Essays on Human Capital Formation over the Life Cycle

Francesca Salvati

A dissertation submitted in fulfilment
of the requirements for the degree of

Doctor of Philosophy

University College London (UCL)

Department of Economics

April, 2021

Declaration

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

Signature: [redacted]

Date: April 9, 2021

Abstract

A large literature in economics has highlighted the importance of human capital as a cause and consequence of economic inequality. In this thesis, I study different determinants as well as dynamics of human capital development at various stages of the life cycle. The chapter titled *"The Effect of Classroom Rank on Learning throughout Elementary School: Experimental Evidence from Ecuador"* investigates the causal impact of classroom rank on the learning of children using experimental data from Ecuador. We find that children with higher classroom rank at the beginning of the academic year have significantly higher math test scores at the end of that grade. The impact of classroom math rank is larger for younger children, and grows substantially over time. Exogenous changes in classroom rank in math also improve cognitive flexibility, non-cognitive skills, and teacher perceptions of students. The chapter titled *"Human Capital Growth and Poverty: Evidence from Ethiopia and Peru"* estimates the production functions of health and cognition from childhood into adolescence, characterizing the nature of persistence and dynamic complementarities between these skills. It shows that differences in investments by parental income lead to large gaps in skills by age 8 that persist through age 15. Finally, the chapter titled *"Health Inequality, Labor Supply and Retirement Policies"* examines the dynamic relationship between health and employment of women at older ages in a rich structural model of consumption, savings, labor supply and health which allows for a two-way interaction between health and employment. The estimated model is used to study welfare implications of increases to the state pension age highlighting the differential effects of the policy due to differences in health, and, conversely, what role the policy plays in shaping health inequalities.

Impact statement

In this thesis, I study different determinants as well as dynamics of human capital development at various stages of the life cycle. I explore the development of different facets of individual human capital, spanning the domains of cognition, socio-emotional skills and health. I document how differences in these skills can lead to inequality in economic outcomes, starting from the early years of life into adulthood. The relevance of the role human capital plays for economic inequality is frequently highlighted in academic, as well as policy debates.

In chapter 2, we leverage unique experimental data from Ecuador to provide novel evidence on the causal effect of classroom rank on learning. Within each school, students were randomly assigned to classrooms in every grade between kindergarten and 6th grade. Therefore, two students with the same ability can have different classroom ranks because of the (random) peer composition of their classroom. We find that children with higher classroom rank at the beginning of the academic year have significantly higher math test scores at the end of that grade. The impact of classroom math rank is larger for younger children, and grows substantially over time. Exogenous changes in classroom rank in math also improve cognitive flexibility, non-cognitive skills, and teacher perceptions of students. This paper is the first to analyse rank effects in education using an experiment with multiple rounds of random assignment.

In chapter 3, we estimate parental investment equations and child production functions for human capital from age 1 to 15, allowing these to be dynamically connected. We use high quality data from Ethiopia and Peru to implement our strategy. We innovate on the existing literature in a number of dimensions, emphasizing the interaction between health and cognition and experimenting with flexible functional forms for the production functions. Our results offer insights into the production of child development across various low income countries which may help us shed light on the process of human capital formation in the early years, and design more effective interventions. This work has been published in a leading economics journal, the Review of Economic Dynamics.

In chapter 4, I examine the dynamic relationship between health and employment of women at older ages in a rich structural model of consumption, savings, labor supply and health which allows for a two-way interaction between health and employment. I use the model to study welfare implications of increases to the state pension age, a highly debated policy across many high income countries. My results offer novel insights into the dynamic relationship between health and labor market outcomes of women around retirement age, that inform the design of public policies that affect employment at this stage of the life cycle. I provide the first evidence of welfare implications of increasing the state pension age that highlight the specific role that health plays in generating different responses to the policy, as well as the role that the policy plays in reinforcing health inequalities.

Acknowledgements

I am profoundly grateful to many people who contributed significantly to my growth and development over the course of the PhD. Orazio Attanasio, Richard Blundell and Monica Costa Dias have been invaluable mentors and supervisors. Orazio has been an incredibly supportive and constructive advisor, and an extremely important influence on me. His seminal work and our interactions spurred my interest in the topics of human capital development, that became the core of my research interests, and this thesis. Orazio introduced me to the world of research and I am incredibly grateful for all the advice he has given me throughout the years. Richard has been an outstanding mentor. He has always been supportive, enthusiastic and friendly, adding to my research with many insightful comments, his kindness and his ability to motivate. Monica is a great role model for me, and I aspire to emulate her in both my professional and my personal life. Her knowledge, generosity and support have been invaluable, and I am very fortunate to have benefited from her guidance. Meeting Monica has been a turning point of my development as a researcher, and for this I will be forever grateful to her. I cannot express my gratitude to Orazio, Richard and Monica enough. It has been an honour to be mentored by them and they will continue to inspire me in my personal and professional life.

I am also very grateful to Pedro Carneiro. Pedro is an impressive researcher and an incredibly generous person. Working with him has been one of the best experiences of my PhD, and I am very fortunate to be able to learn from him.

I am also very grateful to the delightful coauthors I have had the pleasure to work with over the course of the PhD: Sarah Cattan, Gabriella Conti, Yyannu Cruz Aguayo, Costas Meghir, Emily Nix and Norbert Schady. I cannot express how lucky I feel to have benefited from their knowledge and insights.

This thesis benefited greatly from the interactions with many other people. Cormac O'Dea has provided insightful and helpful comments, and he has always been a great motivator. James Banks, Mariacristina De Nardi, Eric French, Fabien Postel-Vinay and Michela Tincani have provided great comments at various stages. A sincere thanks also goes to Giacomo Mason and Pete Spittal, my dearest colleagues in the

PhD. They are both fantastic researchers and generous friends, who were always available to listen to all my worries and doubts, always providing support and helpful advice. I would also like to thank Huw Pill, who has been a generous mentor and constant source of motivation.

This work would not have been possible without generous funding from the Economic and Social Research Council as well as the UCL Department of Economics. Moreover, I acknowledge helpful support from the Institute for Fiscal Studies.

Charlie, my partner, has been the most fantastic companion. He has been by my side with his unconditional love and support, always putting my needs above his own and turning every "bad day" into a positive and hopeful one through his kindness, joy and brilliant sense of humor.

I feel utmost gratitude towards my mother Maria, my father Stefano and my dearest sister Federica. I would like to dedicate this thesis and achievement to you all, who have been by my side from the very beginning and will always be. Thank you for your affection, your generosity and your constant encouragement.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The importance of human capital as a determinant of inequality in economic outcomes is widely acknowledged in the economics literature. Understanding the determinants and dynamics of human capital formation over the life cycle is important for the design of policies that aim to promote efficiency and reduce inequality. Scholars widely agree that human capital is a multi-faceted construct, that encompasses different skills such as cognition, health and socio-emotional skills. This thesis comprises three essays that touch on important topics in human capital research: understanding the determinants of skill formation over the life cycle, modelling the development of said skills, and investigating the interplay between skill formation and economic policy. The essays study the evolution of different components of human capital, covering different periods of the life cycle, spanning from early childhood to older ages, and draw on data from low, middle and high income countries. In chapters 2 and 3 I study determinants of human capital development in early childhood across different low and middle income settings, while the focus of chapter 4 is on older ages in a high income country. The overarching goal across all three chapters is to (i) provide empirical evidence on the determinants and dynamics of skill formation at different stages of the life cycle, and (ii) evaluate the implications of the findings for policy design and welfare. To achieve this objective, I combine reduced form and structural methods.

Chapter 2 focusses on the classroom environment as an input in the process of skill formation of children. In joint work with Pedro Carneiro, Yyannu Cruz Aguayo and Norbert Schady, we investigate the impact of classroom rank on learning and other outcomes throughout elementary school. We use a unique experiment in elementary schools in Ecuador where, at the start of every grade, a cohort of students was randomly assigned across classrooms within a given school. Variation in peer groups resulting from randomly assigning students to classrooms means that two students with the same underlying ability and in the same school and grade will have different classroom ranks. Our unusually rich data on teacher perceptions of student ability, executive function, happiness, depression, self-esteem, grit, growth mindset, allow us to study how different inputs affect the formation of multiple skills in elementary school. We find that children with higher classroom rank at the beginning of the academic year have significantly higher math test scores at the end of that grade. Classroom rank in math, not language, drives our results. The impact of classroom math rank is larger for younger children, and grows substantially over time. Exogenous changes in classroom rank in math also improve cognitive flexibility, non-cognitive skills, and teacher perceptions of students. This paper is the first to analyse rank effects in education using an experiment with multiple rounds of random assignment, with essentially perfect compliance. The fact that we have data on children randomly assigned to different classrooms in every grade in elementary school means that we can convincingly test whether classroom rank has larger effects in some grades than in others. Moreover, because we follow children over the entire elementary school cycle, we can credibly estimate how the effects of classroom rank evolve over time. As we show, both of these considerations—differences across grades and changes in the impact of classroom rank over time—are important, at least in the setting that we study.

In chapter 3, which is joint work with Orazio Attanasio, Costas Meghir and Emily Nix, we use high quality data from Ethiopia and Peru drawn from the Young Lives Survey to estimate flexible specifications of the production functions for health and cognition. For each of these countries, we have observations for two different cohorts spanning most of childhood. In particular, the younger cohort is observed at ages 1, 5 and 8, while the older cohort is observed at ages 8, 12 and 15. These production functions are used to map the interaction of family background, the current skill level of the child across different domains, as well as investments in the child at each age into child development and growth. This approach allows us to identify the degree of persistence of different inputs into development and their influence on subsequent growth. We find that the production of health and cognition is quite similar in both Peru and Ethiopia. Specifically, in both countries we find that both cognitive skills and health are very persistent. We find some evidence that health is cross productive; health positively impacts the production of cognition at early ages. Investments have large impacts on the production of cognition, but the effect decreases with age. We also find in both countries that investments are endogenous and parents compensate for negative shocks. We then use our estimated framework for some counterfactual analysis. First, we examine the impact of increasing either investments alone or investments and health at age 5 for children with cognitive deficits at age 5. We find that this intervention leads to large gains in cognition which are sustained through age 15.

Next, we show that rich and poor children with identical baseline skills will end up with large gaps in cognitive skills by age 15 due to the fact that richer parents invest more in their children. Our paper offers innovations in a number of dimensions, including our particular attention on the functional form for the production functions and the comparison between two countries. Ultimately, the goal behind the characterization of these processes at different ages and across different settings is to identify important regularities that will help us understand the process and design more effective interventions.

In chapter 4, I investigate the dynamic relationship between health and labor market outcomes of older women. I develop a structural framework that allows for feedback effects of employment on health to depend on health status and other characteristics of the workers. The model is especially well suited for welfare analysis of retirement policies such as those incentivizing workers to extend their working lives. Estimation of the model exploits a policy change that increased the legal age of eligibility for state pensions of women in the United Kingdom. I focus on women, who were typically allowed to retire earlier than men and thus have experienced more sizeable changes to their working lives compared to men as a result of these policies. The paper has two main contributions. First, I shed light on the dynamic relationship between health and labor market outcomes of women around retirement age, allowing for a two-way interaction between employment and health. I study the role of health shocks and financial incentives for individual labor supply decisions of women at older ages, and how the impact of these shocks varies across the health distribution and for different education groups. I produce Marshallian and Frisch elasticities, as well as estimates of labor supply sensitivity to changes in health, to study differences in the labor supply response to health and financial shocks across different education groups and by health status. Second, I investigate the welfare consequences of policies that increase the state pension age, i.e. the age at which individuals become eligible to receive state pension benefits. I quantify differences in the welfare implications of the policy across the health distribution and education groups and, conversely, the effect that these policies have on health inequality arising from the effect employment has on health. I show that employment has negative effects on the health of women, and that these effects are stronger for those already in poor health. I also quantify the responses of employment to changes in health and show that these are sizeable when compared to responses to changes in earnings and are larger among women in worse health. Using the model, I show that the effects of extending the state pension age are very heterogeneous and tend to widen inequality in health. In particular, poorer women with low levels of health bear a larger cost from the reform than other groups.

This is because they cannot afford not to work in the absence of retirement benefits, but employment for them is more costly, and damages their already poor health.

Chapter 2

The Effect of Classroom Rank on Learning throughout Elementary School: Experimental Evidence from Ecuador

2.1 Introduction

There are many settings in economics where relative rank concerns are important. They emerge naturally in tournaments (Lazear and Rosen, 1981), they affect job satisfaction (Card, Mas, Moretti, and Saez, 2012), and can interact in interesting ways with social preferences (Bandiera, Barankay, and Rasul, 2005). Although the idea that rank is important in education dates back to Marsh (1987), recent years have seen a growing literature quantifying its importance (including Elsner and Isphording (2017); Elsner, Isphording, and Zölitz (2018); Tincani (2018); Denning, Murphy, and Weinhardt (2020), Murphy and Weinhardt (2020)), and showing medium- and long-term impacts on college attendance, earnings, and the probability of engaging in risky behaviors.[1] In this paper, we investigate the impact of classroom rank on learning and other outcomes throughout elementary school. We use a unique experiment in elementary schools in Ecuador where, at the start of every grade, a cohort of students

[1] Existing papers show causal effects of rank on many domains. Murphy and Weinhardt (2020) and Cicala, Fryer, and Spenkuch (2018) show that rank can have positive effects on test scores in primary and secondary school. Elsner and Isphording (2017) and Elsner, Isphording, and Zölitz (2018) document positive impacts of rank in high school on college enrollment and choice. In recent work, Denning, Murphy, and Weinhardt (2020) show that rank can also have long-term impacts on earnings later in life. Rank has also been shown to affect the likelihood of engaging in risky behaviors, as in Elsner and Isphording (2018).

was randomly assigned across classrooms within a given school. Compliance with the random assignment was almost perfect, 98.9 percent on average over the 7 years of the experiment. Variation in peer groups resulting from randomly assigning students to classrooms means that two students with the same underlying ability and in the same school and grade will have different classroom ranks. We find that increasing a child's classroom rank at the start of given grade, keeping own ability constant, raises end-of-grade achievement. Rank, however, is just a particular form of peer effects, which could influence outcomes in various ways. For this reason, we estimate models that, in addition to classroom rank, include average peer quality. We show that, because rank and peer quality are negatively related, estimates of the effect of classroom rank increase when we control for peer quality using a standard linear-in-means model. Our results are also robust to more flexible ways of incorporating peer quality (such as by using classroom fixed effects). On the other hand, we find that average peer quality only raises achievement in models that also control for classroom rank. We show that the effects of classroom rank are entirely driven by classroom rank in math, rather than language. Focusing on classroom rank in math, we show that rank effects are concentrated in the upper half of the rank distribution. Consistent with this result, we also find that rank effects are larger for children with higher lagged achievement, and for children who started kindergarten with higher levels of vocabulary. On the other hand, we find no evidence that the effect of classroom rank varies with the gender of the child. A novelty of our paper is the focus on differences in the effects of classroom rank by grade, and how these evolve over time. For this purpose, we divide our sample into children in "early" grades (1st and 2nd grades), "middle" grades (3rd and 4th grades) and "late" grades (5th and 6th grades) in elementary school. We first show that classroom math rank effects are largest in the early grades: moving a child from the 50th to the 75th percentile of classroom rank in 1st or 2nd grade increases her end-of-grade achievement by 1 percentile point of national achievement, while classroom rank in 5th and 6th grades has no effect on achievement. We reject the null that rank effects are the same in early, middle, and late grades (p-value<0.01). Our analysis then turns to the evolution of classroom rank effects over time. We show, remarkably, that the effects of early classroom math rank increase substantially as children age. After 4 lags the effect of classroom rank in 1st or 2nd grade is more than twice as large as it was originally. This result stands in sharp contrast with the effects of many other determinants of achievement, which tend to fade out.[2] The fact that the medium-term effects of early classroom rank in math

[2]See, for example, Chetty, Friedman, and Rockoff (2014a), and Jacob, Lefgren, and Sims (2010) for estimates of the fade-out of the effects of teacher quality, measured by teacher value added. For a review and discussion of fade-out in education interventions, with a particular focus on early childhood,

are larger than the short-term effects could occur for a number of reasons. One possibility is that students randomly assigned a high classroom rank in 1st (or 2nd) grade benefit from a virtuous cycle, where a high rank in grade t leads to higher learning and higher rank in grade $t+1$, which in turn lead to higher learning and higher ranks in the following grades. We find, however, that our estimates are more consistent with an alternative model where, in addition to its direct effect on achievement, early math rank affects an unobserved trait—for example, academic self-concept—which affects learning in later grades, and which does not depreciate over time. With this insight, we turn to other child outcomes. Between 1st and 4th grades, we collected data on child executive function (EF). EF refers to a set of skills that allow individuals to plan, focus attention, remember instructions, and juggle multiple tasks. It includes working memory, inhibitory control, and cognitive flexibility (Center for the Developing Child 2019). Executive function in childhood has been shown to predict a variety of outcomes in adulthood, including performance in the labor market, involvement in criminal activities, and health status, even after controlling for socioeconomic status in childhood (Moffitt, Arseneault, Belsky, Dickson, Hancox, Harrington, Houts, Poulton, Roberts, Ross, et al., 2011). We show that randomly assigning a child to a classroom where she is more highly-ranked in math improves her cognitive flexibility, but not attention or working memory. We also have data on a number of non-cognitive outcomes. At the end of 1st grade, we asked children whether they were happy in school, and at the end of 6th grade, we collected data on child depression, self-esteem, grit, and growth mindset. We show that children with higher math classroom rank (at the start of 1st grade) are more likely to say they are always happy (at the end of that grade), and have higher growth mindset scores (at the end of elementary school). Finally, we analyze whether classroom rank affects teacher perceptions of students. These perceptions could be important if, for example, teachers pay more attention to children they consider bright—and if these changes in teacher behaviors, in turn, affect future achievement. We first show that teacher perceptions of children at the top and bottom of the distribution are correlated with actual achievement, but very imperfectly. We then show that, controlling for own lagged achievement, students who have a higher classroom rank in one grade are more likely to be thought to be at the top of the class by their teachers in the next grade.[3] In sum, children with higher classroom rank

see Bailey, Duncan, Cunha, Foorman, and Yeager (2020). For a recent example from a setting similar to ours, see Barrera-Osorio, Gonzalez, Lagos, and Deming (2020) on the fade-out of information on student performance provided to parents in Colombia.

[3]We asked teachers who were the 5 children in the class with the highest achievement, and those with the lowest achievement. Using these data, we regress indicator variables for children said to be in the top or bottom 5 students in their class by their teacher in grade $t+1$, on rank in grade t, controlling for achievement at the end of $t-1$.

have higher end-of-grade achievement. The effects of early classroom rank in math in 1st and 2nd grade increase substantially over time. Children with higher classroom rank in 1st grade also exhibit higher levels of happiness and growth mindset. More highly-ranked children are thought to be brighter by their future teachers. Our paper extends the economics literature on the impacts of rank in important ways. First, ours is the first paper that analyzes rank effects in education using an experiment with multiple rounds of random assignment, with essentially perfect compliance. The fact that we have data on children randomly assigned to different classrooms in every grade in elementary school means that we can convincingly test whether classroom rank has larger effects in some grades than in others. Moreover, because we follow children over the entire elementary school cycle, we can credibly estimate how the effects of classroom rank evolve over time. As we show, both of these considerations—differences across grades and changes in the impact of classroom rank over time—are important, at least in the setting that we study. In addition, because we have unusually rich data on teacher perceptions of student ability, executive function, happiness, depression, self-esteem, grit, growth mindset, we can study how different inputs affect the formation of multiple skills in elementary school. The rest of the paper proceeds as follows. In section 2.2 we describe the setting and data, in section 2.3 we discuss our empirical strategy. Results are in section 2.4, and we conclude in section 2.5.

2.2 Setting and Data

We study the acquisition of math, language, executive function, and non-cognitive skills in Ecuador, a middle-income country in South America. As is the case in most other Latin American countries, educational achievement of young children in Ecuador is low (Berlinski and Schady, 2015). The data we use comes from an experiment in 202 schools.[4] Schools have at least two classrooms per grade (most have exactly two). An incoming cohort of children was randomly assigned to kindergarten classrooms within schools in the 2012 school year. These children were reassigned to 1st grade classrooms in 2013, to 2nd grade classrooms in 2014, to 3rd grade classrooms in 2015, to 4th grade classrooms in 2016, to 5th grade classrooms in 2017, and to 6th grade classrooms in 2018. Compliance with the assignment rules was very high—98.9 percent on average. Random assignment means that we can effec-

[4] Araujo, Carneiro, Cruz-Aguayo, and Schady (2016) discuss in detail the selection of schools in this study. They show that the characteristics of students and teachers in our sample are very similar to those of students and teachers in a nationally-representative sample of schools in Ecuador.

tively deal with concerns about any purposeful matching of students with teachers and peers that often arise in non-experimental settings. We provide further details on the classroom assignment rules and compliance with randomization in Appendix 2.8.1. We have baseline data on maternal education and wealth of households, whether a child attended preschool, and her receptive vocabulary at the beginning of kindergarten. Table 2.1, Panel A, provides summary statistics for children in our sample. The table shows that children were 5 years of age on the first day of kindergarten, and half of them are girls. Mothers were in their early 30s and fathers in their mid-30s. Both parents had on average just under 9 years of schooling, which corresponds to completed middle school. The average receptive vocabulary score of children in the sample is 1.7 SDs below the level of children that were used to norm the sample for the test.[5] Table 2.1, Panel B, summarizes characteristics of classrooms and teachers. Average class size is 36. The average teacher in the sample has 18 years of experience. Eighty-two percent of teachers are women, and 82 percent are tenured. An important consideration in interpreting our results is that there is always one teacher per classroom, without a classroom aide, and that the same teacher teaches all academic subjects (all subjects other than physical education and, when they are available, art and music). We collected data on math and language achievement at the end of each grade between kindergarten and 6th grade. For both subjects, tests were a mixture of material that teachers were meant to have covered explicitly in class—for example, in math, addition or subtraction; material that would have been covered, but probably in a somewhat different format—for example, simple word problems; and material that would not have been covered at all in class but that has been shown to predict current and future math achievement—for example, the Siegler number line task.[6] We aggregate responses in math and, separately, language, by Item Response Theory (IRT), and calculate an average achievement score that gives the same weight to math and language.[7]

[5]To measure baseline receptive vocabulary, we use the Test de Vocabulario en Imágenes Peabody (TVIP) (Dunn et al 1986), the Spanish-speaking version of the much-used Peabody Picture Vocabulary Test (PPVT). The TVIP was normed on samples of Mexican and Puerto Rican children. It has been used widely to measure development among Latin American children. See Paxson and Schady (2007) for a comparison of vocabulary scores between children in Ecuador and the U.S., and Schady, Behrman, Araujo, Azuero, Bernal, Bravo, Lopez-Boo, Macours, Marshall, Paxson, and Vakis (2015) for evidence on levels and socioeconomic gradients in the TVIP in five Latin American countries, including Ecuador.

[6]The number line task works as follows. Children are shown a line with the two clearly marked endpoints—for example, in 1st grade, the left end of the line is marked with a 0, and the right end is marked with a 20. They are then asked to place various numbers on the line—for example, the number 2 or the number 18. The accuracy with which children place the numbers has been shown to predict general math achievement (see Siegler and Booth (2004)).

[7]Our results are very similar if, instead, we calculate a simple sum of correct responses within blocks of questions on the test, and give equal weight to each of these test sections (as in Araujo et al.

In every grade between kindergarten and 4th grade, we tested child executive function. EF includes a set of basic self-regulatory skills which involve various parts of the brain, but in particular the prefrontal cortex.[8] Executive function is generally thought of as having three domains: working memory, inhibitory control, and cognitive flexibility. It is an important determinant of how well young children adapt to and learn in school. Basic EF skills are needed to pay attention to a teacher; wait to take a turn or raise one's hand to ask a question; and remember steps in, and shift from one approach to another, when solving a math problem, among many other tasks that children are expected to learn and carry out in the classroom. Children with high EF levels are able to concentrate, stay on task, focus, be goal-directed, and make good use of learning opportunities. Low levels of EF are associated with low levels of self-control and "externalizing" behavior, including disruptive behavior, aggression, and inability to sit still and pay attention, which affects a child's own ability to learn, as well as that of her classmates (Séguin and Zelazo (2005)).[9] At the end of each grade, we asked teachers who were the 5 children with the highest achievement, and 5 with the lowest achievement.[10] In 1st grade, we asked children whether they were happy in school and in their classroom (two separate questions). In both cases, children had the option of answering "always", "sometimes", or "never". Most children in the sample answered "always" to both questions, so we use their responses to construct a single variable for children who were always happy, almost always happy, or mostly happy. In 6th grade, we collected data on child depression, self-esteem, growth mindset, and grit. To measure child depression, we used the Patient-Reported Outcomes Measurement Information System (PROMIS) Depression Scale for children aged 11-17 years, developed by the American Psychiatric Association.[11] To measure self-

(2016)).

[8] Volumetric measures of prefrontal cortex size predict executive function skills; children and adults experiencing traumatic damage to the prefrontal cortex sustain immediate (and frequently irreversible) deficits in EF (Nelson and Sheridan 2011, cited in Obradovic, Portilla, and Boyce (2012)).

[9] Working memory measures the ability to retain and manipulate information; for example, 2nd grade child were asked to remember (increasingly long) strings of numbers and repeat them in order and then backwards. Cognitive flexibility measures the ability to shift attention between tasks and adapt to different rules; for example, 1st grade children were shown picture cards that had trucks or stars, red or blue, and were asked to first sort cards by shape (trucks versus stars), and then by color (red versus blue). Inhibitory control refers to the capacity to suppress impulsive responses; for example, kindergarten children were quickly shown a series of flash cards that had either a sun or a moon and were asked to say the word "day" when they saw the moon and "night" when they saw the sun. We calculate scores on each of the three domains in executive function, as well as a measure of overall EF scores that gives one-third of the weight to each individual domain. We do not have data on inhibitory control in 1st grade. In this grade, the overall measure of executive function includes only cognitive flexibility and working memory, with equal weight given to both.

[10] Importantly, these data are not disaggregated by subject—that is, we did not ask teachers who they thought were top and bottom performers in math and, separately, in language.

[11] Olino, Yu, McMakin, Forbes, Seeley, Lewinsohn, and Pilkonis (2013), Klein, Dougherty, and Olino

esteem, we selected 5 questions from the National Longitudinal Study of Adolescent to Adult Health (Add Health).[12] To measure Growth Mindset, we selected 10 of the 20 questions on the Dweck "Mindset Quiz"; growth mindset refers to the belief that intelligence is malleable, rather than fixed, and can be increased with effort (Blackwell, Trzesniewski, and Dweck, 2007). Finally, to measure grit, we adapted 4 questions from the 8-item Grit Scale for children (Duckworth and Quinn, 2009); grit refers to the capacity of individuals to persevere at a given task. For each of these 6th grade outcomes, we aggregated responses by factor analysis. We also calculate an overall non-cognitive score that gives the same weight to each of the individual tests. Most of the tests were applied to children individually (as opposed to in a group setting) by specially trained enumerators.[13] All tests, other than the non-cognitive tests applied in 6th grade, were applied in school. In all tests, to choose questions, we piloted the test; made changes as necessary; and selected questions that could be understood by children in our context, and which showed reasonable levels of variability in the pilot. Further details on child assessments are provided in Appendix 2.8.2.

2.3 Empirical Strategy

2.3.1 Main Model

Our goal is to estimate the impact of child i' s classroom rank on her subsequent learning in elementary school. The dataset we use allows us to construct measures of classroom rank, lagged achievement, and achievement at the end of the current grade for children between 1st and 6th grades. With these data, we can investigate the impact of classroom rank in the short- and medium-run, starting as early as 1st grade.[14] To begin our discussion, we focus on two questions. First, how should we measure rank, and which measure of rank is likely to be more relevant for future learning? Second, how do we identify the causal impact of rank on learning? The two issues are interlinked in our setting, so we discuss them together. In the experiment we study, in each school, children were randomly assigned to classrooms at the start

(2005), and Aylward and Stancin (2008) argue that the PROMIS depression scale has superior qualities (greater precision, more internal reliability, and more discriminant validity) than other commonly-used depression scales, including the Beck Depression Inventory (BDI), the Children's Depression Inventory (CDI), and the Center for Epidemiologic Studies-Depression (CES-D) scale.

[12] See Harris and Udry (2018) for a description of the Add Health data. The questions on self-esteem in Add Health build on the much-used Rosenberg Self-Esteem Scale (Rosenberg (2015)).

[13] The only exception is some of the achievement tests in 4th through 6th grades, which were applied in a group setting.

[14] We can construct measures of end-of-grade math and language achievement at the end, but not the beginning, of kindergarten. For this reason, our analysis focuses on 1st through 6th grades.

of every grade. As a result, every student has a randomly-assigned set of peers in each grade, so two students with the same underlying ability can nevertheless have different classroom ranks. Our study exploits this variation by estimating the impact of rank at the beginning of grade t on learning at the end of grade t, as well as on learning in subsequent grades (until the end of 6th grade). Achievement at the start of grade t is measured using tests administered at the end of grade $t-1$. Beginning-of-grade classroom rank for each student is based on her achievement at the end of $t-1$ and the achievement of her randomly-assigned classmates. From now on, we refer to this measure simply as classroom rank.[15] Our approach assumes that students and teachers react to the beginning-of-grade student classroom rank. This makes most sense if they can perceive their rank, and act on it, fairly early during the school year. Furthermore, although school rank may also be important, we are not able to assess its impact as convincingly, since random assignment happens within schools. Throughout the paper we denote $Y_{i,s,c,t}$ as student i' s performance (measured by an index of math and language scores), at the end of grade t, in school s, and classroom c. To be consistent with the literature, we define $Y_{i,s,c,t}$ in terms of percentiles of national rank.[16] $CR_{i,s,c,t}$ is student i' s classroom rank at the start of grade t, when the student is randomly assigned to classroom c. In our simplest model, we pool observations from all grades and estimate:

$$Y_{i,s,c,t} = \beta CR_{i,s,c,t} + g_t(Y_{i,s,c,t-1}) + \delta_{st} + \epsilon_{i,s,c,t} \tag{2.3.1}$$

where δ_{st} is a school (by grade) fixed effect and $\epsilon_{i,s,c,t}$ is a residual. In this model, β is restricted to be the same across all grades, but all other parameters are allowed to be grade-specific. $g_t(Y_{i,s,c,t-1})$ is a third-order polynomial in $Y_{i,s,c,t-1}$.[17] We also estimate models in which, instead of pooling data for all six grades, we estimate separate coefficients on β for 1st and 2nd grades ("early" grades), 3rd and 4th grades ("middle" grades) and 5th and 6th grades ("late" grades).[18] Next, we separately analyze the

[15]One issue we face is that we only observe end of grade $t-1$ scores for students who were in a school in our sample in grade $t-1$, which means that we do not know what these test scores are for students who arrived at the school in grade t. Therefore, we cannot compute grade t ranks for these students, and our measures of ranks for all other children ignore the fact that new entrants are in their classroom. This will introduce random measurement error in rank.

[16]See, for example, Murphy and Weinhardt (2020).

[17]As Murphy and Weinhardt (2020) emphasize, it is important to use flexible specifications for this function, to avoid the potential problem that β does not capture a true rank effect, but is instead an artefact of the misspecification of this function. Our robustness checks show that considering polynomials in lagged scores of order higher than three does not lead to substantial changes in the results. For this reason, in our main empirical specification $g_t(Y_{i,s,c,t-1})$ is a cubic polynomial in its argument.

[18]In Appendix 2.8.3 we also present estimates where β varies by grade in an unrestricted way, and therefore it is also indexed by t. Those estimates are noisier than the ones we focus on in the paper.

effects of classroom rank in math and language:

$$Y^k_{i,s,c,t} = \beta^k CR^k_{i,s,c,t} + g_t(Y^k_{i,s,c,t-1}) + \delta^k_{st} + \epsilon^k_{i,s,c,t} \tag{2.3.2}$$

where the k superscript refers to a subject, math or language. Up to this point, we have assumed that the effect of classroom rank on learning is linear in classroom rank. It is quite possible that this is not the case—it may be, for example, that rank has a different effect at the top and bottom of the (rank) distribution. Therefore, we also consider a version of equation (2.3.1) where the effect of classroom rank on learning is not constrained to be linear:

$$Y_{i,s,c,t} = \beta(CR_{i,s,c,t}) + g_t(Y_{i,s,c,t-1}) + \delta_{st} + \epsilon_{i,s,c,t} \tag{2.3.3}$$

where $\beta(CR_{i,s,c,t})$ is now a flexible function of $CR_{i,s,c,t}$. In our preferred specification we discretize $CR_{i,s,c,t}$ into q values (forcing it to take q=10 values, corresponding to 10 deciles of the distribution), so $\beta(CR_{i,s,c,t}) = \sum_{q=1}^{1} 0(\beta_q CR_{i,s,c,t,q}$ (where $CR_{i,s,c,t,q}$ is an indicator variable that takes value 1 if $CR_{i,s,c,t}$ is in decile q).

2.3.2 Alternative Models

Other papers in this literature use different measures of rank and different specifications. A leading example is Murphy and Weinhardt (2020), who study the impact of school rank at the end of elementary school on learning in secondary school. In contrast, we study the impact of classroom rank at the beginning of a grade on learning occurring in that grade. Because in Murphy and Weinhardt (2020), rank is measured at the end of elementary school, it is a result of a student's position relatively to her peers, but also of the student's reaction to her peers and any subsequent feedback, as well as other school shocks occurring between the beginning and the end of elementary school.[19] If we were to use the Murphy and Weinhardt (2020) specification instead of ours (and using classroom rank instead of school rank) we would estimate:

$$Y_{i,s,c,t} = \beta CR'_{i,s,c,t-1} + g_t(Y_{i,s,c,t-1}) + \delta_{st} + \epsilon_{i,s,c,t} \tag{2.3.4}$$

where $CR'_{i,s,c,t-1}$ is the classroom rank at the end of grade $t-1$, computed using scores at the end of $t-1$ relative to peers in $t-1$. The main reason why we

[19] Murphy and Weinhardt (2020) document the impact of this measure on future learning, when a student moves to another school and experiences a different set of peers. In their model students are motivated to work hard in secondary school because of the rank they experienced and perceived in their past school, as opposed to their rank in the current school.

use equation (2.3.1), rather than (2.3.4), as our preferred specification is that it follows directly from our experimental design. Classroom ranks at the start of a grade are randomly assigned and cannot be modified by student effort or other unobserved shocks, whereas classroom ranks at the end of a grade are both a result of random assignment of peers, student effort, peer effort, and potentially even the responses of teachers and parents. Also, our approach estimates the effects of classroom rank experienced in the same year as we measure learning, which is arguably more relevant in the short run for a "big fish little pond" mechanism (as in Marsh (1987)).[20] That said, students may not know their rank at the start of the grade, and may take some time to learn about it. In contrast, in Murphy and Weinhardt (2020) students are more likely to have a reasonable perception of their rank since it is measured at the end of elementary school. Therefore, we present results from estimating (2.3.4) as a robustness check.

2.3.3 Peer Effects

By construction, the quality of a student's peers is negatively correlated with her rank. Non-parametrically it is impossible to distinguish the impact of rank from that of other impacts of peers. This is a concern faced by every paper focused on rank effects, which are a particular form of peer effects. However, there are some forms of peer effects that can be distinguished from rank effects. For example, in many empirical peer effects papers, a student's outcome depends on the average ability of her peers (also known as the linear-in-means model).[21] We augment our specification to include average peer ability:

$$Y_{i,s,c,t} = \beta CR_{i,s,c,t} + g_t(Y_{i,s,c,t-1}) + \theta \bar{Y}_{i,s,c,t-1} + \delta_{st} + \epsilon_{i,s,c,t} \qquad (2.3.5)$$

where $\bar{Y}_{i,s,c,t-1}$ is average peer ability in the classroom at the beginning of grade t, based on end of grade $t-1$ test scores (using leave-one-out estimates, as is standard in this literature; see, for example, Duflo, Dupas, and Kremer (2011)). One can also allow for more general peer influences as long they have the same impact on all

[20] That said, one advantage of using end-of-classroom rank in our setting is that we can construct rank using everyone in the classroom at the end of the previous grade, while with our preferred measure of beginning of grade rank excludes new school entrants for whom we do not have test scores at the end of the previous grade, introducing measurement error in our preferred measure of classroom rank (on average, 7.1 percent of students are new entrants in each classroom). However, as we show in the robustness checks below, our results are robust to a standard multiple imputation procedure for missing data.

[21] Some examples include Duflo, Dupas, and Kremer (2011), Booij, Leuven, and Oosterbeek (2017), Feld and Zölitz (2017). See Epple and Romano (2011) and Ioannides (2011) for recent surveys of the peer effects literature.

individuals in the same classroom, by including classroom fixed effects in the model:

$$Y_{i,s,c,t} = \beta CR_{i,s,c,t} + g_t(Y_{i,s,c,t-1}) + \delta_{sct} + \epsilon_{i,s,c,t} \tag{2.3.6}$$

where δ_{sct} are classroom (by grade) fixed effects. This is analogous to the main approach proposed by Murphy and Weinhardt (2020). As explained in their paper, there is residual variation in rank across individuals assigned to different classrooms that one can exploit even after accounting for classroom fixed effects, due to differences in the distribution of ability across classrooms.[22] Equations (2.3.5) and (2.3.6) are estimated below.

2.3.4 Dynamics

To estimate the dynamics of rank effects, we begin with specifications of the following form:

$$Y_{i,s,c,t+l} = \beta_{t,l} CR_{i,s,c,t} + g_{t+l}(Y_{i,s,c,t-1}) + \delta_{s,t+l} + \epsilon_{i,s,c,t+l} \tag{2.3.7}$$

We estimate these regressions separately for "early", "middle", and "late" grades, as discussed above. When $l = 0$, equation (2.3.7) is equivalent to (2.3.1), and provides estimates of the short-run impact of classroom rank at the start of grade t, $CR_{i,s,c,t}$, on learning at the end of that same grade, $Y_{i,s,c,t}$. We label this effect $\beta_{t,0}$. When $l > 0$, equation (2.3.7) provides estimates of the medium-term effect of classroom rank at various lags, which we label $\beta_{t,l}$. $\beta_{t,l}$ (medium-run impact) and $\beta_{t,0}$ (short-run impact) are related through three main channels: (i) class rank in grade t affects learning at the end of that grade, and therefore also affects student achievement in grade $t+1$ and in subsequent grades, through the function $g_{t+1}(Y_{i,s,c,t})$ in the $(t+1)$ version of equation (2.3.1); (ii) since learning at the end of grade t is affected, classroom rank in $t+1$ and subsequent grades is also affected, which can have a further impact on learning in those grades, captured by $\beta_{t+1,0}$ in equation (2.3.1); (iii) in addition, class rank in grade t may affect other skills not captured by our tests at the end of that grade (unobserved skills), but which nevertheless can affect learning in grades $t+1$ and beyond. To quantify the importance of these channels, we first estimate an additional equation relating classroom rank at the beginning of grade $t+1$ with learning at the end of grade t, which we will assume can be approximated by the

[22] Note that this is not equivalent to exploring within-classroom variation. In each classroom, individual ability and classroom rank are perfectly correlated, so one would not be able to estimate this model classroom by classroom (allowing for classroom-specific parameters). The model is identified because it imposes that the impact of classroom rank is the same across classrooms. With this assumption, it would be identified even if we allowed for some restricted forms of interactive fixed effects, as in, for example, Bai (2009).

following linear relationship:

$$CR_{i,s,c,t+1} = \delta_{t+1} + \gamma_{t+1} Y_{i,s,c,t} + \tau_{s,t+1} + \eta_{i,s,c,t+1} \tag{2.3.8}$$

where $\tau_{s,t+1}$ is a school fixed effect, and $\eta_{i,s,c,t+1}$ is the variation coming from random assignment of students to different peer groups. In practice, as we show below, when $Y_{i,s,c,t}$ is the national percentile rank, $\gamma_{t+j} \approx 1$. Suppose, for simplicity, that $g_t(Y_{i,s,c,t-1})$ is also linear: $g_t(Y_{i,s,c,t-1}) = \lambda_t Y_{i,s,c,t-1}$. Taking equations (2.3.1) and (2.3.8) together, and assuming that all medium-term impacts of rank on learning operate through observed tests scores and observed rank:

$$\frac{\partial Y_{i,s,c,t}}{\partial CR_{i,s,c,t}} = \beta_{t,0}$$

$$\frac{\partial Y_{i,s,c,t+1}}{\partial CR_{i,s,c,t}} = \left(\frac{\partial Y_{i,s,c,t+1}}{\partial CR_{i,s,c,t+1}} \frac{\partial CR_{i,s,c,t+1}}{\partial Y_{i,s,c,t}} + \frac{\partial Y_{i,s,c,t+1}}{\partial Y_{i,s,c,t}}\right) \frac{\partial Y_{i,s,c,t}}{\partial CR_{i,s,c,t}} = (\beta_{t+1,0}\gamma_{t+1} + \lambda_{t+1})\beta_{t,0}$$

Similarly:

$$\frac{\partial Y_{i,s,c,t+2}}{\partial CR_{i,s,c,t}} = (\beta_{t+2,0}\gamma_{t+2} + \lambda_{t+2})(\beta_{t+1,0}\gamma_{t+1} + \lambda_{t+1})\beta_{t,0}$$

Substituting these expressions, in subsequent grades we get:

$$\frac{\partial Y_{i,s,c,t+l}}{\partial CR_{i,s,c,t}} = \beta_{t,0} \prod_{j=1}^{l} (\beta_{t+j,0}\gamma_{t+j} + \lambda_{t+j}) \tag{2.3.9}$$

Equation (2.3.9) tells us how the medium-term impact of rank in grade t on learning in grade $t + j$ depends on the short-term impact of rank at the beginning of each grade on learning at the end of that grade ($\beta_{t,0}$), the impact of learning in one grade on learning in the subsequent grade (λ_t), and the impact of learning in one grade on classroom rank in the subsequent grade (γ_t). We also note that, because (as we show below) $\gamma_{t+j} \approx 1$, equation (2.3.9) indicates that it is possible that we may see little or no decay of rank effects over time. This is because a high classroom rank early in elementary school can in principle lead to a self-fulfilling cycle, where a high rank produces high learning, which in turn leads to a high rank, which in turn leads to high learning. Observed differences between estimates of actual medium-term impacts of rank ($\beta_{t,l}$), and ($\frac{\partial Y_{i,s,c,t+l}}{\partial CR_{i,s,c,t}}$) from equation (2.3.9) tell us about the importance of unobserved skills as mediators of the medium-term impacts of rank.

2.3.5 Executive Function, Non-cognitive Skills, and Teacher Perceptions

To estimate effects of ability classroom rank on happiness in 1st grade and non-cognitive skills in 6th grade, we run regressions comparable to (2.3.1), replacing achievement in grade t with the relevant outcome.[23] To estimate rank effects on executive function, we also use the model in (2.3.1), but add to this model a third-order polynomial in lagged EF (in addition to the polynomial in lagged achievement). These regressions use information in 1st through 4th grades, where data on current and lagged EF are available. Finally, as discussed above, we have data on teacher perceptions of students—specifically, a list of the 5 students each teacher thought had the highest, and lowest, achievement in their classrooms. We generate indicator variables for children who are seen to be at the top and, separately, bottom of their classroom by their teachers, and use them as outcomes.

2.4 Results

2.4.1 Graphical Evidence

To motivate our analysis, we start with some simple figures. For this purpose, we first sort children into ventiles on the basis of their test scores in math and language at the end of grade $t-1$ (say, end of kindergarten). Then, within each ventile, we calculate average test scores at the end of grade t (end of 1st grade) for two groups of children: those who, relative to other children in that ventile, were randomly assigned to classrooms where their rank at the beginning of t was "high"—classroom rank in the top 25 percent for that ventile—and those in classrooms where their rank was "low"—in the bottom 25 percent for that ventile. If classroom rank has a positive effect on test scores, we would expect the line that corresponds to high-ranked children to be above that which corresponds to low-ranked children. Results are in Figure 2.1. Panel A focuses on the short-term effects of classroom rank in 1st grade. The figure shows that children with high classroom ranks have higher achievement than those with low classroom ranks, but only above the 40th percentile of the distribution

[23]Child happiness is only available in 1st grade, so we run regressions of child happiness in 1st grade on math classroom rank at the beginning of 1st grade and the polynomial on math achievement at the end of kindergarten. Non-cognitive skills are only available at the end of 6th grade, so it is not obvious whether we should regress these skills on rank in 1st grade, 6th grade, or any grade in between. Because (as we argue below) we are most interested in possible medium-term effects of early rank (rank in 1st and 2nd grade), and to be consistent with the results on child happiness, we report the results of 6th grade non-cognitive skills on 1st grade math classroom rank.

of national rank. Panel B compares these two groups of children at the end of 3rd grade. The figure shows that the vertical distance between the two lines is larger than in Panel A, indicating that the effect of 1st grade classroom rank increases over time. Panel C focuses on the short-term effects of classroom rank in 4th grade. Children with higher classroom ranks appear to have higher achievement at the at the end of 4th grade, but the difference is quite small. Panel D focuses on these same children at the end of 6th grade. The lines in this panel are very similar to those in Panel C, suggesting that the (modest) effects of 4th grade classroom rank do not grow over time. In sum, Figure 2.1 suggests that: the effects of classroom rank are larger in 1st grade than in 4th grade; the 1st grade effects are larger in the upper half of the distribution; and these effects increase over time.

