10-26-2016 12:00 AM

Essays on Human Capital and Inequality

Youngmin Park, *The University of Western Ontario*

Supervisor: Lance Lochner, *The University of Western Ontario*
Joint Supervisor: Igor Livshits, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Economics

Abstract

This thesis conducts positive and normative analysis of inequality based on human capital theory. In Chapter 2, we document important differences in early child investments by family income and study four leading mechanisms thought to explain these gaps: intergenerational ability correlation, consumption value of investment, information frictions, and credit constraints. We evaluate whether these mechanisms are consistent with other stylized facts related to the marginal returns on investments and the effects of parental income on child investments and skills.

In Chapter 3, I study optimal higher education subsidies when parents willingness to pay for their children's education differs due to heterogeneity in altruism. I first document substantial heterogeneity in the fraction of college expenditure paid by parents across families and provide evidence that this heterogeneity can be explained by parental altruism. Then I analytically characterize optimal education subsidies when the social planner minimizes distortions generated by borrowing constraints and can observe neither the amount of parental transfers nor parental altruism. I show that redistributing towards constrained students of low altruism parents is socially beneficial, but it involves substantial deadweight loss. The calibrated model suggests that the deadweight loss due to unobservable heterogeneity in parental altruism can be quantitatively large and therefore limit redistribution towards students with low parental transfers.

In Chapter 4, we study the role of returns to unobserved skills in the rising residual earnings inequality for the past few decades in the U.S. We identify and estimate a general model of earnings residuals that incorporates (i) changing returns to unobserved skills, (ii) changing distribution of unobserved skills, and (iii) changing volatility of earnings that is not related to skills. Using data from the PSID, we find that the returns to unobserved skills went down since the mid-1980s despite the steady increase in the residual inequality. Using a simple demand and supply framework, we show that both demand and supply factors contributed to the downward trend in the returns to skills over time.

Keywords: Human capital, inequality, intergenerational mobility, credit constraints, education subsidies, technical change, earnings dynamics

Co-Authorship Statement

This thesis contains co-authored material. Chapter 2 is co-authored by Elizabeth Caucutt (University of Western Ontario) and Lance Lochner (University of Western Ontario). Chapter 4 is co-authored by Lance Lochner and Youngki Shin (University of Technology Sydney). All authors are equally responsible for the work.

In Memory of My Mother

Acknowledgements

It has been my great fortune to study with my advisors. Professor Lance Lochner has shown me how to combine deep economic insight with empirical analysis that will always guide my work. I am also very grateful to him for supporting me financially and providing me with various opportunities to grow up academically. Professor Igor Livshits recruited me and has advised me from my first year at Western. I learned from him that talking about economics should be incredibly exciting and fun. I also want to thank Professor Elizabeth Caucutt who was very involved throughout my study and provided invaluable feedback and advise. I greatly enjoyed interacting with her and working together on our joint paper.

I am also indebted to many professors, fellow graduate students, and supporting staffs at Western. The list is too long to mention, but my classmates Hiro Mori, Jin Zhou, Peter and Zina Foltin, and Brandon Malloy made the past six years enjoyable. We had stimulating discussions and encouraged each other through the whole process.

I also wish to thank Professor Yongsung Chang, who was the first to expose me to frontier macroeconomic research at Yonsei University in Korea and recommended me good Canadian graduate programs including Western. I also thank Korean community at Western. My best friends Hyeongsuk Jin and Youjin Choi, and Professors Tai-Yeong Chung and Youngki Shin and their families treated me like a family member.

Most importantly, I'd like to thank my parents who altruistically invested a lot in me. My father generously supported me through my study despite the sustained financial hardship. It is very regrettable that my mother did not live to see this day.

Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

Introduction

Why do some people earn more than others? According to human capital theory, differences in earnings can be attributed to differences in early life human capital investment, but can also be also be affected by economy-wide factors later in life. In this thesis I conduct positive and normative analysis of inequality based on human capital theory. The thesis consists of three chapters, each focusing on a different stage of the life-cycle. Chapter 2 studies human capital investment during early childhood and asks why children from low income families obtain lower cognitive achievement than those from high income families. Chapter 3 is concerned with efficient financial aid policy for post-secondary education in the presence of credit constraints and unobservable heterogeneity in parents' willingness to invest in their children. Finally, Chapter 4 investigates dynamics of earnings inequality during adulthood and asks whether the rise in U.S. residual earnings inequality over the past few decades is due to increasing returns to unobserved human capital/skills or rising volatility of earnings.

It is well-known that poor children perform significantly worse academically and on achievement tests than their more economically advantaged counterparts. The most immediate explanation for these differences is that poor parents invest less in their young children. While there are many competing theories for these investment and skill gaps, few studies attempt to sort amongst them. In Chapter 2, which is a joint work with Professors Lance Lochner and Elizabeth Caucutt, we systematically study four leading investment-based theories/mechanisms thought to drive income-based skill gaps: an intergenerational correlation in ability, a consumption value of investment, information frictions, and credit constraints. In order to help understand which mechanisms drive family investments in children, we consider the extent to which they also explain other important stylized facts related to the marginal returns to investment and the effects of parental income on child investment and skills. Although all competing explanations are consistent with some of the stylized facts, only the binding credit constraints are consistent with all of the stylized facts we consider. This suggests that the inability of families to borrow against (parents' or children's) future income can be an important source of inequality in skill acquisition.

The importance of credit constraints motivates government policies that aim to lower the cost of higher education for students who receive little support from parents and thus are likely to be constrained. For example, the U.S. federal government provides grants and subsidized loans for students from low income families. But it is challenging to design a financial aid system that specifically targets constrained students, as parental transfers may be imperfectly correlated with family resources and are difficult to observe. In Chapter 3, I first show that American parents with similar income contribute very differently towards their children's college education and that these differences are not easily explained by differences in monetary/psychic returns to college across families. Motivated by this evidence, I next study how education subsidies should be distributed when students receive different amount of parental transfers due to differences in parental altruism and the policymaker can observe neither parental transfers nor altruism. When the social planner minimizes distortions generated by borrowing constraints, it is optimal to give higher subsidies for early years of college than late years, because it helps constrained students who do not stay long in college due to low parental transfers. However, such policy also provides an incentive to attend college and quit too early, which is inefficient due to a fixed cost of college attendance, for those who otherwise would not attend. The calibrated model shows that the efficiency cost due to unobservable heterogeneity in parental altruism can be quantitatively large and therefore severely limit the extent to which subsidy policies can address the inefficiencies due to borrowing constraints.

Skills that are acquired through educational investment are important determinants of earnings, but the returns to skills may change over time, leading to changes in earnings inequality. A change in cross-sectional earnings inequality, however, cannot be simply interpreted as a change in the returns to skills because it could also reflect a change in the volatility of earnings that is unrelated to skills. In Chapter 4, along with Professors Lance Lochner and Youngki Shin, I study the role of returns to unobserved skills in the rising residual earnings inequality for the past few decades in the U.S. We identify and estimate a general model of earnings residuals that incorporates (i) changing returns to unobserved skills, (ii) changing distribution of unobserved skills, and (iii) changing volatility of earnings that is not related to skills. Using data from the PSID, we find that the returns to unobserved skills went down since the mid-1980s despite the steady increase in the residual inequality. Using a simple demand and supply framework, we show that both demand and supply factors contributed to the downward trend in the returns to skills over time.

Chapter 2

Correlation, Consumption, Confusion, or Constraints: Why do Poor Children Perform so Poorly?[1]

2.1 Introduction

Adolescent skill and achievement gaps by parental income explain a substantial share of subsequent differences in educational attainment and lifetime earnings (Cameron and Heckman, 1998, Keane and Wolpin, 1997, Carneiro and Heckman, 2002), suggesting that an important component of intergenerational economic and social mobility is determined by the time children reach adolescence. Perhaps more troubling, sizeable differences in achievement by parental income are already evident by very young ages, persisting throughout childhood (Carneiro and Heckman, 2002, Cunha et al., 2006, Cunha and Heckman, 2007, Cunha, 2013). This raises the possibility that a generation's fate may be sealed by the time it enters school.[2] Altogether, this evidence suggests that a complete understanding of intergenerational mobility and its implications for economic and social policy requires convincing answers to the vexing question: Why do poor children perform so poorly?

Given the importance of family investments for early child development (Todd and Wolpin, 2007, Cunha and Heckman, 2008, Cunha, Heckman, and Schennach, 2010, Del Boca, Flinn, and Wiswall, 2014, Pavan, 2014), we concentrate on understanding why low-income families invest so much less in their young children compared to higher income families (Guryan, Hurst, and Kearney, 2008, Kaushal, Magnuson, and Waldfogel, 2011). We consider four broad

[1] A version of this chapter has been accepted for publication by *Scandinavian Journal of Economics*.

[2] Recent studies show that these early achievement and educational attainment gaps have been growing in the United States for decades (Belley and Lochner, 2007, Reardon, 2011).

mechanisms often thought to explain early investment and achievement gaps by family income.

1. The natural ability of children and parents may be correlated (Becker and Tomes, 1979, 1986). If child achievement is an increasing function of own ability, then a positive intergenerational ability correlation can generate the income – achievement gradients documented in the literature.

2. Parents may enjoy making investments in their children. If investments provide a direct benefit to parents above and beyond the future labor market returns to children, parents will choose to invest more as their income rises like they would purchase more of any other normal good (Lazear, 1977). It is also possible that low- and high-income families place different intrinsic value on investments or human capital more generally (Abbott et al., 2013).

3. Low-income parents may be poorly informed about the productivity of or returns to investments in their children (Cunha, Elo, and Culhane, 2013, Cunha, 2014, Dizon-Ross, 2015). For example, poor parents may incorrectly believe that investments in their young children are unproductive (or poorly rewarded in the labor market), or they may simply face greater uncertainty in the productivity of or returns to investments. Alternatively, poor parents may recognize the importance of investing in their children, but they may not know which types of investment activities/goods are most productive.

4. Poor families may be unable to invest efficiently in their children due to limits on their capacity to borrow against their own future income or against the potentially high returns on investments in their children (Becker and Tomes, 1979, 1986, Caucutt and Lochner, 2006, 2012, Cunha et al., 2006, Cunha and Heckman, 2007, Cunha, 2013, Lee and Seshadri, 2014).

We use a simple framework of dynamic human capital investment to formally examine whether these mechanisms are also able to account for other important stylized facts in the literature on child development. In particular, we focus on four well-established findings related to the marginal returns to early investment and the role of family income: (i) the high marginal returns to early investments in economically disadvantaged children, (ii) lower returns on marginal investments in higher income children, (iii) exogenous increases in family income lead to greater investments in children and improved childhood outcomes, and (iv) the impacts of income on child investments, achievement and educational attainment are greater if the income is earned (or received) when children are young.[3] While our analysis is not intended to

[3]We also briefly discuss other evidence related to specific mechanisms in Sections 2.5-2.8 where those mechanisms are considered in detail.

determine which mechanism is most important for explaining income-based achievement gaps, it is useful for helping understand which mechanisms are needed to provide a more complete picture of the child development process and the role of family income.[4] This is important, because the different mechanisms can have very different policy implications. For example, if investment and achievement gaps are driven only by intergenerational ability correlations or a 'consumption' value of investment, then investments in children are likely to be economically efficient (in the absence of human capital externalities) and policies designed to improve equity will be inefficient.[5] By contrast, either information-based or credit market frictions can lead to inefficiently low investments in economically disadvantaged children. In this case, it may be possible to simultaneously improve both equity and efficiency through well-designed policies.

We organize this chapter in the following way. In Section 2.2, we briefly document differences in child achievement and investment levels by family income using data from the Children of the National Longitudinal Survey of Youth (CNLSY). Then, we summarize evidence on four additional stylized facts from the literature on child development in Section 2.3. Sections 2.4-2.8 develop and analyze a unified framework of dynamic skill investment that incorporates all four potential mechanisms commonly thought to drive investment and achievement gaps by family income. We use this framework to formally examine whether the explanations are consistent with the stylized facts in Section 2.3 as well as other evidence in the child development literature. In Section 2.9, we conclude with a summary of our main results and their implications for future research.

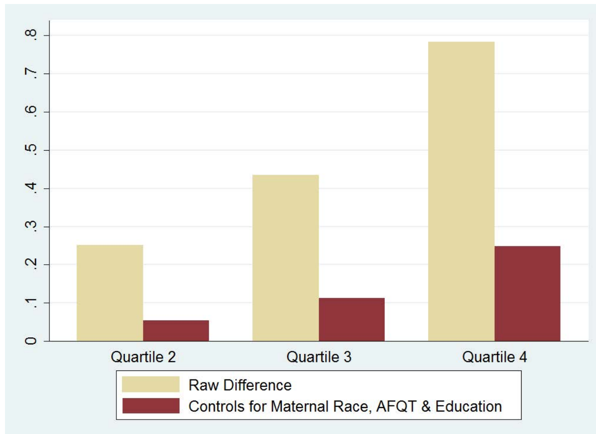
2.2 Child Achievement and Investment Gaps by Family Income

In this section, we document differences in child achievement and investment behavior by family income using data from the CNLSY. A longitudinal survey that links mothers with their children, the CNLSY contains excellent measures of family background and income (starting in 1979), as well as biennial measures of child math and reading achievement and family

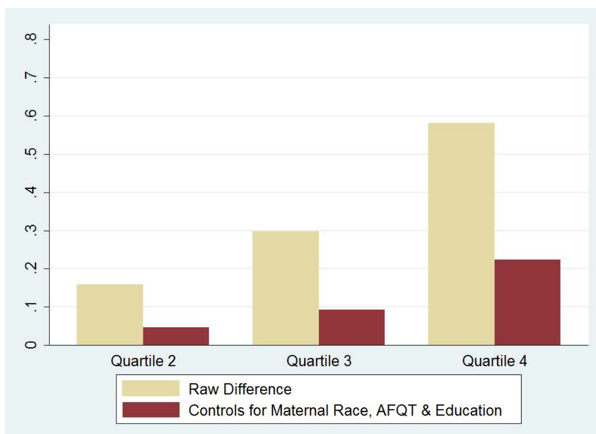
[4]See Cunha (2014) for a novel effort to empirically decompose the relative importance of a similar set of mechanisms using unique data on parental perceptions and stated choices about investments in children under different hypothetical budget sets. While we do not empirically evaluate the relative importance of different mechanisms, our theoretical analysis is based on a more general dynamic human capital investment model. We consider a wide range of information frictions and explicitly model intertemporal borrowing constraints.

[5]Of course, it may be socially desirable to encourage investment beyond the privately optimal amount due to human capital externalities in production (Moretti, 2004a,b) or related to crime (Lochner and Moretti, 2004) or citizenship (Milligan, Moretti, and Oreopoulos, 2004).

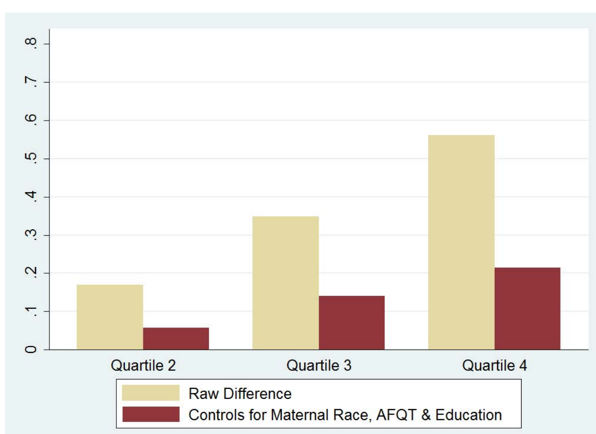
Figure 2.1: Ages 6-7 Achievement Gaps by Parental Income Quartile (Relative to Quartile 1)



(a) Math Achievement

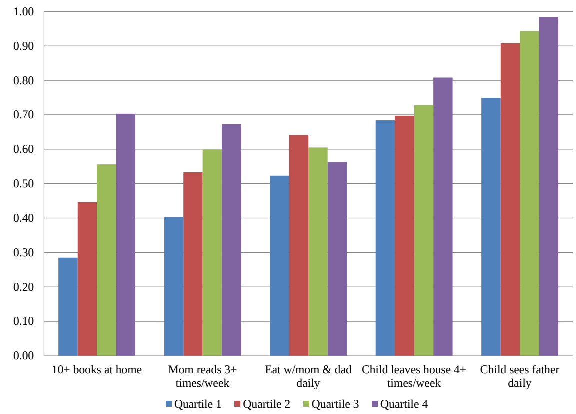


(b) Reading Recognition Achievement

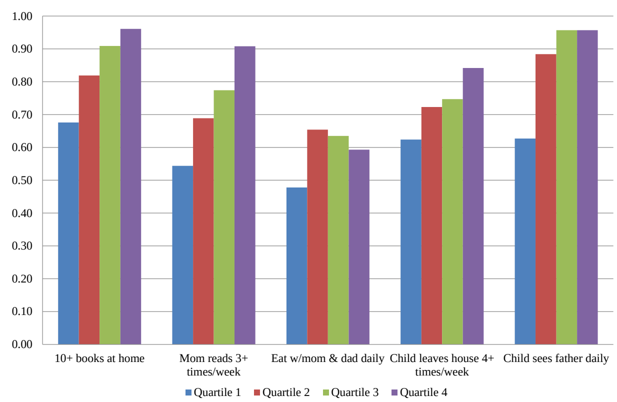


(c) Reading Comprehension Achievement

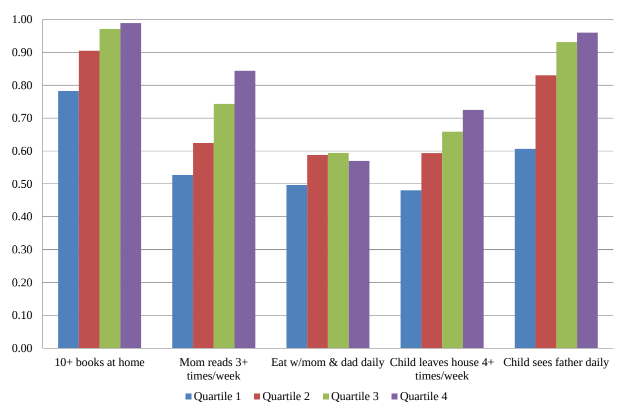
Figure 2.2: Family Investments in Children Ages 0-1, 2-3, and 4-5 by Parental Income Quartile



(a) Child Ages 0-1

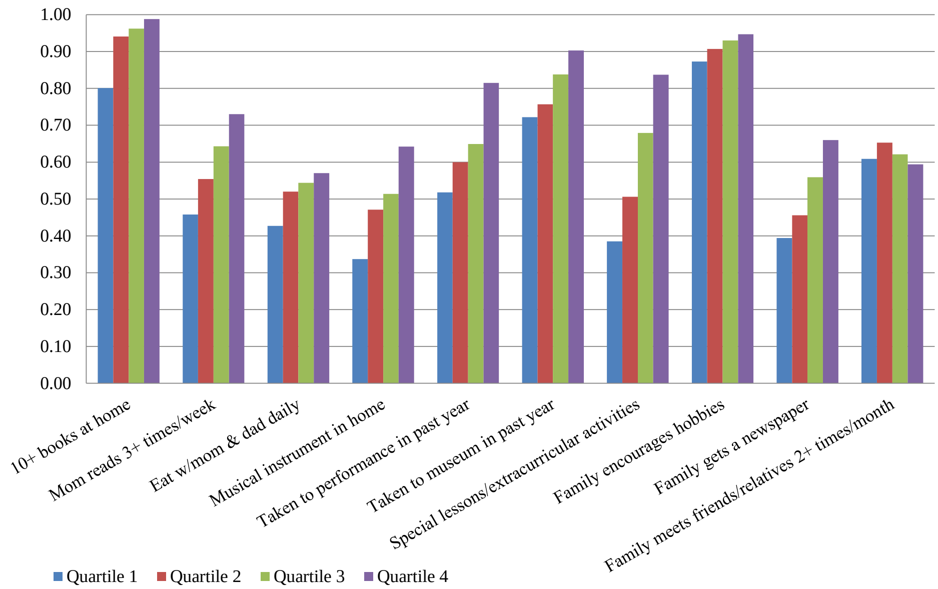


(b) Child Ages 2-3



(c) Child Ages 4-5

Figure 2.3: Family Investments in Children Ages 6-7 by Parental Income Quartile



investments in children (beginning in 1986).[6]

We use background measures of maternal education, race/ethnicity, and "ability" as measured by the Armed Forces Qualifying Test (AFQT).[7] As a medium-run measure of family income, we average all available reports of earnings by the mother and her spouse (if married) from the child's birth through ages 6 or 7 (depending on which of these ages achievement and investments in children were measured).[8] Child achievement is measured by the Peabody Individual Achievement Tests (PIAT) in math, reading recognition and reading comprehension; these measures are standardized to have a mean of zero and standard deviation of one at each age. A number of child investment activities/inputs are also reported in the CNLSY as we discuss below.

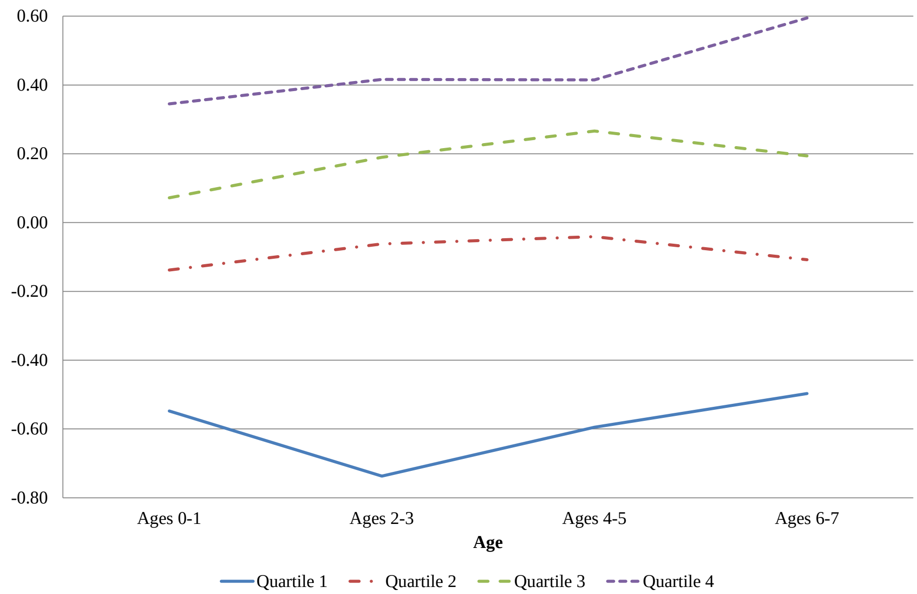
Figure 2.1 documents sizeable differences in ages 6-7 math and reading achievement by family income quartile. The light (beige) bars represent raw differences in achievement between the reported parental income quartile and the bottom income quartile, while the dark

[6] As is standard in studies using the CNLSY, our (unweighted) analysis is based on children born to mothers from the random sample of the NLSY79. Thus, distributions are based on children born to a random sample of American women born between 1957 and 1964, which may differ from distributions for any specific cohort of children.

[7] (Nearly) all CNLSY mothers, born between 1957 and 1964, took the AFQT in 1980 as part of the survey. The AFQT tests basic math and verbal/reading skills.

[8] Before averaging across time, we discount all income back to the child's birth year using a 5% discount rate, so our earnings measure reflects average discounted family earnings from the child's birth through ages 6-7. Individuals are dropped if fewer than 3 income reports are available.

Figure 2.4: Family Investment Factor Scores by Child Age and Parental Income Quartile

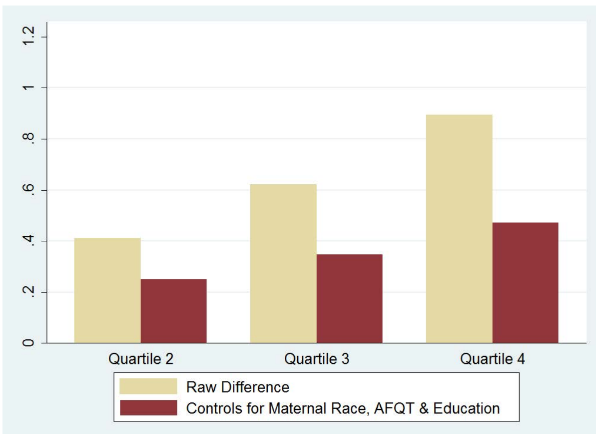


(red) bars report differences after controlling flexibly for maternal race/ethnicity, AFQT, and educational attainment.[9] Raw gaps by income are sizeable: math and reading scores of children with parents in the highest income quartile are all more than half of a standard deviation higher than those with parents in the lowest income quartile. Controlling for other important maternal characteristics substantially reduces these gaps (by as much as three-quarters), but does not eliminate them – parental income still has economically (and statistically) significant effects on child achievement.

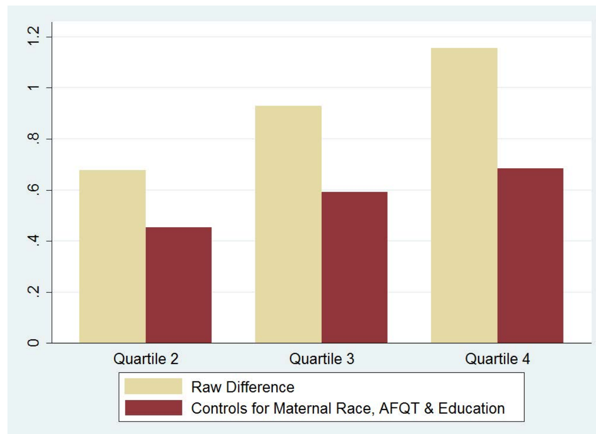
Figures 2.2 and 2.3 document a number of early childhood family investment measures by parental income at different ages. For all measures except ‘eat with mom and dad daily’ (ages 0-1, 2-3, 4-5) and ‘family meets friends/relatives two or more times per month’ (ages 6-7), investments are monotonically increasing in parental income. For a number of measures, the differences are substantial. For example, mothers of young children from the highest income quartile are over 50% more likely to read to their child three or more times per week compared to mothers from the lowest income quartile. High income mothers with children ages 0-1 are more than twice as likely to have 10 or more books in the home. Among children ages 6-7,

[9]Standard multivariate regressions are used to control for maternal race/ethnicity (white, black, hispanic), AFQT quartiles, and educational attainment (high school dropout, high school graduate, some college, completed college) by including three-way interactions of all three sets of indicators along with indicator variables for parental income quartile. Sample sizes for raw differences by income are 3,449 for math, 3,436 for reading recognition, and 3,267 for reading comprehension. Approximately 80 observations are dropped due to missing covariates when controlling for maternal characteristics.

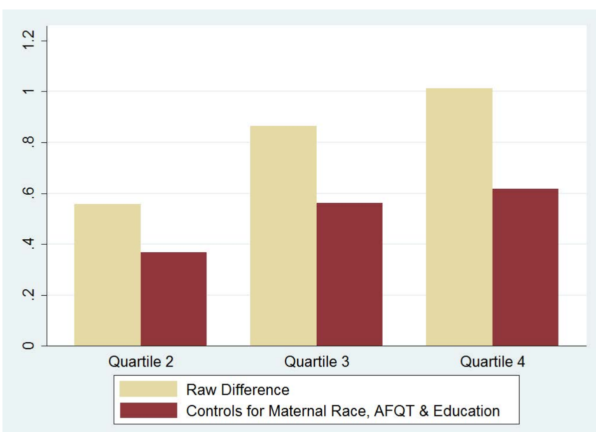
Figure 2.5: Family Investment Gaps by Parental Income Quartile (Relative to Quartile 1)



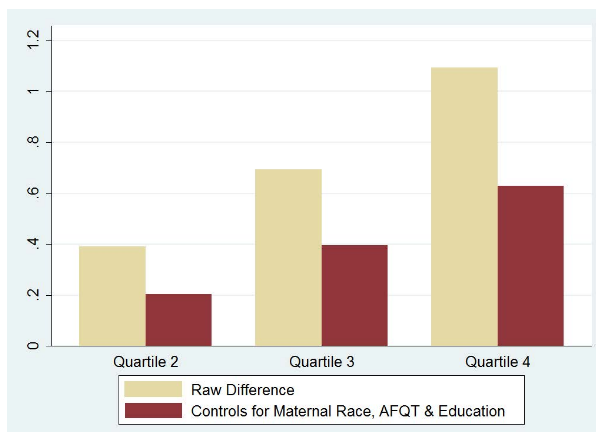
(a) Child Ages 0-1



(b) Child Ages 2-3



(c) Child Ages 4-5



(d) Child Ages 6-7

those from high income families are more than twice as likely to be enrolled in special lessons or extracurricular activities.

One interpretation of the investment measures reported in Figures 2.2 and 2.3 is that they represent different types of investment inputs that influence child development. An alternative interpretation is that they all represent noisy measures of a single underlying 'investment'. Under the latter interpretation, factor analysis can be used to uncover a more precise measure of the latent investment (Cunha and Heckman, 2008, Cunha, Heckman, and Schennach, 2010).[10] Based on this insight, we employ principal factor analysis using the measured inputs reported in Figures 2.2 and 2.3 to create age-specific predicted investment factor scores for each child. (See Appendix A.1.) For interpretation purposes, we normalize scores to have a mean of zero and standard deviation of one, plotting average scores by age and parental income quartile in Figure 2.4. The figure reveals sizeable differences in investment factor scores by parental income that are already evident at very young ages. Investments are roughly a full standard deviation higher among children from high income families relative to low income families. Figure 2.5 shows that these gaps shrink by as much as 50% but remain sizeable when controlling for maternal race, AFQT, and educational attainment.[11] Comparing Figures 2.1 with 2.5 reveals that maternal characteristics explain a greater share of the income-based gaps in achievement than in investments.

2.3 Additional Stylized Facts on Child Development

In this section, we discuss four stylized facts on the marginal returns to investment in children and the role of family income in child development. Because of their general nature, these facts can be compared against the predictions of any investment-based model of skill formation. Other, more mechanism-specific findings in the literature are briefly discussed in the next section.

[10]These interpretations are not necessarily mutually exclusive if families face the same relative prices and productivity of inputs (assuming families maintain correct beliefs about the relative productivity of inputs). In this case, all inputs (in a given period) will be proportional to total child investment expenditure (that period), with all families choosing the same proportional mix. One can then think about the various reported measures as noisy measures of that specific input or of total investment expenditure (multiplied by the factor share for that input). We study the link between multiple early inputs and total early investment expenditures in Section 2.7 along with the consequences of family mis-perceptions about relative and overall early input productivity levels.

[11]Sample sizes for raw income differences are 2,324 (ages 0-1), 2,749 (ages 2-3), 2,813 (ages 4-5), and 3,493 (ages 6-7). When controlling for maternal characteristics, between 50 and 80 observations are dropped due to missing covariates.

Fact 1

The returns to early marginal investments are higher than the return to savings for economically disadvantaged children. A number of comprehensive surveys (Karoly et al., 1998, Blau and Currie, 2006, Cunha et al., 2006, Almond and Currie, 2011, Heckman and Kautz, 2014, Kautz et al., 2014) document both the short- and long-term impacts of numerous early childhood interventions in the U.S. Most notably, an experimental evaluation of the Perry Preschool program followed participants in the early 1960s through age 40, measuring the program's impacts on a wide-array of outcomes, including cognitive achievement, educational attainment, earnings, and crime.[12] Based on program costs and impacts measured from ages 15-40, Heckman et al. (2010) estimate a private internal rate of return to Perry Preschool participants of around 8%.[13] The Abecedarian Project offered high quality full-day preschool to a (randomly assigned) sample of mostly African American children born in the mid-1970s who were at risk for delayed intellectual and social development. Follow-up evaluations of Abecedarian through age 21 reveal significant long-term benefits that exceed the program's costs by a factor of roughly 2.5 (Barnett and Masse, 2007).[14] Researchers have also extensively analyzed the long-term impacts of Chicago's Child-Parent Center (CPC) preschool program, following a sample of low-income, mostly African American participants from the mid-1980s to the present. Rough calculations based on program impacts measured through age 26 suggest an average (private) benefit/cost ratio of 3.6 (Reynolds et al., 2011b).[15] Finally, a number of studies document significant long-term impacts of Head Start (Currie and Thomas, 1995, 1999, Garces, Currie, and Thomas, 2002, Deming, 2009, Ludwig and Miller, 2007, Carneiro and Ginja, 2014); however, its full rate of return has not been systematically estimated.

Other family-based investments aimed at improving mother-child interactions and maternal parenting skills can also serve as productive early investments in the child development process, producing significant long-run benefits for children. For example, the Nurse-Family Partnership provided regular pre- and post-natal (to age 2) home visits to low-income mothers

[12] Perry Preschool provided daily high quality preschool (2.5 hours per day) and weekly home visits for two years to children ages 3 and 4. The randomized control trial sample was drawn from low IQ children from families of low socioeconomic status.

[13] Social returns are even higher, largely due to benefits from crime reduction.

[14] As is common in this literature, these calculations assume an annual real discount rate of 3%. The benefit calculations project lifetime earnings impacts based on average earnings differences by educational attainment and the significant effects of Abecedarian on educational attainment. An age 30 follow-up study (Campbell et al., 2012) estimates a sizeable but statistically insignificant increase in annual earnings for participants ($33,000 vs. $21,000). Estimated effects on employment rates and use of public aid at age 30 are statistically significant.

[15] As with the Abecedarian cost-benefit analysis of Campbell et al. (2012), Reynolds et al. (2011b) use a 3% discount rate and project lifetime earnings benefits from estimated impacts on age 26 educational attainment; they also incorporate benefits from reductions in child care costs and child abuse/neglect. In a subsequent follow-up, Reynolds et al. (2011a) show that the program significantly increased age 28 earnings by 7%.

with the goals of improving pregnancy outcomes and maternal health, improving the health and development of their children through proper care, and enhancing parental life-course development. Studies of the program in three U.S. cities estimate long-term benefits from these investments on a number of child outcomes (Olds et al., 2002, Eckenrode et al., 2010, Kitzman et al., 2010). Long-term benefits of similar family-based interventions have also been documented in Jamaica and Colombia, where home visitation programs provided one-hour weekly visits (for up to 2 years) aimed at improving mother-child interactions and developing child cognitive, language and psychosocial skills (Gertler et al., 2014, Attanasio et al., 2015).

Fact 2

The returns to marginal investments are lower for more economically advantaged children. Because most experimental and government-subsidized early childhood programs serve low-income families, less is known about the lifetime returns to early investments in children from higher income families. However, a number of studies estimate short- and medium-term impacts of early childhood interventions by family income or socioeconomic status (SES). These studies typically report greater benefits for more disadvantaged children. For example, Duncan and Sojourner (2013) estimate that the Infant Health Development Program (IHDP), which provided the Abecedarian preschool curriculum to an economically diverse sample of low-birth weight 1-2 year-olds, yielded significantly greater improvements in age five IQ for the subsample of children from low-income families relative to those from higher income families.[16] A recent analysis of Head Start (Puma et al., 2012) estimates significantly greater impacts on third grade cognitive and learning outcomes for children from 'high risk' (i.e. low SES) households relative to lower risk households. Estimated impacts of the Chicago CPC preschool program on educational attainment and earnings at age 28 are also higher for children from 'high risk' families (Reynolds et al., 2011a). A few studies that estimate the impacts of introducing universal early child care subsidies in Canada and Norway find negligible or even adverse impacts on children from middle- and high-income families, likely due to the substitution of lower quality subsidized/free child care in place of higher quality unsubsidized family or informal care (Baker, Gruber, and Milligan, 2008, Havnes and Mogstad, 2014, Kottelenberg and Lehrer, 2014).[17]

Taking a very different approach, Cunha, Heckman, and Schennach (2010) apply dynamic factor models using multiple noisy measurements of child investments and skill levels to estimate the technology of human capital production from birth through the end of school.

[16] Brooks-Gunn et al. (1992) estimate greater effects of IHDP on age three IQ for families with lower maternal education.

[17] See (Baker, 2011) for a careful discussion of recent universal early child care initiatives.

Their estimated technology suggests that the most efficient allocation of educational investments would provide more to young disadvantaged children. The fact that actual investments are much lower for disadvantaged children (see Figures 2.2-2.5) coupled with diminishing marginal returns, suggests that returns on the margin are higher for the most disadvantaged.

Fact 3

Exogenous increases in parental income lead to greater investments in children and improvements in childhood outcomes. A number of recent studies attempt to address concerns about endogeneity in estimating the effects of exogenous changes in family income on children. Dahl and Lochner (2012) exploit expansions of the Earned Income Tax Credit (primarily over the mid-1990s) to estimate the effects of additional family income on cognitive achievement. Their instrumental variable estimates suggest that an additional $1,000 in family income raises combined math and reading scores by 6 percent of a standard deviation. Estimated effects also appear to be larger for children from more disadvantaged families. Milligan and Stabile (2011) estimate that expansions of child tax benefits in Canada led to similar improvements in child cognitive and educational outcomes as well as improvements in child and maternal health. Combining data from ten welfare and anti-poverty experiments, Duncan, Morris, and Rodrigues (2011) attempt to separately identify the effects of changes in family income from employment and other effects induced by different programs. Their analysis reaches similar conclusions regarding the impacts of income on child achievement as Dahl and Lochner (2012) and Milligan and Stabile (2011). Finally, Løken (2010) and Løken, Mogstad, and Wiswall (2012) estimate the impact of family income on Norwegian children using regional variation in the economic boom following the discovery of oil as an instrument for income. The latter study estimates that income has sizeable impacts on education and IQ for children from low-income families but much weaker effects for children from higher income families.[18]

Changes in income may affect children in many ways. In this chapter, we focus on investment-based theories, so it is important to know whether changes in family income cause families to make different investment choices. A few studies suggest that this is the case. Following the approach of Milligan and Stabile (2011), Jones, Milligan, and Stabile (2015) examine how Canadian parents altered their household expenditures in response to an expansion of child tax benefits. Their estimates suggest that low-income families, on average, spent 13 cents out of every additional dollar in benefits on education-related items (e.g. tuition, computers).[19]

[18] Studies on the effects of parental job displacement on children also suggest that family income may have important effects on child schooling and labor market earnings (Oreopoulos, Page, and ann Huff Stevens, 2008, Stevens and Schaller, 2011); however, parental job displacement may also affect child development through other channels (e.g. family dissolution).

[19] Interestingly, their results for all Canadian families suggest negligible effects on average education-related

Carneiro and Ginja (2014) estimate models of income dynamics in the U.S., examining the extent to which family investments in children respond to permanent and transitory income shocks. Their results suggest modest positive responses to permanent shocks but negligible responses to transitory shocks. Effects appear to be largest for younger children and those with less-educated parents. Among children whose mothers had not attended college, a 10% increase in permanent income is estimated to increase measures of cognitive stimulation and time investments by about .02 standard deviations. Cunha, Heckman, and Schennach (2010) and Pavan (2014) estimate both the technology of skill formation for children (from birth through later school ages) and the extent to which family income, as well as maternal and child skills, affect investments in children. Their estimated investment functions suggest that increases in family income lead to significantly higher investments in children.

Fact 4

The timing of income matters for child development: increases in income at early ages (compared to later ages) lead to larger increases in investments and achievement/educational outcomes. The estimated child investment functions of Pavan (2014) imply significantly greater effects of family income on investments at very early ages relative to older ages. Other studies estimate the effects of family income received at different child ages on adolescent achievement or educational outcomes. For example, Duncan and Brooks-Gunn (1997), Duncan et al. (1998), and Levy and Duncan (1999) all estimate that income received at earlier ages has a greater impact on adolescent achievement than income received at later ages. Carneiro and Heckman (2002) correctly point out, however, that (undiscounted) early income should have a larger effect than (undiscounted) later income due purely to discounting – something not taken into account in previous analyses.[20] More recent studies address this concern by discounting all income measures back to the year of birth (Carneiro and Heckman, 2002, Caucutt and Lochner, 2006, 2012).[21] Caucutt and Lochner (2006) report results consistent with the earlier literature, finding that income received at young ages has a greater effect than income received at older ages on subsequent child achievement. Caucutt and Lochner (2012) further show that family income earned when children are younger has a significantly greater effect on college attendance than does income earned at later ages; however, Carneiro and Heckman (2002) cannot reject that income has the same effects on college enrolment regardless of the age at which it was received. While both of these studies use data from the CNLSY, the former benefits

expenditures, so poor families appear to increase education-related spending much more than the typical family.

[20]That is, with perfect credit markets, income received at age 0 should have an effect that is $(1+r)^a$ times larger than income received at age a, where r is the annual interest rate.

[21]These studies all use a 5% rate to discount income back to the year of a child's birth.

from a sample size that is roughly twice as large, allowing for greater precision. Furthermore, because Carneiro and Heckman (2002) are more concerned with the importance of borrowing constraints at college-going ages, they control for age 12 math achievement levels, which may absorb much of the effect of early income.

2.4 Understanding Investment and Skill Gaps by Family Income

We now develop a general model of dynamic human capital investment that incorporates four mechanisms thought to generate child investment and skill gaps by family income. Within this framework, we explore the extent to which these mechanisms are also capable of explaining the additional stylized facts just discussed. The problem is written as a lifecycle problem in which individuals invest in their human capital, while borrowing and saving (potentially subject to borrowing constraints) in an effort to finance investments and smooth consumption over time. However, the problem can also be interpreted as a 'family' investment problem, where altruistic parents make investments in their children and family borrowing/saving decisions to smooth family consumption.[22]

We assume that people live through three stages in their lives. Human capital investment takes place in the first two stages (i.e. 'childhood'), followed by the final stage, adulthood. Adulthood may last for many periods; however, its length is inconsequential for most of our analysis. We are largely agnostic about the form that investments may take, instead focusing primarily on total investment expenditures at different ages and the dynamic nature of skill production. Conceptually, investments may include various forms of goods inputs like computers and books, parental time in child development activities, formal schooling, and other time inputs by older children.[23] When considering parental time as an investment, if the marginal product of parental time investment (in children) is proportional to the parent's labor market productivity (i.e. human capital), then investment is simply given by the parent's forgone earnings.[24]

[22]See Caucutt and Lochner (2012) for a direct mapping between this lifecycle problem and a more explicit intergenerational problem.

[23]They might also include more general investments in families or mother-child interactions that are designed to facilitate child development (e.g. home visitation programs as discussed in Section 2.3).

[24]To the extent that time investments are important, we implicitly assume that parents can flexibly adjust their labor supply at a fixed wage. In particular, they can adjust their labor supply downwards without incurring an hourly wage penalty. Otherwise, distortions similar to those observed for borrowing constraints (discussed below) might arise if parental time is a key (and non-substitutable) input for young children and if parents cannot fully reduce their work hours to make desired time investments.

2.4.1 Technology for Human Capital Production

Denote a child's ability to learn by $\theta > 0$. Investment expenditures in periods 1 and 2 are given by i_1 and i_2, respectively. Early investment produces an interim level of human capital,

$$h_2 = z i_1, \tag{2.1}$$

where $z > 0$ is the productivity of early investment. Together, late investment and this interim human capital produce stage 3 (adult) human capital:

$$h_3 = \theta f(h_2, i_2). \tag{2.2}$$

The human capital production function $f(\cdot)$ is strictly increasing and strictly concave in both of its arguments. To guarantee appropriate second order conditions hold in the decision problems described below, we assume the following throughout our analysis (without explicit reference):

Assumption 2.1 $f_{12}^2 < f_{11} f_{22}$ *and* $f_{12} > \max\left\{ f_{22}\left(\frac{f_1}{f_2}\right), f_{11}\left(\frac{f_2}{f_1}\right)\right\}$.

The first condition limits the degree of dynamic complementarity in investments and ensures strict concavity of the production function. The second condition implies that the least costly way to produce additional human capital h_3 is to increase both early and late investments. Most plausible specifications for human capital production would entail dynamic complementarity (i.e. $f_{12} \geq 0$), satisfying this condition (Cunha, Heckman, and Schennach, 2010, Caucutt and Lochner, 2012); however, the condition holds much more generally.[25] We also assume standard Inada conditions to ensure interior solutions.[26]

At times, our analysis will employ a CES human capital production function of the form

$$f(h_2, i_2) = \left[a^{1-b} h_2^b + (1-a)^{1-b} i_2^b\right]^{d/b}, \tag{2.3}$$

where $a \in (0,1)$, $b < 1$, and $d \in (0,1)$; however, most of our analysis does not rely on any particular functional form. Assumption 2.1 holds for this production function.

2.4.2 General Decision Problem

We assume a period utility function over consumption, $u(c)$, that is strictly increasing, strictly concave and satisfies standard Inada conditions. Tastes for early educational investment, i_1, are

[25] For example, the condition holds for homothetic functions (e.g., CES) regardless of the degree of complementarity.

[26] That is, $\lim_{h_2 \to 0} f_1(h_2, i_2) = \infty$, $\forall i_2 \geq 0$, and $\lim_{i_2 \to 0} f_2(h_2, i_2) = \infty$, $\forall h_2 \geq 0$.

given by νi_1. The time discount rate is $\beta \in (0,1)$, and the gross rate of return on borrowing and saving is $R > 0$. Assets saved in period j are given by a_{j+1}.

The individual/family receives exogenous income y_j during childhood periods $j = 1,2$. We will sometimes refer to these as (early and late) parental income; although, it may also include government transfers or earnings while older children are still enrolled in school (in period 2).[27]

Children/families allocate their resources to consumption and skill investment, leaving some assets/debt for when the child grows up:

$$\max_{c_1,c_2,i_1,i_2,a_2,a_3} \mathbb{E}\left[u(c_1) + \nu i_1 + \beta u(c_2) + \beta^2 V(a_3,h_3)\right] \tag{2.4}$$

subject to human capital production equations (2.1) and (2.2); budget constraints

$$a_{j+1} = Ra_j + y_j - i_j - c_j \quad \text{for } j = 1,2; \tag{2.5}$$

initial assets a_1 given; and where $V(a_3,h_3)$ represents the child's utility in adulthood given a_3 and h_3.

We now consider the main mechanisms commonly thought to explain income-based gaps in early investment and skill levels. We analyze each mechanism separately, abstracting from the others, in order to highlight the key underlying forces of each mechanism and the extent to which it can explain other stylized facts discussed in Section 2.3. In the next two sections, we abstract from uncertainty and other information problems, considering the consumption value of investment and the intergenerational correlation of ability. In Section 2.7, we study the implications of uncertainty and mis-information, at which point we describe information sets and variables over which expectations are taken in Equation (2.4). Finally, in Section 2.8 we introduce restrictions on borrowing of the form $a_{j+1} \geq -L_j$, where L_j is an upper limit on the amount that can be borrowed in period j. Until then, we assume that borrowing and saving are unrestricted.

2.5 Correlated Ability

We begin by studying the implications of a positive intergenerational correlation in ability, which is likely to generate a positive correlation between a child's ability and lifetime parental income, i.e. $\mathrm{Cov}(\theta, Y) > 0$. Indeed, this is the starting point for many economic theories of

[27] Even if one considers the child to be the sole decision maker with y_1 and y_2 reflecting inter vivos transfers from parents, the interpretations in the text regarding parental income carry through as long as transfers are strictly increasing in parental income.

intergenerational correlations in human capital and earnings (Becker and Tomes, 1979, 1986, Loury, 1981, Restuccia and Urrutia, 2004, Caucutt and Lochner, 2012, Cunha, 2013). To focus on this potential explanation for income-based gaps in investment and achievement, we study the effects of θ on investments, marginal returns to investment, and human capital, while abstracting from any consumption value of schooling, uncertainty, and credit constraints. Specifically, we assume $\nu = 0$ and that families have full and perfect information about the productivity of human capital investments. In the absence of borrowing constraints, the length of adulthood is irrelevant for our analysis, so we simply consider a three-period problem with $V(a_3, h_3) = u(Ra_3 + h_3)$ and a single lifetime budget constraint. For expositional purposes, we normalize $z = 1$.

With these assumptions, the problem can be written as:

$$\max_{c_1, c_2, c_3, i_1, i_2} \{u(c_1) + \beta u(c_2) + \beta^2 u(c_3)\}$$

subject to the lifetime budget constraint:

$$c_1 + R^{-1}c_2 + R^{-2}c_3 = Y - i_1 - R^{-1}i_2 + R^{-2}\theta f(i_1, i_2),$$

where $Y \equiv Ra_1 + y_1 + R^{-1}y_2$, which we will often (loosely) refer to as 'lifetime parental income'. Notice, Y may also include initial family assets and government transfers; however, we can interpret the effects of changes in Y as changes in parental income holding these constant.

Optimal investments must satisfy the following first order conditions:

$$\theta f_1(i_1, i_2) = R^2, \tag{2.6}$$

$$\theta f_2(i_1, i_2) = R. \tag{2.7}$$

When investments are made purely for investment purposes, they are chosen to equate the marginal labor market returns to investment, $\frac{\partial h_3}{\partial i_j} = \theta f_j(i_1, i_2)$ in both periods $j = 1, 2$, with the corresponding return to savings. This is the well-known result of Becker (1975): in the absence of borrowing constraints, uncertainty, and a direct utility value from investment, human capital investments simply maximize discounted lifetime earnings net of investment expenditures. Importantly, this relationship holds regardless of ability or family income. As such, investments are independent of family income, Y, given ability. The role of ability is summarized in the following proposition. (All proofs can be found in Appendix A.2.)

Proposition 2.1 *Optimal investments satisfy the following: (i) the marginal returns to investments are independent of ability; (ii) early and late investments are strictly increasing in ability; (iii) adult human capital h_3 is strictly increasing in ability.*

Not surprisingly, a positive correlation between parental income and child ability would produce a positive correlation between parental income and child investments and skills. Yet, the marginal return on investments should be unrelated to parental income, because investments in all children equate their marginal returns to the interest rate. This is inconsistent with both stylized Facts 1 and 2, which document returns to early investments for poor children that exceed standard interest rates as well as the returns for more economically advantaged children.

Additionally, the model implies no causal relationship between parental income and child investments/skills. Holding the child's ability constant, there should be no correlation between investments and parental income.[28] Furthermore, exogenous changes in parental income should have no effect on investments in children, contradicting stylized Facts 3 and 4.

2.6 Consumption Value of Investment

We next explore the implications of a consumption/utility value associated with investment in children. Lazear (1977) provides an early analysis of education as a joint producer of human capital/earnings and utility. Keane and Wolpin (2001), Cunha, Heckman, and Navarro (2005), and Carneiro, Heckman, and Vytlacil (2011) emphasize and estimate the role of heterogeneity in the 'consumption' value of schooling in explaining differences in schooling behavior, while Abbott et al. (2013) explicitly consider differences in tastes for schooling by parental wealth. To study the implications of this mechanism for investments in young children, we incorporate tastes for schooling $\nu \neq 0$ as in Equation (2.4), while continuing to assume perfect information, no credit constraints with $V(a_3,h_3) = u(Ra_3 + h_3)$, and $z = 1$.

For simplicity, we assume that $\beta = R^{-1}$ so that optimal consumption profiles are flat: $c_t = c = B^{-1}[Y - i_1 - R^{-1}i_2 + R^{-2}\theta f(i_1,i_2)]$ for $t = 1,2,3$, where $B = 1 + R^{-1} + R^{-2}$. In the absence of any consumption value associated with late investment, i_2 is still determined from the first order condition above (Equation (2.7)); however, optimal early investment must now satisfy:

$$\theta f_1(i_1,i_2) = \left[1 - \frac{\nu}{u'(c)}\right] R^2. \tag{2.8}$$

When investment provides a direct consumption or utility value to children or their families, this must be taken into account when making investment decisions, driving a wedge between the marginal labor market return to investment and the return to savings. For a positive consumption value ($\nu > 0$), early investment will have a low labor market return on the margin

[28] Figures 1 and 5 suggest that there is likely a strong intergenerational ability correlation, since the relationship between family income and both achievement and investments is much weaker (though still non-trivial) once we control for maternal characteristics like AFQT and educational attainment.

(i.e. $\theta f_1 < R^2$), because families will want to invest beyond the point where lifetime income is maximized.[29] The opposite is true if families/children dislike investment ($\nu < 0$).

When investment has a non-zero consumption value, the effects of parental income on investments and human capital, as well as the marginal labor market return to early investment, are easily derived from the the first order conditions above and are summarized in the following proposition.

Proposition 2.2 *For $\nu > 0$ ($\nu < 0$), optimal investments satisfy the following: (i) the marginal return to early investment is strictly less (greater) than the return to savings and is strictly decreasing (increasing) in lifetime parental income Y; (ii) early investment is strictly increasing (decreasing) in lifetime parental income Y; (iii) later investment is increasing (decreasing) in lifetime parental income Y if and only if $f_{12} \geq 0$; and (iv) final human capital h_3 is strictly increasing (decreasing) in lifetime parental income Y.*

These results are intuitive. If families enjoy investing in their children ($\nu > 0$), they will invest beyond their income maximizing amounts and will invest more if their income rises. Thus, the positive relationship between family income and early childhood investment and skills requires a positive consumption value. Proposition 2.2 shows that if $\nu > 0$, the marginal labor market return to early investment should be low (inconsistent with Fact 1) and decreasing in lifetime parental income (consistent with Fact 2). A positive consumption value predicts that early investment should rise with exogenous increases in lifetime parental income (consistent with Fact 3); however, it predicts that the timing of that income is irrelevant (inconsistent with Fact 4).

One might reconcile the high estimated returns for early interventions targeted to economically disadvantaged children (Fact 1) by assuming that low-income parents find investment in their children costly (i.e. $\nu < 0$ for low-income families). However, this would then imply that investment in young children and their skill levels should decline when poor parents receive additional income, contradicting Fact 3.

2.7 Confusion

Poor families may face greater uncertainty about the returns to investment, or they may simply maintain inaccurate beliefs about the productivity of various investments. We next examine how uncertainty and mis-information influence investment behavior and highlight the importance of accounting for the dynamic nature of skill production.

[29] This result also holds if parents value the child's final human capital level instead of early investment itself.

We begin this analysis by studying uncertainty about θ. In our general framework, this may reflect uncertainty about the child's ability to learn, parents' abilities to teach, or even the price of skill (including idiosyncratic variation, e.g., due to search frictions) in the labor market. We first consider the role of risk aversion, assuming that θ is revealed only after all investments have been made. We then consider uncertainty about θ, as well as the marginal productivity of early investments z, when that uncertainty is completely resolved after early investments but before late investments have been made. To focus on the nature of skill production and irreversibility of investments, these results abstract from risk aversion. Objective uncertainty (i.e. rational expectations) about θ (reflecting ability or skill prices) is most commonly assumed in the human capital literature; however, a growing number of studies highlight the role of subjective uncertainty, typically about the returns to education (Dominitz and Manski, 1996, Attanasio and Kaufmann, 2009, Stinebrickner and Stinebrickner, 2014, Wiswall and Zafar, 2015). We discuss both forms of uncertainty.

The literature on subjective beliefs about child development also explores biases in those beliefs. For example, Nguyen (2008) and Jensen (2010) document downward biased beliefs (on average) about the returns to education in Madagascar and the Dominican Republic, respectively, further showing that these beliefs respond to newly provided information. Both studies show that actual schooling choices can be influenced by the provision of new information; however, Jensen (2010) estimates no behavioral responses among the most poor who may be financially constrained.[30] In this chapter, we are more interested in the implications of biased beliefs about the productivity/value of earlier investments in children. Cunha, Elo, and Culhane (2013) and Cunha (2014) show that a sample of mothers from Philadelphia under-estimates, on average, the value of time investments for cognitive development in young children. Cunha (2014) further demonstrates that black mothers are more pessimistic about the productivity of these investments than white mothers, arguing that this difference may explain one-fourth of black-white early investment gaps. Dizon-Ross (2015) shows that parents in Malawi hold distorted beliefs about their child's school achievement levels with greater biases held by the least-educated parents. Furthermore, providing accurate information in a simple format for parents to understand leads to a re-allocation in the types of investments they make in their children (e.g. purchasing remedial vs. advanced workbooks) with greater responses observed among less-educated parents. Interestingly, the differential responses by parental education do not lead to corresponding differences in schooling outcomes (e.g. educational expenditures, school attendance).

[30] Consistent with the latter finding, Attanasio and Kaufmann (2009) estimate that subjective expectations about the returns to school are not significant predictors of college attendance for youth at the bottom of the income and wealth distribution in Mexico, while they are significant predictors among those that are not as poor. We discuss the role of borrowing constraints in Section 2.8.

Motivated by this literature on subjective beliefs, we study early investments and their marginal returns when there are many types of early investment activities/inputs, and parents are mis-informed about the productivity of those activities/inputs. In particular, we consider the implications of both *systematic* and *non-systematic* bias. By systematic bias, we mean incorrect beliefs about the marginal productivity of all types of early investments, z, as emphasized by Cunha, Elo, and Culhane (2013) and Cunha (2014). Non-systematic bias refers to incorrect beliefs about the relative productivity of different types of early investments as discussed in Dizon-Ross (2015).

In the following analysis, we use the term *beliefs* to reflect a family's subjective probability distribution for some (productivity) parameter(s). For much of our analysis, it does not matter whether these beliefs reflect actual variation or simply subjective uncertainty. We use the term *purely objective uncertainty* to refer to the case where beliefs coincide with the actual probability distribution for the parameter(s) of interest. This is also commonly referred to as rational expectations. We use the term *purely subjective uncertainty* to refer to the case where beliefs are non-degenerate even though the actual probability distribution is degenerate (i.e. if the true distribution for parameters were known, there would be no uncertainty). The distinction between *purely objective* or *purely subjective* uncertainty is mainly important for the realized marginal labor market returns to investments, because the former implies a distribution of ex post marginal returns for the same investments while the latter does not. We return to this point below.

Throughout this section, we continue to assume $\nu = 0$ and $V(a_3, h_3) = u(Ra_3 + h_3)$ in order to focus on the role of information frictions.

2.7.1 Risk Aversion and Uncertain Returns

We begin with a very natural form of *purely objective uncertainty*: both beliefs and the true distribution of θ are given by $\theta \sim \Phi(\theta)$, with $\bar{\theta} \equiv \mathbb{E}(\theta)$. For this analysis, we assume that the true value of θ is not revealed until after all skill investments have been made. We continue to normalize $z = 1$.

If individuals are risk averse, the expected return to risky investments should exceed the return on safe investments if individuals are to hold risky assets at all. With a concave human capital production technology, this means that skill investments will be lower under uncertainty (Levhari and Weiss, 1974).

The first order conditions for investments satisfy

$$f_1(i_1, i_2) = Rf_2(i_1, i_2).$$

Given the separability between θ and $f(\cdot)$, there is no distortion between early and late investment even if total investment spending is distorted. That is, for a given level of spending $i_1 + R^{-1}i_2$, early and late investments are chosen to maximize $f(i_1, i_2)$. Assumption 2.1 ensures that both i_1 and i_2 increase when total investment spending increases.

The level of total investment spending will equate the expected marginal benefit with the marginal cost of investment, so

$$\bar{\theta} f_1(i_1,i_2) + \frac{\mathrm{Cov}\left(u'(c_3(\theta)),\theta\right)}{\mathbb{E}\left[u'(c_3(\theta))\right]} f_1(i_1,i_2) = R^2,$$

where $c_3(\theta) = Ra_3 + \theta f(i_1,i_2)$ is optimal period 3 consumption in state θ. The expected marginal benefit of investment consists of a monetary return (first term) and a utility cost (second term). Because the marginal utility of consumption is low in states with a high return to investment, uncertain returns produce an additional utility cost of investment as reflected in the (negative) covariance term. As such, the expected marginal labor market returns to investment exceed the return to savings:

$$\bar{\theta} f_1(i_1,i_2) = R\bar{\theta} f_2(i_1,i_2) > R^2.$$

Risk averse individuals facing uncertain returns invest less at all ages relative to those who know the return with certainty.

Moreover, if having greater resources makes people less risk averse, then investments are increasing and marginal labor market returns decreasing in lifetime parental income Y. Because no information about the value of θ is revealed until all investments have been made, choices depend only on the discounted present value of income over all investment years and not the timing of that income. The following proposition summarizes these results.[31]

Proposition 2.3 *When there is uncertainty in the final return to investment θ, optimal investments satisfy the following: (i) expected marginal returns to investment are strictly greater than the return to savings; (ii) expected marginal returns to investment are strictly decreasing in parental income Y if $u(\cdot)$ exhibits decreasing absolute risk aversion; and (iii) early and late investments are strictly increasing in parental income Y if $u(\cdot)$ exhibits decreasing absolute risk aversion.*

[31] These results assume *purely objective uncertainty*; however, they also apply to the case of unbiased *subjective uncertainty*. The only difference in the latter case is that all individuals would experience the same marginal return to investments, given by the expected marginal returns in the case of *purely objective uncertainty*. Thus, the marginal returns to investment exceed the return on savings (and decline with income) regardless of the form of uncertainty.

