Essays on Intergenerational Income Mobility, Geographical Mobility, and Education

Stefanie Heidrich

Department of Economics
Umeå School of Business and Economics
Umeå University, Umeå 2016

Umeå Economic Studies No. 932
ISBN: 978-91-7601-440-0
ISSN: 0348-1018

Cover: Christian Scholz & Anna Heidrich
Electronic version available at http://umu.diva-portal.org/
Printed by: Print & Media at Umeå University
Umeå, Sweden, 2016

Für Papa

ABSTRACT

In **Paper [I]** we analyze the implications of social identity and self-categorization for optimal redistributive income taxation. A two-type model is supplemented by an assumption that individuals select themselves into social categories, in which norms are formed and education effort choices partly depend on these norms. The results show, among other things, that externality correction by a welfarist government leads to an element of tax progression that serves to reduce the discrepancy between the effort norm and the actual effort chosen by low-productivity individuals in the high-effort group. Furthermore, if the preference for social identity is sufficiently strong, increased wage-inequality leads to higher social welfare through a relaxation of the selection constraint. It may thus be desirable to use publicly provided education to induce more wage-inequality, even if higher wage-inequality increases the intrinsic utility of a potential mimicker.

In **Paper [II]** I employ high quality register data to present new facts about income mobility in Sweden. The focus of the paper is regional differences in mobility, using a novel approach based on a multilevel model. This method is well-suited when regions differ greatly in population size as is the case in Sweden. The maximum likelihood estimates are substantially more precise than those obtained by running separate OLS regressions. I find small regional differences in income mobility when measured in relative terms. Regional differences are large when adopting an absolute measure and focusing on children with below-median parent income. On the national level I find that the association between parent and child income ranks has decreased over time, implying increased mobility.

In **Paper [III]** I study the long term effects of inter-municipal moving during childhood on income using Swedish register data. Due to the richness of the data I am able to control for important sources of selection into moving, such as parent separation, parents' unemployment, education, long run income, and immigration background. I find that children's long run incomes are significantly negatively affected by moving during childhood, and the effect is larger for those who move more often. For children who move once, I also estimate the effect of the timing and the quality of the move. I measure the quality of each neighborhood based on the adult outcomes for individuals who never move. The quality of a move is defined as the difference in quality between the origin and

the destination. Given that a family moves, I find that the negative effect of childhood moving on adult income is increasing in age at move. Children benefit economically from the quality of the region they move to only if they move before age 12 (sons) and age 16 (daughters).

Applied research on the association between parent and child lifetime income is relying on income data that covers only part of the life cycle which may lead to misleading estimates of the intergenerational elasticity (IGE). In **Paper [IV]** I study the bias of IGE estimates for different missing-data scenarios based on simulated income processes. Using an income process from the income dynamics and risks literature to generate two linked generations' complete income histories, I use Monte Carlo methods to study the relationship between available data patterns and the bias of the IGE. I find that the traditional approach using the average of the typically available log income observations leads to IGE estimates that are around 40 percent too small. Moreover, I show that the attenuation bias is not reduced by averaging over many father income observations. Using just one income observation for each generation at the optimal age (as discussed in the paper) or using weighted instead of unweighted averages can reduce the bias. In addition, the rank-rank slope is found to be clearly less sensitive to missing data.

ACKNOWLEDGEMENTS

There is an old proverb that says: “Behind every successful man there is a great woman”. If completing this Ph.D. thesis in economics makes me in some sense a successful woman, I can report that there were a significant number of great men who supported me along the way! First of all, I want to thank my supervisor Thomas Aronsson. You have been a fantastic guide through the turbulent waters that constitute the life of a Ph.D. student! Without your positive attitude, your regular high doses of encouragement, all your time spent reading and discussing my research, and your help with all other aspects of Ph.D. student life, I doubt that I would have finished this thesis. I also want to thank my assistant supervisors David Granlund and Magnus Wikström for diversifying my care-package during my time in Umeå and for always having an open door for me to discuss smaller and bigger questions.

I am also greatly indebted to my fiancé Spencer who has been at my side from the beginning of this journey and who has always believed in me, long before I started to believe in myself. I am incredibly grateful for all our discussions and arguments that helped me to reflect about and widen my perspective. You were there when things went well, but also when they didn’t, and you never stopped supporting me.

I learned a great deal about perspective during my time as a visiting fellow at Harvard in Boston, Massachusetts. Experiencing daily the contrasts between Ivy League and the homeless, between insiders and outsiders, strengthened my interest in social justice and the extent we are able to decide over our own fate; topics I want to continue working on even after this thesis. I am extremely grateful having had this opportunity to experience amazing professors, courses and seminars, and having met a bunch of fantastic people from all over the world, united in being curious, ambitious, and full of positive energy.

There are of course many people that helped me arrive at this point, a greater number than I can list here. I want to thank my department for providing an excellent work environment and support during my studies. Some people have made my time here especially memorable. Kalle, thank you for your support (and your unforgettable seminar comments)! Gauthier, at times I felt I had four supervisors, considering all the times you have read and commented on my papers. Thank you!

Tharshini, please know that you made my first months in Umeå so much easier! And even though I do prefer my own office, if I ever have to share a one-person office again with someone I would definitely choose you. I always appreciated our discussions. David, thanks for never, ever turning me down when I had something on my mind! And for opening your home and social life, helping me to enjoy Umeå even outside of work. Maria, thanks for diversifying the department and for being such a good friend! I am also thinking of my friend Sara who I applied with together to the Ph.D. program. I can hardly believe that it is more than half a decade ago since we sat in the sun in Ultuna and celebrated our admission eating ice cream! Who would have thought that we would actually pull it off, and finally defend our theses with less than a week in between? Not to forget, I also owe a big thanks to Umeå Musiksällskapet. Playing music with you has helped me to stay sane, especially during the last couple of months. I will sincerely miss you.

Finally, I want to thank my family. I am forever grateful for the way my parents have raised and supported me, even when my chosen path was not always the expected one. Even though my dad is not with us anymore to experience this point in my life, I fancy to think that he would be proud. I know my mum is; but really, it is to a large part because of what you two did for me that the world can call me Dr. Heidrich now! I also miss and thank my two fantastic sisters and hope to spend a little more time with you, now this thesis is completed.

Stefanie
Umeå, May 2016

CONTENTS

This thesis consists of an introduction, a summary of the papers, and the following four self-contained papers:

INTRODUCTION

This thesis consists of four independent papers in the areas of labor and public economics. A common theme of these papers is the question how long-term economic outcomes are determined by family- and socio economic background, and by individuals' choices. Some policy implications thereof are also addressed.

A strand of the economics literature that focuses on precisely the tension between free choice and background dependence is the literature on equality of opportunity.[1] The concept of equality of opportunity has traditionally been closely connected to anti-discrimination. Every person should be judged according to his/her abilities, and not according to "irrelevant" characteristics such as race, gender, or religion. Milton and Rose Friedman discussed this notion of equality of opportunity in relation to equality of outcomes in their book "Free to Choose" in 1979. They advocated a free market society with only a minimum amount of public regulation. John Roemer (inspired by Rawls, 1971; Sen, 1985; and Dworkin, 1981a and 1981b) developed a more detailed theory of equality of opportunity (1993, 1996), arguing that a society should guarantee its members equal access to advantage, regardless of their circumstances, while holding them responsible for turning that access into actual advantage by the application of effort. The distinction between circumstances that are outside of someone's control on the one side, and personal effort on the other side is argued to be the key to determine the degree of equality of opportunity in a society or in a given situation.

In practice, measuring (in-) equality of opportunity is very difficult, if not to say impossible. The researcher needs to define opportunity sets for groups of individuals with the same circumstances. The relevant circumstances are often approximated by socio-economic background, and socio-economic background itself is often approximated by the education level of the father. Besides the pitfalls arising due to these potentially imprecise approximations, it is in many cases extremely difficult to draw the line between circumstance on the one hand, and own effort or choice, on the other hand. Do an individual's high school grades depend on the genes and study ethics obtained from the parents (and thereby considered circumstances), or should the individual be held accountable for her grades since a 17-year-old person can reasonably be assumed to understand the importance of school grades for adult outcomes, independent of the

[1] For the record, it was also this literature that initially spurred my interest in pursuing a Ph.D. in economics.

parents? How accountable are the parents, how accountable is the student? Most people will tend to agree that it is probably a combination of both which is at work. However, when trying to measure equality of opportunity, the researcher has to make a decision not only about which circumstances to include, but also how to define and measure them. Moreover, the more “circumstances” that are added to the model, the more unequal will the sample under study appear in terms of opportunity (since allowing for finer and finer distinctions between groups will tend to explain variation in income to a higher and higher degree, leaving less leeway for unaccounted differences between individuals which are typically interpreted as personal effort). In the extreme case, individuals cannot be held responsible for any outcome since most events in life can be traced to family background, genes, childhood environment, or immutable characteristics such as gender and age.

Obviously, translating philosophical principles underlying a fair society with equal chances for success for everyone into hard science, and measuring opportunities, effort, fairness, and free choice in numbers is a very ambitious project (some might say hopeless). This does not mean, however, that we should stop trying. A fruitful direction appears to be to study in depth the aspects of childhood which are most crucial for long run adult outcomes. As noted for instance by James Heckman, the accident of birth is a major source of inequality. Since we cannot choose our parents, a fair society might want to support children with less favorable circumstances as early as possible in their lives, in order to improve the children’s opportunities later in life.

James Heckman and his collaborators have produced a tremendous amount of work examining the origins of inequality, the determinants of social mobility, and the links among different stages of the life cycle. A large part of this work which has been produced within the Center for the Economics of Human Development at the University of Chicago. Heckman and co-authors have for example shown that the positive effects of early intervention programs, such as the Perry Preschool program, mostly worked through non-cognitive skill development, and not IQ enhancement (Heckman et al., 2013). Conti, Heckman and Urzua (2013) have highlighted the important role played by the early years in producing health. Since health bequests better educational outcomes, and better outcomes bequest better health, improving the conditions in early life appears to be an effective strategy to promote adult well-being.

The social environment

In very recent work, Chetty, Hendren, and Katz (2015) have shown that moving poor families to richer neighborhoods results in higher high school completion rates and higher adult income for the children. The positive effects were however absent if children moved after the age of 13. Chetty and Hendren (2015) found similar results studying families that move across the US: the earlier a child moves to a better region, the more her adult outcomes improve compared to the outcomes of children in her original region.

However, there is an alternative explanation for this result: instead of benefits being lower the later a child moves, one could equally well interpret this as the costs of moving being higher for children who move later. In addition, neighborhoods might matter for different reasons for child human capital production at different ages. For example, schools, teachers, and direct neighbors can be assumed to matter even for a young child, while peers, social networks and role models might become more important during teenage years. If children's development of identity and sense of self is particularly vulnerable to disruptions from age 13 onwards, we would expect to see large disruption costs even when moving to a very good neighborhood.

Paper [III] in this thesis attempts to shed light on this question by studying families that move across municipalities and counties in Sweden. Comparing children that never moved, to similar children that moved at least once during childhood, I find significant differences in their long-term adult incomes. Moving seems to be costly for Swedish children. Controlling for the type of move (i.e. if a family moves to a relatively better or worse region) as well as the age of the child when moving, I find that the move's quality does matter for adult outcomes, but only when moving occurs before a certain age.

This finding is well in line with research showing that social identity is an important motivating factor and source of well-being for individuals. Akerlof and Kranton (2002) studied how social identity and self-image can affect individuals' outcomes. In their model, students choose effort in school and divide themselves into three social categories with different study norms. Importantly, an individual's utility depends on how well she conforms to the group's social norm, here in terms of study effort. Their paper is based on a large literature in sociology and psychology where the fact that individuals make choices dependent on their social reference group has been acknowledged and studied for a long time. Paper [I] in this thesis contributes to the literature on social identity and schooling by studying optimal redistributive income tax and publicly provided education when individuals compare themselves to a social group.

If social identity matters for individuals' choices, it is important to think about how social groups (and therefore the social norms associated with each such group) are chosen and can be influenced by, for example, parents and society. In the case of moving, we can assume that the social identity that was built up in the original location is lost when the child moves to a new place. Even if it is possible to "join" a new group at the new location, it is reasonable to think that social identity has to be rebuilt. The older a child is, the more social identity will be lost by moving, and the harder it might be to be accepted into a new social group. Social identity theory seems important to be taken into account when studying childhood environment and childhood human capital production.

From inequality to intergenerational income mobility in Sweden

Sweden is usually cited as an example of a coherent society with small differences between the poorest and the richest citizens. Education and health care is publicly provided, free of charge, and cheap student loans are available and accessible for everyone. Nonetheless, Sweden has recently been receiving attention due to its increasing levels of income inequality (OECD, 2011). Real household income increased in Sweden between 1985 and 2000 by around 2.4 percent per year for households in the top decile of the distribution, but only by 0.4 percent annually for households in the bottom decile. Despite the still low absolute level of income inequality, judged by international standards, researchers and politicians are showing great interest in the drivers and effects of this development for different groups in society.

One hypothesis why Americans seem to tolerate more inequality compared to citizens in for example Sweden, is what Bénabou and Ok (2001) termed the prospect of upward mobility. Believing in the American Dream, i.e. a society where you can accomplish almost anything if you just work hard enough, can consistently lead low income individuals to turn against highly redistributive policies. Given the belief that you (or at least your children) will climb up the ladder in society by hard work, there is no reason to support low income individuals with generous social benefits.

Economists have developed the concept of income mobility to refer to the degree to which an individual's income depends on family background and especially parent income, as opposed to just own effort. In a recent summary of the literature, Corak (2013) examined the link between income mobility and income inequality for a number of OECD countries and sheds light on what has been called the "Great Gatsby Curve". This curve illustrates the negative relationship between mobility and inequality found in the data. The less income mobility in a country, the higher is income inequality. Corak writes:

"Inequality lowers mobility because it shapes opportunity. It heightens the income consequences of innate differences between individuals; it also changes opportunities, incentives, and institutions that form, develop, and transmit characteristics and skills valued in the labor market; and it shifts the balance of power so that some groups are in a position to structure policies or otherwise support their children's achievement independent of talent." According to this line of argument, rising income inequality can very well be accompanied by decreasing mobility which gives the accident of birth a larger weight among the drivers of a child's long run outcomes. Consequently, the growth in income inequality has also increased the public interest in the opportunities for Swedish children to improve their financial situation in relation to the financial situation of their parents.

The study of intergenerational income mobility represents a particular perspective on the tension between individuals' own choices and their pre-determined opportunities. In the literature on intergenerational mobility, researchers are interested in the association between parent outcomes and child outcomes. Even though we perhaps ultimately should be interested in how the well-being of parents translate into the well-being of the children, i.e. the transmission of utility across generations, the literature has focused on the transmission of income and wealth across generations since income and wealth is assumed to be a good predictor of well-being. If there is a close relationship between parent and child adult outcome, there is said to be low intergenerational (income) mobility. On the other hand, if parent outcome does not predict child outcome very well, intergenerational (income) mobility is considered high.

One of the very first empirical studies of intergenerational income mobility was conducted by Soltow (1965), who found a very low association between father and son income for a sample of families living in a city in southern Norway. Since then, researchers in all parts of the world have added to the literature by estimating the relationship between parent and child income for different samples. Recent reviews of the literature can be found in Black & Devereux (2011) and Jäntti & Jenkins (2015). A general pattern of those estimates is that the predictability of son income from father income is higher in the US and the UK (implying low mobility), and lower in the Nordic countries (implying a higher mobility).

Importantly, there is no level of mobility that is considered unequivocally optimal. The problem is easy to grasp when thinking about the desirable degree of meritocracy in a society. Most would agree that very talented and knowledgeable individuals should hold positions of power, in the public and the private sector. Moreover, since talent depends on parents' genes, and talent makes it easier to successfully undergo a difficult education and obtain expert skills, a meritocratic society where people are matched to positions depending solely on their skills, will have a strictly positive association between parent and child outcomes. Thus, maximum mobility does not seem desirable based on principles of fairness and economic efficiency. Zero mobility on the other hand, implying that an individual's outcome is completely determined at birth, seems likewise undesirable and we might only want to agree on a degree of mobility that is somewhere in between 0 and 100. Estimates of intergenerational mobility might therefore be most useful when seen in relation to each other, for example in terms of the development over time or a comparison between regions.

Paper [II] in this thesis is contributing to the literature of intergenerational mobility by providing the first regional estimates for two different mobility measures for Sweden. Since all data stems from the same national registers, the resulting mobility estimates are highly comparable.

The measurement problem in the literature on intergenerational mobility

The income path over the life cycle for a given individual consists of a deterministic trend depending on factors such as education, age, and labor market experience, as well as the economic cycle and smaller and larger shocks that can vary in their persistence. Therefore, one annual income observation might not be very informative about an individual's lifetime income. In the best of worlds, estimates of intergenerational income mobility should be based on individuals' complete income histories. However, due to the inability of researchers to observe the complete income history of each individual across their lifecycles, approximations of lifetime income must be used based on the often limited number of income observations available.

These measurement problems make it particularly difficult to compare mobility estimates between different countries, since different countries have different types of data available. In the Nordic countries, population registers with detailed information about annual earned income, capital income, and transfers are available since the 1960's. In other countries, the data is usually less detailed and available for a shorter time span. In order to make meaningful comparisons, the influence of redistributive policies has to be taken into account. Estimates based on before-tax income will give very different results compared to after-tax income. The same holds for income measures including or excluding social insurance transfers such as unemployment benefits or parental benefits. In addition, using different cohorts is problematic due to business cycles affecting different cohorts differently, as well as changing institutions and policies over time. The implications of finding two different values of mobility in two countries are therefore far from clear.

In spite of the challenges discussed here, I believe that measuring intergenerational mobility is a worthy exercise. In a given country, estimates of mobility can be interpreted in the light of existing policies and economic structure. A particularly fruitful direction is to study how mobility differs across regions. This provides a way of relating estimates of mobility to regional factors while not running into the data-related and institutional issues that hamper comparability between cross-country estimates of mobility. It is one potential tool to monitor a country's development in terms of balancing the tradeoff between holding individuals responsible for their outcomes, and acknowledging limits to the opportunity sets imposed by the accident of birth.

Paper [IV] is another contribution to the literature of intergenerational mobility. Here, I employ simulations to illustrate how approximations of lifetime income affect estimates of the intergenerational elasticity as well as the association between long run income ranks between fathers and sons. Because of the typical concave pattern of annual incomes over the lifecycle, income measured during young ages usually underestimates average lifetime income, and income measured during middle ages tends to overestimate average lifetime income. In more or less much all previous papers, life time income has been approximated by using the average over all available income observations in the data. My results show that such approximations may lead to severe measurement errors of lifetime income, and to overestimation of intergenerational mobility.

Measuring the state of income mobility in a society is a very difficult task. Exposing the different channels through which individuals' long-term outcomes are shaped, probably an even harder one. In order to design policies and give advice how to foster equal opportunity, we need to further improve our understanding of the many mechanisms that determine adult outcomes given background, institutions, and effort. My thesis hopefully constitutes a small step towards reaching this goal.

SUMMARY OF THE PAPERS

Paper [I]: Social Identity, Taxation, and Publicly Provided Education

In Paper [I] we analyze the implications of social identity and self-categorization for optimal redistributive income taxation and publicly provided education. In psychology and sociology, social identity theory has long been used to explain human behavior (see, e.g., Hogg & Reid, 2006; Tajfel & Turner, 1979). Social identity can be defined in terms of how a person's sense of self depends on the group (or groups) which the person associates with (e.g., social reference groups such as family, colleagues, friends, social class, etc.).

In economics, education is usually described as an investment that pays off in the future through higher wages. However, if people self-select into social categories where certain types of behavior are desirable, e.g., due to category-specific norms, the incentives underlying effort choices may differ substantially from those that follow from standard economic investment-models. Our study departs from a model of educational choice and social identity presented in Akerlof and Kranton (2002), where study effort depends on such category-specific norms.

We consider a model with two productivity-types, where individual productivity is private information, along the lines of Stern (1982) and Stiglitz (1982). Public education is modeled as a publicly provided input good (financed through the tax system). Each individual lives for two periods; attains education in the first and earns labor income in the second. This model is here extended to accommodate social identity by allowing each individual, when young, to select into one of two social groups, which differ with respect to the prescribed study effort. Furthermore, the social norms are themselves endogenous variables.

The model assumes that individuals differ in two ways. First, we assume that individuals differ in terms of innate ability, and that this ability is private information. Second, we introduce an element of social inertia by assuming that individuals differ with respect to their preferences for the social group they want to be associated with. Our model contains two such social identity groups in which norms regarding study effort are formed. Individuals will then select into one of the groups depending on ability and preferences, and they will make educational (effort) choices based on their ability and the group norm. We study the optimal tax and expenditure policy for both a welfarist and a paternalist that does not share the consumer preference for social identity.

We find that the consumer preference for social identity leads to corrective as well as redistributive motives for income taxation, and also influences the optimal provision of public education. First, externality correction by a welfarist government leads in itself to an element of tax progression that serves to reduce the discrepancy between the effort norm and the actual effort chosen by low-productivity individuals in the high-effort group. In turn, this effect is counteracted by the government's desire to relax the self-selection constraint, since increases in the effort norms contribute to make mimicking less attractive. Second, if the preference for social identity is sufficiently strong, increased wage-inequality leads to higher social welfare through a relaxation of the selection constraint. It may thus be desirable to use publicly provided education to induce more wage-inequality than in the absence of self-categorization, even if such inequality increases the intrinsic utility of a potential mimicker. Third, the policy rule for publicly provided education does not depend on whether the government has a welfarist or paternalist objective. However, the marginal tax policy differs between the two types of government in a fundamental way, primarily because the paternalist government has a motive to correct effort choices to "undo" the effects of self-categorization.

Paper [II]: Intergenerational Mobility in Sweden: a Regional Perspective

This paper analyzes regional differences of intergenerational income mobility in Sweden. My data set allows me to analyze national and regional mobility measures very precisely for the Swedish population born between 1968 and 1976. Income mobility refers broadly to the extent child income is associated with parent income. The by far most commonly employed mobility measure in the literature is the intergenerational elasticity (IGE). This is simply the slope parameter of a regression of log lifetime income of generation t on log lifetime income of generation $t-1$. A small IGE means that it is harder to predict child income using parent income, and that income mobility is higher. Estimates of the IGE in the literature center around 0.4 with higher estimates for the US, and usually smaller estimates for the European and especially the Nordic countries (see Björklund & Jäntti, 1997; Solon, 1992; Solon, 1999; Solon, 2004; Mazumder, 2005).

In addition to traditional IGE measures, I compute regional and national measures of mobility based on income ranks. For the regional analysis I employ two different mobility measures based on income ranks. The first one is "relative mobility" which describes the mean difference in outcomes between children with parents in the top and bottom of the income distribution, respectively. The second one is "upward mobility" which measures the mean absolute outcome of children from families with below-median income levels and, importantly, focuses exclusively on the regional differences in outcomes of children in the poorer half of the population.

The geographical unit I use in the regional analysis is the "local labor market", which is an aggregation of municipalities defined by commuting patterns. In comparison to the commuting zones in the US, there is much more variation between different Swedish local labor markets in terms of population size. I show that this aspect of the data results in imprecise estimates. To remedy this problem, I propose a joint estimation technique using maximum likelihood, referred to as a multilevel (or hierarchical) model. The multilevel model allows me to make a comparison between the different regional mobility measures in a statistically rigorous way. For example, I can test if the mobility estimate of one particular region is statistically significantly different from the national average. For completeness however, I also report and discuss results based on separate OLS regressions by region.

I find that relative mobility is quite homogeneous across Sweden. The difference of mean son income rank between families at the very top and the very bottom of the income distribution, respectively, is 22.2 percentile ranks in most local labor markets. Only 9 areas out of 112 show significantly lower or higher relative mobility. Stockholm ranks at the bottom with the lowest relative mobility, and the Umeå region in northern Sweden shows the highest relative mobility.

Upward mobility, the expected outcome for sons from below-median income families, varies considerably more across Swedish local labor market areas, from 36.32 percentile ranks in Torsby to 50.77 in Hylte. This corresponds to an income difference of 32.842 SEK per year (≈ 3.839 USD). In addition, children who spend a significant part of their childhood in very rural areas of Sweden generally have significantly worse outcomes compared to children growing up in urban areas. This result can be explained in part by the large fraction of children from rural areas that do not move into cities as adults. However, those who do move, do on average even better than the city natives.

Sweden is considered to be a country with exemplary high levels of intergenerational income mobility. My results show that there are large differences in terms of mobility across Sweden and that location does matter. The evidence provided here indicates that there are particularly large differences in the expected outcomes for children from the lower half of the income distribution, depending on childhood region and moving patterns. A general lesson of this study is that country-wide measures of income mobility might say very little about the state of mobility at a particular location within the country. Cross-country comparisons of income mobility should therefore be interpreted with caution.

Paper [III]: The Effect of Moving during Childhood on Long Run Income: Evidence from Swedish Register Data

In Paper [III] I use Swedish register data for nine cohorts born 1968 to 1976 to study the effect of moving between regions during childhood on long run adult income. In addition, I analyzed the effect of the timing and quality of the move. Childhood is a very important phase during human capital development, since many skills that are capitalized during adulthood are obtained early in life. As Cunha & Heckman (2007) pointed out, skills produced at an earlier point in life will not only augment the skills attained at a later stage, but also increase the returns to later skill investments.

Recent evidence stresses the importance of neighborhood quality for children's long run outcomes (see Chetty et al., 2014; Chetty & Hendren, 2015; Chetty et al. , 2016; Heidrich, 2015). Identifying certain neighborhoods that are better or worse for a child's development leads, however, to another question: what is the effect of moving from one neighborhood to another? If neighborhoods matter, moving to a better area can be thought to be beneficial for a child. However, there might also be disruption costs from moving.

The literature on childhood human capital production suggests that shocks to the skill formation process can result in large differences in the stock of adult human capital. Moving is one such potential shock. The child is taken out of her usual environment and needs to re-build a social identity at the new home, get to know her new class mates, teachers, and neighborhoods. Research in sociology for example indicates negative effects of moving on academic performance (Hagan, 1996). Early-intervention program evaluations have shown that the childhood environment strongly affects children's non-cognitive skills and social attachment, which are both very important for adult outcomes (Heckman & Carneiro, 2003).

I define moving costs as the effect of childhood moving on long run adult income. I show not only that such costs exist and that their size is increasing in the number of times a child moves, but also that the timing of the move matters. The results are well in line with predictions from the human capital literature where shocks in the skill formation process can accumulate to large differences in adult outcome (see for instance Cunha & Heckman, 2007; Currie & Almond, 2011; Heckman, 2006; Heckman et al., 2013).

The main difficulty when trying to measure the costs of moving is selection. In observational data, we will suspect differences between families that move different numbers of times, as well as non-randomness in the type of location families move to and from. Parent income, for instance, is an important factor to take into consideration since higher income families are able to better compensate for disruption effects suffered by the child. Other transitions in family life that occur at a similar time as the move have the potential to confound the true effect of moving. Parental separation or unemployment are two common reasons to move, which, however, also affect children's outcomes independently of moving. The richness of my data allows me to control for many factors that would otherwise bias the results, such as parent unemployment and separation during childhood, immigration background, and education level. An additional aspect to consider is that children can be expected to have only a minor say in when or where to move. That is, children are usually tied to their parents which suggests that any selection effect left can reasonably be assumed to be smaller compared to studying the effect of migration on adults (even if some families move because of their child).

My results suggest significant negative effects of moving during childhood on long run adult income. The effects differ by gender and parent income. Studying family moves between Swedish counties instead of municipalities (that is, moves that cover on average longer distances), I find that the costs of moving on child long run outcome are very robust to the geographic unit chosen.

Focusing on a reduced sample containing only families that move exactly once during childhood, I can control for the type of move, i.e. if families move to areas that are relatively better or relatively worse for the child compared to their region of origin. The quality of a region is computed as the average long run adult income of similar children who spent their entire childhood in this region. I find that the negative effect of childhood moving on adult income is increasing in the age at the time of the move. I also show that moving to a region with a higher quality increases a child's long run outcome, given that the family moves. However, sons have to move before the age of 12 in order to benefit from the better region. From age 12, the quality of the new region has no significant effect. Daughters have to move before the age of 16 in order to benefit from a better quality region. In general, the effects on daughters' adult income are smaller compared to sons.

Paper [IV]: A Study of the Missing Data Problem for Intergenerational Mobility using Simulations

Paper [IV] is a simulation study dealing with the measurement of income when estimating the intergenerational elasticity (IGE). In order for the IGE to be unbiased, the researcher needs to observe the income histories over the complete life cycle for two linked generations. In practice, this is impossible today since income records have not been kept for a long enough time span yet. I borrow a well-established income process from the macroeconomics literature to simulate the complete income trajectories for two generations. Based on the incomes generated from this income process I study the bias of the IGE for different missing-data scenarios.

Snapshots of individual incomes are not necessarily a good indicator of permanent or average life time income. Individuals usually experience both a certain pattern of income growth depending on age, education, and other variables, as well as more or less persistent income shocks. Thus, in order for the IGE to be meaningful, some measure of long run income must be employed. The early studies in intergenerational mobility used just the few available income snapshots for sons and fathers to estimate the IGE. Later studies acknowledged that using income snapshots will not give an accurate description of intergenerational mobility, but the deviations between annual and long run average income were assumed to be random. However, as Nybom et al. (2016) pointed out, the shape of the child's income path over the life cycle depends on education and other background variables, which are related to their parents.

My simulated earnings data reproduces all earlier findings in the literature regarding the bias of the IGE, such as left-hand side measurement error which is negative when son income is observed at young ages, and positive when son income is observed at old ages. I show that having left-hand side measurement error based on “nearly complete” income histories of the two generations as in Nybom et al. (2016) still significantly underestimates the IGE. As long as incomes for sons at older ages (and for fathers at young ages) are missing, the estimates will be severely biased even though there are around 30 observations available for each generation.

I find that, in principle, there exist many different age combinations at which the IGE will be unbiased. As a general rule, IGE estimates are only very little biased when income observations for son and father at the same age are used. This is a consequence of the assumption in my data generating process that the income paths of son and father are correlated, by imposing family-specific returns to experience. The fact that the shape of the income profile over the life cycle depends on family background variables has been demonstrated by Nybom et al. (2016) based on Swedish register data.

Moreover, I find no evidence that using the average over a large number of father income observations diminishes attenuation bias, contrary to the result in Solon (1992). This is possibly due to the fact that Solon assumed the deviations between annual and average lifetime income to be independent random errors, which does not hold in general.

I also find that the rank-rank slope (the association between the average life time income ranks of fathers and sons) is much less sensitive to missing data than the IGE.

In the last part of the analysis, I mimic the typical data that is available to a researcher attempting to estimate the IGE or the rank-rank slope. Income is observed for the son and the father generation during the same years and thus during very different ages. Parents get children at different ages which implies that the age of the fathers during the years for which income observations are available differs greatly. I find that using the average of all available log income observations for sons and fathers (as traditionally done in the literature) leads to IGE estimates that are on average 41 percent too small. This approach thus heavily overstates income mobility. Using only the last available observation for the son and only the earliest available observation for the father gives IGE estimates that are on average 1 percentage point less biased. Using a weighted average of the available observations where the later (earlier) observations for the son (father) are weighted more heavily, reduces the bias by two more percentage points.

Using income ranks, the true value of the rank-rank slope is underestimated by just 5 percent when using the simple average. The first-last observation method yields equally good results. A weighted average of the available income ranks is found to perform best, underestimating the true value by just 4 percent.

Bibliography

Akerlof, G. A., & Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, *40*(4), 1167–1201.

Benabou, R., & Ok, E. A. (2001). Social Mobility and the Demand for Redistribution: The Poum Hypothesis. The Quarterly Journal of Economics, 116(2), 447–487. http://doi.org/10.1162/00335530151144078

Black, S. E., & Devereux, P. J. (2011). Recent Developments in Intergenerational Mobility. In Handbook of Labor Economics (Vol. 4, Part B, pp. 1487–1541). Elsevier. http://doi.org/http://dx.doi.org/10.1016/S0169-7218(11)02414-2

Björklund, A., & Jäntti, M. (1997). Intergenerational income mobility in Sweden compared to the United States. *The American Economic Review*, *87*(5), pp. 1009–1018. http://www.jstor.org/stable/2951338

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, 129(4), i1.

Chetty, R., & Hendren, N. (2015). The impacts of neighborhoods on intergenerational mobility: Childhood exposure effects and county-level estimates. Working paper.

Chetty, R., Hendren, N., & Katz, L. (2016). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Project. American Economic Review, 106(4).

Conti, G., Heckman, J. J., & Urzua, S. (2010). The Education-Health Gradient. American Economic Review: Papers & Proceedings, 100, 234–238. doi:10.1257/aer.100.2.234

Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *The Journal of Economic Perspectives*, 27(3), pp. 79–102.

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review*, *97*(2), 31–47. doi:10.1257/aer.97.2.31

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

Dworkin, R. (1981a). What is Equality? Part 1: Equality of Welfare. Philosophy & Public Affairs, 10(3), pp. 185–246.

Dworkin, R. (1981b). What is Equality? Part 2: Equality of Resources. Source: Philosophy & Public Affairs, 10(4), 283–345.

Friedman, M., & Friedman, R. D. (1979). Free to choose: a personal statement (1st ed.). Hartcourt Brace Jovanovich.

Hagan, J., MacMillan, R., & Wheaton, B. (1996). New kid in town: Social capital and the life course effects of family migration on children. *American Sociological Review*, *61*, 368–385.

Heckman, J. J., & Carneiro, P. (2003). Human capital policy. In A. Krueger & B. Friedman (Eds.), *Inequality in America: What Role for Human Capital Policy?* MIT Press.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900–1902.

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, *103*(6), 2052–2086. doi:10.1257/aer.103.6.2052

Heidrich, S. (2015). Intergenerational mobility in Sweden: A regional perspective. *Umeå Economic Studies*, *916*.

Hogg, M. A., & Reid, S. A. (2006). Social identity, self-categorization, and the communication of group norms. *Communication Theory*, *16*(1), 7–30. doi:10.1111/j.1468-2885.2006.00003.x

Jäntti, M., & Jenkins, S. P. (2015). Income Mobility. In A. B. Atkinson & F. Bourguignon (Eds.), Handbook of Income Distribution (Vol. 2, pp. 807–935). Elsevier. http://doi.org/http://dx.doi.org/10.1016/B978-0-444-59428-0.00011-4

Mazumder, B. (2005). Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data. *Review of Economics and Statistics*, *87*(2), 235–255.

Nybom, M., & Stuhler, J. (2016). Heterogeneous income profiles and lifecycle bias in intergenerational mobility estimation. *Journal of Human Resources*, *51*(1), 239–268. doi:10.3368/jhr.51.1.239

OECD. (2011). An Overview of Growing Income Inequalities in OECD Countries: Main Findings.

Rawls, J. (1971). A Theory of Justice. Cambridge, Mass.: Harvard University. Cambridge, Mass.: Harvard University.

Roemer, J. E. (1993). A Pragmatic Theory of Responsibility for the Egalitarian Planner. *Philosophy & Public Affairs*, 22(2), pp. 146–166

Roemer, J. E. (1996). Theories of distributive justice. Harvard University Press.

Sen, A. (1985). The Moral Standing of the Market. *Social Philosophy and Policy*, 2(02), 1. http://doi.org/10.1017/S0265052500003186

Solon, G. (1992). Intergenerational income mobility in the United States. *The American Economic Review*, 393–408.

Solon, G. (1999). Intergenerational mobility in the labor market. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3, Part A, pp. 1761–1800). Elsevier. doi:http://dx.doi.org/10.1016/S1573-4463(99)03010-2

Solon, G. (2004). A model of intergenerational mobility variation over time and place. In M. Corak (Ed.), *Generational income mobility in North America and Europe* (pp. 38–47). Cambridge University Press.

Soltow, L. (1965). Toward income equality in Norway. Madison, U. of Wisconsin P.

Stern, N. (1982). Optimum taxation with errors in administration. *Journal of Public Economics*, *17*(2), 181–211. doi:10.1016/0047-2727(82)90019-6

Stiglitz, J. E. (1982). Self-selection and pareto efficient taxation. *Journal of Public Economics*, *17*(2), 213–240.

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. *The Social Psychology of Intergroup Relations*, *33*(47).

I

SOCIAL IDENTITY, TAXATION, AND PUBLICLY PROVIDED EDUCATION*

THOMAS ARONSSON, STEFANIE HEIDRICH, AND MAGNUS WIKSTRÖM

ABSTRACT

This paper analyzes the implications of social identity and self-categorization for optimal redistributive income taxation. A two-type model is supplemented by an assumption that individuals select themselves into social categories, in which norms are formed and education effort choices partly depend on these norms. The results show, among other things, that externality correction by a welfarist government leads to an element of tax progression that serves to reduce the discrepancy between the effort norm and the actual effort chosen by low-productivity individuals in the high-effort group. Furthermore, if the preference for social identity is sufficiently strong, increased wage-inequality leads to higher social welfare through a relaxation of the selection constraint. It may thus be desirable to use publicly provided education to induce more wage-inequality, even if higher wage-inequality increases the intrinsic utility of a potential mimicker.

JEL classification: D03, H21, I21, Z13
Keywords: Optimal income taxation, education, social identity, self-categorization.

* Department of Economics, Umeå School of Business and Economics, Umeå University, 901 87 Umeå, Sweden. E-mail: Thomas.Aronsson@econ.umu.se. The authors would like to thank Spencer Bastani, David Granlund and Ronald Wendner for helpful comments and suggestions. Research grants from the Swedish Research Council (ref 421-2010-1420) are also gratefully acknowledged.

1 Introduction

In psychology and sociology, social identity theory has long been used to explain human behavior (see, e.g., Tajfel and Turner, 1979; Hogg, 2006). Social identity can be defined in terms of how a person's sense of self depends on the group (or groups) which the person associates with (e.g., social reference groups such as family, colleagues, friends, social class, etc.). Akerlof and Kranton (2000) show that social identity theory is useful in the context of economics, and might be suitable for analyzing a variety of issues where standard economic theory is more or less silent. The present paper deals with policy responses to self-categorization in a model of educational choice, where individuals select into social groups and the government redistributes through taxation and public resources devoted to education. This will be described in greater detail below.

In economics, education is usually described as an investment that pays off in the future through higher wages. As such, the interesting tradeoff when choosing effort is that between leisure (or consumption) at present and increased productivity in the future.[1] However, if people self-select into social categories where certain types of behavior are desirable, e.g., due to category-specific norms, the incentives underlying effort choices may differ substantially from those that follow from standard economic investment-models. Our study departs from a model of educational choice and social identity presented in Akerlof and Kranton (2002), where study effort depends on such category-specific norms. To be more specific, they introduce an identity component into the individual's utility function, such that individuals differ in terms of how close their preferences or attributes are to the ideal prescribed by different social identity groups. Also, since the identity component depends on the behavior of other members in the same social reference group, externalities play an important role in the economics of identity. In other words, deviating too much from the prescribed behavior may both imply a drop in one's own utility and (positive or negative) changes in the utility of other members in the same social identity group.[2] Our study examines policy implications thereof. The overall purpose is to analyze the implications of social identity and self-categorization for optimal redistributive income taxation and publicly provided education.

Literature in other areas of economics shows that social norms are important for individual choices as well as for policy outcomes. Social norms (or customs) may be persistent despite that that they lead to lower intrinsic utility for individuals if disobedience is associated with lost reputation (Akerlof, 1980), and may even lead

[1] See, e.g., Becker (1975) and Willis and Rosen (1979).

[2] These externalities also contribute in explaining the peer-group effects on study achievement discussed in the literature on educational outcomes (e.g., Hanushek, 1971; Wolfe, 1977; Evans, Oates and Schwab, 1992; Sacerdote, 2000).

individuals to conform in terms of behavior (Bernheim, 1994).[3] Based on a political economy model Lindbeck, Nyberg, and Weibull (1999) examine redistributive tax-transfer policies under an employment norm that "one should live off one's own work", and assume that the perceived cost to the individual of deviating from this norm decreases with the share of benefit recipients in society. Among other things, they find that the economy can end up in a low-tax equilibrium supported by the employed or high-tax equilibrium supported by transfer recipients. Lindbeck, Nyberg, and Weibull (2003) use a similar model to examine the implications of social insurance when the preference for leisure varies among individuals. They show that endogenous social norms may lead voters to choose less generous benefits than otherwise, thereby counteracting the free-rider problem, and also that a temporary unemployment shock may result in a persistent increase in the number of beneficiaries.[4] Another line of research on social norms and economic behavior refers to interdependent behavior in labor supply choices (e.g., Blomquist, 1993; Aronsson, Blomquist and Sacklen, 1999), showing that norms give rise to feedback effects on labor supply of clear practical relevance for assessing the effects that taxes have on work hours.

However, there is surprisingly little research on the implications of social norms for optimal taxation.[5] An exception is the study by Aronsson and Sjögren (2010) analyzing optimal redistributive taxation in an economy characterized by two social norms in the labor market: a conformity norm for work hours implying that individuals perceive a cost of deviating too much from the choices made by other people (i.e., interdependent labor supply choices), and a participation norm emanating from the argument that one should earn one's living from work, meaning that they combine the two labor market norms discussed above. Another exception is a recent study by Aronsson and Granlund (2015) dealing with tax policy implications of gender norms in a household model with household production. They assume that men and women face different social norms, where men are subject to a market work norm and women to a norm in terms of household work. Our study differs from these earlier studies in several important ways. First, and foremost, we are concerned with the implications of social identity and norms in the

[3] See also Bruvoll and Nyborg (2004), who analyze a model where norm-adherence is connected to self-image.
[4] See Lindbeck (1995) for more informal discussions of social norms and economic behavior.
[5] The importance of social interaction for optimal taxation has been analyzed in other contexts; in particular, in economies where consumers are concerned with their relative consumption; see, e.g., Boskin and Sheshinski (1978), Oswald (1983), Tuomala (1990), Ljunqvist and Uhlig (2000), Dupor and Liu (2003), Aronsson and Johansson-Stenman (2008, 2010), Wendner and Goulder (2008), and Eckerstorfer and Wendner (2013).

context of education choices; not social norms in the labor market. Second, we consider a broader set of policy instruments by analyzing the role of public expenditure on education in addition to an optimal income tax. Third, since the education choice is fundamentally intertemporal, we use a two-period model to analyze how tax and expenditure policy can be used to implement a socially desirable outcome.

Our paper is also related to literature on education and optimal taxation (e.g., Boadway et al., 1996; Bovenberg and Jacobs, 2005; Maldonado, 2008; Guo and Krause, 2013; Jacobs, 2013), which deals with a variety of aspects of redistributive education policy. Yet, none of these studies addresses the policy implications of social identity and social norms. Our intention is to bridge this gap by introducing a social identity component into the education choice and then examine the implications for optimal redistributive taxation and public resources spent on education.

We consider a model with two productivity-types, where individual productivity is private information, along the lines of Stern (1982) and Stiglitz (1982). Such a framework is often used in theoretical literature on optimal redistributive taxation and enables us to integrate corrective and redistributive tax policy in a relatively simple way. Public education is modeled as a publicly provided input good (financed through the tax system). Each individual lives for two periods; attains education in the first and earns labor income in the second. This model is here extended to accommodate social identity by allowing each individual, when young, to select into one of two social groups, which differ with respect to the prescribed study effort. Furthermore, the social norms are themselves endogenous variables, and our assumptions about norm formation are based on research in social psychology emphasizing that norms typically reflect more extreme attitudes within groups than just group-specific mean values which, in turn, further contributes to polarization between groups. This is further discussed in Section 2.

The outline of the study is as follows. In Section 2, a model of individual choice is described along with the decision-problem faced by the government. The optimal tax and expenditure policy is analyzed in Sections 3. Following convention, the model developed in Sections 2 and 3 assumes that the government is welfarist in the sense of accepting all aspects of consumer preferences, including the preference for social identity that leads individuals to select into social groups characterized by conformity norms for study effort. Yet, although the welfarist approach may seem innocuous, it is not obvious to us that a government would accept preferences for social identity. One reason is that self-categorization based on such preferences is likely to reflect family characteristics, at least to some extent, and policy makers may not want factors correlated with family characteristics (or social background in general) to affect outcomes later in life; such influences run counter to the notion of equality of opportunity. Therefore, in Section 4, we consider a case where the government is paternalist in the sense of not respecting consumer preferences for social identity, although it respects the intrinsic utility that each consumer derives from consumption and effort choices. Section 5 summarizes and concludes.

2 The Model

Consider an economy comprising two productivity-types, *l* and *h*, where type *l* will be referred to as the "low-productivity type" and type *h* the "high-productivity type". This is interpretable to mean that type *h* has higher innate ability than type *l*. We also distinguish between two social identity groups, which differ with respect to the prescribed effort (through a group-specific effort norm) during the education period of the individual's life. In the following, we just refer to these groups as the "low-effort" (*L*) and "high-effort" (*H*) group, respectively. The pre-tax wage facing an individual of productivity-type *i* in social identity group *j* depends on (1) the individual's innate ability, (2) the public resources spent on education, and (3) the individual's study effort, and is given by $w_j^i = \theta^i e_j^i$, where e_j^i reflects the effort level during education. The variable $\theta^i = \theta(\xi^i, g)$ is a measure of productivity per unit of effort, which reflects the individual's innate ability, ξ^i, and the public expenditure per student, g.[6] We assume that $\theta_\xi^i = \partial\theta^i / \partial\xi^i > 0$ and $\theta_g^i = \partial\theta^i / \partial g > 0$. Therefore, since $\xi^h > \xi^l$ by our earlier assumptions, we have $\theta^h > \theta^l$. There are N^i individuals of productivity-type *i*, among which $n_L^i \in [0, N^i]$ belongs to social identity group *L* and $n_H^i = N^i - n_L^i$ to social identity group *H*.

Although social groups may be characterized along several dimensions, the only distinction we focus on here is study effort. As indicated above, we assume that membership in social identity group *L* prescribes less study effort than membership in social identity group *H*, ceteris paribus. The main motivation for this approach is simplicity: study effort is a decision-variable in our model. Our set up is, nevertheless, supported by research in sociology, which indicates great influence of peers and social groups on educational aspirations.[7] This is found to be true in general for countries that do not sort students into different school tracks at an early age (which is the case for the US and also for the Nordic countries).[8]

[6] Maldonado (2008) uses a similar wage equation, although he abstracts from public resources spent on education.
[7] See, e.g., Austin and Draper (1984) and Wentzel and Caldwell (1997).
[8] See Buchmann and Dalton (2002).

We further assume that sorting into these two groups is determined by innate study preference. Akerlof and Kranton (2002) use three different social groups in their description of high school students (leading crowd, nerds, and burnouts). The mechanism leading pupils to sort into different groups, however, is rather ad hoc in their paper. "Looks" for example might in practice not be a reliable predictor for one's choice of social group and there is little if not no evidence in the literature to be found of such effects. Thus, in our model, we base the sorting mechanism on empirical research that points to strong transmission of educational attainment from parents to their children.[9] The sorting into identity group *L* or *H* may, therefore, reflect that family background affects young individuals' selection into social identity groups, i.e., the importance attached to study achievement in the individual's environment, as well as other attributes that the individual would like to be associated with.

