Essays on Inequality and Human Capital

Rachel Yi Tan

Submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

University College London (UCL)
Department of Economics

June 2025

Declaration

I, Rachel Yi Tan, confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis explores the interplay between inequality and human capital across three distinct but connected chapters. Together, these chapters offer insights into how family background, perceptions, and behavioural responses contribute to persistent disparities in human capital development.

Chapter One delves into the mechanisms driving intergenerational earnings transmission among White, Black, and Hispanic families in the United States. It highlights how factors like educational attainment, cognitive skills, and the quality of the home environment contribute differently across racial groups. Notably, these channels account for a greater share of income persistence among Black and Hispanic families compared to Whites. For Black families, educational attainment and cognitive skills play a more important role than for Whites and Hispanics. Home environment emerges as a key factor for Hispanics, but not for Whites and Blacks.

Chapter Two focuses on parental beliefs about their children's skills, exploring the impact of these beliefs on parental investments and how these beliefs evolve over time. The analysis uncovers three main insights: (1) parental beliefs may be misaligned with actual skills but become more accurate as children age; (2) there are no significant effects of beliefs on investments for both the high SES and low SES parents; and (3) while parental beliefs are persistent over time, they are responsive to changes in children's skill levels.

Chapter Three quantifies the contribution of parental beliefs about child skill to the SES skill gap. This is achieved by estimating a dynamic model of parental investment that incorporates belief dynamics. The model features two-way interaction between beliefs and investments, providing a channel for early beliefs to influence future beliefs, investments and skills. The model reveals that differences in parental beliefs contribute only modestly to the socio-economic skill gap, suggesting that other forces are more dominant in shaping inequality.

Impact statement

This thesis examines how human capital factors shape the intergenerational transmission of income and evaluates the influence of parental beliefs on investment decisions and the socio-economic status (SES) skill gap. Results presented offer valuable insights for policymakers seeking to promote equality of opportunity. Together, the findings highlight the interactions between human capital, parental perceptions, and socioeconomic inequality.

The first chapter investigates the pathways through which income is transmitted across generations among Whites, Blacks, and Hispanics in the United States. The chapter reveals that human capital factors — years of education, cognitive skills, non-cognitive skills and home environment quality — account for a substantial portion of income persistence – ranging from 52.7% to 74.9%. Notably, these mechanisms differ by race: human capital explains a larger share of income transmission among Blacks and Hispanics compared to Whites. Educational attainment and cognitive skills explain a greater proportion of income transmission for Blacks than other races. The quality of the home environment is a key channel of transmission for Hispanics, but not for Whites and Blacks.

The second chapter studies parental beliefs about their children's skills, exploring the impact of these beliefs on investments and how these beliefs evolve with time. Findings indicate that although parents may hold inaccurate beliefs, particularly when children are younger, beliefs become more accurate over time. There are no significant effects of beliefs on investments among both high SES and low SES families. Importantly, while beliefs exhibit strong persistence over time, they are responsive to changes in child skill levels. Providing timely information to parents could adjust their beliefs.

The third chapter introduces a dynamic model of parental investment to quantify the contribution of parental beliefs about child skill to the SES skill gap. The model includes feedback between beliefs and investments, enabling early beliefs to affect later beliefs, investments and skills. The analysis reveals that parental beliefs only contribute to the

SES skill gap in a minor way: equalising beliefs barely reduces the gap. Therefore, while parental beliefs matter in investment decisions, they are not a primary driver of skill inequality, suggesting that policy interventions focused solely on beliefs may have limited impact.

Acknowledgements

I am deeply grateful to Pedro Carneiro, Suphanit Piyapromdee and Monica Costa Dias for their time, generous support, and invaluable guidance during my PhD journey. I am especially thankful to Pedro for his thoughtful advice, patience, and steady encouragement. His confidence in me, both academically and personally, has provided a foundation of support that has shaped my development, and has made a lasting difference.

I also sincerely thank the many professors, senior colleagues, and fellow PhD students who shared their insights and experiences along the way. Their inputs helped to shape both my work and my perspectives. Finally, I am profoundly thankful to my friends and family for their constant reassurance, motivation, and belief in me throughout this process.

Contents

List of Figures

List of Tables

Introduction

This thesis is driven by two central questions: (1) What mechanisms underlie the intergenerational transmission of economic inequality? and (2) Do parental beliefs about their children's skills significantly influence parental investments and contribute to the socio-economic status (SES) skill gap?

In Chapter 1, how earnings are transmitted across generations among White, Black, and Hispanic families in the United States is examined, with a focus on human capital channels —- children's educational attainment, cognitive and non-cognitive skills, and the quality of their home environments. Using mediation analysis, the contribution of these factors towards relative mobility is assessed. The analysis reveals that these channels explain between 52.7% and 74.9% of income transmission, with notable differences across racial groups. In total, the human capital channels explain less of the income transmission for Whites than Blacks and Hispanics. Years of education and cognitive skills explain more of the income transmission of Blacks than other races. The quality of the home environment is an important channel in the income transmission of Hispanics, but not for Whites and Blacks. In addition, racial differences in upward mobility (expected child income rank at parent income rank 25) and downward mobility (expected child income rank at parent income rank 75) are also explored. The Black-White gap in upward/downward mobility is between 11-12 ranks, while the Hispanic-White gap in upward/downward mobility is between 3-5 ranks. Controlling for human capital factors (years of education, skills and home environment quality) eliminates the Hispanic-White gap in upward mobility. Furthermore, accounting for these factors decreases the Black-White gap in downward mobility and the Hispanic-White gap in downward mobility, but slightly raises the Black-White gap in upward mobility.

Chapter 2 turns to the role of parental beliefs about child skill and their influence on parental investment behavior. The analysis indicates that while parental beliefs are often inaccurate, they tend to become more accurate as children age. To determine the casual impact of beliefs on investments, the panel data structure is utilised to obtain instruments for belief. Findings reveal that there are no significant effects of beliefs on investments for both high SES and low SES families. To investigate how beliefs evolve, a belief updating model is estimated. Results indicate that while beliefs are persistent, they adjust in response to changes in children's skills. Overall, the chapter highlights that there may be important connections between beliefs, investments and skills.

In Chapter 3, to quantify the contribution of parental beliefs to the SES skill gap, a dynamic model of parental investment that incorporates parental beliefs is introduced. In the model, two-way interaction between beliefs and investments enables beliefs in earlier periods to influence later beliefs, investments and skills. Model estimates indicate that there is a mechanism consistent with a self-fulfilling prophecy: parents with higher beliefs invest more, which leads them to hold higher beliefs in the following period, and so on. Therefore, their children attain higher skill levels. Given that on average, high SES parents hold higher beliefs than low SES parents, it appears that the self-fulfilling prophecy has the potential to reinforce SES skill disparities. However, it turns out that the force of the self-fulfilling prophecy is low, and beliefs are not a primary driver of the SES skill gap. Rather, disparities in initial resources and baseline child skills account for much more of the observed inequality. These findings imply that even if information-based interventions can shift beliefs, they may have limited impact on reducing inequality.

Chapter 1

Race and Intergenerational Mobility in Earnings

1.1 Introduction

In the United States, there is substantial persistence in earnings across generations (Chetty, Hendren, Kline, & Saez, 2014), which indicates that children's earnings may largely depend on their family circumstances, instead of their own effort. Furthermore, there are racial differences in mobility (Bhattacharya & Mazumder, 2011; Mazumder, 2014; Chetty, Hendren, Jones, & Porter, 2020), which may contribute towards persistent racial wage gaps (Neal & Johnson, 1996; Bayer & Charles, 2018; Petre, 2019; Thompson, 2021). What are the channels behind intergenerational earnings transmission? Do these differ by race? If we can uncover the mechanisms, we may discover effective strategies to level the playing field.

In this study, we investigate the role of human capital factors from birth to adulthood — childhood environments, skills and education — in explaining intergenerational earnings transmission for Whites, Blacks and Hispanics. These channels are likely to be important: richer parents may provide better home environments for their children and enable their children to obtain higher skills and/or higher education, which then allows their children to achieve higher income. Indeed, we find that these channels explain a total of 52.7% of income transmission for Whites, 72.0% of income transmission for Blacks and 74.9% of income transmission for Hispanics. Existing studies on race and intergenerational mobility have not examined the joint contribution of these factors. Studies using administrative data (Chetty et al., 2020, 2024) lack measures of home environment and children's skills. Studies using survey data (including Bhattacharya and Mazumder (2011), Mazumder (2014) and Davis and Mazumder (2018)) explore only a subset of these factors, such as cognitive skills and education.

We now elaborate on our methodology and results in greater detail. We begin by examining relative mobility. First, we estimate the baseline relative mobility of different races. Our measure of relative mobility is the rank-rank slope, which is the coefficient on the parent income rank in a regression of the child income rank on the parent income rank. We find that relative mobility is similar across races, like Chetty et al. (2020) and Davis and Mazumder (2018).

To determine the contribution of human capital channels (years of education, skills and home environment quality) towards explaining income transmission, we conduct a mediation analysis. This involves the following procedure. We regress the child income rank on the parent income rank and these mediators. Then, we compute the percentage of intergenerational transmission explained by each mediator. Furthermore, to understand what drives the percentage explained by each mediator, we estimate (1) the return of the mediator to child income rank and (2) the dependence of the mediator on parent income rank.

The main findings from the mediation analysis are as follows. First, human capital channels explain a large proportion of earnings transmission (between 52.7% and 74.9%). Furthermore, the most important channel is education: years of education explains the highest percentage of intergenerational earnings transmission (around 32.3% to 44.1%). Second, there are racial differences in income transmission. Human capital factors explain a greater proportion of the transmission for Blacks and Hispanics than for Whites. Moreover, years of education and cognitive skills explain a higher proportion of earnings transmission for Blacks than Whites and Hispanics. This is because Blacks attain the highest returns to years of education and cognitive skills. In addition, home environment quality explains a significant proportion of earnings transmission for Hispanics (28.4%), but only a small proportion for Whites and Blacks. This finding is interesting, as it indicates that the quality of the home environment contributes to Hispanics' wages through channels other than skills and education. The importance of home environment quality to the income transmission of Hispanics stems from Hispanics (1) receiving the highest returns to home environment quality and (2) experiencing the strongest correlation between home environment quality and parent income rank.

In the following part of the paper, we move to other measures of mobility: upward mobility (expected child income rank at parent income rank 25) and downward mobility (expected child income rank at parent income rank 75). We begin by obtaining baseline estimates of upward/downward mobility. We find that the upward/downward mobility of Hispanics is closer to Whites than Blacks. The Black-White gap in upward/downward mobility is between 11-12 ranks, while the Hispanic-White gap in upward/downward mobility is between 3-5 ranks. Next, we examine how racial gaps in upward/downward mobility change when we control for human capital factors (years of education, skills and home environment quality). Controlling for these factors significantly decreases the Black-White gap in downward mobility, while it slightly raises the Black-White gap in upward mobility. It also decreases the Hispanic-White gap in downward mobility and eliminates the Hispanic-White gap in upward mobility.

This paper relates to the literature investigating the channels of intergenerational income transmission (Altonji & Dunn, 1991; Hertz, 2006; Blanden, Gregg, & Macmillan, 2007; Blanden, Haveman, Smeeding, & Wilson, 2014; Kourtellos, Marr, & Tan, 2020; Bolt, French, Maccuish, & O'Dea, 2024) and the literature on race and mobility in the United States (Hertz, 2005; Bhattacharya & Mazumder, 2011; Chetty et al., 2014; Davis & Mazumder, 2018, 2024; Chetty et al., 2020, 2024; Jácome, Kuziemko, & Naidu, 2025). We contribute by estimating the contribution of a comprehensive set of human capital factors (spanning birth to adulthood) towards intergenerational income transmission. Existing studies on race and mobility have not evaluated the joint contribution of these factors. Studies relying on administrative data (Chetty et al. (2020) and Chetty et al. (2024)) lack measures of home environment and children's skills, while studies utilising survey data (e.g. Bhattacharya and Mazumder (2011), Mazumder (2014), Davis and Mazumder (2018)) have focused on the contribution of cognitive skills and/or education, neglecting the home environment. Furthermore, our analysis includes Hispanics, which only a few studies (Davis and Mazumder (2018), Chetty et al. (2020) and Chetty et al. (2024)) have explored.

This paper also contributes to the literature on race and discrimination in the United States (Neal & Johnson, 1996; Johnson & Neal, 1998; Carneiro, Heckman, & Masterov, 2005; Neal, 2006; Lang & Manove, 2011; Petre, 2019; Thompson, 2021). These papers examine whether racial differences in skills and education (and in returns to skills and education) can explain racial wage gaps. In our analysis, we demonstrate how (1) the return of the human capital factor to child income rank interacts with (2) the dependence of the human capital factor on parent income rank, contributing to intergenerational income transmission.

The rest of the paper is organised as follows. Section 1.2 explains a measure of mobility which we use in our analysis, the rank-rank slope. Section 1.3 describes the data and variables used in this study. Section 1.4 presents the baseline estimates of relative mobility and explores the mechanisms behind relative mobility. Section 1.5 presents the baseline estimates of upward/downward mobility and investigates whether controlling for human capital factors can explain racial gaps in upward/downward mobility. Section 1.6 concludes the paper.

1.2 Measures of Mobility

1.2.1 Relative Mobility

A measure of relative mobility is the rank-rank slope, the correlation between the child income rank and the parent income rank. The rank-rank slope measures the correlation between the child's position in the income distribution and the parent's position in the income distribution. It is the estimate of b in a regression of the child's percentile rank in the income distribution R_i^c on the parent's percentile rank in the distribution R_i^p, as in Equation 1.2.1.

The interpretation is that a one percentage point increase in the parent income rank is associated with a b percentage point increase in the child's mean income rank. A higher value of the slope denotes lower relative mobility. In the United States, the rank-rank relationship is almost linear (Chetty et al., 2014, 2020).

$$R_i^c = a + bR_i^p + \epsilon_i \tag{1.2.1}$$

1.2.2 Upward and Downward Mobility

Upward mobility is defined as the expected income rank of the child, given a parent income rank of 25. Downward mobility denotes the expected income rank of the child, given a parent income rank of 75. These measures indicate whether children are likely to attain higher or lower income ranks, compared to their parents. They may provide an indication of whether the income persistence (as suggested by the relative mobility estimate) is driven by the high income or the low income (Chetty et al., 2014). Higher relative mobility may be driven by the poor achieving better outcomes and/or the rich achieving worse outcomes.

After estimating Equation 1.2.1, upward and downward mobility measures are obtained by predicting the child income rank at parent income rank 25 and 75 respectively. As an example, the predicted child income rank at parent income rank p is equal to:

$$R_p^c = a + bp \tag{1.2.2}$$

Different from relative mobility, which depends solely on the slope b in Equation 1.2.1, these measures of mobility also depend on the intercept a, the expected income rank of children at the bottom of the income distribution.

1.3 Data

We use the National Longitudinal Survey of Youth 1979 (NLSY79) and the National Longitudinal Survey of Youth Child and Young Adult (CYA), the latter of which surveys the children of females in the NLSY79. Parent family income is obtained from the NLSY79, while child income is retrieved from the CYA. By the most recent release of the CYA, the median age of children is over 30 years, making it possible to observe the wage income of a significant proportion of children during adulthood.

In this study, we explore differences in mobility by race. Race of the mother and the child is defined by the survey screener's assignment of race to the child's mother in the NLSY79. Whites are defined as the non-Black and non-Hispanic individuals in the cross-section sample. Note that the cross-section sample of the NLSY79 was constructed to be representative of the U.S. population. In our analysis, Blacks refer to the Blacks in both the cross-section and supplementary samples. Similarly, Hispanics denote the Hispanics in both the cross-section and supplementary samples. We exclude individuals in the supplementary poor white and the military samples.

The CYA provides measures of the child's educational attainment, cognitive skills, non-cognitive skills and home environment quality. We will use these variables to study the channels of income transmission.

1.3.1 Income Variables

Income variables are adjusted to 2018 dollars using the Consumer Price Index series (CPI-U).

Parent Family Income

Family income sums the following components received by respondent and spouse or partner: wage income, farm/business income, military income, unemployment compensation, Aid to Families with Dependent Children (AFDC), food stamps, Supplemental Security Income (SSI)/welfare, child support, alimony, educational benefits and/or scholarships, fellowships and grants, veteran benefits and income from other sources. From survey year 2002, income from worker's compensation, disability and social security are also included. In our analysis, what we refer to as parent family income is the mean family income over the years when the child is between 0 and 17 years of age. The family income is scaled to adjust for the presence of a spouse/partner. In years when the spouse/partner wage income is reported, the family income is divided by 2. Otherwise, we take the family income value as it is.

Child Wage Income

Child wage income includes the amount received by the child from wages, salary, commissions, or tips from all jobs before deductions for taxes. In our analysis, what we refer to as the child wage income is the average of the child's wage income when the child is between 28 to 33 years old. This age range is chosen to strike a balance between minimising life-cycle bias (Haider & Solon, 2006) and maintaining a sufficient sample for analysis. We average over all available wage income reports within the age range. We only use income reported during years when the child indicates that he/she was not enrolled in school full-time. We assume that the child was not enrolled in school full-time in the past calendar year (income is reported for the previous calendar year) if he/she was not enrolled in school during the year of the interview. Wage income of part-time school attendees is included. We excluded wage entries when we could not determine whether the attendee was enrolled full-time or part-time. The median number of income observations is 2 per child[1].

[1] Since there are few observations of income per child, we may be concerned about measurement error in child income and consequently, measurement error in the child income rank. In this draft, this measurement error issue has yet to be addressed. In the future, we could attempt to address it using

Construction of Income Ranks

Ranks for both parent family income (average family income when the child was aged between 0 and 17 years) and child wage income (average wage income between 28 and 33 years) are computed within the sample of children of the females in the cross-section sample of the NLSY79. We pool all birth cohorts in the ranking — we do not rank by birth cohort as children born in earlier years are negatively selected on mother's characteristics. Both males and females are ranked together, as in Chetty et al. (2014). We want males and females to be ranked in the same distribution, because we wish to compare the estimates of males with those of females[2]. Following Chetty et al. (2014), when there are ties in ranks, each is assigned the mean rank. For example, if 10% of income values are zeros, each individual with a zero wage income is assigned an income rank of 5.

1.3.2 Mediators: Child Skills, Child Education and Home Environment Quality

The mediators are the years of education of the child, a measure of the child's cognitive skills, a measure of the child's non-cognitive skills along with the quality of the home environment.

Years of education: This is the highest years of schooling attained by the child by the latest data release (survey round 2020). Since years of schooling is only collected up to the 2012 survey round, after which the survey switched to collecting the highest category of education received, we impute the years of education from survey round 2014 onwards. More information on the imputation is provided in Appendix 1.A.2.

Cognitive skill: This is the average percentile score on achievement tests administered

the method introduced by Nybom and Stuhler (2016). According to Nybom and Stuhler (2016), even random errors in the income measures result in non-classical error in the income ranks, where the errors in ranks are negatively correlated with the true ranks. To deal with the measurement error, Nybom and Stuhler (2016) propose that the relationship between the observed child income rank R^c and the true rank $\tilde{R}^c$ be represented as a generalised errors-in-variables model: $R^c = \alpha_c + \lambda_c \tilde{R}^c + w_c$. By definition, w_c is uncorrelated to the true rank $\tilde{R}^c$. Because R^c and $\tilde{R}^c$ are ranks, $\lambda_c \leq 1$. A greater value of λ_c implies a lower level of measurement error. Given the generalised errors-in-variables representation, the computed rank-rank correlation is equal to the true rank-rank correlation multiplied by λ_c. Since $\lambda_c < 1$ when there is measurement error, the estimated rank-rank correlation is downward biased. The bias may be corrected if we obtain an estimate of λ_c. Perhaps this can be achieved by estimating the generalised errors-in-variables model on alternative data with longer income trajectories, which we require to observe the lifetime income of individuals and obtain the "true" income rank. We need to assume that the λ_c in that population is equal to the value in the CYA sample. Potential data sets include the National Longitudinal Survey of Youth 1979 and the Panel Study of Income Dynamics.

[2]Estimates for females will be added in a future draft.

by the survey between ages 3 and 14. The tests include the Peabody Picture Vocabulary Test-Revised (PPVT-R), Peabody Individual Achievement Test (PIAT) mathematics, PIAT reading recognition and PIAT reading comprehension. A higher percentile score denotes a higher level of cognitive skills, relative to those of the same age.

Non-Cognitive skill: This is the average percentile score on the Behaviour Problems Index (BPI) between ages 4 to 14. The Behaviour Problems Index is a measure of the level and frequency of behaviour problems exhibited by the child. It is computed from responses by the mother of the child to a set of questions about the child's behaviour. Higher percentile scores denote worse behaviour (and lower non-cognitive skills), relative to children of the same age. More information on the BPI is provided in Appendix 1.A.3.

Home environment quality: This is the average percentile score on the Home Observation Measurement of the Environment (HOME) between ages 0 and 14. The HOME measures the cognitive stimulation and emotional support provided to the child. A higher percentile score denotes a higher level of home environment quality, relative to others of the same age. Additional information on the HOME is provided in Appendix 1.A.4.

1.3.3 Sample Selection and Summary Statistics

We focus on male children. To be included in our sample, we require that children have at least one income observation between ages 28 to 33 when they are not in school full-time. We also require that they have at least one observation of parent family income when they are aged between 0 and 17 years. In addition, as we are interested in how childhood skills and environment shape the wage outcomes of children, we further require that children have at least one cognitive score percentile between 3-14 years, one non-cognitive score percentile between 4-14 years and one home environment score percentile between 0-14 years. These requirements are not very restrictive (around 95% of the sample is preserved). However, we note that there is negative selection into missing scores: the mothers of children with missing scores tend to have lower cognitive skills (as measured by percentile score on the Armed Forces Qualification Test) and lower years of education.

Our sample includes male children from birth cohorts 1974 to 1992. Sample summary statistics are presented in Table 1.1. In the latest survey release (survey round 2020), the median age of children in our sample is between 34 and 37. The percentile scores on the Armed Forces Qualification Test (AFQT), a measure of cognitive skill, is close to 50 for White mothers. In contrast, the AFQT percentiles for the Black and Hispanic mothers are significantly lower. The median level of education reached by mothers is 12 years, which corresponds to high school graduation (assuming no repeated grades). On average, children in our sample were born to mothers between 22 and 24 years of age.

From Table 1.1, we observe that White children attain higher years of education, higher cognitive skills, higher non-cognitive skills (lower percentile of behaviour problems) and higher home environment quality than Blacks and Hispanics. The Black-White gap in the cognitive skill percentile is around 20, the Black-White gap in the behaviour problems index percentile is around 5 and the Black-White gap in the home environment quality percentile is around 23. The Hispanic-White gap in the cognitive skill percentile is around 15, the Hispanic-White gap in the behaviour problems index percentile is around 3 and the Hispanic-White gap in the home environment quality percentile is around 16.

Table 1.2 presents summary statistics of the child wage income and parent family income in our sample. Parent family income has been scaled by size — in years when spouse/partner income was reported, the income was divided by 2. The median and mean of child wage income are highest for Whites, followed by Hispanics and then Blacks. The same pattern holds for parent family income. Within each race group, a fraction of individuals received 0 wage income. The proportion of those who report zero wage income is lowest for White males and highest for Black males. If this reflects unemployment, Black males are more likely to be unemployed than the other races.

1.3.4 Discussion on Sample

Though the cross-section of the NLSY79 has been constructed to be representative of the individuals in those birth cohorts, the children of the females in the cross-section may not be representative of their birth cohorts. There is selection into fertility and timing of children. Children in earlier birth cohorts were born to younger mothers and are generally negatively selected in terms of mother's characteristics — their mothers typically have lower years of education and cognitive skills. In addition, they are more likely to grow up in single parent households.

Table 1.1: Summary Statistics

	White-M	Black-M	Hispanic-M
Age of Child at 2020 survey			
Median	34	37	35
Mean	34.7	36.5	35.9
Mother's AFQT Percentile			
Median	51.1	15.3	18.6
Mean	51.6	21.2	24.3
Mother's Years of Education			
Median	12	12	12
Mean	13.4	13.0	12.0
Mother's Age at Birth			
Median	24	22	23
Mean	24.3	22.3	23.2
Child's Years of Education			
Median	13.9	12.5	12.5
Mean	14.0	12.9	13.0
Cognitive Skill Percentile Age 3-14			
Median	57.8	33.8	40.7
Mean	56.2	34.9	40.7
Behaviour Problems Index Percentile Age 4-14			
Median	64.1	70.8	68.7
Mean	61.4	66.5	64.2
Home Percentile Age 0-14			
Median	58.7	29.9	36.6
Mean	56.2	33.3	39.7
Number of children	1,070	830	569

Notes: This table presents summary statistics of White, Black and Hispanic males in the sample. Age of child at 2020 survey is the age of the child during the 2020 survey round, which is the most recent release of the CYA. AFQT stands for the Armed Forces Qualification Test, which is a measure of cognitive skill. Mother's AFQT scores are missing for some children and the AFQT summary statistics are based on non-missing values. We observe AFQT scores for 1,036 Whites, 804 Blacks and 541 Hispanics. The child's cognitive skill, the Behaviour Problems Index and home environment quality have been described in section 1.3.2.

Furthermore, in our analysis, we require at least one income observation between age 28 and 33. This affects the sample composition: children who are younger than age 28 in the latest survey round are excluded from our income sample. This means that our sample includes a higher proportion of children in earlier birth cohorts (born to younger mothers) than children in later birth cohorts (born to older mothers). Overall, our sample includes children in birth cohorts between 1974 and 1992.

When interpreting the results, it is important to keep in mind that the sample is different

Table 1.2: Income Summary Statistics (Deflated to Year 2018 using CPI-U)

	White-M	Black-M	Hispanic-M
Child wage income			
Median	42,340	22,131	32,952
Mean	48,814	26,324	36,252
% Zeros	5.23	16.63	7.73
Parent family income			
Median	31,195	12,951	18,507
Mean	38,562	18,233	22,751
% Zeros	0.00	0.48	0.00
Number of children	1,070	830	569

Notes: This table presents income statistics for White, Black and Hispanic males. Child wage income is the average income from wages and salary when the child was aged between 28 and 33 years. Parent family income is the average family income when the child was aged between 0 and 17 years. Parent family income has been scaled by size — in years when spouse/partner income was reported, the income was divided by 2. The % zeros row provides the proportion of the sample which reported zero values for the respective incomes. All monetary values are measured in 2018 dollars (deflated using the CPI-U).

from the administrative data used in Chetty and Hendren (2018), which is nationally representative of the tax population.

1.4 Relative Mobility

To obtain baseline measures of relative mobility, we estimate Equation 1.2.1 for each race[3]. Estimates are presented in the first row (“Baseline") of Table 1.3. They indicate, for example, that a 1 percentage point increase in the parent family income rank is associated with a 0.388 percentage point increase in the child’s mean wage income rank for White males.

Our rank-rank slope estimates are within the range of the literature for the United States, which is between 0.28-0.4 (Chetty et al., 2014; Mazumder, 2016; Bratberg et al., 2017; Mazumder, 2018; Davis & Mazumder, 2018; Chetty et al., 2020; Davis & Mazumder,

[3] In the main analysis, we assume that the relationship between the child income rank and the parent income rank is linear. In Appendix 1.E, we relax this assumption and estimate the child income rank as a function of a linear spline in parent income rank. Then we assess whether the child income rank and the parent income rank are related in a non-linear manner, by testing whether the slopes between the knots of the spline are equal to each other. We find evidence of non-linearity for Black males and Hispanic males.

2024). Like Chetty et al. (2020) and Davis and Mazumder (2018), the relative mobility estimates do not differ by race — we cannot reject that the rank-rank slopes are equal across races.

This being said, our rank-rank slope estimates are higher than corresponding estimates in Chetty et al. (2020). This may be because our children are from a wider range of birth cohorts than their sample and/or because we use the income of children when they are younger. We note that Davis and Mazumder (2024) obtain a rank-rank slope of 0.38 for parent-son pairs when son's income is measured at age 29, which is a similar age range to our sample (ours is the average income between 28 and 33 years). Moreover, Bratberg et al. (2017) estimate a rank-rank slope of 0.395 for the United States.

Table 1.3: Estimates of the Rank-Rank Slope

	White-M	Black-M	Hispanic-M
1.Baseline			
estimate	0.388	0.370	0.332
se	0.031	0.041	0.048
2.Educ			
estimate	0.223	0.179	0.207
se	0.033	0.041	0.052
3.Educ+Cog			
estimate	0.206	0.137	0.168
se	0.033	0.042	0.054
4.Educ+Cog+Noncog			
estimate	0.191	0.114	0.157
se	0.033	0.042	0.054
5.Educ+Cog+Noncog+Home			
estimate	0.184	0.104	0.083
se	0.035	0.044	0.058
Number of children	1,070	830	569

Notes: This table presents the estimates and corresponding standard errors of the coefficient on the parent income rank in a regression of child wage income rank on parent income rank. Panel 1 presents the baseline estimates. From Panel 2 onwards, additional variables (mediators) are included in the regression as controls. Educ denotes years of education, Cog denotes cognitive skill, Noncog denotes non-cognitive skill and Home denotes the quality of the home environment. The definition of the mediators is provided in the data section.

Having documented the baseline relative mobility, we are now interested to explore what mediates intergenerational income transmission. We concentrate on human capital channels: the child's years of education, cognitive skills (average percentile score on achievement tests between ages 3 and 14), non-cognitive skills (average percentile score on Behaviour Problems Index between ages 4 and 14) and home environment quality (average percentile score on home environment quality between ages 0 and 14). The idea is that parent income rank affects child income rank through these human capital channels, which we call mediators (Black & Devereux, 2011). For instance, children of the rich may receive higher investments, attain higher skills and/or higher years of education, which in turn enables them to attain higher income.

We begin by examining how the rank-rank slope changes when we control for these channels. Table 1.3 presents the coefficients on the parent income rank for each race after adding mediators into the regression sequentially. The row “Educ" presents the rank-rank slopes after controlling for years of education. The row “Educ + Cog" presents the rank-rank slopes after controlling for years of education and cognitive skills of the child. The row “Educ + Cog + Noncog" presents the rank-rank slopes after controlling for years of education, cognitive skills and non-cognitive skills of the child. The row “Educ + Cog + Noncog + Home" presents the rank-rank slopes after controlling for the years of education, cognitive skills, non-cognitive skills and home environment quality of the child. When additional variables are added, the relative mobility estimate decreases, which indicates that the new variables explain a segment of intergenerational earnings transmission independent of the existing variables (Groves, 2005).

As observed in Table 1.3, once the mediators are included, the rank-rank slopes decrease by a higher percentage for the Blacks and Hispanics than the Whites, which indicates that human capital factors explain a greater proportion of income transmission for Blacks and Hispanics than for Whites. This leads to a widening of racial gaps in relative mobility. As 47% of income transmission remains unexplained for Whites, there must be other factors which explain the transmission. What might these channels be? The literature indicates that school quality, neighbourhood and social capital could also influence mobility (Cholli & Durlauf, 2022). It would be interesting to uncover the importance of these channels in future work.

Next, we wish to understand the relative importance of each of the human channels in explaining earnings mobility. To formally determine the contribution of each mediator towards intergenerational income transmission, we conduct a mediation analysis. Before doing so, we explain the mediation analysis procedure. Consider the case of a single variable M, the mediator, which mediates the relationship between parent income rank R^p and child income rank R^c. First, we estimate the relationship between the mediator

M and parent income rank R^p.

$$M_i = \alpha_M + \eta_M R_i^p + \varepsilon_{Mi} \tag{1.4.1}$$

We also estimate the returns of the mediator to child income rank γ_M using a regression of the child income rank R^c on both the parent income rank R^p and the mediator M.

$$R_i^c = \alpha + \gamma_M M_i + \delta R_i^p + u_i \tag{1.4.2}$$

The rank-rank slope can be decomposed as follows:

$$b = \underbrace{\gamma_M \eta_M}_{through\, M\, component} + \underbrace{\delta}_{direct\, component} \tag{1.4.3}$$

The percentage explained by mediator M is the percentage decrease in the coefficient on the parent income rank after mediators are added. It is computed from the estimated coefficients on the parent income rank in Equation 1.2.1 and Equation 1.4.2. Specifically, the percentage explained by the mediator is equal to $\left(1 - \frac{\delta}{b}\right) \times 100\%$. Note that this decomposition is descriptive, rather than causal.

The procedure we have just described is for a single mediator. If there are two mediators (M_1 and M_2), we can extend the mediation procedure by using the following equations (Equation 1.4.4 and Equation 1.4.5).

$$M_{ki} = \alpha_{M_k} + \eta_{M_k} R_i^p + \varepsilon_{M_{ki}}, k = 1, 2 \tag{1.4.4}$$

$$R_i^c = \tilde{\alpha} + \gamma_{M_1} M_{1i} + \gamma_{M_2} M_{2i} + \tilde{\delta} R_i^p + u_i \tag{1.4.5}$$

The rank-rank slope can be decomposed as follows:

$$b = \underbrace{\gamma_{M_1} \eta_{M_1}}_{through\, M_1\, component} + \underbrace{\gamma_{M_2} \eta_{M_2}}_{through\, M_2\, component} + \underbrace{\tilde{\delta}}_{direct\, component} \tag{1.4.6}$$

The percentage of income transmission explained by mediator M_1 is $\left(\frac{\gamma_{M_1}\eta_{M_1}}{b}\right) \times 100\%$ while the percentage explained by mediator M_2 is $\left(\frac{\gamma_{M_2}\eta_{M_2}}{b}\right) \times 100\%$. This framework can be extended to $K > 2$ mediators by including all K mediators in Equation 1.4.5 and estimating Equation 1.4.4 for each of the K mediators.

In our analysis, we have $K = 4$ mediators. We run a regression of the child income rank on the parent income rank and the four mediators: years of education, cognitive skills, non-cognitive skills and home environment quality. Then, we use the method described above to compute the percentage explained by each mediator[4]. Furthermore, as seen from above, the percentage explained by each mediator M_k is based on (1) the return of the mediator to the child income rank γ_{M_k} and (2) the correlation between the mediator and parent income rank η_{M_k}. We also estimate these for each mediator, to understand what is driving the percentage explained.

Table 1.4 presents the percentage explained by each of the mediators and the total percentage explained by all mediators[5]. Table 1.5 presents the corresponding estimates of the returns of the mediators and the correlation between the mediator and parent income rank. From Table 1.4, the combination of years of education, cognitive skills, non-cognitive skills and home environment quality account for between 52.7% and 74.9% of intergenerational earnings transmission. This suggests that human capital channels play crucial roles in income transmission, which is consistent with studies indicating that human capital is a key determinant of children's life outcomes (Keane & Wolpin, 1997; Heckman, Stixrud, & Urzua, 2006; Almlund, Duckworth, Heckman, & Kautz, 2011; Francesconi & Heckman, 2016). In total, the human capital channels explain a higher percentage of transmission for Blacks and Hispanics than Whites.

Among the mediators, years of education explains the highest percentage (between 32.2% and 44.1%) of transmission across all races. Cognitive skills explain between 6.5% and 10.9% of the transmission, while non-cognitive skills explain between 3.3% and 7.6%. Home environment quality generally explains the least of the income transmission (around 3%), except for Hispanics.

There are racial differences in the relative importance of the human capital channels. First, both education and cognitive skills explain a higher percentage of income transmission for Blacks than the other races. As can be seen from Table 1.5, this is driven by Blacks receiving higher returns to education and cognitive skills. This corroborates other studies which find that the return to years of education is higher for Blacks than

[4]In practice, we use the STATA command written by Gelbach (2016) to perform the computation.

[5]This mediation analysis is done under the assumption that the relationship between the child income rank and the parent income rank is linear. In Appendix 1.E, we allow the relationship between the child income rank and the parent income rank to be non-linear, by estimating the child income rank as a function of a linear spline in parent income rank. Using this framework, we examine whether the percentage explained by the mediators depends on parent income rank. For Whites, we find suggestive evidence that the mediators might explain a greater proportion of income transmission at middle values of parent income. However, the difference in percentage explained might not be statistically significant. This needs to be investigated in greater detail.

Whites, or at least as high for Blacks as Whites (Barrow & Rouse, 2005; Heckman, Lochner, & Todd, 2006; Neal, 2006). This is also consistent with studies which find that the return to cognitive skills is higher for Blacks than Whites, or at least as high for Blacks as Whites (Neal & Johnson, 1996; Johnson & Neal, 1998; Neal, 2006; Thompson, 2021).

Table 1.4: Percentage of Intergenerational Transmission Explained by Mediators

	White-M	Black-M	Hispanic-M
Education	35.3	44.1	32.3
Cognitive	6.5	15.5	10.9
Non-Cognitive	7.6	8.7	3.3
Home	3.3	3.6	28.4
Total	52.7	72.0	74.9

Notes: This table presents the percentage of income transmission explained by each of the mediators: years of education, cognitive skill, non-cognitive skill and home environment quality. Total refers to the percentage explained by all mediators.

Table 1.5: Mediation Decomposition

	b	β	Education			Cognitive			Non-Cognitive			Home		
			γ	η	$\gamma \times \eta$	γ	η	$\gamma \times \eta$	γ	η	$\gamma \times \eta$	γ	η	$\gamma \times \eta$
White-M	.388	.184	3.454	.04	.137	.081	.31	.025	-.107	-.276	.029	.035	.364	.013
Black-M	.37	.104	5.207	.031	.163	.171	.336	.057	-.145	-.224	.032	.033	.402	.013
Hispanic-M	.332	.083	2.834	.038	.107	.105	.344	.036	-.049	-.221	.011	.196	.483	.094

Notes: For each mediator, this table presents (1) the return of the mediator to child income rank (γ) and (2) the correlation between the mediator and parent income rank (η). b is the baseline rank-rank slope. β is the rank-rank slope after controlling for all mediators: years of education, cognitive skills, non-cognitive skills and home environment quality. Mediators are defined in the data section.

Second, home environment quality explains a significant proportion of intergenerational income transmission for Hispanics (28.4%), but only a low proportion for the other races (3.3%-3.6%). This is a striking result as it suggests that the quality of the home environment contributes towards Hispanics' labour market outcomes through channels other than skills and education. To the best of our knowledge, ours is the first study to document this. From Table 1.5, this finding arises because Hispanics have the highest returns to home environment quality and the highest correlation between home environment quality and parent income rank.

Now, we examine Table 1.5 in greater detail, beginning with education. Blacks have the highest returns to education. However, the dependence of years of education on the parent income rank is lowest for Blacks (The finding that Blacks have a lower correlation between years of education and parent income than Whites is corroborated by Belley and Lochner (2007)). Therefore, when we multiply the return to education (γ) by the dependence on parent income (η), the resulting figures ($\gamma \times \eta$) are not very different across races.

Next, we turn to cognitive skills. Blacks have higher returns to cognitive skills than Whites and Hispanics. In addition, the dependence of cognitive skills on parent income rank is not very different across races. Consequently, when we multiply the return to cognitive skill by the dependence on parent income, the resulting figure is highest for Blacks.

Moving to non-cognitive skills, both the returns and the dependence on parent income rank are negative because a higher percentile of behaviour problems indicates a lower level of non-cognitive skills. The (negative) return to a higher level of behaviour problems (lower non-cognitive skills) is lowest for Hispanics. Additionally, the correlation between non-cognitive skill and the parent income rank is lowest for Hispanics (though very close to Blacks). Therefore, when we multiply the return to non-cognitive skill by the dependence on parent income, the resulting figure is lowest for Hispanics.

Finally, we move to home environment quality. Hispanics have the highest returns to home environment quality and the highest correlation between home environment quality and parent income rank. Overall, when we multiply the return to home environment quality by the dependence on parent income, the resulting figure is highest for Hispanics.

Since the return to education and cognitive skills is higher for Blacks than Whites, policies which raise skill accumulation and educational attainment of Blacks could be effective in reducing the White-Black gap in earnings. Why is it that Blacks do not achieve higher education and skills, to realise these higher returns? Neal (2006) proposes that Blacks may face barriers. Blacks may face discrimination from their peers in school (Chavez, 2021); the psychological costs from such discrimination may dissuade them from staying in school. Blacks may also face higher opportunity costs of continuing their education because of family circumstances and responsibilities. A recent study by Lumina Foundation-Gallup finds that Blacks are more likely to cite external responsibilities as a reason preventing them from attaining college education (*The State of Higher Education 2023 Report*, 2023). If these are the key reasons, raising skills and education of Blacks may require changing underlying attitudes and perceptions and family circumstances, which is an uphill task.

Why is it that Hispanics have higher returns to home environment quality? Perhaps Hispanic families are distinct from other races in terms of culture, parenting styles and values. For example, Hispanic families exhibit familism which is characterised by behaviours such as strong family ties and family values, fostering close family relationships and relying on the family for social support (Landale, Oropesa, & Bradatan, 2006). In addition, the attitudes of Hispanic parents may be different. A recent Pew Research Center report highlights that Hispanic parents are more concerned than White and Black parents that their children will face challenges such as bullying and drugs (Minkin & Horowitz, 2023). Maybe these factors interact with home environment quality in such a way as to amplify the returns.

1.5 Upward and Downward Mobility

We estimate the upward mobility (expected child income rank at parent income rank 25) and downward mobility (expected child income rank at parent income rank 75) of Whites, Blacks and Hispanics. Compared to relative mobility, a measure of average income persistence, upward/downward mobility measures indicate whether the income ranks of children are generally improving or worsening, relative to their parents.

Table 1.6 presents the estimates. Children at parent income rank 25 are predicted to attain higher income ranks than their parents, while those at parent income rank 75 are predicted to reach lower income ranks than their parents. Furthermore, the predicted income ranks of Hispanics are noticeably closer to Whites than they are to Blacks. For example, the upward mobility gap between Hispanics and Whites is around 3 ranks, while the upward mobility gap between Blacks and Whites is around 11 ranks. The downward mobility gap between Hispanics and Whites is around 5 ranks, while the corresponding gap between Blacks and Whites is around 12 ranks. The magnitude of our Black-White gaps is similar to Chetty et al. (2020), who estimate a gap of 10.0 ranks at parent income rank 25 and a gap of 11.7 ranks at parent income rank 75.

Table 1.6: Upward and Downward Mobility

	White-M	Black-M	Hispanic-M
Parent Income Rank 25	50.54	39.31	47.45
Parent Income Rank 75	69.93	57.82	64.04

Notes: This table presents the predicted child wage income ranks of White, Black and Hispanic males at parent income rank 25 and parent income rank 75.

Next, we examine the extent to which the upward/downward mobility changes when we control for human capital factors: years of education, cognitive skill, non-cognitive skill and home environment quality. To achieve this, we estimate Equation 1.5.1 separately for each race. Then we predict the child income rank at parent income ranks 25 and 75. Since the predicted ranks depend on the value of mediators, we use the same value of mediators for all races. Specifically, we use the 25th and 75th percentile of mediators in the distribution of cross-section males, for parent income rank 25 and 75 respectively.

$$R_i^c = a + b_p R_i^p + c_1 education_i + c_2 cognitive_i + c_3 noncognitive_i + c_4 home_i + e_i \quad (1.5.1)$$

Table 1.7 presents the predicted ranks after controlling for human capital factors. Controlling for the human capital factors, the Hispanic-White gap in downward mobility decreases and the Hispanic-White gap in upward mobility is eliminated. In addition, the Black-White gap in upward mobility widens slightly, while the Black-White gap in downward mobility declines significantly. Overall, this may indicate that differences in human capital explain a greater proportion of Hispanic-White disparities at lower parent income ranks than at higher income ranks. In addition, differences in human capital explain a greater proportion of the Black-White disparities at higher parent income ranks than at lower income ranks.