2.7.2 Uncertain Returns and the Irreversibility of Early Investment

We now consider the case in which families face uncertainty when making early investments in their children; however, that uncertainty is fully resolved before late investments are chosen. While early investments made under uncertainty are irreversible (i.e. families cannot go back in time to modify ex post suboptimal early investment choices), families can base late investment decisions on the realizations of early investments and full knowledge of the human capital production process. Thus, families may be able to partially compensate for ex post suboptimal early investment through their choice of late investment. The extent to which this is effective depends crucially on the intertemporal complementarity/substitutability of investments.[32] It also depends on which features of technology are unknown. We consider uncertainty in the productivity of both investments, θ, as well as in the productivity of early investments alone, z, at the time early investments are made.

To focus on the implications of investment irreversibility and the dynamic nature of human capital productivity, we abstract from risk aversion. We continue to focus on *purely objective uncertainty* with the distribution of beliefs over (θ, z) reflecting the true variation in these productivity parameters; however, we comment briefly on the implications of *purely subjective uncertainty* at the end of our discussion.

It is useful to begin with the second period investment problem, which conditions on early investment and technology once θ and z are known. Let $\hat{i}_2(zi_1, \theta)$ denote the optimal second period investment conditional on i_1 and technology state (θ, z):

$$\hat{i}_2(zi_1, \theta) \equiv \underset{i_2}{\operatorname{argmax}} \left\{ -i_2 + R^{-1} \theta f(zi_1, i_2) \right\}. \tag{2.9}$$

Optimal late investment equates the marginal labor market return with the return to savings, i.e. $\theta f_2(zi_1, i_2) = R$. From this, it is easy to see that second period investment is increasing in θ. However, late investment is increasing in z if and only if early and late investments are gross complements (i.e. $f_{12} \geq 0$), because z only affects the marginal return to late investment indirectly through $h_2 = zi_1$. Here, we begin to see the distinction between neutral and early-specific productivity as well as the importance of intertemporal complementarity/substitutability of investments. These factors are also important in determining the response of early investment to uncertainty about investment productivity.

Taking the late investment policy $\hat{i}_2(zi_1, \theta)$ as given, the net realized (or ex post) return to

[32] See Cunha and Heckman (2007) for a discussion of dynamic complementarity and the difficulty of compensating for low early investments by increasing late investments. Cunha, Heckman, and Schennach (2010) estimate and Caucutt and Lochner (2012) calibrate a strong degree of intertemporal complementarity for investments in children.

early investment for actual productivity parameters (θ, z) is given by:

$$\Pi(i_1, \theta, z) \equiv -i_1 - R^{-1}\hat{i}_2(zi_1, \theta) + R^{-2}\theta f\big(zi_1, \hat{i}_2(zi_1, \theta)\big). \tag{2.10}$$

The following lemma establishes concavity of net realized returns in early investment and is useful for a number of results.

Lemma 2.1 *The net return to early investment* $\Pi(i_1, \theta, z)$ *is strictly concave in* i_1.

Because i_1 must be determined before (θ, z) is realized, optimal early investment maximizes the expected net return:

$$\tilde{i}_1 \equiv \underset{i_1}{\operatorname{argmax}} \, \mathbb{E}\left[\Pi(i_1, \theta, z)\right],$$

where the expectation is taken over the distribution of (θ, z). The first order condition equates the expected marginal labor market return to early investment with the return to savings:

$$\mathbb{E}\left[z\theta f_1\big(z\tilde{i}_1, \hat{i}_2(z\tilde{i}_1, \theta)\big)\right] = R^2. \tag{2.11}$$

With *purely objective uncertainty* and risk neutrality, the expected marginal labor market return to early investment always equals the return to savings regardless of the type (i.e. θ or z) or extent of uncertainty – the expected marginal return is independent of the (θ, z) distribution. In contrast with Facts 1 and 2 of Section 2.3, the average marginal labor market return should equal the interest rate for children from all backgrounds.

We are also interested in understanding how changes in the distribution of (θ, z) affect early investment amounts. We consider two notions of a change in the distribution that are widely used in economics: *first order stochastic dominance* and *mean-preserving spread* (Rothschild and Stiglitz, 1970, 1971). In this framework, how $\tilde{i}_1$ changes with the distribution of productivity parameters depends on how $\partial\Pi(i_1, \theta, z)/\partial i_1$ varies with (θ, z). If $\partial\Pi(i_1, \theta, z)/\partial i_1$ is increasing (decreasing) in θ, a first order stochastic dominance shift in θ will increase (decrease) $\tilde{i}_1$. If $\partial\Pi(i_1, \theta, z)/\partial i_1$ is concave (convex) in θ, a mean-preserving spread in θ decreases (increases) $\tilde{i}_1$. The same is true for changes in the distribution of z. We first consider investment when θ is unknown, then turn attention to the case with z unknown. Some of our results assume a CES production function for human capital as defined in Equation (2.3).

Neutral Productivity Shock We now consider uncertainty in the overall ability of a child (θ) that is fully resolved after early investments have been made but before late investments are chosen. We assume z is known.[33]

Proposition 2.4 *(i) A first order stochastic dominance shift in θ increases early investment. (ii) For the CES production function (2.3), a mean-preserving spread in θ reduces early investment if and only if $b > d$.*

It is not surprising that a first order stochastic shift in θ unambiguously increases early investment, because θ directly raises the marginal return to investment for any given level of early and late investment. The effect of a mean-preserving spread in the distribution of θ is more complicated and depends on the degree of complementarity between investments. For a CES human capital production function, an increase in uncertainty about θ reduces early investment if and only if early and late investments are gross substitutes (i.e. $b > d \Leftrightarrow f_{12} < 0$). With strong intertemporal substitutability, families facing uncertainty about θ will choose to invest little in the first period and wait to learn the productivity of investment. If investment is highly productive, the family can easily compensate for inadequate early investment by investing more in the second period. In the more empirically relevant case where investments are gross complements ($f_{12} > 0$), it is too costly to make up for a lack of early investment by increasing late investment. As a result, early investment increases with the degree of uncertainty.

Early Investment-Specific Productivity Shock Families may be more uncertain about the productivity of early investments in their children than they are about later investments like college attendance. To explore this possibility, we now assume θ is known from birth and consider the case where z is initially unknown but revealed after early investments have been made.

A change in z has two opposing effects on the marginal return to early investment. An increase in z directly increases the productivity of i_1, but it also reduces the marginal return because $h_2 = zi_1$ is subject to diminishing returns in the production of adult human capital h_3. The latter effect is attenuated by adjustments in late investments when $f_{12} \neq 0$. The overall effect of z depends on the following condition:

Condition 2.1 $f_1 > \left(\frac{f_{11}f_{22}-f_{12}^2}{-f_{22}}\right) zi_1$.

If the direct productivity effect (left hand side) is greater than the the diminishing return effect (right hand side), then the marginal return to i_1 is greater for larger z. For the CES production function given in Equation (2.3), this condition holds if $b \geq 0$.

Condition 2.1 is appealing, because it is equivalent to requiring that an increase in z raises the net marginal return to early investment.

[33]This case was originally considered by Hartman (1976) in the analysis of firm investment and labor demand under uncertain output prices.

Lemma 2.2 *The marginal net return to early investment $\partial\Pi(i_1,\theta,z)/\partial i_1$ is strictly increasing in z if and only if Condition 2.1 holds.*

When Condition 2.1 is satisfied, a better distribution of z produces a higher marginal return to i_1, on average, which makes it profitable to increase early investment. The following proposition formalizes this result and characterizes the effects of a mean-preserving spread in z.

Proposition 2.5 *For the CES production function (2.3), (i) a first order stochastic dominance shift in z increases early investment if $b \geq 0$; (ii) a mean preserving spread in z reduces early investment if $b \geq 0$.*

The effect of a mean-preserving spread in z is similar to its counterpart for θ (Proposition 2.4), except that uncertainty in z discourages early investment more than does uncertainty in θ. In contrast with an increase in uncertainty about θ, an increase in uncertainty about z can reduce early investment even when early and late investment are gross complements (e.g. $0 \leq b \leq d$). Intuitively, diminishing marginal returns to $h_2 = zi_1$ in the production of adult human capital lessens the benefits of high z realizations for the marginal return to early investment. This force moderates the costs of under-investment and discourages early investment when z is uncertain, but it is absent with uncertainty in θ.

Purely Subjective Uncertainty, Investments, and Marginal Returns The previous analysis assumes individuals have *purely objective uncertainty* about heterogeneous productivity levels (θ,z). Even if all children have the same productivity levels, they may have different subjective beliefs about the true productivity of investments. For example, the poor may have downward biased beliefs or they may have unbiased beliefs with greater subjective uncertainty. Regardless, Propositions 2.4 and 2.5 characterize the effects of changes in beliefs on early investment choices.

More interestingly, *purely subjective uncertainty* has different implications from *purely objective uncertainty* for observed marginal returns in the labor market. In the latter case, families facing the same distribution of productivity levels make the same early investment choices, but they experience different labor market outcomes due to heterogeneous productivity levels. As discussed earlier (see the discussion surrounding Equation (2.11)), the average realized marginal labor market return under *purely objective uncertainty* always equals the return to savings. The case of *purely subjective uncertainty* is quite different. All families with the same beliefs and ability/productivity will make the same early investment choices and will, therefore, experience the same labor market returns. Strict concavity of $\Pi(i_1,\theta,z)$ in i_1

(Lemma 2.1) directly implies that under *purely subjective uncertainty*, the observed marginal labor market return to early investment is strictly decreasing in the level of early investment. This, together with Propositions 2.4 and 2.5, directly implies the following corollary.

Corollary 2.1 *Assume the CES production function given in Equation (2.3). Under purely subjective uncertainty, a mean-preserving spread in the distribution of beliefs about θ (z) increases the marginal labor market return to early investment if and only if $b > d$ (if $b \geq 0$).*

Families with greater subjective uncertainty about θ will have a higher marginal labor market return to early investment if and only if early and late investments are gross substitutes. Marginal labor market returns will be increasing in the amount of subjective uncertainty about z under a modest amount of dynamic complementarity in investments.

2.7.3 Biased Beliefs about Human Capital Production

Thus far, we have focused on the extent of uncertainty about the productivity of investments. It is also possible that parents may have little subjective uncertainty about the productivity of investments, but their beliefs may be biased. This possibility seems particularly likely to arise when there are many different potential inputs/activities families may engage in to raise the human capital of their children (e.g. reading to children, taking them to museums, teaching them to play musical instruments, playing with them). Even if parents are correct in gauging the average productivity across different inputs, they might easily misjudge their relative productivity. To explore this issue, we now assume homogeneity in the productivity of different early inputs across families; however, we allow for the possibility that families may hold biased beliefs about the productivity of any or all early child inputs. To simplify the analysis, we abstract away from any form of uncertainty – families are certain but may be wrong. We further assume that families learn the true outcomes of their early investments, h_2, before they need to make later investment decisions.

Assume early investment consists of n different 'activities' $\boldsymbol{x} = (x_1, \ldots, x_n)$ that produce h_2 according to the following CES production function:

$$h_2 = z \left(\sum_{j=1}^{n} w_j^{1-\phi} x_j^{\phi} \right)^{\frac{1}{\phi}}, \tag{2.12}$$

where $\phi \in (0,1)$ and $\boldsymbol{w} = (w_1, \ldots, w_n) \geq 0$ satisfies $\sum_{j=1}^{n} w_j = 1$.[34] Here, z reflects the total factor productivity of early investments. Changes in z have no affect on the relative productivity

[34] Vector equality and inequality are defined as follows: (i) $\tilde{\boldsymbol{x}} = \boldsymbol{x}$ if $\tilde{x}_j = x_j$ for all $j = 1, \ldots, n$; (ii) $\tilde{\boldsymbol{x}} \neq \boldsymbol{x}$ if $\tilde{x}_j \neq x_j$ for some $j = 1, \ldots, n$; (iii) $\tilde{\boldsymbol{x}} \leq \boldsymbol{x}$ if $\tilde{x}_j \leq x_j$ for all $j = 1, \ldots, n$.

or optimal composition of different inputs. Productivity weights w_j determine the relative importance of each input as well as their optimal expenditure shares. It is straightforward to show that demand for input x_j conditional on total early investment spending $i_1 = \sum_{j=1}^{n} x_j$ is given by

$$x_j = w_j i_1.$$

By substituting these conditional demands into the production function, we obtain the indirect production function (as a function of total early expenditure i_1) equivalent to that assumed earlier:

$$h_2 = z\left(\sum_{j=1}^{n} w_j^{1-\phi}(w_j i_1)^{\phi}\right)^{\frac{1}{\phi}} = z i_1.$$

To investigate the implications of incorrect beliefs about early investment productivity, it is useful to distinguish beliefs from actual productivity parameters. Let $\tilde{z}$ and $\tilde{\boldsymbol{w}}$ denote a family's beliefs about z and $\boldsymbol{w}$, respectively. Without loss of generality, we assume $\sum_{j=1}^{n} \tilde{w}_j = 1$. We say that a belief is *biased* if $\tilde{z} \neq z$ or $\tilde{\boldsymbol{w}} \neq \boldsymbol{w}$. When $\tilde{z} \neq z$, the bias is *systematic* in the sense that families are, on average, biased about the productivity of early investments. When $\tilde{z} = z$ but $\tilde{\boldsymbol{w}} \neq \boldsymbol{w}$, the bias is *non-systematic*, because beliefs are, on average, correct even though they are wrong about the relative productivity of different early inputs.

Let $(\tilde{\boldsymbol{x}}, \tilde{i}_1, \tilde{h}_2, \tilde{h}_3)$ be the optimally chosen investments and realized human capital of children with family beliefs $(\tilde{z}, \tilde{\boldsymbol{w}})$. Let $(\boldsymbol{x}^*, i_1^*, h_2^*, h_3^*)$ reflect these same variables when beliefs are unbiased. Families first choose investments $\boldsymbol{x}$ and i_1 based on their beliefs $(\tilde{z}, \tilde{\boldsymbol{w}})$:

$$\begin{aligned}\tilde{i}_1 =& \operatorname*{argmax}_{i_1}\ \Pi(i_1, \boldsymbol{\theta}, \tilde{z})\\ \tilde{x}_j =& \tilde{w}_j \tilde{i}_1, \quad \forall j = 1, \ldots, n,\end{aligned}$$

where the child's lifetime income net of investments, $\Pi(\cdot,\cdot,\cdot)$, is defined by equation (2.10). Notice that total early investment spending $\tilde{i}_1$ is only affected by $\tilde{z}$ and not by $\tilde{\boldsymbol{w}}$, because $\boldsymbol{w}$ does not affect total factor productivity.

Next, interim human capital $\tilde{h}_2$ is realized based on investments choices $(\tilde{\boldsymbol{x}}, \tilde{i}_1)$ and the true technology $(z, \boldsymbol{w})$:

$$\tilde{h}_2 = z\left(\sum_{j=1}^{n} w_j^{1-\phi}\tilde{x}_j^{\phi}\right)^{\frac{1}{\phi}} = z\tau(\tilde{\boldsymbol{w}})\tilde{i}_1,$$

where

$$\tau(\tilde{\boldsymbol{w}}) \equiv \left(\sum_{j=1}^{n} w_j^{1-\phi}\tilde{w}_j^{\phi}\right)^{\frac{1}{\phi}} \leq 1$$

reflects the distortion due to a suboptimal allocation of expenditures across inputs. When $\tilde{\boldsymbol{w}} \neq \boldsymbol{w}$, early investment spending is less productive than it should be $(\tau(\tilde{\boldsymbol{w}}) < 1)$, so interim human capital is low $(\tilde{h}_2 < z\tilde{i}_1)$.

We assume that i_2 is chosen knowing the actual realization for $\tilde{h}_2$. That is, families are able to evaluate their child's skill/achievement, effectively learning that their beliefs were mistaken.[35] Given the resulting interim human capital, late investments are determined as in the previous subsection (see Equation (2.9)). Finally, adult human capital $\tilde{h}_3$ is produced based on actual $\tilde{h}_2$ and late investment:

$$\tilde{h}_3 = \theta f\big(\tilde{h}_2, \hat{i}_2(\tilde{h}_2, \theta)\big).$$

where $\hat{i}_2(.,.)$ is defined in Equation (2.9).

We first study how *systematic* bias affects early investment and human capital accumulation.

Proposition 2.6 *(i) $\tilde{i}_1$ is strictly increasing in $\tilde{z}$ if and only if Condition 2.1 holds for $(h_2, i_2) = \big(\tilde{z}\tilde{i}_1, \hat{i}_2(\tilde{z}\tilde{i}_1, \theta)\big)$. (ii) Suppose that $\tilde{\boldsymbol{w}} = \boldsymbol{w}$. If and only if $\tilde{i}_1 \leq i_1^*$, then: $\tilde{\boldsymbol{x}} \leq \boldsymbol{x}^*$, $\tilde{h}_2 \leq h_2^*$, $\tilde{h}_3 \leq h_3^*$, and $z\theta f_1\big(\tilde{h}_2, \hat{i}_2(\tilde{h}_2, \theta)\big) \geq R^2$.*

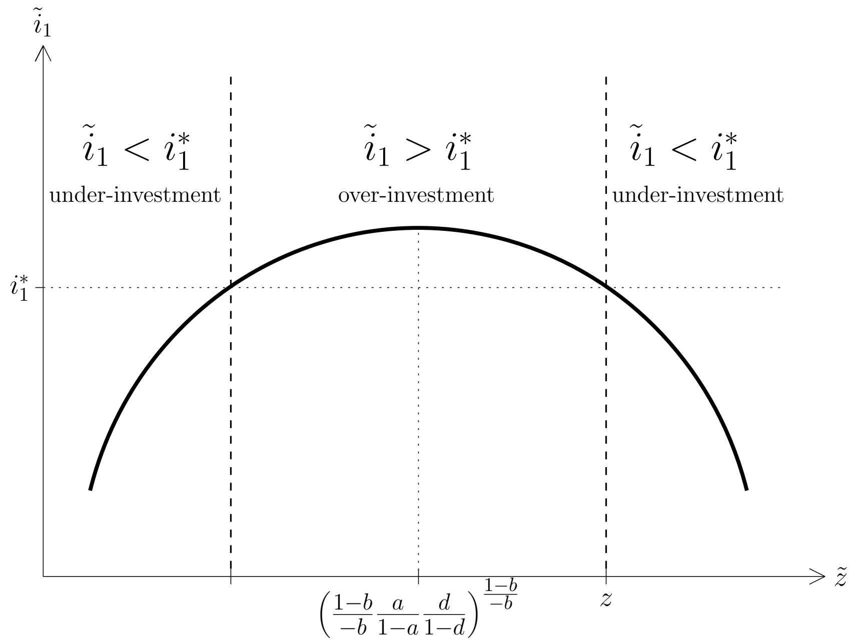
Families with a biased belief about z behave as if the true productivity of early investment is $\tilde{z}$ rather than z. As shown in Lemma 2.1, higher productivity in early investment does not necessarily lead to more early investment due to the diminishing return effect. However, when this effect is weak so Condition 2.1 is satisfied (e.g. modest dynamic complementarity or substitutability), individuals with downward biased beliefs under-invest in all early inputs/activities resulting in low levels of human capital. Moreover, under-investment implies a high observed marginal labor market return to early investment. By contrast, when Condition 2.1 does not hold, families with downward biased beliefs may over-invest (in all inputs) and obtain high levels of human capital. In this case, providing information that shifts beliefs upwards towards the truth would actually *reduce* early investment and human capital. Figure 2.6 demonstrates this possibility with CES production function (2.3) and $b < 0$. In this example, Condition 2.1 does not hold for high values of z, so moving downward biased beliefs $\tilde{z}$ from the middle of the graph towards the true (higher) value of z would result in lower (but more efficient) levels of early investment.

Next, consider the effects of *non-systematic* bias.

Proposition 2.7 *Suppose that $\tilde{z} = z$ and $\tilde{\boldsymbol{w}} \neq \boldsymbol{w}$. Then (i) $\tilde{x}_j \leq x_j^*$ if and only if $\tilde{w}_j \leq w_j$; (ii) $\tilde{i}_1 = i_1^*$, $\tilde{h}_2 < h_2^*$, and $\tilde{h}_3 < h_3^*$; (iii) $z\tau(\tilde{\boldsymbol{w}})\theta f_1\big(\tilde{h}_2, \hat{i}_2(\tilde{h}_2, \theta)\big) < R^2$ if Condition 2.1 holds for all $(h_2, i_2) = \big(z' i_1^*, \hat{i}_2(z' i_1^*, \theta)\big)$ where $z' \in [z\tau(\tilde{\boldsymbol{w}}), z]$.*

[35] Note that this does not necessarily require that parents learn the true productivity values (z, w); although, it is too late to matter.

Figure 2.6: Systematic Bias and Early Investment (CES with $b < 0$)



Non-systematic bias does not affect total early investment spending $\tilde{i}_1$, but it reduces the actual return to early investment due to the misallocation of resources to the wrong inputs. As such, it leads to low levels of human capital. In this case, providing more precise information will not affect total early investment expenditures, but it will lead to more efficient human capital production and, consequently, greater human capital. Because non-systematic bias reduces the productivity of early investment while leaving total investment expenditure unaffected, its effect on the marginal return to early investment depends on Condition 2.1. When Condition 2.1 holds, families with non-systematic bias have lower marginal labor market returns to early investment due to misallocation.

2.7.4 Information Problems and the Stylized Facts

The nature of an information problem is important for understanding its effects on human capital investment behavior. While none of the information problems we study are able to explain why the timing of income is important for human capital investment (Fact 4), some are more consistent with the other stylized facts in Section 2.3.

Uncertainty about child ability θ (or labor market returns to human capital) coupled with risk aversion causes families to under-invest in their children. With decreasing absolute risk aversion, under-investment is worse among the poor, and an increase in lifetime parental income would be met with an increase in child investments (Fact 3). Expected marginal returns exceed the return to savings (Fact 1) and are especially high for children from low-income families (Fact 2). These results apply whether uncertainty is objective or subjective.

We also explore the implications of uncertainty resolved after early investments have been made but in time for late investment choices to respond. Here, we abstract from risk aversion in order to emphasize the role of early investment irreversibility and the technology of skill formation. If early and late investments are mildly complementary, subjective uncertainty about the productivity of early investments, z, can lead to under-investment in young children and high marginal returns to early investment (relative to interest rates). If poor families face greater subjective uncertainty about z, then they will invest less in their young children than higher income families, stopping investment when marginal returns are relatively high.[36] Thus, with modest dynamic complementarity, differences in subjective uncertainty by parental income can help explain Facts 1 and 2. While this form of uncertainty can explain the positive correlation between parental income and child investments, it does not help us understand why changes in parental income lead to contemporaneous changes in child investments and achievement (Facts 3 and 4) unless income brings new information with it.[37] Uncertainty about θ (with risk neutrality) resolved after early childhood is inconsistent with a positive parental income – child investment relationship and other stylized facts unless early and late investments are substitutes.[38]

Finally, we consider the possibility that families are simply mistaken about the productivity of early investment activities as documented in Cunha (2014) and Dizon-Ross (2015). With modest dynamic complementarity, we show that poor families that systematically underestimate the productivity of early investments will under-invest in their young children and have a high marginal return to early investment relative to the return to savings and the marginal return for high income families with accurate beliefs (Facts 1 and 2). By contrast, non-systematic bias (i.e. over-estimation of the productivity of some inputs offset by under-estimation of the productivity of others) has no effect on total early investment expenditures. Instead, it results in a mis-allocation across early inputs, which tends to reduce the marginal return to early expenditures.[39] Thus, non-systematic bias among poor families cannot explain the basic correlation between family income and child investment/achievement, nor can it explain the high marginal

[36] With purely objective uncertainty and risk neutrality, expected marginal returns to investment always equal the return on savings.

[37] Changes in income may lead to changes in information and, therefore, investment behavior, if information about the productivity value of investment can be purchased by families. We do not explicitly model this possibility.

[38] Available evidence suggests dynamic complementarity for investments (Cunha, Heckman, and Schennach, 2010, Caucutt and Lochner, 2012, Attanasio et al., 2015).

[39] The prediction of over-investment in some inputs and under-investment in others conflicts with Figures 2.2 and 2.3, which shows that the poor generally invest less than the rich in nearly all inputs.

return to investment among the poor (Facts 1 and 2). Neither form of bias helps explain the responsiveness of early investment and achievement to changes in income (Facts 3 and 4).[40]

2.8 Borrowing Constraints

Lastly, we consider the possibility that families may be unable to borrow against future earnings to efficiently finance investments in their children. Studies of intertemporal consumption behavior frequently document patterns consistent with borrowing constraints, with many finding stronger evidence of binding constraints among younger households (e.g.Costas Meghir (1996), Alessie, Devereux, and Weber (1997), Stephens (2008)).[41] We analyze the implications of borrowing constraints at different stages of child development for human capital investment within our framework. Consistent with Cunha and Heckman (2007), we demonstrate the importance of dynamic complementarity of investments for the impacts of both borrowing constraints and family resources.

We incorporate borrowing constraints by imposing upper limits on the total debts families can accumulate in any period. Specifically, we restrict assets carried into any period $j+1$ to satisfy the constraint $a_{j+1} \geq -L_j$. To focus on the role of borrowing constraints, we consider the problem of Section 2.4 (with borrowing constraints) assuming $z=1$, $\nu=0$, and perfect information.

Because we are not only concerned with borrowing constraints during the investment period, but also later in life, we interpret the continuation utility $V(a_3, \theta f(h_2, i_2))$ as the solution to the asset allocation problem for individuals entering adulthood, allowing for the possibility of binding future constraints. We assume that individuals live to age T and that adult earnings depend on human capital acquired through childhood investments h_3, growing exogenously thereafter with

$$h_j = \Gamma_j h_3, \qquad j \in \{4, ..., T\}. \tag{2.13}$$

Individuals entering adulthood with human capital h_3 and assets a_3 allocate consumption across their remaining life in the following way:

$$V(a_3, h_3) = \max_{c_3,...,c_T} \sum_{j=3}^{T} \beta^{j-3} u(c_j), \tag{2.14}$$

[40] Systematic bias in z may be consistent with Fact 3 if $b>0$ and beliefs about z respond to changes in family income. See footnote 37.

[41] In related work, Attanasio, Koujianou Goldberg, and Kyriazidou (2008) conclude that many younger and middle-age American households are likely to be borrowing constrained based on differential car loan demand elasticities with respect to interest rates and loan maturity.

subject to budget constraints $a_{j+1} = Ra_j + h_j - c_j$ for $j \in \{3, ..., T\}$, borrowing constraints $a_{j+1} \geq -L_j$ for $j \in \{3, ..., T-1\}$, and $a_{T+1} = 0$.[42]

Given the value function defined in (2.14), families solve the maximization problem (2.4) subject to budget constraints (2.5), initial assets a_1, and borrowing constraints $a_2 \geq -L_1$ and $a_3 \geq -L_2$.

2.8.1 Investment Behavior

Consumption allocations satisfy $u'(c_j) \geq \beta R u'(c_{j+1})$, $\forall j = 1, ..., T-1$, where the inequality is strict if and only if the borrowing constraint for period j ($a_{j+1} \geq -L_j$,) binds. First order conditions for investment are given by

$$u'(c_1) = \beta^2 \theta \frac{\partial V(a_3, h_3)}{\partial h_3} f_1(i_1, i_2), \tag{2.15}$$

$$u'(c_2) = \beta \theta \frac{\partial V(a_3, h_3)}{\partial h_3} f_2(i_1, i_2), \tag{2.16}$$

where $\frac{\partial V(a_3,h_3)}{\partial h_3} = \sum_{j=3}^{T} \beta^{j-3} \Gamma_j u'(c_j) > 0$ with $\Gamma_3 = 1$. Taking the ratio of these equations reveals that optimal investment equates the technical rate of substitution in the production of human capital with the marginal rate of substitution for consumption: $\frac{f_1(i_1,i_2)}{f_2(i_1,i_2)} = \frac{u'(c_1)}{\beta u'(c_2)} \geq R$.

Unconstrained optimal investments, i_1^* and i_2^*, satisfy $\chi \theta f_1(i_1^*, i_2^*) = R^2$ and $\chi \theta f_2(i_1^*, i_2^*) = R$, where $\chi = \sum_{j=3}^{T} R^{3-j} \Gamma_j$ reflects the discounted present value of an additional unit of human capital. As in Section 2.5, unconstrained investments maximize the discounted present value of lifetime earnings net of investment costs. They are independent of the marginal utility of consumption and income/transfers, because individuals can optimally smooth consumption across periods. This is not true when borrowing constraints bind as shown in the next proposition.

Proposition 2.8 *(i) If and only if any borrowing constraint binds, then: optimal early investment is strictly less than the unconstrained amount, the marginal return to early investment is strictly greater than the return to savings, and adult human capital is strictly less than the unconstrained level. (ii) If any borrowing constraint binds and either (a) the period one constraint does not bind or (b) $f_{12} > 0$, then optimal late investment is strictly less than the unconstrained amount. (iii) If and only if any borrowing constraint in period two or later binds, then the marginal return to late investment is strictly greater than the return to savings.*

Early investment is always low and its marginal labor market return high (relative to the unconstrained case) when any borrowing constraints bind. Late investment is also low and its

[42] Appendix A.2 shows that $V_2 > 0$, $V_{22} < 0$, and $V_{21} < 0$. These properties are used repeatedly in proving results below.

marginal return high if constraints at that age or later are binding and either the early constraint does not bind or early and late investments are complementary.

The complementarity of investments across periods plays a central role in determining individual responses to borrowing constraints and changes in parental income. If investments are very substitutable, individuals can shift investment from constrained periods to unconstrained periods with little loss to total acquired human capital. Their ability to do this is diminished as investments become more complementary. In particular, the following dynamic complementarity condition is important for a number of results.

Condition 2.2 $\frac{f_{12}f}{f_1 f_2} > \frac{-V_{22}(-RL_2,h_3)h_3}{V_2(-RL_2,h_3)}$.

If preferences are given by the constant intertemporal elasticity of substitution (IES) form $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$ (where $1/\sigma$ is the IES) and credit constraints are non-binding throughout adulthood, then this condition simplifies to something very intuitive:

$$\underbrace{\frac{f_1 f_2}{f_{12} f}}_{\text{Hicksian elasticity of substitution}} < \underbrace{\frac{1}{\sigma}}_{\text{IES}} \underbrace{\left(1 - \frac{RL_2}{\chi h_3}\right)}_{\text{1-}\frac{\text{maximum debt}}{\text{lifetime income}}} .$$

See Appendix A.2 for details.[43] As the Hicksian elasticity of substitution between early and late investments declines (i.e. investments become more complementary) or the consumption intertemporal elasticity of substitution increases (i.e. individuals become less concerned about maintaining smooth consumption profiles), this inequality is more likely to hold. More generally, when individual preferences for smooth consumption are strong, Condition 2.2 requires strong complementarity between early and late investments.

We are now ready to study how family income during early and late childhood affect investment behavior. As noted above, changes in family income have no effect on investments for unconstrained individuals. The following proposition shows how constraints at different stages of child development determine the responsiveness of investment to changes in income at early and late ages. These results highlight how the timing of income/transfers can impact human capital investments and accumulation when individuals are constrained.[44]

Proposition 2.9

I. If borrowing constraints bind in late childhood, but not early childhood, then:

[43] For the CES production function given in equation (2.3), the Hicksian elasticity of substitution between early and late investments (the left hand side) is simply $\frac{d}{d-b}$. The condition cannot hold for $d \leq b$, but this only rules out very strong substitution between early and late investments such that $f_{12} \leq 0$.

[44] See Cunha and Heckman (2007) for a related analysis of the impacts of early vs. late income on the early-to-late investment ratio i_1/i_2 when the early borrowing constraint binds.

(i) $\frac{\partial i_1}{\partial y_1} = R\frac{\partial i_1}{\partial y_2} = \frac{\partial i_1}{\partial (R^{-1}y_2)} > 0$*;*
(ii) $\frac{\partial i_2}{\partial y_1} = R\frac{\partial i_2}{\partial y_2} = \frac{\partial i_2}{\partial (R^{-1}y_2)} > 0$*;*
(iii) $\frac{\partial h_3}{\partial y_1} = R\frac{\partial h_3}{\partial y_2} = \frac{\partial h_3}{\partial (R^{-1}y_2)} > 0$*.*

II. If borrowing constraints only bind in early childhood, then:
(i) $\frac{\partial i_1}{\partial y_1} > 0$*; and* $\frac{\partial i_1}{\partial y_2} < 0$*;*
(ii) $\frac{\partial i_2}{\partial y_1} > 0 \iff f_{12} > 0$*; and* $\frac{\partial i_2}{\partial y_2} < 0 \iff f_{12} > 0$*;*
(iii) $\frac{\partial h_3}{\partial y_1} > 0$*; and* $\frac{\partial h_3}{\partial y_2} < 0$*.*

III. If borrowing constraints bind during both early and late childhood, then:
(i) $\frac{\partial i_1}{\partial y_1} > 0$*; and* $\frac{\partial i_1}{\partial y_2} > 0 \iff$ *Condition 2.2 holds;*
(ii) $\frac{\partial i_2}{\partial y_1} > 0 \iff$ *Condition 2.2 holds; and* $\frac{\partial i_2}{\partial y_2} > 0$*;*
(iii) $\frac{\partial h_3}{\partial y_1} > 0$ *and* $\frac{\partial h_3}{\partial y_2} > 0$*.*

There are two key implications of this proposition. First, if the late constraint binds but the early constraint does not, then investments depend only on the discounted present value of family income $y_1 + R^{-1}y_2$, not the timing of income (conditional on discounting y_2). Second, when the early constraint binds, the response of investments to changes in income depends on the timing of income, the extent of dynamic complementarity, and whether late constraints are also binding. While constrained early investment is always increasing in y_1, this is not necessarily the case for changes in y_2. Because an increase in late income exacerbates the early borrowing constraint, early investment is unambiguously decreasing in y_2 when the late constraint does not also bind. Intuitively, families would like to consume some of the increased late income in the earlier period as well; however, if they are borrowing constrained, they can only do this by reducing early investments. When only the early constraint binds, the impacts of income on late investment depend entirely on its effect on early investment and whether early investment raises ($f_{12} > 0$) or lowers ($f_{12} < 0$) the marginal return to late investment. Perhaps surprisingly, when $f_{12} > 0$ and only the early constraint binds, then an increase in family income during late childhood *reduces* skill investments in both periods. When constraints are binding throughout (early and late) childhood, increases in income in any period increase investment in both periods if and only if there is sufficient dynamic complementarity.

To better understand the implications of policies aimed at expanding credit for educational investments, we consider the impacts of raising borrowing limits for families at different stages of child development, beginning with limits faced by families with older children (e.g. expanding student loan programs for higher education).

Proposition 2.10 *Assume that the borrowing constraint binds during late childhood (i.e.* $a_3 = -L_2$*).*

(i) If the early borrowing constraint does not bind (i.e. $a_2 > -L_1$), then: $\frac{\partial i_1}{\partial L_2} > 0$, $\frac{\partial i_2}{\partial L_2} > 0$, *and* $\frac{\partial h_3}{\partial L_2} > 0$.

(ii) If the early borrowing constraint also binds (i.e. $a_2 = -L_1$), then: $\frac{\partial i_1}{\partial L_2} > 0$ *if Condition 2.2 holds;* $\frac{\partial i_2}{\partial L_2} > 0$*; and* $\frac{\partial h_3}{\partial L_2} > 0$.

Relaxing the borrowing constraint during late childhood unambiguously increases late investment. If the early constraint is non-binding or if early and late investments are sufficiently complementary, then any increase in late investment encourages additional early investment as well. Even in the case of strong intertemporal substitutability when early investment may decline, individuals acquire more adult human capital when the late constraint is relaxed. Altogether, these results are fairly intuitive.

We next show that relaxing borrowing constraints on families during early childhood (e.g. loans for preschool) can lead to more surprising effects on investment behavior depending on the extent of dynamic complementarity.

Proposition 2.11 *Assume that the borrowing constraint binds during early childhood (i.e. $a_2 = -L_1$).*

(i) If no other borrowing constraint binds, then: $\frac{\partial i_1}{\partial L_1} \in (0,1)$*;* $\frac{\partial i_2}{\partial L_1} > 0 \iff f_{12} > 0$*; and* $\frac{\partial h_3}{\partial L_1} > 0$.

(ii) If the late borrowing constraint also binds (i.e. $a_3 = -L_2$) and Condition 2.2 does not hold, then: $\frac{\partial i_1}{\partial L_1} > 0$ *and* $\frac{\partial i_2}{\partial L_1} < 0$.

When individuals are only constrained during early childhood, relaxing that constraint leads to an increase in early investment, which encourages late investment as long as the marginal productivity of i_2 is increasing in i_1.

When children are constrained in both periods, relaxing the early constraint effectively shifts resources from late to early childhood. If early and late investments are very complementary, they will both tend to move in the same direction. In most cases, investments will increase; however, it is possible that investments could actually *decrease* in both periods. Intuitively, if late investment is very productive, then relaxing the early borrowing constraint can 'starve' that investment. By contrast, if investments are sufficiently substitutable over time, shifting resources from late to early childhood by relaxing the early constraint causes investment to shift from the late to the early period as well.

The stylized facts of Section 2.3 and other evidence in Caucutt and Lochner (2012) are most consistent with binding early and late constraints and sufficient dynamic complementarity (case III of Proposition 2.9). In this case, investments increase with additional family income (Fact 3) and the timing of income matters (Fact 4). Because poor children are more likely to be borrowing constrained, Proposition 2.8 implies that they are likely to have marginal labor market returns that exceed the return to savings as well as the marginal returns for unconstrained children from higher income families (Facts 1 and 2). Finally, Propositions 2.10 and 2.11 suggest that policies designed to expand borrowing opportunities (at either stage of child development) can raise the investment and skill levels of children from constrained (i.e. low-income) families, improving both efficiency and equity.

2.9 Summary and Conclusions

It is well-known that poor children perform much worse academically and on achievement tests than their more economically advantaged counterparts. The most immediate explanation for these differences is that poor parents invest less in their young children. As we document, poor parents have fewer books in the home, read less to their young children, engage in fewer lessons and extracurricular activities, etc. Important differences in investment activities and achievement by family income remain even after controlling for maternal characteristics like education, achievement, and race. In this chapter, we ask the next logical question: why do poor parents invest so much less in their children?

While there are many competing theories for these investment and skill gaps, few studies attempt to sort amongst them.[45] We systematically study four leading investment-based theories/mechanisms thought to drive income-based skill gaps: an intergenerational correlation in ability, a consumption value of investment, information frictions, and credit constraints. In order to help understand which mechanisms drive family investments in children, we consider the extent to which they also explain other important stylized facts related to the marginal returns to investment and the effects of parental income on child investment and skills.

The main lessons from our theoretical analysis are summarized in Table 2.1, which shows considerable differences in the extent to which each mechanism explains important stylized facts about child development. While a positive intergenerational correlation in ability may be partially responsible for the relationship between family income and child investment and achievement, it is not helpful for understanding any of the other important stylized facts. A theory based only on a positive consumption value of investment can explain the positive causal effects of income on investment as well as decreasing marginal labor market returns in family income; however, it predicts over-investment in skills such that the labor market returns to investment should be less than the return to savings. This mechanism offers no explanation for the importance of early income relative to late income.

Uncertainty coupled with risk aversion leads to under-investment in human capital and high

[45] Cunha (2014) is an important recent exception.

Table 2.1: Summary of Results

	Intergen. Ability Correlation	Consum. Value ($v > 0$)	Uncertainty w/Risk Aversion	Poor more subjective uncertainty, risk neutrality, i_1 irreversible: Uncertainty in θ	Poor more subjective uncertainty, risk neutrality, i_1 irreversible: Uncertainty in z	Poor have biased beliefs: Systematic downward bias	Poor have biased beliefs: Non-Systematic bias	Credit Constraints
High MR to i_1 for Poor	No	No	Yes	No^{a}	Yes^{ab}	Yes^{b}	No^{b}	Yes
Lower MR to i_1 for Rich	No	Yes	Yes	No^{a}	Yes^{ab}	Yes^{b}	No^{b}	Yes
Increase in Income causes an increase in i_1	No	Yes	Yes	No	Only if info. changes with income^{b}	Only if info. changes with income^{b}	No	Yes
Timing of Income	No	No	No	No	No	No	No	Yes

Notes: Where relevant, results in the table assume gross complementarity, i.e. $f_{12} \geq 0$. ^{a} Under purely objective uncertainty, expected marginal returns equal return to savings for everyone. ^{b} Assumes that Condition 1 holds or that $b \geq 0$ in the case of CES production function.

marginal returns to additional investment. With decreasing absolute risk aversion, investment disincentives are greater for the poor. Thus, this mechanism can explain the qualitative patterns for marginal returns documented in the literature as well as the evidence on causal effects of income on early child investments and achievement. Neither this mechanism nor any other information-based explanation we explore can explain why the timing of income is important. Even in the absence of risk aversion, subjective uncertainty in the productivity of early investments can lead to under-investment and high marginal returns due to the irreversibility of investments. If poor families face greater subjective uncertainty than rich families, then predicted patterns for marginal returns are consistent with empirical evidence. This is also true if poor families simply under-estimate the productivity of early investments compared to higher income families. Unless changes in income directly improve the information of poor families, these mis-information problems only generate a correlation between family income and investment; they cannot explain why changes in income produce changes in investment or achievement.

The inability of poor families to borrow against future income can lead to under-investment in their children, which can further explain high marginal returns to investment among the poor. For children in constrained families, improvements in income lead to increases in investment and higher skill levels. If constraints are binding for families with young children, the timing of income will be important. Thus, binding credit constraints are consistent with the four main stylized facts we consider.[46]

We caution that our comparison of model predictions with the evidence should not be taken as a score sheet, evaluating the importance of each mechanism by the number of facts it explains. A positive intergenerational correlation in ability is almost certainly important given the extent to which maternal characteristics help explain income-based differences in investment and achievement (see Figures 2.1 and 2.5); yet, it offers no explanation for any of the other stylized facts. A number of recent studies also document important biases in beliefs about the productivity of early investments or labor market returns to education; however, the extent to which these biases explain differences in investment and achievement by parental income remains to be seen. Our results suggest that these biases are unlikely to explain why child achievement improves when family income rises or why the timing of income is important.

Our primary contribution is to help clarify key empirical predictions of different mechanisms that can be useful in thinking about policy options or in future empirical research aimed at quantifying the importance of those mechanisms. For example, our analysis suggests that

[46] Borrowing constraints cannot easily explain why many poor mothers hold biased beliefs about the productivity of investments in children (Cunha, 2013, Dizon-Ross, 2015); however, they may explain why schooling choices among the most poor are relatively unresponsive to differences in those beliefs or to new information (Jensen, 2010, Attanasio and Kaufmann, 2009).

information or credit market frictions are needed to explain the high marginal returns to early investment among the poor. This is important, because it means that appropriately designed policies may be able to reduce inequality while improving economic efficiency. We also show that evidence on the relative importance of early vs. late family income for child investment and achievement is particularly useful for identifying the presence of credit market frictions, since none of the other mechanisms predict that the timing of income matters. More generally, a better understanding of the implications and limits of different mechanisms should be helpful in refining current theories and empirically sorting out their quantitative importance. In particular, future empirical research should attempt to exploit data/evidence on the types of relationships we highlight to aid in the identification of general models that incorporate multiple mechanisms.

Bibliography

Abbott, Brant, Giovanni Gallipoli, Costas Meghir, and Giovanni Violante. 2013. "Education Policy and Intergenerational Transfers in Equilibrium." NBER Working Paper No. 18782.

Alessie, Rob, Michael P. Devereux, and Guglielmo Weber. 1997. "Intertemporal Consumption, Durables and Liquidity Constraints: A Cohort Analysis." *European Economic Review* 41 (1):37 – 59.

Almond, Douglas and Janet Currie. 2011. "Human Capital Development Before Age Five." In *Handbook of Labor Economics*, vol. 4B, chap. 15. Amsterdam: Elsevier, 1315–1486.

Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. 2015. "Estimating the Production Function for Human Capital: Results from a Randomized Control Trial in Colombia." NBER Working Paper No. 20965.

Attanasio, Orazio and Katja Kaufmann. 2009. "Educational Choices, Subjective Expectations, and Credit Constraints." NBER Working Paper No. 15087.

Attanasio, Orazio P., Pinelopi Koujianou Goldberg, and Ekaterini Kyriazidou. 2008. "Credit Constraints in the Market for Consumer Durables: Evidence from Micro Data on Car Loans." *International Economic Review* 49 (2):401–436.

Baker, Michael. 2011. "Innis lecture: Universal early childhood interventions: What is the evidence base?" *Canadian Journal of Economics* 44 (4):1069–1105.

Baker, Michael, Jonathan Gruber, and Kevin Milligan. 2008. "Universal Child Care, Maternal Labor Supply, and Family Well-Being." *Journal of Political Economy* 116 (4):709–745.

Barnett, W. Steven and Leonard Masse. 2007. "Comparative Benefit-Cost Analysis of the Abecedarian Program and its Policy Implications." *Economics of Education Review* 26 (1):113–125.

Becker, Gary S. 1975. *Human Capital, 2nd ed.* New York, NY: Columbia University Press.

Becker, Gary S. and Nigel Tomes. 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." *Journal of Political Economy* 87 (6):1153–1189.

———. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics* 4 (3):S1–S39.

Belley, Philippe and Lance Lochner. 2007. "The Changing Role of Family Income and Ability in Determining Educational Achievement." *Journal of Human Capital* 1 (1):37–89.

Blau, David and Janet Currie. 2006. "Pre-School, Day Care, and After-School Care: Who's Minding the Kids?" In *Handbook of the Economics of Education*, vol. 2, edited by E. Hanushek and F. Welch, chap. 20. Amsterdam: Elsevier, 1163–1278.

Brooks-Gunn, Jeanne, Ruth T Gross, Helena C Kraemer, Donna Spiker, and Sam Shapiro. 1992. "Enhancing the Cognitive Outcomes of Low Birth Weight, Premature Infants: For Whom is the Intervention Most Effective?" *Pediatrics* 89 (6):1209–1215.

Cameron, Stephen and James J. Heckman. 1998. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males." *Journal of Political Economy* 106 (2):262–333.

Campbell, Frances A, Elizabeth P Pungello, Margaret Burchinal, Kirsten Kainz, Yi Pan, Barbara H Wasik, Oscar A Barbarin, Joseph J Sparling, and Craig T Ramey. 2012. "Adult outcomes as a function of an early childhood educational program: An Abecedarian project follow-up." *Developmental Psychology* 48 (4):1033.

Carneiro, Pedro and Rita Ginja. 2014. "Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start." *American Economic Journal: Economic Policy* 6 (4):135–173.

Carneiro, Pedro, James Heckman, and Edward Vytlacil. 2011. "Estimating Marginal Returns to Education." *American Economic Review* 101 (6):2754–81.

Carneiro, Pedro and James J. Heckman. 2002. "The Evidence on Credit Constraints in Post-Secondary Schooling." *Economic Journal* 112 (482):705–734.

Caucutt, Elizabeth and Lance Lochner. 2012. "Early and Late Human Capital Investments, Borrowing Constraints, and the Family." NBER Working Paper No. 18493.

Caucutt, Elizabeth and Lance J. Lochner. 2006. "Borrowing Constraints on Families with Young Children." In *Innovation in Education*. Cleveland: Federal Reserve Bank of Cleveland, 39–48.

Costas Meghir, Guglielmo Weber. 1996. "Intertemporal Nonseparability or Borrowing Restrictions? A Disaggregate Analysis using a U.S. Consumption Panel." *Econometrica* 64 (5):1151–1181.

Cunha, Flavio. 2013. "Investments in Children When Markets are Incomplete." Working Paper, University of Pennsylvania.

———. 2014. "Gaps in Early Investments in Children." Working Paper, University of Pennsylvania.

Cunha, Flavio, Irma Elo, and Jennifer Culhane. 2013. "Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation." NBER Working Paper No. 19144.

Cunha, Flavio and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2):31–47.

Cunha, Flavio, James Heckman, and Salvador Navarro. 2005. "Separating Uncertainty from Heterogeneity in Life Cycle Earnings." *Oxford Economic Papers* 57 (2):191–261.

Cunha, Flavio and James J Heckman. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Journal of Human Resources* 43 (4):738–782.

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." In *Handbook of the Economics of Education*, vol. 1, edited by E. Hanushek and F. Welch, chap. 12. Amsterdam: Elsevier, 697–812.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3):883–931.

Currie, Janet and Duncan Thomas. 1995. "Does Head Start Make a Difference?" *American Economic Review* 85 (3):341–364.

———. 1999. “Does Head Start Help Hispanic Children?” *Journal of Public Economics* 74 (2):235–262.

Dahl, Gordon and Lance Lochner. 2012. “The Impact of Family Income On Child Achievement: Evidence from the Earned Income Tax Credit.” *American Economic Review* 102 (5):1927–1956.

Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall. 2014. “Household Choices and Child Development.” *Review of Economic Studies* 81 (1):137–185.

Deming, David. 2009. “Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start.” *American Economic Journal: Applied Economics* 1 (3):111–34.

Dizon-Ross, Rebecca. 2015. “Parents’ Perceptions and Children’s Education: Experimental Evidence from Malawi.” Working Paper.

Dominitz, Jeff and Charles Manski. 1996. “Eliciting Student Expectations of the Returns to Schooling.” *Journal of Human Resources* 31 (1):1–26.

Duncan, Greg and Jeanne Brooks-Gunn. 1997. *Consequences of Growing Up Poor*. New York: Russell Sage Foundation.

Duncan, Greg, Pamela Morris, and Chris Rodrigues. 2011. “Does Money Really Matter? Estimating Impacts of Family Income on Young Children’s Achievement with Data from Random-Assignment Experiments.” *Developmental Psychology* 47 (5):1263–1279.

Duncan, Greg and Aaron Sojourner. 2013. “Can Intensive Early Childhood Intervention Programs Eliminate Income-Based Cognitive and Achievement Gaps?” *Journal of Human Resources* 48 (4):945–968.

Duncan, Greg, Wei-Jun Yeung, Jeanne Brooks-Gunn, and Judith Smith. 1998. “How Much Does Childhood Poverty Affect the Life Chances of Children?” *American Sociological Review* 63 (3):406–423.

Eckenrode, John, Mary Campa, Dennis Luckey, Charles Henderson, Robert Cole, Harriet Kitzman, Elizabeth Anson, Kimberly Sidora-Arcoleo, Jane Powers, and David Olds. 2010. “Long-Term Effects of Prenatal and Infancy Nurse Home Visitation on the Life Course of Youths.” *Archives of Pediatric Adolescent Medecine* 164 (1):9–15.

Garces, Eliana, Janet Currie, and Duncan Thomas. 2002. “Longer-Term Effects of Head Start.” *American Economic Review* 92 (4):999–1012.

Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M Chang, and Sally Grantham-McGregor. 2014. "Labor market returns to an early childhood stimulation intervention in Jamaica." *Science* 344 (6187):998–1001.

Guryan, Jonathan, Erik Hurst, and Melissa Kearney. 2008. "Parental Education and Parental Time with Children." *Journal of Economic Perspectives* 22 (3):23–46.

Hartman, Richard. 1976. "Factor Demand with Output Price Uncertainty." *American Economic Review* 66 (4):675–681.

Havnes, Tarjei and Magne Mogstad. 2014. "Is Universal Child Care Leveling the Playing Field?" *Journal of Public Economics* 127:100–114.

Heckman, James, Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. 2010. "The Rate of Return to the High/Scope Perry Preschool Program." *Journal of Public Economics* 94 (1-2):114–128.

Heckman, James J. and Tim Kautz. 2014. "Fostering and Measuring Skills Interventions that Improve Character and Cognition." In *The Myth of Achievement Tests: the GED and the Role of Character in American Life*, edited by James J. Heckman, John E. Humphries, and Tim Kautz, chap. 9. Chicago: University of Chicago Press, 341–430.

Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics* 125 (2):515–548.

Jones, Lauren, Kevin Milligan, and Mark Stabile. 2015. "Child Cash Benefits and Family Expenditures: Evidence from the National Child Benefit." NBER Working Paper No. 21101.

Karoly, Lynn et al. 1998. *Investing In Our Children: What We Know And Don't Know About The Costs And Benefits Of Early Childhood Interventions*. Santa Monica: The RAND Corporation.

Kaushal, Neeraj, Katherine Magnuson, and Jane Waldfogel. 2011. "How is Family Income Related to Investments in Children's Learning?" In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg Duncan and Richard Murnane, chap. 9. New York: Russell Sage Foundation, 187–206.

Kautz, Tim, James J. Heckman, Ron Diris, Bas ter Weel, and Lex Borghans. 2014. *Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success*. OECD.

Keane, Michael P. and Kenneth I. Wolpin. 1997. “The Career Decisions of Young Men.” *Journal of Political Economy* 105 (3):473–522.

———. 2001. “The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment.” *International Economic Review* 42 (4):1051–1103.

Kitzman, Harriet, David Olds, Robert Cole, Carole Hanks, Elizabeth Anson, Kimberly Arcoleo, Dennis Luckey, Michael Knudtson, Charles Henderson, and John Holmberg. 2010. “Enduring Effects of Prenatal and Infancy Home Visiting by Nurses on Children.” *Archives of Pediatric Adolescent Medecine* 164 (5):412–418.

Kottelenberg, Michael and Steven Lehrer. 2014. “Reinvestigating who Benefits and Who Loses from Universal Childcare in Canada.” Working Paper, Queens University.

Lazear, Edward. 1977. “Education: Consumption or Production?” *Journal of Political Economy* 85 (3):569–598.

Lee, Sang Yoon and Ananth Seshadri. 2014. “On the Integenerational Transmission of Economic Status.” Working Paper.

Levhari, David and Yoram Weiss. 1974. “The Effect of Risk on the Investment in Human Capital.” *American Economic Review* 64 (6):950–963.

Levy, D. and G. Duncan. 1999. “Using Sibling Samples to Assess the Effect of Childhood Family Income on Completed Schooling.” Working Paper.

Lochner, Lance and Enrico Moretti. 2004. “The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports.” *American Economic Review* 94:155–189.

Løken, Katrine. 2010. “Family Income and Children’s Education: Using the Norwegian Oil Boom as a Natural Experiment.” *Labour Economics* 17 (1):118–129.

Løken, Katrine, Magne Mogstad, and Matthew Wiswall. 2012. “What Linear Estimators Miss: The Effects of Family Income on Child Outcomes.” *American Economic Journal: Applied Economics* 4 (2):1–35.

Loury, Glenn C. 1981. “Intergenerational Transfers and the Distribution of Earnings.” *Econometrica* 49 (4):843–867.

Ludwig, Jens and Douglas Miller. 2007. “Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design.” *Quarterly Journal of Economics* 122 (1):159–208.

Milligan, Kevin, Enrico Moretti, and Philip Oreopoulos. 2004. "Does education improve citizenship? Evidence from the United States and the United Kingdom." *Journal of Public Economics* 88 (9-10):1667–95.

Milligan, Kevin and Mark Stabile. 2011. "Do Child Tax Benefits Affect the Well-Being of Children? Evidence from Canadian Child Benefit Expansions." *American Economic Journal: Economic Policy* 3 (3):175–205.

Moretti, Enrico. 2004a. "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics* 121 (1-2):175–212. Higher education (Annals issue).

———. 2004b. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *The American Economic Review* 94 (3):656–90.

Nguyen, Trang. 2008. "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar." Working Paper.

Olds, David, JoAnn Robinson, Ruth OBrien, Dennis Luckey, Lisa Pettitt, Charles Henderson, Rosanna Ng, Karen Sheff, Jon Korfmacher, Susan Hiatt, and Ayelet Talmi. 2002. "Home Visiting by Paraprofessionals and by Nurses: A Randomized, Controlled Trial." *Pediatrics* 110 (3):486–496.

Oreopoulos, Phyilip, Marianne E. Page, and ann Huff Stevens. 2008. "The Intergenerational Effects of Worker Displacement." *Journal of Labor Economics* 26 (3):455–83.

Pavan, Ronni. 2014. "On the Production of Skills and the Birth Order Effect." Working Paper, Royal Holloway, University of London.

Puma, Mike, Stephen Bell, Ronna Cook, Camilla Heid, Pam Broene, Frank Jenkins, Andrew Mashburn, and Jason Downer. 2012. "Third grade follow-up to the Head Start Impact Study: Final Report." Tech. rep., Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, Washington, DC.

Reardon, Sean. 2011. "The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations." In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg Duncan and Richard Murnane, chap. 5. New York: Russell Sage Foundation, 91–116.

Restuccia, Diego and Carlos Urrutia. 2004. "Intergenerational Persistence of Earnings: The Role of Early and College Education." *American Economic Review* 94 (5):1354–1378.

Reynolds, Arthur J, Judy A Temple, Suh-Ruu Ou, Irma A Arteaga, and Barry AB White. 2011a. "School-based early childhood education and age-28 well-being: Effects by timing, dosage, and subgroups." *Science* 333 (6040):360–364.

Reynolds, Arthur J, Judy A Temple, Barry AB White, Suh-Ruu Ou, and Dylan L Robertson. 2011b. "Age 26 cost-benefit analysis of the Child-Parent Center early education program." *Child development* 82 (1):379–404.

Rothschild, Michael and Joseph E. Stiglitz. 1970. "Increasing Risk: I. A Definition." *Journal of Economic Theory* 2 (3):225 – 243.

———. 1971. "Increasing Risk II: Its Economic Consequences." *Journal of Economic Theory* 3 (1):66 – 84.

Stephens, Melvin. 2008. "The Consumption Response to Predictable Changes in Discretionary Income: Evidence from the Repayment of Vehicle Loans." *The Review of Economics and Statistics* 90 (2):241–252.

Stevens, Ann Huff and Jessamyn Schaller. 2011. "Short-Run Effects of Parental Job Loss on Children's Academic Achievement." *Economics of Education Review* 20 (3):289–299.

Stinebrickner, Ralph and Todd Stinebrickner. 2014. "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout." *Review of Economic Studies* 81 (1):426–472.

Todd, Petra and Kenneth I. Wolpin. 2007. "The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps." *Journal of Human Capital* 1 (1):91–136.

Wiswall, Matthew and Basit Zafar. 2015. "Determinants of college major choice: Identification using an information experiment." *Review of Economic Studies* 82 (2):791–824.

Chapter 3

Heterogeneity in Parental Altruism and Optimal Higher Education Subsidies

3.1 Introduction

In the U.S., parents have no legal obligations to pay for their children's higher education. However, for the purposes of financial aid, most undergraduate students are still considered dependent[1] and their parents' financial resources limit the amount of financial aid they qualify for. Because aid is awarded not based on the actual amount of parental transfers but on parents' fiscal ability to contribute, students from high income families can have a hard time financing college costs if their parents do not support them through college. In the presence of borrowing constraints, it would be socially beneficial to provide financial aid to students who receive little support from parents, regardless of family background. But it is challenging to design a financial aid system that specifically targets those students, because parental transfers and preferences are difficult to observe. In this chapter, I provide empirical evidence that parents' contributions towards college costs are partly driven by heterogeneous parental altruism and study optimal higher education subsidies when the policymaker can observe neither the amount of parental transfers nor altruism.

In the first part of the chapter, I derive testable implications of heterogeneous parental altruism and empirically examine them. I begin with presenting a model similar to Becker and Tomes (1986) and Brown, Scholz, and Seshadri (2012) in which parental transfers and college-going decisions are jointly determined. Students choose years of schooling but face limited borrowing opportunities. Parents care about their children's well-being (i.e., they are altruistic)

[1] When filing for the Free Application for Federal Student Aid (FAFSA), a student is considered independent if one of the following criteria is satisfied: at least 24 years old, married, a graduate or professional student, a veteran, a member of the armed forces, an orphan, a person with dependants, or homeless.

and decide how much to pay for college education ('schooling transfer') and how much to give after schooling has been completed ('non-schooling transfer'). Among wealthy parents, those with sufficiently low altruism give low transfers and their children are constrained. On the other hand, sufficiently altruistic parents not only give enough money for college for their children to be unconstrained but also give additional cash transfers after college. Therefore, heterogeneity in parental altruism generates a *positive* correlation between schooling and non-schooling transfers among families with similar resources. The positive correlation between the two types of parental transfers cannot be generated by other factors that affect parental transfers through net return to schooling, such as ability or tastes for schooling: as the monetary or psychic return to schooling rises, parents will substitute away from non-schooling transfer towards schooling transfers.