2.4.2 Static Model

Table 2.2 reports estimates of the effect of classroom rank on learning, measured by an index of math and language, using equation (2.3.1) above. Column (1) shows that the coefficient on β is 0.018, with a standard error of 0.006.[24] This implies that moving a child from the 50th to the 60th percentile of classroom rank increases her end-of-grade achievement by 0.18 percentiles of the national distribution.[25] Column (2) replaces the measure of classroom rank with the average achievement of classroom peers—the linear-in-means peer effects model. The coefficient on average peer quality is positive but is not significant (0.013, with a standard error of 0.014). In column (3), we include both classroom rank and the baseline achievement of peers, as in equation (2.3.5) above. Because rank and peer quality are negatively related, the coefficient on classroom rank increases to 0.029 (with a standard error of 0.007), and the measure of average peer quality also increases and is now significant (coefficient of 0.046, with a standard error of 0.015). In column (4), finally, we include classroom-by-grade fixed effects. In this specification, the coefficient on β is 0.035, with a standard error of 0.007—more than twice as large in magnitude as that in the specification in column (1). Table 2.3 is based on estimates of equation (2.3.2) above. Specifically, we report estimates of the effects of classroom rank in math on achievement in math and language (Panels A and B, respectively), as well as the effects of classroom rank

[24] Standard errors for all models that pool data across grades are clustered at the student level.

[25] To get a sense of magnitude, we take all children who are in the same school and grade, have the same value of lagged achievement in $t-1$, but are assigned to different classrooms in grade t. The (absolute value) of the median difference in classroom rank between children in these pairs is 5.5 percentiles (that is, on average, in these pairs of identical children, one child has a classroom rank of 47.25 and the other has a rank of 52.75. At the 75th and 90th percentiles of the difference, the values are 9.8 and 14.7 percentiles, respectively.

in language on achievement in language and math (Panels C and D, respectively). The table shows that classroom rank in math affects math achievement (coefficient of 0.026, with a standard error of 0.007) and, to a lesser extent, language achievement; classroom rank in language does not affect either math or language achievement.[26]
Table 2.4 reports the results from a number of robustness checks. Panel A refers to specifications with school-by-grade fixed effects (corresponding to equation (2.3.1) above), while Panel B refers to specifications with classroom-by-grade fixed effects (corresponding to equation (2.3.6) above). To facilitate comparisons, column (1) in Panel A reproduces the coefficient and standard error from column (1) in Panel A of Table 2.3, while column (1) in Panel B corresponds to column (4) in Panel A of Table 2.3. Columns (2) to (4) report estimates where $g_t(Y_{i,s,c,t-1})$ is specified as a polynomial of orders 1, 2, and 4, respectively (as opposed to our main estimates, in which we include a cubic in lagged achievement). The estimated classroom rank effects are larger when we include only linear or quadratic terms in lagged achievement. Reassuringly, however, the coefficient on classroom rank is essentially unchanged when we include a polynomial of order 4 (rather than order 3) in lagged achievement as a control. Column (5) shows that, as expected given the random assignment, including controls for child gender, as well as age and its square, does not affect our results. There are between 14,322 (kindergarten) and 17,529 (5th grade) students per grade in our data. These are not always the same students. Typically, from one year to the next, between 7.5 and 10 percent of students leave our sample of schools (the exception is the transition from kindergarten to 1st grade, where this number is 15 percent).[27] Selective attrition out of our sample of schools could generate a correlation between $CR_{i,s,c,t}$ and $\epsilon_{i,s,c,t}$, which may bias estimates of the effect of classroom rank. To assess whether our estimates are likely to be affected by these considerations, we use a standard inverse probability weighting (IPW) correction, which gives greater weight to observations that had a higher probability of being lost to follow-up.[28] Column (6) in Table 2.4 indicates that our estimates are robust to this correction for

[26]We do not know why math rank, but not language rank, affects achievement. It is in principle possible that classroom rank in math is more visible to students and teachers than is the case with language rank. However, we do not find strong evidence that this is the case in our setting. As we discuss below, teachers appear to place similar weights on math and language achievement in determining which students in their class have the highest and lowest achievement. We note that it is not uncommon in the literature to find larger effects of school-based interventions on math than on language (see the discussion in Fryer (2017)). In the U.S., teachers have larger effects on math than on language achievement (Hanushek and Rivkin, 2010).

[27]Similarly, in any given grade, between 7 and 13 percent of students are new entrants to the sample (with the exception of 1st grade, where this value is 24 percent). New students are randomly assigned to classrooms just like any other students.

[28]We regress attrition on gender, age and its square, a third-order polynomial in lagged test scores, and school-by-grade fixed effects (in Panel A) or classroom-by-grade fixed effects (in Panel B).

missing data. In column (7), finally, we present estimates of equation (2.3.4). As we discuss above, in this specification—which is similar in character to that used by Murphy and Weinhardt (2020)—classroom rank refers to rank at the end of $t-1$, rather than at the beginning of t. Table 2.4 shows that these estimates are substantially larger than those from our preferred specification.[29] The approach to classroom rank we take therefore yields conservative estimates of rank effects on learning.

2.4.3 Heterogeneity

We begin our analysis of heterogeneity by estimating effects at different points in the distribution of classroom rank, again focusing on the effects of math classroom rank on math achievement. In Figure 2.2 we graph coefficients and confidence intervals on deciles 1 through 4, and 7 through 10 from equation (2.3.3) above, with deciles 5 and 6 as the omitted category. The figure shows that there is essentially no impact (or a negative impact) of classroom rank in the bottom half of the distribution. For example, we cannot reject the null that being in the lowest decile of classroom math rank has the same effect on achievement as being in the middle of the distribution of rank. The coefficients that correspond to deciles 7 through 10, on the other hand, are all positive, and are generally larger in the higher deciles (so that the coefficient for the 10th decile is larger in magnitude than that for the 7th decile). In this case, we can reject the null that being in the highest decile of classroom math rank has the same effect on achievement as being in the middle of the distribution. In Table 2.5, we analyze other possible sources of heterogeneity in the effect of classroom rank. We first focus on gender. In theory, if girls have different levels of self-confidence than boys (as in Bordalo, Coffman, Gennaioli, and Shleifer (2019)), or react differently to competition than boys (as in Niederle and Vesterlund (2011)), it is possible that they react differently to rank. In Table 2.5, we present estimates of a specification where all the coefficients in equation (2.3.2) for math are interacted with gender. Girls have significantly lower math scores than boys, but the impact of classroom rank on learning is the same for girls and boys. Finally, we interact classroom rank with vocabulary at the beginning of kindergarten, or with lagged achievement. These results show that classroom rank effects are substantially and significantly larger for children with higher baseline vocabulary levels, as well as for those with higher lagged achievement. In sum, and consistent with the results in Figure 2.2, Table 2.5 shows that classroom

[29] It is interesting that classroom rank effects estimated by (2.3.4) are larger than those estimated by 2.3.1. As mentioned above, this could happen because children are more aware of end-of-grade rank than beginning-of-grade rank and therefore react more to it, or because end-of-grade rank captures other aspects of the school experience besides rank.

rank seems to have larger impacts for brighter children.

2.4.4 Dynamics

We begin our analysis of dynamics by estimating effects of classroom rank separately for children in the "early" grades (1st and 2nd grade), "middle" grades (3rd and 4th grade), and "late" grades, both contemporaneously (without lags, as in equation (2.3.1) above) and at various lags (as in equation (2.3.7)). These results are in Table 2.6.[30] Column (1) shows that the short-term effect of math classroom rank in the early and middle grades are both positive and of a comparable magnitude. On the other hand, classroom rank in 5th and 6th grades has no effect on achievement. Chi-squared tests reject the null that the coefficients on the early, middle, and late grades are the same (p-value: 0.004). We next turn to the evolution of rank effects over time. Specifically, in columns (2) through (5), we report estimates of the effect of classroom math rank on math achievement after (up to) 1, 2, 3, and 4 lags, respectively. In the first row, corresponding to rank in the early grades, the coefficients increase monotonically over time—the impact after 4 lags is 0.093 (with a standard error of 0.018), more than twice as large as the short-term effect. We reject the null that the coefficients for all lags are the same (p-value: 0.04). On the other hand, the coefficients in the second row of the table show that the classroom rank effects in the middle grades decline, although we cannot reject the null that the coefficients for lag=0 and lag=2 are the same (p-value: 0.32).[31] The results in Table 2.6 show that the effects of classroom rank in the early grades increase substantially and significantly over time. Given these results, we now ask the following question: can we account for the increase in the effect of early classroom rank using estimates of $\beta_{t,0}$ (short-term impact of classroom rank), γ_t (impact of learning on future rank), and λ_t (impact of lagged skills on current skills)? Estimates of $\beta_{t,0}$, γ_t, and λ_t for each grade t show that $\gamma_t \approx 1$ but $\beta_t + \lambda_t < 1$ for every grade, which means that, if rank operated primarily through short-term learning gains and the resulting improvement in subsequent rank, we should observe fade-out in the impacts of rank on learning. As we have seen in Table 2.6, this is not the case. Rather, our results suggest that classroom rank operates at least in part by producing

[30] In Appendix 2.8.3 we report results from estimating rank effects by grade, rather than by aggregating estimates into "early", "middle", and "late" grades. In this appendix, we also show that all our results carry through if, instead of using math rank, we use achievement in math and language to calculate classroom rank.

[31] We do not know why the effects of rank in the early grades increase, while those in the middle grades do not. We note, however, that some theories of human capital argue that earlier investments tend to have the largest effects (as in Cunha and Heckman (2007)), in part because earlier investments allow children to better take advantage of later investments. It is possible that early classroom rank affects achievement in this way.

sustainable changes in another unobserved skill, which has independent effects on learning. A useful way to make this point is in Figure 2.3, which plots the implied change in learning in grades t+l, as a response to an exogenous change in early (1st or 2nd grade) achievement (t=0) percentile rank by 10 points, under two scenarios: (i) using the estimates of β_{t+l} (for l =0, 1...4) from equation (2.3.7), labeled reduced form in the figure; (ii) using the estimates of γ_t, β_t and λ_t, (for several values of t) from equations (2.3.1) and (2.3.8), and then simulating the response to a particular change in rank using the equations (2.3.9), labeled structural in the figure. The figure shows that, under scenario (i), the effect of rank grows over time, while in scenario (ii) it does not.[32] In sum, Figure 2.3, suggests that early classroom rank affects future achievement through channels that are not modeled explicitly in our equations, such as through its impact on other unobservable skills. We now turn directly to this question.

2.4.5 Executive Function, Non-cognitive skills, and Teacher Perceptions

As discussed above, there is a large literature in child psychology that argues that executive function is a key determinant of learning (Anderson (2002); Espy, McDiarmid, Cwik, Stalets, Hamby, and Senn (2004); Senn, Espy, and Kaufmann (2004)). In our data, too, EF in a given grade predicts future achievement.[33] It is therefore of interest to analyze whether classroom rank improves executive function. The results from regressions of EF (standardized to have mean zero and unit standard deviation) on classroom rank in math are in Table 2.7. The table shows that the coefficients on classroom math rank are positive for all EF dimensions, as well as for the aggregate measure of executive function. In the case of cognitive flexibility, the impacts are significant, and imply that an increase in classroom rank from the 50th to the 60th percentile of the distribution improves cognitive flexibility by 1.2 percent of a standard deviation.[34] Table 2.8, Panel A, reports the marginal effects from ordered probit re-

[32] For the purpose of illustration, we first normalize the estimate of $\beta_{t,0}$ to be the same across the two scenarios.

[33] We do not have the data to identify the causal effect of executive function on learning. Nevertheless, in a regression of achievement in grade t on executive function in grade $t-1$, including school fixed effects, the coefficient on EF is 0.538 (with a standard error of 0.005). If, in addition, we control for achievement at the end of $t-1$, the coefficient on EF is 0.096 (with a standard error of 0.002).

[34] We did not apply an inhibitory control test in 1st grade because, during the pilot, we found that virtually all 1st graders got a perfect (or close to perfect) score on the “Day-Night” test, but only a minority of children could carry out the inhibitory control test we applied in 2nd grade. In that test, children were shown words that correspond to a color, written in ink of a different color (for example, the word “green” written in red ink). They were then asked to say the name of the color of the ink, thus suppressing the natural reaction, which is to read the word written on the page. The test favors children

gressions of child happiness on math classroom rank in 1st grade. These results show that moving a child from the 50th to the 60th percentile of the distribution of classroom rank increases the probability that a child says she is always happy in school by 1 percentage point. Panel B reports the results from regressions of 6th grade non-cognitive skills on math classroom rank in 1st grade. The coefficients in all of the regressions are positive and one of them, corresponding to growth mindset, is significant, implying that moving a child from the 50th to the 60th percentile of 1st grade classroom rank improves her growth mindset by 2.8 percent of an SD. The table also shows that 1st grade classroom rank significantly improves the aggregate measure of non-cognitive skills. Finally, we turn to teacher perceptions. In our data, teacher perceptions of achievement are only imperfectly correlated with actual achievement: the correlation between being reported as one of the top 5 students in the classroom by a teacher, and actually being one of the top 5 students according to measured achievement is 0.42, while the corresponding correlation between teacher and objective classification of the bottom 5 students is 0.45.[35] In Table 2.9, we show that, in a regression that controls for achievement at the end of $t-1$, moving a child from the 50th to the 60th percentile of classroom rank at the start of grade t increases the probability that she is seen as a top student by her teacher in $t+1$ by 0.56 percentage points, but does not have a significant effect on the probability that she is seen as a bottom-achieving student. The fact that teachers perceive highly-ranked children to be particularly bright—even conditional on their actual ability—may reinforce academic self-concept of highly-ranked children, and thus contribute to the impact of rank on learning outcomes we observe in both the short- and medium-run. In sum, we show that children who, as a result of random assignment, have higher math classroom rank have higher levels of cognitive flexibility, are more likely to be happy, and have higher growth mindset scores. More highly-ranked children are also perceived to be smarter by their future teachers.

who cannot read, or can read only very imperfectly, which is why we did not apply it in 1st grade.

[35]The fact that these two measures are only imperfectly correlated could in part be a result of measurement error in either one of them. It is also possible, however, that in assessing "achievement" teachers are in fact taking account of a broader or somewhat different construct. Assessing the effects of classroom rank on teacher perceptions is therefore of interest. We also note that there is no evidence that teachers make more use of math or language achievement in assessing who are top and bottom students: The correlations between being in the top 5 by measured math achievement and language achievement, on the one hand, and having a teacher report a student as being in the top 5 are 0.38 and 0.33, respectively, while the correlations between being in the bottom 5 by measured math achievement and language achievement, on the one hand, and having a teacher report a student as being in the bottom 5 are 0.38 and 0.41, respectively.

2.5 Conclusion

This paper analyzes the impact of classroom ability rank measured at the start of the academic year on learning during that year, and learning in subsequent years. In our data, which comes from a longitudinal study of students in elementary schools in Ecuador, two students with the same underlying ability and attending the same school can have different classroom ranks because they are randomly assigned to different classrooms, with slightly different peers. We measure classroom rank and learning in math and language. Beginning-of-grade classroom rank and end-of-grade achievement are available for all grades from 1st to 6th grade. We also observe executive function in kindergarten through 4th grade, self-reported child happiness in 1st grade, and non-cognitive skills at the end of 6th grade. In addition, we have data on teacher perceptions of student ability in every grade between kindergarten and 6th grade. We show that classroom rank has modest short-term effects on achievement. Estimated effects of classroom rank can be confounded by the effects of peer quality. Students randomly assigned to classrooms with better peers will in general have lower classroom ability rank, but potentially benefit from better peers. Our estimates of the effect of classroom rank on learning are not affected when we control for average peer quality, or when we include classroom fixed effects. The converse is not true: we only observe peer effects in models that also control for classroom rank. The conflation of rank and peer quality effects is a feature of any study where both change at the same time, such as studies of the impact of selective schools, affirmative action, or neighborhood effects. We also show that classroom rank in math, but not language, affects achievement. The impact of classroom rank in math is larger for younger children and grows substantially over time. Moving a child from the 50th to the 60th percentiles of classroom rank in 1st (2nd) grade increases her achievement in 5th (6th) grade by 1 percentile of the national distribution. The increase in the magnitude of rank effects is remarkable given the evidence that impacts of many other interventions in elementary school fade out over time. Exogenous changes in classroom math rank also improve cognitive flexibility, happiness, growth mindset, and teacher perceptions of students. Changes in these skills, or others that we do not observe, are likely to be important in explaining how classroom rank raises child learning.

2.6 Tables

Table 2.1: Child, teacher, and classroom characteristics

	Mean	Standard deviation	N
Child and household characteristics			
Age of child (months)	60.3	4.9	13,858
Gender of child	0.49	0.5	14,477
Receptive vocabulary score (TVIP)	83.3	16.9	13,733
Mother's years of completed schooling	8.8	3.8	13,627
Father's years of completed schooling	8.5	3.8	10,594
Mother's age	30.2	6.6	13,637
Father's age	34.6	7.9	10,620
Proportion who attended preschool	0.61	0.49	14,472
Household has piped water in home	0.83	0.38	14,407
Household has flush toilet in home	0.46	0.5	14,407
Teacher and classroom characteristics			
Proportion female	0.82	0.38	2830
Proportion tenured	0.82	0.38	2818
Years of experience	18.1	10.5	2820
Class size	36.2	6.4	2838

Notes: Table reports means and standard deviations of the characteristics of children entering kindergarten in 2012, measured at the beginning of the school year, and of the teachers they had between kindergarten and 6th grade. The TVIP is the *Test de Vocabulario en Imágenes Peabody*, the Spanish version of the Peabody Picture Vocabulary Test (PPVT). The test is standardized using the tables provided by the test developers which set the mean at 100 and the standard deviation at 15 at each age.

Table 2.2: Effect of classroom rank and peer quality on achievement

	(1)	(2)	(3)	(4)
Classroom rank	0.018***		0.029***	0.035***
	(0.006)		(0.007)	0.007
Mean of classroom peers		0.013	0.046***	
		(0.014)	(0.015)	
School-by-grade fixed effects	X	X	X	
Classroom-by-grade fixed effects				X

Notes: The table reports estimates from regressions of achievement national rank on classroom rank and the leave-one-out mean of classroom peer achievement, pooling observations across grades. Column (1) shows our main model results. We regress national rank on classroom rank at the beginning of the school year, including a third-order polynomial in lagged national rank. Column (2) regresses national achievement rank on the leave-one-out mean achievement of classroom peers. Column (3) combines classroom rank and the leave-one-out mean of classroom peers. Columns (1)-(3) include school-by-grade fixed effects. Column (4) estimates the main model using classroom-by-grade fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout. Sample size is 87,706 observations in all columns. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 2.3: Classroom rank effects, separating math and language

	(1)	(2)	(3)	(4)
	Panel A: Effect of rank in math on math achievement			
Classroom rank	0.026***		0.036***	0.040***
	(0.007)		(0.009)	(0.009)
Mean of classroom peers		0.004	0.040**	
		(0.014)	(0.016)	
	Panel B: Effect of rank in math on language achievement			
Classroom rank	0.016		0.025**	0.030***
	(0.01)		(0.012)	(0.012)
Mean of classroom peers		0.011	0.036*	
		(0.019)	(0.021)	
	Panel C: Effect of rank in language on language achievement			
Classroom rank	-0.008		-0.006	0.003
	(0.008)		(0.008)	(0.009)
Mean of classroom peers		0.018	0.013	
		(0.016)	(0.018)	
	Panel D: Effect of rank in language on math achievement			
Classroom rank	0.001		0.005	0.017
	(0.01)		(0.012)	(0.012)
Mean of classroom peers		0.016	0.021	
		(0.019)	(0.022)	
School-by-grade fixed effects	X	X	X	
Classroom-by-grade fixed effects				X

Notes: The table reports estimates from regressions of achievement national rank on classroom rank and the leave-one-out mean of classroom peer achievement, pooling observations across grades. For the regressions in Panels A and B, classroom rank is calculated on the basis of math test scores, and a third-order polynomial in lagged math achievement is included as a control, while for the regressions in Panels C and D, classroom rank is calculated on the basis of language test scores, and a third-order polynomial in lagged language achievement is included as a control. Column (1) shows our main model results. Column (2) regresses national achievement rank on the leave-one-out mean achievement of classroom peers. Column (3) combines classroom rank and the leave-one-out mean of classroom peers. Columns (1)-(3) include school-by-grade fixed effects. Column (4) estimates the main model using classroom-by-grade fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout. N ranges from 87,713 (Panel A) to 87,789 (Panel C). *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 2.4: Robustness checks, effects of math classroom rank on math achievement

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Estimations with school-by-grade fixed effects						
Classroom rank	0.026***	0.042***	0.042***	0.026***	0.025***	0.032***	0.109***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)
School-by-grade fixed effects	X	X	X	X	X	X	X
Age and gender					X		
	Panel B: Estimations with classroom-by-grade fixed effects						
Classroom rank	0.040***	0.063***	0.063***	0.040***	0.039***	0.040***	0.107***
	(0.009)	(0.008)	(0.008)	(0.009)	(0.008)	(0.012)	(0.007)
Classroom-by-grade fixed effects	X	X	X	X	X	X	X
Age and gender					X		

Notes: The table reports estimates from regressions of national rank in math on classroom math rank. Observations are pooled across grades. All estimates include school-by-grade fixed effects. Column (1) reproduces the result from column (1) of Panel A in Table 4. In columns (2), (3), and (4), lagged achievement in math enters the regression as a linear term, quadratic term, or fourth-order polynomial (as opposed to a cubic term). Column (5) is comparable to column (1) but adds controls for child gender, age, and its square. In column (6) we use inverse probability weighting to correct for missing data. Column (7) corresponds to estimates of equation (2.3.4) in the main text. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 2.5: Heterogeneity of math classroom rank effects, by gender and ability

	(1) Gender	(2) Baseline vocabulary	(3) Lagged national rank
Classroom rank	0.025***	0.018*	-0.009
	(0.008)	(0.009)	(0.014)
Main covariate effect	-0.006***	0.020***	1.056***
	(0.002)	(0.001)	(0.06)
Interaction (rank*covariate)	0.001	0.007**	0.069***
	(0.004)	(0.003)	(0.023)

Notes: The table reports estimates from regressions of national rank in math on classroom math rank and interactions. Observations are pooled across grades. Column (1) shows the results from a regression of national rank on classroom rank interacted with an indicator variable for girls. Column (2) shows the results from a regression of national rank on classroom rank interacted with baseline vocabulary. Column (3) shows the results from a regression of national rank on classroom rank interacted with lagged national rank. All regressions are limited to schools in which there are at least two classrooms. All regressions include a third order polynomial in lagged national rank in math and school-by-grade fixed effects. Standard errors are clustered at the student level. N is 87,700 for column (1), 61,178 for column (2), and 87,713 for column (3). N is lower for column (2) because we collected vocabulary at the beginning of kindergarten, so it is not available for children who joined schools in our sample after the beginning of kindergarten. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 2.6: Effects of math classroom rank on achievement, by grade and lag

	Lags						
	0	1	2	3	4	F-test 1	F-test 2
Rank, “early” (1st & 2nd grades)	0.042***	0.060***	0.079***	0.091***	0.093***	0.038	0.004
	(0.014)	(0.016)	(0.016)	(0.017)	(0.018)		
Rank, “middle” (3rd & 4th grades)	0.040***	0.035**	0.027*			0.597	0.317
	(0.012)	(0.014)	(0.014)				
Rank, “late” (5th & 6th grades)	-0.005						
	(0.010)						
F-test 3	0.004	0.211	0.014				
F-test 4	0.008						

Notes: The table reports estimates from regressions of national rank in math on classroom math rank for different lags of classroom rank, separately for children in the “early” (1st and 2nd), “middle” (3rd and 4th), and “late” grades (5th and 6th) grades. All regressions are limited to schools in which there are at least two classes. All regressions include a third order polynomial in lagged national rank in math and school-by-grade fixed effects. Standard errors are clustered at the student level. F-tests are calculated after running estimates by Seemingly Unrelated Regressions. F-test 1 is a test that the coefficient for all ranks is the same, and F-test 2 is a test that the coefficient on lag=0 is the same as that on lag=4 (for the “early” grades) or lag=2 (for the “middle” grades). F-test 3 is a test that the coefficients on “early”, “middle”, and “late” grades are the same, and F-test 4 is a test that the “early” and “late” effects are the same. N varies by period and lag, from 21,012 for regressions of the effect of early rank on achievement after 4 lags, to 30,940 for regressions of the effect of late rank on achievement with no lags.
*Significant at 10%, **significant at 5%, ***significant at 1%.

Table 2.7: Math classroom rank effects on executive function

	(1)	(2)	(3)	(4)
	EF aggregate	Cognitive flexibility	Working memory	Inhibitory control
Classroom rank	0.055	0.124**	0.021	0.075
	(0.043)	(0.051)	(0.045)	(0.07)

Notes: The table reports estimates from regressions of executive function, in SDs, on classroom achievement rank in math, pooling observations across grades. All regressions include third order polynomials in lagged national rank in math, a third-order polynomial in lagged executive function, and school-by-grade fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. There are fewer observations for inhibitory control because we did not apply an inhibitory control test in 1st grade because, during the pilot, we found that virtually all 1st graders got a perfect (or close to perfect) score on the kindergarten test, but only a minority of children could carry out the inhibitory control test we applied in 2nd grade. In that test, children were shown words that correspond to a color, written in ink of a different color (for example, the word "green" written in red ink), and were then asked to say the name of the color of the ink, thus suppressing the natural reaction, which is to read the word written on the page. The test favors children who cannot read, or can read only very imperfectly, which is why we did not apply it in 1st grade. Standard errors are clustered at the student level throughout. N is 56,761 for columns (1) through (3), and 30,065 in column (4) because we did not collect data on inhibitory control in 1st grade, as discussed in the text. The EF aggregate in column (1) in 1st grade includes only cognitive flexibility and working memory, with both given equal weight. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 2.8: Math classroom rank effects on happiness and non-cognitive skills

		Panel A: Child happiness (1st grade)			
		Mostly happy	Almost always happy	Always happy	
Classroom rank		-0.055**	-0.042**	0.098**	
		(0.023)	(0.018)	(0.04)	
		Panel B: Non-cognitive skills (6th grade)			
	Non-cognitive aggregate	Depression	Self-esteem	Growth mindset	Grit
Classroom rank	0.255**	0.193	0.135	0.282**	0.188
	(0.128)	(0.128)	(0.129)	(0.128)	(0.132)

Notes: Panel A reports estimates from a regressions of child happiness in 1st grade on classroom achievement rank in math in 1st grade, estimated by ordered probit. Panel B reports estimates from regressions of a given 6th grade non-cognitive skill, or the non-cognitive aggregate of all four skills, on classroom achievement rank in math in 1st grade. All regressions include third order polynomials in kindergarten national rank in math and school fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout. N is 12,062 in Panel A and 7,789 in Panel B. *Significant at 10%, **significant at 5%, ***significant at 1%.

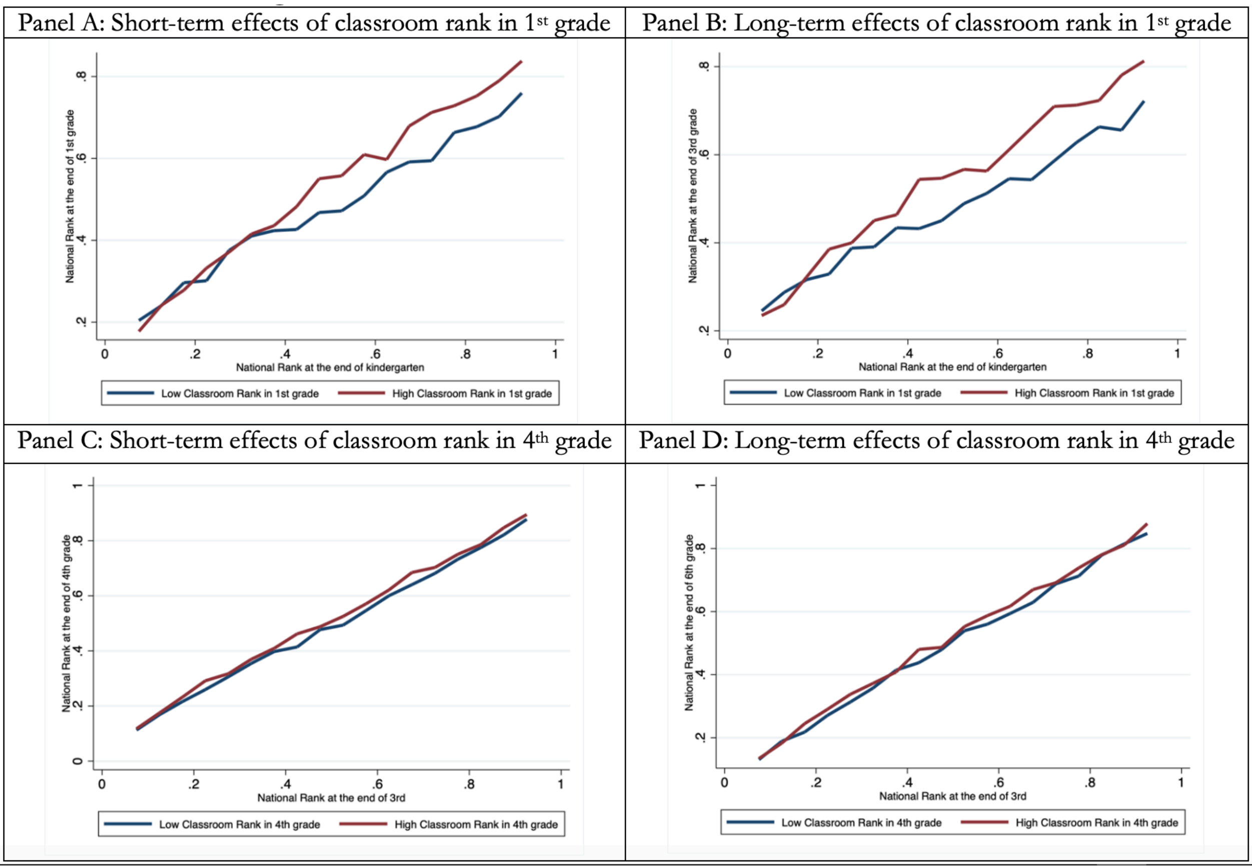
Table 2.9: Math classroom rank effects on teacher perceptions

	Top 5	Bottom 5
Classroom rank	0.056***	-0.022
	(0.017)	(0.016)

Notes: The table reports the results from regressions of a child being reported to be among the top 5 (bottom 5) by achievement by her teachers in grade t+1 on classroom rank in grade t, controlling for a third-order polynomial in national achievement in math in grade t-1, and school-by-grade fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout. N is 68,724. *Significant at 10%, **significant at 5%, ***significant at 1%.

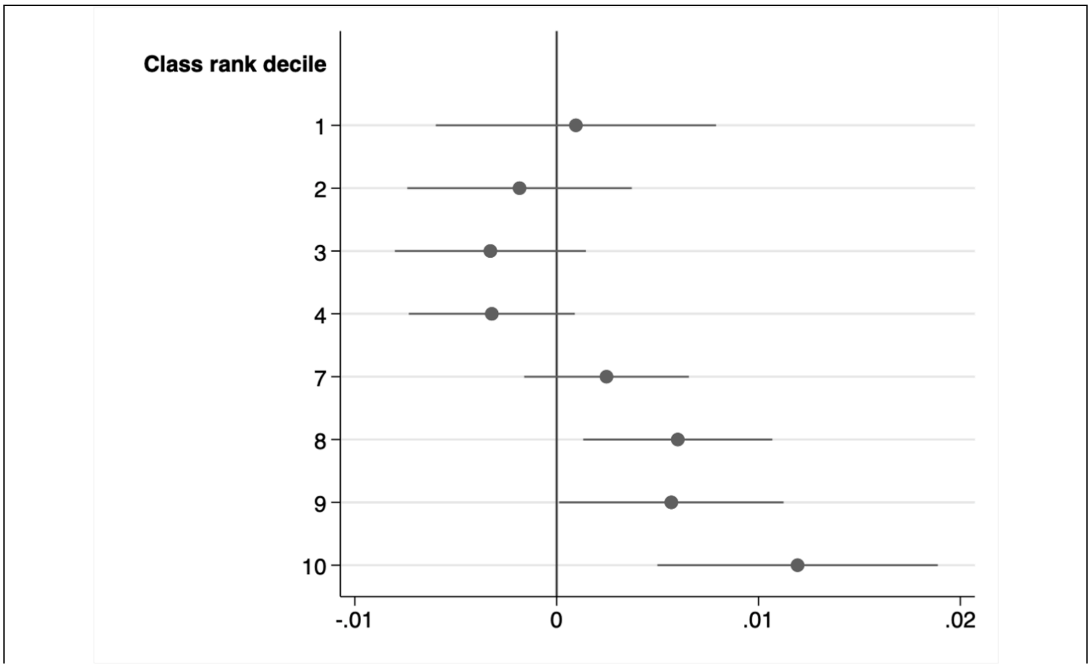
2.7 Figures

Figure 2.1: Visual evidence of classroom rank effects on achievement



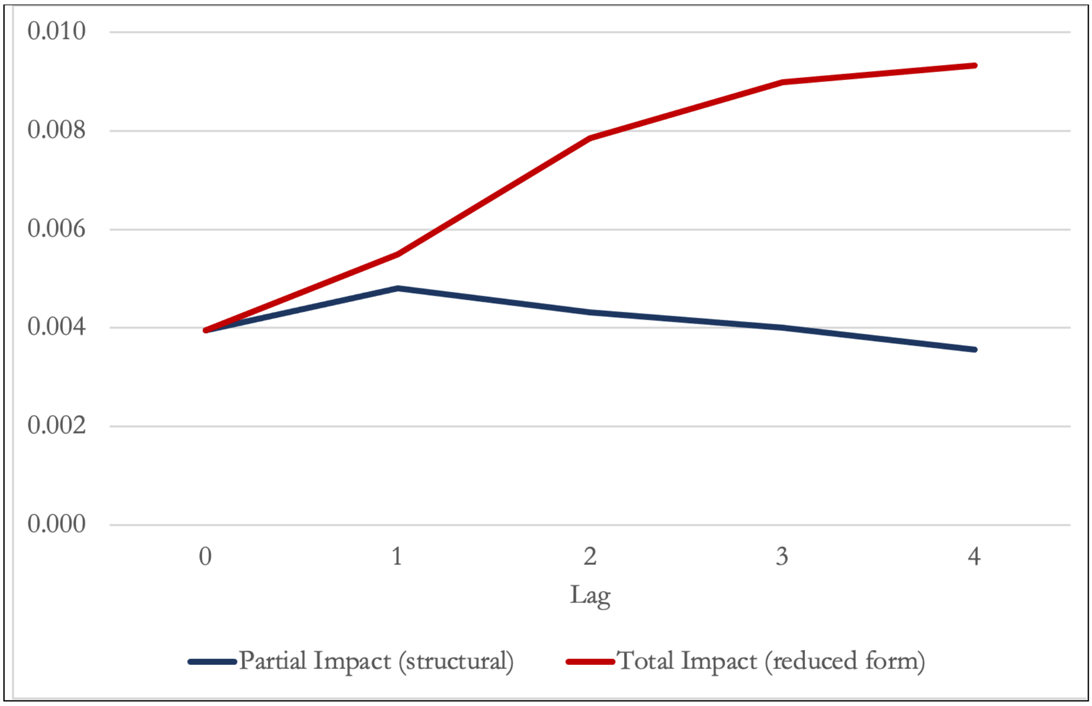
Notes: To generate this figure, we first sort children into ventiles on the basis of their test scores at the end of grade t-1. Then, for each ventile, we calculate average test scores at the end of grade t for two groups of children: those who, relative to other children in that ventile, were randomly assigned to classrooms where their rank at the beginning of t was "high"—classroom rank in the top 25 percent for that ventile—and those in classrooms where their rank was "low"—in the bottom 25 percent for that ventile. Panel A focuses on the short-term effects of classroom rank in 1st grade, and Panel B compares these two groups of children at the end of 3rd grade. Panel C focuses on the short-term effects of classroom rank in 4th grade, and Panel D focuses on these same children at the end of 6th grade.

Figure 2.2: Classroom rank effects at different deciles of the distribution of achievement



Notes: To produce this figure we discretize classroom rank in math into 10 deciles, and run regressions of math achievement on classroom rank deciles. We graph coefficients and 90 percent confidence intervals on deciles 1 through 4, and 7 through 10, with deciles 5 and 6 as the omitted category. All regressions include a third order polynomial in lagged national rank and school-by-grade fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout.

Figure 2.3: Total and partial effects of rank on achievement



Notes: The figure plots the implied change in learning in grades t+l, as a response to an exogenous change in early (1st or 2nd grade) achievement (t=0) percentile rank by 10 points, under two scenarios: (i) using the estimates of β_{t+l} (for l=0, 1...4) from equation 2.3.7 in the main text ("total impact"); and (ii) using the estimates of γ_t, β_t and λ_t, (for several values of t) from equations 2.3.1 and 2.3.8 in the main text, and then simulating the response to a particular change in rank using the equations 2.3.9 ("partial impact"). We normalize the estimate of $\beta_{t,0}$ to be the same across the two scenarios.

2.8 Appendix

2.8.1 Tests of random assignment

An important assumption underlying our empirical strategy is that children's classroom rank at the beginning of a given grade is random, due to random assignment of children to classrooms within schools in every year.[36] Random assignment is closely monitored, and compliance is very high, 98.9 percent on average. In this appendix, we present tests of random assignment using a methodology developed in Jochmans (2020). First, we briefly discuss the procedure outlined in Jochmans (2020). Consider our setting, in which we observe data on S schools, and each school has $n_1, \ldots, n_s$ students. Within each school, children are assigned to a classroom—and therefore their peer group—every year. Let $x_{s,i}$ be an observable characteristic of child i in school s. If assignment to peer groups is random, $x_{s,i}$ will be uncorrelated with $x_{s,j}$, for all j belonging to the set of i's classroom peers. Let $\bar{x}_{s,i}$ be the average value of characteristic x among student i's peers. The procedure tests whether the correlation in a within-school regression of $x_{s,i}$ on $\bar{x}_{s,i}$ is statistically significantly different from zero (a methodology first proposed in Sacerdote (2001)), introducing a bias correction for the inclusion of group fixed effects (in our case, schools). It is important to control for school fixed effects, as randomization happens within schools, but there may be selection into a school based on individual characteristics. Jochmans (2020) shows that a fixed-effects regression of $x_{s,i}$ on $\bar{x}_{s,i}$ will yield biased estimates due to inconsistency of the within-group estimator. The proposed corrected estimator is given by

$$ts = \frac{\sum_{s=1}^{S} \sum_{i=1}^{n_s} \tilde{x}_{s,i}(\bar{x}_{s,i} + \frac{x_{s,i}}{n_s - 1})}{\sqrt{\sum_{s=1}^{S} (\sum_{i=1}^{n_s} \tilde{x}_{s,i}(\bar{x}_{s,i} + \frac{x_{s,i}}{n_s - 1}))^2}} \tag{2.8.1}$$

where $\tilde{x}_{s,i}$ is the deviation of $x_{s,i}$ from its within-school mean. The null hypothesis is thus absence of correlation between i's characteristics and those of her peers. To test the random assignment in our setting, we implement this procedure by testing for the presence of correlation between child i's scores measured at the end of grade $t-1$ and the average end-of-grade scores in $t-1$ of the classroom peers assigned to her

[36] We use the word "random" as shorthand but, as discussed at length in Araujo, Carneiro, Cruz-Aguayo, and Schady (2016), strictly speaking random assignment only occurred in 3rd through 6th grade. In the other grades, the assignment rules were as-good-as-random. Specifically, the assignment rules we implemented were as follows: In kindergarten, all children in each school were ordered by their last name and first name, and were then assigned to teachers in alternating order; in 1st grade, they were ordered by their date of birth, from oldest to youngest, and were then assigned to teachers in alternating order; in 2nd grade, they were divided by gender, ordered by their first name and last name, and then assigned in alternating order; in 3rd through 6th grades, they were divided by gender and then randomly assigned to one or another classroom.

in a given grade t. We do so for each grade. We implement the test for all children in the sample, and restricting the sample to those children who have both end of grade $t-1$ scores as well as end of grade t scores (as these will be the children that end up being included in the estimation of our models). The results are shown in tables 2.10 and 2.11, respectively. Our results show that we cannot reject the null hypothesis that there is no correlation between child i's achievement and that of her classroom peers. This result is true for all grades and both samples. Hence, we conclude that random assignment was successful in our setting.

Table 2.10: Testing for random assignment of children to classrooms, full sample

	Kindergarten	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Test statistic	1.359	-0.383	0.905	0.3	-0.445	-0.222	0.98
P-value	0.174	0.702	0.366	0.764	0.657	0.825	0.327

Notes: In this table, we report results for tests of random assignment of children to classrooms within schools using a methodology proposed by Jochmans (2020). The null hypothesis is absence of correlation between a child's ability measured at the end of the previous grade and the average ability of classroom peers assigned to her at the beginning of a given grade, conditional on school. The sample includes all children.

Table 2.11: Testing for random assignment of children to classrooms, restricted sample

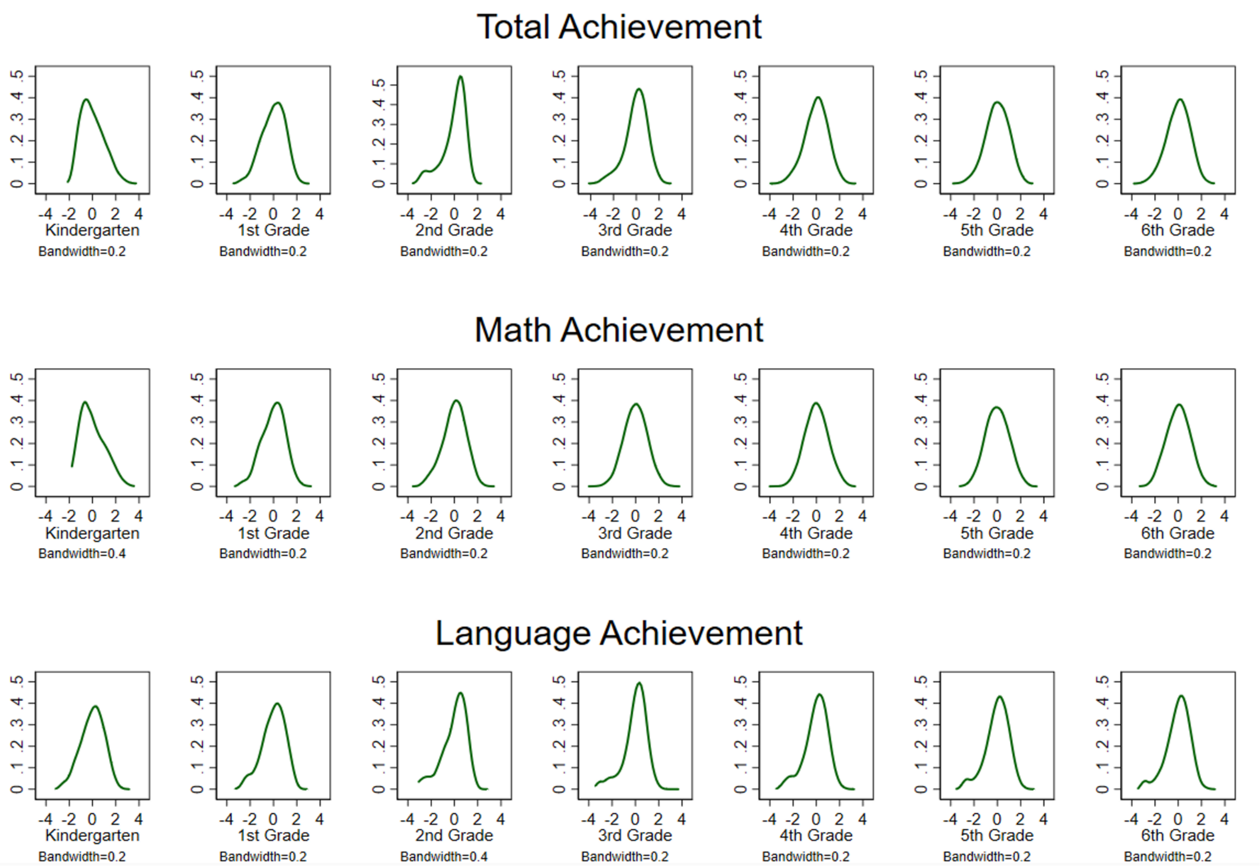
	Kindergarten	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Test statistic	1.392	-0.005	1.425	0.413	-0.043	0.001	1.037
P-value	0.164	0.996	0.154	0.68	0.966	0.999	0.3

Notes: In this table, we report results for tests of random assignment of children to classrooms within schools using a methodology proposed by Jochmans (2020). The null hypothesis is absence of correlation between a child's ability measured at the end of the previous grade and the average ability of classroom peers assigned to her at the beginning of a given grade, conditional on school. The sample is restricted to children who have available both beginning- and end-of-grade scores for a given grade.

