 59–71.

Scholte, R. S., G. J. Van Den Berg, and M. Lindeboom (2015). Long-run effects of gestation during the dutch hunger winter famine on labor market and hospitalization outcomes. *Journal of health economics 39*, 17–30.

Scholz, J. K. and A. Seshadri (2013). Health and wealth in a lifecycle model. *Mimeo*.

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

Stinebrickner, R. and T. Stinebrickner (2008). The effect of credit constraints on the college drop-out decision: A direct approach using a new panel study. *American Economic Review 98*(5), 2163–84.

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

Theil, H. (1958). Economic forecasts and policy. *Amsterdam: North-Holland Publishing Company*.

Theis, D. R. and M. White (2021). Is obesity policy in england fit for purpose? analysis of government strategies and policies, 1992–2020. *The Milbank Quarterly 99*(1), 126–170.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. American Psychological Association.

Tibshirani, R., G. Walther, and T. Hastie (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*(2), 411–423.

Todd, P. E. and W. Zhang (2020). A dynamic model of personality, schooling, and occupational choice. *Quantitative Economics 11*(1), 231–275.

Yum, M. (2020). Parental time investment and intergenerational mobility. *SSRN WP 3862378*.

Zhu, Z., F. Cao, and X. Li (2019). Epigenetic programming and fetal metabolic programming. *Frontiers in endocrinology 10*, 764.