I empirically investigate these predictions using data from the NLSY97, which collects detailed information on how students pay for college and the parental transfers they receive in each year. The data also contains a rich set of family characteristics. I first document that there is substantial heterogeneity across families in the fraction of net college expenditure[2] paid by parents. In each parental income quartile, there exist many parents who paid nothing and many who paid everything. Moreover, I find that conditional on observable family characteristics and a measure of student ability, schooling transfers and non-schooling transfers are positively correlated, suggesting that some variation in parental transfers is driven by differences in altruism. Because only 50% of the total covariance between the two types of transfers are explained by observable family characteristics, there is a potentially important role for heterogeneous parental altruism.

In the second part of the chapter, I characterize the properties of optimal subsidies with heterogeneous parental altruism. Borrowing constraints distort schooling choices as well as intertemporal and intergenerational consumption allocations. The social planner aims to minimize the distortions but cannot directly eliminate the borrowing contraints. Instead, the social planner can distribute subsidies based on schooling choices and observable characteristics of students, such as ability and family income. Neither the amount of parental transfers nor altruism are observable, but students with identical observable characteristics may choose different years of college due to different parental transfers. Thus, the social planner can target specific types of students by linking subsidies to years of college.

The first analytical result shows that the optimal subsidy policy tends to subsidize early years of college more heavily than late years. More precisely, I show that even students from high income families can receive subsidies in early years of college, but the subsidy should

[2]Net college expenditure is defined as the sum of tuition and living expenses net of non-repayable financial aid such as grants and tuition discounts.

phase out as they progress towards late years if there are some unconstrained students. This means that late years in college should be taxed, not subsidized. The subsidy over early years of college relaxes borrowing constraints for students whose parents have low altruism and give insufficient transfers to finish college. Thus, the subsidy can help constrained students stay in school longer. However, it also benefits unconstrained students whose parents are altruistic and give enough transfers to graduate college. Because subsidies to unconstrained students do not improve economic efficiency but are costly, the optimal subsidy is reduced at high levels of schooling.

The second result shows that the amount of subsidies allocated to early years of college is limited due to distortionary costs related to college attendance at the extensive margin. In particular, I show that the optimal subsidy is increasing in years of schooling at the lowest level (i.e., the subsidy *phases in*) when there are some students who do not attend college. Providing a more generous subsidy at early years of college reduces distortions at the intensive margin (i.e., it improves efficiency for constrained families with college attending children), but it also increases distortions at the extensive margin (i.e., it reduces efficiency for families with children who switch from non-college to college). Because marginal families affected by the policy change are indifferent between going to college and not, the public funding spent on them is purely deadweight loss. Therefore, the total amount of subsidy is not maximized at the lowest level of schooling even when the welfare gain of additional subsidy at the intensive margin is the greatest at that level.

The two results show a fundamental trade-off faced by the social planner: redistributing to students of parents with low altruism generates large welfare gains at the intensive margin, but it can also involve large distortionary costs at the extensive margin. This trade-off is also important in determining how subsidies are allocated across income groups. The presence of borrowing constraints means it can be more efficient to offer larger subsidies to low income families compared to high income families. However, the amount of subsidy given to low income families may be limited by relatively large behavioral responses in college attendance among those with low altruism, which generates substantial deadweight loss.

To quantify these mechanisms, I calibrate the model to match college attainment by family income and ability in the NLSY97 under current U.S. financial aid policy. I then solve for budget-neutral optimal subsidy schedules for high ability (80th percentile) students with different income levels. Compared to current policy, the optimal policy provides larger subsidies for children with low altruism parents by offering greater subsidies to low levels of schooling. This reduces the dispersion of education and improves efficiency in human capital production. The results further suggest that students from high income families should receive larger subsidies (at the current level of aggregate spending) than those from low income families. Although the optimal policy significantly changes the allocation of subsidies as well as schooling outcomes, its welfare gain is modest.

The implications of heterogeneous parental altruism has been recently studied in the optimal taxation literature by Weinzierl (2008) and Farhi and Werning (2013). They study optimal consumption allocation within and across generations under various social welfare functions. In their environment, there are no frictions that justify redistribution across individuals, so the social planner's only goal is to reduce inequality. Therefore, the properties of the optimal allocation inevitably depend on how the social planner values consumption inequality and how much weight is put on each individual. By contrast, in this chapter, borrowing constraints generate distortions that can be reduced by proper redistribution. Therefore, there is room for policy intervention even when the social planner does not care about inequality per se. Based on Bénabou (2002), I develop a social objective where concerns for inequality can be completely separated from those for efficiency, and focus on characterizing policies that minimize distortions.

There is a large literature on education policies in the presence of borrowing constraints.[3] A closely related paper is de Fraja (2002), which characterizes optimal education policies when parents have private information. However, as commonly assumed in the literature, parents' willingness to pay for children's education is heterogeneous due to differences in returns to schooling, instead of altruism. In this chapter, I show that heterogeneity in returns to schooling and parental altruism lead to very different conclusions about optimal education policies. When families choose different schooling levels because of heterogeneous returns to schooling, families with high ability students attain more schooling and are more likely to be constrained. Thus, additional subsidies given for late years of college can relax borrowing constraints and have large welfare gains. By contrast, when parental altruism is heterogeneous, students of altruistic parents obtain more schooling and are less likely to be constrained. Therefore, subsidies given to late years of college attendance have smaller welfare effects.

The rest of the chapter is organized as follows. Section 3.2 builds a model of intergenerational transfers and schooling choices, and derives comparative statics results on how different sources of heterogeneity across families generate different patterns of parental transfers. Section 3.3 reports U.S. evidence on the relationship between parental transfers and schooling, and empirically examines the comparative statics results. Section 3.4 theoretically characterizes optimal policies. Section 3.5 calibrates the model and computes the optimal policy for families with different income. Section 3.6 concludes.

[3] For example, Keane and Wolpin (2001), Caucutt and Kumar (2003), Restuccia and Urrutia (2004), Johnson (2013), Abbott et al. (2013), Hanushek, Leung, and Yilmaz (2014), Findeisen and Sachs (2014); to name a few. See Lochner and Monge-Naranjo (2012) for a recent review and references therein.

3.2 Model

This section develops a model of intergenerational transfers and education choice and derives testable implications. I begin by assuming that parents give one-time transfers at the beginning of their children's post-secondary schooling and then consider the case where the transfers occur again when children finish their schooling.

3.2.1 Environment

A family consists of a parent and a child (youth). At time $t = 0$, the child is age 18 and lives until time $t = T_k$. The parent lives until $t = T_p$. Time is continuous. Each family member $i \in \{p,k\}$ accrues a flow of utility from consumption $c_i(t)$ in each period t and they discount future utility flows at a subjective discount rate ρ. The lifetime utility from a consumption profile $c_i(\cdot)$ is

$$U_i = \int_0^{T_i} e^{-\rho t} u\big(c_i(t)\big)dt,$$

where $u(c) = c^{1-\sigma}/(1-\sigma)$ and $1/\sigma \in \mathbb{R}_+$ is the intertemporal elasticity of substitution.

Parents are altruistic towards their children, so they care about both U_p and U_k. Parents' preferences over (U_p,U_k) are represented by the following twice continuously differentiable and strictly increasing function

$$V = \mathscr{W}(U_p,U_k,\delta), \tag{3.1}$$

where δ is the degree of parental altruism. Parents with high δ are more altruistic, and this is captured by the following single-crossing condition:

$$\frac{\partial}{\partial \delta}\left(\frac{\partial \mathscr{W}(U_p,U_k,\delta)/\partial U_k}{\partial \mathscr{W}(U_p,U_k,\delta)/\partial U_p}\right) > 0.$$

Parents with lifetime wealth $W \in \mathbb{R}_{++}$ make one-time transfer $B \geq 0$ to their children at the beginning (time $t = 0$). Taking parental transfer B as given, youth decide when to leave school. Let $s \geq 0$ denote the time when youth quit schooling, so s also reflects years of college education. While all individuals can borrow and save at interest rate $r = \rho$, youth cannot borrow more than an exogenous limit $\overline{D} \geq 0$ during the schooling period. Schooling involves both direct (out-of-pocket) costs and opportunity costs (forgone earnings). The direct flow cost for the t^{th} year of schooling is $p(t) \geq 0$.

Individuals start working at $t = s$, accumulate experience $x(s,t) = t - s$ while working, and retire at $t = T_R$. Those with schooling s, ability a, and experience x earn $y(s,a,x) \geq 0$ in each period, where $y(s,a,x)$ is twice continuously differentiable and strictly increasing in s and a. Let $Y(s,a)$ be the discounted value of lifetime earnings evaluated at the end of the schooling period

$$Y(s,a) = \int_s^{T_R} e^{-r(t-s)} y\big(s,a,x(s,t)\big)dt.$$

High ability individuals have not only high earnings conditional on schooling but also high returns to schooling. The following condition imposes some complementarity between ability and schooling:

$$\frac{\partial^2 \ln Y(s,a)}{\partial s \partial a} \geq 0.$$

This condition, known as log-supermodularity, ensures that high ability students attain more schooling in the absence of borrowing constraints. The condition holds when ability is multiplicatively separable (i.e., $y(s,a,x) = \mu(a)\iota(s,x)$), which is commonly assumed in empirical studies.

3.2.2 Schooling Choice

I begin with analyzing the schooling choice of a youth endowed with (B,a). Because B is endogenous, we can think of this as a youth's choice conditional on B.

For each period in school $t \leq s$, the youth consumes and pays for schooling costs by accumulating debt (or running down savings when $D(t) < 0$)

$$\dot{D}(t) = rD(t) + c_k(t) + p(t) \tag{3.2}$$

with the initial condition $D(0) = -B$ and the end-of-schooling borrowing constraint

$$D(s) \leq \overline{D}. \tag{BC}$$

Youth can fully smooth consumption within the schooling period and after schooling, but consumption will exhibit a jump at the end of the schooling period if the borrowing constraint (BC) binds. The optimality condition for schooling is [4]

$$\frac{\partial \ln Y(s,a)}{\partial s} \geq r + \frac{p(s)}{Y(s,a)}, \tag{3.3}$$

[4] See Appendix B.1.1 for analytical results and derivations in this section.

where the inequality is strict if and only if (BC) binds.

Additional schooling increases lifetime income, but it also involves opportunity cost due to delayed labor market entry, which is reflected in r, as well as the direct cost $p(s)$ (Mincer, 1958, Card, 2001, Heckman, Lochner, and Todd, 2006). In the absence of binding credit constraints, schooling is chosen to maximize net lifetime income. As such, high ability students attain more schooling, while schooling is independent of the parental transfer (Becker, 1967).

Students with high ability or low parental transfers accumulate more debt by the end of the schooling period. When the borrowing constraint binds, the marginal return to schooling is high because schooling incurs an additional utility cost due to low consumption during school (Becker, 1967, Lochner and Monge-Naranjo, 2011). In order to stay longer in school, youth must suffer from lower consumption during the schooling periods and endure low consumption for a longer period of time. For youth with less initial resources (parental transfer), the cost of this distortion is larger, so they attain less schooling. Given any level of transfer, the consumption distortion is larger for high ability students (income effect), but the monetary benefit is also greater (substitution effect). With borrowing constraints, the relationship between schooling and ability depends on which effect dominates (Lochner and Monge-Naranjo, 2011). I discuss this issue in detail in Section 3.2.4.

3.2.3 Parental Transfer

Parents with lifetime wealth W choose B to maximize (3.1) subject to the constraint that transfers should be nonnegative

$$B \geq 0, \tag{TC}$$

their lifetime budget constraint

$$\int_0^{T_p} e^{-rt} c_p(t)dt + B \leq W,$$

and youth's utility U_k conditional on the transfer.

Parents' utility function (3.1) is often parameterized as (e.g. Becker and Tomes, 1986)

$$V = U_p + \delta U_k, \tag{3.4}$$

with $\delta \in \mathbb{R}_+$. This functional form assumes that the desire to smooth consumption over time is equal to the desire to smooth consumption across generations, so $1/\sigma$ measures both the elasticity of substitution across periods and generations. I use a more general specification that allows attitudes towards intertemporal and intergenerational substitution to be distinguished (Córdoba and Ripoll, 2014). As discussed in Section 3.4.2, this generalization facilitates measuring distortions generated by borrowing constraints.

To separate preferences for intertemporal smoothing, I first transform each family member's utility into its consumption value, and then define parents' preferences over the consumption values of each member. Let $\underline{c}_i(U_i)$ be the utility of family member $i \in \{p,k\}$, measured in constant-equivalent value of consumption flow

$$\underline{c}_i(U_i) = \min_{c(\cdot)} \left\{ \frac{1}{\Lambda(T_i)} \int_0^{T_i} e^{-rt} c(t) dt \middle| \int_0^{T_i} e^{-\rho t} u\big(c(t)\big) dt \geq U_i \right\}, \tag{3.5}$$

where $\Lambda(T) = \int_0^T e^{-rt} dt$. $\underline{c}_i(U_i)$ is the minimum flow cost to achieve the utility level U_i. When consumption is *intertemporally distorted*, $\Lambda(T_i)\underline{c}_i(U_i)$ is strictly less than discounted present value of lifetime consumption and the magnitude of this gap depends on the intertemporal elasticity of substitution.

Parents' preferences over (U_p, U_k) are represented by the following utility function

$$V = \int_0^{T_p} e^{-\rho t} v\big(\underline{c}_p(U_p)\big) dt + \delta \int_0^{T_k} e^{-\rho t} v\big(\underline{c}_k(U_k)\big) dt, \tag{3.6}$$

where $v(c) = c^{1-\eta}/(1-\eta)$. The parameter $1/\eta \in \mathbb{R}_+$ is the 'intergenerational elasticity of substitution', because it governs smoothing between $\underline{c}_p$ and $\underline{c}_k$. When $v(\cdot) = u(\cdot)$ (i.e. $\sigma = \eta$) (3.6) is equal to (3.4).

The optimality condition for B is

$$v'\big(\underline{c}_p(U_p)\big) \geq \delta v'\big(\underline{c}_k(U_k)\big), \tag{3.7}$$

where the inequality is strict if and only if (BC) or (TC) is binding. The gap in the first order condition (3.7) represents *distortions in intergenerational consumption smoothing*, and the magnitude of the distortion depends on the intergenerational elasticity of substitution.

Condition (3.7) shows that, in the absence of binding constraints, parental transfers are determined to equate the marginal utilities across generations. Therefore, the amount of parental transfer varies by the relative wealth between the parent and the child as well as the degree of parental altruism. Rich or altruistic parents give higher transfers, and high ability youth receive lower transfers.

When parental transfers are sufficiently low such that (BC) or (TC) binds, condition (3.7) holds as an inequality. In this case, children are relatively better off than parents (i.e., $v'(\underline{c}_p) > \delta v'(\underline{c}_k)$), so transfers should be reduced further to equalize marginal utilities. Parents cannot do so when (TC) binds. But parents who give positive transfers may not want to further reduce transfers (i.e., only (BC) binds), because it would distort the schooling and consumption allocation of children too much.

3.2.4 Comparative Statics

I now study how family characteristics (W, δ, a) affect schooling and parental transfers. The comparative statics results give testable implications about how different sources of heterogeneity shape the relationship between schooling and parental transfers.

I begin by stating a condition that is important in determining the effects of ability with binding (BC):

Condition 3.1

$$\eta \leq \sigma \leq \frac{\frac{Y_{sa}(s,a)}{Y_a(s,a)} - r}{\frac{Y_s(s,a)}{Y(s,a)-\overline{D}} + \frac{e^{-rT_k}}{\int_s^{T_k} e^{-rt}\,dt}}.$$

The relationship between ability and schooling for constrained families is determined by the balance of two opposing forces. On the one hand, there is a substitution effect: more able students earn a higher income return on schooling, so parents give higher transfers so that students can stay in school longer. On the other hand, there is an income effect: more able students have higher earnings, so parents would like to give lower transfers and increase their own consumption, while students would like to increase consumption during school by going for fewer years. If Condition 3.1 holds, ability is sufficiently complementary with schooling compared to the intertemporal/generational consumption smoothing motive, so the substitution effect dominates the income effect.[5]

The next proposition summarizes how differences across families in (W, δ, a) affect both schooling and total parental transfers. See Appendix B.1 for all proofs.

Proposition 3.1 *Let $\hat{s}(W, \delta, a)$ and $\hat{B}(W, \delta, a)$ be the solution to the family's problem. Then*

1. *$\hat{s}(W, \delta, a)$ and $\hat{B}(W, \delta, a)$ are nondecreasing in W and δ.*

2. *If Condition 3.1 holds, then*

   (a) *$\hat{s}(W, \delta, a)$ is nondecreasing in a;*

   (b) *$\hat{B}(W, \delta, a)$ is (i) nonincreasing in a if (BC) does not bind; (ii) nondecreasing in a if (BC) binds.*

[5] A similar condition is derived by Lochner and Monge-Naranjo (2011).

Richer or more generous (altruistic) parents give larger transfers, which positively affect schooling when (BC) binds. In contrast, the relationship between ability and parental transfers is not monotonic. When (BC) does not bind, parental transfers are decreasing in ability due to the intergenerational consumption smoothing motive. High ability students are more likely to be constrained, because they borrow more for a given amount of parental transfer and they receive lower transfers. When (BC) binds, additional parental transfers reduce distortions in schooling and intertemporal consumption allocations. Furthermore, if Condition 3.1 holds, the marginal benefit of transfers strongly increases with ability. Therefore, parental transfers are first decreasing and then increasing with ability.

The non-monotonic relationship between ability and total parental transfers reflects the fact that parental transfers are motivated by two different concerns: efficient investment in human capital, and consumption smoothing within the family. If parental transfers occur repeatedly, the former is likely to affect the amount of transfers while children are enrolled in college ('schooling transfer') and the latter is likely to affect the amount of transfers when they are not in college ('non-schooling transfer').

To derive the relationship between schooling and non-schooling transfers, I extend the model to allow repeated transfers as in Brown, Scholz, and Seshadri (2012). Parents move first by giving a transfer $B_s \geq 0$ that is tied to s years of college. Parents also choose their own savings. After observing parents' decisions, youth go to college for s years. Out of initial wealth B_s, youth consume, pay tuition $p(t)$ and accumulate debt $D(s)$. After schooling is over, parents observe $D(s)$ and give additional transfer $B_n \geq 0$ (discounted back to period 0). The following corollary characterizes the equilibrium allocation of this game.

Corollary 3.1 *Let $\tilde{s}(W,\delta,a)$, $\tilde{B}_s(W,\delta,a)$, and $\tilde{B}_n(W,\delta,a)$ be the subgame-perfect equilibrium outcome of the game with two-time transfers. Suppose that $\overline{D}=0$, $\eta \leq \sigma$, and that $e^{-rs}Y(s,a) - \int_0^s e^{-rt}p(t)dt$ is strictly concave in s. Then the following hold:*

1. *$\tilde{s}(W,\delta,a)$, $\tilde{B}_s(W,\delta,a)$, and $\tilde{B}_n(W,\delta,a)$ are nondecreasing in W and δ.*

2. *Suppose that Condition 3.1 holds. Then*

   (a) $\tilde{s}(W,\delta,a)$ and $\tilde{B}_s(W,\delta,a)$ are nondecreasing in a;

   (b) $\tilde{B}_n(W,\delta,a)$ is nonincreasing in a.

When youth expect parental transfers after schooling, they have incentives to consume and borrow too much while in college, because parents will give higher transfers if youth have larger debt (Buchanan, 1975, Bruce and Waldman, 1991). If youth cannot borrow, parents can prevent this strategic behavior by giving low schooling transfers so that youth have no assets at the end of the schooling period. In this case, the model with one-time transfers and that with two-time transfers give identical allocation in terms of schooling and total amount of parental transfers. The assumption that net earnings are concave in schooling ensures that only youth whose schooling is not distorted receive additional transfers after college.

This result shows that factors that positively affect schooling are also positively related with the amount of schooling transfers, but they have different effects on the amount of non-schooling transfers. Rich or altruistic parents would like to raise the overall welfare of their children, which can be achieved by increasing consumption during schooling as well as non-schooling periods. Parents with higher ability children give larger schooling transfers due to higher returns to schooling, but it makes them relatively poor compared to their children. Thus, they would give less money when schooling is over. As shown in Appendix B.1.3, this conclusion also holds when heterogeneity in the return to schooling derives from heterogeneity in 'psychic' returns (i.e., heterogeneous tastes) instead of monetary returns.

Corollary 3.1 gives testable implications about the nature of heterogeneity among families with identical wealth.[6] If parents give different amounts of schooling transfers due to differences in altruism, then schooling transfers and non-schooling transfers are positively correlated. On the other hand, if schooling transfers are mainly motivated by differences in the return to schooling, the two types of transfers are negatively correlated. In the next section, I empirically examine these implications.

3.3 Empirical Relationship Between Parental Transfers and Schooling

In this section, I provide empirical evidence on the relationship between parental transfers and schooling. From NLSY97 data, I document that parental transfers are important sources of funding to attend college, but the role of parents differs from family to family. Using the comparative statics results, I investigate whether observed heterogeneity in parental transfers can be explained by differences in parental altruism.

Data on Parental Transfers The NLSY97 is a longitudinal survey of 8,984 Americans from the cohort born between 1980 and 1984. They have been sampled annually since 1997. The survey contains extensive information on each youth's educational outcomes, together with detailed information about family background. A unique feature of the NLSY97 data is that

[6]Unobserved differences in wealth will yield results similar to those assuming unobserved heterogeneity in altruism.

it contains a set of questions about how youth pay for college costs in each semester (amount of family transfer, grant, loan, and own earnings) as well as the amount of cash transfers received from family in each year.[7] From these questions, I construct measures of total schooling transfers and total non-schooling transfers, discounted at a 5% annual interest rate back to age 17.

Suppose that youth j was age 17 in year $t = 1$. Let $b_{j,t}$ be the amount of transfer youth j received in year t and $d_{j,t}$ be a dummy variable indicating whether the youth was enrolled in college or not. Then, total schooling transfer $B_{s,j}$ and non-schooling transfer $B_{n,j}$ for youth j are defined as

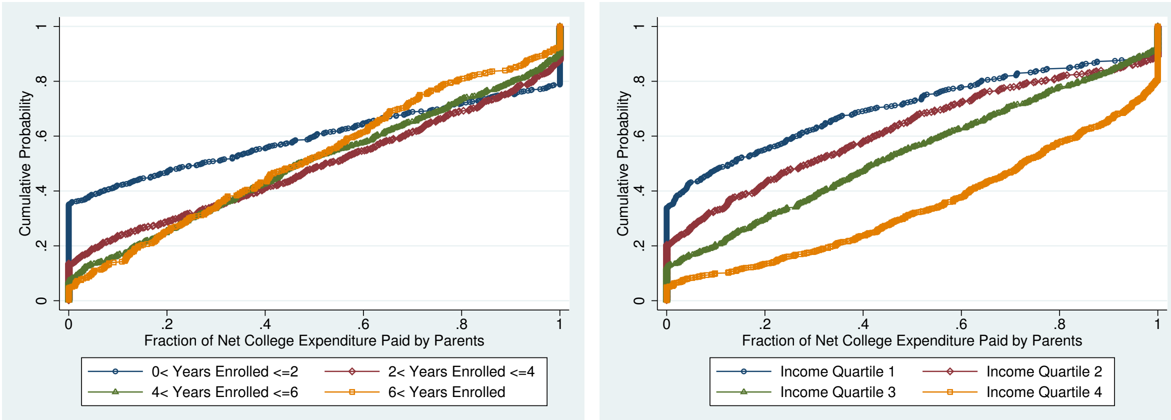
$$B_{s,j} = \sum_{t=1}^{T} \frac{b_{j,t} \mathbb{1}_{\{d_{j,t}=1\}}}{(1+r)^{t-1}}, \quad B_{n,j} = \sum_{t=1}^{T} \frac{b_{j,t} \mathbb{1}_{\{d_{j,t}=0\}}}{(1+r)^{t-1}}.$$

Table 3.1: Present Discounted Value of Parental Transfers and College Costs (2014 U.S. dollars)

Years Enrolled	Total College Cost Paid from Each Source				Total Parental Transfer ($B_s + B_n$)
	Schooling Transfer (B_s)	Grant	Loan	Total Cost	
0	0	·	·	·	1,817
(0,2]	2,390	3,288	3,254	10,333	4,188
(2,4]	21,299	16,857	9,467	52,950	23,170
(4,6]	26,402	21,496	14,237	70,756	28,334
6+	32,229	34,868	20,608	100,561	33,202

Table 3.1 shows the average amount of parental transfers by years enrolled in college. Schooling transfers, as well as total parental transfers, are strongly positively related with years of schooling. On average, about 1/3 of total college expenditure is paid by parents, which is consistent with Sallie Mae and Ipsos (2012). However, the parents' role in paying for college costs is different from family to family. Figure 3.1 shows the cumulative distribution of the fraction of schooling transfer out of net college cost (total cost net of grant). Within each schooling and parental income quartile, the distribution shows large dispersion: there are some parents who gave nothing for college and others who paid for everything. The dispersion is lower among students who attain more schooling, while parents with high income paid a larger fraction of college costs, on average. Large dispersion in parental contributions to college costs across families is also documented in Haider and McGarry (2012), using the data from the Health and Retirement Study.

[7] In this chapter, money received from other family members is counted as part of parental transfers. These variables were used in earlier studies by Johnson (2013) and Abbott et al. (2013).



(a) By Years Enrolled in College

(b) By Parental Income Quartile

Figure 3.1: Distribution of Parental Contribution Towards Net College Expenditure

Source of Heterogeneity Heterogeneity in schooling transfers may be explained by various observable as well as unobservable factors. To systematically investigate the source of this heterogeneity, I utilize the relationship between schooling and non-schooling transfers theoretically predicted by Corollary 3.1. Factors that affect the amount of schooling transfers also affect the amount of non-schooling transfers, and the direction of influence depends on the nature of heterogeneity. Factors that increase total resources available to youth, such as parental wealth and altruism, will have positive effects on both types of transfers, while factors that raise monetary or psychic returns to education will only increase schooling transfers. Therefore, the correlation between the two types of transfers is crucial in identifying the nature of heterogeneity.

Let $\boldsymbol{X}$ be observable family characteristics that determine total parental resources, such as income and assets. Then we can decompose the covariance between schooling and non-schooling transfers into covariances explained by observables and unobservables:

$$\mathrm{Cov}(B_s, B_n) = \mathrm{Cov}\big(\mathbb{E}[B_s|\boldsymbol{X}], \mathbb{E}[B_n|\boldsymbol{X}]\big) + \mathbb{E}\left[\mathrm{Cov}(B_s, B_n|\boldsymbol{X})\right]. \tag{3.8}$$

The first component captures the variation of parental transfers due to differences in parental wealth W; it is likely to be positive if $\boldsymbol{X}$ is a good predictor of W. The second component captures the variation of parental transfers among families with similar wealth, and its sign depends on whether the remaining heterogeneity is mainly driven by differences in return to schooling or parental altruism.

I implement the covariance decomposition (3.8) by regressing B_s and B_n on $\boldsymbol{X}$ and compute the covariance between the residuals of the two equations. For the baseline specification, $\boldsymbol{X}$ includes dummy variables for parental income and asset quartiles, parent age and education, single parenthood, number of siblings, and youth cohort.[8] An alternative specification also includes quartiles of Armed Forces Qualifying Test (AFQT) scores to see whether heterogeneity in returns to schooling has effects that are predicted by Corollary 3.1.[9]

Table 3.2: Correlation Between Schooling and Non-Schooling Transfers

	All			Ever Enrolled in College		
	Raw	Residual		Raw	Residual	
		Baseline[a]	Ability Controls[b]		Baseline	Ability Controls
Correlation Coefficient	0.072	0.038	0.045	0.101	0.063	0.067
	(0.019)	(0.018)	(0.018)	(0.022)	(0.022)	(0.023)
Fraction of Covariance		0.445	0.514		0.528	0.555
Not Explained by Observables		(0.165)	(0.144)		(0.116)	(0.117)

Bootstrapped standard errors are shown in parentheses (1,000 repetitions).

[a] Includes controls for parent income, asset, age, education, single parenthood, number of siblings, youth cohort.

[b] Additionally includes dummy variables for AFQT quartiles.

Table 3.2 shows the estimation results. The correlation between schooling and non-schooling transfers is positive and statistically significant, even after observables are controlled for. Moreover, about 50% of the total covariance between B_s and B_n cannot be explained by observable variables. This suggests that some variation in parental transfers is driven by heterogeneous altruism. Controlling for observable ability makes the correlation stronger, suggesting that heterogeneous returns to schooling generates a negative correlation between schooling transfers and non-schooling transfers, consistent with Corollary 3.1.

Altogether, these results imply considerable unobserved heterogeneity in parental transfers across families and suggest that differences in parental altruism are an important part of the story. However, subsidy policies will in general differ across youth of different abilities and from families with different wealth levels. We devote more attention to these differences in the quantitative analysis of Section 3.5.

[8] Parent education might be correlated with unobserved factors such as ability, tastes for schooling, or altruism. But it can be an important determinant of parental wealth due to, for example, heterogeneity in life-cycle income profile by education. Cohort matters for two reasons. First, parental income and assets are measured in early years of the survey (income 1997-2000, assets 1997), and youth were different ages depending on their birth year. Second, parental transfers are measured until 2010.

[9] The AFQT test scores are widely used as a measure of cognitive ability. Most NLSY97 respondents took the test as part of the survey.

3.4 Theory of Optimal Education Subsidies

In this section, I explore how heterogeneity in parental altruism across families affects the way education should be subsidized based on the model of Section 3.2 (where parents make one-time transfers). To study how college choice at the extensive margin affects the optimal subsidy policy, I introduce a fixed monetary cost $\chi > 0$ for college attendance. Let s_0 and s_1 be arbitrarily small and large schooling levels such that conditional on attending college, nobody chooses schooling outside the interval $[s_0, s_1]$. Let $S = \{0\} \cup [s_0, s_1]$ be the choice set of schooling.

3.4.1 Policy Instruments

The social planner is endowed with budget $E \geq 0$ that can be used to subsidize higher education. Only those who go to college can receive these subsidies, effectively limiting the planner's ability to do lump-sum redistribution. The social planner can condition the amount of subsidies on observable characteristics of students, such as family income and student ability; however, the planner can observe neither parental transfer B nor altruism δ. Thus, the planner specifies subsidy schedules as a function of years of schooling for each group of students with identical observable characteristics (W, δ). In most of this section, I focus on families with identical resources and ability and omit the explicit dependence of the policy on those characteristics.

Let $g(t)$ be the flow subsidy (or 'marginal' subsidy) and $\kappa(t) \geq 0$ be the flow resource cost at schooling t, and denote their present discounted values for those who complete s years of schooling by $G(s)$ and $K(s)$, respectively:

$$G(s) = \int_0^s e^{-rt} g(t) dt, \quad K(s) = \int_0^s e^{-rt} \kappa(t) dt.$$

Define the flow of net payment students face as $p(t) = \kappa(t) - g(t)$. I assume $\kappa(t)$ is continuous and $e^{-rs} Y(s, a) - K(s)$ is strictly concave in s.

Let $\hat{s}(\delta)$ be the schooling choice of a family with parental altruism δ under policy $G(\cdot)$. Parental altruism is distributed over $[\delta_0, \delta_1] \subset \mathbb{R}_+$ with distribution function $F(\delta)$ and density $f(\delta)$ that is strictly positive and continuous. The planner's budget constraint must satisfy

$$\int_{\delta_0}^{\delta_1} G\big(\hat{s}(\delta)\big) dF(\delta) \leq E. \tag{3.9}$$

3.4.2 Social Objectives

The planner sets $G(\cdot)$ in order to minimize the distortions generated by borrowing constraints. In order to separate the concern for inequality from that for efficiency, I assume that the social planner does not care about inequality. Based on Bénabou (2002), I measure distortions by individual's willingness to pay in order to eliminate them (i.e., the amount that makes individuals indifferent between the distorted and undistorted allocations). The degree of aggregate distortions in the economy is equal to the sum of these amounts across individuals.

There are three sources of distortions: (i) schooling; (ii) intertemporal consumption allocations; (iii) intergenerational consumption allocations. I explain the measure of distortions for each of these. Consider a family with characteristic δ who chooses allocation $\left(\hat{s}, \hat{c}_p(\cdot), \hat{c}_k(\cdot)\right)$ under policy $G(\cdot)$, which gives utilities $(\hat{V}, \hat{U}_p, \hat{U}_k)$.

Distortion in Schooling Let s^* be the first-best level of schooling that maximizes total family consumption in the absence of education subsidies ($G(s) = 0$)

$$s^* = \underset{s \in S}{\operatorname{argmax}} \left\{ W + e^{-rs} Y(s,a) - K(s) - \chi \cdot \mathbb{1}_{\{s>0\}} \right\},$$

and denote the family consumption associated with s^* by C^*.

Schooling can be distorted because of borrowing constraints and/or distortionary subsidy policy ($g(t) \neq 0$). Consider families who chose $\hat{s} \neq s^*$ and received subsidy $G(\hat{s})$. If they were given a cash grant equal to the subsidy amount $G(\hat{s})$ and borrowing constraints were removed, they would choose s^* and consume $C^* + G(\hat{s})$, which is is higher than their actual consumption $\sum_{i \in \{p,k\}} \int_0^{T_i} e^{-rt} \hat{c}_i(t) dt$. Therefore, the welfare due to schooling distortion is defined by the gap:

$$\Omega_s = \left(C^* + G(\hat{s})\right) - \sum_{i \in \{p,k\}} \int_0^{T_i} e^{-rt} \hat{c}_i(t) dt. \tag{3.10}$$

Distortion in Intertemporal Consumption Allocation The second source of inefficiency is the distortion in intertemporal consumption allocation due to borrowing constraints, i.e. $u'(\hat{c}_i(t)) \neq u'(\hat{c}_i(t'))$ for some $t \neq t'$. For family member i, welfare loss due to intertemporal consumption distortion can be measured by the gap between the total consumption and the constant-equivalent value of the consumption flow defined by (3.5):

$$\Omega_{c,i} = \int_0^{T_i} e^{-rt} \hat{c}_i(t) dt - \Lambda(T_i) \underline{c}_i(\hat{U}_i). \tag{3.11}$$

Distortion in Intergenerational Consumption Allocation Third, the intergenerational consumption allocation is distorted if (BC) or (TC) is binding, i.e., $v'(\underline{c}_p(\hat{U}_p)) \neq \delta v'(\underline{c}_k(\hat{U}_k))$. In this case, the utility level $\hat{V}$ can be achieved at a lower cost than the actual constant-equivalent consumption $\sum_{i\in\{p,k\}} \Lambda(T_i)\underline{c}_i(\hat{U}_i)$ if parents were allowed to directly choose (c_p, c_k). Let $\underline{C}(V,\delta)$ be the minimum resource cost that delivers a certain level of utility V to parents:

$$\underline{C}(V,\delta) = \min_{c_p,c_k} \left\{ \sum_{i\in\{p,k\}} \Lambda(T_i)c_i \middle| \Lambda(T_p)v(c_p) + \delta\Lambda(T_k)v(c_k) \geq V \right\}.$$

The distortion in intergenerational consumption allocation is measured by the gap between the actual and the minimum cost that delivers the current level of utility $\hat{V}$ to parents:

$$\Omega_g = \sum_{i\in\{p,k\}} \Lambda(T_i)\underline{c}_i(\hat{U}_i) - \underline{C}(\hat{V},\delta). \tag{3.12}$$

Aggregating Distortions The total distortion of the family is the sum of the amount each member of the family is willing to pay to remove all distortions.

$$\Omega = \Omega_s + \sum_{i\in\{p,k\}} \Omega_{c,i} + \Omega_g. \tag{3.13}$$

By using (3.10), (3.11), and (3.12), (3.13) can be written as

$$\underline{C}(\hat{V},\delta) = C^* - \Omega + G(\hat{s}). \tag{3.14}$$

The identity (3.14) gives a clear way to decompose the family's welfare ($\underline{C}$), represented by parents' utility, into its frictionless benchmark (C^*) and the welfare loss generated by frictions (Ω), as well as the amount of transfer received (G). The social planner's budget constraint (3.9) implies that, the planner can minimize aggregate distortions, the sum of Ω, by maximizing aggregate welfare, the sum of $\underline{C}$. Because $\underline{C}(V,\delta)$ is a monotone transformation of parent's utility V, the social planner respects parents' preferences over (U_p, U_k). However, when the social planner cares about inequality across generations, the planer may disagree with parents about how (U_p, U_k) should be allocated (Phelan, 2006, Farhi and Werning, 2010, Ray, 2013).

Welfare Effect of a Lump-sum Transfer The social planner can affect family welfare only through Ω and G, because the potential family consumption C^* is policy invariant. A one dollar increase in transfers G mechanically raises family welfare by one dollar, but total family welfare can be increased by more than a dollar if the additional transfer improves efficiency and reduces the welfare loss Ω. In this case, I say the lump-sum transfer has 'excess welfare gain'.

For a given policy $G(\cdot)$, consider a perturbed policy $\tilde{G}(s) = G(s) + z$ and let $\tilde{V}(z, \delta)$ be parent's utility under the perturbed policy. Then the effect of a small increase in lump-sum transfer on the welfare of families with schooling $s = \hat{s}(\delta)$, denoted by $m(s)$, is

$$m\big(\hat{s}(\delta)\big) = \left.\frac{\partial \underline{C}\big(\tilde{V}(z,\delta),\delta\big)}{\partial z}\right|_{z=0}. \tag{3.15}$$

The following Lemma shows that the excess welfare gain exists only for constrained families.

Lemma 3.1 *Suppose that* (TC) *does not bind and* $\hat{s}(\delta) > 0$ *for families with* δ. *Then*

$$m\big(\hat{s}(\delta)\big) \geq 1,$$

where the inequality is strict if and only if (BC) *binds.*

An increase in z shifts up the subsidy schedule $G(s)$ for all s and increases the amount of subsidies given to students without affecting the marginal cost of schooling or the unconstrained optimal level of schooling. Therefore, for unconstrained families with $\Omega_{c,i} = \Omega_g = 0$, an increase in z is translated one-to-one into an increase in total family consumption and there are no excess welfare gains. On the other hand, a one dollar increase in subsidy improves constrained families' welfare by more than a dollar by relaxing borrowing constraints.

This the key result that determines the direction of redistribution among families with different parental altruism. Redistributing from unconstrained families with altruistic parents to constrained families with low altruism parents improves social welfare. Now, I investigate its implications for the design of education subsidies.

3.4.3 Deriving Optimality Conditions

To understand how optimal policy $G(\cdot)$ and associated marginal subsidy $g(\cdot)$ is determined at the optimum, I derive the optimality condition heuristically, following Diamond (1998), Saez (2001) and Jacquet and Lehmann (2015).[10] The optimality conditions derived in this section will be the basis for characterizing the optimal policy in the next subsection.

Let $\Phi(s)$ be the distribution function of s under a given policy $G(\cdot)$ and let $\phi(s)$ be its density whenever it exists. Because $\hat{s}(\delta)$ is nondecreasing in δ, $\Phi(s)$ is related to the distribution function of altruism $\Phi\big(\hat{s}(\delta)\big) = F(\delta)$. Also let $\underline{s} > s_0$ be the lowest level of schooling chosen by students who attend college.

[10] Formal derivation can be found in Appendix B.1.8

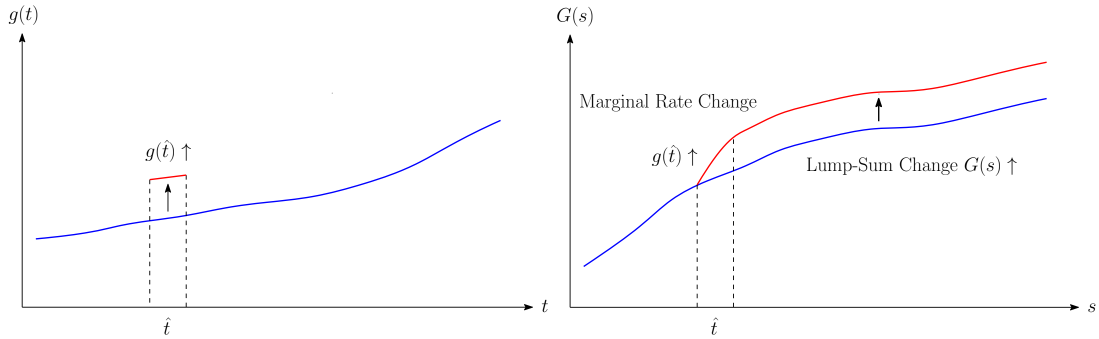


Figure 3.2: Policy Perturbation: Increasing Marginal Subsidy

Raising Marginal Subsidies Consider the effects of a small increase in $g(t)$ by dg around $\hat{t}$ on the social planner's objective function. First, as depicted in Figure 3.2, it reduces the marginal cost of schooling around $\hat{t}$, encouraging those with $\hat{s}(\delta) = \hat{t}$ to stay longer in school by ds via the Hicks substitution effect. While the policy change does not have a first order effect on the welfare of those with schooling around $\hat{t}$, the amount of total subsidies they receive increases by $e^{-r\hat{t}}g(\hat{t})ds$ due to behavioral responses, which reflects the increased cost to the social planner. In this sense, this effect can be seen as the cost of distorting the allocation at $\hat{t}$.

Second, the change dg mechanically affects the present value of subsidies by $dG = e^{-r\hat{t}}dg$ for those who attain more schooling than $\hat{t}$, i.e. $\hat{s}(\delta) = t > \hat{t}$. The lump-sum increase in the total amount of subsidy increases welfare by $d\underline{C} = m \cdot dG$ (m is defined in (3.15)), but it also leads to higher schooling ds via an income effect, which further increases the total amount of subsidies by $e^{-rt}g(t)ds$.

Around the optimal policy, a small change dg cannot improve the planner's objective function, so the total effects of the policy change must be zero at any schooling levels s, which gives the optimality condition

$$\underbrace{g(s)\phi(s)\frac{ds}{dg}}_{\text{substitution effect}} + \int_{(s,s_1]} \Big\{ \underbrace{1}_{\text{mechanical}} - \underbrace{\frac{m(s')}{\lambda}}_{\text{welfare gain}} + \underbrace{e^{-rs'}g(s')\frac{ds}{dG}}_{\text{income effect}} \Big\} d\Phi(s') = 0, \tag{3.16}$$

where λ is the Lagrangian multiplier on the social planner's budget constraint (3.9). This condition (3.16) reveals the key trade-off the social planner faces: increasing flow subsidy $g(t)$ at t incurs distortionary cost for those who choose $s = t$, but it can generate net welfare gains for those who attain more schooling than t. Therefore, the social planner will subsidize schooling only when the net welfare gains are large enough to compensate distortionary costs.

Raising Subsidy Levels Next, consider a small increase in $G(s)$ by dG for all levels of schooling $[s_0, s_1]$, as shown in Figure 3.3. Such a policy change affects everyone who attends college ($\hat{s}(\delta) \geq \underline{s}$) through a lump-sum increase in the level of subsidies. Moreover, it makes attending college cheaper and induces more students to go to college via a substitution effect, increasing the number of college students by $d(1-\Phi(0))$. For every student who switches from non-college to college, it costs $G(\underline{s})$ to the social planner but there are no first order effects on welfare. Therefore, the total additional cost due to behavioral response along the extensive margin, $G(\underline{s}) \times d(1-\Phi(0))$, can be seen as the distortionary cost of increasing subsidy levels. Around the optimal policy, the sum of these effects is zero, giving the following optimality condition:

$$\underbrace{G(\underline{s})\frac{d\big(1-\Phi(0)\big)}{dG}}_{\text{substitution effect}} + \int_{[\underline{s},s_1]} \Bigg\{ \underbrace{1}_{\text{mechanical}} - \underbrace{\frac{m(s')}{\lambda}}_{\text{welfare gain}} + \underbrace{e^{-rs'} g(s') \frac{ds}{dG}}_{\text{income effect}} \Bigg\} d\Phi(s') = 0. \qquad (3.17)$$

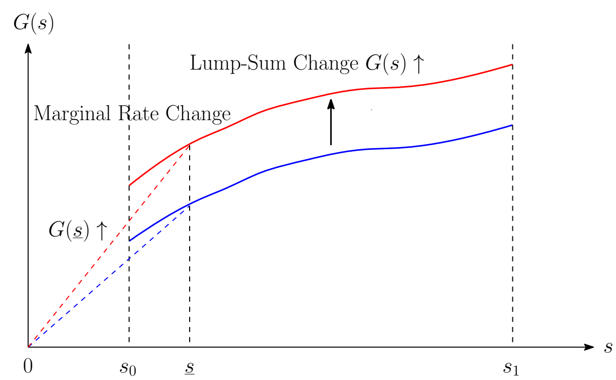


Figure 3.3: Policy Perturbation: Increasing Subsidy Levels

3.4.4 Properties of the Optimal Policy

In this section, I characterize the optimal policy using the optimality conditions (3.16) and (3.17).

Phasing out Towards the Top

The first result shows that there is a tendency to assign higher subsidies, $G(s)$, to lower years of schooling. In particular, the amount of total subsidies is decreasing toward the highest level of schooling that is chosen by unconstrained families. This result is most relevant for high income families, since low income families may not have enough resources to be unconstrained.

Proposition 3.2 *Suppose that under the optimal policy $G(\cdot)$, (i) $\Phi(0) = 0$, (ii) (TC) does not bind for all $\delta \in [\delta_0, \delta_1]$, (iii) (BC) does not bind for δ_1 and it binds for strictly positive measure of families. Then (i) $\lambda > 1$, (ii) there exists $\delta_\lambda = \inf\{\delta \in [\delta_0, \delta_1] \,|\, m(\hat{s}(\delta')) < \lambda, \forall \delta' \in [\delta, \delta_1]\}$, and (iii) there exists $\overline{\delta}$ and $\overline{s}$ such that $\hat{s}(\delta) = \overline{s}$ for almost all $\delta \in [\overline{\delta}, \delta_1]$, (BC) does not bind for all $\delta \in [\overline{\delta}, \delta_1]$, (BC) would bind for all $\delta \in [\delta_0, \overline{\delta})$ if they chose $\overline{s}$. Moreover, if $\hat{s}(\delta_\lambda) < \lim_{\delta \uparrow \overline{\delta}} \hat{s}(\delta)$, then (iv) $\lim_{\delta \uparrow \overline{\delta}} \hat{s}(\delta) < \overline{s} = s^*$, (v) $\lim_{\delta \uparrow \overline{\delta}} G(\hat{s}(\delta)) > G(\overline{s})$, and (vi) $G(\hat{s}(\delta))$ is nonincreasing in δ for all $\delta \in [\delta_\lambda, \delta_1]$ where $\hat{s}(\delta)$ is continuous.*

Proposition 3.2 shows that optimal subsidies have to phase-out towards high schooling levels when there are unconstrained students. When parents are rich enough, sufficiently altruistic parents give enough transfers so that their children are unconstrained. Among students with identical abilities (implicitly assumed thus far), unconstrained students choose the same schooling level ($\hat{s}(\delta) = \hat{s}(\overline{\delta})$ for all $\delta \geq \overline{\delta}$), and constrained students whose parents have low altruism choose lower schooling ($\hat{s}(\delta) < \hat{s}(\overline{\delta})$ for all $\delta < \overline{\delta}$).

The fact that the unconstrained schooling level is the highest means that there is no need to distort schooling incentives ($\hat{s}(\overline{\delta}) = s^*$) at that level. However, as shown in Lemma 3.1, because subsidies given to unconstrained families cannot improve efficiency and thus have lower returns, the level of subsidy decreases toward the unconstrained level of schooling. This can be seen from the optimality condition (3.16). For unconstrained families, there are neither income effects ($ds/dG = 0$) nor excess welfare gains ($m = 1$). So, from (3.16), the net cost of increasing G becomes positive as we approach the unconstrained families $\overline{\delta}$ from below

$$\lim_{\delta \uparrow \overline{\delta}} \int_{(\hat{s}(\delta), s_1]} \left\{1 - \frac{m(s')}{\lambda} + e^{-rs'} g(s') \frac{ds}{dG}\right\} d\Phi(s') = \left(1 - \frac{1}{\lambda}\right)\left(1 - \lim_{\delta \uparrow \overline{\delta}} \Phi\big(\hat{s}(\delta)\big)\right) > 0. \quad (3.18)$$

An additional dollar given to unconstrained families increases their welfare by only one dollar. Because additional subsidies to constrained families provides an excess welfare gain ($\lambda > 1$), the social planner wants to extract revenue from the unconstrained families instead of spending on them. To reduce the total subsidies given to unconstrained families, schooling levels close to the unconstrained amount should not be marginally subsidized even though it discourages schooling for some constrained students. Note that $\Phi(s)$ is discontinuous at s^*, where all unconstrained families are bunched together. This leads to a discontinuity in allocations near s^* (Hellwig, 2010).

Figure 3.4 depicts an example of the optimal subsidy schedule. The amount of total subsidies $G(s)$ may be increasing at low years of schooling, but it declines at high levels of schooling.[11] This property is fundamentally at odds with the current financial aid system in the U.S.,

[11] Note that even though the graph illustrates the case $g(t) > 0$ for some t, it is also possible that $g(t) \leq 0$ for all t.

where additional schooling always leads to nonnegative additional subsidies. This inconsistency can be explained if the current policy is designed to achieve different goals or designed to deal with different types of heterogeneity across families. Next, I study properties of optimal education subsidies under different environments in order to clarify the implication of heterogenous parental altruism on the design of optimal education subsidies.

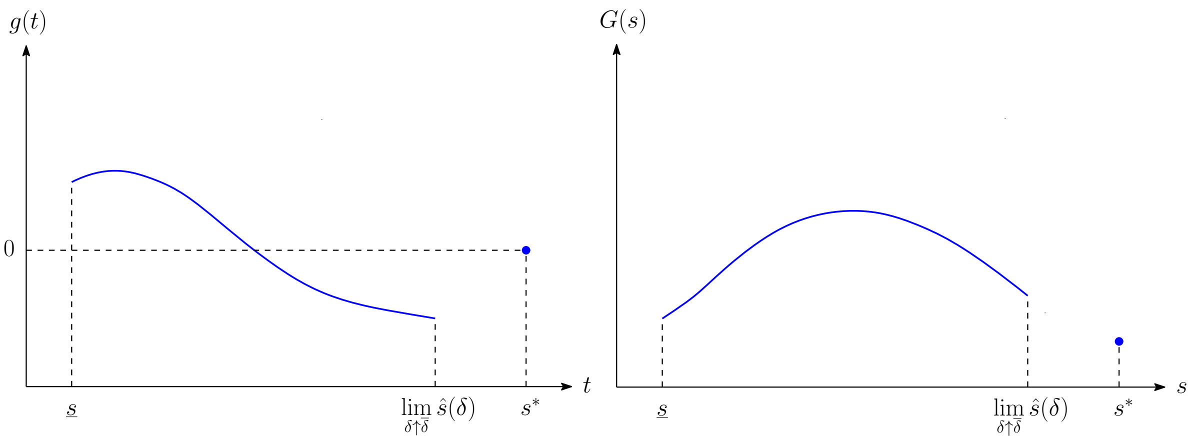


Figure 3.4: Illustration of Optimal Policy

Different Social Objectives Optimal policies are inevitably sensitive to the social planner's objectives. Here, I introduce two elements that are commonly assumed in public finance: inequality aversion and externalities.

I first study the role of inequality aversion. So far, we have assumed that the social planner does not care about inequality across families. That is, when evaluating the welfare gain of a lump-sum transfer to families, the social planner cares about reducing distortions ($m = d\underline{C}/dG$), but not about the welfare level of families ($\underline{C}$). Now assume that the social planner's objective is to maximize the average of $\Gamma(\underline{C})$ instead of $\underline{C}$, where $\Gamma(\cdot)$ is a strictly increasing and strictly concave function. Then the welfare gain of a lump-sum transfer is $\Gamma'(\underline{C}) \cdot m$. In Appendix B.1.4, I show that $\underline{C}$ is nondecreasing in δ. Unconstrained families not only have lower marginal welfare (m), but they also have higher welfare levels ($\underline{C}$) than constrained families. Thus, they have lower $\Gamma'(\underline{C}) \cdot m$ and redistribution towards them is undesirable. Therefore, this type of inequality aversion *strengthens* the result of Proposition 3.2.

Next, I consider the role of externalities. Suppose that the social planner's objective does not coincide with parents' objectives due to various factors such as human capital spillover effect (de Fraja, 2002), social concerns for inequality across generations (Farhi and Werning, 2010), or distortionary taxation (Bovenberg and Jacobs, 2005). In the presence of positive externalities, the private return to education is lower than the social return, so the social planner marginally subsidizes education to correct this private behavior even when there are no borrowing constraints. Let $\tilde{C}$ be the externality that positively depends on schooling, and suppose that the social planner maximizes the average of $\underline{C}+\tilde{C}$ instead of $\underline{C}$. Then the formula for the optimal subsidy becomes

$$\left(g(s)-\frac{1}{\lambda}\frac{d\tilde{C}}{ds}\right)\phi(s)\frac{ds}{dg}+\int_{(s,s_1]}\left\{1-\frac{m(s')}{\lambda}+\left(e^{-rs'}g(s')-\frac{1}{\lambda}\frac{d\tilde{C}}{ds}\right)\frac{ds}{dG}\right\}d\Phi(s')=0. \tag{3.19}$$

The 'Pigouvian' correction term $d\tilde{C}/ds$ captures the social benefit of schooling that is not internalized by parents, and it raises marginal subsidies overall. However, redistributing towards unconstrained families is still undesirable, since $ds/dG=0$ and $m=1$ hold for them. Therefore, the phase-out effect remains as long as the externality is not too large.

Different Sources of Heterogeneity I have assumed that other characteristics of family, i.e. family wealth W and return to schooling a, are observable to the social planner, and thus can be conditioned on when designing the optimal policy. This can be justified by the fact that they are easier to observe (compared to parental altruism) and that government and institutional financial aid currently condition on need and/or merit. However, some of these characteristics might not be observed perfectly, which will affect the nature of the social planner's problem.

First, it is straightforward to see that unobserved heterogeneity in family resources W does not change the conclusions of Lemma 3.1 and Proposition 3.2. Rich families are unconstrained and attain more schooling, while poor families are constrained. Because additional subsidies given to unconstrained families cannot improve efficiency, the optimal subsidy has to phase-out. Thus, Proposition 3.2 applies when there is unobserved heterogeneity in parents' willingness to pay in general, regardless of whether it stems from variation in the actual amount of resources or not.

Second, when families differ only in children's return to schooling, the optimal subsidy is increasing in years of schooling, redistributing to those who attain high levels of schooling. To state this result formally, consider families with identical parental wealth and altruism, but with heterogeneous ability distributed over $[a_0,a_1]$. Let $\hat{s}(a)$ be the schooling choice of a family with student ability a under a given policy $G(\cdot)$.

Proposition 3.3 *Suppose that under the optimal policy $G(\cdot)$, (i) $\Phi(0)=0$, (ii) (TC) is not binding, (iii) (BC) is binding for strictly positive measure of families, and (iv) Condition 3.1 holds. Then $G(\hat{s}(a))$ is nondecreasing in a for all $a\in[a_0,a_1]$ where $\hat{s}(a)$ is continuous.*

This result is similar to that of de Fraja (2002). The crucial difference from the case of heterogenous altruism is that, when families choose different schooling levels because of heterogeneous returns to schooling, families with high ability students attain more schooling and are more likely to be constrained. Thus, additional subsidies given for late years of college attendance can relax borrowing constraints and have large welfare gains. When ability is unobservable, the marginal subsidy should be positive over all levels of schooling in order to give higher subsidies to high ability students. More generally, when there is unobserved heterogeneity in both parental altruism and students' ability, these two types of private information will create conflicting forces for how subsidies should vary with years of schooling.

Phasing in at the Bottom

In the previous section, I showed that unobservable heterogeneity in altruism creates a force to redistribute towards those who attain less schooling, thus assigning higher subsidies to early years of college. In this section, I show that increasing subsidies at early years of college can involve a significant distortionary cost, limiting the degree to which the social planner wants to redistribute towards students with less altruistic families. In particular, I show that the amount of subsidies is increasing at the lowest level of schooling when there are behavioral responses at the college attendance margin. In other words, the subsidy phases in at the bottom of schooling distribution even when the welfare gain of additional subsidy at the intensive margin is the largest at the lowest schooling levels.

Proposition 3.4 *Suppose that (i) $\Phi(0) \in (0,1)$, (ii) $\Phi(\underline{s}) = \lim_{s \uparrow \underline{s}} \Phi(s)$, and (iii) $G(\underline{s}) > 0$. Then $g(\underline{s}) > 0$.*

The result can be derived directly from the optimality conditions (3.16) and (3.17). At the lowest schooling level $\underline{s}$, the marginal subsidy is determined by the optimality condition (3.16):

$$g(\underline{s})\phi(\underline{s})\frac{ds}{dg} = -\int_{(\underline{s},s_1]} \left\{1 - \frac{m(s)}{\lambda} + e^{-rs} g(s) \frac{ds}{dG}\right\} d\Phi(s). \tag{3.20}$$

The right-hand side is equal to the marginal benefit of a lump-sum increase of the subsidy schedule over the whole range $[s_0, s_1]$ when there is no mass point at $\underline{s}$. This is equated with the additional cost due to increased college attendance according to condition (3.17). Then condition (3.20) becomes

$$g(\underline{s})\phi(\underline{s})\frac{ds}{dg} = G(\underline{s})\frac{d\big(1-\Phi(\underline{s})\big)}{dG}. \tag{3.21}$$

When $\Phi(0) > 0$ and $G(\underline{s}) > 0$, we have $g(\underline{s}) > 0$, so additional schooling at $\underline{s}$ leads to higher total subsidies, as shown in Figure 3.4. Note that this result holds even when additional subsidies assigned to lower schooling levels have larger welfare gains at the intensive margin (i.e., $m(s)$ is decreasing at $\underline{s}$). This is due to the distortionary cost $G(\underline{s})d(1-\Phi(\underline{s}))/dG$ at the extensive margin.

To see why it is costly to increase subsidies at low levels of schooling, recall that the family's welfare can be written as $\underline{C} = C^* + G - \Omega$ and consider a small increase in $G(\underline{s})$ that makes a marginal family (i.e., who is indifferent between $s = 0$ and $s = \underline{s}$) to switch from non-college to college. Let $\Delta\underline{C}$, ΔG, and $\Delta\Omega$ be the resulting changes in $\underline{C}$, G, and Ω. Because the marginal family is indifferent between the two options, the welfare gain of switching is zero:

$$\Delta\underline{C} = \Delta G - \Delta\Omega = 0,$$

which implies $\Delta G = G(\underline{s}) - 0 = \Delta\Omega$. That is, an additional subsidy given to a marginal student is translated into a one-to-one increase in distortion. Encouraging college attendance reduces distortions in schooling, but it worsens consumption smoothing. Moreover, the benefit of reduced schooling distortion may be small at low levels of schooling due to the fixed cost.

This result crucially depends on the non-convexity that generates behavioral responses at the extensive margin. The distortionary cost due to the extensive margin of college attendance not only limits the amount of subsidies given to the lowest schooling level within an income group, but it also affects the amount of subsidies given to everyone in that income group. This result is explained in the next subsection.

Allocation of Subsidies Across Income Groups

In this subsection, I explore implications for how public funding should be allocated across income groups. Let $h = 1, \ldots, H$ be the group identity for families with wealth W_h and suppose that each group has mass π_h. The social planner sets policy for each income group separately, which is denoted by $G_h(s)$. And let $\Phi_h(s)$ be the distribution function of families with income W_h over schooling under a given policy. Then, by summing up the budget constraint for each income group (3.9), the economy-wide budget constraint can be derived as follows:

$$\sum_{h=1}^{H} \pi_h \int_{[s_0,s_1]} G_h(s)d\Phi_h(s) \leq E. \tag{3.22}$$

Because family wealth W_h is observable, within each income group, the social planner's problem can be solved as before. In particular, the optimality condition for subsidy level (3.17) holds for each income group. Rewriting the formula, the following condition holds

$$\lambda = \frac{\int_{[\underline{s}_h, s_1]} m_h(s) \frac{d\Phi_h(s)}{1-\Phi_h(\underline{s})}}{\frac{d\ln(1-\Phi_h(\underline{s}))}{d\ln G_h(\underline{s})} + \int_{[\underline{s}_h, s_1]} \left\{1 + e^{-rs} g_h(s) \frac{ds_h}{dG}\right\} \frac{d\Phi_h(s)}{1-\Phi_h(\underline{s})}}, \quad \forall h = 1, \ldots, H. \tag{3.23}$$

This condition states that the social planner allocates resources to each income group such that the average marginal benefit of public funding is equalized across groups. First, for the same level of spending on all income groups, the allocation of low income families would be more distorted. They would, therefore, have a large welfare gain m_h compared to higher income families. It is optimal to increase subsidies to low income families until the average marginal gain is equalized. Second, additional spending on low income families can also entail a large distortionary cost if the college attendance rate is elastic to subsidies. This lowers the average marginal return to spending on low income families. Altogether, the optimal budget allocation across income groups depends on which of these effects dominates. In the next section, I examine this question quantitatively.