2.8.2 Additional information on outcome variables

This appendix presents additional information on test scores, executive function, and non-cognitive skills. Figure 2.4 presents the univariate densities of our achievement measures, separately by grade. The figure shows that most of the distributions appear to have a reasonable spread and are generally symmetric. One clear exception is math achievement in kindergarten, which is left-censored. Figure 2.5 presents comparable densities for executive function. It shows that the distributions of inhibitory control and cognitive flexibility are often highly skewed. This is not surprising given the nature of the tests. As an example, we describe the executive function tests we applied in kindergarten. In the inhibitory control test, kindergarten children were quickly shown a series of 14 flash cards that had either a sun or a moon and were asked to say the word “day” when they saw the moon and “night” when they saw the sun. Just over half (50.8 percent) of all children made no mistake on this test, so there is a concentration of mass at the highest value, while very few children (1.6 percent) answered all prompts incorrectly. The cognitive flexibility test we applied in kindergarten worked as follows. Children were handed a series of picture cards, one by one. Cards had either a truck or a star, in red or blue. The enumerator asked the child to sort cards by color, or by shape. Specifically, in the first half of the test, the enumerator asked the child to play the “colors” game, handed her cards, indicating their color, and asked the child to place them in the correct pile (“this is a red card: where does it go?”). After 10 cards, the enumerator told the child that they would switch to the “shapes” game, and reminded the child that, in this game, trucks should be placed in one pile and stars in another. The enumerator then handed the child cards, indicating the shapes on the card, and asked her to place them in the correct pile (“this is a star: where does it go?”). In both the first and the second part of the test, if the child made three consecutive mistakes, the enumerator paused the test, reminded her what game they were playing (“remember we are playing the shapes game; in the shapes game, all trucks go in this pile, and all stars in this other pile”), and handed the child a new card with the corresponding instruction. A small proportion of children in kindergarten (7.5 percent) did not understand the game, despite repeated examples, and were given a score of 0; just under half of all children (47 percent) answered all prompts correctly in both the “colors” and “shapes” parts of the test; and just over a quarter (27.3 percent) of all children made no mistakes in the first part of the test (the “colors” game), but incorrectly classified every card in the second part of the test (the “shapes” game). These children were unable to switch rules, despite repeated promptings from the enumerator. The distribution of scores for this test therefore has a concentration of mass at two points, with much less mass at other points. The working memory test had two parts. In the first part, children were given 2 minutes to find as many sequences of dog, house, and ball, in that order, on a sheet that has rows of dogs, houses, and balls in various possible sequences. The score on this part of the test is the number of correct sequences found by the child. In the second part of the test, the enumerator recited strings of numbers, and asked the child to repeat them, in the same order or backwards. Figure 2.5 shows that the aggregate working memory score is distributed smoothly, with little evidence of a concentration of mass at particular values. In practice the correlations of the scores across the three dimensions in our sample are low—in the range of 0.21 to 0.32 between cognitive flexibility and working memory, between 0.17 and 0.33 between working memory and inhibitory control, and in the range of 0.12 to 0.15 between cognitive flexibility and inhibitory control—see Appendix Table 2.12.[37] When the scores across the three dimensions are averaged, the distributions of the total executive function score are generally smooth and symmetric. Figure 2.6, finally, shows univariate densities of the four non-cognitive measures we applied in 6th grade. The figure shows that the distribution of the depression and grit scores appear to be right-censored. The distribution for the aggregate measure of non-cognitive outcomes, on the other hand, is smooth and symmetric. Table 2.13 shows that the different non-cognitive outcomes are positively correlated, although the correlations are far from unity—they range from 0.20 (between depression and grit) to 0.49 (between growth mindset and self-esteem).

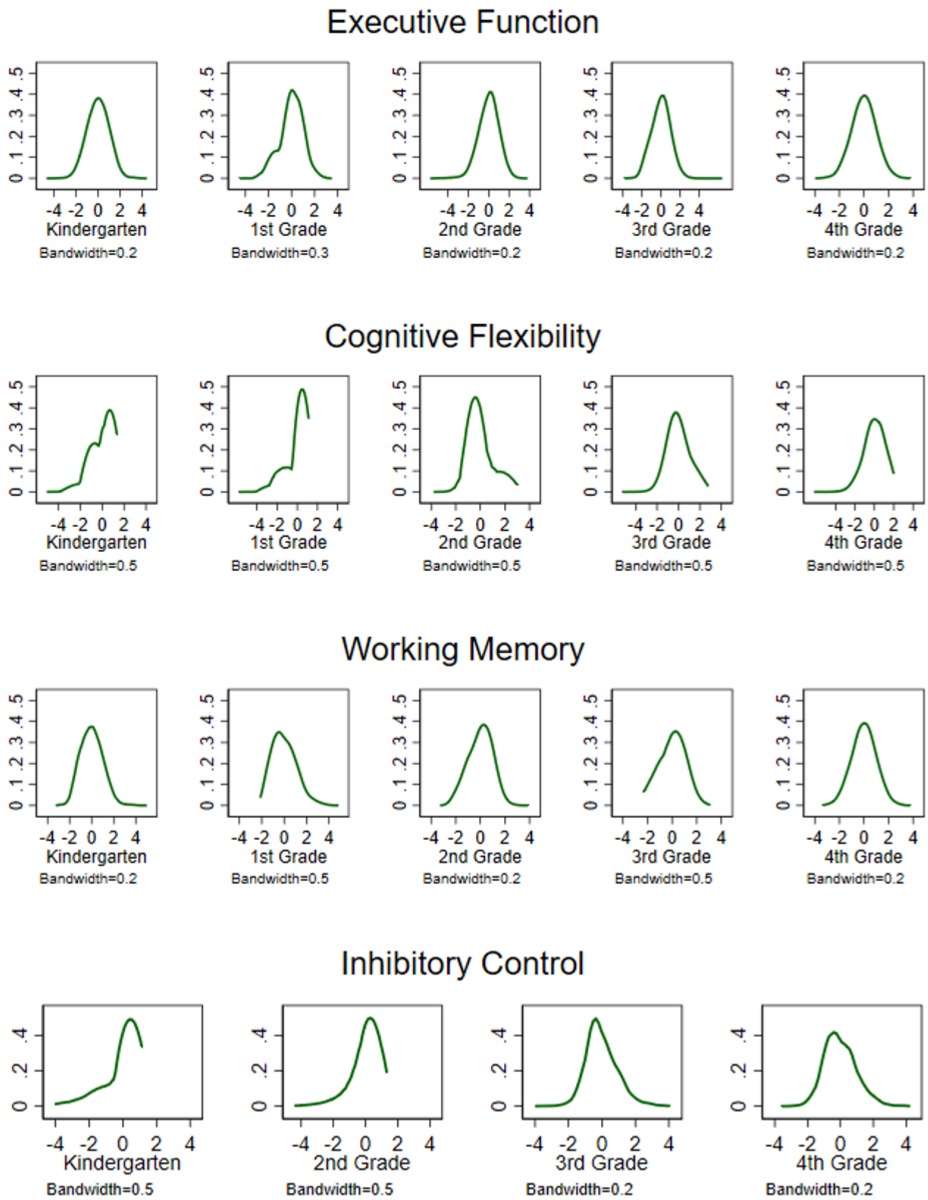
Figure 2.4: Distributions of achievement, by grade



Notes: The figure shows univariate densities of achievement, in z-scores, by grade.

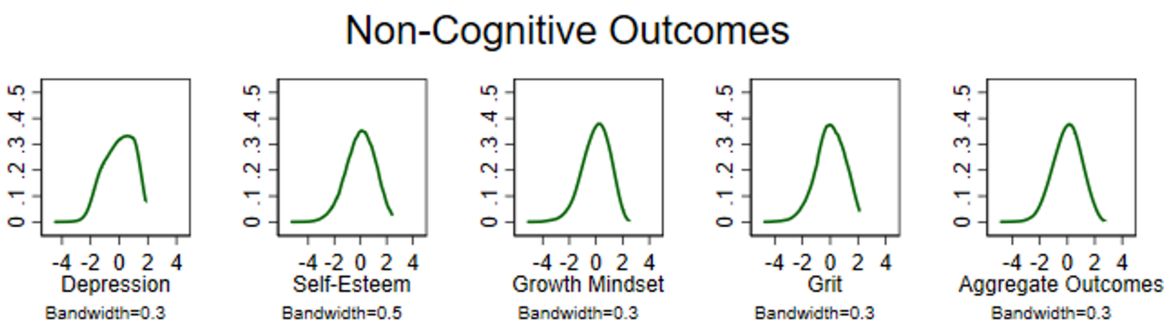
[37] The fact that these correlations are very low is likely to be a result of both measurement error and differences across the constructs that each domain measures.

Figure 2.5: Distributions of executive function, by grade



Notes: The figure shows univariate densities of executive function, in z-scores, by grade.

Figure 2.6: Distributions of non-cognitive outcomes, by grade



Notes: The figure shows univariate densities of non-cognitive outcomes, in z-scores, by grade.

Table 2.12: Correlations across dimensions in executive function

	Inhibitory Control	Cognitive Flexibility
	Kindergarten	
Cognitive Flexibility	0.13	
Working Memory	0.22	0.29
	1st Grade	
Working Memory		0.23
	2nd Grade	
Cognitive Flexibility	0.15	
Working Memory	0.25	0.24
	3rd Grade	
Cognitive Flexibility	0.12	
Working Memory	0.17	0.21
	4th Grade	
Cognitive Flexibility	0.15	
Working Memory	0.33	0.32
	Pooled	
Cognitive Flexibility	0.14	
Working Memory	0.24	0.26

Notes: The table reports the pairwise correlations between executive function dimensions. All the correlations are significant at the 1 percent level.

Table 2.13: Correlations across non-cognitive outcomes

	Depression	Self- Esteem	Growth Mindset
Self- Esteem	0.24		
Growth Mindset	0.26	0.49	
Grit	0.2	0.45	0.38

Notes: Table presents the results from pairwise correlations between non-cognitive outcomes collected in 6th grade. All the correlations are significant at the 1 percent level.

2.8.3 Additional estimates of classroom rank effects

In this Appendix, we report classroom rank effects by grade (for grades 1, 2 . . . 6). We also report the results of estimating Tables 2.4 through 2.9, and Figures 2.2 and 2.3, when rank is calculated on the basis of achievement in math and language, rather than achievement in math only (as in the results in the main body of the paper).

Grade-specific estimates of effects of classroom rank Table 2.14 presents estimates of math classroom rank effects, by grade (rather than when grades are aggregated into “early”, “middle” and “late” periods. These results are consistent with those in the first column of Table 2.6 in the main body of the paper, although they are noisier.

Table 2.14: Grade-specific estimates of effects of classroom rank

	Grade					
	1	2	3	4	5	6
Classroom rank	0.058***	0.026	0.011	0.069***	0.011	-0.02
	(0.022)	(0.019)	(0.018)	(0.016)	(0.015)	(0.015)
N	12,161	14,534	14,823	15,255	15,350	15,590

Notes: The table reports estimates from regressions of national rank in math on classroom math rank, separately by grade. All regressions are limited to schools in which there are at least two classes. All regressions include a third order polynomial in lagged national rank in math and school-by-grade fixed effects. Standard errors are clustered at the student level. *Significant at 10%, **significant at 5%, ***significant at 1%.

B. Estimates of rank effects when rank is calculated on the basis of total achievement (rather than math achievement) In the main body of the paper, Figures 2.2 and 2.3, and Tables 2.4 through 2.9, refer to the effects of classroom rank when rank is calculated on the basis of math achievement, as this follows logically from Table 2.3. In this appendix, we show that results are very similar if, instead, we use overall achievement (on the basis of test scores in math and language) to calculate rank effects.

a. Figure 2.2 in the paper is very similar to Figure 2.7 in the appendix, as are Figure 2.3 in the paper and Figure 2.8 in the appendix.

b. Table 2.4 in the paper and Table 2.15 in the appendix show very similar patterns: Our main result on the effect of classroom rank on achievement is robust to various checks, regardless whether we use overall achievement or only math achievement to calculate rank.

c. Tables 2.5 in the paper and 2.16 in the appendix refer to the analysis of heterogeneity of rank effects. The only important difference is that the main effect of gender is negative and significant in Table 2.5, but positive and (borderline) significant in Table 2.16. The reason for this is that girls have lower math achievement, but higher language achievement, than boys. In any case, the coefficient of interest—the interaction between rank and gender—is very small and insignificant in both cases.

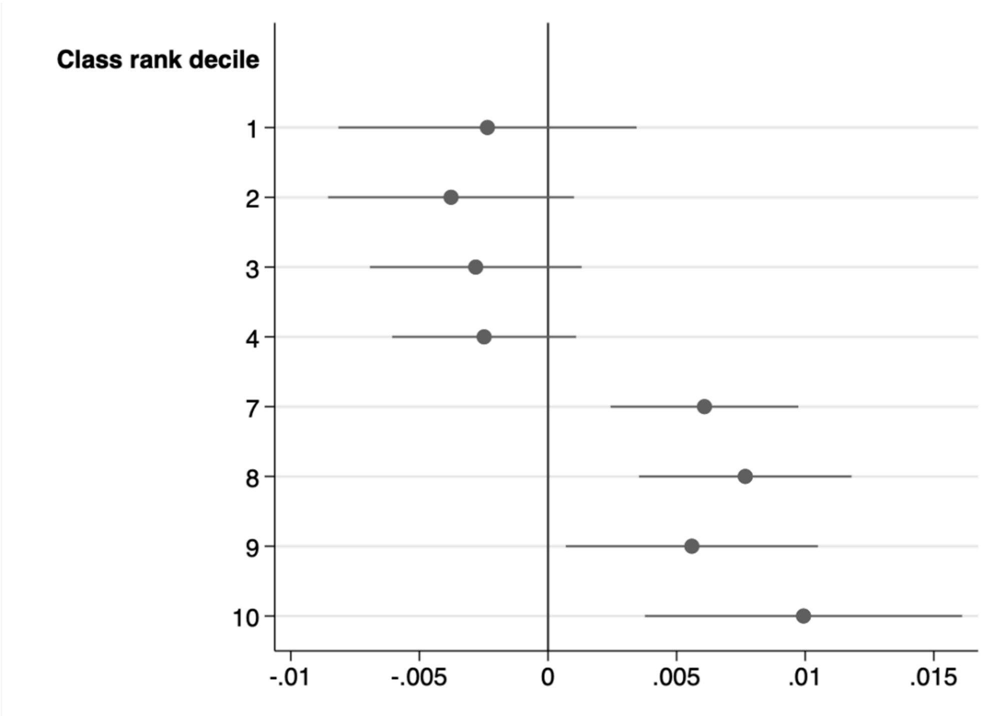
d. Table 2.6 in the paper and Table 2.17 in the appendix are, once again, very similar. Regardless of whether we calculate rank on the basis of overall achievement (as in the appendix) or math achievement (as in the main body of the paper), the effect of "early" rank on achievement increases substantially and significantly over time.

e. In both Tables 2.7 in the paper and 2.18 in the appendix, the coefficients on rank are positive, and significant (or borderline significant) for the measure of cognitive flexibility. In Table 2.18, but not in Table 2.7, we find that classroom rank also has a significant effect on inhibitory control.

f. In both Tables 2.8 in the paper and 2.19 in the appendix, more highly-ranked children report they are happier, and in both cases classroom rank improves growth mindset. In Table 2.8, classroom rank has a positive and significant effect on the aggregate measure of non-cognitive skills, whereas in Table 2.19 this impact is also positive but is not significant.

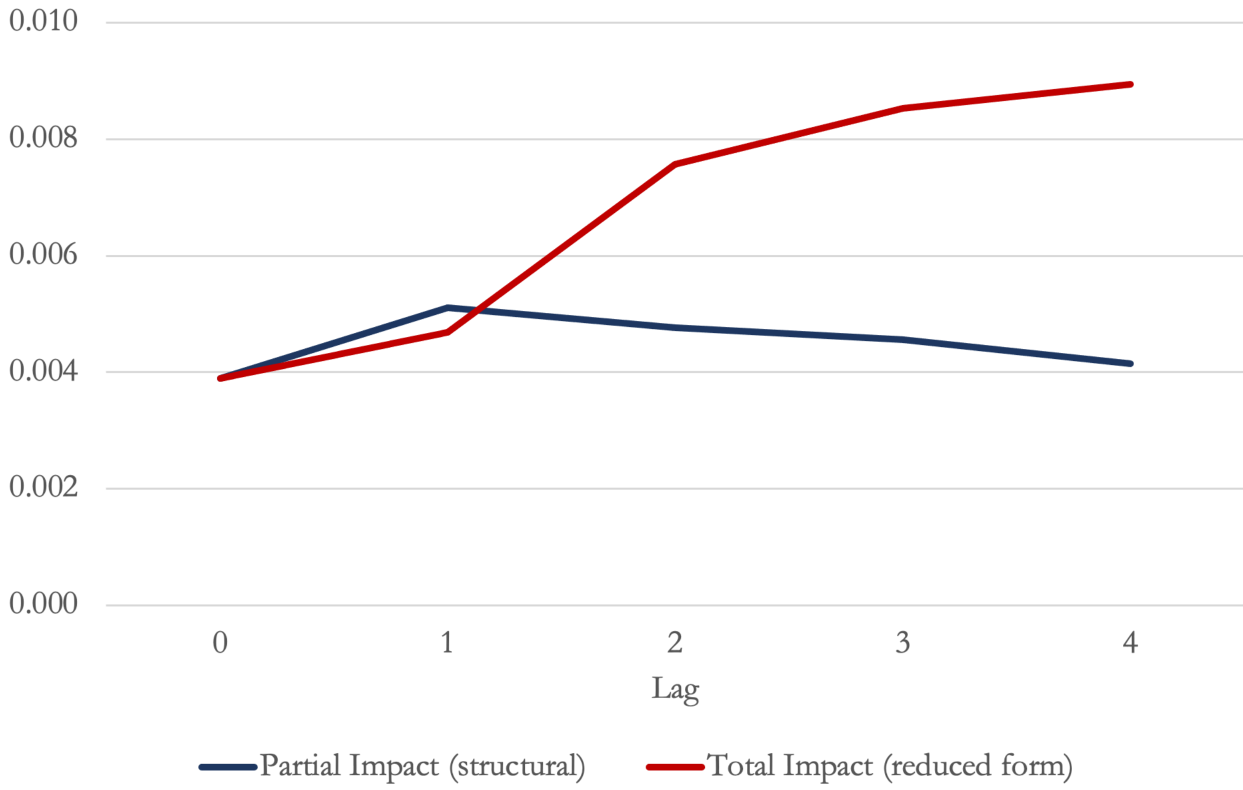
g. Tables 2.9 in the paper and 2.20 in the appendix both show that more highly ranked children are seen to be among the top 5 children by their future teachers.

Figure 2.7: Classroom rank effects at different deciles of the distribution of achievement, with rank and achievement calculated on the basis of test scores in math and language



Notes: To produce this figure we discretize classroom rank in total achievement into 10 deciles, and run regressions of achievement on classroom rank deciles. We graph coefficients and 90 percent confidence intervals on deciles 1 through 4, and 7 through 10, with deciles 5 and 6 as the omitted category. All regressions include a third order polynomial in lagged national rank and school-by-grade fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout.

Figure 2.8: Total and partial effects of rank on achievement, with rank and achievement calculated on the basis of test scores in math and language



Notes: The figure plots the implied change in learning in grades t+l, as a response to an exogenous change in early (1st or 2nd grade) achievement (t=0) percentile rank by 10 points, under two scenarios: (i) using the estimates of β_{t+l} (for l=0, 1...4) from equation (2.3.7) in the main text ("total impact"); and (ii) using the estimates of γ_t, β_t and λ_t, (for several values of t) from equations (2.3.1) and (2.3.8) in the main text, and then simulating the response to a particular change in rank using the equations (2.3.9) ("partial impact"). We normalize the estimate of $\beta_{t,0}$ to be the same across the two scenarios.

Table 2.15: Robustness checks, effects of achievement classroom rank on achievement, with rank and achievement calculated on the basis of test scores in math and language

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Classroom rank	0.018***	0.036***	0.037***	0.018***	0.017***	0.019***	0.090***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
R-squared	0.772	0.771	0.771	0.772	0.774	0.776	0.773
School-by-grade fixed effects	X	X	X	X	X	X	X
Age and gender					X		

Notes: The table reports estimates from regressions of national rank on classroom rank. Observations are pooled across grades. Column (1) reproduces the result from column (1) of Table 2.2. In columns (2), (3), and (4), lagged achievement enters the regression as a linear term, quadratic term, or fourth-order polynomial (as opposed to cubic term). Column (5) is comparable to column (1) but adds controls for child gender, age, and its square. In column (6) we use inverse probability weighting to correct for missing data. Column (7) corresponds to estimates of equation (2.3.4) in the main text. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 2.16: Heterogeneity of classroom rank effects, by gender and ability, , with rank and achievement calculated on the basis of test scores in math and language

	(1) Gender	(2) Baseline vocabulary	(3) Lagged national rank
Classroom rank	0.019***	0.018**	-0.001
	(0.007)	(0.008)	(0.012)
Main covariate effect	0.003*	0.013***	1.290***
	(0.002)	(0.001)	(0.056)
Interaction (rank*covariate)	-0.002	0.014***	0.040**
	(0.003)	(0.003)	(0.02)
N	87,693	61,177	87,706

Notes: The table reports estimates from regressions of national rank on classroom rank and interactions. Observations are pooled across grades. Column (1) shows the results from a regression of national rank on classroom rank interacted with an indicator variable for girls. Column (2) shows the results from a regression of national rank on classroom rank interacted with baseline vocabulary. Column (3) shows the results from a regression of national rank on classroom rank interacted with lagged national rank. All regressions are limited to schools in which there are at least two classrooms. All regressions include a third order polynomial in lagged national rank and school-by-grade fixed effects. Standard errors are clustered at the student level. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 2.17: Effects of classroom rank on achievement, by grade and lag, with rank and achievement calculated on the basis of test scores in math and language

	Lags						
	0	1	2	3	4	F-test 1	F-test 2
Rank, “early” (1st & 2nd grades)	0.040***	0.050***	0.078***	0.086***	0.089***	0.011	0.002
	(0.013)	(0.014)	(0.015)	(0.016)	(0.017)		
Rank, “middle” (3rd & 4th grades)	0.027***	0.039***	0.040***			0.389	0.233
	(0.010)	(0.012)	(0.012)				
Rank, “late” (5th & 6th grades)	-0.012						
	(0.009)						
F-test 3	0.001	0.529	0.043				
F-test 4	0.001						

Notes: The table reports estimates from regressions of national rank on classroom rank for different lags of classroom rank, separately for children in the “early” (1st and 2nd), “middle” (3rd and 4th), and “late” grades (5th and 6th) grades. All regressions are limited to schools in which there are at least two classes. All regressions include a third order polynomial in lagged national rank. Standard errors are clustered at the student level. F-test 1 is a test that the coefficient for all ranks is the same, and F-test 2 is a test that the coefficient on lag=0 is the same as that on lag=4 (for the “early” grades) or lag=2 (for the “middle” grades). F-test 3 is a test that the coefficients on “early”, “middle”, and “late” grades are the same, and F-test 4 is a test that the “early” and “late” effects are the same. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 2.18: Classroom rank effects on executive function, with rank and achievement calculated on the basis of test scores in math and language

	EF aggregate	Cognitive flexibility	Working memory	Inhibitory control
Classroom rank	0.059	0.097*	0.01	0.141**
	(0.042)	(0.05)	(0.044)	(0.067)
N	56,759	56,759	56,759	30,064

Notes: The table reports estimates from regressions of executive function, in SDs, on classroom achievement rank, pooling observations across grades. All regressions include third order polynomials in lagged national rank, a third-order polynomial in lagged executive function, and school-by-grade fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. There are fewer observations for inhibitory control because we did not apply an inhibitory control test in 1st grade because, during the pilot, we found that virtually all 1st graders got a perfect (or close to perfect) score on the kindergarten test, but only a minority of children could carry out the inhibitory control test we applied in 2nd grade. In that test, children were shown words that correspond to a color, written in ink of a different color (for example, the word "green" written in red ink), and were then asked to say the name of the color of the ink, thus suppressing the natural reaction, which is to read the word written on the page. The test favors children who cannot read, or can read only very imperfectly, which is why we did not apply it in 1st grade. Standard errors are clustered at the student level throughout. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 2.19: Classroom rank effects on happiness and non-cognitive skills, with rank calculated on the basis of test scores in math and language

		Panel A: Child happiness (1st grade)			
		Mostly happy	Almost always happy	Always happy	
Classroom rank		-0.052**	-0.040**	0.091**	
		(0.023)	(0.018)	(0.041)	
N		12,062	12,062	12,062	
	Panel B: Non-cognitive skills (6th grade)				
	Non-cognitive aggregate	Depression	Self-esteem	Growth mindset	Grit
Classroom rank	0.164	0.076	0.113	0.275**	-0.027
	(0.131)	(0.128)	(0.135)	(0.129)	(0.135)
N	7,789	7,789	7,789	7,789	7,789

Panel A reports estimates from a regressions of child happiness in 1st grade on classroom achievement rank in 1st grade, estimated by ordered probit. Panel B reports estimates from regressions of a given 6th grade non-cognitive skill, or the non-cognitive aggregate of all four skills, on classroom achievement rank in 1st grade. All regressions include a third order polynomial in kindergarten national rank and school fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout. *Significant at 10%, **significant at 5%, ***significant at 1%.

Table 2.20: Classroom rank effects on teacher perceptions, with rank calculated on the basis of test scores in math and language

	Top 5	Bottom 5
Classroom rank	0.046***	-0.032**
	(0.017)	(0.016)
N	68,720	68,720

The table reports the results from regressions of a child being reported to be among the top 5 (bottom 5) by achievement by her teachers in grade t+1 on classroom rank in grade t, controlling for a third-order polynomial in national achievement in grade t-1, and school fixed effects. All regressions are limited to schools in which there are at least two classrooms per grade. Standard errors are clustered at the student level throughout. *Significant at 10%, **significant at 5%, ***significant at 1%.

Chapter 3

Human Capital Growth and Poverty: Evidence from Ethiopia and Peru

3.1 Introduction

There is an increasing consensus that events and experiences in the early years of childhood, from conception to at least the age of three, have long lasting consequences for an individual's development and productivity.[1] There is also agreement on the fact that human capital is a multidimensional object, with its various constituents (health, cognition, language, socio-emotional skills) interacting over time along the process of its development, giving rise to complex dynamics and complementarities (see Cunha and Heckman (2008), Cunha, Heckman, and Schennach (2010) and Aizer and Cunha (2012)). These dynamic interactions, together with the possible malleability of skills at certain points in the life cycle, give rise to potential 'windows of opportunity' for targeted interventions that can improve the development of vulnerable children.[2]

These issues are particularly relevant in developing countries, where children living in poverty are exposed to a variety of risks, including disease, malnutrition, vi-

[1] For example, Campbell, Conti, Heckman, Moon, Pinto, Pungello, and Pan (2014) documents long run impacts of the Carolina Abecedarian Project on health and Chetty et al. (2014b) documents long run outcomes of Project STAR on earnings. For a review of a number of studies documenting the long run outcomes of other early interventions in the U.S., see Currie and Almond (2011). For examples from a developing country, see Gertler, Heckman, Pinto, Zanolini, Vermeersch, Walker, Chang, and Grantham-McGregor (2014) which documents long run impacts on earnings of an early life health intervention in Jamaica and Walker, Chang, Powell, and Grantham-McGregor (2005) which documents long run impacts of the same intervention in Jamaica on cognitive outcomes.

[2] For just a few examples from a rich literature documenting the malleability of skills in early childhood, see Grantham-McGregor, Cheung, Cueto, Glewwe, Richter, and Strupp (2007), Bharadwaj, Løken, and Neilson (2013), Engle, Black, Behrman, Cabral de Mello, Gertler, Kapiriri, Martorell, and Young (2007), Olds, Henderson Jr, Kitzman, Eckenrode, Cole, and Tatelbaum (1999), Heckman and Kautz (2013), and Hoddinott, Maluccio, Behrman, Flores, and Martorell (2008).

olence and unstimulating environments. Indeed, the developmental gaps between children living in more versus less affluent families have been amply documented (see Fernald, Weber, Galasso, and Ratsifandrihamanana (2011), Carneiro and Heckman (2003), Currie (2008), Rubio-Codina, Attanasio, Meghir, Varela, and Grantham-McGregor (2015), Hart and Risley (1995), and Fernald, Marchman, and Weisleder (2013)).

One way of understanding how these factors interact during childhood to produce life-long outcomes is to estimate production functions for the various dimensions of human capital. These production functions can map the interaction of family background and the current skill level of the individual child, along with investments in the child at each age into child development and growth. This approach is useful because it allows us to identify the degree of persistence of different inputs into development and their influence on subsequent growth. This characterization in turn can be used to identify 'windows of opportunity': periods in the life cycle of the child where investments (including policy interventions) might be particularly fruitful.

In this paper, we use high quality data from Ethiopia and Peru drawn from the Young Lives Survey to implement this approach, by estimating flexible specifications of the production functions for health and cognition, two key components of human capital. For each of these countries, we have observations for two different cohorts spanning most of childhood. In particular, the younger cohort is observed at ages 1, 5 and 8, while the older cohort is observed at ages 8, 12 and 15. Building on earlier work (see, for example, Cunha and Heckman (2008), Cunha, Heckman, and Schennach (2010), Attanasio, Meghir, and Nix (2020), and Attanasio, Cattan, Fitzsimons, Meghir, and Rubio-Codina (2020)) we follow a factor analytic approach to estimate investment equations and production functions for human capital from age 1 to 15, allowing these to be dynamically connected. Our model can be viewed as an approximation to a dynamic model of household choice and investments in children with liquidity constraints. Del Boca, Flinn, and Wiswall (2014) show how such a dynamic model can be specified and estimated structurally.

Our paper offers innovations in a number of dimensions, including our particular attention on the functional form for the production function and the comparison between two countries. Given the results in Attanasio, Meghir, and Nix (2020), we emphasize the interaction between health and cognition, recognizing that disease and malnutrition can have detrimental effects on cognitive development. We also use their approach to allow for the endogeneity of parental investment decisions, which offers some insight on whether parental investments reinforce or compensate shocks experienced by the children. We experiment with flexible functional forms for the production functions. But perhaps the most important innovation in the paper is the empirical focus on two developing countries, where, with the exception of Attanasio, Meghir, and Nix (2020) in India, child development has not been studied before over such extensive age ranges. Ultimately the hope is that by studying child development in various different low income countries we can identify important regularities that will help us understand the process and design more effective interventions.

We find that the production of health and cognition is quite similar in both Peru and Ethiopia. Specifically, in both countries we find that both cognitive skills and health are very persistent, although health is even more persistent than cognition. We find some evidence that health is cross productive; health positively impacts the production of cognition at early ages. Investments have large impacts on the production of cognition, but the effect decreases with age. We also find in both countries that investments are endogenous and parents compensate for negative shocks. Overall, our results are consistent with a growing body of evidence that early investments in children matter and that the pattern of persistence and dynamic complementarities is a complex one that changes over time as children age (see, for example, Cunha, Heckman, and Schennach (2010), Del Boca, Flinn, and Wiswall (2014), Attanasio, Cattan, Fitzsimons, Meghir, and Rubio-Codina (2020) and Attanasio, Meghir, and Nix (2020)).

Last, we perform a number of counterfactuals. First, we examine the impact of increasing either investments alone or investments and health at age 5 for children with cognitive deficits at age 5. We find that this intervention leads to large gains in cognition which are sustained through age 15. Next, we show that rich and poor children with identical baseline skills will end up with large gaps in cognitive skills by age 15 due to the fact that richer parents invest more in their children.

The rest of the paper is organized as follows. In Section 3.2, we develop the conceptual framework that is used in the empirical analysis. In particular, we sketch the model of human capital accumulation and discuss how different functional forms can have different implications. In Section 3.3, we discuss the estimation strategy we use and in Section 3.4 the data. In Section 3.5, we present our main empirical results, focusing on the cross-country comparisons and the differences across the Nested CES and the CES. In Section 3.6 we present counterfactual exercises and Section 3.7 concludes.

3.2 The Process of Human Capital Development from Age 1 to 15

We consider human capital development over a sequence of periods in childhood, driven by the available spacing of the data. In each period, we specify how current levels of health and cognition are separately determined as the result of a process that combines five different inputs: the child's past health ($\theta^H_{i,t}$), child's past cognitive ability ($\theta^C_{i,t}$), parental investments ($I_{i,t}$), parental health (P^H_i), and parental cognition (P^C_i).

In particular, we write the production of a given component of human capital this period, $\theta^k_{i,t+1}$, as the result of a production function f_t which takes as inputs the components discussed above:

$$\theta^k_{i,t+1} = f_t\left(\theta^C_{i,t}, \theta^H_{i,t}, I_{i,t}, P^C_i, P^H_i, X_{i,t}, A^k_t, \varepsilon^k_{i,t}\right), \quad k \in \{C, H\} \tag{3.2.1}$$

Notice that in addition to the five inputs listed above, we also include a number of other factors that might affect the accumulation of human capital, in a vector $X_{i,t}$. These include the gender of the child, the number of siblings in the family, and the number of older siblings. Gender is included to test whether the process of human capital accumulation is different for boys and girls. Note that gender effects could also capture differential investments by parents, if those investments are not fully captured by I_t. For example, we do not have sufficient information on parental time investments in children, so a difference in human capital production by gender could capture differential time investments. The number of siblings may capture differential returns to scale across large and small families: smaller investments in any given child may be overcome by shared investment in siblings.

In addition to observable variables, we allow the production functions for cognition and health to depend on unobserved shocks ($\varepsilon^k_{i,t}$) and a term meant to capture 'total factor productivity' (TFP). The TFP term (A^k_t) is particularly important when estimating different production functions covering several ages since it can flexibly capture growth over time and with age. We discuss issues related to the interpretation of TFP in more detail in the results section.

We pay particular attention to the marginal effects of parental investments and the child's past skills at different points in time. The derivatives of child skills are important as they play a large role in determining the persistence of the process that governs human capital accumulation. The combination of the marginal effects of investments and child's past skills determines the degree to which investment may (or may not) have long run effects. Understanding the persistence as well as the immediate impact of investments is clearly crucial for the design of effective interventions and the identification of what have been called 'windows of opportunities'.[3]

Equation (3.2.1) above is a fairly general representation of the production of child human capital. However, while a fully nonparametric production function is identified under conditions defined in Cunha, Heckman, and Schennach (2010), as well as further conditions on the support of the instruments for investments, the dimensionality of the problem makes this a forbidding task, both computationally and in terms of the amount of data required. Even so, a flexible specification can be important for the economic implications of the results obtained from the estimates. In particular, estimates of the productivity of parental investment and its persistence (or fade-out) can be substantially affected by the functional form assumptions used. In what follows, we strive to combine analytical and empirical tractability with flexibility.

The Functional Form of Children's Human Capital Production

A specification for the production function that has recently been used in the literature on human capital production is the Constant Elasticity of Substitution (CES) production function (see, for instance, Cunha, Heckman, and Schennach (2010)). The CES allows for some degree of flexibility in the way in which inputs interact. In particular, different inputs can be either complements or substitutes. These interactions have important implications for human capital development among children and optimal policies to foster human capital development. Given the two components of human capital we are considering, cognition (C) and health (H), the CES production function is given by:

$$\begin{aligned}\theta^k_{i,t+1} = [\gamma^k_{1t} P^{C\,\rho_{tk}}_i + \gamma^k_{2t} P^{H\,\rho_{tk}}_i + \gamma^k_{3t} I_{i,t}{}^{\rho_{tk}} + \\ \gamma^k_{4t} \theta^{C\ \rho_{tk}}_{i,t} + \gamma^k_{5t} \theta^{H\ \rho_{tk}}_{i,t}]^{\frac{1}{\rho_{tk}}} e^{\phi'_{tk} X_{i,t} + A^k_t + \varepsilon^k_{i,t}}, \quad k \in \{C, H\}\end{aligned} \tag{3.2.2}$$

where the parameters $\gamma^k_{jt}, j = 1, ...5$, are constrained to sum up to 1 in each period. The term A^k_t captures TFP. Notice that the parental cognition and health variables are assumed not to change over time. In contrast, the parameters of the production function and all other inputs vary with the age of the child.

The CES nests as special cases a linear production function, which occurs when

[3] Cunha and Heckman (2008) and Cunha, Heckman, and Schennach (2010) have also emphasized the importance of non-cognitive (or social) skills, which they show to be important determinants of future outcomes. Unfortunately our data does not include a sufficient number of relevant indicators quantifying this dimension of human capital.

$\rho_k = 1$ and implies perfect substitutability among the various inputs. It also nests as a special case the Cobb Douglas, which occurs when $\rho_{tk} = 0$ and implies that the elasticity of substitution between different inputs is 1. An important limitation of the CES functional form, however, is that it imposes an elasticity of substitution that is the same among any pair of inputs of the production technology: ρ_k is the unique parameter capturing the degree of substitutability among any pair of inputs.

A feasible and more flexible alternative to the CES is the so-called nested CES. Such a specification allows one to explore whether subsets of inputs are complements or substitutes in the production of cognitive or health skills. Inputs are grouped in different sets and each set is aggregated with a CES function, with the resulting 'intermediate output' entering another CES function. In principle one can consider many groups of inputs, even pairs, and hence achieve a different degree of substitutability among all different pairs of inputs.[4] There is, of course, a degree of arbitrariness in the way in which groups of inputs are formed and nested within the outside CES functional.

In this paper, we explore the possibility that initial conditions for child development (given by lagged cognition and health) are combined by a simple CES which is then nested in another CES. This expresses the production of human capital outcomes as a function of the aggregated lagged child skill levels, parental background and investments. This gives the following equation:

$$\theta^k_{i,t+1} = [\gamma^k_{1t} {P^C_i}^{\rho_{tk}} + \gamma^k_{2t} {P^H_i}^{\rho_{tk}} + \gamma^k_{3t} {I_{i,t}}^{\rho_{tk}} + \gamma^k_{4t} (\delta^k_{1t} {\theta^C_{i,t}}^{\rho_{skills,tk}} + (1-\delta^k_{1t}) {\theta^H_{i,t}}^{\rho_{skills,tk}})^{\frac{\rho_{tk}}{\rho_{skills,tk}}}]^{\frac{1}{\rho_{tk}}} e^{\phi'_{tk} X_{i,t} + A^k_t + \varepsilon^k_{i,t}}, \quad k \in \{C, H\} \tag{3.2.3}$$

This specification allows us to test for the existence of cognitive and health skill complementarity in the production of future skills. Notice that equation (3.2.3) reduces to equation (3.2.2) when $\rho_{skills,tk} = \rho_{tk}$. The possibility of having different degrees of substitutability among subsets of inputs is relevant since it will affect the marginal productivity of investment in skills. The nested CES allows for greater flexibility in the set of complementarity patterns between initial cognition and health on the one hand and investments on the other, which may be important in understanding how skills develop with age.

[4] For a good overview of the nested CES, see Sato (1967).

3.3 Estimation

There are two main challenges to estimating the production functions. First, the factors are not directly observed and this leads to a complex measurement error problem. Second, parental investments may be endogenous.

3.3.1 Extracting the latent factors from a system of measurements

While PPVT scores, math test scores, and language test scores can all be thought of as measuring underlying cognition and the various health measures (such as height for age) as reflecting the underlying latent health, none are a perfect proxies for latent cognition and health respectively. We can view these measurements as error-ridden indicators of these latent unobserved factors. To extract the latent factors from these measurements and remove the measurement error, we use a dynamic latent factor model as developed for nonlinear models in Cunha, Heckman, and Schennach (2010), building on work from Hu and Schennach (2008) and Schennach (2004). We use their framework to identify our latent factors of interest from the rich set of measurements in our data. As they show, provided we have $2K$ measurements on a set of K factors in which we are interested, we can use the multiplicity of measurements for each factor to extract the true, unobserved latent factor, despite the presence of measurement error. While extracting the true, latent factor from the set of available proxies is a nontrivial exercise, Cunha, Heckman, and Schennach (2010) show that the alternative - to estimate production functions using a proxy for each factor, ignoring measurement error concerns - performs poorly, even when using the most informative proxy for each factor.

Letting $m_{j,k,t}$ denote the jth available measurement relating to latent factor k in time t, we assume a semi-log relationship between measurements and factors

$$m_{j,k,t} = a_{j,k,t} + \lambda_{j,k,t} ln\left(\theta_{k,t}\right) + \epsilon_{j,k,t} \qquad (3.3.1)$$

where $\lambda_{j,k,t}$ is a factor loading, $a_{j,k,t}$ are constants, and $\epsilon_{j,k,t}$ are zero mean measurement errors, which capture the fact that the measurements are imperfect proxies of the underlying factors. We assume that the measurement error is independent of the latent factors, which is necessary for identification. We also assume that the measurement errors are independent of each other (across both contemporaneous latent factors and time).[5]

[5]This assumption can be somewhat relaxed. However, even when allowing for some correlation among the errors of some of the available measures, it will be necessary to have some measures with independent errors. This is not something we explore in this paper.

To identify the model we must also scale and normalize the measurements (Anderson and Rubin (1956)). We use the same normalizations used in Cunha, Heckman, and Schennach (2010), setting the first factor loading for each measurement to 1 and normalizing the mean of the latent factors to be 0, while including a TFP term. An alternative approach, presented in Agostinelli and Wiswall (2016), normalizes the means and loadings of only the initial factors. Cattan and Nix (in progress) explore alternative methods and estimators and provide Monte Carlo evidence that shows that the approach we are following performs well in samples of the size we are considering.

The interpretation of the estimates of the production function depends on the units of measurement of the latent factors. In some cases, the available measurements are test scores which are ordinal in nature. Ordinal variables introduce some arbitrariness to results, given that any monotonic transformation of a test score conveys the same information. A way round this issue is to anchor all latent factors to a measure with meaningful units such as earnings, as elaborated by Cunha, Heckman, and Schennach (2010). Anchoring will define not only the scaling but the entire monotonic transformation from the latent factor to the measure. [6] Here we lack a measure that would make a suitable anchor because we do not possess longitudinal data that will link early test scores to adult outcomes such as earnings. We thus assume that the factor and the measure are related by a semi-log transformation (and the measurement error).

3.3.2 Determinants of parental investment

In our model, parental investments in children are determined by parental resources (current and over the lifecycle), parents' expectations regarding the returns to investments in their children, and the prices of investment goods. More specifically, parents select investments taking into account the child's current level of cognition and health, because the child's current level of human capital may determine the returns to investments (in particular if there is complementarity between child skills and investments). Additionally, the level of investments may be determined in part by the parent's own level of cognition and health, both because this may reflect knowledge about the value of investments and because parental characteristics may capture lifetime resources. Gender may also play a role because of gender preferences or even because of perceived differences in the returns to gender in the labor market. Finally, birth order and the number of siblings impact the available material and time resources that parents

[6] Meghir and Rivkin (2011) discuss the this issue in the context of program evaluation with methods such as difference in differences.

are able to devote to a given child. This characterization of parental investments can be derived from a problem where parents choose investments to maximize a welfare function whose arguments are child human capital and own consumption, subject to a budget constraint and the technology constraint imposed by the production function of human capital. Estimating the investment equation can provide some insight as to how parents choose to invest in children and ultimately will help us understand some key sources of inequality in this paper.

We assume that parental investments are at least partly motivated by a desire to affect long term outcomes through the formation of human capital, where human capital in this paper consists of health and cognition. This is why the production functions depend on investment. However, in order for the estimates of the production function to be unbiased, we have to assume that all determinants of both investments and human capital formation are fully captured by the remaining inputs into the production function for human capital: parental cognition and health, past child health and cognition, and household composition.

Yet there is an important possible source of endogeneity: the unobserved shocks to the child's human capital that parents might react to when choosing their investments. For example, parents may decide to compensate for a negative shock, such as an unexpected episode of ill-health or particularly bad quality of education, by increasing their investments in the child. Alternatively, such a shock may be perceived as lowering the returns to investment, in which case investments may also fall.

To address this potential endogeneity of investments, we follow the control function approach used in Attanasio, Meghir, and Nix (2020).[7] As in that paper, we use household income and regional variation in prices as instruments. These instruments are valid provided differences in prices and income affect future cognition (or health) only through their impacts on investments. The use of prices as instruments requires the assumption that regional variation reflects cost differences rather than differences in demand. Given the longer term determinants of income (such as parental cognition and health) we assume that current income reflects a liquidity shock and not inputs that should also be included in the production function. As a robustness check we also present results where we only use prices as excluded instruments in the appendix. In those specifications, income is included in the production functions in $X_{i,t}$ as well as in the investment equations.

We approximate the parental investment function using a log-linear specification

[7] See Gronau (1974), and Heckman (1979).

which takes the form:

$$lnI_{i,t} = d_0 + d_1 ln\theta^C_{i,t} + d_2 ln\theta^H_{i,t} + d_3 P^C_i + d_4 lnP^H_i + d_5^{'} X_{i,t} + d_6^{'} lnq_{k,t} + d_7 lnY_{i,t} + v_{i,t} \quad (3.3.2)$$

where $\nu_{I,t}$ is the error term, $X_{i,t}$ includes child gender, birth order, and number of children, Y is income, and $q_{k,t}$ are a set of prices in village k at time t. We can then augment the productions functions to include the residual of the investment function, $v_{i,t}$, as a control function. Our estimating equation for the production functions for child health and cognition are then:

$$ln\theta^k_{i,t+1} = ln\left(g\left(\theta^C_{i,t}, \theta^H_{i,t}, I_{i,t}, P^C_i, P^H_i\right)\right) + \phi^k_t X_{it} + A^k_t + \xi^k v_{i,t} + \varepsilon^{*k}_{i,t} \quad k \in \{C, H\} \quad (3.3.3)$$

where the function $g(\dot{)}$ is the CES or the nested CES production function outlined earlier and $C,\ H$ denote cognition and health respectively.

An alternative approach is to solve explicitly for the maximization problem faced by parents and compute the investment function that would result from such a problem, as in Del Boca, Flinn, and Wiswall (2014). Such a function could then be estimated jointly with the production function. Such an approach might be more efficient in certain contexts but does require stronger assumptions about behavior. For example, we do not necessarily require that parents have full knowledge of the parameters of the production function.

3.3.3 A Three-Step Estimator

As discussed above, the identification of the nonlinear factor model we use is provided in Cunha, Heckman, and Schennach (2010) and is based on the idea of dealing with measurement error in a nonlinear context by using multiple measures for each underlying variable, as in Schennach (2004). The estimation approach we use is described in Attanasio, Meghir, and Nix (2020) and consists of three steps. In the first step, we estimate the joint distribution of the measurements. In addition to the measurements for the latent factors that enter the production function, we also include the variables used as controls (gender, number of children, and so on) and the variables used as instruments for investments (prices and income). Even though no measurement error is considered for these variables, we must include them in this first step in order to recover their relationships with the latent variables. In the second step, we use the measurement system to recover the joint distribution of the latent factors, as well as estimates of the factor loadings and measurement error distribution. With the joint distribution of the latent factors and all other relevant variables in hand, we then generate a synthetic data set by drawing observations from the joint distribution of the factors (including the instruments and controls) and estimate the investment functions and production functions in the third and final step, using nonlinear least squares. Experimenting with the functional form of the production function only involves repeating the relatively simple third step, since the joint distribution of the latent factors completely characterizes all the information in the actual data. For more details on this estimation procedure and its performance, see Attanasio, Meghir, and Nix (2020).