Table 1.7: Upward and Downward Mobility After Controlling for Mediators

	White-M	Black-M	Hispanic-M
Parent Income Rank 25			
Educ+Cog+Noncog+Home	44.63	32.79	44.63
Parent Income Rank 75			
Educ+Cog+Noncog+Home	75.58	71.18	72.57

Notes: This table presents the predicted child wage income ranks of White, Black and Hispanic males at parent income rank 25 and parent income rank 75. To compute the predicted child income ranks at the 25th / 75th percentile of parent income, we use the 25th / 75th percentile of mediators in the sample of male children of the NLSY79 cross-section sample. Educ denotes years of education, Cog denotes cognitive skill, Noncog denotes non-cognitive skill and Home denotes the quality of the home environment.

1.6 Conclusion

In this paper, we investigate the mechanisms behind intergenerational income transmission for Whites, Blacks and Hispanics in the United States, to provide insights on how to improve equality of opportunity. We focus on human capital channels: children's years of education, skills and home environment quality. These are likely to be important — richer parents may provide their children with better home environments and enable their children to achieve higher skills and/or education, which in turn allows them to obtain higher income.

Indeed, human capital factors are important channels of intergenerational income transmission. Together, years of schooling, cognitive skills, non-cognitive skills and home environment quality explain between 52.7% and 74.9% of the transmission. Moreover, among the mediators, years of education is the most important channel of intergenerational income transmission, explaining between 32.3% and 44.1% of the transmission.

The analysis reveals racial differences in the mechanisms. First, human capital channels (years of education, skills and home environment quality) explain a higher proportion of intergenerational income transmission for Blacks and Hispanics than Whites. Second, years of education and cognitive skills explain a higher proportion of earnings transmission for Blacks than Whites and Hispanics. This is driven by Blacks receiving higher returns to education and cognitive skills than the other races. Third, home environment quality explains a sizeable proportion of the earnings transmission for Hispanics (28.4%), but only a low proportion for other races (around 3%). This arises as the dependence of home environment quality on parent income rank is higher for Hispanics and Hispanics derive higher returns from home environment quality.

In the following part of the paper, we move to upward mobility (the expected child income rank at parent income rank 25) and downward mobility (expected child income rank at parent income rank 75). We find that Black-White gap in upward/downward mobility is between 11-12 ranks, while the Hispanic-White gap in upward/downward mobility is between 3-5 ranks. After this, we investigate whether differences in years of education, cognitive skills, non-cognitive skills and home environment quality can explain racial gaps in upward/downward mobility. We find that controlling for these factors decreases the Black-White gap in upward mobility and raises the Black-White gap in downward mobility. In addition, controlling for the factors eliminates the Hispanic-White gap in upward mobility and decreases the Hispanic-White gap in downward mobility. This suggests that human capital factors can fully account for Hispanic-White gaps in upward mobility. Furthermore, they can account for some proportion of Black-White gaps in downward mobility and some proportion of the Hispanic-White gap in downward mobility.

A limitation of our analysis is that we cannot account for other potentially important factors such as school quality, neighbourhood and networks, which are probably highly correlated with the mediators we use. Looking forward, it would be informative to disentangle the contribution of school quality, neighbourhood and networks from children's years of education, children's skills and home environment quality.

Appendix to Chapter 1

1.A Data Appendix

1.A.1 Construction of Family Income

To construct parent family income, we sum various income components reported in the NLSY79. Apart from the income received from wages and salary, whenever an income component is missing, we code it as 0. In years when wages and salary income of the NLSY79 respondent's spouse/partner is non-missing (including report of 0 wage income), we divide family income by 2.

Similarly, to construct child family income, we sum various income components reported in the CYA. Apart from the income received from wages and salary, whenever an income component is missing, we code it as 0. During years when the income of the child's spouse/partner is non-missing (including report of 0), we divide family income by 2.

1.A.2 Imputation of Years of Education

In survey years 1994-2012, the highest grade achieved by the respondent was reported in terms of years of education. From survey year 2014 onwards, the highest grade achieved was reported in terms of categories of education. Therefore, we impute the years of education in survey years 2014 onwards.

To achieve this, in each survey round, we first classify the respondents into the following categories of educational attainment: (1) less than high school graduate; (2) high school diploma; (3) some vocational education; (4) completed vocational training (after high school); (5) some college; (6) completed college (associate's degree); (7) completed college (bachelor's degree); (8) some graduate school or completed graduate school (master's, PhD, post-baccalaurete professional education).

The classification is performed by inferring the education category of the respondent based on (1) the dates at which his/her qualifications are attained and (2) whether he/she is enrolled in school during that survey round. For example, if the respondent reports being enrolled in school after receiving a bachelor's degree, we presume that the respondent has attained some graduate education (master's, PhD or professional degree).

Next, using the inferred education categories, we estimate an imputation model: a regression of the years of education on the categories of educational attainment and a quadratic function of the respondent's age. This model is estimated on data from the cross-section sample and the supplemental samples of Blacks and Hispanics, up to survey round 2012. The estimated model is utilised to predict the years of education of respondents in survey rounds 2014 onwards, based on the reported education categories and the respondent's age.

1.A.3 Non-Cognitive Skill Measure: Behaviour Problems Index

The Behaviour Problems Index (BPI) is based on the mother's responses to questions regarding the frequency at which the child exhibits problem behaviours. These questions are asked when the child is aged between 4 and 14. The BPI contains subscales in the following topics: anxious/depressed, antisocial, dependent, headstrong, hyperactive and peer conflicts/withdrawn. Some examples of items included in the subscales are: breaks things deliberately, cheats or tells lies, has sudden changes in mood or feeling, is disobedient at school, has trouble getting along with teachers, is disobedient at home and is not liked by other children. Possible responses are (1) often true, (2) sometimes true and (3) not true. The BPI items are age-specific. For example, being disobedient in school is only asked when children are older than 5 years while breaks things deliberately is only asked when children are younger than 12. A complete list of the items in each of the BPI subscales is provided in Appendix D of the codebook supplement of the CYA[6].

To form the overall BPI score, items on the subscales are recoded to binary variables before being added together. The responses "often" or "sometimes true" are coded as 1, while "not true" is coded as 0. This analysis relies on a normed version of the overall BPI score, the percentile score of the BPI. The BPI percentile provides an indication of

[6]https://www.nlsinfo.org/content/cohorts/nlsy79-children/other-documentation/codebook-supplement/appendix-d-behavior-problems (last assessed: 5 June 2025)

the level of behaviour problems relative to children of the same age: a higher percentile score indicates worse behaviour.

1.A.4 Home Environment Quality: Home Observation Measurement of the Environment

The Home Observation Measurement of the Environment (HOME) measures the cognitive stimulation and emotional support provided to the child. Some items relating to cognitive stimulation are the following: how many children's books the child has, how often the child is brought to the grocery store, how often the mother reads to the child, whether the family encourages the child to start and keep engaging in hobbies, whether there is a musical instrument which the child can use at home and how often the child is brought to the museum. Some items relating to emotional support are the following: how often the child eats a meal with the child's mother and his/her father/stepfather/father-figure, how often the mother talks to the child when working, how often the whole family gets together with relatives or friends, how often the mother spanked the child in the past week, whether the child is expected to clean his/her own room, how often the child is expected to pick up after himself/herself and how often the child is expected to keep shared living areas clean and straight.

Items included in the HOME depend on the age of the child. There is a different set of questions for children under 3 years, 3-5 years, 6-9 years and 10-14 years. For example, how often the mother reads to the child is collected up to age 9. How often the child is brought to the museum is asked when the child is aged 3 and above. A comprehensive list of the items in HOME is provided in Appendix A of the codebook supplement in the CYA[7].

To compute the total raw score for the HOME, all non-binary items are recoded to binary items. Then, the scores on all items are summed together. If children had one or more unanswered items, those items were imputed with the average value, before the total HOME score was calculated. This analysis uses the a normed version of the total HOME score, the HOME percentile score. This percentile score indicates the level of home environment quality relative to children of the same age: a higher percentile score indicates higher home environment quality.

[7] https://www.nlsinfo.org/content/cohorts/nlsy79-children/other-documentation/codebook-supplement/appendix-home-sf-scales (last assessed: 5 June 2025)

1.B Child Family Income Mobility Estimates

The main text provides estimates for child wage income. Here, we provide estimates for child family income.

1.B.1 Construction of Child Family Income and Ranks

We construct a measure of child family income analogous to parent family income. In each survey round, this sums the following income received by the child and the child's spouse/partner: income from wages and salary, farm/business income, military income, child support, unemployment benefits, cash assistance such as Temporary Assistance for Needy Families (TANF) and Aid to Families with Dependent Children (AFDC), food stamps, Supplemental Security Income (SSI)/public assistance/welfare payments, veteran benefits, worker's compensation, disability payments and social security payments. Unlike parent family income, child family income does not include income from educational benefits for veterans and scholarships, fellowships and grants. Though some respondents reported receiving such income, these amounts were not collected. The family income is scaled to adjust for couples. In years when the spouse or partner wage income is collected, the family income is divided by 2. Otherwise, we take the family income value as it is.

We use the average family income of the child when the child is aged between 28 and 33 years. Like the child wage income, we use all available income reports during that age range. In addition, we only use the income during the years when the child indicates that he/she was not enrolled in school full-time. We assume that the child was not enrolled in school full-time in the past calendar year (income is reported retrospectively, for the previous calendar year) if he/she was not enrolled in school during the year of the interview. We excluded income observations when we could not determine whether the respondent was enrolled in school full-time or part-time.

Percentile ranks for child family income (average family income between 28 and 33 years) are computed within the children of the NLSY79 cross-section sample. We pool children from all birth cohorts for the ranking.

1.B.2 Estimates

Table 1.B.1: Income Summary Statistics (Deflated to Year 2018 using CPI-U)

	White-M	Black-M	Hispanic-M
Child family income			
Median	32,175	16,438	23,207
Mean	36,185	20,067	27,096
%Zeros	1.31	8.07	3.87
Number of children	1,070	830	569

Notes: This table presents income statistics for White, Black and Hispanic males. Child family income is the average family income when the child is aged between 28 and 33 years. Child family income has been scaled by size — in years when spouse/partner income was reported, the income was divided by 2. The % zeros row provides the proportion of the sample which reported zero values for child family income. All monetary values are measured in 2018 dollars (deflated using the CPI-U).

Table 1.B.2: Estimates of the Rank-Rank Slope (Child family income)

	White-M	Black-M	Hispanic-M
1.Baseline			
estimate	0.365	0.324	0.328
se	0.033	0.041	0.048
2.Educ			
estimate	0.191	0.147	0.194
se	0.034	0.041	0.052
3.Educ+Cog			
estimate	0.171	0.109	0.148
se	0.035	0.042	0.054
4.Educ+Cog+Noncog			
estimate	0.156	0.083	0.136
se	0.035	0.042	0.054
5.Educ+Cog+Noncog+Home			
estimate	0.141	0.063	0.057
se	0.036	0.044	0.058

Notes: This table presents the estimates and corresponding standard errors of the coefficient on the parent income rank in a regression of child family income rank on parent income rank. Panel 1 presents the baseline estimates. From Panel 2 onwards, additional variables (mediators) are included in the regression as controls. Educ denotes years of education, Cog denotes cognitive skill, Noncog denotes non-cognitive skill and Home denotes the quality of the home environment. The definition of the mediators is provided in the data section.

Table 1.B.3: Percentage of Intergenerational Transmission Explained by Mediators (Child family income)

	White-M	Black-M	Hispanic-M
Education	38.5	46.3	35.1
Cognitive	8.2	14.9	13.2
Non-Cognitive	7.0	11.1	3.3
Home	7.8	8.0	31.1
Total	61.5	80.4	82.7

Notes: This table presents the percentage of income transmission explained by each of the mediators: years of education, cognitive skill, non-cognitive skill and home environment quality. Total refers to the percentage explained by all mediators.

Table 1.B.4: Mediation Decomposition (Child family income)

	b	β	Education			Cognitive			Non-Cognitive			Home		
			γ	η	$\gamma \times \eta$	γ	η	$\gamma \times \eta$	γ	η	$\gamma \times \eta$	γ	η	$\gamma \times \eta$
White-M	.365	.141	3.546	.04	.14	.096	.31	.03	-.093	-.276	.026	.078	.364	.028
Black-M	.324	.063	4.78	.031	.15	.144	.336	.048	-.161	-.224	.036	.065	.402	.026
Hispanic-M	.328	.057	3.044	.038	.115	.126	.344	.043	-.049	-.221	.011	.211	.483	.102

Notes: For each mediator, this table presents (1) the return of the mediator to child family income rank (γ) and (2) the correlation between the mediator and parent income rank (η). b is the baseline rank-rank slope. β is the rank-rank slope after controlling for all mediators: years of education, cognitive skills, non-cognitive skills and home environment quality. Mediators are defined in the data section.

Table 1.B.5: Upward and Downward Mobility (Child family income)

	White-M	Black-M	Hispanic-M
Parent Income Rank 25	44.81	33.77	40.81
Parent Income Rank 75	63.05	49.95	57.21

Notes: This table presents the predicted child family income ranks of White, Black and Hispanic males at parent income rank 25 and parent income rank 75.

Table 1.B.6: Upward and Downward Mobility After Controlling for Mediators (Child family income)

	White-M	Black-M	Hispanic-M
Parent Income Rank 25			
Educ+Cog+Noncog+Home	38.28	27.40	37.81
Parent Income Rank 75			
Educ+Cog+Noncog+Home	69.01	62.84	66.54

Notes: This table presents the predicted child family income ranks of White, Black and Hispanic males at parent income rank 25 and parent income rank 75. To compute the predicted child income ranks at the 25th / 75th percentile of parent income, we use the 25th / 75th percentile of mediators in the sample of male children of the NLSY79 cross-section sample. Educ denotes years of education, Cog denotes cognitive skill, Noncog denotes non-cognitive skill and Home denotes the quality of the home environment.

1.C Estimates for Family Income Between Ages 11 and 22

In the main text, we use the average family income when the child is aged between 0 and 17. To contrast our findings to Chetty et al. (2020), we repeat our analysis using a similar measure of family income to them, the average family income when the child is aged between 11 and 22. Note that the sample we use here is different from that in the main text.

From Table 1.C.1, the baseline estimates of the rank-rank slope are lower than those in the main text (Table 1.3), because we are averaging income across fewer years.

Table 1.C.1: Estimates of the Rank-Rank Slope (Child wage income)

	White-M	Black-M	Hispanic-M
1.Baseline			
estimate	0.352	0.335	0.326
se	0.031	0.039	0.045
2.Educ			
estimate	0.197	0.168	0.212
se	0.032	0.038	0.048
3.Educ+Cog			
estimate	0.180	0.131	0.176
se	0.032	0.039	0.049
4.Educ+Cog+Noncog			
estimate	0.166	0.115	0.168
se	0.032	0.039	0.050
5.Educ+Cog+Noncog+Home			
estimate	0.160	0.105	0.098
se	0.033	0.041	0.054

Notes: This table presents the estimates and corresponding standard errors of the coefficient on the parent income rank in a regression of child wage income rank on parent income rank. Panel 1 presents the baseline estimates. From Panel 2 onwards, additional variables (mediators) are included in the regression as controls. Educ denotes years of education, Cog denotes cognitive skill, Noncog denotes non-cognitive skill and Home denotes the quality of the home environment. The definition of the mediators is provided in the data section.

Table 1.C.2: Estimates of the Rank-Rank Slope (Child family income)

	White-M	Black-M	Hispanic-M
1.Baseline			
estimate	0.352	0.300	0.332
se	0.032	0.038	0.045
2.Educ			
estimate	0.194	0.147	0.217
se	0.033	0.038	0.048
3.Educ+Cog			
estimate	0.175	0.114	0.178
se	0.034	0.039	0.050
4.Educ+Cog+Noncog			
estimate	0.163	0.096	0.169
se	0.034	0.039	0.050
5.Educ+Cog+Noncog+Home			
estimate	0.149	0.078	0.095
se	0.035	0.041	0.055

Notes: This table presents the estimates and corresponding standard errors of the coefficient on the parent income rank in a regression of child family income rank on parent income rank. Panel 1 presents the baseline estimates. From Panel 2 onwards, additional variables (mediators) are included in the regression as controls. Educ denotes years of education, Cog denotes cognitive skill, Noncog denotes non-cognitive skill and Home denotes the quality of the home environment.

Table 1.C.3: Percentage Mediated (Child wage income)

	White-M	Black-M	Hispanic-M
Education	35.8	42.3	30.5
Cognitive	7.8	15.0	10.3
Non-Cognitive	7.6	7.5	2.7
Home	3.3	3.9	26.4
Total	54.6	68.7	69.9

Notes: This table presents the percentage of income transmission explained by each of the mediators: years of education, cognitive skill, non-cognitive skill and home environment quality. Total refers to the percentage explained by all mediators.

Table 1.C.4: Percentage Mediated (Child family income)

	White-M	Black-M	Hispanic-M
Education	35.7	43.2	29.9
Cognitive	8.6	13.7	11.5
Non-Cognitive	6.2	9.4	2.6
Home	7.1	7.6	27.4
Total	57.7	73.9	71.3

Notes: This table presents the percentage of income transmission explained by each of the mediators: years of education, cognitive skill, non-cognitive skill and home environment quality. Total refers to the percentage explained by all mediators.

Table 1.C.5: Mediation Decomposition (Child wage income)

	b	β	Education			Cognitive			Non-Cognitive			Home		
			γ	η	$\gamma \times \eta$	γ	η	$\gamma \times \eta$	γ	η	$\gamma \times \eta$	γ	η	$\gamma \times \eta$
White-M	.352	.16	3.452	.037	.126	.098	.283	.028	-.108	-.25	.027	.033	.35	.011
Black-M	.335	.105	5.164	.027	.142	.168	.299	.05	-.14	-.179	.025	.035	.372	.013
Hispanic-M	.326	.098	3.015	.033	.099	.106	.318	.034	-.046	-.191	.009	.18	.477	.086

Notes: For each mediator, this table presents (1) the return of the mediator to child income rank (γ) and (2) the correlation between the mediator and parent income rank (η). b is the baseline rank-rank slope. β is the rank-rank slope after controlling for all mediators: years of education, cognitive skills, non-cognitive skills and home environment quality. Mediators are defined in the data section.

Table 1.C.6: Mediation Decomposition (Child family income)

	b	β	Education			Cognitive			Non-Cognitive			Home		
			γ	η	$\gamma \times \eta$	γ	η	$\gamma \times \eta$	γ	η	$\gamma \times \eta$	γ	η	$\gamma \times \eta$
White-M	.352	.149	3.44	.037	.126	.107	.283	.03	-.087	-.25	.022	.072	.35	.025
Black-M	.3	.078	4.724	.027	.13	.137	.299	.041	-.157	-.179	.028	.061	.372	.023
Hispanic-M	.332	.095	3.009	.033	.099	.12	.318	.038	-.045	-.191	.009	.191	.477	.091

Notes: For each mediator, this table presents (1) the return of the mediator to child family income rank (γ) and (2) the correlation between the mediator and parent income rank (η). b is the baseline rank-rank slope. β is the rank-rank slope after controlling for all mediators: years of education, cognitive skills, non-cognitive skills and home environment quality. Mediators are defined in the data section.

Table 1.C.7: Upward and Downward Mobility (Child wage income)

	White-M	Black-M	Hispanic-M
Parent Income Rank 25	50.76	38.06	46.61
Parent Income Rank 75	68.36	54.80	62.91

Notes: This table presents the predicted child wage income ranks of White, Black and Hispanic males at parent income rank 25 and parent income rank 75.

Table 1.C.8: Upward and Downward Mobility (Child family income)

	White-M	Black-M	Hispanic-M
Parent Income Rank 25	44.44	32.70	39.77
Parent Income Rank 75	62.03	47.69	56.38

Notes: This table presents the predicted child family income ranks of White, Black and Hispanic males at parent income rank 25 and parent income rank 75.

Table 1.C.9: Upward and Downward Mobility After Controlling for Mediators (Child wage income)

	White-M	Black-M	Hispanic-M
Parent Income Rank 25			
Educ+Cog+Noncog+Home	44.53	32.49	44.29
Parent Income Rank 75			
Educ+Cog+Noncog+Home	74.81	70.54	73.10

Notes: This table presents the predicted child wage income ranks of White, Black and Hispanic males at parent income rank 25 and parent income rank 75. To compute the predicted child income ranks at the 25th / 75th percentile of parent income, we use the 25th / 75th percentile of mediators in the sample of male children of the NLSY79 cross-section sample. Educ denotes years of education, Cog denotes cognitive skill, Noncog denotes non-cognitive skill and Home denotes the quality of the home environment.

Table 1.C.10: Upward and Downward Mobility After Controlling for Mediators (Child family income)

	White-M	Black-M	Hispanic-M
Parent Income Rank 25			
Educ+Cog+Noncog+Home	37.89	27.22	37.44
Parent Income Rank 75			
Educ+Cog+Noncog+Home	68.57	62.69	66.92

Notes: This table presents the predicted child family income ranks of White, Black and Hispanic males at parent income rank 25 and parent income rank 75. To compute the predicted child income ranks at the 25th / 75th percentile of parent income, we use the 25th / 75th percentile of mediators in the sample of male children of the NLSY79 cross-section sample. Educ denotes years of education, Cog denotes cognitive skill, Noncog denotes non-cognitive skill and Home denotes the quality of the home environment.

1.D Neal and Johnson (1996) Regression

Neal and Johnson (1996) find that the Black-White gap in log wages is almost eliminated after controlling for a measure of cognitive skill, the Armed Forces Qualification Test (AFQT). In this section, we assess whether we can obtain a similar result with this data. Our specification is different from Neal and Johnson (1996) in several ways. Instead of log wages, we use the child income rank. In addition, instead of the AFQT, the cognitive skill measure is the average percentile score on cognitive tests between ages 3 to 14. Moreover, we do not control for age, as there is only one measure of child income rank per individual (rank of the average income between ages 28 and 33). We regress the child income rank on race dummies and years of education or a quadratic function of cognitive skill.

Table 1.D.1 and Table 1.D.2 present the estimates for child wage income rank and child family income rank respectively. We find that the cognitive skill measure explains less than 50% of the Black-White gap in child income rank.

Table 1.D.1: Neal and Johnson (1996) regression: Dependent variable is child wage income rank

	(1)	(2)	(3)
Hispanic	-10.645***	-5.014***	-4.641***
	(1.410)	(1.306)	(1.402)
Black	-21.662***	-15.505***	-13.064***
	(1.257)	(1.175)	(1.315)
Years of Education		5.466***	
		(0.241)	
Cognitive Skill			0.656***
			(0.091)
$(\text{Cognitive Skill})^2$			-0.003***
			(0.001)
Observations	2469	2469	2469
R^2	0.108	0.262	0.193

Notes: This table displays the estimated coefficients in a regression of child wage income rank on a dummy for Hispanic, a dummy for Black, years of education or a quadratic function of cognitive skill. Cognitive skill is the average percentile score on cognitive tests between ages 3 and 14.

Table 1.D.2: Neal and Johnson (1996) regression: Dependent variable is child family income rank

	(1)	(2)	(3)
Hispanic	-10.931***	-5.380***	-4.879***
	(1.426)	(1.328)	(1.422)
Black	-20.898***	-14.828***	-12.350***
	(1.271)	(1.195)	(1.334)
Years of Education		5.389***	
		(0.245)	
Cognitive Skill			0.559***
			(0.093)
$(\text{Cognitive Skill})^2$			-0.002*
			(0.001)
Observations	2469	2469	2469
R^2	0.099	0.247	0.180

Notes: This table displays the estimated coefficients in a regression of child family income rank on a dummy for Hispanic, a dummy for Black, years of education or a quadratic function of cognitive skill. Cognitive skill is the average percentile score on cognitive tests between ages 3 and 14.

1.E Linear Spline in Parent Income Rank

In the main text, we assume that the relationship between the child income rank and the parent income rank is linear. Here, we relax this assumption and model the child income rank as a function of a linear spline in terms of parent income rank. The spline has 4 knots at parent income ranks 20, 40, 60 and 80 respectively. The specification is presented in Equation 1.E.1.

$$\begin{aligned} R_i^c = \beta_0 + \beta_1 R_i^p + \beta_2 \left(R_i^p - 20\right) \mathbb{1} \left(R_i^p \geq 20\right) + \beta_3 \left(R_i^p - 40\right) \mathbb{1} \left(R_i^p \geq 40\right) \\ + \beta_4 \left(R_i^p - 60\right) \mathbb{1} \left(R_i^p \geq 60\right) + \beta_5 \left(R_i^p - 80\right) \mathbb{1} \left(R_i^p \geq 80\right) + \epsilon_i \end{aligned} \tag{1.E.1}$$

After estimating the model, we use it to obtain predicted values of the child income rank at each parent income rank. Figure 1.E.1 presents the predicted values and the associated 95% confidence interval band for White, Black and Hispanic males. Estimates of the slopes at different points along the spline and the corresponding standard errors are presented in Table 1.E.1.

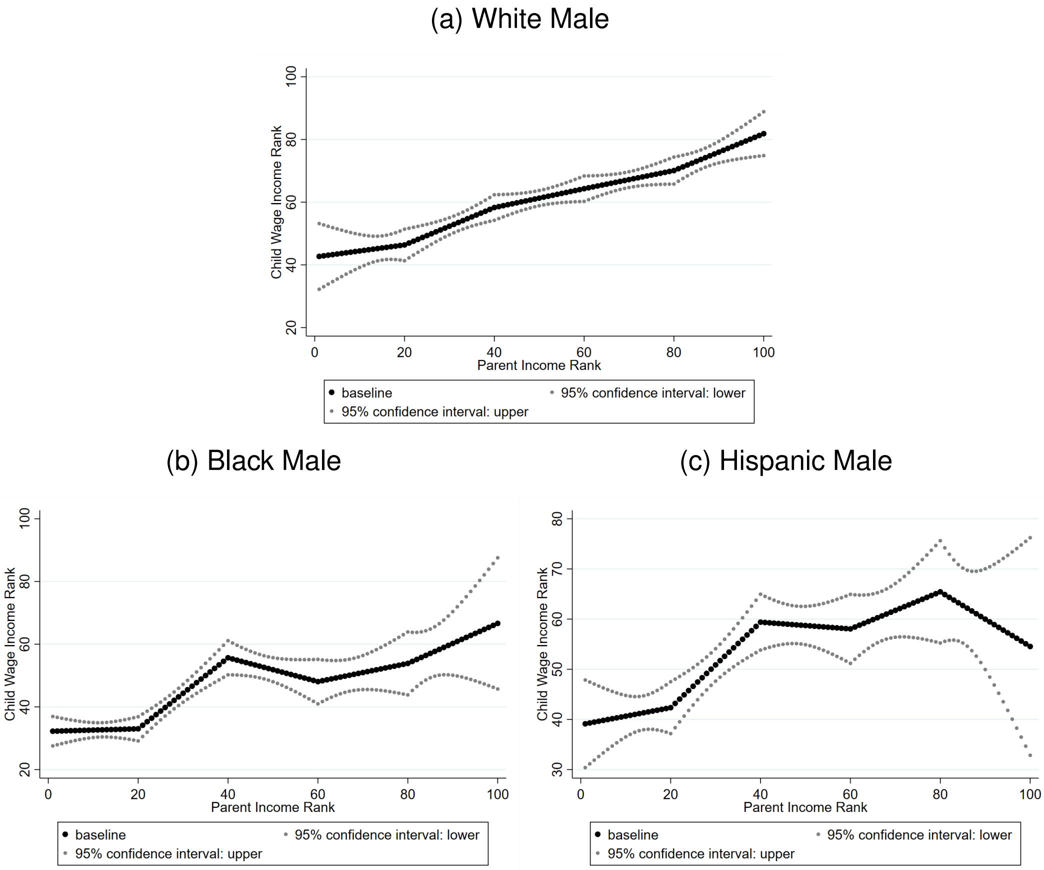
Table 1.E.1: Spline Slope Estimates

	1 to 20	20 to 40	40 to 60	60 to 80	80 to 100
White-M					
estimate	0.193	0.596	0.301	0.289	0.588
se	0.349	0.188	0.166	0.171	0.239
Black-M					
estimate	0.040	1.133	-0.380	0.290	0.638
se	0.194	0.192	0.256	0.346	0.658
Hispanic-M					
estimate	0.169	0.852	-0.067	0.369	-0.545
se	0.321	0.219	0.253	0.350	0.697

Notes: This table presents the estimates of the slopes along the spline from Equation 1.E.1 and the corresponding standard errors.

Using the spline, we assess whether the relationship between child income rank and parent income rank is non-linear, by testing whether the slopes between the knots are equal to each other. Specifically, we test whether the slopes at parent income ranks 1-20, 20-40, 40-60, 60-80 and 80-100 are the same. In practice, this is equivalent to testing whether the marginal slopes are jointly equal to 0. That is, $\beta_2 = \beta_3 = \beta_4 =$

Figure 1.E.1: Spline in Parent Income Rank



Notes: This figure presents the predicted values of the child wage income rank from a regression of the child wage income rank on a linear spline in parent income rank. The spline has 4 knots at parent income ranks 20, 40, 60 and 80 respectively.

$\beta_5 = 0$. We conduct this test for each of the races. The test statistics and associated p-values are presented in Table 1.E.2. We reject the equality of slopes for Black males and Hispanic males. This suggests that the relationship between child income rank and parent income rank is non-linear for these races.

Table 1.E.2: Test that Marginal Slopes are Jointly Equal to Zero

	Test Statistic	p-value
White-M	0.548	0.700
Black-M	4.725	0.001
Hispanic-M	2.563	0.038

Notes: This table presents the test statistics and associated p-values of a test of equality of slopes along the spline. Specifically, we test whether $\beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ in Equation 1.E.1.

Next, we control for the mediators by adding them to the right hand side of Equation 1.E.1. Specifically, we estimate Equation 1.E.2.

$$\begin{aligned} R_i^c = \beta_0 + \beta_1 R_i^p + \beta_2 \left(R_i^p - 20\right) \mathbb{1}\left(R_i^p \geq 20\right) + \beta_3 \left(R_i^p - 40\right) \mathbb{1}\left(R_i^p \geq 40\right) \\ + \beta_4 \left(R_i^p - 60\right) \mathbb{1}\left(R_i^p \geq 60\right) + \beta_5 \left(R_i^p - 80\right) \mathbb{1}\left(R_i^p \geq 80\right) \\ + c_1 education_i + c_2 cognitive_i + c_3 noncognitive_i + c_4 home_i + \epsilon_i \end{aligned} \quad (1.E.2)$$

After including the mediators, we are interested in how the slope of each segment of the spline (parent income rank 1-20, 20-40, 40-60, 60-80 and 80-100) changes. If the slope in a segment decreases significantly after controlling for the mediators, it suggests that within that segment, the mediators explain a significant proportion of income transmission. Figure 1.E.2 presents the predicted values from the baseline spline together with the predicted values after controlling for mediators. Table 1.E.3 presents the estimated slopes along the spline after accounting for the mediators. Table 1.E.4 displays the corresponding test statistics and p-values of a test of equality of marginal slopes. We reject equality of slopes for the Black males and Hispanic males.

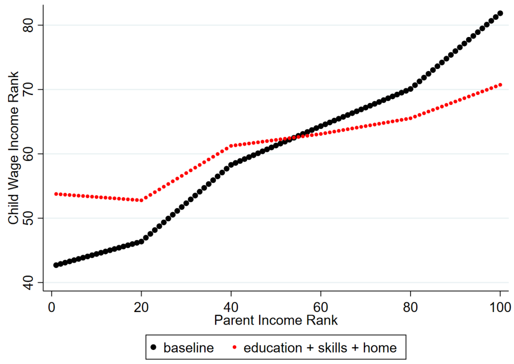
Table 1.E.3: Spline Slope Estimates: Controlling for Mediators

	1 to 20	20 to 40	40 to 60	60 to 80	80 to 100
White-M					
estimate	-0.052	0.423	0.093	0.122	0.260
se	0.328	0.179	0.157	0.161	0.226
Black-M					
estimate	-0.204	0.632	-0.198	-0.287	0.518
se	0.178	0.178	0.233	0.317	0.598
Hispanic-M					
estimate	0.036	0.494	-0.278	0.294	-1.188
se	0.311	0.218	0.247	0.338	0.677

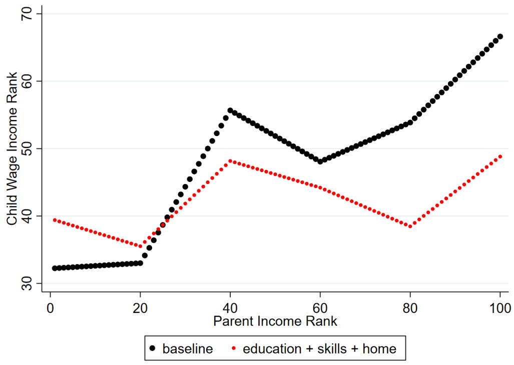
Notes: This table presents the estimates of the slopes along the spline from Equation 1.E.2 and the corresponding standard errors.

Figure 1.E.2: Spline in Parent Income Rank Controlling for Mediators

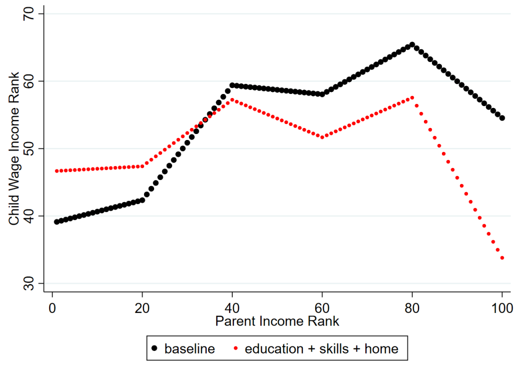
(a) White Male



(b) Black Male



(c) Hispanic Male



Notes: Each sub-figure displays two lines. The black line is the predicted value of the child wage income rank from a regression of the child wage income rank on a linear spline in parent income rank. The red line is the predicted value of the child wage income rank from a regression of the child wage income rank on a linear spline in parent family income and the mediators. Mediators are years of education, cognitive skill, non-cognitive skill and home environment quality. The spline has 4 knots at parent income ranks 20, 40, 60 and 80 respectively.

Table 1.E.4: Test that Marginal Slopes are Jointly Equal to Zero: Controlling for Mediators

	Test Statistic	p-value
White-M	0.559	0.693
Black-M	3.167	0.013
Hispanic-M	2.466	0.044

Notes: This table presents the test statistics and associated p-values of a test of equality of slopes along the spline. Specifically, we test whether $\beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ in Equation 1.E.2.

1.F Quantile Regressions

We estimate quantile regressions with linear splines in parent income rank. The splines have 4 knots at parent income ranks 20, 40, 60 and 80 respectively. Details of the quantile regression model are provided in Equations 1.F.1, 1.F.2 and 1.F.3.

$$\begin{aligned}M_n(\beta;\tau) = \frac{1}{n}\sum_{i=1}^{n}\rho_\tau(Y_i - \beta_0 - \beta_1 R_i^p - \beta_2\left(R_i^p - 20\right)\mathbb{1}\left(R_i^p \geq 20\right) \\ - \beta_3\left(R_i^p - 40\right)\mathbb{1}\left(R_i^p \geq 40\right) - \beta_4\left(R_i^p - 60\right)\mathbb{1}\left(R_i^p \geq 60\right) \\ - \beta_5\left(R_i^p - 80\right)\mathbb{1}\left(R_i^p \geq 80\right))\end{aligned} \quad \text{(1.F.1)}$$

$$\rho_\tau(x) = \begin{cases} -x(1-\tau) \text{ if } x < 0 \\ x\tau \text{ if } x \geq 0 \end{cases} \quad \text{(1.F.2)}$$

The estimator is the minimiser of the function:

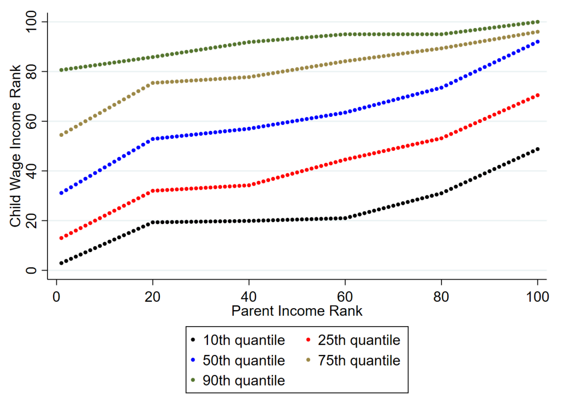
$$\hat{\beta}_\tau = \arg\min_\beta M_n(\beta;\tau) \quad \text{(1.F.3)}$$

Figure 1.F.1 displays the conditional quantile function of child wage income rank as a function of parent income rank, for the 10th, 25th, 50th, 75th and 90th quantiles. The conditional quantile regression functions provide an indication of the conditional distribution of child wage income rank given parent income rank. From Figure 1.F.1, for White males, the variance in child income rank decreases as parent income rank increases. In contrast, for Black males and Hispanic males, the variance in the child income rank remains similar as parent income rank increases.

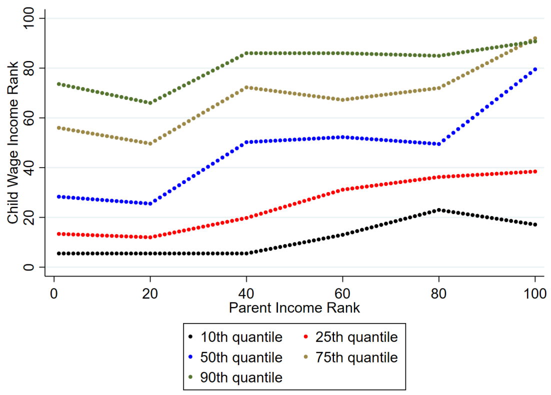
Next, we control for the mediators: years of education, cognitive skill (average percentile score on achievement tests between age 3 and 14), non-cognitive skill (average percentile on Behaviour Problems Index between ages 4 and 14) and home environment quality (average percentile score on home environment quality between ages 0 and 14). The function to minimise is now Equation 1.F.4.

Figure 1.F.1: Quantiles with Spline in Parent Income Rank

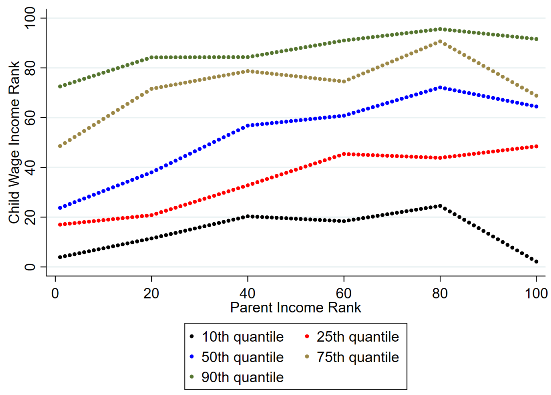
(a) White Male



(b) Black Male



(c) Hispanic Male



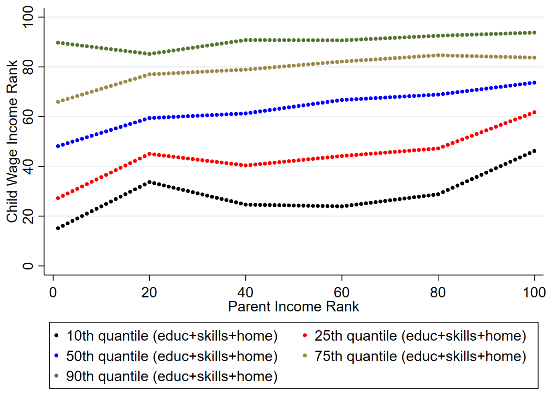
Notes: This figure presents the conditional quantile function of child wage income rank as a function of parent income rank. Each sub-figure displays the 10th, 25th, 50th, 75th and 90th quantiles.

$$
\begin{aligned}
M_n(\beta;\tau) = \frac{1}{n}\sum_{i=1}^{n}\rho_\tau(Y_i - \beta_0 - \beta_1 R_i^p - \beta_2\left(R_i^p - 20\right)\mathbb{1}\left(R_i^p \geq 20\right) \\
- \beta_3\left(R_i^p - 40\right)\mathbb{1}\left(R_i^p \geq 40\right) - \beta_4\left(R_i^p - 60\right)\mathbb{1}\left(R_i^p \geq 60\right) \\
- \beta_5\left(R_i^p - 80\right)\mathbb{1}\left(R_i^p \geq 80\right) \\
- c_1 education_i - c_2 cognitive_i - c_3 noncognitive_i - c_4 home_i) \quad \text{(1.F.4)}
\end{aligned}
$$

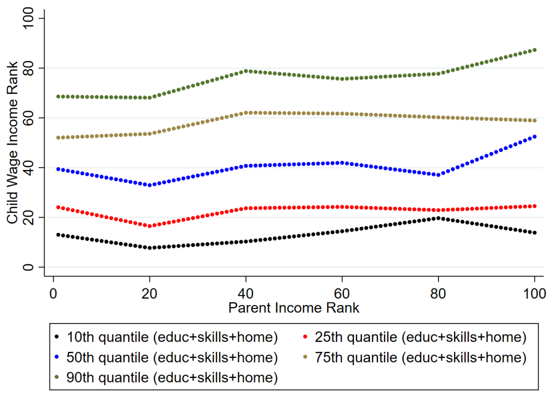
Figure 1.F.2 presents the predicted child income rank evaluated at the average values of the mediators within each racial group.

Figure 1.F.2: Quantiles with Spline in Parent Income Rank: Controlling for Mediators

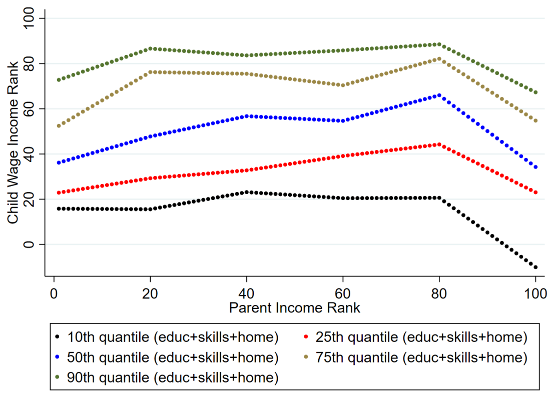
(a) White Male



(b) Black Male



(c) Hispanic Male



Notes: This figure presents the conditional quantile function of child wage income rank as a function of parent income rank. Each sub-figure displays the 10th, 25th, 50th, 75th and 90th quantiles. The quantile regressions include the mediators as independent variables: years of education, cognitive skill, non-cognitive skill and home environment quality. The figures present the predicted child income rank evaluated at the average values of the mediators within each racial group.

