Essays on Educational Choice and Intergenerational Mobility

Martin Nybom
To my favorite girls:
Hillevi, Ingrid and Kerstin
There are many that I wish to thank for their part in my PhD. First of all, I would like to thank my supervisors, Anders Björklund and Markus Jäntti. Anders has been a steady source of inspiration and encouragement for my research. His research skills are impressive and his acquaintance with everything from people in the profession, various literatures and journals, to how to write an academic paper has been of great help to me. I am fascinated by his never-fading curiosity for research, which in itself is a source of inspiration and energy to me and probably many others.

Markus has also been a fantastic supervisor. With his broad knowledge in statistics, econometrics and programming, he has been a perfect complement to Anders. As with Anders, Markus has always been there to answer questions, read drafts and give advice. He is also a professor with a great sense of humor. I deeply appreciate the way in which they both have treated me as an equal and with respect, despite titles and formal hierarchies.

I also want to thank my co-author Jan Stuhler. Our collaboration has given rise to two of the chapters in this thesis and hopefully there is more to come. Jan is very talented and hard working, but also a prestigeless, down-to-earth guy and it is a lot of fun working together. A special thanks also goes to Jan and his colleagues in London for hosting me during the spring 2012.

I would also like to thank Philipp Eisenhauer, Matthew Lindquist, Magne Mogstad, and many others, for helpful advice and discussions; Ann-Sofie Kolm and Jonas Häckner for mentorship during the first years; all the administrative and IT staff at both SOFI and the Department of Economics for their great work; and all other colleagues. Many of my fellow graduate students made the PhD more fun. Lisa Jönsson, Martin Olsson, Lotta Boström, Abdulaziz Shifa, Christian Odendahl, Susan Niknami, and Shon Ferguson deserve a special mention. Financial support from Swedish Council of Social and Working Life Research (FAS) is also gratefully appreciated.

Finally, I want to thank my family. My parents are great role models and have given me unconditional support through all years and all my endeavors. My brothers, my extended family, all great friends, and not the least Ingrid and Kerstin are great sources of non-academic happiness. But the greatest of gratitudes goes to Hillevi: my rock-solid supporter and the joy of my life.

Martin Nybom, Stockholm, December 2012
# Contents

Acknowledgements

1 Introduction

References

2 College Choice, Abilities and Lifetime Earnings: a Local IV Approach with Swedish Registry Data

2.1 Theoretical Illustration

2.2 Econometric Model

2.3 Data and Sample Restrictions

2.4 Empirical Results

2.5 Conclusions

References

3 Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation

3.1 The Intergenerational Mobility Literature

3.2 Measuring Income at a Certain Age

3.3 Empirical Evidence on Life-Cycle Bias

3.4 Other Methods to Address Life-Cycle Bias

3.5 Conclusions

References

4 Interpreting Trends in Intergenerational Income Mobility

4.1 The Literature

4.2 A Simple Model

4.3 Extensions and Sensitivity Analysis

4.4 Conclusions

References
1. Introduction

This thesis consists of four self-contained essays in labor economics. The first essay belongs to the classical literature on educational choice and the (monetary) returns to education. The subsequent three essays contribute to the burgeoning economics literature on intergenerational income mobility. Central themes throughout the thesis are economic inequality and opportunity; how they can be assessed, understood, and related to policy.

Modern research on the economics of education began in the 1950s with work by Jacob Mincer, Gary Becker, Sherwin Rosen, and others. They provided a new perspective on education by viewing it as a conscious investment that carries both costs and returns, and established the concept of human capital.\(^1\) All since, the literature on human capital investments and their returns has been central to the discipline.

Knowledge about the returns to human capital, and how they vary across people, is important for several reasons. It can help explain which people succeed and fail on the labor market and why; in particular by shedding light on which individual skills and background factors that are important for a beneficial post-school outcome. To gain deeper insight about the evolution of earnings inequality and its drivers, it is fruitful to consider the interacting roles of formal schooling and informally produced abilities. Moreover, in order to address policy questions such as whether college education should be expanded or not, it is necessary to consider individual differences and the full distribution of returns. Issues like these are addressed in Chapter 2.

The literature on intergenerational mobility is newer to the discipline. However, social scientists, and sociologists in particular, have long been concerned about the correlation between the economic status of parents and their children. This concern is largely motivated in terms of fairness and "equality of opportunity". The work by Gary Becker and Nigel Tomes on the one hand, and Anthony Atkinson and Arthur Goldberger on the other, came to establish

\(^1\)See Becker (2011) for a review. Becker also notes that the term human capital has been controversial. By some it was considered demeaning to human beings as it was claimed to treat them as machines or buildings rather than as real people with emotions and feelings. Becker, among others, has in subsequent work however also frequently highlighted the many fundamental differences between human and physical capital.
the topic within economics (see Piketty, 2000, for a comprehensive review).

The central research question – common across disciplines – is if economic inequality in one generation leads to inequality of opportunity in the next. Although the term “equality of opportunity” is used in a vague sense here, it is rather obvious that this is a huge question. Most tend to agree that income inequality is more justifiable if it is a product of individuals’ own choices and actions, rather than inherited from previous generations. Related to this is the more instrumental aspect of mobility – maintained by liberal sociological and political theories – that a high degree of mobility and equality of opportunity is necessary to sustain the liberal democracy (and, e.g., avoid radicalization).

Economists typically attempt to quantify the degree of mobility using a descriptive statistic called the "intergenerational earnings (or income) elasticity". This elasticity is, in fact, a measure of persistence, so that a high elasticity corresponds to low mobility, and vice versa. A large number of estimates of this elasticity have been produced in the last two decades as new data that span long time periods have become available. This empirical literature has greatly advanced our knowledge about income mobility and provided many intriguing results.

First, the degree of income mobility has been found to be much lower than what was previously believed. Second, some noteworthy cross-country differences in mobility have been recorded. For example, we know today that US mobility is lower than in many European countries, and in particular as compared to the Nordic ones (Björklund and Jäntti, 2009).

Thus far, much work has also been focused on techniques to measure this elasticity more accurately, and how to interpret various findings. The large-scale upward correction of elasticity estimates in the last decades is not only a consequence of better data, but also of an improved account of individual income dynamics and various measurement issues. A crucial example concerns how to estimate the intergenerational elasticity in lifetime income – which is typically seen as the relevant concept – when only short-run income measures are available. Chapter 3 in this thesis provides a critical analysis of the most commonly used procedure to deal with this.

The finding that the degree of income mobility is lower than we used to believe, and particularly in countries with relatively high levels of income inequality (such as the US and the UK), has been received with some concern. Concepts such as the “American dream” and the “land of opportunity” have come to be questioned in the US, and a fear of falling mobility in times of

---

2There are competing schools of thought. In particular, marxist theories generally postulate that mobility is low, leading to class struggle and instability (for a discussion, see Erikson and Goldthorpe, 1992).
increasing income inequality has sparked heated debates on the topic. There have also been some recent attempts to estimate trends in mobility, i.e., if the degree of income mobility is currently increasing or decreasing. The more reliable studies for the UK and the US indicate a small decrease in income mobility, although there are caveats related to this evidence. But apart from the difficulties associated with estimation, we may ask: how are we to interpret evidence on trends in mobility? In Chapter 4, we pose an intriguing question that directly relates to these recent political and academic debates on trends in mobility: are changes in today’s society responsible for current trends in mobility, or should we rather seek the answer in the past?

The result in the mobility literature that has gained most attention is the large differences in mobility across countries. The intergenerational income elasticity has, for example, been estimated to be about 0.5 in the US, but about half that size (or even less) in the Nordic countries. Persistence in many European countries are found to lie somewhere in between, although the UK is believed to be an outlier and closer to the US. But what are the underlying mechanisms behind these differences? Are they related to policy? It is currently a great research challenge to find out what mechanisms account for these differences and, in particular, whether or not certain policies and institutions are important for enhancing mobility and equality of opportunity. In Chapter 5, we contribute to this line of research by trying to account for some potential sources (e.g., child health, schooling and labor markets) of the mobility difference between Sweden and the UK. I will now briefly summarize each of the four essays in some more detail.

Short Outline of the Essays

The first essay, “College Choice, Abilities and Lifetime Earnings: a Local IV Approach with Swedish Registry Data”, is presented in Chapter 2. In the essay, I analyze the heterogeneity in individual returns to college education (i.e., the degree of variation across people). I build on three separate features that have been highlighted in the recent literature: the issue of self selection into education (or “sorting on the gain”); the importance of cognitive and noncognitive skills for educational and labor market outcomes; and the complex relationship between short- and long-run earnings measures.

3Exemples for the US are Wooldridge (2005), Wessel (2005), and Noah (2012). The political importance of the topic is also exemplified by a recent speech of Alan Krueger, Chairman of Council of Economic Advisers, who warned that intergenerational mobility should be expected to decline further as of the recent rise in income inequality in the US (speech delivered at the Center for American Progress, January 12th, 2012).
Much of the early literature assumed that the return to education was homogeneous, i.e., the same for everyone. However, recent research has showed that if individual returns differ across people, and people choose schooling based on their own return, then conventional empirical methods are likely to produce results that are hard to interpret. The most standard econometric technique, ordinary least squares (OLS), suffers from being based on the problematic assumption that selection into college is as good as random (after having controlled for observed characteristics). Thus, if individuals who choose to go to college systematically differ from those who do not – such that attending college is *endogenous* – OLS is likely to provide biased estimates. The highly popular instrumental variable (IV) method deals with non-random selection by employing a so-called *instrument* – a variable that is correlated with the endogenous variable (i.e., college choice), but not with the relevant outcome (i.e., earnings) – to predict the endogenous variable. Under self selection on individual returns (or “sorting on the gain”), different instruments will however produce different results. Thus, neither of these approaches may be particularly helpful for seeking answers to questions such as what the benefit would be from a marginal expansion of the college sector, or what the average benefit of going to college is for the whole population. I deal with this by using the Local IV framework that has recently been proposed by Heckman and Vytlacil (1999, 2005).

The study also relates to the literature on the role of human skills. I use high-quality data on cognitive and noncognitive ability from Sweden’s mandatory military enlistment to analyze how these relate to selection into college, and returns to college education on the labor market. Apart from having direct effects on earnings, these skills can differentially affect both subjective and time costs of acquiring schooling, and the subsequent returns on the labor market. It might be the case that one type of ability is more valuable in school, while the other proves more valuable on the labor market. For example, a high IQ person may easily pass the most demanding courses in college, but with the college degree in hand, this person may only succeed on the labor market with a sufficient set of social skills. Conversely, a socially competent person may have a bright future on the labor market, but may have a tough time managing school because of insufficient cognitive ability. Whether this is the case is ultimately an empirical question that I seek to examine. My findings indicate that cognitive and noncognitive abilities both have sizeable and comparable effects on the return to college, whereas selection into college is much more strongly related to cognitive ability. This pattern would, for example, be consistent with a finding that cognitive ability also has a larger effect on the non-monetary costs, or utility, from going to college.

For many central topics within labor economics, long-run rather than short-
run income is the variable of interest. However, most researchers tend to have access only to short-run measures spanning one or a few years. If the deviations of short-run from long-run measures were purely random, this would pose little problem in applications. Recent work by Haider and Solon (2006), however, has shown that there is a strong and systematic relationship over age in how well short-run measures approximate long-run measures, which they denote “life-cycle bias”. In the essay in Chapter 2, I use nearly career-long earnings measures to reduce the effect of this problem. I find that the lifetime return to a year of college education is on average close to 5 percent in the population as a whole. The importance of heterogeneity is manifested in that the return is significantly higher in the share of the population that went to college as compared to the (potential) return of those who did not. I also confirm the existence of a systematic age-relationship, or life-cycle bias, by estimating the returns using earnings at different ages as outcome variable.

This leads over to the second essay, “Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation”, that is presented in Chapter 3. In this essay, coauthored with Jan Stuhler, we analyze the consequence of life-cycle bias in the context of intergenerational income mobility. Life-cycle bias is essentially an issue of income growth heterogeneity; those that do well in terms of lifetime income tend to start out with low income but experience rapid income growth, while those that do worse tend to start out acceptably but subsequently suffer from slow income growth. Using short-run income for young people will therefore underestimate differences in lifetime income, while using such data for older people will tend to overestimate differences. Since available data tend to oversample old parents and young offspring, this bias has shown to be large for mobility estimates. Haider and Solon (2006) suggest that the age at which the deviations of short-run earnings are on average zero can be estimated. Using income observations from this age, typically found to be around ages 35-40, would then minimize the impact of life-cycle bias in applications. This approach has in recent years been highly influential in empirical work on intergenerational mobility.

However, we argue that this is based on partly premature interpretations. Although extreme cases of life-cycle bias can be safely avoided by employing the procedure of Haider and Solon, the bias is not necessarily small or minimized at those ages. We show that even at the age where the mis-measurement of long-run income is on average zero, life-cycle bias remains a problem if these mis-measurements are correlated within families. Thus, the problem of heterogeneous income profiles remains, in particular correlated income profiles within families.

We demonstrate this empirically using Swedish data based on income tax registers that cover an unusually long time period of almost 50 years. More
specifically, we construct nearly career-long measures of long-run income of fathers and their sons, and estimate a benchmark (or “true”) intergenerational income elasticity of about 0.27. We compare this benchmark with estimates from using short-run income observations for sons and can thus provide a directly estimated life-cycle profile of the intergenerational elasticity. Most importantly, we compare our benchmark to the outcome from using the approach advocated by Haider and Solon. The bias is found to be negative and substantial (corresponding to about a fifth of the benchmark) even when fully implementing their method. We also critically review a set of related procedures and conclude by presenting some recommendations on how to handle life-cycle bias based on our empirical results.

The third essay, “Interpreting Trends in Intergenerational Income Mobility”, is presented in Chapter 4. In this essay, also coauthored with Jan Stuhler, we examine how intergenerational income mobility responds to structural changes – for example, changes in policies or institutions – in a standard model of intergenerational transmission. More generally, the study sheds new light on the question of how large changes in our societies relate to measures of mobility and equality of opportunity. We believe that this is a question of potentially great importance, especially in light of ongoing debates in the US and elsewhere such as discussed in a previous paragraph. Our focus is on how the intergenerational income elasticity evolves over time. We deviate from the existing theoretical literature by explicitly analyzing the transition path of the elasticity between steady states (or equilibria). This is useful in the intergenerational context that has an inherent long-run time frame, especially if we want be able to interpret evidence on mobility trends which typically only stretch across a few decades, i.e., less than a typical generation.

We find that mobility not only depends on current, but also on past structural mechanisms, such that important societal changes may generate long-lasting mobility trends. Variation in current mobility levels across countries may thus be partly explained by differences in former institutions; current mobility trends may be caused by institutional changes in the past. We further find that transitions between steady states are often non-monotonic.4 For example, a shift towards a less plutocratic and more meritocratic economy will initially raise mobility but later tend to lower mobility.5 Declining mobility today may

4 An example of a non-monotonic transition is one with a trend that initially decreases but then subsequently increases (i.e., following a U-shape), or vice versa. A monotonic transition, on the other hand, implies either a constantly increasing or decreasing trend during the whole transition phase.

5 In a plutocratic economy the economic status of parents are of high importance for child outcomes. In a meritocratic economy, on the other hand, children’s own skills are more important for their own outcomes.
then not reflect a recent deterioration of equality of opportunity, but rather major improvements made in the past. Similarly, changes in the relative returns to different skills – for example, due to industrial or technological change – tend to generate short-term gains in mobility but subsequently also a negative trend. Times of change thus tend to be times of high mobility, while mobility is bound to decrease when the economic environment stabilizes.

Thus, looking beyond the relationship between policy and mobility in equilibrium results in implications that are of practical relevance for analysts of current trends in mobility. We believe that this is an important contribution to the literature. Our more general finding that past societal changes may cause lagged effects today is likely to also have implications for the interpretation of trends in other intergenerational associations (e.g., schooling, crime, etc).

The fourth and final essay, “The Role of Parental Income over the Life Cycle: a Comparison of Sweden and the UK“, is presented in Chapter 5. This essay, coauthored with Anders Björklund and Markus Jäntti, provides a cross-country perspective on intergenerational mobility by comparing Sweden and the UK. Research on intergenerational income mobility has shown stronger persistence (i.e., lower mobility) between parental and offspring’s income in the UK than in Sweden. In the study, we use a child-development perspective to examine this cross-country difference in mobility by accounting descriptively for various potential mechanisms.

We exploit the flexibility of Swedish registry data to mimic (for Sweden) a survey data set covering a British cohort born in 1970. We thus use plausibly comparable data sets for the two countries to explore whether the mobility differences show up early in offspring’s life in outcomes such as birth weight, height during adolescence, school grades, and final educational attainment. We do indeed find significant country differences in the association between parental income and these outcomes, and the associations are stronger in the UK than in Sweden. Therefore, we continue to investigate whether these differences are large enough to account for a substantial part of the difference in intergenerational mobility estimates. We find that the country differences in the intergenerational associations in birth weight and height are too weak to account for hardly any of the UK-Sweden difference in intergenerational income mobility. For the more traditional human-capital variables, grades and final education, however, our results suggest that the country differences can account for a substantial part of the difference in income persistence.

Although our results do not prove that differences in early childhood circumstances are not important for cross-country mobility differences, they suggest that future research should look broadly for possible determinants of these differences. The labor market strikes us as an important arena with many mechanisms that could potentially explain these differences.
References


2. College Choice, Abilities and Lifetime Earnings: a Local IV Approach with Swedish Registry Data*

Introduction

The literature on the returns to education is a classical but still vibrant part of labor economics. The main objective has traditionally been to estimate the return to education, either as the return to an additional year of schooling or to a specific level such as college (see Card, 1999, for a review). With the primary focus on solving the standard omitted-variable problem (Griliches, 1977), a large amount of estimates have been produced using standard quasi-experimental methods. However, both the notion of a homogeneous return, and the usefulness of such conventional methods, have been questioned in recent work that relates back to Willis and Rosen’s (1979) notion of self-selection on heterogeneous gains.

Carneiro et al. (2011) provide evidence that returns vary and that individuals select into college based on their idiosyncratic gain from doing so. In such cases, an additional identification problem arises that is distinct from standard selection bias. Angrist and Imbens (1995) and Heckman and Vytlacil (1999) have clarified that a standard instrumental variable (IV) approach then identifies a local average treatment effect (LATE) of potentially low or unclear external relevance.1 One remedy that enables one to interpret IV estimates is

*I thank Anders Björklund and Markus Jäntti for advice and encouragement. I am also grateful to Magne Mogstad, Philipp Eisenhauer, Tuomo Suhonen and seminar participants at SOFI (Stockholm University) and the Sunstrat workshop (Stockholm) for help and comments. Financial support from Swedish Council of Social and Working Life Research (FAS) is gratefully acknowledged.

1The external relevance of the LATE will vary by context and instruments: in some cases it can be nearly impossible to interpret, in other cases it can be both easily interpretable and of high relevance (e.g., if a certain reform constitutes the instrument and the effect of the reform is of primary interest).
to impose some structural assumptions; in particular, a clear specification of the choice model is crucial. A recent example of such an approach is Heckman and Vytlacil (1999, 2001, 2005) who use the generalized Roy model and Local IV (LIV) estimation.

Despite growing interest in this approach, existing applications are few and have been limited to survey data. A drawback of the method is its heavy data requirements, especially in its semiparametric versions. This paper sheds new light on the LIV approach by estimating the returns to college using a large registry-based data set of Swedish males. The paper has three objectives: (i) to explore the applicability of the LIV approach; (ii) to provide new estimates of the returns to college in Sweden while taking self selection into account; (iii) to examine the extent of heterogeneity and assess the relative importance of heterogeneity that is observable and unobservable to the researcher. In particular, I make use of high-quality data on cognitive and noncognitive ability to analyze their effects on estimated returns.2

Such evidence may shed new light on several important questions. To evaluate the effects of educational policy, it is of central interest to know both who gains from schooling and how much they gain. An example is the debate on the optimal size of the college sector, for which it is necessary to concentrate on the distribution of returns in the population, and in particular if returns for those at the margin of attending exceeds marginal costs. Exposing potential heterogeneity may also deepen the understanding of earnings inequality in general, and provide explanations to changes in returns and earnings inequality across time. A related question is if the rise in the college premium (and overall inequality) since the 1980s (see, e.g., Autor et al., 2008) primarily reflects a general shift in the demand for college-educated workers, or rather changes in the degree of “school-skill complementarity”, i.e., an increased demand for college-educated workers equipped with certain skills that are produced independent of college (Blackburn and Neumark, 1993; Taber, 2001).3

The approach of Heckman and Vytlacil has two cornerstones: (i) a choice-theoretic structure that defines each individual’s margin of indifference towards selecting into treatment and (ii) local identification of marginal treatment effects (MTE). The MTE, first introduced by Björklund and Moffitt (1987), implicitly contains information about heterogeneity, and can be used with appro-

2There is a well-known conceptual distinction between “abilities”, “skills” and, e.g., “test scores”. I will predominantly use the term abilities in this paper, although I recognize that my measures are imperfect measurements of underlying ability. Below when I describe the data I will however argue that these proxies are of unusually high reliability.

3More specifically, this school-skill complementarity refers to the cross-derivatives of formal schooling and “informal” skills or abilities in a typical production function.
appropriate weights to compute summary treatment effect parameters such as the population average treatment effect (ATE).

The idiosyncratic return has two parts: heterogeneity with respect to characteristics that are observable and unobservable to the researcher. Whereas previous work, including Carneiro et al. (2011), have been heavily focused on the role of unobservable heterogeneity, I also seek to address the role of heterogeneity with respect to observable characteristics. I devote special attention to two high-quality measures of cognitive and noncognitive ability that are based on high-stake tests from the mandatory enlistment to the Swedish military. I examine if these abilities influence returns and assess their relative importance. This will ultimately cast new light on the potential complementarity of formal education and different abilities.

The study thus relates to the ongoing debate on the role of cognitive and noncognitive ability for educational and labor market outcomes. The notion of individual ability has recently shifted from a onedimensional concept primarily related to IQ, such as in the single-skill signaling model (Arrow, 1973) and the $g$ factor (Herrnstein and Murray, 1994), to a multidimensional set of skills that especially recognizes the importance of personality or noncognitive ability.\footnote{For example, it has been found that class size affects later performance mainly by increasing noncognitive rather than cognitive ability (Chetty et al., 2011), and that early childhood programs such as Headstart and the Perry Preschool Program are effective in terms of long-run outcomes mainly by increasing noncognitive ability ( Heckman, 2005).} A growing literature concerns the reduced-form earnings returns to various measures of abilities: some focus on the returns to cognitive ability or IQ (e.g., Murnane et al., 1995; Cawley et al., 2001; Zax and Rees, 2002); others on the returns to noncognitive ability and personality traits (e.g., Nyhus and Pons, 2005; Groves, 2005; Mueller and Plug, 2006); and a third group of papers considers both types of ability jointly (e.g., Heineck and Anger, 2010; Lindqvist and Vestman, 2011).

These studies however rarely address how abilities transmit into earnings and wages, in particular via endogenous schooling choices. For example, one could ask if those with high IQ earn more because of their IQ as such, or because high IQ makes acquiring more education less costly or more beneficial, thus indirectly leading to increased earnings. Some evidence suggests that returns to education vary with respect to different measures of cognitive ability. In two recent Swedish studies, Nordin (2008) and Öckert (2012) estimate heterogeneous returns relying on a selection-on-observables assumption. The former finds that the return to a year of schooling increases at a diminishing rate by level of cognitive ability, the latter that the return to a year of college is steadily increasing with respect to secondary school GPA. Carneiro and Lee
(2009) estimate returns to college within the LIV framework using NLSY data and find that cognitive test scores are positively related to the return, and at an increasing rate. These studies do not consider the role of noncognitive ability. In a rare exception, Heckman et al. (2006a) consider multiple abilities jointly. They use NLSY data in a factor structure model and find that both cognitive and noncognitive ability are important in explaining several economic and non-economic outcomes within different schooling groups.

A key difference in the present paper is that I apply a semiparametric LIV estimator, with arguably less restrictive parametric assumptions. The main merit, however, is the data source. Whereas most related studies rely on NLSY data, this paper is based on especially rich registry data that cover a large and representative sample of the Swedish male population. The size of the data set provides more flexibility compared to previous studies. I use measures of close to full lifetime earnings to minimize life-cycle effects in my estimates (Bhuller et al., 2011). In contrast to most previous work, my ability measures are collected at a uniform age and prior to college, thus offsetting much of the concern regarding endogeneity in observed ability measures (Hansen et al., 2004). Moreover, my measure of noncognitive ability is unique in that it is an overall judgement of psychological capability that stems from a semi-structured interview with a certified psychologist. This is in contrast with the NLSY measures that are based on combinations of different self-reported answers about one’s personality.

I estimate the average treatment effect of a year of college to be a 4.8 percent increase in lifetime earnings. My findings are consistent with Carneiro et al. (2011) in that agents select into college based on their individual returns. The positive selection is manifested in that the difference between the average treatment effect on the treated and the untreated is consistently positive and statistically significant across specifications. The evidence on unobserved heterogeneity is however more ambiguous. Overall, the results suggest that observed characteristics do capture a significant part of the total heterogeneity that drives self selection. The heterogeneity in returns with respect to both cognitive and noncognitive ability is substantial, and of comparable magnitude. My findings thus corroborate the idea of substantial school-skill complementarities.

The rest of this paper is structured as follows. Section 2.1 provides a brief theoretical illustration of the college decision. In Section 2.2, I show how the structure of the generalized Roy model is used to define the MTE, and briefly discuss identification and estimation. I describe the data in Section 2.3 and present the results in Section 2.4. I conclude by discussing some implications of my findings.
2.1 Theoretical Illustration

The decision rule in the generalized Roy model can be seen as a reduced form of a more elaborate theoretical model. To fix ideas, I will illustrate the college decision with a discrete-choice model of college attendance that builds on those in Keane and Wolpin (2001) and Keane (2002). I expand the model by introducing heterogeneity such that cognitive and noncognitive ability is allowed to affect the costs and benefits of acquiring college education. In addition to the model assumptions listed in Keane (2002), I therefore assume that all agents are endowed with a set of abilities \( A \) that are allowed to impact on both the indirect time costs and the direct utility (or consumption value) of going to college, as well as college and non-college earnings capacity.\(^5\) The wage rate in period 1 is \( w_1(A) \). In the following periods, the wage rate is \( w_2(A) + \beta(A) \) if the agent attended school, and \( w_2(A) \) otherwise.

For a given ability realization \( A = a \), the value function conditional on college attendance is

\[
V_5 | a = \max_{\{h,b\}} u [y_1 + b + hw_1(a) - t, L - h - s(a)] + \\
\varphi(a) + \rho^{-1}u [w_2(a) + \beta(a) - rb, 1],
\]

(2.1)

and the value function for not attending is

\[
V_0 | a = \max_{\{h,b\}} u [y_1 + b + w_1(a), L - h] + \rho^{-1}u [w_2(a) - rb, 1].
\]

(2.2)

Utility maximization gives

\[
u_1(c_1, l_1) = r\rho^{-1}u_1(c_2, l_2),
\]

(2.3)

as the first-order condition for intertemporal optimality. Moreover, the first-order condition for intratemporal optimality is

\[
w_1(a)u_1(c_1, l_1) = u_2(c_1, l_1),
\]

(2.4)

\(^5\)In short, the assumptions listed by Keane imply that agents: are infinitely lived in discrete time; decide whether to attend college in period 1 with a direct cost of attending (e.g., tuition, transaction or moving costs) denoted by \( t \); face a discount factor \( \rho \) and interest rate \( r \); can borrow or save \( b \) in period 1 with fixed annuity payments \( rb \) from period 2 and onwards; devote time to work denoted by \( h \) and can work while in college; receive an exogenous transfer payment \( y_1 \) from their parents in period 1; receive non-monetary utility from college denoted by \( \varphi(a) \); receive utility from consumption \( c \) and leisure \( l \) through a utility function denoted \( u(c, l) \) which is concave in both arguments and \( L \geq l \geq 0 \); and inelastically supply one unit of labor after period 1 so that utility is \( u(c, 1) \).
given an interior solution. A first-order Taylor expansion of $V_S$ around $V_0$ at the point of indifference gives the (approximate) decision rule to attend college if and only if

$$\varphi(a) + r^{-1} \beta(a) u_1(c_2, l_2) - u_2(c_1, l_1) s(a) - u_1(c_1, l_1) t \geq 0. \quad (2.5)$$

By using the two first-order conditions, equation (2.5) can be rewritten as

$$\lambda_1^{-1} \varphi(a) + r^{-1} \beta(a) \geq w_1(a) s(a) + t, \quad (2.6)$$

where $\lambda_1 = u_1(c_1, l_1)$. This has several implications. First, parental transfers $y_1$ affect the decision only through the marginal utility of consumption. If there is no non-monetary utility from schooling, i.e., $\varphi(a) = 0$, parental transfers do not affect the schooling decision.\(^6\) If $\varphi(a) > 0$, then larger parental transfers increase attendance rates by decreasing the marginal utility from consumption and thereby increasing $\varphi(a)/\lambda_1$.\(^7\) Second, the higher the interest rate as well as the direct and time costs of getting a degree, the lower is attendance.

The focus of this paper is on the role of individual abilities. For simplicity, consider $a = a$ as a uni-dimensional ability realization such as a standard notion of cognitive ability. Differentiating the decision rule with respect to $a$ gives:

$$\frac{\varphi'(a)}{\lambda_1} + \varphi(a) \lambda_1^{-2} + r^{-1} \beta'(a) \geq w_1'(a) s(a) + w_1(a) s'(a). \quad (2.7)$$

Ability thus affects the decision rule through several mechanisms: (i) through the non-monetary utility (or consumption value) of college $\varphi'(a)$; (ii) indirectly through the effect of $w_1(a)$ on the marginal utility of consumption; (iii) through the monetary return $\beta(a)$; (iv) through the time cost of acquiring schooling $s(a)$; and (v), since $s(a)$ implies foregone earnings, through the monetary opportunity cost $w_1(a)$. For example, assume that ability increases the non-monetary utility of college, as well as both first- and second-period earnings (i.e., absolute advantage), and that it lowers the time cost. In this case the only mechanism that works against attending college is $w_1'(a) s(a)$.

\(^6\)This is however overturned if individuals are credit constrained (see, e.g., Keane, 2002). Constrained individuals use the parental transfer to pay the direct costs $t$ and to smooth consumption, and higher parental transfers require less borrowing in order to achieve intertemporal optimality. This dimension has been frequently studied in the US perspective, largely motivated by its perceived importance. It is not explicitly addressed in this paper since the absence of tuition fees suggests it is less important (but not necessarily irrelevant) in Sweden.

\(^7\)This is a well-know result and also pointed out in Keane (2002).
i.e., first-period earnings at a given time cost. If we believe that non-college earnings differ relatively little by ability so that \( w'(a) \) is small or even negative (i.e., comparative advantage), then the effect of ability on the attendance decision is unambiguously positive. Given positive partial ability effects, the first two terms in (2.7) imply that increased ability both increases \( \varphi(a) \), thus directly inducing more consumption of schooling, and lowers \( \lambda_1 \), which indirectly encourages the individual to consume even more schooling through \( \varphi(a) \). The term \( r^{-1}\beta' (a) \) is the effect of ability on the long-run earnings return from attending college, which I will examine in detail in the empirical part below.

Finally, consider the role of parental transfers as ability varies. Given that the marginal utility of consumption is positive and strictly concave, higher ability increases attendance more for individuals with less wealthy parents (if they give smaller transfers). Moreover, the transfer could be thought of as a composite of parental transfers \( y^p_1 \) and public tax-based student support \( y^g_1 \) such that \( y_1 = y^p_1 + y^g_1 \). An important implication is then that higher tax-based student support crowds out the attendance effect of parental transfers. I return in Section 2.4 to some of the implications of this model to interpret my empirical findings.

### 2.2 Econometric Model

Empirical work on the returns to schooling traditionally seeks to estimate variations of the equation

\[
Y = \alpha + \beta S + \epsilon, \tag{2.8}
\]

where \( Y \) denotes the post-school wage or earnings, and \( S \) can be years of schooling, a vector of schooling levels, or an indicator variable of, e.g., college education. Under familiar assumptions, an OLS regression of \( Y \) on \( S \) yields an unbiased estimate of \( \beta \).\(^8\) As discussed above, standard selection bias (\( S \) correlated with \( \epsilon \)) has typically received most attention, but more recently the issue of heterogeneous “sorting on the gain” (\( S \) correlated with \( \beta \)) has been advanced. Heckman et al. (2006b) distinguish between non-essential and essential sources of heterogeneity, where the former implies sorting by observable and the latter by unobservable characteristics. In what follows, I will

\(^8\)For example, if \( S \) is a college dummy, and equation (2.8) includes control variables \( X \), the OLS estimator gives \( E[Y_1 | X,S = 1] - E[Y_0 | X,S = 0] \). The necessary assumptions are: no selection bias, i.e., \( \text{Cov}(\epsilon,S | X) = 0 \); no self selection based on unobserved heterogeneous gains, i.e., \( \text{Cov}(\beta,S | X) = 0 \); and the parametric assumptions of OLS.
use the terms observed and unobserved in reference to the empirical analyst’s perspective. Although a more advanced IV approach may in principle take the former into account, the latter will typically cause IV estimates (i.e., the LATE) to diverge from average and marginal treatment effects.\(^9\)

The econometric method that I apply originates from research by Heckman and Vytlacil (1999, 2001, 2005) in which they use the marginal treatment effect (MTE) to identify and unify different treatment effect parameters under heterogeneous sorting. The MTE, originally introduced by Björklund and Moffitt (1987), is the average treatment effect of those at the margin of indifference for selecting into treatment. This margin of indifference can be identified by imposing the structure of the generalized Roy model on the selection equation.

The Generalized Roy Model

The generalized Roy model offers a discrete-choice framework for policy analysis in which agents self select into treatment based on their expected gains.\(^10\)

The decision rule in the binary version of the model can be seen as the reduced form of an economic model of college attendance such as the one outlined in Section 2.1.

Let \(S\) be a binary choice indicator with \(S = 1\) if the agent selects into treatment and \(S = 0\) if not. Moreover, let the potential outcomes in the two states be

\[
Y_j = \mu_j(X, A) + U_j, \quad \text{for } j = 0, 1
\]

(2.9)

where \(X\) is a set of observed regressors, \(A\) is a set of observed ability measures, \(\mu_j\) are unknown functions, and \(U_j\) are unobserved random variables that need not be orthogonal to \(X\) and \(A\). The observed outcome can be written in switching regression form:

\[
Y = SY_1 + (1 - S)Y_0.
\]

(2.10)

Plugging (2.9) for both states into (2.10) gives

\(^9\)Under observable heterogeneity, IV may recover average and marginal treatment effects provided that the functional forms of the regression equations are sufficiently flexible so that the LATE coincides with these parameters. This would, for example, imply non-linearities such as interactions between “sorting variables” and the endogenous treatment variable. The literature appears unsettled on the issue whether two-stage least squares IV is an appropriate estimator in such cases.

\(^{10}\)The original version of the model is due to Roy (1951). Although different in style and notation, the essence of the model is similar to the one in Willis and Rosen (1979).
The individual benefit of treatment is defined as the difference between potential outcomes $Y_1 - Y_0 = \mu_1(X, A) - \mu_0(X, A) + U_1 - U_0$. Thus, the average treatment effect conditional on $X = x$ is given by $\text{ATE}(x) = \mu_1(x) - \mu_0(x)$, and the average ability-specific treatment effect conditional on $X = x$ and $A = a$ is $\text{ATE}(x, a) = \mu_1(x, a) - \mu_0(x, a)$. Moreover, conditioning on $S = 1$ or $S = 0$ defines the average treatment effect of the treated (ATT) and of the untreated (ATU), respectively.\footnote{We have $\text{ATT}(x) = \text{ATE}(x) + E(U_1 - U_0 \mid S = 1, X = x)$ and $\text{ATU}(x) = \text{ATE}(x) + E(U_1 - U_0 \mid S = 0, X = x)$, and, for the ability-specific effects, $\text{ATT}(x, a) = \text{ATE}(x, a) + E(U_1 - U_0 \mid S = 1, X = x, A = a)$ and $\text{ATU}(x, a) = \text{ATE}(x, a) + E(U_1 - U_0 \mid S = 0, X = x, A = a)$.}

Let $I_S$ denote the (expected) net benefit of selecting into college. An individual’s decision rule can then be written as a standard latent variable discrete choice model (see, e.g., Willis and Rosen, 1979) of observed and unobserved variables:

\[
I_S = \mu_S(Z) - V, \quad S = 1 \text{ iff } I_S \geq 0.
\]

The individual thus selects into college if $I_S \geq 0$, and otherwise not. $Z$ is an observed vector which may include some or all of the components of $(X, A)$, but also components $Z \setminus (X, A)$ that are excluded from $(X, A)$. $V$ is unobserved and represents the individual (latent) resistance to select into college. Moreover, assume that $V$ is a continuous variable with a strictly increasing cumulative distribution $F_V$, and that $(U_0, U_1, V)$ are statistically independent of $Z$ conditional on $(X, A)$. $Z \setminus (X, A)$ thus work as exogenous cost-shifters that affect the outcome only through the college decision. At this stage, no independence condition is required for the common elements of $Z$ and $(X, A)$.

Finally, let the propensity score $P(z) \equiv \Pr(S = 1 \mid Z = z) = F_V[\mu_S(z)]$ denote the probability of college attendance conditional on $Z$, with the conditioning on all common elements in $(X, A)$ held implicit. Define $U_S = F_V(V)$ such that $U_S$ corresponds to the quantiles of $V$ and is by construction uniformly distributed. The latent index can be rewritten using $F_V(\mu_S(Z)) = P(Z)$ so that $S = 1$ if $P(Z) > U_S$. Within this framework, $P(Z)$ and $U_S$ represent the observed and unobserved inducement to college education: the higher is $P(Z)$, the more inducement to attend college from the observables in $Z$; the higher is $U_S$, the larger the unobserved resistance to college. For a person of high $U_S$,
it thus takes a high inducement from $Z$ to attend college. If $P(Z) = U_S$, the individual is indifferent to attending.

Identifying the Marginal Treatment Effect

The marginal treatment effect (MTE) is defined by

$$\text{MTE}(x, a, u_S) \equiv E(Y_1 - Y_0 \mid X = x, A = a, U_S = u_S)$$

and can be identified across the support of $U_S$. Since this is conditional on $(X, A)$, it is the local perturbation of the propensity score that is induced by $Z \backslash (X, A)$ at quantile $u_S$ that provides identification. The return to college can be recovered for persons on the margin of indifference at all quantiles of $U_S$ within the support of $P(Z)$. Persons with high $P(Z)$ identify the return for those with high $U_S$, and vice versa. Local perturbations at high levels of $P(Z)$ induce persons with high $U_S$ (i.e., high unobserved resistance) to change their treatment status. Those with low $U_S$ are already in treatment for such values of $P(Z)$. Local perturbations at low $P(Z)$ induce those with low $U_S$ to change their treatment status while those with high $U_S$ remain out of treatment for such values of $P(Z)$. If the treatment effect is homogeneous with respect to $U_S$, then the MTE as a function of $U_S$ would be flat. If the MTE correlates with $U_S$ conditional on $(X, A)$, then there is unobserved heterogeneity.