Consumers

By ranking the individuals of each ability-type on the basis of preferences for social identity-group, from the person with the highest to the person with the lowest preference for social identity-group L (or, equivalently, from the person with the lowest to the person with the highest preference for social identity-group H), the life-time utility function facing individual $k \in [1, N^i]$ of productivity-type i in social identity group j can be written as

$$U_j^{k,i} = u(c_j^i, x_j^i, e_j^i) + I_j^{k,i} \tag{1}$$

where *c* denotes consumption when young, *x* consumption when old, and *e* denotes education effort. The function $u(\cdot)$ represents the intrinsic utility of consumption and effort, and is assumed to be increasing in its first and second arguments, decreasing in the third, and strictly quasi-concave. Following Akerlof and Kranton (2002), the second term on the right hand side of equation (1) represents the identity utility component defined as

$$I_L^{k,i} = I_L - \beta k - \frac{1}{2}(e_L^i - \overline{e}_L)^2 \tag{2a}$$

$$I_H^{k,i} = I_H - \beta(N^i - k) - \frac{1}{2}(e_H^i - \overline{e}_H)^2 \tag{2b}$$

if the individual belongs to social identity group *L* and *H*, respectively. The term I_j represents a fixed component of the payoff associated with social identity *j*. Equations (2a) and (2b) presuppose that individuals of a certain productivity-type differ according to their preferences for social identity, such that individual $k = 1$ has the strongest

9 See, e.g., Ermisch and Francesconi (2001), Anger and Heineck (2010) and Björklund and Jäntti (2012).

preference for being part of social identity group L and weakest preference for social identity group H, and so on. As such, the parameter β reflects how the payoff differs between individuals. The final component is a perceived cost of deviating from the behavior prescribed by the social group, where $\bar{e}_j$ is interpretable as an identity-group-specific norm for study effort. The two types of heterogeneity (ability and social identity preference) are private information. We assume that ability and preference for social identity are uncorrelated, reflecting the view that talent is unrelated to social background. The reader should note that although the preference for social identity is private information, the choice of social identity group will be observable to the policy maker in our setting, meaning that the tax policy rules can be derived conditional on social identity group choice.[10]

We abstract from any initial wealth in what follows. Adding an exogenous initial income or inherited wealth component would not affect any of the qualitative results below, which means that we refrain from such extensions here. Therefore, an individual of productivity-type i in social identity group j, faces the following life-time budget constraint:

$$-s_j^i = c_j^i \tag{3a}$$

$$s_j^i + w_j^i - T_j^i = x_j^i \tag{3b}$$

where s denotes savings. The variable $T_j^i = T(w_j^i, s_j^i)$ denotes a tax payment (positive or negative), and $T(\cdot)$ is a tax payment function. Without loss of generality, the interest rate is set to zero, and tax payments are made based on earnings and savings. Equations (3a) and (3b) mean that the individual finances his/her studies by borrowing against the future income.[11] When employed in the second period, the individual supplies one unit of labor inelastically. Each individual is small relative to the economy as a whole and acts as an atomistic agent in the sense of treating the identity-group-specific norms, i.e., $\bar{e}_j$ for j=L, H, as exogenous. Therefore, and conditional on being part of social identity group j, the individual behaves as if he/she chooses e_j^i and s_j^i to maximize utility given in equation

[10] The two-period framework used here is a simplification in that the first period summarizes all stages of the educational system. To some extent the model therefore hides the possibility that social identity can be more or less important at some stages of youth, and that young people respond differently to economic incentives depending on age. Our purpose is to characterize these two effects (and the interactions between them) qualitatively in terms of tax and expenditure policy, and the results derived in the paper will hold as long as both social identity and economic incentives affect individual choices.

[11] This is a simple way of addressing "study loans", which are used in many countries.

(1) subject to the budget constraint in equations (3a) and (3b). The first order conditions are

$$u^i_{j,e} - (e^i_j - \bar{e}_j) + u^i_{j,x}\theta^i(1 - T^i_{j,w}) = 0 \tag{4a}$$

$$-u^i_{j,c} + u^i_{j,x}(1 - T^i_{j,s}) = 0 \tag{4b}$$

in which the second subscript attached to the utility function denotes partial derivative, i.e., $u^i_{j,e} = \partial u(c^i_j, x^i_j, e^i_j)/\partial e^i_j$, $u^i_{j,c} = \partial u(c^i_j, x^i_j, e^i_j)/\partial c^i_j$ and $u^i_{j,x} = \partial u(c^i_j, x^i_j, e^i_j)/\partial x^i_j$, while $T^i_{j,w}$ and $T^i_{j,s}$ denote marginal income and savings taxes. The choice of social identity group is then based on utility comparisons between regimes *L* and *H*.

Finally, we assume that equations (2a) and (2b) are such that, for each productivity-type, there is a marginal individual, who is precisely indifferent between the two social identity groups. For productivity-type *i*, this means that

$$\begin{aligned} &u(c^i_L, x^i_L, e^i_L) + I_L - \beta n^i_L - \frac{1}{2}(e^i_L - \bar{e}_L)^2 = \\ &u(c^i_H, x^i_H, e^i_H) + I_H - \beta(N^i - n^i_L) - \frac{1}{2}(e^i_H - \bar{e}_H)^2 \end{aligned}. \tag{5}$$

Equation (5) implicitly defines the number of members of social identity group *L* of productivity-type *i*, n^i_L, as a function of variables characterizing both social identity groups.

To analyze the decision-problem faced by the government, we must specify how the effort norms are determined. Theoretical literature dealing with the policy implications of social comparisons often assumes that people compare their own choices with an in-group mean value[12]. However, research within the social psychology literature suggests that group norms may be more extreme than those based on group-specific averages. One of the first studies on group polarization is Moscovici and Zavalloni (1969) who studied group discussion to consensus and found group judgments to be more extreme than individual attitudes. Tajfel et al. (1971) found that group members tend to follow a maximum difference strategy (vs. an out-group), even at the price of sacrificing other advantages.[13] A review of this literature can be found in Cooper et al. (2001), who also discuss the underlying mechanisms leading to group polarization as well as the relationship between polarization and social norms in general.

To capture the bias away from the in-group averages towards the extremes in a simple way, we consider a modal value comparison by assuming that that the majority of

---

[12] This is the case in much of the literature on optimal taxation in models with relative consumption concerns referred to in the introduction.

[13] Further examples of studies with similar results on group polarization are Turner et al. (1971) and McGarty et al. (1992).

members in social identity group L is of the lower productivity-type, and the majority of members in social identity group H is of the higher productivity-type, such that

$$\bar{e}_L = e_L^l \text{ and } \bar{e}_H = e_H^h . \tag{6}$$

To ensure that e_L^l and e_H^h represent extreme effort choices, we also assume that the inequalities $e_L^l < e_L^h, e_H^l < e_H^h$ hold.[14] Equations (5) and (6) then imply that the number of individuals in social identity group L of productivity-type l and h, respectively, can be written as follows:

$$n_L^l = n_L^l(\overset{+}{c_L^l}, \overset{+}{x_L^l}, \overset{-}{e_L^l}, \overset{-}{c_H^l}, \overset{-}{x_H^l}, \overset{+}{e_H^l}, \overset{+}{\bar{e}_H}) \tag{7a}$$

$$n_L^h = n_L^h(\overset{+}{c_L^h}, \overset{+}{x_L^h}, \overset{-}{e_L^h}, \overset{-}{c_H^h}, \overset{-}{x_H^h}, \overset{+}{e_H^h}, \overset{+}{\bar{e}_L}) , \tag{7b}$$

where the sign above each argument denotes partial derivative. The corresponding number of individuals in social identity group H can then be analyzed simply by recalling that $n_H^i = N^i - n_L^i$ for $i=l, h$.

Social Decision-Problem

The government (or social planner) is assumed to observe income and savings at the individual level, whereas individual productivity is private information. A nonlinear tax attached to earnings and savings means that the government can implement any desired combination of consumption and effort (subject to informational limitations, see below). Therefore, we follow convention in the literature on optimal nonlinear taxation and write the public decision-problem as a direct decision-problem, where the government directly decides upon consumption and effort. The optimal marginal tax policy through which this desired resource allocation can be implemented in a decentralized economy is derived by comparing the first order conditions of the public decision-problem with the individuals' first order conditions for education effort and savings. The government is also assumed to recognize how the social norms are determined, i.e., according to equations (6), and treat them as endogenous.[15]

We consider a Pareto efficient resource allocation, where the government maximizes utility for one sub-group, e.g., low-productivity individuals belonging to social identity group L, subject to minimum utility restrictions for all other sub-groups. The only difference in preferences between individuals of the same productivity-type arises through the identity utility defined in equations (2a) and (2b), according to which

[14] This always holds if the utility function is quasi-linear in the second period consumption.

[15] For purposes of comparison, we will also briefly discuss the case where the government treats the identity-group-specific norms as exogenous (see Section 3 below).

individuals differ in their preferences for social identity groups L and H. Therefore, and conditional on group-choice, we can suppress constant terms and write the utility of productivity-type i in social identity group j as

$$U_j^i = u(c_j^i, x_j^i, e_j^i) - \frac{1}{2}(e_j^i - \bar{e}_j)^2, \tag{8}$$

which is the utility component of any such i-j individual that the government may directly affect through tax policy.

We also assume that the government wants to redistribute from high-productivity to low-productivity individuals. Therefore, since productivity is private information, the government must prevent high-productivity individuals from mimicking low-productivity individuals. This is accomplished by introducing the self-selection constraints

$$U_j^h \geq u\left(c_j^l, x_j^l, \frac{\theta^l}{\theta^h} e_j^l\right) - \frac{1}{2}\left(\frac{\theta^l}{\theta^h} e_j^l - \bar{e}_j\right)^2 = \hat{U}_j^h \text{ for } j{=}L, H. \tag{9}$$

If (9) is satisfied, none of the individuals of productivity-type h has an incentive to become a mimicker, irrespective of the strength of the preference for belonging to social identity group j. In other words, each individual will reveal his/her true type regardless of preferences for social identity. The left hand side of the weak inequality is the utility of the true high-productivity individual, while the right hand side is the utility of the mimicker for whom the hat symbol is attached to the utility function. The variable $\theta^l / \theta^h < 1$ measures the relative productivity, meaning that $\hat{e}_j^h = (\theta^l / \theta^h) e_j^l < e_j^l$ denotes the mimicker's education effort: although the mimicker earns as much income and consumes the same amount as a low-productivity individual, the mimicker needs to exert less effort to reach this income.

The government collects revenue for purposes of redistribution and public consumption. By using the government's budget constraint,

$\sum_i \left[n_L^i T(w_L^i, s_L^i) + (N^i - n_L^i) T(w_H^i, s_H^i) \right] - Ng = 0$, where $N = N^l + N^h$, together with the private budget constraints, the economy's resource constraint becomes

$$\sum_i \left[n_L^i (w_L^i - c_L^i - x_L^i) + (N^i - n_L^i)(w_H^i - c_H^i - x_H^i) \right] - Ng = 0. \tag{10}$$

The resource constraint means that income is used for private consumption and public education.

The decision-problem for the welfarist government can then be written as (if we assume that the government attempts to maximize the utility of the low-ability type in social identity group L)

$$\underset{c_L^l, x_L^l, e_L^l, c_L^h, x_L^h, e_L^h, c_H^l, x_H^l, e_H^l, c_H^h, x_H^h, e_H^h, g}{Max} \quad U_L^l$$

$$\text{s.t. } (i) U_H^l \geq \bar{U}_H^l, U_L^h \geq \bar{U}_L^h, U_H^h \geq \bar{U}_H^h,$$

$$(ii) U_L^h \geq \hat{U}_L^h, \; U_H^h \geq \hat{U}_H^h,$$

$$(iii) \sum_i \left[n_L^i (w_L^i - c_L^i - x_L^i) + (N^i - n_L^i)(w_H^i - c_H^i - x_H^i) \right] - Ng = 0 \text{ for } i = i, h,$$

$$(iv) \text{ equations } (6) \text{ and } (7).$$

where $\bar{U}_H^l$, $\bar{U}_L^h$ and $\bar{U}_H^h$ are minimum utility restrictions. The Lagrangean of the public-decision problem can then be written as

$$\begin{aligned} \mathcal{L} = U_L^l + \delta_L^h \left(U_L^h - \bar{U}_L^h \right) + \sum_{i=l}^{h} \delta_H^i \left(U_H^i - \bar{U}_H^i \right) + \sum_{j=L}^{H} \lambda_j \left(U_j^h - \hat{U}_j^h \right) \\ + \gamma \left[\sum_i \left\{ n_L^i (w_L^i - c_L^i - x_L^i) + (N^i - n_L^i)(w_H^i - c_H^i - x_H^i) \right\} - Ng \right] \end{aligned} \tag{11}$$

where δ_j^i, λ_j and γ are Lagrange multipliers. The first order conditions are presented in the Appendix.

3 Optimal Taxation and Public Education

Throughout, we focus on income taxes and do not present any results for marginal savings taxes in the main text. The reason is that the social identity choices made by the consumers directly affect the optimal marginal income tax rates, while they have no direct influence on the policy incentives underlying the marginal savings taxes. As such, in a first best setting where the self-selection constraints do not bind, the government would not use marginal savings taxes. If the self-selection constraints bind, the marginal savings tax would still be zero for high-productivity individuals, while it would be positive (negative) for low-productivity individuals depending on whether low-productivity individuals have a stronger (weaker) preference for early consumption compared to the corresponding mimicker. These policy incentives have nothing to do with social identity and social norms and are well understood from earlier research (see, e.g., Brett, 1997). We justify this argument by presenting the marginal savings taxes in the Appendix.

First Best Taxation

To simplify the presentation, consider first the special case where individual productivity is observable. In terms of the model set out above, this special case means that the self-selection constraints become redundant and $\lambda_L = \lambda_H = 0$. It also provides a suitable starting point, as it implies that the government can redistribute through productivity-specific lump-sum taxes. The only reason for distorting the effort choice is thus to correct for externalities. Therefore, the optimal marginal income tax rates will solely reflect (i) the welfare contributions of the social norms, and (ii) how each productivity-identity group affects these norms through effort choices. The welfare contribution of each social norm can be derived by differentiating the Lagrangean in equation (11) with respect to $\overline{e}_L$ and $\overline{e}_H$, respectively, as follows:

$$\frac{\partial \mathcal{L}}{\partial \overline{e}_L} = \delta_L^h (e_L^h - \overline{e}_L) + \gamma \frac{\partial n_L^h}{\partial \overline{e}_L}\left(G_L^h - G_H^h\right) \tag{12a}$$

$$\frac{\partial \mathcal{L}}{\partial \overline{e}_H} = \delta_H^l (e_H^l - \overline{e}_H) + \gamma \frac{\partial n_L^l}{\partial \overline{e}_H}\left(G_L^l - G_H^l\right) \tag{12b}$$

where $G_j^i = w_j^i - c_j^i - x_j^i$ is the net contribution to public revenue by productivity-type i in social identity group j, while equations (7a) and (7b) imply $\partial n_L^h / \partial \overline{e}_L > 0$ and $\partial n_L^l / \partial \overline{e}_H > 0$ based on our earlier assumptions. We base most of our interpretations below on the additional (and reasonable) assumption that for each productivity-type, individuals in social identity group H contribute more to the tax revenue than individuals in social identity group L, such that $G_H^i > G_L^i$ for i=l, h.

We have derived the following result:

Proposition 1. *In a first best setting where* $\lambda_L = \lambda_H = 0$, *the marginal income tax policy can be characterized as*

$$T_{L,w}^l = -\frac{\partial \mathcal{L}}{\partial \overline{e}_L}\frac{1}{\gamma n_L^l \theta^l}, \quad T_{L,w}^h = 0, \quad T_{H,w}^l = 0 \quad \textit{and} \quad T_{H,w}^h = -\frac{\partial \mathcal{L}}{\partial \overline{e}_H}\frac{1}{\gamma n_H^h \theta^h},$$

where $\partial \mathcal{L} / \partial \overline{e}_L$ *and* $\partial \mathcal{L} / \partial \overline{e}_H$ *are given by equations (12).*

Proof: see the Appendix.

Note that it is only the low-productivity type in social identity group L and high-productivity type in social identity group H that generate externalities. As such, it is only the education choices in these two sub-groups that will be distorted in the first best optimum. Given the assumptions set out above, we have $e_H^l < \overline{e}_H$. Therefore, if $G_H^l > G_L^l$, it follows that $\partial \mathcal{L} / \partial \overline{e}_H < 0$, since an increase in $\overline{e}_H$ leads to lower utility for low-productivity individuals in social identity group H and to lower tax revenue through an increase in the number of high-productivity individuals that select into social identity group *L*. This means that $T^h_{H,w} > 0$, suggesting that the incentives faced by the government to internalize externalities generate an element of tax progression. However, $\partial\mathcal{L} / \partial\overline{e}_L$ can be either positive or negative because an increase in $\overline{e}_L$ leads to higher utility for high-productivity individuals in social reference group *L* (since $e^h_L > \overline{e}_L$ by our earlier assumptions) and to lower tax revenue (if $G^h_H > G^h_L$). As a consequence, $T^l_{L,w}$ may be either positive or negative at the optimum depending on which effect dominates.

Finally, notice that the only reason for a welfarist government to distort the individual's education choice in a first best setting is to influence $\overline{e}_L$ and $\overline{e}_H$. This is seen from the following corollary to Proposition 1, which characterizes the marginal tax policy that would follow in the special case where the government treats $\overline{e}_L$ and $\overline{e}_H$ as exogenous:

Corollary 1. *In a first best setting, and if the government treats* $\overline{e}_L$ *and* $\overline{e}_H$ *as exogenous, the marginal tax rates on earnings are zero, i.e.,*

$$T_{w^l_L} = T_{w^l_H} = T_{w^h_L} = T_{w^h_H} = 0.$$

The interesting thing to note here is that Corollary 1 applies despite that the selection into social identity groups is endogenous. In other words, a government that treats the identity-group-specific norms as exogenous has no incentive to influence the selection into social identity groups by taxing earnings. As we will see below, this result does not apply for a paternalist government, which would like the individuals to behave as if social identity were of no concerns for them.

Second Best Taxation

With Proposition 1 at our disposal, we are now ready to examine the implications of social identity choices for optimal second best taxation. To shorten the notation, let

$$MRS^i_{j,ex} = \frac{u^i_{j,e} - \left(e^i_j - \overline{e}_j\right)}{u^i_{j,x}} \text{ and } M\hat{R}S^h_{j,ex} = \frac{\left(\hat{u}^h_{j,e} - \left(\frac{\theta^l}{\theta^h} e^l_j - \overline{e}_j\right)\right)}{\hat{u}^l_{j,x}} \tag{13}$$

denote the marginal rate of substitution between effort and second period consumption for productivity-type *i* and the mimicker, respectively, in social identity group *j*.

With binding self-selection constraints, the partial welfare effect of an increase in $\overline{e}_L$ and $\overline{e}_H$, respectively, extends to read

$$\frac{\partial \mathcal{L}}{\partial \overline{e}_L} = \delta_L^h (e_L^h - \overline{e}_L) + \lambda_L \left(e_L^h - \frac{\theta^l}{\theta^h} e_L^l \right) + \gamma \frac{\partial n_L^h}{\partial \overline{e}_L} \left(G_L^h - G_H^h \right) \tag{14a}$$

$$\frac{\partial \mathcal{L}}{\partial \overline{e}_H} = \delta_H^l (e_H^l - \overline{e}_H) + \lambda_H \left(e_H^h - \frac{\theta^l}{\theta^h} e_H^l \right) + \gamma \frac{\partial n_L^l}{\partial \overline{e}_H} \left(G_L^l - G_H^l \right). \tag{14b}$$

The difference between equations (12) and (14) is the second term on the right hand side of equations (14a) and (14b): each such component is positive, meaning that an increase in each social norm contributes to relax one of the self-selection constraints. This is so because an increase in $\overline{e}_L$ leads to lower utility (through greater effort) for the mimicker in social identity group *L*, ceteris paribus, while an increase in $\overline{e}_H$ contributes to increase the distance between the effort norm and the mimicker's effort in social identity group *H*. The second best optimal tax policy is presented in Proposition 2:

Proposition 2. *In a second best setting where the self-selection constraints bind, the marginal income tax policy can be characterized as*

$$T_{L,w}^l = -\frac{\lambda_L \hat{u}_{L,x}}{\gamma n_L^l \theta^l} \left(MRS_{L,ex}^l - \frac{\theta^l}{\theta^h} M\hat{R}S_{L,ex}^h \right) - \frac{\partial \mathcal{L}}{\partial \overline{e}_L} \frac{1}{\gamma n_L^l \theta^l},$$

$$T_{L,w}^h = 0,$$

$$T_{H,w}^l = -\frac{\lambda_H \hat{u}_{H,x}}{\gamma n_H^l \theta^l} \left(MRS_{H,ex}^l - \frac{\theta^l}{\theta^h} M\hat{R}S_{H,ex}^h \right)$$

$$T_{H,w}^h = -\frac{\partial \mathcal{L}}{\partial \overline{e}_H} \frac{1}{\gamma n_H^h \theta^h}$$

where $\partial \mathcal{L} / \partial \overline{e}_L$ *and* $\partial \mathcal{L} / \partial \overline{e}_H$ *are given by equations (14).*

Proof: see the Appendix.

There are two important differences between the tax formulas in Proposition 2 and the corresponding first best policy analyzed in Proposition 1. First, the component

$$-\frac{\lambda_j \hat{u}_{j,x}}{\gamma n_j^l \theta^l} \left(MRS_{j,ex}^l - \frac{\theta^l}{\theta^h} M\hat{R}S_{j,ex}^h \right)$$

in the marginal income tax rate of the low-productivity type in each social identity group is positive, and interpretable as the marginal tax rate that would be implemented for this low-productivity type if $\overline{e}_L$ and $\overline{e}_H$ were exogenous to the government. This tax incentive serves to make mimicking unattractive by exploiting that each low-productivity type and corresponding mimicker differ from one another with respect to the marginal value of leisure. The corresponding component for high-productivity individuals is zero here (because the relative productivity, θ^l / θ^h, is constant). As such, this mechanism is well known from earlier studies (e.g., Stiglitz, 1982).[16] Second, increases in $\bar{e}_L$ and $\bar{e}_H$ contribute to relax the self-selection constraints as explained above, which can be seen from the second term on the right hand side of the expressions for $\partial\mathcal{L} / \partial\bar{e}_L$ and $\partial\mathcal{L} / \partial\bar{e}_H$ given in equations (14): this mechanism works to reduce the marginal tax rates facing the externality generating consumers.

Public Expenditure

In our model, the public resources devoted to education are interpretable as an input in the production of skills. As such, these resources contribute to increased income by boosting individual productivity through increased earnings ability. There is also a redistributive element involved because the productivity increase caused by publicly provided education may vary between individuals, which the government can exploit to relax the self-selection constraint. The cost benefit rule for public expenditure can be written as follows (recalling that $\theta_g^i \equiv \partial\theta^i / \partial g > 0$):

Proposition 3. *The optimal provision of public education satisfies*

$$\sum_i \theta_g^i \sum_j n_j^i e_j^i - N = \sum_j \frac{\lambda_j}{\gamma}\left[\hat{u}_{j,e}^h - \left(\frac{\theta^l}{\theta^h} e_j^l - \bar{e}_j\right)\right] e_j^l \frac{\partial(\theta^l / \theta^h)}{\partial g} .$$

The proof of Proposition 3 follows directly from the first order condition for g and is, therefore, omitted. The left hand side measures the difference between the direct marginal benefit of a higher g (through increased productivity and, therefore, increased output) minus the direct marginal cost, while the right hand side reflects that an increase in g affects the wage distribution and thus welfare via the self-selection constraints.[17] If the self-selection constraints do not bind, i.e., if $\lambda_L = \lambda_H = 0$, in which case the resource

[16] Note that this component would be present also in the absence of any preference for social identity, and works in the direction of regressive taxation, in the sense that the marginal tax rates decline with productivity, ceteris paribus. Guo and Krause (2013) use a model with two productivity-types different from ours, where the individual both supplies work hours and faces direct expenditures on education, and where the policy instruments are labor income taxes and taxes on education expenditure. Although they find that the income tax is regressive in the sense described above, they also find that the optimal education policy is progressive in the sense that the marginal education tax is negative (i.e., a marginal subsidy) for the low-productivity type and zero for the high-productivity type.

[17] Note that we have assumed away production externalities associated with spillover effects of skill formation. Adding such a component to the model would magnify the first term on the left hand side (i.e., increase the direct marginal benefit of public education), ceteris paribus, while the other terms would remain unchanged. As such, this simplification does not change the interpretations discussed below.

allocation is first best, the optimality condition for public expenditure just implies that the direct marginal benefit in terms of the increased output equals the direct marginal cost, i.e., the first best policy rule reads $\sum_i \theta_g^i \sum_j n_j^i e_j^i = N$. Therefore, in a first best setting, consumer preferences for social identity will have no direct effect on the optimal provision of public education; instead, the income tax is adjusted to internalize the associated externalities.

In a second best framework where the self-selection constraints bind, on the other hand, this distinct pattern of adjustment no longer applies. The expression in square brackets on the right hand side can be either positive or negative, as an increase in the relative wage (i) reduces the intrinsic utility of the mimicker (by requiring greater effort to reach the same income as the low-productivity type), and (ii) increases the identity utility of the mimicker (by reducing the discrepancy between the mimicker's effort and the effort norm). For purposes of interpretation, we begin by assuming that

$$\hat{u}_{j,e}^h - \left(\frac{\theta^l}{\theta^h} e_j^l - \overline{e}_j \right) < 0 \tag{15a}$$

in which case an increase in the relative wage leads to lower utility for the mimicker. Proposition 3 then implies that the government may relax the self-selection constraint by implementing policies that lead to lower wage-inequality. We should thus over-provide public education relative to the first best policy rule if increased public expenditure leads to less wage-inequality, and under-provide public education relative to the first best policy rule if increased public expenditure leads to more wage-inequality, ceteris paribus. Yet, we cannot a priori rule out the alternative scenario where an increase in the relative wage rate leads to such a large increase in the mimicker's identity utility that the mimicker's overall utility increases (despite the lost intrinsic utility), i.e.,

$$\hat{u}_{j,e}^h - \left(\frac{\theta^l}{\theta^h} e_j^l - \overline{e}_j \right) > 0 \,. \tag{15b}$$

Inequality (15b) means that increased wage-inequality is associated with a net social benefit instead of a net social cost, as it makes mimicking less attractive. Therefore, compared to the first best policy rule described above, this scenario means that we should under-provide public education if it leads to reduced wage-inequality, and over-provide public education if it leads to increased wage-inequality. As a consequence, the qualitative results largely depend on the strength of the preference for social identity.

Taken together, Propositions 2 and 3 give an interesting characterization of the second best optimal policy mix. The government uses the income tax to correct for the externalities that the preference for social identity gives rise to, while the adjustment of public education depends on (a) the effect that public education has on the wage-inequality and (b) whether the mimicker's gain in identity utility following a more equal wage distribution (an increase in the relative wage rate) is large enough to dominate the lost intrinsic utility.

4 Paternalist Policy

A paternalist government does not share the preferences for social identity; instead, this government wants each individual to behave as if the life-time utility coincides with the intrinsic utility of consumption and leisure. Therefore, the paternalist government imposes the following utility on each individual of productivity-type *i* and social identity group *j*:

$$V_j^i = u(c_j^i, x_j^i, e_j^i)\,. \tag{16}$$

The corresponding decision-problem of the paternalist government becomes

$$\underset{c_L^l, x_L^l, e_L^l, c_L^h, x_L^h, e_L^h, c_H^l, x_H^l, e_H^l, c_H^h, x_H^h, e_H^h, g}{Max} \quad V_L^l$$

s.t. $(i)\, V_H^l \geq \overline{V}_H^l, V_L^h \geq \overline{V}_L^h,\ V_H^h \geq \overline{V}_H^h,$

$(ii)\, U_L^h \geq \hat{U}_L^h,\ U_H^h \geq \hat{U}_H^h,$

$(iii) \sum_i \left[n_L^i (w_L^i - c_L^i - x_L^i) + (N^i - n_L^i)(w_H^i - c_H^i - x_H^i) - Ng \right] = 0$ for $i = i, h,$

(iv) equations (6) and (7).

As explained above the welfarist government recognizes the consumer preferences for social identity, while the paternalist government bases its objective on the intrinsic part of the consumers' utility functions (as represented by the function $u(\cdot)$ in equation (1)). Note also that irrespective of whether the government is welfarist or paternalist, the self-selection constraints (constraints (ii) in the decision-problems characterized above) are always based on the actual consumer objectives: the reason is, of course, that these constraints are used to counteract mimicking and must, therefore, reflect the incentives faced by the consumers. Turning to the paternalist government's decision-problem, the Lagrangean is now given by

$$\begin{aligned}\mathcal{L} = V_L^l + \delta_L^h \left(V_L^h - \overline{V}_L^h \right) + \sum_{i=l}^{h} \delta_H^i \left(V_H^i - \overline{V}_H^i \right) + \sum_{j=L}^{H} \lambda_j \left(U_j^h - \hat{U}_j^h \right) \\ + \gamma \left[\sum_i \left\{ n_L^i (w_L^i - c_L^i - x_L^i) + (N^i - n_L^i)(w_H^i - c_H^i - x_H^i) \right\} - Ng \right]\end{aligned}\,. \tag{17}$$

The first order conditions are presented in the Appendix. Since the optimality condition for publicly provided education takes exactly the same form here as it did in Section 3, Proposition 3 remains valid also under a paternalist government. Therefore, we focus on tax policy here.

Based on equation (17), the partial welfare effects of increases in the social norms, i.e., $\overline{e}_L$ and $\overline{e}_H$, can be written as

$$\frac{\partial \mathcal{L}}{\partial \overline{e}_L} = \lambda_L \left(e_L^h - \frac{\theta^l}{\theta^h} e_L^l \right) + \gamma \frac{\partial n_L^h}{\partial \overline{e}_L} \left(G_L^h - G_H^h \right) \tag{18a}$$

$$\frac{\partial \mathcal{L}}{\partial \overline{e}_H} = \lambda_H \left(e_H^h - \frac{\theta^l}{\theta^h} e_H^l \right) + \gamma \frac{\partial n_L^l}{\partial \overline{e}_H} \left(G_L^l - G_H^l \right). \tag{18b}$$

Compared to the welfarist model, equations (18a) and (18b) imply that the welfare effect of an increase in each such norm is decomposed into two (instead of three) components, i.e., the first term on the right hand side of equation (14a) and (14b), respectively, is absent here since the paternalist government does not share the consumer preference for social identity. The remaining terms are identical to their counterparts in the welfarist case. As explained above, the first term on the right hand side of equation (18a) and (18b), respectively, is positive, since an increase in $\overline{e}_L$ makes mimicking less attractive in social identity group *L* by necessitating greater effort of the mimicker, whereas an increase in $\overline{e}_H$ makes mimicking less attractive in social identity group *H* due to lower identity utility for the mimicker. The second term on the right hand side of equation (18a) and (18b), respectively, is negative if $G_L^i < G_H^i$ and positive otherwise for *i*=*l*, *h*.

First Best Policy

As in subsection 3.1, we begin by considering the special case where the self-selection constraints do not bind, i.e., where $\lambda_L = \lambda_H = 0$. Therefore, if $G_L^i < G_H^i$ for *i*=*l*, *h*, equations (18a) and (18b) imply $\partial \mathcal{L} / \partial \overline{e}_L < 0$ and $\partial \mathcal{L} / \partial \overline{e}_H < 0$. Proposition 3 describes the first best marginal tax policy of a paternalist government:

Proposition 4. *If* $\lambda_L = \lambda_H = 0$*, the marginal income tax rates implemented by a paternalist government can be written as*

$$T_{L,w}^l = -\frac{1}{\gamma \theta^l n_L^l} \frac{\partial \mathcal{L}}{\partial \overline{e}_L} = -\frac{1}{\theta^l n_L^l} \frac{\partial n_L^h}{\partial \overline{e}_L} \left(G_L^h - G_H^h \right)$$

$$T_{L,w}^h = -\frac{1}{\theta^h} \frac{(e_L^h - \overline{e}_L)}{u_{L,x}^h} + \frac{1}{\theta^h n_L^h} \left(G_L^h - G_H^h \right) \frac{(e_L^h - \overline{e}_L)}{2\beta}$$

$$T_{H,w}^l = -\frac{1}{\theta^l} \frac{(e_H^l - \overline{e}_H)}{u_{H,x}^l} - \frac{1}{\theta^l n_H^l} \left(G_L^l - G_H^l \right) \frac{(e_H^l - \overline{e}_H)}{2\beta}$$

$$T_{H,w}^h = -\frac{1}{\gamma n_H^h \theta^h} \frac{\partial \mathcal{L}}{\partial \overline{e}_H} = -\frac{1}{\theta^h n_H^h} \frac{\partial n_L^l}{\partial \overline{e}_H} \left(G_L^l - G_H^l \right).$$

Proof: see the Appendix.

Recall from Proposition 1 that the only reason for a welfarist government to distort the effort choice in a first best world is to influence $\overline{e}_L$ and $\overline{e}_H$ (implying a positive marginal tax for the high-productivity type in social identity group *H*, while the marginal tax imposed on the low-productivity type in social identity group *L* could be either positive or negative). The policy incentive to distort the effort choices of *l-L* and *h-H* individuals is present here as well; yet, in modified form, since the paternalist government gives no weight to the identity component of the utility function. Accordingly, if $G_L^i < G_H^i$ for *i=l*, *h*, the first best tax policy of the paternalist government implies $T_{L,w}^l > 0$ and $T_{H,w}^h > 0$. The intuition is that a decrease in $\overline{e}_L$ increases the number of high-productivity individuals in social identity group *H*, and a decrease in $\overline{e}_H$ leads to an increase in the number of low-productivity individuals in social identity group *H*, ceteris paribus, which contribute to increased tax revenue if $G_L^i < G_H^i$. Conversely, if $G_L^i > G_H^i$ the paternalist policy implies $T_{L,w}^l < 0$ and $T_{H,w}^h < 0$; let be that this outcome seems unlikely.

However, contrary to the welfarist government, a paternalist government imposes non-zero marginal income taxes also on the high-productivity type in social identity group *L* and low-productivity type in social identity group *H*. As can be seen from the second and third formulas in the proposition, there are two reasons for this. The first term on the right hand side of the formula for $T_{L,w}^h$ is negative (since $e_L^h > \overline{e}_L$) and represents a pure paternalist motive for subsidizing the income of the high-productivity type in social identity group *L*. The intuition is that such a marginal subsidy counteracts the incentive for this agent to choose less effort in response to the effort norm. By analogy, the first term on the right hand side of the formula for $T_{H,w}^l$ is positive (since $e_H^l < \overline{e}_H$) and constitutes a pure paternalist motive for marginal income taxation of the low-productivity type in social identity group *H*, who would otherwise exert too much effort in response to the effort norm.

The second reason for imposing non-zero marginal income taxes on the *h-L* and *l-H* individuals is captured by the second term on the right hand side of the formulas for $T_{L,w}^h$ and $T_{H,w}^l$ in Proposition 3. When the government does not share the individual preferences for social identity, the individuals' selection into social identity groups may influence the marginal tax policy: these effects were absent under a welfarist government, which is seen from Corollary 1. To be more specific, there is a paternalist motive for influencing the tax revenue through the selection into social identity groups. We show in the

Appendix that these components are derived from the following expressions

$$\gamma\left(G_L^i - G_H^i\right)\left[\frac{\partial n_L^i}{\partial e_L^i} - \frac{u_{L,e}^i}{u_{L,x}^i}\frac{\partial n_L^i}{\partial x_L^i}\right] = -\gamma\left(G_L^i - G_H^i\right)\frac{(e_L^i - \overline{e}_L)}{2\beta} \tag{19a}$$

$$\gamma\left(G_L^i - G_H^i\right)\left[\frac{\partial n_L^i}{\partial e_H^i} - \frac{u_{H,e}^i}{u_{H,x}^i}\frac{\partial n_L^i}{\partial x_H^i}\right] = \gamma\left(G_L^i - G_H^i\right)\frac{(e_H^i - \overline{e}_H)}{2\beta} \tag{19b}$$

where the right hand side follows from the comparative statics properties of equations (7a) and (7b). Note that the right hand side of equation (19a) is zero for productivity-type *l* and non-zero for productivity-type *h*, while the right hand side of equation (19b) is zero for productivity-type *h* and non-zero for productivity-type *l*. The intuition is that a compensated increase in e_L^h (where the utility-compensation is based on the preferences of the paternalist government) leads to lower identity utility for *h-L* individuals and, therefore, a decrease in the number of individuals of productivity-type *h* is social identity group *L*. Similarly, a compensated increase in e_H^l leads to that fewer individuals of productivity-type *l* choose social identity group *L* (due to that the discrepancy between e_H^l and $\overline{e}_H$ decreases). In turn, this leads to increased tax revenue if $G_L^i < G_H^i$ for *i=l, h*, which constitutes an incentive to subsidize income at the margin for the high-productivity type in social identity group *L* and the low-productivity type in social identity group *H* (the second term on the right hand side of each such tax formula is negative).

We summarize the qualitative implications of Proposition 4 as follows:

Corollary 2. *If* $\lambda_L = \lambda_H = 0$ *and* $G_L^i < G_H^i$ *for i=l, h, the optimal tax policy implemented by a paternalist government satisfies* $T_{L,w}^l > 0$, $T_{L.w}^h < 0$ *and* $T_{H,w}^h > 0$, *while* $T_{H,w}^l$ *can be either positive or negative at the optimum.*

Second Best Policy

We now turn to the second best model, where the self-selection constraints bind. To shorten the notation, let

$$PRS_{j,ex}^i = \frac{u_{j,e}^i}{u_{j,x}^i} \text{ and } P\hat{R}S_{j,ex}^h = \frac{\hat{u}_{j,e}^h}{\hat{u}_{j,x}^h} \tag{20}$$

represent the marginal rate of substitution between effort and second period consumption for productivity-type *i* and the mimicker, respectively, in social identity group *j* based on the preferences imposed on them by the paternalist government. As such, these differ from the individuals' own marginal rates of substitution presented in equations (13). Also, to suppress policy incentives already explained, we use $T_{j,w}^{i,FB}$ to denote the first best marginal tax formula for productivity-type i in social identity group j as defined in Proposition 4, although here evaluated in the second best allocation. The optimal marginal tax policy is characterized in Proposition 5.

Proposition 5. *In a second best optimal resource allocation where the self-selection constraints bind, the tax policy implemented by a paternalist government satisfies*

$$T^l_{L,w} = T^{l,FB}_{L,w} - \frac{\lambda_L}{\gamma n^l_L \theta^l}\left(e^h_L - \frac{\theta^l}{\theta^h}e^l_L\right) - \frac{\lambda_L \hat{u}_{L,x}}{\gamma n^l_L \theta^l}\Delta_L - \frac{\lambda_L}{\gamma n^l_L \theta^h}\left(\frac{\theta^l}{\theta^h}e^l_L - \overline{e}_L\right)$$

$$T^h_{L,w} = T^{h,FB}_{L,w} + \frac{\lambda_L}{\gamma n^h_L \theta^h}\left(e^h_L - \overline{e}_L\right)$$

$$T^l_{H,w} = T^{l,FB}_{H,w} - \frac{\lambda_H \hat{u}_{H,x}}{\gamma n^l_H \theta^l}\Delta_H - \frac{\lambda_H}{\gamma n^l_H \theta^h}\left(\frac{\theta^l}{\theta^h}e^l_H - \overline{e}_H\right)$$

$$T^h_{H,w} = T^{h,FB}_{H,w} - \frac{\lambda_H}{\gamma n^h_H \theta^h}\left(e^h_H - \frac{\theta^l}{\theta^h}e^l_H\right)$$

where $\Delta_j = PRS^l_{j,ex} - \frac{\theta^l}{\theta^h}P\hat{R}S^h_{j,ex}$ *for j=L,H.*

Proof: see the Appendix.

Three additional components arise here compared to the first best policy rules characterized in Proposition 4. First, there is a direct incentive to make mimicking less attractive through marginal taxation of the low-productivity type captured by

$$-\frac{\lambda_j \hat{u}_{j,x}}{\gamma n^l_j \theta^l}\Delta_j - \frac{\lambda_j}{\gamma n^l_j \theta^h}\left(\frac{\theta^l}{\theta^h}e^l_j - \overline{e}_j\right) > 0 \text{ for } j=L,H$$

in the first and third tax formulas in Proposition 5. Therefore, this effect works to increase the marginal tax rates faced by low-productivity individuals (compared to the marginal tax rates under full information). The interpretation is that a decrease in e^l_j relaxes the self-selection constraint both because $\Delta_j < 0$ (as in conventional models of optimal income taxation) and by increasing the distance between the mimicker's effort and the effort norm. Second, there is an incentive to relax the self-selection constraint though marginal income taxation of the high-productivity type in social identity group L, which works to offset the subsidy result derived under first best conditions (see Corollary 2). This effect is captured by the second term on the right hand side of the formula for $T^h_{L,w}$ (which is positive since $e^h_L > \overline{e}_L$), implying that the marginal tax facing *h-L* individuals can be either positive or negative here. The intuition is that a decrease in e^h_L leads to higher identity utility (and, therefore, counteracts the incentive to become a mimicker) for *h-L* individuals by reducing distance between e^h_L and $\overline{e}_L$, ceteris paribus. There is no corresponding effect in the marginal income tax implemented for *h-H* individuals where $e_H^h = \overline{e}_H$.

The third additional component reflects a desire to relax the self-selection constraints through changes in the two social norms, $\overline{e}_L$ and $\overline{e}_H$: this will only affect the marginal income tax rates of the low-productivity type in social identity group *L* and high-productivity type in social identity group *H*. Since the government may relax the self-selection constraint by increasing the social norms, there is an incentive to subsidize *l-L* and *h-H* individuals for this particular reason as reflected by the second term on the right hand side in their tax formulas.

5 Summary and Conclusions

In this paper, we have used a model of education choice to study the implications of social identity and self-categorization for optimal income taxation and publicly provided education. The model assumes that individuals differ in two ways. First, as in a standard optimal taxation framework, we assume that individuals differ in terms of innate ability, and that this ability is private information. Second, we introduce an element of social inertia by assuming that individuals differ with respect to their preferences for the social group they want to be associated with. Our model contains two such social identity groups in which norms regarding study effort are formed. Individuals will then select into one of the groups depending on ability and preferences, and they will make educational (effort) choices based on their ability and the group norm. The social norms are modelled as the modal effort in each social group based on finding in the social psychology literature that norms appear to be more polarized than group averages. In the baseline model developed and analyzed in Sections 2 and 3, the government is welfarist in the sense of accepting all aspects of consumer preferences (including social identity), while the optimal tax and expenditure policy of a paternalist government that does not share the consumer preference for social identity is discussed in Section 4.

The take-home message of the paper is that the consumer preference for social identity leads to corrective as well as redistributive motives for income taxation, and also influences the optimal provision of public education. Our contribution is to characterize the incentives on which these policy responses are based. First, externality correction by a welfarist government leads in itself to an element of tax progression that serves to reduce the discrepancy between the effort norm and the actual effort chosen by low-productivity individuals in the high-effort group. In turn, this effect is counteracted by the government's desire to relax the self-selection constraint, since increases in the effort norms contribute to make mimicking less attractive. Second, if the preference for social identity is sufficiently strong, increased wage-inequality leads to higher social welfare through a relaxation of the selection constraint. It may thus be desirable to use publicly provided education to induce more wage-inequality, even if such inequality increases the intrinsic utility of a potential mimicker. This policy incentive would not be present if the consumers were not concerned with their social identity. Third, the policy rule for publicly provided education does not depend on whether the government has a welfarist or paternalist objective. However, the marginal tax policy differs between the two types of government in a fundamental way, primarily because the paternalist government has a motive to correct effort choices to "undo" the effects of self-categorization.[18]

Let us end by briefly discussing two possible directions of future research. First, our model focuses on effort choices and the policy implications of social identity and self-categorization in that particular context. Yet, preferences for social identity may also affect other aspect of education choices such as the selection into education programs and, therefore, how individuals eventually end up in different occupations. Second, a model with more than two periods would allow us to relate effort choices (and possibly also the social identity component underlying such choices) to time-inconsistent preferences for immediate gratification. If individuals later in life regret their lack of study effort, at least a paternalist government may have incentives to use taxation and public expenditure to correct these behavioral mistakes (e.g., by providing adult education). We hope to address these and other questions in future research.

[18] In other contexts, the policy incentive to correct behavioral failures perceived by a paternalist government and the incentives to correct for externalities perceived by a welfarist government need not necessarily lead to very different policy rules for optimal taxation. Aronsson and Johansson-Stenman (2014) compare a paternalist and a welfarist government from the perspective of optimal taxation under relative consumption concerns and find that the two governments implement surprisingly similar policy rules despite completely different underlying policy incentives. Our results thus stand in sharp contrast to the similarities they found.

Appendix

Social First Order Conditions under Welfarism

Use $G_j^i = w_j^i - c_j^i - x_j^i$, $\hat{u}_j^h = u^h(c_j^l, x_j^l, \frac{\theta^l}{\theta^h} e_j^l)$, and $n_H^i = N^i - n_L^i$.

We have

$$\delta_j^l u_{j,c}^l - \lambda_j \hat{u}_{j,c}^h + \gamma\left[-n_j^l + \frac{\partial n_j^l}{\partial c_j^l}\left(G_L^l - G_H^l\right)\right] = 0 \tag{A1}$$

$$\delta_j^l u_{j,x}^l - \lambda_j \hat{u}_{j,x}^h + \gamma\left[-n_j^l + \frac{\partial n_j^l}{\partial x_j^l}\left(G_L^l - G_H^l\right)\right] = 0 \tag{A2}$$

$$\begin{aligned}&\delta_j^l\left[u_{j,e}^l - (e_j^l - \bar{e}_j)\right] - \lambda_j \frac{\theta^l}{\theta^h}\left[\hat{u}_{j,e}^h - \left(\frac{\theta^l}{\theta^h} e_j^l - \bar{e}_j\right)\right] \\ &+\gamma\left[n_j^l \theta^l + \frac{\partial n_j^l}{\partial e_j^l}\left(G_L^l - G_H^l\right)\right] + \frac{\partial \mathcal{L}}{\partial \bar{e}_j}\frac{\partial \bar{e}_j}{\partial e_j^l} = 0\end{aligned} \tag{A3}$$

$$\left(\delta_j^h + \lambda_j\right) u_{j,c}^h + \gamma\left[-n_j^h + \frac{\partial n_j^h}{\partial c_j^h}\left(G_L^h - G_H^h\right)\right] = 0 \tag{A4}$$

$$\left(\delta_j^h + \lambda_j\right) u_{j,x}^h + \gamma\left[-n_j^h + \frac{\partial n_j^h}{\partial x_j^h}\left(G_L^h - G_H^h\right)\right] = 0 \tag{A5}$$

$$\left(\delta_j^h + \lambda_j\right)\left[u_{j,e}^h - (e_j^h - \bar{e}_j)\right] + \gamma\left[n_j^h \theta^h + \frac{\partial n_j^h}{\partial e_j^h}\left(G_L^h - G_H^h\right)\right] + \frac{\partial \mathcal{L}}{\partial \bar{e}_j}\frac{\partial \bar{e}_j}{\partial e_j^h} = 0 \tag{A6}$$

for j=L, H, and $\delta_L^l = 1$, and

$$\gamma\left[\sum_i \theta_g^i \sum_j n_j^i e_j^i - N\right] - \sum_j \lambda_j\left[\hat{u}_{j,e}^h - \left(\frac{\theta^l}{\theta^h} e_j^l - \bar{e}_j\right)\right] e_j^l \frac{\partial(\theta^l / \theta^h)}{\partial g} = 0 \ . \tag{A7}$$

In all proofs below, we use that the norms are given by $\bar{e}_L = e_L^l$ and $\bar{e}_H = e_H^h$, meaning that the partial welfare effects of increases in these norms will only be part of the social first order conditions for effort of the low-productivity type in social identity group L and high-productivity type in social identity group H, respectively.

Marginal Savings Taxes

Consider the low-productivity type in social identity group j. Combining equations (A1) and (A2) gives

$$MRS^l_{j,cx}\left[\lambda_j\hat{u}^h_{j,x}-\gamma\left(-n^l_j+\frac{\partial n^l_j}{\partial x^l_j}\left(G^l_L-G^l_H\right)\right)\right]=\lambda_j\hat{u}^h_{j,c}-\gamma\left[-n^l_j+\frac{\partial n^l_j}{\partial c^l_j}\left(G^l_L-G^l_H\right)\right]$$

where $MRS^l_{j,cx}=u^l_{j,c}/u^l_{j,x}$. By using $\partial n^l_j/\partial e^l_j=MRS^l_{j,cx}(\partial n^l_j/\partial x^l_j)$ from equation (7a) and $T^l_{j,s}=1-MRS^l_{j,cx}$ from equation (4b), this can be rewritten to read

$$\gamma n^l_j T^l_{j,s}=\lambda_j\hat{u}^h_{j,x}MRS^l_{j,cx}\left[MRS^l_{j,cx}-M\hat{R}S^h_{\mathrm{j,c}x}\right].$$

By using equations (A4), (A5), (6b), and (7b), we can use the same procedure to derive $T^h_{j,s}=0$.