Chapter 2

Parental Beliefs and Parent Investments

2.1 Introduction

Parental investments are widely recognized as central determinants of children's skill development and long-term life outcomes (Attanasio, Cattan, & Meghir, 2022). Based on economic theory, investments are affected by parental resources, skill levels of children and preferences of parents (Caucutt & Lochner, 2020). A growing body of empirical research highlights that investment decisions are also influenced by parental beliefs, including perceptions of child skills (Dizon-Ross, 2019; Kinsler & Pavan, 2021) and perceptions about the returns to investment (Attanasio, Boneva, & Rauh, 2022; Boneva & Rauh, 2018; List, Pernaudet, & Suskind, 2021).

Since investments play a crucial role in child skill development, learning more about parental beliefs, including the accuracy of beliefs, how they interact with investments and how beliefs evolve, is important. First, suppose that parental beliefs affect investments and these beliefs are inaccurate. This implies that parents could be making mistakes in investment decisions. If the consequences are severe, policymakers may wish to correct these biases, possibly through providing parents with information. Second, even before attempting to shift beliefs, it is critical to determine whether beliefs can be changed, how they evolve, and what type of information may trigger belief changes.

This paper makes two primary contributions to the literature. First, it provides estimates of the causal impact of parental beliefs about child skill on parental investments in the context of a developed country. Second, it models the belief updating process, exploring how parents revise their beliefs over time. Additionally, the analysis examines heterogeneity by socio-economic status (SES), offering insight into whether belief-driven investment mechanisms differ between high and low SES families and whether this may partially explain the SES investment gap — rich parents invest more in children than poor parents (Caucutt, Lochner, & Park, 2017; Bolt et al., 2024; Carneiro, Reis, & Toppeta, 2024).

This study utilises longitudinal data from the National Longitudinal Survey of Youth (NLSY), which provides repeated measures of parental beliefs and investments over multiple years. Parental beliefs are estimated using a factor model based on mothers' reports of their child's (1) academic standing in class, (2) expected educational attainment and (3) future prospects. Parental investments are similarly estimated from a factor model based on measures of time and goods investments.

Descriptive analyses reveal several data patterns regarding parental beliefs and parent investments. Beliefs are strongly influenced by family background: around 43.8% of the variation in parental beliefs can be explained by time-invariant characteristics of the family. Furthermore, beliefs show strong persistence over time: around 66.9% of the variation in parental beliefs can be explained by child-specific time-invariant characteristics. Moreover, beliefs are not perfectly aligned with objective skill measures: the correlation between beliefs and skills is less than one. Yet the accuracy of beliefs rises as children grow older, indicating a learning process. Interestingly, there are socio-economic status (SES) differences in beliefs and investments. Low SES parents are more likely to hold inaccurate beliefs. Additionally, on average, high SES parents hold higher beliefs and invest more in their children, even after accounting for child skill levels.

To identify the casual effect of beliefs on investments, the analysis addresses key sources of endogeneity, including omitted variables bias — due to the presence of unobservable factors which jointly affect both beliefs and investments — and reverse causality between beliefs and investments: do beliefs cause investments, or investments cause beliefs? To address the endogeneity, several strategies are employed. First, child-specific time-invariant characteristics and persistent preferences of parents are controlled for, as they may jointly affect both beliefs and investments. Second, the panel data structure is utilised to obtain instruments for parental beliefs. Estimates indicate that there are no significant effects of beliefs on investments for both the high SES and low SES families.

To explore the evolution of parental beliefs, a belief updating model is estimated, where the current belief depends on lags of belief (representing the previous information which parents had about the child), child skill measures and a child fixed effect (capturing the child-specific time-invariant characteristics). Child skills include cognitive skills, non-cognitive skills (level of behaviour problems) and child health (rating of child's health). Results show that beliefs depend on the previous information which parents had about the child: both the first and second lags of belief are positively associated with current belief. This suggests that beliefs are strongly persistent. However, findings also indicate that parents adjust their beliefs when the skills of their child change — beliefs increase with cognitive skills and decline with worsening non-cognitive skills (rising behavioural problems). Since beliefs adjust when skills change, a possible strategy to shift beliefs could be to provide parents with additional information about their child's skills. When the belief updating model is estimated separately for the high SES and low SES, no evidence of SES differences in parent belief updating is found.

Various studies have established an important link between parental beliefs and parent investments (Attanasio & Kaufmann, 2014; Boneva & Rauh, 2018; Attanasio, Cunha, & Jervis, 2019; Attanasio, Boneva, & Rauh, 2019; Dizon-Ross, 2019; Kinsler & Pavan, 2021; List et al., 2021; Cunha, Elo, & Culhane, 2022; Conti, Giannola, & Toppeta, 2022; Page & Ruebeck, 2022). Most of these studies have focused on the correlation between beliefs and investments, instead of deriving the causal impact of beliefs on investments. This is because it is challenging to distinguish the effect of beliefs from other unobservable factors. Only a few studies have succeeded in estimating the causal impact through generating exogenous shifts in beliefs via experiments (Barrera-Osorio, Gonzalez, Lagos, & Deming, 2020; List et al., 2021; Dizon-Ross, 2019). Among these, those which estimate the causal effect of parental beliefs about child skill (Dizon-Ross (2019) and Barrera-Osorio et al. (2020)) have focused on developing countries and it is unclear whether results extend to developed countries. This paper contributes by providing the causal impact of these beliefs in a developed country, the United States. An identification strategy which relies on multiple reports of beliefs and investments being available is introduced, where the panel dimension of the data is exploited to obtain instruments for beliefs.

This paper builds on a growing literature relating to belief formation, learning and belief dynamics (Zafar, 2011; Sanders, 2012; Stinebrickner & Stinebrickner, 2014; Arcidiacono, Aucejo, Maurel, & Ransom, 2025). Of the few studies that estimate the determinants and evolution of beliefs or expectations (Koşar & O'Dea, 2023), only two studies have investigated the determinants of parental beliefs and the evolution of these beliefs: Kinsler and Pavan (2021) and Nicoletti, Sevilla, and Tonei (2022). Nicoletti et al. (2022) provide evidence that the gender-biased beliefs of parents about child skills are affected by information about child skills. Kinsler and Pavan (2021) highlight that parental beliefs about child skill depend on school-level or classroom-level skill. Most studies only have data on parental beliefs for up to two time points, making it difficult to study how beliefs are revised. Overcoming this obstacle with data on up to five waves of parental beliefs, this paper explores the evolution of beliefs, including the persistence of beliefs (whether beliefs depend on earlier lags of beliefs) and whether beliefs are related to child skills on multiple dimensions.

The rest of this paper is organised as follows. Section 2.2 introduces the data. Section 2.3 presents data patterns regarding parental beliefs and parent investments. Section 2.4 estimates the impact of parental beliefs on parent investments. Section 2.5 explores parent belief updating and Section 2.6 concludes the paper.

2.2 Data

Data sources are the National Longitudinal Survey of Youth 1979 (NLSY79) and the NLSY Child and Young Adult (CYA). The NLSY79 is a U.S. longitudinal survey of individuals who were aged 14-22 years in 1979. The CYA is a longitudinal survey of children of females in the NLSY79. By merging the NLSY79 and the CYA, information is obtained about the mother of the child, the family and the child. This includes household demographics (marital status, employment), mother's characteristics, mother's family background and children's characteristics.

This analysis focuses on children of females in the cross-section of the NLSY79. As of the latest survey round, a total of 5,819 children have been recorded. All biological children of the females in the NLSY79 are included — there can be multiple children in the same family.

Barring the attrition of females from the NLSY79 and attrition of children from the CYA, the CYA is representative of the population of children born to females from birth cohorts 1957 to 1964 in the United States. Though the cross-section of the NLSY79 is representative of the individuals in those birth cohorts, the children of the females in the cross-section may not be representative of their birth cohorts. There is selection into fertility and timing of children. Children in earlier birth cohorts were born to younger mothers and are generally negatively selected in terms of mother's characteristics — their mothers typically have lower years of education and cognitive skills. In addition, these children are more likely to grow up in single parent households.

2.2.1 Parental Belief

The parental belief is based on responses of the child's mother to 3 questions in the survey.

The first question is an expectation of the child's future educational attainment. Specifically, the mother is asked "How far do you think your child will go in school?". Possible options are: (1) Leave high school before graduation; (2) Graduate from high school; (3) Get some college or other training; (4) Graduate from college; (5) Get more than 4 years of college/further training after college; and (6) Something else. In practice, very few mothers choose option 1. Thus, option 1 and option 2 are merged together, and henceforth referred to as a single category: "Up to High School". In addition, option 6 (something else) is vague and few mothers chose it. Therefore, it is treated as a missing response. From this point onwards, the expectations of educational attainment have four categories: (1) Up to high school; (2) Some college or other training; (3) Graduate from college; and (4) More than college.

The second question is the rating of the child's academic standing in class. The child's mother indicates whether she perceives the child to be (1) Near the bottom of the class; (2) Below the middle; (3) In the middle; (4) Above the middle; or (5) One of the best students in class. Most children are rated at being in the middle of the class or above. This is a relative ranking rather than an absolute ranking, and there is evidence that relative rankings matter for child achievement (Kinsler & Pavan, 2021; Elsner, Isphording, & Zölitz, 2021).

The third question is the mother's rating of the child's future prospects. Possible options are: (1) Poor; (2) Fair; (3) Good; and (4) Excellent. Few mothers chose the category poor. Therefore, in this analysis, poor and fair are grouped together and treated as a single category.

Overall, mothers tend to hold good opinions about their children. More than 50% of children are expected to attain high school education and more than 50% of the children are rated as being above the middle of their class or one of the top in their class. Also, greater than 50% of children are rated as having excellent future prospects. Detailed proportions of the responses to each of the above measures are provided in Appendix 2.A.1.

Two of these measures are about expectations about the child's future. Therefore, aside from capturing the mother's perception about the child's skills, they may also reflect the mother's expectations about planned future investments and future shocks.

There is within-child variation in these measures. For example, in the sample of children who had beliefs reported three times between the ages of 9 and 14, (1) around 57% had at least one change in expected educational attainment; (2) around 61% had at least one change in rating of academic standing and (3) around 45% had at least one change in the rating of future prospects.

I assume that these three questions reflect a latent belief factor and use a factor model to extract the underlying factor, predicting a factor score for each child. I interpret the factor score which I extract from these three measures as representing the difference between the child's cognitive skill level and the average cognitive skills of children of the same age[1]. A factor model is useful because it adjusts for measurement error which is embedded in the responses to each of the 3 questions. The measure of expected educational attainment is used to link the factor over time, so it can be compared across ages of the child. Since latent factors have no natural scale (Anderson & Rubin, 1956), the belief factor is anchored to the years of education of the child at age 24 and above. This means that a 1 unit increase in the latent factor corresponds to a 1 unit increase in the conditional expectation of the years of education of the child at age 24 and above. To obtain the beliefs of parents in terms of the skill level, rather than the difference from average skills of children of the same age, the mean log cognitive skill at the corresponding age is added to the predicted factor score[2]. This "corrected" factor score

[1]It is an assumption that these measures capture beliefs about cognitive skills relative to children of the same age. It is possible that parental beliefs about the child's absolute performance are influenced by perceptions regarding relative performance. One could imagine that parents believe that their child will compete with peers for limited places in higher education institutions and job opportunities. Therefore, parental beliefs about absolute performance, such as the child's future prospects (including job opportunities) and educational attainment, may depend on the child's standing relative to his/her peers.

[2]Because of the nature of the measurements of belief, on average, the belief factor score will not rise in value when children grow older. In fact, if parents already rate their children in the highest possible categories of all belief measurements when children are young, the belief factor scores can only remain the same or decrease when children grow older. In contrast, the skill factor score will generally rise as children grow because children attain higher values on the measurements of skills, the raw scores on achievement tests. The following narrative could rationalise why the belief factor score does not grow with age while the skill factor score does. To parents, the skill level of the child is equivalent to a child-specific constant term plus an age-specific average skill, which is common to all children. Parents are aware of the age-specific skill value, but they do not observe the child-specific constant term. Consequently, belief measurements and the belief factor score only depend on the parent's perceptions about the child-

is referred to as the anchored belief factor score. This anchored belief factor score is the measure which is used in the analysis. The mean log cognitive skill is the average cognitive skill factor score (anchored) at the corresponding age. The construction of the cognitive skill factor score is discussed below.

Beliefs of agents are usually represented as a distribution (see Stinebrickner and Stinebrickner (2014) and Arcidiacono et al. (2025), for example). In this context, the belief of parents about the skill level of their child could be a distribution over the skill level of their child. The anchored belief factor score is taken as the mean of this distribution.

Appendix 2.G provides some evidence that the anchored belief factor score is informative — it predicts later skills and life outcomes of children. It also shows that on average, high SES mothers make more accurate predictions about the future educational attainment of their children.

2.2.2 Parent Investment

Following Cunha, Heckman, and Schennach (2010), a latent factor of parent investment is constructed from components of the Home Observation Measurement of the Environment (HOME), which proxy time and goods investments. Some components include how often the mother reads to the child, how many books the child has and how often the child is brought to the museum. A complete list of the component measures is provided in Appendix 2.A.2. The measure of how frequently the child is brought to the museum is used to link the factor over time, so that it can be compared across ages. The investment factor does not have a natural location or scale (Anderson & Rubin, 1956). It is standardised within the sample: its mean is 0 and it has a standard deviation of 1.

2.2.3 Child Skill Measures

The measures of cognitive skills, non-cognitive skills and child health used in the analysis are as follows.

Cognitive skill measures are Peabody Individual Achievement Tests (PIAT) in mathematics, reading recognition and reading comprehension, which were administered by

specific constant term. In this way, when child skills grow with age, the belief factor score may remain unaffected. If this narrative holds, then parental belief about the child's skill can be obtained by adding the age-specific skill to the belief factor score (belief about child-specific constant term). It is assumed that the age-specific skill is the average log skill factor score at the specific age.

the survey. These are available when the child is between the ages of 5 and 14. Higher raw scores indicate that the child has a higher skill level, while higher percentile scores indicate that the child performs better, compared to others of the same age. Generally, mothers do not observe these achievement test scores — I assume that these scores are correlated with the child's school grades, which were not collected by the survey, except during one particular year. In this year, I find that the school grades are correlated with the achievement test scores: national percentile ranks in the school achievement tests are correlated with the percentile ranks in the PIAT collected by the survey (see Appendix 2.A.5).

It is assumed that the raw scores on the achievement tests are equivalent to an underlying log cognitive skill of the child plus measurement error. To uncover the underlying skill, I estimate a factor model based on the raw scores on the three achievement tests and predict cognitive skill factors for the children. The raw score on the mathematics achievement test is used to link the latent factor over time, so it can be compared across ages. Latent factors have no natural scale (Anderson & Rubin, 1956). Therefore, the cognitive skill factor is anchored to years of education of the child at age 24 and above. This means that a 1 unit increase in the log cognitive skill corresponds to a 1 unit increase in the conditional expectation of the years of education of the child at age 24 and above. Note that because I also anchored beliefs to years of education, the skill factor is in the same units as the belief factor.

The non-cognitive skill measure is the Behaviour Problems Index (BPI), which is reported when the child is between the ages of 4 and 14. Based on the mother's responses to a set of questions about the child's behaviour, it is a measure of the level and frequency of problem behaviours exhibited by the child. This analysis uses the BPI percentile score. A higher percentile score on the BPI corresponds to a higher level of behaviour problems, relative to children of the same age. More information about the BPI is provided in Appendix 2.A.4. Since the BPI is based on the mother's observations of the child exhibiting specific problem behaviours, the BPI is interpreted as being distinct from the mother's belief about the child skill.

The health measure is a rating of the child's health provided by the child's mother, which is collected when the child is between the ages of 5 and 14. Specifically, the child's mother rates the child's health as one of the following: poor, fair, good or excellent.

Non-cognitive skill and health are included in the analysis as there is reason to expect that they are related to the belief factor score. This is because the component measures used to construct the belief include expectations about the child's future outcomes: expectations of child educational attainment and a rating of the child's future prospects. Studies indicate that non-cognitive skills and health are associated with children's outcomes in later life — non-cognitive skills predict long-term child outcomes (Borghans, Duckworth, Heckman, & Weel, 2008; Almlund et al., 2011), while child health affects skill development (Attanasio, Meghir, & Nix, 2020), education (Glewwe & Miguel, 2007) and earnings (Lundborg, Rooth, & Alex-Petersen, 2022).

2.2.4 Family Income

Family income is the total net family income of the family. It sums the following components received by respondent and spouse or partner: wage income, farm/business income, military income, unemployment compensation, Aid to Families with Dependent Children (AFDC), food stamps, Supplemental Security Income (SSI)/welfare, child support, alimony, educational benefits and/or scholarships, fellowships and grants, veteran benefits and income from other sources. From survey year 2002, income from worker's compensation, disability and social security is also included. In every survey year, the family income is top-coded. Truncated values are set equal to the average value of the top 2%.

2.2.5 Definition of Socio-economic Status

This paper will consistently refer to the concept of socio-economic status (SES). High SES families refer to those with above median level of family income[3] (defined based on average family income when the child is between 11 and 22). Otherwise, families are low SES.

2.2.6 Sample and Summary Statistics

As mentioned above, this analysis focuses on the 5,819 children of females in the cross-section sample of the NLSY79. Most children have siblings — only 550 children do not. Sample statistics of these children are provided in Table 2.1. Whenever there are less than 5,819 observations of certain variables, this is because of missing values.

[3]In the future, I could consider alternative definitions of SES based on mother's education and/or assets.

Table 2.1: Summary Statistics of Children of Females in Cross-Section Sample of NLSY79

	N	Mean	Median	Std. Dev.
Mother's Years of Education	5,819	13.42	12.00	2.57
Mother's AFQT Percentile	5,547	45.89	44.06	28.87
Mother's Age at Birth	5,819	26.79	26.00	6.06
White	5,819	0.77	1.00	0.42
Black	5,819	0.14	0.00	0.35
Hispanic	5,819	0.09	0.00	0.28
Male	5,819	0.52	1.00	0.50
Number of Siblings	5,819	1.87	2.00	1.36

Notes: This table presents summary statistics of the 5,819 children born to females in the cross-section sample of the NLSY79. AFQT refers to Armed Forces Qualification Test and it is a measure of cognitive skill.

On average, mothers attained 13.42 years of education (note that 12 years corresponds to graduating from high school, provided no grades were repeated). The average percentile score on the Armed Forces Qualification Test (AFQT), which is a measure of cognitive skill, is 45.89. The mean age of the mother at birth is 26.79 years. Around 77% of the sample are Whites, 14% are Blacks and 9% are Hispanics. 52% of the children are male and the rest are female. Within the sample, most children have 1 or more siblings.

To maximise the sample, when producing each data pattern, child-year observations are included whenever the key variables of interest are available. In this way, the sample used to produce each data pattern is distinct. The children used to produce each data pattern are a subset of the 5,819 children presented in Table 2.1.

2.3 Data Patterns

This section documents several patterns in the data regarding parental beliefs about the skill level of their child and parent investments.

First, parental beliefs are strongly influenced by family background: around 43.8% of the variation in parental beliefs is explained by family-specific time-invariant characteristics. Furthermore, beliefs do not vary much with time: around 66.9% of the variation in parental beliefs is explained by child-specific time-invariant characteristics.

These percentages are obtained from the R^2 of a regression of the belief factor score (anchored) on a family fixed effect (controlling for time-invariant characteristics of the family) or a child fixed effect (controlling for time-invariant characteristics of the child). Table 2.2 presents the R^2 of these regressions in column 2 and column 4 respectively. The table also presents the R^2 of regressions which additionally include the age of the child as a control.

Table 2.2: Dependent variable is Belief Factor Score (Anchored)

	(1)	(2)	(3)	(4)	(5)
Constant	10.537***	12.961***	9.939***	12.961***	9.955***
	(0.053)	(0.000)	(0.052)	(0.000)	(0.057)
Observations	12726	12726	12726	12726	12726
R^2	0.178	0.438	0.630	0.669	0.813
Number of Children	4827	4827	4827	4827	4827
Age of the Child	Yes	No	Yes	No	Yes
Family FE	No	Yes	Yes	No	No
Child FE	No	No	No	Yes	Yes

Notes: This table presents the constant term and the R^2 value in a regression of the belief factor score (anchored) on a constant term, age of the child and/or family FE or child FE. The belief factor score is explained in the data section. I use all available observations of children between ages 5 and 14.

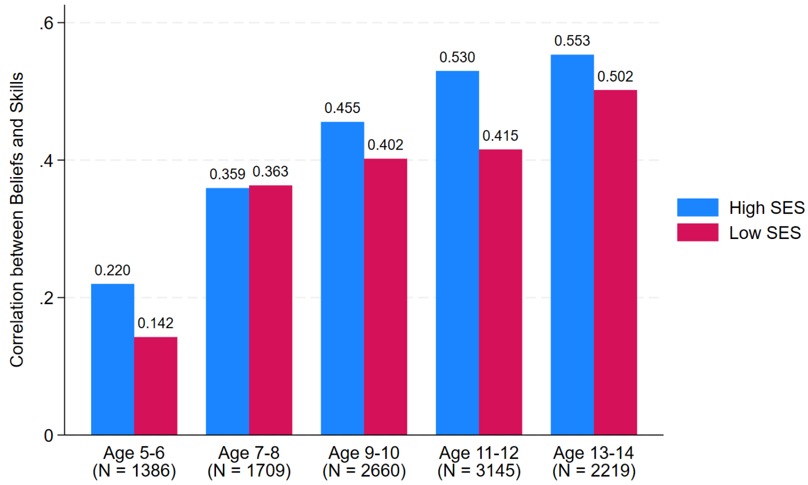
Second, parents may not accurately perceive the skill levels of their children. Figure 2.1 plots the correlation between parental beliefs (anchored) and skills (anchored). Both parental beliefs and skills are in the same units — they have been anchored to years of education of the child at age 24 and above. The correlation is less than 1, indicating that parents may not accurately perceive the skill levels of their child. In addition, parents learn about their child over time: the correlation between beliefs and skills is higher when children are older.

The potential impact of beliefs on investments will be explored later in this paper. If beliefs affect investments, and if beliefs are misaligned with actual skill, parents may be making mistakes in investment decisions.

Third, there are socio-economic status differences in terms of parental beliefs and parent investments. Low SES parents are more likely to hold inaccurate beliefs about the skills of their child: from Figure 2.1, the correlation between beliefs (anchored) and skills (anchored) is smaller for the low SES parents. At ages 9-10, 11-12 and 13-14, the SES differences in correlations are statistically significant at the 10% level, 1% level and the 10% level respectively.

Furthermore, high SES parents hold higher beliefs, even after accounting for child skill

Figure 2.1: Correlation between Belief Factor Score (Anchored) and Cognitive Skill Factor Score (Anchored) by SES



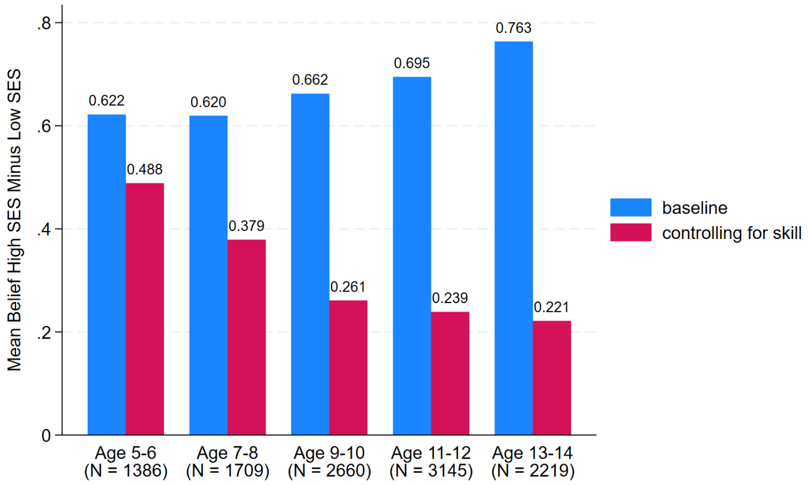
Notes: This figure presents the correlation between the belief factor score (anchored) and the cognitive skill factor score (anchored) at different ages of the child. High socio-economic status (SES) parents refer to parents who have above median family income (defined based on average family income when the child is aged between 11 and 22). Otherwise, parents are low SES.

levels. Figure 2.2 presents the average difference in the belief factor score (anchored) between the high SES and low SES families at different ages of the child. At all ages, the blue bar is positive, which indicates that on average, high SES parents hold higher beliefs. When I condition on the contemporaneous skill measures of the child (pink bars), the cognitive skill factor score and the percentile on the Behaviour Problems Index, the gap remains positive. This indicates that high SES parents hold higher beliefs, even after accounting for child skill levels[4]. Both the unconditional and conditional SES differences in average beliefs are statistically significant.

Moreover, high SES parents invest more in their children, even after accounting for children's skill levels. Figure 2.3 shows the difference between the average investment factor score of the high SES and the low SES parents. The blue bars are positive at all ages, which indicates that on average, the high SES invest more. When I condition on contemporaneous skill measures of the child (pink bars), the cognitive skill factor score and the percentile on the Behaviour Problems Index, the high SES still invest more than low SES. Both the unconditional and the conditional SES differences in mean investment are statistically significant.

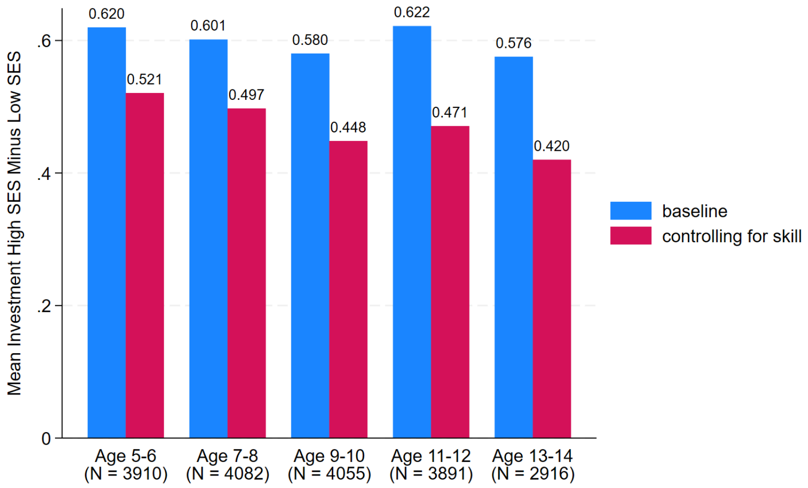
[4]Note that this does not imply that high SES parents over-estimate the skills of their children and low SES parents under-estimate the skills of their children. To determine whether this is true, the belief factor score needs to be compared to the skill factor score in terms of levels.

Figure 2.2: Average Difference in Belief Factor Score (Anchored) Between High SES and Low SES



Notes: This figure presents the mean difference in the belief factor score (anchored) between high SES and low SES parents at different ages of the child. The blue bar is the unconditional difference, while the pink bar is the difference conditional on contemporaneous child skill measures. Child skill measures are the cognitive skill factor score and the percentile on the Behaviour Problems Index. High SES parents refer to parents who have above median family income (defined based on average family income when the child is aged between 11 and 22). Otherwise, parents are low SES. Note that the blue bars do not correspond exactly to the difference in beliefs between the high SES and low SES in Figure 2.2 because the sample in this figure is different: to be included, contemporaneous skill measures must be non-missing.

Figure 2.3: Average Difference in Investment Factor Score (Standardised) Between High SES and Low SES



Notes: This figure presents the average difference in the investment factor score (standardised) between high SES and low SES parents at different ages of the child. The blue bar presents the unconditional difference in investment factor scores, while the pink bar is the difference conditional on contemporaneous skill measures. Child skill measures are the cognitive skill factor score and the percentile on the Behaviour Problems Index. High SES parents refer to parents who have above median family income (defined based on average family income when the child is aged between 11 and 22). Otherwise, parents are low SES.

2.4 What is the Impact of Parental Beliefs on Parent Investments?

This section shows that parental beliefs have an impact on parent investments for high socio-economic status (SES) parents, but not low SES parents.

2.4.1 Empirical Strategy

To estimate the impact of parental beliefs on parent investments, the investment factor score (standardised) is regressed on the belief factor score (standardised) and the coefficient on the belief factor score is examined. An important concern is endogeneity. Endogeneity may arise because of omitted variables bias: there may be unobserved factors which jointly affect both the beliefs and investments. Endogeneity may also arise from reverse causality between beliefs and investments: do beliefs cause investments, or do investments cause beliefs?

In the analysis, I employ several specifications and discuss the key assumptions required for each specification to produce a causal estimate. Each successive specification addresses additional endogeneity concerns from the previous one. The specifications are: (A) Ordinary least squares; (B) Child fixed effects; (C) Child fixed effects with lags of investment and (D) Child fixed effects with lags of investment, where I additionally treat parental beliefs as a potentially endogenous variable. Next, each of the specifications is described in greater detail. In the following, I_{jt} and μ_{jt} denote the investment factor score and belief factor score of child j at time t. y_{jt} denotes the log family income of child j at time t.

<u>A. Ordinary Least Squares</u>

$$\underbrace{I_{jt}}_{\text{investment}} = \alpha + \beta \underbrace{\mu_{jt}}_{\text{belief}} + \gamma_1 X_{jt} + \gamma_2 Z_j + \lambda y_{jt} + \eta_{jt} \tag{2.4.1}$$

This investment equation is motivated by the literature on skill production functions (Cunha et al., 2010; Attanasio et al., 2020; Agostinelli & Wiswall, 2025), which highlights that the investment policy function depends on the state variables such as the child skill level, family income and the skill level of parents. In general, the investment policy function could be non-linear. For simplicity, it is approximated as a linear function of the state variables.

In this setting, because parents do not observe the skill of the child, instead of the child's skill, the state variable is the parent's belief about the skill. Therefore, belief is included in the investment equation. Furthermore, the time-invariant variables can be viewed as proxies of the skill level of parents, or other time-invariant characteristics that influence the investment decision.

Before discussing how the specification addresses endogeneity, it is worth mentioning the assumptions inherent in this baseline model. One key assumption is that the relationship between beliefs and investments does not depend on the age of the child. This is made primarily to maximise the time periods of data which can be used in the dynamic panel data estimator which will be introduced later on. Given that the skill production technology could depend on the child's age, this could be a strong assumption. Another assumption is that there are homogeneous effects of beliefs on investments. Later in this analysis, heterogeneity by SES is explored[5]. This dimension of heterogeneity is chosen because this paper seeks to uncover whether there are SES differences which could possibly contribute towards the SES investment gap.

Addressing endogeneity: Several time-varying controls X_{jt} and time-invariant controls Z_j are included. If these are good proxies for the unobserved factors affecting both beliefs and investments, including them will address the concern of endogeneity due to omitted variables bias.

The time-varying controls include age of the child, dummies for region of residence (north, south, east, west), urban/rural residence, dummies for Standard Metropolitan Statistical Area (SMSA) status (not in SMSA, SMSA but not in central city, SMSA in central city, SMSA in unknown central city), an indicator which takes the value of 1 if the mother of the child is employed, an indicator which takes the value of 1 if the child's biological father lives with the child, the education of mother's spouse and mother's marital status at time t.

Non-time varying controls include race, gender of the child and the age of the mother at birth.

Key assumptions: Control variables X_{jt} and Z_j are good proxies for unobserved factors influencing both beliefs and investments

It is unlikely that this assumption is valid, so in this analysis, we will use this specification to gain an understanding of the baseline correlation between beliefs and investments.

[5]It is possible that there are heterogeneous effects of beliefs on investments along other dimensions like gender, race, parental age at birth or household structure. However, these dimensions of heterogeneity are not the focus of this paper.

B. Child Fixed Effects

$$\underbrace{I_{jt}}_{\text{investment}} = \alpha + \beta \underbrace{\mu_{jt}}_{\text{belief}} + \gamma X_{jt} + \lambda y_{jt} + \underbrace{\delta_j}_{\text{child FE}} + \eta_{jt} \tag{2.4.2}$$

Addressing endogeneity: Including child fixed effects reduces the chance of omitted variables bias because they account for the child-specific time-invariant characteristics which jointly influence both beliefs and investments.

Key assumptions: Strict exogeneity of inputs with respect to the error term, omitted inputs and their effects are constant with child age

The assumption of strict exogeneity may be invalid if there are time-varying shocks which are unobserved by the researcher. For example, an unobserved shock to the skill of the child could shift both beliefs and investments of parents. The assumption will also be invalid if lags of belief influence the current investment.

C. Child Fixed Effects + Lags of Investment

$$\begin{aligned}\underbrace{I_{jt}}_{\text{investment}} = \alpha + \beta \underbrace{\mu_{jt}}_{\text{belief}} + \psi_1 \underbrace{I_{j,t-1}}_{\text{lag investment}} + \psi_2 \underbrace{I_{j,t-2}}_{\text{second lag of investment}} + \psi_3 \underbrace{I_{j,t-3}}_{\text{third lag of investment}} \\ + \gamma X_{jt} + \lambda y_{jt} + \underbrace{\delta_j}_{\text{child FE}} + \eta_{jt}\end{aligned} \tag{2.4.3}$$

Addressing endogeneity: Including lags of investment reduces the likelihood of omitted variables bias, because the lags control for persistent preferences of parents which may jointly affect both beliefs and investments.

This model includes both a lag dependent variable and a child fixed effect. Given the short T panel, the coefficient on the lag dependent variable will be inconsistently estimated with standard fixed effects or first difference estimators. This happens because the lag dependent variable is correlated with the error term (Nickell, 1981). Therefore, to obtain consistent estimates, I use the Blundell and Bond (1998) dynamic panel data estimator (system GMM), which relies on moment conditions relating instruments to the level equation (Equation 2.4.3) and the differenced equation (Equation 2.4.4).

$$\Delta \underbrace{I_{jt}}_{\text{investment}} = \beta\Delta \underbrace{\mu_{jt}}_{\text{belief}} + \psi_1\Delta \underbrace{I_{j,t-1}}_{\text{lag investment}} + \psi_2\Delta \underbrace{I_{j,t-2}}_{\text{second lag of investment}} + \psi_3\Delta \underbrace{I_{j,t-3}}_{\text{third lag of investment}} + \gamma\Delta X_{jt} + \lambda\Delta y_{jt} + \Delta\eta_{jt} \quad (2.4.4)$$

Δ is the first difference operator

The instruments used in the differenced equation are: $\Delta\mu_{jt}$; ΔX_{jt}; Δy_{jt}; $I_{j,t-2}$, $I_{j,t-3}$, $I_{j,t-4}$

The instruments used in the level equations are: $\Delta I_{j,t-1}$

Key assumptions: Weak exogeneity of inputs, limited time dependence of the error term, predetermined initial conditions

Since the model is over-identified, the Sargan test can be used to assess the validity of over-identifying moment conditions (validity of subset of instruments), under the assumptions of homoscedasticity and no serial correlation (Arellano & Bond, 1991).

D. Child Fixed Effects + Lags of Investment + Belief and Income Endogeneous

Addressing endogeneity: To address endogeneity due to reverse causality between beliefs and investments, belief is treated as a potentially endogenous variable. As there may be unobserved shocks which affect both investments and family income, log family income is also treated as a potentially endogenous variable.

The equations of this model are identical to specification C. Like specification C, this model is also estimated with the Blundell and Bond (1998) dynamic panel data estimator. However, the moment conditions used in the estimation are slightly different. Since beliefs and log family income are treated as potentially endogenous variables, in the differenced equation, only beliefs and log family income from period $t-2$ and earlier are instruments.

The instruments used in the differenced equation are: $\mu_{j,t-2}$, $\mu_{j,t-3}$, $\mu_{j,t-4}$; ΔX_{jt}; $y_{j,t-2}$, $y_{j,t-3}$, $y_{j,t-4}$; $I_{j,t-2}$, $I_{j,t-3}$, $I_{j,t-4}$

The instruments used in the level equations are: $\Delta I_{j,t-1}$, $\Delta\mu_{j,t-2}$, $\Delta y_{j,t-2}$

This is the preferred specification[6] because it implements measures to address the endogeneity due to the omitted variables bias and reverse causality between beliefs and investments. It even handles the concern that family income may be endogenous.

[6] An alternative means to deal with endogeneity is to find an instrument for belief, aside from the lags of belief. No convincing instrument has been found.

2.4.2 Estimates

Table 2.3 presents estimates of the coefficient on the belief factor score (standardised) in a regression of the investment factor score (standardised) on belief factor score (standardised). Estimates from specifications A, B, C and D are presented in columns 1, 2, 3 and 4 respectively. The same sample of child-year observations is used to estimate all specifications. This implies that differences in results are driven by differences in the specifications themselves, instead of by sample composition.

Specifications A to C indicate that there is a positive relationship between parental beliefs and investments. The ordinary least squares estimate reveals that there is a strong correlation between the belief factor score and the investment factor score: a one standard deviation increase in the belief factor score is associated with a 0.270 standard deviation increase in the investment factor score. However, the preferred specification (specification D, which is in column 4) indicates that there is no significant effect of parental beliefs on investments.

This finding contrasts with Dizon-Ross (2019), who demonstrates that beliefs affect investments in Malawi. Why might this be the case? One possible reason is that the type of investments is different. Dizon-Ross (2019) focuses on investments relating to school education, whereas the parent investments in this paper relate to engagement with the child outside of the school environment.

The magnitude of the estimates (0.063-0.270) is similar to those in the literature. For example, Attanasio, Cunha, Jervis, and Toppeta (2025) obtain correlations of 0.025 - 0.080 between various dimensions of perceived returns to investment (standardised) and investment (standardised). Dizon-Ross (2019) presents correlations in the range of 0.021 and 0.069 between different types of investments and the parent's perception of the child's academic performance . Kinsler and Pavan (2021) document correlations in the range of -0.009 and 0.123 between various dimensions of standardised investment and an indicator that parents believe their child is above average, relative to children of a similar age.

Table 2.3: Dependent variable is Investment Factor Score (Standardised)

	(1) All	(2) All	(3) All	(4) All	(5) High SES	(6) Low SES
Belief Factor Score	0.270***	0.063***	0.074***	0.102	0.042	0.028
	(0.015)	(0.022)	(0.026)	(0.064)	(0.070)	(0.082)
Observations	5828	5828	5828	5828	3272	2556
Number of Children	3110	3110	3110	3110	1715	1395
Child FE	No	Yes	Yes	Yes	Yes	Yes
3 Lags of Investment	No	No	Yes	Yes	Yes	Yes
Belief Potentially Endogenous	No	No	No	Yes	Yes	Yes
Income Potentially Endogenous	No	No	No	Yes	Yes	Yes
Sargan Test P-value			0.456	0.115	0.029	0.176

Notes: This table reports coefficients on the belief factor score (standardised) in a regression of the investment factor score (standardised) on the belief factor score (standardised). Column 1 is the estimate from an OLS model. Column 2 is the estimate from the child fixed effects model. Columns 3-6 are estimated with the Blundell and Bond (1998) dynamic panel data estimator. The construction of the belief factor score and the investment factor score is provided in the data section. All regressions include time-varying controls which are listed in the description of specification A. Specification A also includes time-invariant controls: child gender, race and age of the mother at birth. Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). Standard errors are clustered at the child level in columns 1 and 2. One-step GMM standard errors are used in columns 3 to 6. High SES parents refer to parents who have above median family income (defined based on average family income when the child is aged between 11 and 22). Otherwise, parents are low SES. All available observations between age 5 and 14 are used.

Even if there is no effect of beliefs on investments on the average, there could be effects on subgroups. Could the relationship between beliefs and investments depend on socio-economic status (SES)? It may be that the investments of low SES are less reactive to beliefs. One reason is that although low SES parents desire to invest more when they hold higher beliefs, resource constraints may prevent them from doing so. Another reason is that the low SES may perceive that the return to investment does not vary significantly with the skill level of the child. This may happen, for instance, if low SES parents believe that investments have low contributions to skill growth[7] or if they derive lower value from child human capital[8]. To determine whether effects depend on SES, specification D is estimated separately for the high SES and the low SES. The corresponding estimates are presented in column 5 and column 6 of Table 2.3, respectively. Results indicate that beliefs do not have significant effects on investments for both the high SES and low SES families.

Various papers document significant correlations between beliefs and investments in developed countries (see, for example, Boneva and Rauh (2018), Attanasio, Boneva, and Rauh (2022) and Kinsler and Pavan (2021)). The results in this paper serve as a reminder that correlation may not imply causation, and the correlations may be driven by some underlying factors which affect both beliefs and investments.

Specifications C to D (columns 3 to 6) are estimated with the Blundell and Bond (1998) dynamic panel data estimator[9]. This estimator relies on the assumptions of the lack of serial dependence in the error term and exogeneity of the instruments. Even though three lags of investments are controlled for, there may still be persistent shocks (4 or more periods ago) which lead to autocorrelated errors. For instance, an unobserved shock to child skills. In addition, some instruments may be invalid. For example, the first lag of belief is treated as an instrument, but it may be endogenous if it has a direct effect on investment. This may happen if beliefs of parents are strongly persistent.

An indication of the validity of the instruments is provided by the Sargan test of over-identifying restrictions. Reassuringly, in columns 3, 4 and 6, the null hypothesis that the over-identifying restrictions are valid is not rejected[10]. However, the validity of the over-

[7] Some evidence that parents' perceived return to investment depends on SES is provided by Boneva and Rauh (2018) along with Attanasio, Boneva, and Rauh (2022).

[8] Caucutt et al. (2017) mention this as a possible reason for the investment gap between rich and poor families.

[9] When using an alternative dynamic panel data estimator by Arellano and Bond (1991), the impact of beliefs on investments is negative for the high SES families.

[10] When I include less than 3 lags of investments, I reject the null hypothesis that the over-identifying restrictions are valid in a Sargan test. In addition, in specification D, when I do not treat the log family

identifying restrictions is rejected in column 5, which may indicate that the instruments are invalid or that the model is misspecified.

The result that beliefs do not have significant effects on investments may not be robust. Versions of specification D with only a single lag of investment or two lags of investment indicate a different result, that beliefs have a positive and significant effect on investment. However, these specifications fail the test of over-identifying restrictions, suggesting that the instruments may be invalid or the model is misspecified. The finding that there are no significant effects of beliefs on investments for both the high SES and the low SES appears to be more robust. This holds even in models with a single lag or only two lags of investments.

To assess whether beliefs and investments might be related in a non-linear manner, I regress the investment factor score on a quadratic function of the belief factor score (see Appendix 2.C). There is no evidence of non-linearity.

2.5 Parent Belief Updating

In this section, I explore the belief updating process of parents. I provide evidence that parental beliefs depend on previous information which parents had about their child (captured by previous beliefs) and child skill measures.

2.5.1 Empirical Strategy

I hypothesise that the beliefs of parents depend on previous information which parents have about the child (captured by lags of belief), the child's skills[11] and child-specific time-invariant characteristics (captured by child fixed effects). Specifically, the belief factor score of child j at period t (μ_{jt}) is a linear function[12] of the belief factor score in period $t-1$ ($\mu_{j,t-1}$), belief factor score in period $t-2$ ($\mu_{j,t-2}$), child skill measures in period t and a child fixed effect δ_j. Child skill measures include the cognitive skill factor score ($cognitive$), the percentile score on the Behaviour Problems Index ($behaviour$) and a dummy which takes the value of 1 when the child's health rating is excellent

income as a potentially endogenous variable, I also reject the null hypothesis that the over-identifying restrictions are valid.