A key virtue of the MTE approach is that summary parameters – e.g., the ATE, ATT, ATU and conventional IV effects (LATE) – can be recovered using estimates of the MTE and appropriate weights (Heckman et al., 2006b). The LATE is in this framework a discrete form of the MTE, defined on a particular section of $U_S$. With full support on $U_S$, the MTE can be estimated at each $u_S$ quantile on the unit interval, with $P(z) = u_S$ defining the margin of indifference in equation (2.12). An average MTE at each level of $U_S$ can be obtained by integrating over the joint distribution of $(X, A)$ conditional on $U_S = u_S$. Integrating over the (uniform) distribution of $U_S$ yields the unconditional ATE. The same procedure, conditioning on $S = 1$ or $S = 0$, gives the unconditional ATT and ATU, respectively.

The ability-specific ATE can be obtained by integrating over $X$ and $U_S$ while conditioning on $A = a$, such that $\text{ATE}(a) = E_{X, U_S \mid A = a} [\text{MTE}(x, a, u_S)]$ traces out the ATE at a given value of $A$. However, the interpretation of $\text{ATE}(a)$ as the contribution of ability to the treatment effect will be confounded if there are heterogeneous treatment effects with respect to variables in $X$ that correlate with $A$. An alternative and potentially more robust procedure is instead to impose a linear and separable version of $\mu_1(X, A) - \mu_0(X, A)$ in equation (2.11) with $\mu_0(X, A) = X\delta_0 + A\gamma_0$ and $\mu_1(X, A) = X\delta_1 + A\gamma_1$. The ability-specific
treatment effect, purged of other heterogeneous treatment effects related to observed covariates, is then given by $\gamma_1 - \gamma_0$.

Estimation using Semiparametric LIV

The MTE can be estimated using local instrumental variables (LIV) as proposed by Heckman and Vytlacil (1999, 2001, 2005). This approach relies on the fact that the expected value of $Y$ depends on the propensity score $P(Z)$, so that $P(Z)$ serves as a local IV. Heckman and Vytlacil show that

$$\Delta_{LIV}(x, a, u_S) = \frac{\partial E(Y \mid X = x, A = a, P(Z) = p)}{\partial p} \bigg|_{p = u_S} = MTE(x, a, u_S).$$

(2.13)

The computation of the MTE thus involves the estimation of the partial derivative of the conditional expectation of $Y$ with respect to $p$. For my empirical analysis, in which I consider the linear and separable version of the model, the expected value in (2.13) can be written as

$$E(Y \mid X = x, A = a, P(Z) = p) = x \delta_0 + a \gamma_0 + p[x(\delta_1 - \delta_0) + a(\gamma_1 - \gamma_0)] + K(p),$$

(2.14)

where $K(p) = E(U_1 - U_0 \mid S = 1, P(Z) = p)$. The expression shows that the expected outcome is determined by three components: non-college earnings, the part of the treatment effect that is attributed to observed characteristics, and $K(p)$, which represents the effect that is attributed to unobserved characteristics. Using equations (2.13) and (2.14), the estimator becomes

$$MTE(x, a, u_S) = x'(\delta_1 - \delta_0) + a'(\gamma_1 - \gamma_0) + \frac{\partial K(p)}{\partial p} \bigg|_{p = u_S}.$$  

(2.15)

In order to compute the MTE, I thus need to estimate $(\delta_1 - \delta_0)$, $(\gamma_1 - \gamma_0)$ and $\partial K(p)/\partial p$. The full estimation procedure that I implement involves several steps. In the first stage, I estimate the college choice equation using a probit model to obtain estimates of $P(Z)$. I then estimate the coefficients in equation (2.14) using a semiparametric version of the double residual regression procedure. Specifically, I estimate separate local linear regressions of each of the regressors and the outcome variable on the predicted propensity score.

12The implementation follows the guidelines for LIV estimation (“Semiparametric Method 1”) presented in Heckman et al. (2006b), and in more detail, at: http://jenni.uchicago.edu/underiv.
Estimates of $\delta_0$, $\gamma_0$, $(\delta_1 - \delta_0)$, and $(\gamma_1 - \gamma_0)$ are then obtained by regressing the residual associated with the outcome on the residuals associated with the variables in $(X, A)$. Finally, with these estimates at hand, $\partial K(p)/\partial p$ can be estimated using standard nonparametric techniques.\footnote{This term is also estimated using local linear regression across the common support of $P(Z)$, i.e., the subset of $P(Z)$ for which I obtain positive frequencies in both $S = 1$ and $S = 0$. I use an Epanechnikov kernel function with a bandwidth of 0.1.}

The LIV estimate of MTE$(x, a, u_S)$ is computed by plugging in the resulting parameter estimates into equation (2.15). I obtain estimates of the summary treatment effects by applying the respective weights obtained from the data (see Appendix A.1).

An alternative to the semiparametric estimation approach is to impose parametric assumptions on the unobservables and derive the expression for the MTE. This approach, relying on joint estimation of the choice and outcome equations as an endogenous switching regression, is more in line with the work of Willis and Rosen (1979) and Björklund and Moffitt (1987). Similar to above, the parametric MTE estimates can be used together with weights to compute summary treatment effects. As a comparison to the results from the semiparametric LIV, I will also present estimates from a parametric version of the model that assumes joint normality of $(U_0, U_1, V)$. In this case, the MTE can be written as

\[
\text{MTE}(x, a, u_S) = x'(\delta_1 - \delta_0) + a'(\gamma_1 - \gamma_0) - (\sigma_{1V} - \sigma_{2V})\Phi^{-1}(u_S), \tag{2.16}
\]

where $E(U_1 - U_0 \mid U_S = u_S) = - (\sigma_{1V} - \sigma_{2V})\Phi^{-1}(u_S)$ and has a variance that is normalized to one.\footnote{Moreover, $\sigma_{1V} = \text{Cov}(U_1, V)$, $\sigma_{0V} = \text{Cov}(U_0, V)$, and $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function.}

\footnote{Notice that, if $n_X + n_A$ denotes the total number of variables in $(X, A)$, this step involves the estimation of in total $2 \times (n_X + n_A) + 1$ regressions. This is since equation (2.14) also contains interaction terms between each of the variables in $(X, A)$ and the propensity score. The local linear regressions are estimated for the set of values of $p$ that is contained in the support of $P(Z)$ using a kernel function with a bandwidth of 0.4.}
2.3 Data and Sample Restrictions

The data set is based on a representative sample of Swedish men born 1951-1957 and is obtained by merging several registers from Statistics Sweden using unique personal identifiers. These registers include data on earnings, ability test scores, educational attainment, personal and local labor market characteristics, and data on family members. The analysis is restricted to men since the ability data come from military enlistment registers.

Measure of Lifetime Earnings

To construct a measure of lifetime earnings, I make use of individual information on annual labor earnings from tax-declaration files for the years 1968-2007. These data come with a number of advantages: they are almost entirely free from attrition; pertain to all jobs; are not right-censored; and are believed to suffer relatively little from reporting errors. I approximate each individual’s log lifetime earnings by the log of the mean of all non-missing earnings observations over ages 20-50 (i.e., ages for which earnings are observed for all cohorts).

That the data allow for a nearly career-long earnings measure is unusual. In estimation of the returns to education in general, and when based on an explicit decision model in particular, the relevant outcome (or maximand in the model) is the stream of earnings across the lifetime. As a contrast, it has been standard in the literature to use single-year or short-run outcome measures, often from around age 30. If the age-earnings relationship is related to the components in equation (2.9), such estimates will not in general be unbiased, because heterogeneous earnings profiles cause non-classical measurement error when short-run earnings measures are used as proxies for lifetime earnings.

---

16 The raw sample is based on a random draw of approximately a third of the Swedish male population. The 1951-1957 cohorts are chosen for two primary reasons: they enable me to use nearly career-long measures to approximate lifetime earnings, and they are the earliest cohorts with data from the military enlistment.

17 The actual measure is “Inkomst av tjänst” in Swedish. The measure includes labor earnings and labor-related benefits such as parental leave benefits. The measure does not include income from self employment. I discount each annual observation to a present value at age 20 using an annual rate of 0.02. Using the slightly different measure “Arbetsinkomst”, which includes income from self employment, yields very similar results.

18 This is particularly true for the numerous studies that rely on NLSY data, including examples such as Heckman et al. (2006a) and Carneiro et al. (2011).

19 The standard practice of using short-run measures from around age 30 appears to be notably precarious when estimating the returns to college. Individuals with longer
such “life-cycle bias” can have large quantitative effects on estimates has been shown in recent studies (e.g., Nybom and Stuhler, 2011; Bhuller et al., 2011). Since I use earnings data that span over 31 years, my estimates should be relatively unaffected by such bias.

Measures of Cognitive and Noncognitive Abilities

An attractive feature of the data set is that it includes information from the mandatory military enlistment’s tests of cognitive and psychological (noncognitive) ability. The enlistment typically takes place at age 18 and includes two days of physical, intellectual, and psychological tests and evaluations.\(^{20}\)

The measure of cognitive ability is based on scores on a test of general intelligence that has been conducted since the 1940s. The test consists of four subtests of logical, verbal, and spatial ability, as well as technical comprehension, each graded on a discrete scale from 1 to 9. The scores on the subtests are transformed to a discrete general variable between 1 and 9 that follows a Stanine scale.\(^{21}\)

The measure of noncognitive ability is based on standardized interview-based evaluations made by certified psychologists. In the interview, the enlistee’s psychological profile and capacity to fulfill the requirements of military duty are evaluated. Central to this is the ability to cope with stress and contribute to group cohesion. Other valued traits include willingness to assume responsibility, independence, emotional stability, outgoing character, persistence, and the ability to take initiatives. Motivation for doing military service is not considered. The interview is semi-structured in the sense that the schooling may then recently have entered the labor market, which is reflected in relatively low, and possibly also noisy earnings as of higher rates of on-the-job investments and job switching. This is especially a concern in a country like Sweden where higher education is on average both commenced and finished at later ages than in comparable countries.

\(^{20}\)For the male cohorts born 1951-57, only a tiny fraction were exempted from the enlistment, mainly because of physical or psychological disability. Although the test scores are drawn from a similar age for all, concerns about the joint causality of schooling, latent skills, and test performance (see Hansen et al., 2004) leads me to restrict the sample to individuals with similar pre-academic educational attainment at the age of enlistment by excluding those without a degree from an academic high school track. I examine the sensitivity of the results by dropping this restriction in Section 2.4.

\(^{21}\)Carlstedt (2000) provides a detailed overview of this test as well as the Swedish military’s history of psychometric testing. Carlstedt also provides evidence that the test is a good measure of general intelligence, thus in contrast with tests that tend to measure the more malleable concept of crystallized intelligence (as noted by Lindqvist and Vestman, 2011).
chologist follows a manual that states topics to discuss and how to grade answers. Scores are given on four subscales and as an overall assessment that follows a Stanine scale between 1 and 9.

This variable is valuable in two ways: it provides a general “omnibus” measure of noncognitive ability and it is based on a psychologist’s experience from a personal encounter with the individual, which is likely to capture more aspects of a personality than what can be deduced from survey questionnaires. Moreover, both the measure of cognitive and noncognitive ability benefit from high comparability and cover large samples.\(^{22}\)

**Data on Education and Background Characteristics**

The education data come from a population register that describes the highest level of education, in what year it was achieved, and from what type of study program. I measure educational attainment up until 1990 when the men in the sample were 33-39 years old. The college indicator takes a value one if an individual had a minimum of three years of college studies.\(^{23}\) To be able to show estimates in annualized values, I impute a measure of years of schooling based on the highest level.\(^{24}\)

From the registers, I also obtain data on personal characteristics and family background. From censuses, I get data on birth date, country of birth, and geographical residency at different ages. I use a multigenerational register to identify (biological) family members and thus obtain data on birth order, number of siblings, father’s and mother’s years of schooling, and father’s earnings.\(^{25}\)

**Instrumental Variables**

As exogenous cost shifters in the choice equation, I use distance from the place of residence to the closest university and short-run fluctuations in unemployment and average earnings in the municipality of residence at the end of high

\(^{22}\)For more information on the military enlistment data, see the excellent data description in Lindqvist and Vestman (2011).

\(^{23}\)In Section 2.4, I analyze the sensitivity of the results by using a less strict definition of the college variable (at least one semester).

\(^{24}\)Around 1970, when most of the men in the sample were about to make their college choice, admission to higher education in Sweden was largely unrestricted. Most higher education was open to anyone with a high school degree and many faculties had no formal application procedure. There were also no tuition fees and the system of student aid was generous. See Erikson and Jonsson (1993) for a detailed account of the history of the Swedish system of higher education.

\(^{25}\)I compute father’s earnings as the average of non-missing annual earnings in the years 1968-1972.
school. Distance to college was first used as an instrument by Card (1995) and has later been frequently utilized, for example in applications of the LIV procedure such as Carneiro et al. (2011). I construct a continuous measure of typical travel distance by car between the central town of the local municipality in which the individual resided in 1965 and the closest university city. This is in contrast with most of the previous literature that has either used dummy variables for whether a college is located in the home county, or, in the case of continuous measures, more crude measures such as “as the crow flies” generated from geographical coordinates.

The unconditional exogeneity of such distance instruments has been questioned in studies based on both US (Cameron and Taber, 2004) and Swedish data (Kjellstrom and Regner, 1999). Given the recommendations in these studies, it is thus crucial that I condition this instrument on measures of ability and family background.

My instruments also include measures of short-run fluctuations in the local labor market at the end of high school, conditioned on permanent local labor market conditions. The underlying idea is that while both current (or short-run) and permanent local labor market conditions are in the individual’s information set at the time of the college decision, the current conditions do not contain any additional information about the future conditioned on the permanent component. If this holds true, then such innovations in the local labor markets can be excluded from the outcome equation. As measures of current conditions, I use the unemployment rate and average earnings in the municipality of residence at age 20. As permanent measures, I use municipal averages of unemployment and earnings over the years 1968-1988. Although this type

26 Other papers that have used variations of this instrument include Kling (2001), Currie and Moretti (2003), Cameron and Taber (2004), and Carneiro and Lee (2009).

27 In 1965 there were 998 local municipalities in my sample, thus ensuring large variation in the distance measure. I consider six university cities: Stockholm, Uppsala, Gothenburg, Lund, Umeå, and Linköping. Up until the late 1970s, nearly all Swedish college students studied in one of these cities. The measure is calculated using the website eniro.se, which is a tool similar to Google Maps. Using a measure of shortest travel time in minutes instead of distance yields similar results.

28 Previous papers that have done so include Cameron and Heckman (1998), Cameron and Taber (2004), Carneiro and Lee (2009), and Carneiro et al. (2011). As Cameron and Taber (2004) argue, the impact of these variables on schooling choice is theoretically ambiguous. On the one hand, better labor market conditions increase the opportunity cost of schooling. On the other hand, a better labor market also increases the resources of credit constrained households, thus promoting educational attainment.

29 In effect, I use “non-employment”, i.e., one minus the employment rate in the local working-age population, as my measure of unemployment.
of instrument has been frequently used in the recent literature, one may still be concerned about whether current conditions actually can be excluded from the outcome equation. If individuals would put a higher weight on current measures when forecasting the benefits of the choice alternatives, then these would also enter the outcome equations. To address this concern, I will also provide estimates that do not include these among the instruments.

Model Specification and Sample Statistics

The linear-in-parameter representations of \((X, A)\) include linear and quadratic terms of cognitive and noncognitive test scores \((A)\), mother’s and father’s years of schooling, father’s log earnings, number of siblings, permanent local earnings and unemployment, as well as region and cohort dummies \((X)\). The exclusion restrictions that enter \(Z \setminus (X, A)\) are linear and quadratic terms in local short-run earnings and unemployment, and a cubic polynomial of the distance measure. Following Carneiro et al. (2011), I interact the instruments with linear terms in cognitive and noncognitive test scores, mother’s years of schooling, and number of siblings. Recognizing that the effect of distance may vary depending on region, I also interact these with the regional dummies.\(^{30}\) The sample statistics are presented in Table 2.1.

2.4 Empirical Results

The main objective of this paper is to examine the importance of heterogeneity in the returns to college using the semiparametric LIV estimator. However, it is instructive to first consider more standard approaches, which will then serve as a comparison later in the paper.

Results using Conventional Methods

A natural point of departure is to consider standard OLS estimates. I estimate different versions of eq. (2.8), with control variables either only entering independently, or also interacted with the college dummy \(S\). Moreover, to illustrate

\(^{30}\) The six regions are defined as “university regions” so that all municipalities that share the same “closest university” constitute a region. Moreover, it has been common in applications of the LIV method to include variables in the outcome equation that are not in the selection equation as a means to increase precision. Carneiro et al. (2011), for example, include experience as well as local earnings and unemployment in the municipality of residence at prime age. As these are post-determined, and thus likely endogenous, I avoid including such variables in the baseline analysis. I instead present estimates from such specifications in the sensitivity analysis.
Table 2.1 Summary Statistics by Treatment Group

<table>
<thead>
<tr>
<th></th>
<th>$S = 1$</th>
<th></th>
<th>$S = 0$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Log lifetime earnings</td>
<td>12.07</td>
<td>0.50</td>
<td>11.93</td>
<td>0.50</td>
</tr>
<tr>
<td>Cognitive test score</td>
<td>0.90</td>
<td>0.78</td>
<td>0.54</td>
<td>0.80</td>
</tr>
<tr>
<td>Noncognitive test score</td>
<td>0.46</td>
<td>0.98</td>
<td>0.33</td>
<td>0.92</td>
</tr>
<tr>
<td>Mother’s years of education</td>
<td>9.99</td>
<td>3.03</td>
<td>8.71</td>
<td>2.38</td>
</tr>
<tr>
<td>Number of siblings</td>
<td>2.84</td>
<td>1.17</td>
<td>2.86</td>
<td>1.26</td>
</tr>
<tr>
<td>Father’s years of education</td>
<td>11.45</td>
<td>3.63</td>
<td>9.66</td>
<td>2.98</td>
</tr>
<tr>
<td>Father’s log earnings</td>
<td>12.36</td>
<td>1.19</td>
<td>12.05</td>
<td>1.27</td>
</tr>
<tr>
<td>Local long-run earnings (SEK/100)</td>
<td>137.94</td>
<td>14.37</td>
<td>136.27</td>
<td>14.82</td>
</tr>
<tr>
<td>Local long-run unemployment</td>
<td>0.21</td>
<td>0.04</td>
<td>0.22</td>
<td>0.04</td>
</tr>
<tr>
<td>Distance to university (km/100)</td>
<td>0.91</td>
<td>0.99</td>
<td>1.00</td>
<td>1.02</td>
</tr>
<tr>
<td>Local short-run earnings (SEK/100)</td>
<td>132.35</td>
<td>19.12</td>
<td>131.06</td>
<td>19.75</td>
</tr>
<tr>
<td>Local short-run unemployment</td>
<td>0.26</td>
<td>0.07</td>
<td>0.26</td>
<td>0.07</td>
</tr>
<tr>
<td>Non-missing earnings observations</td>
<td>30.72</td>
<td>1.36</td>
<td>30.81</td>
<td>1.15</td>
</tr>
<tr>
<td>Years of education</td>
<td>15.90</td>
<td>1.16</td>
<td>12.10</td>
<td>0.98</td>
</tr>
<tr>
<td>Number of observations</td>
<td>23186</td>
<td></td>
<td>31840</td>
<td></td>
</tr>
</tbody>
</table>

Note: Lifetime earnings is computed as the average of all non-missing annual earnings observations for ages 20-50. Test scores are standardized by birth year at the population level (i.e. before any sample restrictions). Father’s earnings are computed as the average of annual non-missing earnings for years 1968-1972. Local permanent labor market characteristics are computed as averages across the years 1968-1990 by municipality of residence at age 20. The short-run measures are for age 20. Unemployment is computed as one minus the local working-age employment rate (i.e. a measure of “non-employment”). Distance to university is measured as the closest route by car from the municipality of residence in 1965 (in total 996 entities) to the closest university city (Stockholm, Uppsala, Linköping, Gothenburg, Lund or Umeå). Included in the set of controls are also regional and birth-year dummies (not reported here).

The (discounted) lifetime return to a year of college is estimated to be around 3.7 percent when not controlling for observed abilities (columns 1-2). It falls to about 3.1 percent when these are included as controls (columns 3-5). This illustrates in a simple way the potential (positive) ability bias in OLS estimates. Allowing for interactions between the control variables and the college dummy, on the other hand, seems to have little effect on the main estimate. Nevertheless, the observed heterogeneity with respect to the two ability measures is quite substantial (columns 4-5). A one standard deviation increase in cognitive (noncognitive) ability increases the return to a year of college by around 1.7 (0.7) percent. The results thus suggest that there may be considerable variation in individual returns, despite the relative stability of the
Table 2.2 OLS Estimates of the Return to a Year of College

<table>
<thead>
<tr>
<th>OLS Coefficients</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>College dummy (S)</td>
<td>0.0374</td>
<td>0.0376</td>
<td>0.0319</td>
<td>0.0312</td>
<td>0.0316</td>
</tr>
<tr>
<td></td>
<td>(0.0012)</td>
<td>(0.0012)</td>
<td>(0.0012)</td>
<td>(0.0012)</td>
<td>(0.0012)</td>
</tr>
<tr>
<td>S*A (Cognitive)</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>0.0186</td>
<td>0.0175</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>(0.0022)</td>
<td>(0.0023)</td>
</tr>
<tr>
<td>S*A (Noncognitive)</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>0.0076</td>
<td>0.0067</td>
</tr>
<tr>
<td></td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>(0.0014)</td>
<td>(0.0013)</td>
</tr>
<tr>
<td>Ability controls (A)</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Interactions S*A</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Interactions S*X</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
</tr>
</tbody>
</table>

Note: This table reports OLS regression coefficients of log lifetime earnings on the college dummy (S). The control variables (X) include region and cohort dummies, as well as linear and quadratic terms of father’s and mother’s years of schooling, father’s log earnings, number of siblings, local long-run unemployment and earnings in the municipality of residence at age 20. Specifications (3)-(5) also include linear and quadratic terms of the measures of cognitive and noncognitive ability (i.e. A). Specifications (2) and (5) include interactions between S and all components of X, and (4) and (5) include interactions between S and all components of A. The interaction terms in rows 2 and 3 (S*A) are reported as average derivatives (standard errors from 100 bootstrap replications). All coefficients are divided by 3.8 to reflect the difference in years of schooling between those with and without college. Standard errors are in parentheses.

As opposed to OLS, standard IV estimates a causal effect without assuming equal potential outcomes for treated and untreated individuals. I report IV estimates for different sets of instruments in Table 2.3. Observed heterogeneity is taken into account by including interaction effects in the second stage. In line with several previous studies, my IV estimates are larger than the OLS estimates. There is also variation in the estimated LATEs across different instruments and first-stage models (linear 2SLS or probit). Since different instruments identify the LATE for different subpopulations, such variation is expected in the presence of self selection on heterogeneous returns. Nevertheless, when I use $P(Z)$ with the full set of instruments, the estimate is in the lower range and close to the semiparametric estimate of the ATE (see below). This is not in itself a rejection of the self-selection hypothesis, but is nevertheless noteworthy.

Results using a Normal Selection Model

The traditional approach to estimate the model in Section 2.2 is to specify a parametric joint distribution for the error terms (e.g., Willis and Rosen, 1979). Björklund and Moffitt (1987), for example, estimate the MTE assuming that...
the error terms are jointly normally distributed. Although my main focus is on the semiparametric method, results based on a normal selection model are useful for purposes of comparison.

Figure 2.1 shows parametric estimates of the MTE by levels of $U_S$, conditioned on mean values of $(X, A)$. The MTE is weakly declining and relatively precisely estimated. A test for selection on unobserved gains is to test whether the slope of the conditional MTE is zero. For the normal selection model this implies testing whether $\sigma_{1V} - \sigma_{2V} = 0$ in eq. (2.16). I estimate that $\sigma_{1V} - \sigma_{2V} = -0.0481$ with a standard error of 0.0291 (obtained using the delta method). Thus, I cannot reject the hypothesis that the slope of the MTE is zero at the 95 percent confidence level, although it is on the border of rejection at the 90 percent level ($z$-statistic = 1.6538). As a comparison, Carneiro et al. (2011) estimate that $\sigma_{1V} - \sigma_{2V} = -0.2388$ with a standard error of 0.0982, and thus reject a flat MTE.

If I also account for observed heterogeneity, i.e., the variation in $X$ and $A$ and their impact on the MTE through $X'(\delta_0 - \delta_1) + A'(\gamma_0 - \gamma_1)$, then the slope becomes much steeper. The magnitude of total heterogeneity is illustrated by the change in the part of the (unconditional) MTE that is attributed to $(X, A)$ when going from the first to the tenth decile of $U_S$, which corresponds to about 10 percentage points in terms of the return to a year of college. Across all individuals, the heterogeneity associated with observed characteristics varies between -0.1089 and 0.2328.

Table 2.4 (column 1) reports estimates of summary treatment parameters based on the parametric MTE estimates and the appropriate weights (reported in Appendix A.2). The estimated ATE implies a return to one year of college

<table>
<thead>
<tr>
<th>Distance to university</th>
<th>Local earnings</th>
<th>Local unempl.</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard 2SLS</td>
<td>0.0798</td>
<td>0.1150</td>
<td>0.0673</td>
</tr>
<tr>
<td></td>
<td>(0.0395)</td>
<td>(0.0461)</td>
<td>(0.0364)</td>
</tr>
<tr>
<td>$P(Z)$ as instr.</td>
<td>0.0501</td>
<td>0.0847</td>
<td>0.0463</td>
</tr>
<tr>
<td></td>
<td>(0.0378)</td>
<td>(0.0376)</td>
<td>(0.0383)</td>
</tr>
</tbody>
</table>
of about 2.4 percent. The corresponding estimates for the ATT and ATU are approximately 2.9 and 1.6 percent, respectively. Table 2.4 also shows tests of equality between ATT and ATE, ATT and ATU, and ATE and ATU, which serve as broad tests for self selection on total heterogeneity. All tests reject equality and support the notion that individuals choose schooling based on their own comparative advantage. These results however rest on the potentially restrictive normality assumption, and it is thus not clear how reliable they are.

Results using Semiparametric LIV

A potentially more robust approach for estimating the MTE is to estimate \( E(Y \mid X = x, A = a, P(Z) = p) \) semiparametrically and then compute its derivative with respect to \( p \), as in eq. (2.13). This is the essence of the LIV approach. If \((X, A)\) is not independent of \((U_0, U_1, V)\), a necessary (and very demanding) condition is that \( P \) has full support at each value of \((X, A)\). For each combination of \((X, A)\), variation in \( P \) can only identify the MTE across small intervals of \( V \). To reduce the dimensionality of \((X, A)\), I therefore use an index of \( X'(\delta_1 - \delta_0) + A'(\gamma_1 - \gamma_0) \). The support of \( P \) for each value of the index is

\[ \text{Note: This figure shows point estimates and 95 percent confidence bands of the MTE from the parametric normal selection model in equation 2.16 estimated by maximum likelihood. All estimates are conditioned on mean values of } X \text{ and } A. \]

\[ \text{Figure 2.1 MTE by } U_5 \text{ Estimated from a Normal Selection Model} \]

[Figure showing MTE estimates with 95% confidence bands]

\[ \text{Table 2.4 also shows tests of equality between ATT and ATE, ATT and ATU, and ATE and ATU, which serve as broad tests for self selection on total heterogeneity. All tests reject equality and support the notion that individuals choose schooling based on their own comparative advantage. These results however rest on the potentially restrictive normality assumption, and it is thus not clear how reliable they are.} \]

\[ \text{Results using Semiparametric LIV} \]

\[ A \text{ potentially more robust approach for estimating the MTE is to estimate } E(Y \mid X = x, A = a, P(Z) = p) \text{ semiparametrically and then compute its derivative with respect to } p, \text{ as in eq. (2.13). This is the essence of the LIV approach. If } (X, A) \text{ is not independent of } (U_0, U_1, V), \text{ a necessary (and very demanding) condition is that } P \text{ has full support at each value of } (X, A). \text{ For each combination of } (X, A), \text{ variation in } P \text{ can only identify the MTE across small intervals of } V. \text{ To reduce the dimensionality of } (X, A), \text{ I therefore use an index of } X'(\delta_1 - \delta_0) + A'(\gamma_1 - \gamma_0). \text{ The support of } P \text{ for each value of the index is} \]

\[ 31 \text{I follow Basu et al. (2007) and condition on demideciles (i.e., 20 uniformly distributed groups) of the scalar index } X(\delta_1 - \delta_0) + A(\gamma_1 - \gamma_0). \text{ The results are robust to conditioning on finer partitions of the index (50 or 100 uniformly distributed groups).} \]
Table 2.4 Returns to a Year of College

<table>
<thead>
<tr>
<th>Model</th>
<th>Normal</th>
<th>Semiparametric</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATE</td>
<td>0.0238</td>
<td>0.0484</td>
</tr>
<tr>
<td></td>
<td>(0.0027 )</td>
<td>(0.0208)</td>
</tr>
<tr>
<td>ATT</td>
<td>0.0321</td>
<td>0.0574</td>
</tr>
<tr>
<td></td>
<td>(0.0029)</td>
<td>(0.0208)</td>
</tr>
<tr>
<td>ATU</td>
<td>0.0178</td>
<td>0.0418</td>
</tr>
<tr>
<td></td>
<td>(0.0027)</td>
<td>(0.0210)</td>
</tr>
<tr>
<td>ATT - ATU</td>
<td>0.0142</td>
<td>0.0155</td>
</tr>
<tr>
<td></td>
<td>(0.0008)</td>
<td>(0.0032)</td>
</tr>
<tr>
<td>ATT - ATE</td>
<td>0.0082</td>
<td>0.0089</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0018)</td>
</tr>
<tr>
<td>ATE - ATU</td>
<td>0.0060</td>
<td>0.0066</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0013)</td>
</tr>
</tbody>
</table>

Note: This table reports estimates of the average treatment effect (ATE), the average treatment effect on the treated (ATT), and average treatment effect on the untreated (ATU). The estimates in column 1 are based on maximum likelihood estimates of MTEs from the normal switching regression model in equation 2.16. The estimates in column 2 are based on the semiparametric model, and thus for \( \tilde{ATE} \), \( \tilde{ATT} \), and \( \tilde{ATU} \) (rather than the true ATE, ATT, ATU), with the twiggle indicating that they are sample specific parameters that are conditional on the estimated support of \( U \).

Rows 4-6 shows the estimated differences between the treatment effect parameters. Standard errors are obtained using the bootstrap (100 replications).

nevertheless small. If I instead follow Carneiro et al. (2011) and invoke the assumption that \((X, A)\) is independent of \((U_0, U_1, V)\), then each of the intervals from the conditional identification can be put together so that the MTE can be identified over almost the entire support of \(V\). It is thus only necessary to examine the marginal support of \(P(Z)\) as opposed to the support of \(P(Z)\) conditional on \((X, A)\). This assumption also legitimizes the use of interactions between \(Z\) and components of \((X, A)\) as instruments in the choice equation.

I estimate \(P(Z)\) in a probit model and present estimated average marginal derivatives in Table 2.5. I also report average marginal effects for each of the polynomials of the instruments. The average effect of the distance instrument on college attendance is negative and highly significant. The average effect of local unemployment at age 20 is also negative and significant, whereas local earnings at age 20 is a weak predictor of college attendance. The instruments are jointly strong predictors of college attendance, as are mother’s and father’s years of schooling, father’s earnings, the measure of noncognitive ability, and, in particular, the measure of cognitive ability. In fact, cognitive ability is in terms of average derivatives around eight times stronger than noncognitive ability as a predictor of going to college.

Figure 2.2 shows the support of the estimated \(P(Z)\). There is a lack of
### Table 2.5 College Decision Model

<table>
<thead>
<tr>
<th>Controls (A, X)</th>
<th>Avg. derivative</th>
<th>Avg. marginal effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive test score</td>
<td>0.1047</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0030)</td>
<td></td>
</tr>
<tr>
<td>Non-cognitive test score</td>
<td>0.0130</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0024)</td>
<td></td>
</tr>
<tr>
<td>Mother’s years of schooling</td>
<td>0.0164</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0011)</td>
<td></td>
</tr>
<tr>
<td>Number of siblings</td>
<td>-0.0024</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0018)</td>
<td></td>
</tr>
<tr>
<td>Father’s years of schooling</td>
<td>0.0173</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0009)</td>
<td></td>
</tr>
<tr>
<td>Father’s log earnings</td>
<td>0.0406</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0031)</td>
<td></td>
</tr>
<tr>
<td>Local long-run earnings</td>
<td>0.0018</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td></td>
</tr>
<tr>
<td>Local long-run unempl.</td>
<td>0.7832</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.3828)</td>
<td></td>
</tr>
<tr>
<td><strong>Instruments (Z)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to university (km/100)</td>
<td>-0.0586</td>
<td>-0.1114</td>
</tr>
<tr>
<td></td>
<td>(0.0159)</td>
<td>(0.0312)</td>
</tr>
<tr>
<td>Distance to univ. (km/100) quadratic</td>
<td>0.2372</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0676)</td>
<td></td>
</tr>
<tr>
<td>Distance to univ. (km/100) cubic</td>
<td>-0.1278</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0410)</td>
<td></td>
</tr>
<tr>
<td>Local short-run earnings</td>
<td>-0.0018</td>
<td>-0.0003</td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td>(0.0040)</td>
</tr>
<tr>
<td>Local short-run earnings quadratic</td>
<td>-0.0000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td></td>
</tr>
<tr>
<td>Local short-run unempl.</td>
<td>-0.7292</td>
<td>-0.5483</td>
</tr>
<tr>
<td></td>
<td>(0.3208)</td>
<td>(0.6411)</td>
</tr>
<tr>
<td>Local short-run unempl. quadratic</td>
<td>-0.3892</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.0025)</td>
<td></td>
</tr>
<tr>
<td><strong>Joint significance test of Z: p-value</strong></td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

Note: The table reports average derivatives and marginal effects from a probit regression of a college indicator on the set of variables listed in the table and cohort and region dummies (see Section 2.3 for exact specification). The average derivatives are obtained by computing for each individual the effect including all polynomial terms of increasing a variable by one unit (keeping all the others constant) on the probability of enrolling in college and then average across all individuals. The average marginal effects (reported for the instruments) are obtained in the same manner but separately for each polynomial term of the respective variable. Standard errors are obtained using the bootstrap (100 replications).
support in the lowest tenth of the interval, whereas the support in the upper part of the interval nearly reaches one.\textsuperscript{32} Given the estimates of $P(Z)$, the next step is to estimate the components of equation (2.15) and compute the MTEs.

**Figure 2.2** Support of $P(Z)$ for untreated ($S = 0$) and treated ($S = 1$)

<table>
<thead>
<tr>
<th>Propensity score</th>
<th>Untreated</th>
<th>Treated</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 .2 .4 .6 .8 1</td>
<td>Untreated</td>
<td>Treated</td>
</tr>
</tbody>
</table>

Note: This figure shows the support of $P(Z)$ for the treated and the untreated. $P(Z)$ is the probability of going to college estimated in a probit regression of the college choice equation (see Table 2.5).

Figure 2.3 shows how the MTE depends on $V$ across the quantiles of $V$ (i.e., across $U_S$), with the components of $(X,A)$ fixed at their sample means. Two main results emerge. First, the variation in unobserved heterogeneity is in terms of point estimates quite substantial. The difference between the sections of $U_S$ with the highest and the lowest MTEs corresponds to about 20 percentage points in the returns to a year of college. For a major part of the $U_S$ interval, however, the results suggest an almost flat MTE and thus little unobserved heterogeneity. Second, the evidence on self selection on unobserved gains is mixed. For values of $U_S$ up until about 0.35, the MTE declines in $U_S$ (i.e., positive selection). For intermediate values of $U_S$, there is not much of a clear pattern and for the top decile of $U_S$, the MTE is even increasing (i.e., negative selection). This indicates that the normal selection model provides an incorrect representation of unobserved heterogeneity. The overall picture is thus mixed, although the evidence on self selection on unobserved gains is clearly weaker than what is reported in Carneiro et al. (2011).\textsuperscript{33}

A simple test of selection on unobserved gains consists of comparing the

\textsuperscript{32}The common support is defined as the intersection of the support of $P(Z \mid S = 0)$ and the support of $P(Z \mid S = 1)$. I trim observations for which the estimated $P(Z)$ is either lower than the minimum, or higher than the maximum value of $P(Z)$ for which
average MTE across equally spaced adjacent intervals along the support of $U_S$, i.e., LATEs defined over different subpopulations (see Heckman et al., 2010). Table 2.6 reports the outcome of the test. I cannot reject the joint hypothesis that all adjacent LATEs are equal.

Lastly, I turn to my estimates of the ATE, ATT, and ATU. Since I do not have full support for $P$, these parameters cannot be estimated in exact accordance with their definitions. I can, however, compute approximations of these there is common support.

Although the semiparametric estimates have larger standard errors than the estimates based on the normal model, the precision of my semiparametric estimates are larger than in Carneiro et al. (2011). The main difference is instead that they find evidence of an MTE with an unambiguously steep negative slope.