Proof of Proposition 1

In the special case where the self-selection constraints are not binding, we have $\lambda_L=\lambda_H=0$ in equations (A1)-(A6). Consider first the marginal income tax rate implemented for the low-productivity type in social identity group L. Combining equations (A2) and (A3) while using equations (13) give

$$(MRS^l_{L,ex}+\theta^l)\gamma n^l_j=\gamma\left(G^l_L-G^l_H\right)\left[MRS^l_{L,ex}\frac{\partial n^l_L}{\partial x^l_L}-\frac{\partial n^l_L}{\partial e^l_L}\right]-\frac{\partial\mathcal{L}}{\partial\overline{e}_L}\qquad\text{(A8)}$$

Making use of the private first order condition from equation (4a) we can replace $MRS^l_{L,ex}+\theta^l$ with $\theta^l T^l_{L,w}$ in equation (A7). Finally, using equations (7) to derive

$$\frac{\partial n^l_L}{\partial x^l_L}=\frac{1}{2\beta}\frac{\partial u^l_L}{\partial x^l_L}\qquad\text{(A9a)}$$

$$\frac{\partial n^l_L}{\partial e^l_L}=\frac{1}{2\beta}\left(\frac{\partial u^l_L}{\partial e^l_L}-\left(e^l_L-\overline{e}_L\right)\right)\qquad\text{(A9b)}$$

we can see that the two terms within square brackets in equation (A8) cancel out, which means that we arrive at the expression for $T^l_{L,w}$ in the proposition. The marginal income tax rate facing productivity-type *h* in social identity group *H* can be derived analogously by using equations (A5) and (A6). Following again the same approach, the formulas for the optimal marginal tax rates for productivity-type *l* in social identity group *H* and productivity-type *h* in social identity group *L* will reduce to zero, since they do not contain the partial welfare effect of any of the two social norms. ■

Proof of Proposition 2

In a second best setting where $\lambda_j \neq 0$ for *j*=*L*, *H*, we see from the social first order conditions that only the low-productivity type will be additionally distorted in its effort choice compared to the first best. The equation analogous to equation (A8) for type *l* individuals in social identity group *L* then becomes

$$\begin{aligned}(MRS^l_{L,ex} + \theta^l)\gamma n^l_j = \gamma\left(G^l_L - G^l_H\right)&\left[MRS^l_{L,ex}\frac{\partial n^l_L}{\partial x^l_L} - \frac{\partial n^l_L}{\partial e^l_L}\right] \\ &- MRS^l_{L,ex}\lambda_L \hat{u}^h_{L,x} + \lambda_L \frac{\theta^l}{\theta^h}\left[\hat{u}^h_{L,e} - \left(\frac{\theta^l}{\theta^h} e^l_L - \bar{e}_L\right)\right] - \frac{\partial \mathcal{L}}{\partial \bar{e}_L}\end{aligned}.$$

Using the definition of $M\hat{R}S^h_{j,ex}$ from equation (13), and $MRS^l_{L,ex} + \theta^l = \theta^i T^l_{L,w}$, we arrive at the marginal tax formula for *l*-*L* individuals in the proposition. The corresponding tax formulas for other combinations of productivity and social identity groups are derived in the same general way. ■

Social First Order Conditions under Paternalism

$$\delta_j^l u_{j,c}^l - \lambda_j \hat{u}_{j,c}^h + \gamma\left[-n_j^l + \frac{\partial n_j^l}{\partial c_j^l}\left(G_L^l - G_H^l\right)\right] = 0 \tag{A10}$$

$$\delta_j^l u_{j,x}^l - \lambda_j \hat{u}_{j,x}^h + \gamma\left[-n_j^l + \frac{\partial n^l}{\partial x_j^l}\left(G_L^l - G_H^l\right)\right] = 0 \tag{A11}$$

$$\delta_j^l u_{j,e}^l - \lambda_j \frac{\theta^l}{\theta^h}\left[\hat{u}_{j,e}^h - \left(\frac{\theta^l}{\theta^h} e_j^l - \overline{e}_j\right)\right] + \gamma\left[n_j^l \theta^l + \frac{\partial n_j^l}{\partial e_j^l}\left(G_L^l - G_H^l\right)\right] + \frac{\partial \mathcal{L}}{\partial \overline{e}_j}\frac{\partial \overline{e}_j}{\partial e_j^l} = 0 \tag{A12}$$

$$\left(\delta_j^h + \lambda_j\right) u_{j,c}^h + \gamma\left[-n_j^h + \frac{\partial n_j^h}{\partial c_j^{2h}}\left(G_L^h - G_H^h\right)\right] = 0 \tag{A13}$$

$$\left(\delta_j^h + \lambda_j\right) u_{j,x}^h + \gamma\left[-n_j^h + \frac{\partial n_j^h}{\partial x_j^h}\left(G_L^h - G_H^h\right)\right] = 0 \tag{A14}$$

$$\left(\delta_j^h + \lambda_j\right) u_{j,e}^h - \lambda_j (e_j^h - \overline{e}_j) + \gamma\left[n_j^h \theta^h + \frac{\partial n_j^h}{\partial e_j^h}\left(G_L^h - G_H^h\right)\right] + \frac{\partial \mathcal{L}}{\partial \overline{e}_j}\frac{\partial \overline{e}_j}{\partial e_j^h} = 0 \tag{A15}$$

for j=L, H, and $\delta_L^l = 1$. The first order condition for the publicly provided good takes the same form as equation (A7).

Proof of Proposition 4

Consider productivity-types l and h in social identity groups group L. By using equations (A11) and (A12) together with the definition of marginal rate of substitution in equations (18), we have

$$\left(PRS_{L,ex}^l + \theta^l\right)\gamma n_L^l = \left(PRS_{L,ex}^l \frac{\partial n_L^l}{\partial x_L^l} - \frac{\partial n_L^l}{\partial e_L^l}\right)\gamma\left(G_L^h - G_H^h\right) - \frac{\partial \mathcal{L}}{\partial \overline{e}_L}. \tag{A16}$$

Similarly, combining equations (A14) and (A15) gives

$$\left(PRS_{L,ex}^h + \theta^h\right)\gamma n_L^h = \left(PRS_{L,ex}^h \frac{\partial n_L^h}{\partial x_L^h} - \frac{\partial n_L^h}{\partial e_L^h}\right)\gamma\left(G_L^h - G_H^h\right). \tag{A17}$$

Note that $\partial \overline{e}_L / \partial e_L^h = \partial \overline{e}_H / \partial e_L^h = 0$, which explains why there are no partial welfare effects of social norms in equation (A17).

Now, due to the difference between the *MRS* and *PRS* measures, we can use the private first order condition for effort to derive

$$PRS_{j,ex}^i + \theta^i = \theta^i T_{j,w}^i + \frac{\left(e_j^i - \overline{e}_j\right)}{u_{j,x}^i}. \tag{A18}$$

Substituting into equations (A16) and (A17) gives

$$\theta^l T_{L,w}^l \gamma n_L^l = \left(PRS_{L,ex}^l \frac{\partial n_L^l}{\partial x_L^l} - \frac{\partial n_L^l}{\partial e_L^l} \right) \gamma \left(G_L^h - G_H^h\right) - \frac{\partial \mathcal{L}}{\partial \overline{e}_L} - \frac{\left(e_L^l - \overline{e}_L\right)}{u_{L,x}^l} \tag{A19}$$

$$\theta^h T_{L,w}^h \gamma n_L^h = \left(PRS_{L,ex}^h \frac{\partial n_L^h}{\partial x_L^h} - \frac{\partial n_L^h}{\partial e_L^h} \right) \gamma \left(G_L^h - G_H^h\right) - \frac{\left(e_L^h - \overline{e}_L\right)}{u_{L,x}^h}. \tag{A20}$$

Note that equations (7a) and (7b) imply

$$PRS_{L,ex}^l \frac{\partial n_L^l}{\partial x_L^l} - \frac{\partial n_L^l}{\partial e_L^l} = 0 \tag{A21}$$

$$PRS_{L,ex}^h \frac{\partial n_L^h}{\partial x_L^h} - \frac{\partial n_L^h}{\partial e_L^h} = \frac{1}{2\beta}\left(e_L^h - \overline{e}_L\right). \tag{A22}$$

The right hand side of equation (A22) is due to the discrepancy between the *MRS* and *PRS* measures; this discrepancy vanishes in equation (A21), since $\overline{e}_L = e_L^l$. Substituting equations (A21) and (A22) into equations (A19) and (A20) and using $\overline{e}_L = e_L^l$ gives the expressions for $T_{L,w}^l$ and $T_{L,w}^h$ in the proposition. The marginal income tax rates implemented for productivity-types *l* and *h* in social identity group *H* can be derived analogously. ■

Proof of Proposition 5

When the self-selection constraints bind (such that $\lambda_L \neq 0$ and $\lambda_H \neq 0$), the analogue to equation (A16) is given by

$$\begin{aligned}\left(PRS^l_{L,ex} + \theta^l\right)\gamma n^l_j = {} & \left(PRS^l_{L,ex}\frac{\partial n^l_L}{\partial x^l_L} - \frac{\partial n^l_L}{\partial e^l_L}\right)\gamma\left(G^h_L - G^h_H\right) \\ & - \lambda_L \hat{u}^l_{L,x}\left(PRS^l_{L,ex} - \frac{\theta^l}{\theta^h}P\hat{R}S^h_{L,ex}\right) \\ & - \lambda_L \frac{\theta^l}{\theta^h}\left(\frac{\theta^l}{\theta^h}e^l_L - \overline{e}_L\right) - \frac{\partial \mathcal{L}}{\partial \overline{e}_L}\end{aligned} \quad . \tag{A23}$$

The difference to the first best setting is the second and third term on the right hand side. Similarly, the analogue to equation (A17) becomes

$$\left(PRS^h_{L,ex} + \theta^h\right)\gamma n^h_j = \left(PRS^h_{L,ex}\frac{\partial n^h_L}{\partial x^h_L} - \frac{\partial n^h_L}{\partial e^h_L}\right)\gamma\left(G^h_L - G^h_H\right) + \lambda_j\left(e^h_L - \overline{e}_L\right) \tag{A24}$$

where the final term on the right hand side constitutes the addition due to the binding self-selection constraints. By proceeding in exactly the same way as in the proof of Proposition 4, we can derive the expressions for $T^l_{L,w}$ and $T^h_{L,w}$ in Proposition 5. Again, the marginal income tax rates implemented for productivity-types *l* and *h* in social identity group *H* are derived in an analogous way. ■

References

Anger, S. and Heineck, G. (2010), Do smart parents raise smart children? The Intergenerational Transmission of Cognitive Abilities. *Journal of Population Economics*, **23** (3), 1105-1132.

Akerlof, G. (1980), A Theory of Social Custom, of which Unemployment may be one Consequence. *Quarterly Journal of Economics* **94**, 749-775.

Akerlof, G. and R. Kranton (2000), Economics and Identity. *The Quarterly Journal of Economics **115***, 715-753.

Akerlof, G. and R. Kranton (2002), Identity and Schooling: Some Lessons for the Economics of Education. *Journal of Economic Literature **40***, 1167-1201.

Aronsson, T., S. Blomquist, and H. Sacklen (1999), How to Identify Interdependent Behaviour in an Empirical Model of Labor Supply. *Journal of Applied Econometrics* **14**, 607-626.

Aronsson, T. and D. Granlund (2015), Gender Norms, Work Hours, and Corrective Taxation. *Journal of Behavioral and Experimental Economics* **56**, 33-39.

Aronsson, T. and O. Johansson-Stenman (2008), When the Joneses' Consumption Hurts: Optimal Public Good Provision and Nonlinear Income Taxation. *Journal of Public Economics* **92**, 986-997.

Aronsson, T. and O. Johansson-Stenman (2010), Positional Concerns in an OLG Model: Optimal Labor and Capital Income Taxation. *International Economic Review* **51**, 1071-1095.

Aronsson, T. and O. Johansson-Stenman (2014) Paternalism against Veblen: Optimal Taxation and Non-Respected Preferences for Social Comparisons. Umeå Economic Studies no 901.

Aronsson, T. and T. Sjögren (2010) Optimal Income Taxation and Social Norms in the Labor Market. *International Tax and Public Finance* **17**, 67-89.

Austin, A. M. B., and Draper, D. C. (1984), The Relationship among Peer Acceptance, Social Impact, and Academic Achievement in Middle Childhood. *American Educational Research Journal*, **21** (3), 597-604.

Becker, G. S. (2009), Human capital: A Theoretical and Empirical Analysis, with special reference to education. University of Chicago Press.

Bernheim, D. (1994), A Theory of Conformity. *Journal of Political Economy* **102**, 1-35841-877.

Björklund, A., and Jäntti, M. (2012), How Important is Family Background for Labor-economic Outcomes? *Labour Economics* **19** (4), 465-474.

Blomquist, S. (1993), Interdependent Behavior and the Effect of Taxes. *Journal of Public Economics* **51**, 211-218.

Boskin, M. J., and E. Sheshinski (1978), Individual Welfare Depends upon Relative Income. *Quarterly Journal of Economics* **92**, 589-601.

Bowles, S., Gintis, H. (2000), Reciprocity, Self-Interest and the Welfare State. *Nordic Journal of Political Economy* **26**, 33-53.

Bruvoll, A., Nyborg, K. (2004), The Cold Shiver of not Giving Enough: On the Social Costs of Recycling Campaigns. *Land Economics* **80**, 540-549.

Buchmann, C., and Dalton, B. (2002), Interpersonal Influences and Educational Aspirations in 12 Countries: The Importance of Institutional Context. *Sociology of Education* **75** (2), 99-22.

Cooper, J., Kelly, K. A., and Weaver, K. (2001), Attitudes, Norms, and Social Groups. *Blackwell Handbook of Social Psychology: Group Processes*, 259-282.

Dupor, B. and W. F. Liu (2003), Jealousy and Overconsumption. *American Economic Review* **93**, 423-428.

Eckerstorfer, P. and R. Wendner (2013), Asymmetric and Non-atmospheric Consumption Externalities, and Efficient Consumption Taxation. *Journal of Public Economics* **106**, 42-56.

Ermisch, J., and Francesconi, M. (2001), Family Matters: Impacts of Family Background on Educational Attainments. *Economica* **68** (270), 137-156.

Guo, J-T. and A. Krause (2011), Optimal Nonlinear Income Taxation with Habit Formation”. *Journal of Public Economic Theory* **13**, 463-480.

Guo, J-T. and A. Krause (2013), Optimal Nonlinear Taxation of Income and Education Expenditures. *Oxford Economic Papers* ***65***, 74-9.

Hogg, M. A. (2006), Social Identity Theory, in: Burke P. J. (ed.), *Contemporary Social Psychology Theories*, pp. 111-135, Stanford University Press, Stanford, CA.

Hogg, M. A. and Reid, S.A. (2006), Social Identity, Self-Categorization, and the Communication of Group Norms. *Communication Theory* **16**, 7-30.

Lindbeck, A. (1995), Welfare State Disincentives with Endogenous Habits and Norms. *Scandinavian Journal of Economics* **97**, 477-494.

Lindbeck, A., S. Nyberg, and J. Weibull (1999), Social Norms and Economic Incentives in the Welfare State. *Quarterly Journal of Economics* **114**, 1-35.

Lindbeck, A., S. Nyberg, and J. Weibull (2003), Social Norms and Welfare State Dynamics. *Journal of the European Economic Association* **1**, 533-542.

Ljungqvist, L. and H. Uhlig (2000), Tax Policy and Aggregate Demand Management Under Catching Up with the Joneses. *American Economic Review* **90**, 356-366.

Maldonado, D. (2008) Education Policies and Optimal Taxation. *International Tax and Public Finance* 15, 131-143.

McGarty, C., Turner, J. C., Hogg, M. A., David, B. and Wetherell, M. S. (1992), Group Polarization as Conformity to the Prototypical Group Member. *British Journal of Social Psychology* **31**: 1–19.

Moscovici, S. and Zavalloni, M. (1969), The Group as a Polarizer of Attitudes. *Journal of Personality and Social Psychology* **12**, 125–135.

Oswald, A. (1983), Altruism, Jealousy and the Theory of Optimal Non-Linear Taxation. *Journal of Public Economics* **20**, 77–87.

Stern, N.H. (1982), Optimum Taxation with Errors in Administration. *Journal of Public Economics* **17**, 181-211.

Stiglitz, J. E. (1982), Self-Selection and Pareto Efficient Taxation. *Journal of Public Economics* **17**, 213-240.

Tajfel, H., Billig, M., Bundy, R. P., and Flament, C. (1971), Social Categorization and Intergroup Behaviour. *European Journal of Social Psychology* **1**, 149–177.

Tajfel, H., and Turner, J. C. (1979), An Integrative Theory of Intergroup Conflict. In Austin, W.G. and S. Worchel (eds.), *The Social Psychology of Intergroup Relations*, pp 33–47, Monterey, CA: Brooks/Cole.

Tuomala, M. (1990), *Optimal Income Tax and Redistribution*, Oxford: Clarendon Press.

Turner, J. C., Wetherell, M. S. and Hogg, M. A. (1989), Referent Informational Influence and Group Polarization. *British Journal of Social Psychology*, **28,** 135–147.

Wendner, R. and L. H. Goulder (2008), Status Effects, Public Goods Provision, and the Excess Burden. *Journal of Public Economics* **92**, 1968-1985.

Wentzel, K. R., and Caldwell, K. (1997), Friendships, Peer Acceptance, and Group Membership: Relations to Academic Achievement in Middle School. *Child Development* **68** (6), 1198-1209.

Willis, R. J., and Rosen, S. (1979), Education and Self-Selection. *Journal of Political Economy* **87** (5), S7-S36.

II

INTERGENERATIONAL MOBILITY IN SWEDEN: A REGIONAL PERSPECTIVE

STEFANIE HEIDRICH*

ABSTRACT

I employ high quality register data to present new facts about income mobility in Sweden. The focus of the paper is regional differences in mobility, using a novel approach based on a multilevel model. This method is well-suited when regions differ greatly in population size as is the case in Sweden. The maximum likelihood estimates are substantially more precise than those obtained by running separate OLS regressions. I find small regional differences in income mobility when measured in relative terms. Regional differences are large when adopting an absolute measure and focusing on children with below-median parent income. On the national level I find that the association between parent and child income ranks has decreased over time, implying increased mobility.

JEL classification: D31, J62, R0
Keywords: intergenerational income mobility, regional analysis, multilevel model

*Department of Economics, Umeå School of Business and Economics, Umeå University.
E-Mail: stefanie.heidrich@umu.se. I would like to thank Thomas Aronsson, Spencer Bastani, David Granlund, Gauthier Lanot, Magnus Wikström, and the seminar participants at ZEW Mannheim and Umeå University for helpful comments on this as well as earlier versions of the paper. The Umeå SIMSAM Lab data infrastructure used in this study was developed with support from the Swedish Research Council and by strategic funds from Umeå University.

1 Introduction

The academic and public interest in the shape and changing patterns of income distributions has been growing steadily over the past decades. For example, the rising top income share in the US has inspired many discussions on everyone's equal opportunity to prosperity through hard work in the formerly known "land of opportunity". In a recent paper, Chetty et al. (2014a) emphasize the importance of regional differences in income mobility and describe the US as being, instead of a *land of opportunity*, a collection of *societies* some of which are *lands of opportunity* with high rates of mobility across generations, and others in which few children escape poverty.[1]

This paper employs high quality register data to present new facts about the state of income mobility in Sweden with a focus on regional differences in income mobility. My data set allows me to analyze national and regional mobility measures very precisely for the Swedish population born between 1968 and 1976.

Income mobility broadly refers to the extent child income can be predicted using parent income. The by far most commonly employed mobility measure in the literature is the intergenerational elasticity (IGE). This is simply the slope parameter of a regression of log lifetime income of generation t on log lifetime income of generation $t-1$. A smaller IGE means that it is harder to predict child income using parent income, and that income mobility is higher. Estimates of the IGE in the literature are generally around 0.4 with higher estimates for the US, and usually smaller estimates for the European and especially the Nordic countries (see Björklund and Jäntti, 1997; Solon, 1992; Solon, 1999; Solon, 2004; or Mazumder, 2005). Recent summaries of economic research in intergenerational mobility are provided by Björklund and Jäntti (2009) and Black and Devereux (2011). Recent extensions include the study of more than two generations such as Lindahl et al. (2015).

The IGE, however, has some well-known drawbacks. One limitation is that zeroes have to be dropped from the regression which can lead to biased estimates. There is also evidence that a linear model does not fit very well the relationship between the logged incomes of two generations (see, i.e., Couch and Lillard, 2004 and Bratsberg et al., 2007).

More recently, scholars have directed their attention to an alternative measure of mobility based on income ranks. This measure, called the rank-rank slope, is obtained by regressing the position in the income distribution (expressed in percentiles) of each member of the child generation on the position in the income distribution of their parents. In addition to traditional IGE measures, in this paper I will also compute regional and national measures of mobility based on income ranks. For the regional analysis I employ two different mobility measures based on income ranks. The first one is "relative mobility"

[1] Chetty et al. (2014a; 2014b) found large differences in mobility across the 741 commuting zones in the US, and that economic mobility in the US has not changed significantly over time for cohorts born between 1971 and 1993 (even though the consequences of this same mobility have increased due to the growth in income inequalities).

which describes the mean difference in outcomes between children with parents in the top and bottom of the income distribution, respectively. The second is "upward mobility" which measures the mean absolute outcome of children from families with below-median income levels and, importantly, focuses exclusively on the regional differences in outcomes of children in the poorer half of the population.

It is important to keep in mind that both the IGE and relative mobility (as defined above) are *relative* measures and therefore do not reveal if an improvement in mobility is driven by better outcomes of some poorer families, or solely by worse outcomes of richer families. Therefore, upward mobility, and other measures based on absolute outcomes, are necessary to obtain a more comprehensive picture of income mobility.

The geographical unit that I will focus on in the regional analysis is the "local labor market", which is an aggregation of municipalities defined by commuting patterns. The local labor market unit is similar to the commuting zone used by Chetty et al. (2014a). However, in comparison to the commuting zones in the US, there is much more variation between different Swedish local labor markets in terms of population size (and thereby the number of observations). As I have shown below, this aspect of the data results in imprecise estimates. To remedy this problem I propose a joint estimation technique using maximum likelihood, referred to as a multilevel (or hierarchical) model. In contrast to the approach taken in Chetty et al. (2014a), where they essentially run a set of distinct regressions, the multilevel model allows me to make a comparison between the different regional mobility measures in a statistically rigorous way. For example, I can test if the mobility estimate of one particular region is statistically significantly different from the national average. For completeness however, I also report and discuss results based on separate OLS regressions by region.

My results can be summarized as follows. I find that relative mobility is relatively homogeneous across Sweden. The difference of mean son income rank between families at the very top and the very bottom of the income distribution, respectively, is 22.2 percentile ranks in most local labor markets. Only 9 areas out of 112 show significantly lower or higher relative mobility, i.e. a larger or smaller difference between sons from families with highest and lowest incomes, respectively. Stockholm ranks at the bottom with the lowest relative mobility, and the Umeå region in northern Sweden shows the highest relative mobility.

Upward mobility, the expected outcome for sons from below-median income families, varies considerably more across Swedish local labor market areas, from 36.32 percentile ranks in Torsby to 50.77 in Hylte. This corresponds to an income difference of 32.842 SEK per year (≈3.839 USD). In addition, children who spend a significant part of their childhood in very rural areas of Sweden generally have significantly worse outcomes compared to children growing up in urban areas. This result can be explained in part by the large fraction of children from rural areas that do not move into cities as adults.

However, those who do move, do on average even better than the city natives.

For Sweden as a whole, the association between parent and son income measured by the relationship between income ranks has declined between 1968 and 1976. The IGE shows the opposite development and is misleading: in addition to the parent-child income association, the IGE also reflects a considerable increase in the ratio of the standard deviations of son over parent income.

The remainder of this paper is organized as follows. Section 2 includes the theoretical background of the IGE, mobility measures based on income ranks, and a description of the multilevel model. The data and variables used are described in Section 3. Section 4 reports on the non-parametric and parametric results for intergenerational mobility on the national level and over time. The regional results are the focus of Section 5 and section 6 concludes.

2 Measuring Intergenerational Mobility

The first part of this section comprises of a rather short review on the estimation of the intergenerational income elasticity. Further details can be found in the cited references. In the second part, I will explain the concepts of relative- and upward mobility which are used later to compare the Swedish local labor markets in terms of mobility. A brief introduction to multilevel modelling and the exact model used in this study is given in the last part of this section.

2.1 The Intergenerational Income Elasticity

The IGE is typically estimated using the following benchmark equation:

$$y_f^C = \alpha + \beta y_f^P + \varepsilon_f^C \tag{1}$$

where y_f^C and y_f^P are a the log of child and parent lifetime earnings in family f, respectively, and ε_f^C is assumed to be an iid error term representing all other influences on child earnings not correlated with parental income. I will use the terms income and earnings interchangeably in this section due to the range of different income/ earning concepts used in this literature. Traditionally, this relationship has been estimated for sons and fathers only. In Sweden, female labor market participation has been close to male participation for more than three decades. Thereby it would be particularly interesting to also study the association between child income and combined parent income, in addition to father or mother income only.

β is the parameter of interest, the elasticity between parent and child income. Equation (1) is a simple Markov model and a lower IGE corresponds to a greater regression toward the mean of income from one generation to the next. Black and Devereux (Chapter 1.2 in 2011) reviewed the results obtained for the IGE in different studies from the past decades.

What makes the estimation of the IGE difficult is the need for lifetime income data for both generations. Approximations made from lack of sufficient data lead to at least two known measurement problems: attenuation bias and life-cycle bias. Attenuation bias is due to measurement error of the regressor, most clearly seen when single year income observations are used to estimate the IGE. This was typical in early studies such as Solon (1992). Assuming a classic error-in-variables-model, measured income y_f then equals the true income $y_f^\star$, plus an error:

$$y_f = y_f^\star + \nu_f \,. \tag{2}$$

The known implication (Hausman, 2001) is a downward inconsistent IGE estimate.[2] The bias can be reduced using an average of T income observations to approximate the average of true lifetime income:

$$y_f^P = \frac{1}{T}\sum_{t=1}^{T}\left(y_{f,t}^{P\star} + \nu_{f,t}^P\right). \tag{3}$$

Björklund and Jäntti (1997) showed that in this case the inconsistency is diminishing in the number of observed years T (assuming the measurement errors/ transitory fluctuations are not serially correlated).[3] Mazumder (2005) used simulations to show that using a five year average (a number of typical magnitude in the literature) to measure father lifetime income still results in a downward bias of around 30 percent.

Life-cycle bias arises when single-year income observations of the child systematically deviate from the average of annual lifetime income (left hand-side measurement error). One can think of a parameter in front of $y_t^\star$ in equation (2) that is time variable. In this case, the inconsistency of the OLS coefficient varies as a function of the age at which annual income is measured. Simply adding age controls as done in earlier literature will not prevent this inconsistency. Life-cycle bias is a serious concern in the Swedish context where sons' individual income trajectories have been shown to be correlated with family characteristics (Nybom and Stuhler, 2016). Using the "correct" age as suggested, for example, by Haider and Solon (2006) to measure son's income can therefore only diminish, but never totally eliminate life cycle bias.

I address attenuation bias by averaging over a very large number of annual income observations where T is 17 for most parents in the sample (see Section 3.1 for more details). Importantly, income is observed for all individuals during the same age span, in the middle of their working lives. Life-cycle bias is handled by measuring child income

[2] This can be seen from the probability limit of β in equation (1) after substituting (2) for y_f^P:
$\operatorname*{plim}_{N\to\infty}\hat{\beta} = \frac{Cov(y_f^C, y_f^P)}{Var(y_f^P)} = \frac{\sigma^2_{y_f^{P\star}}}{\sigma^2_{y_f^{P\star}} + \sigma^2_{\nu_f^P}}\beta < \beta$, assuming $Cov\left(y_f^{P\star}, \nu_f^P\right) = 0$ and $\sigma^2_{\nu_f^P} \neq 0$, where $\sigma^2_{\nu_f^P}$ denotes the variance of the measurement error.

[3] The probability limit of β is here $\operatorname*{plim}_{N\to\infty}\hat{\beta}_1 = \frac{Cov(y_f^C, y_f^P)}{Var\left(y_f^P\right) + Var\left(\frac{1}{T}\sum_{t=1}^{T}\nu_{f,t}^P\right)} = \frac{\sigma^2_{y_f^{P\star}}}{\sigma^2_{y_f^{P\star}} + \frac{\sigma^2_{\nu_f^P}}{T}}\beta < \beta$

at the approximate age where Swedish individuals have been shown to earn just as much as the yearly average over a whole lifetime. [4]

There are two additional problems associated with the IGE measurement, the functional form and the handling of observations with zero income. For US data Chetty et al. (2014a) showed that the relationship between log incomes of children and their parents is not well represented by a simple linear regression model. This point has even been raised by Couch and Lillard (2004) and Bratsberg et al. (2007). One suggested remedy is to use income ranks instead of the log of incomes.

The issue with zero-income observations has long been known and dealt with in different manners such as dropping those observations or recoding zeros as different, usually small values. Dropping individuals with zero income will overstate mobility if children with zero income are over-represented in low-income families. Those families with low mobility will not be part of the sample. Recoding all zeros, on the other hand, leads to highly variable results depending on the replacement values chosen. A detailed analysis of this issue for my data can be found in Appendix A. Income ranks are found to be the preferred choice, and thus are used exclusively in the regional part of this paper.

2.2 The relationship between income ranks

A different approach to measuring intergenerational income mobility is to use income ranks instead of log incomes. Children are ranked based on their average lifetime income relative to other children in the same birth cohort. Parents are ranked similarly, based on their average lifetime incomes relative to other parents with children in the same cohort. Importantly, observations with zero income do not need any special treatment here (Dahl and DeLeire, 2008). The ordered income levels are then transformed into percentile ranks[5] (normalized fractional ranks). The following equation is then estimated by OLS:

$$R_f^c = \alpha + \beta R_f^p + \varepsilon_f^c \tag{4}$$

where R_f^c and R_f^p are the rank of the child and parents in family f, respectively. The coefficient β is equal to the correlation coefficient between the ranks since, by construction, the ranks are uniformly distributed. Both the IGE and the rank-rank slope show the degree of dependence between parent and child average lifetime income. The measures differ conceptually when inequality is larger in the child generation compared to the parent generation: with growing inequality, moving one rank down will correspond to a larger income loss in absolute terms since the distance between ranks increases.

When estimating rank-rank relationships on the regional level below, the national

[4] Recently, Nybom and Stuhler (2015) studied attenuation- and life-cycle bias for the most common measures of income mobility in the literature. Based on nearly career-long income histories for a sample of two generations of Swedish individuals, they found that measurement bias is much less of a problem when using income ranks (as described in the next section), compared to log incomes.

[5] Fractional ranks are calculated by cohort and then normalized to span from 0 to 100. In case of ties all observations with identical incomes share the mean rank for this group. For example, if 20% of a cohort had zero income, they all would be assigned a percentile rank of 10.

ranks assigned to each individual remain the same. If we were to use regional ranks instead, i.e. order individuals within each region, we would have a hard time interpreting the results: what does it mean that sons from low-income families in Stockholm reach on average the 38th percentile rank (within Stockholm), while sons from low-income families in Göteborg reach on average the 35th percentile rank (within Göteborg)? Is the income level at the 38th percentile within Stockholm higher or lower than the 35th percentile within Göteborg? Using national ranks, we create a common unit that makes a regional comparison meaningful.

I analyze two mobility measures on the regional level, relative and absolute mobility. Relative mobility is the difference in mean outcomes of the children from parents at the highest and lowest rank respectively:

$$\bar{R}^c_{100,r} - \bar{R}^c_{0,r} = 100 \times \beta_r \tag{5}$$

where $\bar{R}^c_{p,r}$ is the average child rank at percentile p in region r and β_r is the rank-rank slope parameter in region r. Relative mobility is thus just the re-scaled rank-rank slope coefficient in region r.

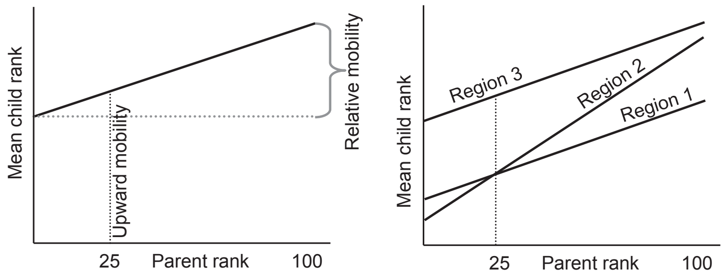
Absolute mobility is defined as the mean rank of children with parents at a certain percentile of the parent distribution (looking not only at the slope but also at the intercept). This measure adds no information on the national level since it is mechanically related to the rank-rank slope (Chetty et al., 2014a, p. 1562). However, by keeping the national income ranks for the regional analysis, however, incomes in one region can be assumed to have no influence on the national income distribution.

This absolute measure can be used to describe upward mobility, defined as the expected income rank of children from families with below-median income. This measure of mobility focuses on the absolute outcomes of the poorer half of the population, and is not affected by relative income changes between children in different percentiles. Since we assume that the rank-rank relationship is linear, this equals the mean child rank of children with parents at the 25th percentile, which we can predict using the estimated regional coefficients:

$$\bar{R}^c_{25,r} = \alpha_r + \beta_r \times 25. \tag{6}$$

The left panel in Figure 1 illustrates relative and upward mobility. The former is given by the difference in mean child rank (Y-axis) between parents with the highest and lowest income rank (X-axis), while the latter is measured by the mean child rank for parents at the 25th percentile. The right panel shows three example regions. Region 1 and Region 3 share the

Figure 1: Relative and Upward Mobility



Note: The left figure illustrates relative- and upward mobility. Relative mobility is the difference between the expected outcome of a child with parents in the top of the income distribution and a child with parents and the bottom of the income distribution. Upward mobility is the expected income rank of a child with below-median parent income. The right figure shows the association between child and parent income for three different regions. Regions 1 and 3 exhibit the same level of relative mobility, and regions 1 and 2 share the same level of upward mobility. Regions 1 and 3 would be indistinguishable from each other when using purely relative measures.

same level of relative mobility, i.e. the mean difference in ranks for children from the top and bottom of the parent income distribution is the same. However, mobility differs in absolute terms: for every parent percentile, the mean child rank is higher in region 3. Region 1 and Region 2 have the same level of upward mobility. Children with parents below the median reach, on average, the same national income rank. However, relative mobility is lower in Region 2 which can be seen by the steeper rank-rank slope. Note that a steeper rank-rank slope means a lower level of relative mobility. The steeper the slope, the stronger the association between the two generations' incomes, and thus the lower the intergenerational mobility.

It is important to be aware of which aspects the mobility measures above can and cannot capture. The IGE, the slope coefficient of a regression of log incomes, takes into account both the correlation between log incomes and the spread of the child and parent income distribution, since it is equal to

$$\beta = \frac{Cov\left(y_f^C, y_f^P\right)}{Var\left(y_f^P\right)} = \frac{Cov\left(y_f^C, y_f^P\right)}{\sigma_P \sigma_C} \frac{\sigma_C}{\sigma_P} = corr\left(y_f^C, y_f^P\right) \frac{\sigma_C}{\sigma_P}, \tag{7}$$

where $\sigma_{C(P)}$ is the standard deviation of the child (parent) distribution. On the other hand the rank-rank slope is just equal to the correlation coefficient between the income ranks because, after transforming income levels into percentile ranks, incomes in all generations are uniformly distributed between 0 and 100 and the ratio of standard deviations cancels out.

If income inequality had grown more from one generation to the next everything else equal (i.e., an increase in σ_C only), the IGE would be larger while the rank-rank slope would not change. A change in the mean of the income distribution (a shift of the complete distribution to the left or right) will show up in neither the IGE or the rank-rank slope. Yet, any measure based on absolute outcomes such as upward mobility will reflect such a shift, since absolute mobility makes use of the regression's constant to predict outcomes.

2.3 Regional Estimation

The estimation of rank-rank slopes and intercepts by region can be implemented in a variety of ways. The simplest one would be to estimate R different equations as in equation 4 for regions $r = 1, \ldots, R$ by OLS, resulting in R different slopes and intercepts (as done in Chetty et al. (2014a)). Let us call this the no-pooling case. Ignoring the regional information completely and estimating the equation for the whole sample as one group would give us one slope estimate and one intercept, i.e. the national estimates. We can call this the complete pooling case, for further reference below.

A third and potentially better alternative is to recognize not only the grouped nature of the problem at hand (individuals are sorted into different regions), but to explicitly model this relationship by taking into account both the within- and the between-region variances using a multilevel (or hierarchical) model. Multilevel models are widely used in political sciences (modelling for instance election turnouts or state-level public opinion, see for example Lax and Phillips (2009), Galbraith and Hale (2008), Shor et al. (2007), or Steenbergen and Jones (2002) for an overview) and in the context of education (students are grouped into class rooms and class rooms into schools and school districts, see for example Koth et al. (2008)). The terminology and notation below follow Gelman and Hill (2006).

The multilevel model is characterized by a level-1 equation for the smallest units, in this case relating child income rank to parent income rank for family f, and a set of level-2 equations for the larger units, here being the regions. The level-2 equation models explicitly the intercept and slope coefficients by region r:

$$R_f^c = \alpha_r + \beta_r R_f^p + \varepsilon_f^c \tag{8}$$

$$\alpha_r = \gamma^\alpha + \eta_r^\alpha \tag{9}$$

$$\beta_r = \gamma^\beta + \eta_r^\beta \tag{10}$$

where ε_f^c, η_r^α and η_r^β are random errors centered around zero and with variances σ_R^2, σ_α^2, and σ_β^2. Another common and equivalent way to write this model is

$$R_f^c \sim N\left(\alpha_r + \beta_r R_f^p,\, \sigma_R^2\right), \text{ for } f = 1,...,F \tag{11}$$

$$\begin{pmatrix} \alpha_r \\ \beta_r \end{pmatrix} \sim N\left(\begin{pmatrix} \gamma^\alpha \\ \gamma^\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix}\right), \text{ for } r = 1,...,R \tag{12}$$

which emphasizes the fact that the coefficients α_r and β_r are given a probability distribution with means and variances estimated from the data. Substituting equations 9 and 10 into equation 8, the model can be re-expressed as a mixed model

$$R_f^c = \gamma^\alpha + \eta_r^\alpha + \gamma^\beta R_f^p + \eta_r^\beta R_f^p + \varepsilon_f^c \tag{13}$$

where, in multilevel terminology, the γ's are "fixed effects" (= averages across all regions) and the η's are "random effects" (= draws from the estimated distributions).[6]

The multilevel model appears similar to a random- or fixed effects model often used in economics, but there are some important differences. We could for instance estimate a fixed effects model by simply adding $2 \times (R-1)$ regional dummies to equation 4, for regional intercepts and slopes. This approach would basically control away all between-region differences. In a multilevel model, the between-regions variance is explicitly estimated from the data and used to predict the regional effects. Also, if there are only a few observations in some regions, the estimates using regional dummies will be inefficient. The multilevel model on the other hand makes use of all observations when estimating the variance components and leads therefore to more precise estimates when there is little within-region variance.

If we were interested in the general effect of parent income on child income but were worried about regional unobservables, adding standard fixed- or random effects would be a good solution. However, since we are particularly interested in the mobility estimates for each of the regions, here the multilevel framework is the better choice.

In the second model, I add five regional types (as described in section 5.1 below) as a regional level predictor in the form of dummies to equations 9 and 10:

$$\alpha_r = \gamma_1^\alpha + \sum_{i=2}^{6} \gamma_i^\alpha T_i + \eta_r^\alpha \tag{14}$$

$$\beta_r = \gamma_1^\beta + \sum_{i=2}^{6} \gamma_i^\beta T_i + \eta_r^\beta. \tag{15}$$

This gives the following mixed model:

$$R_f^c = \gamma_1^\alpha + \eta_r^\alpha + \sum_{i=2}^{6} \gamma_i^\alpha T_i + \gamma_1^\beta R_f^p + \sum_{i=2}^{6} \gamma_i^\beta T_i R_f^p + \eta_r^\beta R_f^p + \varepsilon_f^c \tag{16}$$

[6]Note the important differences in terminology between fixed-effect models in economics and multilevel modelling. While a different intercept per region would be termed a regional-fixed effect in the former, it is a random effect in the latter. Only estimates of the average coefficients across all regions like the γ's are thought of as fixed here.

which allows the type of region during childhood to have an effect on both regional intercepts and slopes via $\sum_{i=2}^{6} \gamma_i^{\alpha}$ and $\sum_{i=2}^{6} \gamma_i^{\beta}$.

In the no-pooling case, the α_r's and β_r's in equation 8 are the OLS estimates from separate regressions, varying completely freely from each other. In the complete pooling case, the α_r's and β_r's are constrained to one common α and β. Here in the multilevel model where equations 8 - 10 are fitted simultaneously by (restricted) Maximum Likelihood estimation, the α_r's and β_r's are given a "soft constraint": they are assigned a probability distribution given in 12, with mean and standard deviation estimated from the data, which pulls the coefficient estimates partially towards their mean (termed shrinking).

The amount of pooling depends on the number of observations in each group as well as the between-regions variance of the parameters. For example, an estimate of a regional intercept can be expressed as a weighted average between the mean across regions γ^{α} (complete pooling), and the average of the R_f^c's within the region $\bar{R}_r^c$ (no-pooling):

$$\hat{\alpha}_r^{\,multilevel} = \omega_r \hat{\alpha}^{complete-pooling} + (1-\omega_r)\,\hat{\alpha}_r^{\,no-pooling}. \tag{17}$$

$$\hat{\alpha}_r^{\,multilevel} = \omega_r \gamma^{\alpha} + (1-\omega_r)\,\bar{R}_r^c \tag{18}$$

where the pooling factor ω_r is calculated according to

$$\omega_r = 1 - \frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_R^2/n_r}. \tag{19}$$

Thus, the intercept in a region with few observations is assumed less reliable and pulled towards the average value of all regions. The estimate for a region with many observations on the other hand will usually coincide with a separate OLS regression.

This is the main argument for using multilevel modelling in this particular study: there are many regions in Sweden with relatively few observations. The large regions have more than 400 times as many observations as the small regions. A separate regression for those small regions leads to extreme mobility estimates with large standard errors. In other words, we would not trust those estimates (even though they might seem appealing since we could report some exceptionally low and high levels of intergenerational mobility). Another useful aspect of multilevel models is that it is possible to include regional-level indicators along with regional-level predictors, which would lead to collinearity in OLS.

The model is built step wise, starting with a random intercept per region and the adding then random slopes and predictors. After each step, the model is assessed using a log-likelihood ratio test to assess if the model is a better fit to the data compared to classical regression (first model), or a better fit compared to the previous step.

Maximum likelihood estimation is used to fit the model. The "fixed effects" (regional average) parameters of intercept and slope given by the gammas in equation 12 are analogous to standard regression coefficients and are directly estimated. The regional effects given by the etas are not directly estimated but summarized in terms of their estimated variance and covariances. The best linear unbiased predictors (BLUPs) of the regional effects and their standard errors are computed based upon those estimated variance components as well as the "fixed effects" estimates. [7]

3 Data and Variable Descriptions

The data in this study comes from the SIMSAM database at Umeå University (Swedish Initiative for Research on Microdata in the Social And Medical Sciences). SIMSAM combines several different Swedish micro data registers; the population, geographic and income registers used in this study are provided by Statistics Sweden. A detailed description of the sample, the income variable used, and the geographical unit used for the regional analysis is given below.

3.1 Sample Selection and Income

My population sample consists of all individuals born in Sweden between 1968 and 1976, in the following termed children (927,008 observations before applying any restrictions). Due to the Swedish centralized registration system 99.5 percent of those children can be linked to their fathers and mothers. The age of the parents at their child's birth is restricted, 16 to 40 years for fathers and 16 to 36 years for mothers.[8] This makes it possible to observe parental income from their early/mid thirties onward while including 95 percent of the sample.

The income variable used here is the sum of taxable income from employment, self-employment, and transfers from the Swedish Social Insurance Agency.[9] The taxable transfers include parental benefits, pension payments and sick pay, and are labor market and -income related. Using this income variable on the individual level, the focus lies on how the child's ability to earn income is related to parental earned income.

Chetty et al. (2014b) instead measure family income for both children and parents (average income of two adults if married). This is potentially problematic since this measure is more affected by assortative mating. What one might be measuring in this case is the relationship between parental income and a child's ability to find a high income partner.[10]

Annual earned income can, in principle, be observed for each individual (children and parents) over the time period 1968 to 2010. All income observations are expressed in

[7] The exact estimation method used in this study is restricted maximum likelihood (REML). The basic idea behind REML estimation in this context is that one can form a set of linear contrasts of the response that do not depend on the estimated "fixed effects" (gammas), but instead depend only on the variance components to be estimated. One can then apply maximum likelihood methods by using the distribution of the linear contrasts to form the likelihood. If estimation is done by REML, the predicted standard errors of the BLUPs account for uncertainty in the estimate of the gammas which leads to slightly larger and more conservative standard errors. See for example Thompson (1962), Bates and Pinheiro (1998), Steenbergen and Jones (2002), and StataCorp (2013) for further technical details of the estimation procedure.

[8] I can choose a slightly smaller age span for mothers without loosing observations since there is less variance in mothers age at birth.

[9] "Sammanräknad förvärvsinkomst"

[10] See Ermisch et al. (2006) for a specific treatment of assortative mating and intergenerational mobility,

2010 SEK. Income and earned income are used interchangeably below.

I follow the literature discussed in section 2 and approximate average parental lifetime income by averaging over a large number of annual incomes. For 96 percent of the parents, I have 17 consecutive income observations available from when they were 34 to 50 years old. The smallest number of income observations available is 10, for fathers born in 1928 (only 0.1 percent of all fathers). On the mothers' side, 99.6 percent have 17 consecutive years of income observations, from 34 to 50 years. Due to the restriction to 36 years, even the oldest mothers born in 1932 have observable income records for 15 years. Parents missing too many income observations are dropped from the sample.[11] The great advantage here compared to earlier studies is that I measure parental income at approximately the same age for each parent, as well as over a very long time span. Averaging instead over the same calendar years for everyone (i.e. 2010 - 2012) as done in many other studies would give a biased measure: we would underestimate average income for young parents and overestimate average income for old parents, and even include some parents who are already retired.

For the children I have naturally fewer income observations available. Following the results by Bhuller et al. (2011) and **?**, I choose to approximate sons lifetime income by taking the average over three years when 32 to 34 years old.[12] All children missing more than one observation are deleted from the sample (3.5 percent).[13]

In my sample, around 45 percent of the daughters in each cohort receive maternal leave payments in their early thirties. Due to the flexibility of parental leave days as well as the income cap used to calculate the amount of parent leave payments in Sweden, it is very difficult to impute a reliable "true" income for the daughters in their early thirties based on the available information in my data set. The observed income is potentially a bad proxy for annual lifetime income for daughters and, as a consequence, I will report only the baseline estimates for them.[14] For those estimates I use the same approximation of average lifetime income as for boys, namely the average income between the age of 32 and 34.

Table A.2 in the Appendix summarizes the sample. The average age at child birth is 26 for mothers and 28 for fathers, and has increased slowly but steadily over the observed

[11] Parents are allowed to miss at most 4 years of income observations when having 15 - 17 observations (most parent cohorts), and then a decreasing number of years until missing at most 1 observation (for parents having only 10 years of income observations). Depending on the cohort, 97 or more percent of the parents fulfill this requirement.

[12] Bhuller et al. showed with Norwegian register data that annual earnings when 32-33 years old most closely reflect men's lifetime earnings. Nybohm and Stuhler (Table 4) showed for Swedish male cohorts 1955-1957 that the correlation between annual and the average of annual incomes over a lifetime is close to one when using a three year average around the age of 33.

[13] Missing observations occur for example when living (temporarily or permanently) abroad or after death,

[14] Böhlmark and Lindquist (2006) studied the development of annual versus lifetime income separately for men and women. While their results for sons was similar to the findings in the above studies, the income trajectories for women follow quite a different pattern that, along with the increasing labor market entry, also changes strongly over time.

time horizon. There are roughly between 80,000 and 90,000 children in each cohort and 774,953 children (and their parents) in total.

Table 1: Income Distributions

N = 781,630	Mean	St.dev.	Min	Max	p50	p90/p10
Son income	312,515	(169,699)	0	9,841,421	297,163	3.3
Daughter income	223,147	(104,796)	0	5,361,351	214,724	3.2
Children income	269,171	(148,835)	0	9,841,421	254,923	3.6
Mother income	157,958	(72,146)	0	5,565,328	156,356	3.3
Father income	269,116	(132,778)	0	15,160,164	244,572	2.5
Parent income	427,074	(161,288)	0	15,288,524	404,252	2.2

Note: All incomes are expressed in 2010 SEK.