We assume the measurement errors are normally distributed and we approximate the joint distribution of the measurements as a mixture of two normals. These assumptions then imply that the joint distribution of the latent factors is a mixture of normals. The departure from normality is important: the normal distribution has a linear conditional mean. Here the conditional mean is the production function, which means that assuming normality of the latent factors would only allow us to estimate production functions where the inputs are perfect substitutes. More generally, the joint distribution of latent factors must be flexible enough to capture the dependencies in the data and general enough to be consistent with the production function one wishes to estimate. By using an increasingly large number of elements in the mixture the joint distribution can be approximated to an arbitrary degree of precision. Cattan and Nix (in progress) show that a mixture of two normals is sufficient to capture a CES production function.

3.4 Data

To estimate the production functions, we use data selected from the Young Lives Survey. This is a longitudinal survey covering 12,000 children in four countries: Ethiopia, India, Peru, and Vietnam.[8] The survey began in 2002 with two cohorts of children in each of these countries. In 2002, the younger cohort was between 6 and 18 months and the older cohort was between 7.5 and 8.5 years of age. The second wave of the survey took place in 2006-2007, and the third wave took place in 2009.[9]

In Ethiopia, the sample selected children from five regions out of nine in the country: Addis Ababa, Amhara, Oromiya, SNNP (Southern Nations, Nationalities and Peo-

[8] Attanasio, Meghir, and Nix (2020) present results for India, in addition to developing the estimation procedure used here.

[9] The survey is scheduled to continue, following the same children, every three years through 2016. Once the final two waves are publicly released, it will be possible to extend the analysis in this paper to look at a broader range of outcomes. In particular, we will be able to anchor the estimates to adult outcomes, such as teenage pregnancy, high school completion, college matriculation, and wages.

ple's region) and Tigray. These five regions account for 96% of Ethiopia's total population. Within each region, three to five districts were selected to obtain a balanced representation of poor rural and urban households, as well as relatively less poor rural and urban households. In total, 20 districts are included in the sample. Within the districts, at least one peasant association (in rural areas) or one kebele (i.e. the lowest level of administration for urban areas) was selected and was included as sites if they contained a sufficient number of households with children in the relevant age range. Note that sites were chosen to oversample areas where food deficiency is particularly relevant, as well as to capture Ethiopia's diversity in terms of ethnicities, in both rural and urban areas. The households participating in the survey were randomly selected within sites.[10]

In contrast to the other countries in Young Lives, Peru chose the 20 sentinel sites using a multi-stage, cluster-stratified random sampling approach. However, as in Ethiopia, the 20 sites considered for the multi-stage, cluster-stratified random sampling were chosen so as to over-represent poorer districts (this was achieved by excluding the richest 5% of districts from the sample according to the poverty map developed in 2000 by the Fondo Nacional de Cooperacion para el Desarrollo). The poverty ranking of districts is constructed taking into account factors such as infant mortality, housing, schooling, infrastructure, and access to services. The sample covers households from the following regions of Peru: Tumbes, Piura, Amazonas, San Martin, Cajamarca, La Libertad, Ancash, Huanuco, Lima, Junin, Ayacucho, Arequipa and Puno.[11]

The surveys are extremely detailed, and we use information from household questionnaires, child questionnaires, and community questionnaires. In Table 3.1, we present statistics on the children and their households. In Ethiopia, the younger cohort includes 1,999 children and the older cohort includes 1,000 children. By age 5, the average number of children in the household is approximately 4. In the older cohort, by age 12 the same figure is around 6. On average, the focus child in the younger cohort has 2 older siblings at age 5 while children in the older cohort have 3 older siblings at age 12. These relatively high numbers reveal demographic characteristics that are typical of developing countries, which often display households with a relatively high number of children.

Consistent with the aims of the Young Lives sampling approach, the households in our sample are relatively poor. This is captured in the summary statistics on average annual income and the wealth index. Average annual income in Ethiopia is computed

[10] For more information on the sampling approach, see Escobal and Flores (2008).

[11] For more information on the sampling approach and the Peru data, see Outes-Leon and Sanchez (2008) and Escobal, Lanata, Madrid, Penny, Saavedra, Suárez, Verastegui, Villar, and Huttly (2003).

taking into account income from both wages and benefits. The wealth index is computed by Young Lives as an average of three indices that measure housing quality, consumer durables, and access to services. Average annual income in round 2 is around 6,000 Ethiopian Birr, and around 13,000 Ethiopian Birr in round 3. The high standard deviation of the wealth index reveals the degree of heterogeneity in terms of economic background in the Ethiopian sample, but could also be indicative of measurement error. Given this fact, we will use both the wealth index and measured income as measurements on latent income.

The younger cohort in Peru includes 2,052 children, whereas the older cohort includes 714 children. Peru is a relatively more developed country, and this is reflected in the summary statistics in Table 1. In particular, the average number of children in the household is 3 by age 5, and 4 by age 12. Also, the average number of older siblings is approximately 2 for both samples. In terms of economic well-being, average annual income is approximately 12,000 Peruvian Sols in rounds 2 and 3. The Young Lives wealth index displays relatively less variance with respect to Ethiopia.

In Tables 3.12-3.13 in the Appendix we report the summary statistics for the child measurements associated with each latent factor at each age. These measurements demonstrate the richness of data available in the Young Lives data set, which we take full advantage of in our approach. For example, the survey administered a number of tests, including the PPVT, a math test, the CDA test, and the Early Grade Reading Assessment (EGRA), which we use to identify latent cognition for ages 5, 8, 12, and 15. For age 1, no measurements for cognition are available, reflecting the fact that it is almost impossible to measure cognition when children are under the age of 1. In addition, the survey measured weight, height, and elicited self-reported (by either the parent or child) health status, which we use to identify latent health for ages 1, 5, 8, 12, and 15. Assignment of measurements to the remaining factors (investments and parental health and cognition) are discussed in Section 3.5.1.

3.5 Results

We organize our results as follows. We first present and discuss the estimates of the measurement system. These estimates identify the joint distribution of the latent factors (child health and cognition, parental investments, and parental health and cognition) and the other variables that make up the production functions. Next, we discuss the determinants of parental investments. Last, we present the results for the production functions. For each set of results, we compare the estimates obtained with the Peruvian and Ethiopian data.

3.5.1 The measurement system across countries

As described in Section 3.3, the estimation of the measurement system is done in two steps. The estimates of the measurement system are then used to generate synthetic data in the third step. The same synthetic data can be used to estimate both the CES and the Nested CES production functions, given that the same measurement system underlies both production functions.

The estimates of the measurement system can be used to summarize how informative each measurement is in terms of the underlying factor it proxies. In particular, for each measurement used to extract the latent factors, we can compute the signal to noise ratio which is equal to the ratio of the variance of the latent factor to the variance of the measurement error.

In Table 3.2, we report signal to noise for the measurements we use to identify latent children's cognitive ability at each age. Also reported in Table 3.2 are the signal to noise ratios for the measurements we use to identify latent parent's cognitive ability for the children from each cohort, which is treated as fixed across the different ages. We find that the measurements we use for cognitive skills are very informative, with only a few exceptions (specifically numeracy and writing level are not very informative measurements, which may be related to the limited variability in these measurements).[12] Finally, this and subsequent tables report the factor loading on the *log* of the factor.[13]

In Table 3.3, we report signal to noise for the measurements used to proxy child health across all ages. We find that with the exception of self-reported health status[14], the measurements on child health are extremely informative regarding latent health: their signal to noise ratios all exceed 50%. The parental health measurements are marginally less informative.

Finally, in Table 3.4, we report signal to noise ratios for the measurements used to proxy investments in children across all ages. The results in the table show that the majority of our information regarding latent investments is related to expenditures on children. The amounts spent on clothing, shoes, and books are particularly informative.

One drawback of our data is that we do not have sufficient information to split latent

[12] Numeracy is the answer to the question: does the child correctly answer what is 2x4? Writing level is a score of 0, 1, or 2.

[13] Notice that because of the log transformation the units relating to the level of the factor are absorbed by the intercept of the measurement equations, which has been eliminated by demeaning all the measures.

[14] The signal to noise for self-reported health status is sometimes very close to 0. When this occurs, it is because the loading is estimated to be 0 and identification is then based off of the other two measurements. If there were only two measurements, the factor would not be identified, but fortunately in both of these cases there are two additional measurements that are used to identify the latent factor for health.

investments into separate factors, representing different types of investment. With rich enough data, one could consider different types of investment, such as investment in time and material (as done in Attanasio, Cattan, Fitzsimons, Meghir, and Rubio-Codina (2020)) and separately identify investments in health and cognition.

Cross Country Comparison

The signal to noise ratio for the same measurement is often very different across the two countries. For example, the PPVT test appears to be much more informative in Peru relative to Ethiopia. Similarly, the number of food groups is more informative in Ethiopia than in Peru. In other cases, there are large differences but without a discernible pattern across countries when considering all ages. This is a very interesting result and of relevance for international comparisons and, more generally, for the construction of measures that can be used for such comparisons. There are a number of possible explanations for the differences we observe.[15]

It may well be possible that the differences are down to data collection: if the enumerators have different skills collecting data in the two countries this could give rise to very different measurement error structures. The difference in signal to noise ratios across contexts could also arise if there is less variability in a given factor in one country versus the other, but the same amount of noise. This may have to do with how heterogeneous the population is in each of the two surveys. A third hypothesis is that some measurements are not comparable across countries because they have not been adapted in the same way (this explanation is particularly relevant for the cognitive tests but arguably less relevant for measurements like height and weight, which can be uniformly applied across countries). This is a general point and cautions about the use of tests to compare across different contexts. It is likely that all three of these explanations are at work in our context. Provided the differences across contexts are not so severe that they invalidate our assumptions on the measurement system, the approach taken here deals with variability in measurement error across countries by removing it and extracting comparable latent factors, albeit with a scale set by the choice of normalization of the loading.

[15] This set of issues is somewhat related to the discussions in Deaton and Heston (2010) and Deaton (2008).

3.5.2 The determinants of parental investments in children across countries

As discussed in Section 3.3.2, parental investment decisions are approximated by a log-linear function. Investments in period t depend on the child's existing stock of human capital, parental cognition and health, birth order, the number of children in the household, child gender, parental income, and prices of investment goods. Birth order, number of siblings, and child gender could be potential determinants of investment both through the budget constraint and by capturing features of parental preferences. Income is clearly an important determinant of investment as it contributes to the relative tightness of the budget constraint that the household faces. Prices of various goods (at the district level each period) that are related to child development may affect investments through the budget constraint. Specifically, we include a price index for food, a price index for clothing, the price of a notebook (related to costs of education), and the prices of the medicines Mebendazol and Amoxicillin.[16] As noted in our description of the data, child cognition in the first wave of the youngest cohort, when the children were 1, was not measured. For this reason the investment equation for age 5 does not include child cognition.

The investment equations are of interest since they reveal potential links through which child's and parental characteristics, as well as other covariates, influence parental investment behavior. In addition, the saliency of prices and income are particularly important since we use them as instruments for investments in the production functions. For this reason, in the tables below we report the results of a joint F-test for income and the price variables. As we will also investigate the possibility of not using income as an instrument for investments when estimating the production functions, since income may itself be endogenous, we also report the F-test for the joint significance of the price variables alone.

Ethiopia

The determinants of parental investment in health and cognition in Ethiopia are reported in Table 3.5, along with 95% bootstrap confidence intervals. We find that parental cognition increases investments at all ages but 15 and parental health affects investment at age 5 and 8, but not at older ages of the child.[17] Perhaps surprisingly,

[16] Mebendazol is used for the treatment of worms in children and Amoxicillin is a penicillin antibiotic used in the treatment of various infections.

[17] To the extent that child cognition at age 1 is correlated with parental cognition it may well be that the impact of the latter on investments at age 5 is overestimated since we do not observe child cognition at age 1, and thus cannot include lagged cognition at that age.

child health and cognition do not affect investments. Birth order and the number of children are also mostly insignificant. However it is remarkable that for children at age 5 investments in males are higher by 16%. Although the male preference is no longer evident at older ages, given the potential importance of investment at early ages this is possibly an important source of gender differences in child development. Consistent with theory, families with higher income invest more. The elasticity is highest at the lowest age and decreases thereafter to just under 0.4. While we cannot establish causality beyond reasonable doubt it is worth remembering that the estimated income effect is *conditional* on parental background and household composition.

Turning to the joint significance of the variables we use as instruments in the production function, we find that when we consider income and prices jointly, the p-values for the null of no joint significance are very low. However, when we consider only the price variables, we find that they are only jointly significant at ages 8 and 15. In these cases what stands out is the price of the notebook (age 8) reflecting educational costs at that critical age, the price of the deworming drug Mebendazol at age 15, and the price of the antibiotic Amoxicillin at age 8. These estimates imply that a 10% increase in the price of a notebook at age 8 would lead to a 1.93% decrease in investments, a 10% increase in the price of Mebendazol at age 15 would lead to a 3.58% reduction in investments, and a 10% increase in the price of Amoxicilin would lead to a 2.08% decrease in investments at age 8.

Peru

The estimated coefficients of the investment equations for Peru are reported in Table 3.6. Here again neither child cognitive skills nor child health are significant determinants of investment at any age. This result is potentially important because, as we will see next, investments and child cognition are complementary which implies that the returns are higher for children with higher initial levels of cognition. Perhaps parents are not aware either of the level of cognition of their child or of the parameters of the production functions for human capital.[18] The exception is that child health has a positive effect on investments at age 8.

As in Ethiopia, parental cognition seems to matter early on at ages 5 and 8, although the effects are not as strong in Peru. Nevertheless, the result we obtain with the Ethiopian data also holds in the Peruvian context: at older ages child investments do not depend on parental cognition. Parental health matters when the child is 5 but not later. The number of children reduces investments at all ages (although it is

[18] Attanasio, Cunha and Jervis (in progress) explore the possibility that parents have distorted beliefs about the production function for human capital.

not significant at age 8). In contrast, investments are higher for children with older siblings at later ages, perhaps because older siblings are independent earners and more resources are available (conditional on the overall lower investments in families with many children). In Peru, we find no evidence of male preference in investments. Finally, income plays a critical role in investments with the impact being highest for younger children, as in Ethiopia. In Peru, we see a much more pronounced decline of the income effect as the child ages. However, what stands out is that in both cases income matters most at the youngest age suggesting that poorer parents invest substantially less at an age where these investments may be critical for child development. We will explore the broader implications of this fact, combined with what we find regarding the production function parameters over these ages, at the end of the paper.

Jointly prices are significant at all ages but the youngest. However, the price of clothing and the price of food enter with the 'wrong sign'. As discussed above, the likely reason for this result is the source of (regional) variation in prices that we are exploiting. What stands out as before is the price of the notebook at all ages except the youngest (although it is only significant at the 5% level at age 12), and the price of Mebendazol at all ages (although at age 12 it is not significant).

Summary

The one unambiguous result across both countries is that investment in children is driven by income conditional on both parental background and child health and cognition. We interpret this as the impact of current resources. Importantly, the effect of income is highest for the youngest children, meaning that differences in investments across income groups are highest at the youngest ages. This holds true in both countries. The quality/quantity trade-off is only apparent in Peru and male preference is evident at younger ages in Ethiopia.

Finally, there is evidence that prices matter, particularly those relating to health care and education. We rely exclusively on spatial variation, which we assume is not driven by to demand differences. However, this is clearly not the ideal data set to identify robust price effects. Some of the perverse price impacts may well be due to a violation of the price exogeneity (due to shifts in preference for child goods across regions). Notice that violation of price exogeneity does not jeopardize our ability to identify the parameters of the production function, as long as the 'shocks' to prices or the demand shifts across regions that are the cause of endogeneity are orthogonal to the shocks to the human capital production function that cause the endogeneity of investment.

3.5.3 Production function estimates across countries

Investments reflects parental choices. The next step is to estimate the way these choices, together with the child's background, affect the production of cognition and health over the various stages of the child's life. We explore different functional forms for the production functions of child human capital. We first report estimates of CES production functions, which until now has been the most flexible functional form estimated in this literature. We then move on to consider a less restrictive specification, given by the nested CES. Specifically, we estimate the parameters of equations 3.2.2 and 3.2.3 with one adjustment at age 5. As in the case of investment, the production functions estimated for age 5 do not include cognitive skills in the previous period as an input, since we have no measures on cognition at age 1.

Constant Elasticity of Substitution (CES) Production Functions: Results and Cross Country Comparisons

In Tables 3.7 and 3.8, we present the results from estimating CES production functions for each country. In both tables, investment is considered as endogenous and the coefficients are obtained with a control function approach, using prices and income as excluded instruments, as described above. The assumption underlying the exclusion of income is that conditional on parental and child background as well as the various demographics, income represents a liquidity shock and does not reflect omitted skills and characteristics of the parents that directly affect the production of child human capital. In the appendix we experiment with using just prices as exclusion restrictions. However, as discussed above, prices are not necessarily exogenous either: we still need to assume that either the demand functions for investment are the same across regions over which we have spatial price variation, or at least that any shocks to demand across regions are orthogonal to shocks to the production of human capital that are correlated with parental investments. In general finding the right instruments is an important but hard task and one hopes that in future we can exploit randomized or quasi-experimental variation.

Parental cognition and health have no significant impact at any age on child cognition in Ethiopia. However, in Peru parental cognition enters from age 8 onwards. The production of health does not depend on parental cognition but for both Ethiopia and Peru parental health matters for child health when the child is 5 and when the child is 15. In both Ethiopia and Peru, child cognition and health are self productive. Most importantly, and similarly to the results for India in Attanasio, Meghir, and Nix (2020), we find that child health matters for cognition at least at age 12 in Peru and at ages 5 and 8 in Ethiopia. This result is important because it highlights the importance of child ill-health, prevalent in poor environments, in generating cognitive deficits. According to these results, improvements in child health will not only increase future health, but will also feed back into child cognitive development.

Turning now to the coefficients on investments: in Ethiopia there is a very strong effect of investment on cognition at all stages of childhood and a strong effect on health at age 5 and at age 15, but not in between. Similarly, in Peru we find a very strong impact of investment on cognition when the child is 5 and 8 but not at later ages where the effect is imprecisely estimated. Moreover the impact of investment on health in Peru is very imprecisely estimated. Looking at the investment residual we see that it is often significant and comes in with a negative sign; the negative correlation of the investment residual with the shock to the production function may be interpreted as compensatory behavior of the parents, i.e. following a negative shock to their child's cognition they tend to increase investments. However, this causal interpretation should be made with some caution, especially in situations in which the joint significance in the investment equations of the excluded instruments is not very strong as is the case at age 15 for Peru. For this reason, in the appendix we report the results obtained by OLS. As can be verified from the results in those tables, the size of most parameters is not affected dramatically.

The OLS results in the appendix are useful for another reason. When the coefficients on the investment residuals are significant (and negative) in the production function, it is useful to compare the OLS estimates of the coefficients on investments to those obtained when taking into account the endogeneity of investments. We find that in these instances, the coefficient on investment increases considerably in size relative to the OLS estimates. This is an additional indication that parental investment may serve a compensatory role relative to shocks received by the children.

In both countries whenever there is a significant effect of the number of children on the production of either cognition or health it is negative. More children in the household seem to be detrimental both in terms of investment but also in terms of the production of human capital (with one exception at 12 in the production of cognition in Ethiopia). Importantly, there does not seem to be a strong effect of gender on the production function. Any significant effects are very small. The notable exception is the effect of Male on health at age 15 in both Ethiopia (with a negative sign) and Peru (with a positive sign). This may have to do with differences in behavioral norms of teenagers in the two countries, although we do not have any concrete evidence relating to this.

The elasticity of substitution is not always very well determined. However, for cognition it is generally close to one. For health it varies between 0.65 and nearly 2. Despite being imprecisely estimated these results exclude very high levels of substitution, suggesting that there is complementarity between the various inputs. We explore this further with the nested CES in the next section.

Nested CES: Results and Cross Country Comparisons

The nested CES allows for a more flexible pattern of substitutability among the different inputs. The inputs we nest are the child's skills (cognition and health) in the previous period. This allows us to investigate whether lagged child cognition and health have a different degree of substitutability in the production of child human capital with respect to all other inputs.

However, cognition does not affect health except at age 12 in Ethiopia and age 8 in Peru. In contrast, health affects cognition at ages 5 and 8 in Ethiopia, age 12 in Peru, and at all other ages health enters positively in the production of cognition and is only marginally insignificant. For this reason, we focus on the nested results for cognition only, and report the estimates for the nested CES for health at age 8 for Peru and age 12 for Ethiopia in the appendix. Since we do not observe measures of cognition at age 1, we can only estimate this specification for ages 8, 12 and 15. The results for Ethiopia are in Table 3.9 and the results for Peru are in Table 3.10. As with the CES, we also report OLS estimates without instruments for investments (see Tables 3.17 - 3.18) and estimates using only prices as instruments for investments, in which case income enters both the investment equations and the production functions (see Tables 3.21 - 3.22).

We discuss the implications of the nested CES coefficients on the pattern of substitution elasticity in more detail below. However, here we note that, whilst the coefficient that determines the elasticity between the two initial endowments is not estimated very precisely, we can reject the hypothesis that it is equal to 1 which would imply the standard CES.

The patterns of the other coefficients is broadly similar to that obtained with the CES. Cognition is highly persistent. Health is cross productive, although only significantly so at ages 8 and 12 in Ethiopia. Parental cognition continues to be an important input in child cognitive skill in Peru, but not in Ethiopia. We confirm that investments have a strong impact on the production of cognition, with the effect decreasing in magnitude as children age. This finding is a key result that survives this more general specification. We now turn to the substitution elasticity (and hence implied complementarity) and how this changes with this more flexible specification.

Comparison between the CES and the Nested CES

To test whether the grouping introduced in the nested CES is important, Table 3.11 gives the difference between the two elasticities in the nested CES production functions and 95% confidence intervals for the differences. We find that the two elasticities are significantly different at all ages and in both countries. This implies that the nested CES does in fact fit the data better. However, given the substantial differences in the production functions, it is difficult to know how this, combined with the differences in the magnitudes of the coefficients, should affect our interpretation of the results.

In Figures 3.1 and 3.2 for Ethiopia and Peru respectively, we plot the marginal product of investment for each decile of baseline cognition, computed using the estimated parameters for both the CES and nested CES at ages 8, 12, and 15 for Ethiopia. In particular, these values are computed for each decile of baseline cognition (recorded at age 5, 8, or 12), keeping all other input values at their sample averages. In all cases the strong complementarity between initial cognition and investments is evident. This illustrates the difficulty with which early deficits can be remedied by later investments and poses the difficult policy challenge of reaching the most deprived populations effectively. The graphs also show the potential importance of allowing for flexible production functions: the degree of complementarity differs substantially between the CES and the nested CES.

3.6 Counterfactual Simulations

3.6.1 Impulse response functions

An important advantage of the longitudinal data we are using is that they allow us to follow children over a long period of time and estimate in a flexible way the degree of persistence and cross productivity of shocks and inputs. One way to study the implications of our sets of parameter estimates is to plot an impulse response function for investment innovations. In this subsection, we present this exercise for Ethiopia and Peru in Figures 3.3 and 3.4.

The experiment we perform is to select the children who are, at age 5, in the bottom 5% in terms of cognition. Using this sample, we compute the median values for all inputs and skills, including cognition. As one would expect, children who are "poor" in terms of cognitive skills, also have poorer health, poorer parental health and poorer parental cognition. Using these values and our estimates of the production functions we can predict the pattern of cognition that would occur for this sample (according to our model) if no changes were made. This is our baseline. In Figures 6.1 and 6.2

below, we plot the pattern of cognition that would occur, according to the model, in two alternative scenarios about parental inputs and initial health conditions, relative to the baseline. In the first scenario, we increase the investments these children receive at age 5 to equal the median investment of children in the top 10% in terms of cognition, and then follow the effect of the increase in investments through age 15 as implied by the production functions we have estimated. In the second scenario, we repeat the exercise but also give the child an initial health level at age 5 equivalent to the median health level of children in the top 10% of cognition at that age. This exercise is meant to capture the effect of increasing both inputs in the presence of dynamic complementarities. Any outcome in the figures above 1 implies an increase in cognition relative to the status quo.

For both countries and both production functions we observe that a positive shock to investment and health, as well as to investments alone, have large positive effects on the evolution of cognition over time. Specifically, the interventions lead to a 15-70% increase in cognition by age 15. For Ethiopia, increasing both investment and health at baseline leads to a larger increase in cognition at age 8 in the CES specification compared to the nested CES case, although by age 15 the differences between the two specifications are less noticeable. While there is a small gain from increasing both health and investments versus investments alone in the CES case for Ethiopia, for the nested CES specification the increase in later cognition due to the two alternative policies is almost identical. In Peru the difference between the two interventions is larger, with greater gains from increasing both health and investments at baseline. This result signifies greater levels of complementarity in Peru. This effect is somewhat larger with the nested CES. The key point demonstrated by these exercises is that the complementarity between initial conditions and investments can lead to substantially different results. Moreover, this exercise emphasizes the importance of health at an early age: with low health levels child outcomes respond less to investments.

3.6.2 Quantifying the importance of parental investments in generating inequality in child outcomes

In this section, we compare the patterns of cognitive skills over time for children who have the same baseline initial skill levels, but whose parents belong to different ends of the income distribution. First, we assign children the median level of skills in our sample, as well as the median amount of all other inputs. Next, we use our estimates of the investment equation to compute parental investments for children whose parents are at the 90th income percentile in our data, and the same investments for children whose parents are at the 10th income percentile in our data (all else equal). We then compute the evolution of cognitive skills over childhood for these two different patterns of investment. After the initial age, investments at later ages are also computed using our investment function estimates. However, at later ages not only will children receive different investments due to the differences in their parent's incomes, but also as a result of the differences in cognition and health that come out of the production functions via inequality in investments accumulated in previous ages. The results from this exercise are reported in Figure 3.5. The dashed line shows the evolution of cognitive skills over time when parental investment is computed using the 90th percentile of income in our data. The dot-dashed line shows the same pattern for investment computed using the 10th percentile of income in our data. Both these lines are relative to the outcome that would occur if median income was used. In addition, we include two more lines. These lines show the evolution of skills when initial conditions are not restricted to be the same. Instead, for the top line, we compute the evolution of skills for children whose parents are at the 90th income percentile AND who have the corresponding top 10 percentile level of initial skills (cognition and health), relative to the median. The bottom line gives the evolution of skills for children whose parents are at the 10th income percentile AND who have the corresponding bottom 10 percentile level of initial skills. To economize on space we just show the results for the nested CES. These graphs clearly show that children of high income families, who enjoy persistently higher levels of parental investments, immediately open a large gap relative to poorer children, and this gap not only persists over time but grows. This is despite the same initial conditions and illustrates the inequalities that arise simply from different investments among the better off and the poorer. What is even more impressive is that while the initial gap in skills is much larger at age 5 when initial skills correspond to the income percentile (which is mechanically expected when compared to the case where initial skills are required to be equal) a substantial part of the gap that persists to age 15 when initial conditions are allowed to differ appears to be consistent with the gaps driven by differences in parental income which drive differences in investments at early ages.

3.7 Conclusion

Understanding the development of human capital from an early age is critical to tracing the origins of inequality, to understanding the mechanisms for the inter-generational transmission of skills, and ultimately to design policies that can reduce poverty in a substantial way.

In this paper, following the approach of Cunha, Heckman, and Schennach (2010) and building on work for Colombia and India (Attanasio, Cattan, Fitzsimons, Meghir, and Rubio-Codina (2020) and Attanasio, Meghir, and Nix (2020), respectively) we estimate the determinants of investments in children and how these affect child outcomes in Peru and Ethiopia. Relative to previous studies we investigate the implications of more flexible functional forms that allow for richer patterns of complementarities among different inputs of the production function and add to the evidence by bringing in data from other developing countries, which helps identify common patterns.

We show that in several contexts, allowing for a flexible functional form is important. While in some cases the relatively simple Constant Elasticity of Substitution production function performs well, we find that in many cases the more flexible patterns of complementarity across different inputs allowed by a Nested CES fit the data better. This has important implications for the interpretation of the results, and we show that the shape of the marginal productivity of investments by decile of cognitive skills can be quite different for the CES and Nested CES.

Our focus is on cognition and health, for which we have rich information in the Young Lives Survey. While there are some important differences in the results from the two countries, some common elements stand out, which are also consistent with earlier work. First, investments increase with parental cognition and income, implying that children from better backgrounds get more resources and those resources may also be used more effectively. While this result is not unexpected, the large differences in cognitive development we document driven only by differences in parental income clearly illustrates that this may be an important source of social inequality and points to the potential importance of interventions boosting investments in children from poor backgrounds at a very early stage.

Second, cognition and health are self productive, implying persistence of past levels of child skills. However, there is some depreciation of skills, particularly at the cognitive level, which points to the potential importance of continuing investments throughout childhood. Third, health is important for cognitive development, a result that is similar to what is found for India by Attanasio, Meghir, and Nix (2020). This is a key result because poor health at early ages is prevalent in developing countries. Our results confirm that such ill-health can cause cognitive deficits, starting at very young ages, which can perpetuate poverty. Fourth, investments are important for cognition; however for both countries their impact declines with age pointing to the importance of early interventions. This result coupled with the fact that the elasticity of investments with respect to income is higher at lower ages points to another source of inequality, since the poor will tend to invest less exactly when such investments are most important. Finally, inputs into the production of human capital are complementary, at least based on the measurements and normalizations used in this paper, and the nested CES, which implies richer patterns of complementarity, generally fits the data better.

Together, these results provide important evidence that it is key to devise and implement health and cognition interventions that account for their complex interactions over childhood in order to promote human development. For example, programs focused on cognitive stimulation, sanitation, and possibly nutrition at very early ages can have long lasting and mutually reinforcing effects. We expect our results to further stimulate experimentation with such interventions. In addition, future work could extend the work done in this paper to additional contexts, and connect skills at adulthood to interesting adult outcomes like completed schooling, criminality, and wages, in order to better understand the mapping from these important skills to other policy relevant outcomes.

3.8 Tables

Table 3.1: Summary Statistics: Demographic Variables

	Younger Cohort			Older Cohort		
	Age 1	Age 5	Age 8	Age 8	Age 12	Age 15
Ethiopia						
Demographics						
Number of Children	3.18	4.14	.	4.62	5.61	.
	(2.02)	(2.31)	.	(2.14)	(2.47)	.
Number Older Siblings	.	2.44	.	.	2.84	.
	.	(2.25)	.	.	(2.45)	.
Measures of Economic Well Being						
Annual Income	.	5,742	12,800	.	5,680	12,585
	.	(17,619)	(35,846)	.	(11,097)	(22,452)
Wealth Index	0.21	0.28	0.33	0.22	0.30	0.35
	(0.17)	(0.18)	(0.18)	(0.17)	(0.17)	(0.17)
Number of Observations		1999			1000	
Peru						
Demographics						
Number of Children	2.51	3.15	.	3.5	4.0	.
	(1.85)	(2.05)	.	(1.9)	(2.14)	.
Number Older Siblings	.	1.61	.	.	1.9	.
	.	(1.95)	.	.	(2.13)	.
Measures of Economic Well Being						
Annual Income	.	12,394	11,691	.	12,843	11,943
	.	(15,737)	(19,782)	.	(11,980)	(18,425)
Wealth Index	0.42	0.47	0.54	0.46	0.52	0.59
	(0.19)	(0.23)	(0.21)	(0.19)	(0.23)	(0.19)
Number of Observations		2052			714	

Table 3.2: Signal to Noise Ratios for Cognition Measures

Factor	Measures	Ethiopia		Peru	
		% Signal	Loading	% Signal	Loading
Child's Cognitive Skills (Age 5)	PPVT Test	43%	1	86%	1
	CDA Test	25%	0.72	28%	0.63
Child's Cognitive Skills (Age 8, Younger Cohort)	PPVT Test	36%	1	73%	1
	Math Test	53%	1.23	42%	0.79
	Egra Test	28%	0.94	36%	0.74
Child's Cognitive Skills (Age 8,Older Cohort)	Ravens Test	-	-	20%	1
	Reading Level	59%	1	62%	1.69
	Writing Level	52%	0.94	44%	1.46
	Numeracy	12%	0.48	5%	0.48
Child's Cognitive Skills (Age 12)	PPVT Test	36%	1.13	59%	1
	Math Test	24%	1	42%	0.86
	Reading Level	15%	0.76	23%	0.63
	Writing Level	-	-	2%	-0.19
Child's Cognitive Skills (Age 15)	PPVT Test	43%	1.02	54%	1
	Math Test	36%	1	38%	0.86
	Cloze Test	39%	1.04	58%	1.03
Parental Cognitive Skill (Younger Cohort)	Mom Education	79%	1	62%	1
	Dad Education	33%	0.65	40%	0.82
	Literacy	43%	0.78	46%	0.87
Parental Cognitive Skill (Older Cohort)	Mom Education	8%	1	27%	1
	Dad Education	28%	0.70	40%	0.93
	Literacy	74%	0.79	64%	0.81

Table 3.3: Signal to Noise Ratios for Health Measures

Factor	Measures	Ethiopia		Peru	
		% Signal	Loading	% Signal	Loading
Child's Health (Age 1)	Height Z-Score	58%	1	60%	1
	Weight Z-Score	77%	1.11	60%	1.00
	Healthy? (0-2)	3%	-0.20	3%	0.24
Child's Health (Age 5)	Height Z-Score	57%	1	62%	1
	Weight Z-Score	68%	1.09	72%	1.13
	Healthy? (0-2)	1%	-0.09	3%	0.23
Child's Health (Age 8, Younger Cohort)	Height Z-Score	65%	1	73%	1
	Weight Z-Score	77%	1.09	72%	1.01
	Healthy? (1-5)	2%	0.17	2%	0.17
Child's Health (Age 8,Older Cohort)	Height Z-Score	45%	1	55%	1
	Weight Z-Score	55%	1.19	66%	1.14
	Healthy? (0-2)	1%	-0.18	1%	0.12
Child's Health (Age 12)	Height Z-Score	72%	1	60%	1
	Weight	72%	1.01	70%	1.11
	Healthy? (0-2)	1%	-0.13	2%	0.17
Child's Health (Age 15)	Height Z-Score	69%	1	56%	1
	Weight	81%	1.06	56%	1.04
	Healthy? (1-5)	0%	0.08 [19]	0%	0.09
Parental Health (Younger Cohort)	Mom Weight	92%	1	34%	1
	Mom Height	5%	0.22	45%	1.11
Parental Health (Older Cohort)	Mom Weight	32%	1	33%	1
	Mom Height	71%	0.30	66%	1.05

Table 3.4: Signal to Noise Ratios for Investment Measures

Factor	Measures	Ethiopia		Peru	
		% Signal	Loading	% Signal	Loading
Investment (Age 5)	Amount Spent on Books	18%	1	5%	0.33
	Amount Spent on Clothing	31%	0.70	60%	1
	Amount Spent on Shoes	42%	1.03	52%	0.95
	Amount Spent on Uniform	1%	0.41	18%	0.54
	Meals in Day	2%	0.22	1%	0.14
	Food Groups in Day	9%	0.47	3%	0.21
Investment (Age 8)	Amount Spent on Books	17%	1.61	38%	0.88
	Amount Spent on Clothing	4%	1	28%	1
	Amount Spent on Shoes	58%	1.72	51%	1.16
	Amount Spent on Uniform	14%	0.75	10%	0.56
	Meals in Day	4%	0.45	2%	0.23
	Food Groups in Day	6%	0.55	1%	0.13
Investment (Age 12)	Amount Spent on Fees	-	-	6%	1
	Amount Spent on Books	11%	1	41%	2.50
	Amount Spent on Clothing	30%	1.62	55%	2.80
	Amount Spent on Shoes	49%	2.06	38%	2.28
	Amount Spent on Uniform	-	-	5%	0.94
	Meals in Day	13%	1.03	1%	0.43
	Food Groups in Day	20%	1.24	3%	0.69
Investment (Age 15)	Amount Spent on Fees	-	-	6%	1
	Amount Spent on Books	20%	1	33%	2.16
	Amount Spent on Clothing	19%	1.01	46%	2.51
	Amount Spent on Shoes	6%	0.63	43%	2.41
	Amount Spent on Uniform	-	-	9%	1.17
	Meals in Day	4%	0.48	1%	0.31
	Food Groups in Day	6%	0.52	1%	0.34

Table 3.5: The Coefficients of the Investment Equations - Ethiopia

	Age 5	Age 8	Age 12	Age 15
Cognition		0.013 [−0.03,0.08]	0.014 [−0.02,0.07]	0.154 [−0.03,0.31]
Health	0.022 [−0.03,0.06]	0.004 [−0.03,0.04]	−0.001 [−0.03,0.07]	0.027 [−0.04,0.07]
Parental Cognition	0.133 [0.07,0.22]	0.196 [0.12,0.23]	0.04 [0.01,0.12]	0.094 [−0.02,0.18]
Parental Health	0.066 [0.02,0.12]	0.052 [0.02,0.1]	0.012 [0,0.05]	0.008 [−0.02,0.07]
Price Clothes	0.164 [0.02,0.29]	0.132 [0.06,0.18]	−0.063 [−0.27,0.12]	0.081 [−0.05,0.17]
Price Notebook	0.045 [−0.35,0.44]	−0.193 [−0.29,−0.05]	0.037 [−0.17,0.44]	0.133 [−0.09,0.34]
Price Mebendazol	−0.007 [−0.04,0.03]	−0.055 [−0.16,0.03]	−0.044 [−0.09,0.01]	−0.358 [−0.47,−0.2]
Price Amoxicillin	0.038 [−0.03,0.09]	−0.208 [−0.28,−0.16]	0.049 [−0.05,0.14]	−0.056 [−0.19,0.06]
Price Food	0.01 [−0.03,0.07]	−0.044 [−0.09,0.01]	0.013 [−0.09,0.09]	−0.024 [−0.1,0.02]
Older Siblings	0.053 [−0.02,0.1]	0.033 [−0.02,0.08]	−0.032 [−0.07,0.01]	0.033 [−0.02,0.07]
Number of Children	−0.076 [−0.12,0]	0.005 [−0.04,0.07]	0.054 [0.01,0.09]	−0.006 [−0.04,0.02]
Gender	0.16 [0.07,0.23]	−0.059 [−0.14,0.03]	0.06 [−0.04,0.17]	−0.033 [−0.14,0.07]
Income	0.644 [0.46,0.83]	0.284 [0.11,0.45]	0.371 [0.14,0.48]	0.367 [0.12,0.47]
Prices and Income (P-values)	0	0	.0010	.0003
Prices (P-values)	.2671	0	.5772	.0003

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

Table 3.6: The Coefficients of the Investment Equations - Peru

	Age 5	Age 8	Age 12	Age 15
Cognition	–	−0.011 [−0.09,0.03]	0.088 [−0.06,0.17]	−0.021 [−0.08,0.04]
Health	0.007 [−0.1,0.03]	0.082 [0.01,0.13]	−0.051 [−0.09,0.02]	−0.024 [−0.08,0.05]
Parental Cognition	0.094 [−0.02,0.18]	0.088 [−0.01,0.15]	−0.048 [−0.14,0.11]	0.018 [−0.09,0.17]
Parental Health	0.072 [0,0.24]	−0.002 [−0.12,0.11]	0.067 [−0.03,0.12]	0.002 [−0.09,0.08]
Price Clothes	−0.003 [−0.16,0.12]	0.099 [−0.01,0.21]	0.304 [0.14,0.43]	0.163 [−0.01,0.29]
Price Notebook	0.217 [−0.05,0.51]	−0.315 [−0.48,0.03]	−0.84 [−1.73,−0.36]	−0.356 [−0.89,0.34]
Price Mebendazol	−0.069 [−0.18,−0.02]	−0.05 [−0.12,−0.01]	−0.075 [−0.2,0.06]	−0.153 [−0.25,−0.05]
Price Amoxicillin	0.053 [−0.01,0.17]	0.058 [−0.03,0.12]	−0.01 [−0.18,0.14]	−0.037 [−0.23,0.08]
Price Food	0.567 [0,1.07]	1.103 [0.73,1.5]	−0.621 [−1.4,0.63]	0.251 [−0.32,0.77]
Older Siblings	0.012 [−0.05,0.06]	0.02 [−0.06,0.05]	0.052 [0,0.09]	0.055 [0.01,0.1]
Number of Children	−0.087 [−0.13,−0.03]	−0.058 [−0.09,0.01]	−0.104 [−0.15,−0.05]	−0.081 [−0.12,−0.04]
Gender	0.009 [−0.08,0.08]	−0.042 [−0.1,0.04]	−0.127 [−0.24,0.03]	−0.183 [−0.29,−0.05]
Income	0.611 [0.48,0.79]	0.567 [0.47,0.85]	0.132 [0.01,0.37]	0.219 [0.07,0.39]
Prices and Income (P-values)	.0001	0	.0113	.0498
Prices (P-values)	.1753	0	.0112	.0354

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

Table 3.7: Production of Human Capital, CES Production Function, Ethiopia

	Cognition				Health			
Age	5	8	12	15	5	8	12	15
Cognition		0.354 [0.26,0.46]	0.468 [0.33,0.57]	0.801 [0.56,0.87]		0.003 [−0.05,0.05]	0.257 [0.13,0.32]	0.014 [−0.07,0.11]
Health	0.09 [0.02,0.18]	0.077 [0.03,0.15]	0.043 [−0.04,0.13]	0.029 [−0.01,0.08]	0.58 [0.52,0.64]	1 [0.92,1.05]	0.893 [0.82,0.99]	0.875 [0.81,0.92]
Parental Cognition	0.024 [−0.06,0.16]	0.02 [−0.04,0.15]	0.017 [−0.03,0.09]	−0.02 [−0.07,0.05]	−0.117 [−0.2,−0.02]	0.007 [−0.04,0.04]	−0.061 [−0.1,0.01]	−0.016 [−0.08,0.02]
Parental Health	−0.003 [−0.04,0.05]	−0.04 [−0.09,0.01]	−0.02 [−0.07,0.01]	−0.016 [−0.04,0.02]	0.13 [0.1,0.25]	−0.015 [−0.03,0.02]	0.032 [0,0.08]	0.033 [0,0.07]
Investment	0.89 [0.65,1.04]	0.59 [0.35,0.7]	0.492 [0.39,0.72]	0.206 [0.06,0.47]	0.408 [0.23,0.53]	0.005 [−0.1,0.14]	−0.121 [−0.26,0.05]	0.094 [0.01,0.24]
Complementarity (ρ)	−0.563 [−0.79,0.72]	0.082 [−0.05,0.47]	−0.54 [−0.76,−0.09]	−0.042 [−0.26,0.6]	−0.244 [−0.37,−0.05]	−0.095 [−0.61,0.49]	0.461 [0.24,0.95]	−0.443 [−0.71,−0.04]
Elasticity of Substitution	0.64 [0.54,3.23]	1.089 [0.95,1.87]	0.649 [0.57,0.92]	0.96 [0.8,2.51]	0.804 [0.73,0.96]	0.913 [0.62,1.97]	1.854 [1.24,4.03]	0.693 [0.59,0.96]
Log TFP ($\log(A_t)$)	0.045 [−0.05,0.05]	0.031 [−0.01,0.06]	0.024 [−0.06,0.07]	−0.049 [−0.12,−0.01]	0.077 [0.04,0.11]	0.004 [−0.03,0.02]	0.011 [−0.06,0.05]	0.133 [0.1,0.17]
Investment Residual (υ)	−0.737 [−0.9,−0.51]	−0.552 [−0.71,−0.27]	−0.497 [−0.78,−0.23]	−0.227 [−0.56,−0.1]	−0.345 [−0.5,−0.17]	−0.081 [−0.2,0.04]	0.268 [−0.19,0.48]	−0.158 [−0.38,0.09]
Number of Children	−0.017 [−0.06,0.01]	−0.044 [−0.07,−0.02]	0.019 [0,0.05]	−0.027 [−0.05,−0.01]	−0.046 [−0.1,−0.03]	−0.007 [−0.03,0.01]	−0.001 [−0.02,0.01]	0 [−0.01,0.02]
Older Siblings	0.018 [0,0.06]	0.02 [0,0.05]	−0.014 [−0.04,0.01]	0.011 [−0.01,0.05]	0.024 [0.01,0.07]	0.012 [0,0.04]	−0.01 [−0.02,0.02]	0.003 [−0.01,0.02]
Gender	−0.003 [−0.01,0.01]	0.004 [−0.01,0.02]	0.007 [−0.02,0.03]	0.044 [0.03,0.07]	−0.006 [−0.02,0.01]	−0.006 [−0.01,0]	−0.011 [−0.02,0.01]	−0.07 [−0.09,−0.06]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