The test is based on 100 bootstrap replications of the MTE, evaluated at mean values of $X$ and $A$. I take the average of the MTE in equally spaced intervals along the support of $U_S$ and compute the statistics $T = | \text{LATE}^j - \text{LATE}^{j+1} |$ (the absolute value of the difference between two adjacent LATEs $j$ and $j+1$) and $T_b = | (\text{LATE}^j_b - \text{LATE}^{j+1}_b) - (\text{LATE}^j - \text{LATE}^{j+1}) |$, where $\text{LATE}^j_b$ is the $b^{th}$ bootstrap replication of LATE$^j$. The corresponding statistics for the joint test are $C = \sum_{j=1}^{J-1} | \text{LATE}^j - \text{LATE}^{j+1} |^2$ and $C_b = \sum_{j=1}^{J-1} \left[ (\text{LATE}^j_b - \text{LATE}^{j+1}_b) - (\text{LATE}^j - \text{LATE}^{j+1}) \right]^2$. The $p$-value of the tests is the proportion of bootstrap replications for which $T_b > T$ (or $C_b > C$ for the joint test).
Table 2.6 Test for Equality of LATEs over Different Intervals

<table>
<thead>
<tr>
<th>Range of LATE(j)</th>
<th>(.125; .200)</th>
<th>(.275; .350)</th>
<th>(.425; .500)</th>
<th>(.575; .650)</th>
<th>(.725; .800)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range of LATE(j+1)</td>
<td>(.275; .350)</td>
<td>(.425; .500)</td>
<td>(.575; .650)</td>
<td>(.725; .800)</td>
<td>(.875; .950)</td>
</tr>
<tr>
<td>LATE(j)-LATE(j+1)</td>
<td>0.0876</td>
<td>0.0138</td>
<td>0.0147</td>
<td>0.0029</td>
<td>0.1079</td>
</tr>
<tr>
<td>(p)-value</td>
<td>0.4600</td>
<td>0.8400</td>
<td>0.6500</td>
<td>0.9300</td>
<td>0.0400</td>
</tr>
<tr>
<td>Joint (p)-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.5000</td>
</tr>
</tbody>
</table>

Note: This table reports a test of essential heterogeneity conducted by testing the equality of LATEs in pairwise adjacent intervals of \(U_S\). I construct intervals of \(U_S\) and average the MTE within these intervals by computing \(E(Y_1 - Y_0 \mid X = \bar{x}, U_S^{j} \leq U_S \leq U_S^{j+1})\), where \(U_S^{j}\) and \(U_S^{j+1}\) are the lower and upper bounds of \(U_S\) in interval \(j\). This gives the different LATEs and the null of the tests are \(H_0: \text{LATE}(U_S^{j}, U_S^{j+1}) - \text{LATE}(U_S^{j+1}, U_S^{j+1}) = 0\). The bottom row reports the outcome of the test that all adjacent LATEs are jointly equal. All tests take the multiple estimation steps into account by using the bootstrap (100 replications).

parameters, denoted \(\tilde{\text{ATE}}\), \(\tilde{\text{ATT}}\), and \(\tilde{\text{ATU}}\), for which I rescale the weights (reported in Appendix A.2) to integrate to one over the common support.

Table 2.4 (column 2) reports the estimates together with a set of simple tests for self selection on total heterogeneity. The semiparametric estimate of the \(\tilde{\text{ATE}}\) suggests a return to one year of college of about 4.8 percent. As expected, the estimated \(\tilde{\text{ATT}}\) is larger and \(\tilde{\text{ATU}}\) smaller, although the differences are relatively small. Nevertheless, the differences are all statistically significant, thus indicating sorting into college based on overall heterogeneity. It is those who actually have selected into college that, on average, also have the highest estimated ex-post returns. What is more surprising is the large and positive returns for those who have chosen not to go to college. This is in sharp contrast with Carneiro et al. (2011), who report an \(\tilde{\text{ATU}}\) that is close to zero. Finally, it is worthwhile to compare with the estimates from the normal selection model (column 1). These are substantially lower, but the pattern in terms of the differences across \(\text{ATE}\), \(\text{ATT}\), and \(\text{ATU}\) is very similar. As unobserved heterogeneity seems to be of modest importance in my sample, I now turn to examine the role of observed heterogeneity, and in particular ability-specific heterogeneity.

Evidence on Observable Heterogeneity and Ability Heterogeneity

Figure 2.4 shows the component of the MTE that is attributable to total observed heterogeneity along the scalar index \(X(\delta_1 - \delta_0) + A(\gamma_1 - \gamma_0)\). First, note that this curve is not comparable to Figure 2.3, which plotted the MTE across the distribution of \(U_S\). \(U_S\) is an unobserved variable, while the scalar index on the x-axis in Figure 2.4 is itself estimated. Both do however govern expected returns and thereby selection into college. Figure 2.4 implies that
the variation in total observed heterogeneity is substantial and the slope of the
curve indicates that observed characteristics impact on returns across the entire
distribution (in terms of point estimates). The curve suggests that those with
the most favorable characteristics (i.e., that complement formal college edu-
cation the most) on average have a return that is around 20 percentage points
higher than those with the least favorable characteristics. In reality, the het-
erogeneity is even larger since there is also considerable variation within the
groups on the x-axis.

Figure 2.4 Average MTE by Total Observed Heterogeneity

Note: This figure plots the average MTE with 95 percent confidence bands across the index of observed het-
erogeneity. The index is computed by estimating $X(\delta_1 - \delta_0) + A(\gamma_1 - \gamma_0)$ for each individual and splitting
the sample into demideciles (i.e., 20 uniformly distributed groups), thus following the procedure of Basu et
al. (2007) and Carneiro et al. (2011).

Figures 2.5a and 2.5b show the ATE conditional on the measures of cog-
nitive and noncognitive ability, respectively. There is a strong relationship be-
tween both measures and the estimated ATE. Moreover, both the pattern and
the magnitude of the heterogeneity are roughly similar for the two measures,
although the negative effects at the low end are more pronounced for cognitive
ability. At the top end, the positive complementarity with college education,
as suggested by the point estimates, is even somewhat larger for noncognitive
ability, although the difference is marginal. Belonging to the top category in
either cognitive or noncognitive ability implies a return to a year of college that
is around 10 percentage points higher than the average.

A potential explanation to the high resemblance across the two measures
would be that they are highly correlated. This correlation is about 0.19 in my
Figure 2.5 Observed Ability Heterogeneity in the Return to a Year of College

(a) ATE by Cognitive Ability

(b) ATE by Noncognitive Ability

Note: The figures show semiparametric estimates of average treatment effects (ATE) with 95 percent confidence bands conditional on levels of cognitive and noncognitive abilities. The ability measures are plotted on the x-axes in their original unstandardized form, although standardized measures are used in all estimations. Standard errors are obtained using the bootstrap (100 replications).
sample, suggesting that this could only be a partial explanation. Moreover, it is possible that the ability measures are correlated with other control variables that impact on observed heterogeneity. In Table 2.7 (column 2), I therefore report the semiparametric estimates of observed heterogeneity with respect to the (standardized) ability measures from the outcome equation. Despite the fact that these estimates are thus conditional on any potential impacts on observed heterogeneity from other covariates, they imply a pattern similar to the comparison of the conditional ATEs.\textsuperscript{35} The estimates, presented as average derivatives, imply that an increase of one standard deviation in the measure of cognitive (noncognitive) ability on average increases the return by about 3.5 (2.5) percentage points. These estimates are larger than the comparable OLS estimates in Table 2.2, but roughly similar in terms of the relative importance of the two measures. Moreover, the semiparametric estimates in Table 2.7 (column 1) imply very modest, or even zero, direct effects from the ability measures in the outcome equation. The evidence thus lends support to the notion of comparative advantage in college education, whereas the support for absolute advantage is weak.

\textbf{Table 2.7 Average Derivatives for Abilities in the Outcome Equation}

<table>
<thead>
<tr>
<th></th>
<th>$\gamma_0$</th>
<th>$\gamma_1 - \gamma_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive ability</td>
<td>-0.0011</td>
<td>0.0353</td>
</tr>
<tr>
<td></td>
<td>(0.0038)</td>
<td>(0.0082)</td>
</tr>
<tr>
<td>Noncognitive ability</td>
<td>0.0069</td>
<td>0.0250</td>
</tr>
<tr>
<td></td>
<td>(0.0017)</td>
<td>(0.0037)</td>
</tr>
</tbody>
</table>

Note: This table reports average derivatives of the (standardized) measures of cognitive and noncognitive ability in the outcome equations for the semiparametric model. The model is estimated by local linear regression. This procedure, and exact specifications of the full set of control variables (not reported here), are further described in Section 2.2. The average derivatives are obtained by computing for each individual the effect of increasing a variable by one unit (keeping all the others constant) on log lifetime earnings and then average across all individuals. Column 1 reports the main effects, whereas column 2 reports the interaction effects (i.e. observable heterogeneity). Standard errors are bootstrapped (100 replications).

The resemblance between the estimates in Table 2.7 (column 2) and the conditional ATEs is not surprising. First, the impact on observed heterogeneity from other covariates is very small as compared to the impact from the measures of cognitive and noncognitive ability.\textsuperscript{36} Most observed heterogene-

\textsuperscript{35}Note however that these estimates are not conditional on unobserved heterogeneity as they rely on the auxiliary assumption that the observed and unobserved heterogeneity components are uncorrelated. The previously discussed estimates of unobserved heterogeneity are derived conditional on estimated observed heterogeneity, I can thus not control for unobserved heterogeneity when I estimate observed heterogeneity.

\textsuperscript{36}The estimates of observed heterogeneity with respect to other control variables
ity thus seems to be captured by these two variables. Second, the previous evidence did not suggest any dramatic effects from unobserved heterogeneity, although estimates were imprecise. What is maybe more surprising is that the heterogeneity with respect to noncognitive ability is so large, and roughly comparable to the one with respect to cognitive ability. The probit estimates of the choice equation implied that cognitive ability is a much stronger predictor of selection into college than noncognitive ability. If selection were purely driven by expected benefits (i.e., monetary returns), then the two ability types should have a more equal predictive power of college attendance. However, the model of college choice in Section (2.1) illustrates some potential explanations for why this need not be the case. Cognitive ability might impact on the cost of going to college more than noncognitive ability, either in terms of time costs (yielding more leisure) or psychic costs (less headache) for a given achievement. It could also be due to heterogeneity in the valuation of college as a consumption good; the level of cognitive ability might influence the direct utility derived from going to college more positively than the level of noncognitive ability. Such explanations could each contribute to the result that cognitive ability seems to trigger selection into college much more strongly than noncognitive ability, despite having comparable effects on monetary returns.37

Sensitivity to Sample, Specifications and Variable Definitions

A simple way to analyze the robustness of my estimates is to examine how $\tilde{\text{ATE}}$, $\tilde{\text{ATT}}$, and $\tilde{\text{ATU}}$ vary across specifications. In addition, I report a straightforward test of selection on returns: a test of the null that $\tilde{\text{ATT}} = \tilde{\text{ATU}}$, i.e., whether the average person attending college has the same return as the average person not attending. Results are reported in panels A, B and C of Table 2.8.

First, the results in panel A concern choice of sample and specification of the outcome equation. I excluded from my baseline sample everyone without an (academic) high school degree. This is in line with Willis and Rosen (1979), whereas Carneiro et al. (2011) include dropouts. Column 2 in panel A indicates that my estimates are relatively unaffected when including those, although estimated heterogeneity ($\tilde{\text{ATT}} - \tilde{\text{ATU}}$) increases. The model that I use has the limitation that it restricts the college variable to be binary. An intuitive critique of estimates of heterogeneous returns is that different people might choose different types of college in terms of quality or field of study. One are not shown here, but are available from the author upon request.

37There are of course other potential explanations that go beyond this simple model, e.g., differences in preferences such as discount factors (as emphasized by Willis and Rosen, 1979) or risk attitudes, and differential forecasting errors.
Table 2.8 Returns to a Year of College: Sensitivity Analyses

(a) Different Samples and the Specification of the Outcome Equation

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Including high school dropouts</th>
<th>Type of college in $X \setminus Z$</th>
<th>$X \setminus Z$ as Carneiro et al. (2011)</th>
<th>$X \setminus Z$ excluding experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{ATE}$</td>
<td>0.0484</td>
<td>0.0528</td>
<td>0.0137</td>
<td>0.0339</td>
<td>0.0408</td>
</tr>
<tr>
<td></td>
<td>(0.0208)</td>
<td>(0.0206)</td>
<td>(0.0201)</td>
<td>(0.0130)</td>
<td>(0.0210)</td>
</tr>
<tr>
<td>$\tilde{ATT}$</td>
<td>0.0574</td>
<td>0.0786</td>
<td>0.0604</td>
<td>0.0378</td>
<td>0.0508</td>
</tr>
<tr>
<td></td>
<td>(0.0208)</td>
<td>(0.0173)</td>
<td>(0.0212)</td>
<td>(0.0123)</td>
<td>(0.0209)</td>
</tr>
<tr>
<td>$\tilde{ATU}$</td>
<td>0.0418</td>
<td>0.0421</td>
<td>-0.0256</td>
<td>0.0311</td>
<td>0.0335</td>
</tr>
<tr>
<td></td>
<td>(0.0210)</td>
<td>(0.0224)</td>
<td>(0.0198)</td>
<td>(0.0130)</td>
<td>(0.0212)</td>
</tr>
<tr>
<td>$\tilde{ATT} - \tilde{ATU}$</td>
<td>0.0155</td>
<td>0.0365</td>
<td>0.0860</td>
<td>0.0067</td>
<td>0.0173</td>
</tr>
<tr>
<td></td>
<td>(0.0032)</td>
<td>(0.0071)</td>
<td>(0.0079)</td>
<td>(0.0028)</td>
<td>(0.0032)</td>
</tr>
</tbody>
</table>

(b) Specification of the Choice Equation

<table>
<thead>
<tr>
<th></th>
<th>Any college as treatment</th>
<th>Only linear terms in $Z$</th>
<th>No interactions with $Z$</th>
<th>Only distance in $Z$</th>
<th>Sample, $X$ as Carneiro et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{ATE}$</td>
<td>0.0418</td>
<td>0.0279</td>
<td>0.0318</td>
<td>0.0207</td>
<td>0.0501</td>
</tr>
<tr>
<td></td>
<td>(0.0204)</td>
<td>(0.0286)</td>
<td>(0.0344)</td>
<td>(0.0349)</td>
<td>(0.0122)</td>
</tr>
<tr>
<td>$\tilde{ATT}$</td>
<td>0.0503</td>
<td>0.0375</td>
<td>0.0413</td>
<td>0.0299</td>
<td>0.0613</td>
</tr>
<tr>
<td></td>
<td>(0.0206)</td>
<td>(0.0284)</td>
<td>(0.0325)</td>
<td>(0.0328)</td>
<td>(0.0077)</td>
</tr>
<tr>
<td>$\tilde{ATU}$</td>
<td>0.0332</td>
<td>0.0209</td>
<td>0.0248</td>
<td>0.0141</td>
<td>0.0463</td>
</tr>
<tr>
<td></td>
<td>(0.0202)</td>
<td>(0.0288)</td>
<td>(0.0359)</td>
<td>(0.0350)</td>
<td>(0.0121)</td>
</tr>
<tr>
<td>$\tilde{ATT} - \tilde{ATU}$</td>
<td>0.0171</td>
<td>0.0166</td>
<td>0.0165</td>
<td>0.0158</td>
<td>0.0150</td>
</tr>
<tr>
<td></td>
<td>(0.0033)</td>
<td>(0.0034)</td>
<td>(0.0031)</td>
<td>(0.0031)</td>
<td>(0.0056)</td>
</tr>
</tbody>
</table>

(c) Definitions of the Outcome Variable and analysis of life-cycle effects

<table>
<thead>
<tr>
<th></th>
<th>Avg. wage ages 20-50</th>
<th>Average earnings ages 26-30</th>
<th>Average earnings ages 36-40</th>
<th>Average earnings ages 46-50</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{ATE}$</td>
<td>0.0432</td>
<td>-0.0403</td>
<td>0.1088</td>
<td>0.1005</td>
</tr>
<tr>
<td></td>
<td>(0.0138)</td>
<td>(0.0312)</td>
<td>(0.0378)</td>
<td>(0.0389)</td>
</tr>
<tr>
<td>$\tilde{ATT}$</td>
<td>0.0479</td>
<td>-0.0288</td>
<td>0.1170</td>
<td>0.1060</td>
</tr>
<tr>
<td></td>
<td>(0.0137)</td>
<td>(0.0312)</td>
<td>(0.0376)</td>
<td>(0.0366)</td>
</tr>
<tr>
<td>$\tilde{ATU}$</td>
<td>0.0398</td>
<td>-0.0487</td>
<td>0.1027</td>
<td>0.0965</td>
</tr>
<tr>
<td></td>
<td>(0.0139)</td>
<td>(0.0313)</td>
<td>(0.0381)</td>
<td>(0.0391)</td>
</tr>
<tr>
<td>$\tilde{ATT} - \tilde{ATU}$</td>
<td>0.0080</td>
<td>0.0198</td>
<td>0.0143</td>
<td>0.0095</td>
</tr>
<tr>
<td></td>
<td>(0.0022)</td>
<td>(0.0047)</td>
<td>(0.0047)</td>
<td>(0.0053)</td>
</tr>
</tbody>
</table>

Note: This table reports estimates of the return to a year of college for the semiparametric model for various samples and specifications. The estimates of the average treatment effect (ATE), the average treatment effect on the treated (ATT), and average treatment effect on the untreated (ATU) are computed such that their weights integrate to one over the respective common support. The table also reports a simple test of self selection: if $\tilde{ATT} - \tilde{ATU} = 0$. Panel A reports estimates for different samples and specifications of the outcome equation, Panel B for different specifications of the outcome equation, and Panel C for the definition of the outcome variable and life-cycle effects. Standard errors are obtained using the bootstrap (100 replications).
could, in principle, extend the method used in this paper to multiple schooling types, but that would require distinct instruments for each schooling transition (Heckman et al., 2006b). To explore this limitation, I report estimates with indicators for three broad categories of college (bachelor or less, master, and doctoral degree) included in the outcome equation. The resulting estimates in column 3 surprisingly indicate larger heterogeneity, which might potentially be because these indicators are not exogenous.

An efficient way of reducing the residual variance in the estimation of the outcome equation is to include additional controls in the outcome equation only. Column 4 shows estimates from using total experience (quadratic) and local unemployment and earnings at about age 35, as such additional controls (i.e., similar to Carneiro et al., 2011). This approach is generally questionable, as each of these variables may be endogenous. Since I use lifetime earnings rather than short-run wage as outcome variable, this is most obviously the case with the experience variable, which I exclude from the estimates in column 5. The estimates in columns 4 and 5 are somewhat smaller than the baseline, and when including experience, both estimated heterogeneity and standard errors are much smaller.

Second, the results in panel B concern the specification of the choice equation. Column 1 shows that the estimates are relatively unaffected by using “any college” (minimum one semester) as the college indicator. Columns 2-4 show alternative specifications of the instruments. Using only linear terms in Z (column 2), or excluding interactions with Z (column 3), produce somewhat smaller estimates but larger standard errors. One might worry that the local labor market instruments affect selection into college by shifting expected returns, despite the fact that I only use the innovations in these variables in Z. In column 4, I therefore report estimates for which only the distance variable is included in Z. The estimates are now substantially smaller than the baseline but too imprecise to draw any firm conclusions. However, the pattern of estimated heterogeneity is remarkably stable across all the different specifications of the choice equation. Column 5 shows estimates for a setup in which I mimic Carneiro et al. (2011) as closely as possible in terms of sample and specification.38 Both the estimates and the extent of heterogeneity are similar to my baseline, while the standard errors are considerably lower.

Finally, in panel C I exploit the nearly career-long earnings data to examine how the estimates vary across different definitions of the outcome variable. This sensitivity analysis relates to the interpretation (or external validity) of the estimated effects, rather than the internal validity. I first use an imputed wage

38I thus include dropouts in the sample, and include the same type of variables in X (i.e., excluding noncognitive ability, and father’s schooling and earnings) and in X \ Z (i.e., local labor market characteristics in prime age and experience) as they do.
measure as the outcome in order to examine the role of labor supply (reported in column 1). The estimates are similar in size, but the standard errors and estimated heterogeneity are somewhat lower. This may suggest that labor supply plays a role for the evidence on selection on returns, as measured by annual earnings. Columns 2-4 show evidence on life-cycle effects in the estimates. In column 2, average earnings across ages 26-30 are used as the outcome and the resulting estimates are now negative (but statistically insignificant). This is not surprising, as many at this age might still be in school, or are recent entrants on the labor market. If earnings are instead observed in their late 30s (column 3) or late 40s (column 4), the estimated effects are positive and considerably larger than the baseline. The estimated heterogeneity is consistent with the baseline for earnings observed at the earlier ages, but somewhat smaller for older ages. This highlights that it is relatively easy to under- or overstate point estimates depending on at what age the outcome variable is observed.

2.5 Conclusions

I applied the LIV approach of Heckman and Vytlacil (1999, 2001, 2005) to a large registry-based data set of Swedish males. My analysis of the returns to college revealed a relatively modest role for heterogeneity in general, and for unobserved heterogeneity in particular, at least in comparison to previous evidence (e.g., Carneiro et al., 2011). Nevertheless, total heterogeneity (mainly via observed characteristics) seems to be an important phenomenon, and this holds across various specifications and sample definitions. However, it is unclear whether the divergence from previous evidence is due to differences in data quality or contextual setting (Sweden vs. the US). A possible explanation for both smaller returns and less heterogeneity could be a lower degree of selection into college in Sweden and most notably a more compressed wage structure. Recent quasi-experimental evidence also lends support to the finding of low returns to college in Sweden (see Öckert, 2010).

Moreover, I provided new evidence on ability heterogeneity using measures of cognitive and noncognitive ability from military enlistment tests. The results implied that both cognitive and noncognitive ability have a large influence on the return, thus indicating that “school-skill complementarities” (i.e., between formal schooling investments and independently produced abilities) are potentially important features of the labor market. Since the effect

39I follow the procedure of Antelius and Björklund (2000) who show that left truncating these data, so that low earnings observations and likely part-time workers are excluded, gives similar estimates of the returns to schooling as when using wage measures. I thus use average annual earnings conditional on having annual earnings above 75 000 SEK (about $10000).
of noncognitive ability is almost as large as that of cognitive ability, it is puzzling that the former has much less influence on the probability of selecting into college. A potential explanation might be that cognitive ability also has a more positive impact on either the costs (e.g., time or psychic costs) or the consumption value of going to school. An intriguing avenue for future research is to enable a more causal interpretation of the cost side of such ability heterogeneity, for example by introducing exogenous return shifters in the Roy model.

Some lessons regarding the applicability of the LIV approach can also be learned from my analysis. In general, sample size seems important for the applicability of the LIV approach, but also the existence of good continuous instruments. My large sample clearly produces more precise estimates of treatment effect parameters compared to previous applications that use smaller survey data. The evidence on unobserved heterogeneity is nevertheless somewhat inconclusive, as the part of the MTE that varies with respect to the unobserved characteristic remains quite imprecisely estimated.

The tendency of a U-shaped pattern of the MTE is also notable, as it differs from the monotonic curve reported in Carneiro et al. (2011). A pessimistic explanation would be that part of this is caused by a failure of the independence assumption. On the other hand, it is also possible that the effect of unobserved heterogeneity is more complex than what is commonly assumed. For example, Brinch et al. (2012) find robust evidence that supports a U-shaped MTE curve when applying the LIV approach to the quantity-quality tradeoff of children.

Several sources can potentially generate a non-monotonic shape of the MTE, including heterogeneity in time or risk preferences, asymmetric information about the costs and benefits of college, and differences in economic resources or access to credit at the time of the college decision. This would be consistent with a population divided into multiple subpopulations that can be represented as a mixture distribution. As Brinch et al. (2012) demonstrate, a non-monotonic MTE can be derived from such underlying data, for example illustrated by a mixture of multiple normal distributions. A methodological implication is that a typical univariate normal selection model will impose a potentially incorrect representation of treatment effect heterogeneity. A practical implication, given that the highest returns are found in the top and bottom of the distribution, is that a mix of targeted policies that absorb students from opposite ends of the spectrum should be preferred over general expansions or contractions of the college sector.
References


Appendix

A.1 Definitions of Weights for ATE, ATT, ATU

Under the LIV approach (and the parametric), all treatment parameters of concern can be identified by using weighted averages of MTE. Heckman et al. (2006b) show that

\[
\text{ATE}(x,a) = E[B \mid X = x, A = a] = \int_0^1 \text{MTE}(x,a,u_S)w_{ATE}(x,a,u_S)du_S
\]

\[
\text{ATT}(x,a) = E[B \mid X = x, A = a, S = 1] = \int_0^1 \text{MTE}(x,a,u_S)w_{ATT}(x,a,u_S)du_S
\]

\[
\text{ATU}(x,a) = E[B \mid X = x, A = a, S = 0] = \int_0^1 \text{MTE}(x,a,u_S)w_{ATU}(x,a,u_S)du_S,
\]

where the weights are

\[
w_{ATE}(x,a,u_S) = 1
\]

\[
w_{ATT}(x,a,u_S) = \frac{\int_{u_S}^1 f(P(Z) = P(z) \mid X = x, A = a)dP(z)}{E[P(Z) \mid X = x, A = a]}
\]

\[
w_{ATU}(x,a,u_S) = \frac{\int_{0}^{u_S} f(P(Z) = P(z) \mid X = x, A = a)dP(z)}{E[1 - P(Z) \mid X = x, A = a]},
\]

and \(f\) is the density function of \(P(Z)\). By integrating the weighted estimates of \(\text{MTE}(x,a,u_S)\) over the joint distribution of \((X,A)\) the estimates of \(\text{MTE}(u_S)\) are obtained. In practice, however, I do not condition on \((X,A)\) nonparametrically. Instead, I follow Basu et al. (2007) and others and condition on, and thus also integrate over, demideciles of the (estimated) scalar index \(X(\delta_1 - \delta_0) + A(\gamma_1 - \gamma_0)\). Lastly, integrating over \(P(z)\) gives the unconditional estimates of ATE, ATT and ATU. The ability-specific ATEs are obtained by evaluating and comparing the treatment parameters at different values of \(A = a\).
A.2 Computed Weights for ATE, ATT, ATU

Figure A. 2.1 Sample Weights for Different Treatment Parameters

(a) Normal Selection Model

(b) Semiparametric Model

Note: The figures show the weights for ATE, ATT, and ATU, plotted across $U_5$, for the normal selection model and the semiparametric model. The weights are defined in Appendix A.1.
3. Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation*

Introduction

Transmission of economic status within families is often measured by the intergenerational elasticity between parents’ and children’s lifetime income. A large and growing literature has estimated this parameter in order to analyze the extent of intergenerational mobility across countries, groups and time.\(^1\) Unfortunately, the estimates in the early literature suffered greatly from measurement error in lifetime income, and successive improvements of the methodology led to large-scale corrections.\(^2\)

While the early estimates were severely attenuated from approximation of lifetime values by noisy single-year income data for parents, Jenkins (1987) identifies systematic deviations of current from lifetime values over the life

---

\(^*\)This is joint work with Jan Stuhler. Financial support from the Swedish Council of Working Life (FAS) and the German National Academic Foundation is gratefully acknowledged. We thank Anders Björklund and Markus Jäntti for advice and encouragement. We are further grateful for comments from Christian Dustmann, Michael Amior, Anders Böhlmark, Thomas Cornelissen, Flavio Cunha, Nathan Grøve, Steven Haider, Stephen Jenkins, Kristian Koerselmann, Matthew Lindquist, Steve Machin, Marieke Schnabel, Gary Solon, Uta Schönberg, Yoram Weiss, and seminar participants at the 2010 ENTER conference in Toulouse, the 2011 ENTER conference in Tilburg, the 2011 ESPE conference in Hangzhou, the 2011 EALE conference in Cyprus, and at SOFI, Stockholm University.

\(^1\)See Solon (1999) for a comprehensive evaluation of the early empirical literature. Recent surveys include Björklund and Jäntti (2009) and Black and Devereux (2011).

\(^2\)For example, the intergenerational elasticity of earnings for fathers and sons in the U.S. was estimated to be less than 0.2 among early studies (surveyed in Becker and Tomes, 1986), ranged between about 0.3 and 0.5 in the studies surveyed in Solon (1999), and is estimated to be around 0.6 or above in more recent studies like Mazumder (2005) and Gouskova et al. (2010).
cycle as an additional source of inconsistency. Haider and Solon (2006) and Grawe (2006) show that the latter is empirically of great importance. Various refined methods to eliminate such life-cycle bias have recently been presented and the “generalized errors-in-variables model” proposed by Haider and Solon has been widely adopted in the literature.

In this paper we conduct an evaluation of these refined methods, which results in a set of contributions. First, we provide intuition why these methods will not eliminate life-cycle bias in intergenerational elasticity estimates. Second, we use Swedish income data to quantify the importance of life-cycle effects empirically. Our data contain nearly complete lifetime income histories of both fathers and sons, allowing us to derive a benchmark estimate and thus to directly analyze the life-cycle bias in both ordinary least squares (OLS) and instrumental variable (IV) estimates that are based on approximations of lifetime by annual income. Third, we illustrate how current standard procedures can be modified to reduce life-cycle bias. Fourth, we discuss our results in the more general context of income dynamics over the life cycle. We conclude from our analysis that (unobserved) heterogeneity in income profiles is substantial and that it can have important consequences in empirical applications.

The main part of our analysis centers on the generalized errors-in-variables model, which suggests that intergenerational elasticities can be consistently estimated if lifetime income is approximated by current income at a certain age. We find that this procedure improves estimates but that the life-cycle bias is substantially larger than the generalized model predicts. The model disregards some of the heterogeneity in income profiles, and can therefore only partially reduce life-cycle bias. The remaining bias from left-side measurement error alone amounts to about 20 percent of the true intergenerational elasticity (0.21 vs. 0.27) under favorable conditions.\(^3\) We also find that these results apply likewise on a variant of the model proposed in Lee and Solon (2009), which is often used for analyses of mobility trends. The bottom line is that the common practice of measuring annual income at a certain age as surrogate for unobserved lifetime income, which is widespread not only in the intergenerational mobility literature, is subject to substantial life-cycle bias.

We also analyze two other methods to address measurement error in lifetime income: we illustrate why the consideration of differential income growth across subgroups will not yield consistent estimates, and show that IV estimates suffer from even greater life-cycle effects than OLS estimates. IV estimators do therefore not provide an upper bound of the true parameter, contrary

\(^3\)Assuming that central parameters of the generalized errors-in-variables model are perfectly observed, so that current income is measured at the exact proposed age. Right-side measurement error aggravates the life-cycle bias further if fathers’ and sons’ incomes are measured at similar ages.
to previous findings. Our results are hence rather pessimistic. They imply that current methods to compensate for incomplete income data are less successful than commonly assumed, casting doubts on the accuracy of mobility estimates as well as on the validity of comparisons across populations. However, observing our benchmark elasticity allows us to describe the direction and magnitude of the bias at different stages in the life cycle for our population, and thus to provide general recommendations for practitioners. We find that annual income at a late age provides a more reliable base for application of the generalized errors-in-variables model, that averaging over multiple income observations reduces life-cycle bias, and that the treatment of missing and zero income observations has important consequences.

Life-cycle bias stems from a more general mechanism involving the interaction of two factors: heterogeneity in income profiles cannot be fully accounted for, and unobserved idiosyncratic deviations from average profiles correlate with individual and family characteristics. This mechanism is of importance for other literatures that depend on measurement of long-run income. Examples include studies on the returns to schooling, and the extensive literature that relates measures of stochastic income shocks to consumption or other outcomes. We discuss these briefly and present evidence that unexplained dispersion in income growth is at least partially due to latent heterogeneity instead of persistent stochastic shocks.

The next section describes the general methodology and identifying assumptions employed in the early literature. We then examine methods based on more recent contributions: the generalized errors-in-variables model theoretically in Section 3.2 and empirically in Section 3.3, IV methods and consideration of income dynamics across subgroups in Section 3.4. Section 5.6 concludes.

3.1 The Intergenerational Mobility Literature

The most common regression model in intergenerational mobility research is

\[ y_{s,i}^* = \beta y_{f,i}^* + \varepsilon_i, \]  

(3.1)

where \( y_{s,i}^* \) denotes log lifetime income of the son in family \( i \), \( y_{f,i}^* \) log lifetime income of his father, \( \varepsilon_i \) is an error term that is orthogonal to \( y_{f,i}^* \), and variables are expressed as deviations from their generational means.\(^4\) The co-

\(^4\)We use the terms earnings and income interchangeably (since the issues that arise are similar), and examine fathers and sons since this has been the baseline case in the literature. A growing literature exists on intergenerational mobility in other family
efficient $\beta$ is interpreted as the intergenerational income elasticity.

Equations akin to (1) appear in two distinctive forms in the literature. First, as a statistical relationship to measure the outcome of interest, i.e. the degree of intergenerational mobility. Second, as a structural relationship to study causal mechanisms of intergenerational transmission, derived from an economic model as in Becker and Tomes (1979). The statistical relationship is typically based on broad ex-post measures of long-run economic status such as lifetime income. The structural relationship instead relates to the ex-ante concept “permanent income”, since expectations on long-run status determine individual behavior. For simplicity, our analysis relates to the statistical relationship, but incomplete measurement of long-run status impedes identification of both types.

Approximation of Lifetime Income

As commonly available data sets do not contain complete income histories for two generations, a major challenge is how to approximate lifetime income. Let $y_t$ be some observed proxy for unobserved log lifetime income of an individual in family $i$, e.g. a single-year observation, an average of multiple annual income observations, or a more complex estimate based on such annual incomes. Observed values are related to true values by

$$y_{s,i} = y^*_s,i + u_{s,i},$$

where $y^*_s,i$ is the unobserved true log lifetime income of the son in family $i$ and $u_{s,i}$ is measurement error. Similarly, for the father we observe

$$y_{f,i} = y^*_f,i + u_{f,i}.$$

The probability limit of the OLS estimator from a linear regression of $y_s$ on $y_f$ can be decomposed into

$$plim \hat{\beta}_{approx} = \frac{Cov(y_f, y_s)}{Var(y_f)} = \frac{\beta Var(y^*_f) + Cov(y^*_f, u_s) + Cov(y^*_s, u_f) + Cov(u_s, u_f)}{Var(y^*_f) + Var(u_f) + 2 Cov(y^*_f, u_f)},$$

dimensions (e.g. mothers, daughters or siblings) and in other income concepts (such as household income), for which the idea behind our analysis is likewise relevant.

For various reasons these concepts are not always clearly distinguished. First, simple economic models assign one time period to each generation, so that the concept of permanent and lifetime income coincide. Second, permanent income is difficult to measure. Empirical analysis of the structural relationship is still based on ex-post measures of (current) income, and is then often similar to the statistical relationship. Third, some of the empirical work in the literature has lately adopted the term “permanent income” even while focusing on the measurement of outcomes.
where we used eq. (3.1) to substitute for $y^*_s, i$ and applied the covariance restriction $\text{Cov}(y^*_f, \epsilon_i) = 0$. It follows that the estimator can be down- or upward biased and that the covariances between measurement errors and lifetime incomes impact on consistency. The empirical strategies employed in the literature in the last decades can be broadly categorized in terms of changes in identifying assumptions about these covariances.

First Two Waves of Studies

The first wave of studies, surveyed in Becker and Tomes (1986), neglected the problem of measurement error in lifetime status. Often just single-year income measures were used as proxies for lifetime income, thereby implicitly assuming that

$$\text{Cov}(y^*_f, u_s) = \text{Cov}(y^*_s, u_f) = \text{Cov}(u_s, u_f) = \text{Cov}(y^*_f, u_f) = 0,$$

and

$$\text{Var}(u_f) = 0.$$

Classical measurement error in lifetime income violates the latter assumption, so that estimates suffered from large attenuation bias. Estimates of the intergenerational elasticity were therefore too low. This problem was recognized in Atkinson (1980) and then frequently addressed in the second wave of studies (surveyed in Solon 1999). But the assumption remained that measurement errors are random noise, independent of each other and of true lifetime income. That life-cycle variation had to be accounted for was recognized, but it was generally assumed that including age controls in the regression equation would suffice. The assumptions were therefore

$$\text{Cov}(y^*_f, u_s) = \text{Cov}(y^*_s, u_f) = \text{Cov}(u_s, u_f) = \text{Cov}(y^*_f, u_f) = 0,$$

and

$$\text{Var}(u_f) \neq 0.$$

If these hold, then the probability limit in eq. (3.2) reduces to

$$\text{plim} \hat{\beta}_{\text{approx}} = \beta \frac{\text{Var}(y^*_f)}{\text{Var}(y^*_f) + \text{Var}(u_f)}.$$

This is the classical errors-in-variables model; inconsistencies are limited to attenuation bias caused by measurement error in lifetime income of fathers. In contrast, measurement error in sons’ lifetime income is not a source of inconsistency in this model. Researchers typically used averages of multiple income observations for fathers to increase the signal-to-noise ratio, but gave less attention to the measurement of sons’ income.
Recent Literature

Recently the focus has shifted towards the importance of non-classical measurement error. An early theoretical discussion can be found in Jenkins (1987). Analyzing a simple model of income formation, he finds that usage of current incomes in eq. (3.1) will bias $\hat{\beta}$ as income growth over the life cycle varies across individuals. He concludes that the direction of this life-cycle bias is ambiguous, that it can be large, and that it will not necessarily be smaller if fathers’ and sons’ incomes are measured at the same age.

Haider and Solon (2006) demonstrate that life-cycle bias can explain the previously noted pattern that intergenerational elasticity estimates increase with the age of sampled sons. They show that the association between current and lifetime income varies systematically over the life cycle, contrary to a classical errors-in-variables model with measurement error independent of true values. Böhlmark and Lindquist (2006) find strikingly similar patterns in a replication study with Swedish data.

Haider and Solon also note that controlling for the central tendency of income growth in the population by including age controls in eq. (3.1) will not suffice, as variation around the average growth rate will bias estimates. Vogel (2006) provides an illustration based on the insight that highly educated workers experience steeper-than-average income growth. Since available data tend to cover annual incomes of young sons and old fathers, lifetime incomes of highly educated sons (fathers) will be understated (overstated), which is likely to bias $\hat{\beta}_{approx}$ substantially downwards if educational achievement is correlated within families. Indeed, the probability limit of $\hat{\beta}_{approx}$ can be negative in extreme cases, as our data will confirm. Various refined estimation procedures have been proposed to address such life-cycle bias. We proceed to examine the most popular one in detail.

3.2 Measuring Income at a Certain Age

Haider and Solon (HS) generalize the classical errors-in-variables model to allow for variation in the association between annual and lifetime income over the life cycle. Their empirical analysis documents that this variation is substantial, with important implications for analyses based on short-term income measures and, in particular, the intergenerational mobility literature.

6For a summary, see Solon (1999). Age-dependency of elasticity estimates could also arise if the dispersion in transitory income and thus the attenuation bias vary over the life cycle. Such variation has been documented in Björklund (1993) for Sweden, but Grawe (2006) finds that the observed age-dependency can be better explained by the existence of life-cycle bias.
The underlying intuition of the model is that, for two individuals with different income trajectories, there will nevertheless exist an age $t^*$ where the difference between their log annual income equals the difference between their log (annuitized) lifetime income. HS argue that the classical errors-in-variables model holds at this age.