Table 1 shows an income summary for the individuals in the data (all cohorts pooled). Mothers earned on average about 60 percent of fathers' incomes (but only 36 percent in terms of the highest income). The daughters distribution has come somewhat closer to their male counterparts compared to the previous generation but a strong difference persists both in level and variance. However, since daughters' incomes are affected heavily by maternity leave during the observed income years the numbers are not fully comparable. The last column shows that income inequality has risen from the father to the son generation, dividing the 90th income percentile by the 10th percentile increased from 2.5 to 3.3.

3.2 Geographic Unit

The geographic unit I choose to work with is the local labor market region, or LLM. An LLM is a self-sufficient area in terms of labor within which individuals live and work, and thus spend most of their time. The aggregation of municipalities into LLMs is taken from Statistics Sweden which measures commuting flows between municipalities. The aggregation into local labor markets corresponds most closely to the commuting zones which are used by Chetty et al. (2014a) for the US.

Studying local labor markets is a first step towards measuring the effect of immediate conditions (family, neighborhood), the local community (school quality, for example), and the larger metro area which is picking up for example labor market conditions. Using smaller geographical units such as municipalities there is a larger risk of selection bias due to residential segregation, i.e. that families sort themselves into certain residential areas and municipalities. A local labor market area contains several municipalities and probably several different residential areas, with different types of families. There are currently 75 LLMs in Sweden (112 in 1990 due to increasing commuting patterns), containing on average 4 municipalities and a population of 90,000. In contrast, there are 741 commuting zones in the US containing on average 4 counties and a population of 380,000.

In addition, I use five different regional types, based upon the “regional families” classification of local labor markets by The Swedish Agency for Economic and Regional Growth. The five regional types (T1-T5) are large cities (such as Stockholm), large regional centers (such as university cities), small regional centers (small cities employing a large share of the population in the surrounding rural areas), sparsely populated regions (less than 6 people per square kilometer), and other small regions (ranking in between small regional centers and sparsely populated regions). A complete list of local labor market regions and their type classifications can be found in table A.3 the Appendix.

Research by Cunha and Heckman (2007), Cunha et al. (2010), and Heckman (2007) indicates that the early environment is important in the human capital formation of children. Early investments generate not only human capital directly but also lead to higher returns to later investments. Other potentially important factors influencing the accumulation of human capital and life time income are the school environment and peers (Lavy et al., 2012), the home and neighborhood environment (Chetty et al., 2015), the availability of adult role models, and receiving guidance when choosing higher education or career paths during the teenage years.

I therefore assign children to the local labor market region in which they lived for at least six years between the age of 6 and 15 (ignoring moves within a local labor market), in order to capture both some influences during earlier as well as some teenage years. Using the strict assignment rule of a minimum of 6 years in the same region, we can be sure that a child was actually exposed to this location a significant portion of her childhood and that studying regional differences in mobility is meaningful.[15] The regional sample now includes 97.5 percent of all sons in the data (393,715 individuals), while 2.5 percent moved too often to determine a childhood region.

4 Mobility on the National Level

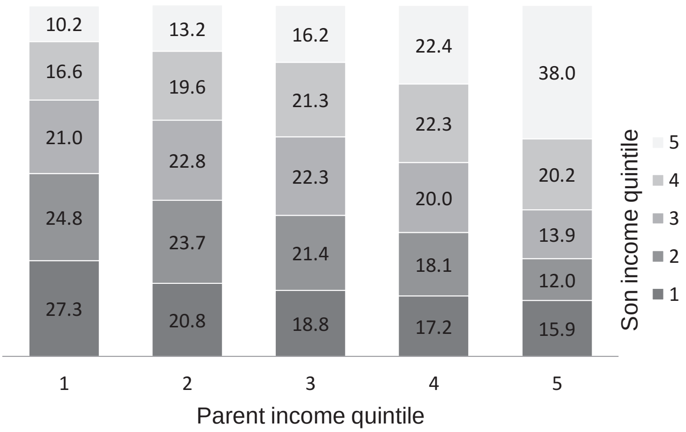
In this section, I summarize the national mobility estimates. First, by simply looking at transition matrices we can see the outcomes for Sweden as a whole in terms of son lifetime income given parent income, both for all cohorts pooled and over time. Next, I present the IGE and rank-rank slope estimates for different family member combinations and discuss how those estimates relate to earlier findings in the literature. Lastly, I decompose the IGE to show that the association between the incomes of two generations has weakened over the observed time period in Sweden.

[15] Sensitivity tests show that, in a country like Sweden with relatively few observations in a large number of regions, the age at which we establish the regional assignment can matter. Assigning children to the region where they lived between 0 and 3 years, or rather at age 15 as done in Chetty et al. (2014), changes the mobility results for the smaller regions. See Figure A.3 in the Appendix for a sensitivity test of the childhood definition.

4.1 Non-Parametric Description of Mobility

A non-parametric description of intergenerational mobility is a good starting point for the analysis of joint distributions of incomes (see for example Fields and Ok (1999) and Jäntti and Jenkins (2015)). Figure 2 shows a visual transition matrix for sons and their parents. 27.3 percent of the sons with parents in the first quintile are themselves located in the first quintile (of their own income distribution) as adults. Just over 10 percent of sons from the poorest fifth of the parents will reach the top quintile. Mobility of sons with parents in the first quintile is higher than in the US (using the quintile transition matrix given in Chetty et al. (2014a, p.1577)) where 33.7 percent of the children stay in the first quintile and only 7.5 percent of those starting at the bottom reach the top quintile.

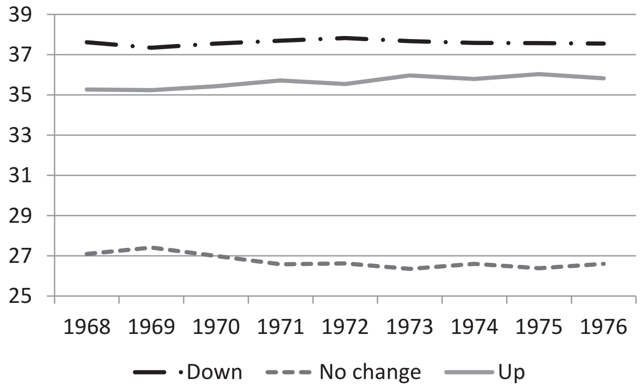
Figure 2: Quintile Transition Matrix: Parents and Sons



Looking at the upper end of the parent distribution, 15.9 percent of their children fall to the first quintile, while 38 percent stay at the very top. This is particularly interesting, again, since the numbers shown here are entirely depicting earned income and not wealth. In the US study, the numbers are 10.9 percent who move down four quantiles, and 36.5 percent who stay in the top, respectively.

Figure 3 shows the fraction of sons who reach the same, a higher, and a lower quintile in their own distribution compared to their parents over time, i.e. for each of the nine cohorts separately. For each year, the three lines add up to 100 percent (for every individual that moves up, another one has to make room and move either also up, or down). The largest group (37.5 percent) represents the fraction of sons who are doing worse compared to their parents. There are slightly less sons in the 1976 cohort compared to the older cohorts who do just as well as their parents (from just above 27 down to 26.6 percent). The group of individuals who change places in the income distribution contains in the later years more sons that do relatively better than their parents, which can be seen in the slight upward trend of the "Up" curve.

Figure 3: Quintile Mobility over Time



4.2 National Mobility Estimates

The base line mobility results for different family member combinations are shown in table 2. Both the IGE and the rank-rank slope show very little dependence between the incomes of mothers and their children. The IGE is largest between sons and parents (and mobility therefore the lowest), while the rank-rank slopes indicate the relation between sons and fathers to be the least mobile. A ten percentile points increase in father's income rank implies on average a 2.45 percentile increase in the son's income rank.

Table 2: Mobility Estimates for the Pooled Sample

	IGE			Rank-Rank slope		
	Child	Son	Daughter	Child	Son	Daughter
Parents	0.302	0.326	0.279	0.198	0.238	0.206
	(0.002)	(0.004)	(0.003)	(0.001)	(0.002)	(0.002)
N obs	766,404	394,157	372,247	774,953	399,095	375,858
Father	0.217	0.253	0.180	0.192	0.245	0.180
	(0.002)	(0.003)	(0.003)	(0.001)	(0.002)	(0.002)
N obs	766,338	394,122	372,216	774,953	399,095	375,858
Mother	0.071	0.061	0.084	0.101	0.101	0.133
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
N obs	765,327	393,590	371,737	774,953	399,095	375,858

Note: Standard errors are given in parentheses.

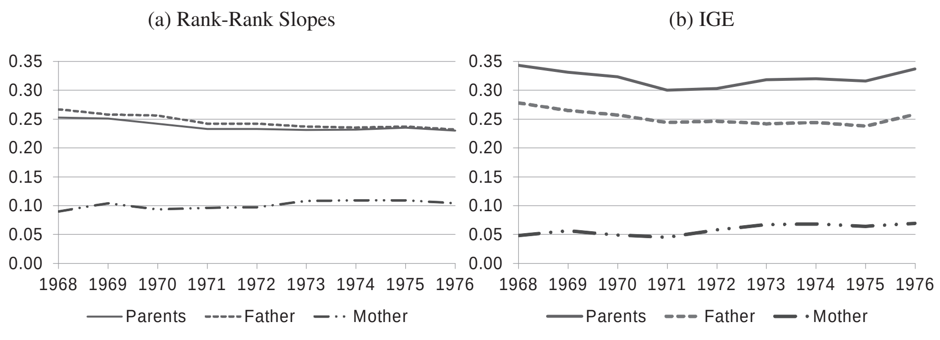
The estimated IGE for sons and fathers, 0.253, is in line with previous results. Nybom and Stuhler (2016) got an estimate of 0.27, based on a sample of 3,504 Swedish sons born between 1955 and 1957. The two main differences to their study are that their income measure is total pre-tax income which includes capital realizations, and fathers older than 28 years at their son's birth are excluded from the sample. The effects of those might however work in opposite directions which could explain the similarity to this study's result.

Björklund and Jäntti (1997) estimated the IGE to be 0.216 between fathers and sons. Their sample was quite different from the one used here: no actual father and son pairs were observed; rather two independent samples for both groups were combined. Their income measure was earnings; a five-year average for the fathers and one single observation for the sons.

Österberg (2000) also presented results for daughters and mothers. There are several ways her sample differed from mine. Incomes were only observed over three calendar years where the parents were up to 65 years old - and thus possibly already retired. Many children in the sample were under 33 years when their income was measured. Her estimate for the IGE between sons/ daughters and fathers (0.13/ 0.071, respectively) as well as for sons/ daughters and mothers (0.022/ 0.036, respectively) are substantially smaller in magnitude than my estimates which might be caused by attenuation and life cycle bias.

The IGE and rank estimates for parents and their children (first row in Table 2) are generally larger than the estimates for father and mother separately. This suggests an important role of the parental income combination, or parent income matching, for income transmission between generations as opposed to just the sum of parents' incomes. Investigating this finding further would be an interesting direction for future research.

Figure 4: Intergenerational Mobility over Time



Figures 4a and 4b show the development of the IGE and rank-rank slopes for sons and their parents separately by cohort. The association between mother and son income ranks has not changed considerably from the 1968 to the 1976 cohort and fluctuates closely around 0.1. The parents and father associations both show a negative trend, with the father estimates initially being larger than the parents estimates, but converging to the same value as for the parents in the 1976 cohort.

The IGE shows a different development. Both the association between fathers and sons, and parents and sons first declined and then increased again, while the IGE between sons and their mothers increased just slightly over the whole time span. Therefore according to the IGE we could conclude that the association between son and parent income for the youngest cohorts is just as high as for the oldest cohort, while the rank-rank slope would suggest a decline in the income association.

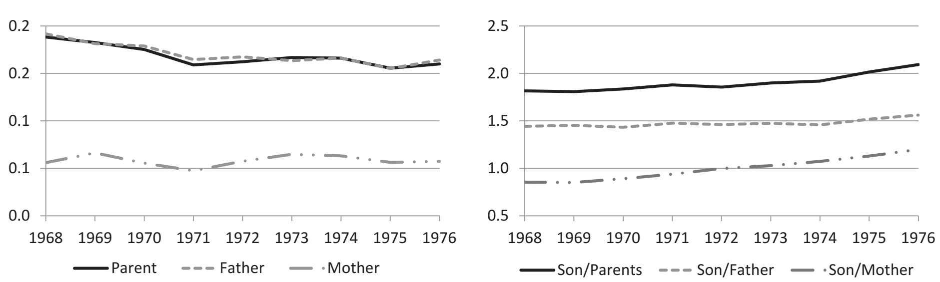
Equation 7 in Section 2.2 can help to explain this apparently contradictory finding: The rank-rank slope is simply the correlation coefficient between the percentile ranks of sons and parents, and the linear dependence between sons and their fathers as well as sons and their parents in terms of income rank has declined over time (and not been subject to a change between sons and their mothers).

The IGE on the other hand is the product of (a) the correlation coefficient between log child income and log parent income and (b) the ratio of their standard deviations. The development of both measures between 1968 and 1976 is shown in figure 5. We can see that the increase in IGE for the later cohorts is driven by an increase in the relative variance of the son income distribution compared to their parents, and not an increase in the linear dependence between the incomes of two generations.

Figure 5: Log Income Distributions

(a) Correlation Coefficients

(b) Ratio of Standard Deviations



5 Mobility across Regions

The multilevel analysis reveals some interesting facts about intergenerational mobility across Sweden. The first part in this section discusses the results from the multilevel model and their implications for relative and upward mobility on the regional level. In the second part, I focus on children growing up in very rural areas who are shown to have the lowest expected income ranks as adults.

5.1 Relative- and Upward Mobility on the Regional Level

The results of the multilevel model (1) from Section 2.3 can be summarized nicely by plotting the predicted slope- and intercept-random effects for each region, relative to the estimated average values across all regions (see Table A.4 in the appendix for the detailed estimation output). The slopes and intercepts are used in a later step to compute relative- and upward mobility according to the formulas in Section 2.2. As shown in figure 6a, at first glance the estimated slope-random effects vary greatly across Sweden. The regional slopes to the left with data points below the horizontal line are smaller than the average, and the regional slopes located above the line to the right are larger. However, most estimates are not significantly different from the average: most of the 95 percent confidence intervals (shown as error bars) include the horizontal line at zero which indicates the average intercept. Of all 112 regions, only 2 show a significantly flatter slope (weaker association between parent and son income rank), and 7 regions show significantly steeper slopes (stronger association). If we had used separate OLS regression for each region, we would probably have overstated the differences in rank-rank slopes over regions since there is no easy way to compare the estimates of many disjoint regressions.

As opposed to the slopes, the regional intercepts (shown in Figure 6b), differ significantly from the average in most local labor markets. Thus, we already know that mobility measures based on absolute outcomes will show large differences between regions, while relative measures might not reveal this information.

The fixed-effects (regional average) slope estimate shows that a ten percentile increase of parent income rank implies, on average, an increase of 2.2 percentile ranks for the son (see Table A.4 in the Appendix for the detailed estimation output). The average slope coefficient across regions is thus slightly smaller than the average slope coefficient across individuals from section 4.2 above (2.38).[16] With a correlation coefficient of -0.54, we also learn that regions with steeper rank-rank slopes on average have lower intercepts.

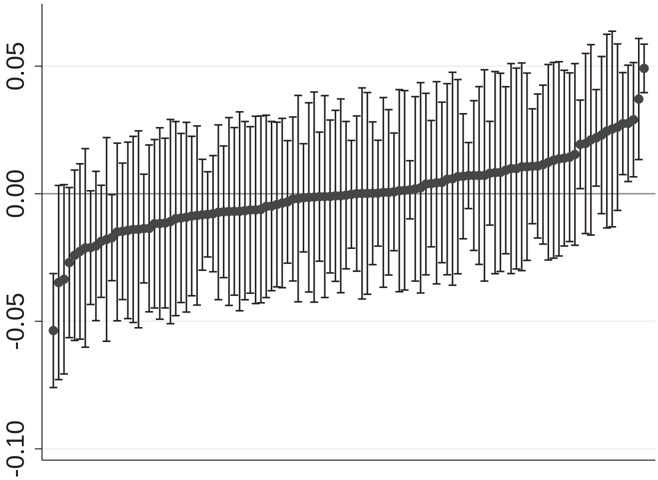
The relationship between the multilevel model, separate OLS regressions for each region, and the completely-pooled estimates are demonstrated in Figure 7. The top panel shows Dorotea, a local labor market region in the north of Sweden with 147 observations. The dotted line shows the mobility estimates from a separate regression: the line is almost completely flat and would indicate extremely high levels of relative income mobility. However, the large spread of the underlying binned scatter plot in gray shows the inefficiency of the estimation and thus how unreliable this result is. The best linear unbiased predictor (BLUP) from the multilevel estimation (given by the solid line) deviates from this extreme result and pulls towards the dashed line above, which represents the average mobility level across all regions.

The bottom panel in Figure 7 displays a similar figure for Stockholm. According to the gray dots of the binned scatter plot, the observations are located much closer to a line. The multilevel estimates (BLUPs) coincide here completely with the estimates from a

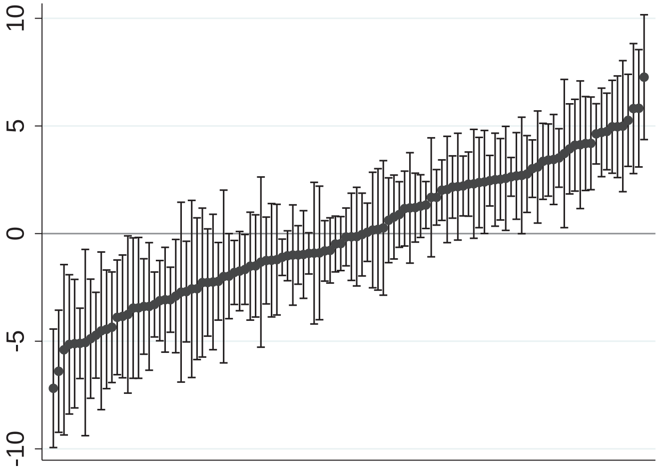
[16] It is possible that this difference is due to the 2.5 percent of sons who cannot be assigned to a childhood region: if sons with outcomes similar to their parents' are over-represented in the group of dropped observations, average mobility will increase (and the slope coefficient decrease).

Figure 6: Regional Effects

(a) Slopes



(b) Intercepts



Note: The upper panel shows the 112 random intercepts relative to the fixed effect, i.e. the average intercept across all regions (the horizontal line), sorted in ascending order from left to right. The error bars indicate 95% confidence intervals. Regions that include the horizontal line in their confidence interval do not differ statistically from the average. The lower panel shows a similar graph for the estimated random rank-rank slopes relative to the fixed effect slope.

regression run exclusively for children grown up in Stockholm (the solid line and the dotted line are indistinguishable from each other). With 67,228 observations there is no shrinkage toward the pooling-result happening at all.

When we add the five regional types (from large city to sparsely populated regions) to the model with large cities as the reference category, we find that only the additional "fixed effect" for sparsely populated regions is significant: growing up in an extremely rural local labor market region reduces the average income rank of sons by 4.4 percentile ranks. The rank-rank

slope is significantly different compared to large cities in both small regional centers and other small regions. Here, the association between son and parent income rank is slightly smaller which implies slightly higher relative mobility. Adding the regional types to the model results in a reduction of the variance between regions for the intercepts, but not for the slopes, as can be seen from the lower part of Table A.4.

I calculate relative- and upward mobility for each region using the predicted region specific intercept and slope if they are significantly different from zero, and the estimated average value otherwise. The results obtained this way can therefore be interpreted as a lower bound of the existing regional differences in mobility. The complete list of results by local labor market can be found in Table A.5 in the Appendix.

Relative mobility is higher than average in the two regions Umeå and Uppsala only. The expected outcome difference between top and bottom income families in Umeå is just 16.83 percentile ranks and 20.48 in Uppsala. Seven regions have a less than average relative mobility, with Stockholm ranking lowest. Here, the average outcome of sons from the highest and lowest income families differs by 27.11 percentile ranks.

Upward mobility varies from 36.32 in Torsby to 50.77 in Hylte[17], with an average of 43.69 across all regions (standard deviation 3). Calculating the percentiles back to income levels, we find that the expected difference in outcome between growing up in Torsby or Hylte for sons from below-median income parents amounts to 32.842 SEK less income per year (≈3.839 USD).

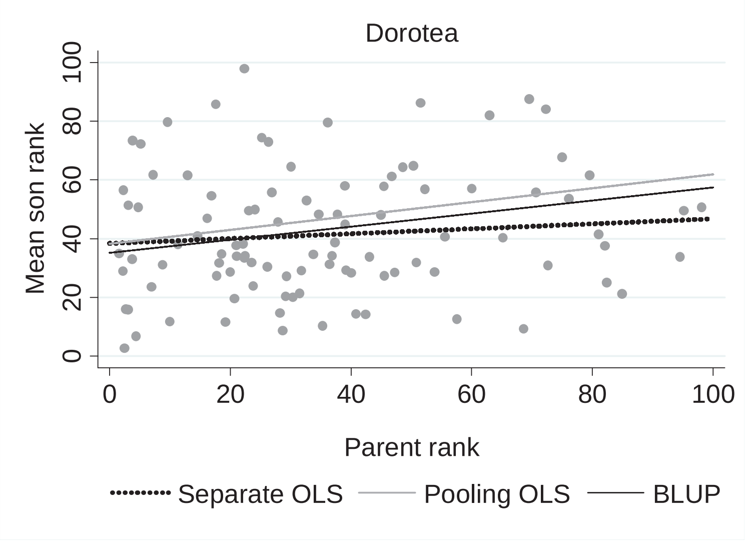
Figure 8 shows relative and upward mobility for all regions. The crossed lines through the center of the plot indicate the average levels of relative and upward mobility, respectively.[18] The arrow-tips indicate the direction in which mobility is increasing (note: high values of relative mobility indicate less mobility due to stronger associations between parent and child income). The quadrant marked with a large plus (minus) sign indicates regions with both above (below) average relative and upward mobility. The data point right in the center represents not one but 45 regions which all have average levels of both mobility measures.

[17] Torsby is located in middle-west Sweden bordering Norway, and Hylte at the west coast south of Göteborg

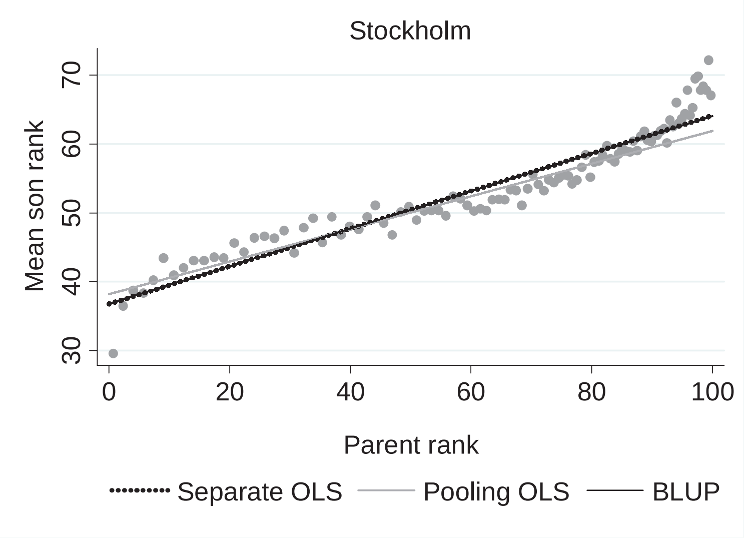
[18] Assuming average values whenever the 95 percent confidence intervals include zero for the regional random effects.

Figure 7: Comparison of Estimation Strategies

(a) Positive shrinkage



(b) No shrinkage

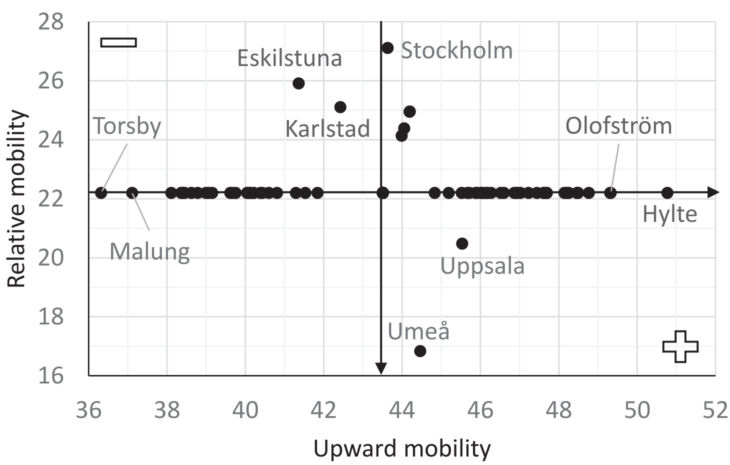


Note: The upper panel shows a binned scatter plot of son and parent income ranks for Dorotea, with three different fitted lines from (1) a separate OLS regression, (2) the national OLS regression, and (3) the Best Linear Unbiased Predictors from the multilevel model. The multilevel estimates are close to the national average and gives less weight on the within-LLM information. The lower panel shows a similar figure for Stockholm. The results from (1) and (3) are here indistinguishable from each other.

All regions with extremely high- or low levels of upward mobility show just average levels of relative mobility. Thus, even though the relative difference between sons from the highest and lowest income families in, for example, Torsby and Hylte, is the same, sons from below-median income families in those regions will end up with very different levels of income. Using the IGE or the rank-rank slope as the only measure for mobility, this difference would go completely unnoticed.

Stockholm and Malmö have very similar levels of upward mobility, i.e. the same expected outcome for below-median income families. However, due to the much steeper rank-rank slope, sons growing up in Stockholm achieve overall better outcomes than sons growing up in Malmö, and the difference is increasing in parent income rank. Thus, sons from high income families do better when growing up in Stockholm compared to Malmö. Figure A.4 in the Appendix shows the expected son outcomes over the whole parent income distribution for some of the local labor markets.

Figure 8: Relative and Upward Mobility by Region



Note: For each region, relative mobility is plotted against upward mobility. The lines of the crosshair indicate the average levels of the measures. The data point right in the center is actually an overlay of 45 regions, all with average mobility. The quadrant marked with a plus (minus) sign indicates areas with above-average (below-average) mobility levels according to both measures.

The estimates from this study can be compared to the results from Chetty et al. (2014a) for the US. There are some important differences between our studies, besides the obvious issues such as population size. The child cohorts in Chetty et al. are younger (born 1980 - 1982) and life time income for the children is measured at a slightly younger age (during 2010 - 2012 for all cohorts). Importantly, their income variable includes capital income and is calculated per family instead of individually. Parents were aged 15 to 40 at child birth and their income is measured between 1996 and 2000. This means that parents are between 29 and 60 when their income is measured (compared to the same age for all parents in this study). The US study includes both sons and daughters.

We can compare the middle 80 percent of the distribution of US commuting zones and Swedish LLMs in terms of upward mobility and relative mobility: upward mobility at the tenth percentile of all regions is 37.4 percentile ranks in the US and 39.1 percentile ranks in Sweden. Commuting zones at the 90th percentile in the US show upward mobility of 52 percentile ranks, Swedish LLMs 47.7. The mean (43.3 in the US and 43.7 in Sweden) and middle 80 percent of the distributions of upward mobility by commuting zone and LLM in the US and Sweden are therefore quite similar, although the tails are thicker in the US with more extreme values at the very top and bottom. Relative mobility on the other hand varies much more in the US, where the outcome difference in percentile ranks for children from families at the top and bottom of the distribution takes on values between 6.8 and 50.8. In Sweden, relative mobility across LLMs varies only between 16.8 and 27.1 percentile ranks.

It is important to keep in mind that Chetty et al. did not discuss how their individual regional estimates relate to each other and in how far they significantly differ from each other (or from the US-average). My results are more conservative both in the sense that my estimation method accounts for the number of observations (which gives less extreme results), and because I choose to compute the mobility measures using the regional predictions solely if they differ significantly from the Swedish average.

Since the income distribution is much more compressed in Sweden compared to the US, the distance between two ranks in terms of income levels is considerably smaller in Sweden. The monetary difference between the top and bottom 10 percent of US commuting zones in terms of upward mobility is 12,600 USD (labor and capital income), while the same difference between Swedish LLMS amounts to just 19,362 SEK (≈2,353 USD) (labor income only).

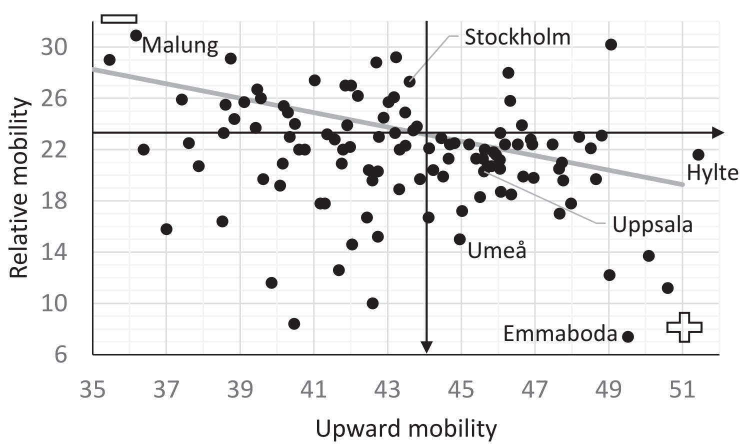
There are four local labor markets that stick out: Eskilstuna and Karlstad with both below average relative- and upward mobility, and Uppsala and Umeå with both above average relative- and upward mobility. Even though an in-depth analysis of the underlying forces driving this result is beyond the scope of this paper, we can look at one known factor correlated with mobility, namely income inequality. Countries with more income inequality have been shown to have less intergenerational income mobility. This relationship has become known as the Great Gatsby Curve, see for instance Corak (2013).

A simple indicator of income inequality is the ratio of median income to mean income level which informs us about the skewness of the income distribution. Across all municipalities in Sweden in 1991, weighted by population size, this measure is 0.9586, i.e. the median income level amounts to 95.86 percent of the mean income and the distribution is thus, as expected, right skewed. Looking at at this indicator separately for the local labor markets Umeå (96.93), Uppsala (95.72), Eskilstuna (96.79) and Karlstad (95.63), we do not find a pattern that could explain the very high/ low combined mobility results. Looking at relative mobility only, however, shows some evidence for the relationship between high income inequality and low relative mobility: The Stockholm region has the most right-skewed income distribution (the median income level is just 92.65 percent of the mean income level) and also has the lowest levels of relative mobility among all Swedish LLMs.

5.2 A Comparison to OLS

Figure 9 visualizes relative and upward mobility similar to Figure 8 in Section 5.1, but here based on 112 separate OLS regressions by region. As expected, the regions differ more in terms of both mobility measures compared to the multilevel approach.

Figure 9: Relative and Upward Mobility by Region using Separate OLS Regressions

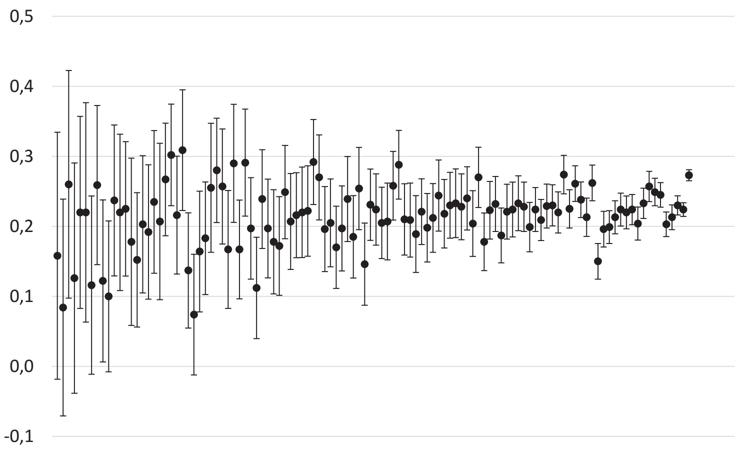


Note: For each region, relative mobility is plotted against upward mobility. The lines of the crosshair indicate the mean of each measure, calculated from the 112 OLS regressions. The gray line shows the fitted values from an OLS regression of the 112 relative mobility results on upward mobility, weighted by the number of observations in each region.

Relative mobility (the difference in mean outcome for sons from the families with the highest and lowest income, respectively) varies the most: from a 7.4 percentile rank difference in Emmaboda to 30.9 percentile ranks in Malung. However, as I have emphasize in this paper, interpreting those differences is not straight forward.

Figure 10 illustrates the rank-rank slope estimates underlying the mobility measures obtained by separate OLS regressions in figure 9 above. The regions are sorted in ascending order by number of observation from left to right. It is clear that the lowest and highest slope estimates are found on the left side of the graph, together with the largest standard errors. In addition, most regional slopes are statistically indistinguishable from each other. There is no obvious way to compare the estimates of the 112 regressions. Thus, based solely on those regressions, an interpretation of regional differences in mobility for Sweden is not very convincing.

Figure 10: Rank-Rank Slopes and their 95% Confidence Intervals



Note: Every black dot represents a point estimate of the rank-rank slope for one region (112 in total). The regions are sorted in ascending order from left to right. In general, the fewer inhabitants in a region (the more to the left in the graph), the less efficient is the slope estimate.

5.3 The Impact of Regional Types

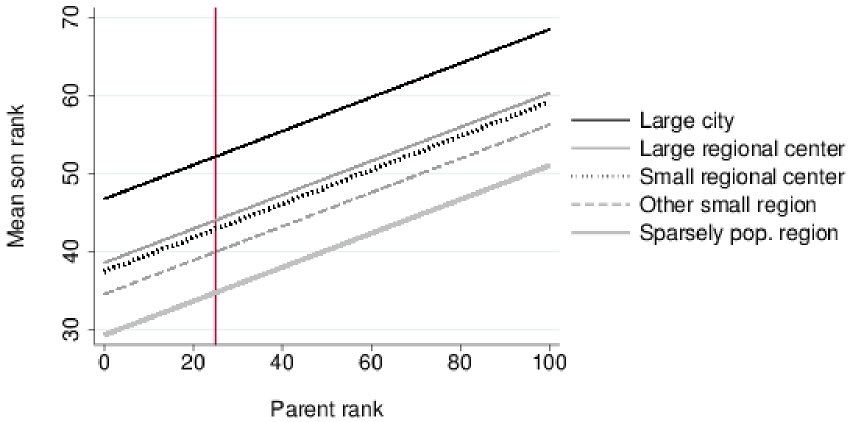
The model including regional types showed that very rural areas exhibit significantly lower intercepts and thus lower levels of upward mobility than other regional types. This can also be seen from Figure 8, where the data points furthest on the left (lowest upward mobility) all represent very rural local labor markets.

An interesting question is whether these lower absolute outcomes for sons growing up in the country-side persist independently of location choice later in life. Put differently, are there differences between sons from rural communities who move to the city and those who remain?

Figure 11 shows fitted regression lines for son income rank on parent income rank for the subsample of sons who grew up in a very rural (Type 5) area, by type of region those individuals lived in at age 33. It is clear that there are large differences depending on where individuals choose to live as adults: the average outcome of sons, for all parent income ranks, are clearly highest when moving to large cities, and smallest when remaining in a very rural area.

Table 3 shows the moving patterns of the individuals in my sample in percentages. The first row, for example, shows that more than 90 percent of the sons who grew up in a

Figure 11: Son and Parent Income Associations by Adult Location
(Childhood: Sparsely Populated Regions (T5))



large city (Type 1) also live there as adults, while less than 1 percent moves to rural areas (Type 4 and Type 5). The last row shows children from very rural areas. A very large share, 52 percent, lives in a very rural area even as an adult. Going back to the results for local labor markets, we can say that sons from rural areas do well when moving to urban areas. However, since only a few did so, the average outcomes for all sons from rural areas overall is low. [19]

Table 3: Moving Patterns between Regional Types

		Adult					
		T1	T2	T3	T4	T5	Total
Childhood	T1	92.24	5.28	1.92	0.30	0.25	100
	T2	31.87	63.58	3.41	0.67	0.47	100
	T3	24.84	25.42	47.02	2.14	0.57	100
	T4	18.98	28.50	13.88	36.09	2.56	100
	T5	16.38	25.44	5.19	1.39	51.61	100
	Total	49.00	34.59	11.72	2.74	1.95	100

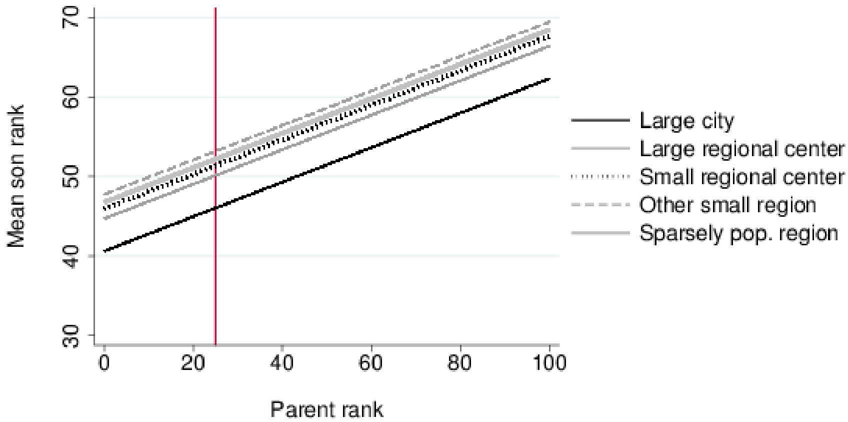
Note: The cells are relative frequencies, by childhood region type. The first row, for instance, indicates that 92.24% of sons who grew up in a large city remain there as adults.

The local labor markets with the highest levels of upward mobility (the data points to the right in Figure 8) are surprisingly also rather rural (Type 4), but not as remote as Type 5 areas. An analysis for the subsample of sons growing up in Type 4 regions by

[19] Note that I am not making a causal statement here: there are possibly very large differences between those sons that move to a city and those that stay (ability, education, occupation, etc.).

adult location shows a very similar patter as Figure 11: the more urban the adult location, the higher the average outcome. Since much fewer Type 4-children choose to stay in the country side compared to sons from Type 5 areas, however, the overall result of high upward mobility follows.

Figure 12: Son and Parent Income Associations by Childhood Location
(Adult location: Large Cities (T1))



Interestingly, we see that children who grew up in the country side actually do much better in absolute terms than those who grew up in the city, when comparing all children that live in large cities as adults (see Figure 12). This could be a result of sorting behavior: more able individuals growing up in rural areas might choose to obtain more education and, as a consequence, they move to larger cities that offer appropriate jobs and career opportunities. However, less able individuals who grow up in urban areas might not relocate (no active choice of relocation), which results in a different mixture of individuals with an urban childhood, compared to those who only moved to the city as adults.

6 Discussion and Concluding Remarks

In this paper I have used detailed population-wide register data on nine Swedish cohorts and their parents to draw a picture of intergenerational mobility in and across Sweden. I focused on estimating the intergenerational elasticity, as well as the relationship between percentile ranks from one generation to the next.

In line with previous literature, I focused on income measurements at ages where annual income is most likely to equal average life-time income. These measures were constructed by averaging over 17 consecutive annual income observations for parents (when they were 34 to 50 years old) and three annual income observations for children (when they were 32 to 34 years old). One advantage of the data is that I am able to measure parental income at approximately the same age for each parent, as well as over a very long time span.

For Sweden as a whole, the estimated IGE between parents and their children is 0.3. This implies that 30 percent of the difference between parent income and the average parent income is inherited by a child. The association between the log incomes of parents and sons is 0.33 and thus even stronger. Using income ranks, I found that a 10 percentile rank increase in parent income implies a 2.38 percentile rank increase of the son's income.

Interestingly, child income is more strongly associated with total parent income than with father income. In the case of the IGE, the child-parent association exceeds the sum of the individual elasticities between child and father and child and mother income. This suggests an important role of parent matching for income transmission, and is an interesting direction for future research.

Focusing on each cohort separately from 1968 to 1976, I found a decrease in the rank-rank association over time (reflecting an increase in mobility over time). This holds true both when considering the association between father income and son income as well as when considering the association between total parent income and son income. The IGE, on the other hand, suggested an *increase* in this association between 1971 and 1976 (and thus a *decrease* in mobility over time). Here the IGE is misleading since it mainly reflects the growing variance in the income distribution of sons over time.

In my regional analysis I estimated mobility measures (relative and upward mobility) for all Swedish local labor markets. This geographical unit is reasonable in the context of intergenerational mobility since it is small enough to account for immediate conditions and the local community, but large enough to capture more general local labor market conditions. The allocation of children to regions was based on the municipality where the child had lived for at least 6 years between the ages of 6 and 15. I also tested the robustness of alternative ways of allocating individuals to regions.

My primary measurement method for regional differences is a multilevel model. However, I also reported the results from separate regional OLS regressions. The multilevel analysis revealed that relative mobility, the difference in outcomes between two sons with parents in the top and bottom one percent of the national distribution, respectively, is 22.2 percentile ranks in most regions. The strongest association (lowest relative mobility) between son and parent income rank was measured in Stockholm, where the relative outcome difference is more than 27 percentile ranks. On the other end of the spectrum we found Umeå and Uppsala, with the weakest association between income ranks and a relative difference of 16.8 and 20.5 percentile ranks between sons from the highest and lowest income families, respectively.

It is fair to say that upward mobility varies greatly across Swedish local labor markets. The lowest upward mobility (the expected outcome for a son with below-median parent income) is 36.3, measured in Torsby. In Hylte, sons with below-median parent income can expect to reach the 50.8th percentile as adults, more than 14 percentile ranks higher than in Torsby. This difference between growing up in the most and least favorable regions corresponds to 32,842 SEK ($\approx$ 3,898 USD) in yearly income. When using only the IGE or rank-rank slopes to study mobility, these differences in upward mobility would be completely invisible.

Even though the Swedish income distribution is considerably more compressed than the US distribution (measured in terms of percentile locations), the distributions of regional upward mobility estimates for the two countries are similar. Relative mobility on the other hand varies considerably less in Sweden than in the US.

In regions that can be classified as very rural, son outcomes given parent income rank are more than 4 percentiles lower as compared to other, less rural areas. A closer look at the group of sons who grew up in those regions revealed that their outcomes differs greatly depending on where they live as adults. Moving to less rural areas lifts incomes significantly. However, less than half of them choose to move to more urban areas. Therefore the low absolute outcomes for this group as a whole remain.

Conversely, by studying the cohort of those who live in a large city as adults, we saw that children who grew up in rural areas actually do better compared to city kids. This could be a consequence of sorting behavior: highly educated individuals with a rural childhood may move to large cities in order to pursue professional careers, while urban kids might stay in the city independently of productivity, education, or occupation (over 90 percent of the sons who grew up in large cities still live there as adults).

Sweden is considered to be a country with exemplary high levels of intergenerational income mobility. My results show that there are large differences in terms of mobility across Sweden and that location does matter. The evidence provided here indicates that there are particularly large differences in the expected outcomes for children from the lower half of the income distribution, depending on childhood region and moving patterns. A general lesson of this study is that country-wide measures of income mobility might say very little about the state of mobility at a particular location within the country. Cross-country comparisons of income mobility should therefore be interpreted with caution.

References

Bates, D. M., and Pinheiro, J. C. (1998). "Computational methods for multilevel models." Technical Memorandum BL0112140-980226-01TM, Bell Labs, Lucent Technologies, Murray Hill, N.J.

Böhlmark, A., and Lindquist, M. J. (2006). "Life-Cycle Variations in the Association between Current and Lifetime Income: Replication and Extension for Sweden." *Journal of Labor Economics*, *24*(4), pp. 879–896.

Bhuller, M., Mogstad, M., and Salvanes, K. (2011). "Life-cycle bias and the returns to schooling in current and lifetime earnings." *NHH Dept. of Economics Discussion Paper*, (4).

Björklund, A., and Jäntti, M. (2009). "Intergenerational income mobility and the role of family background." *Oxford Handbook of Economic Inequality, Oxford University Press, Oxford*, 491–521.

Björklund, A., and Jäntti, M. (1997). "Intergenerational income mobility in Sweden compared to the United States." *The American Economic Review*, *87*(5), pp. 1009–1018.

Black, S. E., and Devereux, P. J. (2011). "Recent developments in intergenerational mobility." In *Handbook of Labor Economics*, vol. 4, Part B, chap. 16, 1487 – 1541, Elsevier.

Bratsberg, B., Røed, K., Raaum, O., Naylor, R., Jäntti, M., Eriksson, T., and Österbacka, E. (2007). "Nonlinearities in intergenerational earnings mobility: Consequences for cross-country comparisons." *The Economic Journal*, *117*(519), pp. C72–C92.

Chetty, R., Hendren, N., and Katz, L. F. (2015). "The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment." NBER Working Paper No. 21156.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014a). "Where is the land of opportunity? The geography of intergenerational mobility in the United States." *The Quarterly Journal of Economics*, *129*(4), i1.

Chetty, R., Hendren, N., Kline, P., Saez, E., and Turner, N. (2014b). "Is the United States still a land of opportunity? Recent trends in intergenerational mobility." *American Economic Review*, *104*(5), 141–47.

Corak, M. (2013). "Income inequality, equality of opportunity, and intergenerational mobility." *The Journal of Economic Perspectives*, *27*(3), pp. 79–102.

Couch, K. A., and Lillard, D. R. (2004). "Nonlinear patterns of intergenerational mobility in Germany and the United States." *Generational income mobility in North America and Europe*, 190–206.

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

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

Dahl, M. W., and DeLeire, T. (2008). *The association between children's earnings and fathers' lifetime earnings: estimates using administrative data*. University of Wisconsin-Madison, Institute for Research on Poverty.

Ermisch, J., Francesconi, M., and Siedler, T. (2006). "Intergenerational mobility and marital sorting." *The Economic Journal*, *116*(513), 659–679.

Fields, G. S., and Ok, E. A. (1999). *The measurement of income mobility: an introduction to the literature*. Springer.

Galbraith, J. K., and Hale, J. T. (2008). "State income inequality and presidential election turnout and outcomes*." *Social Science Quarterly*, *89*(4), 887–901.

Gelman, A., and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Haider, S. J., and Solon, G. (2006). "Life-cycle variation in the association between current and lifetime earnings." *American Economic Review*, *96*(4), 1308–1320.

Hausman, J. (2001). "Mismeasured variables in econometric analysis: Problems from the right and problems from the left." *The Journal of Economic Perspectives*, *15*(4), pp. 57–67.

Heckman, J. J. (2007). "The economics, technology, and neuroscience of human capability formation." *Proceedings of the National Academy of Sciences*, *104*(33), 13250–13255.

Jäntti, M., and Jenkins, S. P. (2015). "Income mobility." In A. B. Atkinson, and F. Bourguignon (Eds.), *Handbook of Income Distribution*, *Handbook of Income Distribution*, vol. 2, chap. 10, 807 – 935, Elsevier.

Koth, C. W., Bradshaw, C. P., and Leaf, P. J. (2008). "A multilevel study of predictors of student perceptions of school climate: The effect of classroom-level factors." *Journal of Educational Psychology*, *100*(1), 96.

Lavy, V., Silva, O., and Weinhardt, F. (2012). "The good, the bad, and the average: Evidence on ability peer effects in schools." *Journal of Labor Economics*, *30*(2), pp. 367–414.

Lax, J. R., and Phillips, J. H. (2009). "How should we estimate public opinion in the states?" *American Journal of Political Science*, *53*(1), 107–121.

Lindahl, M., Palme, M., Massih, S. S., and Sjögren, A. (2015). "Long-term intergenerational persistence of human capital: An empirical analysis of four generations." *Journal of Human Resources*, *50*(1), 1–33.

Mazumder, B. (2005). "Fortunate sons: New estimates of intergenerational mobility in the United States using social security earnings data." *Review of Economics and Statistics*, *87*(2), 235–255.

Nybom, M., and Stuhler, J. (2015). "Biases in standard measures of intergenerational income dependence." IFAU Working paper No. 2015:13.

Nybom, M., and Stuhler, J. (2016). "Heterogeneous income profiles and lifecycle bias in intergenerational mobility estimation." *Journal of Human Resources*, *51*(1), 239–268.

Shor, B., Bafumi, J., Keele, L., and Park, D. (2007). "A bayesian multilevel modeling approach to time-series cross-sectional data." *Political Analysis*, *15*(2), 165–181.

Solon, G. (1992). "Intergenerational income mobility in the United States." *The American Economic Review*, 393–408.