3.5 Quantitative Analysis

I now explore the quantitative implications of heterogeneous altruism for the design of education subsidies. I first calibrate the distribution of parental altruism, costs and returns to college education by student ability and family income to match both the distribution of schooling and the relationship between schooling, ability, and family income in the U.S. I then characterize the optimal policy for different income groups.

Thus far, I have assumed a fixed cost of schooling χ. To better fit the distribution of schooling, I extend this assumption by introducing a more general form of 'psychic' cost of schooling:

$$\Psi(s) = \chi \cdot \mathbb{1}_{\{s>0\}} + \gamma_1 s + \gamma_2 s^2.$$

This cost represents various forms of schooling costs that are not captured by tuition costs or foregone earnings, which could be psychic or monetary. The psychic cost is commonly assumed in models of schooling choices (e.g. Heckman, Lochner, and Taber, 1998, Bils and Klenow, 2000), and is also consistent with empirical studies (Cunha, Heckman, and Navarro, 2005, Heckman, Lochner, and Todd, 2006).

When education investment incurs utility cost, it generates a counterfactual *negative* relationship between income and investment due to income effects (Caucutt, Lochner, and Park, 2015). Thus, a positive correlation between tastes for schooling and family income is sometimes required to match the data (Abbott et al., 2013). Because of this issue, I assume the psychic cost enters youth's utility function as follows:

$$V_k = \Lambda(T_k)\underline{c}_k(U_k) - \Psi(s).$$

Since the utility function is linear in constant-equivalent consumption, there are no income effects (Greenwood, Hercowitz, and Huffman, 1988) and schooling is independent of initial wealth if (BC) does not bind.

To incorporate the psychic cost, parents care about $V_k/\Lambda(T_k) = \underline{c}_k(U_k) - \Psi(s)/\Lambda(T_k)$, instead of $\underline{c}_k(U_k)$. The definition of the schooling distortion Ω_s is also modified accordingly.

3.5.1 Calibration

I normalize time so that a unit interval represents a calendar year. All monetary amounts are denominated in 2004 U.S. dollars using the consumer price index (CPI-U-RS). As a measure of ability, I use standardized AFQT scores that are comparable between NLSY79 and NLSY97 cohorts constructed by Altonji, Bharadwaj, and Lange (2009).

Externally Calibrated Parameters

At $t = 0$, youth choose $s \in \{0, 1, \dots, 8\}$ years of college. All individuals work until age $J_R = 65$ and die at age $J = 80$. Thus, youth retire at time $T_R = J_R - 18 = 47$ and die at time $T_k = J - 18 = 62$. I assume an annual interest rate $r = 0.05$ based on historical average of the riskless rate and the return to capital in the U.S. I also set $\rho = r$.

Table 3.3: Externally Calibrated Parameters

Parameter	Description	Value
J_R	Retirement age	65
J	Lifetime	80
$\rho = r$	Discount rate	0.05
α_0		8.12
α_1		0.19
α_2	Log earnings $\ln y(s, a, x)$	0.17
α_3		0.2
α_4		-0.005

Earnings Function I use data on wage income, education, age, and AFQT scores from the random sample of NLSY79 (1979-2006) to estimate parameters of the following earnings function:

$$\ln y(s,a,x) = \alpha_0 + \alpha_1 \cdot \ln a + \alpha_2 \cdot s + \alpha_3 \cdot x + \alpha_4 \cdot x^2, \tag{3.24}$$

where s is completed years of post-secondary schooling and x is potential labor market experience. The scale of ability is set such that log ability is standard normally distributed. The earnings function (3.24) is estimated by OLS, assuming it is measured by wage income with idiosyncratic errors. The estimated earnings function, shown in Table 3.3, is used to calculate youth's as well as parent's lifetime earnings.

Table 3.4: Tuition Costs and Borrowing Limit by Years of College ($)

Years of College	Corresponds to:	Annual Amount		
		Average Tuition	Stafford Loan Limit	Average Earnings While Enrolled
1	Public 2 Year	1,843	2,625	4,464
2			3,500	
3 4 5	Public 4 Year	4,664	5,500	3,155
6 7 8	Public Graduate/ Professional	7,960	18,500	10,476

Tuition Costs, Financial Aid, and Borrowing Limits Using data from the National Post-secondary Student Aid Study (academic year 2003-2004), I calculate average annual tuition costs and financial aid amounts of full-time students attending public institutions by schooling level and family income.[12] I associate different years of college with different schooling levels when calculating tuition costs $\kappa(t)$ and financial aid amounts, as shown in Table 3.4. I assume students can borrow up to annual limits of the Stafford Loan Program, presented in Table 3.4, for each year of college. This means that the total amount students can borrow depends on years of schooling. Instead of imposing annual borrowing limits, I only impose the total borrowing limit $\overline{D}(s)$, implied by annual limits up to s years of college. I also assume that students

[12] For undergraduate students, the calculation is based on dependent students who applied for federal aid and attended in-state public institutions.

Table 3.5: Average Annual Financial Aid by Family Income ($)

Family Income	Public 2 Year Non-Repayable Financial Aid	Public 2 Year Subsidized Loan	Public 4 Year Non-Repayable Financial Aid	Public 4 Year Subsidized Loan	Public Graduate Non-Repayable Financial Aid	Public Graduate Subsidized Loan
Less than 20K	4,030	298	6,804	2,330		
20K-30K	3,777	470	6,018	2,434		
30K-40K	2,638	448	4,561	2,759		
40K-50K	1,982	572	3,124	2,612		
50K-60K	1,034	481	2,066	2,258		
60K-70K	1,126	282	1,687	1,817	3,199	3,401
70K-80K	673	307	1,595	1,514		
80K-90K	708	170	1,240	926		
90K-100K	666	118	1,299	1,005		
100K-120K	584	94	1,127	699		
More than 120K	686	226	1,304	207		

Table 3.6: Target Statistics and Model Fit

	Schooling Distribution				Schooling Regression Coefficients						
	Years Enrolled				Family Income Quartile			AFQT Quartile			Constant
	0	(0,2]	(2,5]	5+	2	3	4	2	3	4	
Data	0.347	0.187	0.290	0.177	0.203	0.683	1.172	0.904	1.820	2.834	0.728
Model	0.299	0.198	0.387	0.160	0.192	0.645	1.729	1.076	1.629	2.313	0.644

earn part time income during the schooling period.[13]

Table 3.5 shows the amount of subsidy distributed through financial aid. This is calculated separately for 11 family income groups only for undergraduate students, since graduate students are considered independent and their parental income is irrelevant in determining financial aid amounts. The present value of subsidy amount $G(s)$ as well as flow subsidy $g(t)$ are calculated from the sum of all financial aid excluding loans ('non-repayable aid') and the subsidy value of subsidized loans (Perkins and subsidized Stafford loans). The subsidy value of subsidized loans are interest payments while students are still in school. Let $n(t)$ and $l(t)$ be the flow of non-repayable aid and subsidized loans, respectively. Then, the flow subsidy is

$$g(t) = n(t) + r\int_0^t l(t')dt'.$$

Internally Calibrated Parameters

The remaining parameters—parameters of the psychic cost function $(\chi, \gamma_1, \gamma_2)$, consumption smoothing parameters (σ, η), and the distribution of altruism—are calibrated by simulating the model and comparing the resulting allocations with those observed in the data. I assume parental altruism follows a beta distribution that is defined over the interval from 0 to 1. The beta distribution is flexible and can take different shapes depending on the values of the parameters, but it may not have density at 0 and 1, which is problematic when calculating the optimal policy. So I truncate parental altruism to the interval $[\delta_0, \delta_1] = [0.05, 0.95]$.

Simulation Method For each individual in the NLSY97, I draw parental altruism δ from the parameterized distribution and calculate the present discounted value of parental wealth W by adding parents' net worth and present discounted value of income as of 1997. To calculated parents' lifetime income, I first back out the individual-specific intercept of the log earnings equation (3.24) using parental education, age, and parental income, which is averaged over 4 years (1997-2000). Then I calculate the present value of earnings from 1997 to age 65. I numerically solve the family's problem laid out in Section 3.2. Parents give one-time transfers and students choose schooling $s \in \{0, 1, \dots, 8\}$ based on their ability, earnings function (3.24), psychic cost, monetary costs and borrowing limits calibrated in Section 3.5.1. Finally, I calculate the target statistics for the simulated data, take averages over 100 simulations, and compare them with their empirical counterparts. I choose parameters that minimize the weighted sum of squared errors between target statistics (reported in Table 3.6) in the simulated and actual data. In estimating schooling regression using NLSY data, I additionally include dummies for

[13] Teaching and research assistantship for graduate students are included as earnings.

birth year, number of siblings, and non-single parenthood to control for variables not modelled. Note that the model assumes that youth are 18 years old, have no siblings, and have a single parent.

Table 3.7: Internally Calibrated Parameters

Parameter	Description	Value
σ	Intertemporal smoothing	3.18
η	Intergenerational smoothing	0.98
χ	Psychic cost $\Psi(s)$	34,531
γ_1		2,966
γ_2		5,668
$\mathbb{E}[\delta]$	Parental altruism δ	0.39
$\mathrm{Var}[\delta]$		0.09

Calibrated Parameters and Model Fit As Table 3.6 shows, the model can generate the distribution of schooling as well as the relationship between schooling, ability, and family income that are close to the data. The intertemporal elasticity of substitution $1/\sigma = 0.31$ is smaller than 1, which is consistent with Browning, Hansen, and Heckman (1999). The intergenerational elasticity is larger than intertemporal elasticity, consistent with Córdoba and Ripoll (2014). Calibrated density of parental altruism is shown in Figure 3.5.

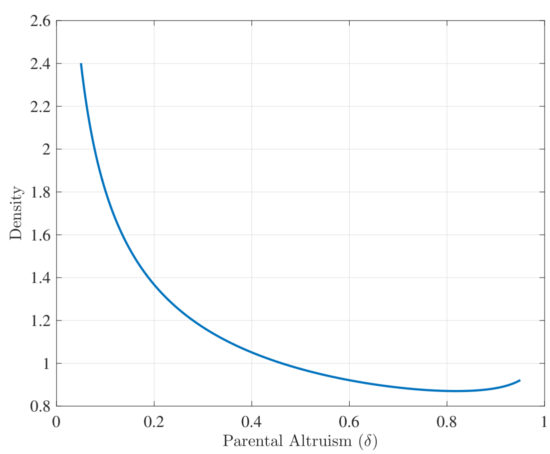


Figure 3.5: Calibrated Density of Parental Altruism

3.5.2 Optimal Policies

I now use the calibrated model to compute optimal subsidies for low (\$20K-\$30K), middle (\$50K-\$60K), and high (\$100K-\$120K) income families. Parents within each income group are assumed to have identical lifetime wealth, which is computed by averaging lifetime wealth of the parents used in calibration. Youth ability is also fixed at the 80th percentile, so families that belong to the same income group are only different in the degree of parental altruism. Conditional on attending college, students face a continuous choice set $[s_0, s_1] = [1, 8]$.

I compute two sets of budget-neutral optimal policies. I first characterize the optimal policy for middle income families that keeps constant the total amount of subsidies given to them as a group. I then solve for the optimal policy when the budget for low, middle, and high income groups is pooled together.

Optimal Policy for Middle Income Families

Figure 3.6 plots the current and the optimal subsidy. Subsidy schedules are shown at schooling levels chosen by any students under each policy. The optimal policy reallocates the budget from late years of college to early years. The total amount of subsidy for those who complete 2 years of college is increased by around $1,500 and, as a result, college attendance increases. On the other hand, a lower marginal subsidy beyond 3 years of college induces students with altruistic parents to attain less schooling. Figure 3.7(a) shows schooling allocations by parental altruism. Note that schooling exhibits jumps due to the fixed psychic cost, discontinuities in the flow cost of schooling, and borrowing limits. Even though there are some unconstrained students under the optimal policy, the amount of subsidy is monotonically increasing in years of schooling (i.e., Proposition 3.2 does not hold), because not everyone goes to college.

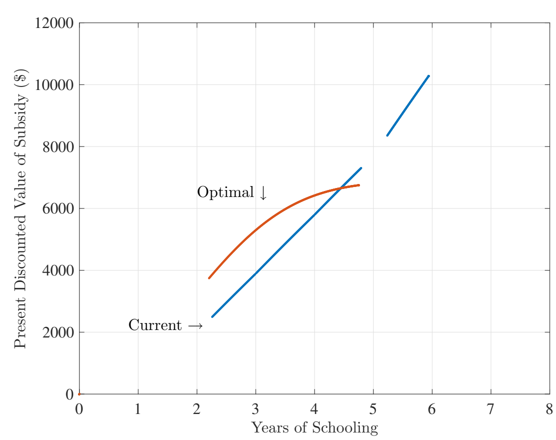
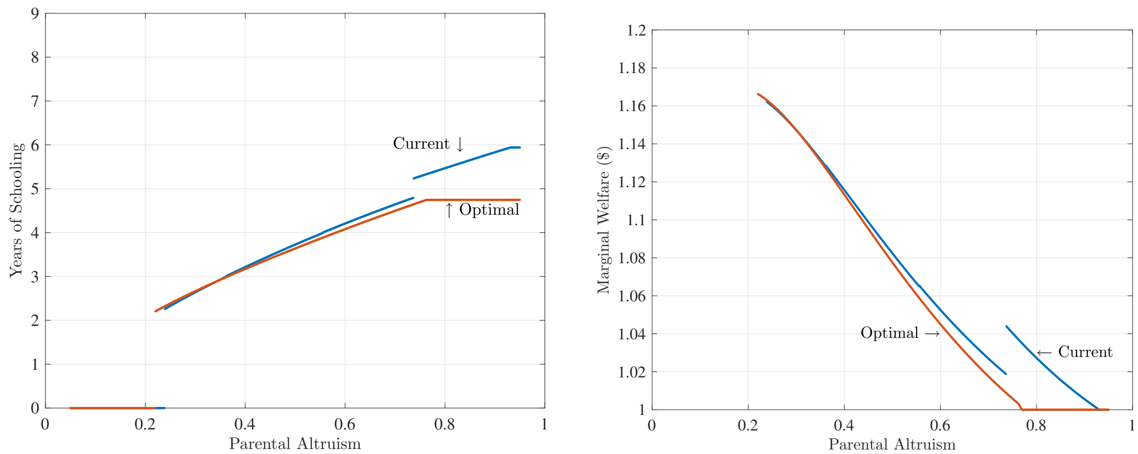


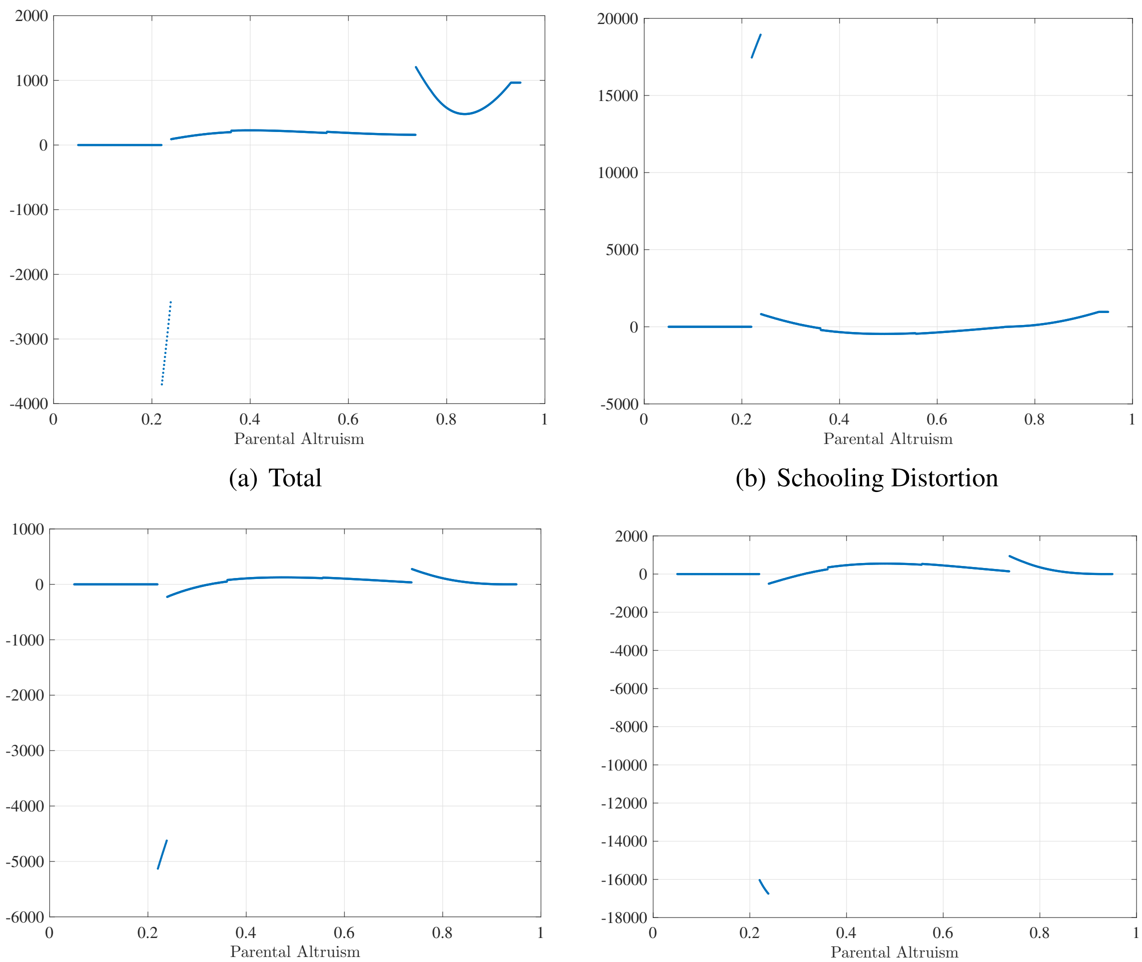
Figure 3.6: Current and Optimal Policy for Middle Income Families

Table 3.8 summarizes average effects of the optimal policy on schooling, welfare, and distortions. The optimal policy improves efficiency in schooling by reducing inequality in educational attainment, but increased college attendance for youth from low altruism families worsens intertemporal and intergenerational consumption allocations. Although the optimal



(a) Schooling (b) Marginal Welfare

Figure 3.7: Schooling and Marginal Welfare for Middle Income Families



(a) Total (b) Schooling Distortion

(c) Intertemporal Distortion (d) Intergenerational Distortion

Figure 3.8: Reduced Distortion for Middle Income Families ($)

policy leads to changes in the subsidy schedule and allocation of schooling, its total welfare effect is negligible. One reason for the small welfare effect is the absence of inequality aversion. Reallocating subsidies to families with less altruistic parents reduces inequality in the level of welfare across families and youth, but this is not valued very much by the social planner.

Another reason for the small welfare effect is that the social planner's ability to redistribute across families using education subsidies is limited. Although the optimal policy is more redistributive to families with low altruism compared to the current policy, the amount of subsidy is still increasing in years of schooling. Consequently, unconstrained families with the most altruistic parents receive the highest amount of subsidy, even though families with lower altruism have higher marginal welfare (see Figure 3.7(b)). This reflects the large distortionary cost due to behavioral responses at the attendance margin, which limits subsidy levels at early years of college. The distortionary cost lowers the average marginal benefit of subsidies, so that one dollar increase in budget only increases social welfare by less than a dollar (i.e., $\lambda < 1$). Therefore, giving subsidies to unconstrained families is not a waste in this case.

Figure 3.8 shows efficiency gains from the optimal policy by parental altruism. For all families with students who attend college under the current policy (δ greater than 0.25), the optimal policy improves efficiency. The efficiency gain is largest among families who over-invest under the current policy (δ greater than 0.8). By reducing their schooling below the first-best level, the optimal policy improves efficiency in schooling as well as consumption allocations. Among families with middle levels of altruism, the efficiency gain in consumption allocation is large enough to compensate for the efficiency loss from reduced schooling. The optimal policy increases schooling for families with low altruism, but it worsens consumption smoothing across time and generations, which results in the smallest efficiency gain among families with college students. For families who switch from non-college to college, the efficiency gain is very negative: attending college improves efficiency in schooling substantially, but it severely distorts consumption allocation. Although relatively few families switch from non-college to college, the magnitude of the efficiency loss is several times larger than the efficiency gain obtained from families who attend college under the current policy.

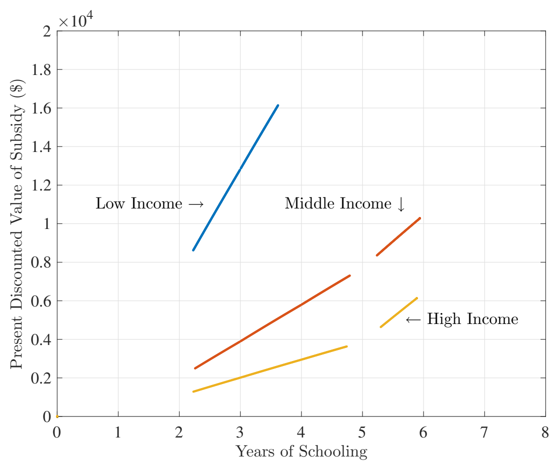
Optimal Policy for Low, Middle, and High Income Families

Now, suppose that the social planner sets the subsidy schedules for all three income groups using pooled budget (3.22). The planner may give different amounts of subsidies to different groups, redistributing resources. Since the planner does not value equity, the objective is only to increase aggregate efficiency.

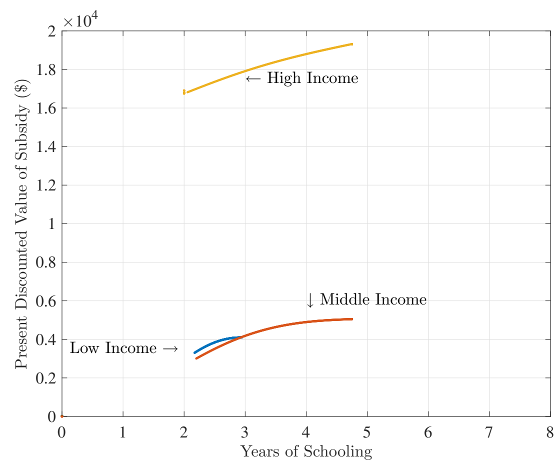
Figure 3.9 plots the current and the optimal policy for each income group. The optimal policy redistributes from low and middle income to high income families. As we can see from

Figure 3.10, among families whose children attend college, low income families have significantly higher marginal welfare than high income families. However, a higher fraction of low income families does not attend college, which raises the total cost of increasing subsidies. Although some high income students do not attend college under the current policy, a sufficiently high subsidy under the optimal policy allows all of them to attend. Even high income students who do not receive any parental transfers finish 2 years of college.

Table 3.9 shows average effects of the optimal policy. Average welfare gain is about \$550, with most of this gain coming from taking money away from low income families. The optimal policy takes \$6,020 and \$1,050 away from low and middle income families, and gives it to high income families. Interestingly, this *reduces* distortions for low income and middle income families by \$1,270 and \$170, while it *increases* distortions for high income families by \$210. Lower schooling among low income families significantly raises distortions in schooling, but it is offset by improved consumption smoothing. For high income families, the opposite is true: improved efficiency in schooling is not large enough to compensate for worsened consumption smoothing.



(a) Current

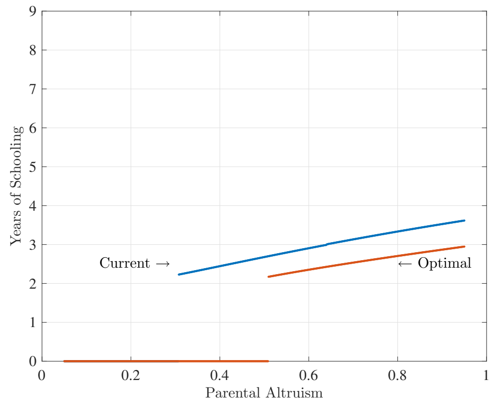


(b) Optimal

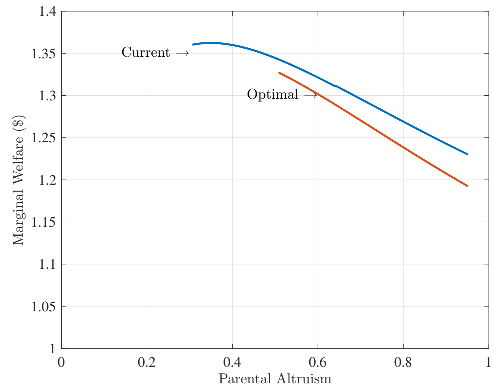
Figure 3.9: Current and Optimal Policy for Low, Middle, and High Income Families

3.6 Conclusion

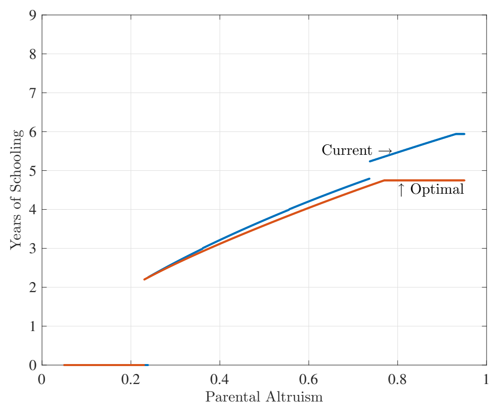
This chapter argues that heterogeneous parental altruism can be an important determinant of parental transfers and youth's education investment, and it significantly affects how education subsidies should be distributed. Parental transfers are important sources of funding for



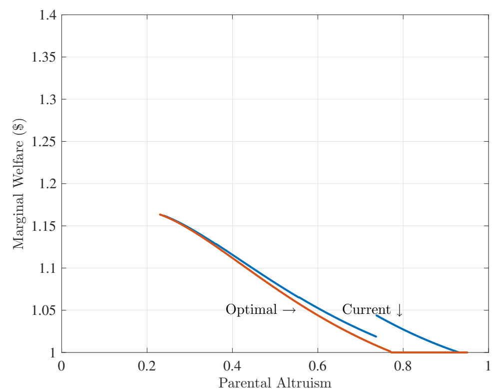
(a) Schooling (Low Income)



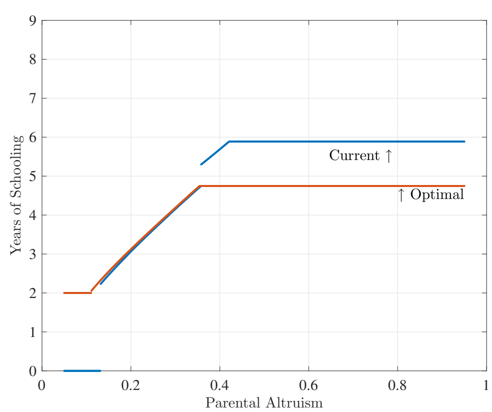
(b) Marginal Welfare (Low Income)



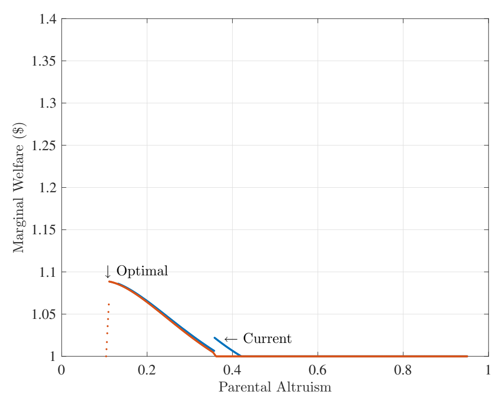
(c) Schooling (Middle Income)



(d) Marginal Welfare (Middle Income)



(e) Schooling (High Income)



(f) Marginal Welfare (High Income)

Figure 3.10: Schooling and Marginal Welfare

Table 3.8: Average Effects of Optimal Policy for Middle Income Families

	Years of Schooling	Welfare ($\underline{C}$) ($1,000)	Subsidy (G) ($1,000)	Distortion ($1,000)			
				Total (Ω)	Schooling (Ω_s)	Intertemporal ($\Omega_{c,k}$)	Intergenerational (Ω_g)
Current	2.84	815.39	4.26	45.23	13.01	0.76	31.47
Optimal	2.70	815.55	4.26	45.08	12.56	0.84	31.68
Change	-0.14	0.15	0.00	-0.15	-0.45	0.08	0.22

Table 3.9: Average Effects of Optimal Policy for Low, Middle, and High Income Families

	Income	Years of Schooling	Welfare ($\underline{C}$) ($1,000)	Subsidy (G) ($1,000)	Distortion ($1,000)			
					Total (Ω)	Schooling (Ω_s)	Intertemporal ($\Omega_{c,k}$)	Intergenerational (Ω_g)
Current	Low	1.78	456.95	7.55	95.44	18.57	3.99	72.88
	Middle	2.84	815.39	4.26	45.23	13.01	0.76	31.47
	High	4.24	1361.83	4.04	20.52	7.29	0.36	12.87
	Average	2.67	774.13	5.60	61.39	14.20	2.04	45.15
Optimal	Low	1.02	452.20	1.53	94.17	25.24	1.12	67.80
	Middle	2.64	814.51	3.20	45.06	12.99	0.73	31.34
	High	3.96	1376.22	18.64	20.73	4.39	0.99	15.35
	Average	2.22	774.68	5.59	60.83	16.42	0.95	43.47
Change	Low	-0.76	-4.75	-6.02	-1.27	6.67	-2.87	-5.07
	Middle	-0.19	-0.88	-1.05	-0.17	-0.02	-0.03	-0.12
	High	-0.27	14.39	14.60	0.21	-2.89	0.63	2.48
	Average	-0.45	0.55	-0.01	-0.56	2.21	-1.09	-1.68

college students, but the amount of parental transfers vary from family to family. Since parents who give higher transfers during schooling also give higher transfers after schooling, the heterogeneity in transfers is likely to be driven by differences in parental altruism rather than differences in returns to schooling. In the presence of borrowing constraints, heterogeneous parental altruism generates dispersion in education attainment among students with similar ability and family income. Those with altruistic parents receive enough transfers to make efficient education investment while those of parents with low altruism are constrained and attain less schooling due to low transfers. Since additional subsidies to unconstrained students do not improve economic efficiency, the social planner redistributes towards students of low altruism parents who attain less schooling. This implies that early years of college should be more heavily subsidized than late years. However, increasing subsidy for early years of college distorts college-going decision at the extensive margin. The distortionary cost can be large and can severely limit the amount of subsidy assigned to early years of college for a given income group as well as the amount of subsidy given to low income families with low college attendance rate.

In this chapter, I focus on one dimension of heterogeneity, parental altruism, across families and policies that care only about efficiency. The implications of different sources of unobserved heterogeneity across families, such as returns to schooling and parental wealth, and different social objectives are briefly discussed in the chapter, but quantitatively evaluating the importance of various heterogeneity across families and different social objectives on the design of education policies is an important extension to be addressed in future work.

Bibliography

Abbott, Brant, Giovanni Gallipoli, Costas Meghir, and Giovanni L Violante. 2013. "Education Policy and Intergenerational Transfers in Equilibrium." *Working Paper* .

Altonji, Joseph G, Prashant Bharadwaj, and Fabian Lange. 2009. "Constructing AFQT Scores that are Comparable Across the NLSY79 and the NLSY97." *Working paper* .

Becker, Gary S. 1967. "Human Capital and the Personal Distribution of Income: An Analytical Approach." Woytinsky Lecture. University of Michigan.

Becker, Gary S. and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics* 4 (3):pp. S1–S39.

Bénabou, Roland. 2002. "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?" *Econometrica* 70 (2):481–517.

Bils, Mark and Peter J. Klenow. 2000. "Does Schooling Cause Growth?" *The American Economic Review* 90 (5):1160–1183.

Bovenberg, A. Lans and Bas Jacobs. 2005. "Redistribution and Education Subsidies are Siamese Twins." *Journal of Public Economics* 89 (1112):2005 – 2035.

Brown, Meta, John Karl Scholz, and Ananth Seshadri. 2012. "A New Test of Borrowing Constraints for Education." *The Review of Economic Studies* 79 (2):511–538.

Browning, Martin, Lars Peter Hansen, and James J. Heckman. 1999. "Chapter 8 Micro data and general equilibrium models." Elsevier, 543 – 633.

Bruce, Neil and Michael Waldman. 1991. "Transfers in Kind: Why They Can be Efficient and Nonpaternalistic." *The American Economic Review* 81 (5):1345–1351.

Buchanan, James. 1975. "The Samaritans Dilemma." In *Altruism, Morality, and Economic Theory*, edited by Edmund Phelps. New York: Russell Sage foundation, 71–85.

Card, David. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69 (5):pp. 1127–1160.

Caucutt, Elizabeth M and Krishna B Kumar. 2003. "Higher Education Subsidies and Heterogeneity: a Dynamic Analysis." *Journal of Economic Dynamics and Control* 27 (8):1459–1502.

Caucutt, Elizabeth M., Lance Lochner, and Youngmin Park. 2015. "Correlation, Consumption, Confusion, or Constraints: Why do Poor Children Perform so Poorly?" Working Paper 21023, National Bureau of Economic Research.

Córdoba, Juan Carlos and Marla Ripoll. 2014. "The Elasticity of Intergenerational Substitution, Parental Altruism, and Fertility Choice." *Working Paper* .

Cunha, Flavio, James Heckman, and Salvador Navarro. 2005. "Separating Uncertainty from Heterogeneity in Life Cycle Earnings." *Oxford Economic Papers* 57 (2):191–261.

de Fraja, Gianni. 2002. "The Design of Optimal Education Policies." *The Review of Economic Studies* 69 (2):437–466.

Diamond, Peter A. 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *The American Economic Review* 88 (1):pp. 83–95.

Farhi, Emmanuel and Iván Werning. 2010. "Progressive Estate Taxation." *The Quarterly Journal of Economics* 125 (2):635–673.

———. 2013. "Estate Taxation with Altruism Heterogeneity." *The American Economic Review* 103 (3):489–95.

Findeisen, Sebastian and Dominik Sachs. 2014. "Designing Efficient College and Tax Policies." *Working Paper* .

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W. Huffman. 1988. "Investment, Capacity Utilization, and the Real Business Cycle." *The American Economic Review* 78 (3):402–417.

Haider, Steven J. and Kathleen M. McGarry. 2012. "Parental Investments in College and Later Cash Transfers." Working Paper 18485, National Bureau of Economic Research.

Hanushek, Eric A., Charles Ka Yui Leung, and Kuzey Yilmaz. 2014. "Borrowing Constraints, College Aid, and Intergenerational Mobility." *Journal of Human Capital* 8 (1):pp. 1–41.

Heckman, James J, Lance Lochner, and Christopher Taber. 1998. "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents." *Review of Economic Dynamics* 1 (1):1 – 58.

Heckman, James J., Lance J. Lochner, and Petra E. Todd. 2006. "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond." In *Handbook of the Economics of Education*, vol. 1, edited by E. Hanushek and F. Welch. Elsevier, 307 – 458.

Hellwig, Martin F. 2010. "Incentive Problems With Unidimensional Hidden Characteristics: A Unified Approach." *Econometrica* 78 (4):1201–1237.

Jacquet, Laurence and Etienne Lehmann. 2015. "Optimal Income Taxation when Skills and Behavioral Elasticities are Heterogeneous." *CESifo Working Paper* .

Johnson, Matthew T. 2013. "Borrowing Constraints, College Enrollment, and Delayed Entry." *Journal of Labor Economics* 31 (4):669–725.

Keane, Michael P and Kenneth I Wolpin. 2001. "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment." *International Economic Review* 42 (4):1051–1103.

Lochner, Lance and Alexander Monge-Naranjo. 2012. "Credit Constraints in Education." *Annual Review of Economics* 4 (1):225–256.

Lochner, Lance J. and Alexander Monge-Naranjo. 2011. "The Nature of Credit Constraints and Human Capital." *American Economic Review* 101 (6):2487–2529.

Mincer, Jacob. 1958. "Investment in Human Capital and Personal Income Distribution." *Journal of Political Economy* 66 (4):pp. 281–302.

Phelan, Christopher. 2006. "Opportunity and Social Mobility." *The Review of Economic Studies* 73 (2):487–504.

Ray, Debraj. 2013. "Hedonistic Altruism and Welfare Economics." *Working Paper* .

Restuccia, Diego and Carlos Urrutia. 2004. "Intergenerational Persistence of Earnings: The Role of Early and College Education." *The American Economic Review* 94 (5):1354–1378.

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *The Review of Economic Studies* 68 (1):205–229.

Sallie Mae and Ipsos. 2012. "How America Pays for College 2012." *Sallie Maes National Study of College Students and Parents Conducted by Ipsos Public Affairs (Washington: Salle Mae, 2012)* .

Weinzierl, Matthew. 2008. "Optimal Policy with Heterogeneous Parental Altruism." In *Essays in Optimal Taxation*, Ph.D. thesis, chap. 3. Harvard University, 166–204.

Chapter 4

Earnings Dynamics and Returns to Skills

4.1 Introduction

In the past few decades, the U.S. has experienced a sharp surge in labor income dispersion. Earnings differentials across workers with different observed characteristics such as education and age/experience all rose substantially, but residual earnings inequality (i.e. inequality conditional on observable characteristics) increased simultaneously. Although there is widespread agreement that the returns to observable skills have increased dramatically (e.g., Katz and Autor, 1999), there are different views about the factors underlying changes in residual earnings inequality. Changes in residual inequality have often been equated with changes in returns to unobserved skills (Juhn, Murphy, and Pierce, 1993) or changes in the distribution of unobserved skills (Lemieux, 2006) in the literature. However, a sizeable portion of the rise in residual inequality is attributable to short-term fluctuations in earnings (Gottschalk and Moffitt, 1994), which are unlikely to reflect changes related to skills. Distinguishing between these explanations is important for understanding the causes and welfare consequences of rising earnings inequality.

In this chapter, we consider a general model of log earnings residuals that incorporates (i) changing returns to unobserved skills, (ii) changing distribution of unobserved skills, and (iii) changing volatility of earnings unrelated to skills. We begin by discussing identification when unobserved skills are subject to permanent shocks and the volatile (non-skill) component of earnings is subject to permanent and transitory shocks as in much of the literature on earnings dynamics. We show that, when the panel data cover long periods, the returns to unobserved skills for all periods are identified and the variances of unobserved skills as well as the volatile component of earnings are identified except for the last few periods. Intuitively, identification of the returns to unobserved skills derives from the fact that correlations in earnings residual changes far enough apart in time are due to unobserved skills and not permanent or transitory shocks unrelated to skills.

We estimate the returns to unobserved skills, the variances of unobserved skills, and the variances of the non-skill component of earnings using data on log annual/weekly earnings for men ages 30-59 in the Panel Study of Income Dynamics (PSID) separately by college attendance. We highlight five main findings. (i) The returns to unobserved skills were stable until the mid-1980s, and then fell afterwards. The decline of the return was more substantial for those who did not attend college ('high school workers'). (ii) The variance of unobserved skill rose throughout the sample period. (iii) The variance of the unobserved skill component (i.e., return to skill multiplied by unobserved skill) rose until the mid-1980s, decreased afterwards, and then rebounded slightly in the early 2000s only for those who attended college ('college workers'). (iv) The variance of permanent shocks increased since the mid-1980s. (v) The variance of transitory shocks increased sharply in the early 1980s among high school workers but increased only modestly in the long-run for both high school and college workers.

Our results demonstrate that accounting for changes in the distribution of unobserved skills and in the volatility of earnings is important in estimating the returns to unobserved skills. Time patterns of the returns to unobserved skills are quite different from those estimated in the existing studies (e.g., Juhn, Murphy, and Pierce, 1993, Moffitt and Gottschalk, 2012) that do not take into account these factors. The evolution of the variance of log earnings residuals after the mid-1980s cannot be easily explained by unobserved skills alone, as the decline in the returns to unobserved skills pushed down the share of skill-related component in total residual variance.

We interpret our empirical findings using a simple demand and supply framework and ask whether the changes in the returns to unobserved skills reflect changes in skill demand or supply. We present an assignment model of labor market that is similar to Sattinger (1979) where the returns to skills are determined by production technology and how workers with heterogeneous skills and jobs with heterogeneous productivity are combined in production. More skilled workers earn more partly because they work at more productive jobs. Skill distribution affects the returns to skills by changing the equilibrium assignment. An increase in the variance of skill reduces the returns to skills by reducing the productivity differential across jobs among workers different skill levels. We recover the demand and supply factors by combining our estimates—returns to skills and variance of skills—with the restrictions implied by the theoretical model. We find that both demand and supply factors played important roles in the decline of the returns to unobserved skills since the mid-1980s for both high school and college workers. The decrease in skill demand was a more important factor than supply changes for high school workers.

This chapter proceeds as follows. In Section 4.2, we provide identification results for our model where the returns to unobserved skills, the variance of unobserved skills, and the variance of non-skill component of earnings change over time. Section 4.3 describes the PSID data used to estimate earnings dynamics for American men and reports our empirical findings. Section 4.4 interprets the time patterns of the estimated returns to unobserved skills in a demand and supply framework and Section 4.5 concludes.

4.2 Identifying the Returns to Unobserved Skills

In this section, we establish conditions under which the returns to unobserved skills, the variances of unobserved skills, and the variance of volatile component of earnings are identified.

Suppose that we observe log earnings residuals of a large number of individuals $i = 1,\ldots,N$ from period $t = \underline{t}$ to period $t = \bar{t}$. We consider log earnings residuals for individual i in period t of the form

$$w_{i,t} = \mu_t \theta_{i,t} + u_{i,t}, \tag{4.1}$$

where $\theta_{i,t}$ represents an unobserved skill, μ_t a return to unobserved skills, and $u_{i,t}$ idiosyncratic shocks. We assume that $\theta_{i,t}$ and $u_{i,t}$ are mean zero and uncorrelated with each other.

The earnings equation (4.1) reflects the classical idea of Friedman and Kuznets (1954) that one's earnings consists of a permanent component that is related to one's intrinsic skills and a transitory component that reflects a short-run variation in earnings that is not related to skills. Although the transitory component can be serially correlated, the correlation between transitory components that are far apart in time is likely to be small. Therefore, we begin with the assumption that there exists $k > 0$ such that $\mathrm{Cov}(u_t, u_{t'}) = 0$ for $|t' - t| \geq k$, which will be generalized later.[1] Then the "long" autocovariance of earnings residuals reflects only the skill-related component (Carroll, 1992, Moffitt and Gottschalk, 2012): for $|t' - t| \geq k$,

$$\mathrm{Cov}(w_t, w_{t'}) = \mu_t \mu_{t'} \mathrm{Cov}(\theta_t, \theta_{t'}). \tag{4.2}$$

Time-invariant Unobserved Skills When unobserved skills do not change over time, i.e., $\theta_{i,t} = \theta_i$, then $\mathrm{Cov}(\theta_t, \theta_{t'}) = \mathrm{Var}(\theta)$. In this case, the ratio of the covariances reveals the ratio of the returns to skills: for $|t' - t| > k$,

$$\frac{\mathrm{Cov}(w_t, w_{t'})}{\mathrm{Cov}(w_{t-1}, w_{t'})} = \frac{\mu_t \mu_{t'} \mathrm{Var}(\theta)}{\mu_{t-1} \mu_{t'} \mathrm{Var}(\theta)} = \frac{\mu_t}{\mu_{t-1}}. \tag{4.3}$$

[1] Let x_t be a random variable and its realization for individual i be $x_{i,t}$. Denote its cross-sectional second moments by $\mathrm{Var}(x_t)$ and $\mathrm{Cov}(x_t, x_{t'})$.

Therefore, if we observe individual earnings for long periods, the returns to skills are identified up to a normalization that $\mu_{t^*} = 1$ for some t^*. Once the returns are identified, we can also recover the variance of unobserved skills $\mathrm{Var}(\theta) = \mathrm{Cov}(w_t, w_{t'})/(\mu_t \mu_{t'})$ as well as the variance of earnings shocks $\mathrm{Var}(u_t) = \mathrm{Var}(w_t) - \mathrm{Var}(\mu_t \theta)$.

Equation (4.3) can be interpreted in the framework of instrumental variable regression. We can obtain μ_t/μ_{t-1} by regressing $w_{i,t}$ on $w_{i,t-1}$ using $w_{i,t'}$ as an instrumental variable. To see why, first note that $w_{i,t}$ and $w_{i,t-1}$ are related as follows:

$$w_{i,t} = \mu_t \theta_i + u_{i,t} = \mu_t \left(\frac{w_{i,t-1} - u_{i,t-1}}{\mu_{t-1}} \right) + u_{i,t}. \tag{4.4}$$

Since $w_{i,t-1} = \mu_{t-1}\theta_i + u_{i,t-1}$ is a noisy measure of unobserved skill θ_i, it is correlated with the measurement error $u_{i,t-1}$ (and also correlated with $u_{i,t}$ if $u_{i,t}$ and $u_{i,t-1}$ are correlated) and running an OLS regression of $w_{i,t}$ on $w_{i,t-1}$ would give a biased estimate of μ_t/μ_{t-1}. To address this endogeneity problem, an earnings residual of distant future or past $w_{i,t'}$ can be used as an instrumental variable since it is correlated with θ_i but uncorrelated with $u_{i,t-1}$ and $u_{i,t}$.

Time-varying Unobserved Skills In more general cases where unobserved skills change over time, let $\nu_{i,t} = \Delta\theta_{i,t} = \theta_{i,t} - \theta_{i,t-1}$ be the skill change between period $t-1$ and period t. Then Equation (4.4) becomes

$$w_{i,t} = \mu_t \left(\frac{w_{i,t-1} - u_{i,t-1}}{\mu_{t-1}} + \nu_{i,t} \right) + u_{i,t}.$$

As long as the skill changes are uncorrelated with past skills,[2] past earnings residuals are still valid instruments and μ_t for $t > \underline{t} + k$ is identified by Equation (4.3). However, future earnings residuals are not valid if the skill change has permanent effects on future skills. To see this, note that, for $t' > t$, $\theta_{t'} = \theta_t + \sum_{\tau=t+1}^{t'} \nu_\tau$ and $\mathrm{Cov}(\theta_t, \theta_{t'}) = \mathrm{Var}(\theta_t)$. Then, for $t' \geq t + k$,

$$\frac{\mathrm{Cov}(w_t, w_{t'})}{\mathrm{Cov}(w_{t-1}, w_{t'})} = \frac{\mu_t \mu_{t'} \mathrm{Var}(\theta_t)}{\mu_{t-1}\mu_{t'} \mathrm{Var}(\theta_{t-1})} = \frac{\mu_t}{\mu_{t-1}} \frac{\mathrm{Var}(\theta_{t-1}) + \mathrm{Var}(\nu_t)}{\mathrm{Var}(\theta_{t-1})}, \tag{4.5}$$

which is larger than μ_t/μ_{t-1} unless $\mathrm{Var}(\nu_t) = 0$. This poses an identification problem for μ_t in early periods where there are no past earnings to be used as instrumental variables. However, they can still be recovered by differencing out the bias generated by permanent skill changes. Let c_i be the period when individual i starts accumulating skills ('cohort') and suppose that there exist two cohorts c and c' such that $\mathrm{Var}(\theta_{t-1}|c) \neq \mathrm{Var}(\theta_{t-1}|c')$ and $\mathrm{Var}(\nu_t|c) = \mathrm{Var}(\nu_t|c')$. The second condition holds when skill shock variance depends only on time, or when there

[2] This is the standard weak exogeneity assumption in dynamic models.

is a non-monotonic age trend in the variance of skill change.[3] In this case, the ratio of the difference of the long autocovariance between cohorts gives the ratio of the returns to skills for early periods: for $t'-t \geq k$,

$$\frac{\mathrm{Cov}(w_t, w_{t'}|c) - \mathrm{Cov}(w_t, w_{t'}|c')}{\mathrm{Cov}(w_{t-1}, w_{t'}|c) - \mathrm{Cov}(w_{t-1}, w_{t'}|c')} = \frac{\mu_t \mu_{t'} \left[\mathrm{Var}(\theta_{t-1}|c) - \mathrm{Var}(\theta_{t-1}|c')\right]}{\mu_{t-1} \mu_{t'} \left[\mathrm{Var}(\theta_{t-1}|c) - \mathrm{Var}(\theta_{t-1}|c')\right]} = \frac{\mu_t}{\mu_{t-1}}.$$

Then the variances of unobserved skills for all $t \leq \bar{t} - k$ are also identified from the covariance between current and future earnings residuals $\mathrm{Var}(\theta_t) = \mathrm{Cov}(w_t, w_{t'})/(\mu_t \mu_{t'})$ for $t'-t \geq k$. Although the return to unobserved skill is identified for all periods, the variance of unobserved skills is not identified for later periods near $\bar{t}$, because we cannot distinguish between unobserved skills and earnings shocks without future earnings residuals.

Permanent Non-skill Component of Earnings We have assumed that all permanent components of earnings are related to skills. However, there may be factors that are orthogonal to skills but also have permanent effects on earnings. For example, workers with identical skills may get paid differently because of the differences in non-pecuniary characteristics of jobs (Rosen, 1986). These differences are likely to be permanent if they reflect workers' preferences over jobs. Also permanent shocks to firm productivity can be translated into permanent earnings shocks among similarly skilled workers when there are labor market frictions that prevent full insurance against firm level shocks (Guiso, Pistaferri, and Schivardi, 2005, Lamadon, 2014).

In this case, a similar identification strategy works under the modified assumption that $\mathrm{Cov}(u_t, \Delta u_{t'}|c) = 0$ for all (c, t, t') such that $t'-t \geq k$. With this assumption, the change in distant future earnings residual $\Delta w_{i,t'}$, instead of level $w_{i,t'}$, can be used as an instrumental variable in a regression of $w_{i,t}$ on $w_{i,t-1}$ unless the return to unobserved skill stays constant (i.e., $\Delta\mu_{t'} = 0$). To see this, note that $\Delta w_{i,t'}$ can be written as

$$\Delta w_{i,t'} = \Delta\mu_{t'} \theta_{i,t'} + \mu_{t'-1} \Delta\theta_{i,t'} + \Delta u_{i,t'}. \tag{4.6}$$

For $t' \geq t-k$, $\Delta w_{i,t'}$ is uncorrelated with $u_{i,t-1}$ and $u_{i,t}$, but it is correlated with $w_{i,t-1}$ through $\theta_{i,t'}$ if $\Delta\mu_{t'} \neq 0$. Therefore, the availability of an instrumental variable hinges on some variation in the returns to unobserved skills over time. In the extreme case where μ_t does not change at all, it is impossible to separately identify the permanent skill component and non-skill component. To simplify exposition, we impose the strong sufficient condition that $\Delta\mu_t \neq 0$ for all

[3]For example, young (old) workers may experience larger skill changes than middle age workers due to training (health shocks). Baker and Solon (2003) and Blundell, Graber, and Mogstad (2015) find this U-shaped age profile in the variance of earnings shocks.

periods.

We summarize the above discussion in the following Proposition. The proofs for this and subsequent results are provided in Appendix C.1.

Proposition 4.1 *Suppose that (i) there exists* $k > 0$ *such that* $\bar{t} - \underline{t} > 2k$ *and* $\mathrm{Cov}(u_t, \Delta u_{t'}|c) = 0$ *for all* (c, t, t') *such that* $t' - t \geq k$*, (ii)* $\Delta\mu_t \neq 0$ *for all* t*, and (iii)* $\theta_{i,t} = \theta_{i,t-1} + \nu_{i,t}$ *where, for all* t*,* $\mathrm{Cov}(\nu_t, \theta_{t-j}|c) = 0$ *for all* c *and* $j > 0$ *and* $\mathrm{Var}(\theta_{t-1}|c) \neq \mathrm{Var}(\theta_{t-1}|c')$ *and* $\mathrm{Var}(\nu_t|c) = \mathrm{Var}(\nu_t|c')$ *for some* $c \neq c'$*. Then (i)* μ_t *is identified for all* t*, up to a normalization* $\mu_{t^*} = 1$ *for some* t^**, and (ii)* $\mathrm{Var}(\theta_t)$ *and* $\mathrm{Var}(u_t)$ *are identified for all* $t < \bar{t} - k$*.*

4.3 Estimation

We estimate the returns to unobserved skills, the variance of unobserved skills, and the variance of earnings shocks using panel data from the Panel Study of Income Dynamics (PSID) on male annual/weekly earnings in the U.S. from 1970 to 2008.

4.3.1 Empirical Specification

Since the relative wages between those who did not attend college and those that did have diverged significantly during the sample periods (Katz and Murphy, 1992), we allow all parameters including the returns to unobserved skills to differ between the two groups, which we call 'sector'. Specifically, let s_i be an indicator variable for college workers for individual i. Then, the log earnings residual for individual i in year t is

$$w_{i,t} = \mu_t(s_i)\theta_{i,t} + u_{i,t}, \tag{4.7}$$

where $\mu_t(s)$ is the return to unobserved skills for an individual with $s_i = s$ and, as before, unobserved skill evolves according to

$$\theta_{i,t} = \theta_{i,t-1} + \nu_{i,t}. \tag{4.8}$$

We further assume that the non-skill earnings shock $u_{i,t}$ is the sum of a unit root process ('permanent component') and a lower order moving average process ('transitory component') as typically assumed in the literature:[4]

$$u_{i,t} = \kappa_{i,t} + \varepsilon_{i,t}, \tag{4.9}$$

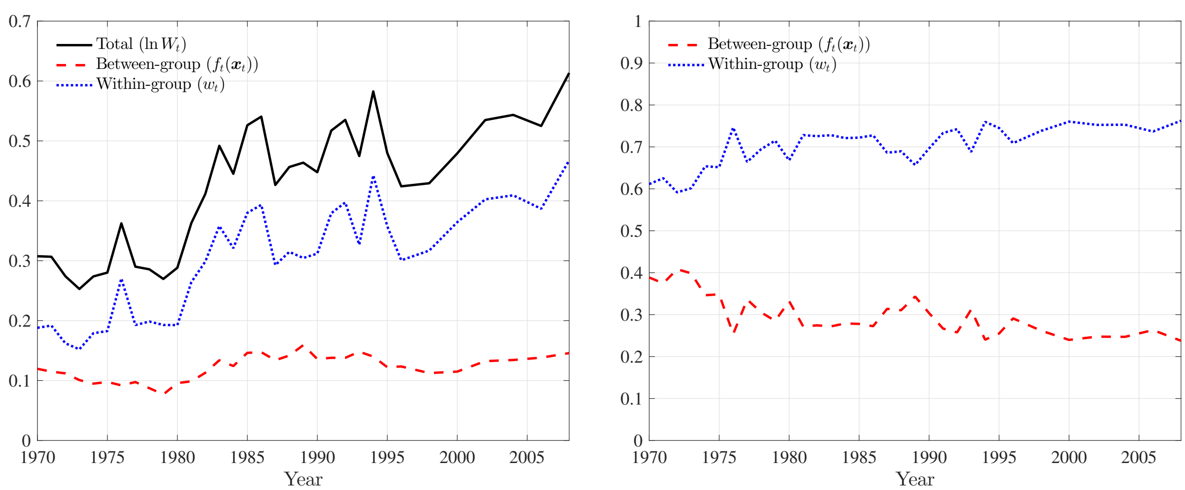
$$\kappa_{i,t} = \kappa_{i,t-1} + \eta_{i,t}, \tag{4.10}$$

$$\varepsilon_{i,t} = \sum_{j=0}^{q} \beta_j(s_i)\xi_{i,t-j}, \tag{4.11}$$

where $\beta_0(s) = 1$. With this specification, the identifying assumption $\mathrm{Cov}(u_t, \Delta u_{t'}|s,c) = 0$ in Proposition 4.1 holds for $k = q+2$, so the returns to unobserved skills for all periods and the variances of unobserved skills for $t < \bar{t} - k = \bar{t} - q - 2$ are identified. Variances of permanent and transitory non-skill components for $t < \bar{t} - q - 2$ are also identified as follows: $\mathrm{Var}(\kappa_t|s) = \mathrm{Cov}(w_t, w_{t'}|s) - \mu_t(s)\mu_{t'}(s)\,\mathrm{Var}(\theta_t|s)$ for $t' \geq t+q+1$ and $\mathrm{Var}(\varepsilon_t|s) = \mathrm{Var}(w_t|s) - \mu_t(s)^2\,\mathrm{Var}(\theta_t|s) - \mathrm{Var}(\kappa_t|s)$.

4.3.2 PSID Data

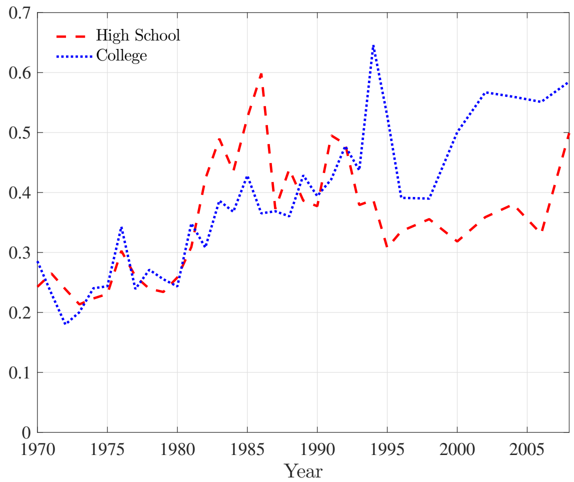
The PSID is a longitudinal survey of a representative sample of individuals and families in the U.S. beginning in 1968. The survey was conducted annually through 1997 and biennially since. We use data collected from 1971 through 2009. Since earnings and weeks of work were collected for the year prior to each survey, our analysis considers annual earnings from 1970 to 2008. (Key findings also hold for weekly earnings as shown in Appendix C.3.)



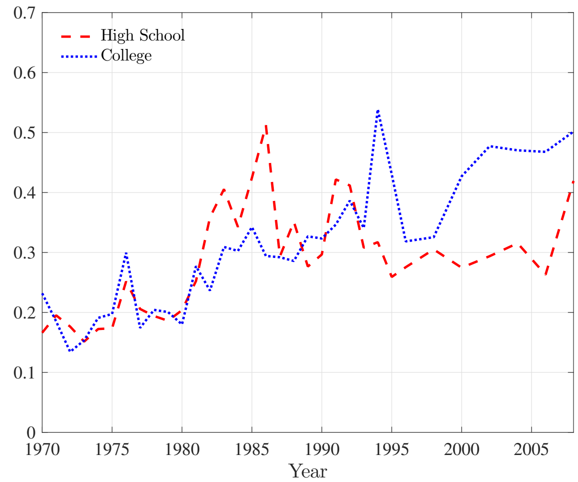
(a) Variance (b) Share of Variance

Figure 4.1: Between- and Within-group Variance of Log Annual Earnings

[4]See, among others, Abowd and Card (1989), Blundell and Preston (1998), Meghir and Pistaferri (2004), Blundell, Pistaferri, and Preston (2008), Heathcote, Perri, and Violante (2010). See MaCurdy (2007) for a review.



(a) Total Variance



(b) Within-group Variance

Figure 4.2: Variance of Log Annual Earnings by Sector

Our sample is restricted to male heads of households from the core (SRC) sample.[5] We use earnings from any year these men were ages 30-59, had positive wage and salary income, worked at least one week, and were not enrolled as a student. Our earnings measure reflects total wage and salary earnings (excluding farm and business income) and is denominated in 1996 dollars using the CPI-U-RS. We trim the top and bottom 1% of all earnings measures within year and sector by ten-year age cells. The resulting sample contains 3,194 men and 32,549 person-year observations – roughly ten observations for each individual.

Our sample is composed of 92% whites, 6% blacks and 1% hispanics with an average age of 42 years old. We create seven education categories based on current years of completed schooling: 1-5 years, 6-8 years, 9-11 years, 12 years, 13-15 years, 16 years, and 17 or more years. College workers are defined as those with more than 12 years of schooling. In our sample, 16% of respondents finished less than 12 years of schooling, 33% had exactly 12 years of completed schooling, 20% completed some college (13-15 years), 21% completed college (16 years), and 10% had more than 16 years of schooling.

Our analysis focuses on log annual earnings residuals after controlling for differences in educational attainment, race, and age. Let $W_{i,t}$ be earnings and $\boldsymbol{x}_{i,t}$ be a vector of observed

[5]We exclude those from any PSID oversamples (SEO, Latino) as well as those with non-zero individual weights. The earnings questions we use are asked only of household heads. We also restrict our sample to those who were heads of household and not students during the survey year of the observation of interest as well as two years earlier. Our sampling scheme is very similar to that of Moffitt and Gottschalk (2012), except that we do not include earnings measures before age 30.

characteristics for individual i in year t. Log earnings residual $w_{i,t}$ can be written as follows:

$$\ln W_{i,t} = f_t(\boldsymbol{x}_{i,t}) + w_{i,t}. \tag{4.12}$$

We run year- and sector-specific regressions of log earnings on age, race, and education indicators, along with interactions between race and education indicators and a third order polynomial in age.