We describe their model here in the context of the intergenerational mobility literature. HS first focus on left-side measurement error and assume that $y^*_{s,i}$ is unobserved and proxied by $y_{s, it}$, log annual income of sons at age $t$. Their generalization of the classical errors-in-variables model is given by

$$y_{s, it} = \lambda_{s, t} y^*_{s, i} + u_{s, it}, \quad (3.3)$$

where $\lambda_{s, t}$ is allowed to vary by age and $u_{s, it}$ is orthogonal to $y^*_{s, i}$. Regressing $y_{s, it}$ on $y^*_{s, i}$ by OLS, and using eqs. (3.3) and (3.1) to substitute for $y_{s, it}$ and $y^*_{s, i}$, yields

$$\text{plim} \hat{\beta}_t = \frac{\text{Cov}(y_{s, t}, y^*_{f, i})}{\text{Var}(y^*_{f, i})} = \beta \lambda_{s, t} + \frac{\text{Corr}(y^*_{f, i}, u_{s, t}) \sigma_{u_{s, t}}}{\sigma_{y^*_{f, i}}}. \quad (3.4)$$

HS assume that

$$\text{Corr}(y^*_{f, i}, u_{s, t}) = 0, \quad (3.5)$$

and thus conclude that left-side measurement error can cause substantial life-cycle bias in intergenerational elasticity estimates, but that it is innocuous for consistency if lifetime incomes of sons are proxied by annual incomes at an age $t^*$ where $\lambda_{s, t}$ is close to one.\(^7\) Their empirical analysis reveals that for an American cohort born in the early 1930s $\lambda_{s, t}$ is below one for young ages, but close to one around midlife.

The model, often referred to as the generalized errors-in-variables (GEiV) model, thus illustrates how life-cycle bias should be expected to vary with age. This result is an important conceptual insight. It is potentially also of great usefulness in applications. In many literatures researchers face the problem that individual long-run outcomes like lifetime income are of theoretical interest, but that available data only contain short snapshots of income. The GEiV model offers a potential remedy since it implies that measurement of income at a certain age might suffice if long-run outcomes are not directly observed. Possible applications are for example the returns to schooling or, as emphasized by HS, the intergenerational mobility literature.

The model has indeed been widely adopted in the latter, where the implied

procedure to measure income at the “right” age has become standard practice. A variation of the model that relies on the same intuition has been presented in Lee and Solon (2009).

But the results of the GEiV model and thus its applicability depend critically on assumption (3.5), as also noted by HS. While the validity of this assumption has to date not been examined or discussed, the current literature nevertheless assumes that the model can eliminate (or nearly eliminate) life-cycle bias in applications. We argue that this is unlikely to be the case since assumption (3.5) or similar assumptions will typically fail to hold.

To understand our reasoning, first note that for more than two workers we will generally not find an age \( t^* \) where annual income is an undistorted approximation of lifetime income. Figure 3.1 illustrates this by plotting log income trajectories for workers 1, 2 (as in Figure 1 in HS) and an additional worker 3. The horizontal lines depict log annuitized lifetime income, and differences in workers’ log lifetime income are given by the vertical distances between these lines. At age \( t^*_1 \) the distance between the annual income trajectories equals the distance between the horizontal lines for workers 1 and 2, and at age \( t^*_2 \) for workers 1 and 3. There exists no age where these distances are equal for all three workers at once. This illustrates that the parameter \( \lambda_{s,t} \) only reflects how differences in annual income and differences in lifetime income relate on average among all workers. Individuals, and groups of individuals, will nevertheless deviate from this average relationship as of dispersion in income profiles, so that their annual income systematically over- or understates their lifetime income compared to the rest of the population.

Among others, in Gouskova et al. (2010) for the US; Björklund et al. (2006, 2009) for Sweden; Nilsen et al. (2011) for Norway; Raaum et al. (2007) for Denmark, Finland, Norway, the UK and the US; Nicoletti and Ermisch (2007) for the UK; Piraino (2007) and Mocetti (2007) for Italy. More examples are covered in the surveys of Björklund and Jäntti (2009) and Black and Devereux (2011).


As an example, Gouskova et al. (2010) conclude: “As Haider and Solon (2006) show in their analysis of the relationship between current income and lifetime income, current income from the early thirties to mid forties generally provides an unbiased estimate of lifetime income, which means it will satisfy the assumptions of the classical error-in-variables model.” Similar conclusions can be found in Grawe (2006) and Lee and Solon (2009).

This result does not depend on a high degree of complexity in income growth processes, but holds for example also for a simple log-linear income formation model as analyzed in HS (see Appendix A.1).

For example, highly educated individuals might experience steeper income growth over the life cycle, such that their annual incomes at different age points systematically over- or understate their lifetime income relative to individuals with less education.
For intergenerational mobility studies it is crucial that such idiosyncratic deviations might correlate within families. There are many reasons to expect such dependency: parents can transmit abilities, or influence their offspring’s educational and occupational choice, all of which can affect income growth and the shape of income profiles over the life cycle. The individual association between annual and lifetime income is thus likely to exhibit an intergenerational correlation itself, and cannot be readily captured by a single population parameter like $\lambda_{s,t}$. Assumption (3.5) is then unlikely to hold, the probability limit of $\hat{\beta}$ does not equal $\lambda_{s,t}\beta$, and knowledge of the exact life-cycle pattern of $\lambda_{s,t}$ cannot eliminate life-cycle bias.\footnote{Corresponding biases arise in the case of right-side measurement error in which unobserved lifetime income of fathers is approximated by annual income (see Appendix A.2) and if approximations are made for both fathers and sons (Appendix A.3).} The illustrative usefulness of the GEiV model is not impaired by these arguments. They however imply that life-cycle bias is harder to address than has been hoped in the literature, and that the search for a “right age” to measure income at might not be an entirely satisfying path to follow.

There are various ways to probe our theoretical arguments. One can examine the validity of assumption (3.5) formally by deriving the elements of $u_{s,it}$
for a given income formation model and analyzing its relation to the regressor $y_{f,i}^*$. While it can be shown that $u_{s,it}$ is correlated with $y_{f,i}^*$ even for a simple log-linear income formation model (see Nybom and Stuhler, 2011), such exercises will not be informative on the magnitude of life-cycle bias that should be expected in practice. In the next section, we will thus instead provide empirical evidence on the size of the bias. In Section 3.4, we will further examine if we indeed find parent-related heterogeneity in the shape of income profiles, which would generate such bias.

3.3 Empirical Evidence on Life-Cycle Bias

We use Swedish panel data containing nearly life-long income histories to provide direct evidence on the life-cycle bias that remains after application of the GEiV model. The size of the bias depends on two factors. First, the complexity of income profiles in the population.\textsuperscript{14} Second, if the dispersion in income profiles is caused by heterogeneity or stochastic shocks. The former more than the latter would cause idiosyncratic deviations from average income profiles to be correlated within families.\textsuperscript{15} Our findings will thus also be indicative about how complex the dispersion in income profiles is, and if its underlying causes are deterministic or stochastic. We return to these issues in Section 3.4.

Data Sources and Sample Selection

To the best of our knowledge, Swedish tax registry data offer the longest panel of income data, covering annual incomes across 48 years for a large and representative share of the population. Moreover, a multi-generational register matches children to parents, and census data provide information on schooling and other individual characteristics. All merged together, the data provide a unique possibility to examine life-cycle bias in intergenerational mobility estimation using actual income histories.

To select our sample, we apply a number of necessary restrictions. As we mainly aim to make a methodological point, we follow the majority of the literature and limit our sample to sons and their biological fathers. To these we

\textsuperscript{14}For example, if individuals merely differ in linear income growth then differences in log lifetime income are well approximated by differences in log current income around midlife for the whole population and the GEiV model would perform relatively well.

\textsuperscript{15}Simulation studies like Stuhler (2010) illustrate these arguments but are not informative about the size of the bias in applications since it varies strongly with unknown characteristics of the income generating process.
merge income data for the years 1960-2007.\textsuperscript{16} Since most other income measures are available only from 1968, we use total (pre-tax) income, which is the sum of an individual’s labor (and labor-related) earnings, early-age pensions, and net income from business and capital realizations.

Our main analysis is based on sons born 1955-1957. Earlier cohorts could be used, but then we would observe fewer early-career incomes for fathers. Conversely, later cohorts are not included, since we want to follow the sons for as long as possible. Moreover, to avoid large differences in the birth year of fathers, we exclude pairs where the father was older than 28 years at the son’s birth. Young fathers and first-born sons are thus over-represented in our sample. Although this is a limitation, we expect any detected bias for this particular sample to understate the bias in the population.\textsuperscript{17} On other sampling issues we adopt the restrictions applied by HS and Böhlmark and Lindquist (2006).\textsuperscript{18}

Our data come with a couple of drawbacks. To maximize the length of the income histories we use the measure total income, whereas e.g. HS use labor earnings. However, total income is a highly relevant measure of economic status, approximation of lifetime status gives rise to the same methodological challenges, and Böhlmark and Lindquist find that total income and earnings yield similar results for the intergenerational mobility of sons. Further, the use of tax-based data could raise concern about missing data in the low end of the distribution if individuals have no income to declare. The Swedish system however provides strong incentives to declare some taxable income since doing so is a requirement for eligibility to most social insurance programs. Hence, this concern most likely only applies to a very small share of the population.

Our data also have many advantages. First, they are almost entirely free from attrition. Second, they pertain to all jobs. Third, in contrast to many other studies, our data are not right-censored. Fourth, we use registry data, which is believed to suffer less from reporting errors than survey data. Fifth, and most important, we have annual data from 1960 to 2007, giving us nearly career-long series of income for both sons and their fathers. Overall, we believe that the data are the best available for the purpose of this study.

Our main sample consists of 3504 father-son pairs, with sons’ income measured from age 22 to age 50 and fathers’ income measured from age 33 to age

\textsuperscript{16} Income data for the year 1967 are missing in the registry.

\textsuperscript{17} Reduced sample heterogeneity will tend to decrease heterogeneity in income profiles, which in turn diminishes the idiosyncratic deviations from sample average relationships between annual and lifetime income that cause life-cycle bias.

\textsuperscript{18} We restrict the sample to fathers and sons who report positive income in at least 10 years. We exclude those who died before age 50, and sons who immigrated to Sweden after age 16 or migrated from Sweden on a long-term basis (at least 10 years).
65, irrespective of birth years. We express all incomes in 2005 prices, apply an annual discount rate of 2 percent, and divide the sums by the number of non-missing income observations to construct our measures of annuitized lifetime income. Table 3.1 reports descriptive statistics. Rows 2 and 3 show that dispersions in lifetime income are of similar magnitudes for fathers and sons. Rows 4 and 5 provide information on the number of positive income observations. On average there are more than 28 observations for sons, and more than 30 for fathers, with relatively low dispersion in both cases.

### Table 3.1 Summary Statistics by Birth Year of Sons

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>1955</th>
<th>1956</th>
<th>1957</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father’s age at birth of son</td>
<td>24.68</td>
<td>24.66</td>
<td>24.77</td>
<td>24.62</td>
</tr>
<tr>
<td></td>
<td>(2.53)</td>
<td>(2.51)</td>
<td>(2.50)</td>
<td>(2.58)</td>
</tr>
<tr>
<td>Log lifetime income (sons)</td>
<td>11.97</td>
<td>11.98</td>
<td>11.98</td>
<td>11.95</td>
</tr>
<tr>
<td></td>
<td>(0.43)</td>
<td>(0.42)</td>
<td>(0.42)</td>
<td>(0.44)</td>
</tr>
<tr>
<td>Log lifetime income (fathers)</td>
<td>11.72</td>
<td>11.73</td>
<td>11.72</td>
<td>11.72</td>
</tr>
<tr>
<td></td>
<td>(0.42)</td>
<td>(0.44)</td>
<td>(0.43)</td>
<td>(0.40)</td>
</tr>
<tr>
<td>Number of pos. income obs. (sons)</td>
<td>28.52</td>
<td>28.57</td>
<td>28.56</td>
<td>28.43</td>
</tr>
<tr>
<td></td>
<td>(1.86)</td>
<td>(1.71)</td>
<td>(1.74)</td>
<td>(2.11)</td>
</tr>
<tr>
<td>Number of pos. income obs. (fathers)</td>
<td>30.32</td>
<td>29.99</td>
<td>30.36</td>
<td>30.59</td>
</tr>
<tr>
<td></td>
<td>(3.76)</td>
<td>(4.13)</td>
<td>(3.62)</td>
<td>(3.48)</td>
</tr>
<tr>
<td>Father-son pairs (N)</td>
<td>3504</td>
<td>1167</td>
<td>1173</td>
<td>1164</td>
</tr>
</tbody>
</table>

Note: The table reports means with standard deviations within parentheses.

### Empirical Strategy

To assess the size of life-cycle bias we compare estimates based on annual incomes with a benchmark estimate that is based on lifetime incomes. As in the theoretical discussion we focus on left-side measurement error (i.e., for sons), although we provide brief evidence on life-cycle bias due to right-side (i.e., for fathers) and measurement error on both sides in a later subsection. We do this for two reasons. First, left-side measurement error has until recently been neglected in the literature. Second, life-cycle bias is not confounded by attenuation bias from classical measurement error on the left-hand side, which simplifies the analysis.

We use our measures of log lifetime incomes $y_{f,i}^*$ and $y_{s,i}^*$ to estimate eq. (3.1) by OLS, which yields our benchmark estimate $\hat{\beta}$.\(^{19}\) We then approximate

\(^{19}\)Of course, this estimate is not exactly true since we still lack some years of income. This is however irrelevant for the validity of our approach to use it as benchmark. The GEiV model is not restricted to any specific population, and should there-
log lifetime income of sons $y_{s,i}^{*}$ by log annual income $y_{s,it}$ (left-side measurement error) to estimate

$$y_{s,it} = \beta_i y_{s,i}^{*} + \epsilon_i$$

separately for each age $t$, to yield a set of estimates $\hat{\beta}_t$. Finally, we estimate eq. (3.3), which provides us with estimates of $\lambda_{s,t}$.

Under the assumptions of the GEiV model, the probability limit of $\hat{\beta}_t$ equals $\lambda_{s,t}\beta$, and using annual income of sons at age $t^*$ where $\lambda_{s,t} = 1$ consistently estimates $\beta$.\(^{20}\) As discussed in the previous section, we anticipate $\hat{\beta}_t$ to be biased even after adjustment by $\hat{\lambda}_{s,t}$. The remaining life-cycle bias after adjustment by the GEiV model, denoted by $b(t) = \hat{\beta}_t / \hat{\lambda}_{s,t} - \tilde{\beta}$, is thus of central interest.\(^{21}\) Note that we assume that $\hat{\lambda}_{s,t}$ is known in order to evaluate the model’s theoretical capability to adjust for life-cycle bias under favorable conditions. A second (known) source of inconsistency can arise in that the age profile of $\lambda_{s,t}$ will typically not be directly estimable by the researcher.

**Empirical Results**

We first present estimates of $\lambda_{s,t}$. Figure 3.2 shows that $\hat{\lambda}_{s,t}$ rises over age and crosses one at around age $t^* = 33$. Largely consistent with others, we find that income differences at young (old) age substantially understate (overstate) differences in lifetime income. We note that $\hat{\lambda}_{s,t}$ is close to one only for a short time around age 33, in contrast to the pattern found for older American and Swedish cohorts in HS and Böhlmark and Lindquist (2006) in which $\hat{\lambda}_t$ remains close to one for an extended period through midlife. A general concern is thus that measuring annual income only a few years earlier or later can cause large differences in elasticity estimates.

Our central estimates are presented in Figure 3.3, which plots $\hat{\beta}$ (the benchmark elasticity), $\hat{\beta}_t$ (estimates based on annual income of sons at age $t$), and

$\tilde{\beta}$ for be applicable to our variant of the Swedish population in which we truncate income profiles at some age. It is nevertheless advantageous that we have long income histories. First, our benchmark estimate will be closer to the true value. Second, since the income profiles contain most of the idiosyncratic heterogeneity that leads to life-cycle bias, our estimate of the bias will be representative for a typical application. We provide evidence that our main findings are not sensitive to the exact length of observed income histories later in this section.

\(^{20}\)Since age is a discrete variable, $\lambda_{s,t}$ will not necessarily equal exactly one at $t^*$. We thus adjust $\hat{\beta}_t$ by $\hat{\lambda}_{s,t}^{-1}$ at all ages, including $t^*$.

\(^{21}\)The arguments of HS relate to the probability limit. In a finite sample we need to consider the distribution of $b(t)$. Reported standard errors for $b(t)$ are based on a Taylor approximation and take the covariance structure of $\hat{\beta}$, $\hat{\beta}_t$, and $\hat{\lambda}_{s,t}$ into account.
Figure 3.2 OLS Estimates of $\lambda_{s,t}$

![Graph of OLS Estimates of $\lambda_{s,t}$](image)

Note: The figure shows estimates of $\lambda_{s,t}$ by sons’ age for cohorts 1955-57. $\lambda_{s,t}$ is the regression coefficient in a regression of son’s annual income on son’s lifetime income, see eq. (3.3).

$\hat{\beta}_t / \hat{\lambda}_{s,t}$ (estimates at age $t$ adjusted by the GEiV model). The sample is balanced within (but not across) each age, such that zero or missing income observations that are not considered for estimation of $\lambda_{s,t}$ and $\beta_t$ are not used to estimate $\beta$. Hence the estimated benchmark elasticity varies slightly by age.

We list our key findings.

**First.** Our benchmark estimate of the intergenerational elasticity of lifetime income for our Swedish cohort is about 0.27 (see also Table 3.2). This is marginally higher than what most previous studies have found for Sweden, and should be closer to the population parameter due to our nearly complete income profiles.\(^{22}\)

**Second.** We confirm that the variation of $\hat{\beta}_t$ over age resembles the pattern in $\hat{\lambda}_{s,t}$, as predicted by HS. We therefore find that $\hat{\beta}_t$ increases with age and that the life-cycle bias is negative for young and positive for old ages of sons. One of the central predictions of the GEiV model, that current income around midlife is a better proxy for lifetime income than income in very young or very old ages, is thus confirmed.

**Third.** The magnitude of life-cycle bias stemming from left-side measurement error alone can be striking. For example, analysis based on annual in-

\(^{22}\)Our benchmark elasticity is nevertheless still likely to understate the true intergenerational elasticity. We lack some early observations of fathers and late observations of sons, which reduces $\sigma_{y_s}$ and increases $\sigma_{y_f}$, thereby reducing the numerator and increasing the denominator of the OLS estimator.
Heterogeneous Income Profiles and Life-Cycle Bias

Figure 3.3 OLS Estimates of Elasticities and Life-Cycle Bias

Note: The figure shows the benchmark estimate of the intergenerational elasticity together with the unadjusted and adjusted estimates based on sons’ annual income. The estimates are for cohort 1955-57, left-side measurement error only (i.e., in sons’ income).

Income of sons only two years below age \( t^* \) yields \( \hat{\beta}_y = 0.191 \), in contrast to a benchmark estimate that is almost 40 percent larger. Moreover, analysis based on income below age 26 yields a *negative* elasticity. We therefore find direct evidence on the importance of life-cycle bias in intergenerational mobility estimates that has been discussed in the recent literature.

**Fourth.** The life-cycle bias is larger than implied by the GEiV model. While the adjustment of estimates according to this model leads on average to sizable improvements, it cannot fully eliminate the bias. This holds true even under the assumption that the central parameters \( \lambda_{s,t} \) are perfectly observed. The remaining bias is overall substantial, and especially large for young ages. Intergenerational elasticity estimates based on income at very young ages are still negative.\(^{23}\)

**Fifth.** The life-cycle bias is not minimized at age \( t^* \), the age at which the current empirical literature aims to measure income, but at an age \( t > t^* \). We report a similar pattern for other cohorts in the subsection with extensions below.

\(^{23}\)The results further imply that even exact knowledge of the pattern of \( \lambda_{s,t} \) over age is not much more useful than the rule of thumb that income should be measured around midlife instead of young or old ages. In fact, for our cohort the “correction” of elasticity estimates by \( \lambda_{s,t} \) worsens elasticity estimates around midlife (age 33 to age 40), but improves estimates at older ages.
Table 3.2 OLS Estimates of Elasticities and Life-Cycle Bias

<table>
<thead>
<tr>
<th>t=Age</th>
<th>$\lambda_{s,t}$</th>
<th>$\hat{\beta}$</th>
<th>$\hat{\beta}_t$</th>
<th>$\hat{\beta}<em>t / \lambda</em>{s,t}$</th>
<th>$b(t)$</th>
<th>$b(t)$ in %</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>0.897</td>
<td>0.266</td>
<td>0.191</td>
<td>0.213</td>
<td>-0.053</td>
<td>19.8</td>
<td>3478</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.016)</td>
<td>(0.023)</td>
<td>(0.029)</td>
<td>(0.021)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.909</td>
<td>0.267</td>
<td>0.246</td>
<td>0.271</td>
<td>0.003</td>
<td>1.3</td>
<td>3476</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.016)</td>
<td>(0.023)</td>
<td>(0.028)</td>
<td>(0.021)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>0.982</td>
<td>0.267</td>
<td>0.203</td>
<td>0.207</td>
<td>-0.061</td>
<td>22.7</td>
<td>3479</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.016)</td>
<td>(0.025)</td>
<td>(0.031)</td>
<td>(0.024)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>1.039</td>
<td>0.256</td>
<td>0.212</td>
<td>0.204</td>
<td>-0.051</td>
<td>20.1</td>
<td>3469</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.016)</td>
<td>(0.025)</td>
<td>(0.031)</td>
<td>(0.023)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>1.114</td>
<td>0.261</td>
<td>0.234</td>
<td>0.210</td>
<td>-0.052</td>
<td>19.7</td>
<td>3460</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.016)</td>
<td>(0.027)</td>
<td>(0.029)</td>
<td>(0.022)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Cohort group 1955-1957, left-side measurement error only (i.e., in sons’ income). Standard errors in parentheses, which for $\hat{\beta}_t / \lambda_{s,t}$ and $b(t)$ are based on Taylor approximations that take the covariance structure of $\lambda_{s,t}$, $\hat{\beta}$, and $\hat{b}_t$ into account. Column (7) displays $b(t)$ in percent of our benchmark estimate $\hat{\beta}$.

Sixth. The remaining life-cycle bias $\hat{b}(t)$ around age $t^*$ is substantial and significantly different from zero. Table 3.2 shows that $\hat{b}(t)$ is on average around 0.05 over ages 31-35, which corresponds to about 20 percent of our benchmark. Furthermore, the large deviation from this average at age 32 indicates that the year-to-year variation can be large. Knowledge of age $t^*$ will thus not eliminate life-cycle bias.

We briefly compare these empirical results with our theoretical discussion of the determinants of $\hat{b}(t)$. Table 3.3 shows the components $\hat{b}(t)$ according to eq. (3.4). Variation of $\hat{b}(t)$ over age stems mostly from variation in the residual correlation $\text{Corr}(y^*_{f}, u_{s,t})$, while the ratio $\sigma_{u_{s,t}} / \lambda_{s,t} \sigma_{y^*_{f}}$ is close to one over most of the life cycle.\(^{24}\) Seemingly small residual correlations can thus translate into substantive biases. For example, a residual correlation of 0.03 translates into a life-cycle bias of more than 10 percent of the benchmark elasticity.

These results provide guidance for applied research, but some remarks about generalizability are warranted. Life-cycle bias will differ quantitatively across populations. The bias is determined by the degree of systematic differences in income profiles between sons from poor and sons from rich families. This mechanism is likely to vary across cohorts and countries. The question is if observed qualitative patterns over age can nevertheless be generalized. Figure 3.3 demonstrates that annual income at late age provides a more reliable base for application of the GEiV model in intergenerational studies than income at young age. The remaining life-cycle bias is large and negative up

\(^{24}\)The previously documented increase in $\lambda_{s,t}$ over age is offset by an increase in $\sigma_{u_{s,t}}$.\)
Table 3.3 Decomposition of Life-Cycle Bias

<table>
<thead>
<tr>
<th>t=Age</th>
<th>$\hat{b}(t)$</th>
<th>$\text{Corr}(y_f^*, \hat{u}_{s,t})$</th>
<th>$\hat{\sigma}<em>{u</em>{s,t}}$</th>
<th>$\hat{\sigma}_{y_f^*}$</th>
<th>$\hat{\sigma}<em>{u</em>{s,t}}/\hat{\lambda}<em>{s,t}\hat{\sigma}</em>{y_f^*}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>-0.053</td>
<td>-0.044</td>
<td>0.455</td>
<td>0.424</td>
<td>1.198</td>
</tr>
<tr>
<td>32</td>
<td>0.003</td>
<td>0.003</td>
<td>0.431</td>
<td>0.423</td>
<td>1.123</td>
</tr>
<tr>
<td>33</td>
<td>-0.061</td>
<td>-0.052</td>
<td>0.485</td>
<td>0.422</td>
<td>1.169</td>
</tr>
<tr>
<td>34</td>
<td>-0.051</td>
<td>-0.050</td>
<td>0.452</td>
<td>0.422</td>
<td>1.031</td>
</tr>
<tr>
<td>35</td>
<td>-0.052</td>
<td>-0.049</td>
<td>0.494</td>
<td>0.422</td>
<td>1.050</td>
</tr>
</tbody>
</table>

Note: The table displays the remaining bias, $\hat{b}(t)$, together with its associated statistical components. Results are for cohort group 1955-1957, left-side measurement error only (i.e., in sons’ income).

until the early forties, but then small for most older ages.\textsuperscript{25} Thus, the relationship between current and lifetime income differs with respect to family background particularly at the beginning of the life cycle. This result is intuitive if one considers potential causal mechanisms of intergenerational transmission. Sons from rich families might acquire more education or face different conditions that particularly affect initial job search (e.g. regarding credit-constraints, family networks, or ex-ante information on labor market characteristics). Such mechanisms are likely to apply to some degree to most populations. Although the size of the life-cycle bias is bound to differ across populations, its pattern over age is thus likely to hold more generally. This conclusion is supported by results for other Swedish cohorts, as we will discuss later on.

Extensions

We proceed to examine alterations of the estimation procedure to reduce the bias, test the sensitivity of our results, and discuss if adjustments according to the GEiV model can eliminate life-cycle bias in other applications.

Multi-Year Averages of Current Income

Some recent studies that reference to the GEiV model (see footnote 8) average over multiple income observations of sons, although without clear theoretical motivation. Our finding that life-cycle bias is substantial even if this age would be known raises the question if and how such practice can help to reduce the bias.

\textsuperscript{25}The latter result cannot easily be exploited. Adjustment of $\hat{\beta}_t$ by $\hat{\lambda}_{s,t}$ can rarely be done in practice due to lack of information on the latter. Importing estimates of $\hat{\lambda}_{s,t}$ from other sources can be misleading since its pattern over age could differ across populations.
We therefore estimate $\beta_t$ using three-, five- and seven-year averages of son’s income centered around age $t^*$. These averages are also used to estimate $\lambda_{s,t}$, and the remaining life-cycle bias after adjustment by $\hat{\lambda}_{s,t}$. The results are summarized in Table 3.4. The remaining life-cycle bias falls in the number of income observations but is not eliminated. For the seven-year average, the estimated bias (in absolute value) is on average slightly below 0.03 at ages 31-35 compared to about 0.05 using one-year measures. The standard deviation of the residuals $\hat{\sigma}_{us,t}$, which is a central component of the bias, decreases by about a third when moving from one- to seven-year measures, and diminishes the estimated bias proportionally. The residual correlation falls only slightly and estimates of $\lambda_{s,t}$ remain stable. These improvements are moderate, but they are generalizable since they are simply driven by the fact that the residual variance decreases when more income observations are used. We believe that this is a clarifying result that provides a rationale for averaging over multiple income observations on the left-hand side when possible.

### Table 3.4 OLS Estimates with Multi-Year Averages of Son’s Income

<table>
<thead>
<tr>
<th>Age</th>
<th>Three-Year Average</th>
<th>Five-Year Average</th>
<th>Seven-Year Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}$</td>
<td>$\hat{\beta}$</td>
<td>$\hat{\beta}$</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>$\beta$</td>
<td>$\beta$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\lambda}$</td>
<td>$\hat{\lambda}$</td>
<td>$\hat{\lambda}$</td>
</tr>
<tr>
<td></td>
<td>$b(t)$</td>
<td>$b(t)$</td>
<td>$b(t)$</td>
</tr>
<tr>
<td>31</td>
<td>0.268 0.218 -0.015</td>
<td>0.268 0.213 -0.017</td>
<td>0.270 0.221 -0.015</td>
</tr>
<tr>
<td></td>
<td>(0.016) (0.020) (0.017)</td>
<td>(0.016) (0.018) (0.014)</td>
<td>(0.016) (0.018) (0.013)</td>
</tr>
<tr>
<td>32</td>
<td>0.267 0.214 -0.041</td>
<td>0.268 0.229 -0.020</td>
<td>0.268 0.227 -0.018</td>
</tr>
<tr>
<td></td>
<td>(0.016) (0.021) (0.017)</td>
<td>(0.016) (0.020) (0.014)</td>
<td>(0.016) (0.018) (0.013)</td>
</tr>
<tr>
<td>33</td>
<td>0.267 0.229 -0.041</td>
<td>0.267 0.226 -0.043</td>
<td>0.268 0.238 -0.023</td>
</tr>
<tr>
<td></td>
<td>(0.016) (0.022) (0.022)</td>
<td>(0.016) (0.021) (0.016)</td>
<td>(0.016) (0.020) (0.013)</td>
</tr>
<tr>
<td>34</td>
<td>0.268 0.229 -0.056</td>
<td>0.267 0.235 -0.044</td>
<td>0.268 0.245 -0.038</td>
</tr>
<tr>
<td></td>
<td>(0.016) (0.024) (0.020)</td>
<td>(0.016) (0.022) (0.018)</td>
<td>(0.016) (0.022) (0.015)</td>
</tr>
<tr>
<td>35</td>
<td>0.262 0.232 -0.059</td>
<td>0.268 0.247 -0.052</td>
<td>0.267 0.251 -0.044</td>
</tr>
<tr>
<td></td>
<td>(0.016) (0.025) (0.020)</td>
<td>(0.016) (0.024) (0.019)</td>
<td>(0.016) (0.023) (0.017)</td>
</tr>
</tbody>
</table>

Note: Cohort group 1955-1957, left-side measurement error only (i.e., in sons’ income). Standard errors are in parentheses.

### Treatment of Outliers in the Income Data

Intergenerational elasticity estimates can be sensitive to how one treats outliers in general, and observations of zero or missing income in particular (Couch and Lillard, 1998; Dahl and DeLeire, 2008). We test the robustness of our results along this dimension by (i) balancing the sample across ages such that only sons with positive income in all ages 31-35 are included, (ii) bottom coding very low non-missing incomes, and (iii) top coding very high incomes.\(^{26}\) We

\(^{26}\)As of the log-specification we do not expect high extremes to have as large influence as low extremes. Top-coding has however been suggested to test the sensitivity
compare the life-cycle bias for ages 31-35 for each of these samples (summarized in Table 3.5) with the results for our main sample in Table 3.2.

Table 3.5 Summary of Robustness Tests

<table>
<thead>
<tr>
<th>Age</th>
<th>Balanced Sample</th>
<th>Bottom-Coded Incomes</th>
<th>Top-Coded Incomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}$</td>
<td>$\hat{\beta}_b$</td>
<td>$b(t)$</td>
</tr>
<tr>
<td>31</td>
<td>0.257</td>
<td>0.184</td>
<td>-0.033</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.021)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>32</td>
<td>0.257</td>
<td>0.227</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.020)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>33</td>
<td>0.257</td>
<td>0.185</td>
<td>-0.053</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.023)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>34</td>
<td>0.257</td>
<td>0.219</td>
<td>-0.029</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.022)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>35</td>
<td>0.257</td>
<td>0.239</td>
<td>-0.027</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.024)</td>
<td>(0.026)</td>
</tr>
</tbody>
</table>

Note: Cohort group 1955-1957, left-side measurement error only (i.e., in sons’ income). Standard errors are in parentheses. The sample in columns (1)-(3) is balanced across ages, hence excluding individuals who have zero or missing incomes at any age 31-35. The sample in columns (4)-(6) is with low non-missing incomes bottom-coded as 10 000 SEK. The sample in columns (7)-(9) is with high incomes top-coded as 2 000 000 SEK.

Estimates of the remaining life-cycle bias are on average about a third lower for the balanced sample than for our main sample (at ages 31-35), but still correspond to more than 10 percent of the benchmark elasticity. Decreases in both the residual correlation and residual variance contribute to this drop.\(^{27}\) Bottom coding has the opposite effect and increases the bias slightly since observations with zero income are now always included. Finally, results for a sample with top-coded incomes are very similar to those for the main sample, implying low sensitivity to the exact measurement of high incomes. While we thus find that zero and missing incomes are influential for the size of life-cycle bias, it is not obvious what the right sampling choice would be. To derive a general measure of mobility one would like to include all individuals, but our analysis shows that doing so comes with the cost of increased life-cycle bias.

\(^{27}\)Excluding those with occasional zeros or missings reduces the number of extreme values and thereby the variation in $u_{s,lt}$. The residual correlation decreases since individuals with frequent zero and missing income observations are likely to experience quite different income profiles than the average population, and therefore amplify the heterogeneity in income profiles that causes the residual correlation.
Length of Observed Income Profiles

Although our data are to our knowledge the best available for our purpose, it might be a concern that our measures of lifetime income are still based on incomplete income histories. We thus perform a number of robustness tests. We consider a younger cohort — sons born 1958-60 — to study the influence of early-age income data of fathers, and an older cohort — born 1952-54 — to study the influence of late-age data of sons.

Age profiles of the life-cycle bias before and after adjustment by \( \hat{\lambda}_{s,t} \) are shown in the Appendix in Figures A.3.1 (main sample), A.3.2 (cohort 1958-60), and A.3.3 (cohort 1952-54) for variations of the age spans. Abstracting from general cohort differences, we find that changes in the fathers’ age span have little effect on the life-cycle bias, probably due to our focus on left-side measurement error. In contrast, changes in the sons’ age span cause noticeable shifts. This is not unexpected since changes in the age span on which our measures of lifetime income are based are likely to alter both \( \sigma_{y_s} \) and \( \lambda_{s,t} \) slightly. While the exact relation between the size of the life-cycle bias and age therefore depends on the definition of the age span, the major facts remain stable: the remaining life-cycle bias after adjustment by \( \hat{\lambda}_{s,t} \) can be large and tends to be negative for young ages and around \( t^* \).

Cohort and Population Differences

We use the same three cohort groups to briefly assess if the magnitude of life-cycle bias can be expected to vary across populations. To separate true cohort differences from differences due to age span definitions, we limit the income profiles of both fathers and sons to the longest age span observed in all three samples. We thus use incomes of sons for ages 22-47, and incomes of fathers for ages 36-65.\(^{28}\) Differences between these samples are hence due to their respective data generating processes.

Table 3.6 presents the most central results around age \( t^* \) for each sample.\(^{29}\) The 1958-60 cohort has an estimated benchmark elasticity \( \hat{\beta} \) that is similar to our main cohort but a slightly larger remaining life-cycle bias \( \hat{b}(t) \). For the 1952-54 cohort both \( \hat{\beta} \) and \( \hat{b}(t) \) are substantially lower.

Figure 3.4 plots estimates of \( \beta_t \) for all three samples over the full age range. While the overall pattern over age are relatively similar, the differences between elasticity estimates at each age are quite volatile. These differences

\(^{28}\)Restricting the age intervals reduces the benchmark estimate. Dropping income observations for sons at old age and fathers at young age decreases \( \sigma_{y_s} \) and increases \( \sigma_{y_f} \), reducing the numerator and increasing the denominator of the OLS estimator.

\(^{29}\)More detailed evidence on cohort differences is also provided in Figures 11 and 12 in Nybom and Stuhler (2011).
Table 3.6 Summary of Cohort Differences, Averages over Ages 31-35

<table>
<thead>
<tr>
<th>Cohort Group</th>
<th>$\lambda_{s,t}$</th>
<th>$\beta$</th>
<th>$\beta_t$</th>
<th>$\beta_t / \lambda_{s,t}$</th>
<th>$b(t)$</th>
<th>$b(t)$ in %</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1958-60</td>
<td>1.071</td>
<td>0.274</td>
<td>0.235</td>
<td>0.220</td>
<td>-0.054</td>
<td>19.9</td>
<td>3427</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.016)</td>
<td>(0.028)</td>
<td>(0.032)</td>
<td>(0.026)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955-57</td>
<td>1.066</td>
<td>0.246</td>
<td>0.216</td>
<td>0.204</td>
<td>-0.042</td>
<td>17.2</td>
<td>3444</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.015)</td>
<td>(0.024)</td>
<td>(0.028)</td>
<td>(0.020)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1952-54</td>
<td>1.059</td>
<td>0.206</td>
<td>0.190</td>
<td>0.179</td>
<td>-0.027</td>
<td>12.8</td>
<td>3160</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.015)</td>
<td>(0.024)</td>
<td>(0.027)</td>
<td>(0.019)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Left-side measurement error only (i.e., in sons’ income). Table displays averages of estimates and standard errors (in parentheses) across ages 31-35. $b(t)$ is significantly different from zero (p-value<0.05) at three ages (out of five) for 1958-60, at four ages for 1955-57, and at two ages for 1952-54. To ensure comparability, lifetime income is restricted to be measured over identical ages for all cohorts: 22-47 for sons, and 36-65 for fathers. Column (7) displays $b(t)$ in percent of our benchmark estimate $\bar{\beta}$ (as average over the age interval).

Figure 3.4 OLS Estimates of Elasticities for Various Cohorts

Note: Cohort 1955-57, left-side measurement error only (i.e., in sons’ income).
Variants in the Intergenerational Mobility Literature

Lee and Solon (2009) present an extension of the GEiV model that allows researchers to use income observations over multiple years. They exploit that the results in Haider and Solon (2006) imply that life-cycle bias is a function of sons’ age at measurement, an implication that is confirmed by our data. This functional relation can thus be explicitly captured in a regression equation such as

\[
y_{s,it} = \alpha' D + \beta y_{f,i}^* + \delta_1 (t - t^*) + \ldots + \delta_4 (t - t^*)^4 + \theta_1 y_{f,i}^* (t - t^*) + \ldots + \theta_4 y_{f,i}^* (t - t^*)^4 + \epsilon_i, \tag{3.6}
\]

This equation contains a vector of year dummies \(D\), a quartic in child’s age (normalized to zero at age \(t^*\)), and an interaction of the child’s normalized age quartic with father’s log lifetime income.\(^{30}\) Intuitively, the latter approximates the life-cycle pattern of \(\beta_t\) as evident in Figure 3.3. The choice of \(t^*\) reflects at which age elasticity estimates are expected to be unbiased based on the predictions of the GEiV model.

This specification provides two important advantages. First, the usage of additional income observations for each cohort can potentially improve statistical efficiency. Second, intergenerational elasticities can be estimated for cohorts for which income is not observed at age \(t^*\).\(^{31}\) These two properties are especially useful in analyses of trends in income mobility, as these typically rely on sparse income data and measurement of income at young or old ages for some cohorts (see Lee and Solon, 2009). Hence this specification has been used in most of the recent research on mobility trends, as for example in Hertz (2007), Pekkala and Lucas (2007), Nicoletti and Ermisch (2007) and Aaronson and Mazumder (2008).