Table 3.8: Production of Human Capital, CES Production Function, Peru

	Cognition				Health			
Age	5	8	12	15	5	8	12	15
Cognition		0.444 [0.35,0.58]	0.644 [0.42,0.74]	0.871 [0.49,0.96]		−0.047 [−0.08,−0.01]	−0.041 [−0.12,0.06]	−0.038 [−0.09,0.03]
Health	0.009 [−0.06,0.09]	0.046 [−0.03,0.11]	0.084 [0,0.16]	0.04 [−0.01,0.18]	0.682 [0.62,0.74]	1.088 [1,1.11]	0.988 [0.88,1.03]	0.954 [0.86,1]
Parental Cognition	−0.025 [−0.11,0.24]	0.183 [0.1,0.27]	0.297 [0.12,0.37]	0.097 [0.06,0.39]	0.045 [−0.02,0.13]	0.019 [−0.03,0.04]	0.077 [−0.02,0.11]	0.006 [−0.02,0.08]
Parental Health	−0.114 [−0.27,0.04]	−0.066 [−0.17,0.01]	−0.106 [−0.28,0]	−0.018 [−0.15,0.03]	0.18 [0.11,0.24]	−0.046 [−0.07,0.01]	0.006 [−0.09,0.08]	0.053 [0.02,0.17]
Investment	1.129 [0.79,1.25]	0.394 [0.26,0.54]	0.081 [−0.06,0.46]	0.01 [−0.1,0.16]	0.093 [−0.05,0.2]	−0.014 [−0.08,0.1]	−0.03 [−0.11,0.19]	0.025 [−0.1,0.09]
Complementarity (ρ)	−0.23 [−0.63,0.97]	−0.019 [−0.14,0.17]	0.117 [−0.15,0.44]	−0.097 [−0.69,0.15]	0.008 [−0.23,0.18]	−0.027 [−0.31,0.24]	0.134 [−0.23,0.4]	−0.517 [−0.87,0.24]
Elasticity of Substitution	0.813 [0.42,1.61]	0.981 [0.88,1.21]	1.133 [0.87,1.78]	0.911 [0.59,1.17]	1.008 [0.81,1.23]	0.973 [0.76,1.32]	1.155 [0.81,1.66]	0.659 [0.54,1.31]
Log TFP ($\log(A_t)$)	−0.045 [−0.11,0]	−0.02 [−0.05,0.01]	0.051 [−0.01,0.11]	−0.068 [−0.11,−0.03]	−0.022 [−0.05,0]	0.017 [0.01,0.04]	−0.037 [−0.07,0]	−0.096 [−0.12,−0.06]
Investment Residual (υ)	−1.161 [−1.27,−0.8]	−0.351 [−0.51,−0.2]	−0.388 [−0.69,−0.08]	0.11 [−0.16,0.32]	−0.022 [−0.13,0.17]	0.02 [−0.1,0.13]	−0.017 [−0.25,0.13]	−0.126 [−0.22,0.13]
Number of Children	0.031 [−0.01,0.05]	−0.016 [−0.04,0.01]	−0.034 [−0.07,0]	0.008 [−0.02,0.02]	−0.011 [−0.02,0.01]	−0.017 [−0.03,0]	−0.013 [−0.03,0.01]	0.018 [0,0.03]
Older Siblings	−0.014 [−0.03,0.02]	0.017 [−0.01,0.04]	−0.004 [−0.03,0.02]	0.021 [0,0.05]	0.002 [−0.02,0.01]	0.013 [0,0.03]	0.024 [0,0.04]	−0.009 [−0.02,0.01]
Gender	0 [−0.01,0.02]	0.015 [0,0.03]	0.005 [−0.01,0.03]	0.011 [0,0.03]	0.022 [0.01,0.03]	−0.007 [−0.02,0]	0.009 [−0.01,0.02]	0.048 [0.03,0.06]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

Table 3.9: Production of Human Capital, Nested CES Production Function, Ethiopia

Age	8	12	15
Cognition	0.892 [0.71,0.99]	0.947 [0.74,1]	0.967 [0.9,1.01]
Health	0.108 [0.01,0.29]	0.053 [0,0.26]	0.033 [−0.01,0.1]
Parental Cognition	0.025 [−0.04,0.13]	0.019 [−0.04,0.09]	−0.02 [−0.07,0.04]
Parental Health	−0.029 [−0.09,0.01]	−0.023 [−0.07,0.01]	−0.016 [−0.04,0.02]
Investment	0.563 [0.28,0.69]	0.497 [0.39,0.71]	0.206 [0.06,0.47]
Coefficient on Nested Skills	0.44 [0.35,0.6]	0.507 [0.36,0.58]	0.83 [0.59,0.92]
Complementarity(ρ)	0.305 [0.08,1.07]	−0.42 [−0.68,0.02]	0.01 [−0.27,0.81]
Elasticity of Substitution	1.439 [−0.18,4.94]	0.704 [0.59,1.02]	1.01 [0.63,3.97]
Complementarity Nested (ρ^{skills})	−1.191 [−3.18,−0.24]	−1.273 [−3.18,−0.06]	−0.435 [−1.16,0.44]
Elasticity of Substitution Nested	0.456 [0.24,0.81]	0.44 [0.24,0.95]	0.697 [0.44,1.6]
Log TFP ($\log(A_t)$)	0.045 [−0.01,0.07]	0.026 [−0.05,0.07]	−0.046 [−0.12,−0.01]
Investment Residual (υ)	−0.521 [−0.68,−0.23]	−0.484 [−0.75,−0.22]	−0.227 [−0.56,−0.09]
Number of Children	−0.043 [−0.07,−0.02]	0.019 [0,0.05]	−0.027 [−0.05,−0.01]
Older Siblings	0.02 [0,0.05]	−0.014 [−0.04,0.01]	0.011 [−0.01,0.05]
Gender	0.003 [−0.01,0.02]	0.007 [−0.02,0.03]	0.044 [0.03,0.07]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

Table 3.10: Production of Human Capital, Nested CES Production Function, Peru

Age	8	12	15
Cognition	0.913 [0.81,1.06]	0.902 [0.75,1.01]	0.958 [0.75,1.01]
Health	0.087 [−0.06,0.19]	0.098 [−0.01,0.25]	0.042 [−0.01,0.25]
Parental Cognition	0.181 [0.1,0.27]	0.296 [0.12,0.37]	0.097 [0.06,0.38]
Parental Health	−0.066 [−0.17,0.01]	−0.108 [−0.28,0]	−0.018 [−0.15,0.03]
Investment	0.394 [0.25,0.54]	0.091 [−0.06,0.47]	0.013 [−0.1,0.16]
Coefficient on Nested Skills	0.491 [0.4,0.62]	0.72 [0.47,0.83]	0.908 [0.61,1]
Complementarity(ρ)	−0.053 [−0.16,0.15]	0.065 [−0.21,0.37]	−0.012 [−0.78,0.2]
Elasticity of Substitution	0.95 [0.86,1.18]	1.069 [0.82,1.59]	0.988 [0.55,1.21]
Complementarity Nested (ρ^{skills})	0.517 [−0.65,1.28]	0.708 [−0.77,2.2]	−0.261 [−0.73,0.61]
Elasticity of Substitution Nested	2.071 [−2.16,4.3]	3.428 [−4.85,6.67]	0.793 [0.54,1.94]
Log TFP ($\log(A_t)$)	−0.025 [−0.05,0.01]	0.041 [−0.02,0.1]	−0.067 [−0.11,−0.03]
Investment Residual (υ)	−0.352 [−0.51,−0.2]	−0.398 [−0.7,−0.09]	0.108 [−0.16,0.32]
Number of Children	−0.016 [−0.04,0.01]	−0.033 [−0.07,0]	0.008 [−0.02,0.02]
Older Siblings	0.017 [−0.01,0.04]	−0.004 [−0.03,0.02]	0.021 [0,0.05]
Gender	0.015 [0,0.03]	0.005 [−0.02,0.03]	0.011 [0,0.03]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

Table 3.11: Test of the Nested CES

	Ethiopia	Peru
	Cognition	Cognition
Age 8	1.496	0.57
	[0.346 , 3.626]	[0.021 , 1.384]
Age 12	0.853	0.644
	[0.133 , 2.67]	[0.077 , 2.225]
Age 15	0.445	0.249
	[0.044 , 1.555]	[0.025 , 1.279]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

3.9 Figures

Figure 3.1: Marginal Product of Investment in Ethiopia

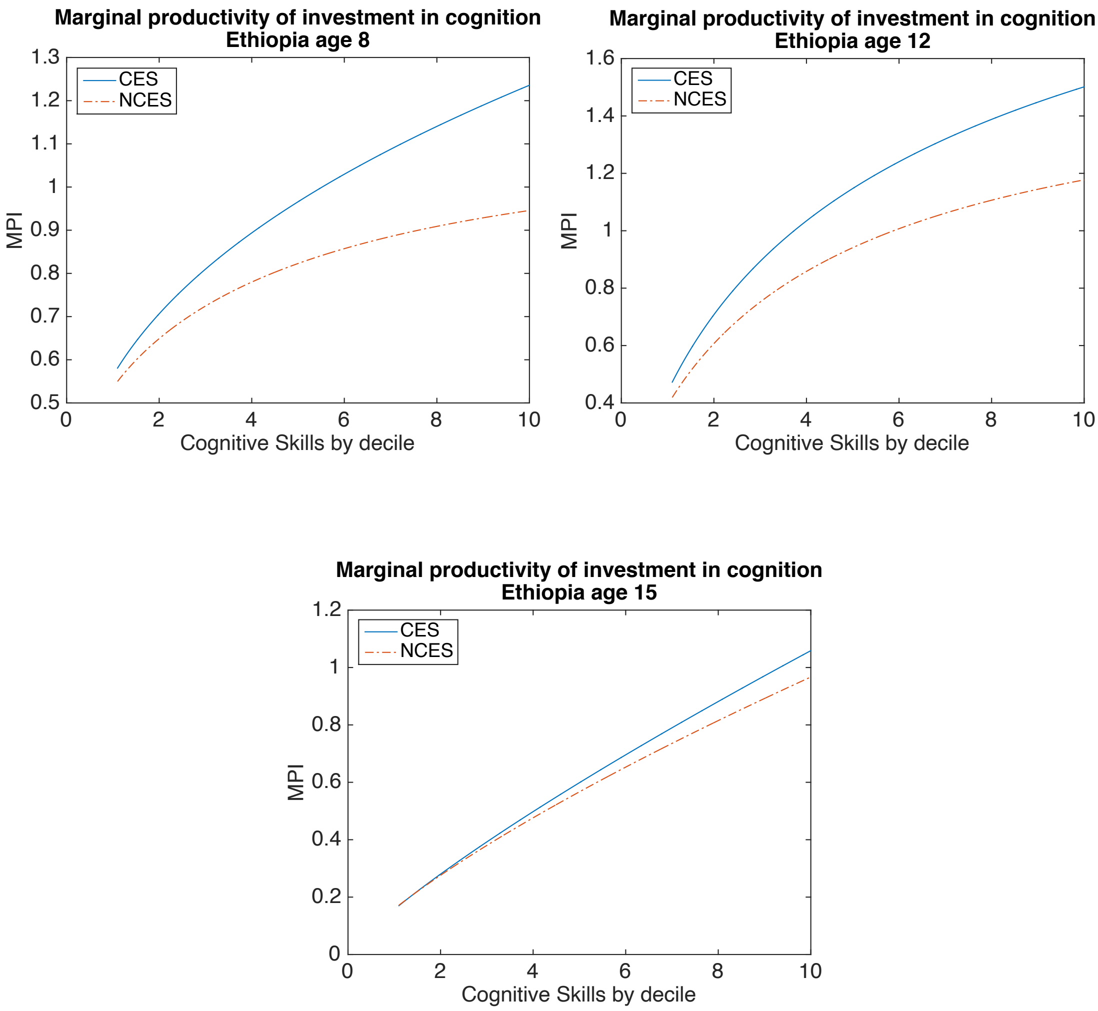


Figure 3.2: Marginal Product of Investment in Peru

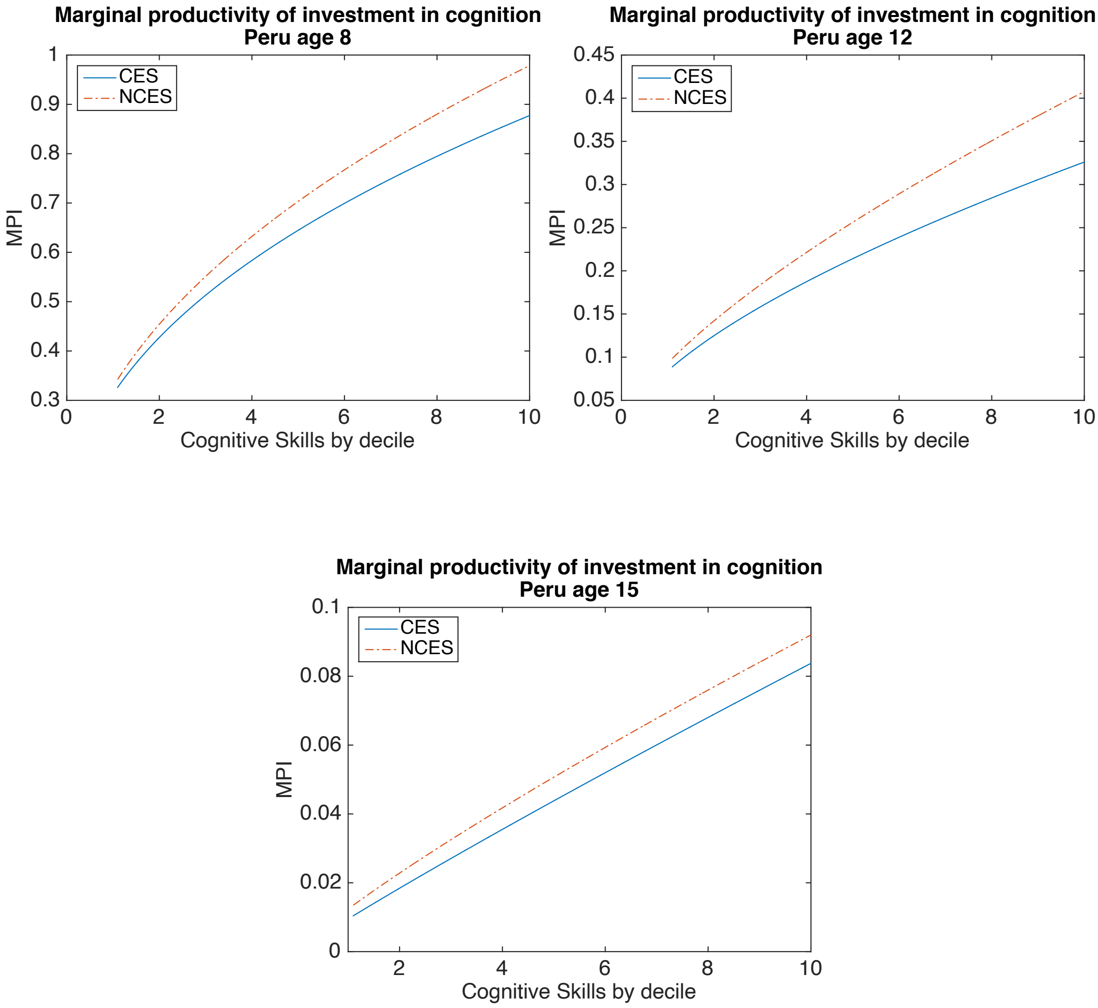


Figure 3.3

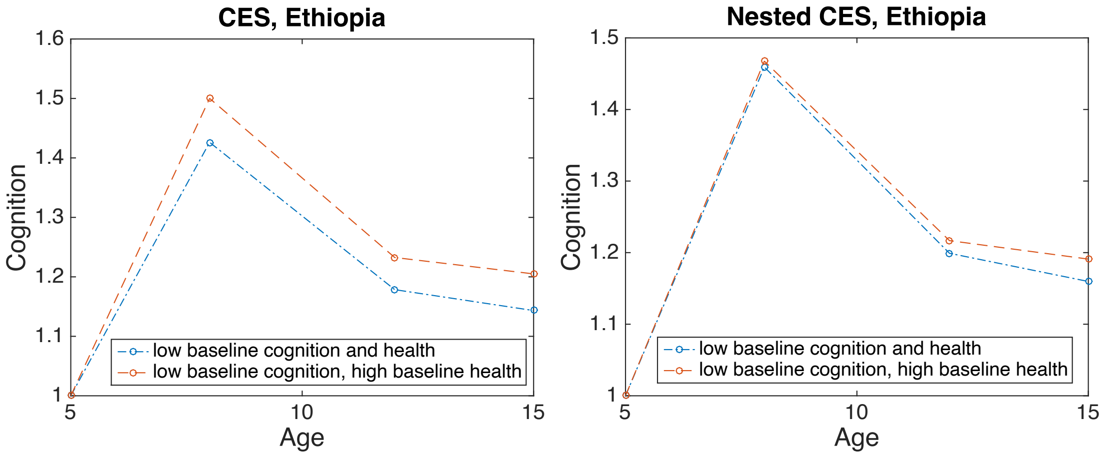


Figure 3.4

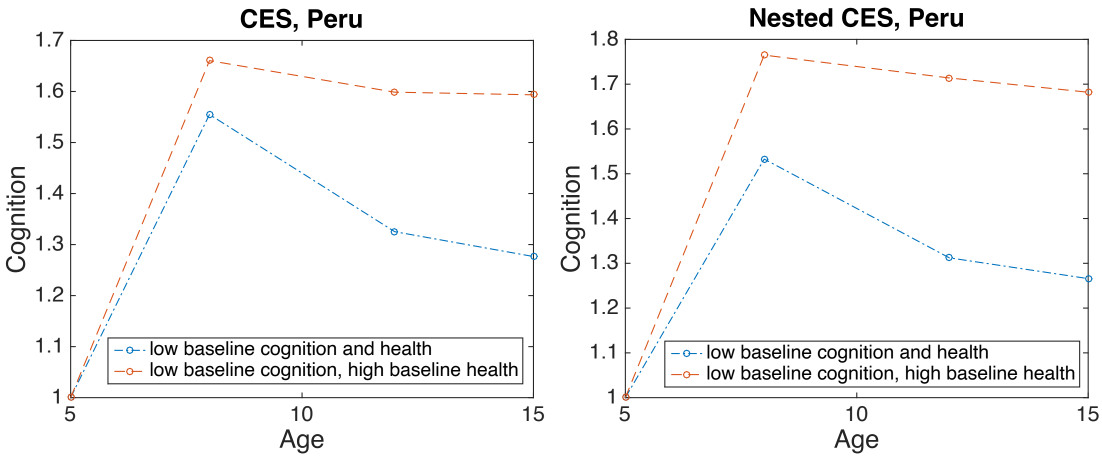
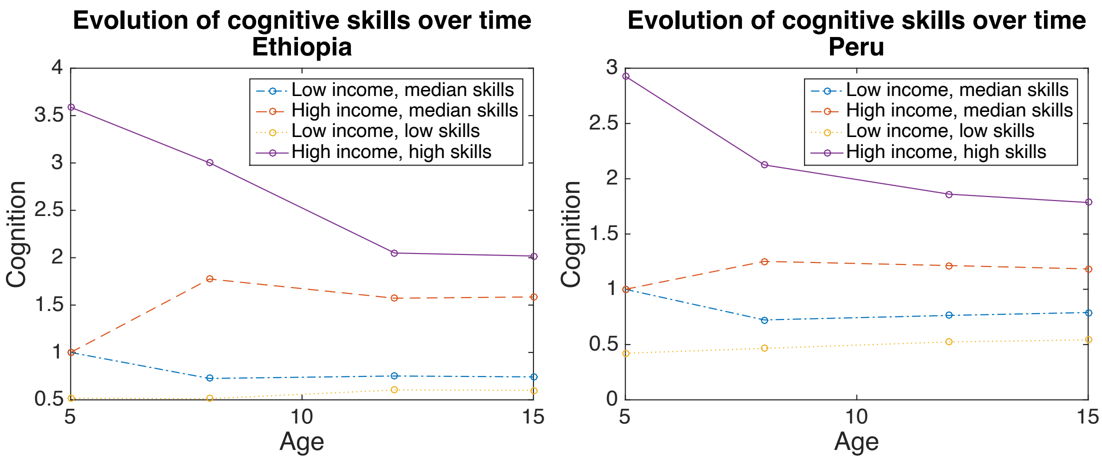


Figure 3.5



3.10 Appendix

3.10.1 Summary Statistics

Table 3.12: Summary Statistics: Child Measurements Ethiopia

	Younger Cohort			Older Cohort		
	Age 1	Age 5	Age 8	Age 8	Age 12	Age 15
Gender (male)	0.53	0.53	0.53	0.51	0.51	0.51
Health Measures						
Height for Age Z-Score	-1.58	-1.45	-1.21	-1.48	-1.39	-1.37
	(1.96)	(1.13)	(1.07)	(1.29)	(1.27)	(1.29)
Weight for Age Z-Score	-1.43	-1.36	-1.63	-2.03	.	.
	(1.52)	(0.93)	(0.94)	(1.34)	.	.
Weight in kg	7.96	15.68	20.49	19.13	30.37	40.47
	(1.47)	(2.02)	(2.89)	(3.37)	(5.83)	(8.07)
How is Child's Health? (0-2)	1.86	1.55	.	1.23	1.23	.
	(0.78)	(0.66)	.	(0.68)	(0.62)	.
How is Child's Health? (1-5)	.	.	4	.	.	4.04
	.	.	(0.87)	.	.	(0.86)
Cognitive Measures						
Number Correct PPVT Test	.	21.42	68.35	.	75.87	124.27
	.	(12.39)	(36.77)	.	(26.16)	(28.36)
Rasch Score Math Test	.	.	300	.	300.00	300.00
	.	.	(14.99)	.	(49.94)	(14.98)
Rasch Score CDA Test	.	300	.	.	.	.
	.	(49.97)	.	.	.	.
Rasch Score Egra Test	.	.	300	.	.	.
	.	.	(14.99)	.	.	.
Rasch Score Cloze Test	.	.	.	.	.	300.00
	.	.	.	.	.	(14.98)
Ravens Total Correct (0-36)	.	.	.	16.79	.	.
	.	.	.	(6.31)	.	.
Child's Reading Level (1-4)	.	.	.	1.94	3.26	.
	.	.	.	(1.20)	(1.04)	.
Child's Writing Level (0-2)	.	.	.	0.66	1.21	.
	.	.	.	(0.83)	(0.63)	.
What is 2x4? (1 if correct)	.	.	.	0.44	.	.
Observations		1,999			1000	

Table 3.13: Summary Statistics: Child Measurements Peru

	Younger Cohort			Older Cohort		
	Age 1	Age 5	Age 8	Age 8	Age 12	Age 15
Gender (male)	0.50	0.50	0.50	0.54	0.54	0.54
Health Measures						
Height for Age Z-Score	-1.30	-1.56	-1.17	-1.41	-1.53	-1.48
	(1.36)	(1.16)	(1.06)	(1.01)	(1.15)	(0.90)
Weight for Age Z-Score	-0.20	-0.54	-0.34	-0.50	.	.
	(1.21)	(1.25)	(1.20)	(0.96)	.	.
Weight in kg	9.10	17.84	24.46	23.74	38.60	50.40
	(1.42)	(3.02)	(4.96)	(3.67)	(8.57)	(8.82)
How is Child's Health? (0-2)	1.26	1.03	.	1.22	1.23	.
	(0.68)	(0.58)	.	(0.62)	(0.57)	.
How is Child's Health? (1-5)	.	.	3.71	.	.	3.75
	.	.	(0.65)	.	.	(0.67)
Cognitive Measures						
Rasch Score PPVT Test	.	300.00	300.00	.	300.00	300.00
	.	(50.00)	(15.01)	.	(50.00)	(15.00)
Rasch Score Math Test	.	.	300.03	.	300.00	300.00
	.	.	(14.98)	.	(50.00)	(15.00)
Rasch Score CDA Test	.	300.00	.	.	.	.
	.	(49.99)	.	.	.	.
Rasch Score Egra Test	.	.	300.01	.	.	.
	.	.	(15.01)	.	.	.
Rasch Score Cloze Test	.	.	.	.	.	300.00
	.	.	.	.	.	(15.00)
Ravens Total Correct (0-36)	.	.	.	20.82	.	.
	.	.	.	(8.06)	.	.
Child's Reading Level (1-4)	.	.	.	3.59	3.93	.
	.	.	.	(0.96)	(0.39)	.
Child's Writing Level (0-2)	.	.	.	1.42	1.12	.
	.	.	.	(0.71)	(0.36)	.
What is 2x4? (1 if correct)	.	.	.	0.75	.	.
Observations		2,052			714	

3.10.2 Estimates of the Nested CES for Health at Age 8 for Peru and Age 12 for Ethiopia

Table 3.14: Production of Human Capital, Nested CES Production Function, Health

	Health	
	Peru, Age 8	Ethiopia, Age 12
Cognition	−0.046 [−0.09,−0.01]	0.201 [0.11,0.24]
Health	1.046 [1.01,1.09]	0.799 [0.76,0.89]
Parental Cognition	0.019 [−0.04,0.04]	−0.089 [−0.13,0]
Parental Health	−0.047 [−0.07,0.01]	0.043 [0,0.09]
Investment	−0.012 [−0.08,0.1]	−0.125 [−0.26,0.07]
Coefficient on Nested Skills	1.04 [0.93,1.09]	1.17 [0.98,1.27]
Complementarity (ρ)	0.025 [−0.78,0.56]	−0.045 [−0.41,0.3]
Elasticity of Substitution	1.026 [0.55,2.12]	0.957 [0.71,1.43]
Complementarity Nested (ρ^{skills})	−0.055 [−0.42,0.14]	0.735 [0.43,1.08]
Elasticity of Substitution Nested	0.948 [0.7,1.14]	3.771 [−11.96,11.06]
log TFP ($\log(A_t)$)	0.017 [0.01,0.04]	−0.023 [−0.07,0.02]
Investment Residual (υ)	0.019 [−0.1,0.13]	0.254 [−0.24,0.48]
Number of Children	−0.017 [−0.03,0]	−0.004 [−0.02,0.01]
Older Siblings	0.013 [0,0.03]	−0.011 [−0.02,0.02]
Gender	−0.007 [−0.02,0]	−0.01 [−0.02,0.01]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

3.10.3 Estimates without Instruments for Investments

Table 3.15: Production of Human Capital, CES Production Function, Ethiopia

	Cognition				Health			
Age	5	8	12	15	5	8	12	15
Cognition		0.438 [0.32,0.54]	0.502 [0.37,0.6]	0.871 [0.7,0.93]		0.015 [−0.03,0.05]	0.237 [0.13,0.29]	0.051 [−0.02,0.12]
Health	0.215 [0.13,0.26]	0.145 [0.08,0.21]	0.101 [0.02,0.22]	0.048 [0,0.1]	0.635 [0.56,0.68]	1.009 [0.94,1.05]	0.865 [0.81,0.96]	0.885 [0.82,0.93]
Parental Cognition	0.314 [0.22,0.41]	0.103 [0.05,0.19]	0.051 [0.02,0.12]	0.009 [−0.04,0.07]	0.014 [−0.04,0.06]	0.019 [−0.03,0.05]	−0.063 [−0.11,0]	0 [−0.05,0.03]
Parental Health	0.087 [0.04,0.16]	−0.002 [−0.04,0.04]	−0.007 [−0.03,0.03]	−0.003 [−0.03,0.04]	0.17 [0.13,0.29]	−0.009 [−0.02,0.02]	0.022 [−0.01,0.07]	0.04 [0.01,0.07]
Investment	0.383 [0.21,0.56]	0.316 [0.2,0.43]	0.353 [0.23,0.47]	0.075 [0,0.24]	0.181 [0.08,0.24]	−0.035 [−0.1,0.05]	−0.06 [−0.2,0.06]	0.024 [−0.06,0.11]
Complementarity (ρ)	−0.278 [−0.52,0.06]	0.013 [−0.18,0.46]	−0.514 [−0.71,−0.1]	−0.186 [−0.42,0.62]	−0.365 [−0.44,−0.12]	−0.094 [−0.5,0.52]	0.516 [0.34,0.89]	−0.435 [−0.74,0.16]
Elasticity of Substitution	0.783 [0.66,1.06]	1.014 [0.85,1.85]	0.661 [0.59,0.91]	0.843 [0.71,2.62]	0.732 [0.69,0.89]	0.914 [0.67,2.07]	2.068 [1.47,5.85]	0.697 [0.58,1.19]
log TFP ($\log(A_t)$)	0.026 [−0.05,0.05]	0.013 [−0.03,0.04]	0.019 [−0.06,0.07]	−0.05 [−0.13,−0.02]	0.081 [0.03,0.12]	0.001 [−0.03,0.02]	0.012 [−0.05,0.05]	0.13 [0.1,0.18]
Number of Children	−0.024 [−0.05,0]	−0.025 [−0.05,−0.01]	0.028 [0.01,0.06]	−0.027 [−0.05,−0.01]	−0.048 [−0.09,−0.03]	−0.004 [−0.03,0.01]	−0.004 [−0.02,0.01]	0.001 [−0.01,0.02]
Older Siblings	0.027 [0.01,0.06]	0.016 [0,0.04]	−0.017 [−0.05,0.01]	0.015 [0,0.05]	0.027 [0.01,0.07]	0.012 [0,0.04]	−0.008 [−0.02,0.02]	0.005 [0,0.02]
Gender	0.013 [0,0.02]	0 [−0.01,0.01]	0.009 [−0.01,0.03]	0.044 [0.02,0.07]	0.001 [−0.01,0.01]	−0.007 [−0.02,0]	−0.011 [−0.02,0.01]	−0.071 [−0.09,−0.06]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

Table 3.16: Production of Human Capital, CES Production Function, Peru

	Cognition				Health			
Age	5	8	12	15	5	8	12	15
Cognition		0.497 [0.41,0.62]	0.694 [0.5,0.77]	0.861 [0.49,0.95]		−0.05 [−0.07,−0.02]	−0.039 [−0.1,0.05]	−0.036 [−0.08,0.02]
Health	0.115 [0.01,0.15]	0.078 [0.01,0.14]	0.091 [0.02,0.17]	0.037 [−0.02,0.17]	0.684 [0.63,0.73]	1.086 [1.01,1.11]	0.988 [0.88,1.03]	0.954 [0.86,0.99]
Parental Cognition	0.479 [0.37,0.64]	0.272 [0.18,0.37]	0.316 [0.17,0.37]	0.086 [0.05,0.38]	0.055 [0.01,0.11]	0.014 [−0.02,0.03]	0.078 [0,0.12]	0.022 [−0.03,0.07]
Parental Health	0.17 [0.03,0.29]	0.007 [−0.08,0.08]	−0.065 [−0.19,0.03]	−0.036 [−0.16,0.02]	0.185 [0.11,0.24]	−0.051 [−0.08,0.01]	0.008 [−0.06,0.09]	0.077 [0.02,0.17]
Investment	0.236 [0.14,0.41]	0.146 [0.07,0.22]	−0.036 [−0.09,0.25]	0.052 [−0.05,0.16]	0.076 [0.02,0.14]	0.001 [−0.02,0.06]	−0.036 [−0.1,0.08]	−0.016 [−0.07,0.05]
Complementarity (ρ)	0.141 [−0.06,1.56]	0.069 [−0.11,0.22]	0.205 [−0.06,0.48]	−0.036 [−0.48,0.25]	−0.001 [−0.18,0.18]	0.011 [−0.23,0.29]	0.14 [−0.25,0.42]	−0.334 [−0.82,0.24]
Elasticity of Substitution	1.164 [−1.94,4.34]	1.074 [0.9,1.28]	1.259 [0.94,1.95]	0.966 [0.68,1.33]	0.999 [0.85,1.23]	1.012 [0.81,1.41]	1.163 [0.8,1.73]	0.75 [0.55,1.32]
log TFP ($\log(A_t)$)	−0.069 [−0.15,−0.04]	−0.031 [−0.06,0]	0.051 [−0.01,0.11]	−0.068 [−0.11,−0.03]	−0.022 [−0.05,0]	0.018 [0.01,0.04]	−0.037 [−0.07,0]	−0.099 [−0.12,−0.07]
Number of Children	0.019 [−0.01,0.03]	−0.019 [−0.05,0]	−0.037 [−0.07,−0.01]	0.007 [−0.02,0.02]	−0.012 [−0.02,0]	−0.017 [−0.03,0]	−0.013 [−0.03,0]	0.02 [0,0.03]
Older Siblings	−0.002 [−0.01,0.03]	0.021 [0,0.05]	0 [−0.03,0.02]	0.02 [0,0.05]	0.002 [−0.02,0.01]	0.013 [0,0.03]	0.025 [0.01,0.04]	−0.007 [−0.02,0.01]
Gender	0.011 [0,0.03]	0.015 [0,0.02]	0.001 [−0.02,0.03]	0.013 [0,0.04]	0.022 [0.01,0.03]	−0.007 [−0.02,0]	0.009 [−0.01,0.02]	0.046 [0.03,0.06]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

Table 3.17: Production of Human Capital, Nested CES Production Function, Ethiopia

	Cognition			Health		
Age	8	12	15	8	12	15
Cognition	0.827 [0.7,0.95]	0.882 [0.68,0.99]	0.949 [0.88,0.99]	0.015 [−0.03,0.05]	0.19 [0.1,0.23]	0.06 [−0.02,0.13]
Health	0.173 [0.05,0.3]	0.118 [0.01,0.32]	0.051 [0.01,0.12]	0.985 [0.95,1.03]	0.81 [0.77,0.9]	0.94 [0.87,1.02]
Parental Cognition	0.084 [0.05,0.16]	0.054 [0.02,0.12]	0.01 [−0.04,0.07]	0.018 [−0.03,0.05]	−0.106 [−0.15,−0.01]	0.003 [−0.04,0.03]
Parental Health	0.006 [−0.03,0.03]	−0.008 [−0.04,0.03]	−0.004 [−0.03,0.04]	−0.009 [−0.02,0.02]	0.033 [−0.01,0.09]	0.021 [0,0.07]
Investment	0.292 [0.15,0.4]	0.369 [0.25,0.47]	0.076 [0.01,0.24]	−0.035 [−0.09,0.05]	−0.053 [−0.19,0.06]	0.021 [−0.06,0.11]
Coefficient on Nested Skills	0.618 [0.52,0.74]	0.586 [0.47,0.69]	0.919 [0.77,0.97]	1.025 [0.94,1.07]	1.126 [0.99,1.2]	0.955 [0.87,1.02]
Complementarity (ρ)	0.618 [0.22,1.2]	−0.2 [−0.59,0.24]	−0.027 [−0.55,1.39]	−0.156 [−0.64,0.66]	−0.064 [−0.49,0.31]	−0.898 [−1.1,0.34]
Elasticity of Substitution	2.617 [−5.66,6.6]	0.833 [0.63,1.31]	0.974 [−1.95,2.38]	0.865 [0.55,2.55]	0.94 [0.66,1.38]	0.527 [0.47,1.43]
Complementarity Nested (ρ^{skills})	−1.149 [−2.56,−0.39]	−1.204 [−2.29,−0.33]	−0.322 [−0.47,0.21]	0.088 [−1.15,0.72]	0.841 [0.52,1.13]	0.28 [−1.12,0.96]
Elasticity of Substitution Nested	0.465 [0.28,0.72]	0.454 [0.3,0.75]	0.756 [0.61,1.25]	1.097 [0.47,3.69]	6.285 [−8.71,18.75]	1.389 [0.32,7.32]
log TFP ($\log(A_t)$)	0.018 [−0.03,0.05]	0.018 [−0.06,0.07]	−0.05 [−0.12,−0.02]	−0.001 [−0.03,0.02]	−0.02 [−0.07,0.03]	0.125 [0.1,0.17]
Number of Children	−0.023 [−0.05,0]	0.027 [0.01,0.06]	−0.027 [−0.05,−0.01]	−0.004 [−0.03,0.01]	−0.008 [−0.03,0.01]	0.002 [−0.01,0.02]
Older Siblings	0.015 [0,0.04]	−0.017 [−0.05,0.01]	0.015 [0,0.05]	0.012 [0,0.04]	−0.01 [−0.02,0.02]	0.005 [0,0.02]
Gender	0 [−0.01,0.01]	0.009 [−0.01,0.03]	0.044 [0.02,0.07]	−0.007 [−0.01,0]	−0.011 [−0.02,0.01]	−0.071 [−0.09,−0.06]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

Table 3.18: Production of Human Capital, Nested CES Production Function, Peru

	Cognition			Health		
Age	8	12	15	8	12	15
Cognition	0.866 [0.76,0.97]	0.892 [0.76,0.98]	0.965 [0.75,1.02]	−0.05 [−0.07,−0.02]	−0.04 [−0.11,0.05]	−0.035 [−0.09,0.03]
Health	0.134 [0.03,0.24]	0.108 [0.02,0.24]	0.035 [−0.02,0.25]	1.05 [1.02,1.07]	1.04 [0.95,1.11]	1.035 [0.97,1.09]
Parental Cognition	0.271 [0.18,0.37]	0.317 [0.17,0.37]	0.083 [0.05,0.38]	0.014 [−0.02,0.03]	0.078 [0,0.12]	0.023 [−0.03,0.07]
Parental Health	0.008 [−0.09,0.09]	−0.065 [−0.19,0.03]	−0.034 [−0.16,0.02]	−0.051 [−0.08,0.01]	0.008 [−0.06,0.08]	0.08 [0.02,0.17]
Investment	0.146 [0.07,0.22]	−0.032 [−0.08,0.25]	0.055 [−0.05,0.16]	0.003 [−0.02,0.06]	−0.036 [−0.09,0.08]	−0.02 [−0.07,0.05]
Coefficient on Nested Skills	0.575 [0.49,0.69]	0.781 [0.57,0.88]	0.895 [0.61,0.99]	1.034 [0.96,1.07]	0.95 [0.85,1.03]	0.918 [0.83,0.97]
Complementarity (ρ)	0.049 [−0.13,0.28]	0.17 [−0.08,0.4]	0.165 [−0.84,0.31]	0.186 [−0.6,0.76]	0.136 [−0.32,0.61]	−0.252 [−0.83,0.24]
Elasticity of Substitution	1.051 [0.89,1.38]	1.205 [0.92,1.68]	1.198 [0.54,1.46]	1.228 [0.62,4.3]	1.158 [0.76,2.59]	0.799 [0.54,1.29]
Complementarity Nested (ρ^{skills})	0.176 [−0.4,0.61]	0.488 [−0.09,1.76]	−0.534 [−0.79,0.61]	−0.053 [−0.35,0.11]	0.188 [−1.1,1.03]	−0.495 [−1.03,0.54]
Elasticity of Substitution Nested	1.213 [0.71,2.57]	1.954 [−10.49,10.24]	0.652 [0.49,1.59]	0.95 [0.74,1.12]	1.231 [−0.39,2.23]	0.669 [0.49,2.17]
log TFP ($\log(A_t)$)	−0.032 [−0.06,0]	0.046 [−0.01,0.11]	−0.066 [−0.11,−0.03]	0.018 [0.01,0.04]	−0.036 [−0.07,0]	−0.102 [−0.12,−0.06]
Number of Children	−0.019 [−0.05,0]	−0.037 [−0.07,−0.01]	0.006 [−0.02,0.02]	−0.017 [−0.03,0]	−0.013 [−0.03,0]	0.02 [0,0.03]
Older Siblings	0.021 [0,0.05]	0 [−0.03,0.02]	0.02 [0,0.05]	0.013 [0,0.03]	0.025 [0.01,0.04]	−0.007 [−0.02,0.01]
Gender	0.015 [0,0.02]	0.001 [−0.02,0.03]	0.013 [0,0.04]	−0.007 [−0.02,0]	0.009 [−0.01,0.02]	0.046 [0.03,0.06]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

3.10.4 Estimates Using Only Prices as Instruments

Table 3.19: Production of Human Capital, CES Production Function, Ethiopia

	Cognition				Health			
Age	5	8	12	15	5	8	12	15
Cognition		0.353 [0.26,0.45]	0.445 [0.32,0.55]	0.79 [0.56,0.86]		−0.001 [−0.05,0.04]	0.266 [0.13,0.34]	0.016 [−0.08,0.11]
Health	0.034 [−0.01,0.13]	0.093 [0.05,0.16]	0.039 [−0.03,0.13]	0.034 [−0.01,0.08]	0.547 [0.49,0.61]	1.007 [0.92,1.06]	0.895 [0.81,0.98]	0.872 [0.79,0.91]
Parental Cognition	0.064 [−0.05,0.17]	−0.054 [−0.07,0.06]	0.006 [−0.06,0.06]	−0.029 [−0.09,0.03]	−0.097 [−0.17,−0.01]	−0.033 [−0.08,0]	−0.038 [−0.1,0.01]	−0.011 [−0.07,0.02]
Parental Health	−0.015 [−0.05,0.04]	−0.051 [−0.1,−0.01]	−0.02 [−0.07,0.01]	−0.021 [−0.05,0.03]	0.122 [0.09,0.23]	−0.018 [−0.03,0.01]	0.042 [0,0.08]	0.029 [0,0.07]
Investment	0.917 [0.68,1.04]	0.659 [0.5,0.75]	0.531 [0.42,0.75]	0.225 [0.11,0.5]	0.429 [0.26,0.53]	0.045 [−0.05,0.16]	−0.165 [−0.31,0.02]	0.094 [−0.04,0.22]
Complementarity (ρ)	0.559 [−0.65,1.08]	−0.042 [−0.15,0.17]	−0.531 [−0.82,−0.09]	−0.04 [−0.31,0.35]	−0.2 [−0.26,−0.02]	0.077 [−0.31,0.42]	0.443 [0.24,0.9]	−0.49 [−0.71,0.17]
Elasticity of Substitution	2.269 [−8.34,7.89]	0.959 [0.87,1.2]	0.653 [0.55,0.91]	0.962 [0.76,1.55]	0.834 [0.8,0.98]	1.084 [0.75,1.52]	1.796 [1.2,6.24]	0.671 [0.59,1.2]
Log TFP	0.182 [−0.02,0.26]	−0.11 [−0.21,0.02]	−0.008 [−0.13,0.06]	−0.1 [−0.25,0]	0.187 [0.06,0.25]	−0.076 [−0.13,−0.04]	0.049 [−0.06,0.12]	0.151 [0.09,0.2]
Investment Residual (υ)	−0.784 [−0.91,−0.54]	−0.656 [−0.76,−0.48]	−0.525 [−0.84,−0.24]	−0.243 [−0.61,−0.08]	−0.361 [−0.49,−0.2]	−0.138 [−0.25,−0.01]	0.295 [−0.18,0.55]	−0.157 [−0.36,0.14]
Number of Children	−0.016 [−0.06,0.01]	−0.044 [−0.08,−0.02]	0.016 [0,0.05]	−0.028 [−0.05,0]	−0.048 [−0.1,−0.03]	−0.006 [−0.03,0.01]	−0.007 [−0.03,0.01]	−0.001 [−0.01,0.01]
Older Siblings	0.017 [0,0.06]	0.018 [0,0.05]	−0.013 [−0.05,0.01]	0.019 [−0.01,0.04]	0.024 [0.01,0.07]	0.011 [0,0.03]	0.001 [−0.02,0.02]	0.004 [−0.01,0.02]
Gender	−0.006 [−0.01,0.01]	0.004 [−0.01,0.02]	0.006 [−0.01,0.03]	0.044 [0.03,0.07]	−0.007 [−0.02,0]	−0.006 [−0.01,0]	−0.007 [−0.02,0.01]	−0.074 [−0.1,−0.06]
Income	−0.144 [−0.22,0]	0.124 [0.01,0.21]	0.028 [−0.05,0.09]	0.037 [−0.03,0.16]	−0.091 [−0.14,−0.01]	0.068 [0.04,0.11]	−0.038 [−0.09,0.05]	−0.006 [−0.06,0.04]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

Table 3.20: Production of Human Capital, CES Production Function, Peru

	Cognition				Health			
Age	5	8	12	15	5	8	12	15
Cognition		0.457 [0.36,0.59]	0.681 [0.46,0.75]	0.875 [0.48,0.96]		−0.062 [−0.1,−0.02]	−0.027 [−0.1,0.08]	−0.023 [−0.08,0.04]
Health	0.041 [−0.05,0.12]	0.045 [−0.03,0.11]	0.004 [−0.09,0.12]	0.044 [−0.01,0.17]	0.702 [0.64,0.76]	1.089 [1.01,1.12]	0.964 [0.85,1.01]	0.975 [0.88,1.02]
Parental Cognition	−0.063 [−0.17,0.12]	0.207 [0.1,0.3]	0.093 [−0.01,0.24]	0.103 [0.04,0.36]	0.008 [−0.06,0.07]	−0.01 [−0.06,0.02]	0.006 [−0.08,0.06]	0.062 [0,0.13]
Parental Health	−0.064 [−0.19,0.08]	−0.086 [−0.18,−0.01]	−0.08 [−0.27,0.03]	−0.021 [−0.15,0.04]	0.21 [0.15,0.27]	−0.023 [−0.04,0.03]	0.01 [−0.06,0.08]	0.054 [0.01,0.13]
Investment	1.086 [0.8,1.25]	0.377 [0.24,0.54]	0.302 [0.14,0.53]	−0.002 [−0.09,0.19]	0.08 [−0.03,0.19]	0.007 [−0.05,0.11]	0.046 [−0.05,0.2]	−0.067 [−0.17,0.04]
Complementarity (ρ)	−0.294 [−0.97,0.76]	0.015 [−0.1,0.17]	−0.196 [−0.45,0.19]	−0.183 [−0.48,0.7]	−0.039 [−0.19,0.15]	−0.003 [−0.23,0.14]	−0.416 [−0.37,0.29]	0.264 [−0.86,0.42]
Elasticity of Substitution	0.773 [0.5,1.44]	1.015 [0.91,1.21]	0.836 [0.69,1.24]	0.845 [0.62,2]	0.963 [0.84,1.17]	0.997 [0.82,1.16]	0.706 [0.73,1.4]	1.359 [0.54,1.73]
Log TFP	−0.238 [−0.38,−0.1]	0.059 [−0.08,0.1]	−0.262 [−0.36,−0.04]	−0.051 [−0.22,0]	−0.159 [−0.24,−0.11]	−0.08 [−0.11,−0.01]	−0.152 [−0.23,−0.08]	0.01 [−0.05,0.1]
Investment Residual (υ)	−1.125 [−1.27,−0.84]	−0.323 [−0.51,−0.19]	−0.663 [−0.9,−0.19]	0.124 [−0.22,0.31]	−0.021 [−0.12,0.11]	−0.013 [−0.11,0.1]	−0.117 [−0.34,0.05]	−0.021 [−0.14,0.2]
Number of Children	0.04 [0,0.06]	−0.021 [−0.05,0.01]	−0.031 [−0.06,0]	0.008 [−0.02,0.02]	−0.004 [−0.02,0.01]	−0.011 [−0.03,0]	−0.012 [−0.02,0.01]	0.02 [0,0.03]
Older Siblings	−0.02 [−0.03,0.01]	0.02 [0,0.04]	−0.015 [−0.04,0.01]	0.022 [0,0.05]	−0.003 [−0.03,0.01]	0.009 [0,0.02]	0.02 [0,0.03]	−0.006 [−0.01,0.02]
Gender	0.001 [−0.01,0.02]	0.015 [0,0.03]	0 [−0.02,0.03]	0.01 [0,0.04]	0.022 [0.01,0.03]	−0.007 [−0.02,0]	0.008 [−0.01,0.02]	0.043 [0.03,0.06]
Income	0.153 [0.06,0.3]	−0.065 [−0.1,0.04]	0.284 [0.08,0.38]	−0.012 [−0.05,0.12]	0.11 [0.07,0.16]	0.077 [0.03,0.1]	0.104 [0.06,0.16]	−0.087 [−0.16,−0.04]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