Figure 4.1 shows total variance, between-group variance, and within-group variance (variance of residuals) of log annual earnings. The variance of log annual earnings increases sharply in the early 1980s and after the late 1990s. The evolution of the within-group variance closely mirrors the evolution of the total variance. The within-group variance explains a larger share of total variance than the between-group variance, and it also explains an increasing share of the total variance over time. Although the within- and between-group variance follow similar time patterns, they show a slight divergence in the mid- and late 1970s and the late 1990s.

The total and within-group variance of log annual earnings evolved quite differently across sectors, as shown in Figure 4.2. The rise in inequality in the early 1980s is stronger among high school workers, while the increasing inequality among college workers is the main driver of the surge in inequality in the early 2000s. Between the mid-1980s and mid-2000s, earnings inequality fell among high school workers and increased among college workers, which is consistent with the literature on 'polarization' in the U.S. labor market (Autor, Levy, and Murnane, 2003, Autor, Katz, and Kearney, 2008, Acemoglu and Autor, 2011, Autor and Dorn, 2012).

4.3.3 Minimum Distance Estimation

We use minimum distance estimation and the residual moments described above to estimate the model by choosing parameters that minimize the distance between the sample covariances and the theoretical covariances implied by the model. We assume that individuals enter labor market at age 20 with initial skill ψ_i and no other prior shocks, which implies that unobserved skill in year t for individual i born in year c_i ('birth cohort') can be written as

$$\theta_{i,t} = \psi_i + \sum_{\tau=c_i+20}^{t} \nu_{i,\tau}.$$

Variance of initial skill does not depend on birth cohort c_i and the variances of all shocks depend on only time and not on cohort or age. We also assume that the variances of all shocks prior to 1970 are the same as those in year 1970. We impose smooth time trend (cubic polynomials in year) in the variances of permanent skill and non-skill shocks, while the variances of transitory shocks are unrestricted. We choose a normalization that $\mu_{1985}(s) = 1$.

The parameters we estimate are, for each s, $\text{Var}(\psi|s)$, $\rho(s)$, $\{\beta_j(s)\}_{j=1}^{q}$, and parameters for $\text{Var}(\nu_t|s)$ and $\text{Var}(\eta_t|s)$, and for each (s,t), $\mu_t(s)$ (except for $t = 1985$) and $\text{Var}(\xi_t|s)$. For a given parameter vector $\mathbf{\Lambda}$, we can compute theoretical counterpart for $\text{Cov}(w_t, w_{t'}|s,c)$ implied by the model (4.7)-(4.11) and compare them with the sample covariances. Since some cohort (or, equivalently, age $a_t = t - c$) cells have few observations when calculating residual covariances, we partition the age set $A = \{30,\ldots,59\}$ into 10-year age groups A_1, A_2, and A_3, each corresponding to ages 30-39, 40-49, and 50-59, and aggregate within age groups.

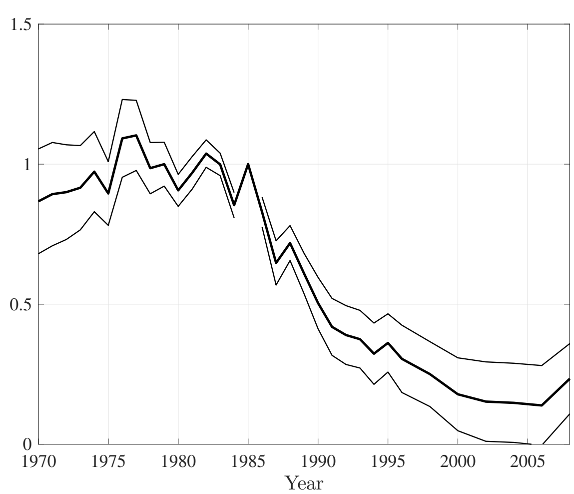
The minimum distance estimator $\hat{\mathbf{\Lambda}}$ solves

$$\min_{\mathbf{\Lambda}} \sum_{(s,j,t,t')\in\Gamma} \left\{ \widehat{\text{Cov}}(w_t, w_{t'}|s, A_j) - \widetilde{\text{Cov}}(w_t, w_{t'}|s, A_j, \mathbf{\Lambda}) \right\}^2,$$

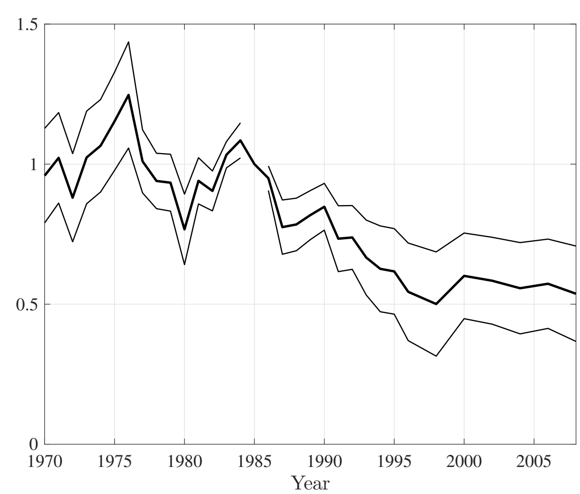
where $\Gamma = \{s,j,t,t'|1970 \le t \le t' \le 2008, s \in \{0,1\}, j \in \{1,2,3\}\}$, $\widehat{\text{Cov}}(w_t, w_{t'}|s,A_j)$ is the sample covariance conditional on sector s and age group A_j in year t', and $\widetilde{\text{Cov}}(w_t, w_{t'}|s, A_j, \mathbf{\Lambda})$ is the corresponding theoretical covariance.

4.3.4 Estimation Results

We discuss results for log annual earnings residuals with MA(3) transitory shocks in the text; however, conclusions are quite similar for log weekly earnings (Appendix C.3) and different specifications for the transitory component (Appendix C.2).



(a) High School



(b) College

Figure 4.3: Estimated $\mu_t(s)$ (thick lines) with 95% Confidence Interval (thin lines)

Estimated Returns Figure 4.3 reports the estimated returns and 95% confidence intervals over time. For high school workers, the return to unobserved skills increases in the early 1970s, levels off in the late 1970s and early 1980s, then falls about 80% between 1985 and 2000. It remains fairly constant thereafter. For college workers, the return to unobserved skills increases more sharply until the mid-1970s, decreases about 75% between 1976 and 1998, and stays constant afterwards. Overall, the returns to unobserved skills fell substantially after 1985 in both sectors despite the increasing residual inequality during the period. Although the returns for high school and college workers display similar patterns, the return for high school workers declines 30 percentage points more than that of college workers after 1985. Indeed, we can reject the hypothesis that the returns to unobserved skills are identical across sectors at the 5% significance level.[6] The relatively larger decline in the returns to unobserved skills for high school workers after 1985 is in line with the falling relative wages of high school workers during that period (e.g., Autor, Katz, and Kearney, 2008). This suggests that the changing returns to unobserved skills can be an important determinant of the evolution of between-group inequality as well as within-group inequality.

Our finding that the returns to unobserved skills have fallen since the mid-1980s is consistent with the work of Castex and Dechter (2014), who showed that the returns to cognitive ability as measured by AFQT scores fell in the 2000s relative to the 1980s by comparing NLSY79 and NLSY97 cohorts. The decreasing returns to unobserved skills, however, is inconsistent with the conclusions reached in both the PSID- and CPS-based literatures. The CPS-based literature, which implicitly ignores earnings shocks and equates changes in the total variance log wages residuals with changes in the returns to unobserved skills, concludes that the returns to unobserved skills increased steadily after the early 1970s (Juhn, Murphy, and Pierce, 1993, Katz and Autor, 1999, Acemoglu, 2002).[7]

The PSID-based literature explores the relative importance of permanent and transitory shocks, typically ignoring variation in the returns of unobserved skills. A few exceptions are Haider (2001) and Moffitt and Gottschalk (2012); however, they reached different conclusions. Haider (2001) estimated that the returns to unobserved skills were stable in the 1970s and then increased throughout the 1980s. Moffitt and Gottschalk (2012)'s estimates suggest that the returns increased until the mid-1980s, stabilized, and then increased again in the mid-1990s.[8] We show that their very different conclusions are due to restrictions imposed in their

[6]The Wald test is used to test the hypothesis $H_0 : \mu_t(0) = \mu_t(1)$ for all $t \neq 1985$.

[7]One exception is Lemieux (2006), who finds that, after controlling for composition effects, the return to unobserved skills declined in the 1970s and 1990s. Autor, Katz, and Kearney (2008) show that the decline in the returns to unobserved skills between 1989 and 2005 is driven by decreasing returns among workers in the lower tail of the wage distribution. Lochner and Shin (2014) explore identification and estimation of the returns to unobserved skills allowing for differential changes in the returns across skill levels.

[8]There are other studies estimating models similar to Haider (2001) and Moffitt and Gottschalk (2012) using

models. Although assuming more persistent ARMA(1,1) transitory shocks, they assume that earnings processes are identical across sectors and also abstract from permanent skill and non-skill shocks with time-varying variances. Figure 4.4 shows how these restrictions affect the estimated returns.

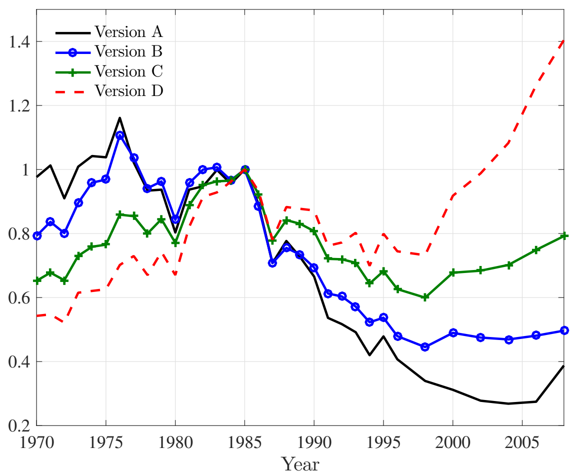


Figure 4.4: Estimated μ_t under Different Restrictions

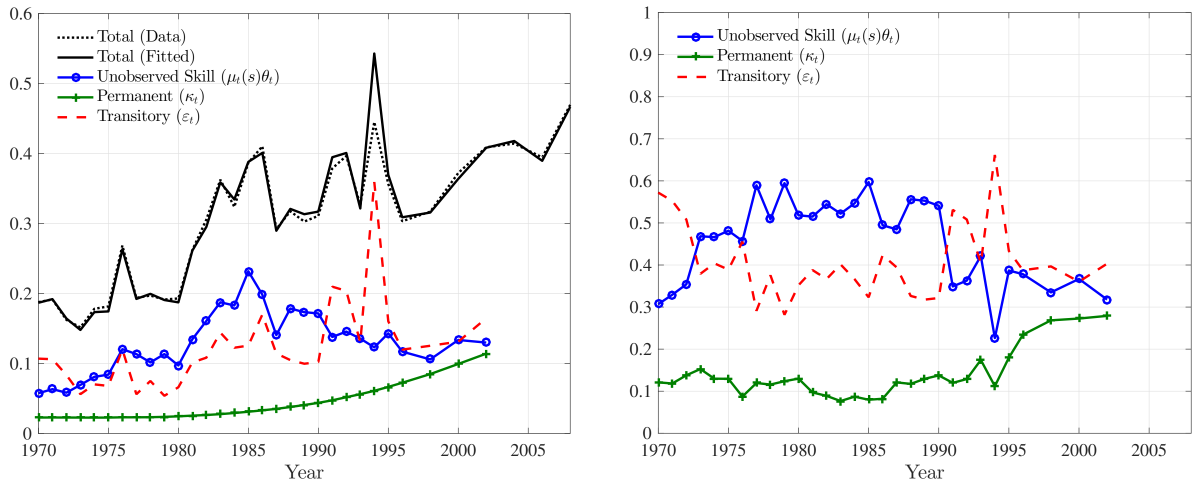
In version A, we assume that transitory shocks follow an ARMA(1,1) process and the returns are identical across sectors.[9] The estimates show similar patterns to our baseline estimates, implying that different assumptions about the persistence of transitory shocks do not affect the results significantly. In version B, we further constrain all other parameters to be identical across sectors. Version C additionally imposes that there are no permanent non-skill shocks and version D further assumes that the variances of skill shocks are constant over time. As we move from version A to version D, the estimated return changes counter-clockwise, generating increasing trends before the mid-1980s and after the mid-1990s with a muted decline in the late 1980s and the early 1990s.

Version D is essentially identical to the models of Haider (2001) and Moffitt and Gottschalk (2012)[10] and the estimates are similar, suggesting that accounting for permanent skill and non-skill shocks with time-varying variances greatly affects the estimates of the returns to unobserved skills. In the absence of permanent shocks with time-varying variances, the model's ability to match increasing residual variance is limited, and all permanent increases in residual inequality is explained by increasing returns to skills.

different panel data. Using the U.S. tax returns data from 1987 to 2009 (DeBacker et al., 2013) and Canadian tax returns data from 1976 to 1992 (Baker and Solon, 2003), they reach similar conclusions that the returns to unobserved skills increased during the sample period.

[9]Transitory shocks that follow an ARMA(1,1) process fade out only asymptotically, so we need a different set of assumptions for identification, as shown in Appendix C.1.3.

[10]They also have heterogenous growth rate of unobserved skills that is correlated with initial unobserved skills. Including heterogenous income profiles to version D has negligible effects on the estimates.



(a) Variance Decomposition

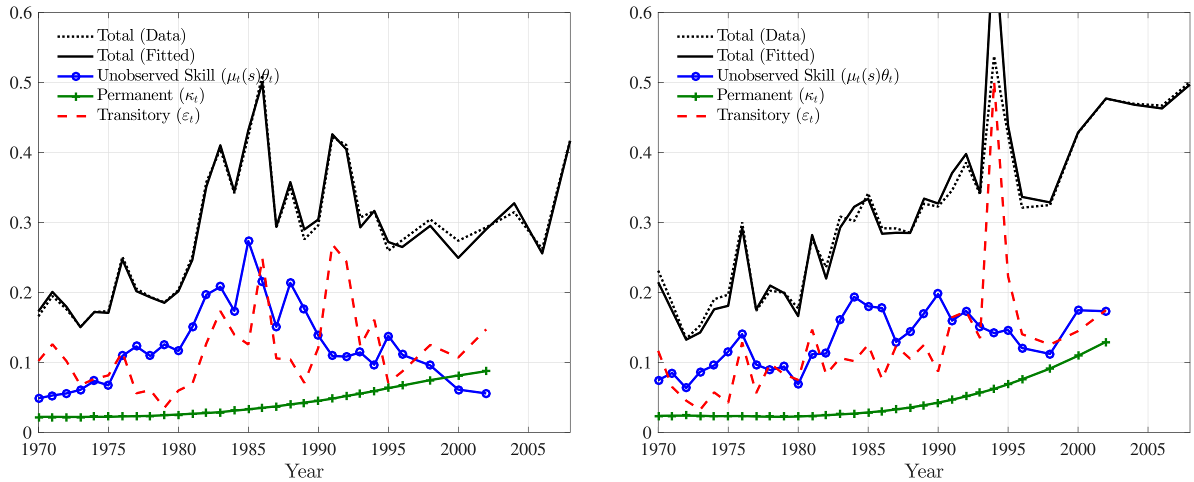
(b) Share of Each Component

Figure 4.5: Variance and Share of Each Component

Role of Unobserved Skills in Explaining Residual Variance The fact that estimated returns have evolved quite differently from the residual inequality implies that the role of unobserved skills in explaining residual inequality has also changed. Figure 4.5 decomposes the residual variance of all sectors into its three components: unobserved skills ($\mu_t(s)\theta_t$), permanent (κ_t) and transitory (ε_t) components of non-skill earnings shocks.[11] Initially quite low, the variance of the unobserved skill component rises over the 1970s and early 1980s before falling after 1985. The variance of transitory component rises in the early 1980s and the variance of permanent non-skill component stays constant throughout the 1970s and then rises continuously afterwards. As a share of the total variance in log earnings residuals, the transitory component plays the largest role in the early 1970s, after which the unobserved skill component dominates. The unobserved skill component explains about 60% of the total residual variance at its peak in the mid-1980s and its share decreases afterwards. The permanent non-skill component explains an increasing share of the total residual variance after the mid-1980s. Between 1985 and 2002, the share of the permanent non-skill component increases from 10% to 30%.

Figure 4.6 shows the decomposition of the residual variance by sector. The broad time patterns—the reversal of the importance of the unobserved skill component around 1985 and the increasing importance of permanent non-skill component afterwards—can also be found in each sector. However, the long-term fluctuation in the variance of unobserved skill component is more pronounced for high school workers: it quadruples between 1970 and 1985, then falls back to its original level by the early 2000s. For college workers, the variance of unobserved skill component decreases only modestly during the 1990s and then increases again in the early

[11] As shown in Section 4.2, the variances of unobserved skills and non-skill earnings components are not identified for the last few years of our panel. Our figures report these variances trough 2002.



(a) High School (b) College

Figure 4.6: Variance Decomposition by Sector

2000s.

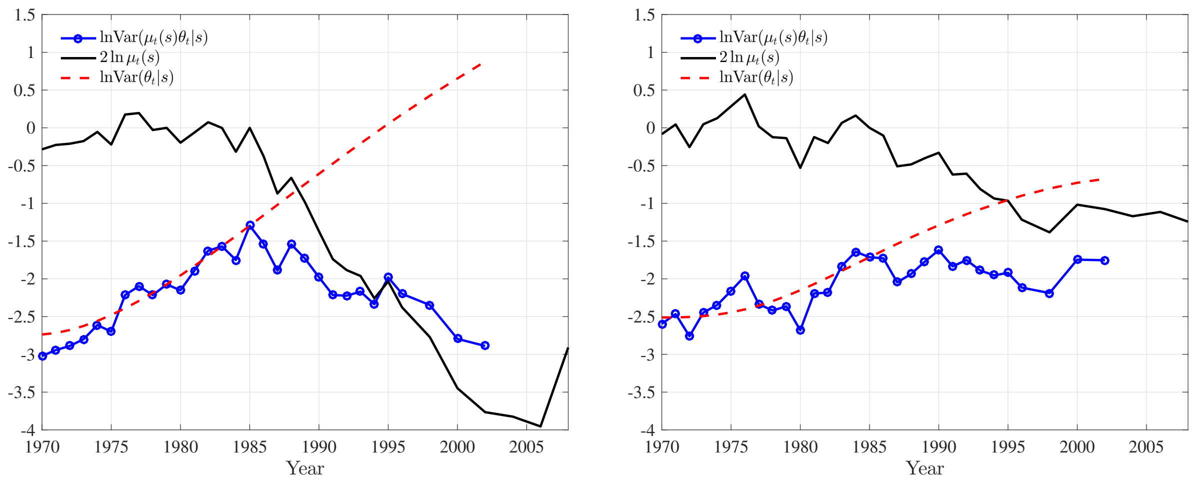
To better understand the evolution of the variance of unobserved skill component, note that the variance of residual earnings due to unobserved skills depends both on the returns to unobserved skills and the variance of unobserved skills: $\mathrm{Var}(\mu_t(s)\theta_t|s) = \mu_t(s)^2\,\mathrm{Var}(\theta_t|s)$. By taking logs, we get the following additive decomposition:

$$\ln \mathrm{Var}\left(\mu_t(s)\theta_t|s\right) = 2\ln \mu_t(s) + \ln \mathrm{Var}(\theta_t|s).$$

Figure 4.7 shows that the variance of unobserved skills increases continuously throughout our sample periods in both sectors, and that the increase is more dramatic for high school workers. Since the returns to unobserved skills stayed relatively constant until 1985, the rising variance of unobserved skills entirely explains the increasing variance of unobserved skill component before 1985 in both sectors. However, after 1985, the sharp decline in returns reduces the variance of unobserved skill component despite the sustained growth of the variance of unobserved skills.

Figure 4.8 shows that the variance of skill shocks increases substantially for high school workers after 1985 and the variance of permanent non-skill shocks for college workers increases more than high school workers. The variance of transitory shocks sharply increases for high school workers in the 1980s and early 1990s, and it has increased continuously for college workers since the mid-1970s, with a sharp spike in 1994.

To summarize empirical findings, the returns to unobserved skills fell after 1985 and the decline was larger for high school workers than college workers. The variance of earnings



(a) High School (b) College

Figure 4.7: Decomposing the Log Variance of Unobserved Skill Component

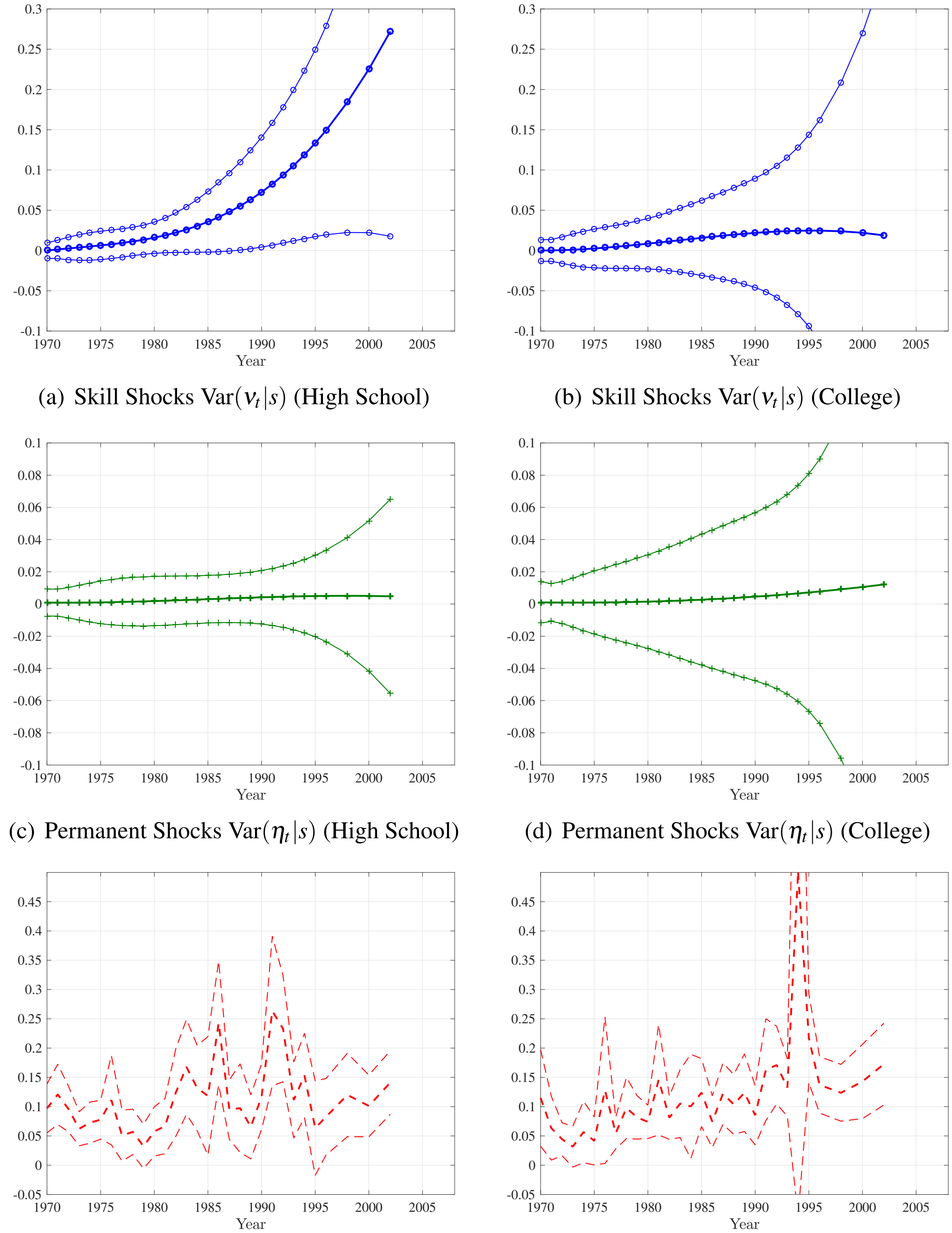
residuals due to unobserved skills increased until 1985 and fell afterwards despite the sustained growth in the variance of unobserved skills. The reversal of residual inequality for high school workers around 1985 mainly reflects the change in the variance of unobserved skill component, although the transitory component non-skill shock also contributed. The rise in residual inequality since the mid-1990s is due to increasing variances of permanent and transitory non-skill shocks as well as increasing variance of unobserved skills among college workers.

4.4 Interpreting the Returns to Skills in a Demand and Supply Framework

Our analysis shows that the changing returns to unobserved skills are important in understanding the evolution of residual inequality. The changes in the returns to skills documented in the previous section could reflect shifts in the demand for skills. However, the concurrent changes in the distribution of skills suggests that the supply of skills may also play a role. In this section, we assess the contributions of skill demand and supply in the evolution of returns to skills using a simple demand and supply framework based on an assignment model of Sattinger (1979).[12]

In the model, the return to skills is generated by differences in the productivity of skills, but is amplified by differences in the productivity of jobs for which skills are employed. Het-

[12] Assignment models have been a standard theory of income inequality in labor economics. See Sattinger (1993) for an early review. Recent theoretical and empirical studies of income inequality based on this framework include Terviö (2008), Gabaix and Landier (2008), Costinot and Vogel (2010), Lindenlaub (2014), Burstein, Morales, and Vogel (2015), Ales, Kurnaz, and Sleet (2015), Scheuer and Werning (2015).



(a) Skill Shocks Var($\nu_t|s$) (High School)

(b) Skill Shocks Var($\nu_t|s$) (College)

(c) Permanent Shocks Var($\eta_t|s$) (High School)

(d) Permanent Shocks Var($\eta_t|s$) (College)

(e) Transitory Shocks Var($\xi_t|s$) (High School)

(f) Transitory Shocks Var($\xi_t|s$) (College)

Figure 4.8: Variances of Shocks (thick lines) with 95% Confidence Interval (thin lines)

erogeneous productivity across jobs can be thought of as the different amount of resources under each worker's control, such as the quantity of capital or hierarchy, or different vintages of technology embodied in capital. As in the models of 'span-of-control' (Lucas, 1978) and 'superstar' (Rosen, 1981), production technology determines how much differences in skills are magnified to differences in earnings. In contrast to these models, however, the demand for skills is not infinitely elastic because jobs are in fixed supply, which allows the supply of skills to also affect the return to skills through its effect on the equilibrium assignment of skills and jobs.

Our model is repeated static, where the distributions of skills and jobs exogenously change over time, abstracting from endogenous skill formation and job creation.[13] To focus on the role of skills and returns to skills, we also abstract from elastic labor supply and dynamics of non-skill component of earnings in this theoretical framework.[14]

4.4.1 Assignment Model of Labor Market

Endowment We consider an economy populated by a continuum of measure 1 of workers and jobs operating in different sectors. In each period t, the fraction of workers in sector s is the same as the fraction of jobs in that sector. In each sector, there exist workers endowed with heterogeneous skills $\Theta_t \geq 0$ distributed with cdf $F_t(\Theta_t|s)$ and jobs with heterogeneous productivity $z_t \geq 0$ distributed with cdf $G_t(z_t|s)$.

Technology In each sector, production takes place through one-to-one matching between workers and jobs. If a worker with skill Θ_t works at a job with productivity z_t in sector s, $y_t(s, \Theta_t, z_t) \geq 0$ units of output is produced. We assume that, for each s, $y_t(s, \cdot, \cdot)$ is twice continuously differentiable and strictly increasing, and strictly supermodular:

$$\frac{\partial y_t(s, \Theta_t, z_t)}{\partial \Theta_t \partial z_t} > 0.$$

[13] There is a literature explaining the trends in between- and within-group wage inequality in the framework of endogenous human capital accumulation with heterogeneous learning ability and imperfect substitution of human capital across groups (e.g., Heckman, Lochner, and Taber, 1998, Guvenen and Kuruscu, 2010). This literature does not take into account idiosyncratic earnings shocks that are not related to human capital, equating all changes in residual wage inequality with changes in the inequality of human capital quantity and time spent on training. Using a model that is similar to ours, Jovanovic (1998) studies the effects of technical changes on long-run inequality when the distributions of skills and jobs endogenously evolve over time in a theoretical framework.

[14] Violante (2002) develops a theory of how technical changes may affect transitory shocks. An acceleration of technological progress increases the productivity gaps across jobs, which is translated into more volatile earnings in a frictional labor market.

These assumptions imply that high skill workers are more productive (i.e., they have absolute advantages) than low skill workers at all jobs, but the productivity gap between high and low skill workers is larger at more productive jobs due to complementarity between skills and jobs. Therefore, the efficient assignment that maximizes aggregate output features positive assortative matching where more skilled workers work at more productive jobs (Becker, 1973).

Profit Maximization All markets are perfectly competitive. Workers in sector s with skill Θ_t earn $W_t(s,\Theta_t)$ and output in sector s is sold at price $p_t(s)$. Producers maximize profits, solving

$$\max_{\Theta_t}\left\{p_t(s)y_t(s,\Theta_t,z_t)-W_t(s,\Theta_t)\right\}. \tag{4.13}$$

Denote the solution by $\hat{\Theta}_t(s,z_t)$, which is monotone in z_t due to the supermodularity of $y_t(s,\cdot,\cdot)$. The necessary first order condition is

$$p_t(s)\frac{\partial y_t\left(s,\hat{\Theta}_t(s,z_t),z_t\right)}{\partial\Theta_t}=\frac{\partial W_t\left(s,\hat{\Theta}_t(s,z_t)\right)}{\partial\Theta_t}. \tag{4.14}$$

Market Clearing The labor market clears if the measure of producers demanding skills $\hat{\Theta}_t(s,z_t)$ or less, which equals $G_t(z_t|s)$, coincides with the measure of workers supplying skills $\hat{\Theta}_t(s,z_t)$ or less, which equals $F_t\left(\hat{\Theta}_t(s,z_t)|s\right)$:

$$F_t\left(\hat{\Theta}_t(s,z_t)|s\right)=G_t(z_t|s). \tag{4.15}$$

We also assume that both producers and workers have an option not to operate, in which case they earn zero. Therefore, (4.15) is the the market clearing condition provided that profits and earnings are nonnegative.

Equilibrium Earnings A competitive equilibrium in each sector consists of the assignment rule $\hat{\Theta}_t(s,z_t)$ and the earnings function $W_t(s,\Theta_t)$ that satisfy the first order condition for profit maximization (4.14) and the market clearing condition (4.15). Let $\hat{z}_t(s,\Theta_t)$ be the inverse of $\Theta_t=\hat{\Theta}_t(s,z_t)$ with respect to z_t. Then the equilibrium skill elasticity of earnings can be written as follows:[15]

$$\frac{\partial\ln W_t(s,\Theta_t)}{\partial\ln\Theta_t}=\frac{\partial\ln y_t\left(s,\Theta_t,\hat{z}_t(s,\Theta_t)\right)}{\partial\ln\Theta_t}\Bigg/\frac{W_t(s,\Theta_t)}{p_t(s)y_t\left(s,\Theta_t,\hat{z}_t(s,\Theta_t)\right)}. \tag{4.16}$$

[15]The same condition is derived by Costinot and Vogel (2010) for the case when workers take the whole revenue (i.e., labor share is 1) due to free entry of jobs. In their model, skill demand is still not infinitely elastic because outputs of different jobs are imperfect substitutes.

The numerator is the partial skill elasticity (that is, holding productivity of job constant) of output at the equilibrium assignment and the denominator is the worker's share of revenue, or labor share. The following Proposition shows that, when the production function is Cobb-Douglas and skills and jobs are log-normally distributed, each of these terms is constant for all skill levels and the skill elasticity of earnings, which we call the 'return to skill' $\mu_t(s)$, can be derived in a simple closed form.

Proposition 4.2 *Suppose that (i)* $\ln y_t(s,\Theta_t,z_t) = \lambda_t(s)\ln\Theta_t + \gamma_t(s)\ln z_t$ *and (ii)* Θ_t *and* z_t *are log-normally distributed, conditional on s. Then the labor share in each sector is*

$$\frac{W_t(s,\Theta_t)}{p_t(s)y_t\big(s,\Theta_t,\hat{z}_t(s,\Theta_t)\big)} = \frac{\lambda_t(s)}{\lambda_t(s)+\gamma_t(s)\frac{\sigma(\ln z_t|s)}{\sigma(\ln\Theta_t|s)}} \tag{4.17}$$

and the return to skill in each sector is

$$\mu_t(s) \equiv \frac{\partial \ln W_t(s,\Theta_t)}{\partial \ln\Theta_t} = \lambda_t(s)+\gamma_t(s)\frac{\sigma(\ln z_t|s)}{\sigma(\ln\Theta_t|s)}, \tag{4.18}$$

where $\sigma(x|s)$ *is conditional standard deviation of* x*.*

More skilled workers are more productive at all jobs, which is captured by the partial skill elasticity of output $\lambda_t(s)$. But they also work for jobs with higher productivity, so part of the productivity increase comes from the productivity of jobs. Taking into account this sorting effect, the total skill elasticity of output is

$$\frac{\partial \ln y_t\big(s,\Theta_t,\hat{z}_t(s,\Theta_t)\big)}{\partial \ln\Theta_t} + \frac{\partial \ln y_t\big(s,\Theta_t,\hat{z}_t(s,\Theta_t)\big)}{\partial \ln z_t}\frac{\partial \ln \hat{z}_t(s,\Theta_t)}{\partial \ln\Theta_t} = \lambda_t(s)+\gamma_t(s)\frac{\partial \ln \hat{z}_t(s,\Theta_t)}{\partial \ln\Theta_t}. \tag{4.19}$$

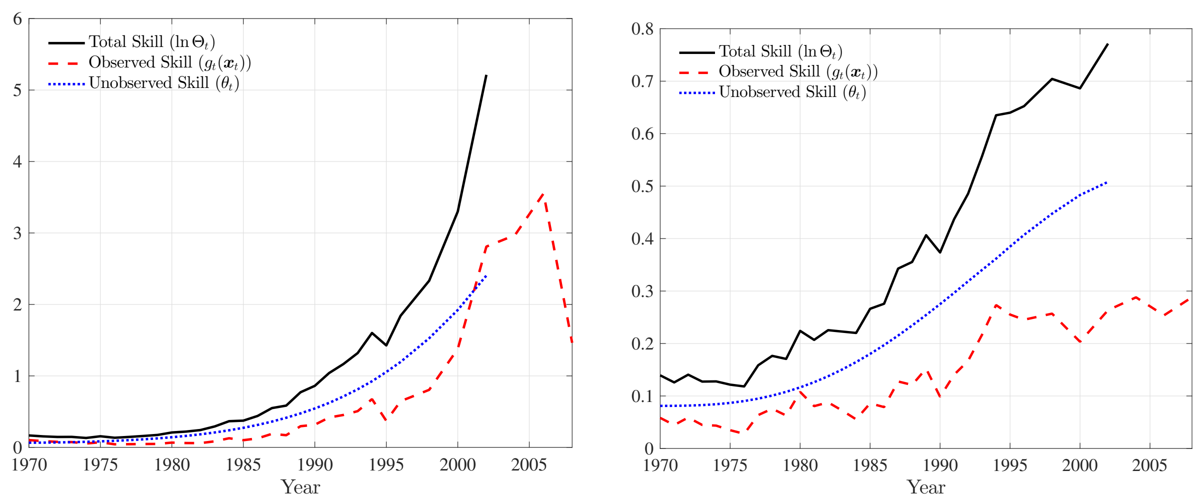
With the log-normality assumption, the market clearing condition (4.15) implies

$$\frac{\partial \ln \hat{z}_t(s,\Theta_t)}{\partial \ln\Theta_t} = \frac{\sigma(\ln z_t|s)}{\sigma(\ln\Theta_t|s)}, \tag{4.20}$$

which shows that relative heterogeneity between workers and jobs is important for the equilibrium assignment. When workers are relatively homogeneous compared to jobs, a slightly more skilled worker is assigned to a much more productive job. This generates a large total skill elasticity of output, amplifying the innate productivity differences across workers. On the other hand, when jobs are relatively homogeneous compared to workers, the sorting effect is small, and the total skill elasticity of output mainly reflects the skill differences.

Equation (4.17) shows that the labor share is determined by the fraction of the worker's marginal contribution, partial skill elasticity of output, out of the sum of worker's and job's marginal contribution to output. By combining (4.16) and (4.17), we get (4.18), which equates the return to skill with the total skill elasticity of output.[16] The return to skill is determined by production technology $\lambda_t(s)$ and $\gamma_t(s)$, distribution of jobs $\sigma(\ln z_t|s)$, and distribution of skills $\sigma(\ln \Theta_t|s)$. An increase in the demand for high skills—an increase in $\lambda_t(s)$ and/or $\gamma_t(s)\sigma(\ln z_t|s)$—can be considered a skill-biased technical change in the sense that it increases the return to skills and widens residual earnings inequality. On the other hand, an increase in the dispersion of skills $\sigma(\ln \Theta_t|s)$ reflects a skill supply change that reduces the return to skills and residual earnings inequality.[17] In the next subsection, we discuss how we can decompose the changes in the returns to skills into the changes in each factor.

4.4.2 Recovering Supply and Demand Factors



(a) High School (b) College

Figure 4.9: Variance of Observed and Unobserved Skills

Variance of Log Skill We assume that skills are correlated with the workers' observed characteristics but are not fully observed. Specifically, log skill of individual i in year t is

$$\ln \Theta_{i,t} = g_t(\boldsymbol{x}_{i,t}) + \theta_{i,t}.$$

[16]To our knowledge, the formula (4.18) is novel. Gabaix and Landier (2008) derived a similar formula with Pareto distribution.

[17]Note that all of these factors, in addition to output price $p_t(s)$ and the average skill and job productivity $\mathbb{E}[\ln \Theta_t|s]$ and $\mathbb{E}[\ln z_t|s]$ of each sector, also affect between-group inequality (i.e., average earnings differential across sectors). This raises a possibility that between- and within-group inequality are driven by a common factor, which may affect them differently.

From (4.18), the equilibrium earnings function can be written as $\ln W_t(s,\Theta_t) = \alpha_t(s) + \mu_t(s)\ln\Theta_t$ for some constant $\alpha_t(s)$. Assuming that individual log earnings $\ln W_{i,t}$ is the sum of a skill-related component that reflects the equilibrium earnings function and a non-skill component of earnings $u_{i,t}$,

$$\ln W_{i,t} = \alpha_t(s_i) + \mu_t(s_i)\ln\Theta_{i,t} + u_{i,t} = \underbrace{\alpha_t(s_i) + \mu_t(s_i)g(\boldsymbol{x}_{i,t})}_{\equiv f_t(\boldsymbol{x}_{i,t})} + \underbrace{\mu_t(s_i)\theta_{i,t} + u_{i,t}}_{= w_{i,t}}$$

gives the first stage regression equation (4.12). Although the time- and sector-specific constant $\alpha_t(s)$ is not separately identified from the location of the function $g_t(\boldsymbol{x}_t)$, the variance of log skill is identified by taking variance conditional on s,

$$\mathrm{Var}(\ln\Theta_t|s) = \frac{\mathrm{Var}(\ln W_t|s) - \mathrm{Var}(u_t|s)}{\mu_t(s)^2}.$$

Therefore, the second stage estimates of $\mu_t(s)$ and $\mathrm{Var}(u_t|s)$ can be used to compute the variances of log skills in each period except for the last $k = q+2$ periods when $\mathrm{Var}(u_t|s)$ is not identified. Figure 4.9 shows that the variances of observed and unobserved skills increased over time, and the increase was much larger for high school workers.

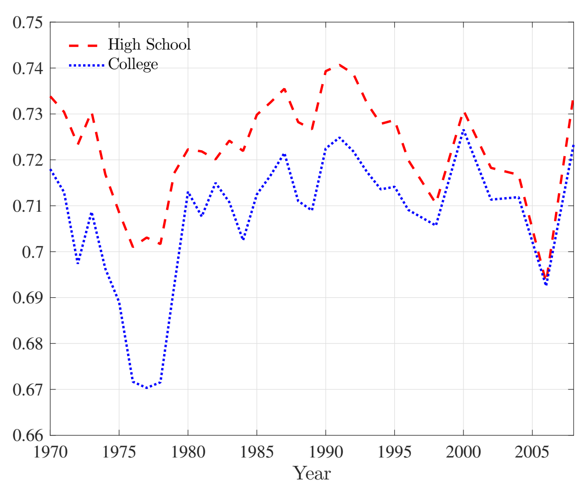
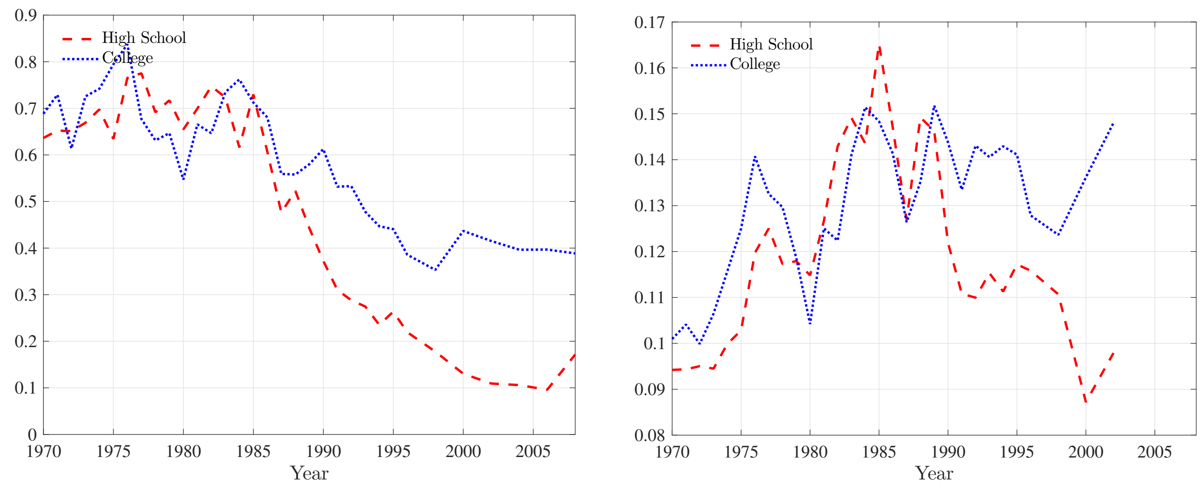


Figure 4.10: Average Labor Share

Production Function Parameters By combining (4.17) and (4.18), the labor share can be written as $\lambda_t(s)/\mu_t(s)$. Therefore, the output elasticity of skill $\lambda_t(s)$ can be identified from $\mu_t(s)$ and the labor share of each sector. Then we can also get $\gamma_t(s)\sigma(\ln z_t|s) = \sigma(\ln\Theta_t|s)(\mu_t(s) - \lambda_t(s))$, but we cannot separately identify $\gamma_t(s)$ and $\sigma(\ln z_t|s)$. Recovering the output elasticity from factor income share is a commonly used method since Cobb and Douglas (1928), but



(a) $\lambda_t(s)$ (b) $\gamma_t(s)\sigma(\ln z_t|s)$

Figure 4.11: Production Function Parameters

implementing it in this case is complicated by the need to know the sector-specific labor shares. The difficulty arises from the fact that data on value added by workers with different levels of education is not available.

To overcome this difficulty, we assume that the labor share in our model can be approximated by average industry-level labor share for workers in each sector. Suppose that there are $j = 1,\ldots,J$ industries with value added $V_{j,t}$ and labor compensation $L_{j,t}$ in year t. Let $E_{j,t}(s)$ be the number of workers in industry j, sector s, and year t, and let $E_t(s) = \sum_{j=1}^{J} E_{j,t}(s)$ be the total number of workers in all industries. Then the average labor share in sector s and year t is

$$\sum_{j=1}^{J} \frac{E_{j,t}(s)}{E_t(s)} \frac{L_{j,t}}{V_{j,t}}.$$

We use data on labor share and number of high school and college workers in the U.S. by 31 ISIC[18] industries from the World KLEMS dataset.[19] The number of workers is computed by counting male workers ages 25-64. Figure 4.10 shows the average labor share by sector. Overall, average labor shares in the two sectors tend to comove. The average labor share is lower for college workers until the late 1990s, but the gap is closed afterwards. The low average labor share for college workers means that college workers are more likely to work in industries with low labor share compared to high school workers. In our theoretical framework, the lower labor share in the college sector means that college workers enjoy a relatively

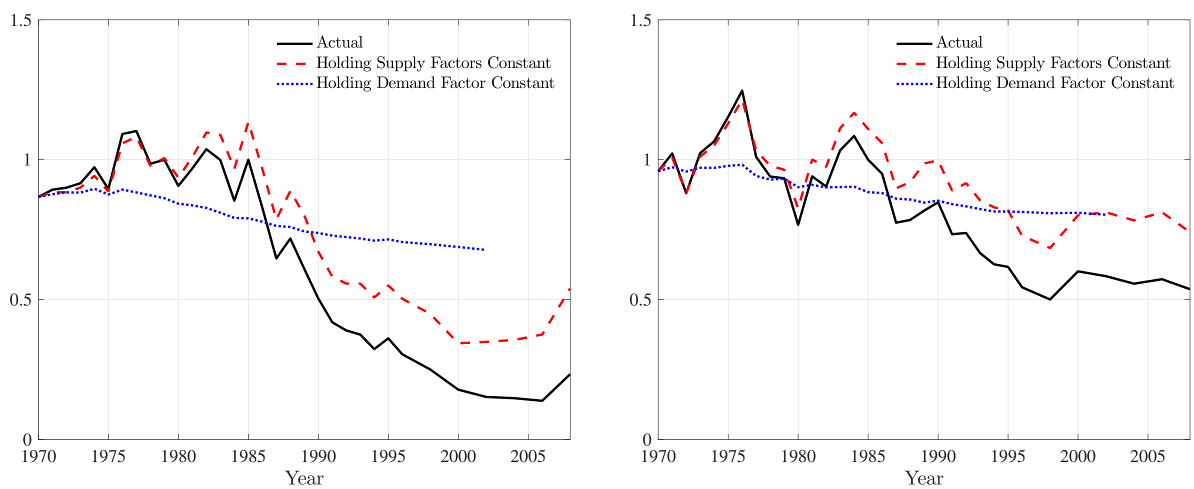
[18]The International Standard Industrial Classification of All Economic Activities (ISIC) is a United Nations industry classification system.

[19]Available at http://www.worldklems.net. See Jorgenson, Ho, and Samuels (2012) for description.

larger productivity gain from jobs (i.e., $\lambda_t(s)$ is smaller compared to $\gamma_t(s)\sigma(\ln z_t|s)/\sigma(\ln\Theta_t|s)$) compared to high school workers.

Figure 4.11 shows the production function parameters calculated from the average labor share and the returns to skills. The time patterns of the partial skill elasticity of output $\lambda_t(s)$ are similar to those of the returns to skills $\mu_t(s)$: they are stable until 1985, decrease afterwards, and the decline is larger for high school workers. The productivity differential across jobs $\gamma_t(s)\sigma(\ln z_t|s)$, however, increases until the mid-1980s and is stable afterwards for college workers. For high school workers, $\gamma_t(s)\sigma(\ln z_t|s)$ decreases between the mid-1980s and 2000, showing divergence from that of college workers in the early and late 1990s.

4.4.3 Decomposing the Returns to Skills



(a) High School (b) College

Figure 4.12: Actual and Counterfactual Returns ($\mu_t(s)$)

We assess the contributions of each factor recovered in the previous subsection in the evolution of the returns to skills. Figure 4.12 shows the actual returns as well as counterfactual returns when demand factors ($\lambda_t(s)$ and $\gamma_t(s)\sigma(\ln z_t|s)$) and supply factor ($\sigma(\ln\Theta_t|s)$) are held constant at their values in 1970. Both demand and supply factors play important roles in the evolution of the returns to skills in the sense that each factor alone can explain the general decreasing trend of the returns to skills. Rising variance of observed and unobserved skills over time reduced the returns to skills by lowering productivity differentials across jobs among workers with different skills. For high school workers, the demand for skills modestly increased until 1985, which mainly reflects the change in the productivity of jobs ($\gamma_t(s)\sigma(\ln z_t|s)$) rather than the productivity of skills ($\lambda_t(s)$). The demand for skills for high school workers undergoes a steep decline between 1985 and 2000, explaining most of the decline in the returns to skills during that period, then slightly increases afterwards. For college workers, the decline of the demand for skills was not as drastic as that for high school workers, explaining about half of the decrease in the return to skills between 1970 and 2000.

4.5 Conclusion

In this chapter, we show that the returns to unobserved skills and the variance of unobserved skills are separately identified from time-varying variances of non-skill earnings shocks. Using panel data from the PSID on male earnings in the U.S. from 1970-2008, we show that accounting for changes in the distributions of skills and the volatility of earnings is important for the estimation of the evolution of the returns to unobserved skills. The time patterns of the estimated returns to unobserved skills are different from those estimated in the literature. The returns to unobserved skills were stable in the 1970s and early 1980s, but they sharply decreased after 1985, and the decline was more sizable for high school workers than for college workers. Unobserved skills explain substantial and increasing share of the total variance of log annual earnings residual, but their importance declined after the mid-1980s despite the continuous increase in the variance of unobserved skills. Much of the increase in the residual variance after the late 1990s is explained by the rise in the variance of permanent non-skill component of earnings.

Using a simple demand and supply framework, we show that the decrease in the returns to unobserved skills after the mid-1980s is driven by both supply and demand factors. The decline in the demand for skills was more drastic for high school workers than for college workers. Our finding that the demand for skills decreased since the mid-1980s seem to reinforce the challenges (e.g., Card and DiNardo, 2002) faced by the skill-biased technical change (SBTC) hypothesis, which typically states that the steady increase in the demand for skills explains increasing inequality. Further work is required to understand why the skill demand declined during the 1990s, precisely when technological progress had accelerated (Cummins and Violante, 2002).

Bibliography

Abowd, John M. and David Card. 1989. “On the Covariance Structure of Earnings and Hours Changes.” *Econometrica* 57 (2):411–445.

Acemoglu, Daron. 2002. "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature* 40 (1):7–72.

Acemoglu, Daron and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*, vol. 4, edited by O. Ashenfelter and D. Card, chap. 12. Elsevier, 1043–1171.

Ales, Laurence, Musab Kurnaz, and Christopher Sleet. 2015. "Technical Change, Wage Inequality, and Taxes." *American Economic Review* 105 (10):3061–3101.

Autor, David and David Dorn. 2012. "The Growth of Low Skill Service Jobs and the Polarization of the U.S. Labor Market." IZA Discussion Papers 7068, Institute for the Study of Labor (IZA).

Autor, David, Frank Levy, and Richard Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118 (4):1279–1333.

Autor, David H, Lawrence F Katz, and Melissa S Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *The Review of Economics and Statistics* 90 (2):300–323.

Baker, Michael and Gary Solon. 2003. "Earnings Dynamics and Inequality among Canadian Men, 19761992: Evidence from Longitudinal Income Tax Records." *Journal of Labor Economics* 21 (2):289–321.

Becker, Gary S. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy* 81 (4):pp. 813–846.

Blundell, Richard, Michael Graber, and Magne Mogstad. 2015. "Labor Income Dynamics and the Insurance from Taxes, Transfers, and the Family." *Journal of Public Economics* 127:58 – 73.

Blundell, Richard, Luigi Pistaferri, and Ian Preston. 2008. "Consumption Inequality and Partial Insurance." *The American Economic Review* 98 (5):pp. 1887–1921.

Blundell, Richard and Ian Preston. 1998. "Consumption Inequality and Income Uncertainty." *The Quarterly Journal of Economics* 113 (2):603–640.

Burstein, Ariel, Eduardo Morales, and Jonathan Vogel. 2015. "Accounting for Changes in Between-Group Inequality." Working Paper 20855, National Bureau of Economic Research.

Card, David and John E. DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics* 20 (4):733–783.

Carroll, Christopher D. 1992. "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence." *Brookings Papers on Economic Activity* 1992 (2):61–156.

Castex, Gonzalo and Evgenia Kogan Dechter. 2014. "The Changing Roles of Education and Ability in Wage Determination." *Journal of Labor Economics* 32 (4):685–710.

Cobb, Charles W and Paul H Douglas. 1928. "A Theory of Production." *The American Economic Review* 18 (1):139–165.

Costinot, Arnaud and Jonathan Vogel. 2010. "Matching and Inequality in the World Economy." *Journal of Political Economy* 118 (4):pp. 747–786.

Cummins, Jason G. and Giovanni L. Violante. 2002. "Investment-Specific Technical Change in the United States (19472000): Measurement and Macroeconomic Consequences." *Review of Economic Dynamics* 5 (2):243 – 284.

DeBacker, Jason, Bradley Heim, Vasia Panousi, Shanthi Ramnath, and Ivan Vidangos. 2013. "Rising inequality: transitory or persistent? New evidence from a panel of US tax returns." *Brookings Papers on Economic Activity* 2013 (1):67–142.

Friedman, Milton and Simon Kuznets. 1954. *Income from Independent Professional Practice*. National Bureau of Economic Research.

Gabaix, Xavier and Augustin Landier. 2008. "Why has CEO Pay Increased So Much?" *The Quarterly Journal of Economics* 123 (1):49–100.

Gottschalk, Peter and Robert Moffitt. 1994. "The Growth of Earnings Instability in the U.S. Labor Market." *Brookings Papers on Economic Activity* 25 (2):217–272.

Guiso, Luigi, Luigi Pistaferri, and Fabiano Schivardi. 2005. "Insurance within the Firm." *Journal of Political Economy* 113 (5):1054–1087.

Guvenen, Fatih and Burhanettin Kuruscu. 2010. "A Quantitative Analysis of the Evolution of the US Wage Distribution, 1970-2000." In *NBER Macroeconomics Annual 2009, Volume 24*. University of Chicago Press, 227–276.

Haider, Steven J. 2001. "Earnings Instability and Earnings Inequality of Males in the United States: 19671991." *Journal of Labor Economics* 19 (4):799–836.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante. 2010. "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 19672006." *Review of Economic Dynamics* 13 (1):15 – 51. Special issue: Cross-Sectional Facts for Macroeconomists.

Heckman, James J, Lance Lochner, and Christopher Taber. 1998. "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents." *Review of Economic Dynamics* 1 (1):1 – 58.

Jorgenson, Dale W, Mun S Ho, and Jon D Samuels. 2012. "A Prototype Industry-Level Production Account for the United States, 1947-2010." Second World KLEMS Conference, Harvard University, August 9, 2012.

Jovanovic, Boyan. 1998. "Vintage Capital and Inequality." *Review of Economic Dynamics* 1 (2):497 – 530.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy* 101 (3):410–42.

Katz, Lawrence F. and David H. Autor. 1999. "Changes in the Wage Structure and Earnings Inequality." In *Handbook of Labor Economics*, *Handbook of Labor Economics*, vol. 3, edited by O. Ashenfelter and D. Card, chap. 26. Elsevier, 1463–1555.

Katz, Lawrence F. and Kevin M. Murphy. 1992. "Changes in Relative Wages, 19631987: Supply and Demand Factors." *The Quarterly Journal of Economics* 107 (1):35–78.

Lamadon, Thibaut. 2014. "Productivity Shocks, Dynamic Contracts and Income Uncertainty." *Working Paper* .

Lemieux, Thomas. 2006. "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *The American Economic Review* 96 (3):461–498.

Lindenlaub, Ilse. 2014. "Sorting Multidimensional Types: Theory and Application." *Working Paper* .

Lochner, Lance and Youngki Shin. 2014. "Understanding Earnings Dynamics: Identifying and Estimating the Changing Roles of Unobserved Ability, Permanent and Transitory Shocks." Working Paper 20068, National Bureau of Economic Research.

Lucas, Robert E. 1978. "On the Size Distribution of Business Firms." *The Bell Journal of Economics* 9 (2):508–523.

MaCurdy, Thomas. 2007. “A Practitioner’s Approach to Estimating Intertemporal Relationships Using Longitudinal Data: Lessons from Applications in Wage Dynamics.” In *Handbook of Econometrics*, *Handbook of Econometrics*, vol. 6, Part A, edited by James J. Heckman and Edward E. Leamer, chap. 62. Elsevier, 4057 – 4167.

Meghir, Costas and Luigi Pistaferri. 2004. “Income Variance Dynamics and Heterogeneity.” *Econometrica* 72 (1):1–32.

Moffitt, Robert A. and Peter Gottschalk. 2012. “Trends in the Transitory Variance of Male Earnings Methods and Evidence.” *Journal of Human Resources* 47 (1):204–236.

Rosen, Sherwin. 1981. “The Economics of Superstars.” *The American Economic Review* 71 (5):845–858.

———. 1986. “The Theory of Equalizing Differences.” In *Handbook of Labor Economics*, *Handbook of Labor Economics*, vol. 1, chap. 12. Elsevier, 641 – 692.

Sattinger, Michael. 1979. “Differential Rents and the Distribution of Earnings.” *Oxford Economic Papers* 31 (1):60–71.

———. 1993. “Assignment Models of the Distribution of Earnings.” *Journal of Economic Literature* 31 (2):831–880.

Scheuer, Florian and Iván Werning. 2015. “The Taxation of Superstars.” Working Paper 21323, National Bureau of Economic Research.

Terviö, Marko. 2008. “The Difference That CEOs Make: An Assignment Model Approach.” *The American Economic Review* 98 (3):642–668.

Violante, Giovanni L. 2002. “Technological Acceleration, Skill Transferability, and the Rise in Residual Inequality.” *The Quarterly Journal of Economics* 117 (1):297–338.

Chapter 5

Conclusion

This thesis is concerned with the sources of earnings inequality and their policy implications. Differences in early human capital investment across family income is one of the most important determinant in life-time earnings inequality. Chapter 2 sorts amongst different potential explanations by comparing their predictions with stylized facts on child development. We find that only the explanation based on credit constraints—children cannot invest themselves by borrowing against their or their parents' future income—is consistent with all facts, suggesting that credit constraints can be an important source of inequality in human capital investment.

One way to address the inefficiency in human capital investment due to credit constraints is to redistribute towards constrained students. The federal need-based financial aid for higher education in the U.S. aims to achieve this goal by providing larger amount of aid for students from lower income families. However, Chapter 3 shows that such policy cannot specifically target constrained students because parents with similar financial resources give very different amounts to their children. I find that the unobservable heterogeneity in parental transfers, which I model as heterogeneous parental altruism, reduces efficiency of the education subsidy program and limits redistribution across families with different income. These findings suggest that it is important to investigate the sources of unobservable heterogeneity across students when designing the financial aid policy.

The inequality in skills due to market frictions such as borrowing constraints will have more important consequences if differences in skills are translated into larger differences in earnings. Chapter 4 shows that the returns to unobserved skills fell in the U.S. since the mid-1980s despite continuing rise in the overall residual earnings inequality, and that this reflects a decline in the demand for skills. Our findings are not consistent with existing theories of skill-biased technical change. Reconciling this inconsistency is an important avenue for future research.

Appendix A

Chapter 2 Appendix

A.1 Factor Score Weights

We employ principal factor analysis using the measured inputs reported in Figures 2.2 and 2.3 (separately by age) to create age-specific predicted investment factor scores for each child using the Thomson (1951) method. Estimated weights for each factor used in constructing the scores are reported in Table A.1. For interpretation purposes, scores are normalized to have a mean of zero and standard deviation of one.

Table A.1: Weights used to Construct Factor Scores

	Age Group			
Early Investment Measure	0-1	2-3	4-5	6-7
10+ Books at Home	0.33	0.24	0.21	0.12
Mom Reads 3+ Times/Week	0.32	0.26	0.22	0.09
Child Eats with Mom and Dad Daily	0.08	0.16	0.19	0.03
Child Leaves House 4+ Times/Week	0.18	0.14	0.15	
Child Sees Father Daily	0.09	0.20	0.24	
Musical Instrument in Home				0.10
Child Taken to a Performance in Past Year				0.18
Child Taken to a Museum in Past Year				0.15
Child Takes Special Lessons/Extracurricular Activities				0.14
Family Gets a Daily Newspaper				0.08
Family Encourages Hobbies				0.08
Family Meets Friends/Relatives 2+ Times/Month				0.02

A.2 Proofs and Technical Results

A.2.1 Proof of Proposition 2.1

Part (i): Follows directly from the first order conditions (2.6) and (2.7).