However, specification (3.6) requires that elasticity estimates are unbiased at age \(t^*\), and thus relies on the same assumption as the GEiV model. We therefore expect it to be subject to a similar degree of life-cycle bias. To probe this conjecture we estimate (3.6) by OLS, separately for 10-year intervals of

\(^{30}\)Our specification is a simplified version of the specification used in Lee and Solon (2009). First, we do not control for father’s age since we observe comparable measures of lifetime income for all fathers. Second, we only estimate one elasticity parameter \(\beta\) instead of elasticity parameters for each cohort or year since we are testing for life-cycle bias instead of estimating mobility trends.

\(^{31}\)Both these advantages however hinge on assumptions, namely (i) that the pattern of life-cycle bias over the age of sons can be well approximated by a fourth-order polynomial, and (ii) that this age pattern is stable across cohorts. The latter seems problematic given the results presented in our previous subsection. From Figure 3.3 we however expect that the first assumption is indeed valid.
The resulting estimates of $\beta$ are all in the range $0.205-0.218$, close to our previously reported estimate $\hat{\beta}_t = 0.203$ that is based on sons’ annual income at age $t^* = 33$. Statistical precision does however indeed rise, as standard errors of the elasticity estimates shrink by almost a half. Both findings also hold when using a randomly selected number of income observations for each son in a given age range. We conclude that Lee and Solon’s objective to improve statistical efficiency by pooling income observations over multiple years has been fulfilled, but that estimates suffer from a similar level of life-cycle bias as other estimates that are based on the GEiV model. Estimates still differ by almost 20 percent from our benchmark elasticity based on lifetime incomes, even when based on a large number of income observations per son. This remaining life-cycle bias can differ by cohort (see previous subsection) and may thus mask gradual changes of mobility over time or generate a false impression of such trends. These results may partly explain why the recent literature on intergenerational mobility trends has produced wildly diverging estimates (see Lee and Solon, 2009).

The GEiV Model in Other Applications

The GEiV model can be applied to other literatures that use short-run income to proxy for unobserved long-run values. However, our explanation why the model will fail to eliminate life-cycle bias generalizes to literatures in which interest lies on other explanatory variables. The GEiV model captures changes in the average association between annual and lifetime income in the population over age. But in most applications we compare subgroups of the population. The GEiV model fails to eliminate life-cycle bias since the association between annual and lifetime income does not only vary over age, it also varies over subgroups defined by parental income, years of schooling, ethnicity, gender, or other characteristics.

To provide brief evidence, we examine if the residuals from eq. (3.3) correlate with a range of characteristics that are of interest in various literatures, specifically (i) father’s age at birth of his son, (ii) father’s education, (iii) son’s education, (iv) son’s cognitive ability, and (v) son’s country of birth. Table 3.7

32 We do not report estimates that are based on income spans at very young or very old ages of sons since estimates of $\hat{\beta}$ become very erratic if not at least some observations around age $t^*$ are included in the regression. This result is due to the fourth-order polynomial approximation of life-cycle patterns in (3.6). If researchers have to rely on extrapolation from observations at very young or very old ages then usage of a quadratic instead of a quartic provides more reliable results, as our empirical findings confirm.
describes how each variable is measured and presents the results. As expected, most of these correlations are significantly different from zero. Importantly, the residual correlations are non-zero also around age \( t^* \). Knowing this age, or the age profile of \( \lambda_{s,t} \), does therefore not allow researchers to fully control for life-cycle effects. The correlations tend to be smaller when sons’ income is measured at later ages, again supporting our argument that the GEiV performs better when applied to current income at ages \( t > t^* \).

**Table 3.7 Correlations Between Residuals and Characteristics**

<table>
<thead>
<tr>
<th>Age Interval of Sons</th>
<th>26-30</th>
<th>31-35</th>
<th>36-40</th>
<th>41-45</th>
<th>46-50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Father’s log lifetime income</td>
<td>-0.057*</td>
<td>-0.050*</td>
<td>-0.063*</td>
<td>-0.020</td>
<td>-0.007</td>
</tr>
<tr>
<td>Father’s age at birth of son</td>
<td>-0.054*</td>
<td>0.014</td>
<td>0.045*</td>
<td>0.017</td>
<td>-0.006</td>
</tr>
<tr>
<td>Father’s education</td>
<td>-0.158*</td>
<td>-0.061*</td>
<td>-0.045*</td>
<td>0.035</td>
<td>0.028</td>
</tr>
<tr>
<td>Son’s education</td>
<td>-0.278*</td>
<td>-0.112*</td>
<td>-0.002</td>
<td>0.085*</td>
<td>0.088*</td>
</tr>
<tr>
<td>Son’s cognitive ability</td>
<td>-0.108*</td>
<td>-0.073*</td>
<td>-0.050*</td>
<td>0.022</td>
<td>-0.004</td>
</tr>
<tr>
<td>Son’s country of birth</td>
<td>-0.040*</td>
<td>-0.026</td>
<td>-0.002</td>
<td>-0.032</td>
<td>0.028</td>
</tr>
</tbody>
</table>

Note: The table reports, for cohort 1955-1957, correlations between characteristics listed in the first column and sons’ income residuals (as average in each five-year age interval) from eq. (3). The education variables are years of education measured at about age 35, "Son’s country of birth" is an indicator for being born outside Sweden, and "Son’s cognitive ability" is a standardized cognitive ability measure from the military enlistment test at age 18. Star superscripts indicate correlations with \( p \)-value<0.05.

The residuals correlate much stronger with son’s and father’s education than with father’s log lifetime income, indicating that the GEiV model might perform worse in applications in which education plays a central role, such as the returns to schooling literature.\(^{33}\) Knowledge of the model’s central parameter \( \lambda_t \) is insufficient, since the association between annual and lifetime income of different educational groups deviates substantially from the population-wide association that is summarized by \( \lambda_t \). Bhuller et al. (2011) present an analysis of life-cycle bias in returns to schooling estimates, and also discuss the applicability of the GEiV model in this context.

Measurement Error on the Right-Hand Side or Both Sides

Although our findings on left-side measurement error are conceptually interesting, evidence on the combined effects of life-cycle bias from both sides is more relevant for practitioners. The questions arise whether we find similar life-cycle effects from the right-hand side, and whether these tend to cancel

\(^{33}\)Correlation with the schooling variables is large in particular for young age, reflecting that income growth of the highly educated is relatively strong while initial income is low.
out or aggravate the effects from left-side measurement error. Our data allow us to directly examine these questions. We now base estimates of $\beta_t$ on lifetime income of sons and approximation of lifetime income by annual income for fathers (right-side measurement error) or approximation for both fathers and sons (measurement error on both sides). The probability limit of $\hat{\beta}_t$ is then affected by attenuation and life-cycle bias. We adjust for both according to the GEiV model. Results are shown in Figures 3.5 and 3.6.\(^{34}\)

**Figure 3.5 OLS Estimates of Elasticities with Right-Side Measurement Error**

![Figure 3.5 OLS Estimates of Elasticities with Right-Side Measurement Error](image)

Note: Cohort 1955-57, right-side measurement error only (i.e., in fathers’ income).

Figure 3.5 demonstrates the additional large attenuating effects from right-side measurement error. The remaining life-cycle bias after adjustment by the GEiV model follows a similar qualitative pattern over age as for the case of left-side measurement error. Figure 3.6 shows the remaining life-cycle bias in the case of measurement error on both sides with fathers’ and sons’ incomes

\(^{34}\)Adjustment is based on separate estimates of $\lambda_t$ for both fathers and sons, denoted $\hat{\lambda}_{f,t}$ and $\hat{\lambda}_{s,t}$. According to the GEiV model the probability limit of $\hat{\beta}_t$ equals $\theta_{f,t}\beta = \left(\hat{\lambda}_{f,t} \sigma^2_{y^*f} / (\lambda^2_{f,t} \sigma^2_{y^*f} + \sigma^2_u)\right)\beta$ for right-side and $\lambda_{s,t} \theta_{f,t}\beta$ for both-side measurement error (assuming $\text{Corr}(u_{s,it}, u_{f,it}) = 0$ in addition to assumption 3.5). Therefore the remaining life-cycle biases equal $b(t) = \hat{\beta}/\hat{\theta}_{f,t} - \hat{\beta}$ and $b(t) = \hat{\beta}/\hat{\lambda}_{s,t} \hat{\theta}_{f,t} - \hat{\beta}$, respectively. See Appendix A.2 and A.3 for a detailed derivation of the components of these biases. For presentational purpose we use only one age subscript $t$ and display combinations of annual income for sons and fathers with equal distances to their respective $t^*$ in Figure 3.6.
measured at similar ages. It is overall larger than for left-side measurement error alone, thus indicating aggravating effects of measurement error on both sides.\footnote{This holds true if estimates are only adjusted for attenuation bias but not for life-cycle effects according to the GEiV model (see Figure 13 in our working paper Nybom and Stuhler, 2011). These results confirm and substantiate the theoretical predictions of Jenkins (1987) that measuring fathers’ and sons’ income at similar ages might not necessarily reduce life-cycle bias, and contradict arguments in the recent literature that such “life course matching” generally leads to smaller biases than asymmetric age combinations.} Importantly, this is also the case when fathers’ and sons’ incomes are measured at their respective $t^*$. We again find that the GEiV model is less successful in eliminating the bias for early ages and around $t^*$ than for later ages. Moreover, the estimates suffer from strong year-to-year variability, supporting the argument that comparisons of elasticity estimates across populations are likely to be difficult. Reducing this variability is an additional motive for averaging over multiple income observations on both sides, apart from our previous finding that it reduces the size of the bias.

Note: Cohort 1955-57, measurement error on both sides (i.e., in fathers’ and sons’ income). To stay in two dimensions, we only display results for annual incomes at the same distance from $t^*$ for sons and fathers. At $s=0$ both are measured at their respective $t^*$, at $s=5$ both are measured five years after $t^*$, etc.
3.4 Other Methods to Address Life-Cycle Bias

We briefly examine two other methods employed in the intergenerational mobility literature. We show that instrumental variable (IV) estimators do not yield an upper bound for $\beta$ due to life-cycle effects, and discuss why the consideration of differential income growth rates across subgroups will not suffice to yield consistent estimates.

Instrumental Variable Methods

IV methods have been proposed as an alternative way to tackle attenuation bias that stems from right-side measurement error in eq. (3.1). Furthermore, in the form of two-sample IV (TSIV) they are heavily relied on for countries with less rich data.\textsuperscript{36} Little is however known about life-cycle effects in IV estimates.

Instruments used in the literature include indices of father’s socioeconomic status (Zimmerman, 1992) and father’s education (Solon, 1992). We here relate to the latter, and analyze the probability limit of the IV estimator $\hat{\beta}_{IV,approx}^*$ that follows from a two-stage least-squares regression of $y_{s,i}$ (with $y_{s,i} = y_{s,i}^* + u_{s,i}$) on $y_{f,i}$ (with $y_{f,i} = y_{f,i}^* + u_{f,i}$), using father’s education $E_{f,i}$ as an instrument for father’s income $y_{f,i}$.\textsuperscript{37}

As shown by Solon (1992), the IV estimator is an upper bound for $\beta$ if $E_f$ is uncorrelated with $u_f$ and $u_s$ (assumption IV1, classical errors-in-variables model), if father’s education has a non-negative relation with son’s income (ass. IV2), and if its correlation with father’s lifetime income is bounded between zero and one (ass. IV3).

However, we argue that assumption IV1 does typically not hold, for example since education should be expected to correlate with income growth rates. Typical panel data cover income observations of young sons and old fathers. Then lifetime income of highly educated sons (fathers) will be understated (overstated), even if we control for the central tendency of income growth in the population, and $\hat{\beta}_{IV,approx}^*$ will not necessarily be an upper bound for $\beta$.\textsuperscript{38}

The results from Grawe (2006) and Haider and Solon (2006) indicate that IV estimates might bound $\beta$ if they are adjusted by a certain parameter ($\lambda_t$ in the GEiV model), or if they are based on income at an age at which this param-

\textsuperscript{36}TSIV was first applied to the intergenerational mobility literature by Björklund and Jäntti (1997).
\textsuperscript{37}For generality we here use the notation from Section 3.1 and exclude age subscripts.
\textsuperscript{38}Our working paper Nybom and Stuhler (2011) contains a formal derivation of the probability limit of the IV estimator given that assumption IV1 does not hold.
eter equals one. Our results imply that this is not the case since measurement errors in lifetime income are not orthogonal to education in either case (see section 3.3).

**Empirical Evidence on Life-Cycle Bias in IV Estimates**

We use our Swedish data to examine life-cycle bias in IV estimates. As most of the literature, we instrument for father’s income by his years of education.\(^{39}\)

We first derive a benchmark estimate \(\hat{\beta}^\text{IV}\) using our measures of lifetime income \(y^*_f\) and \(y^*_s\), to assess if IV methods provide an upper bound if lifetime incomes would be truly observed. The IV benchmark estimate (\(\hat{\beta}^\text{IV} = 0.309, \hat{\sigma}_{\hat{\beta}^\text{IV}} = 0.037\)) is larger than our OLS benchmark estimate (\(\hat{\beta} = 0.273, \hat{\sigma}_{\hat{\beta}} = 0.017\)), indicating that assumptions IV2 and IV3 hold, and that the IV estimator could indeed provide an upper bound if based on lifetime incomes. However, the difference \(\hat{\beta}^\text{IV} - \hat{\beta}\) is statistically insignificant.

To probe the validity of assumption IV1, we again focus on left-side measurement error (using son’s annual income at age \(t\)) in order to abstract from attenuation bias and to directly compare life-cycle bias in IV estimates with our results for OLS estimates. Figure 3.7 plots these estimates together with the OLS benchmark \(\hat{\beta}\) and reveals two important results.

First, life-cycle effects from left-side measurement error are substantially larger in IV than in OLS estimates.\(^{40}\) Adjustment by \(\lambda_t\) would thus improve IV estimates only modestly. Usage of education as an instrument aggravates the life-cycle bias since income profiles differ strongly with education — the correlations between parental education and measurement errors in sons’ and fathers’ income are thus relatively large (c.f. Section 3.3). Second, IV estimates are well below the benchmark also at \(t^*\) at age 33 (\(\hat{\beta}^\text{IV}_{t^*} = 0.183, \hat{\sigma}_{\hat{\beta}^\text{IV}_{t^*}} = 0.056\), whereas \(\hat{\beta} = 0.270, \hat{\sigma}_{\hat{\beta}} = 0.017\)).

We therefore conclude that absent life-cycle effects, IV estimates bound the true parameter from above. But since applications are typically based on current income, IV estimates do not bound \(\beta\) in practice. Given the large sensitivity of IV estimates to the age at which sons’ incomes are measured (ranging between 0.08 and 0.53 over ages 30-45), we argue that IV estimates need to be interpreted with caution. Comparisons of IV estimates across populations are not reliable if based on short spans of income data.

\(^{39}\)We impute years of education from data on level of educational attainment as recorded in the 1970 census, i.e. when the fathers were around 40 years old. Using level dummies yields similar results.

\(^{40}\)In contrast, life-cycle effects from right-side measurement error are not particularly strong in IV estimates (figure available from the authors upon request).
Figure 3.7 IV Estimates Compared with OLS and Benchmark

Note: Cohort 1955-57, left-side measurement error only (i.e., in sons’ income).

Unobserved Heterogeneity in Income Profiles

An alternative method to address life-cycle bias is to model life-cycle income processes across subgroups, instead of assuming a uniform growth rate in the population. Income growth over the life-cycle can be predicted based on a set of observable characteristics in a first step, as proposed by Vogel (2006) and Hertz (2007).

Distinguishing income growth rates across subgroups defined by education, as in Vogel (2006), can reduce measurement error in lifetime income that arises from idiosyncratic deviations from population average income growth. However, after accounting for differential income growth across educational groups, other determinants of income will lead to deviations from the mean income growth rate within any given group. Since such determinants might often be shared by members of the same family, measurement errors in income growth rates are still likely to be correlated within families. For example, if a father holds an occupation that typically leads to steeper than average income growth, then his son’s income growth might also be steeper because he is relatively more likely to enter the same occupation.41

Evidence for within-family correlation in the choice of profession or employer can be easily found, for example in the list of presidents of the United States. More comprehensive evidence on the intergenerational transmission of employers is given in Corak and Piraino (2010).
Estimates could be improved by considering additional individual characteristics for the estimation of growth rates in more specific subgroups, as in Hertz (2007). But the crucial problem remains that we will not be able to sufficiently project life-cycle trajectories of income, since individual growth rates are determined by both observable and unobservable characteristics that can correlate within families. Unexplained dispersion in income growth is large: for example, Jenkins (2009) finds substantial deviations of individual income trajectories from average trajectories of groups defined by education, sex, and birth cohort in British data. The remaining life-cycle bias caused by within-family correlation of the unexplained part of individual income growth rates should therefore not be expected to be negligible.

**Figure 3.8 Life-Cycle Patterns in Income Across Subgroups**

Note: The trajectories depict average growth in log income over the life cycle for sons born in 1955-57, separately for sons with fathers above and below median lifetime income.

Our data allow us to provide empirical evidence in support of this argument. We derive average growth in log income for various groups of sons by regressing current log income on a polynomial in age. Figure 3.8 depicts such income trajectories for four groups of sons defined by education (non-college/college) and their father’s lifetime income (below/above median). While income trajectories are simply shifted for the two groups without college education, the difference in income growth over the life cycle is substantial for the other two groups: college-educated sons of richer fathers have much stronger income growth than college-educated sons of poorer fathers. We thus find evidence for parent-related heterogeneity in income profiles even after
controlling for a range of observable characteristics (sex, cohort, age, country of birth and education). By additionally holding lifetime income constant we can further confirm why the generalized errors-in-variables model does not eliminate life-cycle bias: we find that college-educated sons of richer fathers have lower initial incomes and steeper income growth over the life-cycle than sons of poorer fathers also for a given level of sons’ lifetime incomes. The generalized model does not consider this type of heterogeneity, and thus underestimates the intergenerational elasticity when sons’ incomes are observed at younger ages.

*Income Dynamics and Stochastic Shocks*

These observations are of interest beyond the intergenerational mobility literature, in particular for the extensive literature on income dynamics and stochastic shocks. A lively debate is ongoing in this field on whether idiosyncratic differences in income profiles of otherwise observationally equivalent individuals are mainly due to deterministic heterogeneity or persistent stochastic shocks. These two sources are generally hard to distinguish from one another, and the debate often centers around the capability of certain statistical tests to detect patterns in the autocorrelation of residual income growth.

An intergenerational dimension provides a novel perspective on this debate, as has recently been argued by Mayer (2010). Systematic life-cycle patterns in income growth that relate to parental background cannot stem from unexpected stochastic shocks, but have to be caused by deterministic factors. Observation of long series of income for two generations allows us to provide simple evidence on the existence of a parent-related component of income growth, as in Figure 3.8. It seems implausible that such a deterministic component should be interpreted as persistent shocks that arrive unexpectedly to an individual. The pattern is thus not consistent with income processes that do not account for the possibility of unobserved heterogeneity in income profiles.

42 See Meghir and Pistaferri (2011) for a recent summary of contributions to this debate.

43 Guvenen (2009) shows that the main body of evidence against profile heterogeneity in the existing literature – that the autocorrelations of income changes are small and negative – can also be replicated by income processes with profile heterogeneity, suggesting that this evidence may have been misinterpreted.

44 Note that the existence of a parent-related component of income growth is problematic even if the individual is not aware of it, so that the information set of the individual and the econometrician coincide. While our measures of income shocks would then indeed represent unexpected innovations from the point of view of the individual, these shock measures would nevertheless correlate with unobserved determinants of income like ability.
We instead conclude that unobserved deterministic heterogeneity in income profiles should be expected to be important, at least among highly educated individuals.\textsuperscript{45}

The distinction between persistent stochastic shocks and unobserved heterogeneity is important, since these alternatives have fundamentally different economic implications. Furthermore, if measures of income shocks are easily confounded by unobserved determinants of income growth (like ability or parental background), as our results imply, then one should worry about a potential relation between these unobserved determinants and the outcome of interest that one aims to explain structurally. Measurement error in income shocks should thus be expected to have non-classical characteristics, and its impact can not be restricted to attenuation bias, as is often assumed in the literature. We think that recent contributions in the intergenerational mobility literature illustrate well how serious the impact of such non-classical measurement error can be.

### 3.5 Conclusions

Using snapshots of income over shorter periods in the estimation of intergenerational income elasticities causes a so-called life-cycle bias if the snapshots cannot mimic lifetime outcomes (Jenkins, 1987). We document that recent methodological advances, despite their significant conceptual contribution, fail to fully eliminate this bias in applications. The life-cycle bias that persists in our Swedish data even after application of the generalized errors-in-variables model is strongly negative when using annual income below age thirty and remains negative up until the early forties. Estimates substantially understate the true elasticity also when income is measured around the “preferred age” as predicted by the generalized model. Since the recent literature has relied on the predictions of this model, we may expect that the resulting estimates tend to (still) understate the intergenerational elasticity.

Comparisons of intergenerational elasticity estimates across countries, groups or cohorts may thus be of limited reliability if based on short-run income data.\textsuperscript{46} Still, some of the major conclusions from cross-country studies are

\textsuperscript{45}This finding is consistent with Guvenen (2009), who finds that heterogeneous income profiles explain a larger fraction of income inequality among college-educated individuals.

\textsuperscript{46}One might hope that the bias is of similar magnitude across populations, such that the validity of comparative studies is not affected. Cross-country comparisons would for example be reliable if both the dispersion and the intergenerational correlation in the shape of income profiles is of the same magnitude in each country. But since the intergenerational correlation in income levels varies across countries we suspect that it
not put into question. For example, the findings that income mobility is much lower than found by the early literature, and that mobility differs strongly across countries (e.g. being lower in the U.S. than in the Nordic countries and Canada), are robust to even sizable revisions in the underlying estimates. It might however be necessary to revisit those conclusions that are based on more marginal cross-country differences, such as Denmark vs. the other Nordic countries and the relative positions of Germany and Canada.

Studies on mobility trends are potentially more strongly affected since even moderate life-cycle biases are sufficient to mask gradual changes of mobility over time. It is thus noteworthy that we find similar levels of life-cycle bias in estimates from an extension of the GEiV model presented by Lee and Solon (2009), which has been applied in most of the recent research on mobility trends. Further, comparisons across subgroups of a population that rely on the GEiV model will be compromised when the age pattern in income profiles differs across groups, which may for example be the case when groups are classified by education, sex or immigration status. Finally, our findings are most consequential for studies based on instrumental variable estimators. We find IV estimates based on sparse income data to be highly sensitive to the exact age at which offspring income is measured. In particular they do not provide upper bounds. Recent OLS estimates of the intergenerational elasticity in the U.S. that are close to or in excess of 0.6. (e.g. Mazumder, 2005, and Gouskova et al., 2010) are thus not at odds with lower IV estimates from the earlier literature (Solon, 1992; Zimmerman, 1992).

While these results are mostly negative, our analysis does provide some guidance for applied research. We find evidence that incomes at later ages (e.g. age 40-50) provide a more reliable base for application of the GEiV model. The extension presented in Lee and Solon (2009) improves statistical efficiency, although it does not decrease the life-cycle bias further. The bias can instead be reduced by averaging over multiple income observations from midlife (if available) for both fathers and sons. Finally, the treatment of zero and missing income observations has important consequences. To derive a general measure of mobility one would like to include such observations, but doing so comes with the cost of increased vulnerability to life-cycle bias.

Further refinements of empirical practice with restricted use of income observations around a specific age can thus improve upon previous estimates, but will not eliminate life-cycle bias. Development of a more structured approach that aims to capitalize on all available income data seems desirable. Future research could in particular benefit from a more comprehensive exploitation of partially observed income growth patterns. Intergenerational mobility also differs in other dimensions of income profiles. Our finding that the life-cycle bias varies even across Swedish cohorts born in the same decade supports this conclusion.
estimates are often based on multiple income observations per individual, but researchers typically disregard the idiosyncratic income growth across these observations. Such partially observed growth patterns are determined by both observable and unobservable characteristics of the individual and hence contain more information on lifetime income than what current income levels and observable characteristics can provide.

Our results add to a general conclusion that can be drawn from the intergenerational mobility literature: addressing heterogeneity in income profiles is an important, difficult and recurrently underestimated task. The central problem is that idiosyncratic deviations from average income profiles correlate with a wide range of individual and family characteristics. We hope that our discussion also underlines the potential importance of this issue for other literatures that rely on measurement of long-run income or income dynamics. In particular we conclude that the widespread practice of measuring annual income at a certain age as a surrogate for unobserved lifetime income is prone to life-cycle bias, since the most appropriate age for measurement cannot be predicted and estimates can be quite sensitive to small age changes.
References


Appendix

A.1 Annual and Lifetime Values Over the Life Cycle

As in Haider and Solon (2006), suppose that log annual income of worker $i$ at age $t$ is given by

$$y_{it} = \eta_i + \gamma_i t$$

(3.7)

For simplicity assume infinite lifetimes and a constant real interest rate $r > \gamma_i$.

**Proposition.** (i) For all age $t$, the difference between log annual income $y_{it}$ and the log of the annuitized value of the present discounted value of lifetime income varies with respect to the individual’s income growth rate $\gamma_i$. (ii) For any given age $t$, the difference will be equal for at most two different realizations of $\gamma_i$.

**Proof.** The annuitized value of the present discounted value of lifetime income, denoted $B_i$, is

$$\sum_{s=0}^{\infty} \exp(\eta_i + \gamma_i s)(1 + r)^{-s} = \sum_{s=0}^{\infty} B_i(1 + r)^{-s} = \frac{1 + r}{r} B_i$$

Hence the log of the annuitized value equals

$$\log B_i = \log \left( \frac{r}{1 + r} \sum_{s=0}^{\infty} \exp(\eta_i + \gamma_i s)(1 + r)^{-s} \right)$$

$$= \log r + \eta_i - \log(r - \gamma_i)$$

The difference $D_{it}$ between log annual income $y_{it}$ and the log of the annuitized value of the present discounted value of lifetime income $\log B_i$ is thus

$$D_{it} = \gamma_i t - \log r + \log(r - \gamma_i)$$

Depending on $t$, $D_{it}$ decreases or increases in individuals’ income growth rates $\gamma_i$,

$$\frac{\partial D_{it}}{\partial \gamma_i} = t - \frac{1}{r - \gamma_i}$$

The second derivative with respect to $\gamma_i$ is negative,

$$\frac{\partial^2 D_{it}}{\partial^2 \gamma_i} = -(r - \gamma_i)^{-2} < 0$$

$D_{it}$ is therefore a strictly concave function of $\gamma_i$ conditional on $t$ given, and a specific value of $D_{it}$ can stem from at most two different values of $\gamma_i$. $\square$
A.2 Life-Cycle Bias: Right-Side Measurement Error

Assume that we wish to estimate the regression model (3.1), but that log lifetime income of fathers $y_{f,i}^*$ is approximated by $y_{f,it}$, log annual income at age $t$. Sons’ log lifetime income $y_{s,i}^*$ is observed. We express the linear projection of $y_{f,it}$ on $y_{f,i}^*$ as

$$y_{f,it} = \lambda_{f,t} y_{f,i}^* + u_{f,it}$$

The probability limit of the OLS estimator of a linear regression of $y_{s,i}^*$ on $y_{f,it}$ is then

$$\text{plim} \hat{\beta}_t = \frac{\text{Cov}(y_{f,it}, y_{s,i}^*)}{\text{Var}(y_{f,it})} = \theta_{f,t} \beta + \theta_{f,t} \frac{\text{Cov}(u_{f,it}, y_{s,i}^*)}{\lambda_{f,t} \text{Var}(y_{f,i}^*)}$$

where $\theta_t = \lambda_{f,t} \text{Var}(y_{f,i}^*) / \left( \lambda_{f,t}^2 \text{Var}(y_{f,i}^*) + \text{Var}(u_{f,it}) \right)$ is the slope coefficient in the reverse regression of $y_{f,i}^*$ on $y_{f,it}$. This “reliability ratio” reduces to the familiar attenuation bias if $y_{f,it}$ is measured at age $t^*$ such that $\lambda_{f,t} = 1$. The GEiV model is based on the assumption that $u_{f,it}$ is uncorrelated to $y_{s,i}^*$. It can thus account for the reliability ratio, but not for the remaining life-cycle bias that stems from correlation in the shape of income profiles within families.

A.3 Life-Cycle Bias: Left- and Right-Side Measurement Error

Assume that we wish to estimate the regression model (3.1), but that log lifetime incomes of fathers $y_{f,i}^*$ and sons $y_{s,i}^*$ are not observed and thus approximated by $y_{f,it}$ and $y_{s,it}$, log annual incomes at age $t$. We express the linear projection of $y_{f,it}$ on $y_{f,i}^*$ as

$$y_{f,it} = \lambda_{f,t} y_{f,i}^* + u_{f,it}$$

and the linear projection of $y_{s,it}$ on $y_{s,i}^*$ as

$$y_{s,it} = \lambda_{s,t} y_{s,i}^* + u_{s,it}$$

The probability limit of the OLS estimator of a linear regression of $y_{s,it}$ on $y_{f,it}$ is then

$$\text{plim} \hat{\beta}_t = \frac{\text{Cov}(y_{s,it}, y_{f,it})}{\text{Var}(y_{f,it})} = \frac{\beta \lambda_{s,t} \lambda_{f,t} \text{Var}(y_{f,i}^*) + \lambda_{f,t} \text{Cov}(u_{s,it}, y_{f,i}^*) + \lambda_{s,t} \text{Cov}(y_{s,i}^*, u_{f,it}) + \text{Cov}(u_{s,it}, u_{f,it})}{\lambda_{f,t}^2 \text{Var}(y_{f,i}^*) + \text{Var}(u_{f,it})}$$

$^{47}$Note that for notational simplicity we here do not distinguish the age subscripts for fathers and sons.
If incomes are measured at ages such that $\lambda_{s,t} = \lambda_{f,t} = 1$ the probability limit reduces to

$$\text{plim} \hat{\beta}_t = \frac{\beta \text{Var}(y_{f,i}^*) + \text{Cov}(u_{s,it}, y_{f,i}^*) + \text{Cov}(y_{s,i}^*, u_{f,it})}{\text{Var}(y_{f,i}^*) + \text{Var}(u_{f,it})}$$

an expression akin (except for the subscript $t$) to the general eq. (3.2).

A.4 Figures

**Figure A. 3.1** Bias Estimates for Different Age Spans (Cohort 1955-57)

Note: Left-side measurement error only (i.e., in sons’ income). The age span of observed incomes of sons (fathers) varies along the horizontal (vertical) dimension.
**Figure A. 3.2** Bias Estimates for Different Age Spans (Cohort 1958-60)

Note: Left-side measurement error only (i.e., in sons’ income). The age span of observed incomes of sons (fathers) varies along the horizontal (vertical) dimension.
Figure A. 3.3 Bias Estimates for Different Age Spans (Cohort 1952-54)

Note: Left-side measurement error only (i.e., in sons’ income). The age span of observed incomes of sons (fathers) varies along the horizontal (vertical) dimension.
4. Interpreting Trends in Intergenerational Income Mobility∗

Introduction

The level of inequality and its evolution over time is an important topic in the social sciences, but the current level of interest in both academic research and public debate is nevertheless remarkable. Two central dimensions of interest are the degree of cross-sectional income inequality and intergenerational mobility in income. The first captures inequality in a population, the latter to what degree this inequality is being transmitted across generations. Both have immediate and important welfare implications, but they also relate in fundamental ways to the functioning of political and economic systems. Mobility is for example seen to contribute to the stability of liberal democracies, by legitimating prevailing class and status inequalities and by reducing the potential for radicalization (see Erikson and Goldthorpe, 1992). We still know very little how policies and institutions affect either dimension of inequality, but it is clear that both are closely related: rising inequality not only makes (the lack of) mobility more consequential, it may affect the poor further if the underlying causes for rising cross-sectional inequality also reduce mobility.

Cross-sectional income inequality and skill differentials in earnings have been increasing substantially since the late 1970s in the US and other OECD countries, with many factors such as technological change, the role of unions, immigration and trade being discussed as potential causes.1 While we know

∗This is joint work with Jan Stuhler. Financial support from the Swedish Council of Working Life (FAS) and the German National Academic Foundation is gratefully acknowledged. We thank Anders Björklund and Markus Jäntti for advice and encouragement. We are further grateful for comments from Magne Mogstad and Stephen Jenkins, and seminar participants at University College London and SOFI, Stockholm University.

1Autor and Katz (1999) discuss trends in wage inequality across countries. Atkinson et al. (2011) find a substantial rise in top income shares in the US and various other countries.
much less about trends in intergenerational income mobility, we do observe that mobility differs substantially across countries and that these differences are negatively correlated with cross-sectional inequality. A central theme in the recent literature is thus if income inequality has not only increased but also become more persistent across generations. This question is much debated particularly in countries that recently experienced a strong rise in cross-sectional inequality, such as the US, where commentators argue that low mobility threatens social cohesion and the notion of “American exceptionalism”.

But not only is the empirical evidence on mobility trends inconclusive, the literature has also provided little guidance to how trends should be interpreted. Most of the existing theoretical work examines the relationship between causal transmission mechanisms and steady-state levels of intergenerational mobility (e.g., Conlisk, 1974; Solon, 2004). In this paper, we argue that an understanding of mobility trends requires a dynamic perspective. Transitions between steady states are of particular importance in intergenerational research since a single transmission step corresponds to a whole generation. Structural changes may thus generate long-lasting mobility trends even if transition towards the new steady state is completed within few generations. For example, an educational reform that leads to a new steady state after only three or four generations would affect mobility trends over approximately a full century. Transition paths between steady states are then important determinants of mobility trends, especially if we expect countries to experience multiple structural changes over the course of such long time frames. We thus contribute to the literature by examining these dynamic implications of standard models of intergenerational transmission.

Moreover, in contrast to the previous mobility literature we specify individual human capital to consist of multiple skill types, in accordance with the growing evidence showing that human capital – including informally produced skills – is best regarded as multidimensional (see, e.g., Carneiro and Heckman, 2003; Heckman et al. (2006)). A notion of multidimensional skills is especially motivated by our long-run perspective, which covers such dramatic changes as the shift from mainly physical labor within agriculture to manufacturing and white-collar jobs.

We show in our model that the level of intergenerational mobility depends not only on contemporaneous transmission mechanisms, but also on the distribution of income and skills in the parent generation – and thus on past mechanisms. This has a number of implications. First, changes in transmission mechanisms can generate long-lasting mobility trends. Current trends might thus not be entirely caused by current or recent structural changes, but also

---

2 See the cross-country surveys in Björklund and Jäntti (2009), and Blanden (2011).
by major events in the past. Second, differences in mobility across countries might reflect differences not only in current but also past transmission mechanisms.

A particularly important result from our model is that a fairly general class of changes in transmission mechanisms cause non-monotonic transitions between steady states. First, changes in the relative returns to productive characteristics, or in their relative heritability, tend to increase mobility initially, followed by a decreasing trend that lasts over multiple generations. Times of changes (e.g., as of industrial or technological change) thus tend to be times of high mobility, while mobility is likely to decrease when the economic environment stabilises. Second, changes from more plutocratic to more meritocratic societies – i.e., parental status becomes less and own skill more important – will be characterized by an initial increase in mobility, but also by a negative trend for subsequent generations. Even structural changes that are clearly mobility-enhancing in the long-run can thus cause negative trends.

Declining intergenerational mobility today might then not signal that current policies and institutions promote equality of opportunity less effectively, but might instead be an aftermath of major improvements in the past. Relating current policies and institutions to current mobility levels may therefore result in misleading conclusions about their long-run effects on mobility; the interpretation of mobility trends becomes a rather complex matter. But a dynamic analysis does not only reveal pitfalls that complicate the interpretation of trends, it may also expose complimentary evidence on related matters in the literature. For example, it may aid in the identification of causal mechanisms, as different types of structural shocks have different dynamic implications. And it may extend our comprehension of empirical findings, by indicating how the initial impact of a policy may compare to its long-run effect on mobility.

The rest of the paper is structured as follows. In the next section we discuss the related literature. We present our structural model of intergenerational transmission in Section 4.2. We derive current and steady-state levels of intergenerational mobility in terms of structural parameters and analyze the dynamic content of the model. We examine some typical extensions of the standard model in Section 4.3, thereby also addressing the robustness of our main results. Some further implications and conclusions are found in Section 4.4.

4.1 The Literature

While previous theoretical research has frequently covered the relationship between causal transmission mechanisms and steady-state mobility levels, there exists little work on transition paths between those steady states. In the stan-
standard simultaneous equation approach as developed by Conlisk (e.g., in Conlisk, 1974) only Jenkins (1982) and Atkinson and Jenkins (1984) consider a system that is not in steady state. While Atkinson and Jenkins show how failure of the steady-state assumption impedes identification of invariable parameters of the structural model, we instead consider how changes in structural parameters affect intergenerational mobility in subsequent generations. Solon (2004) notes that the interpretation of mobility trends would benefit from a theoretical perspective, and examines how structural changes (such as in the return to human capital and the progressivity of public investment) affect mobility in the first affected generation. Davies et al. (2005) compare mobility and cross-sectional inequality under private and public education in a simple model of human capital accumulation. They note that the transition path of mobility may help to distinguish between alternative causes of rising inequality. Deviations from steady-state mobility rates arise in their model from changes in cross-sectional inequality over time, while we instead explore the interaction of multiple transmission channels.

There has been much recent empirical work on mobility trends in the US and other countries. A vast literature is concerned with trends in occupational and class mobility, with the recent US-UK comparison by Long and Ferrie (forthcoming) being particularly noteworthy. They argue that in the nineteenth century occupational mobility was higher in the US than in the UK, but that a subsequent decline in mobility in the US had erased this difference by the 1950s. Their findings are partly in contrast with previous studies of mobility in the US, which document little change across this time period (e.g., Grusky, 1986; Hauser, 2010).

The emerging evidence on more recent trends in income mobility is also conflicting. A reason for this lack of consensus may be the substantial data requirements that studies of intergenerational income mobility face. Measuring income mobility ideally requires income data that fully span two generations, but often only sparse data are available or exploited. Lee and Solon (2009)


See Solon (1999) and Haider and Solon (2006) for a discussion of the early empirical literature and the currently preferred method to estimate mobility parameters based on incomplete income data. Nybom and Stuhler (2011) argue that the standard methods employed in the recent literature still suffer from substantial life-cycle bias. The bias can differ by cohort and may thus mask gradual changes of mobility over time, or generate a false impression of such trends.
aim to provide more reliable evidence for the US by making more efficient use of the available income data. They find no evidence of major changes in mobility across cohorts born in 1952-1975, but cannot due to imprecise estimates reject more gradual changes over time. Recent evidence on income mobility in the UK has been more consistent, showing a decreasing trend in mobility across corresponding cohorts (Nicoletti and Ermisch, 2007; Blanden et al., 2004). A central concern in many of these papers, policy-related outlets, and the public press is if mobility has declined in conjunction with the recent rise in income inequality, fueled by the observation that income mobility correlates negatively with inequality across countries. Many potential causal factors for observed mobility trends – such as educational expansion, rising returns to education, or changes in welfare policies – have been discussed in the literature (e.g., Levine and Mazumder, 2007, and further articles in the same issue). Common to all explanations is that they relate trends to recent events that may have directly affected the respective cohorts. We aim to illustrate why the key to understand current mobility trends might instead lie in the more distant past.