[11]Readers may be wondering why this paper focuses on skills, instead of other time-varying factors such as marital status, assets and family income. I have explored whether beliefs are sensitive to these variables and find that after controlling for child skills, beliefs may only be sensitive to changes in marital status.

[12]A linear belief updating model is also used in Stinebrickner and Stinebrickner (2014).

($healthexcellent$), and 0 otherwise. I include time-varying controls X_{jt} which are the child's age and log family income.

$$\mu_{jt} = \alpha + \beta_1 \mu_{j,t-1} + \beta_2 \mu_{j,t-2} + \beta_3 cognitive_{jt} + \\ + \beta_4 behaviour_{jt} + \beta_5 healthexcellent_{jt} + \gamma X_{jt} + \delta_j + \epsilon_{jt} \quad (2.5.1)$$

What motivates this specification of the belief updating model? There is evidence that parental beliefs are based on child skills (Kinsler & Pavan, 2021; Nicoletti et al., 2022). Although for simplicity, the belief factor score is assumed to reflect the cognitive skills of the child, it may also be influenced by other dimensions of human capital such as non-cognitive skills or health. The component measures used to construct the belief include the expected educational attainment of the child and a rating of the child's future prospects. Non-cognitive skills are predictive of long-term child outcomes (Borghans et al., 2008; Almlund et al., 2011) and may also produce cognitive skills (Cunha et al., 2010). Child health affects skill development (Attanasio et al., 2020), education (Glewwe & Miguel, 2007) and earnings (Lundborg et al., 2022). Therefore, it seems reasonable to expect that changes in non-cognitive skills or health could be related to changes in parental beliefs.

Furthermore, as mentioned in the data section, I interpret the belief factor score (anchored) as the mean of the parent's belief distribution over the skill level of their child. If the distribution over the skills is normal, this model (Equation 2.5.1) can be loosely motivated by Bayesian updating of the mean of a normal random variable, in which the posterior (updated) mean is a weighted average of the prior mean and the signal (new information received). Note that this model does not satisfy Bayesian updating because (1) the weights on the prior mean and the signal cannot change with time[13], as they do in Bayesian updating and (2) I include the second lag of belief[14][15].

Equation 2.5.1 contains both a lag dependent variable and a child fixed effect. Given the short T panel, estimating this equation via fixed effects or first differences will yield

[13]Ideally, the coefficients would be allowed to depend on time. Unfortunately, this is not possible with the Blundell and Bond (1998) estimator.

[14]When only the first lag of belief is included, the null hypothesis that the over-identifying restrictions are valid is rejected in a Sargan test.

[15]Readers may wonder why there are two lags of the dependent variable in the belief updating model, but three lags of the dependent variable in the investment model. This is because there are insufficient observations to employ a specification with three or more lags of beliefs, since there are few children with four consecutive reports of beliefs.

inconsistent estimates of the coefficient on the lag dependent variable. This is because the lag dependent variable is endogenous, as it is correlated with the error term (Nickell, 1981). To obtain consistent estimates, the Blundell and Bond (1998) dynamic panel data estimator[16], which is a system generalised method of moments estimator, is employed. This estimator uses moment conditions relating instruments to the level equation (Equation 2.5.1) and the differenced equation (Equation 2.5.2).

$$\Delta\mu_{jt} = \beta_1\Delta\mu_{j,t-1} + \beta_2\Delta\mu_{j,t-2} + \beta_3\Delta cognitive_{jt} + \\ + \beta_4\Delta behaviour_{jt} + \beta_5\Delta healthexcellent_{jt} + \gamma\Delta X_{jt} + \Delta\epsilon_{jt} \quad (2.5.2)$$

Δ is the first difference operator

Key assumptions: weak exogeneity of inputs, limited time dependence of the error term and predetermined initial conditions.

As mentioned before, the lag dependent variable is endogenous, because it is correlated with the error term. Consequently, only values of the dependent variable from time period $t-2$ and earlier are used as instruments in the differenced equation. Furthermore, one could imagine that the skill measures may also be endogenous: there may be unobserved shocks which jointly affect both the child skill measures and parental beliefs. Therefore, I also treat the skill measures as potentially endogenous variables. Like the lag dependent variable, only skill measures from time period $t-2$ and earlier are used as instruments in the differenced equation.

The instruments used in the differenced equation are: $\mu_{j,t-2}$, $\mu_{j,t-3}$, $\mu_{j,t-4}$; $cognitive_{j,t-2}$, $cognitive_{j,t-3}$, $cognitive_{j,t-4}$; $behaviour_{j,t-2}$, $behaviour_{j,t-3}$, $behaviour_{j,t-4}$;

$healthexcellent_{j,t-2}$, $healthexcellent_{j,t-3}$, $healthexcellent_{j,t-4}$; ΔX_{jt}

The instruments used in the level equations are: $\Delta\mu_{j,t-1}$, $\Delta cognitive_{j,t-1}$,

$\Delta behaviour_{j,t-1}$, $\Delta healthexcellent_{j,t-1}$

Since the model is over-identified, the validity of over-identifying restrictions can be assessed with the Sargan test.

[16]The sample size is too small to obtain meaningful results from an alternative dynamic panel data estimator by Arellano and Bond (1991).

2.5.2 Estimates

Table 2.4 presents the estimated coefficients on the lags of the belief factor score and the child skill measures. From column 1, there is evidence that parents use the previous information which they had about their child to form their current beliefs: both the first and second lags of belief are positive predictors of current belief. This also suggests that beliefs are strongly persistent.

In many papers, beliefs are modelled as normal random variables and they are assumed to be updated in a Bayesian way. In such a framework, the updated mean belief depends only on the prior mean belief and the signal (new information). In this paper, the fact that the belief factor score (proxy for belief) is predicted by the second lag of the belief factor score (proxy for earlier beliefs than the prior belief) may indicate that parents do not use Bayesian updating.

Second, the beliefs of parents change when the skills of their child change[17]. When cognitive skills improve, beliefs are revised upwards: a one unit increase in the cognitive skill is associated with a 0.811 increase in the belief factor score (anchored). A one percentile increase in the level of behaviour problems (decline in non-cognitive skill) is associated with a 0.014 decrease in the belief factor score (anchored). When children move from poor/fair/good health to excellent health, the belief factor score (anchored) rises by 0.196. The estimated coefficients on the cognitive skill and the level of behaviour problems are statistically significant. Overall, these findings suggest that parental beliefs can be adjusted by providing parents with information about child skills, which is consistent with experimental studies by Dizon-Ross (2019) and Barrera-Osorio et al. (2020).

To the best of my knowledge, this is the first estimate of a parent belief updating model. Consequently, it is difficult to compare the magnitude of the results to other studies. The closest estimates are by Kinsler and Pavan (2021), who find that within-child changes in beliefs are correlated with within-child changes in child math skill (standardised) and the value of the correlation is 0.143. This number is lower than the estimated correlation between cognitive skill and beliefs in this paper, but the scale and type of beliefs and skills are different.

To assess whether there are SES differences in belief updating, the same model is estimated separately for the high SES (column 2) and the low SES (column 3). The

[17]Ideally, to obtain the casual effect of the skill on belief, one would have an exogenous shock to parent information about child skills. This could happen with an information intervention experiment.

Table 2.4: Dependent variable is Belief Factor Score (Anchored)

	(1) All	(2) High SES	(3) Low SES
Lag Belief	0.260***	0.246***	0.288***
	(0.052)	(0.060)	(0.057)
Second Lag Belief	0.096**	0.093*	0.088*
	(0.040)	(0.049)	(0.050)
Cognitive Skill	0.811***	0.785***	0.856***
	(0.082)	(0.092)	(0.081)
Bad Behaviour	-0.014***	-0.014***	-0.014***
	(0.004)	(0.004)	(0.004)
Health: excellent	0.196	-0.142	0.212
	(0.219)	(0.241)	(0.222)
Observations	2842	1731	1111
Number of Children	1954	1153	801
Child FE	Yes	Yes	Yes
Skills Potentially Endogenous	Yes	Yes	Yes
Sargan Test P-value	0.238	0.621	0.817

Notes: This table presents the estimated coefficients in a regression of the belief factor score (anchored) on the lag belief factor score (anchored), second lag of the belief factor score (anchored) and child skill measures. Child skill measures are the cognitive skill factor score (anchored), percentile score on the Behaviour Problems Index and a dummy which takes the value of 1 when the child's health is rated as excellent (takes value of 0 when health rating is poor/fair/good). The construction of the belief factor score is described in the data section. The model is estimated with the Blundell and Bond (1998) dynamic panel data estimator. All regressions control for the age of the child and log family income. Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). One-step GMM standard errors are used. I use all available observations of the child between ages 5 and 14.

estimated coefficients are similar, aside from the coefficient on the excellent health dummy which is not precisely estimated. Overall, there is no evidence of SES differences in parent belief updating. This contrasts with the finding by Dizon-Ross (2019) that low SES parents are more responsive to information.

Note that ideally, the belief updating model would include the information which parents receive about their child. Because this information is unobserved, child skill measures are included are substitutes, under the assumption that they are correlated with the information. It is possible that the skill measures fail to capture differences in the quality and/or frequency of information received by the high SES and low SES. Therefore, even though this study does not provide evidence that parents update differently by SES, it does not mean that SES differences do not exist.

The validity of the estimator depends on the assumptions of the lack of serial dependence in the error term and the exogeneity of instruments. Despite including two lags of beliefs, there may still be persistent shocks (3 or more periods ago) which lead to the error term being autocorrelated. Furthermore, some instruments may be invalid. For example, lag skill measures are used as instruments, but these could be endogenous if there are unobserved shocks to the child which affect both beliefs and skills across multiple time periods. For example, unobserved shocks to other dimensions of child skills which influence beliefs, cognitive and/or non-cognitive skills.

In each of the columns, I fail to reject the null hypothesis that the over-identifying restrictions are valid in a Sargan test, which provides some assurance that at least a subset of the instruments is valid.

The number of observations in this section is significantly lower than the number of observations in the analysis of the causal impact of beliefs on investments. This is because there are fewer children who have beliefs reported in three consecutive time periods. I note that the sample of children used in this analysis is selected — children with beliefs reported in three consecutive survey rounds are more likely to have mothers with higher cognitive skill (measured by the Armed Forces Qualification Test) and higher education.

In models where only a single lag of belief is included, the patterns of the findings are similar. Both the lag belief factor score and cognitive skill are positive predictors of belief. Bad behaviour is a negative predictor of belief, while being in excellent health is not predictive of belief. This suggests that these findings may be robust to changes in model specification.

In Appendix 2.E, to assess whether there may be non-linear relationships between beliefs and skill measures, I model beliefs as a quadratic function of child skill measures. I find evidence of non-linearity: the coefficient on the quadratic term is significant. Furthermore, I investigate whether beliefs may depend on lag skill measures, even after accounting for current skill measures. Results indicate that the lag cognitive skill of the child also predicts belief[18].

2.6 Conclusion

Using the National Longitudinal Survey of Youth, this paper examines how parental beliefs about their child's skill level influence parental investments, and how these beliefs are updated over time. Attention is paid to differences across socio-economic status (SES) groups, which may help explain the SES investment gap (Caucutt et al., 2017; Bolt et al., 2024; Carneiro et al., 2024) – the tendency of high SES parents to invest more in their children than low SES parents.

Parental beliefs are estimated using a factor model derived from mothers' assessments of: (1) their child's academic standing in class, (2) expected future educational attainment, and (3) future prospects. Several empirical patterns regarding parental beliefs and parent investments are documented. First, parental beliefs are shaped by family background and show strong persistence over time. Second, parental beliefs about the skill levels of their child may be inaccurate, though accuracy improves as children age. Third, SES differences are evident in both beliefs and investments: high SES parents tend to hold higher beliefs and invest more, even after accounting for child skill levels.

Next, to estimate the causal effect of parental beliefs on investments, parent investments are regressed on parental beliefs. The analysis controls for unobserved time-invariant characteristics of children and persistent preferences of parents, which may confound the effects of beliefs on investments. In addition, the panel structure of the data is exploited to obtain instruments for parental beliefs. Results indicate that there are no significant effects of beliefs on investments for both the high SES families and the low SES families.

Finally, parent belief updating is explored. Estimates of a belief updating model indicate that beliefs are influenced by past beliefs and by children's skill levels. Current

[18] Since both the current cognitive skill and lag cognitive skill predict belief, it may be that beliefs are responsive to changes in skills, instead of the level of skill. This can be explored in the future.

beliefs are positively associated with lag beliefs, suggesting that parents use previous information which they had about their child to form current beliefs. When cognitive skills improve, beliefs rise. When non-cognitive skills worsen, beliefs fall. There is no evidence of SES differences in belief updating.

Overall, findings suggest that parents may hold inaccurate beliefs about the skill levels of their child. In addition, providing parents with more information about the skill level of their child could correct these misperceptions. Given that beliefs do not affect investments, correcting these inaccurate beliefs may not be a key concern for policymakers, since they are unlikely to result in investment mistakes.

This paper highlights that there may be important links between parental beliefs, investments and skills. Building on these findings, a framework which unifies these components is introduced in the next paper: a dynamic model of parent investments which incorporates parental beliefs about skill and the evolution of child skills. In the model, as there is feedback between beliefs and investments, early beliefs can affect later beliefs, investments and skills. The model is used to quantify the contribution of parental beliefs about child skill to the SES skill gap.

Appendix to Chapter 2

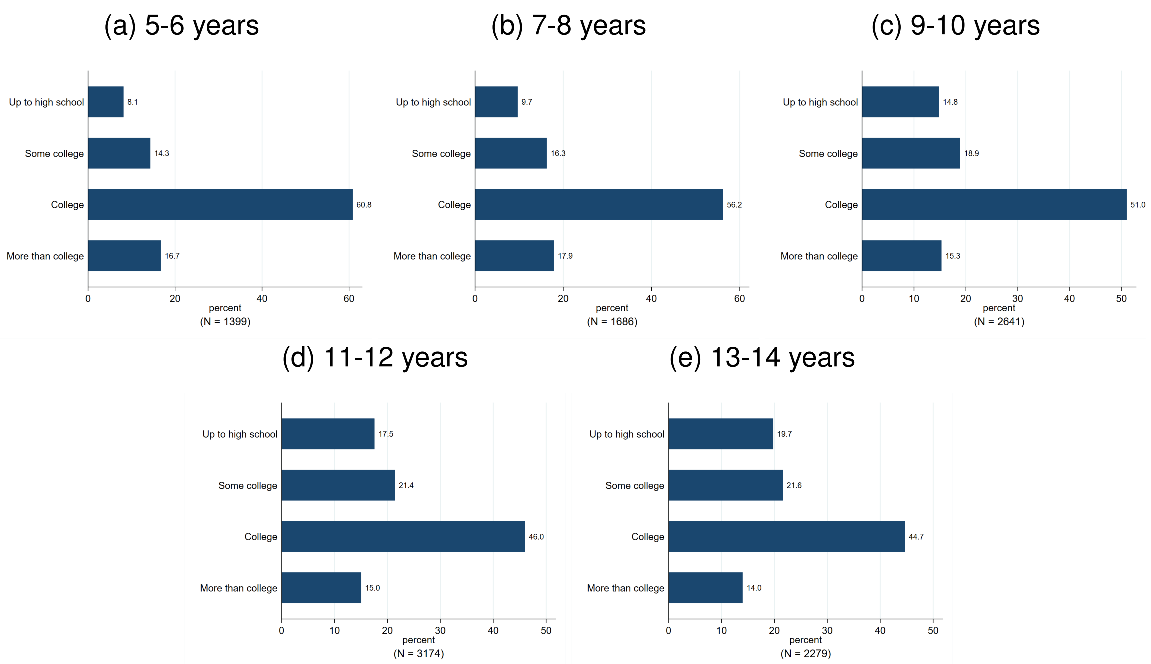
2.A Data Appendix

2.A.1 Measures of Latent Belief

Expectations of Educational Attainment

Figure 2.A.1 presents the mother's responses to the question "How far do you think your child will go in school?".

Figure 2.A.1: Parent Expectations of Educational Attainment

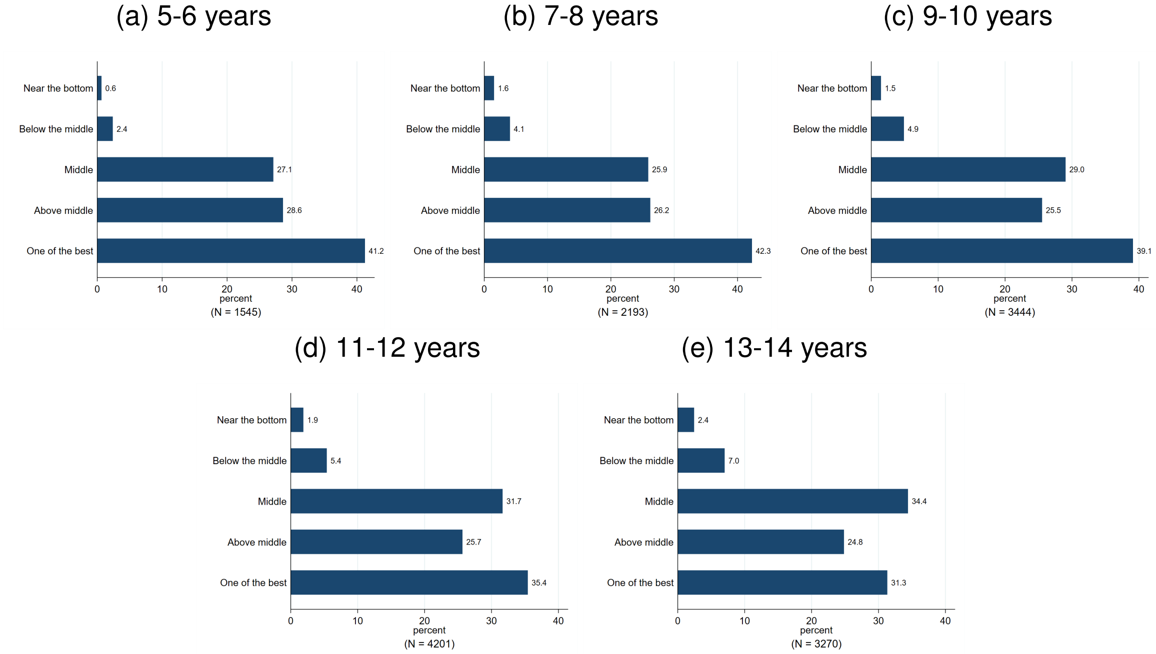


Notes: These figures present the percentage of responses in each of the 4 categories: up to high school, some college, college and more than college.

Rating of Child's Academic Standing in Class

Figure 2.A.2 presents the mother's rating of the child's academic standing in class.

Figure 2.A.2: Rating of Child's Academic Standing in Class

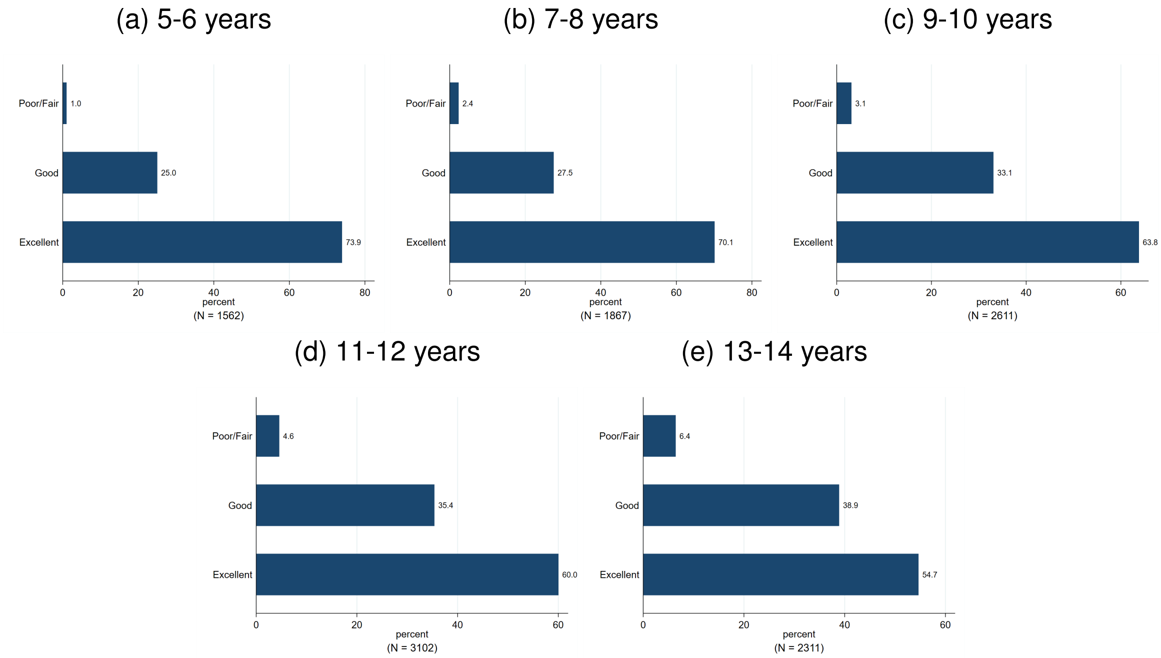


Notes: These figures present the percentage of responses in each of the 5 categories: near the bottom, below the middle, middle, above the middle and one of the best.

Rating of Child's Future Prospects

Figure 2.A.3 presents the mother's rating of the child's future prospects.

Figure 2.A.3: Rating of the Child's Future Prospects



Notes: These figures present the percentage of responses in each of the 3 categories: poor/fair, good, excellent. In the survey, poor and fair were separate options. As few mothers chose the option of poor, in this figure, the poor and fair options have been merged into a single category.

2.A.2 Measures of Parent Investments

Components of the Home Observation Measurement of the Environment (HOME) used to construct the parent investment factor are provided here. The components used to construct the investment factor score are age-specific. For example, how often the mother reads to the child is only used up to age 9. Summary statistics of these measures are presented in Table 2.A.1 and Table 2.A.2.

Table 2.A.1: Summary Statistics of Investment Measures Part 1

Measure	High SES Mean	Low SES Mean	Min	Max
How Often Child Goes to Museum Age 3-5	2.269	1.968	1	5
How Often Goes on Outing Age 3-5	3.686	3.372	1	5
Number of Books Child Has Age 3-5	3.935	3.692	1	4
How Often Mom Reads to Child Age 3-5	5.041	4.461	1	6
Number of Magazines at Home Age 3-5	3.428	2.606	1	5
Child has CD Player Age 3-5	0.874	0.701	0	1
How Often Child Eats with Mom/Dad Age 3-5	2.248	2.719	1	6
How Often Child Goes to Museum Age 6-9	2.392	2.145	1	5
Number of Books Child Has Age 6-9	3.959	3.749	1	4
How Often Mom Reads to Child Age 6-9	4.247	3.925	1	6
Child Has Musical Instrument Age 6-9	0.613	0.375	0	1
Family Subscribes to Daily Newspapers Age 6-9	0.576	0.396	0	1
Family Encourages Hobbies Age 6-9	0.944	0.888	0	1
Child Has Special Lessons Age 6-9	0.745	0.466	0	1
How Often Child Goes to Theatre Age 6-9	2.033	1.768	0	5
How Often Child Attends Family Gatherings Age 6-9	3.759	3.495	1	5
How Often Child Eats with Mom/Dad Age 6-9	2.400	3.094	1	6

Notes: This table presents summary statistics of the investment measures used to estimate the investment factor model. High SES parents refer to parents who have above median family income (defined based on average family income when the child is aged between 11 and 22). Otherwise, parents are low SES.

Table 2.A.2: Summary Statistics of Investment Measures Part 2

Measure	High SES Mean	Low SES Mean	Min	Max
How Often Child Goes to Museum Age 10-14	2.231	2.054	1	5
Number of Books Child Has Age 10-14	3.757	3.373	1	4
Child Has Musical Instrument Age 10-14	0.722	0.451	0	1
Family Subscribes to Daily Newspapers Age 10-14	0.585	0.371	0	1
Family Encourages Hobbies Age 10-14	0.959	0.912	0	1
Child Has Special Lessons Age 10-14	0.814	0.569	0	1
How Often Child Goes to Theatre Age 10-14	2.056	1.768	1	5
How Often Child Attends Family Gatherings Age 10-14	3.529	3.336	1	5
How Often Child Eats with Mom/Dad Age 10-14	2.546	3.458	1	6

Notes: This table presents summary statistics of the investment measures used to estimate the investment factor model. High SES parents refer to parents who have above median family income (defined based on average family income when the child is aged between 11 and 22). Otherwise, parents are low SES.

2.A.3 Cognitive Skill Measure: Achievement Tests

The Peabody Individual Achievement Test (PIAT) is a well-established assessment of children's academic achievement. The NLSY Child and Young Adult (CYA) administered three subtests of the PIAT to children aged between 5 and 14 years: mathematics, reading recognition and reading comprehension.

2.A.4 Non-Cognitive Skill Measure: Behaviour Problems Index

The Behaviour Problems Index (BPI) is based on the mother's responses to questions regarding the frequency at which the child exhibits problem behaviours. These questions are asked when the child is aged between 4 and 14. The BPI contains subscales in the following topics: anxious/depressed, antisocial, dependent, headstrong, hyperactive and peer conflicts/withdrawn. Some examples of items included in the subscales are: breaks things deliberately, cheats or tells lies, has sudden changes in mood or feeling, is disobedient at school, has trouble getting along with teachers, is disobedient at home and is not liked by other children. Possible responses are (1) often true, (2) sometimes true and (3) not true. The BPI items are age-specific. For example, being disobedient in school is only asked when children are older than 5 years while breaks things deliberately is only asked when children are younger than 12. A complete list of the items in each of the BPI subscales is provided in Appendix D of the codebook supplement of the CYA[19].

To form the overall BPI score, items on the subscales are recoded to binary variables before being added together. The responses “often" and “sometimes true" are coded as 1, while “not true" is coded as 0. This analysis relies on a normed version of the overall BPI score, the percentile score of the BPI. The BPI percentile provides an indication of the level of behaviour problems relative to children of the same age: a higher percentile score indicates worse behaviour.

2.A.5 Correlation Between PIAT Scores and School Transcript

Generally, the NLSY Child and Young Adult (CYA) does not collect information on school grades. There is one exception: school transcript information was collected in the year 1995-1996 for around 3,000 children. Transcripts provide the national percentile ranks

[19] https://www.nlsinfo.org/content/cohorts/nlsy79-children/other-documentation/codebook-supplement/appendix-d-behavior-problems (last assessed: 5 June 2025)

of the child in vocabulary, reading comprehension, reading, language and math. Table 2.A.3 presents evidence that these percentile scores are correlated with PIAT percentile scores.

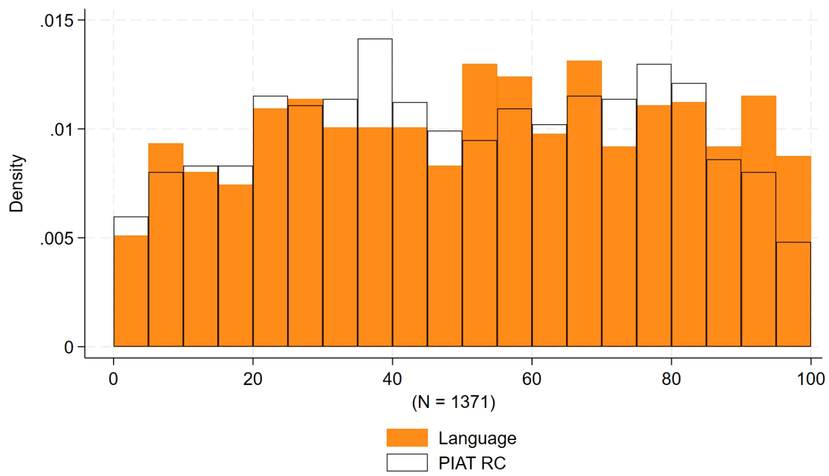
Table 2.A.3: Dependent Variable is Percentile on National Achievement Test

	(1) Vocabulary	(2) Reading Comprehension	(3) Reading	(4) Language	(5) Math
PIAT Mathematics Percentile	0.199***	0.185***	0.235***	0.270***	0.483***
	(0.037)	(0.041)	(0.035)	(0.039)	(0.036)
PIAT Reading Comprehension Percentile	0.262***	0.314***	0.265***	0.203***	0.134***
	(0.034)	(0.039)	(0.033)	(0.038)	(0.035)
PIAT Reading Recognition Percentile	0.390***	0.357***	0.398***	0.382***	0.184***
	(0.037)	(0.040)	(0.036)	(0.045)	(0.041)
Observations	997	881	1026	911	1077
R^2	0.525	0.546	0.579	0.529	0.445

Notes: This table presents the estimated coefficients on the percentile scores on Peabody Individual Achievement Tests (PIAT), in a regression of the respective dependent variable on these tests. There are several dependent variables: percent ranking on national achievement tests in vocabulary, reading comprehension, reading, language and math. Standard errors are clustered at the family level in all regressions. Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

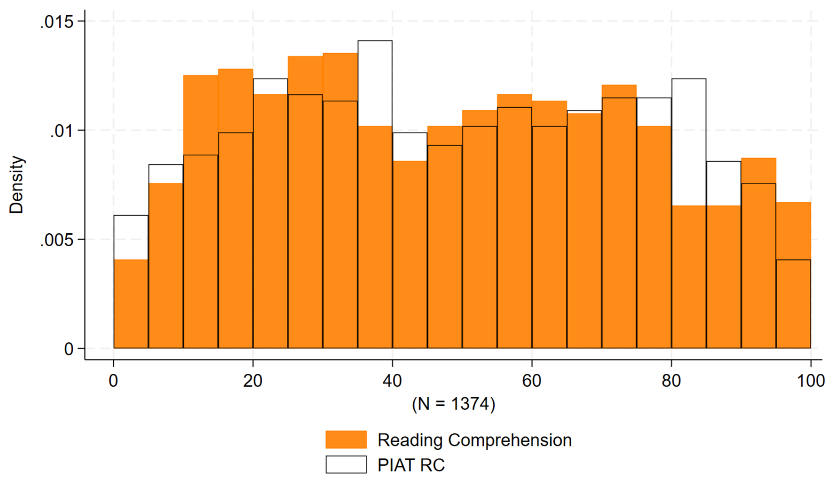
Several histograms (Figures 2.A.4, 2.A.5, 2.A.6, 2.A.7 and 2.A.8) showcase the extent of overlap between the percentile on a national achievement test and a relevant PIAT percentile score.

Figure 2.A.4: Language Test



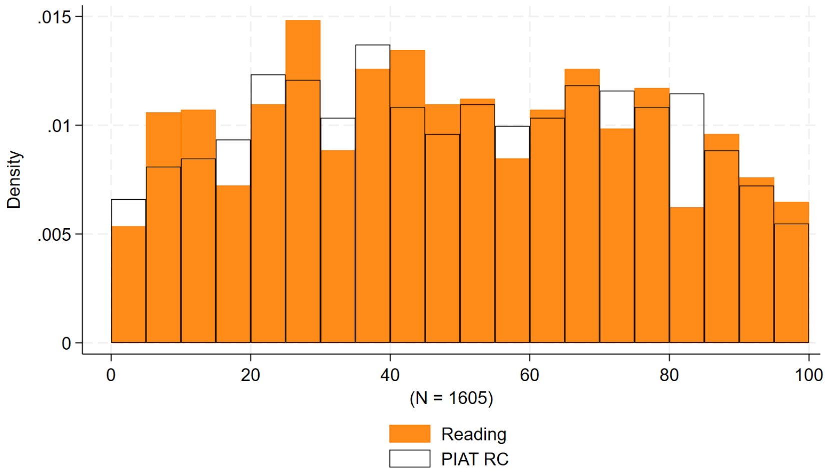
Notes: This figure displays the overlap between the national percentile on the language test and the percentile on the PIAT reading comprehension test (PIAT RC).

Figure 2.A.5: Reading Comprehension Test



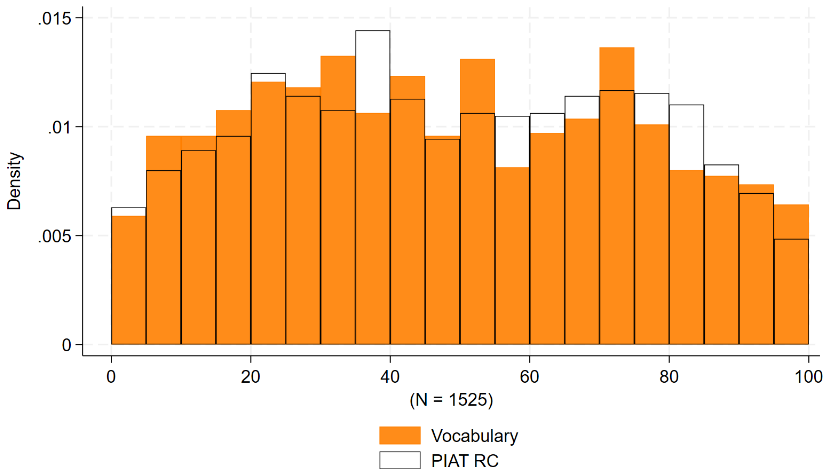
Notes: This figure displays the overlap between the national percentile on the reading comprehension test and the percentile on the PIAT reading comprehension test (PIAT RC).

Figure 2.A.6: Reading Test Total Score



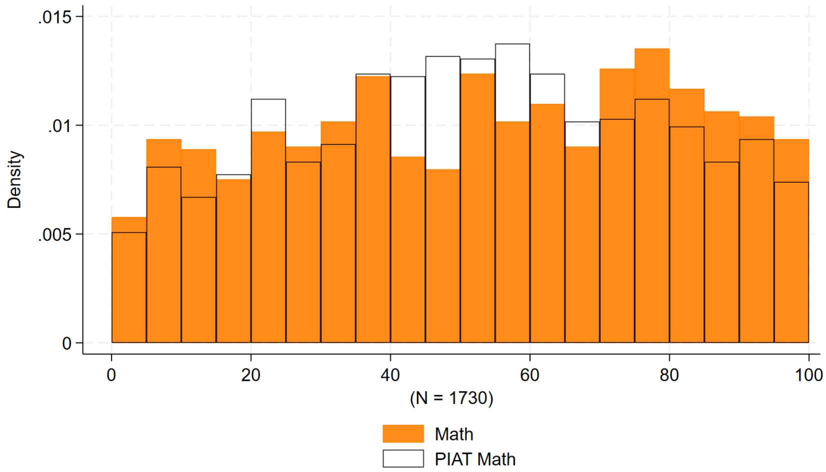
Notes: This figure displays the overlap between the national percentile on the reading test total score and the percentile on the PIAT reading comprehension test (PIAT RC).

Figure 2.A.7: Vocabulary Test



Notes: This figure displays the overlap between the national percentile on the vocabulary test and the percentile on the PIAT reading comprehension test (PIAT RC).

Figure 2.A.8: Math Test



Notes: This figure displays the overlap between the national percentile on the mathematics test and the percentile on the PIAT mathematics test (PIAT Math).

2.B Factor Models

2.B.1 Factor Model for Investment

The measurements of the investment latent factor are binary variables or ordered discrete variables.

For the binary variables, it is assumed that the measurement m_{ijt} depends on an unobserved continuous latent variable m^*_{ijt} according to a threshold model in the following way.

$$m_{ijt} = \begin{cases} 0 \text{ if } m^*_{ijt} \leq 0 \\ 1 \text{ if } m^*_{ijt} > 0 \end{cases}$$

For ordered discrete variables, it is assumed that each of these categorical measurements m_{ijt} depend on an unobserved continuous latent variable m^*_{ijt} according to a threshold model. As examples, when the measurements take 4 or 5 values, the threshold models are the following.

Suppose measurement m_{ijt} takes 4 values. It is assumed to be mapped to unobserved latent continuous variable m^*_{ijt} according to a threshold model in this way:

$$m_{ijt} = \begin{cases} 0 \text{ if } m^*_{ijt} \leq \tau_{1,jt} \\ 1 \text{ if } m^*_{ijt} \in (\tau_{1,jt}, \tau_{2,jt}] \\ 2 \text{ if } m^*_{ijt} \in (\tau_{2,jt}, \tau_{3,jt}] \\ 3 \text{ if } m^*_{ijt} > \tau_{3,jt} \end{cases}$$

Suppose measurement m_{ijt} takes 5 values. It is assumed to be mapped to unobserved latent continuous variable m^*_{ijt} according to a threshold model in this way:

$$m_{ijt} = \begin{cases} 0 \text{ if } m^*_{ijt} \leq \tau_{1,jt} \\ 1 \text{ if } m^*_{ijt} \in (\tau_{1,jt}, \tau_{2,jt}] \\ 2 \text{ if } m^*_{ijt} \in (\tau_{2,jt}, \tau_{3,jt}] \\ 3 \text{ if } m^*_{ijt} \in (\tau_{3,jt}, \tau_{4,jt}] \\ 4 \text{ if } m^*_{ijt} > \tau_{4,jt} \end{cases}$$

The investment latent factor in period t is represented by I_t. Note that for illustrative purposes, the factor model is presented with only three measurements of the latent factor of investment. In reality, the factor model is estimated with more than three measurements in each time period.

Let m_1^*, m_2^* and m_3^* represent unobserved error-ridden latent measurements of the latent factor. The unobserved measurements are assumed to be related to the latent variable in a linear way. For example, measurement m_{1t}^* is related to the latent factor I_t by an intercept $\alpha_{m1,t}$ and a factor loading $\lambda_{m1,t}$.

Equations for the latent factor model of investment for $t = 1, 2, 3, 4, 5$

$$m_{1t}^* = \alpha_{m1,t} + \lambda_{m1,t} I_t + \epsilon_{m1,t} \tag{2.B.1}$$

$$m_{2t}^* = \alpha_{m2,t} + \lambda_{m2,t} I_t + \epsilon_{m2,t} \tag{2.B.2}$$

$$m_{3t}^* = \alpha_{m3,t} + \lambda_{m3,t} I_t + \epsilon_{m3,t} \tag{2.B.3}$$

Measurement errors in the equations are assumed to be normally distributed, uncorrelated across measurements and uncorrelated across time. One of the measurements — in this case, the frequency of bringing the child to the museum — is used to link the factor over time. This is achieved by restricting the value of the first threshold and the factor loading of this measurement to be the same in all time periods. The intercepts and the factor loadings of other measurements are unrestricted.

Table 2.B.1 presents the signal-noise ratio of the measurements of investment.

2.B.2 Factor Model for Cognitive Skills

Latent log cognitive skill at period t is represented by $\ln \theta_t$. Q represents an adult outcome, which in this case is the years of education at age 24 and above. m_1, m_2 and m_3 represent error-ridden measurements of the latent log cognitive skill. These are raw scores on the achievement tests in mathematics, reading recognition and reading comprehension, which are continuous variables. The measurements are assumed to be related to the latent variable in a linear way. For example, measurement m_{1t} is related to the latent factor $\ln \theta_t$ by an intercept $\alpha_{m1,t}$ and a factor loading $\lambda_{m1,t}$.

Equations for the latent factor model of skills for $t = 1, 2, 3, 4, 5$

$$Q = \alpha + \ln \theta_T + \eta_Q \tag{2.B.4}$$

$$m_{1t} = \alpha_{m1,t} + \lambda_{m1,t} \ln \theta_t + \epsilon_{m1,t} \tag{2.B.5}$$

$$m_{2t} = \alpha_{m2,t} + \lambda_{m2,t} \ln \theta_t + \epsilon_{m2,t} \tag{2.B.6}$$

$$m_{3t} = \alpha_{m3,t} + \lambda_{m3,t} \ln \theta_t + \epsilon_{m3,t} \tag{2.B.7}$$

Table 2.B.1: Percentage of Total Variance in Investment Measurements Due to Signal and Noise

Measure	% Signal	% Noise
How Often Child Goes to Museum Age 3-5	0.437	0.563
How Often Goes on Outing Age 3-5	0.202	0.798
Number of Books Child Has Age 3-5	0.573	0.427
How Often Mom Reads to Child Age 3-5	0.368	0.632
Number of Magazines at Home Age 3-5	0.277	0.723
Child has CD Player Age 3-5	0.356	0.644
How Often Child Eats with Mom/Dad Age 3-5	0.018	0.982
How Often Child Goes to Museum Age 6-9	0.437	0.225
Number of Books Child Has Age 6-9	0.257	0.743
How Often Mom Reads to Child Age 6-9	0.110	0.890
Child Has Musical Instrument Age 6-9	0.174	0.826
Family Subscribes to Daily Newspapers Age 6-9	0.136	0.864
Family Encourages Hobbies Age 6-9	0.201	0.799
Child Has Special Lessons Age 6-9	0.331	0.669
How Often Child Goes to Theatre Age 6-9	0.536	0.464
How Often Child Attends Family Gatherings Age 6-9	0.042	0.958
How Often Child Eats with Mom/Dad Age 6-9	0.025	0.975
How Often Child Goes to Museum Age 10-14	0.437	0.563
Number of Books Child Has Age 10-14	0.248	0.752
Child Has Musical Instrument Age 10-14	0.244	0.756
Family Subscribes to Daily Newspapers Age 10-14	0.117	0.883
Family Encourages Hobbies Age 10-14	0.202	0.798
Child Has Special Lessons Age 10-14	0.292	0.708
How Often Child Goes to Theatre Age 10-14	0.580	0.420
How Often Child Attends Family Gatherings Age 10-14	0.039	0.961
How Often Child Eats with Mom/Dad Age 10-14	0.040	0.960

Notes: This table presents the signal-noise ratio of each of the measurements used to estimate the investment factor model.

Measurement errors in the equations are assumed to be normally distributed, uncorrelated across measurements and uncorrelated over time.

The first equation, which relates adult outcome Q to the latent log skill in the last period $\ln\theta_T$, is the anchoring equation. In this equation, the loading on $\ln\theta_T$ is set equal to 1 so that a 1 unit increase in the latent log skills $\ln\theta_T$ corresponds to a 1 unit increase in the conditional expectation of the adult outcome Q, which is years of education at age 24 and above.

One of the measurements (in this case, the raw score on PIAT mathematics) is used to link the factor over time. This is achieved by restricting the intercept and the factor loading of this measurement to be the same in all time periods. In this way, the growth in the measurement is informative about the change in the level of the log latent skill as the child ages. The intercepts and the factor loadings of the other measurements are unrestricted.

2.B.3 Factor Model for Beliefs

The measurements of the belief latent factor are ordered discrete variables. It is assumed that each of these categorical measurements depend on an unobserved continuous latent variable according to a threshold model. There are three measurements. The first measurement (rating of the child's future prospects) takes 3 possible values, the second measurement (expected educational attainment) takes 4 possible values and the third measurement (rating of the child's academic standing in class) takes 5 possible values.