Solon, G. (1999). “Intergenerational mobility in the labor market.” In O. C. Ashenfelter, and D. Card (Eds.), *Handbook of Labor Economics*, vol. 3, Part A, chap. 29, 1761 – 1800, Elsevier.

Solon, G. (2004). “A model of intergenerational mobility variation over time and place.” In M. Corak (Ed.), *Generational Income Mobility in North America and Europe*, 38–47, Cambridge University Press.

StataCorp (2013). “Multilevel mixed-effects reference manual.” In *Stata: Release 13. Statistical Software*, College Station, TX: StataCorp LP.

Steenbergen, M. R., and Jones, B. S. (2002). “Modeling multilevel data structures.” *American Journal of Political Science*, *46*(1), pp. 218–237.

Österberg, T. (2000). “Intergenerational mobility: What do tax data show?” *Review of Income and Wealth*, *46*(4), 421–436.

Thompson, W. A. (1962). “The problem of negative estimates of variance components.” *The Annals of Mathematical Statistics*, *33*(1), 273–289.

Appendix

A Ranks versus Logged Incomes

In order to judge the use of logged incomes against income ranks, we can look at the fit of the data to a linear model, the distribution of zero income observations, and the sensitivity of the mobility estimates to how zero income observations are treated. Figure A.1a shows a binned scatter plot of logged incomes for parents and sons, as well as the fraction of sons with zero income. The first part of the plot is created by binning the parents into 100 equally sized groups by log income (percentiles), and plotting the mean parent income versus the child mean income for each bin. The vertical lines show the 10th and 90th percentile of parent income, respectively. The reported regression coefficients and standard errors are estimated by OLS using the underlying micro data.

The relationship between log incomes of parents and sons does not appear entirely linear. The slope at the bottom and top ten percent of the distribution is less steep than in the middle of the distribution. As can be seen in Figure A.1b, the non-linearity is even more pronounced in the relationship between log incomes of sons and fathers.

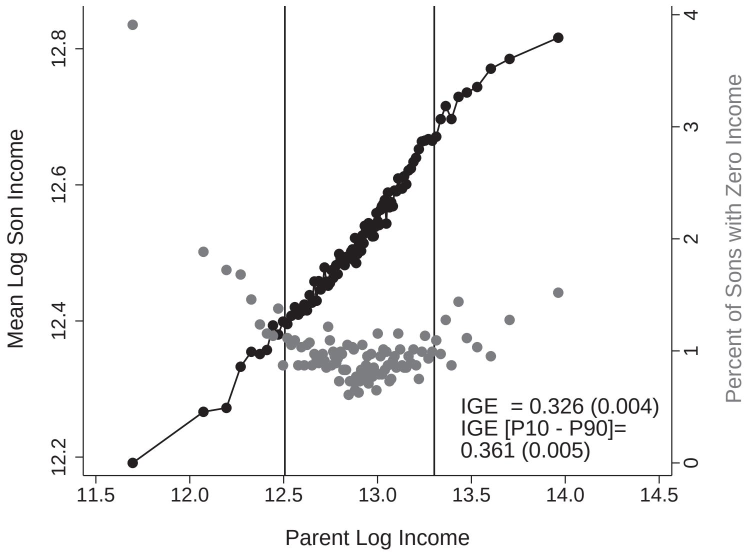
Figure A.2a shows a similar graph using percentile ranks instead of logged incomes, such that the mean son rank is plotted against mean parent percentile rank. For the pooled cohorts of sons and parents, the ends of the distribution diverge again visibly from the fitted regression line. The same is true for the relation between sons and fathers, although the fit in this case appears slightly better, see Figure A.2b. The finding by Chetty et al. (2014a) that rank-rank slopes have a better linear fit than the log-log relationship cannot be said to hold in general.[20]

Next, we can also see from Figure A.1a that the number of children with zero income is clearly highest for parents in the lowest ten percent of the distribution. Even the top ten percent of parents show a slightly higher number of sons with zero income compared to the middle 80 percent. An explanation for this might be the lack of information on capital income. Capital income tends to be concentrated at the top of the income distribution and could be a (here invisible) substitute for earned income. Since zero income children are over represented in the group of low-income parents (i.e., a strong association between parent and child income in the lower part of the distribution), dropping these observations would lead to an upward biased mobility estimate.

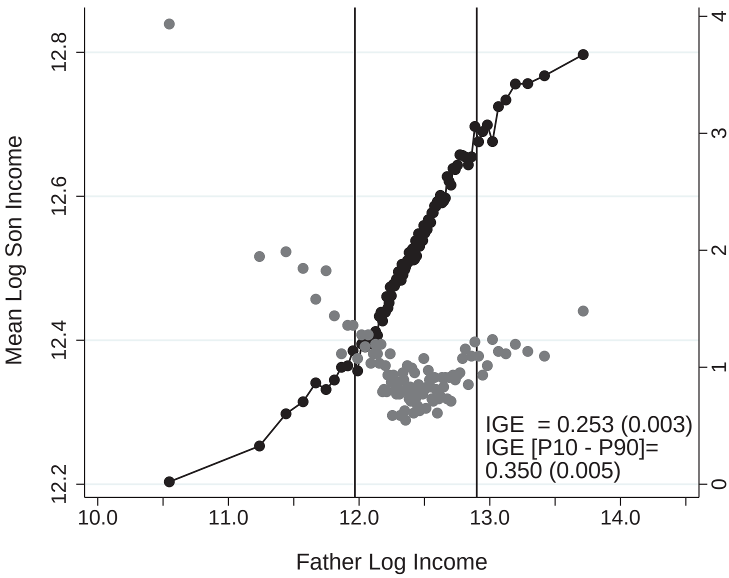
[20] Interestingly, a similar figure for Denmark shown in Chetty et al. (2014a, p. 1576) in panel B of figure II exhibits equally large deviations from the fitted lines especially at the bottom of the distribution, just as in Sweden. Small differences in low parental income rank result in sizable outcome differences for the children in the two Nordic countries Sweden and Denmark.

Figure A.1: Log Incomes

(a) Log Son Income vs. Log Parent Income



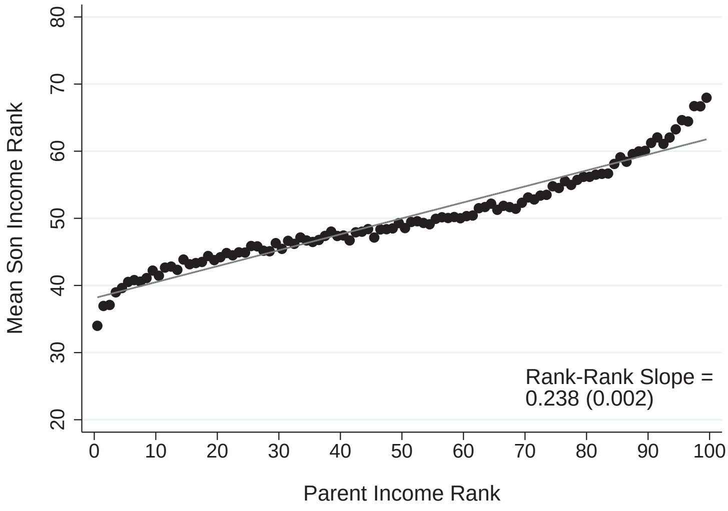
(b) Log Son Income vs. Log Father Income



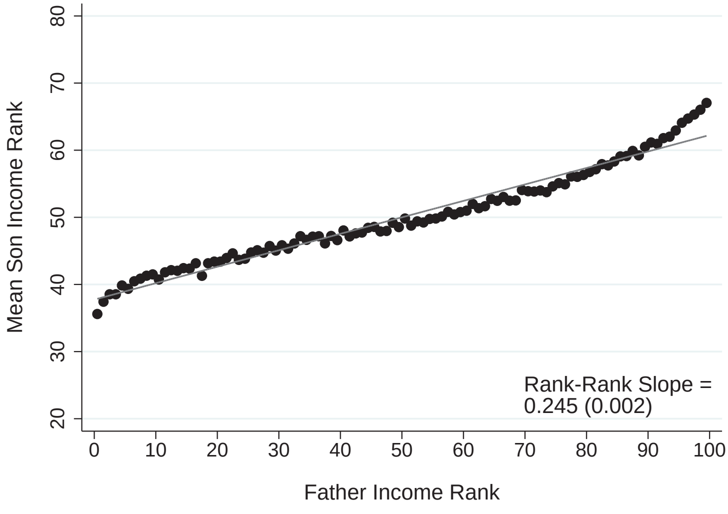
Note: These figures show a visual regression of log son income on log parent (father) income. Each black dot shows the mean log son income plotted against the mean log parent (father) income for one percentile of log parent (father) income. The gray dots (measured on the right y-axis) show the share of sons with zero income for each percentile of log parent income. The estimated IGE as well as the IGE between the 10th and 90th percentile are given in the figure.

Figure A.2: Income Ranks

(a) Mean Son Rank vs. Parent Rank



(b) Mean Son Rank vs. Father Rank



Note: These figures show a visual regression of son income rank on parent (father) income rank. Each dot shows the mean son income rank plotted against the mean parent (father) income rank for one percentile of parent (father) income ranks. The estimated rank-rank slopes are reported in the figure.

Lastly, Table A.1 shows the sensitivity of IGE estimates to the way zero incomes are treated. In the first column, all zero incomes are dropped from the data set. In columns two and three, zeros are replaced by 1 and 1000 SEK, respectively. In all specifications (rows), the IGE is very sensitive to how zeros are handled. The largest difference is found in the bottom ten percent of the father's income distribution (second row) where the IGE almost doubles depending on the treatment. This sensitivity is also present within each cohort separately as shown in the last two rows for cohorts 1968 and 1972 only.

Given the data at hand, using income ranks instead of logged incomes is the preferred option. Importantly, the linear fit of the rank relationship is not found to be superior to logged incomes overall and should be carefully examined on a case-by-case basis.

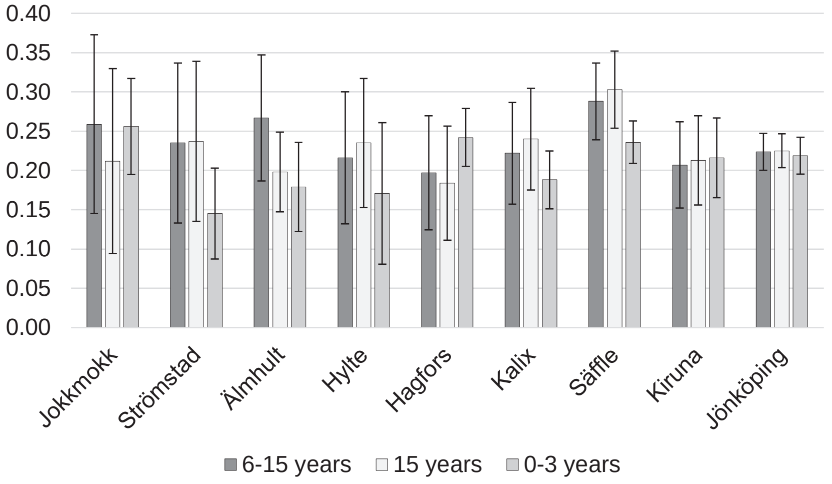
Table A.1: Sensitivity with Respect to the Treatment of Zero Incomes
(IGE between Fathers and Sons)

	Exclude zeros	Replace by 1	Replace by 1000
Pooled	0.253	0.311	0.280
	(0.003)	(0.008)	(0.004)
< P10	0.109	0.259	0.201
	(0.011)	(0.03)	(0.016)
P10 - P90	0.350	0.369	0.356
	(0.005)	(0.013)	(0.007)
> P90	0.132	0.040	0.088
	(0.017)	(0.04)	(0.024)
1968	0.278	0.317	0.296
	(0.009)	(0.024)	(0.012)
1972	0.246	0.318	0.280
	(0.008)	(0.02)	(0.012)

Note: Standard errors are given in parentheses.

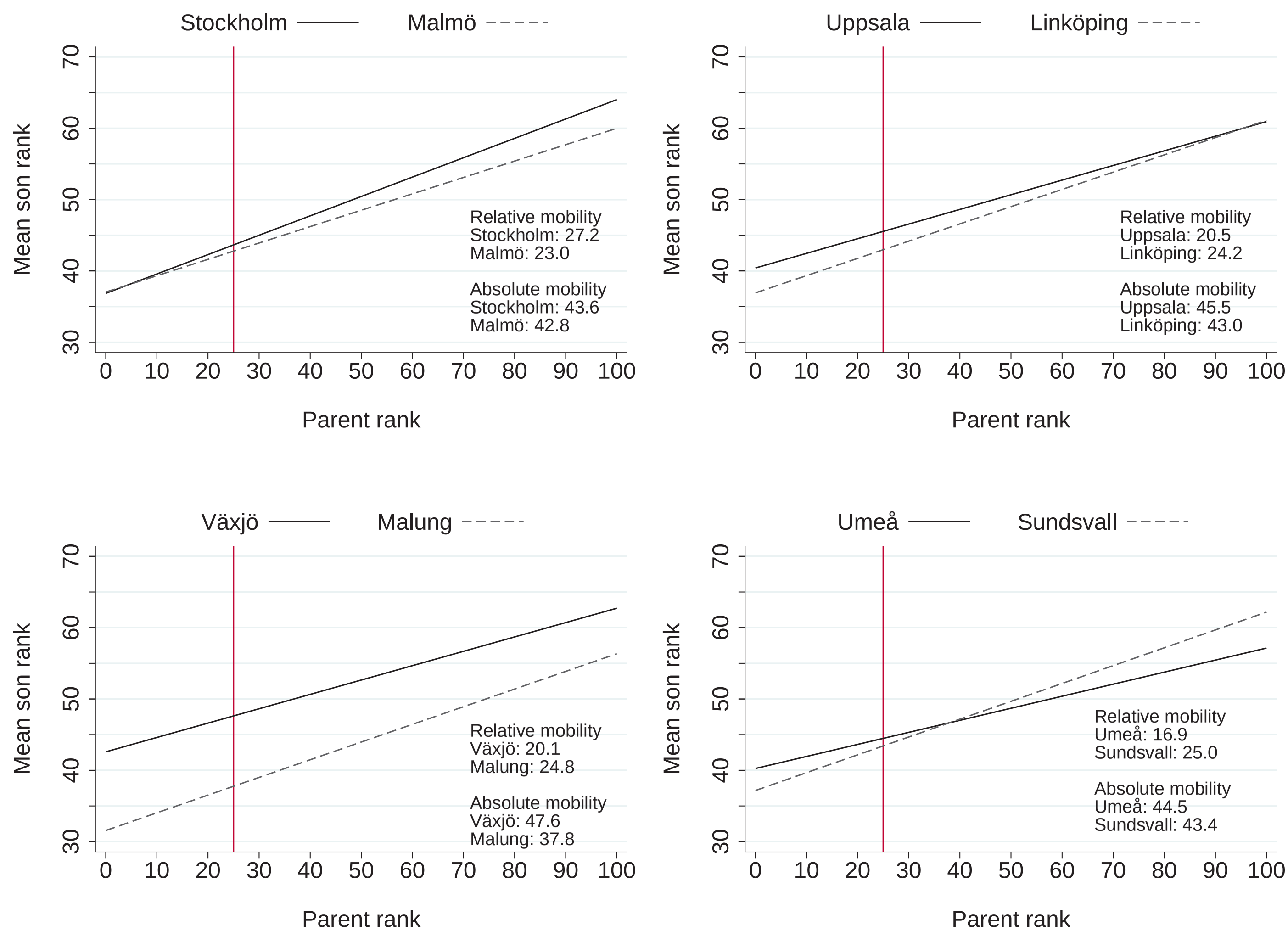
B Figures

Figure A.3: The Effect of Different Childhood Definitions on Mobility Estimates



Note: This figure illustrates the difference in rank-rank slope estimates for some small and medium-sized regions when using different childhood definitions. The parameter estimates are obtained by separate OLS regressions by region. The size of the error bars shows 95% confidence intervals. The regions are shown in ascending order from left to right by number of observations. The smaller the region, the larger the differences between the three childhood definitions, and the larger the standard errors.

Figure A.4: Rank-Rank Relationships for some Regions



C Tables

Table A.2: Sample Summary

Child cohort	Fathers age at birth (St.Dev.)	Mothers age at birth (St.Dev.)	N sons	N daughters	Total
1968	28.0 (4.8)	25.2 (4.1)	43.662	41.573	85.235
1969	28.2 (4.7)	25.5 (4.2)	42.948	40.001	82.949
1970	28.3 (4.7)	25.7 (4.2)	44.003	41.739	85.742
1971	28.3 (4.6)	25.8 (4.2)	46.637	43.633	90.270
1972	28.3 (4.5)	25.8 (4.1)	46.110	43.350	89.460
1973	28.5 (4.4)	26.0 (4.2)	45.534	42.840	88.374
1974	28.5 (4.4)	26.1 (4.2)	45.930	43.284	89.214
1975	28.7 (4.4)	26.3 (4.2)	43.278	40.848	84.126
1976	28.9 (4.4)	26.4 (4.2)	40.993	38.590	79.513
Total	28.4 (4.6)	25.9 (4.2)	399.095	375.858	774.953

Table A.3: Classification of Local Labor Markets

T1 Large cities	T2 Large regional centers	T3 Small regional centers		T4 Other small regions		T5 Sparsely populated regions
Göteborg	Borås	Arboga	Nyköping	Bengtsfors	Sävsjö	Arjeplog
Malmö	Eskilstuna	Arvika	Oskarshamn	Emmaboda	Tidaholm	Arvidsjaur
Stockholm	Falun	Avesta	Simrishamn	Fagersta	Vara	Dorotea
	Gävle	Bollnäs	Skellefteå	Filipstad	Vimmerby	Gällivare
	Halmstad	Eksjö	Strömstad	Gnosjö		Haparanda
	Helsingborg	Falkenberg	Söderhamn	Hagfors		Härjedalen
	Jönköping	Gislaved	Tranås	Hedemora		Jokkmokk
	Kalmar	Gotland	Uddevalla	Hofors		Ljusdal
	Karlstad	Hudiksvall	Varberg	Hultsfred		Lycksele
	Kristianstad	Härnösand	Vetlanda	Hylte		Malung
	Linköping	Hässleholm	Värnamo	Hällefors		Pajala
	Luleå	Höör	Västervik	Kalix		Sollefteå
	Norrköping	Karlshamn	Älmhult	Kramfors		Sorsele
	Skövde	Karlskoga	Örnsköldsvik	Laxå		Storuman
	Sundsvall	Karlskrona		Ludvika		Strömsund
	Trollhättan	Katrineholm		Markaryd		Torsby
	Umeå	Kiruna		Munkfors		Vansbro
	Uppsala	Kristinehamn		Nässjö		Vilhelmina
	Västerås	Köping		Olofström		Ånge
	Växjö	Lidköping		Perstorp		Årjäng
	Örebro	Mariestad		Sunne		Åsele
	Östersund	Mora		Säffle		Övertorneå

Table A.4: Results from the Multilevel Model

N obs: 393,715 N groups: 112	Model 1		Model 2	
Fixed effects (average effects)				
Intercept	37.956***	(0.351)	38.105***	(1.634)
R_f^p: Parent income rank	0.222***	(0.003)	0.243***	(0.013)
T2: Large regional centers			0.444	(1.746)
T3: Small regional centers			0.973	(1.709)
T4: Other small regions			1.545	(1.768)
T5: Sparsely populated regions			-4.411***	(1.816)
T2$\times R_f^p$			-0.018	(0.014)
T3$\times R_f^p$			-0.025*	(0.014)
T4$\times R_f^p$			-0.038***	(0.015)
T5$\times R_f^p$			-0.014	(0.017)
Variance components				
Parent income rank, σ_β^2	0.0005***		0.0005***	
Intercept, σ_α^2	11.5436***		7.9117***	
Cov $\left(\sigma_\alpha^2, \sigma_\beta^2\right)$	-0.0414***		-0.0324***	
Residual, σ_R^2	781.0998***		781.0906***	
LR test M1 vs. linear regression: p = 0.000			LR test M1 vs. M2: p = 0.000	

Table A.5: Relative and Upward Mobility by Local Labor Market

	LLM	Relative Mobility		LLM	Upward Mobility
1	Umeå	16.84	1	Hylte	50.77
2	Uppsala	20.48	2	Olofström	49.33
3	Torsby	22.20	3	Gnosjö	49.31
4	Malung	22.20	4	Ljungby	48.77
5	Överkalix	22.20	5	Hofors	48.50
6	Vilhelmina	22.20	6	Avesta	48.47
7	Årjäng	22.20	7	Värnamo	48.47
8	Gotland	22.20	8	Varberg	48.25
9	Arjeplog	22.20	9	Oskarshamn	48.21
10	Sunne	22.20	10	Växjö	48.14
11	Mora	22.20	11	Vara	47.70
12	Jokkmokk	22.20	12	Gislaved	47.69
13	Härjedalen	22.20	13	Emmaboda	47.63
14	Strömsund	22.20	14	Älmhult	47.61
15	Hagfors	22.20	15	Vetlanda	47.44
16	Filipstad	22.20	16	Perstorp	47.23
17	Övertorneå	22.20	17	Skövde	47.02
18	Hällefors	22.20	18	Nässjö	46.95
19	Storuman	22.20	19	Nyköping	46.92
20	Söderhamn	22.20	20	Örnsköldsvik	46.86
21	Ånge	22.20	21	Hultsfred	46.60
22	Östersund	22.20	22	Jönköping	46.52
23	Bollnäs	22.20	23	Lidköping	46.27
24	Härnösand	22.20	24	Tranås	46.18
25	Lycksele	22.20	25	Göteborg	46.14
26	Ljusdal	22.20	26	Fagersta	46.07
27	Hudiksvall	22.20	27	Falkenberg	46.03
28	Västervik	22.20	28	Karlshamn	46.01
29	Skellefteå	22.20	29	Lysekil	45.90
30	Arboga	22.20	30	Köping	45.88
31	Arvidsjaur	22.20	31	Kristianstad	45.72
32	Arvika	22.20	32	Halmstad	45.67
33	Bengtsfors	22.20	33	Uppsala	45.53
34	Dorotea	22.20	34	Gävle	45.52
35	Falun	22.20	35	Borås	45.19

Continues on next page

Table A.5: Relative and Upward Mobility by Local Labor Market

	LLM	Relative Mobility		LLM	Upward Mobility
36	Gällivare	22.20	36	Helsingborg	44.83
37	Haparanda	22.20	37	Umeå	44.46
38	Hedemora	22.20	38	Sävsjö	44.20
39	Hässleholm	22.20	39	Eksjö	44.19
40	Höör	22.20	40	Vimmerby	44.05
41	Kalix	22.20	41	Kiruna	43.99
42	Kalmar	22.20	42	Stockholm	43.63
43	Karlskoga	22.20	43	Arboga	43.51
44	Karlskrona	22.20	44	Arvidsjaur	43.51
45	Katrineholm	22.20	45	Arvika	43.51
46	Kramfors	22.20	46	Bengtsfors	43.51
47	Kristinehamn	22.20	47	Dorotea	43.51
48	Laxå	22.20	48	Falun	43.51
49	Linköping	22.20	49	Gällivare	43.51
50	Ludvika	22.20	50	Haparanda	43.51
51	Luleå	22.20	51	Hedemora	43.51
52	Malmö	22.20	52	Hässleholm	43.51
53	Mariestad	22.20	53	Höör	43.51
54	Markaryd	22.20	54	Kalix	43.51
55	Munkfors	22.20	55	Kalmar	43.51
56	Norrköping	22.20	56	Karlskoga	43.51
57	Pajala	22.20	57	Karlskrona	43.51
58	Simrishamn	22.20	58	Katrineholm	43.51
59	Sollefteå	22.20	59	Kramfors	43.51
60	Sorsele	22.20	60	Kristinehamn	43.51
61	Strömstad	22.20	61	Laxå	43.51
62	Sundsvall	22.20	62	Linköping	43.51
63	Säffle	22.20	63	Ludvika	43.51
64	Tidaholm	22.20	64	Luleå	43.51
65	Trollhättan	22.20	65	Malmö	43.51
66	Uddevalla	22.20	66	Mariestad	43.51
67	Vansbro	22.20	67	Markaryd	43.51
68	Västerås	22.20	68	Munkfors	43.51
69	Åsele	22.20	69	Norrköping	43.51
70	Örebro	22.20	70	Pajala	43.51

Continues on next page

Table A.5: Relative and Upward Mobility by Local Labor Market

	LLM	Relative Mobility		LLM	Upward Mobility
71	Helsingborg	22.20	71	Simrishamn	43.51
72	Borås	22.20	72	Sollefteå	43.51
73	Gävle	22.20	73	Sorsele	43.51
74	Halmstad	22.20	74	Strömstad	43.51
75	Kristianstad	22.20	75	Sundsvall	43.51
76	Köping	22.20	76	Säffle	43.51
77	Lysekil	22.20	77	Tidaholm	43.51
78	Karlshamn	22.20	78	Trollhättan	43.51
79	Falkenberg	22.20	79	Uddevalla	43.51
80	Fagersta	22.20	80	Vansbro	43.51
81	Göteborg	22.20	81	Västerås	43.51
82	Tranås	22.20	82	Åsele	43.51
83	Lidköping	22.20	83	Örebro	43.51
84	Jönköping	22.20	84	Karlstad	42.43
85	Hultsfred	22.20	85	Skellefteå	41.84
86	Örnsköldsvik	22.20	86	Västervik	41.53
87	Nyköping	22.20	87	Eskilstuna	41.36
88	Nässjö	22.20	88	Hudiksvall	41.29
89	Skövde	22.20	89	Ljusdal	40.81
90	Perstorp	22.20	90	Lycksele	40.60
91	Vetlanda	22.20	91	Härnösand	40.43
92	Älmhult	22.20	92	Bollnäs	40.39
93	Emmaboda	22.20	93	Östersund	40.21
94	Gislaved	22.20	94	Ånge	40.12
95	Vara	22.20	95	Söderhamn	40.12
96	Växjö	22.20	96	Storuman	40.06
97	Oskarshamn	22.20	97	Hällefors	40.05
98	Varberg	22.20	98	Övertorneå	39.75
99	Värnamo	22.20	99	Filipstad	39.66
100	Avesta	22.20	100	Hagfors	39.62
101	Hofors	22.20	101	Strömsund	39.16
102	Ljungby	22.20	102	Härjedalen	39.06
103	Gnosjö	22.20	103	Jokkmokk	38.99
104	Olofström	22.20	104	Mora	38.78
105	Hylte	22.20	105	Sunne	38.62

Continues on next page

Table A.5: Relative and Upward Mobility by Local Labor Market

	LLM	Relative Mobility		LLM	Upward Mobility
106	Kiruna	24.14	106	Arjeplog	38.44
107	Vimmerby	24.39	107	Gotland	38.41
108	Eksjö	24.95	108	Årjäng	38.39
109	Sävsjö	24.96	109	Vilhelmina	38.36
110	Karlstad	25.10	110	Överkalix	38.11
111	Eskilstuna	25.91	111	Malung	37.11
112	Stockholm	27.11	112	Torsby	36.32

End

III

THE EFFECT OF MOVING DURING CHILDHOOD ON LONG RUN INCOME: EVIDENCE FROM SWEDISH REGISTER DATA

STEFANIE HEIDRICH*

ABSTRACT

In this paper I study the long-term effects of inter-municipal moving during childhood on income using Swedish register data. Due to the richness of the data I am able to control for important sources of selection into moving, such as parent separation, parents' unemployment, education, long run income, and immigration background. I find that children's long run incomes are significantly negatively affected by moving during childhood, and the effect is larger for those who move more often. For children who move once, I also estimate the effect of the timing and the quality of the move. I measure the quality of each neighborhood based on the adult outcomes for individuals who never move; the quality of a move follows as the difference in quality between the origin and the destination. Given that a family moves, I find that the negative effect of childhood moving on adult income is increasing in age at move. Children benefit economically from the quality of the region they move to only if they move before age 12 (sons) and age 16 (daughters).

JEL classification: D31, J17, J24, J62, R23
Keywords: long-term effects of moving, disruption costs, neighborhood effects, human capital, child development

*Department of Economics, Umeå School of Business and Economics, Umeå University.
E-Mail: stefanie.heidrich@umu.se. I would like to thank Thomas Aronsson, David Granlund and David Sundström for helpful comments on this as well as earlier versions of the paper. The Umeå SIMSAM Lab data infrastructure used in this study was developed with support from the Swedish Research Council and by strategic funds from Umeå University.

1 Introduction

Levelling the playing field and ensuring that individual outcomes depend more on effort than family background is an outspoken goal of many western societies. However, differences in income and wealth have been growing significantly during the past decades (Organisation for Economic Co-operation and Development, 2015). In order to determine the most equitable and efficient form of public policy intervention to cope with this development, it is essential to understand the drivers of an individual's income, i.e. the human capital production function. Childhood is a very important phase during human capital development, since many skills that are capitalized during adulthood are obtained early in life. As Cunha and Heckman (2007) pointed out, skills produced at an earlier point in life will not only augment the skills attained at a later stage, but also increase the returns to later skill investments.

Recent evidence stresses the importance of neighborhood quality for children's long run outcomes (Chetty et al., 2015; Chetty and Hendren, 2015; Heidrich, 2015). However, identifying certain neighborhoods that are better or worse for a child's development leads to another question: what is the effect of moving from one neighborhood to another? If neighborhoods matter, moving to a better area can be thought to be beneficial for a child. However, there might also be disruption costs from moving. Being taken out of the home environment and having to adjust to a new neighborhood and school, finding new friends and leaving the old ones behind can negatively impact the human capital development of the child. Research in sociology for example indicates that moving can have negative effects on academic performance (Hagan et al., 1996). Disruptions during childhood can potentially accumulate to long-term effects, when taking into account the dynamic nature of human capital development. An important question is then whether these effects are economically important.

My contribution to the literature consists of the first estimates of the long run costs incurred by individuals due to childhood moving. I define those costs as the effect of childhood moving on long run adult income. I show not only that such costs exist and that they increase in the number of times a child moves, but also that the timing of the move matters. The results are well in line with predictions from the human capital literature where shocks in the skill formation process can accumulate to large differences in adult outcome (see for instance Cunha and Heckman 2007; Currie and Almond 2011; Heckman 2006; Heckman et al. 2013).

The main difficulty when trying to measure the costs of moving is selection. In observational data, we will suspect differences between families that move a different numbers of times, as well as non-randomness in the type of location that families move to and from. Parent income, for instance, is an important factor to take into consideration since higher income families are able to better compensate for disruption effects suffered by the child. Other transitions in family life that occur at a similar time as the move have the potential to confound the true effect of moving. Parental separation or unemployment are two common reasons to move, and these also affect children's outcomes independent of moving.

Based on earlier studies on child human capital production I estimate the effect of moving on long-term incomes with help of a detailed panel data set of Swedish families. The richness of my data allows me to control for many factors that would otherwise bias the results. The estimated moving costs are robust to the model specification, the covariates included, and the definition of moving used. An additional aspect to consider is that children can be expected to have only a minor say in when or where to move. That is, children are usually tied to their parents which suggests that any selection effect left can reasonably be assumed to be smaller compared to studying the effect of migration on adults (even if some families move because of their child).

This paper relates to the literature on childhood neighborhood effects by focusing specifically on the cost of transitioning between two regions. Since moving is a form of disruption in a child's development, this study naturally relates to the field of human capital development, an overview of which can be found in Becker (2009). Early childhood development and skill formation in particular have been studied by Cunha et al. (2010), Currie and Almond (2011), Heckman et al. (2013) and others. The common conclusion is that children are highly sensitive to their environment in terms of developing both cognitive and non-cognitive skills.

This study also relates closely to the literature on migration and the family which goes back to Mincer (1978). While the decision to move usually boils down to a comparison of expected utilities in different locations, many studies have focused on labor market participation effects of married women. If both spouses work in the labor market, the probability of moving tends to be significantly smaller compared to a family with a stay-at-home wife (Rabe, 2011). A few studies specifically focus on moving within in the Nordic countries. Pekkala and Tervo (2002) studied unemployment and migration in Finland and showed there is no causal effect of moving on the probability of finding a job. Axelsson and Westerlund (1998) estimated the effect of moving within Sweden on household disposable income. They found no effect on income and no selection effects, i.e. no significant differences between the families that move and those who do not.

One of the few papers looking at the effects of moving on child outcome is Hanushek et al. (2004). They studied the effects of switching schools, both for the children that move and the externality of the move on classmates who experience others leaving or joining the class during the year. They find significant negative effects on test scores both for the children that move between schools, as well as negative externalities for the class mates.

In the first part of this paper, I study a large sample of children and their parents that move between zero and sixteen times between municipalities during childhood. My results suggest significant negative effects of moving during childhood on long run adult income. The effects differ by gender and parent income. Studying family moves between

Swedish counties instead of municipalities, I find that the costs of moving on child long run outcome are very robust to the geographic unit chosen.

The second part of the paper focuses on a reduced sample containing only families that move exactly once during childhood. This way I can control for the type of move, i.e. if families move to areas that are relatively better or relatively worse for the child compared to their region of origin. Taking the type of move into account is important since recent research has shown that neighborhood quality matters for children's long run outcome, as well as what age the child starts living in the new neighborhood (see Chetty et al., 2015; Chetty and Hendren, 2015; Kling et al., 2007). I find that the negative effect of childhood moving on adult income is increasing in the age at the time of the move. I also show that if a family does move, then moving to a higher-quality region increases a child's long run outcome. However, sons have to move before the age of 12 in order to benefit from the better region. From age 12, the quality of the new region has no significant effect. Daughters have to move before the age of 16 in order to benefit from a better region. In general, the effects on daughters' adult income are smaller compared to sons.

The rest of the paper is structured as follows. In Section 2, I present the empirical framework first for the estimation of moving costs and then for the effect of timing and quality of the childhood move. Section 3.1 contains the description of the data set and the variables used in this study. The family moving pattens in Sweden are described in Section 3.2. In Section 4, I present and discuss the results of the two parts of the empirical analysis as well as some robustness checks. Finally, Section 5 concludes.

2 Empirical approach

The human capital literature offers some guidelines on how to think about the effect of childhood moving on adult incomes. Currie and Almond (2011), for instance, used a model based on Grossman (1972) on the demand for health to study the effect of shocks in early childhood on long run human capital. In their model, adult human capital is a function of the sum of parent investments during childhood. Adult outcomes such as earnings or the health capital stock are determined by the human capital or health capital accumulated during childhood, following the investments and possibly depreciation of the existing stock. The impact of an early-life shock on adult outcomes will then depend on the substitutability of investments in different childhood periods and the share of total investment in the time period of the shock. The possibility to remedy a negative shock in a later period is dependent again on the substitutability of investments between time periods. The effects of a shock can either be reinforced or compensated for by a child's parents, depending on the parent's utility function.

In his work on childhood human capital developmentHeckman (2007) stresses the importance of both cognitive and non-cognitive skills in determining adult outcomes. Cognitive skills can foster non-cognitive skills and vice versa, and different childhood phases can be more or less important for each skill type's development. He also emphasizes the two skill formation characteristics dynamic complementarity and self-productivity, which imply that more skills accumulated in one time period increase the productivity of skill investments well as the size of skill stocks in later time periods.

In light of the existing models, moving during childhood represents a shock to the child's skills formation. The accumulated effect on human capital measured during adulthood will depend on the degree to which parents were able to compensate for this disruption. When the disruption happens at critical ages, however, compensation might hardly be possible. Thus, in the first part I will estimate the effect of moving (the disruption) on long run income. In the next part, I will analyze the effect of the timing at the move in order to learn the critical ages that this particular disruption has on human capital production.

2.1 Estimating the effect of moving on long run income

If long run adult income is assumed to be a function of human capital, we can expect to find lower incomes for individuals that moved during childhood, ceteris paribus. If we had an experimental setup at our disposal where we could pick a random sample of families and move them a random number of times to and from to random locations, we could to estimate

$$y_i^c = \beta_0 + \sum_{m=1}^{M} \beta_m I\left(\text{Moving}_i = m\right) + \varepsilon_i \tag{1}$$

by ordinary least squares, where y_i^c is child long run income as adult in family i, $I\left(\text{Moving}_i = m\right)$ is an indicator equaling one if the child moved m times during childhood, and β_m is the effect of moving m times during childhood on long run income. However, in absence of such an experiment, I have to make sure to eliminate potential sources of selection that could bias the OLS estimator.

Any causal interpretation of β_m with observational data relies on the conditional independence assumption, also called selection on observables. That is, moving has to be as good as randomly assigned, conditional on the covariates (Angrist and Pischke, 2008). If the vector X_i includes all relevant controls, the true, "long" regression is

$$y_i^c = \beta_0 + \sum_{m=1}^{M} \beta_m I\left(\text{Moving}_i = m\right) + X'\gamma + \varepsilon_i. \tag{2}$$

The omitted variable formula shows that the regression coefficient for moving in the "short" equation (1) above (where we did not include A') is actually equal to

$$\frac{Cov\left(y_i^c, \beta_m\right)}{V\left(\beta_m\right)} = \beta_m + \gamma' \delta_{AM}, \tag{3}$$

where δ_{AM} is the regression slope coefficient when regressing the omitted variables on moving. That is, omitting variables that are correlated both with moving and with child long run income will give a regression coefficient that is too large or too small, depending on the signs of the correlations.

There is some guidance in the literature regarding which factors simultaneously affect if and how often a family moves, and the long run income of the child in the family. Starting with the drivers of migration, families will move if the family's net real discounted income gain from moving is positive (Mincer (1978)). Since individuals made a rational choice about their current place of residence, only an unforeseen change in any of the variables determining the family's expected utility of a location will lead to a reevaluation of their choices. It is reasonable to assume that in most cases children follow their parents, and that it is the parents who decide if and where to move. Thus, I focus on the drivers of moving from the parents perspective. One typical reason is job search (see Greenwood (1997) for an overview of research on the causes and effects of internal migration). *Parent unemployment* history is therefore important to explain selection into moving. Young parents are especially likely to move, after completing their education and becoming employed. Thus, *parent age* at child birth should be included. *Parent income* can potentially also affect the moving decision: A better paying job enables a family to move to better regions, and career decisions leading to a steeper income growth might even require the family to relocate regularly. At the same time, low wages might cause a family to relocate to a region with higher relative wages. *Parent education* is an additional reason for moving which is closely related to parent income and unemployment. Highly-educated parents for example might need to move to specific regions in order to find qualified work. A different type of shock which influences a move is a disruption of the family itself. *Divorce* is quite likely to induce a move of one or both parents and the children. Lastly, there might also be a difference in moving patterns between families who live in urban or rural areas *(birth municipality type)*.

All of the above reasons that cause families to moving potentially affect a child's long run income. Experiencing unemployment can be a difficult time for a parent both financially and in terms of well being and health. The children can be negatively affected directly by lower investments in their human capital, and indirectly by suffering psychologically. Younger parents that might move after finishing their education will on average also have lower income compared to older parents (who are already higher up on the career ladder), which lowers the investment into their young children and thus children's long run incomes. Studies have shown that children with divorced parents score significantly lower on measures of academic achievement, conduct, psychological adjustment, self-concept, and social relations (Amato, 2001). Parent education and parent income is usually correlated with child income (see Black and Devereux (2011) or Jäntti and Jenkins (2015) for a survey of the literature on intergenerational income mobility). Parents can also compensate for any negative effects from the move depending on their income, as discussed in the beginning of this section. Growing up in a rural or urban area can also be correlated to child long run income, since individuals from large cities are usually less likely to move to the country side and because individuals in urban areas on average have higher wages than individuals living in the country side.

Even though my sample of children only includes individuals born in Sweden, the parents' country of birth still might impact a child's long run outcome. Children with foreign-born parents have been shown to perform worse academically compared to children with two Swedish parents (Skolverket, 2004). Moreover, we cannot exclude the possibility that immigrant parents differ in their moving patterns from native-Swedish parents. Therefore, I will include *parent immigration* status as a covariate.

The parents' decision to move can be driven by regional characteristics that affect all families in the region similarly. When studying families that move many times, one would have to trace the region-specific incentives that made a family move the fourth, third, second, and first time, and I will use differences between regions in the second part of this study only (where I focus on one-time movers).

Children's individual characteristics such as the highest education level and work experience are important when modelling earnings (Mincer, 1974). However, they cannot reasonably be assumed to cause the family to move during childhood and so are not included here.[1] The birth year of the children on the other hand might be important, since there could be different shocks in the economy making individuals more or less likely to move by a certain child age. There may reasonably also be cohort effects on long run income and birth *cohort* is therefore added as a covariate.

The child's *gender* is also potentially important since the human capital production function might look different for girls than for boys (irrespective of whether those differences are inherent or due to social norms and the way girls and boys are raised). Gender might even play a role in how resilient a child is to shocks and how parents compensated for them. Therefore, I will estimate the effect of moving separately by gender.

2.2 The effect of the quality and the timing of the move

Using a sample of families who move exactly once during a child's upbringing enables me to study the effect of the timing of the move (i.e. the age at which a child moves), as well as the type of move (i.e. does the family move to an area which is relatively better or worse for the child, compared to the region of origin?). If one part of the families moves to better areas and the other part to worse-off areas, the moving costs and age effects might be masked by neighborhood quality effects. Moreover, in a specification with one time movers only it is also possible to include regional fixed- or random effects, which can absorb unobserved regional differences that have an impact on the families' decisions to move.

Chetty et al. (2015) have recently focused on the effect of age at the time of the move. Their study shows that children become similar to the children in the region they move to in terms of adult outcomes. The earlier children move to a certain area, the less differences between permanent residents and the new-comers can be observed as adults. This so called 'exposure effect' implies that the positive effect of a better neighborhood on a mover child will be larger if the child moves at a younger age.

Therefore in order to study the effect of timing and quality of the move, I will construct a variable that proxies for the difference in neighborhood quality between the region of origin and the region of destination. I define the quality of a region by the average adult

[1] When repeating the analysis below including child education, the results are qualitatively very similar.

income achieved for an individual who lived her entire childhood in that region. Using only the sample of stayers, S, who never moved during childhood and thus spent their whole childhood in one region r, I estimate the parameters of the following income equation for each of the R regions by OLS :

$$\begin{aligned} y^{c}_{s,r} &= \beta_{0,r} + \beta_{1,r} y^{p}_{s,r} + \beta_{2,r}\text{Educ}_{i,r} + \beta_{3,r}\text{Metro}_{i,r} + \beta_{4,r}\text{Cohort}_{i,r} \\ &\quad + \beta_{5,r}\text{Cont.Family}_{i,r} + \beta_{6,r}\text{Immigr}_{i,r} + \beta_{7,r}\text{Gender}_{i,r} + \varepsilon_{s,r} \; \forall\, r = 1,...R. \end{aligned} \tag{4}$$

where $y^{c}_{s,r}$ and $y^{p}_{s,r}$ are child and parent income of stayer family s in region r, respectively, and the other covariates are child education level, living in a metro area as adult, cohort effects, a dummy for continuous family, parent immigration, and child gender. Next, I use the R sets of estimated regression coefficients obtained from the stayer sample to predict the long run income for all one-time movers, had they grown up completely in their region of origin (denoted by $\hat{y}_{i,o}$). I also use the estimated coefficient to predict the long run income of the mover children had they grown up completely in their region of destination (denoted by $\hat{yi}_{i,d}$). Now, I compute the change in municipal quality for each one-time mover i: $\triangle_i = \hat{y}_{i,d} - \hat{y}_{i,o}$, which is positive if the family moves to an area where children with the same parent income have, on average, higher long run outcomes than in the region they moved away from. The construction of the variable $\triangle_i$ follows Chetty and Hendren (2015). However, these authors use parent income as the only predictor for child income in equation (4).

In a second step, I estimate the parameters of the following equation on the sample of one-time movers by OLS:

$$\begin{aligned} y^{c}_{i} &= \alpha_{op} + \sum_{ag=2}^{4} \beta_a I\left(\text{MoveAgeGroup}_i = ag\right) + \gamma_0 \triangle_i \\ &\quad + \sum_{ag=2}^{4} \lambda_a I\left(\text{MoveAge}_m = a\right) \times \triangle_i + X'\eta + \varepsilon_i \end{aligned} \tag{5}$$

where I group the mover children into 4 groups by age at move (age 1-5, age 6-11, age 12-15, and age 16-17). Due to the relatively small number of observations I will focus on the results using age groups. Results using 17 individual age-at-moves can be found in the Appendix. α_{po} indicates parent income quartile by origin fixed- or random effects. I also test using two separate fixed effects but the interaction of region and parent income is significant which is why I focus on the specification as given in equation (5). The identification is here based on the variation in children moving from the same region to different regions, with different region qualities. Using those fixed- or random effects will “control away” unobserved characteristics that are shared by families in the same region and with similar parent income.

If I assume that moving to a better region increases long run outcomes, I expect γ_0 to be positive. The β_a’s indicate the average effect of moving at a particular age given no change in the neighborhood quality. The β_as are going to tell us if there are heterogeneous effects of moving depending on age. The interaction effects λ_a show if there are ages at which a child's human capital development seems particularly sensitive for region quality. I also include a vector X of similar control variables as in in Section 2.1. However, since I use parent income both in the calculation of $\triangle_i$ and in the fixed effects, it is not contained in X.

3 Data

In the first part of this section I explain the construction of the variables used in the study. The second part gives an overview of the moving patterns in my data. All data were obtained from the Umeå SIMSAM lab[2]. The population registers used are collected and supplied by Statistics Sweden. My sample contains all individuals born in Sweden between 1968 and 1976 (called children), as well as their parents. I exclude families with adopted children from the sample to avoid additional sources of heterogeneity.

3.1 Variables

I use *long run average income* as the main outcome variable. Borrowing from the intergenerational mobility literature, I use this concept of income which aims at comparing individuals' earnings over their whole life cycle. For the children I use the average annual income at age 32 to 34, which has been shown to be a good approximation to the average annual income over the whole life cycle for individuals in Sweden if incomes at older ages are not observed (see Nybom and Stuhler (2016) for a discussion of this approximation). There is more data available for the parent generation and I use the average annual income measured between ages of 34 and 50 as a proxy for their long run average income. Importantly, I can measure the parent income for everyone at the same age, as opposed to during the same calendar years (where some parents might be very young, and others much older).

The income measure contains earning from employment and self employment, as well as earnings related transfers such as sick leave payments and parental benefits. Earlier studies on earnings formation and human capital have traditionally been using the log of income as the outcome variable. However, this implies that observations with average income of zero have to be dropped from the analysis. More recent studies in the intergenerational mobility and neighborhood effect literature have used income expressed in terms of percentile ranks instead of log incomes. Children are ranked relative to other children in the same birth cohort while parents are ranked relative to other parents with children in the same cohort.[3] I will use the long run average incomes transformed into percentile ranks as the main measure of income. However, I also show results based on log incomes for comparison. In order to facilitate interpretation of the percentile ranks, Figure A.1 in the Appendix shows the average child income in SEK for each percentile

[2] SIMSAM is short for "Swedish Initiative for Research on Microdata in the Social And Medical Sciences", an initiative supported by the Swedish Research Council.

[3] A discussion of using income ranks as opposed to logged incomes can be found in Chetty et al. (2014) and Heidrich (2015).

rank (right Y-axis). The second plot in this figure (left Y-axis) indicates how many SEK one percentile rank represents, at each position in the distribution. The distance between two ranks is largest at the top and bottom of the distribution, while it is very similar for the middle of the distribution.

A large proportion of the daughters in my data (around 40 percent in each cohort) receives maternity leave benefits during the years I measure adult incomes. Due to the cap on this earnings-based transfer as well as the flexible ways to take out the benefits there is significantly more measurement error in the daughters long run income compared to the sons which we need to keep in mind during the analysis.

Cohort is a dummy variable where 0 indicates that the child is born between 1968 and 1971, and 1 if the child is born 1972 - 1976. Using individual year of birth dummies, the results showed that there are only statistically significant differences between the two groups of birth cohorts as indicated above.

For both generations the data includes four indicator variables for *educational level*. The first takes the value 1 for individuals with at most at most primary school education and 0 otherwise; the second takes the value 1 for individuals with more than primary school education but at most a high school degree and 0 otherwise; and the third and fourth dummy similarly indicate some college education and a college degree, respectively.

Birth municipality type is another categorical variable indicating if the child was born in a rural area (1), large city (2), or otherwise (0). The classification is taken from the Swedish Association of Local Authorities and Regions municipality grouping based on regional characteristics in 1993.