Parts (ii) and (iii): Applying Cramer's rule to first order conditions (2.6) and (2.7) yields

$$\begin{aligned}\frac{di_1}{d\theta} &= \frac{f_2(i_1,i_2)f_{12}(i_1,i_2)-f_1(i_1,i_2)f_{22}(i_1,i_2)}{\theta[f_{11}f_{22}-f_{12}^2]}>0\\ \frac{di_2}{d\theta} &= \frac{f_1(i_1,i_2)f_{12}(i_1,i_2)-f_2(i_1,i_2)f_{11}(i_1,i_2)}{\theta[f_{11}f_{22}-f_{12}^2]}>0,\end{aligned}$$

where these effects can be signed using Assumption 1. These imply that the final human capital level, h_3, is increasing in θ.

A.2.2 Proof of Proposition 2.2

Part (i): The difference between marginal returns to early investment and the return to savings is immediate from Equation (2.8).

Consider the effects of parental income on the marginal value of investment, $\frac{\partial h_3}{\partial i_1}=\theta f_1(i_1,i_2)$. Differentiating $f_1(i_1,i_2)$ with respect to Y at an optimum yields:

$$\begin{aligned}\frac{df_1(i_1,i_2)}{dY} &= f_{11}\frac{di_1}{dY}+f_{12}\frac{di_2}{dY}\\ &= \left[\frac{\nu u''(c)}{[u'(c)]^2B}\right]\left[f_{11}(i_1,i_2)f_{22}(i_1,i_2)-f_{12}^2(i_1,i_2)\right]/\Delta,\end{aligned}$$

where $\Delta=\left[\frac{\nu^2u''(c)}{[u'(c)]^3B}\right]f_{22}+\beta^2\theta[f_{11}f_{22}-f_{12}^2]>0$ by Assumption 1. $\frac{df_1(i_1,i_2)}{dY}$ has the opposite sign as ν, because the second term in brackets is positive by Assumption 2.1.

Parts (ii) and (iii): We can determine the effects of Y on child investments from first order conditions (2.7) and (2.8) using Cramer's rule:

$$\begin{aligned}\frac{di_1}{dY} &= \left[\frac{\nu u''(c)}{[u'(c)]^2B}\right]f_{22}(i_1,i_2)/\Delta\\ \frac{di_2}{dY} &= \left[\frac{-\nu u''(c)}{[u'(c)]^2B}\right]f_{12}(i_1,i_2)/\Delta.\end{aligned}$$

If investment has no consumption value (i.e. $\nu=0$), then investments are independent of parental income Y. Otherwise, the effects of parental income on investments depend on whether investment has a positive (i.e. $\nu > 0$) or negative (i.e. $\nu < 0$) consumption value and, in the case of second period investment, the complementarity of early and late investments $f_{12}(i_1, i_2)$.

Part (iv): The effect of parental income on final human capital levels is given by

$$\begin{aligned}\frac{dh_3}{dY} &= \theta\left[f_1\frac{di_1}{dY} + f_2\frac{di_2}{dY}\right] \\ &= \theta\left[\frac{\nu u''(c)}{[u'(c)]^2 B}\right][f_1(i_1,i_2)f_{22}(i_1,i_2) - f_2(i_1,i_2)f_{12}(i_1,i_2)]/\Delta,\end{aligned}$$

which has the same sign as ν, because the second term in brackets is negative by Assumption 2.1.

A.2.3 Proof of Proposition 2.3

When the return to investment is realized after all investments have been made, we can combine the period 1 and 2 budget constraints:

$$\underbrace{(c_1 + R^{-1}c_2)}_{\equiv c} + \underbrace{(i_1 + R^{-1}i_2)}_{\equiv e} + R^{-1}a_3 \leq Y.$$

Notice that only discounted lifetime income Y enters the budget constraint, so the timing of income, y_1 and y_2, is irrelevant.

The problem can be decomposed into 3 pieces: investment allocation between periods 1 and 2 (choosing i_1 and i_2 given e), consumption allocation between periods 1 and 2 (choosing c_1 and c_2 given c), and portfolio choice between human capital investment and savings (choosing c, e, and a_3). In proving our results, we first show how i_1 and i_2 (and their marginal returns) depend on the level of total investment spending $e = i_1 + R^{-1}i_2$ (and its marginal return) and then focus on properties of e and its marginal return.

First, consider the investment problem for a given level of total investment expenditure e. Let $g(e)$ be the maximum $f(i_1, i_2)$ produced by spending e:

$$g(e) \equiv \max_{i_1,i_2}\left\{f(i_1,i_2)\middle| i_1 + R^{-1}i_2 \leq e\right\}.$$

Note that $g(\cdot)$ is strictly increasing and strictly concave because $f(\cdot)$ is strictly increasing, strictly concave, and the constraint set is convex. The first order conditions for i_1 and i_2, combined with the envelope condition, are

$$g'(e) = f_1(i_1,i_2) = Rf_2(i_1,i_2), \quad i_1 + R^{-1}i_2 = e. \tag{A.1}$$

Applying the Implicit Function Theorem, we can determine the effects of e on investments:

$$\frac{di_1}{de} = \frac{f_{12} - \frac{f_1}{f_2} f_{22}}{2 f_{12} - \frac{f_2}{f_1} f_{11} - \frac{f_1}{f_2} f_{22}}, \quad \frac{di_2}{de} = \frac{\frac{f_1}{f_2}\left(f_{12} - \frac{f_2}{f_1} f_{11}\right)}{2 f_{12} - \frac{f_2}{f_1} f_{11} - \frac{f_1}{f_2} f_{22}},$$

where we use the first order condition $R = f_1/f_2$. Both derivatives are strictly positive by Assumption 2.1, so spending more on overall education means that both early and late investments increase. We state this result as a Lemma for future reference.

Lemma A.1 *Both early and late investments are strictly increasing in total investment spending if they are optimally chosen to maximize human capital.*

Next, we consider the intertemporal consumption allocation problem. Define the indirect utility of total consumption spending c on consumption in periods 1 and 2:

$$U(c) \equiv \max_{c_1, c_2} \left\{ u(c_1) + \beta u(c_2) \middle| c_1 + R^{-1} c_2 \leq c \right\}.$$

Note that $U(\cdot)$ is strictly increasing and strictly concave, because $u(\cdot)$ is strictly increasing, strictly concave, and the constraint set is convex.

Using the indirect functions $U(\cdot)$ and $g(\cdot)$, we can write the saving and human capital investment problem as follows:

$$\max_{e,s} \left\{ U(Y - s) + \beta^2 \mathbb{E}\left[u\big(R^2(s-e) + \theta g(e)\big) \right] \right\},$$

where $s = e + R^{-1} a_3$ is total savings.

The first order conditions are

$$U'(c) = (R\beta)^2 \mathbb{E}\left[u'(c_3(\theta))\right] = \beta^2 \mathbb{E}\left[u'(c_3(\theta))\theta\right] g'(e), \tag{A.2}$$

where $c_3(\theta) = R^2(s-e) + \theta g(e)$.

Part (i): Combining the first order conditions in (A.2), we get

$$R^2 = \frac{\mathbb{E}\left[u'(c_3(\theta))\theta\right]}{\mathbb{E}\left[u'(c_3(\theta))\right]} g'(e) = \bar{\theta} g'(e) + \frac{\operatorname{Cov}\left(u'(c_3(\theta)), \theta\right)}{\mathbb{E}\left[u'(c_3(\theta))\right]} g'(e).$$

Because $c_3(\theta)$ is strictly increasing in θ and $u(\cdot)$ is strictly concave, the covariance term is strictly negative. This, combined with Equation (A.1), implies that the expected marginal returns to i_1 and i_2 are greater than the returns on savings.

Parts (ii) and (iii): We first study how the portfolio choice between the safe asset a_3 and the risky human capital investment e is affected by total savings s. Specifically, we show that optimal total investment e is increasing in total savings s under decreasing absolute risk aversion. We then show that total savings s is increasing in lifetime parental income Y.

Optimal human capital investment for a given s solves

$$\hat{e}(s) \equiv \underset{e}{\operatorname{argmax}}\, \mathbb{E}\left[u\big(R^2(s-e)+\theta g(e)\big)\right].$$

Because the objective function is strictly concave in e, the Implicit Function Theorem reveals that the sign of $\hat{e}'(s)$ is equal to the sign of

$$\frac{\partial^2}{\partial e \partial s}\left(\mathbb{E}\left[u\big(R^2(s-e)+\theta g(e)\big)\right]\right)\Big|_{e=\hat{e}(s)} = \mathbb{E}\left[u''\big(R^2(s-\hat{e}(s))+\theta g(\hat{e}(s))\big)\big(-R^2+\theta g'(\hat{e}(s))\big)R^2\right].$$

With decreasing absolute risk aversion, this term is strictly positive and $\hat{e}'(s) > 0$. To see this, let $\hat{\theta} = R^2/g'\big(\hat{e}(s)\big)$, so

$$\begin{aligned}
&\mathbb{E}\left[u''\big(R^2(s-\hat{e}(s))+\theta g(\hat{e}(s))\big)\big(-R^2+\theta g'(\hat{e}(s))\big)R^2\right] \\
&= R^2 \int \left[\frac{u''\big(R^2(s-\hat{e}(s))+\theta g(\hat{e}(s))\big)}{u'\big(R^2(s-\hat{e}(s))+\theta g(\hat{e}(s))\big)}\right] u'\big(R^2(s-\hat{e}(s))+\theta g(\hat{e}(s))\big)\big(-R^2+\theta g'(\hat{e}(s))\big)d\Phi(\theta) \\
&> R^2 \left[\frac{u''\big(R^2(s-\hat{e}(s))+\hat{\theta} g(\hat{e}(s))\big)}{u'\big(R^2(s-\hat{e}(s))+\hat{\theta} g(\hat{e}(s))\big)}\right]\left\{\int_{\theta \le \hat{\theta}} u'\big(R^2(s-\hat{e}(s))+\theta g(\hat{e}(s))\big)\big(-R^2+\theta g'(\hat{e}(s))\big)d\Phi(\theta)\right. \\
&\quad \left. + \int_{\theta \ge \hat{\theta}} u'\big(R^2(s-\hat{e}(s))+\theta g(\hat{e}(s))\big)\big(-R^2+\theta g'(\hat{e}(s))\big)d\Phi(\theta)\right\} \\
&= R^2 \left[\frac{u''\big(R^2(s-\hat{e}(s))+\hat{\theta} g(\hat{e}(s))\big)}{u'\big(R^2(s-\hat{e}(s))+\hat{\theta} g(\hat{e}(s))\big)}\right] \mathbb{E}\left[u'\big(R^2(s-\hat{e}(s))+\theta g(\hat{e}(s))\big)\big(-R^2+\theta g'(\hat{e}(s))\big)\right] = 0,
\end{aligned}$$

where the inequality used the assumption that $u''(\cdot)/u'(\cdot)$ is increasing.

Next, we show that total savings s is increasing with Y. Define the value of total savings s:

$$\mathscr{V}(s) \equiv \mathbb{E}\left[\max_{e,a}\left\{u\big(R^2 a+\theta g(e)\big)\Big| e+a \le s\right\}\right],$$

and let $\hat{a}(s) \equiv s - \hat{e}(s)$ reflect the optimal investment in the safe asset conditional on s. $\mathscr{V}(s)$ is strictly increasing and strictly concave, because $u(\cdot)$ is strictly increasing and strictly concave in (a,e) and the constraint set is convex.

Optimal total savings $\hat{s}(Y)$ for a given income Y solves

$$\hat{s}(Y) \equiv \underset{s}{\operatorname{argmax}} \left\{ U(Y-s) + \beta^2 \mathscr{V}(s) \right\}.$$

Applying the Implicit Function Theorem to the first order condition for this problem yields

$$\hat{s}'(Y) = \frac{U''\big(Y - \hat{s}(Y)\big)}{U''\big(Y - \hat{s}(Y)\big) + \beta^2 \mathscr{V}''\big(\hat{s}(Y)\big)} > 0,$$

because both $U(\cdot)$ and $\mathscr{V}(\cdot)$ are strictly concave.

With decreasing absolute risk aversion, both $\hat{s}(Y)$ and $\hat{e}(s)$ are strictly increasing, so e is strictly increasing in Y. From (A.1), expected marginal returns to investment are equal to the expected marginal return to total investment spending: $\bar{\theta} f_1(i_1,i_2) = \bar{\theta} R f_2(i_1,i_2) = \bar{\theta} g'(e)$. This implies that the expected marginal returns to investment are strictly decreasing in Y (with decreasing absolute risk aversion), because $g(\cdot)$ is strictly concave.

A.2.4 Proof of Lemma 2.1

We first show the properties of $\hat{i}_2(zi_1,\theta)$. The first order condition is:

$$\theta f_2\big(zi_1, \hat{i}_2(zi_1,\theta)\big) = R.$$

By applying the Implicit Function Theorem, we get the partial derivatives:

$$\frac{\partial \hat{i}_2(zi_1,\theta,z)}{\partial \theta} = \frac{f_2}{-\theta f_{22}} > 0, \quad \frac{\partial \hat{i}_2(zi_1,\theta,z)}{\partial (zi_1)} = \frac{f_{12}}{-f_{22}} \geq 0 \text{ if and only if } f_{12} \geq 0. \tag{A.3}$$

Then the partial derivatives of $\Pi(i_1,\theta,z)$ are

$$\begin{aligned}
\frac{\partial \Pi(i_1,\theta,z)}{\partial i_1} =& -1 + R^{-2} z \theta f_1, \\
\frac{\partial^2 \Pi(i_1,\theta,z)}{\partial i_1^2} =& R^{-2} z \theta \left(f_{11} z + f_{12} \frac{\partial \hat{i}_2(zi_1,\theta)}{\partial i_1} \right) = -R^{-2} z^2 \theta \left(\frac{f_{11} f_{22} - f_{12}^2}{-f_{22}} \right) < 0.
\end{aligned}$$

A.2.5 Proof of Lemma 2.2

The cross-partial derivative is

$$\frac{\partial^2 \Pi(i_1,\theta,z)}{\partial i_1 \partial z} = R^{-2} \theta \left(f_1 + z f_{11} i_1 + z f_{12} \frac{\partial \hat{i}_2(zi_1,\theta)}{\partial z} \right) = R^{-2} \theta \left\{ f_1 - \left(\frac{f_{11} f_{22} - f_{12}^2}{-f_{22}} \right) z i_1 \right\},$$

which is positive if and only if Condition 2.1 holds.

A.2.6 Proof of Proposition 2.4

See Hartman (1976) for proof.

A.2.7 Proof of Proposition 2.5

Part (i): A first order stochastic dominance shift in z will increase $\tilde{i}_1$ if $\partial\Pi(i_1,\theta,z)/\partial i_1$ is increasing in z. For the CES production function (2.3), the sign of $\partial^2\Pi(i_1,z)/(\partial i_1 \partial z)$ is determined by the sign of

$$b\left(\frac{1-d}{1-b}\right)(1-a)^{1-b}(zi_1)^{1-b}i_2 + a^{1-b}d(zi_1)i_2^{1-b},$$

which is positive if $b \geq 0$.

Part (ii): A mean-preserving spread in z reduces $\tilde{i}_1$ if $\partial\Pi(i_1,\theta,z)/\partial i_1$ is concave in z. For the CES production function (2.3), the sign of $\partial^3\Pi(i_1,z)/(\partial i_1 \partial^2 z)$ depends on the sign of

$$-\left\{b\left(\frac{1-d}{1-b}\right)^2\left(\frac{a}{1-a}\right)^{b-1}\left(\frac{zi_1}{i_2}\right)^{-b} + d\left(\frac{a}{1-a}\right)^{1-b}\left(\frac{zi_1}{i_2}\right)^{b} + \left(\frac{1-d}{1-b}\right)\left(\frac{1-d}{1-b}-(1-2d)\right)\right\},$$

which is negative if $b \geq 0$.

A.2.8 Proof of Corollary 2.1

Follows directly from Lemma 2.1 and Propositions 2.4 and 2.5.

A.2.9 Proof of Proposition 2.6

Part (i): Follows from Lemma 2.1 and Lemma 2.2.

Part (ii): $\tilde{\boldsymbol{x}} \leq \boldsymbol{x}$ and $\tilde{h}_2 \leq h_2^*$ follows immediately from $\tilde{x}_j = w_j\tilde{i}_1$ and $\tilde{h}_2 = z\tilde{i}_1$. $\tilde{h}_2 \leq h_2^*$ also implies $\tilde{h}_3 \leq h_3^*$. To show this, we differentiate $\tilde{h}_3$ with respect to $\tilde{h}_2$ to get

$$\frac{d\tilde{h}_3}{d\tilde{h}_2} = \frac{d}{d\tilde{h}_2}\left(\theta f\left(\tilde{h}_2,\hat{i}_2(\tilde{h}_2,\theta)\right)\right) = \theta\left(f_1 + f_2\frac{\partial\hat{i}_2(\tilde{h}_2,\theta)}{\partial\tilde{h}_2}\right) = \theta\left(f_1 - f_2\frac{f_{12}}{f_{22}}\right) > 0,$$

where we use Equation (A.3), and the inequality holds by Assumption 2.1. We state this result for future reference.

Lemma A.2 *$\theta f\left(h_2, \hat{i}_2(h_2,\theta)\right)$ is strictly increasing in h_2.*

Next, the marginal return to early investment $z\theta f_1\left(z\tilde{i}_1, \hat{i}_2(z\tilde{i}_1,\theta)\right)$ is strictly decreasing in $\tilde{i}_1$ by Lemma 2.1. So we have

$$z\theta f_1\left(z\tilde{i}_1, \hat{i}_2(z\tilde{i}_1,\theta)\right) \geq z\theta f_1\left(zi_1^*, \hat{i}_2(zi_1^*,\theta)\right) = R^2$$

if and only if $\tilde{i}_1 \leq i_1^*$.

A.2.10 Proof of Proposition 2.7

Because $\tilde{i}_1$ depends only on $\tilde{z}$, $\tilde{i}_1 = i_1^*$ holds when $\tilde{z} = z$. From the conditional demand function $\tilde{x}_j = \tilde{w}_j\tilde{i}_1 = \tilde{w}_j i_1^*$, we can see that $\tilde{x}_j \leq x_j^*$ if and only if $\tilde{w}_j \leq w_j$. $\tilde{h}_2 < h_2^*$ directly follows from $\tilde{i}_1 = i_1^*$ and the definition of optimization, but we explicitly prove this by showing that $\tau(\tilde{\boldsymbol{w}}) < 1$.

$$\tau(\tilde{\boldsymbol{w}})^\phi = \sum_{j=1}^n w_j \left(\frac{\tilde{w}_j}{w_j}\right)^\phi < \left(\sum_{j=1}^n w_j \left(\frac{\tilde{w}_j}{w_j}\right)\right)^\phi = \sum_{j=1}^n \tilde{w}_j = 1,$$

where the inequality holds due to Jensen's inequality and $\phi \in (0,1)$.

As shown in Lemma A.2, $\tilde{h}_2 < h_2^*$ also implies $\tilde{h}_3 < h_3^*$.

Finally, the marginal return to i_1 can be written as

$$z\tau(\tilde{\boldsymbol{w}})\theta f_1\left(\tilde{h}_2, \hat{i}_2(\tilde{h}_2,\theta)\right) = z\tau(\tilde{\boldsymbol{w}})\theta f_1\left(z\tau(\tilde{\boldsymbol{w}})i_1^*, \hat{i}_2\left(z\tau(\tilde{\boldsymbol{w}})i_1^*,\theta\right)\right).$$

Because $z\tau(\tilde{\boldsymbol{w}}) < z$, this quantity is smaller than $R^2 = z\theta f_1\left(zi_1^*, \hat{i}_2(zi_1^*,\theta)\right)$ if Condition 2.1 holds for all values of productivity z' between $z\tau(\tilde{\boldsymbol{w}})$ and z (Lemma 2.2).

A.2.11 Simplifying Condition 2.2

Condition 2 simplifies nicely if borrowing constraints are non-binding throughout adulthood and the consumption intertemporal elasticity of substitution is constant. Notice that when constraints are non-binding, we can write $V(a,h) = v(Ra + \chi h)$ where

$$v(z) = \max_{c_3,\ldots,c_T} \sum_{j=3}^T \beta^{j-3} u(c_j) \text{ subject to } \sum_{j=3}^T R^{3-j} c_j = z. \tag{A.4}$$

With $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$, it is straightforward to show that $v(z) = \Psi \frac{z^{1-\sigma}}{1-\sigma}$ where Ψ is a positive constant that depends on β, R and T.

With these assumptions, Condition 2.2 can be re-written as:

$$\frac{f_{12}f}{f_1 f_2} > -\frac{v''(-RL_2+\chi h_3)\chi^2 h_3}{v'(-RL_2+\chi h_3)\chi} = \frac{\sigma \chi h_3}{-RL_2+\chi h_3}.$$

Taking the inverse of both sides produces the simplified condition in the text.

A.2.12 Properties of $V(a_3, h_3)$

Individuals entering adulthood with human capital h_3 and assets a_3 allocate consumption and saving across their remaining life in the following way:

$$V(a_3,h_3) = \max_{c_3,\ldots,c_T,a_4,\ldots a_T} \sum_{j=3}^{T} \beta^{j-3} u(c_j),$$

subject to budget constraints $a_{j+1} = Ra_j + h_3\Gamma_j - c_j$ for $j \in \{3,...,T\}$, borrowing constraints $a_{j+1} \geq -L_j$ for $j \in \{3,...,T-1\}$, and $a_{T+1} = 0$. Let $\{\hat{c}_j(a_3,h_3)\}_{j=3}^{T}$ and $\{\hat{a}_j(a_3,h_3)\}_{j=4}^{T}$ be the solution to the problem. We sometimes omit the arguments of the policy functions in order to simplify notations.

The value function can be rewritten as:

$$V(a_3,h_3) = u(Ra_3+\Gamma_3-\hat{a}_4(a_3,h_3)) + \sum_{j=4}^{T} \beta^{j-3} u(R\hat{a}_j(a_3,h_3)+h_3\Gamma_j-\hat{a}_{j+1}(a_3,h_3)).$$

Differentiate this with respect to initial human capital level, h_3:

$$V_2(a_3,h_3) = \sum_{j=3}^{T} \beta^{j-3} u'(\hat{c}_j)\Gamma_j + \sum_{j=3}^{T-1} \beta^{j-3} \frac{\partial \hat{a}_{j+1}}{\partial h_3}[\beta R u'(\hat{c}_{j+1}) - u'(\hat{c}_j)] = \sum_{j=3}^{T} \beta^{j-3} u'(\hat{c}_j)\Gamma_j > 0,$$

where the second equality uses the fact that

$$\frac{\partial \hat{a}_{j+1}}{\partial h_3}[\beta R u'(\hat{c}_{j+1}) - u'(\hat{c}_j)] = 0.$$

This follows from the FOC: when the borrowing constraint binds, the derivative $\frac{\partial \hat{a}_{j+1}}{\partial h_3}$ is zero; and when it does not bind, the Euler equation is zero.

Differentiating V_2 with respect to the initial asset level, a_3, yields:

$$V_{21}(a_3,h_3) = \sum_{j=3}^{T} \beta^{j-3}\Gamma_j u''(\hat{c}_j)\frac{\partial \hat{c}_j}{\partial a_3}.$$

If optimal consumption in every period increases with an increase in initial assets, then $\frac{\partial \hat{c}_j}{\partial a_3} \geq 0$ and $V_{21} \leq 0$ given strict concavity of $u(\cdot)$. If $\frac{\partial \hat{c}_j}{\partial a_3} > 0$ for at least one $j \in \{3,...T\}$, then $V_{21} < 0$. We have two cases:

I. Suppose no borrowing constraint ever binds. In this case, we can rewrite $V(a_3,h_3)$ as $v(ra_3 + \chi h_3)$, where v is defined above. All of the proofs carry through with $V_{21} = \chi v''$, $V_{22} = \chi^2 v''$ and $V_2 = \chi v'$.

II. Let $j_b \in \{3,...,T-1\}$ be the first period that any borrowing constraint binds, i.e. $a_{j_b+1} = -L_{j_b}$. In this case, as soon as the constraint binds, an increase in initial assets has no impact on consumption, $\frac{\partial \hat{c}_j}{\partial a_3} = 0,\ j \in \{j_b+1,...,T\}$. Prior to the binding constraint, the increase in initial assets is spread across periods, strictly increasing consumption in all, $\frac{\partial \hat{c}_j}{\partial a_3} > 0,\ j \in \{3,...,j_b\}$. This must be the case in order for the Euler equations to hold in all periods: $u'(c_j) = \beta R u'(c_{j+1}),\ j \in \{3,...,j_b-1\}$. If any c_j increases, they all must. Note that even if the constraint immediately binds, the increase in initial assets will increase consumption in period 3, and therefore (minimally) $\frac{\partial \hat{c}_j}{\partial a_3} > 0$, so $V_{21} < 0$.

In a similar fashion:

$$V_{22}(a_3,h_3) = \sum_{j=3}^{T} \beta^{j-3}\Gamma_j u''(\hat{c}_j)\frac{\partial \hat{c}_j}{\partial (h_3)} < 0.$$

Here, if an individual is constrained in a period and the Euler equation does not hold with equality, the fact that income rises due to the increase in h_3 implies that consumption will rise as well. Again, if the individual faces no future constraints we can rewrite $V(a_3,h_3)$ as $v(ra_3 + \chi h_3)$. All of the proofs carry through with the $V_{21} = \chi v''$, $V_{22} = \chi^2 v''$ and $V_2 = \chi v'$.

A.2.13 Proof of Proposition 2.8

Combining FOCs for assets we have:

$$u'(c_1) \geq (\beta R)^{j-1} u'(c_j), \quad j \in \{2,3,...T\}, \tag{A.5}$$

where inequalities are strict when the relevant borrowing constraint binds. We can write this as:

$$u'(c_j) \le (\beta R)^{1-j} u'(c_1), \quad j \in \{2,3,...T\}. \tag{A.6}$$

Part (i): Using equations (2.15) and (A.6) we have:

$$u'(c_1) = \beta^2 \theta f_1(i_1,i_2)\left[\sum_{j=3}^{T} \beta^{j-3}\Gamma_j u'(c_j)\right] \le \beta^2 \theta f_1(i_1,i_2)\left[\sum_{j=3}^{T} \beta^{j-3}\Gamma_j u'(c_1)(\beta R)^{1-j}\right],$$

which implies $\theta\chi f_1(i_1,i_2) \ge R^2$, with strict inequality if any borrowing constraint binds, and equality if no borrowing constraint binds.

We next prove $\tilde{i}_1 \le i_1^*$. Towards a contradiction, suppose $\tilde{i}_1 > i_1^*$ and let $\tilde{h}_3 = \theta f(\tilde{i}_1,\tilde{i}_2)$ and $h_3^* = \theta f(i_1^*,i_2^*)$. Then the followings hold.

Claim 1: $\tilde{h}_3 \ge h_3^*$.

$\tilde{h}_3 < h_3^*$, together with $\tilde{i}_1 > i_1^*$, implies $f_1(\tilde{i}_1,\tilde{i}_2) < Rf_2(\tilde{i}_1,\tilde{i}_2)$, which contradicts the first order condition $f_1(\tilde{i}_1,\tilde{i}_2) \ge Rf_2(\tilde{i}_1,\tilde{i}_2)$. To show this, let $\left(\underline{i}_1(h_3),\underline{i}_2(h_3)\right)$ be the cost-minimizing investment profile to produce h_3 units of human capital. That is,

$$\left(\underline{i}_1(h_3),\underline{i}_2(h_3)\right) \equiv \underset{i_1,i_2}{\operatorname{argmin}}\left\{i_1 + R^{-1} i_2 | \theta f(i_1,i_2) \ge h_3\right\}.$$

By Assumption 2.1, both $\underline{i}_1(h_3)$ and $\underline{i}_2(h_3)$ are strictly increasing in h_3 (dual of Lemma A.1). So $\tilde{h}_3 < h_3^*$ implies $\underline{i}_1(\tilde{h}_3) < \underline{i}_1(h_3^*) = i_1^*$ and we have $\underline{i}_1(\tilde{h}_3) < \tilde{i}_1$ by the assumption $i_1^* < \tilde{i}_1$. Both investment profiles $\left(\underline{i}_1(\tilde{h}_3),\underline{i}_2(\tilde{h}_3)\right)$ and $(\tilde{i}_1,\tilde{i}_2)$ produce the same amount of human capital $\tilde{h}_3$, but the former is efficient (because it achieves the minimum cost by definition) while the latter uses too much i_1 and too little i_2 compared to the former. Because the isoquant is strictly convex (Assumption 2.1 implies strict quasi-concavity of $f(\cdot)$, which in turn implies strict convexity of the isoquant), the movement along the isoquant in the direction of higher i_1 and lower i_2 decreases the marginal rate of technical substitution:

$$R = \frac{f_1\left(\underline{i}_1(\tilde{h}_3),\underline{i}_2(\tilde{h}_3)\right)}{f_2\left(\underline{i}_1(\tilde{h}_3),\underline{i}_2(\tilde{h}_3)\right)} > \frac{f_1(\tilde{i}_1,\tilde{i}_2)}{f_2(\tilde{i}_1,\tilde{i}_2)}.$$

The inequality $f_1(\tilde{i}_1,\tilde{i}_2) < Rf_2(\tilde{i}_1,\tilde{i}_2)$ contradicts the first order condition $f_1(\tilde{i}_1,\tilde{i}_2) \ge Rf_2(\tilde{i}_1,\tilde{i}_2)$, so it must be $\tilde{h}_3 \ge h_3^*$ if $\tilde{i}_1 > i_1^*$.

Claim 2: $\tilde{i}_2 \geq i_2^*$.
First, by Claim 1, $\tilde{h}_3 \geq h_3^*$ and this implies $\underline{i}_2(\tilde{h}_3) \geq \underline{i}_2(h_3^*) = i_2^*$ (Lemma A.1). Second, the first order condition $f_1(\tilde{i}_1, \tilde{i}_2) \geq R f_2(\tilde{i}_1, \tilde{i}_2)$ implies $\tilde{i}_1 \leq \underline{i}_1(\tilde{h}_3)$ and $\tilde{i}_2 \geq \underline{i}_2(\tilde{h}_3)$ due to the strict convexity of the isoquant. From these two inequalities, we get $\tilde{i}_2 \geq i_2^*$.

Combining Claim 1 and Claim 2 together, we have $\tilde{i}_1 > i_1^*$ and $\tilde{i}_2 \geq i_2^*$. This means that the income-maximizing investment (i_1^*, i_2^*) is affordable, but not chosen by the individual. This contradicts the assumption that $(\tilde{i}_1, \tilde{i}_2)$ solves the individual's problem. For example, reducing investment to (i_1^*, i_2^*) and saving the rest to increase only the period 3 consumption is a better strategy, because over-investing in human capital makes the return lower than the return to savings. Thus, if $\tilde{i}_1$ is the optimal early investment, then it must be smaller than i_1^*.

Next, we show that $\tilde{i}_1 \leq i_1^*$ implies $\tilde{h}_3 \leq h_3^*$. Let $\hat{h}_3(i_1) \equiv \theta f\big(i_1, \hat{i}_2(i_1, \theta)\big)$ be the unconstrained optimum human capital conditional on i_1, where $\hat{i}_2(\cdot)$ is defined in (2.9). By Lemma A.2, $\tilde{i}_1 \leq i_1^*$ implies $\hat{h}_3(\tilde{i}_1) \leq \hat{h}_3(i_1^*) = h_3^*$. Moreover, $\tilde{h}_3 \leq \hat{h}_3(\tilde{i}_1)$ holds because, conditional on i_1, over-investing in i_2 is never optimal ($\tilde{i}_2 \leq \hat{i}_2(\tilde{i}_1, \theta)$). From these two inequalities, we have $\tilde{h}_3 \leq h_3^*$.

We have proved that $\tilde{i}_1 \leq i_1^*$ and $\tilde{h}_3 \leq h_3^*$ when $f_1(\tilde{i}_1, \tilde{i}_2) \geq R f_2(\tilde{i}_1, \tilde{i}_2)$. We next show that when some constraints are binding, the inequalities are strict, i.e. $\tilde{i}_1 < i_1^*$ and $\tilde{h}_3 < h_3^*$. Suppose that $\theta \chi f_1(\tilde{i}_1, \tilde{i}_2) > R^2 = \theta \chi f_1(i_1^*, i_2^*)$ but $\tilde{i}_1 = i_1^*$. Obviously, the inequality $f_1(\tilde{i}_1, \tilde{i}_2) > f_1(i_1^*, i_2^*)$ cannot hold when $f_{12} = 0$. When $f_{12} > 0$, the inequality implies $\tilde{i}_2 > i_2^*$, which is a contradiction because this, together with $\tilde{i}_1 = i_1^*$, means $\tilde{h}_3 > h_3^*$. When $f_{12} < 0$, the inequality implies $\tilde{i}_2 < i_2^*$, which in turn implies $f_1(\tilde{i}_1, \tilde{i}_2) < R f_2(\tilde{i}_1, \tilde{i}_2)$ due to the strict convexity of the isoquant. This contradicts the first order condition $f_1(\tilde{i}_1, \tilde{i}_2) \geq R f_2(\tilde{i}_1, \tilde{i}_2)$. Thus, it must be $\tilde{i}_1 < i_1^*$ when $\theta \chi f_1(\tilde{i}_1, \tilde{i}_2) > R^2$. In this case, $\tilde{h}_3 < h_3^*$ also holds because $\hat{h}_3(i_1)$ is strictly increasing in i_1.

Part (ii): By Part (i), when any borrowing contraint binds, $\tilde{i}_1 < i_1^*$ and $\tilde{h}_3 < h_3^*$. (a) If the period one constraint does not bind, $f_1(\tilde{i}_1, \tilde{i}_2) = R f_2(\tilde{i}_1, \tilde{i}_2)$, so $\tilde{i}_1 = \underline{i}_1(\tilde{h}_3)$ and $\tilde{i}_2 = \underline{i}_2(\tilde{h}_3)$ hold. By Lemma A.1, $\tilde{i}_2 = \underline{i}_2(\tilde{h}_3) < \underline{i}_2(h_3^*) = i_2^*$. (b) If $f_{12} > 0$, $\tilde{i}_1 < i_1^*$ implies $\hat{i}_2(\tilde{i}_1, \theta) < \hat{i}_2(i_1^*, \theta) = i_2^*$. Because $\tilde{i}_2 \leq \hat{i}_2(\tilde{i}_1, \theta)$ always holds, we get $\tilde{i}_2 < i_2^*$.

Part (iii): A similar analysis to Part (i) shows that $\theta \chi f_2(i_1, i_2) \geq R$, with strict inequality if any borrowing constraint in period 2 or later binds, and equality otherwise.

A.2.14 Proof of Proposition 2.9

In the decision problem described in Section 2.8, substitute in for $V(a_3, h_3)$ using equation (2.14) and in for c_1 and c_2 using the budget constraints in equation (2.5).

Part I: Assuming this person is constrained as an old child, let $a_3 = -L_2$ and re-write the decision problem as:

$$\max_{i_1,i_2,a_2} u(y_1 - i_1 - a_2) + \beta u(Ra_2 + y_2 - i_2 + L_2) + \beta^2 V(-RL_2, \theta f(i_1,i_2)).$$

First order conditions for i_1, i_2 and a_2 are:

$$\begin{aligned} -u'(c_1) + \beta^2 V_2(-RL_2, \theta f(i_1,i_2))\theta f_1(i_1,i_2) &= 0 && \text{(A.7)} \\ -\beta u'(c_2) + \beta^2 V_2(-RL_2, \theta f(i_1,i_2))\theta f_2(i_1,i_2) &= 0 && \text{(A.8)} \\ -u'(c_1) + \beta R u'(c_2) &= 0. && \text{(A.9)} \end{aligned}$$

Together, these first order conditions imply $f_1 = Rf_2$ at an optimum. Using this with Cramer's rule yields (dropping arguments of $f(\cdot)$ and $V(-RL_2,\cdot)$ for expositional purposes):

$$\begin{aligned} \frac{\partial i_1}{\partial y_1} &= \frac{R\beta^3 u''(c_1)u''(c_2)\theta V_2(Rf_{22} - f_{12})}{\Delta_2} > 0, \\ \frac{\partial i_2}{\partial y_1} &= \frac{R\beta^3 u''(c_1)u''(c_2)\theta V_2(f_{11} - Rf_{12})}{\Delta_2} > 0, \end{aligned}$$

$\frac{\partial i_1}{\partial y_2} = R^{-1}\frac{\partial i_1}{\partial y_1}$ and $\frac{\partial i_2}{\partial y_2} = R^{-1}\frac{\partial i_2}{\partial y_1}$, where

$$\begin{aligned} \Delta_2 \equiv\ & \beta^4 V_2 \theta^2 [u''(c_1) + \beta R^2 u''(c_2)][V_2(f_{11}f_{22} - f_{12}^2) + V_{22}\theta(f_1^2 f_{22} + f_2^2 f_{11} - 2f_1 f_2 f_{12})] \\ & + \beta^3 u''(c_1)u''(c_2)V_2\theta(f_{11} + R^2 f_{22} - 2Rf_{12}) < 0. \qquad \text{(A.10)} \end{aligned}$$

All of these expressions are signed using assumptions on V, Assumption 2.1 and $f_1 = Rf_2$.

Finally, $\frac{\partial h_3}{\partial y_j} = f_1\frac{\partial i_1}{\partial y_j} + f_2\frac{\partial i_2}{\partial y_j} > 0$ for $j = 1,2$, because all terms in this expression are strictly positive; $\frac{\partial h_3}{\partial y_1} = R\frac{\partial h_3}{\partial y_2}$ follows directly from the fact that $\frac{\partial i_j}{\partial y_1} = R\frac{\partial i_j}{\partial y_2}$ for $j = 1,2$.

Part II: Let $a_2 = -L_1$. With no future constraints, we can rewrite the continuation value $V(a,h) = v(Ra + \chi h)$ as defined in equation (A.4). Substitute this into the problem:

$$\max_{i_1,i_2,a_3} u(y_1 - i_1 + L_1) + \beta u(-RL_1 + y_2 - i_2 - a_3) + \beta^2 v(Ra_3 + \chi\theta f(i_1,i_2)),$$

where $\chi > 0$ is defined above. First order conditions for i_1, i_2 and a_3 are:

$$\begin{aligned}
-u'(c_1)+\beta^2 v'(Ra_3+\chi\theta f(i_1,i_2))\chi\theta f_1(i_1,i_2) &= 0\\
-\beta u'(c_2)+\beta^2 v'(Ra_3+\chi\theta f(i_1,i_2))\chi\theta f_2(i_1,i_2) &= 0\\
-\beta u'(c_2)+\beta^2 R v'(Ra_3+\chi\theta f(i_1,i_2)) &= 0.
\end{aligned}$$

Combining first order conditions, we have $\chi\theta f_2 = R$. However, $f_1 > Rf_2 = R^2/(\chi\theta)$ because L_1 binds (see Proposition 2.8).

Cramer's rule yields (dropping arguments of $f(\cdot)$ and $v(\cdot)$):

$$\begin{aligned}
\frac{\partial i_1}{\partial y_1} &= \frac{\beta^3 u''(c_1)v'\chi\theta f_{22}[u''(c_2)+\beta R^2 v'']}{\Delta_1} > 0\\
\frac{\partial i_1}{\partial y_2} &= \frac{-\beta^5 Ru''(c_2)v'v''\chi^2\theta^2 f_1 f_{22}}{\Delta_1} < 0\\
\frac{\partial i_2}{\partial y_1} &= \frac{-\beta^3 u''(c_1)v'\chi\theta f_{12}[u''(c_2)+\beta R^2 v'']}{\Delta_1}\\
\frac{\partial i_2}{\partial y_2} &= \frac{\beta^5 Ru''(c_2)v'v''\chi^2\theta^2 f_1 f_{12}}{\Delta_1}
\end{aligned}$$

where

$$\begin{aligned}
\Delta_1 \equiv\ & \beta^3 u''(c_1)v'[u''(c_2)+\beta R^2 v'']\chi\theta f_{22}+\beta^5 u''(c_2)v'v''\chi^3\theta^3 f_1^2 f_{22}\\
&+\beta^5[u''(c_2)+\beta R^2 v''](v')^2\chi^2\theta^2[f_{11}f_{22}-f_{12}^2] < 0 \qquad \text{(A.11)}
\end{aligned}$$

by Assumption 2.1. Clearly, $\frac{\partial i_2}{\partial y_1} > 0 \iff f_{12} > 0$, and $\frac{\partial i_2}{\partial y_2} < 0 \iff f_{12} > 0$. Because $\frac{\partial h_3}{\partial y_j} = f_1\frac{\partial i_1}{\partial y_j} + f_2\frac{\partial i_2}{\partial y_j}$ for $j = 1,2$, Assumption 2.1 implies that $\frac{\partial h_3}{\partial y_1} > 0$ and $\frac{\partial h_3}{\partial y_2} < 0$.

Part III: Because we are assuming this person is constrained in both childhood periods, let $a_2 = -L_1$ and $a_3 = -L_2$. The decision problem can be written as:

$$\max_{i_1,i_2}\ u(y_1-i_1+L_1)+\beta u(-RL_1+y_2-i_2+L_2)+\beta^2 V(-RL_2,\theta f(i_1,i_2)).$$

The first order conditions for i_1 and i_2 are given by equations (A.7) and (A.8), where $c_1 = y_1 - i_1 + L_1$ and $c_2 = -RL_1 + y_2 - i_2 + L_2$.

Cramer's rule yields (dropping arguments of $f(\cdot)$ and $V(-RL_2,\cdot)$):

$$
\begin{aligned}
\frac{\partial i_1}{\partial y_1} &= \frac{\beta u''(c_1)[u''(c_2)+\beta V_2\theta f_{22}+\beta V_{22}\theta^2 f_2^2]}{\Delta_{12}} > 0 \\
\frac{\partial i_1}{\partial y_2} &= \frac{-\beta^3 u''(c_2)\theta[V_2 f_{12}+V_{22}\theta f_1 f_2]}{\Delta_{12}} \\
\frac{\partial i_2}{\partial y_1} &= \frac{-\beta^2 u''(c_1)\theta[V_2 f_{12}+V_{22}\theta f_1 f_2]}{\Delta_{12}} \\
\frac{\partial i_2}{\partial y_2} &= \frac{\beta u''(c_2)\theta[u''(c_1)+\beta^2 V_2\theta f_{11}+\beta^2 V_{22}\theta f_1^2]}{\Delta_{12}} > 0
\end{aligned}
$$

where

$$
\begin{aligned}
\Delta_{12} \equiv\ & \beta u''(c_1)u''(c_2)+\beta^2 V_2\theta[u''(c_1)f_{22}+\beta u''(c_2)f_{11}]+\beta^2 V_{22}\theta^2[u''(c_1)f_2^2+\beta u''(c_2)f_1^2] \\
& +\beta^4(V_2)^2\theta^2(f_{11}f_{22}-f_{12}^2)+\beta^4 V_2 V_{22}\theta^3[f_2(f_2 f_{11}-f_1 f_{12})+f_1(f_1 f_{22}-f_2 f_{12})] > 0. \quad \text{(A.12)}
\end{aligned}
$$

Assumptions on $V(\cdot,\cdot)$ and Assumption 2.1 ensure that $\frac{\partial i_1}{\partial y_1}$, $\frac{\partial i_2}{\partial y_2}$, and Δ_{12} are strictly positive. Both $\frac{\partial i_1}{\partial y_2}$ and $\frac{\partial i_2}{\partial y_1}$ are strictly positive if and only if Condition 2.2 holds. Using these results for investments, Assumption 2.1 implies that $\frac{\partial h_3}{\partial y_j} > 0$ for $j = 1, 2$.

A.2.15 Proof of Proposition 2.10

Part (i): Based on the problem discussed in the proof of Proposition 2.9 part (I), we can apply Cramer's rule obtaining:

$$
\begin{aligned}
\frac{\partial i_1}{\partial L_2} &= \frac{\beta^4 R V_2 V_{21}\theta^2 f_2[u''(c_1)+\beta R^2 u''(c_2)](f_1 f_{22}-f_{12})+\beta^3 u''(c_1)u''(c_2)V_2\theta(Rf_{22}-f_{12})}{\Delta_2} > 0 \\
\frac{\partial i_2}{\partial L_2} &= \frac{\beta^4 R V_2 V_{21}\theta^2 f_2[u''(c_1)+\beta R^2 u''(c_2)](f_{11}-Rf_{12})+\beta^3 u''(c_1)u''(c_2)V_2\theta(f_{11}-Rf_{12})}{\Delta_2} > 0,
\end{aligned}
$$

where $\Delta_2 < 0$ is defined previously by equation (A.10). All three of these expressions are signed using assumptions on V, Assumption 2.1 and $f_1 = Rf_2$. Finally, $\frac{\partial h_3}{\partial L_2} = f_1\frac{\partial i_1}{\partial L_2} + f_2\frac{\partial i_2}{\partial L_2} > 0$, because all terms in this expression are positive.

Part (ii): Based on the problem used in the proof of Proposition 2.9 part (III), Cramer's rule yields:

$$
\begin{aligned}
\frac{\partial i_1}{\partial L_2} &= \frac{\beta^4 R V_2 V_{21}\theta^2(f_1 f_{22} - f_2 f_{12}) + \beta^3 R u''(c_2) V_{21}\theta f_1 - \beta^3 u''(c_2)\theta(V_2 f_{12} + \theta V_{22} f_1 f_2)}{\Delta_{12}} \\
\frac{\partial i_2}{\partial L_2} &= \frac{\beta u''(c_2)[u''(c_1) + \beta^2 V_2 \theta f_{11} + \beta^2 V_{22}\theta^2 f_1^2] + \beta^2 R u''(c_1) V_{21}\theta f_2 + \beta^4 R V_2 V_{21}\theta^2 (f_2 f_{11} - f_1 f_{12})}{\Delta_{12}} > 0,
\end{aligned}
$$

where $\Delta_{12} > 0$ is defined previously by equation (A.12). Using assumptions on V, and Assumption 2.1, it is clear that $\frac{\partial i_1}{\partial L_2} > 0$ if Condition 1 holds, and that $\frac{\partial i_2}{\partial L_2} > 0$. Assumption 2.1 further implies that $\frac{\partial h_3}{\partial L_2} = f_1 \frac{\partial i_1}{\partial L_2} + f_2 \frac{\partial i_2}{\partial L_2} > 0$.

A.2.16 Proof of Proposition 2.11

Part (i): Based on the problem discussed in the proof of Proposition 2.9 part (II), we can apply Cramer's rule obtaining:

$$
\begin{aligned}
\frac{\partial i_1}{\partial L_1} &= \frac{f_{22}\left\{\beta^3 u''(c_1) v' \chi \theta [u''(c_2) + \beta R^2 v''] + \beta^5 R^2 u''(c_2) v' v'' \chi^2 \theta^2 f_1\right\}}{\Delta_1} > 0 \\
\frac{\partial i_2}{\partial L_1} &= \frac{-f_{12}\left\{\beta^3 u''(c_1) v' \chi \theta [u''(c_2) + \beta R^2 v''] + \beta^5 R^2 u''(c_2) v' v'' \chi^2 \theta^2 f_1\right\}}{\Delta_1}
\end{aligned}
$$

where $\Delta_1 < 0$ is defined previously by equation (A.11). Clearly, $\frac{\partial i_2}{\partial L_1} > 0 \iff f_{12} > 0$. Finally, $\frac{\partial h_3}{\partial L_2} = f_1 \frac{\partial i_1}{\partial L_2} + f_2 \frac{\partial i_2}{\partial L_2} > 0$ by Assumption 2.1.

Part (ii): Based on the problem used in the proof of Proposition 2.9 part (III), Cramer's rule yields:

$$
\begin{aligned}
\frac{\partial i_1}{\partial L_1} &= \frac{\beta u''(c_1)[u''(c_2) + \beta V_2 \theta f_{22} + \beta V_{22}\theta^2 f_2^2] + \beta^3 R u''(c_2)\theta[V_2 f_{12} + V_{22}\theta f_1 f_2]}{\Delta_{12}} \\
\frac{\partial i_2}{\partial L_1} &= \frac{-\beta R u''(c_2)[u''(c_1) + \beta^2 V_2 \theta f_{11} + \beta^2 V_{22}\theta^2 f_1^2] - \beta^2 u''(c_1)\theta[V_2 f_{12} + V_{22}\theta f_1 f_2]}{\Delta_{12}}
\end{aligned}
$$

where $\Delta_{12} > 0$ is defined previously by equation (A.12). Assumptions on V, and Assumption 2.1 imply that $\frac{\partial i_1}{\partial L_1} < 1$. If Condition 2.2 does not hold, then $V_2 f_{12} + V_{22}\theta f_1 f_2 < 0$, which implies that $\frac{\partial i_1}{\partial L_1} > 0$ and $\frac{\partial i_2}{\partial L_1} < 0$.

A.3 Permission from the Copyright Holder

The following 5 pages contain permission obtained from John Wiley and Sons, the publisher of *Scandinavian Journal of Economics*.

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Sep 10, 2016

This Agreement between Youngmin Park ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	3940450497730
License date	Sep 01, 2016
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Scandinavian Journal of Economics
Licensed Content Title	Correlation, Consumption, Confusion, or Constraints: Why Do Poor Children Perform so Poorly?
Licensed Content Author	Elizabeth M. Caucutt,Lance Lochner,Youngmin Park
Licensed Content Date	Jul 18, 2016
Licensed Content Pages	1
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Title of your thesis / dissertation	Essays on Human Capital and Inequality
Expected completion date	Oct 2016
Expected size (number of pages)	200
Requestor Location	Youngmin Park [redacted] [redacted] Attn: Youngmin Park
Publisher Tax ID	EU826007151
Billing Type	Invoice
Billing Address	Youngmin Park [redacted] [redacted] Attn: Youngmin Park
Total	0.00 CAD

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner.**For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and

Bibliography

Hartman, Richard. 1976. "Factor Demand with Output Price Uncertainty." *American Economic Review* 66 (4):675–681.

Thomson, Godfrey H. 1951. *The Factorial Analysis of Human Ability*. London: University of London Press.

Appendix B

Chapter 3 Appendix

B.1 Proofs and Analytical Details

B.1.1 Proof of Proposition 3.1

I prove Proposition 3.1 in two steps. First, I analyze the youth's problem conditional on parental transfers. Second, I analyze the parents' problem. Combining theses two results completes the proof.

Throughout this section, I assume that the choice set of schooling is $S = [0, T_R]$, and that family characteristics are $(W, \delta, a) \in \mathscr{W} \times \mathscr{D} \times \mathscr{A}$ for some interval $\mathscr{W}, \mathscr{D}, \mathscr{A}$ in $\mathbb{R}_{++}$.

Youth's Problem

I analyze youth's behavior conditional on parental transfers $B \geq 0$. The following lemma characterizes the schooling and borrowing decision of unconstrained youth.

Lemma B.1 *If* (BC) *does not bind, then (i) schooling is nondecreasing in a, and independent of B, (ii) debt at the end of schooling is nondecreasing in a and nonincreasing in B.*

Proof When (BC) does not bind, youth first maximize lifetime income and then fully smooth consumption over their life cycle. Since $\rho = r$, a constant path of consumption maximizes utility.

Schooling The lifetime wealth of youth with parental transfers B, ability a, and schooling s is $B + e^{-rs}Y(s,a) - P(s)$, where $P(s) = \int_0^s e^{-rt} p(t) dt$. Since B is additively separable, wealth-maximizing schooling does depend only on ability. Define the net lifetime earnings of youth with (s,a):

$$\mathscr{F}(s,a) = e^{-rs}Y(s,a) - P(s).$$

Let $s^*(a)$ be the schooling that maximizes lifetime wealth. $s^*(a)$ is nondecreasing in a if a family of functions $\{\mathscr{F}(\cdot,a)\}_{a\in\mathscr{A}}$ obeys single crossing differences (Milgrom and Shannon, 1994), that is, for all $s'' > s'$ and $a'' > a'$,

$$\mathscr{F}(s'',a') - \mathscr{F}(s',a') \geq (>) 0 \Rightarrow \mathscr{F}(s'',a'') - \mathscr{F}(s',a'') \geq (>) 0. \tag{B.1}$$

To show this condition holds, note that $\mathscr{F}(s'',a) - \mathscr{F}(s',a) = \int_{s'}^{s''} \frac{\partial \mathscr{F}(s,a)}{\partial s} ds$. Its derivative with respect to a is

$$\begin{aligned}
\frac{\partial}{\partial a}[\mathscr{F}(s'',a) - \mathscr{F}(s',a)] &= \int_{s'}^{s''} \frac{\partial^2 \mathscr{F}(s,a)}{\partial s \partial a} ds \\
&= \int_{s'}^{s''} e^{-rs}[Y_{sa}(s,a) - rY_a(s,a)] ds \\
&\geq \int_{s'}^{s''} e^{-rs} \frac{\partial \ln Y(s,a)}{\partial a}[Y_s(s,a) - rY(s,a)] ds \quad \left(\because \frac{\partial^2 \ln Y(s,a)}{\partial s \partial a} \geq 0\right) \\
&\geq \frac{\partial \ln Y(s',a)}{\partial a} \int_{s'}^{s''} e^{-rs}[Y_s(s,a) - rY(s,a)] ds \quad \left(\because \frac{\partial^2 \ln Y(s,a)}{\partial s \partial a} \geq 0\right) \\
&\geq \frac{\partial \ln Y(s',a)}{\partial a} \int_{s'}^{s''} \frac{\partial \mathscr{F}(s,a)}{\partial s} ds \quad (\because p(s) \geq 0) \\
&\geq \frac{\partial \ln Y(s',a)}{\partial a} [\mathscr{F}(s'',a) - \mathscr{F}(s',a)].
\end{aligned}$$

This implies that $\mathscr{F}(s'',a) - \mathscr{F}(s',a)$ is increasing in a whenever it is positive. Thus, the condition (B.1) is satisfied.

For $a' < a''$, let $s' \in s^*(a')$ and $s'' \in s^*(a'')$. I want to show that $\min\{s',s''\} \in s^*(a')$ and $\max\{s',s''\} \in s^*(a'')$ so that $s^*(a'') \geq s^*(a')$ (in the strong set order). This is satisfied if $s' \leq s''$. It remains to show this for the case $s' > s''$. Since $s' \in s^*(a')$, we have $\mathscr{F}(s',a') \geq \mathscr{F}(s'',a')$. By single crossing differences, $\mathscr{F}(s',a'') \geq \mathscr{F}(s'',a'')$, which implies $s' \in s^*(a'')$. Furthermore, $\mathscr{F}(s',a') = \mathscr{F}(s'',a')$ should hold so that $s'' \in s^*(a')$. If not, $\mathscr{F}(s',a') > \mathscr{F}(s'',a')$, which implies that $\mathscr{F}(s',a'') > \mathscr{F}(s'',a'')$, contradicting the assumption that $s'' \in s^*(a'')$.

Borrowing From (3.2), the amount of debt at the end of schooling can be derived as follows:

$$\begin{aligned}
D(s) =& \left(D(0) + \int_0^s (c_k(t) + p(t)) \exp\left(-\int_0^t r dt'\right) dt \right) \exp\left(\int_0^s r dt\right) \\
=& e^{rs} \left(-B + \int_0^s e^{-rt} (c_k(t) + p(t)) dt \right).
\end{aligned}$$

For unconstrained youth who attain unconstrained optimal schooling, the constant flow of consumption is

$$c_k(t) = \frac{B + e^{-rs^*(a)} Y(s^*(a), a) - P(s^*(a))}{\Lambda(T_k)}.$$

Since $y(s,a,x)$ is strictly increasing in a, $e^{-rs}Y(s,a) - P(s)$ is also strictly increasing in a. Therefore, $c_k(t)$ is also strictly increasing in a.

The debt at the end of schooling is

$$e^{rs^*(a)} \left\{ -B + \Lambda(s^*(a)) \left(\frac{B + e^{-rs^*(a)} Y(s^*(a), a) - P(s^*(a))}{\Lambda(T_k)} \right) + P(s^*(a)) \right\}. \tag{B.2}$$

Since $\Lambda(s^*(a))$ and $P(s^*(a))$ are nondecreasing in a and $c_k(t)$ is strictly increasing in a, higher ability youth borrow more.

Moreover, by differentiating (B.2) with respect to B, we get

$$e^{rs^*(a)} \left\{ -1 + \frac{\Lambda(s^*(a))}{\Lambda(T_k)} \right\} < 0.$$

Thus, youth with higher parental transfers borrow less.

Next, I consider the case where (BC) binds. I first state a condition that is important for the effects of ability.

Condition B.1

$$\sigma \leq \frac{\frac{Y_{sa}(s,a)}{Y_a(s,a)} - r}{\frac{Y_s(s,a)}{Y(s,a) - \overline{D}} + \frac{e^{-rT_k}}{\int_s^{T_k} e^{-rt} dt}}.$$

Lemma B.2 *If* (BC) *binds, then condition* (3.3) *holds as a strict inequality and schooling is (i) nondecreasing in a if Condition B.1 holds, and (ii) nondecreasing in B.*

Proof When (BC) binds for a given level of schooling, consumption exhibits a jump as follows:

$$c_k(t) = \begin{cases} \frac{B-P(s)+e^{-rs}\overline{D}}{\Lambda(s)} & \text{for } t < s \\ \frac{Y(s,a)-\overline{D}}{\Lambda(T_k-s)} & \text{for } t \geq s \end{cases}.$$

Let $c_{k,0}$ and $c_{k,1}$ be the consumption during and after schooling. Then we can write the objective function of the constrained youth as follows:

$$\mathscr{F}(s,a,B) = \Lambda(s)u(c_{k,0}) + e^{-rs}\Lambda(T_k - s)u(c_{k,1}).$$

The first order condition for s is

$$\mathscr{F}_s(s,a,B) = \Lambda(s)u'(c_{k,0})\frac{\partial c_{k,0}}{\partial s} + e^{-rs}\Lambda(T_k - s)u'(c_{k,1})\frac{\partial c_{k,1}}{\partial s} + e^{-rs}(u(c_{k,0}) - u(c_{k,1})) = 0,$$

where

$$\begin{aligned} \frac{\partial c_{k,0}}{\partial s} &= -\frac{\Lambda'(s)}{\Lambda(s)^2}(B - P(s) + \overline{D}) - \frac{P'(s)}{\Lambda(s)} < 0 \\ \frac{\partial c_{k,1}}{\partial s} &= \frac{\Lambda'(T_k - s)}{\Lambda(T_k - s)^2}(Y(s,a) - \overline{D}) + \frac{Y_s(s,a)}{\Lambda(T_k - s)}. \end{aligned}$$

Deriving Condition (3.3) The first order condition $\mathscr{F}_s = 0$ can be written as

$$\begin{aligned} Y_s(s,a) = & rY(s,a) + p(s) + \left(\frac{u'(c_{k,0}) - u'(c_{k,1})}{u'(c_{k,1})}\right)(r\overline{D} + c_{k,0} + p(s)) + \left(\frac{u(c_{k,1}) - u(c_{k,0})}{u'(c_{k,1})}\right) \\ & - (c_{k,1} - c_{k,0}). \end{aligned}$$

When (BC) binds, $c_{k,0} < c_{k,1}$. Thus, $u'(c_{k,0}) > u'(c_{k,1})$ and $u(c_{k,1}) - u(c_{k,0}) > u'(c_{k,1})(c_{k,1} - c_{k,0})$ hold due to the strict concavity of $u(\cdot)$. This gives the inequality $Y_s(s,a) > rY(s,a) + p(s)$.

Schooling and Ability By differentiating $\mathscr{F}_s$ with respect to a, we get

$$\begin{aligned} \mathscr{F}_{sa}(s,a,B) = \frac{e^{-rs}u'(c_{k,1})}{\Lambda(T_k - s)}\Bigg\{ & -\frac{\sigma}{Y(s,a) - \overline{D}}\Lambda(T_k - s)Y_a(s,a)\left(\frac{\Lambda'(T_k - s)}{\Lambda(T_k - s)}(Y(s,a) - \overline{D}) + Y_s(s,a)\right) \\ & + \big(\Lambda'(T_k - s)Y_a(s,a) + \Lambda(T_k - s)Y_{sa}(s,a)\big) - Y_a(s,a)\Bigg\}. \end{aligned} \tag{B.3}$$

Schooling is nondecreasing in ability if $\mathscr{F}_{sa} \geq 0$. This condition holds if and only if

$$Y_{sa}(s,a) - rY_a(s,a) \geq \sigma Y_a(s,a)\left(\frac{re^{-r(T_k-s)}}{1-e^{-r(T_k-s)}} + \frac{Y_s(s,a)}{Y(s,a)-\overline{\overline{D}}}\right).$$

Therefore, if Condition (B.1) holds, schooling is nondecreasing in ability.