Suppose measurement m_{ijt} takes 3 values. We assume that it is mapped to unobserved latent continuous variable m^*_{ijt} according to a threshold model in this way:

$$m_{ijt} = \begin{cases} 0 \text{ if } m^*_{ijt} \leq \tau_{1,jt} \\ 1 \text{ if } m^*_{ijt} \in (\tau_{1,jt}, \tau_{2,jt}] \\ 2 \text{ if } m^*_{ijt} > \tau_{2,jt} \end{cases}$$

Suppose measurement m_{ijt} takes 4 values. We assume that it is mapped to unobserved latent continuous variable m^*_{ijt} according to a threshold model in this way:

$$m_{ijt} = \begin{cases} 0 \text{ if } m^*_{ijt} \leq \tau_{1,jt} \\ 1 \text{ if } m^*_{ijt} \in (\tau_{1,jt}, \tau_{2,jt}] \\ 2 \text{ if } m^*_{ijt} \in (\tau_{2,jt}, \tau_{3,jt}] \\ 3 \text{ if } m^*_{ijt} > \tau_{3,jt} \end{cases}$$

Suppose measurement m_{ijt} takes 5 values. We assume that it is mapped to unobserved latent continuous variable m^*_{ijt} according to a threshold model in this way:

$$m_{ijt} = \begin{cases} 0 \text{ if } m^*_{ijt} \leq \tau_{1,jt} \\ 1 \text{ if } m^*_{ijt} \in (\tau_{1,jt}, \tau_{2,jt}] \\ 2 \text{ if } m^*_{ijt} \in (\tau_{2,jt}, \tau_{3,jt}] \\ 3 \text{ if } m^*_{ijt} \in (\tau_{3,jt}, \tau_{4,jt}] \\ 4 \text{ if } m^*_{ijt} > \tau_{4,jt} \end{cases}$$

The belief latent factor in period t is represented by μ_t. Q represents an adult outcome, which in this case is the years of education at age 24 and above. m^*_1, m^*_2 and m^*_3 represent unobserved error-ridden latent measurements of the latent factor. They are the latent measurements corresponding to the rating of the child's future prospects, the expected educational attainment and the rating of the child's academic standing in class. These unobserved latent measurements are assumed to be related to the latent variable in a linear way. For example, measurement m^*_{1t} is related to the latent factor μ_t by an intercept $\alpha_{m1,t}$ and a factor loading $\lambda_{m1,t}$.

Equations for the latent factor model of belief for $t = 1, 2, 3, 4, 5$

$$Q = \alpha + \mu_T + \nu_Q \tag{2.B.8}$$

$$m^*_{1t} = \alpha_{m1,t} + \lambda_{m1,t}\mu_t + \epsilon_{m1,t} \tag{2.B.9}$$

$$m^*_{2t} = \alpha_{m2,t} + \lambda_{m2,t}\mu_t + \epsilon_{m2,t} \tag{2.B.10}$$

$$m^*_{3t} = \alpha_{m3,t} + \lambda_{m3,t}\mu_t + \epsilon_{m3,t} \tag{2.B.11}$$

Measurement errors in the equations are assumed to be normally distributed, uncorrelated across measurements and uncorrelated across time.

The first equation, which relates adult outcome Q to the latent belief in the last period μ_T, is the anchoring equation. In this equation, the loading on μ_T is set equal to 1, so that a 1 unit increase in the latent factor μ_T corresponds to a 1 unit increase in the conditional expectation of the adult outcome Q, which is years of education at age 24 and above.

One of the measurements (in this case, the expected educational attainment) is used to link the factor over time. This is achieved by restricting the intercept and the factor loading of this measurement to be the same in all time periods. The intercepts and the factor loadings of the other measurements are unrestricted.

Table 2.B.2 presents the estimated factor loadings on the unobserved latent measurements of the belief latent factor. Table 2.B.3 presents the signal-noise ratio of measures of the belief factor score.

Table 2.B.2: Factor Loadings of Measurements

	Coefficient
Educational Attainment Age 5-6	0.696
Academic Standing Age 5-6	0.459
Future Prospects Age 5-6	0.660
Educational Attainment Age 7-8	0.696
Academic Standing Age 7-8	0.501
Future Prospects Age 7-8	0.678
Educational Attainment Age 9-10	0.696
Academic Standing Age 9-10	0.582
Future Prospects Age 9-10	0.643
Educational Attainment Age 11-12	0.696
Academic Standing Age 11-12	0.629
Future Prospects Age 11-12	0.637
Educational Attainment Age 13-14	0.696
Academic Standing Age 13-14	0.761
Future Prospects Age 13-14	0.641

Notes: This table presents the factor loadings on the measurements in the belief factor model.

Table 2.B.3: Percentage of Total Variance in Measurements Due to Signal and Noise

	% Signal	% Noise
Educational Attainment Age 5-6	0.540	0.460
Academic Standing Age 5-6	0.339	0.661
Future Prospects Age 5-6	0.514	0.486
Educational Attainment Age 7-8	0.540	0.460
Academic Standing Age 7-8	0.378	0.622
Future Prospects Age 7-8	0.527	0.473
Educational Attainment Age 9-10	0.540	0.460
Academic Standing Age 9-10	0.451	0.549
Future Prospects Age 9-10	0.501	0.499
Educational Attainment Age 11-12	0.540	0.460
Academic Standing Age 11-12	0.489	0.511
Future Prospects Age 11-12	0.496	0.504
Educational Attainment Age 13-14	0.540	0.460
Academic Standing Age 13-14	0.584	0.416
Future Prospects Age 13-14	0.499	0.501

Notes: This table presents the signal-noise ratio of each of the measurements in the belief factor model.

2.C Relationship between Parental Beliefs and Investments: Non-Linearity

To assess whether the relationship between beliefs and investments is non-linear, I regress the investment factor score on a quadratic function of the belief factor score. Specifically, I estimate the following model for child j at time t:

$$\underbrace{I_{jt}}_{\text{investment}} = \tilde{\alpha} + \tilde{\beta}_1 \underbrace{\mu_{jt}}_{\text{belief}} + \tilde{\beta}_2 \underbrace{\mu^2_{j,t-1}}_{\text{square of belief}} + \tilde{\psi}_1 \underbrace{I_{j,t-1}}_{\text{lag investment}} + \tilde{\psi}_2 \underbrace{I_{j,t-2}}_{\text{second lag of investment}} + \tilde{\psi}_3 \underbrace{I_{j,t-3}}_{\text{third lag of investment}} + \tilde{\gamma} X_{jt} + \tilde{\lambda} y_{jt} + \underbrace{\delta_j}_{\text{child FE}} + \tilde{\eta}_{jt} \quad (2.C.1)$$

y_{jt} denotes the log family income of child j at time t. X_{jt} includes the same variables described in specification A of the main text. As in Specification D of the main text, I estimate the model with the Blundell and Bond (1998) dynamic panel data estimator. The quadratic function of belief and the family income are treated as potentially endogenous variables.

Estimates are presented in Table 2.C.1. The coefficient on the squared belief term is not significant: there is no evidence of non-linearity. For ease of interpretation, in Table 2.C.2, I present the marginal effect of the belief factor score at the mean value in the sample.

Table 2.C.1: Dependent variable is Investment Factor Score (Standardised)

	(1)
Belief Factor Score	0.211***
	(0.061)
$(\text{Belief Factor Score})^2$	0.017
	(0.039)
Observations	5828
Number of Children	3110
Child FE	Yes
3 Lags of Investment	Yes
Belief Potentially Endogenous	Yes
Income Potentially Endogenous	Yes
Sargan Test P-value	0.128

Notes: This table reports coefficients on the parent belief factor score (standardised) in a regression of the investment factor score (standardised) on a quadratic function of the belief factor score (standardised). The construction of the belief factor score is described in the data section. The model is estimated with the Blundell and Bond (1998) estimator. Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). One-step GMM standard errors are used.

Table 2.C.2: Marginal Effect on Investment Factor Score

	Belief Factor Score
Estimate	0.22163
Se	0.03979

Notes: Standard errors computed using the delta method.

2.D Relationship between Parental Beliefs and Investments: Robustness

One may be concerned that expectations about the child's future outcomes (expected educational attainment of the child and the rating of the child's future prospects) might not just capture the belief about the skill level of the child, but may also be influenced by the mother's expectations regarding the future (e.g. about family income, school supply and/or labour market conditions). In contrast, ratings about the child in the present, such as the rating of the child's academic standing in class, could provide a "cleaner" indication of the belief about the child's skill. This section examines the impact of the rating of the child's academic standing in class on the investment factor score.

The academic standing in class is a discrete ordered variable with the following categories: (1) Near the bottom of the class; (2) Below the middle; (3) In the middle; (4) Above the middle and (5) One of the best students in class.

To determine the impact of academic standing on the investment factor score, the model presented in Equation 2.D.1 is estimated. The baseline (omitted) category of academic standing in class is "near the bottom of the class". I_{jt} represents the investment factor score of child j at time t. y_{jt} denotes the log family income of child j at time t. The time-varying controls X_{jt} are described in specification A of the main text. δ_j is a child fixed effect. The estimated specifications are analogous to specification C and D in the main text.

$$\begin{aligned}\underbrace{I_{jt}}_{\text{investment}} &= \alpha + \beta_1 belowmiddle_{jt} + \beta_2 middle_{jt} + \beta_3 abovemiddle_{jt} + \beta_4 best_{jt} \\ &\quad + \psi_1 \underbrace{I_{j,t-1}}_{\text{lag investment}} + \psi_2 \underbrace{I_{j,t-2}}_{\text{second lag of investment}} + \psi_3 \underbrace{I_{j,t-3}}_{\text{third lag of investment}} \\ &\quad + \gamma X_{jt} + \lambda y_{jt} + \underbrace{\delta_j}_{\text{child FE}} + \eta_{jt} \qquad (2.D.1)\end{aligned}$$

Table 2.D.1 presents the estimates. All specifications indicate that there is a positive relationship between the academic standing in class and the investment factor score. However, when I treat the academic standing dummy variables and log family income as potentially endogenous variables (column 2), I reject the null hypothesis that the over-identifying restrictions are valid, which could indicate that some instruments are

invalid.

Table 2.D.1: Dependent variable is Investment Factor Score (Standardised)

	(1)	(2)
Below the middle	0.191*	0.194
	(0.109)	(0.227)
Middle	0.278***	0.304
	(0.106)	(0.212)
Above middle	0.339***	0.416*
	(0.109)	(0.212)
One of the best	0.384***	0.357*
	(0.112)	(0.215)
Observations	7075	7075
Number of Children	3619	3619
Child FE	Yes	Yes
3 Lags of Investment	Yes	Yes
Academic Standing Potentially Endogenous	No	Yes
Income Potentially Endogenous	No	Yes
Sargan Test P-value	0.200	0.001

Notes: This table reports the coefficients on the dummies for category of academic standing in class in a regression of the investment factor score (standardised) on these dummies. The omitted category of academic standing in class is the category near the bottom of the class. The Blundell and Bond (1998) estimator is used. The construction of the belief factor score is described in the data section. All regressions include time-varying controls which are listed in the description of specification A in the main text. Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). One-step GMM standard errors are used.

2.E Parent Belief Updating: Alternative Models

This section examines (1) whether the relationship between the belief factor score and the skill measures might be non-linear and (2) whether lag skill measures predict current belief. To achieve this, alternative specifications of the belief updating model are estimated.

To assess whether there may be non-linearity in the relationship between beliefs and child skill measures, the belief factor score is regressed on a quadratic function of the cognitive skill factor score and the percentile score on the Behaviour Problems Index. The full model is presented in Equation 2.E.1. μ_{jt} denotes the belief factor score of child

j at period t. $cognitive_{jt}$ represents the cognitive skill factor score of child j at time t and $behaviour_{jt}$ represents the percentile score on the Behaviour Problems Index of child j at time t. $healthexcellent_{jt}$ is a dummy which takes the value of 1 when the child's health rating is excellent at time t, and 0 otherwise. Time-varying controls X_{jt} are the child's age and log family income. δ_j denotes the child fixed effect.

$$\begin{aligned} \mu_{jt} = \alpha &+ \beta_1 \mu_{j,t-1} + \beta_2 \mu_{j,t-2} + \beta_3 cognitive_{jt} + \beta_4 cognitive^2_{jt} \\ &+ \beta_5 behaviour_{jt} + \beta_6 behaviour^2_{jt} + \beta_7 healthexcellent_{jt} + \gamma X_{jt} + \delta_j + \epsilon_{jt} \end{aligned} \quad \text{(2.E.1)}$$

The estimated coefficients on the lag beliefs and the skill measures are presented in column 1 of Table 2.E.1. The coefficient on the squared term of the cognitive skill factor score is significant, suggesting that beliefs are related to the cognitive factor score in a non-linear manner. For interpretability, in Table 2.E.2, I present the marginal effect of the cognitive factor score and the Behaviour Problems Index at their mean values in the sample.

Next, to examine whether beliefs are dependent on lag skill measures, I include the lag skill measures as additional regressors. These are treated as potentially endogenous variables. The model with lag skill measures is presented in Equation 2.E.2.

$$\begin{aligned} \mu_{jt} = \alpha &+ \beta_1 \mu_{j,t-1} + \beta_2 \mu_{j,t-2} + \beta_3 cognitive_{jt} + \beta_4 behaviour_{jt} + \beta_5 healthexcellent_{jt} \\ &+ \beta_6 cognitive_{j,t-1} + \beta_7 behaviour_{j,t-1} + \beta_8 healthexcellent_{j,t-1} \\ &+ \gamma X_{jt} + \delta_j + \epsilon_{jt} \end{aligned} \quad \text{(2.E.2)}$$

The estimated coefficients on the lag beliefs, skill measures and lag skill measures are presented in column 2 of Table 2.E.1. There is a lower number of observations than column 1 because some children do not have lag skill measures. The lag cognitive skill factor score is a positive and significant predictor of the belief factor score.

Table 2.E.1: Dependent variable is Belief Factor Score (Anchored)

	(1)	(2)
Lag Belief Factor Score	0.288***	0.322***
	(0.047)	(0.047)
Lag 2 Belief Factor Score	0.127***	0.092**
	(0.037)	(0.038)
Cognitive Skill	0.821***	0.332***
	(0.081)	(0.121)
$(\text{Cognitive Skill})^2$	-0.008**	
	(0.003)	
Bad Behaviour	-0.008	-0.021***
	(0.013)	(0.004)
$(\text{Bad Behaviour})^2$	-0.000	
	(0.000)	
Health: excellent	0.267	0.303
	(0.196)	(0.251)
Lag Cognitive Skill		0.495***
		(0.115)
Lag Bad Behaviour		0.006***
		(0.002)
Lag Health Excellent		-0.057
		(0.074)
Observations	2842	2589
Number of Children	1954	1769
Child FE	Yes	Yes
Skills Potentially Endogenous	Yes	Yes
Sargan Test P-value	0.131	0.802

Notes: This table presents the estimated coefficients in a regression of the belief factor score (anchored) on lag belief factor score (anchored), second lag of belief factor score (anchored) and (1) a quadratic function of child skill measures or (2) child skill measures and lag child skill measures. Child skill measures are the cognitive skill factor score (anchored), percentile score on the Behaviour Problems Index and a dummy which takes the value of 1 when the child's health is rated as excellent (takes value of 0 when health rating is poor/fair/good). The construction of the belief factor score is described in the data section. The model is estimated with the Blundell and Bond (1998) dynamic panel data estimator. Both contemporaneous child skill measures and lag child skill measures are treated as potentially endogenous variables. All regressions control for the age of the child and log family income. Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). One-step GMM standard errors are used.

Table 2.E.2: Marginal Effect on Belief Factor Score (Anchored)

	Cognitive Skill	Behaviour Problems
Estimate	0.61509	-0.01091
Se	0.09085	0.00303

Notes: Standard errors computed using the delta method.

2.F Parent Belief Updating: Robustness

In the belief updating model in the main text, the belief factor score is a function of both objective skill measures (cognitive skill factor score, which is constructed from achievement test scores) and subjective skill measures (the Behaviour Problems Index and the rating of the child's health). In this section, the belief updating model is re-estimated, excluding the subjective measures. The revised model is presented in Equation 2.F.1. μ_{jt} denotes the belief factor score of child j at period t. $cognitive_{jt}$ represents the cognitive skill factor score of child j at time t. Time-varying controls X_{jt} are the child's age and log family income. δ_j denotes the child fixed effect.'

Estimates are presented in Table 2.F.1. Most patterns remain unchanged. An exception is for low SES families, the second lag of belief no longer predicts current belief. Moreover, in column 1 and column 3, we reject the null hypothesis that the over-identifying restrictions are valid, which may indicate that the model is incorrectly specified or that the instruments are invalid.

$$\mu_{jt} = \alpha + \beta_1\mu_{j,t-1} + \beta_2\mu_{j,t-2} + \beta_3 cognitive_{jt} + \gamma X_{jt} + \delta_j + \epsilon_{jt} \quad \text{(2.F.1)}$$

Table 2.F.1: Dependent variable is Belief Factor Score (Anchored)

	(1) All	(2) High SES	(3) Low SES
Lag Belief	0.317***	0.289***	0.259***
	(0.052)	(0.058)	(0.066)
Second Lag Belief	0.115***	0.093*	0.075
	(0.040)	(0.050)	(0.054)
Cognitive Skill	0.714***	0.719***	0.874***
	(0.072)	(0.084)	(0.092)
Observations	2842	1731	1111
Number of Children	1954	1153	801
Child FE	Yes	Yes	Yes
Skills Potentially Endogenous	Yes	Yes	Yes
Sargan Test P-value	0.012	0.332	0.067

Notes: This table presents the estimated coefficients in a regression of the belief factor score (anchored) on lag belief factor score (anchored), second lag of belief factor score (anchored) and the cognitive sklil factor score (anchored). The construction of the parent belief factor score is described in the data section. Specifications are estimated with the Blundell and Bond (1998) dynamic panel data estimator. All regressions control for the age of the child and log family income. Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). One-step GMM standard errors are used.

2.G Informativeness of Belief Measures

This section provides evidence that the belief measures are informative. The belief factor score predicts the later skills and education of the child, even after accounting for earlier measures of skills. In addition, the subjective belief measures are correlated with objective measures.

2.G.1 Beliefs Predict Skills and Education

Skill measures of children at age 12-14 are regressed on the belief factor score (standardised) at age 9-10. There are cognitive skill measures, which are Peabody Individual Achievement Test (PIAT) percentile scores in mathematics, reading recognition and reading comprehension. There is also a measure of non-cognitive skills — the percentile score on the Behaviour Problems Index (BPI). Estimates are presented in Table 2.G.1. Higher percentile scores on the PIAT indicate better performance, and higher percentile scores on the Behaviour Problems Index indicate lower non-cognitive skills (worse behaviour). Beliefs at age 9-10 are positively correlated with percentile scores in PIAT mathematics, PIAT reading recognition and PIAT reading comprehension at age 12-14. Beliefs at age 9-10 are also negatively correlated with the percentile score on the Behaviour Problems Index at age 12-14 (higher beliefs associated with lower level of behaviour problems). Furthermore, these associations remain statistically significant, even after accounting for family-specific time-invariant characteristics.

Next, I regress measures of children's skills at 15-24 years and an indicator for ever attending college on the belief factor score at age 13-14. Estimates are presented in Table 2.G.2. The first measure of skill is the average Pearlin Mastery Score between ages 15 and 24. A higher score means that the individual believes that their life chances are under their own control to a higher extent. The second measure of skill is the average CES-D depression scale between ages 15 and 24. A higher score means that the individual is more depressed. The belief factor score at age 13-14 is positively associated with the Pearlin Mastery Scale score, negatively associated with CES-D and positively associated with ever attending college. Even after controlling for family-specific time-invariant characteristics, beliefs remain predictive of children's skills at age 15-24 along with educational attainment.

Table 2.G.1: Dependent variable is Skills of Child at Age 12-14

	PIAT Math		PIAT Read Recog		PIAT Read Comp		BPI	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Belief Age 9-10	3.589***	4.193***	2.466***	3.232**	2.373***	3.564**	-2.118***	-3.647**
	(0.536)	(1.419)	(0.528)	(1.409)	(0.569)	(1.556)	(0.610)	(1.584)
Observations	2345	2345	2345	2345	2345	2345	2345	2345
R^2	0.593	0.857	0.671	0.887	0.543	0.833	0.509	0.825
Family FE	No	Yes	No	Yes	No	Yes	No	Yes

Notes: This table presents the coefficient on the belief factor score at age 9-10 in a regression of a child skill measure at age 12-14 on the belief factor score at age 9-10. Child skill measures include the percentile score on the Peabody Individual Achievement Test (PIAT) in mathematics, reading recognition and reading comprehension, along with the percentile score on the Behaviour Problems Index (BPI). All regressions include the gender, birth order of the child, average percentiles on PIAT tests in mathematics, reading recognition and reading comprehension between 5 to 10 years and average percentile on the Behaviour Problems Index between 4 to 10 years. The OLS regressions additionally include race, mother's characteristics and mother's family background. Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). Standard errors are clustered at the family level in both the ordinary least squares and the family fixed effects models.

Table 2.G.2: Child Outcomes: Skills at Age 15-24, Ever Attended College

	Pearlin 15-24 yr		CES-D 15-24 yr		Ever College	
	(1)	(2)	(3)	(4)	(5)	(6)
Belief Age 13-14	0.465***	0.354*	-0.466***	-0.094	0.125***	0.096***
	(0.073)	(0.197)	(0.086)	(0.217)	(0.012)	(0.030)
Observations	2368	2368	2368	2368	2368	2368
R^2	0.167	0.732	0.140	0.722	0.320	0.773
Family FE	No	Yes	No	Yes	No	Yes

Notes: This table presents the coefficient on the belief factor score at age 13-14 in a regression of a child skill measure at age 15-24 or an indicator for ever attending college on the belief factor score at age 13-14. All regressions control for the gender, birth order of the child, average percentiles on PIAT tests in mathematics, reading recognition and reading comprehension between 5 to 14 years and average percentile on the Behaviour Problems Index between 4 to 14 years. The OLS regressions additionally control for race, mother's characteristics and mother's family background. Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). Standard errors are clustered at the family level in both the ordinary least squares and the family fixed effects models.

2.G.2 Measures of Belief are Correlated with Objective Measures

This section assesses the accuracy (and therefore informativeness) of two component measures used to construct the belief factor score: (1) the expected educational attainment of the child and (2) the rating of the child's academic standing in class. This is possible as there are objective measures which can be compared to these (subjective) measures.

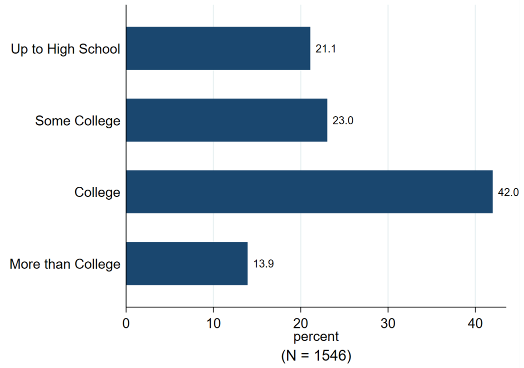
Expectations of Educational Attainment

Here, the accuracy of the expectations of the child's educational attainment is assessed by comparing it to the realised educational attainment of the child.

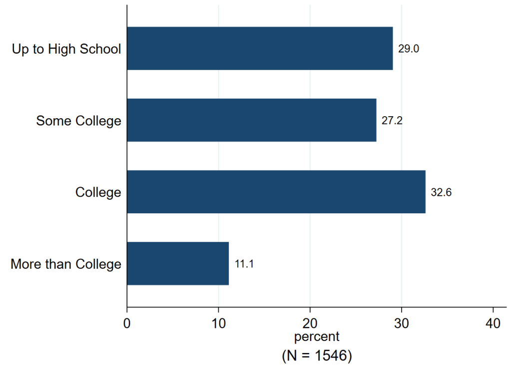
First, I focus on the sub-sample of children whose education outcomes were observed at age 27 and above. Figure 2.G.1 presents the expectations of educational attainment when the child was aged 13-14 and the child's educational attainment at age 27 and above. On average, parents are optimistic about their children and this is consistent with other studies (Dizon-Ross, 2019; Attanasio et al., 2020; Nicoletti et al., 2022; Kinsler & Pavan, 2021; Bergman, 2021). When children were aged between 13-14 years, 55.9% were expected to graduate with college education and above. However, only 60.65% among this 55.9% actually attained college education and above by age 27.

Figure 2.G.1: Expectations at 13-14 years versus Realised Educational Attainment

(a) Expectations at age 13-14 years



(b) Educational Attainment at Age 27 and above



Notes: The figure on the left presents the expected educational attainment of the child at age 13-14. The figure on the right shows the realised educational attainment of the child at age 27 and above. In each figure, there are 4 categories: up to high school, some college, college and more than college. Both figures present the percentages in each of the categories.

One concern is that the optimism is specific to the sample of children who are observed at age 27 and above. Based on the structure of the survey, this sample is negatively selected — these children must have been born to younger mothers. Given this concern, I turn to assess the accuracy of an outcome at an earlier age of the child: college graduation status by age 24 and above. This enables me to include children who were born to slightly older mothers, reducing the selectiveness of the sample.

Mothers are not explicitly asked whether they expect their child to be a college graduate. If mothers selected the options college or more than college, I assume that they expected their child to be a college graduate. Otherwise, I assume that they did not expect their child to be a college graduate. Using this alternative outcome, there is also evidence of inaccurate predictions. When children were aged 13-14, 58.93% were expected to graduate from college. However, only 59.93% out of the 58.93% actually graduated from college.

Table 2.G.3 presents the percentage of accurate predictions about the child's college graduation status at age 24 and above. This is done separately by the education level of the mother (up to median/above median) and the level of family income (up to median/above median). Above median education refers to more than 12 years of education. Median family income is defined based on the average family income when the child was aged between 11 and 22. On average, mothers in families with above median family income made more accurate predictions.

Table 2.G.3: Percentage of Accurate Predictions about Child Being College Graduate at Age 24 and Above by Mother's Education and Income

	Up to Median	Above Median
Education	67.2	68.1
Income	65.2	70.2

Rating of Child Academic Standing in Class

Without objective information on class performance, one cannot determine whether the rating of the child's academic standing in class is accurate. Despite this, one might expect that children who are at the top (bottom) of the class are also more likely to be ranked at the top (bottom) of the national distribution of skills, as measured by percentile scores on the PIAT tests.

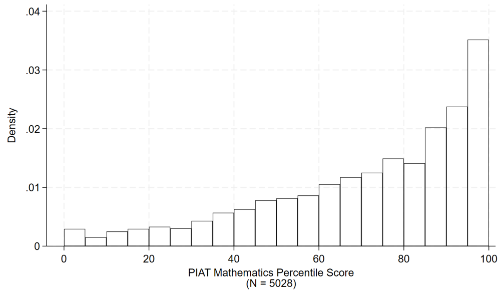
Figure 2.G.2 presents histograms of the percentile scores on the PIAT tests for children who were rated by their mothers as being among the top of their class. Figure 2.G.3 presents the corresponding histograms for children who were rated as being near the bottom of their class. These figures include all available observations of children aged between 5 and 14.

Children who are ranked near the bottom of the class are indeed more likely to have lower percentile scores on the PIAT tests. However, some children are ranked near the bottom, even though they perform well on the achievement tests (score 80 percentile and above). Similarly, children who were ranked among the best in the class are more likely to have attained high percentile scores on the PIAT tests. However, some children who are ranked among the top perform poorly on the achievement tests (score 20 percentile and below).

The discrepancy between the rating of academic standing and performance on the achievement tests could imply that some mothers hold inaccurate perceptions. However, because there is no objective information on class quality, the evidence provided here is only suggestive.

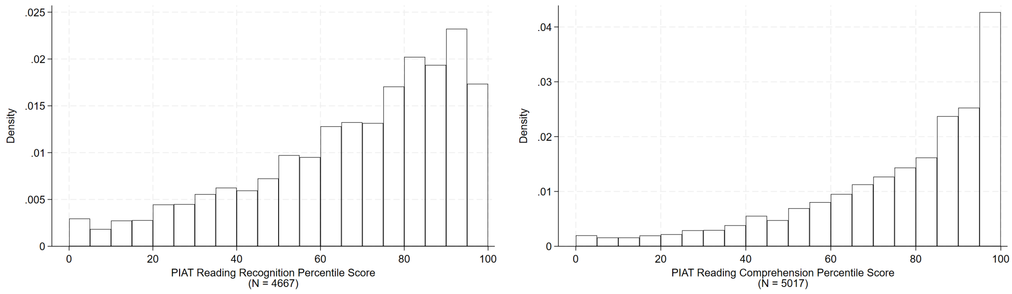
Figure 2.G.2: Histograms of Achievement Test Percentile Scores for Children Ranked Among the Top of the Class

(a) PIAT Mathematics



(b) PIAT Reading Recognition

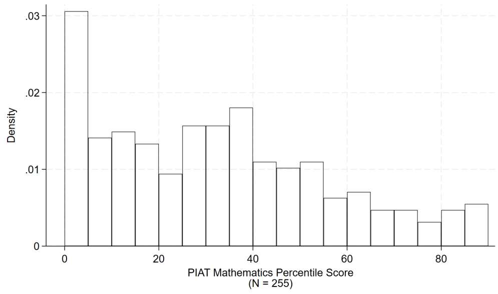
(c) PIAT Reading Comprehension



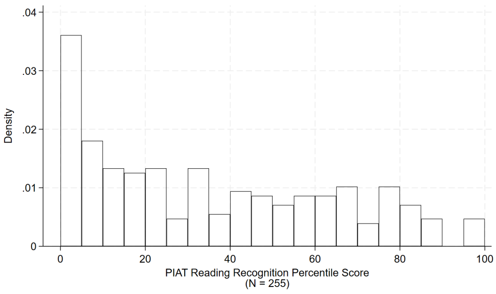
Notes: This figure presents histograms of the percentile scores on (a) Peabody Individual Achievement Test (PIAT) mathematics, (b) PIAT reading recognition and (c) PIAT reading comprehension. The sample comprises children who were ranked at the top of the class by their mothers. All available observations between ages 5 and 14 are used.

Figure 2.G.3: Histograms of Achievement Test Percentile Scores for Children Ranked Near the Bottom of the Class

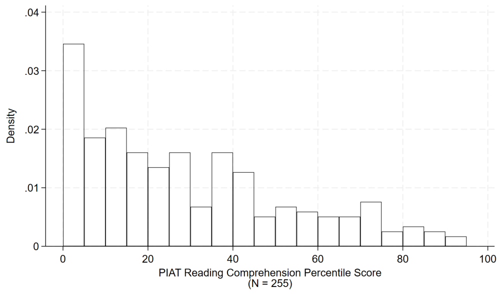
(a) PIAT Mathematics



(b) PIAT Reading Recognition



(c) PIAT Reading Comprehension



Notes: This figure presents histograms of the percentile scores on (a) Peabody Individual Achievement Test (PIAT) mathematics, (b) PIAT reading recognition and (c) PIAT reading comprehension. The sample comprises children who were ranked near the bottom of the class by their mothers. All available observations between ages 5 and 14 are used.

Chapter 3

A Dynamic Analysis of Parental Beliefs and Investments

3.1 Introduction

A well-documented gap exists in parental investments across socio-economic status (SES), with lower-income parents investing significantly less in their children compared to wealthier parents (Caucutt et al., 2017; Bolt et al., 2024; Carneiro et al., 2024). This SES investment gap is concerning because it contributes to persistent disparities in children's human capital and life outcomes, such as educational attainment and earnings (Becker & Tomes, 1979, 1986; Solon, 1999; Attanasio, Cattan, & Meghir, 2022). Typical explanations for the SES investment gap include differences in the resources of parents, skills of children — which affect the marginal returns to investment — and preferences of parents (Cunha, 2014; Caucutt et al., 2017).

Could the SES investment (and therefore skills) gap also be driven by differences in parental beliefs? The data suggests that SES disparities extend to beliefs: on average, high SES parents hold higher beliefs in their children and invest more, even after accounting for child skill levels. Emerging evidence suggests that parental beliefs meaningfully influence investment decisions (Dizon-Ross, 2019; List et al., 2021). While experimental studies confirm short-term effects of beliefs on investments and skills (e.g. Dizon-Ross (2019), List et al. (2021)), whether these effects persist over time remains unknown. Long-term effects of SES differences in beliefs may exist if there is a self-fulfilling prophecy[1] regarding parental beliefs about the skill level of their

[1] The idea of a self-fulfilling prophecy and the importance of beliefs/perceptions in shaping individual outcomes has been acknowledged by other economists. In the EEA Presidential Lecture, La Ferrara

child. Rich parents may over-estimate the skills of their children, while poor parents under-estimate the skills. This leads the rich to invest more than the poor, resulting in the poor having lower skills than the rich, validating the initial prophecy. If the self-fulfilling prophecy exists, correcting parent misperceptions may be a policy lever to narrow the SES skill gap. Investigating the self-fulfilling prophecy requires a framework which connects beliefs, investments and skills across time.

This paper introduces a dynamic model of parent investments which incorporates beliefs of parents about the cognitive skill of their child. The model features feedback between beliefs and investments, providing a channel for the self-fulfilling prophecy, and therefore enabling me to investigate whether beliefs have long-term effects. The model is used to (1) quantify the role of parental beliefs about child cognitive skill in explaining the SES skill gap and to (2) assess the potential effect of correcting inaccurate parental beliefs on the SES skill gap.

The model has the following features. Parents have imperfect information: they do not observe the cognitive skill of their child. Instead, they hold a belief about their child's cognitive skill. Over time, parents learn about their child's skill level as they receive signals (noisy measures of the skill) and update their belief. The investment decisions of parents affect the skills of the child: when parents invest, the unobserved skills of the child grow. The timing and evolution of beliefs is as follows. In each period, after parents invest, they form prior beliefs about the skill of their child in the next period. In the following period, they enter with this prior belief and then they receive a signal, a noisy measure of the skill of their child. Given the signal, they revise their beliefs. This updated belief is used to make their investment decision, and so on.

Crucially for the determination of the long-term effects of beliefs, there is two-way interaction between beliefs and investments. Parental beliefs about their child's skill influence their investments because they determine the parent's perception of returns to investment. In turn, parent investments also influence their beliefs, both directly and indirectly. Investments have a direct effect on beliefs: investing more in this period leads parents to hold higher prior beliefs about next period skill, because parents know that investments produce skills. Investments also have an indirect effect on beliefs, which acts through the signal received in the following period and subsequent revision of beliefs. Investing more leads to higher skills in the following period. Since the signal

(2019) mentions that the poor may aspire less than the rich, leading them to invest less than the rich, causing them to remain in poverty. In the Tanner Lectures, Duflo (2012) states that a lack of hope could result in people putting in less effort and failing to reach their potential.

received in the next period depends on the underlying skill level, on average, this leads to parents receiving a higher signal, which is a force to revise beliefs upwards. These beliefs will then affect investment decisions, and so on and so forth.

The model incorporates both observable and unobservable heterogeneity across families. Parents differ by education level. High education mothers may update beliefs differently, be more productive in converting investments to skills and are more likely to earn higher income. Parents also differ in terms of unobserved preferences. Each parent has a probability of being one of two discrete types, with one type holding a higher relative preference for child skill over consumption. The unobserved preference heterogeneity serves two main functions: controlling for factors that simultaneously influence beliefs and investments, and addressing endogeneity in initial conditions such as income, assets, beliefs, and child skill levels.

The model is estimated using the method of simulated moments, based on longitudinal data on parental beliefs, investments and child skills from the National Longitudinal Survey of Youth. Parental beliefs are estimated from a factor model based on responses by the mother of the child about (1) the child's ranking in class; (2) the child's future educational attainment and (3) the child's future prospects. Parent investments are derived from a factor model based on measures of goods and time investments. Child skills are estimated from a factor model based on achievement test scores.

The results reveal a mechanism consistent with a self-fulfilling prophecy. Parents with higher beliefs invest more, because they perceive a higher marginal return to investment, since skills and investments are complements in the production function. If they invest more, in the following period, on average, they hold higher beliefs, which will incentivise them to invest more for the same reasons, and so on.

The model is used to evaluate the relative contribution of parental beliefs about child cognitive skill (versus parental resources, skills of children and parent preferences) in explaining the SES skill gap. Although parental beliefs affect investments, differences in the initial beliefs of parents explain less of the SES skill gap than differences in the initial resources of parents (income, assets) and differences in the initial skills of children. Equalising parental beliefs in the first period lowers the SES skill gap by less than 2%. In contrast, equalising the initial skills of children or the initial resources of parents decreases the SES skill gap by a significantly larger percentage. In general, these results suggest that parental beliefs about child skill do not explain much of the SES skill gap. The strength of the self-fulfilling prophecy is low and the long-term effects of these beliefs are limited.

Finally, to understand the impact of inaccurate beliefs about the skills of children, the model is used to estimate a counterfactual scenario where parental beliefs match the skills of their child (no information frictions) and parents do not update their beliefs. Since it is the case that on average, high SES parents over-estimate the skills of their child and low SES parents under-estimate the skills of their child, high SES (low SES) parents now have lower (higher) beliefs in the first period than the baseline. This lowers (raises) their perceived marginal return to investment, because skills and investments are complements in the production function, which encourages them to invest less (more). This in turn lowers (raises) their beliefs for the following period, which provides an incentive to invest less (more) in the following period, for the same reasons, and so on. Overall, skills of the high SES (low SES) children decrease (increase). Therefore, the SES skill gap decreases. However, the decrease is less than 1%, which indicates that targeting parental beliefs about the skill level of their child (e.g. by providing them with more information) may not be an effective strategy to reduce the SES skill gap.

This paper relates to the literature on subjective beliefs and information frictions within Economics of Education. Studies have established that subjective beliefs of parents may be inaccurate (Dizon-Ross, 2019; Nicoletti et al., 2022; Bergman, 2021), parental beliefs are associated with parent investments (Attanasio & Kaufmann, 2014; Attanasio, Cunha, & Jervis, 2019; Attanasio, Boneva, & Rauh, 2019; Dizon-Ross, 2019; Bergman, 2021; List et al., 2021; Cunha et al., 2022; Lekfuangfu & Odermatt, 2022; Conti et al., 2022; Page & Ruebeck, 2022) and parents revise their beliefs in response to information (Nicoletti et al., 2022; Greaves, Hussain, Rabe, & Rasul, 2023). Dizon-Ross (2019) shows that providing parents with information can shift their beliefs and investments in the short-term. This paper contributes by determining whether parental beliefs about child skill have long-term effects.

In terms of modelling, this paper draws on the body of work related to beliefs and learning (Ackerberg, 2003; Erdem & Keane, 1996; Crawford & Shum, 2005; Zafar, 2011; Sanders, 2012; Stinebrickner & Stinebrickner, 2014; Kinsler & Pavan, 2016; Arcidiacono et al., 2025; Chikhale, 2024). Two other papers have estimated dynamic models of parent investments which incorporate parental beliefs, but they have different goals. Chikhale (2024) investigates the impact of adjusting parental beliefs about the return to investment on intergenerational mobility. Kinsler and Pavan (2016) explores the determinants of parental beliefs about the average level of child skill in the population, highlighting that these beliefs are distorted by the average skills in the child's school. In contrast, this paper focuses on parental beliefs about child skill, using a dynamic model to quantify the role which these beliefs play in explaining the socio-economic status skill gap. Unlike the models in Chikhale (2024) and Kinsler and Pavan (2016), the model in this paper accounts for both observable and unobservable differences across parents. Since these characteristics affect investment decisions and are likely to be correlated with parental beliefs, explicitly controlling for these factors makes the estimated impact of beliefs on investments more credible. Notably, belief updating is not assumed to follow Bayesian rules, as Bayesian updating leads to unrealistically fast learning relative to the data. Instead, the model adopts a non-Bayesian structure that better matches observed persistence in beliefs.

The rest of this paper is organised as follows. Section 3.2 introduces the data. Section 3.3 presents patterns in the data which motivate the research question and the model. Section 3.4 introduces the dynamic model of parent investments and belief updating. Section 3.5 describes the identification and estimation. Section 3.6 presents the model estimates and features of the baseline. Section 3.7 presents a decomposition of the socio-economic status skill gap and a counterfactual where parental beliefs match the skills of their child and parents do not update their beliefs. Section 3.8 concludes the paper.

3.2 Data

Data sources are the National Longitudinal Survey of Youth 1979, also known as the NLSY79, and the NLSY Child and Young Adult (CYA). This analysis focuses on children of the females in the cross-section of the NLSY79, a total of 5,819 children. Since all biological children of the females are reported, if children have siblings, they are also in the data.

Barring the attrition of females from the NLSY79 and attrition of children from the CYA, the CYA is representative of the population of children born to females from birth cohorts 1957 to 1964 in the United States. Though the cross-section of the NLSY79 is representative of the individuals in those birth cohorts, the children of the females in the cross-section may not be representative of their birth cohorts. There is selection into fertility and timing of children. Children in earlier birth cohorts were born to younger mothers and are generally negatively selected in terms of mother's characteristics — their mothers typically have lower years of education and cognitive skills. In addition, these children are more likely to grow up in single parent households.

3.2.1 Parental Belief

The parental belief is based on responses of the child's mother to 3 questions in the survey.

The first question is an expectation of the child's future educational attainment. Specifically, the mother is asked "How far do you think your child will go in school?". Possible options are: (1) Leave high school before graduation; (2) Graduate from high school; (3) Get some college or other training; (4) Graduate from college; (5) Get more than 4 years of college/further training after college; and (6) Something else. In practice, very few mothers choose option 1. Thus, option 1 and option 2 are merged together, and henceforth referred to as a single category: "Up to High School". In addition, option 6 (something else) is vague and few mothers chose it. Therefore, it is treated as a missing response. From this point onwards, the expectations of educational attainment have four categories: (1) Up to high school; (2) Some college or other training; (3) Graduate from college; and (4) More than college.

The second question is the rating of the child's academic standing in class. The child's mother indicates whether she perceives the child to be (1) Near the bottom of the class; (2) Below the middle; (3) In the middle; (4) Above the middle; or (5) One of the best students in class. Most children are rated at being in the middle of the class or above. This is a relative ranking rather than an absolute ranking, and there is evidence that relative rankings matter for child achievement (Kinsler & Pavan, 2021; Elsner et al., 2021).

The third question is the mother's rating of the child's future prospects. Possible options are: (1) Poor; (2) Fair; (3) Good; and (4) Excellent. Few mothers chose the category poor. Therefore, in this analysis, poor and fair are grouped together and treated as a single category.

Overall, mothers tend to hold good opinions about their children. More than 50% of children are expected to attain high school education and more than 50% of the children are rated as being above the middle of their class or one of the top in their class. Also, greater than 50% of children are rated as having excellent future prospects. Detailed proportions of the responses to each of the above measures are provided in Appendix 2.A.1.

Two of these measures are about expectations about the child's future. Therefore, aside from capturing the mother's perception about the child's skills, they may also reflect the mother's expectations about planned future investments and future shocks.

There is within-child variation in these measures. For example, in the sample of children who had beliefs reported three times between the ages of 9 and 14, (1) around 57% had at least one change in expected educational attainment; (2) around 61% had at least one change in rating of academic standing and (3) around 45% had at least one change in the rating of future prospects.