4.2 A Simple Model

Measuring intergenerational mobility. A popular descriptive measure of persistence in economic status is the intergenerational income elasticity. It is defined as the slope coefficient in a linear OLS regression of the statistical relationship

\[ y_{i,t} = \beta_t y_{i,t-1} + \epsilon_{i,t}. \]

(4.1)

We consider a simplified one-parent one-offspring family structure, with \( y_{i,t} \) as log lifetime income of the offspring in generation \( t \) of family \( i \) and \( y_{i,t-1} \) as log lifetime income of the parent. The error \( \epsilon_{i,t} \) is uncorrelated with the regressor by construction. Under stationarity of \( y_{i,t} \) the elasticity \( \beta_t \) equals the intergenerational correlation parameter, which adjusts the elasticity for changes in cross-sectional inequality. The elasticity captures to what degree percentage differences in parents’ incomes tend to be transmitted to the next generation. A low elasticity or correlation thus indicates high mobility.

\(^5\)Exemplary for the US are Bernstein (2003); Wooldridge (2005); Wessel (2005); Scott and Leonhardt (2005); Acs and Zimmerman (2008), and Noah (2012). The political importance of the topic is exemplified by a recent speech of Alan Krueger, Chairman of Council of Economic Advisers, who warned that intergenerational mobility should be expected to decline further as of the recent rise in income inequality in the US (speech delivered at the Center for American Progress, January 12th, 2012).
A model of intergenerational transmission. We consider a stochastic linear difference-equation model of intergenerational transmission in the tradition of the simultaneous equation approach developed by Conlisk (see Conlisk, 1969, 1974) and also considered in, for example, Atkinson and Jenkins (1984). Although we will not explicitly address optimizing behavior as in Becker and Tomes (1979) or Solon (2004), the mechanistic pathways represented by the structural equations can be derived from an underlying utility-maximization framework (see Goldberger, 1989).

We assume a simple causal model of intergenerational transmission that captures the core mechanisms of the models considered in the previous literature,

\[ y_{i,t} = \gamma y_{i,t-1} + \rho'_t e_{i,t} + u_{i,t} \]  
\[ e_{i,t} = \Lambda e_{i,t-1} + v_{i,t}. \]

The parameter \( \gamma \) captures the effect of parental status (as measured by their income) on the income of offspring that is independent from their actual productivity. It may arise as of nepotism, credit constraints influencing choices on the labor market, parental information and networks, statistical discrimination under imperfect information on individual productivity, or (if total market income is considered) returns to bequests. The exact mechanism and the distinction between earnings and income are not central for our purposes. We denote the human capital of the offspring of family \( i \) in generation \( t \) by \( e_{i,t} \), a \( J \times 1 \) vector with elements \( e_{1,t}, e_{2,t}, \ldots, e_{J,t} \) that reflect distinct productive characteristics such as health, physical attributes, cognitive and non-cognitive abilities. These characteristics are valued on the labor market according to a \( J \times 1 \) price vector \( \rho_t \) with elements \( \rho_{1,t}, \rho_{2,t}, \ldots, \rho_{J,t} \). The random shock term \( u_{i,t} \) captures factors that do not relate to parental background. For our analysis, it makes no difference if these are interpreted as (labor market) luck or as the impact of other characteristics that are not transmitted within families. Income is thus assumed to be determined by parental income, individual characteristics, and chance. In our baseline model we assume that other family members influence offspring only indirectly. We consider independent effects from grandparents instead separately in Section 4.3.

The elements of the \( J \times J \) matrix \( \Lambda_t \) govern to what degree parental characteristics determine the characteristics of the offspring generation. We consider \( \Lambda_t \) to represent a broad concept of heredity potentially working through both nature and nurture. For simplicity we assume no cross-correlations between characteristics, so that \( \Lambda_t \) is diagonal with elements \( \lambda_{1,t}, \lambda_{2,t}, \ldots, \lambda_{J,t} \). Both \( u_{i,t} \) and the elements of the \( J \times 1 \) shock vector \( v_{i,t} \) are assumed to be uncorrelated with each other and to past values of \( \{y_{i,t}, e_{i,t}, u_{i,t}, v_{i,t}\} \). For some of our examples it will suffice to consider a single characteristic \( e_{i,t} \) and thus scalar versions
of equations (4.2) and (4.3).

To keep the model simple we capture both the direct and indirect (via offspring human capital) impact of parental income in only one parameter, even though one may of course argue that parental income should also be included in equation (4.3). Distinctions between direct and indirect income effects may not be sharp in practice. For example, parental credit constraints might affect educational attainment and skill acquisition of offspring, but might also affect their career choices for a given set of skills. The consideration of more evolved models would however complicate the analysis while making little difference for our key findings. The general implications that we highlight stem from the interaction of different transmission mechanisms, and thus extend from simple to more complex transmission frameworks (as we will exemplify in Section 4.3). In principle, they extend also to mobility in other outcomes (such as education, occupation, or class), but are most interesting in the context of income mobility as it contains an additional layer of complexity from the role of prices or returns to human capital.

For convenience we drop the individual subscript \(i\) in the subsequent analysis and make a few simplifying assumptions. Assume that \(y_t\) and all \(e_{jt}\) are measured as trendless indices with constant mean zero (as in Conlisk, 1974) such that we do not need to include constants. To avoid the need for case distinctions, assume further that the indices measure positive characteristics (so that the elements of \(\rho_t\) are non-negative); that parent and offspring characteristics are not negatively correlated (so that the elements of \(\Lambda_t\) are non-negative); and that all slope parameters are between zero (if not noted to the contrary) and one (for stability) for all \(t\). The reduced form of equations (4.2) and (4.3) is

\[
\begin{pmatrix}
  y_t \\
  e_t 
\end{pmatrix}
= 
\begin{pmatrix}
  \gamma_t & \rho_t' \Lambda_t \\
  0_{j \times 1} & \Lambda_t 
\end{pmatrix}
\begin{pmatrix}
  y_{t-1} \\
  e_{t-1} 
\end{pmatrix}
+ 
\begin{pmatrix}
  u_t + \rho_t' v_t \\
  v_t 
\end{pmatrix},
\]

which we may shorten to

\[
x_t = A_t x_{t-1} + w_t. \tag{4.4}
\]

The stability condition \(\lim_{s \to \infty} A_t^s = 0\) is satisfied as the parameter constraints assure that all eigenvalues of \(A_t\) are non-negative and below one, which also ensures that the transitions of the first and second moments of \(x_t\) towards their steady state values are monotonic (Jenkins, 1982). This property does however not extend to the transition path of the intergenerational elasticity, as we will discuss in the next section.

Finally, since our focus is on intergenerational transmission we will initially only consider changes in the relative strength of transmission mechanisms by holding the cross-sectional inequality in income and characteristics
(as measured by their variances) constant. We further normalize the variance of all variables to one by choosing appropriate values for the variances of \( u_t \) and elements of \( \mathbf{v}_t \). These restrictions simplify the discussion and lead to more illustrative derivations by abstracting from additional sources of dynamics that stem from the transition path of cross-sectional inequality. We discuss implications from the latter instead in Section 4.3. While fewer properties of the transition path of the intergenerational elasticity can be generalized under time-varying cross-sectional inequality, our main qualitative results still hold.

In contrast to previous models cited above, we assume that income depends on human capital through a vector of distinct productive characteristics instead of only a single factor. This generalization will prove to be central for some of our findings. Our second deviation from previous work is simply the addition of explicit \( t \) subscripts on all parameters, since we want to consider the effects of changes in the transmission framework over time. Similarity to the existing literature is advantageous for our purposes since it should support our claim that our findings are general and do not arise due to non-standard modelling choices.

The Importance of Past Transmission Mechanisms

The intergenerational income elasticity, which coincides with the intergenerational correlation by virtue of the assumption of constant cross-sectional inequality, is derived by plugging equations (4.2) and (4.3) from our model of intergenerational transmission into equation (4.1):

\[
\beta_t = \frac{\text{Cov}(y_t, y_{t-1})}{\text{Var}(y_t)} = \gamma + \rho_t \lambda_t \text{Cov}(e_{t-1}, y_{t-1}). \tag{4.5}
\]

Thus, \( \beta_t \) depends on current transmission mechanisms (parameters \( \gamma_t, \rho_t \) and \( \lambda_t \)) and on the cross-covariance between income and productive characteristics in the parent generation. Expression (4.5) illustrates that two populations that

\[\text{Taking the covariance of (4.4) and denoting the covariance matrices of } x_t \text{ and } w_t \text{ by } S_t \text{ and } W_t, \text{ respectively, gives}
\]

\[S_t = \lambda_t S_{t-1} \lambda_t' + W_t.\]

Expansion yields that \( \text{Var}(e_{j,t}) = 1 \ \forall j, t \ \text{iff} \ \text{Var}(e_{j,0}) = 1 \ \forall j \ \text{and} \ \text{Var}(w_t) = I_{j,j} - \lambda_t \lambda_t'. \) Likewise, \( \text{Var}(y_t) = 1 \ \forall j, t \ \text{iff} \ \text{Var}(y_0) = 1 \ \text{and} \ \text{Var}(u_t) = 1 - \gamma_t^2 - 2 \gamma \text{Cov}(y_t, e_{t-1}) \lambda_t \rho_t - \rho_t' \rho_t. \) The requirement for these variances to be non-negative is satisfied iff \( \lambda_{j,t} \leq 1, \rho_t' \rho_t \leq 1, \) and \( \gamma_t \leq 1/2(-2 \text{Cov}(y_t, e_{t-1}) \lambda_t \rho_t + ((2 \text{Cov}(y_t, e_{t-1}) \lambda_t \rho_t)^2 + 4(1 - \rho_t' \rho_t))^{1/2}) \) for all \( j, t. \) To avoid case distinctions, we assume that those inequalities hold strictly.
are subject to similar transmission mechanisms (e.g., exposed to similar institutions and policies) can nevertheless differ in their levels of intergenerational mobility, since current mobility depends on the distribution of productive characteristics in the parent generation.

The cross-covariance between income and productive characteristics in the parent generation is in turn determined by past transmission mechanisms, and thus depends on past values of \( \{ \gamma_t, \rho_t, \Lambda_t \} \). We can iterate backwards to express \( \beta_t \) in terms of parameter values,

\[
\beta_t = \gamma_t + \rho_t \Lambda_t \text{Cov}(e_{t-1}, y_{t-1}) \\
= \gamma_t + \rho_t \Lambda_t [\Lambda_{t-1} \text{Cov}(e_{t-2}, y_{t-2}) \gamma_{t-1} + \rho_{t-1}] \\
= \ldots \\
= \gamma_t + \rho_t \Lambda_t \rho_{t-1} + \rho_t \Lambda_t \left( \sum_{r=1}^{\infty} \left( \prod_{s=1}^{r} \gamma_{t-s} \Lambda_{t-s} \right) \rho_{t-r-1} \right),
\]

where for simplicity we assume that the process is infinite.\(^7\) The level of intergenerational mobility today thus depends on current and past transmission mechanisms. If parameters have been constant for past generations, \( \{ \gamma_s = \gamma, \rho_s = \rho, \Lambda_s = \Lambda \}_{s \leq t} \), then equation (4.7) simplifies to the steady-state intergenerational elasticity

\[
\beta = \gamma + \rho \Lambda \sum_{s=0}^{\infty} (\gamma \Lambda)^s \rho \\
= \gamma + \rho \Lambda (I_{JxJ} - \gamma \Lambda)^{-1} \rho,
\]

where the second line follows since the geometric series \( \sum_{s=0}^{\infty} (\gamma \Lambda)^s \) converges (the absolute value of each eigenvalue of \( \gamma \Lambda \) is below one). The literature has almost exclusively focused on how changes in structural parameters affect intergenerational mobility in steady state given by (4.7). We will instead analyze the path of transition towards the new steady state as determined by equation (4.6). Some of its properties can be readily generalized. The transition path of \( \text{Cov}(e_{t-1}, y_{t-1}) \) is governed by the eigenvalues of the reduced form coefficient matrix in equation (4.4), and is thus monotonic. From (4.6) it however follows that income mobility in the \emph{first} generation subject to a structural change is directly affected by changes in the parameter values, not indirectly by changes in the covariance between parental income and characteristics. Trends in income mobility are thus not necessarily monotonic, as we will show in the next section. Other properties, such as the speed of convergence, depend on the parameterization of the model and can thus not be generalized.

\(^7\)For a finite process, \( \beta_t \) will depend on past parameter values and the initial condition \( \text{Cov}(e_{t,0}, y_{t,0}) \).
From Simple Examples to Non-Monotonic Trends

For our first examples it will be sufficient to consider a single skill, \( e_t \), and thus scalar versions of equations (4.2) and (4.3),

\[
y_t = \gamma y_{t-1} + \rho_t e_t + u_t
\]

and

\[
e_t = \lambda_t e_{t-1} + v_t.
\]

Our main findings do not rely on specific parameter choices, but the quantitative implications of our numerical examples will be more relevant if we choose values that are consistent with empirical evidence. Unfortunately, intergenerational data that would allow for empirical estimation of our model (i.e., spanning many generations) are not available. Nevertheless, rough orders of magnitude of the parameters discussed above can be inferred from the empirical literature. The evidence in the literature, and our cross-validations within the model, suggest the following parameter values for the US case:

\[
0.45 \leq \beta \leq 0.55, \quad 0.15 \leq \gamma \leq 0.25, \quad 0.60 \leq \rho \leq 0.70, \quad 0.50 \leq \lambda \leq 0.65
\]

We provide a detailed motivation of our choices of parameter values in Appendix A.1. It will however be useful to first look at an even simpler case in which parental income has no effect on offspring income.

**EXAMPLE 1: A SIMPLE MERITOCRATIC ECONOMY.** Assume that parental income has no direct effect on child income (\( \gamma = 0 \forall t \)). Changes in the heredity of characteristics then shift mobility within one generation. In contrast, changes in the returns to skills affect mobility over the course of two generations.

From equation (4.6) it follows that a change in the heredity of characteristics from \( \lambda_{t< T} = \lambda_1 \) to \( \lambda_{t \geq T} = \lambda_2 \) leads to a one-time shift in \( \beta_t \) equal to

\[
\Delta \beta_{T} = \beta_{T} - \beta_{T-1} = \rho (\lambda_2 - \lambda_1) \rho.
\]

Mobility remains constant afterwards. A change in returns from \( \rho_1 \) to \( \rho_2 \) at time \( T \) instead leads to a transition in \( \beta_t \) lasting over two generations. The first shift is equal to

\[
\Delta \beta_{T} = \beta_{T} - \beta_{T-1} = (\rho_2 - \rho_1) \lambda \text{Cov}(e_{T-1}, y_{T-1})
\]

\[
= (\rho_2 - \rho_1) \lambda \rho_1,
\]
and is induced by the change in returns for the offspring generation in $T$. The second shift equals

$$\Delta \beta_{T+1} = \beta_{T+1} - \beta_T = \rho_2 \lambda \{(\text{Cov}(e_T, y_T) - \text{Cov}(e_{T-1}, y_{T-1}))\} = \rho_2 \lambda (\rho_2 - \rho_1),$$

and is induced by the change in the correlation between income and skills among the parents of the offspring generation $T+1$, in turn caused by changing returns to those skills in the previous period. If returns increase ($\rho_2 - \rho_1 > 0$) then the second shift will be larger than the first.\(^8\) Mobility remains constant afterwards. Figure 4.1 gives a numerical example.

**Figure 4.1** A Change in the Heredity of, or Returns to, Skills

Hence, even in this simple example we find that the effect on the intergenerational income elasticity from changes in the returns to characteristics cannot be fully evaluated until after both the parent and child generations have experienced the new price regime. This result extends to the variable cross-sectional

\(^8\)Co-movements in the cross-sectional variance of income (see section 4.3) can affect the relative size of the first and second elasticity shifts further. Allowing for variable cross-sectional inequality and for a direct effect of parental income ($\gamma \neq 0$) does not affect the implication that current mobility levels do not fully reflect a change in prices before at least two generations.
inequality case, which we consider in Section 4.3 and the Appendix. Relating our example to the evidence on rising skill differentials in wages from the late 1970s in the US and UK (and more recently in other OECD countries) might serve to illustrate its practical implications. Many authors have argued that widening wage differentials could decrease intergenerational mobility (see, for example, Blanden et al., 2004; Solon, 2004; and Aaronson and Mazumder, 2008), a hypothesis that is also one of the main motivations for the recent interest in mobility trends. But recent trend studies do not yet observe offspring cohorts whose parents have fully experienced the changing wage regime. The impact of rising wage differentials on mobility levels may thus become fully evident only in future empirical studies that observe more recent cohorts.

The example further illustrates how the dynamic response of the elasticity between steady states can be informative on the type of structural shock that occurred. Changes in the heritability of skills will have a more immediate effect than changes in the returns to those skills, since income mobility is potentially affected by both the returns to parent and offspring skills.

We proceed with two simple examples on “equalizing opportunities”, in which the outcome of offspring becomes less dependent upon parental income.

**Example 2: Equalizing Opportunities (Type I).** Assume that the relative importance of parental income diminishes and that the importance of factors that do not relate to parental background increases. Mobility increases monotonically in subsequent generations, at a decreasing rate.

Assume that the direct effect of parental income declines from \( \gamma_{<T} = \gamma_1 \) to \( \gamma_{\geq T} = \gamma_2 \), such that \( \gamma_1 > \gamma_2 \). From (4.6) income mobility of subsequent

\[9\] While under variable cross-sectional inequality a change in prices also affects the elasticity over the course of two generations, the size of the second shift may not be larger than the first shift since it depends on the transition path of cross-sectional inequality. The second shift however still exceeds the first shift if the intergenerational correlation instead of elasticity is considered.

\[10\] For example, the youngest offspring cohort observed in Lee and Solon (2009) was born in 1975. The early careers of many of their parents will not have been subject to the widening skill differential.

\[11\] Conlisk (1974) notes that “opportunity equalization” is an ambiguous term that may relate to different types of structural changes in intergenerational transmission models.

\[12\] Constant cross-sectional inequality then requires \( \sigma_{u,t<T} < \sigma_{u,t\geq T} \).
generations changes according to

\[ \Delta \beta_T = \gamma_2 - \gamma_1 \]
\[ \Delta \beta_{T+1} = (\gamma_2 - \gamma_1)\rho \lambda^2 \text{Cov}(e_{T-1}, y_{T-1}) \]

Mobility will thus be increasing monotonically in subsequent generations, at a decreasing rate. Figure 4.2 shows a numerical example. Mobility adapts over the course of multiple generations and not instantly since the current correlation between productive characteristics and income depends on the direct impact of income in the past. A decline in the strength of the direct income mechanism diminishes that correlation in subsequent generations, thereby increasing the intergenerational income mobility.

**Figure 4.2** A Decline in the Importance of Parental Income

The example illustrates that changes in structural mechanisms should be expected to affect mobility trends over multiple generations. Mobility trends today are thus not necessarily indicative of a changing effectiveness of current policies and institutions in the promotion of equality of opportunity, but could be the lagged effect of major institutional changes in the past.

From equation (4.6) it follows that other types of structural changes have similar dynamic implications. Mobility will increase monotonically if the re-
turns to partially inherited characteristics decrease (a fall in $\rho$) or if the heredity of these characteristics decreases (a fall in $\lambda$). Convergence occurs through two channels. A lower degree of heredity, for example, decreases the intergenerational correlation of characteristics and thus the intergenerational persistence in income. The direct effect of parental income ($\gamma$) then becomes less detrimental to income mobility since the offspring from rich families are less likely to inherit productive characteristics. The correlation between income and characteristics within a given generation declines. This latter indirect effect follows a transition path that can last over multiple generations, thus causing the convergence to the new steady state.

Mobility levels converged fast to the new steady state in this example, with the initial drop in $\beta_t$ in the first generation dominating the subsequent trend. We will see in the next example that subsequent shifts may however be quite sizable for other types of structural change.

**Example 3: Equalizing Opportunities (Type II).** Assume that the importance of parental status diminishes and that productive characteristics that are partially inherited within families are instead more strongly rewarded. Mobility responds non-monotonically: it increases in the first generation, but then decreases (at a decreasing rate) in subsequent generations.

In other words, assume that the economy becomes less plutocratic and more meritocratic, which in our model corresponds to the assumption that $\gamma_1 > \gamma_2$ and $\rho_1 < \rho_2$.

From (4.5) it follows that

\[
\Delta \beta_T = (\gamma_2 - \gamma_1) + (\rho_2 - \rho_1)\lambda \text{Cov}(e_{T-1}, y_{T-1})
\]

\[
\Delta \beta_{T+1} = \rho_2 \lambda (\rho_2 - \rho_1) + \rho_2 \lambda^2 (\gamma_2 - \gamma_1) \text{Cov}(e_{T-1}, y_{T-1})
\]

\[
\vdots
\]

\[
\beta_\infty - \beta_{T-1} = (\gamma_2 - \gamma_1) + \frac{\rho_2^2 \lambda}{1 - \gamma_2 \lambda} - \frac{\rho_1^2 \lambda}{1 - \gamma_1 \lambda}.
\]

The first line illustrates that a change to a more meritocratic society tends to increase mobility initially, specifically iff

\[
\frac{\gamma_1 - \gamma_2}{\rho_2 - \rho_1} > \lambda \text{Cov}(e_{T-1}, y_{T-1}).
\]  \hspace{1cm} (4.10)

However, mobility then decreases in subsequent generations iff

\[
\frac{\rho_2 - \rho_1}{\gamma_1 - \gamma_2} > \lambda \text{Cov}(e_{T-1}, y_{T-1}).
\]  \hspace{1cm} (4.11)
Conditions (4.10) and (4.11) will be satisfied for any changes $\gamma_1 - \gamma_2$ and $\rho_2 - \rho_1$ that are sufficiently symmetric, i.e., relatively equal in absolute terms. Intuitively, when the economy becomes more meritocratic, highly productive individuals become more likely to rise to the top of the income distribution. This process initially increases mobility since these individuals are not necessarily from previously rich families. Individuals from highly productive families will subsequently however tend to keep their higher ranks in the income distribution, thus decreasing mobility. Conditions (4.10) and (4.11) are more likely to be satisfied when heritability ($\lambda$) is low since high-productivity individuals are then less likely to come from previously rich families.

**Figure 4.3** A Decline in the Importance of Parental Income and Increasing Returns to Skills

![Graph showing mobility trends](image)

Note: Numerical example with $\lambda = 0.6$ and a decline in $\gamma$ from $\gamma_1 = 0.4$ to $\gamma_2 = 0.2$ as well as a rise in $\rho$ from $\rho_1 = 0.5$ to $\rho_2 = 0.7$ at generation $T$.

Figure 4.3 illustrates that the response in mobility trends can be nonmonotonic and long-lasting. A decline in the importance of parental status increases mobility initially, but mobility declines in subsequent generations. Mobility trends become insignificant only in the third generation, which may be more than half a century after the structural change. These are important implications for the interpretation of mobility trends. First, structural changes that are mobility-enhancing in the long-run may nevertheless cause decreasing mobility trends. Second, these negative mobility trends can last over many decades.
In the numerical example, mobility changes much more strongly in the first and second generation than in the subsequent ones. One might thus expect that the lagged impact of more distant structural changes plays only a negligible role for current mobility trends. That depends however on the relative magnitude of those past changes. For example, in the late 19th and early 20th century the US experienced rapid industrialization, massive immigration, internal migration, and urbanization, a decline in the share of agriculture and self employment, and an expansion of public secondary schooling that affected a large part of the population. The country participated in two world wars and went through a highly turbulent interwar period. These events may have affected intergenerational transmission to a greater degree than more recent changes that have been considered as potential determinants of current mobility trends (such as an increase in private schooling or the increase in higher education).\textsuperscript{13} The recent empirical literature measures trends in income mobility for offspring cohorts born from around 1960, cohorts that are separated only by a few generations from the events of the early 20th century. We thus suspect that mobility trends observed over these cohorts may not only reflect contemporaneous changes in policies or institutions, but also (and maybe primarily) the lagged impact of major changes in past generations.

\textbf{Intergenerational Mobility in Times of Change}

Our finding that a change from plutocracy to meritocracy can lead to long-lasting and non-monotonic mobility trends could be particularly relevant for the interpretation of recent mobility trends. It relates however to a rather specific example of structural change, and one may thus expect that non-monotonic adjustments are more of an exception than a rule. With our next examples, we aim to show that they are instead quite typical, as they tend to occur whenever the relative heritability of characteristics or their relative returns change.

Note that these examples are based on a transmission framework with multiple characteristics, as in equations (4.2) and (4.3). The notion of individual ability has recently shifted from a one-dimensional concept primarily related

\textsuperscript{13} The hypothesis that past structural changes may cause lagged effects is largely consistent with the sociological literature on long-lasting trends in intergenerational occupational mobility (an alternative measure of economic mobility, see Hauser 2010 for a discussion), which implies that mobility has increased substantially between the late 19th and early 20th century and declined thereafter in the US (see Grusky, 1986; and Long and Ferrie, forthcoming). The hypothesis that industrialization tends to increase mobility is also consistent with evidence on mobility trends in Finland, where mobility increased strongly for cohorts born around 1950 compared to cohorts born in the early 1930s (see Pekkala and Lucas, 2007).
to IQ, such as in the single-skill signaling model (Arrow, 1973) and the g factor (see, e.g., Herrnstein and Murray, 1994), to a multidimensional set of traits that for example recognizes the importance of noncognitive skills. A stream of evidence has supported this idea, showing that several distinct types of skills are important for various labor market and social outcomes (e.g., Heckman et al. (2006); Lindqvist and Vestman, 2011). Although typically not discussed in the intergenerational context (an exception is Bowles and Gintis, 2002), our analysis illustrates that such multiplicity may provide additional implications that cannot be captured by models that are based on a single inheritable characteristic.

**Example 4: Changing Labor Market Conditions.** Assume that the returns to individual characteristics change on the labor market ($\rho_1 \neq \rho_2$). Mobility tends to respond non-monotonically, increasing in the first generation, but decreasing (at a decreasing rate) in subsequent generations.

Changes in the returns to characteristics could stem from changes in the demand for skills (e.g., as of industrial or technological change) or changes in relative supply (e.g., as of immigration or changes in the production function of these skills). A specific example is the decrease in the demand for physical ability (strength, endurance, etc) relative to cognitive ability as a labor market moves from agricultural to white-collar jobs. It has been documented in the literature that returns to characteristics can indeed change rapidly even in periods that are much shorter than the time scale underlying our intergenerational analysis.\(^{14}\)

To grasp the intuition consider first a simple symmetric case in which two characteristics $k$ and $l$ are equally transmitted within families ($\lambda_k = \lambda_l = \lambda$), but their prices on the labor market are swapping at time $T$ ($\rho_{2,k} = \rho_{1,l} \neq \rho_{1,k} = \rho_{2,l}$). From (4.5) it follows that

$$\Delta \beta_T = (\rho_{k,2} - \rho_{k,1})\lambda \text{Cov}(e_{k,T-1}, y_{T-1}) - (\rho_{l,1} - \rho_{l,2})\lambda \text{Cov}(e_{l,T-1}, y_{T-1})$$

$$= -\frac{(\rho_{k,2} - \rho_{k,1})^2 \lambda}{1 - \gamma \lambda} < 0$$

and

$$\Delta \beta_{T+1} = \rho_{k,2}\lambda (\rho_{k,2} - \rho_{k,1}) - \rho_{l,2}\lambda (\rho_{l,1} - \rho_{l,2})$$

$$= (\rho_{k,2} - \rho_{k,1})^2 \lambda > 0$$

\(^{14}\)A typical example is the job-polarization literature which highlights how the IT revolution has implied a dramatic shift in demand from substitutable manual skills to complementary abstract skills (e.g., Autor et al., 2003).
We provide a numerical example in Figure (4.4).

**Figure 4.4** A Swap in Prices

Note: Numerical example with $\gamma = 0.2$, $\lambda = 0.6$ and a change in prices from $\rho_{k,1} = 0.3$ to $\rho_{k,2} = 0.6$ and $\rho_{l,1} = 0.6$ to $\rho_{l,2} = 0.3$ at generation $T$.

Intuitively, those characteristics that have been more strongly rewarded in past generations are also more strongly correlated with parental income. As a consequence, mobility tends to initially increase if prices change, since characteristics for which prices increase from low levels are less prevalent among the rich than characteristics for which prices decrease from high levels. In subsequent generations, the characteristics for which prices increased become increasingly correlated with parental income, leading to decreasing mobility trends.

We can derive that such v-shaped responses in mobility are typical for the general case in which the prices of any number of characteristics change. From (4.5) and (4.7) we have

$$2\beta_T = \rho_1'\Lambda(I-\gamma\Lambda)^{-1}\rho_2 + \rho_1'\Lambda(I-\gamma\Lambda)^{-1}\rho_2$$

$$= \rho_1'\Lambda(I-\gamma\Lambda)^{-1}\rho_2 + \rho_1'\lambda(I-\gamma\Lambda)^{-1}\rho_2 - (\rho_2' - \rho_1')\Lambda(I-\gamma\Lambda)^{-1}(\rho_2 - \rho_1).$$

$$= \beta_{T-1} = \beta_\infty$$

(4.12)

The quadratic form in the last term is greater than zero (for $\rho_2 \neq \rho_1$) since $\Lambda(I-\gamma\Lambda)^{-1}$ is positive-semidefinite, indicating that price changes tend to in-
crease intergenerational mobility initially. From (4.12) it follows that $\beta_T$ is below both the previous steady state $\beta_{T-1}$ and the new steady state $\beta_\infty$ if the shift in the steady-state elasticity ($\Delta \beta = \beta_\infty - \beta_{T-1}$) is not too large, specifically if

$$|\Delta \beta| < (\rho'_2 - \rho'_1)\Lambda (I - \gamma \Lambda)^{-1}(\rho_2 - \rho_1),$$

or, plugging in for steady-state values,

$$\rho'_2\Lambda (I - \gamma \Lambda)^{-1}(\rho_2 - \rho_1) > 0 > \rho'_1\Lambda (I - \gamma \Lambda)^{-1}(\rho_2 - \rho_1).$$

(4.13)

Any symmetric changes (as in the numerical example) fulfill these conditions and will thus lead to non-monotonic adjustments as in Figure 4.3. More generally, changes in the returns to individual characteristics that do not affect long-run mobility much (e.g., some prices go down while others go up) will increase mobility in the short-run but cause a decreasing trend in mobility in subsequent generations. We believe that these are new results. While some authors have shown that specific events can lead to non-monotonic mobility trends through repeated changes in structural parameters (such as returns to characteristics), we find that even a one-time reversal in returns may generate a non-monotonic trend.

**Example 5: Changes in the Heritability of Characteristics.**

Assume that the heritability of multiple characteristics change ($\Lambda_1 \neq \Lambda_2$). Mobility responds non-monotonically: it increases in the first generation, but then decreases (at a decreasing rate) in subsequent generations.

For example, changes in the school system may affect the heritability of formal education, and reforms of the health care system may affect intergenerational transmission of health. As with changes in prices, when some characteristics become more and some less transmitted within families, mobility tends to increase initially but then follows a negative trend in subsequent generations. The intuition is similar as well.\textsuperscript{16}

\textsuperscript{15}For example, Galor and Tsiddon (1997) consider how the life-cycle of technological progress might lead to repeated changes in the relative returns to ability and parent-related human capital, and thus to non-monotonic trends in cross-sectional inequality and intergenerational mobility over time.

\textsuperscript{16}Characteristics that are more strongly inherited are also more strongly correlated with parental income. Characteristics for which heredity increase from low levels are thus less prevalent among the rich than characteristics for which heredity decrease from high levels, and changes in the heredity of characteristics will tend to increase mobility initially. The characteristics for which heredity increased then become increasingly correlated with parental income in subsequent generations, leading to a decreasing mobility trend.
From (4.7) it follows that the mobility trend will follow such v-shape 
$$(\Delta \beta_T < 0 < \Delta \beta_{T+1})$$ iff

$$\rho(\Lambda_2 - \Lambda_1) (I - \gamma \Lambda_1)^{-1} \rho < 0 < \rho'(\Lambda_2 - \Lambda_1) (I - \gamma \Lambda_1)^{-1} \rho$$

(4.14)

Again, any symmetric parameter changes, as considered in the previous numerical example, satisfy this condition.

These last two examples have quite general implications that do not depend much on the nature of changes in transmission mechanisms. Relative changes in the returns to or heritability of characteristics will tend to raise intergenerational mobility in the short run. Times of changes thus tend to be times of high mobility. Second, such mobility gains will be succeeded by longer-lasting negative trends in intergenerational income mobility if no further structural changes occur. Countries experiencing a period of stable economic conditions will thus tend to be characterized by negative mobility trends if they were preceded by more turbulent times.

4.3 Extensions and Sensitivity Analysis

Our model is broadly in line with the previous literature, but some of its simplifying assumptions deserve further discussion. To match the empirical literature on mobility trends, we introduce a cohort dimension into our model. We then discuss the sensitivity of our results to the way we model the influence of parental income. In the third subsection, we relax the assumption of constant cross-sectional inequality and consider how the transition path of income dispersion adds an additional source of dynamics. Lastly we consider how more recursive causal mechanisms, such as the presence of an independent effect from grandparents, affect our conclusions. For simplicity we consider the scalar case (a single skill) throughout the section.

From Generations to Cohorts

While the theoretical literature considers how intergenerational mobility evolves over generations, the empirical literature instead typically estimates mobility trends over cohorts. These two dimensions, which do not match if parental age at birth varies across families or time, have to our knowledge not previously been linked in the literature.

Mobility measures are usually indexed to offspring cohorts, a convention that we will follow here.
To link our theoretical implications to empirical trends, we thus introduce a cohort (or birth-year) dimension into our model of intergenerational transmission. We adopt the following notation to distinguish between cohorts and generations. Let the random variable $C_t$ denote the cohort into which a member of generation $t$ of a family is born. Let $A_{t-1}(C_t)$ be a random variable that denotes the age of the parent at birth of the offspring born in cohort $C_t$. For simplicity we assume $A_{t-1}(C_t)$ to be independent of parental income and characteristics, but we allow for dependence on $C_t$ so that the distribution of parental age at birth can change over time. Member $t-j$ of a family is then born in cohort

$$C_{t-j} = C_t - A_{t-1}(C_t) - ... - A_{t-j}(C_{t-j+1}) = C_{t-j}(C_t, A_{t-1}, ..., A_{t-j}).$$

(4.15)

Member $t-j$ becomes parent at age

$$A_{t-j} = A_{t-j}(C_{t-j+1}) = A_{t-j}(C_t, A_{t-1}, ..., A_{t-j+1}).$$

(4.16)

Denote realizations of these random variables by lower case letters. Our model for intergenerational transmission between offspring born into cohort $C_t = c$ and a parent born in cohort $C_{t-1} = c_{t-1}$ is then given by

$$y_t(c) = \gamma_c y_{t-1}(c_{t-1}) + \rho_c e_t(c) + u_t(c)$$

(4.17)

and

$$e_t(c) = \lambda_c e_{t-1}(c_{t-1}) + v_t(c),$$

(4.18)

where we assume a single productive characteristic and keep the same simplifying assumptions on parameters and variables as in our generations-only model in equations (4.2) and (4.3).

By considering a single set of equations per cohort we abstract from life-cycle effects within a given cohort. The transmission parameters in (4.17) and (4.18) can thus be interpreted as representing an average of effective transmission mechanisms over the life-cycle. For example, the price parameter $\rho_c$ does not reflect returns to characteristics in year $c$, but average returns throughout the working life of an individual born in year $c$.

Using (4.17) and (4.18), the intergenerational income elasticity for offspring of generation $t$ born in cohort $C_t = c$ is then

$$\beta_c = \frac{\text{Cov}(y_t(c), y_{t-1}(C_{t-1}))}{\text{Var}(y_t(c))} = \gamma_c + \rho_c \lambda_c \text{Cov}(e_{t-1}(C_{t-1}), y_{t-1}(C_{t-1})).$$

(4.19)

where we do not explicitly note that all random variables are conditional on $C_t = c$ in order to keep the notation short. Income mobility for a given cohort
thus depends on cohort-specific transmission mechanisms ($\gamma_c$, $\rho_c$ and $\lambda_c$) and the cross-covariance of income and characteristics in the parent generation. This cross-covariance may vary with parental age, since different cohorts of parents might have been subject to different policies and institutions. Using (4.15) and the law of iterated expectations we can rewrite

$$\beta_c = \gamma_c + \rho_c \lambda_c E_{A_t-1} \left[ \text{Cov}(e_{t-1}(c-A_{t-1}), y_{t-1}(c-A_{t-1})|A_{t-1}) \right]$$

for $\beta_c$ in terms of parameter values,

$$\beta_c = \gamma_c + \rho_c \lambda_c \sum_{a_{t-1}} \left( f_c(a_{t-1}) \text{Cov}(e_{t-1}(c-a_{t-1}), y_{t-1}(c-a_{t-1})) \right), \quad (4.20)$$

where $f_c$ is the probability mass function for parental age at birth of cohort $c$. Income mobility thus depends on current transmission mechanisms and a weighted average of the cross-covariance of income and characteristics in previous cohorts, where the weights are given by the cohort-specific distribution of parental age in the population.

As before we can iterate backwards to express $\beta_c$ in terms of parameter values,

$$\beta_c = \gamma_c + \rho_c \lambda_c E_{A_{t-1}} \left[ \text{Cov}(e_{t-2}(C_{t-2}), y_{t-2}(C_{t-2})|A_{t-2}) \right] \gamma_{C_{t-1}} + \rho_{C_{t-1}} |A_{t-1}|$$

$$= \ldots$$

$$= \gamma_c + \rho_c \lambda_c \sum_{a_{t-1}} \left( f_c(a_{t-1}) \rho_{c-a_{t-1}} \right) + \rho_c \lambda_c \times$$

$$\sum_{r=1}^\infty \left( \sum_{a_{t-1}} \left( f_c(a_{t-1}) \sum_{a_{t-2}} \left( f_c(a_{t-2}) \ldots \sum_{a_{t-r-1}} \left( f_c(a_{t-r-1}) \prod_{s=1}^r \left( \gamma_{c_{t-s}}, \lambda_{c_{t-s}} \right) \right) \right) \right) \right).$$

Given assumptions or data on the time series of parental age at birth we can thus use equation (4.21) to analyze the dynamic response of mobility trends over cohorts to parameter changes. The insights from the more simple generations-only model still hold (e.g., past transmission mechanisms affect mobility of cohorts today), but comparison of equations (4.21) and (4.6) leads to a number of additional implications.