Table 3.21: Production of Human Capital, Nested CES Production Function, Ethiopia

	Cognition			Health		
Age	8	12	15	8	12	15
Cognition	0.847 [0.69,0.98]	0.953 [0.76,1.07]	0.958 [0.9,1.02]	0 [−0.05,0.05]	0.211 [0.11,0.26]	0.033 [−0.09,0.11]
Health	0.153 [0.02,0.31]	0.047 [−0.07,0.24]	0.042 [−0.02,0.1]	1 [0.95,1.05]	0.789 [0.74,0.89]	0.967 [0.89,1.09]
Parental Cognition	−0.047 [−0.06,0.06]	0.006 [−0.06,0.06]	−0.029 [−0.09,0.03]	−0.033 [−0.08,0]	−0.057 [−0.12,0.01]	−0.005 [−0.07,0.02]
Parental Health	−0.045 [−0.1,0]	−0.023 [−0.07,0.01]	−0.021 [−0.05,0.03]	−0.018 [−0.03,0.01]	0.047 [0.01,0.1]	0.021 [0,0.07]
Investment	0.643 [0.47,0.73]	0.536 [0.42,0.73]	0.225 [0.11,0.49]	0.045 [−0.05,0.16]	−0.173 [−0.33,0.02]	0.072 [−0.07,0.22]
Coefficient on Nested Skills	0.45 [0.35,0.56]	0.481 [0.36,0.58]	0.825 [0.6,0.91]	1.007 [0.9,1.08]	1.184 [0.97,1.29]	0.912 [0.77,0.98]
Complementarity (ρ)	0.119 [−0.06,0.43]	−0.428 [−0.75,0]	−0.028 [−0.32,0.52]	0.079 [−0.33,0.53]	−0.29 [−0.44,0.27]	−0.717 [−1.12,0.42]
Elasticity of Substitution	1.135 [0.94,1.76]	0.7 [0.57,1]	0.973 [0.76,2.09]	1.086 [0.75,2.15]	0.775 [0.7,1.37]	0.583 [0.47,1.72]
Complementarity Nested (ρ^{skills})	−0.957 [−3.11,−0.16]	−1.358 [−2.73,0.17]	−0.117 [−0.76,0.74]	−3.646 [−0.82,0.75]	0.724 [0.43,1.13]	0.616 [−1,1.21]
Elasticity of Substitution Nested	0.511 [0.24,0.86]	0.424 [0.27,1.21]	0.895 [0.42,1.48]	0.215 [0.51,2.61]	3.628 [−7.63,11.86]	2.607 [−3.73,4.63]
log TFP ($\log(A_t)$)	−0.09 [−0.21,0.04]	−0.003 [−0.13,0.07]	−0.1 [−0.24,0]	−0.076 [−0.13,−0.04]	0.007 [−0.08,0.09]	0.148 [0.09,0.2]
Investment Residual (υ)	−0.636 [−0.74,−0.44]	−0.514 [−0.79,−0.24]	−0.243 [−0.6,−0.07]	−0.138 [−0.25,−0.01]	0.296 [−0.2,0.54]	−0.143 [−0.36,0.12]
Number of Children	−0.043 [−0.07,−0.02]	0.016 [0,0.05]	−0.028 [−0.05,0]	−0.006 [−0.03,0.01]	−0.009 [−0.03,0.01]	−0.001 [−0.01,0.01]
Older Siblings	0.017 [0,0.05]	−0.013 [−0.05,0.01]	0.019 [−0.01,0.04]	0.011 [0,0.03]	0 [−0.02,0.02]	0.005 [−0.01,0.02]
Gender	0.004 [−0.01,0.02]	0.006 [−0.01,0.03]	0.044 [0.03,0.07]	−0.006 [−0.01,0]	−0.005 [−0.02,0.01]	−0.074 [−0.1,−0.06]
Income	0.123 [0.01,0.21]	0.027 [−0.05,0.09]	0.037 [−0.03,0.16]	0.068 [0.04,0.11]	−0.038 [−0.09,0.05]	−0.008 [−0.06,0.03]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

Table 3.22: Production of Human Capital, Nested CES Production Function, Peru

	Cognition			Health		
Age	8	12	15	8	12	15
Cognition	0.917 [0.81,1.06]	0.993 [0.79,1.14]	0.952 [0.77,1.01]	−0.061 [−0.1,−0.02]	−0.031 [−0.11,0.08]	−0.025 [−0.09,0.05]
Health	0.083 [−0.06,0.19]	0.007 [−0.14,0.21]	0.048 [−0.01,0.23]	1.061 [1.02,1.1]	1.031 [0.92,1.11]	1.025 [0.95,1.09]
Parental Cognition	0.206 [0.1,0.3]	0.095 [−0.01,0.23]	0.103 [0.04,0.36]	−0.006 [−0.05,0.02]	0.006 [−0.08,0.06]	0.062 [0,0.13]
Parental Health	−0.085 [−0.18,−0.01]	−0.086 [−0.27,0.03]	−0.021 [−0.15,0.04]	−0.024 [−0.04,0.03]	0.011 [−0.06,0.08]	0.054 [0.01,0.13]
Investment	0.378 [0.24,0.54]	0.3 [0.14,0.53]	−0.003 [−0.09,0.19]	0.003 [−0.05,0.11]	0.046 [−0.05,0.2]	−0.067 [−0.17,0.04]
Coefficient on Nested Skills	0.502 [0.4,0.62]	0.691 [0.47,0.78]	0.92 [0.6,0.99]	1.027 [0.93,1.07]	0.937 [0.82,1.01]	0.952 [0.84,1.02]
Complementarity (ρ)	−0.015 [−0.14,0.16]	−0.189 [−0.42,0.16]	−0.209 [−0.51,0.78]	0.268 [−0.91,0.8]	−0.4 [−0.42,0.38]	0.26 [−0.89,0.43]
Elasticity of Substitution	0.986 [0.88,1.18]	0.841 [0.7,1.19]	0.827 [0.51,2.13]	1.366 [0.47,2.55]	0.714 [0.71,1.62]	1.351 [0.53,1.75]
Complementarity Nested (ρ^{skills})	0.521 [−0.53,1.04]	1.809 [−1.14,1.28]	−0.133 [−1.37,0.54]	−0.025 [−0.31,0.12]	−0.179 [−0.94,0.76]	−0.232 [−0.88,0.77]
Elasticity of Substitution Nested	2.09 [−1.25,8.4]	−1.236 [−3.49,4.38]	0.883 [0.35,1.5]	0.975 [0.76,1.14]	0.848 [0.45,2.05]	0.812 [0.48,2.56]
log TFP ($\log(A_t)$)	0.054 [−0.08,0.1]	−0.265 [−0.37,−0.04]	−0.051 [−0.22,0]	−0.079 [−0.11,−0.01]	−0.151 [−0.23,−0.08]	0.005 [−0.06,0.1]
Investment Residual (υ)	−0.325 [−0.51,−0.18]	−0.661 [−0.9,−0.19]	0.126 [−0.22,0.31]	−0.01 [−0.11,0.1]	−0.116 [−0.34,0.05]	−0.02 [−0.14,0.2]
Number of Children	−0.021 [−0.05,0.01]	−0.03 [−0.06,0]	0.008 [−0.02,0.02]	−0.011 [−0.03,0]	−0.012 [−0.02,0.01]	0.02 [0,0.03]
Older Siblings	0.02 [0,0.04]	−0.015 [−0.04,0.01]	0.022 [0,0.05]	0.009 [0,0.02]	0.02 [0,0.03]	−0.006 [−0.01,0.02]
Gender	0.015 [0,0.03]	0 [−0.02,0.03]	0.01 [0,0.04]	−0.007 [−0.02,0]	0.008 [−0.01,0.02]	0.043 [0.03,0.06]
Income	−0.065 [−0.1,0.04]	0.281 [0.08,0.38]	−0.012 [−0.05,0.12]	0.076 [0.03,0.1]	0.104 [0.06,0.16]	−0.087 [−0.15,−0.04]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

3.10.5 Estimates Using Prices, Income, and their Interactions as Instruments

Table 3.23: Production of Human Capital, CES Production Function, Ethiopia

	Cognition				Health			
Age	5	8	12	15	5	8	12	15
Cognition		0.374 [0.27,0.48]	0.472 [0.33,0.57]	0.813 [0.58,0.89]		0.008 [−0.04,0.05]	0.25 [0.13,0.3]	0.011 [−0.07,0.1]
Health	0.125 [0.05,0.21]	0.092 [0.04,0.17]	0.041 [−0.03,0.14]	0.045 [−0.01,0.08]	0.593 [0.53,0.64]	1.004 [0.92,1.05]	0.902 [0.81,0.97]	0.884 [0.81,0.92]
Parental Cognition	0.104 [0.02,0.25]	0.038 [−0.01,0.15]	0.012 [−0.03,0.09]	−0.004 [−0.07,0.06]	−0.08 [−0.15,0]	0.012 [−0.03,0.04]	−0.065 [−0.11,0]	−0.023 [−0.07,0.02]
Parental Health	0.015 [−0.02,0.08]	−0.032 [−0.07,0.01]	−0.022 [−0.07,0.02]	−0.021 [−0.05,0.03]	0.14 [0.11,0.25]	−0.012 [−0.02,0.02]	0.045 [0,0.08]	0.026 [0.01,0.07]
Investment	0.756 [0.51,0.91]	0.527 [0.33,0.63]	0.497 [0.38,0.71]	0.167 [0.06,0.43]	0.347 [0.18,0.45]	−0.011 [−0.09,0.11]	−0.132 [−0.25,0.05]	0.102 [−0.02,0.24]
Complementarity (ρ)	−0.519 [−0.83,0.37]	0.049 [−0.1,0.47]	−0.516 [−0.83,−0.1]	0.005 [−0.32,0.56]	−0.285 [−0.4,−0.06]	−0.122 [−0.59,0.55]	0.431 [0.26,0.98]	−0.583 [−0.64,0.06]
Elasticity of Substitution	0.658 [0.54,1.44]	1.052 [0.91,1.88]	0.659 [0.55,0.91]	1.005 [0.76,2.27]	0.778 [0.72,0.94]	0.891 [0.63,2.21]	1.756 [−0.16,3.71]	0.632 [0.61,1.07]
log TFP ($\log(A_t)$)	0.052 [−0.05,0.06]	0.028 [−0.02,0.05]	0.014 [−0.06,0.07]	−0.069 [−0.11,−0.01]	0.079 [0.04,0.11]	0.002 [−0.03,0.02]	0.012 [−0.05,0.04]	0.16 [0.1,0.17]
Investment Residual (υ)	−0.583 [−0.74,−0.33]	−0.517 [−0.65,−0.26]	−0.519 [−0.77,−0.22]	−0.157 [−0.52,−0.03]	−0.285 [−0.43,−0.14]	−0.06 [−0.18,0.03]	0.263 [−0.18,0.52]	−0.178 [−0.36,0.13]
Number of Children	−0.019 [−0.06,0.01]	−0.039 [−0.07,−0.02]	0.025 [0,0.05]	−0.031 [−0.05,−0.01]	−0.046 [−0.1,−0.03]	−0.005 [−0.03,0.01]	−0.014 [−0.03,0.01]	−0.002 [−0.02,0.01]
Older Siblings	0.02 [0,0.06]	0.018 [0,0.05]	−0.017 [−0.05,0.01]	0.024 [0,0.05]	0.024 [0.01,0.07]	0.012 [0,0.04]	0.001 [−0.02,0.02]	0.004 [−0.01,0.02]
Gender	0.002 [−0.01,0.02]	0.003 [−0.01,0.02]	0.006 [−0.01,0.03]	0.051 [0.03,0.07]	−0.004 [−0.02,0.01]	−0.007 [−0.02,0]	−0.007 [−0.02,0.01]	−0.082 [−0.09,−0.06]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

Table 3.24: Production of Human Capital, CES Production Function, Peru

	Cognition				Health			
Age	5	8	12	15	5	8	12	15
Cognition		0.458 [0.36,0.59]	0.65 [0.44,0.74]	0.87 [0.49,0.96]		−0.048 [−0.08,−0.01]	−0.04 [−0.12,0.06]	−0.038 [−0.09,0.03]
Health	0.023 [−0.04,0.1]	0.054 [−0.02,0.11]	0.085 [0,0.16]	0.04 [−0.01,0.18]	0.681 [0.63,0.74]	1.088 [1.01,1.11]	0.988 [0.88,1.03]	0.954 [0.86,1]
Parental Cognition	0.025 [−0.08,0.25]	0.204 [0.12,0.3]	0.299 [0.13,0.37]	0.096 [0.06,0.38]	0.042 [−0.01,0.13]	0.018 [−0.03,0.04]	0.077 [−0.01,0.11]	0.007 [−0.02,0.08]
Parental Health	−0.079 [−0.22,0.06]	−0.047 [−0.15,0.03]	−0.102 [−0.27,0]	−0.02 [−0.15,0.03]	0.178 [0.11,0.24]	−0.047 [−0.07,0.01]	0.007 [−0.08,0.08]	0.053 [0.02,0.17]
Investment	1.031 [0.75,1.18]	0.331 [0.21,0.51]	0.067 [−0.07,0.43]	0.014 [−0.1,0.15]	0.098 [−0.04,0.2]	−0.011 [−0.08,0.1]	−0.033 [−0.12,0.16]	0.023 [−0.1,0.09]
Complementarity (ρ)	−0.421 [−0.76,1.03]	−0.018 [−0.15,0.17]	0.112 [−0.15,0.39]	−0.087 [−0.65,0.15]	0.011 [−0.22,0.18]	−0.023 [−0.29,0.21]	0.137 [−0.23,0.43]	−0.52 [−0.88,0.22]
Elasticity of Substitution	0.704 [−1,8.27]	0.982 [0.87,1.21]	1.127 [0.87,1.64]	0.92 [0.61,1.17]	1.011 [0.82,1.23]	0.978 [0.77,1.27]	1.159 [0.82,1.74]	0.658 [0.53,1.28]
log TFP ($\log(A_t)$)	−0.039 [−0.11,0]	−0.02 [−0.05,0.01]	0.051 [−0.01,0.11]	−0.068 [−0.11,−0.03]	−0.022 [−0.05,0]	0.017 [0.01,0.04]	−0.037 [−0.07,0]	−0.095 [−0.12,−0.06]
Investment Residual (υ)	−1.069 [−1.2,−0.75]	−0.272 [−0.47,−0.14]	−0.382 [−0.66,−0.01]	0.103 [−0.16,0.33]	−0.03 [−0.14,0.17]	0.017 [−0.1,0.13]	−0.011 [−0.22,0.16]	−0.123 [−0.21,0.12]
Number of Children	0.03 [−0.01,0.04]	−0.017 [−0.04,0.01]	−0.034 [−0.07,0]	0.008 [−0.02,0.02]	−0.011 [−0.02,0]	−0.017 [−0.03,0]	−0.013 [−0.03,0.01]	0.018 [0,0.03]
Older Siblings	−0.014 [−0.02,0.02]	0.018 [0,0.05]	−0.003 [−0.03,0.02]	0.021 [0,0.05]	0.002 [−0.02,0.01]	0.013 [0,0.03]	0.025 [0,0.04]	−0.009 [−0.02,0.01]
Gender	0.001 [−0.01,0.02]	0.015 [0,0.03]	0.005 [−0.01,0.03]	0.011 [0,0.04]	0.021 [0.01,0.03]	−0.007 [−0.02,0]	0.009 [−0.01,0.02]	0.048 [0.03,0.06]

Notes: 95% confidence intervals based on 100 bootstrap replications in square brackets.

Chapter 4

Health Inequality, Labor Supply and Retirement Policies

4.1 Introduction

Public finances in high income countries are increasingly under strain due to gains in longevity that have not been accompanied by commensurate extensions in working lives. Policies incentivizing work among older workers are being widely considered and legislated, but their full effects are complex and not well understood.

Health will play a key role in shaping the effects of retirement policies. The interactions between health and labor market outcomes are especially meaningful around retirement age, a time that sees health deterioration speeding up. These interactions go in two directions: health problems reduce productivity and labor supply, but in turn working may also affect health. These two-way effects may be very heterogeneous and particularly relevant for those who are already in poorer health.

In this paper I investigate the dynamic relationship between labor market and health outcomes for older women. I develop a structural framework that allows for feedback effects of employment on health to depend on health status and other characteristics of the workers. The model is especially well suited for welfare analysis of retirement policies such as those incentivizing workers to extend their working lives. I focus on women, who were typically allowed to retire earlier than men and thus have experienced more sizeable changes to their working lives compared to men as a result of these policies.[1]

The paper has two main contributions. First, I shed light on the dynamic relation-

[1] For example in the United Kingdom, which is the focus of this study, the female state pension age increased from 60 to 66 over the 2010-2020 period. In comparison, the male state pension age in the UK increased from 65 to 66 over the same period, starting in 2019.

ship between health and labor market outcomes of women around retirement age, allowing for a two-way interaction between employment and health. I study the role of health shocks and financial incentives for individual labor supply decisions of women at older ages, and how the impact of these shocks varies across the health distribution and for different education groups. I produce Marshallian and Frisch elasticities, as well as estimates of labor supply sensitivity to changes in health, to study differences in the labor supply response to health and financial shocks across different education groups and by health status. Second, I investigate the welfare consequences of policies that increase the state pension age, i.e. the age at which individuals become eligible to receive state pension benefits. I quantify differences in the welfare implications of the policy across the health distribution and education groups and, conversely, the effect that these policies have on health inequality arising from the effect employment has on health.

I do this by developing and estimating a rich structural model of consumption, savings, labor supply and health of women at older ages. The relationship between labor supply and health is complex and dynamic. The model embeds various channels through which health and the labor supply decision can interact. In the model, women choose savings and labor supply at the extensive margin in every period. Health impacts earnings, the utility cost of working and mortality. Simultaneously, I carefully model the health process, allowing health to be persistent and to depend on the woman's labor supply. The latter effect is allowed to be heterogeneous and depend on health status. I construct a broad measure of health by combining information from subjective and objective measures. The model embeds correlated unobserved heterogeneity in earnings and health to account for any unobserved factors that affect both. Women in the model are subject to uncertainty over earnings, health and mortality. Moreover, they face uncertainty from exogenous partner mortality, his labor supply and earnings. I estimate the model separately for different education groups, allowing me to capture differences in the relationship between health and employment across education groups.

For estimation, I exploit a policy change that increased the state pension age for women in the UK. Importantly, in the UK an individual's date of birth is the only determinant of the eligibility date for state pension benefit receipt. The reform provides a source of identification that helps to credibly disentangle the two-way relationship between employment and health, by isolating exogenous variation in the employment decision. I leverage the policy to validate the model, by showing that the model reproduces well effects generated by the policy that are not targeted in the estimation. Another relevant feature of the UK setting is that it provides tax-funded, free-at-the-time-of-delivery, universal access to health care services, allowing to abstract from the role of out-of-pocket medical expenses. The data is drawn from eight bi-annual waves of the English Longitudinal Study of Ageing (ELSA), covering the years 2002 to 2016. I combine the data with a detailed simulation of the tax and benefit system to incorporate all taxes and benefits and the way the policy changes them across cohorts.

Model estimates show that working has heterogeneous effects on future health depending on the woman's education and underlying health status. Low educated women in poor health suffer sizeable negative effects from working: at median health, one additional year of work reduces future health by 8% of a standard deviation. The negative effect levels off for women in better health. The effects for high educated women are less heterogeneous and smaller in magnitude, but non-negligible (-1% of a standard deviation at the median). Health has a large, positive effect on earnings of both low and high educated women, and women across education groups in better health enjoy lower utility costs of working. Moreover, women in better health suffer lower mortality risk.

I show that heterogeneity in health generates substantial differences in labor supply responses to changes in earnings and health. To investigate labor supply responses to changes in earnings, I compute Frisch and Marshallian elasticities. I find that both elasticities are substantially higher for women in the bottom health quartile than women in the top. To gauge the labor supply response to changes in health, I compute how labor supply changes in response to a health increase that generates a 1% increase in earnings according to the model estimates. The labor supply response with respect to this health shock is large and increases with age. Moreover, it is more than twice as large as the response implied by Frisch and Marshallian elasticities for both education groups, suggesting that health shocks have a greater impact on female labor supply decisions than changes in earnings at older ages. I find that the labor supply response to health shocks is substantially larger for women in poor health, for both education groups.

Overall, the results point towards three main findings. First, health shocks at older ages have large effects on labor supply behavior near retirement, and this effect increases with age. Second, health shocks have larger impacts on labor supply behavior at older ages than changes in earnings. Third, the extent to which women respond to health shocks and earnings changes depends on their health, with those in poor health responding more strongly than those in better health.

Using the model, I simulate a policy counterfactual that increases the female state pension age. I find that the reform has heterogeneous effects and tends to widen inequality in health. In particular, low educated women in poor health bear a larger cost from the reform than other groups, measured using the compensating variation as a proportion of the present value of consumption at baseline. This is because these women cannot afford not to work in the absence of pension benefits, but working for them is more costly, as poor health causes lower earnings and higher utility costs from work. In addition, increased employment damages their already poor health. This increases their mortality and, since health is persistent, also makes future employment more costly.

This paper relates to three main strands of the literature. First, it contributes to the literature that has studied the role of health risk for economic behavior. In particular, several papers have studied the role of health as a determinant of labor supply decisions near retirement. For example, French (2005), Bound, Stinebrickner, and Waidmann (2010) and French and Jones (2011) develop structural models of retirement that explicitly include health risk as a driver of retirement decisions. The literature on health risk and economic outcomes has also looked at the role that health plays for wealth, earnings and consumption inequality over the life cycle (see, for example, Capatina (2015), De Nardi, Pashchenko, and Porapakkarm (2017), Hosseini, Kopecky, and Zhao (2020), Blundell, Borella, Commault, and De Nardi (2020)). I contribute to this literature by allowing for a two-way interaction between health and employment decisions at older ages, while these papers take health as exogenous. This allows me to study the role of health inequality for welfare implications of retirement policies that increase the age of eligibility for state pension benefits, including the role that these policies play in widening or reducing health inequality through their impact on labor supply.

This paper also relates to studies that allow for the evolution of health to be endogenous (see, for example, Grossman (1972), Gilleskie (1998), Ozkan (2017), Cole, Kim, and Krueger (2018), Margaris and Wallenius (2020)). Only a few papers in the structural literature considering endogenous health have allowed for an explicit two-way relationship between health and labor supply (see Jacobs and Piyapromdee (2016), Papageorge (2016), Harris (2019), Capatina, Keane, and Maruyama (2020), Jolivet and Postel-Vinay (2020)).[2] My paper differs substantially from these in terms of focus and methodology. First, I provide novel evidence on the role of health inequality for employment decisions of women at older ages, and on welfare implications of delaying state pension benefit receipt, as well as the role that the policy plays in shaping health inequality. Second, I model the tax and benefit system in detail, and I leverage

[2]See Currie and Madrian (1999) for a survey of the health economics literature on the interaction between health and employment.

an exogenous policy change in women's incentives to work to more credibly disentangle the two-way relationship between employment and health. Third, my model allows for richer heterogeneity than is typically allowed in these models: it is estimated separately for different education groups, and it allows for heterogeneity in the effects of employment on health. This allows to understand the heterogeneous welfare implications of increasing the state pension age for women of different socio-economic backgrounds and with different health.

Finally, there is a large reduced form literature that attempts to identify the causal impact of health on employment and vice-versa. Some examples include Bound (1991), Siddiqui (1997), Bound, Schoenbaum, Stinebrickner, and Waidmann (1999), Blau and Gilleskie (2001), Smith (2004), Disney, Emmerson, and Wakefield (2006)), Coe and Zamarro (2011), Fitzpatrick and Moore (2018), Banks, Cribb, Emmerson, and Sturrock (2019), Blundell, Britton, Costa Dias, and French (2020), Kuhn, Staubli, Wuellrich, and Zweimüller (2020) and Rose (2020). This paper complements this literature by developing a structural framework that allows to quantify the effect of health on employment and vice-versa, as well as conducting counterfactual experiments.

The rest of the paper proceeds as follows. Section 4.2 describes the data, institutional setting and reform I use to estimate the model. Section 4.3 describes the model used in the paper. Section 4.4 details the estimation procedures. Section 4.5 presents parameter estimates. Section 4.6 discusses model fit and implications for behavior. Section 4.7 discusses counterfactual analysis. Section 4.8 concludes.

4.2 Data, Institutional Context and Reform

4.2.1 The panel data sample

To estimate the model, I use data from waves 1 through 8 of the English Longitudinal Study of Ageing (ELSA). The ELSA is a longitudinal survey that focuses on individuals aged 50 or older. It is the English equivalent of the HRS for the US or the SHARE for Europe. The first wave was collected in 2002 and new information on respondents is collected every two years. In waves 3, 4, 6 and 7 new individuals were included in the survey to replenish the sample due to mortality of existing interviewees. All members of the household above the age threshold are interviewed. ELSA collects extensive information regarding individuals' labor supply, earnings, household savings and other demographic and educational aspects. Importantly, ELSA contains rich and repeated information on the survey respondents' health. I exploit this information to generate a single, continuous index measuring health, using a methodology proposed in Blundell, Britton, Costa Dias, and French (2020). The measures included in the construction of the health index are aimed at capturing a broad measure of health concerning work capacity, i.e. the ability and desire to work from the perspective of one's health status, as laid out in Currie and Madrian (1999). Further details on the procedure I use to construct the health index are provided below.

The unit of observation in the model is a woman, to which I link information on her partner from the ELSA survey if she has one. To construct the sample to be used for the model estimation, I exclude all women who are self-employed. The full dataset is an unbalanced panel of 7,018 women aged between 50 and 75 who are observed at some point during the 2002-2017 period. To limit the measurement error in earnings, the distribution is trimmed at percentiles 2 and 99 from below and above. Average career earnings are an important component of the model, as they determine the amount of public and private (if available) pension income women and their male partners receive once they reach the state pension age (following O'Dea (2018)). Information on average career earnings is not collected in the ELSA survey. Instead, there is information available about the employment history of a subset of the individuals in the sample. Thus, I first construct average career earnings for the individuals for whom there is information on their employment history. I impute their earnings history based on earnings growth rates computed on the British Household Panel Survey data conditional on gender, education, age, region of residence and marital status. For individuals who do not have information on employment history, I impute average career earnings using propensity score matching conditional on gender, education, age, health, assets, employment status, earnings if employed, and whether they are enrolled in a private pension scheme.

Constructing Health In ELSA, information about an individual's health is collected in every wave. Some measures of health are self-reported, and others are objective health measures. For example, individuals are asked about physician diagnosed conditions, or to evaluate their own mobility according to standardised scales. In the interest of parsimony, I model health as a single, continuous index. I construct the index using a methodology proposed in Blundell, Britton, Costa Dias, and French (2020), which combines both subjective and objective measures of health.[3] The key idea is to instrument subjective health measures with objective health measures to deal with measurement error and justification bias in subjective health measures. Put differently, this allows to avoid capturing fluctuations in subjective health that are not

[3]Using ELSA, they show that a single index for health that combines subjective and objective health measures is sufficient to capture the variation in health that is relevant for employment.

related with fluctuations in objective health measures. I adapt the methodology in Blundell, Britton, Costa Dias, and French (2020) to my setting in the following way. First, I construct a factor using three subjective health measures: a question on self-reported levels of pain perceived, a question on self-reported overall health status and a question on whether the individual perceives that their health limits their activities. I instrument this factor using binary measures on physician diagnosed conditions (for example, whether the person has been diagnosed with a heart attack, diabetes, high blood pressure, arthritis, osteoporosis, Parkinson's disease, etc.), sight and hearing problems, as well as a number of questions on mobility difficulties, according to the standardized Activities of Daily Living (ADL) and Instrumental ADLs (IADL) scales.[4] A complete variable description of the measures used to construct the health index is provided in appendix 4.11.1. To give a sense of how the objective health conditions affect the health index, I regress the health index on the onset of a new diagnosis, controlling for the lagged value of the index, age dummies and marital status. For example, the onset of a heart attack reduces the index by about 65% of a standard deviation. Suffering a stroke reduces the index by about 40% of a standard deviation.

4.2.2 Institutional context and reform

Pension provision in the UK is a mix of private and publicly provided pensions. State pension benefits in the UK consist of state pension income and a means-tested income floor called pension credit.[5] The state pension age reform consists in a change in the age women become eligible for state pension benefits. Importantly, the state pension benefit eligibility date in the UK is only conditional on an individual's birth date and is not tied to labor force participation requirements. Hence, the reform changes women's financial incentives to work by delaying the receipt of state pension benefits. Many papers have shown that the female state pension age has a substantial effect on women's labor supply in the UK (see, for example, Cribb, Emmerson, and Tetlow (2013), Banks, Cribb, Emmerson, and Sturrock (2019) and Rose (2020)).

The 1995 Pensions Act mandated an increase in the state pension age of women from 60 to 65, to take place between 2010 and 2020. As a result of the reform, the state pension age for women born before April 1950 is 60, whereas for women born in or after April 1950, it increases by 1 month for each month of birth. The 2011 Pensions Act accelerated this increase to 65 and increased the overall state pension

[4] The ADL and IADL scales have been devised by health care professionals in order to measure an individual's functional status (see Katz et al. (1963)). They range from fundamental activities such as the ability of feeding oneself, bathing, dressing, to activities related to the ability of functioning independently, such as cleaning the house, shopping for groceries and preparing meals.

[5] O'Dea (2018) provides a detailed description of the UK pension system.

age to 66. Figure 4.1 reproduces a graph shown in Banks, Cribb, Emmerson, and Sturrock (2019). It shows the female state pension age in the UK by a woman's date of birth and how the reform changes it.

The variation in the state pension age induced by the reform is exploited in the estimation of the model. The model embeds a detailed description of the UK tax and benefits system. For computational simplicity, I model two tax and benefits system, one pre-reform (2006/2007) with state pension age 60 and one post-reform (2012/2013) with state pension age 63. Women in the sample are assigned to either one depending on whether they are born before or after April 1950. The estimation exploits the policy change by comparing the behavior of different cohorts who are subject to different state pension ages.

At this stage, it is worth discussing two important aspects of the institutional context in the UK that guide some of the modeling choices. First, the UK health care system is one where everyone is insured by one single public insurer body, the National Health Service (NHS). The NHS is funded through taxpayers' money, and there are no insurance premiums, fees or other out-of-pocket medical expenses. No one can opt-out. As such, the model abstracts from health insurance considerations.[6] Second, disability benefits in the UK are a flat-rate payment to qualifying individuals. Their value is less than 15% of average earnings, and as such they explain very little of individual retirement behavior in the UK (see Banks, Emmerson, and Tetlow (2014) and Banks, Blundell, and Emmerson (2015) for detailed analyses of this topic). In appendix 4.11.2, I show that disability benefits do not represent a relevant pathway to retirement for women in the sample. Given these empirical facts, the model abstracts from disability benefits entirely.

4.3 Model

4.3.1 Key features

The main agent making optimal decisions is a woman, between the ages of $t_0 = 50$ (first time observed in ELSA) and a terminal period set at age $T = 90$. The woman makes a period-by-period optimal consumption, saving and binary labor supply decision (work full time or not work). During her working years, she is allowed to freely exit and re-enter the labor force. Retirement is exogenous from age 75 and the woman

[6] A small proportion of individuals in the UK has additional private health insurance. Private health insurance in the UK has the benefit of accelerating access to services, rather than providing additional health care that is not available through the NHS. In the sample, about 15% of women report having private health insurance.

lives from her accumulated savings and pension. The model is estimated separately by the woman's education level, which is either i) secondary education (i.e. high school drop out), or ii) high school or college degree.[7] Earnings from work are modeled as a function of the woman's stock of health, a second-order polynomial in age, as well as an unobserved heterogeneity component. All parameters characterizing the earnings equation and the distribution of unobserved heterogeneity vary by the woman's education. Earnings are subject to persistent shocks, and the distribution of these shocks depends on education. An important feature of the model is the woman's health accumulation process, which depends (among other things) on the woman's labor supply choice. Health is modeled as a single index, and it is a continuous variable. Health is persistent, and is allowed to depend on age, the woman's marital status, on the woman's lagged labor supply decision, and an interaction between the labor supply decision and lagged health. Thus, the effect of working on future health is heterogeneous depending on the woman's health stock. Moreover, health depends on an unobserved heterogeneity component, and an iid shock. All parameters governing the health process and the distribution of unobserved heterogeneity vary by education. I allow for the unobserved heterogeneity in health and earnings to be correlated, to control for unobservable differences across earnings-health types combinations. Both unobserved heterogeneity components are assumed to be discrete and can take two values, which can be interpreted as being low or high productivity and low or high health. Thus, individuals can be one of four types.

Health affects the decision to work via a number of channels. First, health affects the woman's probability of survival to the next period. Second, health affects the woman's period utility by affecting the cost of working. Finally, as described above, health affects the amount the woman can earn when in work. Hence, the model allows for multiple channels through which health and labor supply can interact: the woman's current health stock affects her labor supply decision through its impact on earnings, utility, and survival, and labor supply decisions affect health directly. Married women face additional risks over a number of dimensions related to their partners. For computational reasons, partners are modeled in an exogenous fashion, and the parameters characterizing their equations are estimated outside of the model. In each period, partners may work with a given probability that depends on the woman's age and education. If they work, they receive annual earnings that depend on a second-order polynomial in age and on persistent shocks. The distribution of these shocks varies by education. Partners face mortality risk in every period, the survival prob-

[7]This assumption is consistent with what I observe in the data, where the split across these two groups is roughly 50/50.

abilities vary by age and education. All choices are affected by the tax and benefit system, which varies by cohort. For computational simplicity, I model two cohorts and two tax and benefit systems, capturing a pre- and a post-state pension age reform period. The tax and benefit system determines the household's disposable income under each choice of employment. To simulate the tax and benefit system, I use FORTAX, a tax and benefit micro-simulation tool to construct detailed budget constraints that replicate the tax and benefit system in place at a certain point in time.[8] FORTAX simulates the tax and welfare system for people in working ages only. Thus, I augment the simulation tool with specific features of the tax and benefit system targeted towards the elderly. Some women also receive a private pension income, if they report having a private pension scheme in the data. Both public and private pensions are modeled as functions of individual average career earnings accumulated up to the state pension age, following O'Dea (2018). Average career earnings are endogenous to the model and depend on the labor supply decisions made up to the state pension age. Importantly, the state pension age reform exogenously changes the age at which women are entitled to receive state pension benefits. This exogenous change in the woman's financial incentives to work is exploited in the model to estimate the model, by comparing the behavior of the two cohorts. I now explain the model in detail.

4.3.2 Working years

State variables In every period of her working life, the woman maximizes expected utility, taking as given her characteristics and the economic environment. These are given by her age (t), education (s), assets (a), health stock (h), average career earnings (ae), whether she is enrolled in a private pension scheme (pp), productivity shock (ϵ), and unobserved heterogeneity earnings and health type (θ and κ respectively). Moreover, they include family circumstances such as the presence of a partner (m) and information related to her partner if she has one: his labor supply (l^m, either full-time or out of work), productivity shock (ϵ^m) and average career earnings (ae^m). The vector X_t denotes the state variables in period t.

Utility Utility is intertemporally separable. I model instantaneous utility in a similar fashion to Blundell, Costa Dias, Meghir, and Shaw (2016), which depends on equivalised consumption, female labor supply, health and family circumstances. In period t, it is given by

$$u(c_{i,t}, l_{i,t}; h_{i,t}, Z_{i,t}) = U(c_{i,t}, l_{i,t}; h_{i,t}, Z_{i,t}) + \pi^s \tag{4.3.1}$$

[8] see Shephard (2009), Shaw (2011) and Blundell, Costa Dias, Meghir, and Shaw (2016)

with

$$U(c_{i,t}, l_{i,t}; h_{i,t}, Z_{i,t}) = \frac{(c_{i,t}/n_{i,t})^{\mu}}{\mu} * exp(1 + l_{i,t}(\phi_1^s + \phi_2^s h_{i,t} + Z'_{i,t}\psi^s)) \tag{4.3.2}$$

where c is total household consumption, n is the equivalence scale, l is binary female labor supply (full-time or out of work), h is female health and Z is a vector capturing family circumstances, concerning whether the woman has a partner and whether he works in period t. The exponential term in U captures the marginal utility of consumption changes with the woman's labor supply decision. This effect is different depending on the woman's health stock and her family circumstances. The parameter μ determines the curvature of the utility function. It is negative, and as such a positive value of the term in brackets implies that working reduces the utility from consumption. In particular, the parameter ϕ_2 captures the effect that health has on the cost of working. If ϕ_2 is estimated to be negative (positive), this implies that being better health reduces (increases) the cost of working. This parameter is estimated within the model along with the parameters ϕ_1 and ψ.[9] A caveat is that in this model, health affects individual utility only through the cost of working. I thus cannot capture how the marginal utility of consumption changes with health.[10] All parameters related to the cost of working differ by the woman's education. Finally, following Hall and Jones (2007), I add a positive constant π to instantaneous utility to capture the value of life. When survival probabilities are endogenous, as is the case in this model, it is necessary to add a positive constant to the otherwise negative instantaneous utility function to avoid that individuals derive greater utility from death than from being alive. I calibrate this constant by setting it to the positive of the minimum attainable instantaneous utility value, which ensures that period utility is always positive. Further details about the calibration are provided in section 4.4.

Earnings When in work, women receive earnings that evolve according to the following equation

$$log(y_{i,t}) = \alpha_0^s + \alpha_1^s age_{i,t} + \alpha_2^s age_{i,t}^2 + \alpha_3^s h_{i,t} + \theta_i^s + \epsilon_{i,t} \tag{4.3.3}$$

[9]Male health is excluded from the model for computational reasons, and the interaction between female and male health for couples' retirement behavior is left for future work.

[10]The existing empirical evidence on the effect of health on the marginal utility of consumption is inconclusive (see Finkelstein, Luttmer, and Notowidigdo (2009)). The results in Blundell, Borella, Commault, and De Nardi (2020) show that the effect of health on the marginal utility of consumption could vary depending on the type of consumption considered. This suggests the need for a more elaborate model of consumption/health to understand how health interacts with consumption, which is beyond the scope of this paper.

where h is the woman's stock of health, and θ is a discrete unobserved heterogeneity component that can take one of two values, indicating high or low productivity. The idiosyncratic productivity shock ϵ follows an AR(1) process given by

$$\epsilon_{i,t} = \rho^s \epsilon_{i,t-1} + \eta_{i,t} \tag{4.3.4}$$

with normally distributed innovations

$$\eta \sim N(0, \sigma^s_\eta) \tag{4.3.5}$$

The initial productivity shock is distributed as a normal. All parameters governing the earnings process differ by education.[11]

Health The woman's health process at any given period t is given by

$$h_{i,t} = \lambda^s_0 + \lambda^s_1 h_{i,t-1} + \lambda^s_2 l_{i,t-1} + \lambda^s_3 h_{i,t-1} * l_{i,t-1} + \lambda^s_4 age_{i,t} + \lambda^s_5 m_{i,t} + \kappa^s_i + u_{i,t} \tag{4.3.6}$$

where l is binary labor supply and m denotes the presence of a partner.[12] Health is modelled as a persistent process, where current health depends on lagged health. Moreover, current health depends on the lagged labor supply decision, and an interaction between lagged labor supply and health. This allows for potential heterogeneity in the effect of working on health. Health also depends on an unobserved heterogeneity component κ and an iid shock u that is distributed as a normal. κ is a discrete unobserved heterogeneity component that can take one of two values, indicating high or low health. It is correlated with the unobserved heterogeneity in earnings θ.

Unobserved heterogeneity Unobserved heterogeneity components θ and κ are correlated. Both components are assumed to be discrete and can take on two values respectively. Thus, an individual can be one of four combination of health-earnings types: {low, low}, {low, high}, {high, low} or {high, high}. The values of θ and κ and the probabilities of being one of each type are estimated alongside other parameters of the model. More details on the assumptions for identification of correlated unobserved heterogeneity in earnings and health are provided in the estimation section.

[11] In the model, the labor supply decision is binary (work full time or not work). However, the variation in annual earnings will partly capture the variation in hours. If the differences are permanent, this will be captured by the unobserved heterogeneity component θ. If the differences are time-varying, they will be captured by the age and health effects, and the earnings shock.

[12] The presence of a partner is included in the health process to reflect findings in the medical literature showing that marital status may influence health, see Goldman, Korenman, and Weinstein (1995).

Assets The asset evolution equation is described by the budget constraint as

$$\begin{cases} a_{i,t+1} = (1+r)a_{i,t} + l_{i,t}y_{i,t} + m_{i,t}l^m_{i,t}y^m_{i,t} - T(X_{i,t}, l_{i,t}, l^m_{i,t}) - c_{i,t} \\ a_{i,SPA+j} \geq 0, \quad \text{for all } \ j \geq 0 \end{cases} \tag{4.3.7}$$

with terminal condition $a_{T+1} = 0$, where a are assets, r is the risk-free interest rate, y and y^m are female and male earnings respectively, l and l^m are female and male binary labor supply. The borrowing limit is such that by the time agents reach their state pension age, they must pay off any outstanding debt. This is to reflect the fact that individuals usually cannot borrow against their pension wealth. The function T represents the tax and benefit system. It captures an individual's financial incentives to work for all levels of income, as a function of their labor supply, earnings and family structure. Households can face one of two tax and benefit systems depending on the woman's date of birth. In particular, the tax and benefit system function T contains pension benefits that the agent receives from her state pension age onwards. The exogenous policy variation in the female state pension age thus induces variation in the financial incentives to work of women born in different years, that are captured by the function T.

Male employment and earnings Up to the age of 75, men in couples can either work full-time or be out of work. Their earnings and labor supply behavior are exogenous and given by

$$Prob(l^m_{i,t} = 1|X_{i,t}) = \gamma(age_{i,t}, s_i, l^m_{i,t-1}) \tag{4.3.8}$$

where the probabilities are estimated separately by age, education and lagged employment status. If the partner retires at or after his state pension age (which is 65, the UK statutory level for men over this period), I assume he does so permanently. If he works, he earns

$$log(y^m_{i,t}) = \beta^s_0 + \beta^s_1 age_{i,t} + \beta^s_2 age^2_{i,t} + \epsilon^m_{i,t} \tag{4.3.9}$$

where the innovation to male earnings follows an AR(1) process given by

$$\epsilon^m_{i,t} = \rho^{m,s}\epsilon^m_{i,t-1} + \eta^m_{i,t} \tag{4.3.10}$$

and

$$\eta^m \sim N(0, \sigma^{m,s}_\eta) \tag{4.3.11}$$

All parameters characterizing the male earnings process are indexed by education. To avoid including male education and age as additional state variables, and to account for the high correlation between couples' education and age, I use female age and education in the regressions for male earnings and employment.

Average career earnings Average career earnings determine the amount of pension benefits the woman and her partner receive from their respective state pension ages, following O'Dea (2018). Average career earnings data are not collected in ELSA. In section 4.2, I explain the procedure I follow to impute average career earnings of women and men in the first wave of data. Given the initial conditions in average career earnings, they are updated in the model according to the endogenous female labor supply decision. If the woman works, her average career earnings are updated by adding the amount earned, and dividing by the total number of years worked including the current period's decision. If she does not work, the amount earned is counted as a 0. [13] The functions mapping average career earnings into pension benefits are described below.

Mortality Female mortality is endogenous and depends on the woman's health, age and education. The probability of surviving to age $t+1$ conditional on being alive at t is given by

$$surv_{i,t} = \frac{exp(\delta_0^s + \delta_1^s h_{i,t} + \delta_2^s age_{i,t})}{(1 + exp(\delta_0^s + \delta_1^s h_{i,t} + \delta_2^s age_{i,t}))} \tag{4.3.12}$$

Women in couples face additional risk over their partner's mortality. Male mortality is given by

$$surv_{i,t}^m = \frac{exp(\delta_0^{m,s} + \delta^{m,s} age_{i,t})}{(1 + exp(\delta_0^{m,s} + \delta_2^{m,s} age_{i,t}))} \tag{4.3.13}$$

I use female age and education in the equation for male survival probabilities.

Pension benefits While I use the FORTAX routine to capture the tax and benefit structure in the UK, FORTAX does not include the components of the welfare system that are specifically geared towards the elderly. In this paragraph, I describe the way I model pension income, further details on additional features of the welfare system for the elderly are described in appendix 4.11.3. Pension benefits are a key component of the model, as the state pension age reform I exploit for identification shifts the age at which women become eligible to receive said benefits. Hence, I augment the tax and benefit regime as described by FORTAX to include pension benefits and

[13] Analogously, male average career earnings are updated taking into account his exogenous labor supply behavior and earnings.

other specific features of the welfare system for the elderly. Public pension benefits in the UK are a complex function of an individual's economic circumstances. I model public pension benefits following O'Dea (2018). In particular, public pension benefits are composed of public pension income and pension credit. Public pensions are a taxable social-security style source of income payable from the state pension age until death, pension credit is a means-tested income floor for the elderly. As in O'Dea (2018), public pension income is modeled as a quadratic function through the origin of average career earnings accumulated up to the state pension age:

$$sp_{i,t} = \begin{cases} \alpha_0 ae_{i,SPA} + \alpha_1 ae^2_{i,SPA} & \text{if ae } \leq \bar{\text{ae}}_{\text{sp}} \\ \alpha_0 \bar{ae}_{sp} + \alpha_1 \bar{ae}_{sp}{}^2 & \text{if ae } > \bar{\text{ae}}_{\text{sp}} \end{cases} \quad \text{for all } t \geq \text{SPA} \tag{4.3.14}$$

where $\bar{ae}_{sp}$ is the level of average career earnings at which the quadratic starts to decrease.