Metro area shows if a child lives in greater Stockholm, Gothenburg, or Malmö at age 33. The coding of this variable is also based on a classification of municipalities from the Swedish Association of Local Authorities and Regions. The variable is not exogenous to moving and I only use it to predict adult income in the second part of the analysis.

Father (mother) age at birth is the difference between father's (mother's) year of birth and the birth year of their child.

Parent immigration status is coded based on the birth country variable. Even though the data allows for a finer distinction, several tests showed that a very detailed grouping does not increase explanatory power or precision compared with using only the number of parents born outside of Sweden. This variable can therefore take only one out of three values for each child, namely zero, one, or two parents born abroad.

I also construct the variable *continuous family* which is 1 if the parents of a child never separated during childhood. Since actual relationship status is unobservable, especially in register data, I need to make a judgment based on the recorded observations of civil status and parish of the children's mother and father. More specifically, I compare both civil status and parish of the parents in each year from before birth until the child turns 18. If both parents get married during the same year before the child is born, never get a divorce, or get a divorce at the same time (after the child has turned 18), I define the parents as married.

Living in the same parish as from child birth for at least 18 years and having every year the civil status unmarried, I categorize the parents as cohabiting (a common living arrangement in Sweden). The last group of parents considered *continuous family* are parents who were unmarried and living together (cohabiting), but who were later married at some point during their child's upbringing. Even though I do not observe the true family status of the parents, my classification appears to be quite robust. In order to be falsely classified as *continuous family*, parents would have had to get married at the same time with different partners and stay married the same number of years, while living in the same place. Families who are not classified as continuous are instead coded as *separated.*

Father unemployment is a dummy variable which indicates whether the child's father experienced any unemployment before the child turned 18. This variable is mainly based on having received a positive amount of unemployment benefits during at least one year. However, we only observe this variable starting in 1972. For the year 1970 we can use an employment indicator from a different register, but I do not observe any unemployment indicator during 1968, 1969, and 1971. Therefore, for children born before 1972, and especially for the ones born before 1970, this variable contains some measurement error. For example, I do not observe if a father was unemployed when his child, born in 1968, was one year old. However, assuming that the effect on a child of parent unemployment is larger when the child is older, we probably do not have to be too concerned about this. I do not include mother's unemployment since it turned out to not being predictive of children's long run outcome.

Table 1: Summary of the variables

	Mean	sd	p50	Min	Max	N
Child long run average income	269,984	148,979	255,511	0	10,503,468	766,271
Son long run average income	313,644	169,945	297,868	0	10,503,468	394,450
Daughter long run average income	223,667	104,564	215,069	0	5361,351	371,821
Parents long run average income	427,537	158,613	404,740	0	8061,467	766,271
Daughter education	2.82	0.98	3	1	4	368,458
Son education	2.55	0.94	2	1	4	390,245
Mother education	2.03	0.93	2	1	4	766,271
Father education	1.98	0.98	2	1	4	766,271
Father unemployment	0.22	0.41	0	0	1	766,271
Immigrated parents	0.15	0.45	0	0	2	766271
Continuous family	0.69	0.46	1	0	1	766271
Mother age at birth	25.87	4.22	26	16	36	766271
Father age at birth	28.36	4.54	28	16	40	766271
Birth mun. type	0.55	0.65	0	0	2	766271
Metro area	0.44	0.5	0	0	1	766271

Note: All income measures are expressed in 2010 SEK.

Table 1 gives a summary of the variables described above. Mean son long run average income is around 90,000 SEK (≈10,600 USD) higher than daughter income, and also has a larger variance. In both generations there are observations with zero incomes (slightly more than one percent of all observations). Daughters have on average a little more education than sons. More than two thirds of the children in my sample grow up in continuous families and 22 percent experience father unemployment.

3.2 Family moving patterns in Sweden

The variable containing the number of moves during childhood is constructed by comparing the municipality of residence from one year to the next from birth until the year a child turns 17. I do not consider moves at age 18 or older since the focus of this study is on childhood moves. Also, moves at age 18 are potentially different from earlier moves both due to high school graduation and because children are not necessarily tied to their parents any longer.

I mainly consider moves between municipalities. Changing municipality in Sweden usually entails having to switch schools and adjust to a new physical and social environment. However, I will also show the results using moves between counties instead of municipalities. If we believe that moving affects children at least to some extend because of severed social ties, we would expect the effects of moving to be similar when considering counties compared to municipalities. Moving at least to the next county instead of the next municipality will on average entail a larger distance between the origin and the destination and thus less opportunities to keep up social ties in the place of origin.

The municipality structure in Sweden has undergone many changes since 1970. Many municipalities have been merged with or broken out of other municipalities, sometimes several times. All original municipality coding in the data has been carefully revised and updated in order to not wrongfully classify a change in a municipality code as a move. This has been done using official documentation of the Swedish municipality reforms from Statistics Sweden. The residence is recorded only once every year in December, so that the number of moves observed in the data can be seen as a lower limit to the true number of moves.

Moving at age 6 then means that the child was registered in a different location in December in the year she turned 6 compared to the previous year. Table 2 shows the number of moves during childhood, both on the municipality and the county level. Focusing on the municipality level, we find that roughly 65 percent of the children never move, and 99 percent of all families moves 4 times or less. The maximum number of moves recorded is 16.

Table A.1 in the Appendix shows the age at the time of the move for all children that move exactly once between municipalities during their childhood (160.871 individuals). More than half of those families move when their child is four years old or younger. Age 15 is when the smallest number of families move (only 1.61 percent of this sample).

Table 2: Childhood moves

Municipality level			County level		
Number of moves	Frequency	Percent	Number of moves	Frequency	Percent
0	512,797	64.96	0	634,138	80.33
1	160,871	20.38	1	95,869	12.14
2	72,668	9.21	2	42,129	5.34
3	26,665	3.38	3	11,607	1.47
4	10,213	1.29	4	3,924	0.50
5	3,774	0.48	5	1,175	0.15
6	1,475	0.19	6	383	0.05
7	544	0.07	7	119	0.02
8	235	0.03	8	47	0.01
9	96	0.01	$\geq$9	34	0.00
10	43	0.01			
11	25	0.00			
$\geq$12	19	0.00			
Total		100.00			100.00

Table 3 shows the distribution of moving quality by parent income quartile. We can see that parents in the top income quartile are overrepresented among the one time movers (if all families were equally likely to move once, the number of observations would be very similar in each row). The average move quality is decreasing in parent income. This could be driven by the fact that high income parents probably already live in rather good neighborhoods which leaves less room for improvement. Interestingly, for all parent income levels, move quality is centered around just above zero. This means that families move on average to regions that are just slightly better for their children compared to their region of origin.

Table 3: Distribution of moving quality

Parent inc. quartile	Mean	sd	Min	p25	p50	p75	Max	N
1	0.63	5.75	-46.2	-2.65	0.55	3.97	42.4	34,053
2	0.48	5.18	-38.5	-2.47	0.44	3.44	36.1	34,531
3	0.33	4.88	-41.6	-2.54	0.33	3.25	41.0	39,008
4	0.27	4.58	-41.5	-2.54	0.27	3.07	40.1	52,366
Total	0.41	5.05	-46.2	-2.55	0.38	3.36	42.4	159,958

Note: Moving quality is measured as the difference in quality between the destination and the origin region. The quality of a region is based on the long run adult outcome of the children of permanent residents in each region.

Table A.2 in the Appendix shows the number of moves by parent income quartile. As already noted above, parents in the top income quartile are overrepresented among the one time movers. In addition, they also constitute the group of families who move most often two or three times during their kid's childhood. On the other hand, parents in the bottom part of the income distribution are the ones moving most often (four or more times). Thus, it seems that fewer moves a small is associated with high income for the parent generation, while moving often is associated with low parent income.

4 Results

The first part of this section discusses the results for the effect of childhood moves on long run average adult income separately by gender. The second part focuses on how the quality of the move effects a child, i.e. controlling for the fact that some families move to areas which are better or worse for their child, as well as on the timing of the move.

4.1 Moving cost estimates

The results of the baseline regressions from equations (1) and (2) for the children in my data set are summarized in Table 4. The first column in Table 4 shows the estimates and their standard errors pretending we have perfect experimental data. Son income rank is regressed on six dummies for the number of moves an individual has experienced during childhood. The last dummy groups together 6 and more moves, since there are few families moving more often than that. According to the estimates, moving once does not affect long run income while moving twice or more often is associated with larger and larger costs in terms of long run income.

In Model (2) for each gender I add parent income as well as the interaction between parent income and moving. The moving effects are now larger, for sons ranging from 2.88 to 16.3 percentile ranks less in long run income depending on the number of childhood moves given a parent income rank of 0. Given parents with income rank 100, one time movers earn approximately just as much as non-movers. Sons whose parents are in the top of the income distribution will have lower long run incomes if they move more than once during childhood (3,88 to 13.29 percentile ranks less). The effects for daughters show a similar pattern but are in general smaller. The parameter estimate on parent income rank is 0.18 and thus smaller than in the intergenerational mobility literature (0.24 for Sweden, see Heidrich, 2015).

The third model for sons and daughters also includes the covariates discussed in Section 2.1, i.e. parent immigration status, unemployment, education, age at birth, as well as continuous family, starting out living in a rural area or a large city, and child cohort group. The parameter estimates of the covariates are all statistically significant and have the expected signs: parents born outside of Sweden, father unemployment, and being born in a rural area decrease child long run income, while growing up in a continuous family and having educated parents increases child income. We can see that parent education is

Table 4: Baseline results

	Sons			Daughters		
Income rank	(1)	(2)	(3)	(1)	(2)	(3)
Moves=1	0.15	-2.88***	-1.53***	1.36***	-1.46***	-1.01***
	(0.115)	(0.234)	(0.234)	(0.108)	(0.205)	(0.206)
Moves=2	-1.93***	-6.88***	-3.86***	0.28	-3.26***	-2.06***
	(0.167)	(0.336)	(0.338)	(0.151)	(0.281)	(0.283)
Moves=3	-3.96***	-9.43***	-5.85***	-1.21***	-4.56***	-3.29***
	(0.275)	(0.517)	(0.516)	(0.240)	(0.414)	(0.415)
Moves=4	-6.02***	-8.60***	-4.47***	-2.43***	-4.49***	-2.72***
	(0.440)	(0.780)	(0.779)	(0.377)	(0.614)	(0.613)
Moves=5	-10.25***	-14.55***	-10.10***	-5.07***	-4.27***	-2.58**
	(0.729)	(1.150)	(1.146)	(0.611)	(0.959)	(0.953)
Moves≥6	-14.81***	-16.29***	-12.02***	-10.12***	-8.25***	-6.77***
	(0.934)	(1.310)	(1.311)	(0.648)	(0.874)	(0.876)
P. income rank		0.18***	0.17***		0.18***	0.14***
		(0.002)	(0.002)		(0.002)	(0.002)
Moves=1 # PIR		0.03***	0.02***		0.03***	0.02***
		(0.004)	(0.004)		(0.004)	(0.004)
Moves=2 # PIR		0.07***	0.04***		0.04***	0.02***
		(0.006)	(0.005)		(0.005)	(0.005)
Moves=3 # PIR		0.08***	0.05***		0.04***	0.03***
		(0.009)	(0.009)		(0.007)	(0.007)
Moves=4 # PIR		0.05***	0.02		0.03**	0.02
		(0.014)	(0.013)		(0.012)	(0.011)
Moves=5 # PIR		0.12***	0.10***		-0.00	-0.01
		(0.022)	(0.022)		(0.020)	(0.020)
Moves≥6 # PIR		0.12***	0.11***		0.03	0.03
		(0.031)	(0.030)		(0.023)	(0.023)
Imgr. parents=1			-1.96***			-0.21
			(0.169)			(0.153)
Imgr. parents=2			-4.20***			1.03***
			(0.250)			(0.223)

	Sons			Daughters		
	(1) cont.	(2) cont.	(3) cont.	(1) cont.	(2) cont.	(3) cont.
Cont. fam.=1			5.29***			2.13***
			(0.101)			(0.091)
Unemployed			-2.20***			-1.09***
			(0.109)			(0.097)
F. educ.=2			0.64***			1.05***
			(0.099)			(0.090)
F. educ.=3			1.29***			2.53***
			(0.175)			(0.166)
F. educ.=4			0.42*			3.92***
			(0.184)			(0.175)
M. educ.=2			1.09***			1.52***
			(0.100)			(0.091)
M. educ.=3			1.68***			3.22***
			(0.160)			(0.152)
M. educ.=4			0.82***			4.76***
			(0.190)			(0.183)
Born Metro			-0.21*			1.31***
			(0.097)			(0.091)
Born Rural			-2.38***			-0.72***
			(0.152)			(0.138)
F. age/birth			-0.06***			0.04**
			(0.015)			(0.013)
M. age/birth			0.30***			0.31***
			(0.016)			(0.015)
Cohort			-0.98***			0.47***
			(0.087)			(0.081)
Constant	61.21***	52.48***	43.36***	38.39***	29.62***	17.83***
	(0.053)	(0.104)	(0.356)	(0.050)	(0.094)	(0.324)
Observations	394450	394450	394450	371821	371821	371821
Adjusted R2	0.003	0.046	0.061	0.001	0.051	0.068

Standard errors in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001

the only variable in the model that has a stronger effect on long run income for daughters than for sons.
The estimates of the moving effects decrease as a result of introducing the controls. There is hardly any effect on the precision of the estimates while the share of the variance explained by this model rises, as expected. The marginal effects of moving depend on parent income. In order to simplify interpretation of the results, I redo the analysis using grouped moves instead: children who move one to three times during their childhood are grouped as "movers" and children who move more than three times are classified as "extreme movers". Children who spend their complete childhood in one region are referred to as "stayers". The output for this regression can be found in Table A.3 in the Appendix.
Figures 1 and 2 show the marginal effects of being a mover or an extreme mover (i.e. compared to being stayers) on long run outcome for sons and daughters, by parent income rank.

The negative effect on long run adult income in terms of percentile ranks is largest for those children who grow up in low income families. At the bottom of the parent income distribution, a mover son can expect a 2.6 ranks lower income as adult compared to a stayer, and an extreme mover has a 7.2 percentile ranks lower expected income. The reduction in percentile ranks in the middle of the parent income distribution are 1.2 and 4.2 percentile ranks. Being a mover does not lower the expected long run income of a son if his parents are located in the top ten percent of the income distribution. However, those sons with top income parents still have on average lower incomes compared to stayers if they move more than three times.

Focusing on the daughters in Figure 2, we can see that the effects are in general smaller (the largest reduction in long run income is -3.6 percentile ranks for extreme movers with parents in the bottom of the income distribution). Also, with parents in the highest income quintile, there is either no effect or a positive effect of being a mover compared to being a stayer. Similar to the sons, however, daughters with high income parents have on average lower incomes than stayers if they move more than three times during childhood.

In order to get a better understanding of the results we can translate some of the percentile ranks into monetary values. Since the relationship between all observed incomes in the distribution and the ranks 1-100 is not too far from being linear, each percentile rank is actually quite close to representing one percent of income. To highlight this I take the average income rank of stayers by parent income decile and compute the marginal effects of moving and extreme moving in terms of ranks based on the results in Table A.3. Then, I transform those resulting ranks for stayers, movers, and extreme movers for the ten parent income groups back into monetary values using the child income distribution.

Figure 1: Marginal effects of moving (Sons)

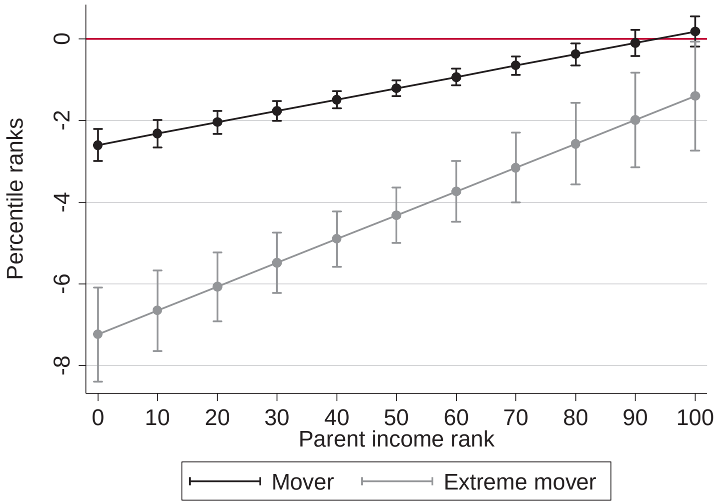
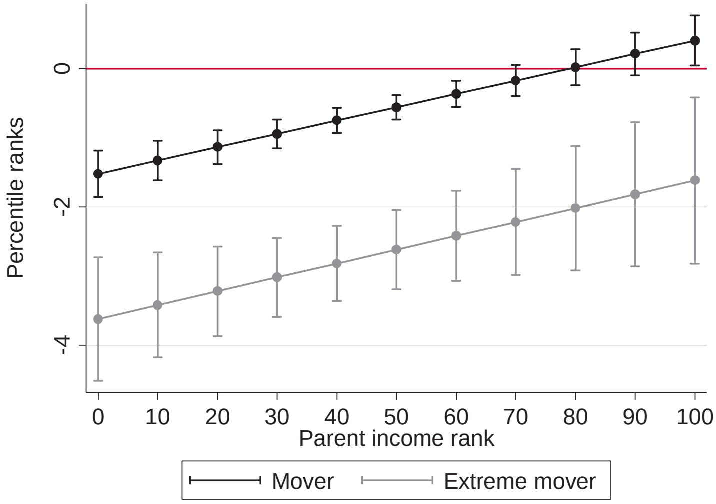


Figure 2: Marginal effects of moving (Daughters)



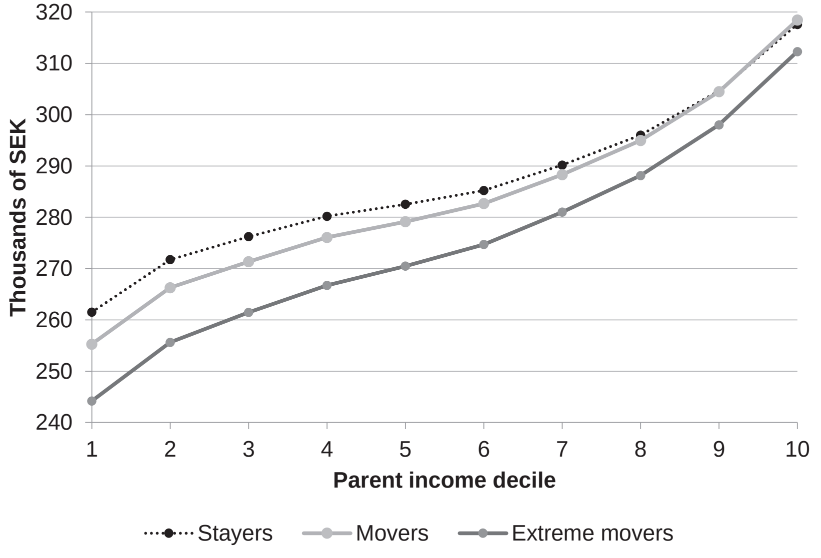
Note: The two figures show the estimated effects of moving 1-3 times (Movers) and more than 3 times (Extreme movers) on long run average income for sons and daughters, given parent income. The bars indicate 95%-confidence intervals.

Figure 3a shows the average son income by parent income decile for stayers, movers, and extreme movers in terms of SEK. Figure 3b shows the same information in terms of percentages. In the bottom decile of the parent distribution, a mover son can expect a 2.4 percent lower income compared to stayers. For an extreme mover, this difference amounts to 6.6 percent lower income. At the fifth decile of the parent distribution, the differences between stayers and movers, and stayers and extreme movers, are approximately 1.2 percent and 4.3 percent, respectively. Focusing on the top decile of families, we can see that moving a few times on average slightly increases incomes (+0.2 percent) while moving often still leads to a reduction of expected income by about 1.7 percent. Comparing those numbers with the results in percentile ranks we see that the two are approximately equivalent to each other.

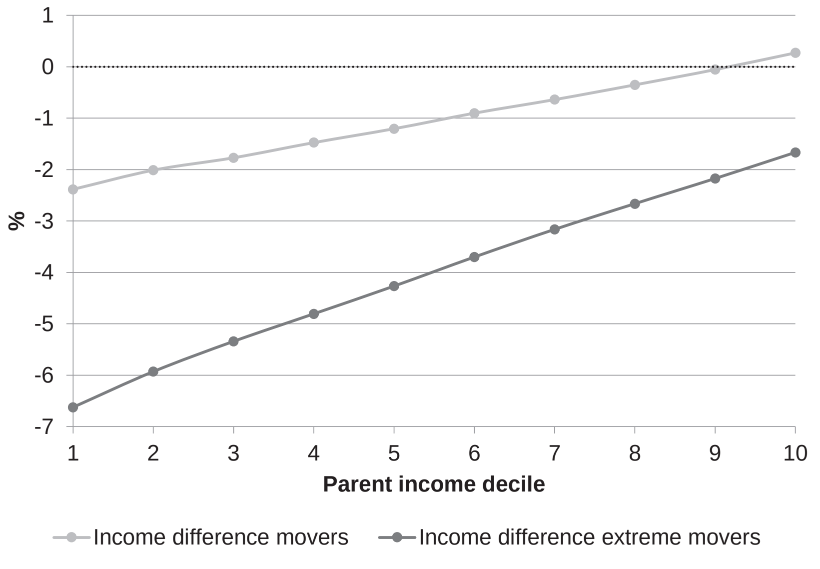
According to my results childhood moving has statistically significant effects on long run adult income. The economic significance of the results is less easily determined. However, a reduction of average annual income by 1.2 to 4.3 percent for children with median-income parents, and up to 6.6 percent lower income for children with low-income parents, can hardly be categorized as negligible.

Figure 3: Effects of moving on son long run adult income

(a) Levels of income



(b) Percent of income



4.1.1 Robustness checks

One could think that moves between municipalities are not informative enough in order to attribute the effects of moving to disruptions in child human capital development. Families can in practice just move a very short distance, across the border to the next municipality. Maybe there are different underlying reasons that make families move short distances compared to long distances. In order to strengthen the argument that moving between municipalities indeed causes damage due to the disruption in the social and physical environment, I check the effect of moves between larger geographic units.

As shown in Table 2 in Section 3.2 above, there is less movement between counties than between municipalities. Table 5 shows the total number of family moves broken down by the number of county moves included in the municipality moves. For example, of all 69,928 families that move two times during childhood between municipalities, 24,692 (35%) move within the same county while 29,129 (42%) move to a different county every time they move.

Table 5: Frequency of moves between municipalities and between counties

		Municipality							
		0	1	2	3	4	5	6	Total
County	0	499,408	84,909	24,692	5,485	1,566	400	176	616,636
	1	0	71,022	16,107	4,237	932	296	89	92,683
	2	0	0	29,129	7,953	2,417	662	337	40,498
	3	0	0	0	7,851	2,288	698	257	11,094
	4	0	0	0	0	2,468	807	450	3,725
	5	0	0	0	0	0	682	412	1,094
	6	0	0	0	0	0	0	541	541
	Total	499,408	155,931	69,928	25,526	9,671	3,545	2,262	766,271

The first robustness check is to redo the analysis in Section 4.1 using county moves instead of municipality moves. The results are shown in the second columns in Tables A.4 (sons) and A.5 (daughters) in the Appendix. The first columns just repeat the earlier results when using moves between municipalities, in order to facilitate the comparison. The estimates using inter-county moves are very similar to the estimates from municipality moves, some effects are slightly larger and some slightly smaller. The standard errors are larger using county moves, probably due to the smaller number of observations. The estimated parameters for the covariates are almost identical, as is the adjusted R^2.

A possible explanation for why the results are so similar is that children experience similar disruptions in their development when moving between counties or between municipalities. The important point is that even a municipality change in Sweden normally implies a change in school (or pre-school) and thus leads to a disruption in social and educational networks. Therefore, for a child it does not matter much if she moves 50 or 500 kilometers.

Since the estimates in Model (2) are not all significantly larger than in Model (1), it is unlikely that inter-county moves are driving all the results even in the municipality analysis. I show this by repeating the analysis using moves between municipalities, but this time excluding all families that move across counties. That is, I use only the observations in the first row of Table 5 (616,636 families in total). If the moves between counties were driving the results of the moving costs using inter-municipality moves, the effects should change notably when we exclude just those county moves. The results are shown in column (3). Some estimates are larger and some are smaller than the ones from Models (2) and (3) (and less precisely estimated), but the findings are in general again very similar.

These results show that the negative effects on long run income from moving during childhood are very robust in terms of the geographic unit chosen. We can be quite convinced that long distance moves are not driving the estimated moving costs on the municipality level.

One additional question is if the chosen measure of outcome, namely the rank of children's long run average outcome, is somehow affecting the results. The last column in Tables A.4 (sons) and A.5 (daughters) therefore presents the estimates of the moving costs equation using logged incomes instead of the income ranks for sons. The parent income on the right hand side is kept in terms of ranks in order to facilitate comparison. Assuming parent income to be at the median, the estimates fro the logged income specification indicate that one move during childhood reduces son long run income by approximately 1.5 percent. From Figure 3b we can see that the effect of moving a few times at parent income decile 5 is on average 1.2 percent lower income. The results are therefore quite similar to each other, taking into account that there are no observations with zero-incomes in the log model, and that the rank-percentage conversion is based on averages by parent deciles. The parameter estimates for the covariates all have the expected sign and are highly significant.

4.2 Effects of quality and timing of the move

The results in Section 4.1 suggest that moving during childhood is costly in terms of adult long run income. In this section, I test whether the timing of the move or the type of move matter, given that a family moves.

First, I estimate the parameters of the income equation (4) for all municipalities on the sample of stayers. There are 285 municipalities and the average adjusted coefficient of determination for those 285 equations is 0.25 (standard deviation 0.04, min. 0.14 and max. 0.4). Thus, the included variables explain on average one quarter of the variance in the long run child incomes. With those estimates I then predict the long run income for children who move once during their childhood in two regions each, once in their origin and once in their destination region. The difference of those two predictions for each child is the change in municipal quality $\triangle_i$. In some regions there are too few stayers in order to sensibly make predictions and the number of observations is therefore slightly smaller than than the number of one time movers in the sample (the number of observations of stayers for included region is at least 100 and the mean is 1,780) .

I estimated equation (5) using four age groups: 1-5 years old, 6-11 years, 12-15 years, and 16-17 years. The results are summarized in Table 6 (the results using 17 individual years of age can be found in Table A.6 in the Appendix). The estimates shown in this table are from a model using parent income quartile by origin municipality fixed effects which is supposed to absorb unobserved differences between the regions that affect the moving behavior of families. The correlation between the fixed effects and the explanatory variables is small but significant, 0.2 for sons and 0.3 for daughters which is an indication to use fixed effects rather than random effects. I perform Hausman's specification test to compare the estimator assumed to be consistent and inefficient (fixed-effects) to the efficient but possibly inconsistent estimator (random-effects) (Hausman, 1978). I reject the null (that the second estimator is efficient and consistent) on the 0.001% significance level.

Comparing Models (1) and (2) for the sons (and Model (3) and (4) for the daughters), we see that including the covariates reduces the main age effects by about one third (more for the daughters). The baseline category here is moving at age 1-5 and the average long run income for one-time movers is clearly decreasing in age at move for sons. The main age effects are less precisely estimated for the daughters but we see a similar trend here as well. The estimate of the effect of $\triangle_i$ for sons is 0.3, which means that sons who move at ages 1-5 to a one percentile better neighborhood (i.e. a region where sons very similar to them obtain long run average incomes as adults that are on average 1 percentile rank higher compared to children in the mover's region of origin), will increase their expected long run income by 0.3 percentile ranks. About one third of a region's quality is thus 'taken up' by a son if he moves there before the age of six.

The interaction effect for ages 6-11 is not significant, but it is significantly negative for the older two age groups. Figure 4 shows the marginal effects of $\triangle_i$ by age group as well as their 95 percent confidence intervals. From age 12, it is uncertain if moving to a region that usually produces higher income adults is beneficial for sons. From age 16, moving to a higher quality region has either no effect, or a negative effect on long run incomes (the marginal effect is negative on the 5 percent significance level). The picture looks slightly different for daughters. The confidence intervals are larger and there is no clear negative trend of the benefits of better regional quality over age. Just as in Section 4.1, the effects are also smaller in size compared to the sons. The marginal effect for the oldest age group for girls is not significantly different from zero, while it was negative for boys.

These results show that children who move at younger ages have on average higher adult incomes than children who move when they are older. Secondly, there is a strong negative trend of the effect of neighborhood quality on long run incomes over sons' age. In particular, even if a family moves to a significantly better area, their son will potentially have a lower adult income if he is over 15 at the time of the move, compared to if

Table 6: Timing and quality of the move

Child income	(1) Sons	(2) Sons+cov	(3) Daughters	(4) D.+cov
Moving age 6-11	-0.97***	-0.57*	-0.74***	-0.35
	(0.243)	(0.244)	(0.221)	(0.223)
Moving age 12-15	-3.10***	-1.94***	-1.66***	-0.94**
	(0.385)	(0.389)	(0.348)	(0.354)
Moving age 16-17	-5.40***	-3.88***	-3.22***	-2.26***
	(0.486)	(0.493)	(0.400)	(0.408)
Change in municipality quality △	0.33***	0.30***	0.18***	0.19***
	(0.026)	(0.026)	(0.029)	(0.029)
Moving age 6-11 # △	-0.08	-0.05	0.07	0.07
	(0.043)	(0.044)	(0.049)	(0.050)
Moving age 12-15 # △	-0.29***	-0.22***	0.03	0.03
	(0.065)	(0.066)	(0.074)	(0.076)
Moving age 16-17 # △	-0.53***	-0.46***	-0.16	-0.16
	(0.079)	(0.081)	(0.083)	(0.084)
Immigrated parents=1		-2.27***		-0.42
		(0.355)		(0.332)
Immigrated parents=2		-3.77***		0.32
		(0.510)		(0.479)
Continuous family=1		4.95***		1.89***
		(0.219)		(0.203)
Father was unemployed		-2.97***		-0.96***
		(0.266)		(0.248)
Father's education=2		0.76**		1.48***
		(0.251)		(0.233)
Father's education=3		1.87***		3.60***
		(0.378)		(0.353)
Father's education=4		1.42***		5.09***
		(0.376)		(0.350)
Mother's education=2		1.10***		1.17***
		(0.256)		(0.237)
Mother's education=3		1.77***		2.86***
		(0.351)		(0.326)
Mother's education=4		1.70***		4.43***
		(0.391)		(0.364)
F. age/birth		-0.15***		0.04
		(0.034)		(0.031)
M. age/birth		0.29***		0.26***
		(0.037)		(0.034)
Cohorts 72-76		-1.17***		0.26
		(0.204)		(0.189)
Constant	61.53***	54.99***	39.85***	27.45***
	(0.127)	(0.769)	(0.118)	(0.716)
Observations	82,236	80,139	77,722	75,589
R2	0.005	0.036	0.003	0.048

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: These estimates are based on the sample of one-time movers only and include origin-by-parent income quartile fixed effects. The base category is moving at age 1-5. The municipality type at birth has a zero effect and is omitted from the output. Including parent income rank in the equation yields nearly identical results.

they moved when their son was younger. If we think that the marginal effect of moving quality is actually negative for boys who move at age 16 or 17, this could be explained by the fact that children move down in the local distribution of school performance. If a family moves to a better area (measured in future adult income of the permanent child residents), their son will on average rank lower in the new, higher achieving municipality compared to the origin municipality. This can potentially lead to lower self esteem, a different peer group, or studying less - all of which can ultimately lead to lower adult earnings.

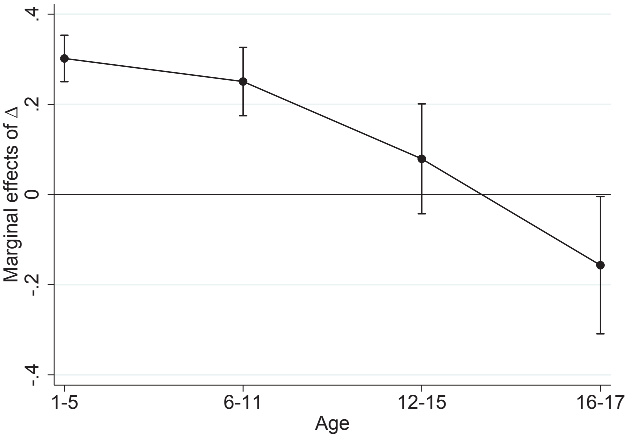
The interpretation of these results depends on the assumptions we are willing to make on both the human capital production function and moving costs. If we for simplicity assume that region quality affects human capital production in the same way and amount at each age, a child who moves earlier to a better neighborhood will, ceteris paribus, have higher human capital as an adult simply because she spent more time in this better region. If we abstract from any moving costs, long run adult income would then be a linearly decreasing function of child age at move, for a given $\triangle_i$. If moving costs in terms of long run income were positive (as the results in Section 4.1 strongly suggest), but constant over age, i.e. if the disruption of moving would have the same effect on adult human capital at any age, long run adult income would still be a linearly decreasing function of child age at move for a given $\triangle_i$, but on a lower level compared to zero moving costs. The results in this section however indicate that adult human capital is actually exponentially decreasing in age at childhood move, due to the decreasing marginal effect of moving quality over child age.

There are (at least) two different explanations for this. One possibility is that moving hurts only a certain part of human capital production, and this part is especially sensitive to disruptions during the teenage years. The effect of moving on adult income may work through weakening a child's social identity and network of friends, peers, and teachers. In this case, moving at older ages hurts long run income more because a strong social identity and a reliable social network are critical for human capital development at age 16 and 17, but not before that.

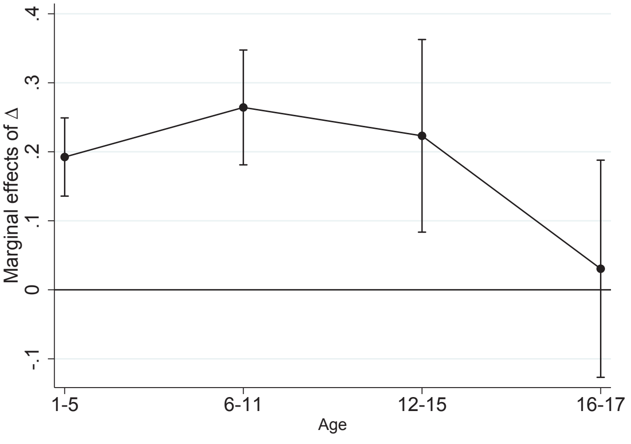
A second explanation is the compensation property of the human capital production function. If we assume that parents compensate for the disruption caused by childhood moving, my results simply indicate that there is less time to compensate for any harm done when there are only one or two years left of childhood human capital production. In practice, both explanation could explain part of the findings.

Figure 4: Marginal effects of moving quality

(a) Sons



(b) Daughters



Note: The plots show the marginal effect of change in municipal quality on adult long run income by age at childhood move for all one-time movers. The bars indicate 95 percent confidence intervals.

5 Discussion and concluding remarks

I used Swedish register data for nine cohorts born 1968 to 1976 to study the effect of moving between regions during childhood on long run adult income. In addition, I analyzed the effect of the timing and quality of the move. The literature on childhood human capital production suggests that shocks to the skill formation process can result in large differences in the stock of adult human capital. Moving is one such potential shock. The child is taken out of her usual environment and needs to re-build a social identity at the new home, get to know her new class mates, teachers, and neighborhoods. Early-intervention program evaluations have shown that the childhood environment strongly affects children's non-cognitive skills and social attachment, which are both very important for adult outcomes (Heckman and Carneiro, 2003).

The analysis of moving patterns revealed that 35 percent of all children born between 1968 and 1976 moved at least once between municipalities during childhood. Children with high income parents are over-represented among the families that move 1-3 times. Low income families are more likely than high income families to move four or more times with their children. On average families moved to regions that are very similar to their origin in terms of expected long run income for their child.

I estimated the effect of moving during childhood on long run income controlling for factors that affect both selection into moving as well as child long run income as adult. My results suggest statistically and economically significant effects of childhood moving on adult income that depend strongly on parent income. At the bottom of the parent distribution, a son that moves one to three times during childhood can expect to have an adult income that is 2.6 percentile ranks lower than if he had never moved. Sons with parents located at the median of the income distribution loose 1.2 and 4.2 percentile ranks by moving and extreme moving, respectively, while sons with parents in the very top only experience a loss in adult income when moving more than three times. I also show that percentile ranks can approximately be interpreted as percent of income. For daughters, the results are similar but smaller in size. Also, moving up to three times is not associated with a negative effect on adult long run income for daughters if their parents are located in the top twenty percent of the parent income distribution.

In order to test the robustness of my results, I repeated the analysis using only moves between counties instead of municipalities, using moves between municipalities but only within counties, and moves between municipalities using log son income as outcome. The results are neither sensitive to the definition of moving nor to the transformation of the income variable used.

In the next part, I analyzed if the timing of move or the quality of the move matter. In order to do so, I constructed a variable called change in municipality quality for all children that move once. Municipal quality is measured by the outcome of children with similar characteristics who spend their entire childhood in one region. The change in quality for the mover children is then the difference in the predicted outcome in the region of destination and region of origin, if they had spent all their childhood in one region.

My results show that, given that a family moves to a region that is similar to their region of origin, the loss in adult income is increasing in age at the move. Moving at age 16 compared to moving at 4, for example, reduces the expected adult income for sons by about 4 percentile ranks. It is therefore preferable in terms of child long run outcome to move when the child is young. Taking into account moving quality, i.e. the change in municipal quality from before and after the move, we saw that there are sizable benefits for children of moving to higher quality regions. Moving to a 10 percentile rank better region between ages 1-5, for instance, improves sons' adult long run income by 3 percentile ranks. However, those benefits disappear and then potentially become negative for sons if the move happens after he turned 12 and 16, respectively.

Further research is needed to determine what is driving these results. If we assume that childhood moving affects long run adult incomes by hampering the creation of social networks and identity, my results suggest that the later years of childhood are particularly crucial for building this type of human capital which is then capitalized during adulthood. It is also important to better understand how parents compensate for the disruptions caused by moving a child into a new environment. If it is relatively easy to compensate for the disruption we would also expect my results to show a decrease in the benefit of a higher quality region with child age, since the older a child is the less time there is to compensate.

Half the families, however, do not move to regions that are better for their child. This could reflect the fact that parents put on average very little weight on their child's expected utility when making a moving decision, compared to their own expected utility. Alternatively, parents could be unaware of the effects of moving on the utility of their children. Parents could also have the wrong expectations or insufficient understanding of the quality of different regions and their child's human capital production function. Therefore, a fruitful direction for future research is to deepen our understanding of which aspects of the childhood human capital production function are driving the estimated effects of moving on adult income, how parents' beliefs about the child human capital production function are formed, and how the preferences of parents look like when they decide to move their family.

References

Amato, P. R. (2001). "Children of divorce in the 1990s: An update of the Amato and Keith (1991) meta-analysis." *Journal of family psychology*, *15*(3), 355.

Angrist, J. D., and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Axelsson, R., and Westerlund, O. (1998). "A panel study of migration, self-selection and household real income." *Journal of Population Economics*, *11*(1), 113–126.

Becker, G. S. (2009). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press.

Black, S. E., and Devereux, P. J. (2011). "Recent developments in intergenerational mobility." In *Handbook of Labor Economics*, vol. 4, Part B, chap. 16, 1487 – 1541, Elsevier.

Chetty, R., and Hendren, N. (2015). "The impacts of neighborhoods on intergenerational mobility: Childhood exposure effects and county-level estimates." *Working paper*.

Chetty, R., Hendren, N., and Katz, L. F. (2015). "The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment." NBER Working Paper No. 21156.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). "Where is the land of opportunity? The geography of intergenerational mobility in the United States." *The Quarterly Journal of Economics*, *129*(4), i1.

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

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

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

Greenwood, M. J. (1997). "Internal migration in developed countries."

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

Hagan, J., MacMillan, R., and Wheaton, B. (1996). "New kid in town: Social capital and the life course effects of family migration on children." *American Sociological Review*, *61*, 368–385.

Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2004). "Disruption versus tiebout improvement: The costs and benefits of switching schools." *Journal of Public Economics*, *88*(9), 1721–1746.

Hausman, J. A. (1978). "Specification tests in econometrics." *Econometrica: Journal of the Econometric Society*, 1251–1271.

Heckman, J., and Carneiro, P. (2003). "Human capital policy." In A. Krueger, and B. Friedman (Eds.), *Inequality in America: What Role for Human Capital Policy?*, MIT Press.

Heckman, J., Pinto, R., and Savelyev, P. (2013). "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes." *American Economic Review*, *103*(6), 2052–86.

Heckman, J. J. (2006). "Skill formation and the economics of investing in disadvantaged children." *Science*, *312*(5782), 1900–1902.

Heckman, J. J. (2007). "The economics, technology, and neuroscience of human capability formation." *Proceedings of the National Academy of Sciences*, *104*(33), 13250–13255.

Heidrich, S. (2015). "Intergenerational mobility in Sweden: A regional perspective." *Umeå Economic Studies*, *916*.

Jäntti, M., and Jenkins, S. P. (2015). "Income mobility." In A. B. Atkinson, and F. Bourguignon (Eds.), *Handbook of Income Distribution*, *Handbook of Income Distribution*, vol. 2, chap. 10, 807 – 935, Elsevier.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). "Experimental analysis of neighborhood effects." *Econometrica*, *75*(1), 83–119.

Mincer, J. (1978). "Family migration decisions." *Journal of Political Economy*, *86*(5), 749–773.

Mincer, J. A. (1974). "Schooling and earnings." In *Schooling, Experience, and Earnings*, 41–63, Columbia University Press.

Nybom, M., and Stuhler, J. (2016). "Heterogeneous income profiles and lifecycle bias in intergenerational mobility estimation." *Journal of Human Resources*, *51*(1), 239–268.

Organisation for Economic Co-operation and Development (2015). *In It Together: Why Less Inequality Benefits All*. OECD Publishing, Paris.

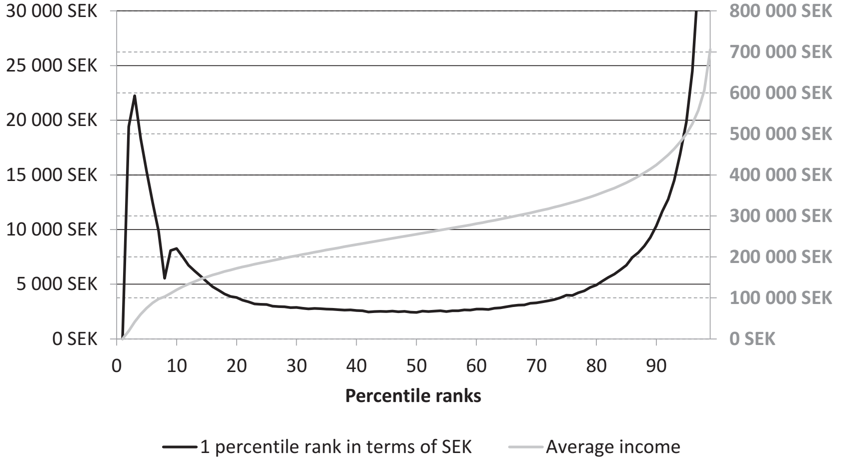
Pekkala, S., and Tervo, H. (2002). "Unemployment and migration: does moving help?" *The Scandinavian Journal of Economics*, *104*(4), 621–639.

Rabe, B. (2011). "Dual-earner migration. earnings gains, employment and self-selection." *Journal of Population Economics*, *24*(2), 477–497.

Skolverket (2004). "Elever med utländsk bakgrund." *Rapport till regeringen*, (Dnr 75-2004:545).

Appendix
A Figures

Figure A.1: Child income distribution: ranks and SEK



Note: The black line shows the number of SEK that each percentile in the child income distribution corresponds to (measured on the left Y-axis). The gray line shows the average adult long run income by percentile rank (measured on the right Y-axis). The distance between two percentile ranks measured in SEK is large in the tails of the distribution and around 2,500 SEK in the middle of the distribution.

B Tables

Table A.1: Child age at move
(one time movers only)

Age at move	Frequency	Percent
1	29,812	18.61
2	23,315	14.56
3	19,092	11.92
4	15,397	9.61
5	12,460	7.78
6	10,256	6.40
7	8,494	5.30
8	6,094	3.80
9	5,342	3.34
10	4,826	3.01
11	4,084	2.55
12	3,659	2.28
13	3,440	2.15
14	2,895	1.81
15	2,578	1.61
16	4,228	2.64
17	4,196	2.62
Total	160,168	100.00

Table A.2: Moving and parent income

	Parent income quartile					
	1	2	3	4	Total	
Moves	Row percent	Row percent	Row percent	Row percent	Frequency	Row percent
0	26.11	26.89	25.60	21.40	511,065	100.00
1	21.29	22.35	25.06	31.30	160,168	100.00
2	22.44	20.57	23.44	33.55	72,236	100.00
3	26.06	19.65	21.37	32.92	26,474	100.00
4	31.20	20.16	19.33	29.31	10,113	100.00
5	39.46	20.24	18.39	21.91	3,725	100.00
6	47.05	19.71	17.99	15.25	1,456	100.00
≥7	61.65	18.59	10.68	9.08	936	100.00
Total	25.00	25.00	25.00	25.00	786,173	100.00

Note: This table shows the distribution of families by parent income for each number of move. If parent income was independent of the number of moves, each cell would contain a quarter of the families (25 percent). Instead, we can see that parents in the top quartile are over represented among families that move one to 3 times, while families in the bottom quartile are over represented among the families that move four and more times.