First, note that both expressions simplify to the same steady-state elasticity given in equation (4.7). The explicit consideration of cohorts in intergenerational transmission models has thus only consequences for transitions between steady states, which presumably explains why the theoretical literature, with its focus on steady states, has not yet been linked to measures of cohort-specific mobility. Second, from (4.21) it follows that the importance of past transmission mechanisms (and thus of past institutions and policies) on current mobility rises with parental age at birth.18 Likewise, the impact of structural changes

---

18 A consideration of life-cycle effects (as in Conlisk, 1969; or Cunha and Heckman, 2007) would be interesting, but the general implications that we discuss here would hold as long as some intergenerational transmission mechanisms tend to be effective in early life (e.g., genetic transmission, childhood environment, and education).
on mobility trends will die out faster in populations in which individuals tend to become parents at younger ages.

These findings might be of interest for cross-country comparisons, especially between developed and developing countries. Equation (4.21) implies that cross-country mobility differentials are not only driven by differences in current and past transmission mechanisms, but also by different weights on those past mechanisms. Various theoretical reasons have been proposed to why developing countries might be characterized by lower levels of intergenerational mobility than developed countries (see Levine and Jellema, 2007). But our results also imply that current mobility levels in developing countries are less dependent on past institutions if parents tend to be younger. Differences in mobility levels between developed and developing countries will then underestimate differences in current institutions if countries share a common trend towards more meritocratic institutions and policies.

Parental Investments in Offspring Human Capital

Parental income may affect offspring via various direct and indirect causal pathways. Its primary effect may relate to human-capital investments, a mechanism emphasized in Becker and Tomes (1979, 1986). Our baseline model is consistent with the “mechanistic” transmission equations that are derived from such frameworks, even while we do not model the utility-maximizing investment behavior of parents explicitly. However, accumulated human capital is perishable in these models; none of it is inherited to the next generation, as parental income does not feature in the actual autoregressive endowment process.

To examine if our results depend on this assumption, we introduce a new parameter that governs the impact of parental income \( y_{t-1} \) on offspring human capital \( e_t \), such that equation (4.3) becomes

\[
e_t = \lambda_t e_{t-1} + \phi_t y_{t-1} + u_t.
\] (4.22)

The degree to which parental income and investments matter may for example depend on the educational system or on health care policies. In this extended

\[\text{For example, the assumption of log-linear preferences in Solon (2004) neutralizes interactions involving the additional parameters that govern the investment decision (e.g., the “altruism” parameter); the resulting transmission equations are similar to our baseline model. More evolved models of utility-maximizing behavior of parents, for example involving public human-capital investments (also Solon, 2004), poverty traps, or alternative assumptions regarding parental preferences, may provide additional implications.}\]
model, the intergenerational elasticity at time $t$ is given by

$$\beta_t = \gamma_t + \rho_t \phi_t + \rho_t \lambda_t \text{Cov}(e_{t-1}, y_{t-1}), \quad (4.23)$$

and the corresponding steady-state elasticity becomes

$$\beta = \gamma + \phi \rho + \frac{\rho \lambda (\rho + \phi \gamma)}{1 - \lambda \gamma}. \quad (4.24)$$

Parental income may thus have a more cursory ($\gamma_t$) or more persistent ($\phi_t$) effect. To explore how the implications of these channels differ we revisit Example 3, in which we documented a non-monotonic mobility trend after an increase in the return to human capital and a decrease in the relevance of parental income. For illustration we consider parameterizations that lead to the same pre-shock and long-run elasticities, but that give different weights on each of the two income channels. Figure 4.5 shows the transition paths for various initial levels of $\gamma_t$.

**Figure 4.5** A Decline in the Importance of Parental Income and Increasing Returns to Skills (Various Cases)

Note: Numerical example with $\lambda = 0.6$ and a rise in $\rho$ from $\rho_1 = 0.6$ to $\rho_2 = 0.7$ at generation $T$. The four cases considered are: a decline in $\gamma$ from $\gamma_1 = 0.3$ to $\gamma_2 = 0.2$ (parental income has only a cursory effect); from $\gamma_1 = 0.2$ to $\gamma_2 = 0.1$; from $\gamma_1 = 0.1$ to $\gamma_2 = 0$; and for $\gamma_1 = \gamma_2 = 0$ (parental income has only a persistent effect). By restricting the steady-state elasticities, the value of $\phi$ implicitly follows from our choice of $\gamma$ such that $\phi = (\beta + \gamma^2 \lambda - \gamma - \lambda \rho^2 - \beta \gamma \lambda)/\rho$.

The mobility trends are similar for the baseline case ($\phi = 0$) and the extreme alternative case ($\gamma = 0$). Mobility does not trend much after two generations. The initial increase in mobility is larger if parental income has only
a persistent effect, and potentially smaller if both channels play a role. Most importantly, all mobility trends follow a non-monotonic pattern.

Co-Movements in Cross-Sectional Inequality

In the previous section we considered examples in which the marginal distributions of income and characteristics remained constant. Various researchers have however illustrated that cross-sectional inequality and intergenerational mobility are likely interdependent. Interest in this interdependence is further warranted because intergenerational persistence is more consequential when cross-sectional inequality is high, and empirically because countries with large cross-sectional inequality tend to be characterized by lower mobility (see Björklund and Jäntti, 2009; and Blanden, 2009).

A detailed analysis of the transition path of the intergenerational elasticity under variable cross-sectional inequality is presented in Appendix A.2 for all our examples. We summarize here briefly our results. First, elasticity shifts in the first generation after a structural change are always equal to the corresponding shift under constant cross-sectional inequality, since the movements in the variance of $u_t$ that ensured constant cross-sectional inequality in Section 4.2 do not affect the elasticity in the first period after the change. Subsequent shifts are larger or smaller than the corresponding shifts under constant cross-sectional inequality depending on if inequality is expanding or contracting. The transition path of cross-sectional inequality generates long-lasting mobility trends also in Example 1, where we assumed that the direct effect of parental income on offspring is zero. The shift in the elasticity in the second generation after an increase in returns to characteristics might be smaller than the shift in the first generation if cross-sectional inequality increases. The second shift however still exceeds the first shift if the intergenerational correlation is considered instead. In Example 2, the elasticity responses after the first generation are smaller when cross-sectional inequality is decreasing. The conditions for mobility trends to be non-monotonic in Examples 3, 4 and 5 are only slightly altered – they tend to be weaker than the corresponding conditions under constant inequality if cross-sectional inequality is decreasing and stronger if inequality is increasing.

Beyond Parents

We so far assumed that intergenerational transmission works exclusively through the parent generation, such that earlier characteristics of the family tree have no

---

20See, for example, Solon (2004), Davies et al. (2005), and Hassler et al. (2007).
effect on offspring. But how are our conclusions affected when such characteristics are important, so that our model understates the degree of recursiveness? Grandparents may directly affect their grandchildren, for example via reputation, networks, through direct bequests, or assistance in their upbringing (most notably if parents are partly or fully absent). To explore this we will assume that grandparents have an effect on offspring income that is independent from their indirect effect through the parent generation, such that

\[
y_t = \gamma_{Pt} y_{t-1} + \gamma_{GP,t} y_{t-2} + \rho_t e_t + u_t \tag{4.25}
\]
\[
e_t = \lambda_t e_{t-1} + v_t \tag{4.26}
\]

where \(\gamma_{Pt}\) and \(\gamma_{GP,t}\) are the respective impacts of fathers and grandfathers on offspring cohort in generation \(t\). Derivation of the transition path and steady state of the intergenerational elasticity follows the same steps as in our simple model. The first is given by

\[
\beta_t = \gamma_{Pt} + \gamma_{GP,t} \beta_{t-1} + \rho_t \lambda_t Cov(e_{t-1}, y_{t-1}), \tag{4.27}
\]

while the latter equals

\[
\beta = \frac{\gamma_P}{1 - \gamma_{GP}} + \frac{\rho^2 \lambda}{(1 - \lambda \gamma_P - \lambda^2 \gamma_{GP})(1 - \gamma_{GP})} \tag{4.28}
\]

**Figure 4.6** AR(1) vs AR(2) Transmission

Note: Numerical example with \(\rho = 0.6, \lambda = 0.6\) and a decline in \(\gamma\) from \(\gamma_1 = 0.3\) to \(\gamma_2 = 0.2\); a decline in \((\gamma_P, \gamma_{GP}) = (0.25, 0.0878)\) from \((\gamma_{P1}, \gamma_{GP1})\) to \((\gamma_{P2}, \gamma_{GP2}) = (0.17, 0.0641)\) at generation \(T\).
We revisit Example 2 to illustrate how the transition path is affected. Figure 4.6 plots the transition path from our original AR(1) model together with a parameterization of our extended AR(2) model that yields the same steady state elasticities before and after the shock. The latter is characterised by somewhat slower convergence to the new steady state, thus illustrating a very intuitive point: transmission mechanisms that allow for a greater role of past family characteristics also imply a greater role for past institutions. An extension from our baseline model to contain more evolved family effects will thus yield more persistent dynamics, strengthening our argument that the effect of past institutions on current mobility trends are important.

4.4 Conclusions

Starting from the notion that transitions between steady states are of particular importance in intergenerational mobility research, we examined the dynamic relationship between income mobility and its underlying structural factors. We based our analysis on a simple theoretical model that fits into the standard framework used in the literature, deviating only in our emphasis of its dynamic properties and our consideration of a multidimensional skill vector. We showed that structural changes such as policy or institutional reform can lead to long-lasting mobility trends, and that the transition path between two steady states may often be non-monotonic. Some implications are surprising, especially our finding that negative mobility trends today can stem from gains in equality of opportunity in the past. Other implications, for example that mobility will tend to be higher in times of structural changes, may have a more intuitive appeal.

We further discussed why a dynamic view may help to distinguish different types of structural shocks, and illustrated how it can be used to interpret existing empirical evidence on the impact of certain policies. While we thus focused on general properties of the relationship between causal transmission mechanisms and mobility trends, we also related our findings to a number of practical cases. One example is our finding that the long-run effect of changing returns to skills tends to be fully reflected in mobility levels only after both the parent and offspring generations have been exposed to the new price regime. The full impact of the recent rise in wage differentials on intergenerational mobility levels may thus not yet have been realized. This is one potential reason why little evidence of falling mobility trends has so far been found for countries that experienced widening skill gaps (see, for example, Lee and Solon, 2009) – a result that has been surprising for the literature since there are both theoretical reasons (see Solon, 2004) and suggestive empirical evidence (from cross-country comparisons) to expect a negative relationship between intra-
generational inequality and intergenerational mobility.

This may be a reason for concern for mobility proponents since it implies that a decline in mobility might yet to be uncovered by empirical research. On the other hand, our results also imply that under certain conditions a contemporaneous decline in mobility may have a rather innocuous explanation. We showed that a decline in the influence of parental income and a rise in the importance of own skills as determinants of earnings (as for example caused by a change from plutocratic to more meritocratic institutions) will likely generate a non-monotonic mobility trend – a sudden increase in mobility being followed by a longer-lasting negative trend. Slight contemporaneous declines in mobility might thus just be a rebounding effect that we should expect to occur in any country that became more meritocratic and mobile in the 20th century.

A final case of interest relates to the empirical literature on changes in mobility levels over longer time intervals, such as the recent work by Long and Ferrie (forthcoming) on the degree of intergenerational occupational mobility in Britain and the US in the 19th and 20th century. While the robustness of their claim that the US experienced extraordinary levels of mobility in the 19th century is contested (see the comments on Long and Ferrie’s work in the same volume, and Hauser, 2010, for a summary of trend analyses in the sociological literature), we illustrated a mechanism that supports their intuition that high levels of geographical mobility might have contributed to intergenerational mobility. Our model implies that large variations in the local demand and thus prices for specific skills over time (presumably a major reason for high levels of geographic mobility) will tend to raise intergenerational mobility since inheritable skills that were beneficial for parents are then less likely to be of similar value for their offspring. Our model is of course highly stylized, and a thorough analysis of these or other applications would require a careful discussion of various issues that we only touched upon.

Co-movements in cross-sectional inequality and the timing of intergenerational transmission over the life-cycle are two dimensions that we abstracted from in the main analysis, which may however be important in some applications. We also did not comment on the considerable difficulties that hinder reliable estimation of the intergenerational elasticity – observation of snapshots of income may generally not be sufficient to detect gradual changes of mobility over time (e.g., as of heterogeneity in income profiles). Analysis of trends in sibling correlations, with its weaker data requirements, may instead be a more promising route (see Björklund et al. (2009); Levine and Mazumder, 2007). But we hope that our model illustrates what type of issues that need to be considered for an analysis of mobility trends, and why a dynamic view of the underlying transmission framework is important. It underlined that additional moments, such as the covariance between human capital and income
in the parental generation, are central for an understanding of both levels and trends in current income mobility. We thus hope that our analysis, while implying that the interpretation of trends in intergenerational income mobility is difficult, will not discourage work on the topic but rather serve as a resource for future studies.
References


Appendix

A.1 Choices of Parameter Values

The parameters of our model reflect total effects of broad concepts of parental economic status ($\gamma$), parental human capital ($\lambda$), and offspring human capital ($\rho$). All these are imperfectly captured by actual data, existing evidence can thus only provide indications of what these parameters may be. Lefgren et al. (2012) examine the relative importance of different mechanisms in a transmission framework that is similar to ours. Using imperfect instruments that are differentially correlated with parental human capital and income they estimate that in Sweden the effect from parental income explains about a third of the intergenerational elasticity, while the effect from parental human capital explains the remaining two thirds. Given that we are considering mobility in total income, a reasonable lower bound may also be given by the contribution of wealth to the intergenerational correlation, which is reported to be about 25 percent in Bowles and Gintis (2002).

The literature provides more guidance on the mechanistic transmission of human capital ($\lambda$). All since the classic work of Galton to more recent evidence concerning transmission of genes, schooling, and measures of ability, the results imply intergenerational correlations of about 0.3-0.4 and much higher correlations when considering both parents. While estimates may capture to various degrees direct effects of parental income they nevertheless provide an upper bound for $\lambda$ if we assume that the effects of parental income are non-negative. Values of $\lambda$ in the range 0.5-0.8 seem thus reasonable.

Finally, a reasonable lower-bound estimate of $\rho$ can be approximated by evidence on the explanatory power of earnings equations. Studies that observe richer sets of covariates, including measures of cognitive and non-cognitive ability, typically yield estimates of $R^2$ in the neighborhood of 0.40. On the one hand, all the reported estimates are likely to understate the explanatory power of human capital (broadly defined) as of imperfect measurement and omitted variables. On the other hand, these estimates may be overestimates for

---

21 For estimates of correlations in measures of cognitive ability, see Bowles and Gintis (2002) and the studies they cite. For measures of both cognitive ability and non-cognitive ability, see Grönqvist et al. (2010).

22 Examples include Zax and Rees (2002), Mueller and Plug (2006) for the US; Groves (2005) for US and UK, Lindqvist and Vestman (2011) for Sweden, Heineck and Anger (2010) for Germany. Fixed-effects models (e.g., individual or job level) as employed by Mueller and Plug (2006) yield even higher estimates, although some of the difference may be capturing persistent luck rather than unobserved characteristics. Moreover, the explanatory power for earnings in general seem to be somewhat lower than wages.
our purposes if we want to only capture the component of human capital that is orthogonal to parental income. Values of $\rho$ in the range of 0.6-0.8 should however be at least roughly consistent with the empirical evidence.\footnote{Given our standardization that $\text{Var}(y) = \text{Var}(e) = 1$ the $R^2 = 0.4$ translates into $\rho \approx 0.63$.}

These parameter values imply an intergenerational income elasticity in steady-state that is roughly consistent with recent US evidence, which typically report estimates of 0.45-0.55 (see Mazumder, 2005; Lee and Solon, 2009). Conversely, given reliable estimates of $\beta$ we can cross-validate the chosen values for the structural parameters of the model. This both enables us to evaluate the robustness of our parameter choices, and potentially narrow down the implied ranges for our parameters. We can write each parameter as a function of the others in steady state:

\begin{align*}
\beta &= \gamma + \frac{\rho^2 \lambda}{1 - \gamma \lambda} \\
\lambda &= \frac{\beta - \gamma}{\beta \gamma + \rho^2 - \gamma^2} \\
\rho &= \sqrt{\frac{(\beta - \gamma)(1 - \gamma \lambda)}{\lambda}} \\
\gamma &= \frac{\beta \lambda + 1 \pm \sqrt{\beta^2 \lambda^2 - 2\beta \lambda + 4 \lambda^2 \rho^2 + 1}}{2\lambda}
\end{align*}

Plugging in the values discussed above on the right-hand sides of the above equations results in imputations of the parameters that are consistent with our reading of the empirical literature. We can also narrow down some of the implied ranges (most notably we rule out too high values of $\lambda$ and $\rho$, as they cause $\gamma$ to approach zero). We arrive at the following implied ranges:

\begin{align*}
0.45 \leq \beta \leq 0.55, \quad 0.15 \leq \gamma \leq 0.25, \quad 0.60 \leq \rho \leq 0.70, \quad 0.50 \leq \lambda \leq 0.65
\end{align*}

These estimates of course need to be taken with a grain of salt, but should be sufficient to provide a reasonable illustration of the potential quantitative implications of our findings.

**A.2 Transition Dynamics under Variable Cross-Sectional Inequality**

We analyze the transition path of the intergenerational elasticity when cross-sectional inequality in income and characteristics are allowed to vary after...
Interpreting Trends in Intergenerational Income Mobility

133

changes of the slope parameters in our structural equations. In contrast to Section 4.2, we thus assume that the variances of the error terms in equations (4.2) and (4.3) remain constant for \( t \geq T \). As before we normalize the variance of our variables before the change in parameters occurs, such that \( \text{Var}(y_t) = \text{Var}(e_{k,t}) = 1 \) for all \( k, t < T \).

Elasticity shifts in the first generation after a structural change are always equal to the corresponding shift under constant cross-sectional inequality since

\[
\Delta \beta_T = \beta_T - \beta_{T-1} = \frac{\rho \lambda_2 \text{Cov}(e_{T-1}, y_{T-1})}{\text{Var}(y_{T-1})} - \frac{\rho \lambda_1 \text{Cov}(e_{T-2}, y_{T-2})}{\text{Var}(y_{T-2})},
\]

such that changes in the variance of \( u_t \) do not affect the elasticity in the first period after the change. For the first example we will also discuss how the transition path of the intergenerational correlation and the intergenerational elasticity differ under variable cross-sectional inequality.

**Example 1a.** Assume that \( \gamma \) is zero. A change in the heredity parameter from \( \lambda_1 \) in \( t < T \) to \( \lambda_2 \) in \( t \geq T \) shifts the intergenerational elasticity according to

\[
\Delta \beta_T = \beta_T - \beta_{T-1} = \frac{\rho \lambda_2 \text{Var}(e_T) + \rho \lambda_1 \text{Var}(e_{T-1})}{\rho^2 \text{Var}(e_T) + \text{Var}(u_T - 1)} - \frac{\rho^2 \lambda_2 \text{Var}(e_{T-1})}{\rho^2 \text{Var}(e_{T-1}) + \text{Var}(u_{T-1})},
\]

which is equal to the response under constant cross-sectional inequality considered in Section 4.2. Mobility however trends also in subsequent generations:

\[
\Delta \beta_{T+1} = \beta_{T+1} - \beta_T = \frac{\rho^2 \lambda_2 \text{Var}(e_T)}{\rho^2 \text{Var}(e_T) + \text{Var}(u_T)} - \frac{\rho^2 \lambda_2 \text{Var}(e_{T-1})}{\rho^2 \text{Var}(e_{T-1}) + \text{Var}(u_{T-1})}.
\]

The trend is negative if \( \lambda_2 < \lambda_1 \) and \( \text{Var}(u_T) = \text{Var}(u_{T-1}) \neq 0 \) since \( \text{Var}(e_t) \) increases in \( \lambda_t \). The convergence of cross-sectional inequality to its new steady state thus constitutes another source of dynamics.

For empirical analysis of mobility trends it might be more informative to consider the *intergenerational correlation* in log lifetime incomes,

\[
r_t = \text{Corr}(y_t, y_{t-1}) = \frac{\beta_t \sigma_{y_{t-1}}}{\sigma_{y_t}},
\]

to abstract from such changes in the marginal distributions across generations.24 First note that the implied long-run shifts in (steady-state) elasticities and correlations are equal since

\[
\Delta r_\infty = \lim_{k \to \infty} \beta_{T+k} \frac{\sigma_{y_{T+k-1}}}{\sigma_{y_{T+k}}} - \beta_{T-1} \frac{\sigma_{y_{T-2}}}{\sigma_{y_{T-1}}} = \lim_{k \to \infty} \beta_{T+k} - \beta_{T-1} = \Delta \beta_\infty.
\]

24Björklund and Jäntti (2009) note that intergenerational correlations might for the same reasons be preferred in cross-country comparisons.
The transition paths however differ. For example, the initial shift in the correlation,
\[
\Delta r_T = r_T - r_{T-1} = \beta_T \frac{\sigma_{yT-1}}{\sigma_y} - \beta_{T-1} \frac{\sigma_{yT-2}}{\sigma_{yT-1}}
= \rho (\lambda_2 \frac{\sigma_{yT-1}}{\sigma_y} - \lambda_1) \rho,
\]
is smaller (larger) than \(\Delta \beta_T\) when the heritability and thus cross-sectional inequality increases (decreases).

**Example 1b.** Assume that \(\gamma\) is zero. A change in prices from \(\rho_1\) to \(\rho_2\) shifts the elasticity according to
\[
\Delta \beta_T = \beta_T - \beta_{T-1} = \frac{\rho_2 \lambda \text{Cov}(e_{T-1}, y_{T-1})}{\text{Var}(y_{T-1})} - \frac{\rho_1 \lambda \text{Cov}(e_{T-2}, y_{T-2})}{\text{Var}(y_{T-2})}
= (\rho_2 - \rho_1) \lambda \rho_1,
\]
which is equal to the response under constant cross-sectional inequality considered in the previous section. The second shift equals
\[
\Delta \beta_{T+1} = \beta_{T+1} - \beta_T = \frac{\rho_2 \lambda \text{Cov}(e_T, y_T)}{\text{Var}(y_T)} - \frac{\rho_2 \lambda \text{Cov}(e_{T-1}, y_{T-1})}{\text{Var}(y_{T-1})}
= \rho_2 \lambda \left( \frac{\rho_2}{\text{Var}(y_T)} - \rho_1 \right).
\]
Since \(\text{Var}(y_T) = \rho_2^2 + \text{Var}(u_T) = 1 + \rho_2^2 - \rho_1^2\) we have that \(\Delta \beta_{T+1}\) is smaller in absolute value than the corresponding shift under constant variance. Its size relative to \(\Delta \beta_T\) depends on the parameter values. By plugging in for \(\text{Var}(y_T)\) and reformulating equation 4.37 we can show that \(\Delta \beta_{T+1}\) may in principal be negative even when prices increase (\(\rho_2 > \rho_1\)), specifically iff \(1 - \rho_1(\rho_2 + \rho_1) < 0\). This condition can only be satisfied for \(\rho_2 > \sqrt{0.5}\). The effect of a stronger correlation between characteristics and income in the parent generation, which increases the elasticity, may thus be dominated by the diminishing effect of cross-sectional inequality changing from an expanding \((\sigma_{yT} > \sigma_{yT-1})\) to a stable \((\sigma_{yT+1} = \sigma_{yT})\) state.

The intergenerational correlation shifts according to
\[
\Delta r_T = \beta_T \frac{\sigma_{yT-1}}{\sigma_y} - \beta_{T-1} \frac{\sigma_{yT-2}}{\sigma_{yT-1}} = (\frac{\rho_2}{\sigma_y} - \rho_1) \lambda \rho_1,
\]
and
\[
\Delta r_{T+1} = \beta_{T+1} \frac{\sigma_y}{\sigma_{yT+1}} - \beta_T \frac{\sigma_{yT-1}}{\sigma_{yT}} = \frac{1}{\sigma_y} \rho_2 \lambda \left( \frac{\rho_2}{\sigma_y} - \rho_1 \right).
\]
Both shifts are smaller in absolute value than the corresponding shifts under constant cross-sectional inequality. The second shift however still exceeds the first shift if prices and cross-sectional inequality increase \((\rho_2 > \rho_1)\) since it then holds that \((1/\sigma_y)\rho_2 > \rho_1\).
Example 2. A decline in the importance of parental income from $\gamma_1$ to $\gamma_2$ shifts the elasticity according to

$$
\Delta \beta_T = \beta_T - \beta_{T-1} = \left[ \gamma_T + \frac{\rho \lambda \text{Cov}(e_{T-1}, y_{T-1})}{\text{Var}(y_{T-1})} \right] - \left[ \gamma_{T-1} + \frac{\rho \lambda \text{Cov}(e_{T-2}, y_{T-2})}{\text{Var}(y_{T-2})} \right]
$$

$$
= \gamma_2 - \gamma_1,
$$

and

$$
\Delta \beta_{T+1} = \frac{\rho \lambda \text{Cov}(e_T, y_T) - \rho \lambda \text{Cov}(e_{T-1}, y_{T-1})}{\text{Var}(y_T)} - \frac{\rho \lambda \text{Cov}(e_{T-1}, y_{T-1} + \rho e_T) - \rho \lambda \text{Cov}(e_{T-1}, y_{T-2} + \rho e_{T-1})}{\text{Var}(y_{T-1})}
$$

$$
= \rho \lambda \rho \left( \frac{1}{\text{Var}(y_T)} - 1 \right) + (\gamma_2 - \gamma_1) \rho \lambda^2 \text{Cov}(e_{T-1}, y_{T-1}).
$$

Again we find that the first shift is equal and the second shift is smaller (since $1/\text{Var}(Y_T) > 1$ for $\gamma_2 < \gamma_1$) than the corresponding shifts under constant cross-sectional inequality.

Example 3. A decrease in the relative importance of parental status from $\gamma_1$ to $\gamma_2$ and an increase in the returns to characteristics from $\rho_1$ to $\rho_2$ shift the elasticity according to

$$
\Delta \beta_T = \left[ \gamma_T + \frac{\rho_T \lambda \text{Cov}(e_{T-1}, y_{T-1})}{\text{Var}(y_{T-1})} \right] - \left[ \gamma_{T-1} + \frac{\rho_{T-1} \lambda \text{Cov}(e_{T-2}, y_{T-2})}{\text{Var}(y_{T-2})} \right]
$$

$$
= (\gamma_2 - \gamma_1) - (\rho_2 - \rho_1) \lambda \text{Cov}(e_{T-1}, y_{T-1}),
$$

and

$$
\Delta \beta_{T+1} = \left[ \gamma_{T+1} + \frac{\rho_{T+1} \lambda \text{Cov}(e_T, y_T)}{\text{Var}(y_T)} \right] - \left[ \gamma_T + \frac{\rho_T \lambda \text{Cov}(e_{T-1}, y_{T-1})}{\text{Var}(y_{T-1})} \right]
$$

$$
= \rho_2 \lambda \left( \frac{\rho_2}{\text{Var}(y_T)} - \rho_1 \right) + \rho_2 \lambda \left( \frac{\gamma_2}{\text{Var}(y_T)} - \gamma_1 \right) \text{Cov}(e_{T-1}, y_{T-1}).
$$

The first shift is equal to the corresponding response under constant cross-sectional inequality. The transition path is v-shaped if if the first shift is negative (see condition 4.10) and the second shift is positive, which is the case if

$$
\left( \frac{\rho_2}{\text{Var}(y_T)} - \rho_1 \right) < \lambda \text{Cov}(e_{T-1}y_{T-1}).
$$

These conditions hold for a greater range of parameters if $\text{Var}(y_T) < 1$ and for a smaller range if $\text{Var}(y_T) > 1$ than the corresponding condition (4.11). We likewise find that the second shift is larger than the corresponding shift under constant cross-sectional inequality if $\text{Var}(y_T) < 1$ and smaller if $\text{Var}(y_T) > 1$. 
Example 4. Changes in the returns to individual characteristics on the labor market ($\boldsymbol{\rho}_1 \neq \boldsymbol{\rho}_2$) shift the elasticity according to

$$
\Delta \beta_T = \frac{\rho'_2 \Lambda \text{Cov}(e_{T-1}, y_{T-1}) - \rho'_1 \Lambda \text{Cov}(e_{T-2}, y_{T-2})}{\text{Var}(y_{T-1})} - \frac{\rho'_1 \Lambda \text{Cov}(e_{T-1}, y_{T-1})}{\text{Var}(y_{T-2})}
= (\rho'_2 - \rho'_1) \Lambda (I - \gamma \Lambda)^{-1} \rho_1,
$$

(4.45)

and

$$
\Delta \beta_{T+1} = \frac{\rho'_{T+1} \Lambda \text{Cov}(e_{T}, y_T) - \rho'_T \Lambda \text{Cov}(e_{T-1}, y_{T-1})}{\text{Var}(y_T)}
= \rho'_2 \Lambda \left( \frac{\rho_2}{\text{Var}(y_T)} - \rho_1 \right) + \rho'_2 \Lambda \gamma \Lambda (I - \gamma \Lambda)^{-1} \left( \frac{\rho_1}{\text{Var}(y_T)} - 1 \right).
$$

(4.46)

We thus find that $\Delta \beta_T < 0 < \Delta \beta_{T+1}$ iff

$$
(\rho'_2 - \rho'_1) \Lambda (I - \gamma \Lambda)^{-1} \rho_1 < 0 < \rho'_2 \Lambda \left( \frac{1}{\text{Var}(y_T)} - 1 \right) + \rho'_2 \Lambda \gamma \Lambda (I - \gamma \Lambda)^{-1} \rho_1 \left( \frac{1}{\text{Var}(y_T)} - 1 \right).
$$

(4.47)

These conditions are weaker than the corresponding conditions under fixed variance if cross-sectional inequality is decreasing ($\text{Var}(y_T) < 1$) and stronger if inequality is increasing ($\text{Var}(y_T) > 1$). It follows that symmetric (or close to symmetric) parameter changes, in which the returns for two characteristics that have the same (or similar) heritability are interchanged, will necessarily satisfy conditions (4.47) if they satisfy the corresponding condition (4.13) under constant variance. To see this, rewrite the variance as

$$
\text{Var}(y_T) = \gamma^2 \text{Var}(y_{T-1}) + \rho'_2 \text{Var}(e_T) \rho_2 + \text{Var}(u_T) + 2 \gamma \rho'_2 \Lambda \text{Cov}(e_{T-1}, y_{T-1})
= \text{Var}(y_{T-1}) + \rho'_2 \rho_2 - \rho'_1 \rho_1 + 2 \gamma (\rho'_2 - \rho'_1) \Lambda (I - \gamma \Lambda)^{-1} \rho_1,
$$

(4.48)

and note that the last term in (4.48) is negative as of the first part in condition (4.13), while $\rho'_2 \rho_2 - \rho'_1 \rho_1$ will be zero for symmetric parameter changes, such that $\text{Var}(y_T) < 1$.

Example 5. Changes in the heredity of characteristics ($\boldsymbol{\Lambda}_1 \neq \boldsymbol{\Lambda}_2$) shift the elasticity according to

$$
\Delta \beta_T = \frac{\rho'_T \Lambda \text{Cov}(e_{T-1}, y_{T-1}) - \rho'_T \Lambda \text{Cov}(e_{T-2}, y_{T-2})}{\text{Var}(y_{T-1})} - \frac{\rho'_T \Lambda \text{Cov}(e_{T-1}, y_{T-1})}{\text{Var}(y_{T-2})}
= \rho' (\Lambda_2 - \Lambda_1) (I - \gamma \Lambda_1)^{-1} \rho_1,
$$

(4.49)
Interpreting Trends in Intergenerational Income Mobility

\[
\Delta \beta_{T+1} = \frac{\rho' \Lambda_{T+1} \text{Cov}(e_T, y_T)}{\text{Var}(y_T)} - \frac{\rho' \Lambda_T \text{Cov}(e_{T-1}, y_{T-1})}{\text{Var}(y_{T-1})}
\]

\[
= \rho' \Lambda_2 \left( \frac{\text{Var}(e_T)}{\text{Var}(y_T)} - 1 \right) \rho + \rho' \Lambda_2 \gamma \left( \Lambda_2 \frac{1}{\text{Var}(y_T)} - \Lambda_1 \right) (I - \gamma \Lambda_1)^{-1} \rho.
\]

We thus find that \( \Delta \beta_T < 0 < \Delta \beta_{T+1} \) iff

\[
\rho' (\Lambda_2 - \Lambda_1) (I - \gamma \Lambda_1)^{-1} \rho < 0 < \rho' \Lambda_2 \left( \frac{\text{Var}(e_T)}{\text{Var}(y_T)} - 1 \right) \rho + \\
\rho' \Lambda_2 \gamma \left( \Lambda_2 \frac{1}{\text{Var}(y_T)} - \Lambda_1 \right) (I - \gamma \Lambda_1)^{-1} \rho.
\]

Symmetric parameter changes, in which the heritability for two characteristics that yield the same returns are interchanged, will necessarily satisfy condition (4.51) if they satisfy the corresponding condition (4.14) under constant cross-sectional inequality. To see this, rewrite the variance as

\[
\text{Var}(y_T) = \gamma^2 \text{Var}(y_{T-1}) + \rho' \text{Var}(e_T) \rho + \text{Var}(u_T) + 2\gamma \rho' \Lambda_2 \text{Cov}(e_{T-1}, y_{T-1})
\]

\[
= \text{Var}(y_{T-1}) + \rho' (\Lambda_2 \Lambda_2 - \Lambda_1 \Lambda_1) \rho + \\
2\gamma \rho' (\Lambda_2 - \Lambda_1) (I - \gamma \Lambda_1)^{-1} \rho,
\]

and note that the last term in (4.52) is negative as of the first part in condition (4.14), while \( \rho' (\Lambda_2 \Lambda_2 - \Lambda_1 \Lambda_1) \rho \) will be zero for symmetric parameter changes, such that \( \text{Var}(y_T) < 1 \) (and \( \text{Var}(y_\infty) < 1 \)). The appearances of \( \text{Var}(e_T) \), or \( \text{Var}(e_\infty) \), in condition (4.51) is inconsequential for symmetric changes as \( \text{Var}(e_T) = I + \Lambda_2 \Lambda_2 - \Lambda_1 \Lambda_1 \) and, again, \( \rho' (\Lambda_2 \Lambda_2 - \Lambda_1 \Lambda_1) \rho \) will be zero.
5. The Role of Parental Income over the Life Cycle: a Comparison of Sweden and the UK*

Introduction

The literature on intergenerational earnings and income mobility (or its inverse transmission), which has developed since the early 1990s, has revealed some striking cross-national differences. The United States has, to the surprise of many observers, come out as the country with the strongest intergenerational transmission among developed countries, whereas the Nordic countries have come out as high-mobility countries. The estimates for the United Kingdom are more mixed, but generally reveal lower mobility than in the Nordic countries.

What are the underlying mechanisms behind these differences? Are they related to policy? And if policies matter, what are the most important ones? No doubt, it is a great challenge for research to find out what mechanisms account for these cross-country differences. The most popular theoretical framework in the economics literature – the so-called Becker-Tomes model (1979, 1986) – suggests that a number of quite different mechanisms can account for country differences. Solon’s (2004) parameterization of this model points at four broad factors, namely (i) “mechanical” (for example, genetic) transmission of income-generating traits, (ii) the efficacy of investment in children’s human capital, (iii) the earnings return to human capital, and (iv) the progressivity of

*This is joint work with Anders Björklund and Markus Jäntti. We thank Jo Blanden, Peter Fredriksson, Libertad Gonzalez, Magne Mogstad and seminar and workshop participants at, CHESS and SOFI (Stockholm University), DIW and Free University (Berlin), IFAU (Uppsala), Oslo University, Universitat Autònoma de Barcelona, University of Chicago, University of Mannheim for useful comments. Financial support from Swedish Council of Social and Working Life Research (FAS) is gratefully acknowledged.

1See Solon (2002), Corak (2006), Jäntti et al. (2006), Björklund and Jäntti (2009), and Blanden (2013) for alternative cross-national surveys of intergenerational income and earnings estimates.
public investment in children’s human capital.

The Becker-Tomes model focuses on parental investment in the health and skills of their children. The model also takes into account how these investments interact with public-policy investments in the form of education and health-care systems. Our reading of the literature is that most scholars, who try to understand intergenerational income and earnings correlations, have a child-development focus of the type highlighted by the Becker-Tomes model. Studies focusing on the labor market as the scene that generates intergenerational correlations are less frequent. Yet, one of the factors that follow from the model is the earnings return to human capital – (iii) above – and we know that such returns vary widely across countries’ labor markets.

Our contribution is to account for cross-national differences in intergenerational income correlations by examining the importance of family income in different phases of the offspring’s life. We compare Sweden, a high-mobility country, with the United Kingdom, a country with relatively low mobility, to study at what stage during the life cycle that the country differences emerge. Do they show up already early in life so that parental income matters more for early childhood characteristics such as health and school performance in the UK compared to Sweden? Or do the country differences not appear until the offspring generation has reached the labor market?

We use the British Cohort Study (BCS) of children born in 1970 and Swedish register data to explore the role of parental income for (1) birth weight, (2) height at late teen ages, (3) grades at the end of compulsory school at age 16, (4) final educational attainment and (5) long-run earnings during adulthood. We start with (5), that is, we start by presenting comparable estimates of intergenerational income transmission for the two countries. Our estimates show that comparable intergenerational income elasticities and correlations are significantly higher in the UK than in Sweden; for sons the elasticities are 0.271 vs. 0.199 and the correlations are 0.227 vs. 0.119; for daughters the difference in elasticities is slightly higher but the difference in correlations slightly lower. These results are in conformity with those reported in the previous literature.

Then we continue to explore whether these differentials show up already early in the life cycle in outcomes (1)-(4) mentioned above. We find that there

---

2 An interesting exception with focus on labor-market mechanisms is Corak and Piraino (2011), who examine intergenerational transmission of employers.

3 See, e.g., Blanden et al. (2004) for the UK. As we explain below, we use income and earnings definitions to suit the UK data and as a consequence the Swedish estimates are not comparable to previous ones. In a companion paper, we compare the gradients in child outcomes with respect to parental education in Sweden and the UK using the same data sources (Björklund et al., 2012).
are indeed significant country differences already early in the life cycle, and the associations are stronger in the UK than in Sweden. We then ask whether these differences are large enough to account for the differences in intergenerational income persistence that motivated our study. For this purpose, we perform a decomposition analysis of the intergenerational income elasticity that gives the covariance between parental income and the child outcomes as well as the earnings returns to the child outcome a role. We experiment with this decomposition and let the UK get the Swedish covariance (and vice versa) and the UK get the Swedish returns (and vice versa). This exercise – although mechanical in some respects – suggests that the country differences in birth-weight and height associations are not strong enough to account for hardly any fraction of the UK-Sweden difference in intergenerational income mobility. For grades and final education, on the other hand, we find that country differences in intergenerational associations can account for a substantial part of the difference in income mobility.