Schooling and Parental Transfer The derivative of $\mathscr{F}_s(s,a,B)$ with respect to B is

$$\mathscr{F}_{sB}(s,a,B) = -u''(c_{k,0})\frac{1}{\Lambda(s)}\left(\frac{\Lambda'(s)}{\Lambda(s)}(B-P(s)+\overline{D}) + P'(s)\right) > 0,$$

which implies that schooling is nondecreasing in B.

Parents' Problem

Let $\hat{s}(B,a)$, $\hat{c}_{k,0}(B,a)$, and $\hat{c}_{k,1}(B,a)$ be the schooling and consumption during and after school of youth with (B,a). Also let $\hat{\underline{c}}_k(B,a)$ be the minimum cost

$$\hat{\underline{c}}_k(B,a) = u^{-1}\left(\frac{\Lambda(\hat{s}(B,a))}{\Lambda(T_k)}u(\hat{c}_{k,0}(B,a)) + \frac{e^{-r\hat{s}(B,a)}\Lambda(T_k-\hat{s}(B,a))}{\Lambda(T_k)}u(\hat{c}_{k,1}(B,a))\right).$$

Its derivative with respect to B is

$$\frac{\partial \hat{\underline{c}}_k(B,a)}{\partial B} = \frac{1}{\Lambda(T_k)}\frac{u'(\hat{c}_{k,0}(B,a))}{u'(\hat{\underline{c}}_k(B,a))} \geq \frac{1}{\Lambda(T_k)},$$

where the inequality is strict if and only if (BC) binds.

The derivative with respect to a is

$$\frac{\partial \hat{\underline{c}}_k(B,a)}{\partial a} = \frac{e^{-r\hat{s}(B,a)}}{\Lambda(T_k)}\frac{u'(\hat{c}_{k,1}(B,a))}{u'(\hat{\underline{c}}_k(B,a))}Y_a(\hat{s}(B,a),a) > 0.$$

Parental Transfer Write the objective function of the parent with (W,δ,a) as follows:

$$\mathscr{F}(B,W,\delta,a) = \Lambda(T_p)v\left(\frac{W-B}{\Lambda(T_p)}\right) + \delta\Lambda(T_k)v(\hat{\underline{c}}_k(B,a)).$$

The first order condition with respect to B is

$$\mathscr{F}_B(B,W,\delta,a) = -v'\left(\frac{W-B}{\Lambda(T_p)}\right) + \delta v'(\hat{\underline{c}}_k(B,a))\frac{u'(\hat{c}_{k,0}(B,a))}{u'(\hat{\underline{c}}_k(B,a))} = 0.$$

The derivatives with respect to w and δ are

$$\mathscr{F}_{Bw}(B,W,\delta,a) = -v''\left(\frac{W-B}{\Lambda(T_p)}\right)\frac{1}{\Lambda(T_p)} > 0,$$
$$\mathscr{F}_{B\delta}(B,W,\delta,a) = v'(\underline{\hat{c}}_k(B,a))\frac{u'(\hat{c}_{k,0}(B,a))}{u'(\underline{\hat{c}}_k(B,a))} > 0.$$

Therefore, $\hat{B}(W,\delta,a)$ is nondecreasing in w and δ.

When (BC) does not bind, $\hat{c}_{k,0}(B,a) = \underline{\hat{c}}_k(B,a)$, and the derivative of $\mathscr{F}_B(B,W,\delta,a)$ with respect to a is

$$\mathscr{F}_{Ba}(B,W,\delta,a) = \delta v''(\underline{\hat{c}}_k(B,a))\frac{\partial \underline{\hat{c}}_k(B,a)}{\partial a} < 0.$$

When (BC) binds,

$$\begin{aligned}
&\mathscr{F}_{Ba}(B,W,\delta,a) \qquad\qquad \text{(B.4)}\\
=&\frac{\partial}{\partial a}\left(\delta v'(\underline{\hat{c}}_k(B,a))\frac{u'(\hat{c}_{k,0}(B,a))}{u'(\underline{\hat{c}}_k(B,a))}\right)\\
=&\delta \underline{\hat{c}}_k(B,a)^{\sigma-\eta-1}\hat{c}_{k,0}(B,a)^{-\sigma-1}\left\{(\sigma-\eta)\frac{\partial \underline{\hat{c}}_k(B,a)}{\partial a}\hat{c}_{k,0}(B,a) - \sigma\frac{\partial \hat{c}_{k,0}(B,a)}{\partial a}\underline{\hat{c}}_k(B,a)\right\}. \qquad \text{(B.5)}
\end{aligned}$$

When Condition B.1 is satisfied, $\partial \hat{s}(B,a)/\partial a \geq 0$ and $\partial \hat{c}_{k,0}(B,a)/\partial a \leq 0$, so $\mathscr{F}_{Ba}(B,W,\delta,a) \geq 0$ if $\sigma \geq \eta$.

Thus, $\hat{B}(W,\delta,a)$ is nonincreasing in a at low levels of a where (BC) does not bind, and nondecreasing in a at high levels of a where (BC) binds.

Schooling Because $\hat{s}(B,a)$ is nondecreasing in B and $\hat{B}(W,\delta,a)$ is nondecreasing in w and δ, $\hat{s}(\hat{B}(W,\delta,a),a)$ is nondecreasing in w and δ. $\hat{s}(B,a)$ is nondecreasing in a. For low levels of ability where (BC) does not bind, $\hat{B}(W,\delta,a)$ is nonincreasing in a but B does not affect s, so $\hat{s}(\hat{B}(W,\delta,a),a)$ is nondecreasing in a. For high levels of ability where (BC) binds, $\hat{B}(W,\delta,a)$ is nondecreasing in a, so $\hat{s}(\hat{B}(W,\delta,a),a)$ is also nondecreasing in a.

B.1.2 Proof of Corollary 3.1

I first show that the game with one-time transfers and that with two-time transfers lead to identical equilibrium allocations in terms of schooling and total parental transfers. Then the comparative statics result is straightforward.

Equivalence of Equilibrium Allocations

Lemma B.3 *Suppose that* $\overline{D} = 0$ *and* $\eta \leq \sigma$*. Then* $\hat{s}(W,\delta,a) = \tilde{s}(W,\delta,a)$ *and* $\hat{B}(W,\delta,a) = \tilde{B}_s(W,\delta,a) + \tilde{B}_n(W,\delta,a)$.

Proof Let d_p and d_k be the debt of the parent and the youth at the end of schooling period. Let $(\hat{s},\hat{B})$ be the equilibrium allocation of the game with one-time transfers and let $(\hat{c}_p,\hat{c}_{k,0},\hat{c}_{k,1})$ be the associated flow consumption of the parent and the child. From these allocations, construct $(\tilde{s},\tilde{B}_s,\tilde{B}_n,\tilde{D}_p,\tilde{D}_k)$ as follows.

$$
\begin{aligned}
\tilde{s} =& \hat{s}, \\
\tilde{B}_s =& \Lambda(\hat{s})\hat{c}_{k,0} + P(\hat{s}), \\
\tilde{B}_n =& \hat{B} - \tilde{B}_s, \\
\tilde{D}_p =& \tilde{B}_s + \Lambda(\hat{s})\hat{c}_p - W, \\
\tilde{D}_k =& 0.
\end{aligned}
$$

The new allocation generates the same consumption path for both the parent and the youth. Therefore, it maximizes the parent's utility if it is achievable. Now I show that the youth does not have an incentive to deviate from the allocation, so this allocation is the equilibrium allocation. Because the youth cannot borrow, it suffices to show that the youth does not have an incentive to save (i.e., $d_k < 0$ is not a profitable deviation).

Consider the final stage of the game where (s,B_s,d_p,d_k) is pre-determined. I first show that B_n is nondecreasing in d_k at the pre-determined allocation $(\tilde{s},\tilde{B}_s,\tilde{D}_p,\tilde{D}_k)$. The parent's problem is

$$
\max_{B_n \geq 0} \left\{ \Lambda(T_p)v(\underline{c}_p) + \delta\Lambda(T_k)v(\underline{c}_k) \right\},
$$

where

$$
\begin{aligned}
\underline{c}_i =& u^{-1}\left(\frac{\Lambda(s)}{\Lambda(T_i)}u(c_{i,0}) + \frac{e^{-\rho s}\Lambda(T_i - s)}{\Lambda(T_i)}u(c_{i,1}) \right) \text{ for } i \in \{p,k\}, \\
c_{p,0} =& \frac{W + e^{-rs}d_p - B_s}{\Lambda(s)}, \quad c_{k,0} = \frac{B_s + e^{-rs}d_k - P(s)}{\Lambda(s)}, \\
c_{p,1} =& \frac{-d_p - e^{rs}B_n}{\Lambda(T_p - s)}, \quad c_{k,1} = \frac{Y(s,a) - d_k + e^{rs}B_n}{\Lambda(T_k - s)}.
\end{aligned}
$$

The first order condition for B_n is

$$-v'(\underline{c}_p)\frac{u'(c_{p,1})}{u'(\underline{c}_p)}+\delta v'(\underline{c}_k)\frac{u'(c_{k,1})}{u'(\underline{c}_k)}\leq 0,$$

where the inequality is strict (i.e., $B_n \geq 0$ binds) if and only if $\hat{c}_{k,0} < \hat{c}_{k,1}$.

I establish conditions under which the choice of B_n is strictly increasing in d_k. It is obvious that B_n is constant when $B_n \geq 0$ binds. Thus, consider the case where $B_n \geq 0$ does not bind.

By differentiating the left hand side with respect to d_k,

$$\delta v'(\underline{c}_k)\frac{u'(c_{k,1})}{u'(\underline{c}_k)\underline{c}_k}\left\{(\sigma-\eta)\frac{\partial \underline{c}_k}{\partial d_k}+\frac{\sigma}{\Lambda(T_k-s)}\right\},$$

where

$$\frac{\partial \underline{c}_k}{\partial d_k}=\frac{e^{-rs}}{\Lambda(T_k)}\left(\frac{u'(c_{k,0})-u'(c_{k,1})}{u'(\underline{c}_k)}\right).$$

First note that $\partial \underline{c}_k/\partial d_k \geq 0$ at the allocation $(\tilde{s},\tilde{B}_s,\tilde{D}_p,\tilde{D}_k)$ because $\hat{c}_{k,0} \leq \hat{c}_{k,1}$. Therefore, if $\eta \leq \sigma$, an increase of d_k from $\tilde{D}_k$ will lead to strictly higher B_n than $\tilde{B}_n$ (Edlin and Shannon, 1998).

Next, I show that $d_k < \tilde{D}_k = 0$ is not a profitable deviation for the child. Consider the saving decision of the child when $(\tilde{s},\tilde{B}_s,\tilde{D}_p)$ is pre-determined. Let $\hat{B}_n(d_k)$ be the non-schooling transfer conditional on d_k. Then we can write the objective function of the child as follows:

$$\mathscr{F}(d_k)=\Lambda(\tilde{s})u\left(\frac{\tilde{B}_s+e^{-r\tilde{s}}d_k-P(\tilde{s})}{\Lambda(\tilde{s})}\right)+e^{-\rho\tilde{s}}\Lambda(T_k-\tilde{s})u\left(\frac{Y(\tilde{s},a)-d_k+e^{r\tilde{s}}\hat{B}_n(d_k)}{\Lambda(T_k-\tilde{s})}\right).$$

First, consider the case where $\hat{B}_n(0)>0$. In this case, $\hat{c}_{k,0}=\hat{c}_{k,1}$. Therefore, for $d_k<0$,

$$\begin{aligned}\mathscr{F}(d_k)\leq&\max_{d'_k}\left\{\Lambda(\tilde{s})u\left(\frac{\tilde{B}_s+e^{-r\tilde{s}}d'_k-P(\tilde{s})}{\Lambda(\tilde{s})}\right)+e^{-\rho\tilde{s}}\Lambda(T_k-\tilde{s})u\left(\frac{Y(\tilde{s},a)-d'_k+e^{r\tilde{s}}\hat{B}_n(d_k)}{\Lambda(T_k-\tilde{s})}\right)\right\}\\ <&\max_{d'_k}\left\{\Lambda(\tilde{s})u\left(\frac{\tilde{B}_s+e^{-r\tilde{s}}d'_k-P(\tilde{s})}{\Lambda(\tilde{s})}\right)+e^{-\rho\tilde{s}}\Lambda(T_k-\tilde{s})u\left(\frac{Y(\tilde{s},a)-d'_k+e^{r\tilde{s}}\hat{B}_n(0)}{\Lambda(T_k-\tilde{s})}\right)\right\}\\ =&\mathscr{F}(0),\end{aligned}$$

where the strict inequality follows from $\hat{B}_n(d_k)<\hat{B}_n(0)$.

Next, consider the case where $\hat{B}_n(0)=0$. In this case, $\hat{c}_{k,0}\leq\hat{c}_{k,1}$ and, for all $d_k<0$,

$\hat{B}_n(d_k) = 0$ and $c_{k,0} < c_{k,1}$. Therefore, for $d_k < 0$,

$$\begin{aligned}\mathscr{F}(d_k) <& \max_{d_k'} \left\{ \Lambda(\tilde{s}) u \left(\frac{\tilde{B}_s + e^{-r\tilde{s}} d_k' - P(\tilde{s})}{\Lambda(\tilde{s})} \right) + e^{-\rho \tilde{s}} \Lambda(T_k - \tilde{s}) u \left(\frac{Y(\tilde{s}, a) - d_k' + e^{r\tilde{s}} \hat{B}_n(d_k)}{\Lambda(T_k - \tilde{s})} \right) \right\} \\ \leq& \mathscr{F}(0).\end{aligned}$$

Since $\mathscr{F}(d_k) < \mathscr{F}(0)$ for all $d_k < 0$, $d_k = \tilde{D}_k = 0$ is the optimal choice of the youth. This proves that the allocation $(\tilde{s}, \tilde{B}_s, \tilde{B}_n, \tilde{D}_p, \tilde{D}_k)$ is achievable, and therefore, the equilibrium allocation.

Comparative Statics

Define

$$X(a) = e^{-rs^*(a)} Y(s^*(a), a) - P(s^*(a)).$$

When $X(a)$ is strictly concave in a, for all values of (W, δ, a) such that $\tilde{s}(W, \delta, a) < s^*(a)$, $\tilde{B}_n(W, \delta, a) = 0$. This is proved as follows. Suppose not. Then there exists (W, δ, a) such that $\tilde{s}(W, \delta, a) < s^*(a)$ but $\tilde{B}_n(W, \delta, a) > 0$. In this case, intertemporal and intergenerational consumption smoothing is fully achieved, but schooling is not chosen to maximize total family consumption. Then a small increase in s increases net family income and improves the parent's utility.

First, consider the case where $\tilde{s}(W, \delta, a) < s^*(a)$ and $\tilde{B}_n(W, \delta, a) = 0$. Then $\tilde{B}_s(W, \delta, a) = \hat{B}(W, \delta, a)$, which is nondecreasing in (W, δ) and also nondecreasing in a if Condition 3.1 holds.

When $\tilde{s}(W, \delta, a) = s^*(a)$, both $\tilde{B}_s(W, \delta, a)$ and $\tilde{B}_n(W, \delta, a)$ are nondecreasing in w and δ. To see this, note that

$$\tilde{B}_s(W, \delta, a) = \Lambda(s^*(a)) \underline{\hat{c}}_k(W, \delta, a) + P(s^*(a)),$$

where

$$\underline{\hat{c}}_k(W, \delta, a) = \frac{\hat{B}(W, \delta, a) + e^{-rs^*(a)} Y(s^*(a), a) - P(s^*(a))}{\Lambda(T_k)}.$$

Since $\hat{B}(W, \delta, a)$ is nondecreasing in w and δ, so is $\underline{\hat{c}}_k(W, \delta, a)$ and $\tilde{B}_s(W, \delta, a)$. Moreover,

$\tilde{B}_n(W,\delta,a)$ is also nondecreasing in w and δ because it can be written as follows:

$$\begin{aligned}\tilde{B}_n(W,\delta,a) =& \hat{B}(W,\delta,a)-\tilde{B}_s(W,\delta,a)\\ =& \left(1-\frac{\Lambda(s^*(a))}{\Lambda(T_k)}\right)\left(\hat{B}(W,\delta,a)-P(s^*(a))\right)-\frac{\Lambda(s^*(a))}{\Lambda(T_k)}e^{-rs^*(a)}Y(s^*(a),a).\end{aligned}$$

Next, I show that $\tilde{B}_s(W,\delta,a)$ is nondecreasing in a. This holds when youth's consumption is nondecreasing in a.

$$\begin{aligned}\frac{\partial \underline{\hat{c}}_k(W,\delta,a)}{\partial a} =& \frac{1}{\Lambda(T_k)}\left(\frac{\partial \hat{B}(W,\delta,a)}{\partial a}+X'(a)\right)\\ =& \left\{\frac{-v''\left(\frac{W-\hat{B}(W,\delta,a)}{\Lambda(T_p)}\right)\frac{1}{\Lambda(T_p)}}{-\left[v''\left(\frac{W-\hat{B}(W,\delta,a)}{\Lambda(T_p)}\right)\frac{1}{\Lambda(T_p)}+\delta v''\left(\frac{\hat{B}(W,\delta,a)+X(a)}{\Lambda(T_k)}\right)\frac{1}{\Lambda(T_k)}\right]}\right\}\frac{X'(a)}{\Lambda(T_k)}\\ >& 0,\end{aligned}$$

where

$$\frac{\partial \hat{B}(W,\delta,a)}{\partial a} = \frac{\delta v''\left(\frac{\hat{B}(W,\delta,a)+X(a)}{\Lambda(T_k)}\right)\frac{X'(a)}{\Lambda(T_k)}}{-\left[v''\left(\frac{W-\hat{B}(W,\delta,a)}{\Lambda(T_p)}\right)\frac{1}{\Lambda(T_p)}+\delta v''\left(\frac{\hat{B}(W,\delta,a)+X(a)}{\Lambda(T_k)}\right)\frac{1}{\Lambda(T_k)}\right]}$$

is used.

Because $\underline{\hat{c}}_k(W,\delta,a)$ is nondecreasing in a, $\tilde{B}_s(W,\delta,a)$ is also nondecreasing in a, which implies that $\tilde{B}_n(W,\delta,a)$ is nonincreasing in a.

B.1.3 Comparative Statics with Heterogeneous Tastes for Schooling

In this section, I derive comparative statics results, that are similar to Proposition 3.1, about how heterogeneous tastes affect schooling and parental transfers.

Let $\Psi(s,\tau)\geq 0$ be the psychic cost of schooling, where τ is the taste parameter. $\Psi(s,\tau)$ is twice continuously differentiable, strictly increasing and strictly convex in s, and $\Psi(0,\tau)=0$ for all τ. Youths with higher τ have higher 'psychic return' to schooling in the sense that they face lower total as well as marginal psychic cost:

$$\frac{\partial \Psi(s,\tau)}{\partial \tau}<0, \quad \frac{\partial^2 \Psi(s,\tau)}{\partial s\partial \tau}<0.$$

The psychic cost enters the utility function of youth as follows:

$$V_k = \Lambda(T_k)\underline{c}_k(U_k) - \Psi(s,\tau)$$

and parents care about $V_k/\Lambda(T_k) = \underline{c}_k(U_k) - \Psi(s,\tau)/\Lambda(T_k)$ instead of $\underline{c}_k(U_k)$.

Proposition B.1 *Let $\hat{s}(W,\delta,a,\tau)$ and $\hat{B}(W,\delta,a,\tau)$ be the solution to the family's problem. If* (B.7) *holds, then*

1. *$\hat{s}(W,\delta,a,\tau)$ is nondecreasing in τ;*
2. *$\hat{B}(W,\delta,a,\tau)$ is (i) nonincreasing in τ if* (BC) *does not bind; (ii) nondecreasing in τ if* (BC) *binds.*

Proof

Youth's Problem I begin with analyzing the unconstrained youth's problem. When (BC) does not bind, the youth's objective function is

$$\mathscr{F}(s,a,\tau,B) = B + e^{-rs}Y(s,a) - P(s) - \Psi(s,\tau).$$

The cross partial is

$$\mathscr{F}_{s\tau}(s,a,\tau,B) = -\Psi_{s\tau}(s,\tau) > 0.$$

Thus, schooling is nondecreasing in taste.

Let $s^*(a,\tau)$ be unconstrained optimal schooling. For unconstrained youth, constant flow of consumption is

$$c_k(t) = \frac{B + e^{-rs^*(a,\tau)}Y(s^*(a,\tau),a) - P(s^*(a,\tau))}{\Lambda(T_k)}.$$

Note that $e^{-rs^*(a,\tau)}Y(s^*(a,\tau),a) - P(s^*(a,\tau))$ is nondecreasing in τ because $\Psi(s,\tau)$ is increasing in s. Therefore, $c_k(t)$ is also nondecreasing in τ.

The debt at the end of schooling is

$$e^{rs^*(a,\tau)}\left\{-B + \Lambda(s^*(a,\tau))\left(\frac{B + e^{-rs^*(a,\tau)}Y(s^*(a,\tau),a) - P(s^*(a,\tau))}{\Lambda(T_k)}\right) + P(s^*(a,\tau))\right\}. \tag{B.6}$$

Since $\Lambda(s^*(a,\tau))$, $P(s^*(a,\tau))$, and $c_k(t)$ are nondecreasing in τ, youth with higher τ borrow more.

The objective function of constrained youth is

$$\begin{aligned}
&\mathscr{F}(s,a,\tau,B)\\
=&\Lambda(T_k)u^{-1}\left(\frac{\Lambda(s)}{\Lambda(T_k)}u\left(\frac{B-P(s)+e^{-rs}\overline{D}}{\Lambda(s)}\right)+\frac{e^{-rs}\Lambda(T_k-s)}{\Lambda(T_k)}u\left(\frac{Y(s,a)-\overline{D}}{\Lambda(T_k-s)}\right)\right)-\Psi(s,\tau).
\end{aligned}$$

The derivative with respect to B is

$$\begin{aligned}
&\mathscr{F}_s(s,a,\tau,B)\\
=&\frac{1}{u'(\underline{c}_k)}\left\{\Lambda(s)u'(c_{k,0})\frac{\partial c_{k,0}}{\partial s}+e^{-rs}\Lambda(T_k-s)u'(c_{k,1})\frac{\partial c_{k,1}}{\partial s}+e^{-rs}(u(c_{k,0})-u(c_{k,1}))\right\}-\Psi_s(s,\tau).
\end{aligned}$$

It is straightforward to see $\mathscr{F}_{s\tau}(s,a,\tau,B)>0$. Taking another derivative with respect to B,

$$\begin{aligned}
&\mathscr{F}_{sB}(s,a,\tau,B)\\
=&\frac{-u''(\underline{c}_k)}{u'(\underline{c}_k)}\frac{\partial \underline{c}_k}{\partial B}\left\{\Lambda(s)u'(c_{k,0})\frac{\partial c_{k,0}}{\partial s}+e^{-rs}\Lambda(T_k-s)u'(c_{k,1})\frac{\partial c_{k,1}}{\partial s}+e^{-rs}(u(c_{k,0})-u(c_{k,1}))\right\}\\
&+\frac{1}{u'(\underline{c}_k)}\left\{-u''(c_{k,0})\frac{1}{\Lambda(s)}\left(\frac{\Lambda'(s)}{\Lambda(s)}(B-P(s)+\overline{D})+P'(s)\right)\right\}>0
\end{aligned}$$

so schooling is nondecreasing in parental transfers and taste.

Parent's Problem Write the objective function of the parent with (W,δ,a,τ) as follows:

$$\mathscr{F}(B,W,\delta,a,\tau)=\Lambda(T_p)v\left(\frac{W-B}{\Lambda(T_p)}\right)+\delta\Lambda(T_k)v\left(\underline{\hat{c}}_k(B,a,\tau)-\frac{\Psi(\hat{s}(B,a,\tau),\tau)}{\Lambda(T_k)}\right).$$

The first order condition with respect to B is

$$\mathscr{F}_B(B,W,\delta,a)=-v'\left(\frac{W-B}{\Lambda(T_p)}\right)+\delta v'\left(\underline{\hat{c}}_k(B,a,\tau)-\frac{\Psi(\hat{s}(B,a,\tau),\tau)}{\Lambda(T_k)}\right)\frac{u'(\hat{c}_{k,0}(B,a,\tau))}{u'(\underline{\hat{c}}_k(B,a,\tau))}=0.$$

When (BC) does not bind, $\hat{c}_{k,0}(B,a,\tau)=\underline{\hat{c}}_k(B,a,\tau)$, and the derivative of $\mathscr{F}_B(B,W,\delta,a,\tau)$ with respect to τ is

$$\mathscr{F}_{B\tau}(B,W,\delta,a,\tau)=\delta v''\left(\underline{\hat{c}}_k(B,a,\tau)-\frac{\Psi(\hat{s}(B,a,\tau),\tau)}{\Lambda(T_k)}\right)\frac{-\Psi_\tau(\hat{s}(B,a,\tau),\tau)}{\Lambda(T_k)}<0.$$

When (BC) binds,

$$
\begin{aligned}
&\mathscr{F}_{B\tau}(B,W,\delta,a,\tau)\\
=&\delta v''\left(\underline{\hat{c}}_k(B,a,\tau)-\frac{\Psi(\hat{s}(B,a,\tau),\tau)}{\Lambda(T_k)}\right)\frac{\partial}{\partial\tau}\left(\underline{\hat{c}}_k(B,a,\tau)-\frac{\Psi(\hat{s}(B,a,\tau),\tau)}{\Lambda(T_k)}\right)\frac{u'(\hat{c}_{k,0}(B,a,\tau))}{u'(\underline{\hat{c}}_k(B,a,\tau))}\\
&+\delta v'\left(\underline{\hat{c}}_k(B,a,\tau)-\frac{\Psi(\hat{s}(B,a,\tau),\tau)}{\Lambda(T_k)}\right)\frac{\partial}{\partial\tau}\left(\frac{u'(\hat{c}_{k,0}(B,a,\tau))}{u'(\underline{\hat{c}}_k(B,a,\tau))}\right),
\end{aligned}
$$

which is positive if and only if

$$
\eta \leq \frac{\frac{\partial \ln\left(\frac{u'(\hat{c}_{k,0}(B,a,\tau))}{u'(\underline{\hat{c}}_k(B,a,\tau))}\right)}{\partial\tau}}{\frac{\partial \ln\left(\underline{\hat{c}}_k(B,a,\tau)-\frac{\Psi(\hat{s}(B,a,\tau),\tau)}{\Lambda(T_k)}\right)}{\partial\tau}}. \tag{B.7}
$$

Thus, $\hat{B}(W,\delta,a,\tau)$ is nonincreasing in τ at low levels of τ where (BC) does not bind, and nondecreasing in τ at high levels of τ where (BC) binds if (B.7) is satisfied.

For low levels of τ where (BC) does not bind, $\hat{B}(W,\delta,a,\tau)$ is nonincreasing in τ but B does not affect s, so $\hat{s}(\hat{B}(W,\delta,a,\tau),a,\tau)$ is nondecreasing in τ. For high levels of τ where (BC) binds, $\hat{B}(W,\delta,a,\tau)$ is nondecreasing in τ, so $\hat{s}(\hat{B}(W,\delta,a,\tau),a,\tau)$ is also nondecreasing in τ.

B.1.4 Properties of $C(V,\delta)$

The family welfare level is

$$
\underline{C}(V,\delta)=\min_{c_p,c_k}\left\{\Lambda(T_p)c_p+\Lambda(T_k)c_k|\Lambda(T_p)v(c_p)+\delta\Lambda(T_k)v(c_k)\geq V\right\}.
$$

This standard cost minimization problem has the solution

$$
\underline{C}(V,\delta)=\left\{\Lambda(T_p)+\Lambda(T_k)\delta^{\frac{1}{\eta}}\right\}^{\frac{-\eta}{1-\eta}}\left[(1-\eta)V\right]^{\frac{1}{1-\eta}},
$$

and its derivative with respect to V is

$$
\frac{\partial\underline{C}(V,\delta)}{\partial V}=\left\{\Lambda(T_p)+\Lambda(T_k)\delta^{\frac{1}{\eta}}\right\}^{\frac{-\eta}{1-\eta}}\left[(1-\eta)V\right]^{\frac{\eta}{1-\eta}}.
$$

Let $\hat{V}(W,a,\delta)$ be the utility under a policy $G(\cdot)$ and let $\hat{c}_i=\underline{c}_i(\hat{U}_p(W,a,\delta))$ be associated constant-equivalent consumption. Then

$$\frac{1}{\delta}\left(\frac{\hat{c}_k}{\hat{c}_p}\right)^{\eta}\begin{cases} >1 & \text{if (BC) or (TC) binds,} \\ =1 & \text{if (BC) and (TC) do not bind.} \end{cases}$$

Comparing Welfare Levels

Then the derivative of $\underline{C}$ with respect to δ is

$$\begin{aligned} &\frac{d\underline{C}(\hat{V}(W,\delta,a),\delta)}{d\delta} \\ =&\frac{\partial\underline{C}(\hat{V}(W,\delta,a),\delta)}{\partial V}\frac{\partial\hat{V}(W,\delta,a)}{\partial\delta}+\frac{\partial\underline{C}(\hat{V}(W,\delta,a),\delta)}{\partial\delta} \\ =&\left\{\Lambda(T_p)+\Lambda(T_k)\delta^{\frac{1}{\eta}}\right\}^{\frac{-1}{1-\eta}}\left[(1-\eta)\hat{V}(W,\delta,a)\right]^{\frac{\eta}{1-\eta}}\Lambda(T_k)\Lambda(T_p)\hat{c}_k^{1-\eta} \\ &\times\frac{1}{1-\eta}\left\{1-\left[\frac{1}{\delta}\left(\frac{\hat{c}_k}{\hat{c}_p}\right)^{\eta}\right]^{\frac{1-\eta}{-\eta}}\right\}. \end{aligned}$$

Thus,

$$\frac{d\underline{C}(\hat{V}(W,\delta,a),\delta)}{d\delta}\begin{cases} >0 & \text{if (BC) or (TC) binds,} \\ =0 & \text{if (BC) and (TC) do not bind.} \end{cases}$$

Welfare Effects of a Lump-sum Transfer

Consider a perturbed policy $\tilde{G}(s)=G(s)+z$ and let $\tilde{V}(z,W,\delta,a)$ and $\tilde{U}_i(z,W,\delta,a)$ be the utility, and $\tilde{c}_i=\underline{c}_i(\tilde{U}_p(z,W,\delta,a))$ be associated constant-equivalent consumption under the perturbed policy.

Non-binding (TC) (Proof of Lemma 3.1) In this case, the marginal value of a lump-sum transfer given to youth is equivalent to the marginal value of parental wealth. From the Envelope condition,

$$\frac{\partial\underline{C}(\tilde{V}(z,W,\delta,a),\delta)}{\partial V}\frac{\partial\tilde{V}(z,W,\delta,a)}{\partial z} \tag{B.8}$$

$$=\left\{\Lambda(T_p)+\Lambda(T_k)\delta^{\frac{1}{\eta}}\right\}^{\frac{-\eta}{1-\eta}}\left\{\Lambda(T_p)+\Lambda(T_k)\delta^{\frac{1}{\eta}}\left[\frac{1}{\delta}\left(\frac{\tilde{c}_k}{\tilde{c}_p}\right)^{\eta}\right]^{\frac{1-\eta}{\eta}}\right\}^{\frac{\eta}{1-\eta}}. \tag{B.9}$$

Thus, we have

$$\left.\frac{\partial \underline{C}(\tilde{V}(z,W,\delta,a),\delta)}{\partial V}\frac{\partial \tilde{V}(z,W,\delta,a)}{\partial z}\right|_{z=0}\begin{cases} >1 & \text{if (BC) binds at } z=0, \\ =1 & \text{if (BC) does not bind at } z=0. \end{cases}$$

Moreover, $\partial \underline{C}(\tilde{V}(z,W,\delta,a),\delta)/\partial V \cdot \partial \tilde{V}(z,W,\delta,a)/\partial z|_{z=0}$ is nondecreasing in a. By Proposition 3.1, $\hat{B}(W,\delta,a)$ is nondecreasing in a, which implies that $\hat{c}_p$ is nonincreasing in a and $\hat{c}_k$ is nondecreasing in a. Therefore, $\hat{c}_k/\hat{c}_p$ is nondecreasing in a.

Binding (TC) and Non-binding (BC) In this case, the marginal value of a subsidy given to youth is

$$\frac{\partial \underline{C}(\tilde{V}(z,W,\delta,a),\delta)}{\partial V}\frac{\partial \tilde{V}(z,W,\delta,a)}{\partial z}$$
$$=\left\{\Lambda(T_p)+\Lambda(T_k)\delta^{\frac{1}{\eta}}\right\}^{\frac{-\eta}{1-\eta}}\left\{\Lambda(T_p)\left[\frac{1}{\delta}\left(\frac{\tilde{c}_k}{\tilde{c}_p}\right)^{\eta}\right]^{\frac{\eta-1}{\eta}}+\delta^{\frac{1}{\eta}}\Lambda(T_k)\right\}^{\frac{\eta}{1-\eta}}.$$

Therefore, when only (TC) binds,

$$\left.\frac{\partial \underline{C}(\tilde{V}(z,W,\delta,a),\delta)}{\partial V}\frac{\partial \tilde{V}(z,W,\delta,a)}{\partial z}\right|_{z=0}<1.$$

B.1.5 Formulating the Planning Problem and Deriving Optimality Conditions

I first formulate the planner's problem as a direct mechanism. The planner assigns schooling $s(\delta)$ and net payment $P(\delta)$ to students whose parents report δ. After reporting δ, parents give a one-time transfer $B\geq 0$ and youth borrow and pay $P(\delta)$ to attain $s(\delta)$ years of schooling. Conditional on B, parents' and youth's objectives coincide because parents are altruistic, so youth do not have incentives to deviate from the allocation $(s(\delta),P(\delta))$ that is chosen by parents. Attending college involves a fixed cost $\chi>0$. I model this as a negative monetary payment for schooling level $s=0$.

Single-Crossing Property

I define an indirect utility function conditional on schooling and payment (s,P). For $s \geq 0$,

$$\mathscr{U}(s,P,\delta) = \max_{B} \underline{C}\Big(\Lambda(T_p)v\left(\underline{c}_p\right) + \delta\Lambda(T_k)v\left(\underline{c}_k\right), \delta\Big), \tag{B.10}$$

subject to (TC) and

$$\begin{aligned}
\underline{c}_p &= \max_{c(\cdot)} \left\{ \frac{1}{\Lambda(T_p)} u^{-1}\left(\int_0^{T_p} e^{-rt}c(t)dt\right) \Big| \int_0^{T_p} e^{-rt}c(t)dt \leq W - B \right\}, \\
\underline{c}_k &= \max_{c(\cdot)} \left\{ \frac{1}{\Lambda(T_k)} u^{-1}\left(\int_0^{T_k} e^{-rt}c(t)dt\right) \Big| \int_0^{T_k} e^{-rt}c(t)dt \leq B - P + e^{-rs}Y(s,a), \right. \\
&\qquad \left. \int_0^{s} e^{-rt}c(t)dt + P - B \leq e^{-rs}\overline{D} \right\}.
\end{aligned}$$

Let $\mathscr{B}(s,P,\delta)$ be the solution to (B.10). $\mathscr{B}(s,P,\delta)$ is strictly increasing in δ as long as (TC) does not bind.

Lemma B.4 *$\mathscr{U}_s(s,P,\delta)/|\mathscr{U}_P(s,P,\delta)|$ is nondecreasing in δ.*

Proof First, note that, for a given level of (s,P), there exists $\underline{B}$ such that (BC) binds for $B < \underline{B}$ and (BC) does not bind for $B > \underline{B}$.

$$\underline{B} = P + \frac{\Lambda(s)Y(s,a) - \Lambda(T_k)\overline{D}}{\Lambda(T_k - s)}$$

Let $\overline{\delta}$ be the altruism level for which $\underline{B} = \mathscr{B}(s,P,\overline{\delta})$, that is, $\overline{\delta}$ solves

$$v'\left(\frac{W - \underline{B}}{\Lambda(T_p)}\right) = \overline{\delta}v'\left(\frac{\underline{B} - P + e^{-rs}Y(s,a)}{\Lambda(T_k)}\right). \tag{B.11}$$

Then for all $\delta \geq \overline{\delta}$,

$$\mathscr{U}(s,P,\delta) = W - P + e^{-rs}Y(s,a).$$

Therefore, the marginal rate of substitution is

$$\frac{\mathscr{U}_s(s,P,\delta)}{|\mathscr{U}_P(s,P,\delta)|} = e^{-rs}[Y_s(s,a) - rY(s,a)].$$

For all $\delta < \overline{\delta}$, (BC) binds, and

$$\frac{\mathscr{U}_s(s,P,\delta)}{|\mathscr{U}_P(s,P,\delta)|} = \frac{\Lambda(s)u'(c_{k,0})\frac{\partial c_{k,0}}{\partial s} + e^{-rs}\Lambda(T_k - s)u'(c_{k,1})\frac{\partial c_{k,1}}{\partial s} + e^{-rs}(u(c_{k,0}) - u(c_{k,1}))}{\Lambda(s)u'(c_{k,0})|\frac{\partial c_{k,0}}{\partial P}|},$$

where

$$\begin{aligned}
c_{k,0} =& \frac{\mathscr{B}(s,P,\delta) - P + e^{-rs}\overline{D}}{\Lambda(s)}, \\
c_{k,1} =& \frac{Y(s,a) - \overline{D}}{\Lambda(T_k - s)}, \\
\frac{\partial c_{k,0}}{\partial s} =& -\frac{\Lambda'(s)}{\Lambda(s)^2}(\mathscr{B}(s,P,\delta) - P + \overline{D}) < 0, \\
\frac{\partial c_{k,0}}{\partial P} =& -\frac{1}{\Lambda(s)} < 0, \\
\frac{\partial c_{k,1}}{\partial s} =& \frac{\Lambda'(T_k - s)}{\Lambda(T_k - s)^2}(Y(s,a) - \overline{D}) + \frac{Y_s(s,a)}{\Lambda(T_k - s)}.
\end{aligned}$$

Note that by the envelope theorem, first derivatives are taken by holding $\mathscr{B}(s,P,\delta)$ fixed.

Altruism affects $\mathscr{U}_s(s,P,\delta)/|\mathscr{U}_P(s,P,\delta)|$ only through $\mathscr{B}(s,P,\delta)$. I first show that $\mathscr{U}_s(s,P,\delta)/|\mathscr{U}_P(s,P,\delta)|$ is strictly increasing in B. Since $\mathscr{B}(s,P,\delta)$ is strictly increasing in δ, this proves that $\mathscr{U}_s(s,P,\delta)/|\mathscr{U}_P(s,P,\delta)|$ is strictly increasing in δ. By differentiating the numerator with respect to B,

$$\Lambda(s)u''(c_{k,0})\frac{\partial c_{k,0}}{\partial B}\frac{\partial c_{k,0}}{\partial s} > 0. \tag{B.12}$$

Moreover, the denominator of $\mathscr{U}_s(s,P,\delta)/|\mathscr{U}_P(s,P,\delta)|$ is also strictly decreasing in B, so $\mathscr{U}_s(s,P,\delta)/|\mathscr{U}_P(s,P,\delta)|$ is strictly increasing in B. Therefore, for all $\delta < \overline{\delta}$, $\mathscr{U}_s(s,P,\delta)$ $/|\mathscr{U}_P(s,P,\delta)|$ is strictly increasing in δ.

Planning Problem and Optimality Conditions

The social planner chooses an allocation $(s(\cdot),P(\cdot))$, taking into account the fact that families have an option not to go to college. The value of not attending college can be written as $\mathscr{U}(0,-\chi,\delta)$. The single crossing condition implies that the allocation is incentive compatible if and only if the first order condition of the reporting problem is satisfied and the allocation is monotone in δ. Therefore, if some students do not attend college, they are from families with low δ. Let $\underline{\delta}$ be the threshold value such that all families with $\delta \leq \underline{\delta}$ do not attend college.

It is convenient to deal with a single control variable by implicitly defining a function $\mathscr{P}(\cdot)$

satisfying

$$\mathscr{V} = \mathscr{U}(s, \mathscr{P}(s, \mathscr{V}, \delta), \delta).$$

Then the planner's problem can be written as follows:

$$\begin{aligned}
\max_{\mathscr{V}(\cdot), s(\cdot), q(\cdot), \underline{\mathscr{V}}, \underline{s}, \underline{\delta}} & \int_{\delta_0}^{\underline{\delta}} \mathscr{U}(0, -\chi, \delta) f(\delta) d\delta + \int_{\underline{\delta}}^{\delta_1} \mathscr{V}(\delta) f(\delta) d\delta \\
\text{subject to } & \int_{\underline{\delta}}^{\delta_1} [K(s(\delta)) - \mathscr{P}(s(\delta), \mathscr{V}(\delta), \delta)] f(\delta) d\delta \leq E, \\
& \mathscr{V}(\delta) = \mathscr{U}(s(\delta), \mathscr{P}(s(\delta), \mathscr{V}(\delta), \delta), \delta), \\
& \dot{\mathscr{V}}(\delta) = \mathscr{U}_\delta(s(\delta), \mathscr{P}(s(\delta), \mathscr{V}(\delta), \delta), \delta), \quad \mathscr{V}(\underline{\delta}) = \underline{\mathscr{V}}, \\
& \dot{s}(\delta) = q(\delta), \quad s(\underline{\delta}) = \underline{s}, \quad q(\delta) \geq 0, \\
& \underline{\mathscr{V}} = \mathscr{U}(0, -\chi, \underline{\delta}), \quad \underline{\delta} \in [\delta_0, \delta_1].
\end{aligned}$$

The Hamiltonian is

$$\mathscr{H}(\mathscr{V}, s, q, \delta, \psi, \varphi, \lambda) = [\mathscr{V} + \lambda(\mathscr{P}(s, \mathscr{V}, \delta) - K(s))] f(\delta) + \psi \mathscr{U}_\delta(s, \mathscr{P}(s, \mathscr{V}, \delta), \delta) + \varphi q,$$

and the small Lagrangian is

$$\mathscr{L}(\underline{\mathscr{V}}, \underline{s}, \underline{\delta}) = \int_{\delta_0}^{\underline{\delta}} \mathscr{U}(0, -\chi, \delta) f(\delta) d\delta + \zeta \left\{ \underline{\mathscr{V}} - \mathscr{U}(0, -\chi, \underline{\delta}) \right\}$$

Following Hellwig (2009) and Weber (2011), the optimality conditions are as follows:

Adjoint Equation For almost all $\delta \in [\underline{\delta}, \delta_1]$,

$$\dot{\psi}(\delta) = -\mathscr{H}_{\mathscr{V}}\big(\mathscr{V}(\delta), s(\delta), q(\delta), \delta, \psi(\delta), \varphi(\delta), \lambda\big), \tag{B.13}$$

$$\dot{\varphi}(\delta) = -\mathscr{H}_s\big(\mathscr{V}(\delta), s(\delta), q(\delta), \delta, \psi(\delta), \varphi(\delta), \lambda\big), \tag{B.14}$$

where

$$\begin{aligned}
\mathscr{H}_{\mathscr{V}}(\mathscr{V}, s, q, \delta, \psi, \varphi, \lambda) &= \left(1 + \frac{\lambda}{\mathscr{U}_P(s, \mathscr{P}(s, \mathscr{V}, \delta), \delta)}\right) f(\delta) + \psi \frac{\mathscr{U}_{\delta P}(s, \mathscr{P}(s, \mathscr{V}, \delta), \delta)}{\mathscr{U}_P(s, \mathscr{P}(s, \mathscr{V}, \delta), \delta)}, \\
\mathscr{H}_s(\mathscr{V}, s, q, \delta, \psi, \varphi, \lambda) &= \lambda \left(\frac{\mathscr{U}_s(s, \mathscr{P}(s, \mathscr{V}, \delta), \delta)}{|\mathscr{U}_P(s, \mathscr{P}(s, \mathscr{V}, \delta), \delta)|} - K'(s) \right) f(\delta) \\
&\quad + \psi |\mathscr{U}_P(s, \mathscr{P}(s, \mathscr{V}, \delta), \delta)| \left(\frac{\partial}{\partial \delta} \frac{\mathscr{U}_s(s, \mathscr{P}(s, \mathscr{V}, \delta), \delta)}{|\mathscr{U}_P(s, \mathscr{P}(s, \mathscr{V}, \delta), \delta)|} \right).
\end{aligned}$$

Transversality

$$\psi(\underline{\delta}) = -\mathscr{L}_{\underline{\mathscr{V}}}(\underline{\mathscr{V}}, \underline{s}, \underline{\delta}), \quad \zeta\left\{\underline{\mathscr{V}} - \mathscr{U}(0, -\chi, \underline{\delta})\right\} = 0, \quad \psi(\delta_1) = 0, \tag{B.15}$$

$$\varphi(\underline{\delta}) = \varphi(\delta_1) = 0, \tag{B.16}$$

where

$$\mathscr{L}_{\underline{\mathscr{V}}}(\underline{\mathscr{V}}, \underline{s}, \underline{\delta}) = \zeta.$$

For almost all $\delta \in [\underline{\delta}, \delta_1]$, $\varphi(\delta) \leq 0$ and $\varphi(\delta) = 0$ if $s(\delta)$ is strictly increasing at δ. Moreover,

$$\int_{\underline{\delta}}^{\delta_1} \varphi(\delta) ds(\delta) = 0.$$

Endpoint Optimality For interior solution $\underline{\delta} \in [\delta_0, \delta_1]$,

$$\mathscr{H}\left(\underline{\mathscr{V}}, \underline{s}, q(\underline{\delta}), \underline{\delta}, \psi(\underline{\delta}), \varphi(\underline{\delta}), \lambda\right) = \mathscr{L}_{\underline{\delta}}(\underline{\mathscr{V}}, \underline{s}, \underline{\delta}), \tag{B.17}$$

where

$$\mathscr{L}_{\underline{\delta}}(\underline{\mathscr{V}}, \underline{s}, \underline{\delta}) = \mathscr{U}(0, -\chi, \underline{\delta}) f(\underline{\delta}) - \zeta \mathscr{U}_{\delta}(0, -\chi, \underline{\delta}).$$

B.1.6 Proof of Proposition 3.2

I first show how the optimality conditions in Section B.1.5 are modified when all students attend college. In this case, $\underline{\delta} \geq \delta_0$ is binding and $\mathscr{V}(\delta_0) > \mathscr{U}(0, -\chi, \delta_0)$. Therefore, $\psi(\delta_0) = -\zeta = 0$ and the endpoint optimality condition (B.17) does not hold with equality at δ_0.

Lemma B.5 $\lambda > 1$ *and* δ_λ *exists.*

Proof From the conditions $\psi(\delta_0) = \psi(\delta_1) = 0$ and (B.13),

$$\int_{\delta_0}^{\delta_1} \left(1 + \frac{\lambda}{\mathscr{U}_P(s(\delta), P(\delta), \delta)}\right) \exp\left(\int_{\delta_0}^{\delta} \frac{\mathscr{U}_{\delta P}(s(\delta'), P(\delta'), \delta')}{\mathscr{U}_P(s(\delta'), P(\delta'), \delta')} d\delta'\right) f(\delta) d\delta = 0.$$

By rewriting this condition, the expression for λ can be derived as follows:

$$\lambda = \frac{\int_{\delta_0}^{\delta_1} \exp\left(\int_{\delta_0}^{\delta} \frac{\mathscr{U}_{\delta P}(s(\delta'), P(\delta'), \delta')}{\mathscr{U}_P(s(\delta'), P(\delta'), \delta')} d\delta'\right) f(\delta) d\delta}{\int_{\delta_0}^{\delta_1} \frac{1}{-\mathscr{U}_P(s(\delta), P(\delta), \delta)} \exp\left(\int_{\delta_0}^{\delta} \frac{\mathscr{U}_{\delta P}(s(\delta'), P(\delta'), \delta')}{\mathscr{U}_P(s(\delta'), P(\delta'), \delta')} d\delta'\right) f(\delta) d\delta}.$$

Note that $-\mathscr{U}_P(s(\delta),P(\delta),\delta) = m(s(\delta))$. Since (TC) does not bind for all δ, $-\mathscr{U}_P(s(\delta),P(\delta),\delta)$ ≥ 1 holds for all δ by Lemma 3.1. Moreover, $-\mathscr{U}_P(s(\delta),P(\delta),\delta) > 1$ holds for a strictly positive measure of families for which (BC) binds. So $\lambda > 1$ follows. Next, for families with $\delta = \delta_1$, $-\mathscr{U}_P(s(\delta),P(\delta),\delta) = 1 < \lambda$. Thus, δ_λ exists.

Lemma B.6 *There exists $\overline{\delta}$ and $(\bar{s},\overline{P})$ such that $(s(\delta),P(\delta)) = (\bar{s},\overline{P})$ for almost all $[\overline{\delta},\delta_1]$ and (BC) does not bind for all $\delta \in [\overline{\delta},\delta_1]$ and (BC) would bind for all $\delta \in [\delta_0,\overline{\delta})$ if they chose $(\bar{s},\overline{P})$.*

Proof Let $\overline{\delta}$ be the minimum type such that (BC) does not bind at the allocation $(s(\delta_1),P(\delta_1))$ (as defined in (B.11)). I first show that (BC) does not bind for all $\delta \in [\overline{\delta},\delta_1]$ at $(s(\delta),P(\delta))$. Suppose not. Then there exists $\delta \in [\overline{\delta},\delta_1]$ such that, for δ, (BC) binds at $(s(\delta),P(\delta))$ but it does not bind at $(s(\delta_1),P(\delta_1))$. By incentive compatibility, the type δ weakly prefers $(s(\delta),P(\delta))$ to $(s(\delta_1),P(\delta_1))$:

$$e^{-rs(\delta_1)}Y(s(\delta_1),a) - P(\delta_1) \leq \mathscr{U}(s(\delta),P(\delta),\delta).$$

Moreover, as shown in Section B.1.4, $\mathscr{U}(s(\delta),P(\delta),\delta) < \mathscr{U}(s(\delta),P(\delta),\delta_1)$ since $\mathscr{U}(s,P,\delta)$ is strictly increasing in δ if (BC) binds, which gives

$$e^{-rs(\delta_1)}Y(s(\delta_1),a) - P(\delta_1) < \mathscr{U}(s(\delta),P(\delta),\delta_1).$$

Therefore, the type δ_1 strictly prefers $(s(\delta),P(\delta))$ to $(s(\delta_1),P(\delta_1))$, which violates incentive compatibility.

Next, I show that $s(\delta)$ and $P(\delta)$ are constant almost everywhere on $[\overline{\delta},\delta_1]$. Suppose not. Then there exist δ' and δ'' such that $\overline{\delta} < \delta' < \delta'' < \delta_1$, $s(\delta') < s(\delta'')$, and $P(\delta') < P(\delta'')$. Since (BC) does not bind for δ' and δ'', by incentive compatibility,

$$e^{-rs(\delta')}Y(s(\delta'),a) - P(\delta') = e^{-rs(\delta'')}Y(s(\delta''),a) - P(\delta'').$$

Define $(\bar{s},\overline{P})$ as follows:

$$\begin{aligned}\bar{s} =& \operatorname*{argmax}_{s\in[s(\overline{\delta}),s(\delta_1)]} \left\{e^{-rs}Y(s,a) - K(s)\right\},\\ \overline{P} =& e^{-r\bar{s}}Y(\bar{s},a) - [e^{-rs(\delta')}Y(s(\delta'),a) - P(\delta')].\end{aligned}$$

By the strict concavity of $e^{-rs}Y(s,a) - K(s)$, $\bar{s}$ is unique. By definition, for all $\delta \geq \overline{\delta}$, $\overline{P} - K(\bar{s}) \geq P(\delta) - K(s(\delta))$, and the inequality is strict if $s(\delta) \neq \bar{s}$.

Consider an alternative allocation $(\tilde{s}(\cdot), \tilde{P}(\cdot))$.

$$\tilde{s}(\delta) = \begin{cases} s(\delta) & \text{for } \delta < \overline{\delta} \\ \overline{s} & \text{for } \delta \geq \overline{\delta} \end{cases}, \quad \tilde{P}(\delta) = \begin{cases} P(\delta) & \text{for } \delta < \overline{\delta} \\ \overline{P} & \text{for } \delta \geq \overline{\delta} \end{cases}.$$

This allocation does not change the social welfare, but reduces the total spending by

$$\int_{\overline{\delta}}^{\delta_1} [K(s(\delta)) - P(\delta)] dF(\delta) - [K(\overline{s}) - \overline{P}][1 - F(\overline{\delta})] > 0,$$

where the strict inequality follows from the fact that $s(\delta) \neq \overline{s}$ for a strictly positive measure of families. Therefore, the original allocation that is not constant over $[\overline{\delta}, \delta_1]$ is not optimal.

Thus, there exists $\overline{\delta}$ such that $s(\delta)$ and $P(\delta)$ are constant almost everywhere on $[\overline{\delta}, \delta_1]$ and (BC) does not bind for all $\delta \in [\overline{\delta}, \delta_1]$. Without loss of generality, assume that $s(\delta)$ and $P(\delta)$ are constant on $[\overline{\delta}, \delta_1]$ and let $(\overline{s}, \overline{P})$ be the allocation. By the definition of $\overline{\delta}$, for all types with $\delta < \overline{\delta}$, (BC) would bind if they chose $(\overline{s}, \overline{P})$.

Lemma B.7 $\psi(\delta) < 0$ *for all* $\delta \in [\delta_\lambda, \delta_1)$.

Proof By condition (B.13) with $\psi(\delta_1) = 0$, $\psi(\delta)$ for $\delta \in [\delta_0, \delta_1]$ can be written as

$$\psi(\delta) = \int_{\delta}^{\delta_1} \left(1 + \frac{\lambda}{\mathscr{U}_P(s(\delta'), P(\delta'), \delta')}\right) \exp\left(\int_{\delta}^{\delta'} \frac{\mathscr{U}_{\delta P}(s(\delta''), P(\delta''), \delta'')}{\mathscr{U}_P(s(\delta''), P(\delta''), \delta'')} d\delta''\right) f(\delta') d\delta'.$$

By the definition of δ_λ, $-\mathscr{U}_P(s(\delta), P(\delta), \delta) < \lambda$ for all $\delta \in [\delta_\lambda, \delta_1]$. Thus, $\psi(\delta) < 0$ holds for all $\delta \in [\delta_\lambda, \delta_1)$.

Now I proceed with additional assumption that $s(\delta_\lambda) < \lim_{\delta \uparrow \overline{\delta}} s(\delta)$.

Lemma B.8 $s(\delta)$ *is strictly increasing at* $\overline{\delta}$[1] *and* $\overline{s} = s^*$.

Proof Suppose not. Then there exists $\tilde{\delta} < \overline{\delta}$ such that $s(\tilde{\delta}) = \overline{s}$ and $s(\tilde{\delta} - \varepsilon) < s(\tilde{\delta})$ for all $\varepsilon > 0$. Note that $\delta_\lambda < \tilde{\delta}$ due to the assumption that $s(\delta_\lambda) < \lim_{\delta \uparrow \overline{\delta}} s(\delta)$. Therefore, $\psi(\delta) < 0$ holds for all $\delta \in [\tilde{\delta}, \delta_1)$.

Note that, by the definition of $\overline{\delta}$, (BC) binds for all types with $\delta \in [\tilde{\delta}, \overline{\delta})$. By the condition (B.16), we have $\varphi(\tilde{\delta}) = \varphi(\delta_1) = 0$, and $\varphi(\delta) \leq 0$ holds for all $\delta \in [\tilde{\delta}, \delta_1]$. Therefore,

$$\varphi(\overline{\delta}) = \varphi(\delta_1) - \int_{\overline{\delta}}^{\delta_1} \dot{\varphi}(\delta) d\delta = \lambda \left\{ \frac{\partial}{\partial s} \left[e^{-r\overline{s}} Y(\overline{s}, a) - K(\overline{s}) \right] \right\} [1 - F(\overline{\delta})] \leq 0,$$

[1] That is, $s(\overline{\delta} - \varepsilon) < s(\overline{\delta} + \varepsilon)$ for all $\varepsilon > 0$.

which implies $\bar{s} \geq s^*$. Moreover, the Spence-Mirrless condition implies that, for all $\delta < \overline{\delta}$

$$\frac{\mathscr{U}_s(\bar{s},\overline{P},\delta)}{|\mathscr{U}_P(\bar{s},\overline{P},\delta)|} < \frac{\mathscr{U}_s(\bar{s},\overline{P},\overline{\delta})}{|\mathscr{U}_P(\bar{s},\overline{P},\overline{\delta})|} = \frac{\partial}{\partial s}\left[e^{-r\bar{s}}Y(\bar{s},a)\right].$$

Using this inequality and the condition $\varphi(\tilde{\delta}) = 0$,

$$\begin{aligned}
0 = \varphi(\tilde{\delta}) =& \varphi(\overline{\delta}) - \int_{\tilde{\delta}}^{\overline{\delta}} \dot{\varphi}(\delta) d\delta \\
<& \lambda \left\{ \frac{\partial}{\partial s}\left[e^{-r\bar{s}}Y(\bar{s},a) - K(\bar{s})\right] \right\} (1 - F(\overline{\delta})) \\
&+ \int_{\tilde{\delta}}^{\overline{\delta}} \psi(\delta) |\mathscr{U}_P(\bar{s},\overline{P},\delta)| \left(\frac{\partial}{\partial \delta} \frac{\mathscr{U}_s(\bar{s},\overline{P},\delta)}{|\mathscr{U}_P(\bar{s},\overline{P},\delta)|} \right) d\delta \\
<& 0,
\end{aligned}$$

which is a contradiction. Thus, all families choosing $(\bar{s},\overline{P})$ are unconstrained and $s(\delta)$ is strictly increasing at $\overline{\delta}$. $\bar{s} = s^*$ follows from the conditions $\varphi(\overline{\delta}) = \varphi(\delta_1) = 0$.

Lemma B.9 *$s(\delta)$ is not continuous at $\overline{\delta}$.*

Proof Suppose not. Then $s(\overline{\delta} - \varepsilon) < s(\overline{\delta})$ for all ε and $\lim_{\delta \uparrow \overline{\delta}} s(\delta) = s(\overline{\delta})$. Because $\varphi(\delta) = 0$ whenever $s(\delta)$ is strictly increasing, we have $\lim_{\delta \uparrow \overline{\delta}} \dot{\varphi}(\delta) = 0$, which gives

$$\begin{aligned}
&\lim_{\delta \uparrow \overline{\delta}} \left\{ \lambda \left(\frac{\mathscr{U}_s(s(\delta),P(\delta),\delta)}{|\mathscr{U}_P(s(\delta),P(\delta),\delta)|} - K'(s(\delta)) \right) f(\delta) \right. \\
&\left. + \psi(\delta) |\mathscr{U}_P(s(\delta),P(\delta),\delta)| \left(\frac{\partial}{\partial \delta} \frac{\mathscr{U}_s(s(\delta),P(\delta),\delta)}{|\mathscr{U}_P(s(\delta),P(\delta),\delta)|} \right) \right\} = 0.
\end{aligned}$$

Because families with $\delta < \overline{\delta}$ are constrained if they were offered $(\bar{s},\overline{P})$,

$$\lim_{\delta \uparrow \overline{\delta}} \frac{\partial}{\partial \delta} \frac{\mathscr{U}_s(s(\delta),P(\delta),\delta)}{|\mathscr{U}_P(s(\delta),P(\delta),\delta)|} = \lim_{\delta \uparrow \overline{\delta}} \frac{\partial}{\partial \delta} \frac{\mathscr{U}_s(\bar{s},\overline{P},\delta)}{|\mathscr{U}_P(\bar{s},\overline{P},\delta)|} > 0.$$

Therefore, we have

$$\begin{aligned}
&\lim_{\delta \uparrow \overline{\delta}} \lambda \left(\frac{\mathscr{U}_s(s(\delta),P(\delta),\delta)}{|\mathscr{U}_P(s(\delta),P(\delta),\delta)|} - K'(s(\delta)) \right) f(\delta) \\
= &-\lim_{\delta \uparrow \overline{\delta}} \psi(\delta) |\mathscr{U}_P(s(\delta),P(\delta),\delta)| \left(\frac{\partial}{\partial \delta} \frac{\mathscr{U}_s(s(\delta),P(\delta),\delta)}{|\mathscr{U}_P(s(\delta),P(\delta),\delta)|} \right) > 0.
\end{aligned}$$

This condition, combined with local incentive compatibility constraints,

$$\mathscr{U}_s(s(\delta),P(\delta),\delta)\dot{s}(\delta)+\mathscr{U}_P(s(\delta),P(\delta),\delta)\dot{P}(\delta)=0, \tag{B.18}$$

gives

$$\lim_{\delta\uparrow\overline{\delta}}\left\{\dot{P}(\delta)-K'(s(\delta))\dot{s}(\delta)\right\}>0.$$

Then unconstrained families have incentives to deviate downward. To see this, consider the highest type δ_1 to ensure that (BC) does not bind in the neighborhood of $(\overline{s},\overline{P})$. For this type, the gain from downward deviation is

$$\begin{aligned}
-\lim_{\delta\uparrow\overline{\delta}}\left\{\frac{\partial}{\partial s}\left[e^{-rs(\delta)}Y(s(\delta),a)\right]\dot{s}(\delta)-\dot{P}(\delta)\right\}=&\lim_{\delta\uparrow\overline{\delta}}\dot{P}(\delta)-\frac{\partial}{\partial s}\left[e^{-r\overline{s}}Y(\overline{s},a)\right]\lim_{\delta\uparrow\overline{\delta}}\dot{s}(\delta)\\
=&\lim_{\delta\uparrow\overline{\delta}}\dot{P}(\delta)-K'(\overline{s})\lim_{\delta\uparrow\overline{\delta}}\dot{s}(\delta)\\
=&\lim_{\delta\uparrow\overline{\delta}}\left\{\dot{P}(\delta)-K'(s(\delta))\dot{s}(\delta)\right\}\\
>&0.
\end{aligned}$$

Therefore, $s(\delta)$ is not continuous at $\overline{\delta}$.