Table A.3: Effect of moving on income, using mover groups

Income rank	Sons	Daughters		Sons cont.	Daughters cont.
Mover	-2.60***	-1.52***	F. educ.=4	0.35	3.88***
	(0.199)	(0.172)		(0.184)	(0.175)
Extreme mover	-7.24***	-3.62***	M. educ.=2	1.10***	1.52***
	(0.588)	(0.455)		(0.100)	(0.091)
Parent income rank	0.17***	0.14***	M. educ.=3	1.67***	3.22***
	(0.002)	(0.002)		(0.160)	(0.152)
Mover # PIR	0.03***	0.02***	M. educ.=4	0.80***	4.75***
	(0.003)	(0.003)		(0.190)	(0.183)
Extreme mover=2 # PIR	0.06***	0.02*	Born Metro	-0.21*	1.32***
	(0.011)	(0.009)		(0.097)	(0.091)
Imgr. parents=1	-1.98***	-0.22	Born Rural	-2.39***	-0.72***
	(0.169)	(0.153)		(0.152)	(0.138)
Imgr. parents=2	-4.20***	1.03***	F. age/birth	-0.06***	0.04***
	(0.250)	(0.223)		(0.015)	(0.013)
Cont. fam.=1	5.40***	2.20***	M. age/birth	0.31***	0.31***
	(0.101)	(0.091)		(0.016)	(0.015)
Unemployed	-2.24***	-1.11***	Cohort	-0.97***	0.48***
	(0.109)	(0.097)		(0.087)	(0.081)
F. educ.=2	0.63***	1.05***	Constant	43.11***	17.67***
	(0.099)	(0.090)		(0.356)	(0.324)
F. educ.=3	1.26***	2.51***	Observations	394,450	371,821
	(0.175)	(0.166)	Adjusted R2	0.061	0.068

Standard errors in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001

Table A.4: Robustness check (Sons)

	(1)	(2)	(3)	(4)		(1)	(2)	(3)	(4)
Income rank	Municipality	County	Municipality (within county)	Log		cont.	cont.	cont.	cont.
Moves=1	-1.53***	-1.77***	-1.38***	-0.0401***	Cont. fam.=1	5.29***	5.53***	5.22***	0.1011***
	(0.234)	(0.288)	(0.300)	(0.00584)		(0.101)	(0.100)	(0.113)	(0.00244)
Moves=2	-3.86***	-3.64***	-4.35***	-0.0925***	Unemployed	-2.20***	-2.24***	-2.12***	-0.0394***
	(0.338)	(0.437)	(0.529)	(0.00907)		(0.109)	(0.109)	(0.120)	(0.00262)
Moves=3	-5.85***	-5.99***	-6.89***	-0.1433***	F. educ.=2	0.64***	0.59***	0.59***	0.0083***
	(0.516)	(0.753)	(1.054)	(0.01438)		(0.099)	(0.099)	(0.107)	(0.00230)
Moves=4	-4.47***	-5.41***	-3.76*	-0.1148***	F. educ.=3	1.29***	1.21***	1.19***	0.0204***
	(0.779)	(1.209)	(1.899)	(0.02244)		(0.175)	(0.175)	(0.198)	(0.00409)
Moves=5	-10.10***	-8.38***	-6.09	-0.2339***	F. educ.=4	0.42*	0.31	-0.10	0.0106*
	(1.146)	(1.941)	(3.179)	(0.03473)		(0.184)	(0.185)	(0.217)	(0.00444)
Moves$\geq$6	-12.02***	-17.94***	-12.40*	-0.2887***	M. educ.=2	1.09***	1.08***	1.06***	0.0176***
	(1.311)	(2.428)	(5.011)	(0.04040)		(0.100)	(0.100)	(0.108)	(0.00232)
P. income rank	0.17***	0.17***	0.17***	0.0030***	M. educ.=3	1.68***	1.65***	1.45***	0.0247***
	(0.002)	(0.002)	(0.002)	(0.00005)		(0.160)	(0.160)	(0.180)	(0.00380)
Moves=1 # PIR	0.02***	0.02***	0.02***	0.0005***	M. educ.=4	0.82***	0.77***	0.61**	0.0081
	(0.004)	(0.005)	(0.005)	(0.00010)		(0.190)	(0.190)	(0.218)	(0.00459)
Moves=2 # PIR	0.04***	0.04***	0.03***	0.0010***	Born Metro	-0.21*	-0.34***	-0.28*	-0.0092***
	(0.005)	(0.007)	(0.009)	(0.00014)		(0.097)	(0.096)	(0.110)	(0.00232)
Moves=3 # PIR	0.05***	0.06***	0.06***	0.0014***	Born Rural	-2.38***	-2.38***	-2.73***	-0.0360***
	(0.009)	(0.012)	(0.019)	(0.00023)		(0.152)	(0.152)	(0.162)	(0.00353)
Moves=4 # PIR	0.02	0.04*	-0.03	0.0007*	F. age/birth	-0.06***	-0.05***	-0.04*	-0.0012***
	(0.013)	(0.020)	(0.036)	(0.00037)		(0.015)	(0.015)	(0.016)	(0.00036)
Moves=5 # PIR	0.10***	0.11**	-0.03	0.0028***	M. age/birth	0.30***	0.31***	0.30***	0.0055***
	(0.022)	(0.036)	(0.070)	(0.00058)		(0.016)	(0.016)	(0.018)	(0.00039)
Moves$\geq$6 # PIR	0.11***	0.12	0.07	0.0033***	Cohort	-0.98***	-0.94***	-0.87***	0.0505***
	(0.030)	(0.067)	(0.133)	(0.00076)		(0.087)	(0.087)	(0.096)	(0.00204)
Imgr. parents=1	-1.96***	-2.00***	-2.05***	-0.0435***	Constant	43.36***	42.27***	42.80***	12.1874***
	(0.169)	(0.169)	(0.191)	(0.00420)		(0.356)	(0.348)	(0.388)	(0.00847)
Imgr. parents=2	-4.20***	-4.22***	-4.13***	-0.0924***	Observations	394,450	394,450	318,148	389,854
	(0.250)	(0.250)	(0.284)	(0.00655)	Adjusted R2	0.061	0.061	0.055	0.041

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.5: Robustness check (daughters)

Income rank	(1) Municipality	(2) County	(3) Municipality (within county)	(4) Log		(1) cont.	(2) cont.	(3) cont.	(4) cont.
Moves=1	-1.01***	-0.74**	-1.30***	-0.03***	Cont. fam.=1	2.13***	2.25***	2.09***	0.06***
	(0.206)	(0.250)	(0.262)	(0.006)		(0.091)	(0.090)	(0.102)	(0.002)
Moves=2	-2.06***	-2.36***	-2.11***	-0.07***	Unemployed	-1.09***	-1.11***	-1.07***	-0.02***
	(0.283)	(0.361)	(0.432)	(0.008)		(0.097)	(0.097)	(0.107)	(0.003)
Moves=3	-3.29***	-3.08***	-2.04*	-0.10***	F. educ.=2	1.05***	1.04***	1.02***	0.01***
	(0.415)	(0.621)	(0.827)	(0.013)		(0.090)	(0.090)	(0.098)	(0.002)
Moves=4	-2.72***	-4.45***	-1.47	-0.12***	F. educ.=3	2.53***	2.52***	2.56***	0.03***
	(0.613)	(0.925)	(1.417)	(0.020)		(0.166)	(0.167)	(0.191)	(0.004)
Moves=5	-2.58**	-3.39*	-3.43	-0.14***	F. educ.=4	3.92***	3.92***	3.82***	0.03***
	(0.953)	(1.573)	(2.784)	(0.032)		(0.175)	(0.175)	(0.207)	(0.004)
Moves≥6	-6.77***	-8.91***	-2.80	-0.26***	M. educ.=2	1.52***	1.51***	1.53***	0.03***
	(0.876)	(1.576)	(3.528)	(0.035)		(0.091)	(0.091)	(0.099)	(0.002)
P. income rank	0.14***	0.15***	0.14***	0.00***	M. educ.=3	3.22***	3.21***	3.26***	0.04***
	(0.002)	(0.002)	(0.002)	(0.000)		(0.152)	(0.152)	(0.172)	(0.004)
Moves=1 # PIR	0.02***	0.01	0.02***	0.00***	M. educ.=4	4.76***	4.75***	4.78***	0.05***
	(0.004)	(0.004)	(0.005)	(0.000)		(0.183)	(0.183)	(0.212)	(0.004)
Moves=2 # PIR	0.02***	0.02***	0.02**	0.00***	Born Metro	1.31***	1.25***	1.47***	0.00
	(0.005)	(0.006)	(0.008)	(0.000)		(0.091)	(0.090)	(0.103)	(0.002)
Moves=3 # PIR	0.03***	0.03*	0.01	0.00***	Born Rural	-0.72***	-0.72***	-0.72***	-0.01*
	(0.007)	(0.011)	(0.016)	(0.000)		(0.138)	(0.138)	(0.147)	(0.003)
Moves=4 # PIR	0.02	0.06**	-0.02	0.00**	F. age/birth	0.04**	0.05***	0.03*	0.00
	(0.011)	(0.018)	(0.029)	(0.000)		(0.013)	(0.013)	(0.015)	(0.000)
Moves=5 # PIR	-0.01	0.01	0.03	0.00	M. age/birth	0.31***	0.31***	0.32***	0.01***
	(0.020)	(0.035)	(0.058)	(0.001)		(0.015)	(0.015)	(0.016)	(0.000)
Moves≥6 # PIR	0.03	0.05	-0.03	0.00***	Cohort	0.47***	0.49***	0.51***	0.07***
	(0.023)	(0.048)	(0.079)	(0.001)		(0.081)	(0.081)	(0.090)	(0.002)
Imgr. parents=1	-0.21	-0.23	-0.21	-0.02***	Constant	17.83***	17.21***	17.64***	11.84***
	(0.153)	(0.153)	(0.173)	(0.004)		(0.324)	(0.316)	(0.354)	(0.008)
Imgr. parents=2	1.03***	1.02***	1.14***	-0.01	Observations	371821	371821	298488	368475
	(0.223)	(0.223)	(0.258)	(0.006)	Adjusted R2	0.068	0.068	0.063	0.036

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.6: The effect of timing and quality of the move (17 separate moving ages)

Child income	(1) Sons	(2) Sons+cov	(3) Daughters	(4) D.+cov		(1) cont.	(2) cont.	(3) cont.	(4) cont.
age=2	0.48	0.25	0.43	0.32	age=3 # △	-0.02	-0.01	-0.21	-0.22*
	(0.351)	(0.347)	(0.327)	(0.322)		(0.099)	(0.098)	(0.115)	(0.113)
age=3	0.36	0.16	0.33	0.20	age=4 # △	0.06	0.06	-0.10	-0.08
	(0.371)	(0.368)	(0.349)	(0.344)		(0.104)	(0.103)	(0.122)	(0.120)
age=4	0.67	0.44	0.10	0.09	age=5 # △	-0.04	-0.01	-0.20	-0.21
	(0.396)	(0.393)	(0.374)	(0.368)		(0.111)	(0.110)	(0.129)	(0.127)
age=5	0.10	0.04	0.00	0.07	age=6 # △	0.17	0.18	-0.17	-0.17
	(0.427)	(0.423)	(0.401)	(0.395)		(0.120)	(0.119)	(0.138)	(0.136)
age=6	-0.05	0.06	-0.45	-0.37	age=7 # △	-0.08	-0.05	-0.18	-0.16
	(0.461)	(0.456)	(0.429)	(0.422)		(0.131)	(0.129)	(0.149)	(0.147)
age=7	-0.36	-0.11	0.05	0.18	age=8 # △	-0.13	-0.09	-0.17	-0.17
	(0.502)	(0.497)	(0.457)	(0.450)		(0.148)	(0.147)	(0.169)	(0.166)
age=8	-0.74	-0.42	-0.71	-0.43	age=9 # △	-0.21	-0.12	-0.13	-0.14
	(0.568)	(0.562)	(0.523)	(0.515)		(0.157)	(0.155)	(0.177)	(0.174)
age=9	-0.82	-0.40	-0.95	-0.78	age=10 # △	-0.08	-0.02	-0.02	-0.05
	(0.607)	(0.601)	(0.546)	(0.538)		(0.162)	(0.161)	(0.181)	(0.178)
age=10	-1.49*	-0.86	-0.71	-0.50	age=11 # △	0.03	0.07	0.10	0.08
	(0.627)	(0.621)	(0.574)	(0.566)		(0.173)	(0.171)	(0.199)	(0.196)
age=11	-1.67*	-0.90	-0.82	-0.40	age=12 # △	-0.37*	-0.34	0.08	0.07
	(0.673)	(0.667)	(0.619)	(0.610)		(0.181)	(0.179)	(0.208)	(0.205)
age=12	-1.41*	-0.60	-1.46*	-1.20	age=13 # △	-0.06	0.04	0.06	0.08
	(0.707)	(0.700)	(0.648)	(0.639)		(0.184)	(0.182)	(0.210)	(0.206)
age=13	-2.71***	-1.72*	-0.50	0.07	age=14 # △	-0.28	-0.20	0.08	0.12
	(0.728)	(0.721)	(0.662)	(0.653)		(0.202)	(0.200)	(0.226)	(0.223)
age=14	-4.44***	-3.50***	-1.14	-0.61	age=15 # △	-0.63**	-0.56**	0.39	0.40
	(0.791)	(0.783)	(0.715)	(0.705)		(0.218)	(0.216)	(0.243)	(0.239)
age=15	-3.33***	-2.29**	-2.60***	-1.91**	age=16 # △	-0.33	-0.23	-0.04	0.02
	(0.844)	(0.836)	(0.745)	(0.735)		(0.178)	(0.176)	(0.190)	(0.187)
age=16	-3.94***	-3.01***	-2.02***	-1.68**	age=17 # △	-0.80***	-0.69***	-0.24	-0.15
	(0.672)	(0.667)	(0.598)	(0.590)		(0.187)	(0.185)	(0.185)	(0.182)
age=17	-6.71***	-5.49***	-3.95***	-3.25***	Constant	61.06***	56.18***	39.78***	33.11***
	(0.730)	(0.724)	(0.563)	(0.555)		(0.234)	(0.322)	(0.216)	(0.296)
△	0.74***	0.60***	0.57***	0.24**	Controls	No	Yes	No	Yes
	(0.067)	(0.067)	(0.075)	(0.075)	Observations	82,236	82,236	77,722	77,722
age=2 # △	0.06	0.07	-0.39***	-0.40***	R2	0.006	0.037	0.003	0.048
	(0.093)	(0.092)	(0.107)	(0.106)	corr(u_i, Xb)	-0.182	0.075	-0.103	0.204

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

IV

A STUDY OF THE MISSING DATA PROBLEM FOR INTERGENERATIONAL MOBILITY USING SIMULATIONS

STEFANIE HEIDRICH*

ABSTRACT

Applied research on the association between parent and child lifetime income is relying on income data that covers only part of the life cycle which may lead to misleading estimates of the intergenerational elasticity (IGE). In this paper I study the bias of IGE estimates for different missing-data scenarios based on simulated income processes. Using an income process from the income dynamics and risks literature to generate two linked generations' complete income histories, I use Monte Carlo methods to study the relationship between available data patterns and the bias of the IGE. I find that the traditional approach using the average of the typically available log income observations leads to IGE estimates that are around 40 percent too small. Moreover, I show that the attenuation bias is not reduced by averaging over many father income observations. Using just one income observation for each generation at the optimal age (as discussed in the paper) or using weighted instead of unweighted averages can reduce the bias. In addition, the rank-rank slope is found to be clearly less sensitive to missing data.

JEL classification: E24, E27, J62
Keywords: intergenerational mobility, IGE, income process, lifetime income, simulations, Monte Carlo methods

*Department of Economics, Umeå School of Business and Economics, Umeå University.
E-Mail: stefanie.heidrich@umu.se. I would like to thank Thomas Aronsson, Spencer Bastani, David Granlund, Gauthier Lanot and David Sundström for helpful comments on this as well as earlier versions of the paper.

1 Introduction

There is a large literature on intergenerational mobility, i.e. the association between the lifetime income of a child and the lifetime income of her parents. The traditional measure of mobility is the intergenerational elasticity (IGE), which is the slope coefficient in a regression of the log of child income on the log of parent income. Importantly, the IGE is merely a summary measure that does not allow to discriminate between the different channels through which parent income affects child income (such as education, the home environment and neighborhood, or parent networks). Despite its rather vague implications for policy makers, the IGE is still a widely used measure in applied research. In order for the IGE to be unbiased[1], the researcher needs to observe the life time incomes for two linked generations. In practice, this is impossible today due to insufficient data. The measures that have been proposed to correct for the resulting bias (such as to use a single income observation at an age at which, on average, individuals' annual incomes are equal to their annual lifetime incomes) have been shown to be insufficient (see Nybom and Stuhler 2016). In this paper, I borrow a well-established income process from the macroeconomics literature to simulate the complete income trajectories for two generations. In this literature on income and consumption risks over the life cycle, the estimation of income processes given panel data has been a major focus for many decades. Based on the incomes generated from this income process I study the bias of the IGE for different missing-data scenarios.

The association between the economic achievements of an individual and his/ her parents has been studied by economists for a long time. Becker and Tomes (1979) modeled the decision problem of parents who care about their own consumption and about their child's future wealth. Becker and Tomes assumed that the adult income of children is a function of market luck plus genetic endowments and luck, as well as human and non-human capital investments received from their parents. This results in a positive correlation between parent and child incomes. Since this seminal work, the empirical evidence of intergenerational income mobility, i.e. the degree to which parent income predicts child adult income, has been studied extensively. Higher income mobility implies a lower association between parent and child income and a lower IGE. Recent overviews of the intergenerational mobility literature can be found in Björklund and Jäntti (2009) and Black and Devereux (2011).

Snapshots of individual incomes are not necessarily a good indicator of permanent or average life time income. Individuals usually experience both a certain pattern of income growth depending on age, education, and other variables, as well as more or less persistent income shocks. Thus, in order for the IGE to be meaningful, some measure of long run income must be employed. Nonetheless, the early studies of intergenerational mobility

[1] I (ab)use the term bias in this paper following the intergenerational mobility literature, talking about the estimates being biased instead of the proper way of referring to the OLS-estimator for β being biased due to measurement error.

used just the available income snapshots for sons and fathers to estimate the IGE.

Soltow (1965) studied the income distribution in the city of Sarpsborg in Norway and found an IGE of 0.14, using the log income in 1960 for both fathers and sons in his regression. Atkinson (1980) tracked down the children of the respondents of a survey conducted in York in 1950 to estimate the IGE. Their sample size was 307 and the main income measure used was weekly earnings reported at some point between 1975 and 1978 for sons, and during 1950 for fathers. Using income observations at some point in calendar time is obviously problematic since children and parents will be at very different stages of their life cycles. A summary of additional early studies can be found in Becker and Tomes (1986).

The next group of studies acknowledged that using income snapshots will not give an accurate description of intergenerational mobility. Solon (1992), for example, assumed a classic errors-in-variables model for the observed incomes and used both up to five-year-averages and age controls in their empirical application. In recent years, researchers have turned away from the classic measurement error assumption. Haider and Solon (2006) pointed out that the relationship between annual income and average lifetime income varies over the life cycle, which is not consistent with a classic measurement error. They made use of panel income data from social security records to show how yearly income observations relate to long-run income. Regressing each yearly observation on the long-run income, Haider and Solon focused on finding the year which, on average, lies closest to the long-run measure. Using this particular year was considered a way to minimize the measurement error in estimating the IGE.

Nybom and Stuhler (2016) (N&S) have recently noted that using the annual income at the "correct" age to proxy long-run income according to Haider and Solon (2006) will still not estimate the true IGE. The reason for this is that for the method of Haider and Solon (2006) to work, the deviations of annual income from the average lifetime income of the child have to be uncorrelated with parent income. The shape of the child's income path over the life cycle depends, however, on education and other background variables, which are related to their parents. N&S's study is based on the hitherto most complete data in the IGE-literature: they make use of 29 income years for the sons (age 22 - 50) and 33 income years for the fathers (age 33 - 65) from the Swedish tax registry. The data is, however, incomplete; the typical pattern of observing early income years for the sons and later years for the fathers is still present even in the study by N&S. Since the annual income observations of one generation fan out over age and become more unpredictable, lacking the income observations at older ages for the younger generation is especially problematic when approximating average life time income as I will show below.

Since it will take a long time until the complete life time income histories for two linked generations will be available in the registers (around 15 years in Sweden and longer in most other countries), there is scope for a simulation study. This is the first paper studying the missing-data problem for the IGE using simulations. I use a well-established model and estimated parameters for the earnings process over the life cycle from the earnings dynamics literature to generate life time income data for two linked generations. This earnings data reproduces all earlier findings in the literature regarding the bias of the IGE, such as left-hand side measurement error which is negative when son income is observed at young ages, and positive when son income is observed at old ages. I also show that having left-hand side measurement error based on "nearly complete" income histories of the two generations as in Nybom and Stuhler (2016) still significantly underestimates the IGE. As long as incomes for sons at older ages (and for fathers at young ages) are missing, the estimates will be severely biased even though there are around 30 observations available for each generation.

I find that the IGE is highly sensitive to missing observations over the life cycle. Using only one income observation for the son and the father each to estimate the IGE, I show that, in principle, there exist many different possible age combinations at which the IGE will be unbiased. As a general rule, IGE estimates are only very little biased when income observations for son and father at the same age are used. This is a consequence of the assumption in my data generating process that the income paths of son and father are correlated, by imposing family-specific returns to experience. This assumption is reasonable and has been shown to hold for Swedish register data (N&S). In this paper, I have only assumed two different family-background types in my data, based on education level. In practice, there will be many groups with different life cycle patterns depending on a whole range of variables. Still, as long as father and son life cycle income patterns are correlated, measuring their income at the same age may give IGE estimates that are very close to the true value.

In addition to the above, I also find the following. First, I find no evidence that using the average over a large number of father income observations diminishes attenuation bias, contrary to the result in Solon (1992). This is possibly due to the fact that Solon assumed the deviations between annual and average lifetime income to be independent random errors, which does not hold in general as discussed above.

Second, estimates of the rank-rank slope are less sensitive to the income measure than estimates of the IGE. This result may be driven by the fact that individuals change income rank less often over their life course than dollar income. An income increase for an individual, for example, does not necessarily imply a change in her *relative* position in the income distribution. Average life time rank is therefore easier to approximate correctly than average life time income.

In the last part of the analysis, I mimic the typical data that is available to a researcher attempting to estimate the IGE. Usually, income is observed during a limited time span over a number of calendar years. That is, income is observed for the son and the father generation during the same years and thus usually during very different ages. An additional problem is that parents get children at different ages. The father age during the calendar years for which incomes are available for the researcher therefore differs greatly. I find that using the average of all available log income observations for sons and fathers (as traditionally done in the literature) leads to IGE estimates that are on average 41 percent too small. This approach thus heavily overstates income mobility. Using only the last available observation for the son and the first available observation for the father, or weighted averages where the later (earlier) observations for the son (father) are weighted more heavily, gives estimates that are on average 1 and 3 percentage points less biased, respectively. These results show that the effect of measurement error of lifetime income leads to severe measurement error of the IGE, respectably of the way the available income observations are used. There is not found to be any advantage of using the average over all available years as usually done in the literature.

I also estimate the association between income ranks given the typical data. The rank-rank slopes are only underestimating the true value by approximately 5 percent when using the simple mean over all available income ranks. Using only the last available observation for the son and only the earliest available observation for the father is found to perform equally well. Employing instead a weighted average is found to perform slightly better, the estimates are on average only 4 percent too small.

The rest of the paper is organized as follows. I present the details of the income generating process in Section 2. In the first part of Section 3, I show how my simulated data matches characteristics of income histories in the literature. The second part focuses on the bias of the IGE based on single year and several year averages of income observations based on my data. In the last part of Section 3, I analyze a typical snapshot of income data often found in the literature and present the preferred income measures for both the IGE and the rank-rank slope. Section 4 concludes.

2 Generating income paths

There is a large literature modelling individuals' earnings over the life cycle for a variety of purposes. For example, in the context of modelling labor supply (see Abowd and Card, 1989), to study individuals' earnings risk (Gourinchas and Parker, 2002), the incidence and persistence of low income spells (Atkinson et al., 1992), and the calibration of consumption and saving models and dynamic general equilibrium models (Browning et al., 1999).

There are at least two different strands of earnings process literature. In the first one, researchers following Lillard and Weiss (1979) allow for heterogeneity in individuals' income profiles, while exogenous shocks are only modestly persistent. The second strand following MaCurdy (1982) assumes that individuals instead face very similar life cycle income profiles, but different and extremely persistent shocks (often modeled as everyone having a unit root). The income process I use here follows the first strand and is based upon Guvenen (2009) who estimated a model featuring both an individual specific age profile and stochastic income shocks for PSID data. Guvenen (2009) also showed that MaCurdy's results on the homogeneity of income profiles and persistence of shocks were based on a test with low power, which strengthens the case for heterogeneous income profiles and modest shock persistence. Below, I mainly follow Guvenen's notation. The basic characteristics of the model are very common in the literature and similar versions can be found for example in Storesletten et al. (2004) or Karahan and Ozkan (2013). I start by describing the statistical model for earnings of what I call the father generation. I will then explain how the process from the second generation (the sons) differs from the fathers' process and how the two generations are linked.

$$y_{h,t}^{i,e} = g(\phi, \delta^e, X_t) + f(\alpha^{i,e}, \beta^{i,e}, X_{h,t}^i) + z_{h,t}^{i,e} + \varepsilon_{h,t}^i \tag{1}$$

Equation (1) shows that the log income y of an individual i in year t with experience h and education level e is assumed to be a linear combination of four components: two life cycle components and two different shocks. The first life cycle component is identical for everyone with parameters that depend only on group variables such as birth cohort or education, while the second is heterogeneous across all individuals (possibly with different distributions of the parameters across subgroups). The function g, the first life cycle component, is here defined as

$$g(\phi, \delta^e, X_t) = \phi + \delta^e x_t \tag{2}$$

where x_t is a deterministic age-dependent sequence that reflects the age-efficiency profile of labor. This variable was first used in Hansen (1993) and the particular values were obtained from Conesa et al. (2009)'s FORTRAN code. Those values have been used in several studies, see for instance Huggett and Ventura (1999) or De Nardi (2004). Both ϕ and δ^e are assumed to be identical for all individuals within one cohort, but not necessarily between cohorts (for readability I refrain from using an additional cohort-index in the formulas). As noted by Guvenen (2009), this part of the income process can be used to capture a changing skill premium over time.

The second term on the right hand side of equation (1) is the main contribution of Guvenen (2009), the explicit modelling of heterogeneous returns to experience over the life cycle. The function is defined to be linear in experience,

$$f(\alpha^{i,e}, \beta^{i,e}, X_{h,t}^i) = \alpha^{i,e} + \beta^{i,e} h. \tag{3}$$

The parameters in this function thus vary both across individuals and education level, with distributions $\alpha^{i,e} \sim N\left(0, \sigma_{\alpha,e}^2\right)$, $\beta^{i,e} \sim N\left(0, \sigma_{\beta,e}^2\right)$, and $Cov\left(\alpha_i^e, \beta_i^e\right) = \sigma_{\alpha,\beta}$.

The stochastic part of earnings is modeled as a purely transitory shock $\varepsilon_{h,t}^i$ and an AR(1) process $z_{h,t}^{i,e}$,

$$z_{h,t}^{i,e} = \rho^e z_{h,t-1}^i + \eta_{h,t}^i, \quad z_{0,t}^i = 0\,. \tag{4}$$

The innovations $\eta^i_{h,t}$ and $\varepsilon^i_{h,t}$ are assumed to be independent of each other and normally distributed with mean zero and variances σ^2_η and σ^2_ε.

I make the following simplifying assumptions when generating the income streams of the fathers. Everyone is part of the same cohort (in order to fix ideas we can think of a sample of male individuals born in 1950 in the US). Those individuals will have one of two education levels, namely either a high school degree or a college degree. I use the share of male individuals between 25 and 29 years who have a college degree in 1976 as reported by the U.S. Census Bureau (2016) for the father generation and in 2010 for the son generation. I follow the literature cited above and define experience h as the *potential* experience, that is, the number of years since entering the labor market after finishing education. Thus, I will only have one cohort and two different experience-by-age streams in my sample of fathers. I generate incomes for 45 years, from age 22 where the college group enters the labor market, up to age 66. The values for the parameters and parameter distributions are taken from Guvenen (2009) who used PSID data to fit their model.[2]

The income process of the second generation is in principle the same as the process described above for the first generation. The difference lies in the parameters ϕ and δ^e of the function g (the first life cycle component), which capture changes in the average income as well as the returns to skills over generations. I assume both average income and the returns to skills to be higher in the second generation (see Abel and Deitz (2014) on the wage premium for college degrees over time).

I construct two links between the father and the son generation. First, I assume that the probability of a son having a college degree depends on the education level of the father. More precisely, a son whose father has a college degree has a four times higher chance of going to college himself. The second link is via the individual return to experience given in function f. The parameters α and β in this function have a different distribution in each of the two education levels of the fathers. However, the fact that the return to experience is differently distributed among fathers with and without college degree does not necessarily mean that this is due to the actual education. It could also be the case that the returns differ due to inherited family traits such as ability, opportunities, networks, and non-cognitive skills which they got from their parents, which on average differ by education level, and which affect the returns to experience. In order to mimic the inheritance of those marketable traits, I draw the sons' values for α and β from the same distributions as for the fathers and, in addition, impose a correlation of 0.15 between the parameters of the sons and the fathers. A son with a college educated father does not necessarily go to college himself, but he might still benefit from his father's skills and network. The data I generate is summarized in table 1. All parameters used in my simulations can be found in Table A.1 in the Appendix.

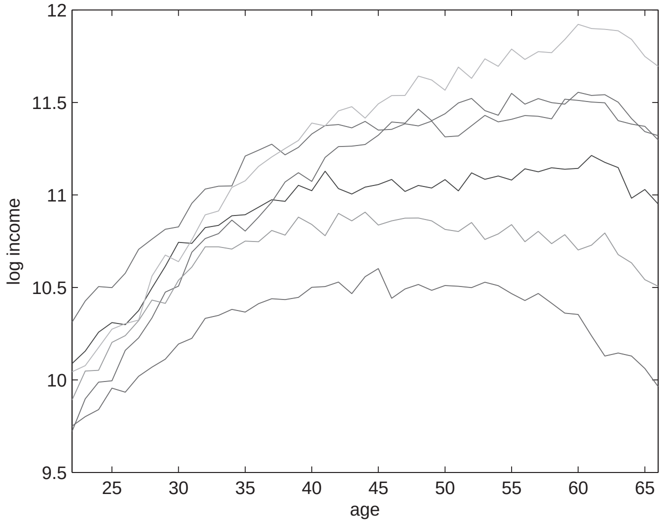
[2]Guvenen (2009) used minimum distance estimation to fit the model to the data, see the paper for more details.

Figure 1 shows the earnings over they life cycle for 10 fathers in my sample. We see the typical hump-shape over the years, with earnings increasing quickly at young ages in order to level off at around mid age, and finally decreasing when hitting retirement. Importantly, the ages in this study are not to be interpret literally. The age span chosen here to define life time earnings could be changed to include even younger and older ages.

Table 1: Summary of the data

	Fathers	Sons
Number of observations	10,000	10,000
Observed age span	22-66	22-66
Share holding college degree	0.24	0.26
Average log life time income	10.94	12.14
Variance of average income	0.25	0.52
Relative skill premium college/high school (δ^e)	1.2	1.6
Intercept (ϕ)	9	10

Figure 1: Income over the life cycle



3 Data availability, data usage, and the IGE

The true IGE for my sample is the slope coefficient β_1 from a regression of the average log lifetime income of the son in family $m = 1, \ldots, N$ ($\bar{y}^s_m$) on the average log lifetime income of his father ($\bar{y}^f_m$):

$$\bar{y}^s_m = \beta_0 + \beta_{IGE}\bar{y}^f_m + \varepsilon_m \,, \tag{5}$$

$$\bar{y}^s_m = \frac{1}{45}\sum_{t=22}^{66} y^s_{m,t} \tag{6}$$

$$\bar{y}^f_m = \frac{1}{45}\sum_{t=22}^{66} y^f_{m,t} \,. \tag{7}$$

Sometimes, researchers have assume a discount rate in order to construct the so called annuitized average lifetime income. Since this does not change the qualitative results I choose to use the simple averages as shown in equations (6) and (7). The true IGE for my sample is 0.63, which means that roughly two thirds of the father's deviation of his lifetime income from his generation's average lifetime income is inherited by the son. There is a strong association between father and son lifetime incomes.

The IGE can be written in terms of the model parameters from the underlying data generating process. The average over 45 years of the income process for son and father can be written as follows:

$$\bar{y}^s_m = \phi^s + \delta^{s,e}\bar{x} + \alpha^{s,e} + \beta^{s,e}\bar{h}_e \tag{8}$$

$$\bar{y}^f_m = \phi^f + \delta^{f,e}\bar{x} + \alpha^{f,e} + \beta^{f,e}\bar{h}_e \tag{9}$$

where $\bar{x}$ and $\bar{h}_e$ are the averages of labor efficiency units and experience, respectively ($\bar{h}_e$ depends on the education level). The mean of the AR(1)-process z and the purely transitory shock ε are zero (large sample assumption, $T > 30$). The IGE can then be expressed as

$$\beta_{IGE} = \frac{Cov\left(\bar{y}^s_m, \bar{y}^f_m\right)}{Var\left(\bar{y}^f_m\right)} = \frac{Cov\left(\alpha^{s,e} + \beta^{s,e}\bar{h}_e, \alpha^{f,e} + \beta^{f,e}\bar{h}_e\right)}{Var\left(\alpha^{f,e} + \beta^{f,e}\bar{h}_e\right)}. \tag{10}$$

The random variables α and β are drawn from two different distributions within each generation (depending on education level), and α and β are correlated with each other within each education level, but not across. Moreover, the α's of sons and fathers are correlated, as well as the β's, but again not across education levels. The IGE therefore depends on the variance-covariance matrix between the α's and β's of each generation- and education level, the average experience, and the share of fathers and sons in each education level.

In the next part of this section, I show how my data matches findings from the IGE literature in terms of measurement error and bias of the IGE. In Section 3.2, I rigorously study the bias of the IGE based upon the complete income histories in my sample. In Section 3.3 I focus on a snapshot of data that is typically available to a researcher and show which income measure gives the least biased estimate given this incomplete data.

3.1 Matching the important characteristics of income data

The data I generated with the model in Section 2 shows the typical relationship between annual income and average lifetime income. Annual income at younger ages lies below the lifetime average, while annual income at older ages lies above the lifetime average. Haider and Solon (2006) (H&S) suggested that one could improve the estimate of the IGE by using one annual income observation for the sons at which, on average, the annual income is closest to the life time average income (termed the Generalized Errors-in-Variables model). To find this age, they estimate the following equation:

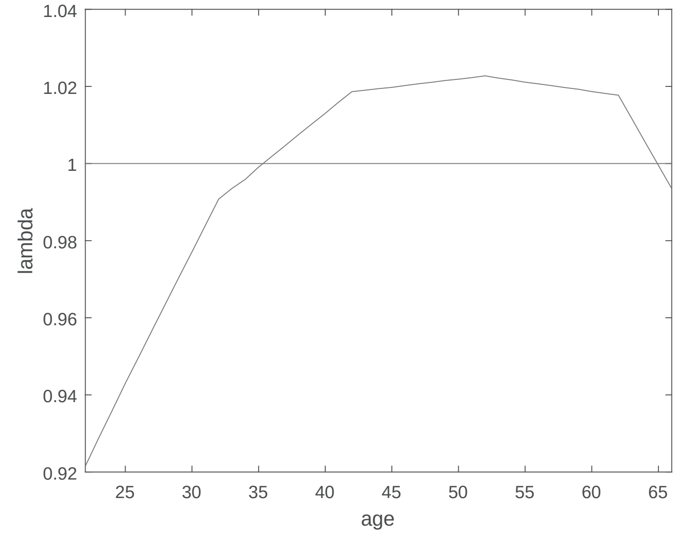
$$y^s_{m,t} = \lambda_t \bar{y}^s_m + \nu_{m,t} \tag{11}$$

where $\nu_{m,t}$ is an i.i.d. error. Note that implementing this strategy actually requires observing income over the whole life cycle in order to compute $\bar{y}^s_m$. Figure 2 shows the estimates of lambda for my sample of sons. The reference line has the value 1 and we can see that, on average, annual earnings are closest to the average life cycle earnings at age 35, as well as at age 65 The difference between my results and earlier findings is that annual earnings and average lifetime earnings actually cross twice. This is because I observe son income around retirement age which is not available in empirical work.

The bias of the IGE that is due to using an incorrect proxy for son average lifetime income is in the literature typically referred to as left-hand side measurement error. Using single annual income observations as a proxy for average lifetime income will lead to estimates of the IGE that are negatively biased if incomes during young ages of the son are used, and to a positive bias if incomes during older ages of the son are used since average lifetime income is under- and overestimated, respectively. This empirical fact holds also true in my simulated data as demonstrated in Figure 3. Here, 45 different IGE estimates are plotted, obtained by regressing single income observations for the son at each of the 45 observed years on the true average life time income of the father. The horizontal reference line indicates the true IGE at 0.63. We see the typical pattern discussed in H&S, Böhlmark and Lindquist (2006), N&S and others that the IGE is indeed underestimated when income during young ages are used to measure lifetime income, and that the IGE is overestimated when income during older ages are used.

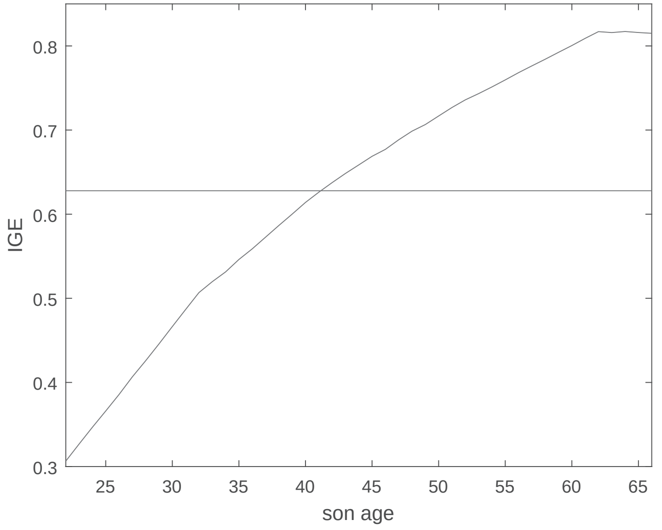
In this data, using the incomes observed at son age 41 gives an IGE that is closest to the true IGE. This means that the criticism by N&S against a proxy of lifetime income based on λ in equation (11) is also shared by my findings based on the simulated data. As we

Figure 2: Estimates of lambda over the sons' life cycle



Note: This figure shows the slope coefficients from regressing 45 annual income observations of the sons on their average lifetime income. At age 35 and age 65, annual incomes are on average closest to average lifetime income (where the slope coefficient equals one).

Figure 3: IGE estimates over the life cycle of the sons

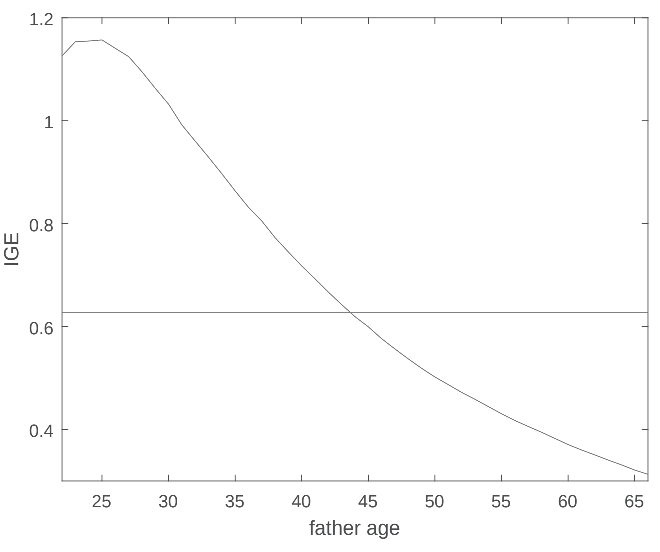


Note: This figure shows the IGE estimates from regressing 45 annual log income observations of the son on the average log lifetime income of the father. At age 43 the estimate is equal to the true value of the IGE.

can see from Figure 3, using annual income at age 35 (age 65) as suggested, will actually give a significantly downward (upward) biased IGE. As N&S pointed out, this is due to the fact that at the age where, on average, annual income is closest to life time income, the individual incomes still vary so much in regard to being below, at, or above their individual life time averages, that the resulting IGE at the age where λ is approximately one cannot actually eliminate the bias. The Generalized Errors-in-Variables model only generates unbiased estimators if idiosyncratic deviations from the average income path are independent of individual and family characteristics which cannot be expected to hold in general, and might be especially problematic when studying intergenerational mobility. In addition, there is no connection between the construction of lambda shown in equation (11) and the covariance between average son and average father income, and the variance of father income, and there is therefore no reason why lambda should be able to indicate the year that will give the least biased estimate of the IGE.

Right-hand side measurement error occurs when the annual father income used to proxy for average lifetime income is measured at the wrong age, i.e. at an age where the annual income or the average of several observed annual incomes is not equal to the true average lifetime income. Figure 4 shows again 45 different IGE estimates based on my data, this time obtained by regressing the true average life time income of the sons on each of the 45 observed incomes of the father. The horizontal reference line indicates again the

Figure 4: IGE estimates over the life cycle of the fathers



Note: This figure shows the IGE estimates from regressing the average log lifetime income of the son on 45 annual log income observations of the father. At age 43 the estimate is equal to the true value of the IGE.

true IGE at 0.63. We see that the IGE is too large when income during young ages of the fathers are used, and that the IGE is too small when income during older ages of the fathers are used, just as in the applied literature. The best age to measure father income in my sample is 44. Combining left- and right-hand side measurement error (observing young sons and old parents as typically found in the literature), the available estimates of the IGE are potentially heavily downward biased. Overall, my simulated data matches the important characteristics of son and father income over the life cycle that are reported in the literature very well.

3.2 The effect of missing data on the bias of the IGE estimator

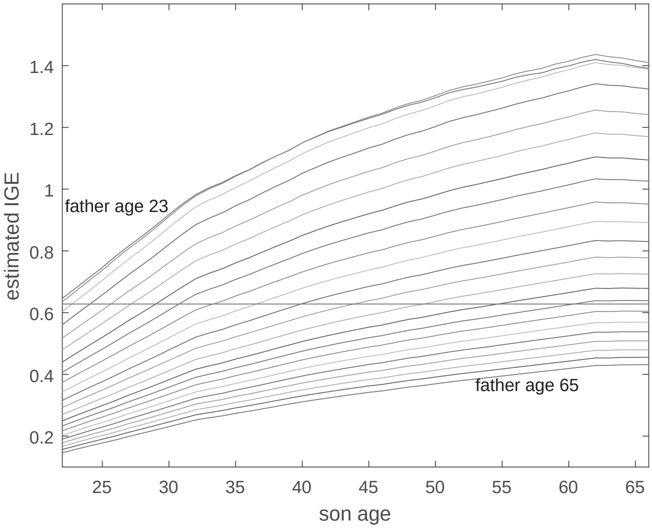
In this section, I study in detail the bias of the IGE arising when individual or several annual income observations are available for sons and fathers. In the first part I use income measured in log incomes. In the second part I focus on income expressed in terms of percentile ranks which has become more and more common in the literature.

3.2.1 Log income

I start by demonstrating the effect of using just one single income observation for each generation as a proxy for lifetime income, based on my simulated father and son sample with complete income histories. Figure 5 shows the estimated IGE for all possible combinations of father and son individual income observations. The horizontal reference line indicates the value of the true IGE (for readability I only plotted every other father age, i.e. income observed at age 23, 25, 27, etc.). The highest line represents estimates of the IGE over the full range of possible ages at which son income is observed, while holding the age at which father income is observed constant at 23. Moving down from line to line I hold constant the father income observation at higher and higher ages. The line at the bottom then represents the estimated IGEs over the full range of son ages, while holding father age constant at 65. The figure shows that, in principle, there are many different combinations of ages to observe income for the two generations that will give the correct IGE (wherever the reference line crosses the estimated plots). In fact, the estimates will be close to the true value whenever the incomes are observed at similar ages. Observing income at age 30 for sons, for instance, requires measuring annual father income at age 36 to estimate the true IGE. This is shown in Figure 6, where one income observation for sons and fathers each *at the same age* is used to estimate the IGE. According to my data, measuring annual income at age 45 for both sons and fathers will give the true IGE. Still, using an earlier or a later age to measure both generations' incomes will result in an estimate of the IGE that is surprisingly close to the true value (compare the scale of the y-axes in Figure 5 and Figure 6). The largest difference is found at age 32, but the estimate is still just 26 percent too large.

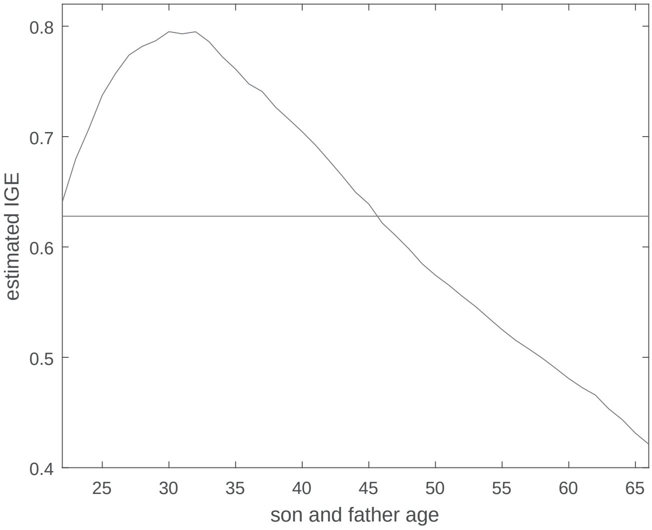
Figure 5 also illustrates why the IGE is severely downward biased in most studies. If incomes are observed at a mixture of medium to high father ages, the IGEs one could

Figure 5: IGE estimates using single income observations



Note: This figure shows half of the IGE estimates using all possible combinations of single annual income observations for both generations. The horizontal line indicates the true IGE. Each line shows the IGE estimates over the life cycle of the son, holding constant the age at which father income is observed.

Figure 6: IGE estimates using single income observations at the same age



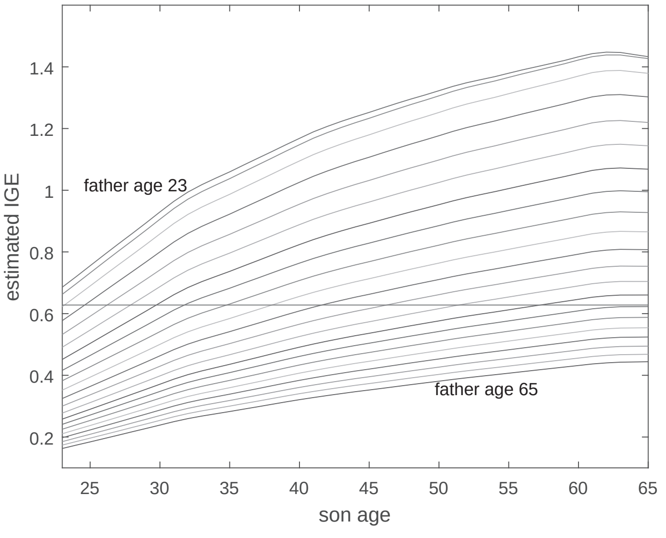
Note: This figure shows the IGE estimates using a single annual income observation for both generations at the same age. Using income at age 45 for both father and son results in an unbiased IGE estimate.

possibly estimate are located on the lower half of the lines representing estimates by father's age. Restricting son income to be observed at an age up to around 40, the possible IGEs that can be "reached" reduce to the points in the lower-left corner of the plot. Since all those points are located below the true IGE, no combination of the available income observation could ever give the true value. Note also that a portion of the IGEs in the upper right part of the figure are above 1 in size, against the usual assumption of regression to the mean according to which the IGE should be bound between zero and one. If income is affected strongly by education, education is inheritable, and if the skill premium is rising over generations, there is no reason why the IGE should not be able to be larger than one.

A common practice in the literature is to use the average of several observed annual income observations. Using the average over several observations for the parents is assumed to be beneficial since it in theory decreases attenuation bias (Solon, 1992; Björklund and Jäntti, 1997). The argument for using the average over several annual income observations for the sons is to reduce noise. Figure 7 repeats the analysis in Figure 5, this time using three year averages around each age from 23 to 65 for both generations. The results are very similar to using only single observations. Using the three year average of incomes around age 23 for the sons requires the average income around age 28 for fathers in order to estimate the correct IGE. Son average income around age 30 requires using the three year average of father income around age 36. The effect of using average income around the same age for both generations is illustrated in Figure 8. The curve is as expected smoother than in Figure 6. Using the average of three annual income observations around age 46 for both generations gives again the best IGE estimates, just as in Figure 6.

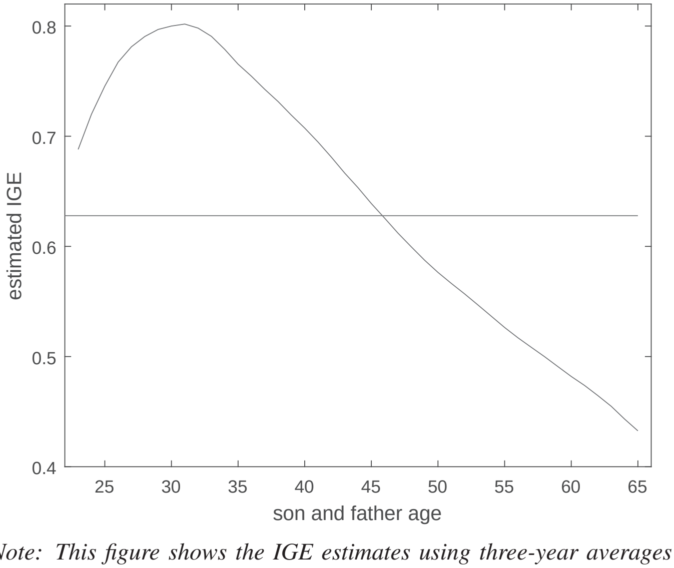
In order to test for the theoretical prediction by Solon (1992) that averaging over many parent income observations decreases attenuation bias, I compute three-, five-, eleven-, and twenty one-year averages for the fathers over all possible ages and then estimate the IGE (using the true average lifetime income for the son). As shown in Figure 9, there are only very small differences in the estimated IGE when averaging over an increasing number of father income observations. Zooming in on the lower right part of the graph around father age 55 (which is usually available in terms of real data), we see that the average over eleven annual income observation is slightly lower and thus furthest from the true value (this average can be observed up to age 60). There is therefore no evidence that attenuation bias is reduced by averaging father incomes. One possible reason for this finding is that the theoretical result in Solon (1992) was based on the assumption that each annual income is the sum of average life time income and a random transitory fluctuation that is uncorrelated with each other and true average lifetime income. However, as H&S and N&S showed, deviations from the average income actually follow a pattern over the life cycle, which also depends on family background (and therefore are not random). Hence, there is no apparent reason to continue with the practice of blindly averaging over all parent income observations in applied work.