Pension credit is a function of income, assets and household composition. Details on the way pension credit is modelled are illustrated in appendix 4.11.3.

Some women are enrolled in a private pension scheme. They receive a taxable private pension income from the state pension age onwards. They accrue entitlements to the private pension while working. In particular, they must make pension contribution p_t in every period, equal to a fraction ω of their pre-tax earnings, until the state pension age is reached. [14] After that, they receive a taxable private pension income which is a function of their average career earnings accumulated up to the state pension age:

$$pp_{i,t} = \begin{cases} \tau^s_0 ae_{i,SPA} + \tau^s_1 ae^2_{i,SPA} & \text{if ae } \leq \bar{\text{ae}}_{\text{pp}} \\ \tau^s_0 \bar{ae}_{pp} + \tau^s_1 \bar{ae}_{pp}{}^2 & \text{if ae } > \bar{\text{ae}}_{\text{pp}} \end{cases} \quad \text{for all } t \geq \text{SPA} \tag{4.3.15}$$

where $\bar{ae}_{pp}$ is the level of average career earnings at which the quadratic starts to decrease.

Maximization problem At any age t during the working years, the woman's decision problem is given by

$$V_t(X_t) = max_{\{c_t, l_t\}} u(c_t, l_t; h_t, Z_t) + \beta surv_t E_t(V_{t+1}(X_{t+1})|X_t)$$

[14] The proportion of earnings that women with private pensions must pay (ω) is set at 5%, following O'Dea (2018).

subject to the earnings process (4.3.3), the health process (4.3.6) and the budget constraint (4.3.7), where the expectation is taken over future events conditional on what is known in period t.

Retirement years From age 75, when retirement occurs exogenously, the woman lives off her accumulated savings. During these years, the state variables are given by age, education, health, health unobserved heterogeneity, assets, average career earnings, enrolment in a private pension scheme, and partner presence and average career earnings. The woman chooses optimal consumption to solve

$$V_t(t,s,h,\kappa,a,ae,pp,m,ae^m) = max_{\{c_t,\}} u(c_t) + \beta surv_t E_t(V_{t+1}(t+1,s,h',\kappa,a',ae',pp,m',ae^{m\prime})$$

subject to the health process (4.3.6) and budget constraint (4.3.7).

4.4 Estimation

I estimate the model parameters using a two-step procedure.[15] In the first step, I estimate some parameters outside the model, and set some with reference to the existing literature. In the second step, I estimate the earnings process, health process, correlated unobserved heterogeneity in earnings and health and the parameters relating to the utility cost of working using the method of simulated moments.

4.4.1 Parameters estimated or set outside of the model

I estimate the predetermined components of the model, such as male labor supply and earnings, survival probabilities and the mapping from average career earnings to public and private pension income of women and their male partners outside the model. Details and estimates are provided in appendix 4.11.4. I set the utility curvature coefficient μ to -1, implying a risk aversion coefficient of 2, a widely used value.[16] The discount factor β is set to 0.98 and the annual risk-free interest rate r is 0.015.[17] [18]

[15] A two-step procedure for estimation is standard in papers that develop and estimate structural life cycle models. Examples include Gourinchas and Parker (2002), French (2005), and Blundell, Costa Dias, Meghir, and Shaw (2016)

[16] See, for example, Capatina, Keane, and Maruyama (2020)

[17] For the discount factor see, for example, Attanasio, Low, and Sánchez-Marcos (2008) and Blundell, Costa Dias, Meghir, and Shaw (2016).

[18] For the risk-free interest rate see Blundell, Costa Dias, Meghir, and Shaw (2016).

4.4.2 Parameters estimated within the model

I estimate the earnings process, health process, correlated unobserved heterogeneity in earnings and health and preference parameters related to the utility cost of working using the method of simulated moments. Estimation exploits the policy change over the period, that changes the state pension age of some women in the sample. The key assumption is that the policy affects health only through its impact on employment. As a result, I use it to isolate exogenous variation in the female employment decision, which aides identification of the impact of employment on health. To exploit this source of exogenous variation, I construct some moments conditional on the woman's cohort. Put differently, the model is estimated comparing the behavior of different cohorts, who are subject to different state pension age regimes. To the best of my knowledge, my paper is the first in the structural literature with endogenous health that exploits a policy reform to aide identification of the effect of employment on health.

To aide identification of the other direction of causality, i.e. the impact of health on employment (through the channels incorporated in the model), I exploit a strategy proposed in Blundell, Britton, Costa Dias, and French (2020). The health index is constructed by instrumenting subjective health with objective health measures such as diagnosed conditions. The key intuition is that, conditional on initial conditions in health, the onset of new health conditions (for example, having a heart attack) provides exogenous variation that can be used to identify the effect of health on employment. These newly diagnosed conditions are assumed to affect employment only through their impact on health. An in-depth discussion of the argument is provided in Blundell, Britton, Costa Dias, and French (2020).

Estimation in the second step relies on an iterative procedure. In every iteration, I start by solving the model under a given set of parameter estimates, taking initial conditions, economic circumstances and exogenously set parameters as given. The solution algorithm is based on a methodology proposed in Blundell, Costa Dias, Meghir, and Shaw (2016).[19] It accounts for the well-known difficulty in solving mixed continuous-discrete dynamic problems, where indifference points in future choices generate discontinuities in the value function. The central idea of this algorithm is to rely on uncertainty to "concavify" the expected continuation value. I adopt this approach relying on the rich characterization of uncertainty in my model. I then simulate the life cycle behavior of 35,090 women (i.e. 5 replications of 7,018 observations in the ELSA sample) under the specific tax and benefit regime they face. Women in the pre-reform cohort face the 2006/2007 regime with state pension age 60, women in the

[19] This is a modified version of algorithms in Fella (2014) and Iskhakov, Jørgensen, Rust, and Schjerning (2017).

post-reform cohort face the 2012/2013 regime with state pension age 63. For each woman, I select an observation window to ensure that the simulated sample exactly reproduces the age structure of the observed data. I use 124 moments to estimate 46 parameters. I compute the moments on the simulated data that are equivalent to those on the observed data. The parameter estimates $\hat{\Theta}$ are defined by

$$\hat{\Theta} = argmin_{\Theta}\left\{\sum_{k=1}^{K}[\frac{(M_{kN}^{d} - M_{ks}^{m}(\Theta))^2}{Var(M_{kN}^{d})}]\right\} \tag{4.4.1}$$

where M_{kN}^{d} is the k-th data moment estimated over N observations, $M_{ks}^{m}(\Theta))$ is the same moment evaluated at parameter Θ over s simulations, and the sum is over K moments. All moments are constructed separately by education group. Appendix 4.11.5 shows the complete list of data moments used in the estimation, along with their simulated counterparts and the difference between the two, normalized by the data standard error. I use moments including employment rates (including information on employment rates by cohort), coefficients from log earnings regressions and the distribution of log earnings over the life cycle and in the initial period, coefficients from health regressions and the distribution of health over the life cycle, correlations between employment and age and health, as well as moments capturing the change in employment and health between ages 60 and 62 for both cohorts. Of course, all estimated model parameters affect all moments, but some particular moments bear greater weight for the identification and estimation of certain parameters. I now discuss identification of each parameter.

Health process

To estimate the health process, I simulate data from the model and iterate on the health process parameters to match a number of moments related to the health distribution. The moments that bear greater weight for the estimation of the health process are the average of the health distribution; the coefficients of an auxilliary regression of health that replicates the health process in the model; the variance, skewness and kurtosis of residuals from the aforementioned regression. I also include moments that use the policy variation, such as employment rates by age and cohort, as well as the difference between the two cohorts in the change in employment and health between ages 60 and 62.

Earnings process

I estimate the earnings process parameters iterating to match a number of moments related to the earnings distribution. These include the moments related to the distribution of earnings at the beginning of the model (age 50), and over the working life. The former are the coefficients of an auxilliary regression of initial earnings on health, and the variance, skewness and kurtosis of residuals from this regression. Moreover, I include average earnings, and coefficients from an auxilliary regression of earnings on health, a second-order polynomial in age and the residuals from the health regression described above. I also include the variance, skewness and kurtosis of the residuals of this regression, and the autocorrelation between lags of these residuals.

Unobserved heterogeneity

Identification of correlated unobserved heterogeneity in earnings and health relies on two key assumptions. The first is that the heterogeneity components are discrete random variables. The second is that the time-varying component of the residuals of earnings and health are continuous, normally distributed and mutually independent. The higher order moments of the distribution of residuals of earnings and health, as well as the correlation between earnings and health residuals bear significant weight in estimating the values and joint probabilities of the unobserved heterogeneity distribution.

Utility cost of working

To estimate the utility cost of working and how it varies by health and family circumstances I target different sets of moment related to employment rates by age, health, marital status and whether the woman has a working partner. Finally, the constant π capturing the value of life in the instantaneous utility function is calibrated to ensure period utility is always positive, by setting the constant equal to the positive of the instantaneous utility function evaluated at the minimum consumption floor for a woman in full-time employment with the minimum level of health. This ensures that agents prefer life to death in all possible states. Since the utility parameters that determine the cost of working are estimated with the method of simulated moments, which relies on an iterative procedure, the constant π is updated with every iteration.

4.5 Parameter Estimates

I present results for the parameters estimated within the model. Table 4.1 shows the estimated parameters for the female earnings process. Health has a large, positive effect on the earnings of both low and high educated women. The estimated coefficients imply that a one standard deviation increase in health increases earnings of low educated women by about 42%, and those of high educated women by about 36%. Table 4.1 also shows estimates of the stochastic process of earnings. The autocorrelation coefficient ρ is estimated to be relatively low when compared to other results in the literature. This is not surprising however, given that persistence in earnings is partly captured by health, which is a highly persistent process itself, and permanent unobserved heterogeneity component θ. The standard deviation of the shocks implies a high degree of uncertainty for next period's earnings draw, and there is heterogeneity in earnings in the initial period, more so for the low educated women.

A key element of the model is the health process. Health is modelled to be persistent, and an important feature of the model is to allow health to be impacted by the woman's labor supply choice. Moreover, the effect of working on future health is heterogeneous depending on the woman's current health status. Table 4.2 reports estimates for the health process. Health is persistent for both education groups. The effect of working on health is very heterogeneous across education groups, and also within education group as a function of the woman's current health. Figure 4.2 shows the effect of working an additional year on the woman's future health as a function of her current health, for both education groups. The left panel shows results for low educated women, the right panel for high educated women. To construct this figure, I evaluate the partial derivative of future health with respect to labor supply at different percentiles of the health distribution, and normalize it by the standard deviation of the health distribution across the whole population. The effect of working on future health of low educated women is very heterogeneous across the distribution of current health. For women in poor health, the effect of working is negative and large. It ranges between -25 % and -12% of a standard deviation for individuals below the 25th percentile of the health distribution. For women at the median, the negative effect of working on health is about 8% of a standard deviation. For low educated women at the very top of the health distribution, the effect of working on health levels off to 0. Overall, the figure shows that the effect of working on health for low educated women is heterogeneous, but mostly negative across the health distribution with the exception of women in very good health. For high educated women the picture looks remarkably different. First, the effect is significantly less heterogeneous across the health distribution. Second, the effects are smaller in magnitude. Nonetheless, a similar pattern emerges where the effect of working is negative for women in poor health, ranging between -3% and -1% of a standard deviation for women in the lower half of the distribution. The effect levels off to about 0 for women in the top 10 percent of the distribution. The standard deviation of the shocks to health implies a considerable degree of uncertainty in next period's health draw.

In Table 4.4, I show the preference parameters that determine the instantaneous utility function in equation (4.3.1). It should be noted that since the function U is negative (because the curvature parameter μ is negative), positive and larger values of the estimated coefficients imply that working is less attractive. The estimated coefficients imply that working full time is costly for all groups. However, being in better health lowers the cost of working for both education groups. This is a standard result in the literature that considers the effect of health on the cost of working (see, for example, De Nardi, Pashchenko, and Porapakkarm (2017)). The utility cost of working is higher for single women than for women in couples. For married women, having a working partner further reduces the utility cost of working, implying a degree of complementarity between the labor supply of couples. This result is analogous to Blundell, Costa Dias, Meghir, and Shaw (2016) and Blundell, Pistaferri, and Saporta-Eksten (2016).

4.6 Model Fit and Implications for Behavior

4.6.1 Model fit

Figures 4.3 (earnings), 4.4 (health) and 4.5 (labor supply) show that the model fits the data well. Earnings are relatively flat over the model period, and decline slightly at older ages. Both health and labor supply decline with age. A well-known puzzle in the retirement literature concerns the difficulty of matching labor market exits around legislated state pension ages, as the financial incentives to retire at these ages are usually small (see, for example, Behaghel and Blau (2012) and Cribb, Emmerson, and Tetlow (2013)). However, my model does a good job at replicating the drop in labor supply at the state pension age, especially so for low educated women. This is visible in the upper panel of Figure 4.5, which shows the employment profiles of women pre-reform for both education groups, for whom the state pension age is 60. I investigate this result further by running a regression that relates a woman's employment status to whether she is above her state pension age.[20] This exercise also serves the purpose

[20] This regression is similar to one proposed in Banks, Cribb, Emmerson, and Sturrock (2019). They use a similar specification as a first stage regression to estimate the effects of employment on a disability index and cognition in a reduced form framework, using ELSA data.

of validating the model. In particular, I run the following regression:

$$E_{i,t} = \alpha^s aboveSPA_{i,t} + \sum_a \delta_a^s (age_{i,t} = a) + \mu_t^s + \gamma^s X_{i,t} + \epsilon_{i,t} \tag{4.6.1}$$

where E is a dummy for being employed, $aboveSPA$ is a dummy equal to 1 if the woman is above the state pension age, and I control for age and interview year dummies, as well as health and marital status in vector X. I run this regression separately for both education groups. Note that the dummy variable $aboveSPA$ is an interaction between the woman's age and the time at which she is interviewed. Thus, the effect of being above the state pension age on employment is identified using a difference-in-differences strategy, comparing women who are the same age but differ in whether or not they are above the state pension age due to the reform. I run this regression on both simulated and observed data, to assess how the model fares in replicating this feature of the data. Table 4.5 shows results for this regression. The model replicates the labor supply response to the state pension age of low educated women almost exactly. The magnitude of the effect is not replicated as well for high educated women, however the model still generates a negative employment response of high educated women to the state pension age. Although I use the reform in the model estimation as a source of variation, I don't target this effect specifically, and the fact that I can replicate the results is encouraging for the model.

The fact that the model can replicate the labor supply response to the state pension age is an interesting result on its own. I investigate the drivers of the large labor supply response to the state pension age seen in the model. Using the structure, I simulate a scenario in which I shift the state pension age of the pre-reform cohort, whose state pension age is 60, to 63. Then, I use this data along with the baseline simulations (with state pension age 60) to estimate the labor supply response of this cohort to the state pension age, for different levels of assets and health in the baseline regime. Specifically, I use this data to run a specification analogous to (4.6.1). I estimate this regression separately for women with positive assets and women with zero or negative assets in the baseline regime, and for women with health below the median or above the median in the baseline regime, combining these categories in order to explore how health and wealth interact in driving the labor supply response to the state pension age in the model.

Table 4.6 shows the effect of being above the state pension age on employment resulting from these regressions. The estimated coefficient captures the effect of being above the state pension age on employment. In the table, I also report the mean employment rate one year prior to the state pension age for all groups, and the ratio between the estimated coefficient and this mean. Panel A refers to low educated women, and panel B refers to high educated women. Columns 1-4 show the response to the state pension age by asset holdings and health status jointly. Poorer women are more likely to adjust their labor supply in response to the state pension age, among both education groups. Among those with low assets (columns 1 and 2), the largest response to the state pension age is from women in poor health. For the high educated, the effect is entirely driven by those in poor health. Among women with positive assets (columns 3 and 4), the drop in employment in percentage terms is substantially higher among women in low health. The drop is larger among the low educated. The results suggest that it is poorer women in poor health who respond the most to the limited financial incentives provided by the state pension by adjusting their labor supply behavior.

4.6.2 Elasticities of labor supply and labor supply response to changes in health

A key contribution of this paper is to provide a framework to study the way in which female health and labor supply interact at older ages. To further investigate the importance of health as a driver of labor supply decisions in the model, I investigate how changes in health affect labor supply decisions. Moreover, I compute labor supply elasticities implied by the model parameters and how they vary across the health distribution. Since labor supply in the model is binary (work full time or not work), all measures of labor supply sensitivity are on the extensive margin.

Frisch elasticities are computed by comparing labor supply profiles in the baseline model with those following a transitory, compensated, anticipated change in net income when in work. Since the change is compensated, the Frisch elasticity isolates the pure income effect (i.e. excluding wealth effects) of the change. Marshall elasticities are computed by perturbing the whole profile of earnings starting from a given age, and comparing simulated labor supply profiles with those of the baseline model. The change in income when working for computing the Marshall elasticity is not anticipated by the agent, and it is uncompensated. As a result of the latter, this measure accounts for wealth effects. Finally, I also compute the response of employment to changes in health. I do this by comparing labor supply profiles in the baseline model with those following an unanticipated change in health (note that due to the persistence in health, a change in health in one period has a permanent effect on the evolution of health over the life cycle). To compute the employment response with respect to changes in health that is comparable to the employment response with respect to a change in earnings captured by the Frisch and Marshallian elasticities, I increase health by the amount that corresponds to a 1% increase in earnings using the model estimates. This amounts to increasing health by 3% of a standard deviation. Thus, the Frisch and Marshallian elasticities show the labor supply response to an increase in earnings that is unrelated to health changes. Instead, the labor supply response with respect to changes in health shows how individuals adjust their labor supply in response to a change in health that generates a change in earnings of the same magnitude, but accounting for other knock on effects that health changes may have on behavior. Tables 4.7 and 4.8 show results for Frisch and Marshall elasticities and the labor supply response to changes in health respectively. Table 4.9 shows the labor supply response in percentage points as implied by the Frisch and Marshallian elasticities and the change in health. I find that both Frisch and Marshall elasticities are in line with existing structural estimates for individuals at older ages (see Keane (2011) for a review of the literature), at about 0.9 and 0.7 respectively. Both elasticities are higher for low educated women and increase with age, as shown in Figure 4.6. The distribution of elasticities by health quartile shows considerable heterogeneity in how agents respond to changes in earnings. Women in poor health respond more strongly, and this is true for both education groups. Overall, the results indicate that labor supply of older women on the extensive margin is responsive to changes in earnings, and that women in poor health are more sensitive to these changes.

Concerning the labor supply response with respect to changes in health, I find that health shocks have a considerably larger effect on the labor supply of older women as compared to what is implied by the Frisch and Marshall elasticities. Table 4.8 shows that labor supply increases by about 3% among the low educated and 2% among the high educated in response to a health shock of a size that generates a 1% increase in earnings. In terms of percentage points, table 4.9 shows that the change in health increases labor supply of low and high educated women by about 1 pp, whereas the change implied by Frisch and Marshallian elasticities is about 0.4 pp for the low educated and about 0.3 pp for the high educated. Moreover, the labor supply response with respect to changes in health rises substantially with age, as shown in figure 4.7. This implies that as people age, health shocks have an increasing impact on individual employment decisions. Table 4.8 also shows that there is considerable heterogeneity in the response to health shocks across the health distribution. For both education groups, agents in the bottom health quartile increase their labor supply around 2 times more than agents in the top health quartile.

Overall, the results suggest a number of findings. First, shocks to health have large effects on female labor supply at older ages. Women respond to changes in health to a greater extent than they do to changes in earnings, for changes of a comparable magnitude. Second, changes in earnings and health shocks lead to larger changes in the labor supply of women who are in poor health relatively to those who are in better health. All in all, health matters more for the employment decision of those who are already in poor health, meaning that health has nonlinear impact on employment. This is mainly because of two reasons. First, poor health increases the probability of dying, with an associated drop in utility from positive when alive to zero when dead. Second, better health decreases the utility cost of working, causing those in poor health to be closer to being indifferent between working and not working.

4.7 Welfare Implications of Increasing the State Pension Age

So far, results have shown that there is significant heterogeneity in the relationship between health and labor supply of women at older ages. On the one hand, employment negatively affects health of women in poor health. On the other hand, health affects employment decisions of women in poor health differently than for women in better health. Thus, we may expect to see differences between women in poor and good health in the way they react to a policy that delays state pension benefits receipt. As a result, health inequality may play a role in generating differences in the welfare consequences of policies that affect employment at older ages. Moreover, since employment has heterogeneous effects on health, the policy may contribute to shaping health inequality. I use the model to produce counterfactual simulations of a revenue-neutral reform that increases the state pension age to 66. This implies that eligibility for state pension benefits is shifted to later in life with respect to the status quo. To achieve revenue neutrality, the public budget gains made by delaying pension benefit payments are redistributed by means of a lump sum transfer to all agents at age 66, i.e. when they become eligible for state pension benefits in the counterfactual scenario. I simulate this reform for women for whom the state pension age is 63 in the baseline regime. To measure the welfare consequences the policy, I compute the compensating variation, i.e. the payment after the reform that would leave the woman as well off as before the reform. I use a measure that is based on the forward-looking value function, as in De Nardi, French, and Jones (2016). In particular, the compensating variation at age 50, CV_{50} is given by

$$V_{50}(a_{50}, \Omega_{50}; \text{current SPA}) = V_{50}(a_{50} + CV_{50}, \Omega_{50}; \text{reformed SPA})$$

where Ω_{50} is the set of state variables at age 50 excluding assets, and $V_{50}(a_{50}, \Omega_{50}; \cdot)$ is the value function evaluated at the state variables, either in the scenario with the current state pension age at 63 or in the scenario with the increased state pension age at 66. I relate the compensating variation to the present value at age 50 of consumption that agents enjoy in the baseline scenario. Moreover, I investigate how employment, consumption and health change in the counterfactual scenario. Tables 4.10 and 4.11 show results for low and high educated women respectively. The tables show how employment and consumption change between ages 50 and 65 in the counterfactual scenario (columns 1 and 2), as well as the change in health at 65 as a percent of a standard deviation of health at that age (column 3). Moreover, the tables show the compensating variation and the ratio of the compensating variation to the present value of consumption at baseline (columns 4 and 5). The results are shown by education group and initial health quartile. Tables 4.10 and 4.11 show that labor supply in the counterfactual scenario increases for most groups, and it increases most for low educated women in poor health. The labor supply of high educated women in good health changes the least with respect to the baseline scenario. Consumption drops for all groups. Low educated women in poor health experience a deterioration in health measured at age 65. It is about 1.4% of a standard deviation of health at that age. This is the biggest health deterioration across education and health groups as a result of the reform, showing that the policy reinforces inequalities in health. Importantly, the health of high educated women in good health does not change in response to the policy counterfactual. The ratio between the compensating variation and the present value of consumption in the baseline regime shows that the reform has negative welfare implications for all women. This is because state pension benefit payments provide insurance against bad earnings shocks at older ages. However, the greatest welfare cost is borne by low educated women in poor health. For them, the compensating variation is about 4% of the present value of consumption in the baseline regime. This is because these women cannot afford to retire before being entitled to pension benefits, but working for them is more costly, as poor health translates into lower earnings, a higher utility cost of working and higher mortality. Moreover, increased employment damages their already poor health. This increases their mortality and since health is persistent, it also makes future labor supply more costly by further reducing earnings capacity and increasing the utility cost of working. Overall, the results show that inequality in health results in differences in the welfare costs of increasing the state pension age, and that this policy may reinforce health inequalities.

4.8 Conclusion

Policies that target employment decisions at older ages may have heterogeneous effects depending on an individual's health status and socio-economic background. Evaluating these policies requires to account for the welfare consequences of health inequality, with the goal of informing welfare-improving policy design.

In this paper, I develop and estimate a rich dynamic structural model of choices over consumption, savings and labor supply of women around retirement age, allowing for a two-way interaction between employment and health. In the model, health may affect employment decisions through its impact on utility, earnings and mortality, while in turn employment may also affect health. I estimate the model using panel data from the UK, exploiting a reform that increased the state pension age of women as a source of exogenous variation.

I find that employment has negative effects on the health of women, and that these effects are stronger for those already in poor health. The negative effect is particularly sizeable among low educated women. Using the model, I estimate women's Frisch and Marshallian labor supply elasticities, as well as the labor supply response to health shocks at older ages. I show that the responses vary substantially across education groups and by health status. Women in poor health respond more strongly to both health shocks and changes to financial incentives. I also show that health shocks cause a larger employment response than changes in financial incentives to work do.

I use the model to simulate a reform that increases the state pension age of women, which reflects recent changes to the UK pension rules. I show that the effects of extending the state pension age are very heterogeneous and tend to widen inequality in health. In particular, the greatest welfare cost is borne by low educated women in poor health. This is because these women cannot afford to retire before being entitled to pension benefits, but working is more costly when in poor health. Moreover, continuing to work is especially damaging for their already poor health, thereby increasing their mortality and making future labor supply more costly.

Overall, the results suggest that there is scope for reducing welfare gaps induced by the state pension age reform by means of redistributive policies that reduce health inequality at older ages.

4.9 Tables

Table 4.1: Female earnings equation

	Low educated	High educated
Intercept	7.347	8.158
Age	0.030	0.010
Age^2	-0.007	-0.006
Health	1.010	0.873
ρ	0.527	0.558
SD innovation in productivity	0.357	0.440
SD initial productivity	0.241	0.105

Notes: Low educated refers to women without a high school degree. High educated refers to women with a high school degree or higher. Age is scaled such that the initial age $t = 50$ is equal to 0. The population mean initial productivity is zero.

Table 4.2: Female health process

	Low educated	High educated
Intercept	0.441	0.221
Health	0.739	0.899
Employment	-0.105	-0.010
Employment X Health	0.054	0.004
Age	-0.002	-0.002
Has partner	0.018	0.018
SD innovation	0.120	0.109

Notes: Low educated refers to women without a high school degree. High educated refers to women with a high school degree or higher. Age is scaled such that the initial age $t = 50$ is equal to 0.

Table 4.3: Permanent unobserved heterogeneity in earnings and health

	Low educated	High educated
θ_{low}	-1.001	-1.715
κ_{low}	-0.210	-0.001
Pr($\theta = \theta_{low}, \kappa = \kappa_{low}$)	0.197	0.166
Pr($\theta = \theta_{low}, \kappa = \kappa_{high}$)	0.077	0.057
Pr($\theta = \theta_{high}, \kappa = \kappa_{low}$)	0.046	0.056
Pr($\theta = \theta_{high}, \kappa = \kappa_{high}$)	0.680	0.721

Notes: Low educated refers to women without a high school degree. High educated refers to women with a high school degree or higher.

Table 4.4: Utility cost of working

	Low educated	High educated
Cost of working	0.520	0.601
Health	-0.050	-0.064
Partner	-0.061	-0.069
Partner works	-0.220	-0.202

Notes: Low educated refers to women without a high school degree. High educated refers to women with a high school degree or higher.

Table 4.5: Effect of being above state pension age on employment: Model vs. data

	Employment			
	Low educated		High educated	
	Model	Data	Model	Data
Above SPA	-0.110***	-.116***	-0.039***	-.139***
	(0.004)	(0.045)	(0.004)	(0.034)

Notes: This table shows results from regressing a binary variable for employment on whether a woman is above her state pension age, comparing results between simulated and observed data for both education groups. Other controls: health, marital status, dummies for age, dummies for interview year (the latter in the observed data). Regression includes all women aged 60 to 63. Low educated refers to women without a high school degree. High educated refers to women with a high school degree or higher.

Table 4.6: Effect of being above state pension age on employment: Model

		(1)	(2)	(3)	(4)
Low educated	aboveSPA	-0.118***	-0.539***	-0.038***	-0.097***
		(0.004)	(0.009)	(0.002)	(0.004)
	Mean E 1 yr before SPA	0.142	0.734	0.063	0.486
	Ratio coeff to mean E	-0.829	-0.734	-0.595	-0.199
High educated	aboveSPA	-0.183***	0.023	-0.039***	-0.013**
		(0.013)	(0.021)	(0.005)	(0.005)
	Mean E 1 yr before SPA	0.264	0.276	0.250	0.564
	Ratio coeff. to mean E	-0.694	0.082	-0.155	-0.023
	Group	Low assets, low health	Low assets, high health	High assets, low health	High assets, high health

Notes: This table shows results from regressing a binary variable for employment on whether a woman is above her state pension age, using simulated data. Low/high health refers to women with health below/above the population median in the baseline regime. Low/high assets refers to women with negative or zero/positive assets in the baseline regime. Panel A shows results for women with no high school degree; panel B shows results for women with a high school degree or higher. All regressions control for health, assets, dummies for age and marital status. Regressions include all women aged 60 to 63.

Table 4.7: Elasticities of labor supply

	Frisch		Marshall	
All women	0.855		0.690	
	Low educated	High educated	Low educated	High educated
	1.114	0.657	0.969	0.477
Health quartile				
Bottom	3.508	1.510	2.369	1.069
Third	1.854	0.845	1.482	0.569
Second	0.890	0.613	0.863	0.468
Top	0.790	0.335	0.733	0.262

Notes: The table shows Frisch and Marshall elasticities of labor supply for women by education group, and by health quartile. All effects are measured in the year the change in earnings occurs. Low educated refers to women without a high school degree. High educated refers to women with a high school degree or higher.

Table 4.8: Percentage change in employment in response to change in health

All women	2.207	
	Low educated	High educated
	2.675	1.850
Health quartile		
Bottom	5.513	3.048
Third	3.522	1.826
Second	2.545	1.804
Top	2.211	1.582

Notes: The table shows the female labor supply response in percentage terms to an increase in health by education group, and by health quartile. The health increase is of a magnitude that would generate a 1% increase in earnings using the model estimates. This amounts to increasing health by 3% of a standard deviation. All effects are measured in the year the change in health occurs. Low educated refers to women without a high school degree. High educated refers to women with a high school degree or higher.

Table 4.9: Labor supply response in pp to change in earnings vs. health

Low educated			High educated		
Frisch	Marshall	Health	Frisch	Marshall	Health
0.478	0.415	1.146	0.375	0.272	1.056

Notes: The table shows percentage point changes in labor supply in response to a 1% increase in net earnings (columns 1-2 and 4-5), and in response to a health increase of magnitude that would generate a 1% increase in earnings using the model estimates (column 3 and 6) (corresponding to 3% of a standard deviation). All effects are measured in the year the change in earnings/health occurs. Low educated refers to women without a high school degree. High educated refers to women with a high school degree or higher.

Table 4.10: Counterfactual experiment increasing the state pension age, revenue neutral reform, low educated women

		Employment ages 50-65	Consumption ages 50-65	Health at 65 (SD)	CV (£)	CV/PV consumption in bl
All women		2.9%	-1.4%	-0.8%	£ 7,019	0.034
Health quartile	Bottom	5.2%	-1.9%	-1.4%	£ 6,559	0.043
	Third	2.1%	-1.4%	-0.5%	£ 7,029	0.027
	Second	2.2%	-1.3%	-0.5%	£ 7,670	0.030
	Top	1.2%	-0.9%	-0.3%	£ 7,164	0.028

Notes: Counterfactual experiment under revenue neutrality achieved with a lump sum payment distributed at age 66. Columns 1 and 2 show the employment and consumption changes between ages 50 and 65 in the counterfactual scenario versus the baseline regime. Column 3 shows the change in health at age 65 as a percentage of the standard deviation of health at age 65 in the counterfactual scenario. Columns 4 and 5 show the compensating variation in pounds (2012 prices), and the ratio between the compensating variation and the present value of consumption in the baseline regime. The health quartiles are computed based on the distribution of initial health in the baseline regime. Low educated refers to women without a high school degree.

Table 4.11: Counterfactual experiment increasing the state pension age, revenue neutral reform, high educated women

		Employment ages 50-65	Consumption ages 50-65	Health at 65 (SD)	CV (£)	CV/PV consumption in bl
All women		0.2%	-1.8%	0.0%	£15,182	0.036
Health quartile	Bottom	1.6%	-2.0%	-0.1%	£12,958	0.037
	Third	-0.2%	-1.8%	0.0%	£14,640	0.036
	Second	0.1%	-1.8%	0.0%	£16,957	0.035
	Top	0.0%	-1.6%	0.0%	£15,569	0.035

Notes: Counterfactual experiment under revenue neutrality achieved with a lump sum payment distributed at age 66. Columns 1 and 2 show the employment and consumption changes between ages 50 and 65 in the counterfactual scenario versus the baseline regime. Column 3 shows the change in health at age 65 as a percentage of the standard deviation of health at age 65 in the counterfactual scenario. Columns 4 and 5 show the compensating variation in pounds (2012 prices), and the ratio between the compensating variation and the present value of consumption in the baseline regime. The health quartiles are computed based on the distribution of initial health in the baseline regime. High educated refers to women with a high school degree or higher.

4.10 Figures

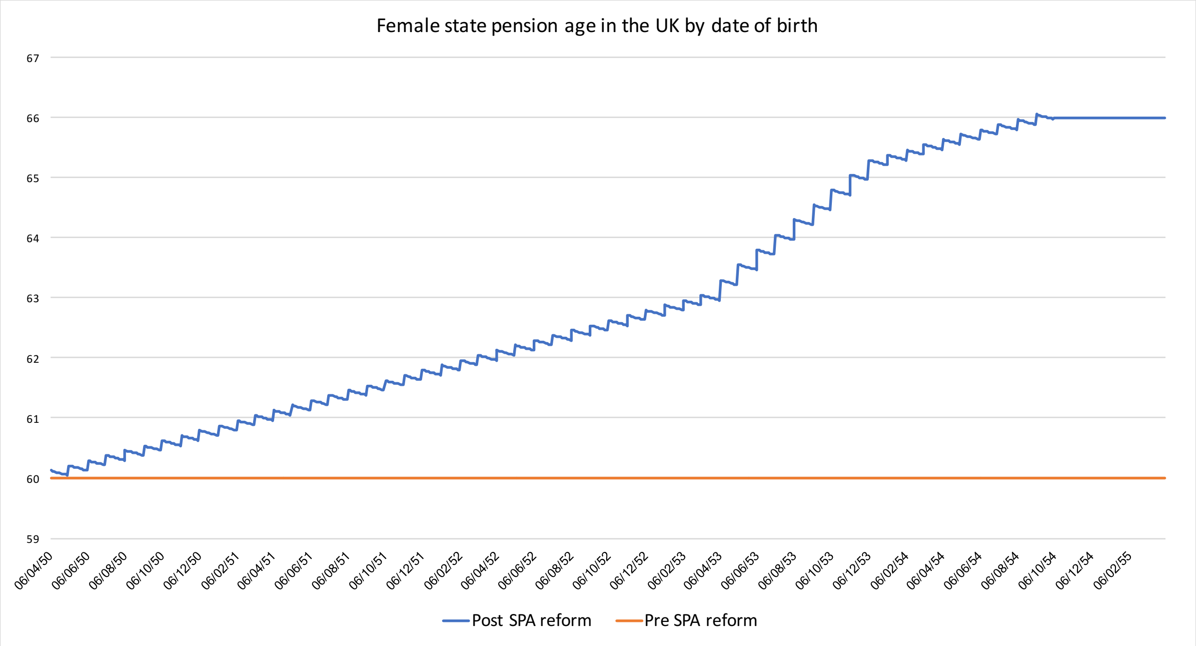


Figure 4.1: Female state pension age following the state pension age reform
Notes: The figure shows the female state pension age as a function of a woman's date of birth. The state pension age increases following a 'sawtooth' pattern. This is because women born in a given month become eligible for the state pension on the same date, irrespective of their day of birth. As a result, women born later in the month have a slightly lower state pension age relative to those born early in the same month.

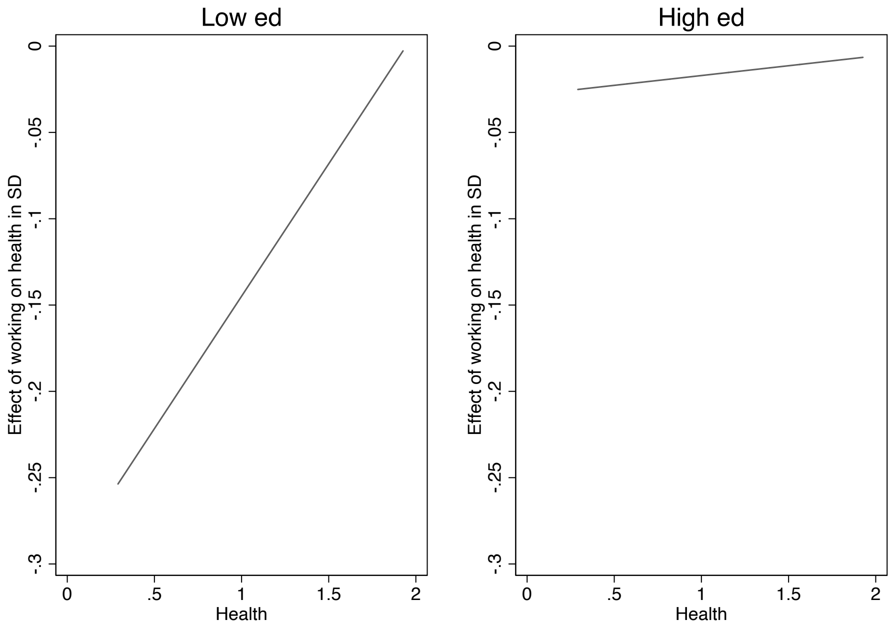


Figure 4.2: Effect of one additional year of work on health as a function of the woman's health stock
Notes: Low educated refers to women without a high school degree. High educated refers to women with a high school degree or higher.

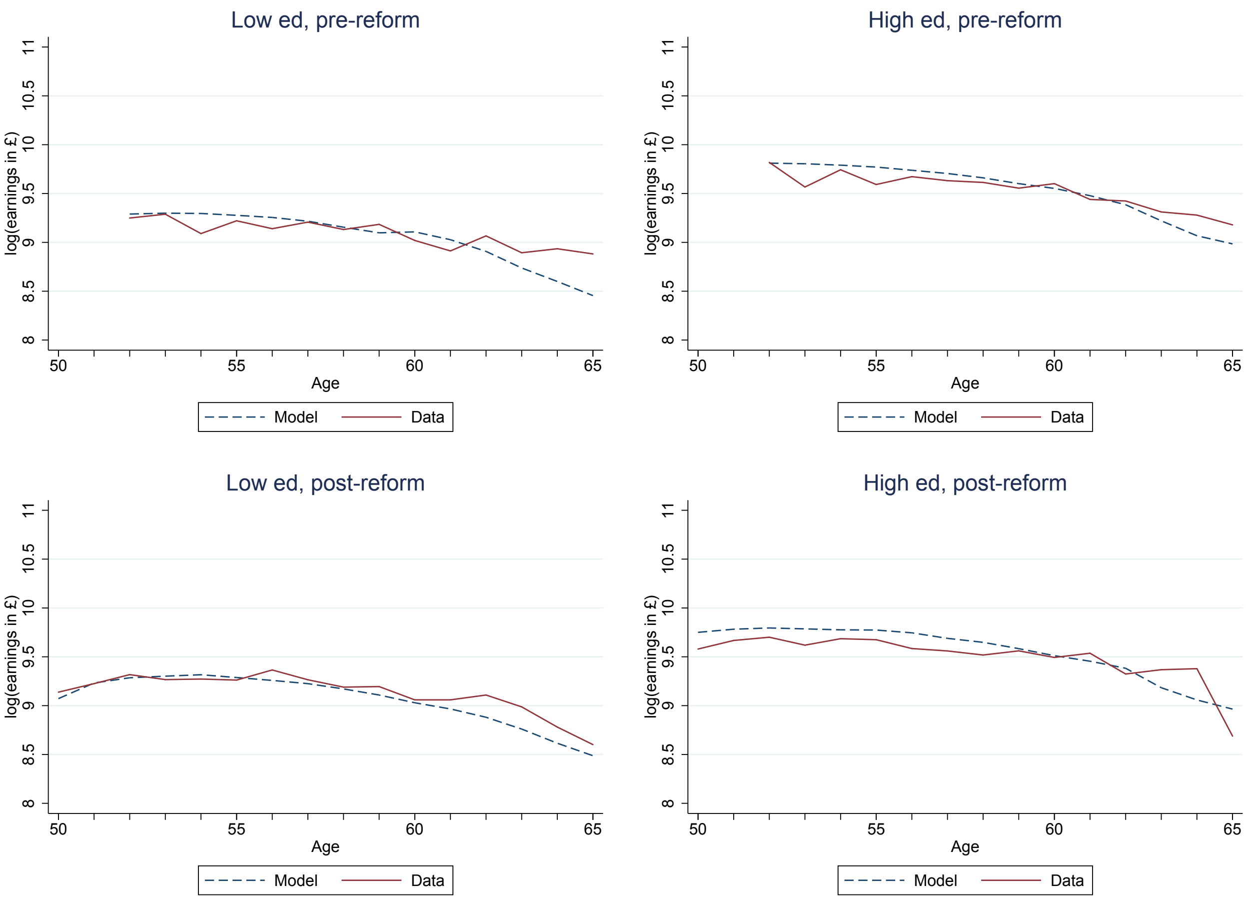


Figure 4.3: Mean log earnings for working women over the life-cycle by education and cohort: data versus model.
Notes: ELSA versus simulated data, in solid and dashed lines, respectively. 2012 prices. Low educated refers to women without a high school degree. High educated refers to women with a high school degree or higher.

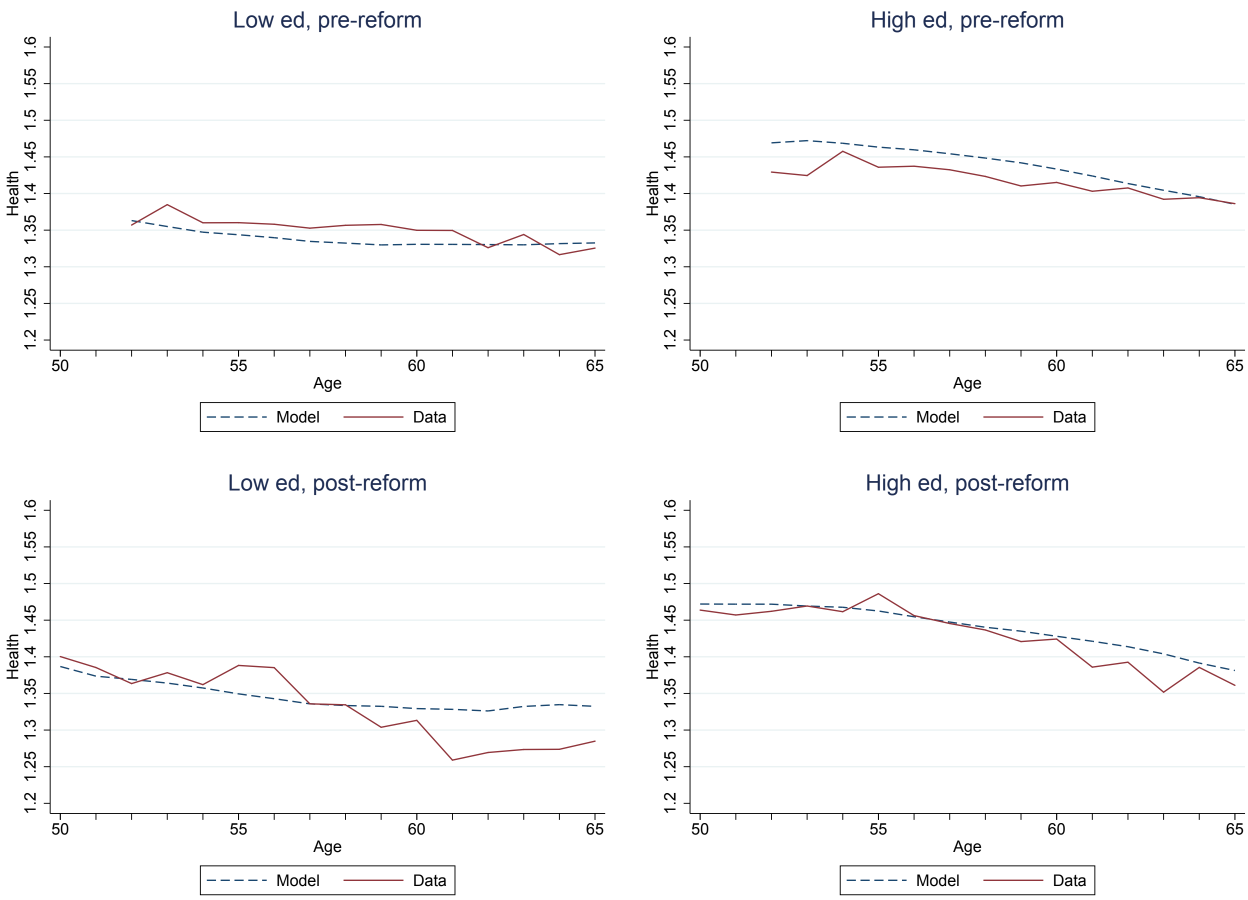


Figure 4.4: Mean health over the life-cycle by education and cohort: data versus model.
Notes: ELSA versus simulated data, in solid and dashed lines, respectively. Low educated refers to women without a high school degree. High educated refers to women with a high school degree or higher.