I assume that these three questions reflect a latent belief factor and use a factor model to extract the underlying factor, predicting a factor score for each child. I interpret the factor score which I extract from these three measures as representing the difference between the child's cognitive skill level and the average cognitive skills of children of the same age[2]. A factor model is useful because it adjusts for measurement error which is embedded in the responses to each of the 3 questions. The measure of expected educational attainment is used to link the factor over time, so it can be compared across ages of the child. Since latent factors have no natural scale (Anderson & Rubin, 1956), the belief factor is anchored to the years of education of the child at age 24 and above. This means that a 1 unit increase in the latent factor corresponds to a 1 unit increase in the conditional expectation of the years of education of the child at age 24 and above. To obtain the beliefs of parents in terms of the skill level, rather than the difference from average skills of children of the same age, the mean log cognitive skill at the corresponding age is added to the predicted factor score[3]. This "corrected" factor score

[2]It is an assumption that these measures capture beliefs about cognitive skills relative to children of the same age. It is possible that parental beliefs about the child's absolute performance are influenced by perceptions regarding relative performance. One could imagine that parents believe that their child will compete with peers for limited places in higher education institutions and job opportunities. Therefore, parental beliefs about absolute performance, such as the child's future prospects (including job opportunities) and educational attainment, may depend on the child's standing relative to his/her peers.

[3]Because of the nature of the measurements of belief, on average, the belief factor score will not rise in value when children grow older. In fact, if parents already rate their children in the highest possible categories of all belief measurements when children are young, the belief factor scores can only remain the same or decrease when children grow older. In contrast, the skill factor score will generally rise as children grow because children attain higher values on the measurements of skills, the raw scores on achievement tests. The following narrative could rationalise why the belief factor score does not grow with age while the skill factor score does. To parents, the skill level of the child is equivalent to a child-specific constant term plus an age-specific average skill, which is common to all children. Parents are aware of the age-specific skill value, but they do not observe the child-specific constant term. Consequently, belief measurements and the belief factor score only depend on the parent's perceptions about the child-specific constant term. In this way, when child skills grow with age, the belief factor score may remain

is referred to as the anchored belief factor score. This anchored belief factor score is the measure which is used in the analysis. The mean log cognitive skill is the average cognitive skill factor score (anchored) at the corresponding age. The construction of the cognitive skill factor score is discussed below.

Beliefs of agents are usually represented as a distribution (see Stinebrickner and Stinebrickner (2014) and Arcidiacono et al. (2025), for example). In this context, the belief of parents about the skill level of their child could be a distribution over the skill level of their child. The anchored belief factor score is taken as the mean of this distribution.

3.2.2 Parent Investment

Following Cunha et al. (2010), a latent factor of parent investment is constructed from components of the Home Observation Measurement of the Environment (HOME), which proxy time and goods investments. Some components include how often the mother reads to the child, how many books the child has and how often the child is brought to the museum. A complete list of the component measures is provided in Appendix 3.A.1. The measure of how frequently the child is brought to the museum is used to link the factor over time, so that it can be compared across ages. The investment factor does not have a natural location or scale (Anderson & Rubin, 1956). For the data patterns, it is standardised within the sample: its mean is 0 and it has a standard deviation of 1. For the estimation of the model, the original investment factor score is used.

3.2.3 Child Cognitive Skill

Measurements of cognitive skill are raw scores on the Peabody Individual Achievement Tests (PIAT) in mathematics, reading recognition and reading comprehension, which were administered by the survey. These are available when the child is between the ages of 5 and 14. Higher raw scores indicate that the child has a higher skill level. Generally, mothers do not observe these achievement test scores. These scores are assumed to be correlated with the child's school grades, which are not collected by the survey, except during 1995-1996[4].

---

unaffected. If this narrative holds, then parental belief about the child's skill can be obtained by adding the age-specific skill to the belief factor score (belief about child-specific constant term). It is assumed that the age-specific skill is the average log skill factor score at the specific age.

[4] In 1995-1996, the school grades are correlated with the achievement test scores: national percentile ranks in the school achievement tests are correlated with the percentile ranks in the PIAT collected by the survey.

Following Cunha et al. (2010), it is assumed that the raw scores on the achievement tests are equivalent to an underlying log cognitive skill of the child plus measurement error. To uncover the underlying skill, I estimate a factor model based on the raw scores on the three achievement tests and predict cognitive skill factors for the children. The raw score on the mathematics achievement test is used to link the latent factor over time, so it can be compared across ages. Latent factors have no natural scale (Anderson & Rubin, 1956). Therefore, the cognitive skill factor is anchored to years of education of the child at age 24 and above. This means that a 1 unit increase in the log cognitive skill corresponds to a 1 unit increase in the conditional expectation of the years of education of the child at age 24 and above. Note that because I also anchored beliefs to years of education, the skill factor is in the same units as the belief factor.

3.2.4 Family Income and Assets

Family income is the total net family income of the family. It sums the following components received by respondent and spouse or partner: wage income, farm/business income, military income, unemployment compensation, Aid to Families with Dependent Children (AFDC), food stamps, Supplemental Security Income (SSI)/welfare, child support, alimony, educational benefits and/or scholarships, fellowships and grants, veteran benefits and income from other sources. From survey year 2002, income from worker's compensation, disability and social security is also included. In every survey year, the top 2% of net family income values are top-coded.

Family assets are the total net wealth of the family. This is the sum of asset values minus debts. Assets include the market value of owned residential property, money assets like savings accounts, market value of farm/business/other real estate and the total market value of all other assets worth more than $500. Debts include mortgages and back taxes owed on residential property, debts on farm/business/other property, debts on vehicles including automobiles and the total amount of other debts over $500. In each survey year, the top 2% of total net family wealth values are top-coded.

3.2.5 Definition of Socio-economic Status and Mother's Education

This paper will consistently refer to socio-economic status and mother's education. These terms are defined here.

For the data patterns, high socio-economic status (SES) families refer to those with above median level of family income (defined based on average family income when the child is aged 9-10). Otherwise, families are low SES. The median is defined within the children of females in the cross-section sample of the NLSY79. In the model estimation, high/low SES is also defined based on average family income between age 9-10, but the threshold for median income is set within this smaller sample.

In the model, parameters are allowed to depend on mother's education, which is distinct from SES. High education is defined as mothers with above median education (more than 12 years). Otherwise, mothers have low education. Both high and low SES families can have mothers with high/low education.

3.2.6 Sample for Data Patterns

As mentioned above, this analysis focuses on the 5,819 children of females in the cross-section sample of the NLSY79. All biological children of the females are included. Some children have siblings. Sample statistics of these children are provided in Table 3.1. There are less than 5,819 observations of certain variables because of missing values.

On average, mothers attained 13.42 years of education (note that 12 years corresponds to graduating from high school, provided no grades were repeated). The average percentile score of mothers on the Armed Forces Qualification Test (AFQT), which is a measure of cognitive skill, is 45.89. The mean age of the mother at birth is 26.79 years. Around 77% of the sample are Whites, 14% are Blacks and 9% are Hispanics. 52% of the children are male and the rest are female. Several statistics are provided at age 9-10, to compare with the sample used in the model estimation. Average family income at age 9-10 is around 83,054.44 in 2015 U.S. dollars, while average assets at age 9-10 is around 171,402.31 U.S. dollars.

In this paper, several data patterns are presented to motivate the research question and the model. To maximise the sample, when producing each data pattern, child-year observations are included whenever the key variables of interest are available. In this way, the sample used to produce each data pattern is distinct, and the number of children used to produce each data pattern is a subset of the 5,819 children presented in Table 3.1. Furthermore, although the model focuses on children aged between 9 and 14, data from age 5-8 will also be used to produce some data patterns, particularly when it is important for the estimator to have multiple time periods of data.

Table 3.1: Summary Statistics of Children of Females in Cross-Section Sample of NLSY79

	N	Mean	Median	Std. Dev.
Belief Factor Score Age 9-10	2,888	13.04	13.03	1.25
Skill Factor Score Age 9-10	4,116	13.07	13.14	0.80
Family Income Age 9-10	4,274	83,054.44	62,549.62	107,052.05
Asset Age 9-10	3,425	171,402.31	34,110.26	432,365.27
Mother's Years of Education	5,819	13.42	12.00	2.57
Mother's AFQT Percentile	5,547	45.89	44.06	28.87
Mother's Age at Birth	5,819	26.79	26.00	6.06
White	5,819	0.77	1.00	0.42
Black	5,819	0.14	0.00	0.35
Hispanic	5,819	0.09	0.00	0.28
Male	5,819	0.52	1.00	0.50

Notes: This table presents summary statistics of the 5,819 children born to females in the cross-section sample of the NLSY79. AFQT refers to Armed Forces Qualification Test and it is a measure of cognitive skill. Income and assets are adjusted to 2015 dollars using the Consumer Price Index series (CPI-U).

3.2.7 Sample for Model Estimation

Like the sample used to produce the data patterns, this is a subset of the children born to females in the cross-section sample of the NLSY79. However, this sample is different from that used to produce the data patterns because a different selection criteria (including that beliefs, skills, income and assets are available at age 9-10, the start of the model) is employed. A description of the sample selection is provided in Section 3.5, after the model has been introduced. After the selection, there are 655 parent-child pairs. Currently, siblings in the data are treated as independent observations.

This table presents summary statistics of the high SES and low SES parent-child pairs in the sample. Examining Table 3.2 and Table 3.1 reveals that the sample used in model estimation is positively selected, compared to the full sample of children of the females in the cross-section. This can be seen as the mean and median values of the belief factor score, skill factor score, family income and assets are higher in Table 3.2 than in Table 3.1. In addition, mothers have higher education and there is a higher proportion of Whites and a lower proportion of Blacks.

Table 3.2 also presents summary statistics for the sub-samples of high SES and low SES families. High SES families are defined as those with above median family income at age 9-10, otherwise families are low SES. The median is defined within the 655 parent-child pairs. On average, high SES families have higher average belief factor scores, skill factor scores, family income and assets at age 9-10. In addition, on average, mothers of high SES families also have higher average years of education, AFQT percentile scores and age at birth than those in low SES families.

Appendix Table 3.D.1 provides additional information on the number of children dropped because of the sample selection criteria.

Table 3.2: Summary Statistics of Parent-Child Pairs in Model Estimation

	N	Mean	Median	Std. Dev.
All				
Belief Factor Score Age 9-10	655	13.28	13.43	1.15
Skill Factor Score Age 9-10	655	13.27	13.28	0.74
Family Income Age 9-10	655	105,356.37	85,373.08	97,655.73
Asset Age 9-10	655	303,565.68	141,866.05	528,464.77
Mother's Years of Education	655	14.03	13.00	2.35
Mother's AFQT Percentile	641	56.47	58.40	27.27
Mother's Age at Birth	655	27.01	27.00	3.98
White	655	0.88	1.00	0.33
Black	655	0.06	0.00	0.23
Hispanic	655	0.07	0.00	0.25
Male	655	0.50	1.00	0.50
High SES				
Belief Factor Score Age 9-10	325	13.49	13.49	1.08
Skill Factor Score Age 9-10	325	13.35	13.35	0.76
Family Income Age 9-10	325	152,588.78	118,156.34	120,233.69
Asset Age 9-10	325	474,705.81	263,054.19	680,766.18
Mother's Years of Education	325	14.87	15.00	2.41
Mother's AFQT Percentile	321	65.27	71.37	24.53
Mother's Age at Birth	325	28.27	28.00	3.60
White	325	0.91	1.00	0.29
Black	325	0.04	0.00	0.20
Hispanic	325	0.05	0.00	0.22
Male	325	0.52	1.00	0.50
Low SES				
Belief Factor Score Age 9-10	330	13.06	13.16	1.18
Skill Factor Score Age 9-10	330	13.18	13.22	0.72
Family Income Age 9-10	330	58,839.60	61,182.12	18,620.71
Asset Age 9-10	330	135,018.58	74,817.84	203,292.58
Mother's Years of Education	330	13.21	12.00	1.98
Mother's AFQT Percentile	320	47.64	45.24	27.06
Mother's Age at Birth	330	25.77	26.00	3.95
White	330	0.85	1.00	0.36
Black	330	0.08	0.00	0.27
Hispanic	330	0.08	0.00	0.27
Male	330	0.49	0.00	0.50

Notes: This table presents summary statistics of the 655 parent-child pairs used to estimate the model. AFQT refers to Armed Forces Qualification Test and it is a measure of cognitive skill. Mother's AFQT scores are missing for some children and the AFQT summary statistics are based on the non-missing values. Income and assets are adjusted to 2015 dollars using the Consumer Price Index series (CPI-U). High socio-economic status (SES) parents refer to parents who have above median family income (defined based on average family income when the child is aged between 9 and 10). Otherwise, parents are low SES. The median level of income is defined within the sample of 655 parent-child pairs.

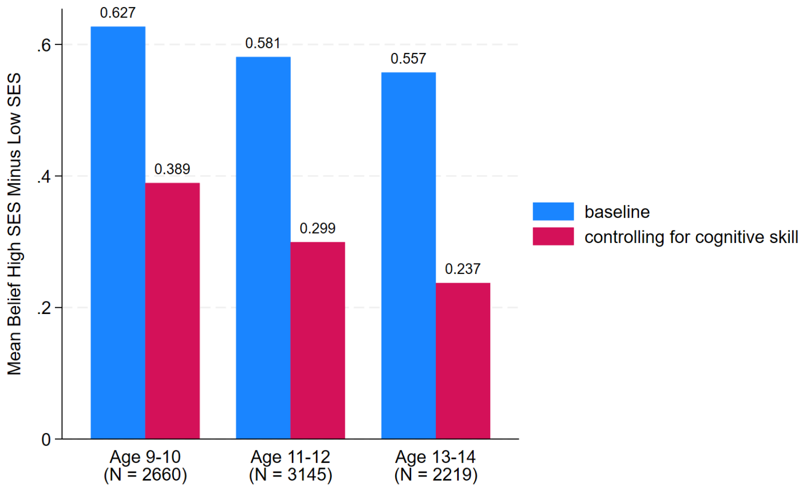
3.3 Data Patterns

This section presents several data patterns which motivate the research question and the model.

3.3.1 SES Differences in Parental Beliefs and Investments

On average, high socio-economic status (SES) parents hold higher beliefs in their children, even after accounting for children's skill levels. Figure 3.1 presents the average difference in the belief factor score (anchored) between the high SES and the low SES parents at different ages of the child. The blue bar is the unconditional average difference, while the red bar is the average difference after controlling for the child's cognitive skill factor score at the same age. At all ages, both the blue and the red bars are positive, indicating that the high SES parents hold higher beliefs about their child, even after accounting for child cognitive skills. For example, the unconditional difference in the belief factor score at age 9-10 is 0.627 units, which is 0.627 years of education at age 24 and above. The SES differences in average beliefs are statistically significant.

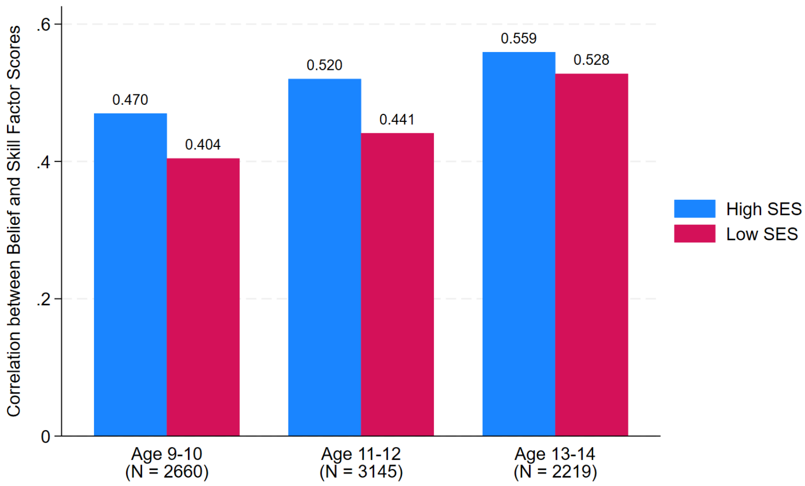
Figure 3.1: Average Belief Factor Score (Anchored) by SES



Notes: This figure presents the mean belief factor score (anchored) for high SES parents and low SES parents at different ages of the child. High socio-economic status (SES) parents refer to parents who have above median family income (defined based on average family income when the child is aged between 9 and 10). Otherwise, parents are low SES. At each age, all children with non-missing belief factor scores, cognitive skill factor scores and average family income between age 9-10 are included in the computation.

Parents may not know the skill level of their child: from Figure 3.2, the correlation between the belief factor scores (anchored to years of education at age 24 and above) and cognitive skill factor scores (anchored to years of education at age 24 and above) is less than 1. Furthermore, the correlation between skills and beliefs is lower for low SES parents, suggesting that on average, low SES parents hold more inaccurate beliefs. At ages 9-10 and 11-12, the SES differences in the correlation between beliefs and skills are statistically significant at the 5% level and 1% level respectively.

Figure 3.2: Correlation between Belief Factor Score (Anchored) and Cognitive Skill Factor Score (Anchored) by SES



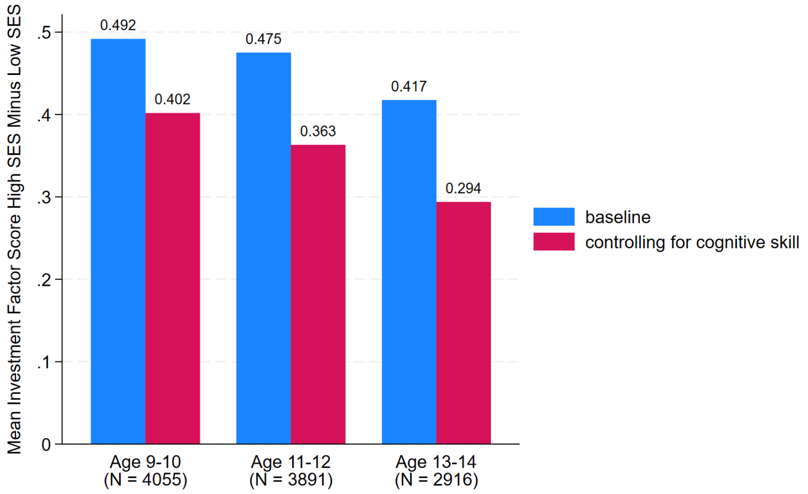
Notes: This figure presents the correlation between the belief factor score (anchored) and the cognitive skill factor score (anchored) at different ages of the child. High socio-economic status (SES) parents refer to parents who have above median family income (defined based on average family income when the child is aged between 9 and 10). Otherwise, parents are low SES. At each age, all children with non-missing belief factor scores, cognitive skill factor scores and average family income between age 9-10 are included in the computation.

In addition, on average, high SES parents invest more in their children, even after controlling for differences in children's cognitive skills. Figure 3.3 displays the average difference in the investment factor score (standardised) between the high SES and the low SES parents. Blue bars indicate the unconditional average difference, while the red bars display the average difference after accounting for the child's cognitive skill factor score at that age. Both the blue and red bars are positive, implying that high SES parents invest more in their children, even after accounting for child cognitive skills. The average difference in investment factor scores between high SES and low SES parents is 0.492 standard deviations at age 9-10. The SES differences in mean investment factor scores are statistically significant.

3.3.2 Parents Revise Beliefs When Cognitive Skills Change

This section provides suggestive evidence that parents revise their beliefs when the cognitive skill of their child changes: estimation of a belief updating model reveals that

Figure 3.3: Average Difference in Investment Factor Score (Standardised) Between High SES and Low SES



Notes: This figure presents the average difference in the investment factor score (standardised) between high SES and low SES parents at different ages of the child. The blue bar presents the unconditional difference in investment factor scores, while the pink bar is the difference after controlling for the cognitive skill factor score. High socio-economic status (SES) parents refer to parents who have above median family income (defined based on average family income when the child is aged between 9 and 10). Otherwise, parents are low SES. At each age, all children with non-missing investment factor scores, cognitive skill factor scores and average family income between age 9-10 are included in the computation.

within-child changes in the belief factor score are correlated with within-child changes in the cognitive skill factor score.

Empirical Strategy: The belief updating model is presented in Equation 3.3.1. The belief factor score of child j at period t (μ_{jt}) is assumed to depend linearly on the belief factor score in period $t-1$ ($\mu_{j,t-1}$) and the cognitive skill factor score of the child at period t (similar to Stinebrickner and Stinebrickner (2014)). A set of controls denoted by X_{it} are included. This set contains dummies for region of residence (north, south, east, west), urban/rural residence, SMSA status (not in SMSA, SMSA but not in central city, SMSA in central city, SMSA in unknown central city), an indicator for whether mother of the child is employed, an indicator for whether the child's biological father lives with the child, the education of mother's spouse, mother's marital status, log family income and the child's age at time t. A child fixed effect δ_i is introduced to account for time-invariant characteristics of the child which parents use to form beliefs. This means that the coefficients are estimated based on within-child variation in the lag belief factor score and the cognitive skill factor score.

$$\mu_{jt} = \alpha + \beta_1 \mu_{j,t-1} + \beta_2 cognitive_{jt} + \gamma X_{jt} + \delta_i + \epsilon_{jt} \quad (3.3.1)$$

As mentioned in the data section, the belief factor score (anchored) is interpreted as the mean of the parent's belief distribution over the cognitive skill level of their child. If the distribution over the skill is normal, this model (Equation 3.3.1) can be loosely motivated by Bayesian updating of the mean of a normal random variable, in which the posterior (updated) mean is a weighted average of the prior mean and the signal (new information received). The cognitive skill factor score is assumed to be a proxy for the signal which parents receive about the child skill: parents do not observe the cognitive skill measures collected by the survey, but these are likely to be correlated with school grades, which provide information about skill. Note that this model does not satisfy Bayesian updating because the weights on the prior mean and the signal cannot change with time, as they do in Bayesian updating.

Given the short T panel, since Equation 3.3.1 contains both a lag dependent variable and a fixed effect, estimating this equation via fixed effects or first differences will yield inconsistent estimates of the coefficient on the lag dependent variable. To obtain consistent estimates, the Blundell and Bond (1998) dynamic panel data estimator is employed. This is a system generalised method of moments estimator which relies on moment conditions relating instruments to both the level equation (Equation 3.3.1) and the differenced equation (Equation 3.3.2).

$$\Delta\mu_{jt} = \beta_1\Delta\mu_{j,t-1} + \beta_2\Delta cognitive_{jt} + \gamma\Delta X_{jt} + \Delta\epsilon_{jt} \quad (3.3.2)$$

Δ is the first difference operator

Because the lag dependent variable is endogenous, only values of the dependent variable from time period $t-2$ and earlier are used as instruments in the differenced equation. Since there may be unobserved shocks which jointly affect both the child cognitive skills and the belief factor score, the cognitive factor score is also treated as a potentially endogenous variable. Consequently, like the lag dependent variable, only values of the cognitive skill factor score from time period $t-2$ and earlier are used as instruments in the differenced equation.

The instruments used in the differenced equation are: $\mu_{j,t-2}$, $\mu_{j,t-3}$, $\mu_{j,t-4}$; $cognitive_{j,t-2}$, $cognitive_{j,t-3}$, $cognitive_{j,t-4}$; ΔX_{jt}

The instruments used in the level equations are: $\Delta\mu_{j,t-1}$, $\Delta cognitive_{j,t-1}$

Since the model is over-identified, the validity of over-identifying restrictions can be assessed with the Sargan test.

Sample: All available observations of children between ages 5 and 14 are used.

Estimates: Table 3.3 presents estimates of the coefficients on the lag belief factor score and the cognitive skill factor score in the belief updating model. First, estimates suggest that parents are using the previous information which they had about the child to form their current belief — the lag belief factor score is a positive and significant predictor of current belief. A one unit increase in the lag belief factor score is associated with a 0.211 increase in the current belief factor score. Second, the belief factor score is positively correlated with the cognitive skill factor score: a one unit increase in the cognitive skill factor score is associated with a 0.821 increase in the belief factor score. This suggests that parental beliefs are sensitive to changes in child cognitive skill.

A note of caution: the test of over-identifying restrictions is rejected at the 10% level, suggesting that the model is misspecified and/or the instruments are invalid.

Table 3.3: Dependent variable is Parent Belief Factor Score (Anchored)

	(1)
Lag Belief Factor Score	0.211***
	(0.034)
Cognitive Skill	0.821***
	(0.047)
Observations	5378
Number of Children	3141
Child FE	Yes
Cognitive Skill Potentially Endogenous	Yes
Sargan Test P-value	0.065

Notes: This table presents the estimated coefficients in a regression of the belief factor score (anchored) on the lag belief factor score (anchored) and the cognitive skill factor score. The construction of the belief factor score and the skill factor score is described in the data section. The model is estimated with the Blundell and Bond (1998) dynamic panel data estimator. The regression includes time-varying controls described in the empirical strategy. Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). One-step GMM standard errors are used. All available observations of the child between ages 5 and 14 are used.

3.3.3 Investments are Correlated with Beliefs

This section shows that parent investments are correlated with parental beliefs, even after accounting for time-invariant characteristics of children and investment preferences of parents.

Empirical Strategy: To understand the relationship between beliefs and investments, the investment factor score I_{jt} of child j at time t is regressed on the belief factor score μ_{jt}. The model is presented in Equation 3.3.3. Child fixed effects are introduced to control for child-specific time-invariant characteristics which jointly influence beliefs and investments. Furthermore, the lag investment factor score is included to account for persistent investment preferences of parents. In addition, a set of time-varying controls denoted by X_{it} are used. This set contains dummies for region of residence (north, south, east, west), urban/rural residence, SMSA status (not in SMSA, SMSA but not in central city, SMSA in central city, SMSA in unknown central city), an indicator for whether mother of the child is employed, an indicator for whether the child's biological father lives with the child, the education of mother's spouse, mother's marital status, log family income and the child's age at time t.

$$\underbrace{I_{jt}}_{\text{investment}} = \alpha + \beta \underbrace{\mu_{jt}}_{\text{belief}} + \gamma X_{jt} + \psi \underbrace{I_{j,t-1}}_{\text{lag investment}} + \underbrace{\delta_j}_{\text{child FE}} + \eta_{jt} \tag{3.3.3}$$

Despite the attempt to account for endogeneity due to omitted variables bias, the coefficient on the belief factor score is interpreted as a correlation, because there may still be unobserved shocks (e.g. to child skills) which affect both beliefs and investments. In addition, there may also be endogeneity due to reverse causality between beliefs and investments. The purpose of this exercise is to examine whether beliefs are related to investments, even after controlling for observables and time-invariant unobservables.

The equation contains both a lag dependent variable and a child fixed effect. Given the short T panel, the coefficient on the lag dependent variable will be inconsistently estimated with fixed effects or first difference estimators. To obtain consistent estimates, the Blundell and Bond (1998) dynamic panel data estimator (system GMM) is used. This system generalised method of moments estimator relies on moment conditions relating instruments to both the level equation (Equation 3.3.3) and the differenced equation (Equation 3.3.4).

$$\Delta \underbrace{I_{jt}}_{\text{investment}} = \beta\Delta \underbrace{\mu_{jt}}_{\text{belief}} + \gamma\Delta X_{jt} + \psi\Delta \underbrace{I_{j,t-1}}_{\text{lag investment}} + \Delta\eta_{jt} \tag{3.3.4}$$

The instruments used in the differenced equation are: $I_{j,t-2}$, $I_{j,t-3}$, $I_{j,t-4}$; $\Delta\mu_{jt}$; ΔX_{jt}

The instruments used in the level equations are: $\Delta I_{j,t-1}$

Since the model is over-identified, the validity of over-identifying restrictions can be assessed with the Sargan test.

Sample: All available observations of children between ages 5 and 14 are used. This sample is larger than the one used to estimate the belief updating model, because children tend to have investments reported over a greater number of time periods than they do for beliefs.

Estimates: Table 3.4 presents estimates of the coefficient on the belief factor score (anchored) in a regression of the investment factor score (standardised) on the belief factor score (anchored). The estimate suggests that a one unit increase in the belief factor score (anchored) is associated with a 0.082 standard deviation increase in the investment factor score (standardised). This correlation suggests that there is an important relationship between these variables — it could be that parental beliefs influence investment decisions. Furthermore, the positive correlation suggests that on average, parents invest more when they hold higher beliefs, as in Dizon-Ross (2019). Despite this finding, the model does not impose that parents invest more when they hold higher beliefs. Instead, the model is flexible enough to inform whether parents invest more or less when they hold higher beliefs in their child.

The test of over-identifying restrictions is rejected at the 5% level, suggesting that the model is misspecified and/or the instruments are invalid.

3.3.4 Summary of Data Patterns and Link to Model

Three data patterns are presented. First, there are socio-economic status differences in beliefs and investments. Low SES parents are more likely to hold inaccurate beliefs. In addition, high SES parents hold higher beliefs and invest more, even after accounting for children's skill levels. The model will be used to explore whether differences in beliefs can explain differences in investments and skills across families. Second, there is suggestive evidence that parental beliefs change when the cognitive skill of their child changes. Therefore, the model incorporates belief updating. Third, parent investments

Table 3.4: Dependent variable is Investment Factor Score (Standardised)

	(1)
Belief Factor Score	0.082***
	(0.017)
Observations	9557
Number of Children	4136
Child FE	Yes
Lag Investment	Yes
Sargan Test P-value	0.019

Notes: This table reports coefficients on the belief factor score (anchored) in a regression of the investment factor score (standardised) on the belief factor score (anchored). The Blundell and Bond (1998) dynamic panel data estimator is used. The construction of the belief factor score and the investment factor score is provided in the data section. The regression includes time-varying controls which are explained in the empirical strategy. Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). One-step GMM standard errors are used. All available observations of the child between age 5 and 14 are used.

are positively correlated with parental beliefs, even after accounting for time-invariant characteristics of children. This indicates that there is an important relationship between beliefs and investments, and it could be that beliefs are influencing investment choices.

On their own, these data patterns cannot inform us how adjusting beliefs in earlier periods will affect beliefs, skills and investments in later periods. A framework which unifies beliefs, investments and skills is required. In the next section of the paper, such a framework is introduced: a dynamic model of parent investments with belief updating. The model connects belief updating with how beliefs affect investments and how investments affect skills.

3.4 Model

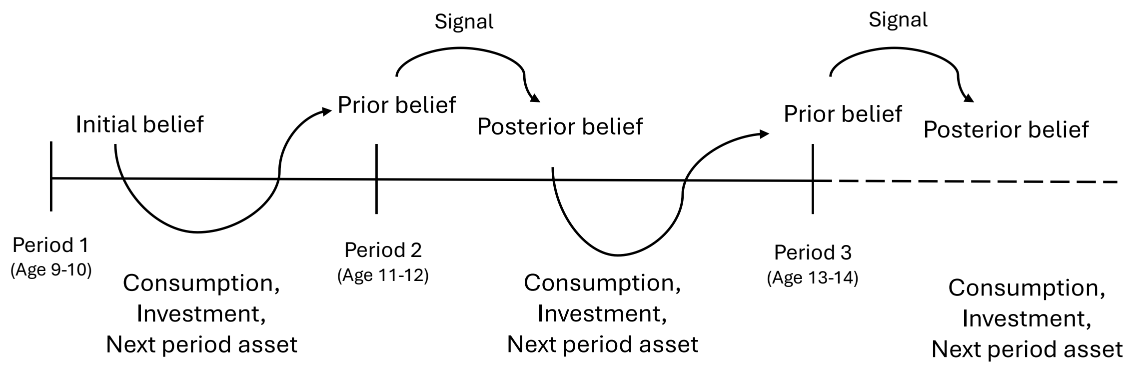
3.4.1 Model Overview

This section introduces a dynamic model of parent investments which incorporates parental beliefs about the skill level of their child. The model connects parent belief updating with how beliefs influence investments and how investments in turn affect child cognitive skill.

In the model, each parent has a single child. The model has 3 periods which are each 2 years in length, corresponding to the child being aged 9-10, 11-12 and 13-14. In the first period, the parent enters with an initial belief about the child's cognitive skill and then makes choices for the period: consumption, investment and next period assets. In periods 2 and 3, parents enter with a prior belief about the cognitive skill of the child. Then, they receive a signal (noisy measure of the child's cognitive skill) and update their belief, obtaining a posterior belief. Given this updated belief, parents make their choices for the period: consumption, investment and next period assets. After this, they form their prior belief about the skill for the following period, based on the skill production function, their investment and their updated belief. The sequence of events is presented in Figure 3.4.

Parents are forward looking and they are expected utility maximisers. Their utility depends on consumption and the perceived cognitive skill of their child (since skill is unobserved). Parents are motivated to invest, because investments produce skills, raising their utility (Solon, 1999). They choose consumption, investments and next period as-

Figure 3.4: Model Overview



sets to maximise their objective function, subject to the following processes: the child skill production function, the belief updating process, the resource constraint and the family income process. Parents face uncertainty in the signal generation process, the skill production process and the income process.

The model features observed heterogeneity driven by mother's education (high/low). Mother's education affects the skill production function, belief updating process and income process. High education mothers are more likely to earn higher income. In addition, they may update beliefs differently and have a different level of productivity in terms of converting investments to skills.

The model also features unobserved preference heterogeneity. Parents can be one of two discrete types, $h \in \{1, 2\}$. One type has a higher relative value for child skill over consumption and is more likely to hold a higher prior belief in the first period. The probability of being either type depends on the initial conditions of the parent-child pair (initial skills of child). The unobserved heterogeneity is introduced for two reasons. First, to control for unobservable factors which jointly affect both parental beliefs and parent investments. Second, to address the problem of endogeneous initial conditions: it is unlikely that the beliefs, income, assets and skills of children at age 9-10 (when the model starts) are randomly assigned.

Parent-child pairs are heterogeneous in terms of their beliefs about the cognitive skill level of their child, income, assets, mother's education and the unobserved discrete type.

Next, I will explain the utility function, the skill production function, beliefs and the belief updating process, the resource constraint and the income process in greater detail. In what follows, the parent-child pair will be indexed by j. Time t indexes the age of the child. Mother's education category is denoted by s, where $s = 1$ represents low education (up to median education: up to 12 years education) and $s = 2$ represents high education (above median education: above 12 years education). The unobserved discrete type is represented by $h \in \{1, 2\}$.

3.4.2 Utility

Parent utility depends on consumption c_{jt} and the perceived cognitive skill level of the child $\tilde{\theta}_{jt}$ (since skill is unobserved). Parents are motivated to invest because investments will raise the perceived skill of the child, which in turn raises their utility. Including the skill level of the child in the flow utility of the parent follows Del Boca, Flinn, and Wiswall (2014) and Del Boca, Flinn, Verriest, and Wiswall (2019).

I assume that the utility is additive and separable in terms of logs in both consumption and the perceived skill level of the child. The weight κ_h on the utility from the perceived skill is the relative value of child skill over consumption. This relative value depends on the unobserved discrete type $h \in \{1, 2\}$ — one type of parent has a higher relative value of child skill over consumption. The tilde over the expectations operator indicates that the expectations are computed using the parent's (subjective) belief about the skill level of their child. The belief is a distribution over the skills of the child. In a later section, it will be explained in greater detail.

$$u\left(\underbrace{c_{jt}}_{\text{consumption}}, \underbrace{\tilde{\theta}_{jt}}_{\text{perceived skill}}\right) = \ln c_{jt} + \underbrace{\kappa_h}_{\substack{\text{relative preference for} \\ \text{child skill over consumption}}} \times \underbrace{\tilde{E}_t \ln \theta_{jt}}_{\substack{\text{utility from} \\ \text{perceived skill}}} \tag{3.4.1}$$

3.4.3 Child Skill Production Technology

Next period log cognitive skills of child j, represented by $\ln \theta_{j,t+1}$, depend on current period log skills $\ln \theta_{jt}$ and current period log investments $\ln i_{jt}$ according to a translog production function (Equation 3.4.2).

$$\ln \theta_{j,t+1} = A_t + \gamma_{s,1} \ln \theta_{jt} + \gamma_{s,2} \ln i_{jt} + \gamma_{s,3} \ln \theta_{jt} \times \ln i_{jt} + \eta_{j,t+1} \tag{3.4.2}$$

There is an age-specific total factor productivity A_t which captures exogenous time-varying factors which influence child skills aside from parent investment (e.g. school education). This total factor productivity is assumed to be equal at age 11-12 and age 13-14. The production of skills is subject to an idiosyncratic shock η_{t+1}, which is normally distributed with mean 0 and variance $\sigma^2_{s,\eta}$.

In the production function, the inclusion of an interaction between the log current skills and the log investment allows the productivity of investment to depend on the current skill of the child. For example, if $\gamma_{s,3}$ is positive, investments are more productive for children with higher skills.

$\gamma_{s,3}$ is a key parameter in the model affecting the feedback between beliefs and investments. Given the log utility and the translog production function, if $\gamma_{s,3}$ is positive (negative), parents will invest more (less) in their child when they hold higher beliefs. This is because when beliefs are higher, a positive (negative) $\gamma_{s,3}$ raises (lowers) the perceived marginal return to investment[5].

Except for the total factor productivity, parameters of the skill production function depend on mother's education $s \in \{1, 2\}$. This allows, for example, for high education mothers to be more productive in producing child skills from investments.

Though parents know the skill production technology (including the return to investment), they do not observe the skill level of their child, which is an input in the production function. Therefore, they need to take expectations with respect to the current log skill $(\ln \theta_{jt})$ to predict how their investments will affect the next period log skill.

3.4.4 Belief about Child Cognitive Skill and Belief Updating

Parents do not observe the log cognitive skills of the child $\ln \theta_{jt}$. They enter each period with a belief about the log skill, which is a distribution over the log skill of their child $\ln \theta_{jt} \sim \mathcal{N}(\mu_{jt}, \Delta_{jt})$.

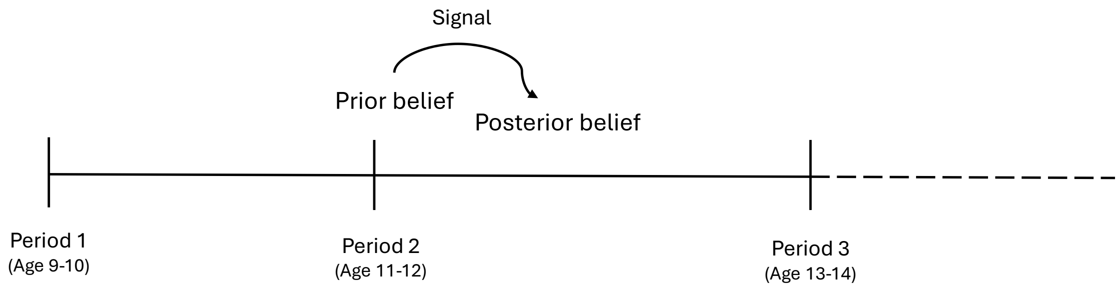
Though the belief is represented as a distribution with both a mean and a variance, this section focuses on how the mean of the distribution evolves. Variances in the belief distributions are assumed to be exogenous. They depend on mother's education and time, but are not individual-specific. In the estimation, they are set equal to the variance of the belief factor score at the corresponding ages. Parents know the variance they face in future periods when make their decisions.

[5] This can be seen from the first order condition with respect to investment in a simplified version of this model.

Next, these two steps are explained: (1) belief updating and (2) formation of prior belief about skill after making investment choice.

Step One: Receive Signal and Update Belief

Figure 3.5: Receiving Signal and Updating Belief



Parents receive a signal g_{jt}, which is a noisy measure of the log skill $\ln\theta_{jt}$. The idiosyncratic error term in the signal $\epsilon_{j,t+1}$ is normally distributed with mean 0 and variance σ_ϵ^2.

$$g_{j,t+1} = \ln\theta_{j,t+1} + \epsilon_{j,t+1}, \epsilon_{j,t+1} \sim \mathcal{N}(0,\, \sigma_\epsilon^2) \tag{3.4.3}$$

The assumption that all parents have the same signal precision may be a strong one. For example, high SES parents could receive more precise signals if they spend more time with their children or their children attend high quality schools, which provide more frequent feedback on performance.

After receiving the signal, parents update their belief and obtain a posterior belief, which is a posterior distribution over the log skill of the child $\ln\theta_{jt} \sim \mathcal{N}(\mu_{jt}^{int}, \Delta_{jt}^{int})$.

$$\mu_{jt}^{int} = \psi_s \mu_{jt} + (1-\psi_s)\, g_{jt} \tag{3.4.4}$$

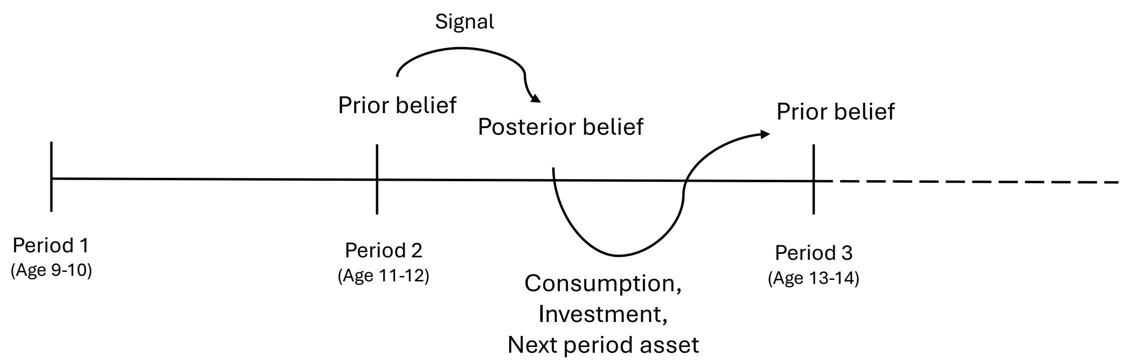
Beliefs are updated in the following way: The mean posterior belief μ_{jt}^{int} is a weighted average of the mean prior belief μ_{jt} and the signal g_{jt}. High education mothers may update beliefs differently from low education mothers[6] — the weights on the mean prior belief and the signal depend on mother's education category s.

Step Two: Choose Investment and Form Prior Belief About Next Period Skill

Given the posterior belief, the parent makes choices for the period: consumption, in-

[6]Dizon-Ross (2019) provides evidence that belief updating depends on SES.

Figure 3.6: Making Decisions and Forming Prior Belief About Next Period Skill



vestment and next period assets. Then, the parent forms their prior belief about the next period log skills $\ln\theta_{j,t+1} \sim \mathcal{N}(\mu_{j,t+1}, \Delta_{j,t+1})$.

The mean of this belief distribution is the parent's expected value of the next period skills. The parent faces uncertainty about the next period log skills because the parent does not observe the current log skills and there is an idiosyncratic error in the production function. Thus, to obtain an expected value, the parent must take expectations over the current period log skill and the error term in the production function.