Lemma B.10 *Let* $(\overline{s}_-,\overline{P}_-)=\lim_{\delta\uparrow\overline{\delta}}(s(\delta),P(\delta))$*. Then* $K(\overline{s}_-)-\overline{P}_->K(\overline{s})-\overline{P}$*.*

Proof I first show that families with $\delta\in[\overline{\delta},\delta_1]$ are indifferent between $(\overline{s},\overline{P})$ and $(\overline{s}_-,\overline{P}_-)$. Note that the following inequality holds for all $\delta<\overline{\delta}$

$$\mathscr{V}(\delta)=\mathscr{U}(s(\delta),P(\delta),\delta)\leq\mathscr{U}(s(\delta),P(\delta),\overline{\delta})\leq\mathscr{U}(\overline{s},\overline{P},\overline{\delta})=\mathscr{V}(\overline{\delta}),$$

where the first inequality holds due to the fact that $\mathscr{U}(s,P,\delta)$ is nondecreasing in δ as long as (TC) does not bind, and the second inequality is the incentive compatibility condition for the type $\overline{\delta}$.

Taking the limit from below,

$$\lim_{\delta\uparrow\overline{\delta}}\mathscr{V}(\delta)=\mathscr{U}(\overline{s}_-,\overline{P}_-,\overline{\delta})\leq\mathscr{U}(\overline{s},\overline{P},\overline{\delta})=\mathscr{V}(\overline{\delta}).$$

Because $\mathscr{V}(\delta)$ is absolutely continuous, $\lim_{\delta\uparrow\overline{\delta}}\mathscr{V}(\delta)=\mathscr{V}(\overline{\delta})$, which implies that

$$\mathscr{U}(\overline{s}_-,\overline{P}_-,\overline{\delta})=\mathscr{U}(\overline{s},\overline{P},\overline{\delta}).$$

Moreover, the following also hold.

$$
\begin{aligned}
e^{-r\bar{s}}Y(\bar{s},a)-\overline{P} \leq& e^{-r\bar{s}_-}Y(\bar{s}_-,a)-\overline{P}_- \\
e^{-r\bar{s}}Y(\bar{s},a)-K(\bar{s}) >& e^{-r\bar{s}_-}Y(\bar{s}_-,a)-K(\bar{s}_-)
\end{aligned}
$$

The first inequality holds due to $\mathscr{U}(\bar{s},\overline{P},\overline{\delta}) = \mathscr{U}(\bar{s}_-,\overline{P}_-,\overline{\delta}) \leq e^{-r\bar{s}_-}Y(\bar{s}_-,a)-\overline{P}_-$ and the second inequality holds due to the fact that $\bar{s}=s^*$. Combining these inequalities, we get $K(\bar{s}_-)-\overline{P}_- > K(\bar{s})-\overline{P}$.

Lemma B.11 *$K(s(\delta))-P(\delta)$ is nonincreasing in δ for all $\delta\in[\delta_\lambda,\delta_1]$ where $s(\delta)$ is continuous in δ.*

Proof Consider the case where $s(\delta)$ is differentiable. By differentiating $K(s(\delta))-P(\delta)$ with respect to δ,

$$
K'(s(\delta))\dot{s}(\delta)-\dot{P}(\delta) = \left(K'(s(\delta)) - \frac{\mathscr{U}_s(s(\delta),P(\delta),\delta)}{|\mathscr{U}_P(s(\delta),P(\delta),\delta)|}\right)\dot{s}(\delta).
$$

When $\dot{s}(\delta)=0$, this is equal to zero. When $\dot{s}(\delta)>0$, $\varphi(\delta)=\dot{\varphi}(\delta)=0$ and

$$
K'(s(\delta)) - \frac{\mathscr{U}_s(s(\delta),P(\delta),\delta)}{|\mathscr{U}_P(s(\delta),P(\delta),\delta)|} = \frac{\psi(\delta)|\mathscr{U}_P(s(\delta),P(\delta),\delta)|}{\lambda f(\delta)}\left(\frac{\partial}{\partial\delta}\frac{\mathscr{U}_s(s(\delta),P(\delta),\delta)}{|\mathscr{U}_P(s(\delta),P(\delta),\delta)|}\right). \tag{B.19}
$$

Therefore, the sign of $K'(s(\delta))\dot{s}(\delta)-\dot{P}(\delta)$ depends on the sign of $\psi(\delta)$. Because $\psi(\delta)\leq 0$ for all $\delta\in[\delta_\lambda,\delta_1]$, $K'(s(\delta))\dot{s}(\delta)-\dot{P}(\delta)\leq 0$. The case where $s(\delta)$ is not differentiable can be treated similarly by taking left or right derivatives.

B.1.7 Proof of Proposition 3.4

Suppose that $\underline{\delta}\in(\delta_0,\delta_1)$ and $\zeta\neq 0$. Then $\underline{\mathscr{V}}=\mathscr{U}(0,-\chi,\underline{\delta})$. Suppose that $s(\delta)$ is strictly increasing at $\underline{\delta}$ so that $q(\underline{\delta})>0$. Then $\varphi(\underline{\delta})=\dot{\varphi}(\underline{\delta})=0$ and condition (B.19) holds at $\underline{\delta}$. Therefore, the sign of $g(\underline{s})$ depends on the sign of $\psi(\underline{\delta})$. Rewriting the endpoint optimality condition (B.17),

$$
\psi(\underline{\delta})\left\{\mathscr{U}_\delta(\underline{s},\underline{P},\underline{\delta})-\mathscr{U}_\delta(0,-\chi,\underline{\delta})\right\} = \lambda[K(\underline{s})-\underline{P}]f(\underline{\delta}),
$$

where $\underline{P}=\mathscr{P}(\underline{s},\underline{\mathscr{V}},\underline{\delta})$. Because $\underline{s}>0$ and the type $\underline{\delta}$ is indifferent between $(\underline{s},\underline{P})$ and $(0,-\chi)$, all types with $\delta>\underline{\delta}$ strictly prefer the former to the latter due to the single-crossing condition.

Therefore, we have $\mathscr{U}_\delta(\underline{s}, \underline{P}, \underline{\delta}) > \mathscr{U}_\delta(0, -\chi, \underline{\delta})$ and the sign of $\psi(\underline{\delta})$ depends on the sign of $K(\underline{s}) - \underline{P} = G(\underline{s})$.

B.1.8 Deriving Formula Based on Behavioral Responses

I first define behavioral responses to policy changes and then rewrite the formula for the optimal policy based on these responses. Throughout this section, I assume that the first order condition for schooling holds at all levels of schooling, $\mathscr{U}(s, P, \delta)$ is twice differentiable, $s(\delta)$ and $P(\delta)$ are continuous and differentiable.

Behavioral Responses

For a given policy $G(\cdot)$, consider a family δ with schooling choice $\hat{s} > 0$. Consider a perturbed policy that changes the slope as well as the level of $G(s)$:

$$\tilde{G}(s, \nu, z) = G(s) + z + \nu \cdot e^{-r\hat{s}} \Lambda(s - \hat{s}).$$

Then the objective function of the family facing the perturbed policy is

$$\mathscr{F}(s, \nu, z, \delta) = \mathscr{U}(s, K(s) - \tilde{G}(s, \nu, z), \delta).$$

The derivative with respect to s is

$$\mathscr{F}_s(s, \nu, z, \delta) = \mathscr{U}_s + \mathscr{U}_P \cdot [K'(s) - \tilde{G}_s(s, \nu, z)].$$

Suppose that the first order condition for s holds for family δ at $\hat{s}$. Then $\mathscr{F}_s = 0$ at $\nu = z = 0$ and $s = \hat{s}$, which gives

$$e^{-r\hat{s}}(\kappa(\hat{s}) - g(\hat{s})) = \frac{\mathscr{U}_s}{|\mathscr{U}_P|}. \tag{B.20}$$

Substitution Effect

The response of schooling to an increase in ν is due to the substitution effect:

$$\varepsilon = \frac{\partial s}{\partial \nu} = \frac{\mathscr{F}_{s\nu}}{-\mathscr{F}_{ss}}\Big|_{s=\hat{s}, \nu=z=0},$$

where

$$\mathscr{F}_{s\nu}|_{s=\hat{s},\nu=z=0} = \left\{\mathscr{U}_{sP} + \mathscr{U}_{PP}\cdot(K'(s) - \tilde{G}_s(s,\nu,z))\right\}(-\tilde{G}_\nu) + \mathscr{U}_P(-\tilde{G}_{s\nu})\Big|_{s=\hat{s},\nu=z=0}$$
$$= -e^{-r\hat{s}}\mathscr{U}_P.$$

Income Effect The response of schooling to an increase in z is due to the income effect:

$$\upsilon = \frac{\partial s}{\partial z}\Big|_{s=\hat{s},\nu=z=0} = \frac{\mathscr{F}_{sz}}{-\mathscr{F}_{ss}}\Big|_{s=\hat{s},\nu=z=0},$$

where

$$\mathscr{F}_{sz}|_{s=\hat{s},\nu=z=0} = \left\{\mathscr{U}_{sP} + \mathscr{U}_{PP}(K'(s) - \tilde{G}_s(s,\nu,z))\right\}(-\tilde{G}_z) + \mathscr{U}_P(-\tilde{G}_{sz})|_{s=\hat{s},\nu=z=0}$$
$$= -\mathscr{U}_{sP} + \mathscr{U}_{PP}\frac{\mathscr{U}_s}{\mathscr{U}_P}.$$

Altruism Effect The response of schooling to an increase in δ is

$$\frac{\partial s}{\partial \delta}\Big|_{s=\hat{s},\nu=z=0} = \frac{\mathscr{F}_{s\delta}}{-\mathscr{F}_{ss}}\Big|_{s=\hat{s},\nu=z=0},$$

where

$$\mathscr{F}_{s\delta}|_{s=\hat{s},\nu=z=0} = \mathscr{U}_{s\delta} + \mathscr{U}_{P\delta}e^{-r\hat{s}}(\kappa(\hat{s}) - g(\hat{s})) = \mathscr{U}_{s\delta} - \mathscr{U}_{P\delta}\frac{\mathscr{U}_s}{\mathscr{U}_P} = \mathscr{U}_P\cdot\left(\frac{\partial}{\partial\delta}\frac{\mathscr{U}_s}{\mathscr{U}_P}\right).$$

$\partial s/\partial\delta$ and ε are related in the following way:

$$\frac{\partial s}{\partial \delta} = \upsilon\left(\frac{\partial}{\partial\delta}\frac{\mathscr{U}_s}{\mathscr{U}_P}\right)\left(\frac{\mathscr{U}_P}{-\mathscr{U}_{sP} + \mathscr{U}_{PP}\frac{\mathscr{U}_s}{\mathscr{U}_P}}\right) = -\varepsilon\left(\frac{\partial}{\partial\delta}\frac{\mathscr{U}_s}{\mathscr{U}_P}\right)e^{r\hat{s}}. \tag{B.21}$$

Extensive Margin If the family is indifferent between going to college and not going to college, an increase in z leads to responses along the extensive margin of college attendance. Suppose that

$$\mathscr{U}(\underline{s}, K(\underline{s}) - G(\underline{s}) - z, \underline{\delta})\Big|_{z=0} = \mathscr{U}(0, -\chi, \underline{\delta}).$$

Consider the effect of an increase in z on the threshold type $\underline{\delta}$. By differentiating both sides with respect to z,

$$\left.\frac{\partial \underline{\delta}}{\partial z}\right|_{z=0} = \frac{\mathscr{U}_P(\underline{s}, K(\underline{s}) - G(\underline{s}), \underline{\delta})}{\mathscr{U}_\delta(\underline{s}, K(\underline{s}) - G(\underline{s}), \underline{\delta}) - \mathscr{U}_\delta(0, -\chi, \underline{\delta})}.$$

Then the change of college attendance with respect to subsidy level is

$$\left.\frac{\partial(1 - F(\underline{\delta}))}{\partial z}\right|_{z=0} = \frac{-\mathscr{U}_P(\underline{s}, K(\underline{s}) - G(\underline{s}), \underline{\delta})}{\mathscr{U}_\delta(\underline{s}, K(\underline{s}) - G(\underline{s}), \underline{\delta}) - \mathscr{U}_\delta(0, -\chi, \underline{\delta})} f(\underline{\delta}).$$

Derivation of Formula

Now I rewrite the optimality conditions stated in Section B.1.5 in terms of behavioral responses.

First, define

$$X(\delta) = \frac{-\mathscr{U}_P(s(\delta), P(\delta), \delta)}{\lambda} \psi(\delta).$$

From the condition $\dot{\varphi}(\delta) = 0$,

$$K'(s) - \frac{\mathscr{U}_s}{|\mathscr{U}_P|} = \frac{\psi|\mathscr{U}_P|}{\lambda f(\delta)} \left(\frac{\partial}{\partial \delta} \frac{\mathscr{U}_s}{|\mathscr{U}_P|}\right) = \frac{X(\delta)}{f(\delta)} \left(\frac{\partial}{\partial \delta} \frac{\mathscr{U}_s}{|\mathscr{U}_P|}\right). \tag{B.22}$$

By differentiating $X(\delta)$ with respect to δ and using (B.13), (B.20), (B.18), (B.21), and (B.22),

$$\begin{aligned}
\dot{X}(\delta) &= -\frac{\psi(\delta)}{\lambda}(\mathscr{U}_{Ps}\dot{s}(\delta) + \mathscr{U}_{PP}\dot{P}(\delta) + \mathscr{U}_{P\delta}) - \frac{\mathscr{U}_P}{\lambda}\dot{\psi}(\delta), \\
&= \left\{\left(1 + \frac{\mathscr{U}_P}{\lambda}\right) + e^{-rs(\delta)} g(s(\delta))\upsilon(\delta)\right\} f(\delta).
\end{aligned}$$

Using $X(\delta_1) = 0$,

$$\begin{aligned}
X(\delta) &= -\int_\delta^{\delta_1} \dot{X}(\delta')d\delta', \\
&= -\int_\delta^{\delta_1} \left\{\left(1 + \frac{\mathscr{U}_P(\delta')}{\lambda}\right) + e^{-rs(\delta')} g(s(\delta'))\upsilon(\delta')\right\} f(\delta')d\delta'.
\end{aligned}$$

Also note that

$$\frac{\partial}{\partial \delta} \frac{\mathscr{U}_s}{|\mathscr{U}_P|} = e^{-r\hat{s}} \frac{\dot{s}(\delta)}{\varepsilon(\delta)}.$$

Then condition (B.22) can be rewritten as

$$e^{-rs(\delta)}g(s(\delta)) = -e^{-rs(\delta)}\frac{\dot{s}(\delta)}{\varepsilon(\delta)f(\delta)}\int_{\delta}^{\delta_1}\left\{\left(1+\frac{\mathscr{U}_P(\delta')}{\lambda}\right)+e^{-rs(\delta')}g(s(\delta'))\upsilon(\delta')\right\}f(\delta')d\delta'.$$

From the identity $\Phi(s(\delta)) = F(\delta)$, we get

$$\phi(s(\delta))\dot{s}(\delta) = f(\delta).$$

Then the above equation is

$$g(s(\delta)) = \frac{-1}{\varepsilon(\delta)\phi(s(\delta))}\int_{\delta}^{\delta_1}\left\{\left(1+\frac{\mathscr{U}_P(\delta')}{\lambda}\right)+e^{-rs(\delta')}g(s(\delta'))\upsilon(\delta')\right\}f(\delta')d\delta'.$$

Finally, by changing the variable of integration from δ to s, we get (3.16)

$$g(s) = \frac{-1}{\varepsilon(s)\phi(s)}\int_{s}^{s_1}\left\{\left(1-\frac{m(s')}{\lambda}\right)+e^{-rs'}g(s')\upsilon(s')\right\}d\Phi(s').$$

Next, condition (3.17) can be derived from $X(\underline{\delta})$:

$$X(\underline{\delta}) = -\int_{\delta}^{\delta_1}\left\{\left(1+\frac{\mathscr{U}_P(\delta')}{\lambda}\right)+e^{-rs(\delta')}g(s(\delta'))\upsilon(\delta')\right\}f(\delta')d\delta' \tag{B.23}$$

First, note that

$$X(\underline{\delta}) = \frac{-\mathscr{U}_P}{\lambda}\psi(\underline{\delta}) = \mathscr{U}_P\frac{[K(\underline{s})-\underline{P}]f(\underline{\delta})}{\mathscr{U}_\delta(\underline{s},\underline{P},\underline{\delta})-\mathscr{U}_\delta(0,0,\underline{\delta})} = \frac{\partial(1-F(\underline{\delta}))}{\partial z}[\underline{P}-K(\underline{s})],$$

where the following conditions are used:

$$\psi(\underline{\delta}) = -\frac{\lambda[K(\underline{s})-\underline{P}]f(\underline{\delta})}{\mathscr{U}_\delta(\underline{s},\underline{P},\underline{\delta})-\mathscr{U}_\delta(0,0,\underline{\delta})},$$
$$\left.\frac{\partial(1-F(\underline{\delta}))}{\partial z}\right|_{z=0} = \frac{-\mathscr{U}_P(\underline{s},K(\underline{s})-G(\underline{s}),\underline{\delta})}{\mathscr{U}_\delta(\underline{s},K(\underline{s})-G(\underline{s}),\underline{\delta})-\mathscr{U}_\delta(0,0,\underline{\delta})}f(\underline{\delta}).$$

Then (B.23) can be written as:

$$\left.\frac{\partial(1-F(\underline{\delta}))}{\partial z}\right|_{z=0}[\underline{P}-K(\underline{s})] = -\int_{\delta}^{\delta_1}\left\{\left(1+\frac{\mathscr{U}_P(\delta')}{\lambda}\right)+e^{-rs(\delta')}g(s(\delta'))\upsilon(\delta')\right\}f(\delta')d\delta', \tag{B.24}$$

which gives (3.17).

B.1.9 Proof of Proposition 3.3

The definition of the indirect utility function as well as the planning problem and the optimality conditions can be derived by replacing δ with a. I first show that the indirect utility function $\mathscr{U}(s,P,a)$ satisfies the single-crossing condition. The result largely depends on the proofs of Proposition 3.1 and Lemma B.4.

Lemma B.12 *$\mathscr{U}_s(s,P,a)/|\mathscr{U}_P(s,P,a)|$ is strictly increasing in a if Condition 3.1 is satisfied.*

Proof Let $\mathscr{B}(s,P,a)$ be the parental transfer conditional on (s,P,a). It is straightforward to apply Proposition 3.1 to show that $\mathscr{B}(s,P,a)$ is nonincreasing in a when (BC) does not bind and $\mathscr{B}(s,P,a)$ is nondecreasing in a when (BC) binds and $\eta \leq \sigma$ (see (B.5)).

When (BC) does not bind, the marginal rate of substitution between s and P is

$$\frac{\mathscr{U}_s(s,P,a)}{|\mathscr{U}_p(s,P,a)|} = e^{-rs}[Y_s(s,a) - rY(s,a)],$$

which is strictly increasing in a due to the properties of $Y(s,a)$.

When (BC) binds,

$$\frac{\mathscr{U}_s(s,P,a)}{|\mathscr{U}_P(s,P,a)|} = \frac{\Lambda(s)u'(c_{k,0})\frac{\partial c_{k,0}}{\partial s} + e^{-rs}\Lambda(T_k - s)u'(c_{k,1})\frac{\partial c_{k,1}}{\partial s} + e^{-rs}(u(c_{k,0}) - u(c_{k,1}))}{\Lambda(s)u'(c_{k,0})|\frac{\partial c_{k,0}}{\partial P}|}.$$

Ability affects $\mathscr{U}_s(s,P,a)/|\mathscr{U}_P(s,P,a)|$ directly through $Y(s,a)$ as well as indirectly through $\mathscr{B}(s,P,a)$. Holding B constant, $\mathscr{U}_s(s,P,a)/|\mathscr{U}_P(s,P,a)|$ is strictly increasing in a if and only if Condition B.1 holds (see (B.3)). Moreover, holding a constant, $\mathscr{U}_s(s,P,a)/|\mathscr{U}_P(s,P,a)|$ is strictly increasing in B (see (B.12)). Since $\mathscr{B}(s,P,a)$ is nondecreasing in a, $\mathscr{U}_s(s,P,a)$ / $|\mathscr{U}_P(s,P,a)|$ is strictly increasing in a.

Next, I show that $\psi(a) \geq 0$ for all $a \in [a_0,a_1]$. From $\psi(a_0) = \psi(a_1) = 0$ and (B.13), we get

$$\lambda = \frac{\int_{a_0}^{a_1} \exp\left(\int_{a_0}^{a} \frac{\mathscr{U}_{aP}(s(a'),P(a'),a')}{\mathscr{U}_P(s(a'),P(a'),a')}da'\right) f(a)da}{\int_{a_0}^{a_1} \frac{1}{-\mathscr{U}_P(s(a),P(a),a)} \exp\left(\int_{a_0}^{a} \frac{\mathscr{U}_{aP}(s(a'),P(a'),a')}{\mathscr{U}_P(s(a'),P(a'),a')}da'\right) f(a)da}. \tag{B.25}$$

When (BC) does not bind, $-\mathscr{U}_P(s(a),P(a),a) = 1$, so $-\mathscr{U}_P(s(a),P(a),a)$ does not depend on a. When (BC) binds, $-\mathscr{U}_P(s(a),P(a),a) > 1$ and it is nondecreasing in a, as shown in Section B.1.4. This, combined with (B.25), implies that there exists $\tilde{a}$ such that, $-\mathscr{U}_P(s(a),P(a),a) \leq$

λ for all $a \leq \tilde{a}$ and $-\mathscr{U}_P(s(a), P(a), a) \geq \lambda$ for all $a > \tilde{a}$. Thus, from the condition

$$\psi(a) = -\int_{a_0}^{a} \left(1 + \frac{\lambda}{\mathscr{U}_P(s(a'), P(a'), a')}\right) \exp\left(-\int_{a'}^{a} \frac{\mathscr{U}_{aP}(s(a''), P(a''), a'')}{\mathscr{U}_P(s(a''), P(a''), a'')} da''\right) f(a') da'$$

it is straightforward to see that $\psi(a) \geq 0$. Using the same argument as in Lemma B.11, the conclusion follows.

Bibliography

Edlin, Aaron S. and Chris Shannon. 1998. "Strict Monotonicity in Comparative Statics." *Journal of Economic Theory* 81 (1):201 – 219.

Hellwig, Martin. 2009. "A Maximum Principle for Control Problems with Monotonicity Constraints." *Working Paper, Max Planck Institute for Research on Collective Goods* .

Milgrom, Paul and Chris Shannon. 1994. "Monotone Comparative Statics." *Econometrica* 62 (1):pp. 157–180.

Weber, Thomas A. 2011. *Optimal Control Theory with Applications in Economics*. MIT Press.

Appendix C

Chapter 4 Appendix

C.1 Proofs and Analytical Details

C.1.1 Proof of Proposition 4.1

First, note that, from (4.6), $\mathrm{Cov}(w_t, \Delta w_{t'}|c) = \mu_t \Delta\mu_{t'} \mathrm{Var}(\theta_t|c)$ for $t' - t \geq k$. For $t' - t \geq k$, the covariances conditional on cohort c are

$$\mathrm{Cov}(w_t, \Delta w_{t'}|c) = \mu_t \Delta\mu_{t'} \left[\mathrm{Var}(\theta_{t-1}|c) + \mathrm{Var}(\nu_t|c)\right], \tag{C.1}$$

$$\mathrm{Cov}(w_{t-1}, \Delta w_{t'}|c) = \mu_{t-1} \Delta\mu_{t'} \mathrm{Var}(\theta_{t-1}|c). \tag{C.2}$$

By taking a difference between cohort c and c' such that $\mathrm{Var}(\nu_t|c) = \mathrm{Var}(\nu_t|c')$, we get

$$\frac{\mathrm{Cov}(w_t, \Delta w_{t'}|c) - \mathrm{Cov}(w_t, \Delta w_{t'}|c')}{\mathrm{Cov}(w_{t-1}, \Delta w_{t'}|c) - \mathrm{Cov}(w_{t-1}, \Delta w_{t'}|c')} = \frac{\mu_t}{\mu_{t-1}}$$

In this way, μ_t is identified for all $t < \bar{t} - k$.

To identify μ_t for $t \geq \bar{t} - k$, note that, for $t' - t \geq k$,

$$\frac{\mathrm{Cov}(w_t, \Delta w_{t'+1}|c)}{\mathrm{Cov}(w_t, \Delta w_{t'}|c)} = \frac{\mu_t \Delta\mu_{t'+1} \mathrm{Var}(\theta_t|c)}{\mu_t \Delta\mu_{t'} \mathrm{Var}(\theta_t|c)} = \frac{\Delta\mu_{t'+1}}{\Delta\mu_{t'}}.$$

Therefore, $\mu_{\bar{t}-k}$ is identified using $\mathrm{Cov}(w_{\bar{t}-2k-1}, \Delta w_{\bar{t}-k-1}|c)$, $\mathrm{Cov}(w_{\bar{t}-2k-1}, \Delta w_{\bar{t}-k}|c)$, $\mu_{\bar{t}-k-2}$, and $\mu_{\bar{t}-k-1}$ provided that $\bar{t} - 2k - 1 \geq \underline{t}$. Similarly, μ_t for $t > \bar{t} - k$ are also identified.

Once we know μ_t, then $\mathrm{Var}(\theta_t)$ for all $t < \bar{t} - k$ can be identified from (C.2): for $t' - t \geq k$,

$$\mathrm{Var}(\theta_t) = \frac{\mathrm{Cov}(w_t, \Delta w_{t'})}{\mu_t \Delta\mu_{t'}}.$$

Finally, for all $t < \bar{t} - k$, $\mathrm{Var}(u_t) = \mathrm{Var}(w_t) - \mathrm{Var}(\mu_t \theta_t)$ is identified.

C.1.2 Proof of Proposition 4.2

With the assumption of log-normality, the market clearing condition (4.15) becomes

$$\frac{\ln \hat{\Theta}_t(s, z_t) - \mathbb{E}[\ln \Theta_t | s]}{\sigma(\ln \Theta_t | s)} = \frac{\ln z_t - \mathbb{E}[\ln z_t | s]}{\sigma(\ln z_t | s)},$$

which gives

$$\ln \hat{\Theta}_t(s, z_t) = \mathbb{E}[\ln \Theta_t | s] + \frac{\sigma(\ln \Theta_t | s)}{\sigma(\ln z_t | s)} (\ln z_t - \mathbb{E}[\ln z_t | s]).$$

Let $\hat{z}_t(s, \Theta_t)$ be the inverse of $\Theta_t = \hat{\Theta}_t(s, z_t)$ with respect to z_t, that is,

$$\ln \hat{z}_t(s, \Theta_t) = \mathbb{E}[\ln z_t | s] + \frac{\sigma(\ln z_t | s)}{\sigma(\ln \Theta_t | s)} (\ln \Theta_t - \mathbb{E}[\ln \Theta_t | s]).$$

Then the first order condition (4.14) becomes

$$\begin{aligned}
\frac{\partial W_t(s, \Theta_t)}{\partial \Theta_t} =& p_t(s) \lambda_t(s) \Theta_t^{\lambda_t(s)-1} \hat{z}_t(s, \Theta_t)^{\gamma_t(s)} \\
=& p_t(s) \lambda_t(s) \exp\left(\gamma_t(s) \left\{ \mathbb{E}[\ln z_t | s] - \frac{\sigma(\ln z_t | s)}{\sigma(\ln \Theta_t | s)} \mathbb{E}[\ln \Theta_t | s] \right\} \right) \Theta_t^{\mu_t(s)-1},
\end{aligned}$$

where

$$\mu_t(s) = \lambda_t(s) + \gamma_t(s) \frac{\sigma(\ln z_t | s)}{\sigma(\ln \Theta_t | s)}.$$

Next, note that $W_t(s, 0) = 0$ because $W_t(s, 0) > 0$ would give negative profits for producers with $z_t = 0$. Then, by integrating the above equation, we get

$$\begin{aligned}
W_t(s, \Theta_t) =& W_t(s, 0) + \int_0^{\Theta_t} \frac{\partial W_t(s, \Theta_t')}{\partial \Theta_t} d\Theta_t' \\
=& p_t(s) \frac{\lambda_t(s)}{\mu_t(s)} \exp\left(\gamma_t(s) \left\{ \mathbb{E}[\ln z_t | s] - \frac{\sigma(\ln z_t | s)}{\sigma(\ln \Theta_t | s)} \mathbb{E}[\ln \Theta_t | s] \right\} \right) \Theta_t^{\mu_t(s)},
\end{aligned}$$

which gives (4.18).

Note that

$$\begin{aligned} p_t(s)y_t\big(s,\Theta_t,\hat{z}_t(s,\Theta_t)\big) =& p_t(s)\Theta_t^{\lambda_t(s)}\hat{z}_t(s,\Theta_t)^{\gamma_t(s)} \\ =& p_t(s)\exp\left(\gamma_t(s)\left\{\mathbb{E}[\ln z_t|s]-\frac{\sigma(\ln z_t|s)}{\sigma(\ln\Theta_t|s)}\mathbb{E}[\ln\Theta_t|s]\right\}\right)\Theta_t^{\mu_t(s)} \\ =& W_t(s,\Theta_t)\frac{\mu_t(s)}{\lambda_t(s)}. \end{aligned}$$

Therefore,

$$\frac{W_t(s,\Theta_t)}{p_t(s)y_t\big(s,\Theta_t,\hat{z}_t(s,\Theta_t)\big)}=\frac{\lambda_t(s)}{\mu_t(s)}.$$

C.1.3 Identification with ARMA(1,1) Transitory Shocks

Suppose that there exist (t,t',c) such that $\mathrm{Var}(\theta_t|c)=\mathrm{Var}(\theta_t|c+1)$ and $\mathrm{Var}(\theta_{t'}|c)=\mathrm{Var}(\theta_{t'}|c+1)$. Then the AR(1) coefficient ρ is identified. The assumption is satisfied if, for example, there is no cohort trend in initial skill and $\mathrm{Var}(\nu_t)=0$ for some t. Once ρ is identified, the rest of the identification is simliar with that of the MA(q) process.

Identification of ρ Let c_i denote the year when individual i enters the labor market. Then

$$\begin{aligned} \varepsilon_{i,t} =& \xi_{i,t}+(\beta+\rho)\xi_{i,t-1}+\rho(\beta+\rho)\xi_{i,t-2}+\ldots,+\rho^{t-c_i-1}(\beta+\rho)\xi_{i,c_i}, \\ \varepsilon_{i,t+k} =& \xi_{i,t+k}+(\beta+\rho)\xi_{i,t+k-1}+\rho(\beta+\rho)\xi_{i,t+k-2}+\ldots \\ &+\rho^{k-1}(\beta+\rho)\xi_{i,t}+\rho^{k}(\beta+\rho)\xi_{i,t-1}+\ldots+\rho^{t+k-c_i-1}(\beta+\rho)\xi_{i,c_i}. \end{aligned}$$

Then the autocovariance of the transitory component is

$$\begin{aligned} &\mathrm{Cov}(\varepsilon_t,\varepsilon_{t+k}|c) \\ =& \rho^{k-1}\Big\{(\beta+\rho)\,\mathrm{Var}(\xi_t)+\rho(\beta+\rho)^2\,\mathrm{Var}(\xi_{t-1})+\ldots+\rho^{2(t-c)-1}(\beta+\rho)^2\,\mathrm{Var}(\xi_c)\Big\}, \end{aligned}$$

which gives the following autocovariance of earnings residual:

$$\begin{aligned} &\mathrm{Cov}(w_t,\Delta w_{t+k}|c) \\ =& \mu_t\Delta\mu_{t+k}\,\mathrm{Var}(\theta_t|c)+\mathrm{Cov}(\varepsilon_t,\varepsilon_{t+k}|c)-\mathrm{Cov}(\varepsilon_t,\varepsilon_{t+k-1}|c) \\ =& \mu_t\Delta\mu_{t+k}\,\mathrm{Var}(\theta_t|c) \\ &+\rho^{k-2}(\rho-1)\Big\{(\beta+\rho)\,\mathrm{Var}(\xi_t)+\rho(\beta+\rho)^2\,\mathrm{Var}(\xi_{t-1})+\ldots+\rho^{2(t-c)-1}(\beta+\rho)^2\,\mathrm{Var}(\xi_c)\Big\}. \end{aligned}$$

By differencing across neighboring cohorts, we get

$$\begin{aligned}&\operatorname{Cov}(w_t,\Delta w_{t+k}|c)-\operatorname{Cov}(w_t,\Delta w_{t+k}|c+1)\\=&\mu_t\Delta\mu_{t+k}\big[\operatorname{Var}(\theta_t|c)-\operatorname{Var}(\theta_t|c+1)\big]+\rho^{k-2+2(t-c)-1}(\rho-1)(\beta+\rho)^2\operatorname{Var}(\xi_c)\end{aligned}$$

Therefore, if there exists (t,t',c) such that $\operatorname{Var}(\theta_t|c)=\operatorname{Var}(\theta_t|c+1)$ and $\operatorname{Var}(\theta_{t'}|c)=\operatorname{Var}(\theta_{t'}|c+1)$, ρ is identified as follows:

$$\rho^{k-k'}=\frac{\operatorname{Cov}(w_t,\Delta w_{t+k}|c)-\operatorname{Cov}(w_t,\Delta w_{t+k}|c+1)}{\operatorname{Cov}(w_t,\Delta w_{t+k'}|c)-\operatorname{Cov}(w_t,\Delta w_{t+k'}|c+1)}.$$

Identification of μ_t Note that

$$\begin{aligned}\Delta w_{t+1}-\rho\Delta w_t=&[\Delta\mu_{t+1}\theta_{t+1}+\mu_t\nu_{t+1}+\eta_{t+1}+\Delta\varepsilon_{t+1}]-\rho[\Delta\mu_t\theta_t+\mu_{t-1}\nu_t+\eta_t+\Delta\varepsilon_t]\\=&[\Delta\mu_{t+1}\theta_{t+1}+\mu_t\nu_{t+1}+\eta_{t+1}]-\rho[\Delta\mu_t\theta_t+\mu_{t-1}\nu_t+\eta_t]+\Delta\xi_{t+1}+\beta\Delta\xi_t.\end{aligned}$$

Then

$$\operatorname{Cov}(w_{t-2},\Delta w_{t+1}-\rho\Delta w_t)=\mu_{t-2}(\Delta\mu_{t+1}-\rho\Delta\mu_t)\operatorname{Var}(\theta_{t-2}),$$

and the difference across cohorts is

$$\begin{aligned}&\operatorname{Cov}(w_{t-2},\Delta w_{t+1}-\rho\Delta w_t|c)-\operatorname{Cov}(w_{t-2},\Delta w_{t+1}-\rho\Delta w_t|c')\\=&\mu_{t-2}(\Delta\mu_{t+1}-\rho\Delta\mu_t)\Big\{\operatorname{Var}(\theta_{t-2}|c)-\operatorname{Var}(\theta_{t-2}|c')\Big\}.\end{aligned}$$

Therefore, if $\operatorname{Var}(\nu_{t-1}|c)=\operatorname{Var}(\nu_{t-1}|c')$, then $\operatorname{Var}(\theta_{t-2}|c)-\operatorname{Var}(\theta_{t-2}|c')$
$=\operatorname{Var}(\theta_{t-3}|c)-\operatorname{Var}(\theta_{t-3}|c')$ and

$$\frac{\operatorname{Cov}(w_{t-2},\Delta w_{t+1}-\rho\Delta w_t|c)-\operatorname{Cov}(w_{t-2},\Delta w_{t+1}-\rho\Delta w_t|c')}{\operatorname{Cov}(w_{t-3},\Delta w_{t+1}-\rho\Delta w_t|c)-\operatorname{Cov}(w_{t-3},\Delta w_{t+1}-\rho\Delta w_t|c')}=\frac{\mu_{t-2}}{\mu_{t-3}}.$$

C.1.4 Calculating Standard Errors

Let $m=1,\dots,M$ be the index of moments. Let $d_{i,m}$ be the indicator of whether individual i contributes to the m^{th} moment $\operatorname{Cov}(w_t,w_{t'}|s,A_j)$. That is, both $w_{i,t}$ and $w_{i,t'}$ are non-missing and $s_i=s$ and $a_{i,t}\in A_j$. Also let $p_m(\mathbf{\Lambda})=\widetilde{\operatorname{Cov}}(w_t,w_{t'}|s,A_j,\mathbf{\Lambda})$. Then we can write

$$h_m(\boldsymbol{z}_i,\mathbf{\Lambda})=d_{i,m}\big[w_{i,t}w_{i,t'}-p_m(\mathbf{\Lambda})\big],$$

where $\boldsymbol{z}_i$ includes $w_{i,t}\ d_{i,m}$ for all t and m for individual i. Let $\boldsymbol{h}(\boldsymbol{z},\boldsymbol{\Lambda}) = (h_1(\boldsymbol{z},\boldsymbol{\Lambda}),\ldots,h_M(\boldsymbol{z},\boldsymbol{\Lambda}))$. Then the following moment condition holds for the true parameter $\boldsymbol{\Lambda}_0$:

$$\mathbb{E}[\boldsymbol{h}(\boldsymbol{z},\boldsymbol{\Lambda}_0)] = 0.$$

The minimum distance estimator $\hat{\boldsymbol{\Lambda}}$ is equivalent to the GMM estimator that solves

$$\min_{\boldsymbol{\Lambda}} \left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{h}(\boldsymbol{z}_i,\boldsymbol{\Lambda})\right]' \boldsymbol{W} \left[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{h}(\boldsymbol{z}_i,\boldsymbol{\Lambda})\right],$$

where $\boldsymbol{W} = \mathrm{diag}[\frac{N^2}{N_1^2},\ldots,\frac{N^2}{N_M^2}]$ and $N_m = \sum_{i=1}^{N} d_{i,m}$.

The GMM estimator $\hat{\boldsymbol{\Lambda}}$ is asymptotically normal with a variance matrix $\boldsymbol{V} = (\boldsymbol{H}'\boldsymbol{W}\boldsymbol{H})^{-1} \times (\boldsymbol{H}'\boldsymbol{W}\Omega\boldsymbol{W}\boldsymbol{H}) \times (\boldsymbol{H}'\boldsymbol{W}\boldsymbol{H})^{-1}$ where $\boldsymbol{H}$ is the Jacobian of the vector of moments, $\mathbb{E}[\partial\boldsymbol{h}(\boldsymbol{z},\boldsymbol{\Lambda}_0)/\partial\boldsymbol{\Lambda}']$, and $\boldsymbol{\Omega} = \mathbb{E}[\boldsymbol{h}(\boldsymbol{z},\boldsymbol{\Lambda}_0)\boldsymbol{h}(\boldsymbol{z},\boldsymbol{\Lambda}_0)']$. Both expectations are replaced by sample averages and evaluated at the estimated parameter:

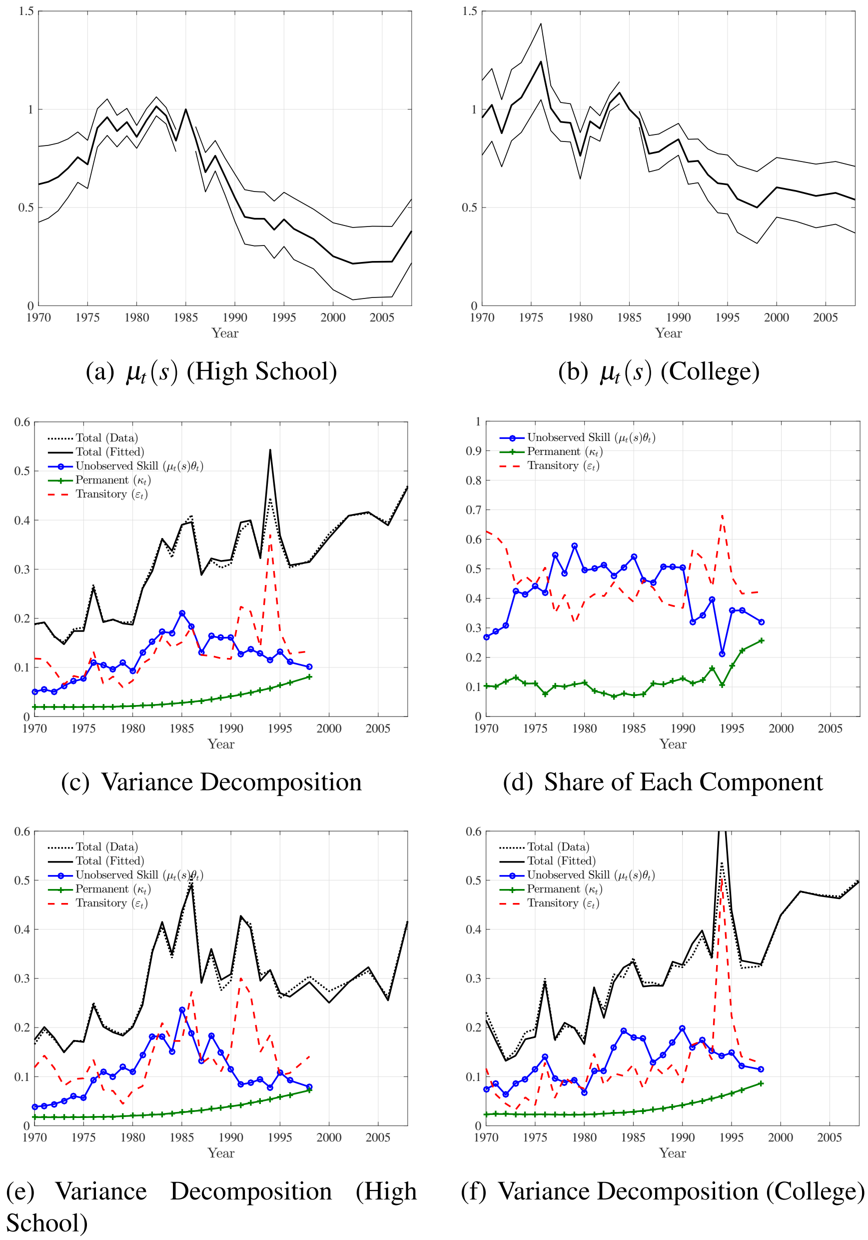
$$\begin{aligned}
\hat{\boldsymbol{H}} =& \frac{1}{N}\sum_{i=1}^{N}\frac{\partial\boldsymbol{h}(\boldsymbol{z}_i,\hat{\boldsymbol{\Lambda}})}{\partial\boldsymbol{\Lambda}'} = \boldsymbol{W}^{-\frac{1}{2}}\frac{\partial\boldsymbol{p}(\hat{\boldsymbol{\Lambda}})}{\partial\boldsymbol{\Lambda}'},\\
\hat{\boldsymbol{\Omega}} =& \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{h}(\boldsymbol{z}_i,\hat{\boldsymbol{\Lambda}})\boldsymbol{h}(\boldsymbol{z}_i,\hat{\boldsymbol{\Lambda}})',
\end{aligned}$$

where $\boldsymbol{W}^{-\frac{1}{2}} = \mathrm{diag}[\frac{N_1}{N},\ldots,\frac{N_M}{N}]$.

We can test r linear parameter restrictions $H_0 : \boldsymbol{R}\boldsymbol{\Lambda} = 0$ using Wald test statistic:

$$\mathrm{Wald} = N(\boldsymbol{R}\hat{\boldsymbol{\Lambda}})'(\boldsymbol{R}\hat{\boldsymbol{V}}\boldsymbol{R}')^{-1}\boldsymbol{R}\hat{\boldsymbol{\Lambda}} \xrightarrow{d} \chi_r^2.$$

C.2 Additional Estimates for Log Annual Earnings



(a) $\mu_t(s)$ (High School)

(b) $\mu_t(s)$ (College)

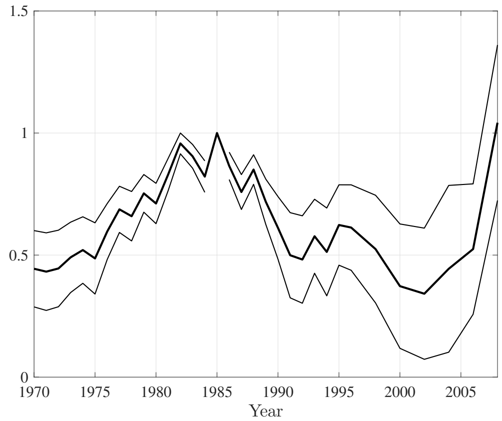
(c) Variance Decomposition

(d) Share of Each Component

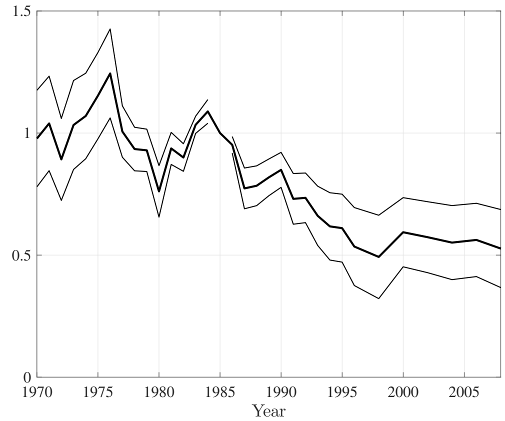
(e) Variance Decomposition (High School)

(f) Variance Decomposition (College)

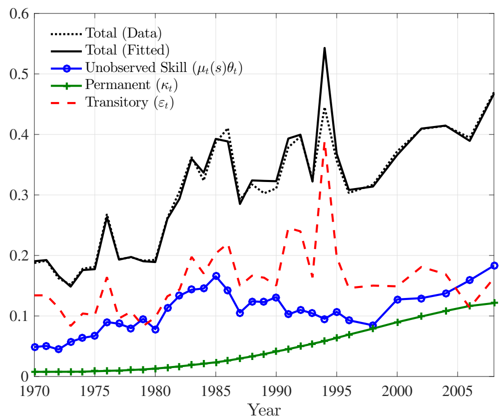
Figure C.1: Estimates with MA(5) Transitory Shocks



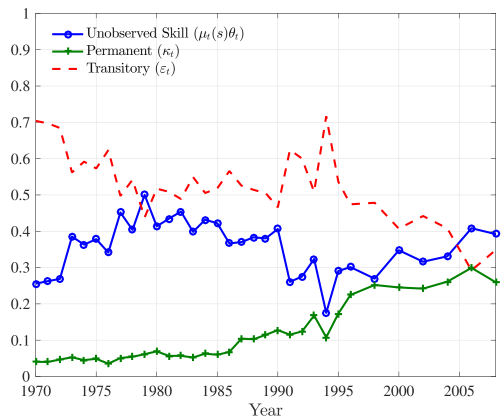
(a) $\mu_t(s)$ (High School)



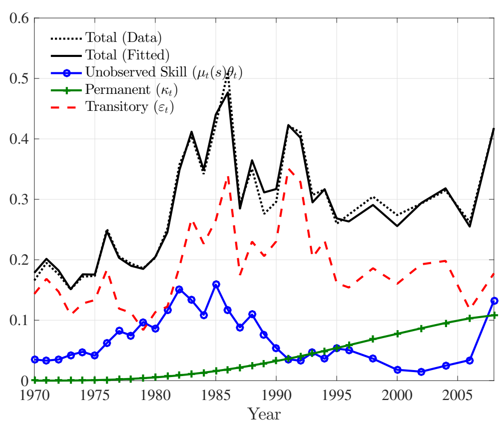
(b) $\mu_t(s)$ (College)



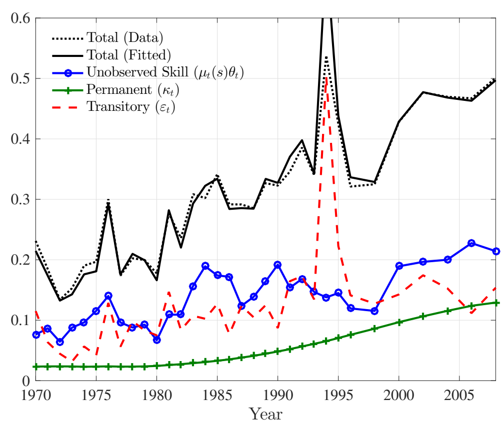
(c) Variance Decomposition



(d) Share of Each Component



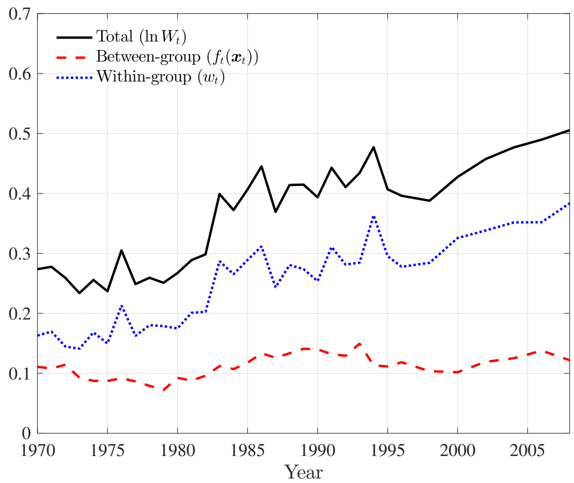
(e) Variance Decomposition (High School)



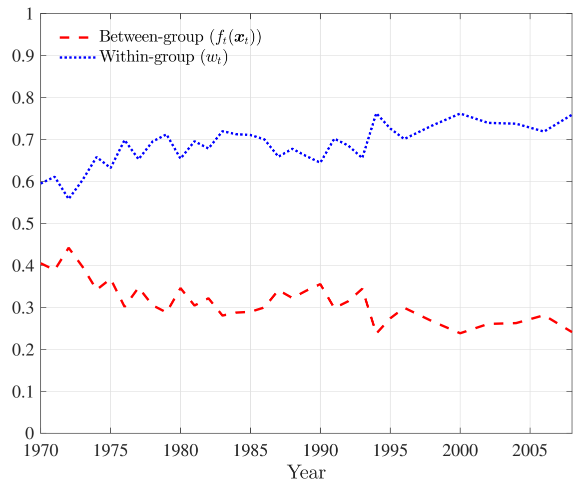
(f) Variance Decomposition (College)

Figure C.2: Estimates with ARMA(1,1) Transitory Shocks

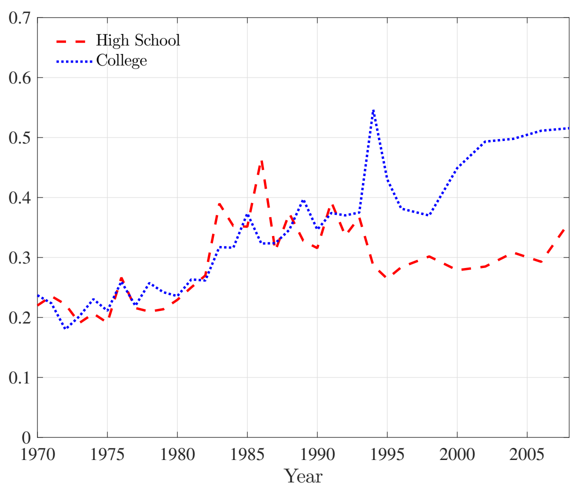
C.3 Estimates for Log Weekly Earnings



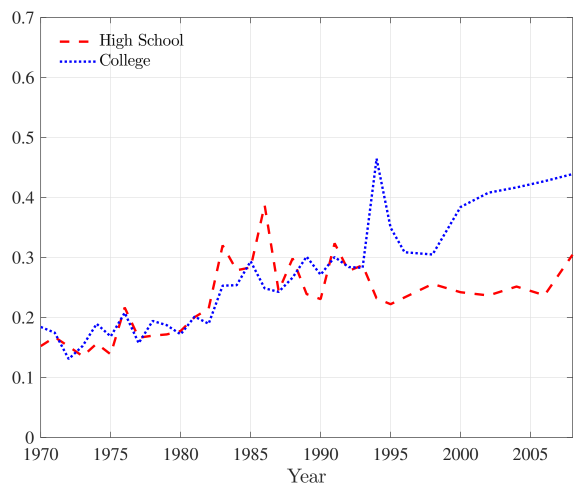
(a) Variance



(b) Share of Variance

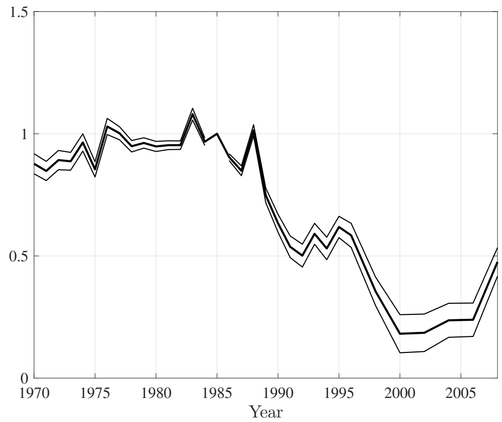


(c) Total Variance by Sector

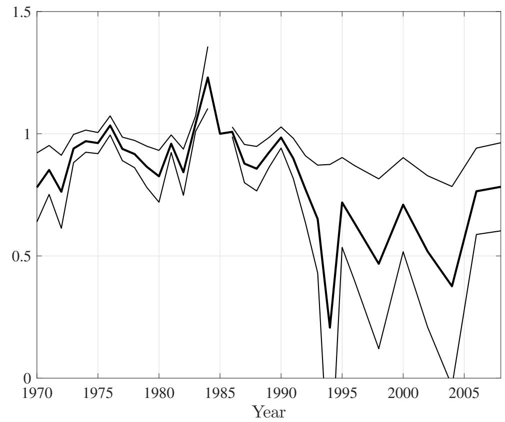


(d) Within-group Variance by Sector

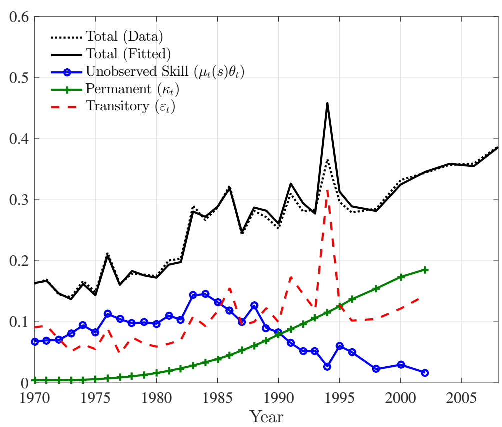
Figure C.3: Total, Between-, and Within-group Variance of Log Weekly Earnings



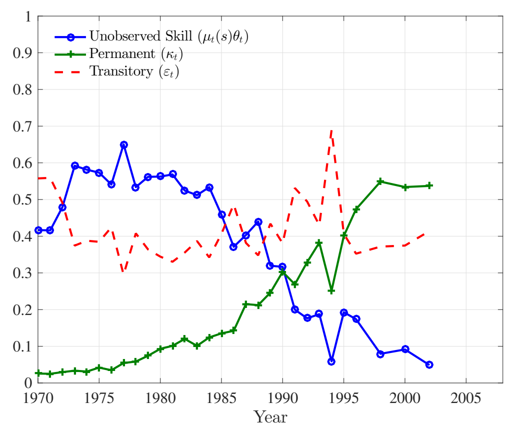
(a) $\mu_t(s)$ (High School)



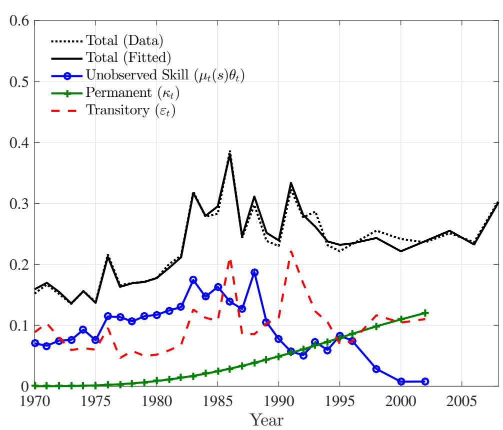
(b) $\mu_t(s)$ (College)



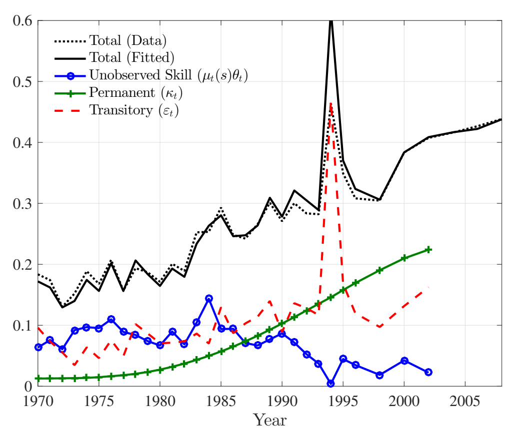
(c) Variance Decomposition



(d) Share of Each Component

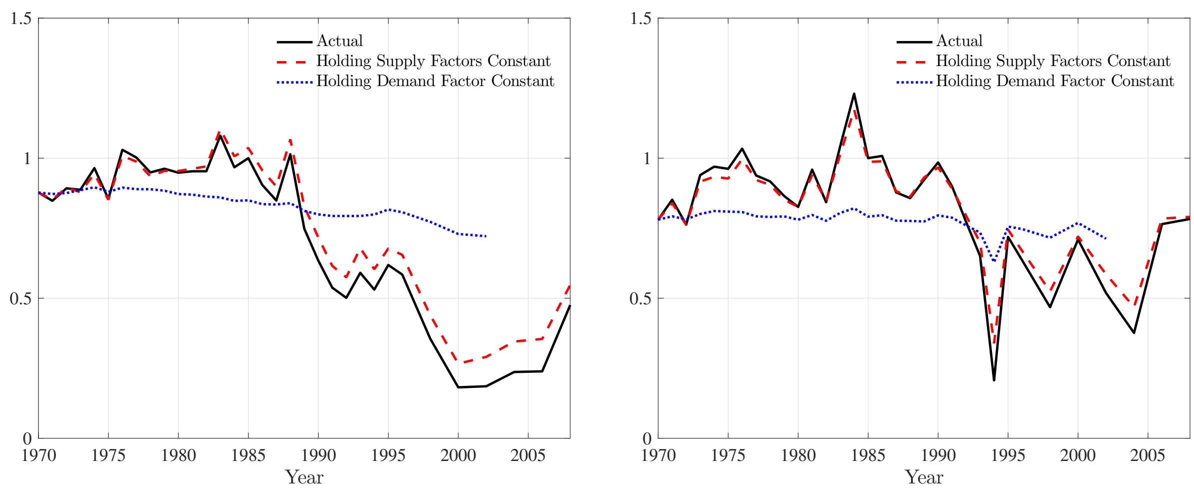


(e) Variance Decomposition (High School)



(f) Variance Decomposition (College)

Figure C.4: Estimates with Log Weekly Earnings and MA(3) Transitory Shocks



(a) High School

(b) College

Figure C.5: Actual and Counterfactual Returns ($\mu_t(s)$)

Curriculum Vitae

Name:	Youngmin Park
Post-Secondary Education and Degrees:	Yonsei University Seoul, South Korea 2001-2007, B.A. in Economics
	Yonsei University Seoul, South Korea 2007-2010, M.A. in Economics
	University of Western Ontario London, Ontario, Canada 2010-2016, Ph.D. in Economics
Related Work Experience:	Teaching Assistant University of Western Ontario 2010-2014
	Research Assistant University of Western Ontario 2012-2016
	Senior Economist Bank of Canada 2016-