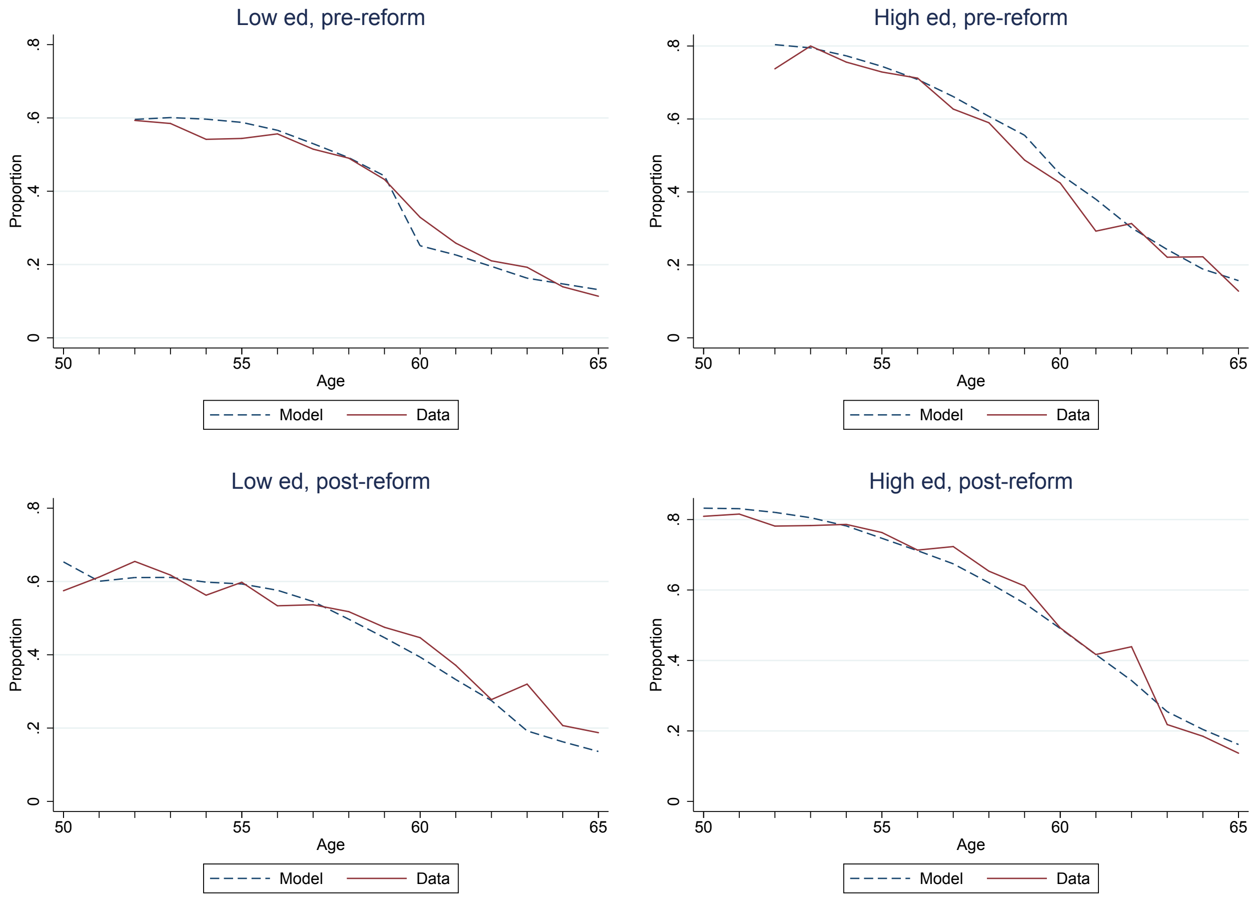


Figure 4.5: Mean employment rates over the life-cycle by education and cohort: data versus model.
Notes: ELSA versus simulated data, in solid and dashed lines, respectively. Low educated refers to women without a high school degree. High educated refers to women with a high school degree or higher.

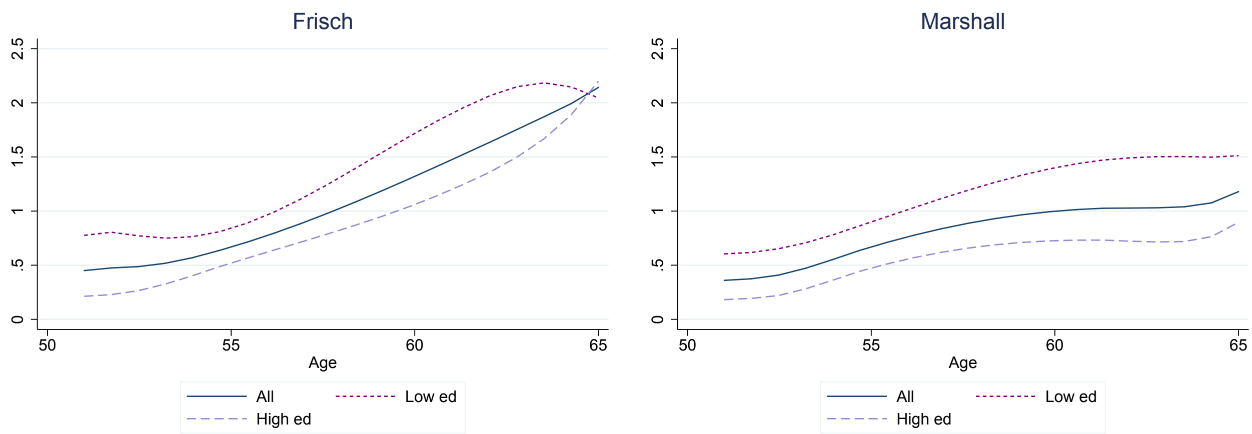


Figure 4.6: Frisch and Marshallian elasticities over the life-cycle of women by education, based on simulated data
Notes: Low educated refers to women without a high school degree. High educated refers to women with a high school degree or higher.

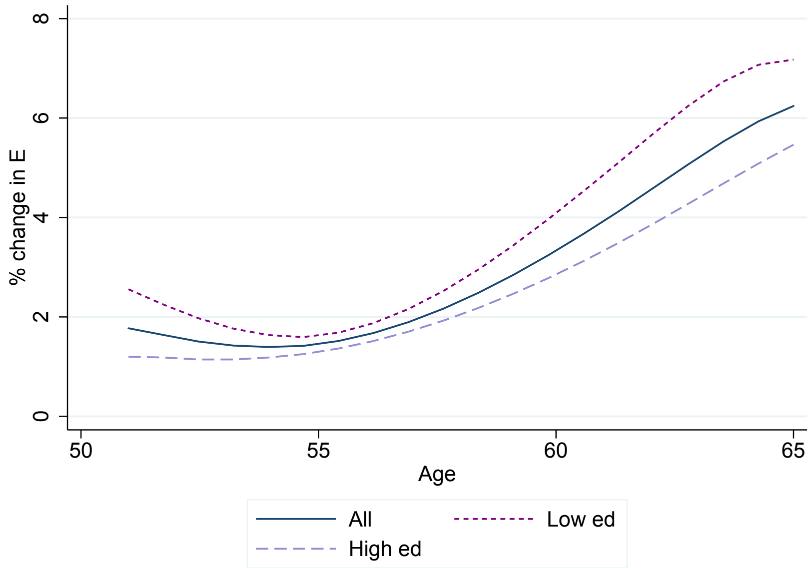


Figure 4.7: Percent change in labor supply with respect to changes in health over the life-cycle of women by education, based on simulated data
Notes: Low educated refers to women without a high school degree. High educated refers to women with a high school degree or higher.

4.11 Appendix

4.11.1 Construction of health index

In this paper, health is a single, continuous index constructed by combining information on subjective and objective health measures collected in every wave of the ELSA survey. To construct the index, I use both subjective and objective health measures, following a procedure proposed in Blundell, Britton, Costa Dias, and French (2020) which I describe in detail in the main body of the paper. Table 4.12 reports all measures included to derive the health index.

Table 4.12: Subjective and objective health measures for health index

Subjective measures	Self-reported pain level on a 4-item scale (no pain-severe pain)
	Self-reported health status on a 5-item scale (very good-very bad)
	Dummy for whether considers health to be limiting of activities
Objective measures	Doctor diagnosed chronic lung disease; asthma; arthritis; osteoporosis; cancer;
	Parkinson's disease; Alzheimer's disease; dementia; high blood pressure;
	angina pectoris; heart attack; congestive heart failure; heart murmur;
	abnormal heart rhythm; diabetes; stroke
	Eyesight on a scale 1-5 (excellent-poor)
	Hearing on a scale 1-5 (excellent-poor)
	ADL: difficulty bathing or showering; difficulty getting in and out of bed;
	difficulty dressing; difficulty eating; difficulty doing work around house;
	difficulty using a map to figure out how to get around a place;
	difficulty taking medications; difficulty managing money;
	difficulty making telephone calls; difficulty preparing a hot meal;
	difficulty shopping for groceries; difficulty walking across a room;
	difficulty using the toilet; difficulty getting up from chair after sitting long time;
	difficulty climbing one flight of stairs without resting;
	difficulty climbing several flights of stairs without resting;
	difficulty lifting or carrying weights over 10 pounds;
	difficulty picking up 5p coin from table;
	difficulty pulling or pushing large objects;
	difficulty reaching or extending arms above shoulder levels;
	difficulty sitting 2 hours; difficulty stooping, kneeling or crouching;
	difficulty walking 100 yards

4.11.2 Evidence on the role of disability benefits

In figures 4.8 to 4.9, I show the prevalence of pathways to retirement of women in my sample that highlight the small role disability benefit receipt plays. Figure 4.8 shows the prevalence of pathways to retirement of women who are not on disability benefits in wave 1. They represent 96% of the sample. A very small proportion of these women retires on disability benefits. In particular, even among women in this group who are out of work in wave 1, relatively few end up retiring using the disability benefit path, as shown in figure 4.9. Finally, figure 4.10 shows the of pathways to retirement among women who are on disability benefits in wave 1 (4% of the total sample). Among them, very few get back to work over the period.

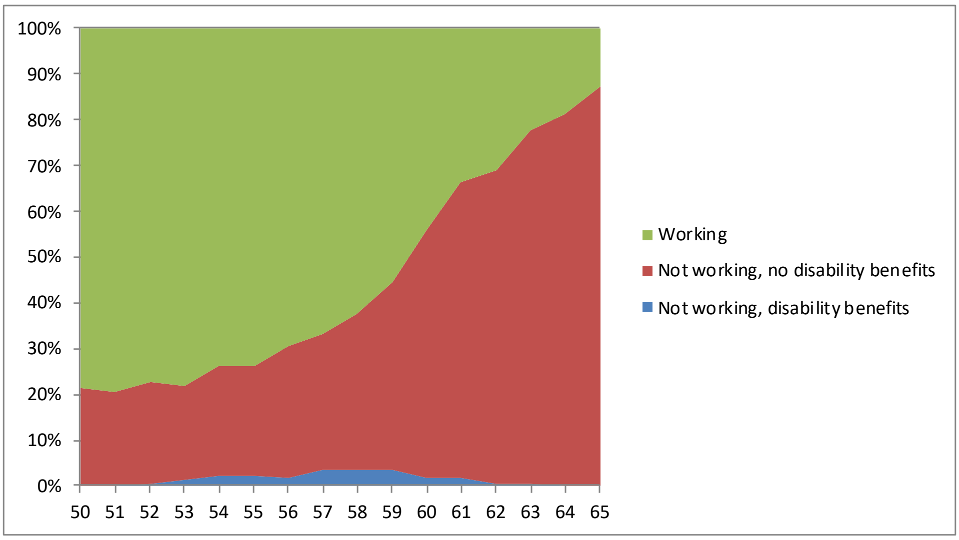


Figure 4.8: Prevalence of pathways to retirement of women not on disability benefits in Wave 1 of ELSA

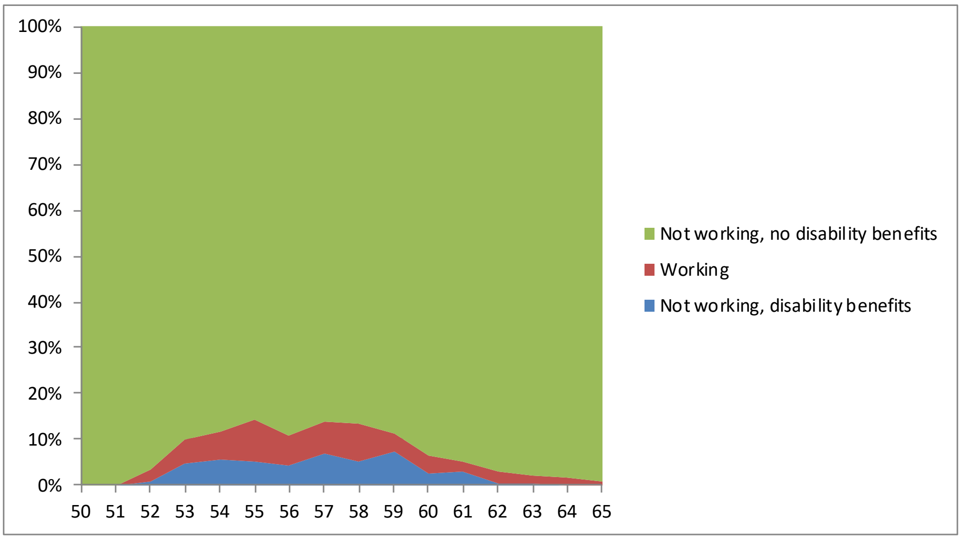


Figure 4.9: Prevalence of pathways to retirement of women not on disability benefits and not in work in Wave 1 of ELSA

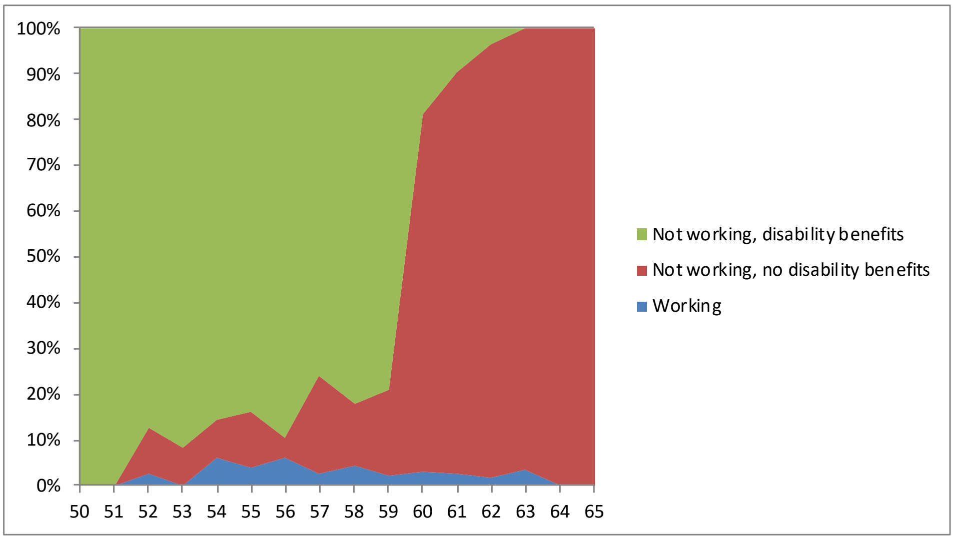


Figure 4.10: Prevalence of pathways to retirement of women on disability benefits in Wave 1 of ELSA

4.11.3 Taxes and benefits for the elderly

To capture the budget constraint in detail, I use FORTAX, a tax and benefit simulation tool which contains a detailed description of all taxes and benefits in the UK targeted towards individuals before they reach the state pension age. However, FORTAX does not account for the special treatment that the elderly receive. In this section, I describe the specific features of the tax and benefit system for the elderly that I augment FORTAX with.

Taxes

Employee National Insurance Employee National Insurance contributions are levied on earnings from work. Only those aged less than the state pension age pay National Insurance contributions. Hence, these are set to 0 for individuals at or past the state pension age.

Income tax Income tax is levied on the sum of earnings, pension income (state and private if available), less any contributions to private pensions (if available). Taxes are levied at the individual level. In the UK, only income above the so-called personal allowance is taxed. The generosity of the personal allowance changes with age, with a more generous treatment for older individuals. Table 4.13 shows the weekly personal allowance for women under the 2006/07 regime and under the 2012/13 respectively, in 2012 prices.

Table 4.13: Income tax personal allowance (weekly), in 2012 prices

	Age		
	<65	65-74	>75
2006/07	119.5	172.8	176.1
2012/13	155.9	201.9	205

Benefits

State pension benefits in the UK comprise state pension income and a means-tested income floor called pension credit. Both are payable from the state pension age until death. I model both benefits following O'Dea (2018). State pension income is modelled as a function of average career earnings accumulated by the state pension age. Pension credit is a function of 'notional' income, which is used to assess entitlement to the benefit. Notional income includes earnings, state and private pension income, as well as an imputed stream of income from non-pension wealth (computed as 10% of the stock of non-pension wealth, less the first 10,000 £). Entitlement to pension credit in the model is given by:

$$pc(y^{notional}) = max(GC - min(y^{notional}, SC) - t(max(y^{notional} - SC, 0)), 0) \quad \text{if age} \geq \text{spa} \tag{4.11.1}$$

$$pc(y^{notional}) = 0 \quad \text{if age} < \text{spa} \tag{4.11.2}$$

where GC is the 'Guarantee Credit Level', i.e. the minimum income guaranteed to all individuals above the state pension age; SC is the 'Savings Credit Threshold', i.e. the income level up to which pension credit is withdrawn at a 100% tax rate; t is the taper rate, i.e. the effective tax rate applied on notional income above SC (equal to 40%). Table 4.14 shows the values of GC and SC for single and married women.

Table 4.14: Pension credit parameters, in 2012 prices

	Singles	Couples
GC	7,400	11,300
SC	5,800	9,200

4.11.4 Parameters estimated outside the model

In this section, I present results for parameters estimated outside the model. These include i) all processes regarding the male partner, ii) survival probabilities of women and men, iii) functions relating average career earnings to public and private pension income (if enrolled in a private pension scheme).

Male employment and labor supply

Table 4.15 reports estimated coefficients from a probit regression of male employment. Table 4.16 reports parameters governing the male earnings process. This is only relevant for women in couples, for whom male employment and income affect the household budget constraint and thereby the woman's employment decision. Male employment is estimated to be very persistent and decreasing with age. In order to obtain the persistence parameter in earnings shocks ρ, I take the square root of the bi-annual persistence parameter estimated by regressing earnings residuals at time t on earnings residuals at time $t-2$. To derive the standard deviation of the innovation to earnings shocks, I use relationship

$$\sigma_\eta = \sqrt{\frac{\sigma^2_{\tilde{\eta}}}{1+\rho^2}} \tag{4.11.3}$$

where $\tilde{\eta}$ is the residual of the earnings shocks regression over two periods.

Table 4.15: Exogenous parameters: male partner employment by woman's education

	Low educated	High educated
Intercept	2.35	2.90
	(0.37)	(0.30)
Previously employed	2.57	2.51
	(0.07)	(0.06)
Woman's age	-0.07	-0.08
	(0.01)	(0.004)

Table 4.16: Exogenous parameters: male partner earnings by woman's education

	Low educated	High educated
Intercept	0.87	10.75
	(3.13)	(2.57)
Woman's age	0.34	0.001
	(0.11)	(0.09)
Woman's age^2	-0.003	-0.0002
	(0.001)	(0.0008)
ρ	0.72	0.69
St. deviation innovation to productivity	0.47	0.52

Survival probabilities

In the model, women face mortality risk. Married women also face risk over their partner's survival. ELSA contains information on whether or not an individual is alive at every wave, and on their date of death. I use this information to run logit regressions of being alive on a woman's health and age. For male partners, I run these regressions on the woman's age. All regressions are estimated separately by education group. The estimated parameters are shown in Tables 4.17 and 4.18.

Table 4.17: Female survival equation

	Low educated	High educated
Intercept	3.963	3.328
Age	-0.065	-0.066
Health	1.529	2.631

Table 4.18: Male survival equation

	Low educated	High educated
Intercept	8.559	8.469
Woman's age	-0.066	-0.054

4.11.5 Model fit

In this section, I show the complete list of data moments used in the estimation, along with their simulated counterparts and the difference between the two, normalized by the data standard error.

Table 4.19: Log earnings at the start of working life, auxilliary regression

Moment	Data	Simulated	SE data	No. SE diff	Data	Simulated	SE data	No. SE diff
	Low educated				High educated			
Intercept	8.845	8.977	0.215	0.615	9.385	9.470	0.168	0.506
Health	0.269	0.088	0.148	1.225	0.188	0.195	0.113	0.056

Table 4.20: Distribution of log earnings at the start of working life

Moment	Data	Simulated	SE data	No. SE diff	Data	Simulated	SE data	No. SE diff
	Low educated				High educated			
Variance of residuals	0.393	0.215	0.026	6.854	0.477	0.388	0.021	4.175
Skewness of residuals	-0.438	-0.141	0.106	2.807	-0.864	-0.641	0.066	3.361
Kurtosis of residuals	3.373	0.220	0.223	14.167	3.973	1.293	0.198	13.571

Table 4.21: Log earnings, auxilliary regression

Moment	Data	Simulated	SE data	No. SE diff	Data	Simulated	SE data	No. SE diff
	Low educated				High educated			
Intercept	8.655	8.634	0.302	0.068	8.860	9.778	0.249	3.694
Age-50	0.035	0.038	0.028	0.135	0.048	0.031	0.021	0.812
$(\text{Age-}50)^2$	-0.004	-0.006	0.002	1.610	-0.005	-0.005	0.001	0.605
Health	0.354	0.402	0.194	0.246	0.442	-0.011	0.159	2.840
Health residuals	-0.302	-0.134	0.214	0.784	-0.219	-0.147	0.165	0.436

Table 4.22: Distribution of log earnings

Moment	Data	Simulated	SE data	No. SE diff	Data	Simulated	SE data	No. SE diff
	Low educated				High educated			
Variance of residuals	0.398	0.249	0.020	7.586	0.505	0.468	0.019	1.925
Skewness of residuals	-0.080	-0.092	0.025	0.485	-0.239	-0.329	0.024	3.690
Kurtosis of residuals	0.508	0.284	0.053	4.240	0.903	1.131	0.063	3.591
Autocorrelation of residuals	0.205	0.108	0.022	4.529	0.238	0.225	0.020	0.598
Mean	9.151	9.128	0.019	1.222	9.576	9.653	0.014	5.286

Table 4.23: Health, auxilliary regression

Moment	Data	Simulated	SE data	No. SE diff	Data	Simulated	SE data	No. SE diff
	Low educated				High educated			
Intercept	0.369	-0.220	0.053	11.206	0.451	0.396	0.039	1.430
Lagged health	0.759	1.023	0.018	14.865	0.815	0.808	0.013	0.482
Age	-0.001	0.003	0.001	5.455	-0.004	-0.003	0.001	2.145
Married	0.014	0.018	0.007	0.522	0.019	0.028	0.004	2.189

Table 4.24: Health distribution

Moment	Data	Simulated	SE data	No. SE diff	Data	Simulated	SE data	No. SE diff
	Low educated				High educated			
Variance of residuals	0.029	0.031	0.001	1.825	0.018	0.022	0.001	4.432
Skewness of residuals	-0.004	-0.001	0.001	5.527	-0.003	-0.000	0.000	11.063
Kurtosis of residuals	0.005	0.003	0.001	2.374	0.002	0.001	0.000	4.009
Autocorrelation of residuals	-0.007	-0.002	0.001	4.954	-0.004	0.000	0.000	10.764
Mean	1.343	1.346	0.006	0.530	1.426	1.438	0.004	2.724
Correlation health, age	-0.005	-0.002	0.001	3.577	-0.006	-0.007	0.001	1.766

Table 4.25: Difference across cohorts in change in employment and health between ages 62-60

Moment	Data	Simulated	SE data	No. SE diff	Data	Simulated	SE data	No. SE diff
	Low educated				High educated			
Employment	-0.050	-0.054	0.056	0.071	0.058	-0.001	0.037	1.584
Health	-0.020	-0.003	0.029	0.567	-0.024	0.006	0.016	1.885

Table 4.26: Employment

Moment	Data	Simulated	SE data	No. SE diff	Data	Simulated	SE data	No. SE diff
	Low educated				High educated			
Ages 50-54	0.596	0.594	0.018	0.121	0.788	0.806	0.011	1.646
Ages 55-57	0.545	0.558	0.016	0.786	0.716	0.706	0.012	0.829
Ages 58-60	0.434	0.401	0.015	2.173	0.543	0.544	0.012	0.115
Ages 61-64	0.216	0.195	0.011	1.935	0.285	0.287	0.010	0.225
Ages 65-69	0.084	0.098	0.007	2.082	0.097	0.087	0.007	1.385
Single, ages 50-54	0.496	0.603	0.040	2.667	0.762	0.775	0.021	0.631
Married, ages 50-54	0.627	0.591	0.020	1.761	0.798	0.818	0.013	1.556
Single, ages 55-57	0.480	0.613	0.033	4.069	0.702	0.696	0.022	0.304
Married, ages 55-57	0.568	0.539	0.019	1.559	0.721	0.710	0.013	0.826
Single, ages 58-60	0.409	0.424	0.031	0.492	0.543	0.551	0.023	0.336
Married, ages 58-60	0.443	0.392	0.018	2.851	0.543	0.542	0.013	0.061
Single, ages 61-64	0.223	0.134	0.020	4.339	0.320	0.290	0.021	1.501
Married, ages 61-64	0.213	0.219	0.013	0.526	0.272	0.286	0.012	1.172
Single, ages 65-69	0.104	0.067	0.013	2.953	0.123	0.080	0.015	2.938
Married, ages 65-69	0.074	0.113	0.008	4.960	0.085	0.090	0.008	0.563
Ages 50-54, partner works	0.695	0.635	0.023	2.624	0.848	0.855	0.012	0.623
Ages 55-57, partner works	0.656	0.599	0.022	2.627	0.800	0.778	0.013	1.586
Ages 58-60, partner works	0.554	0.507	0.023	2.072	0.662	0.645	0.016	1.086
Ages 61-64, partner works	0.359	0.340	0.024	0.819	0.437	0.386	0.020	2.560
Ages 65-69, partner works	0.199	0.194	0.027	0.169	0.207	0.154	0.025	2.105
Pre-reform, ages 50-54	0.568	0.590	0.029	0.763	0.765	0.800	0.024	1.456
Post-reform, ages 50-54	0.607	0.613	0.023	0.233	0.793	0.815	0.012	1.830
Pre-reform, ages 55-57	0.537	0.555	0.021	0.829	0.686	0.704	0.019	0.974
Post-reform, ages 55-57	0.556	0.570	0.026	0.544	0.734	0.710	0.014	1.645
Pre-reform, ages 58-60	0.411	0.392	0.018	1.072	0.496	0.537	0.016	2.566
Post-reform, ages 58-60	0.481	0.443	0.029	1.304	0.592	0.555	0.017	2.147
Pre-reform, ages 61-64	0.197	0.187	0.011	0.929	0.259	0.277	0.012	1.510
Post-reform, ages 61-64	0.310	0.239	0.030	2.329	0.342	0.302	0.019	2.087
Pre-reform, ages 65-69	0.083	0.099	0.007	2.518	0.096	0.085	0.007	1.372
Post-reform, ages 65-69	0.176	0.087	0.054	1.650	0.116	0.088	0.026	1.057
First health quartile	0.114	0.038	0.009	8.142	0.257	0.259	0.012	0.184
Second health quartile	0.288	0.302	0.012	1.133	0.442	0.445	0.011	0.205
Third health quartile	0.356	0.390	0.014	2.473	0.497	0.564	0.011	6.150
Fourth health quartile	0.441	0.578	0.014	9.539	0.611	0.703	0.012	7.396

Table 4.27: Correlation employment and age

Moment	Data	Simulated	SE data	No. SE diff	Data	Simulated	SE data	No. SE diff
	Low educated				High educated			
Correlation	-0.040	-0.043	0.002	1.514	-0.052	-0.053	0.001	1.001

Table 4.28: Correlation employment and health

Moment	Data	Simulated	SE data	No. SE diff	Data	Simulated	SE data	No. SE diff
	Low educated				High educated			
Correlation	0.536	0.470	0.031	2.157	0.606	0.604	0.035	0.075

Chapter 5

Bibliography

Agostinelli, F. and M. Wiswall (2016). Estimating the technology of children's skill formation. Technical report, National Bureau of Economic Research.

Aizer, A. and F. Cunha (2012). The production of child human capital: Endowments, investments, and fertility. *NBER Working Paper 18429*.

Anderson, P. (2002). Assessment and development of executive function (ef) during childhood. *Child neuropsychology 8*(2), 71–82.

Anderson, T. W. and H. Rubin (1956). Statistical inference in factor analysis. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, Volume 5, pp. 1.

Araujo, M. C., P. Carneiro, Y. Cruz-Aguayo, and N. Schady (2016). Teacher quality and learning outcomes in kindergarten. *The Quarterly Journal of Economics 131*(3), 1415–1453.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020). Estimating the production function for human capital: results from a randomized controlled trial in colombia. *American Economic Review 110*(1), 48–85.

Attanasio, O., H. Low, and V. Sánchez-Marcos (2008). Explaining changes in female labor supply in a life-cycle model. *American Economic Review 98*(4), 1517–52.

Attanasio, O., C. Meghir, and E. Nix (2020). Human capital development and parental investment in india. *The Review of Economic Studies 87*(6), 2511–2541.

Aylward, G. P. and T. Stancin (2008). Measurement and psychometric considerations. *Developmental-Behavioral Pediatrics*.

Bailey, D. H., G. J. Duncan, F. Cunha, B. R. Foorman, and D. S. Yeager (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest 21*(2), 55–97.

Bandiera, O., I. Barankay, and I. Rasul (2005). Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics 120*(3), 917–962.

Banks, J., R. Blundell, and C. Emmerson (2015). Disability benefit receipt and reform: reconciling trends in the united kingdom. *Journal of Economic Perspectives 29*(2), 173–90.

Banks, J., J. Cribb, C. Emmerson, and D. Sturrock (2019). The impact of work on cognition and physical disability: Evidence from English women. IFS Working Paper.

Banks, J., C. Emmerson, and G. C. Tetlow (2014). Effect of pensions and disability benefits on retirement in the uk. Technical report, National Bureau of Economic Research.

Barrera-Osorio, F., K. Gonzalez, F. Lagos, and D. J. Deming (2020). Providing performance information in education: An experimental evaluation in colombia. *Journal of Public Economics 186*, 104185.

Behaghel, L. and D. M. Blau (2012). Framing social security reform: Behavioral responses to changes in the full retirement age. *American Economic Journal: Economic Policy 4*(4), 41–67.

Berlinski, S. and N. Schady (2015). The early years: Child well-being and the role of public policy.

Bharadwaj, P., K. V. Løken, and C. Neilson (2013). Early life health interventions and academic achievement. *The American Economic Review 103*(5), 1862–1891.

Blackwell, L. S., K. H. Trzesniewski, and C. S. Dweck (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child development 78*(1), 246–263.

Blau, D. M. and D. B. Gilleskie (2001). The effect of health on employment transitions of older men. *Worker Wellbeing in a Changing Labor Market*, 35–65.

Blundell, R., M. Borella, J. Commault, and M. De Nardi (2020). Why Does Consumption Fluctuate in Old Age and How Should the Government Insure It? NBER Working Papers 27348, National Bureau of Economic Research, Inc.

Blundell, R., J. Britton, M. Costa Dias, and E. French (2020). The impact of health on labor supply near retirement. *forthcoming in Journal of Human Resources*.

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

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2016). Consumption inequality and family labor supply. *American Economic Review 106*(2), 387–435.

Booij, A. S., E. Leuven, and H. Oosterbeek (2017). Ability peer effects in university: Evidence from a randomized experiment. *The review of economic studies 84*(2), 547–578.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019). Beliefs about gender. *American Economic Review 109*(3), 739–73.

Bound, J. (1991). Self-reported versus objective measures of health in retirement models. *Journal of Human Resources 26*(1).

Bound, J., M. Schoenbaum, T. R. Stinebrickner, and T. Waidmann (1999). The dynamic effects of health on the labor force transitions of older workers. *Labour Economics 6*(2), 179–202.

Bound, J., T. Stinebrickner, and T. Waidmann (2010). Health, economic resources and the work decisions of older men. *Journal of Econometrics 156*(1), 106–129.

Campbell, F., G. Conti, J. J. Heckman, H. Moon, Soon, R. Pinto, E. Pungello, and Y. Pan (2014). Early childhood investments substantially boost adult health. *Science 343*(6178), 1478–1485.

Capatina, E. (2015). Life-cycle effects of health risk. *Journal of Monetary Economics 74*, 67–88.

Capatina, E., M. Keane, and S. Maruyama (2020). Health shocks and the evolution of earnings over the life-cycle.

Card, D., A. Mas, E. Moretti, and E. Saez (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review 102*(6), 2981–3003.

Carneiro, P. M. and J. J. Heckman (2003). Human capital policy. *IZA Discussion paper series, No. 821*.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American economic review 104*(9), 2633–79.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014b, September). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review 104*(9), 2633–79.

Cicala, S., R. G. Fryer, and J. L. Spenkuch (2018). Self-selection and comparative advantage in social interactions. *Journal of the European Economic Association 16*(4), 983–1020.

Coe, N. B. and G. Zamarro (2011). Retirement effects on health in europe. *Journal of health economics 30*(1), 77–86.

Cole, H. L., S. Kim, and D. Krueger (2018, 03). Analysing the Effects of Insuring Health Risks: On the Trade-off between Short-Run Insurance Benefits versus Long-Run Incentive Costs. *The Review of Economic Studies 86*(3), 1123–1169.

Cribb, J., C. Emmerson, and G. Tetlow (2013). Incentives, shocks or signals: labour supply effects of increasing the female state pension age in the UK. Technical report, IFS working paper.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

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

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

Currie, J. (2008). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. Technical report, National Bureau of Economic Research.

Currie, J. and D. Almond (2011). Human capital development before age five. *Handbook of Labor Economics 4*, 1315–1486.

Currie, J. and B. C. Madrian (1999). Health, health insurance and the labor market. *Handbook of labor economics 3*, 3309–3416.

De Nardi, M., E. French, and J. B. Jones (2016). Medicaid insurance in old age. *American Economic Review 106*(11), 3480–3520.

De Nardi, M., S. Pashchenko, and P. Porapakkarm (2017). The lifetime costs of bad health. Technical report, National Bureau of Economic Research.

Deaton, A. (2008). Income, health, and well-being around the world: Evidence from the gallup world poll. *Journal of Economic Perspectives 22*(2), 53–72.

Deaton, A. and A. Heston (2010). Understanding ppps and ppp-based national accounts. *American Economic Journal: Macroeconomics 2*(4), 1–35.

Del Boca, D., C. Flinn, and M. Wiswall (2014). Household choices and child development. *The Review of Economic Studies 81*(1), 137–185.

Denning, J. T., R. Murphy, and F. Weinhardt (2020). Class rank and long-run outcomes.

Disney, R., C. Emmerson, and M. Wakefield (2006). Ill health and retirement in britain: A panel data-based analysis. *Journal of health economics 25*(4), 621–649.

Duckworth, A. L. and P. D. Quinn (2009). Development and validation of the short grit scale (grit–s). *Journal of personality assessment 91*(2), 166–174.

Duflo, E., P. Dupas, and M. Kremer (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review 101*(5), 1739–74.

Elsner, B. and I. E. Isphording (2017). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics 35*(3), 787–828.

Elsner, B. and I. E. Isphording (2018). Rank, sex, drugs, and crime. *Journal of Human Resources 53*(2), 356–381.

Elsner, B., I. E. Isphording, and U. Zölitz (2018). Achievement rank affects performance and major choices in college. *University of Zurich, Department of Economics, Working Paper* (300).

Engle, P. L., M. M. Black, J. R. Behrman, M. Cabral de Mello, P. J. Gertler, L. Kapiriri, R. Martorell, and M. E. Young (2007). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *The Lancet 369 (9557)*, 229–242.

Epple, D. and R. E. Romano (2011). Peer effects in education: A survey of the theory and evidence. In *Handbook of social economics*, Volume 1, pp. 1053–1163. Elsevier.

Escobal, J. and E. Flores (2008). An assessment of the young lives sampling approach in peru. *Young Lives Technical Note 3*.

Escobal, J., C. Lanata, S. Madrid, M. Penny, J. Saavedra, P. Suárez, H. Verastegui, E. Villar, and S. Huttly (2003). Young lives preliminary country report: Peru. *Young Lives Technical Note*.

Espy, K., M. McDiarmid, M. Cwik, M. Stalets, A. Hamby, and T. Senn (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental neuropsychology 26*(1).

Feld, J. and U. Zölitz (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics 35*(2), 387–428.

Fella, G. (2014). A generalized endogenous grid method for non-smooth and non-concave problems. *Review of Economic Dynamics 17*(2), 329–344.

Fernald, A., V. A. Marchman, and A. Weisleder (2013). Ses differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science 16*(2), 234–248.

Fernald, L. C., A. Weber, E. Galasso, and L. Ratsifandrihamanana (2011). Socioeconomic gradients and child development in a very low income population: evidence from madagascar. *Developmental science 14*(4), 832–847.

Finkelstein, A., E. F. Luttmer, and M. J. Notowidigdo (2009). Approaches to estimating the health state dependence of the utility function. *American Economic Review 99*(2), 116–21.

Fitzpatrick, M. D. and T. J. Moore (2018). The mortality effects of retirement: Evidence from social security eligibility at age 62. *Journal of Public Economics 157*, 121–137.

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

French, E. and J. B. Jones (2011). The effects of health insurance and self-insurance on retirement behavior. *Econometrica 79*(3), 693–732.

Fryer, R. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experimentsa. In A. V. Banerjee and E. Duflo (Eds.), *Handbook of Economic Field Experiments*, Volume 2 of *Handbook of Economic Field Experiments*, pp. 95–322. North-Holland.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, and S. Grantham-McGregor (2014). Labor market returns to an early childhood stimulation intervention in jamaica. *Science 344*(6187), 998–1001.

Gilleskie, D. B. (1998). A dynamic stochastic model of medical care use and work absence. *Econometrica*, 1–45.

Goldman, N., S. Korenman, and R. Weinstein (1995). Marital status and health among the elderly. *Social science & medicine 40*(12), 1717–1730.

Gourinchas, P.-O. and J. A. Parker (2002). Consumption over the life cycle. *Econometrica 70*(1), 47–89.

Grantham-McGregor, S., Y. B. Cheung, S. Cueto, P. Glewwe, L. Richter, and B. Strupp (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet 369*(9555), 60 – 70.

Gronau, R. (1974). Wage comparisons–a selectivity bias. *Journal of Political Economy 82*(6), 1119–1143.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political economy 80*(2), 223–255.

Hall, R. E. and C. I. Jones (2007). The value of life and the rise in health spending. *The Quarterly Journal of Economics 122*(1), 39–72.

Hanushek, E. A. and S. G. Rivkin (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review 100*(2), 267–71.

Harris, K. M. and J. R. Udry (2018). National longitudinal study of adolescent to adult health (add health), 1994-2008 [public use](icpsr21600).

Harris, M. C. (2019). The impact of body weight on occupational mobility and career development. *International Economic Review 60*(2), 631–660.

Hart, B. and T. R. Risley (1995). *Meaningful Differences in the Everyday Experience of Young American Children.* Paul H Brookes Publishing.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153–161.

Heckman, J. J. and T. Kautz (2013). Fostering and measuring skills: Interventions that improve character and cognition. Technical report, National Bureau of Economic Research.

Hoddinott, J., J. A. Maluccio, J. R. Behrman, R. Flores, and R. Martorell (2008). Effect of a nutrition intervention during early childhood on economic productivity in guatemalan adults. *The Lancet 371*(9610), 411–416.

Hosseini, R., K. Kopecky, and K. Zhao (2020). How important is health inequality for lifetime earnings inequality?

Hu, Y. and S. M. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica 76*(1), 195–216.

Ioannides, Y. M. (2011). Neighborhood effects and housing. In *Handbook of Social Economics*, Volume 1, pp. 1281–1340. Elsevier.

Iskhakov, F., T. H. Jørgensen, J. Rust, and B. Schjerning (2017). The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks. *Quantitative Economics 8*(2), 317–365.

Jacob, B. A., L. Lefgren, and D. P. Sims (2010). The persistence of teacher-induced learning. *Journal of Human resources 45*(4), 915–943.

Jacobs, L. and S. Piyapromdee (2016). Labor force transitions at older ages: Burnout, recovery, and reverse retirement.

Jochmans, K. (2020). Testing random assignment to peer groups. Cambridge working papers in economics, Faculty of Economics, University of Cambridge.

Jolivet, G. and F. Postel-Vinay (2020). A structural analysis of mental health and labor market trajectories. IZA Discussion Paper No. 13518.

Katz, S., A. B. Ford, R. W. Moskowitz, B. A. Jackson, and M. W. Jaffe (1963). Studies of illness in the aged: the index of ADL: a standardized measure of biological and psychosocial function. *Jama 185*(12), 914–919.

Keane, M. P. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature 49*(4), 961–1075.

Klein, D. N., L. R. Dougherty, and T. M. Olino (2005). Toward guidelines for evidence-based assessment of depression in children and adolescents. *Journal of Clinical Child and Adolescent Psychology 34*(3), 412–432.

Kuhn, A., S. Staubli, J.-P. Wuellrich, and J. Zweimüller (2020). Fatal attraction? extended unemployment benefits, labor force exits, and mortality. *Journal of Public Economics 191*, 104087.

Lazear, E. P. and S. Rosen (1981). Rank-order tournaments as optimum labor contracts. *Journal of political Economy 89*(5), 841–864.

Margaris, P. and J. Wallenius (2020). Can Wealth Buy Health? A Model of Pecuniary and Non-Pecuniary Investments in Health.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of educational psychology 79*(3), 280.

Meghir, C. and S. Rivkin (2011). Econometric methods for research in education. *Handbook of the Economics of Education, by Eric A. Hanushek, Stephen Machin and Ludger Woessmann eds. 3*, 1–87.

Moffitt, T. E., L. Arseneault, D. Belsky, N. Dickson, R. J. Hancox, H. Harrington, R. Houts, R. Poulton, B. W. Roberts, S. Ross, et al. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the national Academy of Sciences 108*(7), 2693–2698.

Murphy, R. and F. Weinhardt (2020). Top of the class: The importance of ordinal rank. *The Review of Economic Studies, (forthcoming)*.

Niederle, M. and L. Vesterlund (2011). Gender and competition. *Annu. Rev. Econ. 3*(1), 601–630.

Obradovic, J., X. Portilla, and T. Boyce (2012). Executive functioning and developmental neuroscience: Current progress and implications for early childhood education. *Handbook of Early Childhood Education*, 325–351.

O'Dea, C. (2018). Insurance, efficiency and the design of public pensions.

Olds, D. L., C. R. Henderson Jr, H. J. Kitzman, J. J. Eckenrode, R. E. Cole, and R. C. Tatelbaum (1999). Prenatal and infancy home visitation by nurses: Recent findings. *The future of Children*, 44–65.

Olino, T., L. Yu, D. McMakin, E. Forbes, J. Seeley, P. Lewinsohn, and P. Pilkonis (2013). Comparisons across depression assessment instruments in adolescence and young adulthood: an item response theory study using two linking methods. *Journal of Abnormal Child Phsychology 41*(8), 1267–77.

Outes-Leon, I. and A. Sanchez (2008). An assessment of the young lives sampling approach in ethiopia. *Young Lives Technical Note 1*, 1–37.

Ozkan, S. (2017). Preventive vs. Curative Medicine: A Macroeconomic Analysis of Health Care over the Life Cycle. mimeo, University of Toronto.

Papageorge, N. W. (2016). Why medical innovation is valuable: Health, human capital, and the labor market. *Quantitative Economics 7*(3), 671–725.

Paxson, C. and N. Schady (2007). Cognitive development among young children in ecuador the roles of wealth, health, and parenting. *Journal of Human resources 42*(1), 49–84.

Rose, L. (2020). Retirement and health: Evidence from england. *Journal of Health Economics 73*.

Rosenberg, M. (2015). *Society and the adolescent self-image*. Princeton university press.

Rubio-Codina, M., O. Attanasio, C. Meghir, N. Varela, and S. Grantham-McGregor (2015). The socioeconomic gradient of child development: Cross-sectional evidence from children 6-42 months in bogota. *Journal of Human Resources 50*(2), 464–483.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly journal of economics 116*(2), 681–704.

Sato, K. (1967). A two-level constant-elasticity-of-substitution production function. *The Review of Economic Studies 34*(2), 201–218.

Schady, N., J. Behrman, M. C. Araujo, R. Azuero, R. Bernal, D. Bravo, F. Lopez-Boo, K. Macours, D. Marshall, C. Paxson, and R. Vakis (2015). Wealth gradients in early childhood cognitive development in five latin american countries. *Journal of Human Resources 50*(2), 446–463.

Schennach, S. M. (2004). Estimation of nonlinear models with measurement error. *Econometrica 72*(1), 33–75.

Séguin, J. R. and P. D. Zelazo (2005). Executive function in early physical aggression.

Senn, T. E., K. A. Espy, and P. M. Kaufmann (2004). Using path analysis to understand executive function organization in preschool children. *Developmental neuropsychology 26*(1), 445–464.

Shaw, J. (2011). Fortax: Uk tax and benefit system documentation. Technical report, IFS Working Paper.

Shephard, A. (2009). Fortax: Reference manual. Technical report, Unpublished manuscript.

Siddiqui, S. (1997). The impact of health on retirement behaviour: empirical evidence from west germany. *Health Economics 6*(4), 425–438.

Siegler, R. S. and J. L. Booth (2004). Development of numerical estimation in young children. *Child development 75*(2), 428–444.

Smith, J. P. (2004). Unraveling the ses: health connection. *Population and development review 30*, 108–132.

Tincani, M. M. (2018). Heterogeneous peer effects in the classroom. Technical report.

Walker, S. P., S. M. Chang, C. A. Powell, and S. M. Grantham-McGregor (2005). Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted jamaican children: prospective cohort study. *The Lancet 366(9499)*, 1804–1807.

Statement of Conjoint Work

Note on the joint work in Francesca Salvati's thesis "Essays on Human Capital Formation over the Life Cycle".

Chapter 1, "The Effect of Classroom Rank on Learning throughout Elementary School: Experimental Evidence from Ecuador", was undertaken as joint work with Pedro Carneiro, Yyannu Cruz Aguayo and Norbert Schady.

Chapter 2, "Human Capital Growth and Poverty: Evidence from Ethiopia and Peru", was undertaken as joint work with Orazio Attanasio, Costas Meghir and Emily Nix.

Chapter 3, "Health Inequality, Labor Supply and Retirement Policies", is single-authored by Francesca Salvati.