Specifically, the mean of the distribution of next period log skill $\mu_{j,t+1}$ equals the parent's expected value of log skill ($\ln\theta_{j,t+1}$), as in Equation 3.4.5 below. The tilde over the expectation operator indicates that the expectations are computed based on parent's posterior belief about the skill level of the child, which is a subjective belief. From Equation 3.4.5, parents will hold a higher next period prior mean belief $\mu^{int}_{j,t+1}$ if they invest more. Investing more leads to higher beliefs, because investments produce skills, and this is a crucial part of the interaction between beliefs and investments in this model.

$$\begin{aligned}
\mu_{j,t+1} &\equiv \tilde{E}\left[\ln\theta_{j,t+1}|\mu^{int}_{jt}, i_{jt}\right] \\
&= \tilde{E}\left[A_t + \gamma_{s,1}\ln\theta_{jt}|\mu^{int}_{jt}, i_{jt}\right] + \tilde{E}\left[\gamma_{s,2}\ln i_{jt} + \gamma_{s,3}\ln\theta_{jt}\times\ln i_{jt} + \eta_{j,t+1}|\mu^{int}_{jt}, i_{jt}\right] \\
&= A_t + \gamma_{s,1}\tilde{E}\left[\ln\theta_{jt}|\mu^{int}_{jt}, i_{jt}\right] + \gamma_{s,3}\ln i_{jt}\tilde{E}\left[\ln\theta_{jt}|\mu^{int}_{jt}, i_{jt}\right] \\
&\quad + E\left[\gamma_{s,2}\ln i_{jt} + \eta_{j,t+1}|\mu^{int}_{jt}, i_{jt}\right] \\
&= A_t + (\gamma_{s,1} + \gamma_{s,3}\ln i_{jt})\,\tilde{E}\left[\ln\theta_{jt}|\mu^{int}_{jt}, i_{jt}\right] + \gamma_{s,2}\ln i_{jt} \qquad (3.4.5)
\end{aligned}$$

Unlike many learning models, the belief updating structure is not Bayesian. To see this, note that in Bayesian updating of a normal random variable, the mean of the posterior is a weighted average of the mean prior belief and the signal, where the weights depend on the relative informativeness of the prior versus the signal. The updating weights are different in every period, because the variance of the prior changes each period. Since in this model, the updating weights remain the same over time, this model does not satisfy Bayesian updating.

Readers may be wondering why this model departs from Bayesian updating, especially as other dynamic models with parental beliefs impose this assumption (see Kinsler and Pavan (2016) and Chikhale (2024)). An earlier version of this paper assumed that parents updated their beliefs in a Bayesian way. The model fit was poor. The model predicted that parents learn at a faster rate than the data suggests: the estimated parameters imply a significantly higher correlation between beliefs and skills in the later periods, compared to the data.

3.4.5 Resource Constraint and Income Process

In every period t, parents face a budget constraint. The sum of consumption, investment and savings (in terms of assets in the next period) is equal to the available resources.

$$c_{jt} + i_{jt} + a_{j,t+1} = (1+r)\, a_{jt} + y_{jt} \tag{3.4.6}$$

The income process of parents is as follows. Log family income is a cubic function of the age of the child. The residuals of this equation follow an AR(1) process. These equations describe two-year income processes, given that each period is two years long in the model. The parameters of the family income process depend on mother's education category $s \in \{1, 2\}$, because families with high education mothers tend to earn higher income.

$$\ln(y_{j,t+1}) = \tau_{s,0} + \tau_{s,1} age_{j,t+1} + \tau_{s,2} age^2_{j,t+1} + \tau_{s,3} age^3_{j,t+1} + \nu_{j,t+1} \tag{3.4.7}$$

$$\nu_{j,t+1} = \rho_s \nu_{jt} + u_{y,jt}, u_{y,jt} \sim \mathcal{N}(0,\, \sigma^2_u) \tag{3.4.8}$$

The combination of concave utility and income uncertainty generates a motive for precautionary savings.

3.4.6 Value Function

The state variables of the parent are the following: (1) belief about the child cognitive skill — this is a distribution with mean μ_t^{int} and variance Δ_t^{int}, (2) income y_t, (3) assets a_t, (4) mother's education category $s \in \{1, 2\}$, where $s = 1$ denotes low education (up to 12 years education) and $s = 2$ denotes high education (above 12 years education) and (5) unobserved discrete type $h \in \{1, 2\}$. Type $h = 2$ is what I refer to as the "high" type parent. This type holds a higher relative preference for child skill over consumption.

In the first and second period, the value function of the parent is shown in Equation 3.4.9. Parents choose a sequence of consumption, investment and next period assets to maximise the sum of their flow utility and the discounted expected value of the next period value function. The tilde over the expectations operator indicates that the expectation of the parent depends on their belief about the child's skill level, which is subjective.

$$\begin{aligned} V_t\left(\mu_t^{int}, \Delta_t^{int}, a_t, y_t, s, h\right) = \max_{c_t, i_t, a_{t+1}} & \ln c_t + \kappa_h \tilde{E}_t \ln \theta_t \\ & + \beta \tilde{E}_t V_{t+1}\left(\mu_{t+1}^{int}, \Delta_{t+1}^{int}, a_{t+1}, y_{t+1}, s, h\right) \end{aligned} \tag{3.4.9}$$

The value function in the third and final period is shown in Equation 3.4.10. The investment decision in this period determines the child's final human capital θ_{T+1}. As before, the tilde over the expectations operator indicates that the expectation is computed based on the parent's belief about the child's skill, which is subjective. It is assumed that the resources which have not been consumed by the final period will be consumed and discounted by factor β. Since this is the final period, no income or signal about the child's skill is received in the following period. Therefore, the parent does not face uncertainty stemming from the income process or the signal generation process.

$$\begin{aligned} V_T\left(\mu_T^{int}, \Delta_T^{int}, a_T, y_T, s, h\right) = \max_{c_T, i_T, a_{T+1}} & \ln c_T + \kappa_h \tilde{E}_T \ln \theta_T + \beta \kappa_h \tilde{E}_T \ln \theta_{T+1} \\ & + \beta \ln\left((1+r)\, a_{T+1}\right) \end{aligned} \tag{3.4.10}$$

3.4.7 Initial Conditions and Unobserved Heterogeneity in Preferences

To understand the role of parental beliefs, the investments and skills of children whose parents hold higher beliefs are compared to those with lower beliefs. One concern is that parents who hold higher beliefs may be systematically different from those who hold lower beliefs. For instance, parents who hold higher beliefs may also have higher income and assets, and/or children with higher skills. These systematic differences are controlled for in the following ways.

Firstly, following French and Jones (2011), the initial conditions of the parent-child pair (beliefs of parents, income of parents, assets of parents and initial skills of children) are obtained directly from the data. This means that if parents who hold higher beliefs are more likely to have higher income, assets and children with higher skills, this will also be the case in the distribution of simulated individuals used to estimate the model.

Secondly, unobserved preference heterogeneity in the style of Heckman and Singer (1984), Keane and Wolpin (1997) and Van der Klaauw and Wolpin (2008) is introduced. It may be that differences in unobservable preferences are driving differences in both the beliefs and investments. Furthermore, these underlying preferences are likely to be correlated with the initial conditions of parents and children. This is because parents who held different preferences would have invested differently in earlier periods, before the model begins at age 9-10. A way to adjust for this selection is to solve the optimisation problem from an earlier age (before age 9-10 when the model in this paper begins) and then assess whether the model can replicate the correlations between preferences and conditions in the first period of the model. However, this approach would require assumptions such as the skill production function parameters being identical at earlier ages, which contradicts findings in the literature (see for example, Cunha et al. (2010) and Attanasio et al. (2020)). In addition, this approach is computationally intensive (French & Jones, 2011). Therefore, I instead follow French and Jones (2011) and model the correlation between preferences and a subset of initial conditions. Next, I explain how I model this in greater detail.

In the model, unobservable heterogeneity takes the form of discrete types. Parents can be one of two discrete types, where one type has a higher relative preference for child skill over consumption. The probability of being either type is determined by a logistic function of the log skill of the child in the first period[7]. The idea is that parents with higher preferences would have invested more in earlier periods, leading to the child having higher log skills in the first period. The parameter in this type probability function π_0 is estimated together with other parameters in the method of simulated moments step explained in the next section.

$$Pr\left(\text{Unobserved Type } h = 2|\ln\theta_{j1}\right) = \frac{exp\left(\pi_0 \ln\theta_{j1}\right)}{1 + exp\left(\pi_0 \ln\theta_{j1}\right)} \tag{3.4.11}$$

3.5 Identification and Estimation

This section discusses identification of the model, data preparation followed by the estimation of the model.

3.5.1 Discussion of Identification

The family income process can be identified directly from moments in the data. Next, the identification of the distribution of the latent factors (beliefs, log skills and investment) and the skill production function parameters is discussed.

At each time period t, there are $L_t^I \geq 3$ noisy measurements of investments for child j. Each measurement b_{jkt} for child j at time period t is assumed to be related to the latent investment according to the following equation:

$$b_{jkt} = \alpha_{kt} + \lambda_{kt}\Phi_{jt} + v_{jkt}, k \in \left\{1, 2, ...L_t^I\right\} \tag{3.5.1}$$

The errors v_{jkt} are assumed to be independent across individuals and measures. In a similar way, there are $L_t^\theta \geq 3$ measurements of the latent log cognitive skills $\ln\theta_{jt}$ and $L_t^\mu \geq 3$ measurements of the latent parental belief μ_{jt}^{int} for $t = 1, 2, 3$. Given this, the distribution of $\left(\left\{\Phi_{jt}\right\}_{t=1}^3, \left\{\ln\theta_{jt}\right\}_{t=1}^3, \left\{\mu_{jt}^{int}\right\}_{t=1}^3\right)$ can be identified according to arguments in Cunha et al. (2010).

The belief factors and log skill factors do not have natural location and scale. I follow Cunha et al. (2010) and anchor these factors to the years of education of the child at age 24 and above. In addition, the investment factor score Φ_{jt} is not in monetary units, while the investment in the model is in monetary units. To make both measures

[7] In an earlier version of the model, the type probability depended on the first period beliefs, income of parents, assets of parents and the skill of the child. But the coefficients on the first period beliefs, income and assets were very small, and some were not well identified.

comparable, I adopt the following method from Caucutt and Lochner (2020). I assume that there is a parametric mapping function $\phi(.)$ which connects the investment factor Φ_{jt} to the monetary investment i_{jt} each period. That is, $\Phi_{jt} = \phi(i_{jt})$, where the known function $\phi(.)$ is monotone, so that higher values of the investment factor correspond to higher values of monetary investment. Consequently, given the parametric assumption, the function can be inverted to obtain investment in monetary units i_t. And therefore we can retrieve $\ln i_t$ and its associated distribution.

Since the $\phi(.)$ mapping function is known, we can substitute $i_{jt} = \phi^{-1}(\Phi_{jt})$ into the production function. Given the distribution $\left(\{\Phi_{jt}\}_{t=1}^{3}, \{\ln\theta_{jt}\}_{t=1}^{3}, \{\mu_{jt}^{int}\}_{t=1}^{3}\right)$ and the parametric form of the production function, we can identify the parameters in the skill production function and investment mapping function.

$$
\begin{aligned}
&E\left[\ln\theta_{j,t+1} | \ln\theta_{jt}, \Phi_{jt}\right] \\
&\quad = E\left[A_t + \gamma_{s,1}\ln\theta_{jt} + \gamma_{s,2}\ln\phi^{-1}(\Phi_{jt}) + \gamma_{s,3}\ln\theta_{jt}\times\ln\phi^{-1}(\Phi_{jt}) + \eta_{jt} | \ln\theta_{jt}, \Phi_{jt}\right]
\end{aligned}
$$

The variance of the error term in the skill production technology can be identified as $\sigma_{s,\eta}^2 = Var(\ln\theta_{j,t+1}) - Var(\gamma_{s,1}\ln\theta_{jt} + \gamma_{s,2}\ln\phi^{-1}(\Phi_{jt}) + \gamma_{s,3}\ln\theta_{jt}\times\ln\phi^{-1}(\Phi_{jt}))$. The variances on the right hand side of the equation are identified from the distribution of the latent factors of log skills and investments, along with the mapping from the investment factor to monetary investment (which enables us to retrieve the distribution of the log investment).

However, identification will fail if the error term in the production function is correlated with any of the input arguments. This may occur, for example, if parent investments respond to an unobserved shock to the skill of the child.

Now, I discuss the remaining parameters to be identified. In theory, all parameters are jointly identified by the moments. However, some moments may be particularly informative for certain parameters (parameters have first order effects on these moments, rather than only second order effects).

The relative preference for child skill over consumption $\{\kappa_1, \kappa_2\}$ will be particularly informed by moments of the ratio of next period assets to the current period investment factor score. This is because parents who have a higher preference for child skill over consumption will invest more and save less (lower next period assets). This can be derived from the optimal solution to a simplified static version of the model. Note that this also works because I assume that the price of monetary investment is equal to the price of consumption (both normalised to 1), so that the relative price of investment will not enter the first order condition. The relative preferences are also particularly informed by moments of the investment factor score, since parents with higher preferences will invest more.

The weight on the prior belief in belief updating $\{\psi_{s=1}, \psi_{s=2}\}$ and the variance of the error term in the signal σ_ϵ^2 are particularly informed by the coefficients and residuals of a regression of the belief factor score on the lag skill factor score, lag investment factor score and the skill factor score. This regression is derived from an approximation of the belief updating rule. Details are provided in Appendix 3.D.3.

The type probability parameter π_0 is particularly informed by two categories of moments. First, the standard deviation of the residuals in a regression of the investment factor score on the belief factor score. The idea is that conditional on observables, the variation in investment is driven by unobserved heterogeneity. Second, the autocorrelation of the investment factor score conditional on the quartile of log skills of the child, which provides information on the “fixed effect" of unobserved preferences.

Table 3.5 summarises the data features which are most informative for the estimation of each of the parameters. The belief factor scores, log skill factor scores and investment factor scores are treated as data in the estimation of the model, even though they are quantities estimated from the raw data with factor models[8].

3.5.2 Data Preparation

Sample selection for model estimation and calculation of data moments: The focus is on the children of the females in the cross-section sample of the National Longitudinal Survey of Youth 1979 (NLSY79). To be included in the estimation, the belief factor score, income, assets and child skills must be observed in the first period of the model, which is age 9-10. These variables can be missing in other periods. Assuming that the values are missing at random, the data moments are computed from the relevant non-missing values. For example, the average belief factor score at age 11-12 is the average of all non-missing belief factor scores at age 11-12. As the model does not allow parents to borrow (assets must be weakly positive), observations

[8] Treating the predicted factor scores as data in the estimation affects the standard errors. In the future, this could be addressed by a bootstrap procedure where each individual's factor score is drawn from the estimated distribution.

Table 3.5: Summary of Identification of Estimated Parameters

Parameter	Identifying Data Features
$\kappa_{h=1}, \kappa_{h=2}$	Ratio of next period assets to investment factor score, moments of investment factor score
$A_1, A_2, \gamma_{s=1,1}, \gamma_{s=1,2}, \gamma_{s=1,3}, \gamma_{s=2,1}, \gamma_{s=2,2}, \gamma_{s=2,3}, \sigma^2_{s=1,\eta}, \sigma^2_{s=2,\eta}$	Auxiliary model: Regression of log skill factor score on previous period log skill factor score, previous period investment factor score and interaction between previous period log skill factor score and previous period investment factor score
$\psi_{s=1}, \psi_{s=2}, \sigma^2_{\epsilon}$	Auxiliary model: Regression of the belief factor score on previous period log skill factor score, previous period investment factor score and the current log skill factor score (this model is derived from approximation of belief updating rule)
π_0	Residuals of regression of investment factor score on belief factor score and the autocorrelation of investment factor score conditional on quartile of log skill factor score

Notes: This table provides a summary of the moments which are particularly informative of the parameters estimated in the method of simulated moments step. I treat the belief factor scores, log skill factor scores and investment factor scores as data in the method of simulated moments procedure, even though they are predicted values from factor models.

with negative values of assets at age 9-10, 11-12 or 13-14 are excluded. Additionally, children in families which reported zero family income at age 9-10 are dropped. On top of this, only children whose mothers were married when the child was aged between 5 and 14 are included[9]. After selection, there are 655 parent-child pairs.

Factor models and anchoring: As described in the previous chapter, factor models are used to extract the latent parental beliefs, investments and log cognitive skills of the child from multiple noisy measures of these latent variables. After estimating the factor models, predicted factors of parental beliefs, investments and log cognitive skills are obtained. These are treated as data in the two-step estimation procedure described in Section 3.5.3. Note that it is possible to compare the belief factor score and the skill factor score in terms of levels, because they have both been anchored to years of education of the child at age 24 and above.

The belief factor score is assumed to be equal to the mean of the parent's posterior belief distribution (updated belief), rather than the prior belief distribution. This is because when beliefs were reported, it is likely that the parent had already received signals about the skill level of the child.

Converting investment factor scores to investment in monetary units: Since investment is in monetary units in the model, to bring the model to the data, we need a way to translate the investment factor score (not in monetary units) to investment in monetary units. To achieve this, as in Caucutt and Lochner (2020), we assume that there is a quadratic function $\phi(.)$, which maps the investment factor score Φ_{jt} to investment in monetary units i_{jt} for child j at time t. That is, $\Phi_{jt} = \phi(i_{jt}) = \phi_0 + \phi_1 i_{jt} + \phi_2 i_{jt}^2$. The function is monotone, so that $\phi'(.) > 0$, which means higher factor scores correspond to higher monetary investment. For the time being, the parameters of this function are set at the following values: $\phi_0 = -18.23 \times 10^{-2}$, $\phi_1 = 4.46 \times 10^{-6}$ and $\phi_2 = 5.62 \times 10^{-14}$.

3.5.3 Estimation

Estimation proceeds in two steps. In the first step, some parameters are calibrated or estimated directly from the data. These parameters are then used as inputs in the second step, where the remaining parameters are estimated via method of simulated moments.

[9]The reason is that I want to focus on intact families. Furthermore, allowing a change in marital status could result in the parent income process becoming too volatile.

Step One: Calibration and External Estimation

Following Del Boca et al. (2014), the discount factor for households is set to $\beta = 0.95$. The corresponding interest rate is $r = \frac{1}{\beta} - 1$. The family income process is estimated directly from the data. Estimates of the following parameters are obtained and they are used as inputs in the second step of the estimation.

$$\left\{\tau_{s=1,0}, \tau_{s=1,1}, \tau_{s=1,2}, \tau_{s=1,3}, \rho_{s=1}, \sigma^2_{s=1,u}, \tau_{s=2,0}, \tau_{s=2,1}, \tau_{s=2,2}, \tau_{s=2,3}, \rho_{s=2}, \sigma^2_{s=2,u}\right\}$$

Step Two: Method of Simulated Moments

In step two, the remaining parameters are estimated via method of simulated moments: the relative preference for child skill over consumption, the parameters in the skill production function, the parameters relating to parental beliefs and belief updating and the parameter in the type probability function of the unobserved discrete types. The set of parameters to estimate is Γ defined below. Mother's education is denoted by s, where $s = 1$ represents low education and $s = 2$ represents high education. The unobserved discrete type is represented by $h \in \{1, 2\}$.

$$\Gamma \equiv \left\{\kappa_{h=1}, \kappa_{h=2}, A_1, A_2, \gamma_{s=1,1}, \gamma_{s=1,2}, \gamma_{s=1,3}, \sigma^2_{s=1,\eta}, \ldots \right.$$
$$\left. \gamma_{s=2,1}, \gamma_{s=2,2}, \gamma_{s=2,3}, \sigma^2_{s=2,\eta}, \sigma^2_{\epsilon}, \psi_{s=1}, \psi_{s=2}, \pi_0\right\}$$

Given a guess of parameters, the optimisation problem of the parent is solved numerically by backward induction, beginning from the final period T. This provides the policy functions (choices of investment, consumption and next period assets given the values of the state variables). Next, I forward simulate the trajectories of 5 $\times$ N parent-child pairs. At the beginning of the simulation, an unobserved type is drawn for each parent. Then, in each period, shocks are drawn and the policy functions are used to determine the choice variables and the values of state variables in the following period. After running the simulation, moments are computed from the simulated data ($M(\Omega)$). Finally, the GMM criterion function (right hand side of Equation 3.5.2) is computed, using the distance between the moments from the simulated data ($M(\Omega)$) and the corresponding moments from the sample data (M_D). The steps are repeated until the parameter values which minimise the GMM criterion function are found.

The method of simulated moments estimator is given by Equation 3.5.2, where M_D are the data moments and $M(\Omega)$ are the corresponding moments derived from the simulated data, which are generated from the parameters Ω.

$$\hat{\Omega} = \arg\min_{\Omega} \left(M_D - M(\Omega)\right)' W \left(M_D - M(\Omega)\right) \tag{3.5.2}$$

The variance-covariance matrix of the moments S is computed via bootstrap with $K = 200$ replications according to Equation 3.5.3. The formula involves the moments computed from the data (M_D) and the corresponding moments computed from K resamples of the data $(M_1, M_2, ...M_K)$.

$$S = \frac{1}{K} \sum_{k=1}^{K} \left(M_k - M_D\right) \left(M_k - M_D\right)' \tag{3.5.3}$$

The asymptotically optimal weighting matrix is the variance-covariance matrix of the moments. However, using the optimal matrix may not be ideal in small samples (Altonji & Segal, 1996). Instead, the weighting matrix W used in the estimation is the matrix which retains the diagonal entries of the inverse of S and has zeros on the off-diagonal elements. Standard errors for the parameter estimates are computed using the formula derived by Gourieroux, Monfort, and Renault (1993), which adjust for the simulation noise.

Moments relating beliefs, skills and investments are used to estimate the model. For example, the mean and standard deviation of skill factor scores, investment factor scores and assets. Both unconditional and conditional moments (conditional on quartile of family income, quartile of skills in the previous period) are included. Contemporaneous covariances or correlations between variables (skills factor score, investment factor score, belief factor score) and auto-correlations of variables are also included. Statistics from auxiliary models are also used. An example of an auxiliary model is a regression of the skill factor score on the lag skill factor score, lag investment factor score and an interaction between the lag skill factor score and the lag investment factor score. A comprehensive list of the moments used in the estimation is provided in Appendix 3.E.

3.6 Model Estimates

In this section, I present the parameter estimates, discuss the model fit and explore features of the baseline.

3.6.1 Parameter Estimates

The parameters in the income process are estimated directly from the data (step one). The remaining parameters are estimated via method of simulated moments (step two). Standard errors for the estimates in the second step are computed using the formula in Gourieroux et al. (1993), which adjusts for simulation noise.

Preferences and Type Probabilities: Table 3.6 presents the estimates of the parameters related to preferences and the unobserved heterogeneity. In the model, parents are one of two unobserved discrete types, where one type (the “high" type, which is type $h = 2$) has a higher relative preference for child skill over consumption. In the simulated data, parents are more likely to be the “high" type (around 63.1%). The estimate of $\kappa_{h=1}$ is statistically different from $\kappa_{h=2}$. Table 3.C.1 in Appendix 3.C provides the corresponding test statistic and p-value.

Parents of children with higher first period log skills are more likely to be the “high" type: the estimate of π_0, the coefficient on the log skill in the type probability, is positive. Intuitively, parents with higher relative preferences for skills would have invested more in earlier periods, which means that their children have higher log skill in the first period of the model.

Table 3.6: Preference Parameters and Type Probability Parameter

Parameter	Description	Estimate	S.E.
$\kappa_{h=1}$	Relative preference for skill over consumption: Type 1	2.021	0.01972
$\kappa_{h=2}$	Relative preference for skill over consumption: Type 2	2.776	0.11714
π_0	Type probability logistic: coefficient on log skill	0.500	1.22743

Skill Production Function: Table 3.7 presents the estimates of the parameters in the skill production function. Cognitive skills at age 9-12 are rather persistent: the weight on the current log skills in the production function is around 0.7 and above. This corroborates estimates of child skill production functions (see for example Cunha et al. (2010) and Attanasio et al. (2020)).

Skill production parameters depend on mother's education. First, cognitive skills are more persistent for children of high education mothers — the weight on log skill is greater for high education mothers. Second, high education mothers are more productive in converting investments to cognitive skills — the weight on log investment is greater for high education mothers. However, this difference is not statistically significant. Third, complementarity between skills and investments is higher for low education mothers — the coefficient on the interaction term between log skill and log investment is greater for low education mothers. Table 3.C.2 in Appendix 3.C provides the test statistics and corresponding p-values for statistical tests of the equality of skill production function parameters between high and low education mothers.

Table 3.7: Child Skill Production Function Parameters

Parameter	Description	Estimate	S.E.
A_1	Age 9-10 total factor productivity	0.337	0.02427
A_2	Age 11-12, 13-14 total factor productivity	0.255	0.04731
Low Education ($s = 1$)			
$\gamma_{s=1,1}$	Weight on log skill	0.703	0.01045
$\gamma_{s=1,2}$	Weight on log investment	0.038	0.01245
$\gamma_{s=1,3}$	Weight on interaction between log skill and log investment	0.025	0.00015
$\sigma^2_{s=1,\eta}$	Variance of idiosyncratic shock	0.108	0.01781
High Education ($s = 2$)			
$\gamma_{s=2,1}$	Weight on log skill	0.789	0.00132
$\gamma_{s=2,2}$	Weight on log investment	0.053	0.00195
$\gamma_{s=2,3}$	Weight on interaction between log skill and log investment	0.014	0.00039
$\sigma^2_{s=2,\eta}$	Variance of idiosyncratic shock	0.118	0.02132

The coefficient on the interaction term between log skills and log investment $(\gamma_{s,3})$ is positive for both high education mothers ($s = 2$) and low education mothers ($s = 1$), indicating that skills and investments are complements in the production function and that investments are more productive in children with higher skills. Given the log utility and the translog skill production function, this means that higher beliefs lead to higher investments, because of the following. Parents with higher beliefs perceive a higher return to investment, as investments are more productive in children with higher skills. However, complementarity between skills and investments in the production function is not very high, which is similar to other studies (see Cunha et al. (2010), Carneiro, Cruz-Aguayo, Pachon, and Schady (2022), Goff, Malamud, Pop-Eleches, and Urquiola (2023), Agostinelli and Wiswall (2025)).

Belief Updating: Table 3.8 presents the estimates of the parameters related to belief updating. The weight on the prior belief is around 0.66-0.67, which indicates that parents place a higher weight on their prior belief than the signal (noisy) measure of the child's skill. Furthermore, belief updating does not depend on mother's education: the weight on the prior belief is very similar for both high education and low education mothers.

Income Process: Table 3.9 presents the estimates of the family income process, which

Table 3.8: Belief Parameters

Parameter	Description	Estimate	S.E.
σ_{ϵ}^2	Variance of error in signal	2.381	0.43087
$\psi_{s=1}$	Weight on prior (Low education)	0.678	0.03163
$\psi_{s=2}$	Weight on prior (High education)	0.685	0.03756

depends on mother's education. Families with high education mothers ($s = 2$) are more likely to earn higher income.

Table 3.9: Family Income Process Parameters

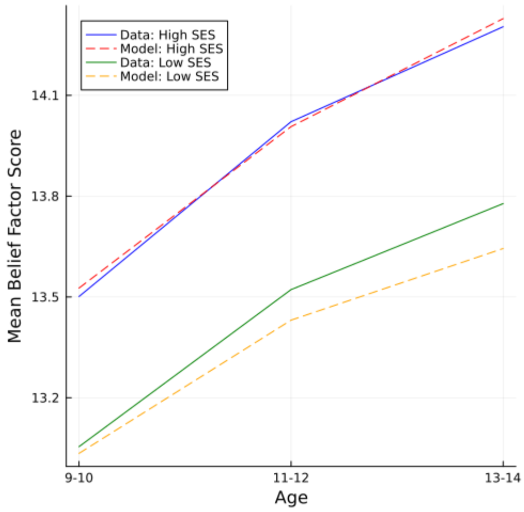
Parameter	Description	Estimate	S.E.
Low Education ($s = 1$)			
$\tau_{s=1,0}$	Intercept term	10.747	0.44918
$\tau_{s=1,1}$	Coefficient on age	0.088	0.15816
$\tau_{s=1,2}$	Coefficient on age squared	-0.009	0.01749
$\tau_{s=1,3}$	Coefficient on age cubed	3.605×10^{-4}	0.00061
$\rho_{s=1}$	AR(1) residual persistence parameter	0.633	0.03862
$\sigma_{s=1,u}^2$	AR(1) residual variance	0.230	0.03265
High Education ($s = 2$)			
$\tau_{s=2,0}$	Intercept term	11.725	0.44912
$\tau_{s=2,1}$	Coefficient on age	-0.129	0.15652
$\tau_{s=2,2}$	Coefficient on age squared	0.018	0.01719
$\tau_{s=2,3}$	Coefficient on age cubed	-6.781×10^{-4}	0.00060
$\rho_{s=2}$	AR(1) residual persistence parameter	0.525	0.01326
$\sigma_{s=2,u}^2$	AR(1) residual variance	0.372	0.03942

Notes: Robust standard errors reported for the log family income regression. Standard errors for the AR(1) process computed via bootstrap with 100 replications.

3.6.2 Model Fit

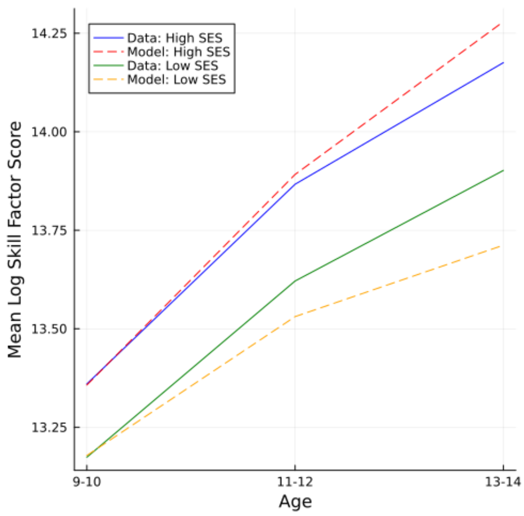
To illustrate the model fit, the trajectories of the belief factor scores and log skill factor scores are presented in Figure 3.7 and Figure 3.8 respectively. Age 9-10 values are included in the figures as a benchmark: by construction, the values in the model are very close to those in the data. The figures indicate that at ages 11-12 and 13-14, the model underpredicts the belief factor score for the low SES. Furthermore, the model overpredicts the skills of the high SES and underpredicts the skills of the low SES. For a more comprehensive understanding of the model fit, a complete list of moments used in the estimation is provided in Appendix 3.E.

Figure 3.7: Average Belief Factor Score by Age and Socio-economic Status



Notes: This figure presents the average belief factor score at age 9-10, 11-12 and 13-14. There are four lines which correspond to the (1) data value for the high SES, (2) model value for the high SES, (3) data value for the low SES and (4) model value for the low SES. Age 9-10 values are presented as a benchmark — by construction, the model values are very close to the data values. High socio-economic status (SES) parents refer to parents who have above median family income (defined based on average family income when the child is aged between 9-10). Otherwise, parents are low SES.

Figure 3.8: Average Log Skill Factor Score by Age and Socio-economic Status



Notes: This figure presents the average skill factor score at age 9-10, 11-12 and 13-14. There are four lines which correspond to the (1) data value for the high SES, (2) model value for the high SES, (3) data value for the low SES and (4) model value for the low SES. Age 9-10 values are presented as a benchmark — by construction, the model values are very close to the data values. High socio-economic status (SES) parents refer to parents who have above median family income (defined based on average family income when the child is aged between 9-10). Otherwise, parents are low SES.

3.7 Counterfactuals

3.7.1 Decomposition of Socio-economic Status Skill Gap

This section investigates the relative contribution of parental beliefs about child cognitive skill towards the socio-economic status (SES) skill gap. The SES skill gap is defined in Equation 3.7.1, where the terminal skill is the child's skill after the parent makes the final investment decision (in third and last period).

$$\text{SES skill gap} = \text{Mean(high SES terminal skill)} - \text{Mean(low SES terminal skill)} \quad (3.7.1)$$

In the model, various factors contribute towards the SES skill gap. These include differences in beliefs of parents, skill levels of children, resources of parents (income, assets) and unobserved preferences (relative preference for child skill over consumption). There are also differences in the belief updating process (different weights placed on the prior versus the signal), income trajectories and skill production functions.

To determine the contribution of each of these factors, several counterfactuals are conducted. Each of the counterfactuals equalises a specific factor across families: beliefs of parents, skill levels of children, resources of parents, preferences of parents. Table 3.10 describes the counterfactuals and the respective SES skill gap. In Table 3.10, an X in the column represents that that factor remains the same as the baseline. In the counterfactuals, equalising the belief updating process, equalising the income trajectory and equalising the skill production function means that all parent-child pairs have the high education belief updating process, income trajectory and skill production function. Equalising initial beliefs, initial income, initial assets and initial skills of children is achieved by assigning all parent-child pairs the median value of these variables (in the baseline) in the first period, respectively. Equalising preferences denotes assigning all parent-child pairs the “high" type preference: all parents are the unobserved discrete type which has the higher relative value for child skill over consumption.

The first counterfactual (C1) in Table 3.10 is the model baseline. Parent-child pairs are heterogeneous in terms of initial beliefs about their child's cognitive skill, income, assets and children's skills. These values are obtained directly from the data. Parent-child pairs also differ in terms of the belief updating process, income trajectory and the skill production function which they face. In subsequent counterfactuals, the sources of inequality are progressively removed.

When the inequality in beliefs is eliminated (parents have the same beliefs in the first period) in counterfactual C1.a, the SES skill gap hardly changes. In counterfactuals C1.b to C1.d, when inequality in initial skills, initial income or initial assets is removed, the SES skill gap decreases by a greater percentage. In the C2 class of counterfactuals, differences in belief updating, income trajectories and skill production function parameters are eliminated. All parent-child pairs face the belief updating process, income trajectory and skill production function of high education mothers. The patterns in the SES skill gap in the C2 counterfactuals are similar to the C1 counterfactuals. When inequality in beliefs is removed in counterfactual C2.a, the SES skill gap hardly decreases from C2. In counterfactuals C2.b to C2.d, when the inequality in child skills, income or assets is removed, the SES skill gap decreases by a greater percentage.

To obtain an idea of how heterogeneity in the dynamic processes of the model (belief updating process, income trajectory and skill production function) affect the SES skill gap, C2 counterfactuals are compared to the corresponding C1 counterfactuals. At first glance, it appears that the differences in the dynamic processes generally act to raise the SES skill gap, since the SES skill gap in counterfactual C2 (when differences in dynamic processes are eliminated) is lower than the model baseline C1. However, a closer examination reveals the relationship between the dynamic processes and the SES skill gap is more nuanced. For example, the SES skill gaps are noticeably higher in counterfactuals C2.c and C2.d than C1.c and C1.d respectively. This suggests that the way in which the differences in the dynamic processes influence the SES skill gap depends on the level of inequality in initial income and assets: when there is no inequality in income or assets, the differences in the dynamic processes act to lower the SES skill gap.

Table 3.10: SES Skill Gap in Counterfactual Scenarios which Eliminate Inequality

Scenario	Belief Updating	Income Trajectory	Skill Production Function	Initial Belief	Initial Skill	Initial Income	Initial Assets	Preferences	SES Skill Gap	Percentage of Baseline SES Skill Gap
C1	X	X	X	X	X	X	X	X	0.702	100.000
C1.a	X	X	X		X	X	X	X	0.694	98.768
C1.b	X	X	X	X		X	X	X	0.542	77.124
C1.c	X	X	X	X	X		X	X	0.544	77.425
C1.d	X	X	X	X	X	X		X	0.266	37.928
C1.e	X	X	X	X	X	X	X		0.700	99.624
C2				X	X	X	X	X	0.743	105.785
C2.a					X	X	X	X	0.737	105.029
C2.b				X		X	X	X	0.589	83.834
C2.c				X	X		X	X	0.646	91.954
C2.d				X	X	X		X	0.414	58.988
C2.e				X	X	X	X		0.738	105.117

Notes: This table presents the SES skill gap in the counterfactual scenarios. An "X' in the belief updating, income trajectory or skill production function columns means that the belief updating process, income trajectory or skill production function depends on mother's education, respectively. When the "X" is absent, all families have the high education belief updating process, income trajectory and skill production function. An "X" in the initial beliefs, initial skill, initial income and/or initial assets means that these are heterogeneous across families. When the "X" is absent, these values are set at the median value. An "X" in the preferences means that parents can be one of two preference types, with one type ("high" type) having a higher relative preference for child skill over consumption. The absence of an "X" means that all families have the "high" type relative preference for child skill over consumption. Low SES is defined as families with up to median income when the child is age 9-10, otherwise families are high SES.

Next, I examine whether there are interactions between parental beliefs and other sources of inequality: children's skills, parent resources and unobserved preferences. This is assessed by estimating counterfactuals where in the first period, inequality in beliefs is eliminated jointly with the inequality in another factor. The SES skill gaps in these counterfactuals are presented in Appendix Table 3.B.1. To assess whether there are interactions, the SES skill gaps in these counterfactuals are contrasted to the counterfactuals where (1) only inequality in beliefs is eliminated and (2) only inequality in the other factor is eliminated. For example, to determine whether beliefs interact with skills, the SES skill gaps in counterfactual C1.a (inequality in initial beliefs removed) and counterfactual C1.b (inequality in initial skills removed) are compared to the gap in counterfactual D1.b (inequality in beliefs and initial skills removed). If the sum of the decrease in the SES skill gap in counterfactuals C1.a and C1.b is equal to the decrease in the SES skill gap in counterfactual D1.b, then there are no interactions between beliefs and skills. Analogous comparisons can be made to assess whether there are interactions between beliefs and other factors. In general, estimates indicate that beliefs interact with initial skills, income, assets and preferences. However, the strength of the interaction is low.

On balance, the results in this section indicate that differences in parental beliefs do not contribute much to the SES skill gap.

3.7.2 Eliminating Belief Updating and Imperfect Information about Child Skill

To understand the potential impact of an information intervention which corrects parental beliefs, a counterfactual which is akin to "perfect information" is estimated. In this counterfactual, the beliefs of parents are equal to the skill levels of their child. In addition, parents do not update their beliefs. In practice, this is achieved by simulating parents' decisions when (1) posterior mean beliefs are equal to the log skills of the child, (2) posterior belief variances are set close to 0 (they are set equal to 0.05 in all periods) and (3) the weight on the prior belief is 1 in the belief updating (this means that the weight on the signal is 0). All other components of the model remain the same.

Table 3.11 presents the percentage change from the baseline (with imperfect information on child skills and belief updating) in the log skills, beliefs (which are now equal to the skills of the child by construction), investments, consumption and assets in the high

SES and low SES groups.

Table 3.11: Percentage Change from Baseline

	Low SES	High SES
Mean log skill age 11-12	0.009	-0.009
Mean log skill age 13-14	0.023	-0.005
Mean log skill terminal	0.038	-0.004
Mean posterior belief age 9-10	1.103	-1.251
Mean posterior belief age 11-12	0.753	-0.831
Mean posterior belief age 13-14	0.522	-0.346
Mean investment age 9-10	0.206	-0.777
Mean investment age 11-12	0.921	1.421
Mean investment age 13-14	0.205	-0.358
Mean consumption age 9-10	-0.533	0.166
Mean consumption age 11-12	-0.218	0.204
Mean consumption age 13-14	-0.284	-0.227
Mean next period assets choice age 9-10	0.094	0.360
Mean next period assets choice age 11-12	-0.842	-0.540
Mean next period assets choice age 13-14	-0.715	-0.462

Notes: This table presents the percentage change from the baseline in the mean log skills, mean posterior beliefs, mean investment, mean consumption and mean next period assets choice in the counterfactual which eliminates belief updating and imperfect information about the skill level of children. Low SES is defined as families with up to median income at age 9-10, otherwise families are high SES.

On average, high SES over-estimate the skills of their child and low SES under-estimate the skills of their child. Therefore, when parental beliefs are adjusted to match the skill levels of their children, low SES (high SES) will raise (lower) beliefs in the first period, which is age 9-10. Higher (lower) beliefs in turn lead to higher (lower) investments, because of higher (lower) perceived returns to investments, because skills and investments are complements in the skill production function. However, this may not happen in all periods because there is a counteracting force: there are decreasing marginal returns to child skills. The overall effect is that the average skills of low SES increase and the average skills of high SES decrease. Therefore, correcting the beliefs of parents lowers the SES skill gap. However, the gap decreases by less than 1%.

Next, we look more closely at the subgroups which are most affected by the correction of parental beliefs. These are the parent-child pairs which had the most inaccurate beliefs. Specifically, we focus on the following subgroups: (1) 50th percentile and above among those who underestimated their child's skills and (2) 50th percentile and above among those who overestimated their child's skills. Table 3.12 presents the percentage change from the baseline in the log skills, beliefs (which are now equal to the skills of the child by construction), investments, consumption and assets.

Table 3.12: Percentage Change from Baseline

	Under-estimate	Over-estimate
Mean log skill age 11-12	0.067	-0.057
Mean log skill age 13-14	0.181	-0.134
Mean log skill terminal	0.192	-0.124
Mean posterior belief age 9-10	11.392	-8.846
Mean posterior belief age 11-12	6.965	-5.898
Mean posterior belief age 13-14	4.608	-3.900
Mean investment age 9-10	1.926	-2.633
Mean investment age 11-12	7.849	-2.923
Mean investment age 13-14	-0.609	-0.228
Mean consumption age 9-10	-4.327	3.154
Mean consumption age 11-12	-2.142	2.126
Mean consumption age 13-14	-3.160	1.943
Mean next period assets choice age 9-10	0.328	0.548
Mean next period assets choice age 11-12	-4.426	2.044
Mean next period assets choice age 13-14	-3.048	1.256

Notes: This table presents the percentage change from the baseline in the mean log skills, mean posterior beliefs, mean investment, mean consumption and mean next period assets choice in the counterfactual which eliminates belief updating and imperfect information about the skill level of children. Underestimate refers to parent-child pairs which were at the 50th percentile and above among those who underestimated their child's skills. Overestimate refers to parent-child pairs who were at the 50th percentile and above among those who overestimated their child's skills.

The feedback between beliefs and investments is evident in the subgroups which are most affected when parental beliefs are corrected. In the first period, parents who underestimated (overestimated) the skills of their child invest more (less) because they perceive the marginal returns to investment to be higher (lower) than before. This is because investments are more productive for children with higher skills, since skills and investments are complements in the production function. Because they invest more (less), beliefs are higher (lower) in the second period, which leads them to invest more (less) again for the same reasons, and so on.

Most models assume that parents perfectly observe the skills of their child. This counterfactual suggests that making this assumption could lead to inaccurate predictions of parent investments among parents with inaccurate beliefs.

3.8 Conclusion

This paper investigates the role of parental beliefs in shaping the socio-economic status (SES) gap in children's skills, using a dynamic model of parent investments that incorporates parental beliefs about their child's skill level.

The results reveal that disparities in parental beliefs account for only a small fraction of the SES skill gap. Equalising beliefs across families reduces the gap by less than 2%, suggesting that beliefs have limited long-term impact relative to other factors such as parental resources and initial child skills. Similarly, a counterfactual scenario where parents hold accurate beliefs about their child's skills yields negligible reductions in the SES skill gap. These findings imply that belief-based information interventions are unlikely to effectively reduce SES skill disparities.

However, limitations of the model suggest caution in interpreting these results as definitive. First, the model abstracts from other potentially important dimensions of parental beliefs, such as beliefs about the skill production function (including how investments are converted to skills, which I shall refer to as the return to investment) or non-cognitive skills. It may be that parental beliefs about the child's cognitive skill are being counteracted by parental beliefs on one of these other dimensions. For example, suppose it is the case that parents invest less when they perceive lower returns to investment. And it is also the case that parents with higher beliefs about skills are more likely to perceive lower returns to investment. Then, since beliefs about returns are missing from the analysis, the impact of this dimension of beliefs may be erroneously attributed to beliefs about skills. Consequently, the estimated effect of parental beliefs about skills on investments will be smaller. Second, the model omits school and neighborhood inputs, factors that may be substitutes or complements to parent investment. This could affect the estimated relationship between beliefs and investments, since the model does not account for how parent investments react to changes in these alternative inputs.