Figure 7: IGE estimates using three-year averages



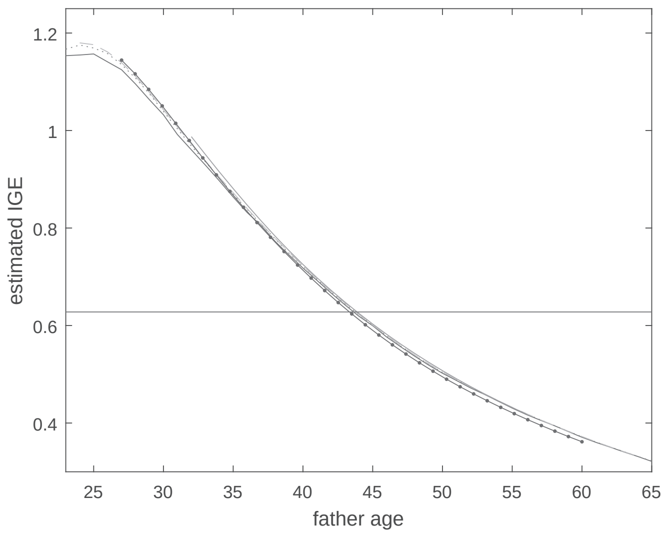
Note: This figure shows half of the IGE estimates using all possible combinations of three-year averages of annual incomes for both generations. The horizontal line indicates the true IGE. Each line shows the IGE estimates over the life cycle of the son, holding constant the age around which father income is observed.

Figure 8: IGE estimates using three-year averages at the same age



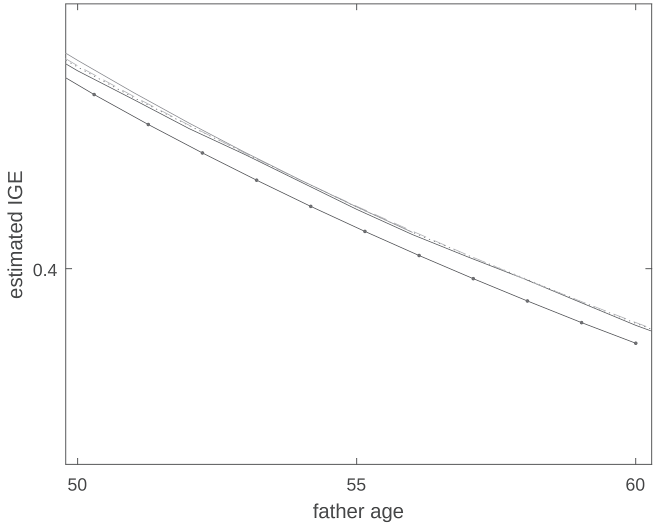
Note: This figure shows the IGE estimates using three-year averages of annual income observations for both generations at the same age. Using the average around age 46 for both father and son results in an unbiased IGE estimate.

Figure 9: Attenuation bias for different parent income averages



Note: This figure shows the bias of the IGE when using the true son average lifetime income, and 3-, 5-, 11-, and 21-year averages of father income over all possible ages.

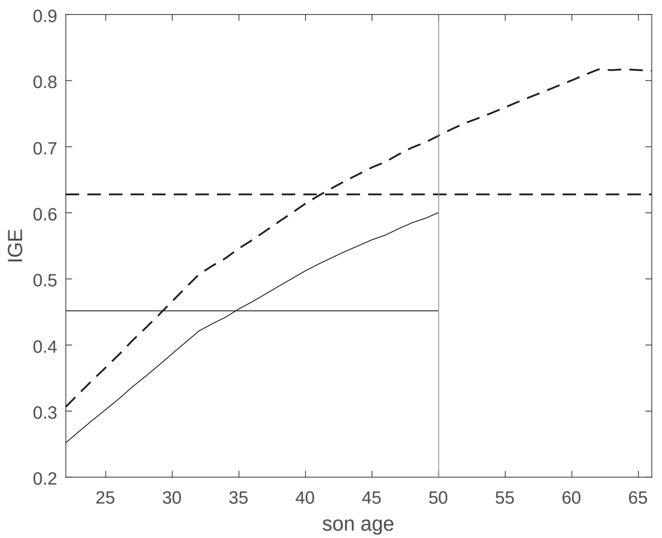
Figure 10: Attenuation bias for different parent income averages (zoom)



Note: This figure shows the bias of the IGE when using the true son average lifetime income, and 3-, 5-, 11-, and 21-year averages of father income over all possible ages.

N&S studied the bias of the IGE claiming "nearly complete income histories of both fathers and sons", having access to Swedish register data covering 29 annual income observations for the sons (age 22 to 50) and 33 observations for the fathers (age 33 to 65). I replicate their analysis of left hand side measurement bias by "ignoring" all income observations in my sample that exceed N&S's data. Figure 11 shows the IGE based on an N&S sample (the solid horizontal reference line, 0.49), as well as 29 IGE estimates using the average of all income observations available to N&S for the father, and one annual income observation for the son, across all ages in the N&S data set (shown as the increasing solid line crossing the solid reference line at around 35). It is clear that lacking income observations during sons' older ages and during fathers' younger ages gives misleading results. The true IGE based on the complete income histories of both generations (shown as a dashed horizontal reference line) is severely underestimated by the nearly-complete data in N&S. Their estimates suggest here using annual income at 35 to approximate son average lifetime income, while income at age 50 or higher is needed to get close to the true

Figure 11: Comparison of measurement error: complete income history vs. nearly complete income history



Note: This figure shows the IGE estimates from regressing all individual annual log income observations over the complete income history of the son on the average log lifetime income of the father for two cases: (i) the nearly complete income history is defined as age 22 to age 50 for the son and age 33 to age 65 for the father, (ii) the complete income history is defined as age 22 to age 66 for both son and father. The vertical line indicates the last year at which son income is observed in case (i).

IGE. In this example, even son income at age 50 will underestimate the IGE. As shown in the figure, there is no combination of observed ages at which the estimate reaches 0.63. Thus, even nearly complete income histories will lead to substantial downward bias of the IGE, as long as we do not observe son income at old ages and parent income at young ages.

Summarizing the results so far, we see that using single log income observations for the son and father around age 45 gives an IGE estimate that coincides with the true IGE (based on the information of the whole life cycle of incomes for two generations). Using annual income observations at similar ages gives in general results that are quite close to the true IGE. This is not really surprising, since income paths across generations within one family are correlated (based on both empirical findings and the assumptions in my data generating process reflecting those findings). Highly educated parents are likely to have a highly educated child, and their income paths over the life cycle will therefore tend to look similar. Thus, in order to estimate just how similar parent and child income trajectories are, looking at the income a son achieved the year he was, for instance, 40 and comparing it with the income his father had at age 40 will already reveal much about the association of *lifetime income* between father and son. The closer the lifetime incomes are linked, the closer the incomes at the same age are usually linked. Although simple, this fact has to my knowledge not been discussed in the literature until now. In addition, I find no evidence that attenuation bias is decreased when averaging over a large number of father income observations as predicted in Solon (1992). Moreover, nearly complete income histories still underestimate the true IGE severely if son (father) income is not observed at older (younger) ages.

3.2.2 Ranks

One relatively new development in the literature of intergenerational income associations has been to use percentile ranks instead of log incomes. The relationship between the ranks of parent and child have, for US data, been shown to be well represented by a linear relationship (Chetty et al., 2014a), much better indeed than logged incomes (the focus has still been on a linear model to estimate the IGE, instead of changing the functional form of the relationship to better match the data). For Swedish data, however, it is not true that the relationship between percentile ranks is necessarily a better fit to a linear model than the relationship between log incomes (Heidrich, 2015).

It is important to point out that the rank-rank slope does not measure the same thing as the IGE. The relationship between income ranks informs us about the association between relative positions in the income distribution only. The income difference between, for instance, percentile ranks 3 and 4 (or 98 and 99) in terms of dollars is presumably much higher than between ranks 50 and 51. Ranks in the tails of the distribution are further apart from each other in dollar terms than ranks in the middle of the distribution, which implies that the marginal effect of climbing or falling one rank depends on the initial position.

The association between ranks is measured by the rank-rank slope β_{rank} in the following equation:

$$R_m^s = \beta_0^{rank} + \beta_{rank} R_m^f + \varepsilon_m^{rank} \tag{12}$$

where R_m^s and R_m^f are the average lifetime income ranks for sons and fathers, respectively. All ranks are normalized to lie between 0 and 100.

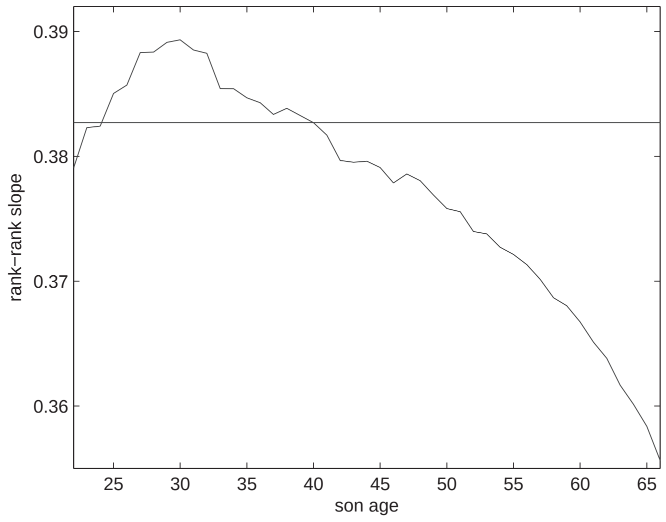
Figure 12 below shows the left hand side measurement error when using the true average lifetime income for the fathers but only single annual income observations for the sons to estimate equation (12). The graph over son ages looks different than Figure 3 using log incomes: the estimates are too large at young son ages and too small at older ages. The bias is however smaller in size compared to the log estimates. The true rank-rank slope in my data (0.38) is estimated when using annual income ranks at ages 24 or 40. The difference could be due to the fact that individuals with higher education start working when they are older, leading to a relatively low income rank at young ages. Income ranks of the largest group of the sample (high school degree) are then overestimated when looking just at the early years.

The right hand side measurement error from varying the age of the father when the income rank is observed, while using the true son average lifetime income rank, is shown in Figure 13. The bias of the estimates follows a similar pattern as the right hand side measurement error using log incomes, but it is again smaller in absolute size compared to logs in Figure 4. Since son income ranks at young ages overestimate the rank-rank slope and father income ranks at older ages underestimate it, the total bias might here actually be reduced when working with typical incomplete income histories.

Figure 14 presents the estimates of the rank-rank slope from all possible combinations of annual income observations, this time expressed in percentile ranks. Interestingly, all estimates are much closer to the true value than when using log incomes (compare to the scale used in Figure 5). Using son income rank at age 30, for instance, the estimated slope coefficient is overestimated by at most 14 percent when father income rank is observed at age 32, and at most 21 percent underestimated when father ranks are observed at age 66. For completeness, Figure 15 shows rank-rank slopes using rank observations at the same age.

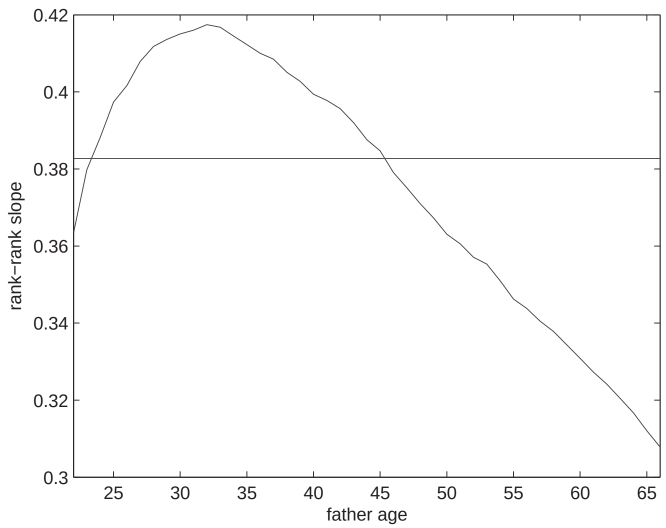
The bias of the rank-rank slope differs in size and form from the bias of the IGE. Annual incomes increase on average over the life cycle, independently if measured in levels or logs. This is why annual income at young ages underestimate the lifetime average income, while annual income at older ages overestimate the true average. The relationship between annual and average lifetime *ranks* is very different. Income ranks are not necessarily increasing in age and on average much less volatile over the life course. Even though income increases for one individual over the life course, it does so for most others as well. The relative positions change therefore much less over time compared to log incomes.

Figure 12: Rank-rank slope estimates over the life cycle of the sons



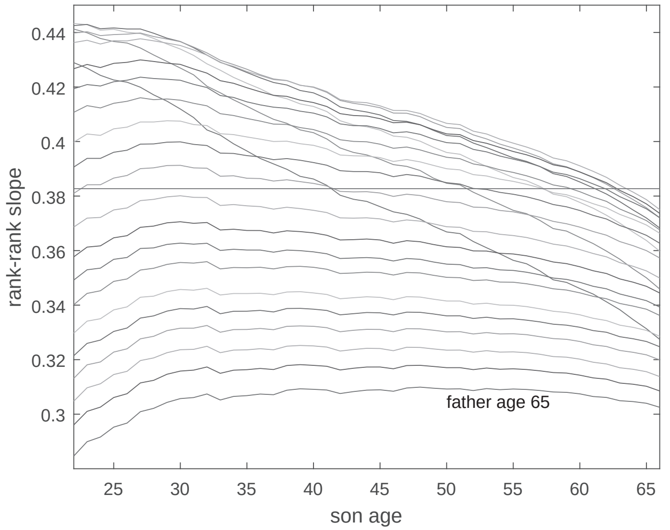
Note: This figure shows the rank-rank slope estimates from regressing 45 annual income rank observations of the son on the average lifetime income rank of the father. The estimates are equal to the true value of the rank-rank slope at around ages 24 and 40.

Figure 13: Rank-rank slope estimates over the life cycle of the fathers



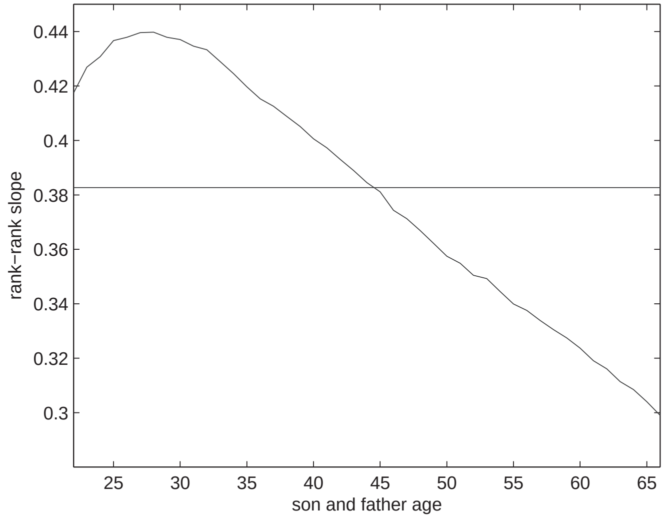
Note: This figure shows the rank-rank slope estimates from regressing the average lifetime income rank of the son on 45 annual income rank observations of the father. The estimates are equal to the true value of the rank-rank slope at around ages 24 and 45.

Figure 14: Rank-rank slope estimates using single income observations



Note: This figure shows half of the rank-rank slope estimates using all possible combinations of single annual income rank observations for both generations. The horizontal line indicates the size of the true rank-rank slope. Each line shows the slope estimates over the life cycle of the son, holding constant the age at which father income rank is observed.

Figure 15: Rank-rank slope estimates using single income observations at the same age



Note: This figure shows the IGE estimates using a single annual income rank observation for both generations at the same age. Using the income rank at age 45 for both father and son results in an unbiased IGE.

3.3 Reducing the bias in applied work

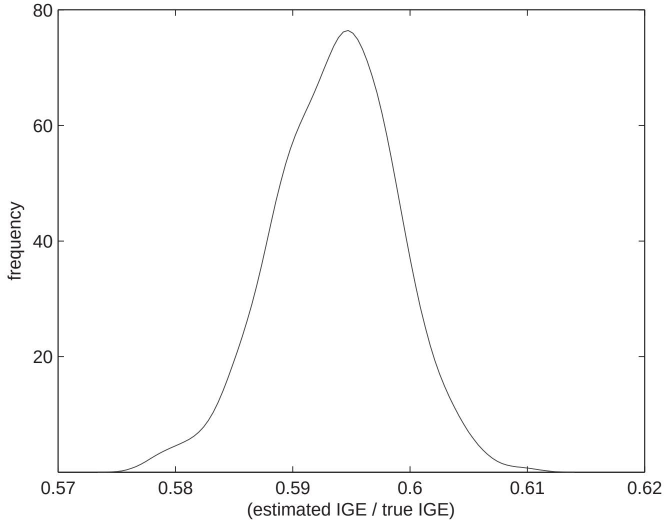
Next, I turn to the typical data situation a researcher faces. The base for most studies are (earned) income observations from either survey data or from register data, usually based on individuals' annual income tax reports. If the data is available for a short time span only and we want to study the income of two linked generations, this implies directly that we cannot observe the income of the two generations at the same ages. Chetty et al. (2014a) and Chetty et al. (2014b) for instance use federal income tax records spanning from 1996 to 2012 in their papers. Their core sample consists of children born between 1980 and 1982 whose parents are between 15 and 40 years old when their children are born. Child income is defined as the average income in 2011 and 2012, when children are between 29 and 32 years old. Parent income is defined as the average of five income observations between 1996 - 2000. Since the parents' age when their child is born differs greatly over the sample, there is also a great span of parent ages between 1996 and 2000 when parent income is measured. That is, parent income is measured for some families when parents are in their early thirties, while others are already in their late sixties.

I simulate this typical data situation by assuming that parents are on average age 50 when income is observed, with symmetric and decreasing probabilities to be younger or older than 50 (a discrete distribution with a bell-shaped histogram centered at 50). Father income is calculated as the five-year average around the given age, while son income is the average of incomes observed at age 31 and 32.

As shown in Figure 16, using this typical type of data and income measured in logs leads to OLS estimates that are on average just 59 percent of the size of the true IGE (based on 1,000 simulations of the income data and computation of the true as well as the "missing-data" IGE). The IGE estimates reported by studies measuring income in this way can therefore be assumed to seriously underestimate the IGE (and thus overestimate intergenerational mobility). According to the analysis above, the IGE will be least biased when using similar ages for sons and fathers. The easiest way to follow this rule-of-thumb when faced with the typical data situation as in Chetty et al. (2014a), is to use the latest income observation available for the son and the earliest income observation available for the father.

In addition, since the important information is hidden in the later years that are observed for the son and in the earlier years that are observed for the father, I also test using a weighted average of the available income observations, instead of the normal mean used in the literature. More specifically, I weigh the income at age 31 for the son with 0.25 and the income at age 32 with 0.75, and the father's five income observations in a decreasing fashion using the five weights 0.5, 0.25, 0.13, 0.07, 0.05 (the numbers are arbitrarily chosen and the results are similar as long as the weights increase and decrease in age, respectively).

Figure 16: Monte Carlo simulation of typical data usage (log)

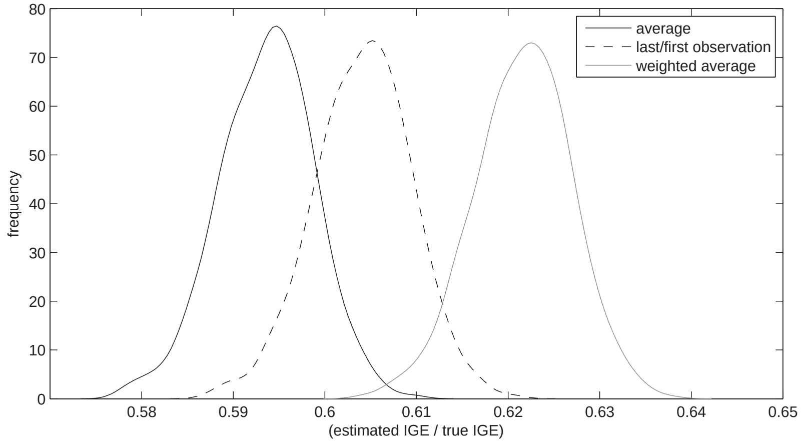


Note: This figure shows the distribution of the fraction of the true IGE that is estimated when using averaged log income at mixed parent ages around 50, and at son ages 31 and 32, to approximate life time income. The mean is 0.59, i.e. the estimated IGE is on average only 59 percent of the true IGE and thus 46 percent too small. The number of simulations is 1,000.

The results are collected in Figure 17. The black, solid line reproduces Figure 16 to facilitate comparison. The dashed line indicates the distribution of IGE estimates from 1,000 simulations using only the last observation for the son and the earliest observation for the father. The mean of those estimates is 0.6. The gray line shows the results based on the weighted averages. The average IGE is now on average 0.62 and thus 62 percent of the true IGE.

Using a two-sample Kolmogorov-Smirnov test (Massey, 1951), I can reject the hypotheses that the results based on the three different methods (simple averages, one observation each, and weighted averages) are from populations with the same distribution on the 0.1 percent significance level. I also test the equality of the means of the three distributions using two-sample t-tests, assuming both equal and unequal variance (Welch's t-test,Welch (1947)). The hypotheses that the estimates are from normal distributions with the same mean are all rejected on the 0.1 percent significance level. Therefore, I conclude that both using one observation each and using weighted averages are preferred to the simple mean of all observed incomes, with weighted averages performing best.

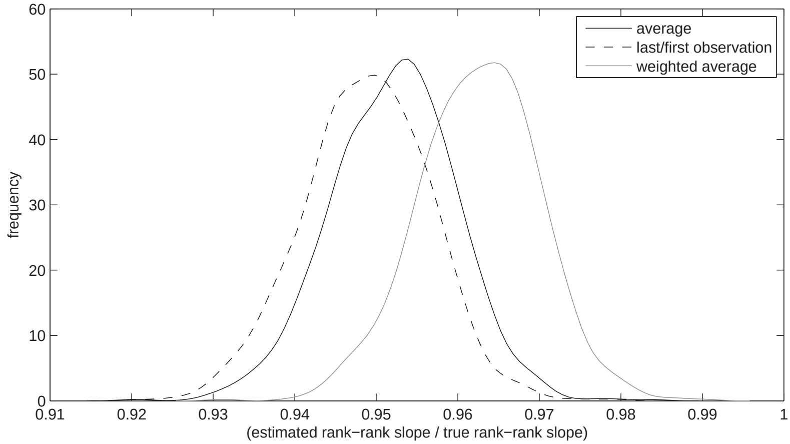
Figure 17: Monte Carlo simulation of typical data: three ways to use the data (log)



Note: This figure shows the distribution of the fraction of the true IGE that is estimated when using one of the following methods to approximate average lifetime income: (i) the average over all available observations, (ii) the latest available income observation for the son and the earliest available income observation for the father, (iii) a weighted average of all available income observations, with increasing weights in age for the son and decreasing weights in age for the father. The estimated means are (i) 0.59, (ii) 0.60, and (iii) 0.62. The number of simulations is 1,000 each.

The results based on ranks are shown in Figure 18. Both the simple mean and the first-last observation method give rank-rank slope estimates that are on average 95 percent of the true value. Using a weighted average results in estimates that are on average 1 percentage point less biased than using the other two methods. The difference is statistically significant on the 0.1 percent level. Note that the spread of the estimates is larger in the rank case compared to using log incomes. Still, even the estimates in the lower tails have no more than a 10 percent bias. As shown in Section 3.2.2, the measurement errors due to the young age at which son income is observed, and the old age at which father income is observed, go in opposite directions which benefits the estimates. In addition, the overall bias is relatively small due to less variation in ranks over the life cycle. Percentile ranks might therefore be a good choice of income measure not only due to functional form considerations or the carefree handling of zero-income observations, but because it results in estimates of a rank-rank slope that are very close to the true value even when large parts of the income histories are missing.

Figure 18: Monte Carlo simulation of typical data: three ways to use the data (rank)



Note: This figure shows the distribution of the fraction of the true rank-rank slope that is estimated when using one of the following methods to approximate average lifetime income rank: (i) the average rank over all available observations, (ii) the latest available income rank observation for the son and the earliest available income rank observation for the father, (iii) a weighted average of all available income rank observations, with increasing weights in age for the son and decreasing weights in age for the father. The estimated means are (i) 0.95, (ii) 0.95, and (iii) 0.96. The number of simulations is 1,000 each.

4 Concluding remarks

Estimates of intergenerational mobility are highly sensitive to the available income data and the chosen income measure. In most studies, the choice of the income measure is heavily constrained by the available data. Most often, the averages of all available observations of log annual income for the two generations have been used as a proxy for lifetime income. Sometimes, parent income is defined as the average over just the earliest years available. Chetty et al. (2014a) argue that this is done in order to reflect the income level that parents had during the time they were raising their child. However, this argument is misleading since one is usually not primarily interested in the relationship between parent income during child-rearing years, and child lifetime income. If we want to measure the association between parent and child lifetime income, the focus has to be on how to best approximate this lifetime income for each generation.

My contribution is to systematically analyze, by the means of simulation methods, when and how to measure income in order to produce estimates of the intergenerational elasticity that are as close to the true value as possible. I have employed a generated sample of income observations over two generations' life cycles that match the typical characteristics of data used in earlier register based studies. Missing income observations for sons at older ages and for fathers at younger ages give heavily downward biased IGE estimates, even when around thirty income observations for each generation are available. The rank-rank slope estimates are only very little biased when incomplete data is used. This is probably due to the fact that an individual's ranks are more stable over her lifetime compared to income levels.

In this paper I have shown that, in principle, there exist many different age combinations at which the IGE will be unbiased. In general it holds true that the closer the age of the son and the father when their respective income is observed, the closer is the IGE to its true value. This is due the fact that the income path of the son is correlated with family background variables. The association between, for instance, son income at age 40 and father income at age 40, already carries a lot of information about the degree of income mobility in a society.

In addition, I have studied the bias of the IGE for typical data availability and usage found in the literature. The most common procedure has been to use the average over all available log income observations for both the father and the son. I have shown that this approach leads to heavily downward biased estimates of the IGE. On average, those estimates are around 40 percent too small. There is no evidence that using the average over as many as income observations as possible for the father is reducing attenuation bias. In contrast, using only one income observation at the latest observed age for the son, and one income observation at the earliest possible age for the father, gives IGE estimates that are statistically significantly less biased than the ordinary average over all available years. A weighted average is the preferred choice, even though the IGE is still severely under estimated.

Focusing on income ranks, I have shown that the estimates of the rank-rank slope based on the simple mean are on average just 5 percent too small. Using only the latest available income observation for the son and the earliest available income observation for the father is found to produce equally good results. A weighted average of the available income observations is the preferred method, resulting in estimates of the IGE with a measurement error of just 4 percent.

Researchers attempting to estimate the IGE based on incomplete income histories (even based on nearly complete income histories) should therefore focus on the available observations for son and father that are as close as possible to each other in terms of age. The crucial point is that one does not need as many income observations as possible in order to estimate the true IGE, but one observation at the right age. Estimates of the rank-rank slope are much less sensitive to missing income observations.

References

Abel, J. R., and Deitz, R. (2014). "Do the benefits of college still outweigh the costs?" In *Current Issues in Economics and Finance*, vol. 20, Federal Reserve Bank of New York.

Abowd, J. M., and Card, D. (1989). "On the covariance structure of earnings and hours changes." *Econometrica*, *57*(2), 411–445.

Atkinson, A. (1980). "On intergenerational income mobility in Britain." *Journal of Post Keynesian Economics*, *3*(2), 194.

Atkinson, A. B., Bourguignon, F., and Morrisson, C. (1992). *Empirical studies of earnings mobility*, vol. 52. Taylor & Francis.

Becker, G. S., and Tomes, N. (1979). "An equilibrium theory of the distribution of income and intergenerational mobility." *Journal of Political Economy*, *87*(6), pp. 1153–1189.

Becker, G. S., and Tomes, N. (1986). "Human capital and the rise and fall of families." *Journal of Labor Economics*, *4*(3), pp. S1–S39.

Böhlmark, A., and Lindquist, M. J. (2006). "Life-Cycle Variations in the Association between Current and Lifetime Income: Replication and Extension for Sweden." *Journal of Labor Economics*, *24*(4), pp. 879–896.

Björklund, A., and Jäntti, M. (2009). "Intergenerational income mobility and the role of family background." *Oxford Handbook of Economic Inequality, Oxford University Press, Oxford*, 491–521.

Björklund, A., and Jäntti, M. (1997). "Intergenerational income mobility in Sweden compared to the United States." *The American Economic Review*, *87*(5), pp. 1009–1018.

Black, S. E., and Devereux, P. J. (2011). "Recent developments in intergenerational mobility." In *Handbook of Labor Economics*, vol. 4, Part B, chap. 16, 1487 – 1541, Elsevier.

Browning, M., Hansen, L. P., and Heckman, J. J. (1999). "Micro data and general equilibrium models." *Handbook of macroeconomics*, *1*, 543–633.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014a). "Where is the land of opportunity? The geography of intergenerational mobility in the United States." *The Quarterly Journal of Economics*, *129*(4), i1.

Chetty, R., Hendren, N., Kline, P., Saez, E., and Turner, N. (2014b). "Is the United States still a land of opportunity? Recent trends in intergenerational mobility." *American Economic Review*, *104*(5), 141–47.

Conesa, J. C., Kitao, S., and Krueger, D. (2009). "Taxing capital? Not a bad idea after all!" *American Economic Review*, *99*(1), 25–48.

De Nardi, M. (2004). "Wealth inequality and intergenerational links." *The Review of Economic Studies*, *71*(3), 743–768.

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

Guvenen, F. (2009). "An empirical investigation of labor income processes." *Review of Economic Dynamics*, *12*(1), 58–79.

Haider, S. J., and Solon, G. (2006). "Life-cycle variation in the association between current and lifetime earnings." *American Economic Review*, *96*(4), 1308–1320.

Hansen, G. D. (1993). "The cyclical and secular behaviour of the labour input: Comparing efficiency units and hours worked." *Journal of Applied Econometrics*, *8*(1), 71–80.

Heidrich, S. (2015). "Intergenerational mobility in Sweden: A regional perspective." *Umeå Economic Studies*, *916*.

Huggett, M., and Ventura, G. (1999). "On the distributional effects of social security reform." *Review of Economic Dynamics*, *2*(3), 498–531.

Karahan, F., and Ozkan, S. (2013). "On the persistence of income shocks over the life cycle: Evidence, theory, and implications." *Review of Economic Dynamics*, *16*(3), 452–476.

Lillard, L. A., and Weiss, Y. (1979). "Components of variation in panel earnings data: American scientists 1960-70." *Econometrica: Journal of the Econometric Society*, 437–454.

MaCurdy, T. E. (1982). "The use of time series processes to model the error structure of earnings in a longitudinal data analysis." *Journal of Econometrics*, *18*(1), 83–114.

Massey, F. J. (1951). "The Kolmogorov-Smirnov Test for goodness of fit." *Journal of the American Statistical Association*, *46*(253), 68–78.

Nybom, M., and Stuhler, J. (2016). "Heterogeneous income profiles and lifecycle bias in intergenerational mobility estimation." *Journal of Human Resources*, *51*(1), 239–268.

Solon, G. (1992). "Intergenerational income mobility in the United States." *The American Economic Review*, 393–408.

Soltow, L. (1965). *Toward income equality in Norway*. Madison, U. of Wisconsin P.

Storesletten, K., Telmer, C. I., and Yaron, A. (2004). "Consumption and risk sharing over the life cycle." *Journal of monetary Economics*, *51*(3), 609–633.

U.S. Census Bureau (2016). "College completion by cohort, age and gender, 1967 to 2015." Working Paper 2016-04, Social, Housing and Economic Statistics Division, Washington, DC.

Welch, B. L. (1947). "The generalization of 'Student's' problem when several different population variances are involved." *Biometrika*, *34*(1/2), 28–35.

Appendix

A Tables

Table A.1: Parameter values used in my simulations

Parameter	
ρ^{e1}	0.829
ρ^{e2}	0.805
$\mu^{\alpha(e1)}$	0.0
$\mu^{\alpha(e2)}$	0.2
$\mu^{\beta(e1)}$	0.0
$\mu^{\beta(e1)}$	0.01
$\sigma^2_{\alpha(e1)}$	0.038
$\sigma^2_{\alpha(e2)}$	0.023
$\sigma^2_{\beta(e1)}$	0.0002
$\sigma^2_{\beta(e2)}$	0.00049
$\text{Corr}\left(\alpha_f, \beta_f\right)^{e1}$	-0.25
$\text{Corr}\left(\alpha_f, \beta_f\right)^{e2}$	-0.7
$\text{Corr}\left(\alpha_f, \alpha_s\right)$	0.15
$\text{Corr}\left(\beta_f, \beta_s\right)$	0.15
$\sigma^2_{\varepsilon(e1)}$	0.034
$\sigma^2_{\varepsilon(e2)}$	0.025
$\sigma^2_{\eta(e1)}$	0.022
$\sigma^2_{\eta(e2)}$	0.025

Note: $e1$ and $e2$ denote the two education levels high school *and* college. *The correlation between the α's and β's between the generations is not taken from Guvenen (2009) but was chose by me to generate the correlation between father and son income path.*

List of Dissertations at the Department of Economics, Umeå University

Holmström, Leif (1972) Teorin för företagens lokaliseringsval. UES 1. PhLic thesis

Löfgren, Karl-Gustaf (1972) Studier i teorin för prisdiskriminering. UES 2. PhLic thesis

Dahlberg, Åke (1972) Arbetsmarknadsutbildning - verkningar för den enskilde och samhället. UES 3. PhD thesis

Stage, Jørn (1973) Verklighetsuppfattning och ekonomisk teori. UES 4. PhLic thesis

Holmlund, Bertil (1976) Arbetslöshet och lönebildning - kvantitativa studier av svensk arbetsmarknad. UES 25. PhD thesis

Löfgren, Karl-Gustaf (1977) En studie i neokeynesiansk arbetslöshets- och inflationsteori. UES 34. PhD thesis

Lundberg, Lars (1976) Handelshinder och handelspolitik - Studier av verkningar på svensk ekonomi. Industrins Utredningsinstitut, Stockholm. PhD thesis

Johansson, Per-Olof (1978) Sysselsättning och samhällsekonomi - En studie av Algots etablering i Västerbotten. UES 53. PhD thesis

Wibe, Sören (1980) Teknik och aggregering i produktionsteorin. Svensk järnhantering 1850-1975; en branschanalys. UES 63. PhD thesis

Ivarson, Lars (1980) Bankers portföljvalsbeteende. En teoretisk studie. UES 64. PhD thesis

Batten, David (1981) Entropy, Information Theory and Spatial Input-output Analysis. UES 92. PhD thesis

Hårsman, Björn (1982) Housing Demand Models and Housing Market Models for Regional and Local Planning. Swedish Council for Building Research, D13:1981. PhD thesis

Holm, Magnus (1983) Regionalekonomiska modeller för planering och samordning i en decentraliserad ekonomi. Byggforskningsrådet, R118:1981 and R5:1983. PhD thesis

Ohlsson, Henry (1986) Cost-Benefit Analysis of Labor Market Programs - Applied to a Temporary Program in Northern Sweden. UES 167. PhLic thesis

Sarafoglou, Nikias (1987) A Contribution to Population Dynamics in Space. UES 179. PhD thesis

Ohlsson, Henry (1988) Cost-Benefit Analysis of Labor Market Programs - Applied to a Temporary Program in Northern Sweden. UES 182. PhD thesis

Anderstig, Christer (1988) Applied Methods for Analysis of Economic Structure and Change. CERUM 1988:2, Umeå University. PhD thesis

Karlsson, Charlie (1988) Innovation Adoption and a Product Life Cycle. UES 185. PhD thesis

Löfström, Åsa (1989) Diskriminering på svensk arbetsmarknad - En analys av löneskillnader mellan kvinnor och män. UES 196. PhD thesis

Axelsson, Roger (1989) Svensk arbetsmarknadsutbildning - En kvantitativ analys av dess effekter. UES 197. PhD thesis

Zhang, Wei-Bin (1989) Theory of Economic Development - Nonlinearity, Instability and Non-equilibrium. UES 198. PhD thesis

Hansson, Pär (1989) Intra-Industry Trade: Measurements, Determinants and Growth - A study of Swedish Foreign Trade. UES 205. PhD thesis

Kriström, Bengt (1990) Valuing Environmental Benefits Using the Contingent Valuation Method: An Econometric Analysis. UES 219. PhD thesis

Aronsson, Thomas (1990) The Short-Run Supply of Roundwood under Nonlinear Income Taxation - Theory, Estimation Methods and Empirical Results Based on Swedish Data. UES 220. PhD thesis

Westin, Lars (1990) Vintage Models of Spatial Structural Change. UES 227. PhD thesis

Wikström, Magnus (1992) Four Papers on Wage Formation in a Unionized Economy. UES 287. PhD thesis

Westerlund, Olle (1993) Internal Migration in Sweden - The Role of Fiscal Variables and Labor Market Conditions. UES 293. PhLic thesis

Bergman, Mats A. (1993) Market Structure and Market Power. The Case of the Swedish Forest Sector. UES 296. PhD thesis

Johansson, Per (1993) Count Data Models - Estimator Performance and Applications. UES 315. PhD thesis

Roson, Roberto (1994) Transport Networks and the Spatial Economy - A General Equilibrium Analysis. UES 340. PhD thesis

Li, Chuan-Zhong (1994) Welfare Evaluations in Contingent Valuation - An Econometric Analysis. UES 341. PhD thesis

Østbye, Stein (1994) Regional Labour and Capital Subsidies - Theory and Evidence of the Impact on Employment under Wage Bargaining. UES 344. PhLic thesis

Westerlund, Olle (1995) Economic Influences on Migration in Sweden. UES 379. PhD thesis

Mortazavi, Reza (1995) Three Papers on the Economics of Recreation, Tourism and Property Rights. UES 396. PhLic thesis

Østbye, Stein (1995) Regional Labour and Capital Subsidies. UES 397. PhD thesis

Hussain-Shahid, Imdad (1996) Benefits of Transport Infrastructure Investments: A Spatial Computable General Equilibrium Approach. UES 409. PhD thesis

Eriksson, Maria (1996) Selektion till arbetsmarknadsutbildning. UES 410. PhLic thesis

Karlsson, Niklas (1996) Testing and Estimation in Labour Supply and Duration Models. UES 413. PhD thesis

Olsson, Christina (1996) Chernobyl Effects and Dental Insurance. UES 428. PhLic thesis

Vredin, Maria (1997) The African Elephant - Existence Value and Determinants of Willingness to Pay. UES 441. PhLic thesis

Eriksson, Maria (1997) To Choose or not to Choose: Choice and Choice Set Models. UES 443. PhD thesis

Widerstedt, Barbro (1997) Employer Change and Migration. Two Papers on Labour Mobility in Sweden. UES 444. PhLic thesis

Lundberg, Sofia (1997) The Economics of Child Auctions in 19th Century Sweden. UES 445. PhLic thesis

Lundberg, Johan (1997) Two Papers on Revenue and Expenditure Decisions in the Swedish Local Public Sector. UES 454. PhLic thesis

Widerstedt, Barbro (1998) Moving or Staying? Job Mobility as a Sorting Process. UES 464. PhD thesis

Bask, Mikael (1998) Essays on Exchange Rates: Deterministic Chaos and Technical Analysis. UES 465. PhD thesis

Löfgren, Curt (1998) Time to Study Students: Two Essays on Student Achievement and Study Effort. UES 466. PhLic thesis

Sjögren, Tomas (1998) Union Wage Setting in a Dynamic Economy. UES 480. PhD thesis

Mortazavi, Reza (1999) Essays on Economic Problems in Recreation, Tourism and Transportation. UES 483. PhD thesis

Rudholm, Niklas (1999) Two Essays on Generic Competition in the Swedish Pharmaceuticals Market. UES 485. PhLic thesis

Olsson, Christina (1999) Essays in the Economics of Dental Insurance and Dental Health. UES 494. PhD thesis

Marklund, Per-Olov (1999) Environmental Regulation and Firm Efficiency. UES 504. PhLic thesis

Berglund, Elisabet (1999) Regional Entry and Exit of Firms. UES 506. PhD thesis

Hellström, Jörgen (1999) Count Data Autoregression Modelling. UES 507. PhLic thesis

Nordström, Jonas (1999) Tourism and Travel: Accounts, Demand and Forecasts. UES 509. PhD thesis

Johansson Vredin, Maria (1999) Economics Without Markets. Four papers on the Contingent Valuation and Stated Preference Methods. UES 517. PhD thesis

Schei, Torbjørn (2000) Natural recreation resources: production and a diversity of interests related to the management of grouse as an outfield resource in Finnmark, Norway, in the Euro-Arctic Barents region. UES 523. PhLic thesis

Backlund, Kenneth (2000) Welfare Measurement, Externalities and Pigouvian Taxation in Dynamic Economies. UES 527. PhD thesis

Andersson, Linda (2000) Job Turnover, Productivity and International Trade. UES 530. PhLic thesis

Ylvinger, Svante (2000) Essays on Production Performance Assessment. UES 531. PhD thesis

Bergkvist, Erik (2001) Freight Transportation. Valuation of Time and Forecasting of Flows. UES 549. PhD thesis

Rudholm, Niklas (2001) The Swedish Pharmaceuticals Market - Essays on Entry, Competition and Antibiotic Resistance. UES 552. PhD thesis

Lundberg, Johan (2001) Local Government Expenditures and Regional Growth in Sweden. UES 554. PhD thesis

Lundberg, Sofia (2001) Going Once, Going Twice, SOLD! The Economics of Past and Present Public Procurement in Sweden. UES 557. PhD thesis

Eliasson, Kent (2001) University Enrollment and Geographical Mobility: The Case of Sweden. UES 558. PhLic thesis

Samakovlis, Eva (2001) Economics of Paper Recycling. Efficiency, policies, and substitution possibilities. UES 563. PhD thesis

Daunfeldt, Sven-Olov (2001) Essays on Intra-Household Allocation and Policy Regime Shifts. UES 570. PhD thesis

Hellström, Jörgen (2002) Count Data Modelling and Tourism Demand. UES 584. PhD thesis

Andersson, Linda (2002) Essays on Job Turnover, Productivity and State-Local Finance. UES 586. PhD thesis

Rashid, Saman (2002) Invandrarinkomster, förvärvsdeltagande och familj. UES 588. PhLic thesis

Hanes, Niklas (2003) Empirical Studies in Local Public Finance: Spillovers, Amalgamations, and Tactical Redistributions. UES 604. PhD thesis

Stenberg, Anders (2003) An Evaluation of the Adult Education Initiative Relative Labor Market Training. UES 609. PhD thesis

Stage, Jesper (2003) Mixing Oil and Water. Studies of the Namibian Economy. UES 611. PhD thesis

Marklund, Per-Olov (2004) Essays on Productive Efficiency, Shadow Prices, and Human Capital. UES 621. PhD thesis

Rashid, Saman (2004) Immigrants' Income and Family Migration. UES 625. PhD thesis

Sandberg, Krister (2004) Hedonic Prices, Economic Growth, and Spatial Dependence. UES 631. PhD thesis

Sjöström, Magnus (2004) Factor Demand and Market Power. UES 633. PhD thesis

Nilsson, William (2005) Equality of Opportunity, Heterogeneity and Poverty. UES 652. PhD thesis

Quoreshi, Shahiduzzaman (2005) Modelling High Frequency Financial Count Data. UES 656. Ph Lic thesis

Ankarhem, Mattias (2005) Bioenergy, Pollution, and Economic Growth. UES 661. PhD thesis

Quoreshi, Shahiduzzaman (2006) Time Series Modelling of High Frequency Stock Transaction Data. UES 675. PhD thesis

Ghalwash, Tarek (2006) Income, Energy Taxation, and the Environment. An Econometric Analysis. UES 678. PhD thesis

Westerberg, Thomas (2006) Two Papers on Fertility – The Case of Sweden. UES 683. Ph Lic thesis

Simonsen, Ola (2006) Stock Data, Trade Durations, and Limit Order Book Information. UES 689. PhD thesis

Eliasson, Kent (2006) College Choice and Earnings among University Graduates in Sweden. UES 693. PhD thesis

Selander, Carina (2006) Chartist Trading in Exchange Rate Theory. UES 698. PhD thesis

Humavindu, Michael N (2007) Essays on Public Finance and Environmental Economics in Namibia. UES 705. Ph Lic thesis

Norberg-Schönfeldt, Magdalena (2007) The Phase-Out of the Nuclear Family? Empirical Studies on the Economics and Structure of Modern Swedish Families. UES 708. PhD thesis

Granlund, David (2007) Economic Policy in Health Care: Sickness Absence and Pharmaceutical Costs. UES 710. PhD thesis

Jonsson, Thomas (2007) Essays on Agricultural and Environmental Policy. UES 719. PhD thesis

Broberg, Thomas (2007) The Value of Preserving Nature – Preference Uncertainty and Distributional Effects. UES 720. PhD thesis

Witterblad, Mikael (2008) Essays on Redistribution and Local Public Expenditures. UES 731. PhD thesis

Thunström, Linda (2008) Food Consumption, Paternalism and Economic Policy. UES 739. PhD thesis

Humavindu, Michael N (2008) Essays on the Namibian Economy. UES 745. PhD thesis

Persson, Lars (2008) Environmental Policy and Transboundary Externalities - Coordination and Commitment in Open Economies. UES 755. PhD thesis

Sahlén, Linda (2008) Essays on Environmental and Development Economics - Public Policy, Resource Prices and Global Warming. UES 762. PhD thesis

Lönnbark, Carl (2009) On Risk Prediction. UES 770. PhD thesis Norin, Anna (2009) Worker Safety and Market Dynamics. UES 772. PhLic thesis

Holmlund, Linda (2009) Essays on Child Care and Higher Education. UES 783. PhD thesis

Landström, Mats (2009) Two essays on Central Bank Independence Reforms. UES 792. PhLic thesis

Åström, Johanna (2009) Marriage, Money and Migration. UES 790. PhD thesis

Birkelöf, Lena (2009) Spatial Interaction and Local Government Expenditures for Functionally Impaired in Sweden. UES 798. PhD thesis

Mannberg, Andrea (2010) Risk and Rationality – Effects of contextual risk and cognitive dissonance on (sexual) incentives. UES 806. PhD thesis

Andersson, Camilla (2010) Changing the Risk at the Margin: Smallholder Farming and Public Policy in Developing Countries. UES 810. PhD thesis

Sörensson, Robert (2010) Marshallian Sources of Growth and Interdependent Location of Swedish Firms and Households. UES 815. PhD thesis

Soultanaeva, Albina (2011) Back on the Map - Essays on the Financial Markets in the three Baltic States. UES 820. PhD thesis

Holmberg, Ulf (2012) Essays on Credit Markets and Banking. UES 840. PhD thesis

Lundgren, Jens (2012) Market Liberalization and Market Integration - Essays on the Nordic Electricity Market. UES 852. PhD thesis

Karimu, Amin (2013) Essays on Energy Demand and Household Energy Choice. UES 864. PhD Thesis

Landström, Mats (2013) Determinants and Effects of Central Bank Independence Reforms. UES 869. PhD Thesis

Cialani, Catia (2014) Essays on Growth and Environment. UES 875. PhD thesis

Tano, Sofia (2014) Migration and Regional Sorting of Skills. UES 882. PhD Thesis

Hakobyan, Lilit (2014) Essays on Growth and Political Transition. UES 893. PhD Thesis

Brockwell, Erik (2014) State and Industrial Actions to Influence Consumer Behavior. UES 894. PhD Thesis

Westéus, Morgan (2014) Essays on Temporary Work Agencies and the Economic Analysis of Law. UES 897. PhD Thesis

Zetterdahl, Emma (2015) Take a Risk –Social Interaction, Gender Identity, and the Role of Family Ties in Financial Decision-Making. UES 908. PhD Thesis

Strömbäck, Elon (2015) Policy by Public Procurement: Opportunities and Pitfalls. UES 915. PhD Thesis

Eriksson, Mathilda (2016) The Role of the Forest in Climate Policy. UES 927. PhD Thesis

Liu, Yuna (2016) Essays on Stock Market Integration – On Stock Market Efficiency, Price Jumps and Stock Market Correlations. UES 926. PhD Thesis