The paper proceeds as follows. Section 5.1 offers a short literature background. We explain our data in Section 5.2. Section 5.3 reports our estimates of intergenerational associations. In Section 5.4, we perform our decomposition analyses to investigate the importance of the various outcomes. We report some robustness checks in Section 5.5 and conclude in Section 5.6.

5.1 Literature Background

Birth Weight

A large recent literature, efficiently surveyed by Currie (2009), has shown that birth weight in general and low birth weight in particular are related to parental socio-economic status. There are several reasons to expect such an association. The quality of the nutritional intake during pregnancy is one obvious candidate explanation. Parents in higher socioeconomic groups may also be better informed about health-related hazards that impact on the growth of the fetus. Such mechanisms have come to be known as the “fetal origin hypothesis”, associated with the British epidemiologist David J. P. Barker. In addition, one can expect that biological mechanisms (genetic inheritance) cause a relationship between parental socio-economic status and birth weight. The country’s health care system might very well have an impact on how strong these mechanisms are. A more compensatory system for care of pregnant mothers may attenuate both the nature and nurture of these relationships.

Birth weight, in turn, has been shown to predict several later and adult outcomes. The literature surveyed by Currie is full of examples showing that birth weight predicts outcomes related to health, including mortality, but also cog-
nitive and noncognitive skills and thus also labor market performance. These relationships may show up because of the specific health problems that are related to low birth weight. However, it is also possible that low birth weight have direct effects on the acquisition of skills. Parents’ reactions to the problems related to the underlying health problems are also likely to affect the impact of early health problems on subsequent acquisition of skills. The same applies to the health-care and school systems in the country.

Much of the recent literature focuses on the issue whether the associations between parental resources and birth weight and between birth weight and adult outcomes are causal or not. For our purposes it is not a major concern whether the associations are causal or not. We treat this variable as an indicator of several traits that are related to parental income and own performance later in life. Our research question is cross-national: we ask whether parental income is more strongly related to this set of traits in one country than in another, and whether this set of traits is a stronger predictor of adult outcomes in one country than in another.

Height in Adolescence

As explained below, we have access to data on height measured at ages 16, 26 and 29 for the UK, and at age 18 for Sweden. Obviously, these height measures reflect the cumulative growth up to these ages. This means that the combined genetic and environmental factors that contribute to birth weight also affect our height measures. Indeed, Black et al. (2007) find not only strong associations between birth weight and height during adolescence, but also suggestive evidence of causal effects. But, in addition, our height measures are also sensitive to a number of environmental conditions experienced during childhood. In their thorough survey of height determinants, Case and Paxson (2008) emphasize that the period from birth to age 3 is considered as the postnatal period that is most critical to height. Nutritional needs are greatest at this point in life as is sensitivity to infections of different types. Such factors are, of course, likely to be related to family background and in ways that might differ across countries.

Case and Paxson also emphasize that conditions during childhood affect the timing of children’s growth. The timing of the typical pubertal growth spurt has been found to be sensitive to the child’s health conditions and therefore also to parental background. Case and Paxson demonstrate that the growth spurt comes earlier for children of high socio-economic background. Thus, at some stages during adolescence, the pubertal growth spurt tends to magnify height differences between economic classes. Observing height at different ages in adolescence across countries could therefore give misleading estimates
of country differences. In our analysis below, we avoid this problem by not only comparing height at age 16 for the UK and age 18 for Sweden – the first ages at which we have data – but also complementing this analysis with results for height at ages 26 and 29 for the UK.

The bivariate correlation between height and labor market earnings is quite strong. Using US PSID data, Case and Paxson report that the observed difference of 4 inches in men’s heights at the 25th and the 75th percentiles is associated with an expected earnings differential of 9.2 percent. Furthermore, this association is observed throughout the whole height distribution. Lundborg et al. (2009), using the same Swedish data source as we do in this study, find only somewhat weaker bivariate associations.

Case and Paxson and Lundborg et al. also explore the mechanisms behind this bivariate height-earnings association. Indeed, the candidate mechanisms are several and include self-esteem, social dominance and discrimination. It might also be that height captures omitted variables such as cognitive and noncognitive skills and strength. Case and Paxson find, using US and UK data, that most of the height-earnings association is eliminated when cognitive skills are controlled for, whereas Lundborg et al. find that physical strength is a more important underlying factor. Just as for birth weight, it is not crucial for our purposes whether height is important per se, or if it is an indicator of several underlying productivity traits.

School Grades

Our next mediating variable is grades at age 16 at the end of compulsory school. This variable has a straightforward motivation for our purposes: grades are strongly associated with parental resources and early predictors of adult outcomes. The grades at the end of compulsory school are particularly important since they determine access to both study fields and schools at the upper-secondary level. A common result from research about the predictive performance of such grades is that they not only capture cognitive skills as measured by typical test scores, but also noncognitive skills such as motivation and persistence (see, e.g., Borghans et al., 2011 for recent research).

Final Education

As compulsory school grades, final education is a key variable when trying to understand intergenerational income correlations. Parental and public investment in the human capital of children is a central transmission mechanism in the theoretical work of Becker and Tomes (1979, 1986). From the perspective of intergenerational mobility, it is also important to stress that final education captures both performance in school, as also measured by our grade variable,
and the set of choices of further education made after compulsory school. A large literature in sociology (see, e.g., Erikson et al., 2005) has shown that family background has a strong influence on both school performance (primary social origin effects) and school choices (secondary social origin effects).

5.2 Data, Sample Restrictions and Variable Definitions

Sources and Sample Restrictions

The British Cohort Study (BCS) is a survey of all children born in England, Scotland and Wales in one particular week in April 1970. The BCS is a very rich data set with surveys performed right after birth and at ages 5, 10, 16, 26, 30, and 34. The first sweep covered the births and families of about 17200 children. In the two last sweeps the number of observations fell to 11200 (in 2000) and 9600 (in 2004). With each sweep, the scope of enquiry has broadened from a strictly medical focus at birth to encompass physical and educational development during the child’s growth, and later on economic and labor market outcomes as adults.

For Sweden we have access to register data from various sources, which have been merged by Statistics Sweden using unique personal identifiers. For intergenerational research purposes, this is a very flexible data source. In this study, we use the available information for Sweden to “mimic” the UK data set as closely as possible. Our main sample for Sweden consists of all who were born in the country in 1973. We restrict our analysis to this cohort since it is the first one for which birth-weight data are accessible. In short, a number of data sets are merged together in order to obtain our variables of interest: the Swedish census in 1985 is used to identify each child’s rearing parents in the fall of this year; birth-weight data are obtained from the National Board of Health and Welfare (in Swedish, Socialstyrelsen); height data are obtained from the compulsory military enlistments tests administered by The Swedish National Service Administration (Plåtverket); and separate registers at Statistics Sweden provide data on compulsory school grades at age 16 (Årskurs-9 registret), final education of both parents and offspring (the Education Register, or Utbildningsregistret), and income and earnings based on tax declarations for both parents and offspring.

Variables

**Parental income.** In the BCS data set, parental income is defined as “all earned and unearned income of both father and mother” in 1980 and 1986. Moreover, it refers to the income of present parents (i.e., not necessarily bio-
logical parents) and is measured before taxes. To mimic this variable with the Swedish data, we use a measure of total income that includes income before taxes from all sources except means-tested benefits and universal child benefits (sammanräknad inkomst). We use data from the years 1983 and 1989 for Sweden since we want to measure parental income at identical ages of the child. In order to take into account only the income of present parents, we define parental income for those adults who lived in the same household as the child according to the census in 1985.\(^4\)

An important difference between the parental income data in the two countries is that the UK data are available only as discrete intervals of the income distribution. Thus, we are restricted to use the center of each of the respective intervals as the measure of parental income. Moreover, the top interval is not bounded from above and consequently there is no center for this interval. We therefore obtain a proxy for those in the top interval by calculating the median gross family income in this top interval according to the Luxemburg Income Study (LIS). To ensure comparability, we use the Swedish data on parental income in the same way, i.e., by dividing the data into corresponding intervals and using the center values. For the unbounded highest interval, we apply the actual median. These income measures are used in our main analyses, but we also conduct a sensitivity analysis with the Swedish data by comparing the main results based on intervals with results from using the actual data.

**Offspring earnings.** The BCS includes data on offspring labor earnings for ages 30 and 34. From the Swedish tax register we extract information on labor earnings (arbetsinkomst) for the same ages. We use the log of the average of these two annual earnings observations as our measure of long-run earnings.

For the UK, both income and earnings information refers to weekly data at the time of the interview, whereas Swedish data sets offer annual measures of income and earnings. We divide the Swedish numbers by 52 in order to show comparable data in our descriptive tables, which of course does not affect our estimates. More important are the potential differences in selection that come with these differing data definitions. In the UK data, an individual needs to record positive earnings in a given week to be included in the sample. In the Swedish data, an individual only needs to record positive earnings sometime during the year to be included. Our UK sample is therefore likely to be more

\(^4\) This choice creates a slight discrepancy from the UK data, which defines present parents at each ages 10 and 16 of the child. We could have identified parents in both the 1980 and 1990 censuses but in that case the children would have been 17 years old in the latter year and some might have moved out of their parents’ home. We could also have used the 1980 census but then the child would have been quite young compared to UK.
selective with a lower share of individuals with intermittent labor-market behavior. We have considered testing our results’ sensitivity to this by restricting the Swedish data to individuals with less signs of intermittency in their earnings data. However, we have not found a satisfactory way of implementing such a test.

To sum up, the income and earnings measures we end up using are not ideal. They are most certainly causing some attenuation bias in our estimated intergenerational associations since the parental income measure is an average of only two annual income observations. Further, we measure offspring’s earnings a couple of years too early in order to minimize the so-called lifecycle bias, at least according to evidence for Sweden. This probably also contributes to some downward bias in our intergenerational income estimates. Our maintained assumption in this study is that these two sources of bias are of the same magnitude in the two countries.

**Birth weight.** For the UK, the birth-weight data stem from the initial sweep of the BCS and are based on reports from hospitals. For the Swedish data set, the birth-weight variable is also obtained from hospital reports delivered to the National Board of Health and Welfare. In our econometric specifications, we use as dependent variables birth weight in kilograms, log of birth weight and a dummy for low birth weight defined as less than 2500 grams.

**Height.** The UK height data stem from a professional medical examination of the survey respondents at age 16. For Sweden, we obtain information on height from data collected at the military enlistment that is compulsory for all Swedish men. Most men do these tests at age 18. Thus we cannot do this analysis for women. Because we measure height at different ages in the two countries, we standardize the variable when we use it as dependent variable

5 See, e.g., Mazumder (2005) for an examination of the relationship between the number of annual observations and the bias due to transitory income variation.

6 For Swedish cohorts born around 1950, Böhlmark and Lindquist (2006) show that annual earnings approximate lifetime earnings at around age 33. With similar data on cohorts born 1955-57, Nybom and Stuhler (2011) however show that the intergenerational elasticity of lifetime income is best approximated when using income of the offspring from somewhat older ages. There is no comparable evidence on this for the UK.

7 Another issue is related to the fact that we choose to measure parental income at a specific age of the child and thus at different ages of the parents depending on the age of parents at child’s birth. We therefore control for parental age throughout, and additionally assume that any remaining bias is similar in Sweden and the UK such that it does not affect our cross-country comparisons. An alternative with perfect data would have been to use parental lifetime income.
The Role of Parental Income over the Life Cycle

in our regressions. As noted in the previous section, we also complement our analysis for the UK by looking at height data from ages 26 and 29 as a robustness check.

School grades. Comparing grades across countries is not unproblematic. In order to be able to include grades on a comparable basis, we transform each grade of selected subjects of every person into a percentile rank, and then take the average percentile rank across all subjects as each individual’s grade measure. For the UK, we use the grades in the O-level (or CSE) examinations in the English language, English literature, mathematics, science, physics, chemistry, biology, history, geography, French, German and business studies. We have spliced together the O-levels and CSE according to the following ranking: (1) O level A, (2) O level B, (3) O level C or CSE 1, (4) O level D or CSE 2, (5) O level E or CSE 3, (6) CSE 4, (7) CSE 5, and (8) fail.\(^8\) We invert the scale so that higher grades get coded higher, and assign to each person for each subject the percentile that corresponds to the tabulated percentile at the grade he or she received. For Sweden, we use the grades at the end of compulsory school (i.e., at age 16 after nine years of compulsory schooling) in English, biology, physics, chemistry, technology, geography, history, religion, social studies and Swedish. We choose these subjects since they are taken by all students according to an identical curriculum. Next, we tabulate the distribution of grades for each subject and assign to each person the percentile rank of the grade they received. We then take the average of a person’s rank across all subjects as our measure of their grade. This procedure has been used before, for example by Björklund et al. (2003), who studied the importance of family background for school performance in Sweden across time. We have experimented with variations of these procedures, and while the exact estimates vary, the qualitative conclusions we reach do not. However, we must concede that the results with respect to the parental background gradient in grades are in some sense likely to be the least comparable, for the simple reason that the schooling institutions in Sweden and the UK differ. As an example, student tracking by field of study occurs at an earlier age in UK than in Sweden. Although this is on the one hand problematic for our analysis, it is on the other hand exactly these types of institutional differences that motivate our study and make the UK-Sweden comparison interesting.

Final education. Since educational systems often differ across countries it is not as straightforward how to use data on final education in cross-country

---

\(^8\)This splicing together of the grades for the UK was implemented for Björklund et al. (2012) after conversations with Jo Blanden and John Ermisch.
studies as, e.g., data on income and earnings. In this paper, we apply the International Standard Classification of Education (ISCED) developed by UNESCO to order highest educational attainment into hierarchical categories. The ISCED was designed with the purpose to make educational levels internationally comparable. We have combined the two lowest categories into a single one, labelled as 2 in the result tables, since category 1 is nearly empty for the cohort we study. The Swedish data come from Statistics Sweden’s education register. The UK data are self reported and we use the BCS wave from 2004. Thus, final education is measured at age 34 for the UK sample. For the offspring in the Swedish sample, we employ data on final education from 2007, i.e., also when the subjects were 34 years old.

5.3 Results

Intergenerational Income Associations

Although the results from previous studies are our point of departure, we start by estimating new intergenerational measures (elasticities and correlations) on our own sample. Table 5.1 shows the associated descriptive statistics. UK parents are about one year younger when their income is measured, a rather small difference. For both countries, we note that the standard deviation of log of parental income is lower than the standard deviation of the log of offspring’s earnings. This difference implies that our estimated intergenerational correlations will differ from our estimated intergenerational elasticities.

<table>
<thead>
<tr>
<th>Table 5.1 Descriptive Statistics for Income Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sons</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Log earnings, offspring</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Log income, parents</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Parental age</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Fraction with two parents</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>n</td>
</tr>
</tbody>
</table>

Note: Table shows means with standard deviations within parentheses. We use weekly income and earnings, PPP-adjusted to 2005 dollars. Swedish annual income and earnings are divided by 52 in order make the data comparable. We condition on offspring’s earnings > 10 dollars/week. Parental age (the mean of both parents) and fraction with two parents are measured at the time parental income is measured. The latter is for descriptive purpose only and not included in the regressions.

For a few observations with missing data in 2004, we instead used information from the year 2000 sweep.
We report estimates of standard intergenerational associations in Table 5.2. Since there is age variation among parents, we have in every case controlled for parental age (and its square) both here and in the subsections below. As expected from previous research, the intergenerational income associations are lower for Sweden than for the UK. The standard errors are reasonably small, so the null hypothesis of equality across countries is strongly rejected. For men, the point estimates of the elasticities are .199 for Sweden and .271 for the UK, and the corresponding correlations – regressions with standardized income measures for both parents and offspring – are .119 and .227. The differences for women are of similar magnitude. As discussed above, all these estimates are downward biased but hopefully equally much for both countries’ estimates.

Table 5.2 Estimates of Intergenerational Income Associations

<table>
<thead>
<tr>
<th></th>
<th>Sons</th>
<th></th>
<th>Daughters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sweden</td>
<td>UK</td>
<td>F-test</td>
<td>Sweden</td>
</tr>
<tr>
<td>Regression coef</td>
<td>0.199 (0.008)</td>
<td>0.271 (0.021)</td>
<td>4.5 [0.034]</td>
<td>0.179 (0.008)</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.119 (0.005)</td>
<td>0.227 (0.017)</td>
<td>29.4 [0.000]</td>
<td>0.108 (0.005)</td>
</tr>
<tr>
<td>Quadratic (lin)</td>
<td>−0.432 (0.114)</td>
<td>−0.956 (0.297)</td>
<td>9.0 [0.003]</td>
<td>−0.375 (0.118)</td>
</tr>
<tr>
<td>(sq.)</td>
<td>0.050 (0.009)</td>
<td>0.101 (0.024)</td>
<td>0.044 (0.009)</td>
<td>0.096 (0.035)</td>
</tr>
</tbody>
</table>

Note: We use the average of the parents’ age and its square as controls in the regression models in this and the subsequent tables. Standard errors are within parentheses and p-values within brackets.

In row 3 of Table 5.2, we also report estimates for a model that adds a quadratic term for the log of parental income (unstandardized). The estimates suggest that the functional form is not very different in the two countries. Thus, we stick to the simple linear framework to explore the cross-national differences in factors underlying intergenerational income persistence.

Birth Weight

We estimate simple models with birth weight as dependent variable and the log of long-run parental income as independent variable. In the literature, dif-

10 Because we use cohort data, we do not have to control for offspring’s age. Since we use family income for parents, we control for the average age of the two parents.
11 An evaluation of the quadratic model shows that the elasticity is rising with parental income in both countries and for both genders. The difference in the elasticity between the countries is constant for daughters but rising with parental income for sons. In our decomposition analysis in Section 5.4, we ignore this different pattern for sons.
ferent measures of birth weight have been used, for example, raw birth weight in kilograms, log of birth weight, a dummy indicator for low birth weight (typically less than 2.5 kilos), and fetal growth (defined as birth weight divided by weeks of gestation). Black et al. (2007) examine the explanatory power of these variables and report that log of birth weight provide the best fit for their outcome variables, which included adult earnings. Nevertheless, we report results for birth weight in kilos, the log of birth weight and low birth weight.

Panels A-C in Table 5.3 report sample descriptives. Here and in subsequent analysis of the other traits, we use the maximum sample size in the data set. In Section 5.5 below, we report sensitivity analysis based on balanced samples for all outcomes. We note that birth weight for the 1973 cohort in Sweden exceeds the birth weight for the 1970 cohort in the UK by around 180-210 grams. Otherwise the standard deviations are similar. Thus, not surprisingly, the prevalence of low birth weight is higher in the UK than in Sweden; around 6 percent compared to around 3 percent.

Estimates for birth weight and low birth weight, respectively, are shown in Panels A-C of Table 5.4. Our basic conclusions are robust with respect to the two measures, and also whether we standardize the variables or not. A first basic conclusion is that there is indeed a significant parental income gradient in birth weight for both sons and daughters, and in both countries. A second conclusion is that the associations are significantly stronger in the UK than in Sweden. Nonetheless, the magnitude of all the estimated intergenerational associations is quite low. For example, the highest coefficient reported in Panel B is .095 for UK daughters, and the correlation between the birth weight and the log of parental income is only .099.

Height

Panel D in Table 5.3 reports sample descriptives for our height samples. Because in the base case we measure height at age 16 for the UK and at age 18 for Swedish sons, it is natural that Swedish sons are taller than the UK sons in our data. Because of this difference in measurement, we must treat estimates of country differences based on unstandardized height with caution. Thus, we should pay most attention to the estimates denoted as correlations.

Panel D in Table 5.4 reports our estimates for age 16 in the UK. All single estimates are strongly significantly different from zero. Again, we also find a significant country difference with larger coefficients for the UK. Yet, the magnitude of the correlations between parental income and height is modest, at most .142 at age 16 in the UK. We have also estimated the same models on

---

12 We note though that their conclusion refers to twin fixed-effects regressions.
Table 5.3 Descriptive Statistics for Outcome Samples

<table>
<thead>
<tr>
<th></th>
<th>Sons</th>
<th>Daughters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sweden</td>
<td>UK</td>
</tr>
<tr>
<td>A. Birth weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth weight</td>
<td>3.55 (0.54)</td>
<td>3.37 (0.55)</td>
</tr>
<tr>
<td>Log parental inc.</td>
<td>6.51 (0.46)</td>
<td>6.13 (0.54)</td>
</tr>
<tr>
<td>Parental age</td>
<td>40.83 (5.02)</td>
<td>39.09 (5.88)</td>
</tr>
<tr>
<td>n</td>
<td>49875</td>
<td>5356</td>
</tr>
<tr>
<td>B. Log birth weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log birth weight</td>
<td>1.25 (0.17)</td>
<td>1.20 (0.18)</td>
</tr>
<tr>
<td>Log parental inc.</td>
<td>6.51 (0.46)</td>
<td>6.13 (0.54)</td>
</tr>
<tr>
<td>Parental age</td>
<td>40.83 (5.02)</td>
<td>39.09 (5.88)</td>
</tr>
<tr>
<td>n</td>
<td>49875</td>
<td>5356</td>
</tr>
<tr>
<td>C. Low birthweight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low birthweight</td>
<td>0.03 (0.17)</td>
<td>0.06 (0.24)</td>
</tr>
<tr>
<td>Log parental inc.</td>
<td>6.51 (0.46)</td>
<td>6.13 (0.54)</td>
</tr>
<tr>
<td>Parental age</td>
<td>40.83 (5.02)</td>
<td>39.09 (5.88)</td>
</tr>
<tr>
<td>n</td>
<td>49875</td>
<td>5356</td>
</tr>
<tr>
<td>D. Height</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>1.80 (0.06)</td>
<td>1.74 (0.09)</td>
</tr>
<tr>
<td>Log parental inc.</td>
<td>6.52 (0.46)</td>
<td>6.20 (0.52)</td>
</tr>
<tr>
<td>Parental age</td>
<td>40.81 (5.01)</td>
<td>39.91 (5.76)</td>
</tr>
<tr>
<td>n</td>
<td>45851</td>
<td>2822</td>
</tr>
<tr>
<td>E. Average grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average grade</td>
<td>0.47 (0.21)</td>
<td>0.50 (0.13)</td>
</tr>
<tr>
<td>Log parental inc.</td>
<td>6.52 (0.46)</td>
<td>6.16 (0.53)</td>
</tr>
<tr>
<td>Parental age</td>
<td>40.82 (5.01)</td>
<td>39.70 (5.82)</td>
</tr>
<tr>
<td>n</td>
<td>48915</td>
<td>4212</td>
</tr>
<tr>
<td>F. Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.51 (0.26)</td>
<td>0.26 (0.35)</td>
</tr>
<tr>
<td>4</td>
<td>0.14 (0.08)</td>
<td>0.08 (0.14)</td>
</tr>
<tr>
<td>5</td>
<td>0.26 (0.35)</td>
<td>0.35 (0.39)</td>
</tr>
<tr>
<td>Parental age</td>
<td>40.84 (5.02)</td>
<td>39.48 (5.83)</td>
</tr>
<tr>
<td>n</td>
<td>48811</td>
<td>3277</td>
</tr>
</tbody>
</table>

Note: Table shows means (with standard deviations within parentheses) of child outcomes and log parental income for the different outcome samples. Parental age (the mean of both parents) and fraction with two parents are measured at the time parental income is measured. Birth weight is measured in kilograms, low birth weight is an indicator for less than 2.5 kilograms. Height is in meters and education is measured as ISCED levels, with levels 1 and 2 merged into one (the omitted category).
<table>
<thead>
<tr>
<th></th>
<th>Sweden</th>
<th>UK</th>
<th>Sweden</th>
<th>UK</th>
<th>Sweden</th>
<th>UK</th>
<th>Sweden</th>
<th>UK</th>
<th>Sweden</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F. Education &amp; correlate</td>
<td>E. Average grade &amp; correlate</td>
<td>D. Height &amp; correlate</td>
<td>C. Low birth weight &amp; correlate</td>
<td>B. Log birth weight &amp; correlate</td>
<td>A. Birth weight &amp; correlate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>corr</td>
<td>regr. coef.</td>
<td>corr</td>
<td>regr. coef.</td>
<td>corr</td>
<td>regr. coef.</td>
<td>corr</td>
<td>regr. coef.</td>
<td>corr</td>
<td>regr. coef.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>(0.00)</td>
<td>(100.00)</td>
<td>(0.00)</td>
<td>(100.00)</td>
<td>(0.00)</td>
<td>(100.00)</td>
<td>(0.00)</td>
<td>(100.00)</td>
<td>(0.00)</td>
<td>(100.00)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.95</td>
<td>1.52</td>
<td>0.87</td>
<td>1.75</td>
<td>0.87</td>
<td>0.90</td>
<td>0.77</td>
<td>0.77</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 5.4: Estimates of Intergenerational Associations – Child Outcomes on Log Parental Income

Notes: Standard errors are within parentheses and p-values within brackets. Birth weight is measured in kilograms, low birth weight is an indicator for less than 2.5 kilograms. Height is in meters and education is measured as ISCED.
UK data for ages 26 and 29 with similar results; the results are available upon request.

School Grades

We report sample descriptives for our grade samples in Panel E of Table 5.3. The overall mean of the average percentile is slightly above 0.5 for daughters and slightly below for sons. This reflects a general and well-known gender gap in school performance. A difference between the countries, however, is that the standard deviation in grades is markedly higher in Sweden. This reflects the fact that the grade systems are different in the two countries; the variation is affected by the number of fields that are graded and by the number of steps for each graded field. Because of this difference, we focus on the results with standardized variables in Panel E of Table 5.4.

It is not surprising that all single parameters are strongly significantly different from zero. Interestingly, however, we also find significant country differences in the intergenerational transmission estimates, and again the UK ones are the largest for the standardized variables.

Final Education

Our basic education information is available in the form of levels of education. One option would be to transform these levels into years of schooling in order to estimate very simple models with years of schooling as dependent variable for both countries. However, we prefer to use the original data since the transformation from levels to years might create country-specific errors. Thus, we estimate ordered-probit models with the four ISCED levels as dependent variable (recall that we have combined levels 1 and 2).

We report sample descriptives in panel F in Table 5.3. The distribution of individuals across the five ISCED levels differs somewhat across the countries. This could both reflect institutional differences and the fact that the individuals in the UK sample are likely to be less representative of the population. The mode for the UK sample is level 5, whereas the mode for the Swedish sample is level 3. This reflects the classification of the standard post-compulsory (high school) track in each country. We also note a quite clear gender gap in educational attainment in Sweden (to the advantage of women), but not in the UK.

We report our estimates in panel F of Table 5.4, and offer a visual presentation of the implications of the estimates in Figure 5.1. We have used standardized parental income. As expected, the parental income gradients in educational attainment are significantly different from zero in both countries.
As for the other traits, the point estimates of the intergenerational transmissions coefficients are higher for the UK than for Sweden. However, in this case only the difference among daughters is significantly different from zero. When looking at the visual presentation in Figure 5.1, we can see that the different gradients are most evident between Swedish and UK offspring at levels 2 and 3. The visual impression is quite similar for men and women, even if the difference among men is insignificant.

**Figure 5.1 Regression of Offspring Final Education on Log Parental Income**

Note: Graph shows the gradient of probability of different final education outcomes (i.e., ISCED 2, 3, 4, and 5) of the offspring with respect to standardized log parental income.
5.4 Interpretation of the Results

So far we have found that there is indeed a UK-Sweden difference in the association between parental income and child outcomes already early in life. But how important are these differences? Can they possibly account for a considerable part of the country difference in intergenerational income associations that we reported in Section 5.3? Or is it more likely that returns to the productive traits that we have examined account for the cross-national differences? To explore this we use a straightforward analytical framework that distinguishes between, on the one hand, the intergenerational covariances between parental income and the productivity traits, and, on the other hand, the returns to these productivity traits.\textsuperscript{13}

Denote a single productivity trait for the child generation by \(X^c\), and the relevant income measures for parents and children by \(Y^p\) and \(Y^c\), respectively. We make use of a linear regression of the child’s earnings on each productivity trait that we study, for example

\[
Y^c = \alpha + \beta X^c + \epsilon, \tag{5.1}
\]

where \(X^c\) in the first case represents birth weight. Using equation (5.1) and \(Y^p\), the estimated intergenerational income elasticity (IGE) becomes

\[
\text{IGE} = \frac{\text{Cov}(Y^p, Y^c)}{\sigma^2} = \frac{\beta \text{Cov}(Y^p, X^c)}{\sigma^2} + \frac{\text{Cov}(Y^p, \epsilon)}{\sigma^2}, \tag{5.2}
\]

where \(\sigma^2\) denotes the variance of parental income. By dividing \(Y^p\) and \(Y^c\) by their respective standard deviations, we instead obtain a decomposition of the intergenerational income correlation (IGC).

We can examine the contribution of each of our productivity traits to the intergenerational income correlation by estimating equation (5.1), retrieving the residuals, and computing the components of equation (5.2). It is the first component – the product of the monetary return to the trait and the covariance between the trait and parental income – that can be attributed to the trait we consider, whereas the second component represents everything else. We can also separately examine the importance of the monetary returns \(\beta\) and the intergenerational covariance between parental income and the trait.

In Table 5.5, we report the results from this accounting exercise for Sweden and the UK.\textsuperscript{14} We restrict the analysis to sons. Neither birth weight nor height – our measures of early childhood and late teen health traits – can account

\textsuperscript{13}For similar exercises, see Österbacka (2001) and Björklund et al. (2005).

\textsuperscript{14}Note that the IGCs differ slightly from the ones in Table 5.2. The reason is that we have estimated the IGCs in Table 5.5 on samples that also contain valid measures of the respective traits.
for hardly anything of the Sweden-UK difference in the IGC. We can see that
the component that is attributed to these traits is very small compared to the
total correlation in both the countries. Thus, the small absolute cross-national
difference in the component cannot explain a substantial part of the country-
national difference in the IGC.

When we turn to grades, another pattern emerges. First, we see that the
components attributed to grades make up about one third of the IGC in both
countries. In absolute terms, this component is .081 for the UK and .035 for
Sweden. Thus, one can say that .046 (=.081-.035) of the cross-national dif-
ference can be attributed to factors associated with grades at age 16. When
considering this difference, it is interesting to look separately at the contribu-
tions of the monetary returns to grades ($\beta$) and the covariance between the trait
and parental income. The country difference obviously pertains about equally
to both these components, with a slightly higher contribution from the mon-
etary return. When we let Sweden have the UK return, its IGC goes up to
.138 (from .116), and when we let Sweden have the UK covariance it goes
up to .130. Similarly, when we let the UK have the Swedish return, its IGC
goes down to .218 (from .250), and when we let the UK have the Swedish
covariance it goes down to .226.

Finally, we have the results for education. For this outcome, the component
attributed to the trait also makes up about one third of the total IGC. However,
the UK IGC is not reduced that much neither by giving the UK the Swedish
return nor by giving the UK the Swedish covariance. The Swedish IGC, in
turn, goes up more with UK returns than with the UK covariances (from .120
to .149 and .143, respectively). The results are consequently somewhat mixed
for education.

5.5 Robustness Checks

So far we have used the largest possible sample for each separate analysis.
By so doing, we have maximized the precision and minimized the sample se-
lection bias for each estimated parameter. It could, however, be argued that
our comparisons of the relative importance of different traits becomes flawed
by the fact that we compare estimates from different samples. Therefore, we
have also reestimated all parameters in Table 5.4 and the decompositions in
Table 5.5 on balanced samples. The requirement that all variables are simul-
taneously available reduces the samples by some 10 percent for Sweden, but
from 3153 to 1511 for UK daughters and from 3304 to 719 for UK sons. As
a consequence, the precision of the estimates falls. However, the main pattern
of the results remain the same. The differences between the intergenerational
income associations in Table 5.2 remain about the same. The country differ-
Table 5.5 Decompositions of the Intergenerational Income Correlation for Sons

<table>
<thead>
<tr>
<th>Birth weight</th>
<th>IGC</th>
<th>$\beta \cdot \text{Cov}(Y^p, X^c)$</th>
<th>Cov($Y^p, \varepsilon$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swedish estimates</td>
<td>0.119</td>
<td>0.028*0.020=0.001</td>
<td>0.119</td>
</tr>
<tr>
<td>UK estimates</td>
<td>0.233</td>
<td>0.045*0.080=0.004</td>
<td>0.230</td>
</tr>
<tr>
<td>UK with Sw. returns</td>
<td>0.232</td>
<td>0.002</td>
<td>0.230</td>
</tr>
<tr>
<td>UK with Sw. cov.</td>
<td>0.231</td>
<td>0.001</td>
<td>0.230</td>
</tr>
<tr>
<td>Sw. with UK returns</td>
<td>0.120</td>
<td>0.001</td>
<td>0.119</td>
</tr>
<tr>
<td>Sw. with UK cov.</td>
<td>0.121</td>
<td>0.002</td>
<td>0.119</td>
</tr>
<tr>
<td>Log birth weight</td>
<td>IGC</td>
<td>$\beta \cdot \text{Cov}(Y^p, X^c)$</td>
<td>Cov($Y^p, \varepsilon$)</td>
</tr>
<tr>
<td>Sw. estimates</td>
<td>0.119</td>
<td>0.030*0.023=0.001</td>
<td>0.118</td>
</tr>
<tr>
<td>UK estimates</td>
<td>0.233</td>
<td>0.049*0.076=0.004</td>
<td>0.230</td>
</tr>
<tr>
<td>UK with Sw. returns</td>
<td>0.232</td>
<td>0.002</td>
<td>0.230</td>
</tr>
<tr>
<td>UK with Sw. cov.</td>
<td>0.231</td>
<td>0.001</td>
<td>0.230</td>
</tr>
<tr>
<td>Sw. with UK returns</td>
<td>0.120</td>
<td>0.001</td>
<td>0.118</td>
</tr>
<tr>
<td>Sw. with UK cov.</td>
<td>0.120</td>
<td>0.002</td>
<td>0.118</td>
</tr>
<tr>
<td>Low birth weight</td>
<td>IGC</td>
<td>$\beta \cdot \text{Cov}(Y^p, X^c)$</td>
<td>Cov($Y^p, \varepsilon$)</td>
</tr>
<tr>
<td>Sw. estimates</td>
<td>0.119</td>
<td>-0.021*-0.023=0.000</td>
<td>0.119</td>
</tr>
<tr>
<td>UK estimates</td>
<td>0.233</td>
<td>-0.031*-0.035=0.001</td>
<td>0.232</td>
</tr>
<tr>
<td>UK with Sw. returns</td>
<td>0.233</td>
<td>0.001</td>
<td>0.232</td>
</tr>
<tr>
<td>UK with Sw. cov.</td>
<td>0.233</td>
<td>0.001</td>
<td>0.232</td>
</tr>
<tr>
<td>Sw. with UK returns</td>
<td>0.120</td>
<td>0.001</td>
<td>0.119</td>
</tr>
<tr>
<td>Sw. with UK cov.</td>
<td>0.120</td>
<td>0.001</td>
<td>0.119</td>
</tr>
</tbody>
</table>

Note: Each panel shows the intergenerational income correlation for sons in each country, decomposes this into a covariance of the intermediate offspring outcome and its income return, as well as the covariance of the error with parental income. These terms are used to calculate counterfactual income correlations by imputing the returns and observed covariance. See equations 5.1 and 5.2 in text.
ences in the parameters of the trait variables are also the same except for height which are no longer significantly different across the two countries. The decompositions reported in Table 5.5 also tell the same story with the balanced sample. A full set of estimates based on the balanced sample is available from the authors upon request.

To ensure comparability across countries, we used the Swedish data on parental income in the same way as for the UK, i.e., by dividing the data into intervals of the parental income distribution and using the center values in each interval. We briefly examine the sensitivity of the results to this particular feature of the data by comparing our main results with results from using the actual data (i.e., continuous measures for Sweden). We report descriptive statistics and estimates for this sensitivity analysis in Appendix Table A. 5.1. The discrepancies are small and do not affect any of our conclusions.

5.6 Conclusions

We have presented an approach to exploring the mechanisms behind cross-national differences in intergenerational income transmission. We applied this approach on data from Sweden and the UK, but we think that it can serve as a model also for other cross-national comparisons.

We first explored the importance of two variables that have received much attention in recent research by economists, namely birth weight and height during adolescence. Although we found a significant cross-national difference in the association between these traits and parental income, only a trivial magnitude of the intergenerational income correlation (and the country difference in this correlation) could be attributed to these traits. The product of the monetary return in adulthood to these traits and the covariance between these traits and parental income were simply too small to be important for the observed intergenerational mobility parameters. This is not to say that early childhood is not important for intergenerational income mobility, but our results suggest that these variables do not capture enough of important mechanisms.

When we turned to the more conventional human-capital variables, grades and final education, the results were different. Especially for grades at the end of compulsory school, we found that factors associated with this variable could account for about a third of the intergenerational income correlation. Further, a substantial part of the country difference in intergenerational income mobility could be accounted for by such factors. Nonetheless, it was striking that the higher monetary return to grades in the UK labor market compared to the Swedish was the most important factor behind this result.

Although our results do not prove that differences in early childhood circumstances are not important for cross-national differences in intergenera-
tional income mobility, they suggest that future research should look broadly for possible determinants of these differences. The labor market strikes us as an important arena with lots of mechanisms that can explain these differences.
References


Appendix

A.1 Continuous Parental Income Data for Sweden

Table A. 5.1 Descriptives and Estimates of Intergenerational Income Associations - Continuous Parental Income for Sweden

<table>
<thead>
<tr>
<th></th>
<th>Sons</th>
<th>Daughters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sweden</td>
<td>UK</td>
</tr>
<tr>
<td>Log earnings, offspring</td>
<td>6.25</td>
<td>6.57</td>
</tr>
<tr>
<td></td>
<td>(0.77)</td>
<td>(0.61)</td>
</tr>
<tr>
<td>Log income, parents</td>
<td>6.51</td>
<td>6.18</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(0.51)</td>
</tr>
<tr>
<td>Parental age</td>
<td>40.80</td>
<td>39.45</td>
</tr>
<tr>
<td></td>
<td>(4.99)</td>
<td>(5.77)</td>
</tr>
<tr>
<td>Fraction with two parents</td>
<td>0.88</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>48424</td>
<td>3304</td>
</tr>
</tbody>
</table>

B. Estimates

<table>
<thead>
<tr>
<th></th>
<th>Sons</th>
<th>Daughters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sweden</td>
<td>UK</td>
</tr>
<tr>
<td>Regression coef</td>
<td>0.186</td>
<td>0.271</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.120</td>
<td>0.227</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Quadratic (lin)</td>
<td>−0.143</td>
<td>−0.956</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.297)</td>
</tr>
<tr>
<td>(sq.)</td>
<td>0.027</td>
<td>0.101</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.024)</td>
</tr>
</tbody>
</table>

Note: Estimates correspond to specifications in Tables 5.1 and 5.2 except that these rely on continuous rather than grouped parental income as the regressor for Sweden. Standard errors are within parentheses and p-values within brackets.