Future research should explore these dynamics by extending the model to include multiple skill dimensions, a richer set of parental beliefs, and institutional factors like school inputs. It would also be valuable to examine whether similar patterns hold in different settings, particularly in developing countries or among younger age cohorts. Given that some types of skills are more malleable at younger ages (Kautz, Heckman, Diris, ter Weel, & Borghans, 2014), it may be that adjusting beliefs and investments in earlier periods could have a greater impact on child skills and the SES skill gap. When extending the dynamic framework to younger ages, it may be important to allow the skill production function parameters and the signal generation process to depend on age. Skill production may be more dynamic at earlier ages. Furthermore, parents may receive less information about their child before the child begins attending school.

Doing so could offer a more comprehensive understanding of how beliefs, resources, and contextual factors jointly shape the development of human capital across socio-economic groups.

Appendix to Chapter 3

3.A Data Appendix

3.A.1 Measures of Parent Investments

Components of the Home Observation Measurement of the Environment (HOME) used to construct the parent investment factor are as follows:

- How many magazines the family gets regularly
- Whether the child has the use of a record player, tape deck or CD player at home and at least 5 children's records or tapes
- The number of books the child has
- How often the mother reads to child
- Whether the family has a musical instrument which the child can use at home
- Whether the family receives a daily newspaper
- Whether the family encourages the child to pursue hobbies
- Whether the child attends special lessons
- How often a family member gets a chance to take the child on any kind of outing
- How often the child is brought to the museum
- How often the child is brought to music/theatre performance
- How often the family gets together
- How often the child eats with mom and dad

3.B Decomposition of SES Skill Gap: Additional Counterfactuals

Table 3.B.1: SES Skill Gap in Counterfactual Scenarios

Scenario	Belief Updating	Income Trajectory	Skill Production Function	Initial Belief	Initial Skill	Initial Income	Initial Assets	Preferences	SES Skill Gap	Percentage of Baseline SES Skill Gap
D1.b	X	X	X			X	X	X	0.533	75.918
D1.c	X	X	X		X		X	X	0.533	75.908
D1.d	X	X	X		X	X		X	0.258	36.752
D1.e	X	X	X		X	X	X		0.690	98.279
D2.b						X	X	X	0.583	83.003
D2.c					X		X	X	0.640	91.118
D2.d					X	X		X	0.407	58.008
D2.e					X	X	X		0.731	104.153

Notes: This table presents the SES skill gap in the counterfactual scenarios. An "X' in the belief updating, income trajectory or skill production function columns means that the belief updating process, income trajectory or skill production function depends on mother's education, respectively. When the "X" is absent, all families have the high education belief updating process, income trajectory and skill production function. An "X" in the initial beliefs, initial skill, initial income and/or initial assets means that these are heterogeneous across families. When the "X" is absent, these values are set at the median value. An "X" in the preferences means that parents can be one of two preference types, with one type ("high" type) having a higher relative preference for child skill over consumption. The absence of an "X" means that all families have the "high" type relative preference for child skill over consumption. Low SES is defined as families with up to median income when the child is age 9-10, otherwise families are high SES.

3.C Parameter Equality Test

Table 3.C.1: Test of Equality of Relative Preference for Child Skill Over Consumption Between Unobserved Discrete Types

Test	Description	Test Statistic	p-value
$\kappa_{h=1} = \kappa_{h=2}$	Relative preference for skill over consumption	6.362	0.00000

Table 3.C.2: Test of Equality of Skill Production Function Parameters Between High and Low Education

Test	Description	Test Statistic	p-value
$\gamma_{s=1,1} = \gamma_{s=2,1}$	Weight on log skill	8.170	0.00000
$\gamma_{s=1,2} = \gamma_{s=2,2}$	Weight on log investment	1.229	0.21899
$\gamma_{s=1,3} = \gamma_{s=2,3}$	Weight on interaction between log skill and log investment	25.441	0.00000
$\sigma^2_{s=1,\eta} = \sigma^2_{s=2,\eta}$	Variance of idiosyncratic shock	0.349	0.72697

3.D Further Details of Estimation

3.D.1 Sample Selection

The sample consists of children of the females in the cross-section sample of the National Longitudinal Survey of Youth 1979 (NLSY79). To be included in the estimation, the following variables must be non-missing in the first period (age 9-10): belief factor score, income, assets and skill factor score. Furthermore, because the model does not allow parents to borrow (assets must be weakly positive), observations with negative values of assets in any of the 3 periods are excluded. In addition, children in families which reported zero family income at age 9-10 are dropped. Moreover, only children whose mothers were married when the child was aged between 5-14 are included.

Table 3.D.1 displays the sample dropped due to the various sample selection criteria. The starting sample is all children of females in the cross-section of the NLSY79. High SES families are those with above median family income (based on average family income at age 9-10), otherwise families are low SES. Married age 5-14 means that the child's mother was married between ages 5-14. Initial conditions available refers to the starting conditions of the model being non-missing: belief factor score at age 9-10, family income at age 9-10, assets at age 9-10 and skill factor score at age 9-10. Assets non-negative means that the assets take non-negative values when the child is age

9-10, 11-12 and 13-14. Family income positive age 9-10 refers to family income at age 9-10 being strictly greater than 0.

From Table 3.D.1, a large proportion of the original sample is dropped after the selection. Furthermore, the attrition is higher for the low SES than the high SES.

Table 3.D.1: Sample Selection for Model Estimation

	All	High SES	Low SES
Original Sample	5,819	3,682	2,137
Married Age 5-14	1,988	1,552	436
Initial Conditions Available	731	510	221
Assets Non-Negative	657	485	172
Family Income Positive Age 9-10	655	485	170

Notes: This table presents the observations remaining after each sample selection criterion is applied. Married age 5-14 means that the child's mother was married when the child was age 5-14. Initial conditions refer to the belief factor score at age 9-10, family income at age 9-10, assets at age 9-10 and skill factor score at age 9-10. High socio-economic status (SES) parents refer to parents who have above median family income (defined based on average family income when the child is aged between 9 and 10). Otherwise, parents are low SES. This is the definition used in the data patterns section.

3.D.2 Method of Simulated Moments

This section provides additional details of the method of simulated moments estimation step.

Discretisation: There are 3 continuous state variables: mean of posterior belief distribution, assets and income. A 4-point grid is used for the mean of posterior belief and 7-point grids are used for the assets and family income. There are more grid points at lower levels of assets, where responses may be more non-linear. The consumption grid has 50 points, while the investment grid has 10 points. Since parents do not observe the skill of the child, they need to take expectations over the child's skill in the computation of the next period value function. To compute the expectations, a 5-point grid of log skill is used, which depends on the belief distribution which the parent has over the child skill.

Interpolation/Extrapolation: Linear interpolation is used to evaluate points within the grid (between grid points). Linear extrapolation is used to evaluate points outside of the grid. Interpolation/extrapolation is used to compute the expected value of the next period value function given the current state. In the simulation, optimal choices are obtained by interpolating/extrapolating policy functions, since the state variables may not fall on the grid.

Integration: 5-node Gauss-Hermite quadrature is used to integrate over the log skill and the shocks (shock in signal generation process, shock in skill production process).

Solving for Value Functions and Policy Functions by Backward Induction: The following procedure is used to compute the value function and policy functions. Starting from the final period, for every possible combination of state variables on the grid, I evaluate the objective function over a range of consumption and investment choices defined by the grids of these choice variables. The consumption and investment grids are endogenous. Consumption is in the outer loop, while investment is in the inner loop. The maximum value on the endogenous consumption grid is the value of resources available. The maximum value on the endogenous investment grid is the value of resources less the consumption choice. The combination of consumption and investment which maximises the objective function is chosen. Given consumption and investment choices, assets are determined from the resource constraint.

After performing the computation in the final period (period T), I proceed to the period before (period $T-1$) and perform a similar computation. This process is repeated all the way to the first period of the model.

Optimisation Algorithm: The minimisation algorithm used in the optimisation is Bound Optimisation by Quadratic Approximation (BOBYQA) by Powell (2009), which is implemented with Julia's NLopt package.

Programming Language: The program is written and implemented in Julia.

3.D.3 Auxiliary Model for Belief Parameters

To identify the weight in belief updating and the variance of the error term in the signal, an approximation of the belief updating equation (Equation 3.4.4) is used. In the approximation, as the prior mean belief is unobserved, it is substituted by the skill of the child. In the following, for ease of exposition, I omit the j subscript which indexes the child.

$$
\begin{aligned}
\mu_t^{int} &= \psi\mu_t + (1-\psi)g_t \\
&\approx \psi\left[A_{t-1} + \gamma_{s,1}\ln\theta_{t-1} + \gamma_{s,2}\ln i_{t-1} + \gamma_{s,3}\ln\theta_{t-1}\times\ln i_{t-1}\right] + (1-\psi)\left(\ln\theta_t + \epsilon_t\right) \\
&\approx \psi A_{t-1} + \psi\gamma_{s,3}\Phi_{t-1}\ln\theta_{t-1} + \psi\gamma_{s,2}\Phi_{t-1} + (1-\psi)\ln\theta_t \\
&+ (1-\psi)\,\epsilon_t - \psi\gamma_{s,3}\left(\Phi_{t-1}\ln\theta_t - \ln\theta_t\times\ln i_{t-1}\right) - \psi\gamma_{s,2}\left(\Phi_{t-1} - \ln i_{t-1}\right) \quad \text{(3.D.1)}
\end{aligned}
$$

Based on Equation 3.D.1, information on the belief updating weight ψ is provided by the coefficients in a regression of the belief factor score on a constant term, the interaction between lag investment factor score and lag skill factor score, the lag investment factor score and the skill factor score.
Furthermore, the residuals of this regression provide information on the variance of the error term in the signal equation. Though the error term in the regression is autocorrelated, this is not an issue, because if the model is correct, this autocorrelation will also feature in the corresponding data moments.

3.E Model Fit

Table 3.E.1 to Table 3.E.9 display the moments used in the estimation. Each table provides (1) the data moment, (2) the corresponding moment in the simulated data (model moment), (3) the standard error of the data moment and (4) the normalised (by the standard error) difference between the data moment and the model moment. A total of 112 moments were used to estimate the model. These include moments regarding the beliefs (Table 3.E.1 and Table 3.E.2), the investment factor scores (Table 3.E.3 and Table 3.E.4), the skills of the child (Table 3.E.5 and Table 3.E.6), the assets of the family (Table 3.E.7), correlations and covariances (Table 3.E.8) and the ratio of next period assets to investment factor score (Table 3.E.9).

Table 3.E.1: Belief Moments Part 1

Moment	Data	Model	SE Data	SE Diff
Mean belief age 11-12: Family income Q1	13.329	13.470	0.097	1.466
Mean belief age 11-12: Family income Q2	13.643	13.662	0.101	0.180
Mean belief age 11-12: Family income Q3	13.889	13.797	0.114	0.807
Mean belief age 11-12: Family income Q4	14.242	13.922	0.113	2.813
Mean belief age 13-14: Family income Q1	13.775	13.745	0.129	0.231
Mean belief age 13-14: Family income Q2	13.755	13.959	0.152	1.338
Mean belief age 13-14: Family income Q3	14.170	14.010	0.143	1.112
Mean belief age 13-14: Family income Q4	14.509	14.200	0.168	1.842
Correlation (low education): Belief age 9-10, Belief age 11-12	0.651	0.821	0.044	3.878
Correlation (high education): Belief age 9-10, Belief age 11-12	0.665	0.836	0.039	4.433
Correlation (low education): Belief age 11-12, Belief age 13-14	0.701	0.840	0.047	2.939
Correlation (high education): Belief age 11-12, Belief age 13-14	0.735	0.836	0.033	3.120
Correlation (low education): Belief age 11-12, Skill age 11-12	0.499	0.607	0.056	1.915
Correlation (high education): Belief age 11-12, Skill age 11-12	0.527	0.592	0.050	1.324
Correlation (low education): Belief age 13-14, Skill age 13-14	0.574	0.765	0.054	3.559
Correlation (high education): Belief age 13-14, Skill age 13-14	0.532	0.716	0.067	2.745

Notes: Belief refers to belief factor score. Data refers to the data moment, Model refers to the moment from the simulated data, SE Data refers to the standard error of the data moment and SE Data is the normalised (by standard error) difference between the data moment and the model moment.

Table 3.E.2: Belief Moments Part 2

Moment	Data	Model	SE Data	SE Diff
Regression of belief on lag investment, lag skill and current skill (low education): Constant	3.020	2.777	1.347	0.180
Regression of belief on lag investment, lag skill and current skill (low education): Coefficient on lag investment x lag skill	0.015	-0.165	0.127	1.421
Regression of belief on lag investment, lag skill and current skill (low education): Coefficient on lag investment	0.238	2.586	1.666	1.409
Regression of belief on lag investment, lag skill and current skill (low education): Coefficient on current skill	0.756	0.787	0.100	0.311
Std residual of regression of belief on lag investment, lag skill and current skill (low education)	1.023	0.827	0.043	4.558
Regression of belief on lag investment, lag skill and current skill (high education): Constant	2.579	3.619	1.529	0.680
Regression of belief on lag investment, lag skill and current skill (high education): Coefficient on lag investment x lag skill	0.134	0.019	0.098	1.171
Regression of belief on lag investment, lag skill and current skill (high education): Coefficient on lag investment	-1.277	-0.121	1.345	0.860
Regression of belief on lag investment, lag skill and current skill (high education): Coefficient on current skill	0.806	0.739	0.110	0.617
Std residual of regression of belief on lag investment, lag skill and current skill (high education)	0.994	0.779	0.039	5.495

Notes: Belief, skill and investment refer to belief factor score, skill factor score and investment factor score, respectively. Data refers to the data moment, Model refers to the moment from the simulated data, SE Data refers to the standard error of the data moment and SE Data is the normalised (by standard error) difference between the data moment and the model moment.

Table 3.E.3: Investment Factor Score Moments Part 1

Moment	Data	Model	SE Data	SE Diff
Std investment age 9-10	0.589	0.569	0.018	1.131
Std investment age 11-12	0.566	0.590	0.016	1.514
Std investment factor Age 13-14	0.636	0.360	0.021	13.289
Mean investment age 9-10: Family income Q1	0.257	0.073	0.051	3.628
Mean investment age 9-10: Family income Q2	0.453	0.222	0.046	5.041
Mean investment age 9-10: Family income Q3	0.519	0.429	0.042	2.160
Mean investment age 9-10: Family income Q4	0.673	0.899	0.043	5.275
Mean investment age 11-12: Family income Q1	0.273	0.069	0.046	4.420
Mean investment age 11-12: Family income Q2	0.487	0.194	0.051	5.780
Mean investment age 11-12: Family income Q3	0.491	0.399	0.045	2.029
Mean investment age 11-12: Family income Q4	0.683	0.687	0.042	0.107
Mean investment age 13-14: Family income Q1	0.225	-0.012	0.063	3.765
Mean investment age 13-14: Family income Q2	0.306	0.038	0.059	4.532
Mean investment age 13-14: Family income Q3	0.420	0.117	0.054	5.632
Mean investment age 13-14: Family income Q4	0.626	0.362	0.056	4.708

Notes: Investment refers to investment factor score. Data refers to the data moment, Model refers to the moment from the simulated data, SE Data refers to the standard error of the data moment and SE Data is the normalised (by standard error) difference between the data moment and the model moment.

Table 3.E.4: Investment Factor Score Moments Part 2

Moment	Data	Model	SE Data	SE Diff
Std residual of investment conditional on belief age 9-10	0.545	0.561	0.016	0.960
Std residual of investment conditional on belief age 11-12	0.537	0.569	0.018	1.808
Std residual of investment conditional on belief age 13-14	0.577	0.966	0.022	17.473
Correlation (Skill Q1): investment age 9-10, investment age 11-12	0.616	0.878	0.051	5.147
Correlation (Skill Q2): investment age 9-10, investment age 11-12	0.560	0.890	0.058	5.662
Correlation (Skill Q3): investment age 9-10, investment age 11-12	0.609	0.873	0.060	4.390
Correlation (Skill Q4): investment age 9-10, investment age 11-12	0.602	0.857	0.069	3.710
Correlation (Skill Q1): investment age 11-12, investment age 13-14	0.657	0.817	0.055	2.923
Correlation (Skill Q2): investment age 11-12, investment age 13-14	0.596	0.855	0.061	4.266
Correlation (Skill Q3): investment age 11-12, investment age 13-14	0.661	0.894	0.064	3.646
Correlation (Skill Q4): investment age 11-12, investment age 13-14	0.609	0.840	0.065	3.556

Notes: Investment and skill refer to investment factor score and skill factor score respectively. Data refers to the data moment, Model refers to the moment from the simulated data, SE Data refers to the standard error of the data moment and SE Data is the normalised (by standard error) difference between the data moment and the model moment.

Table 3.E.5: Skill Moments Part 1

Moment	Data	Model	SE Data	SE Diff
Std skill age 11-12	0.734	0.846	0.025	4.387
Std skill age 13-14	0.787	1.013	0.034	6.571
Mean skill age 11-12: Family income Q1	13.454	13.516	0.070	0.889
Mean skill age 11-12: Family income Q2	13.684	13.702	0.059	0.317
Mean skill age 11-12: Family income Q3	13.755	13.756	0.066	0.020
Mean skill age 11-12: Family income Q4	14.050	13.855	0.066	2.938
Mean skill age 13-14: Family income Q1	13.801	13.770	0.077	0.405
Mean skill age 13-14: Family income Q2	13.959	13.963	0.075	0.056
Mean skill age 13-14: Family income Q3	14.107	14.033	0.074	1.002
Mean skill age 13-14: Family income Q4	14.324	14.194	0.076	1.710
Correlation (low education): Skill age 9-10, Skill age 11-12	0.860	0.886	0.021	1.282
Correlation (high education): Skill age 9-10, Skill age 11-12	0.790	0.883	0.036	2.548
Correlation (low education): Skill age 11-12, Skill age 13-14	0.882	0.901	0.020	0.958
Correlation (high education): Skill age 11-12, Skill age 13-14	0.786	0.906	0.037	3.198
Correlation (low education): Skill age 9-10, Skill age 13-14	0.817	0.739	0.029	2.702
Correlation (high education): Skill age 9-10, Skill age 13-14	0.742	0.777	0.031	1.161

Notes: Skill refers to skill factor score. Data refers to the data moment, Model refers to the moment from the simulated data, SE Data refers to the standard error of the data moment and SE Data is the normalised (by standard error) difference between the data moment and the model moment.

Table 3.E.6: Skill Moments Part 2

Moment	Data	Model	SE Data	SE Diff
Regression skill production function (low education): Constant 1	1.798	-0.388	0.409	5.345
Regression skill production function (low education): Constant 2	1.663	-0.529	0.419	5.230
Regression skill production function (low education): Coefficient on skill	0.894	1.057	0.031	5.215
Regression skill production function (low education): Coefficient on investment	0.137	2.889	0.435	6.324
Regression skill production function (low education): Coefficient on skill x investment	-0.003	-0.166	0.033	4.871
Std residual skill production function (low education)	0.387	0.351	0.023	1.581
Regression skill production function (high education): Constant 1	3.481	0.459	1.012	2.984
Regression skill production function (high education): Constant 2	3.420	0.345	1.029	2.989
Regression skill production function (high education): Coefficient on skill	0.775	0.983	0.075	2.761
Regression skill production function (high education): Coefficient on investment	0.300	0.886	0.875	0.670
Regression skill production function (high education): Coefficient on skill x investment	-0.017	-0.044	0.065	0.419
Std residual skill production function (high education)	0.419	0.347	0.031	2.318

Notes: Investment and skill refer to investment factor score and skill factor score respectively. Data refers to the data moment, Model refers to the moment from the simulated data, SE Data refers to the standard error of the data moment and SE Data is the normalised (by standard error) difference between the data moment and the model moment.

Table 3.E.7: Asset Moments

Moment	Data	Model	SE Data	SE Diff
Mean asset age 11-12	0.48	0.44	0.05	0.884
Std asset age 11-12	0.85	0.84	0.11	0.147
Mean asset age 13-14	0.79	0.28	0.06	7.785
Std asset age 13-14	1.23	0.54	0.13	5.088

Notes: Data refers to the data moment, Model refers to the moment from the simulated data, SE Data refers to the standard error of the data moment and SE Data is the normalised (by standard error) difference between the data moment and the model moment.

Table 3.E.8: Correlations and Covariances

Moment	Data	Model	SE Data	SE Diff
Correlation (low education): Belief age 9-10, Skill age 11-12	0.523	0.419	0.049	2.098
Correlation (high education): Belief age 9-10, Skill age 11-12	0.491	0.419	0.044	1.646
Correlation (low education): Belief age 11-12, Skill age 13-14	0.486	0.570	0.069	1.216
Correlation (high education): Belief age 11-12, Skill age 13-14	0.487	0.570	0.066	1.256
Covariance: Investment age 9-10, Belief age 9-10	0.254	0.110	0.031	4.710
Covariance: Investment age 11-12, Belief age 11-12	0.223	0.169	0.030	1.785
Covariance: Investment age 13-14, Belief age 13-14	0.339	0.134	0.041	4.967
Covariance: Investment age 9-10, Skill age 9-10	0.113	0.082	0.020	1.546
Covariance: Investment age 11-12, Skill age 11-12	0.104	0.157	0.017	3.216
Covariance: Investment age 13-14, Skill age 13-14	0.124	0.134	0.025	0.413
Covariance: Investment age 9-10, Belief age 11-12	0.278	0.183	0.033	2.895
Covariance: Investment age 11-12, Belief age 13-14	0.297	0.273	0.038	0.628

Notes: Belief, skill and investment refer to belief factor score, skill factor score and investment factor score, respectively. Data refers to the data moment, Model refers to the moment from the simulated data, SE Data refers to the standard error of the data moment and SE Data is the normalised (by standard error) difference between the data moment and the model moment.

Table 3.E.9: Ratio of Next Period Assets to Investment Factor Score

Moment	Data	Model	SE Data	SE Diff
Ratio of Assets Age 11-12 to Investment age 9-10: Skill Q1	-1.04	0.13	1.10	1.072
Ratio of Assets Age 11-12 to Investment age 9-10: Skill Q2	-2.02	0.43	2.01	1.223
Ratio of Assets Age 11-12 to Investment age 9-10: Skill Q3	-42.67	1.36	43.28	1.017
Ratio of Assets Age 11-12 to Investment age 9-10: Skill Q4	0.20	0.39	0.10	1.840
Ratio of Assets Age 13-14 to Investment age 11-12: Skill Q1	-0.29	-0.23	0.58	0.094
Ratio of Assets Age 13-14 to Investment age 11-12: Skill Q2	1.12	0.15	1.31	0.742
Ratio of Assets Age 13-14 to Investment age 11-12: Skill Q3	-0.18	0.01	0.47	0.408
Ratio of Assets Age 13-14 to Investment age 11-12: Skill Q4	0.48	0.48	0.14	0.047

Notes: Investment refers to investment factor score. Data refers to the data moment, Model refers to the moment from the simulated data, SE Data refers to the standard error of the data moment and SE Data is the normalised (by standard error) difference between the data moment and the model moment.

References

Ackerberg, D. A. (2003). Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination. *International Economic Review*, *44*(3), 1007–1040.

Agostinelli, F., & Wiswall, M. (2025, March). Estimating the Technology of Children's Skill Formation. *Journal of Political Economy*, *133*(3), 846–887. doi: 10.1086/733396

Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011, January). Personality Psychology and Economics. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 4, pp. 1–181). Elsevier. doi: 10.1016/B978-0-444-53444-6.00001-8

Altonji, J. G., & Dunn, T. A. (1991, June). *Relationships Among the Family Incomes and Labor Market Outcomes of Relatives* (Working Paper No. 3724). National Bureau of Economic Research. doi: 10.3386/w3724

Altonji, J. G., & Segal, L. M. (1996). Small-Sample Bias in GMM Estimation of Covariance Structures. *Journal of Business & Economic Statistics*, *14*(3), 353–366. doi: 10.2307/1392447

Anderson, T. W., & Rubin, H. (1956, January). Statistical Inference in Factor Analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry* (Vol. 3.5, pp. 111–151). University of California Press.

Arcidiacono, P., Aucejo, E., Maurel, A., & Ransom, T. (2025, January). College Attrition and the Dynamics of Information Revelation. *Journal of Political Economy*, *133*(1), 53–110. doi: 10.1086/732526

Arellano, M., & Bond, S. (1991, April). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, *58*(2), 277–297. doi: 10.2307/2297968

Attanasio, O., Boneva, T., & Rauh, C. (2019, January). *Parental Beliefs about Returns* *to Different Types of Investments in School Children* (Working Paper No. 25513). National Bureau of Economic Research. doi: 10.3386/w25513

Attanasio, O., Boneva, T., & Rauh, C. (2022, November). Parental Beliefs about Returns to Different Types of Investments in School Children. *Journal of Human Resources*, *57*(6), 1789–1825. doi: 10.3368/jhr.58.2.0719-10299R1

Attanasio, O., Cattan, S., & Meghir, C. (2022, August). Early Childhood Development, Human Capital, and Poverty. *Annual Review of Economics*, *14*, 853–892.

Attanasio, O., Cunha, F., & Jervis, P. (2019, December). *Subjective Parental Beliefs. Their Measurement and Role* (Tech. Rep. No. w26516). National Bureau of Economic Research. doi: 10.3386/w26516

Attanasio, O., Cunha, F., Jervis, P., & Toppeta, A. (2025, March). *Parental Beliefs and Learning.*

Attanasio, O., & Kaufmann, K. M. (2014, July). Education Choices and Returns to Schooling: Mothers' and Youths' Subjective Expectations and Their Role by Gender. *Journal of Development Economics*, *109*, 203–216. doi: 10.1016/j.jdeveco.2014.04.003

Attanasio, O., Meghir, C., & Nix, E. (2020, November). Human Capital Development and Parental Investment in India. *The Review of Economic Studies*, *87*(6), 2511–2541. doi: 10.1093/restud/rdaa026

Barrera-Osorio, F., Gonzalez, K., Lagos, F., & Deming, D. J. (2020, June). Providing performance information in education: An experimental evaluation in Colombia. *Journal of Public Economics*, *186*, 104185. doi: 10.1016/j.jpubeco.2020.104185

Barrow, L., & Rouse, C. E. (2005, May). Do Returns to Schooling Differ by Race and Ethnicity? *American Economic Review*, *95*(2), 83–87. doi: 10.1257/000282805774670130

Bayer, P., & Charles, K. K. (2018, August). Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men Since 1940. *The Quarterly Journal of Economics*, *133*(3), 1459–1501. doi: 10.1093/qje/qjy003

Becker, G. S., & Tomes, N. (1979). An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy*, *87*(6), 1153–1189.

Becker, G. S., & Tomes, N. (1986). Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*, *4*(3), S1-S39.

Belley, P., & Lochner, L. (2007, October). *The Changing Role of Family Income and Ability in Determining Educational Achievement* (Working Paper No. 13527). National Bureau of Economic Research. doi: 10.3386/w13527

Bergman, P. (2021, January). Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment. *Journal of Political Economy*, *129*(1), 286–322. doi: 10.1086/711410

Bhattacharya, D., & Mazumder, B. (2011). A Nonparametric Analysis of Black–White Differences in Intergenerational Income Mobility in the United States. *Quantitative Economics*, *2*(3), 335–379. doi: 10.3982/QE69

Black, S. E., & Devereux, P. J. (2011, January). Recent Developments in Intergenerational Mobility. *Handbook of Labor Economics*, *4*, 1487–1541. doi: 10.1016/S0169-7218(11)02414-2

Blanden, J., Gregg, P., & Macmillan, L. (2007, March). Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education. *The Economic Journal*, *117*(519), C43-C60. doi: 10.1111/j.1468-0297.2007.02034.x

Blanden, J., Haveman, R., Smeeding, T., & Wilson, K. (2014). Intergenerational Mobility in the United States and Great Britain: A Comparative Study of Parent–Child Pathways. *Review of Income and Wealth*, *60*(3), 425–449. doi: 10.1111/roiw.12032

Blundell, R., & Bond, S. (1998, November). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, *87*(1), 115–143. doi: 10.1016/S0304-4076(98)00009-8

Bolt, U., French, E., Maccuish, J. H., & O'Dea, C. (2024, May). The Intergenerational Elasticity of Earnings: Exploring the Mechanisms. *Institute for Fiscal Studies Working Paper*.

Boneva, T., & Rauh, C. (2018, December). Parental Beliefs about Returns to Educational Investments—The Later the Better? *Journal of the European Economic Association*, *16*(6), 1669–1711. doi: 10.1093/jeea/jvy006

Borghans, L., Duckworth, A. L., Heckman, J. J., & Weel, B. T. (2008). The Economics and Psychology of Personality Traits. *Journal of Human Resources*, *43*(4), 972–1059. doi: 10.3368/jhr.43.4.972

Bratberg, E., Davis, J., Mazumder, B., Nybom, M., Schnitzlein, D. D., & Vaage, K. (2017). A Comparison of Intergenerational Mobility Curves in Germany, Norway, Sweden, and the US. *The Scandinavian Journal of Economics*, *119*(1), 72–101. doi: 10.1111/sjoe.12197

Carneiro, P., Cruz-Aguayo, Y., Pachon, R. H., & Schady, N. (2022, March). *Dynamic*

Complementarity in Elementary Schools: Experimental Estimates from Ecuador.

Carneiro, P., Heckman, J. J., & Masterov, D. V. (2005, April). Labor Market Discrimination and Racial Differences in Premarket Factors. *The Journal of Law and Economics*, *48*(1), 1–39. doi: 10.1086/426878

Carneiro, P., Reis, H., & Toppeta, A. (2024, February). *Parental Investments and Socio-Economic Gradients in Learning across European Countries* (IZA Discussion Paper No. 16785).

Caucutt, E. M., & Lochner, L. (2020, March). Early and Late Human Capital Investments, Borrowing Constraints, and the Family. *Journal of Political Economy*, *128*(3), 1065–1147. doi: 10.1086/704759

Caucutt, E. M., Lochner, L., & Park, Y. (2017). Correlation, Consumption, Confusion, or Constraints: Why Do Poor Children Perform so Poorly? *The Scandinavian Journal of Economics*, *119*(1), 102–147. doi: 10.1111/sjoe.12195

Chavez, N. (2021, December). *'It's everywhere.' Fed up with racist bullying, students across the US are walking out and protesting hate.* https://www.cnn.com/2021/12/05/us/racist-bullying-school-incidents.

Chetty, R., & Hendren, N. (2018, August). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. *The Quarterly Journal of Economics*, *133*(3), 1107–1162. doi: 10.1093/qje/qjy007

Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2020, May). Race and Economic Opportunity in the United States: An Intergenerational Perspective. *The Quarterly Journal of Economics*, *135*(2), 711–783. doi: 10.1093/qje/qjz042

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014, November). Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States*. *The Quarterly Journal of Economics*, *129*(4), 1553–1623. doi: 10.1093/qje/qju022

Chetty, R., University, H., Insights, O., Dobbie, W., University, H., Insights, O., ... Insights, O. (2024, September). *Changing Opportunity: Sociological Mechanisms Underlying Growing Class Gaps and Shrinking Race Gaps in Economic Mobility.*

Chikhale, N. (2024, January). *Parent Beliefs, Social Learning, and Intergenerational Mobility.*

Cholli, N. A., & Durlauf, S. N. (2022, February). *Intergenerational Mobility* (Working Paper No. 29760). National Bureau of Economic Research. doi: 10.3386/w29760

Conti, G., Giannola, M., & Toppeta, A. (2022, November). Parental Beliefs, Perceived

Health Risks, and Time Investment in Children: Evidence from COVID-19. *Institute for Fiscal Studies Working Paper*.

Crawford, G. S., & Shum, M. (2005). Uncertainty and Learning in Pharmaceutical Demand. *Econometrica*, *73*(4), 1137–1173. doi: 10.1111/j.1468-0262.2005.00612.x

Cunha, F. (2014). *Gaps in Early Investments in Children.*

Cunha, F., Elo, I., & Culhane, J. (2022, November). Maternal Subjective Expectations about the Technology of Skill Formation Predict Investments in Children One Year Later. *Journal of Econometrics*, *231*(1), 3–32. doi: 10.1016/j.jeconom.2020.07.044

Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, *78*(3), 883–931. doi: 10.3982/ECTA6551

Davis, J., & Mazumder, B. (2018, March). *Racial and Ethnic Differences in the Geography of Intergenerational Mobility* (SSRN Scholarly Paper No. ID 3138979). Rochester, NY: Social Science Research Network. doi: 10.2139/ssrn.3138979

Davis, J., & Mazumder, B. (2024, February). The Decline in Intergenerational Mobility after 1980. *The Review of Economics and Statistics*, 1–47. doi: 10.1162/rest_a_01413

Del Boca, D., Flinn, C., & Wiswall, M. (2014, January). Household Choices and Child Development. *The Review of Economic Studies*, *81*(1), 137–185. doi: 10.1093/restud/rdt026

Del Boca, D., Flinn, C. J., Verriest, E., & Wiswall, M. J. (2019, February). *Actors in the Child Development Process* (Working Paper No. 25596). National Bureau of Economic Research. doi: 10.3386/w25596

Dizon-Ross, R. (2019, August). Parents' Beliefs about Their Children's Academic Ability: Implications for Educational Investments. *American Economic Review*, *109*(8), 2728–2765. doi: 10.1257/aer.20171172

Duflo, E. (2012, May). *Human Values and the Design of the Fight Against Poverty* [Tanner Lectures].

Elsner, B., Isphording, I. E., & Zölitz, U. (2021, November). Achievement Rank Affects Performance and Major Choices in College. *The Economic Journal*, *131*(640), 3182–3206. doi: 10.1093/ej/ueab034

Erdem, T., & Keane, M. P. (1996). Decision-Making under Uncertainty: Capturing

Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. *Marketing Science*, *15*(1), 1–20.

Francesconi, M., & Heckman, J. J. (2016, December). Child Development and Parental Investment: Introduction. *The Economic Journal*, *126*(596), F1-F27. doi: 10.1111/ecoj.12388

French, E., & Jones, J. B. (2011). The Effects of Health Insurance and Self-Insurance on Retirement Behavior. *Econometrica*, *79*(3), 693–732. doi: 10.3982/ECTA7560

Gelbach, J. B. (2016, April). When Do Covariates Matter? And Which Ones, and How Much? *Journal of Labor Economics*, *34*(2), 509–543. doi: 10.1086/683668

Glewwe, P., & Miguel, E. A. (2007, January). Chapter 56 The Impact of Child Health and Nutrition on Education in Less Developed Countries. In T. P. Schultz & J. A. Strauss (Eds.), *Handbook of Development Economics* (Vol. 4, pp. 3561–3606). Elsevier. doi: 10.1016/S1573-4471(07)04056-9

Goff, L., Malamud, O., Pop-Eleches, C., & Urquiola, M. (2023, April). Interactions Between Family and School Environments: Access to Abortion and Selective Schools. *Journal of Human Resources*. doi: 10.3368/jhr.0419-10143R2

Gourieroux, C., Monfort, A., & Renault, E. (1993). Indirect Inference. *Journal of Applied Econometrics*, *8*, S85-S118.

Greaves, E., Hussain, I., Rabe, B., & Rasul, I. (2023, August). Parental Responses to Information about School Quality: Evidence from Linked Survey and Administrative Data. *The Economic Journal*, *133*(654), 2334–2402. doi: 10.1093/ej/uead015

Groves, M. O. (2005). Personality and the Intergenerational Transmission of Economic Status. In M. O. Groves, S. Bowles, & H. Gintis (Eds.), *Unequal Chances* (pp. 208–231). Princeton University Press.

Haider, S., & Solon, G. (2006, September). Life-Cycle Variation in the Association between Current and Lifetime Earnings. *American Economic Review*, *96*(4), 1308–1320. doi: 10.1257/aer.96.4.1308

Heckman, J. J., Lochner, L. J., & Todd, P. E. (2006, January). Chapter 7 Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. In E. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education* (Vol. 1, pp. 307–458). Elsevier. doi: 10.1016/S1574-0692(06)01007-5

Heckman, J. J., & Singer, B. (1984). A Method for Minimizing the Impact of Distributional

Assumptions in Econometric Models for Duration Data. *Econometrica*, *52*(2), 271–320. doi: 10.2307/1911491

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, *24*(3), 411–482. doi: 10.1086/504455

Hertz, T. (2005). Rags, Riches and Race: The Intergenerational Economic Mobility of Black and White Families in the United States. In S. Bowles, H. Gintis, & M. O. Groves (Eds.), *Unequal Chances* (pp. 165–191). Princeton University Press.

Hertz, T. (2006, April). *Understanding Mobility in America* (Tech. Rep.). Center for American Progress.

Jácome, E., Kuziemko, I., & Naidu, S. (2025, January). Mobility for All: Representative Intergenerational Mobility Estimates over the Twentieth Century. *Journal of Political Economy*, *133*(1), 306–354. doi: 10.1086/732527

Johnson, W., & Neal, D. (1998, February). *Basic Skills and the Black-White Earnings Gap.*

Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., & Borghans, L. (2014, November). Fostering and Measuring Skills.
doi: 10.1787/5jxsr7vr78f7-en

Keane, M. P., & Wolpin, K. I. (1997). The Career Decisions of Young Men. *Journal of Political Economy*, *105*(3), 473–522. doi: 10.1086/262080

Kinsler, J., & Pavan, R. (2016, November). *Parental Beliefs and Investment in Children: The Distortionary Impact of Schools* (Working Paper No. 2016-029). Human Capital and Economic Opportunity Global Working Group.

Kinsler, J., & Pavan, R. (2021, January). Local Distortions in Parental Beliefs over Child Skill. *Journal of Political Economy*, *129*(1), 81–100. doi: 10.1086/711347

Koşar, G., & O'Dea, C. (2023, January). Expectations data in structural microeconomic models. In R. Bachmann, G. Topa, & W. van der Klaauw (Eds.), *Handbook of Economic Expectations* (pp. 647–675). Academic Press. doi: 10.1016/B978-0-12-822927-9.00029-X

Kourtellos, A., Marr, C., & Tan, C. M. (2020, July). Local Intergenerational Mobility. *European Economic Review*, *126*, 103460. doi: 10.1016/j.euroecorev.2020.103460

La Ferrara, E. (2019, December). Presidential Address: Aspirations, Social Norms, and

Development. *Journal of the European Economic Association*, *17*(6), 1687–1722. doi: 10.1093/jeea/jvz057

Landale, N. S., Oropesa, R. S., & Bradatan, C. (2006). Hispanic Families in the United States: Family Structure and Process in an Era of Family Change. In *Hispanics and the Future of America.* National Academies Press (US).

Lang, K., & Manove, M. (2011, June). Education and Labor Market Discrimination. *American Economic Review*, *101*(4), 1467–1496. doi: 10.1257/aer.101.4.1467

Lekfuangfu, W. N., & Odermatt, R. (2022, September). All I Have to Do Is Dream? The Role of Aspirations in Intergenerational Mobility and Well-Being. *European Economic Review*, *148*. doi: 10.1016/j.euroecorev.2022.104193

List, J. A., Pernaudet, J., & Suskind, D. (2021, October). *It All Starts with Beliefs: Addressing the Roots of Educational Inequities by Shifting Parental Beliefs* (Working Paper No. 29394). National Bureau of Economic Research. doi: 10.3386/w29394

Lundborg, P., Rooth, D.-O., & Alex-Petersen, J. (2022, March). Long-Term Effects of Childhood Nutrition: Evidence from a School Lunch Reform. *The Review of Economic Studies*, *89*(2), 876–908. doi: 10.1093/restud/rdab028

Mazumder, B. (2014). Black–White Differences in Intergenerational Economic Mobility in the U.S. - Federal Reserve Bank of Chicago. *Economic Perspectives*, *38*.

Mazumder, B. (2016, February). Estimating the Intergenerational Elasticity and Rank Association in the United States: Overcoming the Current Limitations of Tax Data. In *Inequality: Causes and Consequences* (Vol. 43, pp. 83–129). Emerald Group Publishing Limited. doi: 10.1108/S0147-912120160000043012

Mazumder, B. (2018, November). Intergenerational Mobility in the United States: What We Have Learned from the PSID. *The ANNALS of the American Academy of Political and Social Science*, *680*(1), 213–234. doi: 10.1177/0002716218794129

Minkin, R., & Horowitz, J. (2023, January). *Parenting in America Today* (Tech. Rep.). Pew Research Center.

Neal, D. (2006, January). Chapter 9 Why Has Black–White Skill Convergence Stopped? In E. Hanushek & F. Welch (Eds.), *Handbook of the Economics of Education* (Vol. 1, pp. 511–576). Elsevier. doi: 10.1016/S1574-0692(06)01009-9

Neal, D., & Johnson, W. (1996, October). The Role of Premarket Factors in Black-White Wage Differences. *Journal of Political Economy*, *104*(5), 869–895. doi: 10.1086/262045

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, *49*(6), 1417–1426. doi: 10.2307/1911408

Nicoletti, C., Sevilla, A., & Tonei, V. (2022). Gender Stereotypes in the Family. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4294398

Nybom, M., & Stuhler, J. (2016, December). Biases in Standard Measures of Intergenerational Income Dependence. *Journal of Human Resources*. doi: 10.3368/jhr.52.3.0715-7290R

Page, L., & Ruebeck, H. (2022, September). Childhood Confidence, Schooling, and the Labor Market: Evidence from the PSID. *Journal of Human Resources*. doi: 10.3368/jhr.0621-11743R3

Petre, M. (2019, September). Contributions of Skills to the Racial Wage Gap. *Journal of Human Capital*, *13*(3), 479–518. doi: 10.1086/704322

Powell, M. J. D. (2009). The BOBYQA Algorithm for Bound Constrained Optimization without Derivatives. *Report DAMTP 2009/NA06, University of Cambridge*.

Sanders, C. (2012, October). *Skill Uncertainty, Skill Accumulation, and Occupational Choice.*

Solon, G. (1999, January). Chapter 29 - Intergenerational Mobility in the Labor Market. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3, pp. 1761–1800). Elsevier. doi: 10.1016/S1573-4463(99)03010-2

The State of Higher Education 2023 Report (Tech. Rep.). (2023). Lumina Foundation-Gallup.

Stinebrickner, R., & Stinebrickner, T. (2014, July). Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model. *Journal of Labor Economics*, *32*(3), 601–644. doi: 10.1086/675308

Thompson, O. (2021, March). *Human Capital and Black-White Earnings Gaps, 1966-2017* (Working Paper No. 28586). National Bureau of Economic Research. doi: 10.3386/w28586

Van der Klaauw, W., & Wolpin, K. I. (2008, July). Social Security and the Retirement and Savings Behavior of Low-Income Households. *Journal of Econometrics*, *145*(1), 21–42. doi: 10.1016/j.jeconom.2008.05.004

Zafar, B. (2011, April). How Do College Students Form Expectations? *Journal of Labor Economics*, *29*(2), 301–348. doi: 10.1086/658091

Statement of Conjoint Work

Note on the joint work in Rachel Yi Tan's thesis "Essays on Inequality and Human Capital."

The chapter, "Race and Intergenerational Mobility", is co-authored work with Pedro Carneiro.

The chapters "Parental Beliefs and Parent Investments" and "A Dynamic Analysis of Parental Beliefs and Investments" are single-authored by Rachel Yi Tan.