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: 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, 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 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).

[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.

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

$\alpha^{i,e} \sim N\left(0, \sigma^2_{\alpha,e}\right),\ \beta^{i,e} \sim N\left(0, \sigma^2_{\beta,e}\right)$, and $Cov\left(\alpha^e_i, \beta^e_i\right) = \sigma_{\alpha,\beta}$.

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

$$z^{i,e}_{h,t} = \rho^e z^i_{h,t-1} + \eta^i_{h,t}, \quad z^i_{0,t} = 0\ . \tag{4}$$

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.

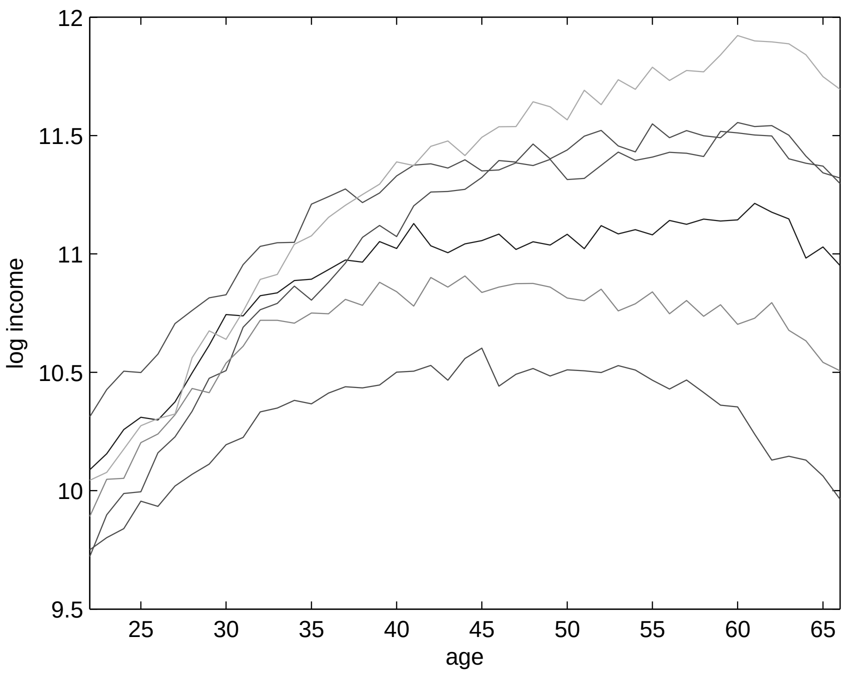
Figure 1 shows the earnings over they life cycle for 10 fathers in my sample. We see the typical hump-shape over the years, with earnings increasing quickly at young ages in order to level off at around mid age, and finally decreasing when hitting retirement. Importantly, the ages in this study are not to be interpret literally. The age span chosen here to define life time earnings could be changed to include even younger and older ages.

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

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 $m = 1,...,N$ ($\bar{y}^s_m$) 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

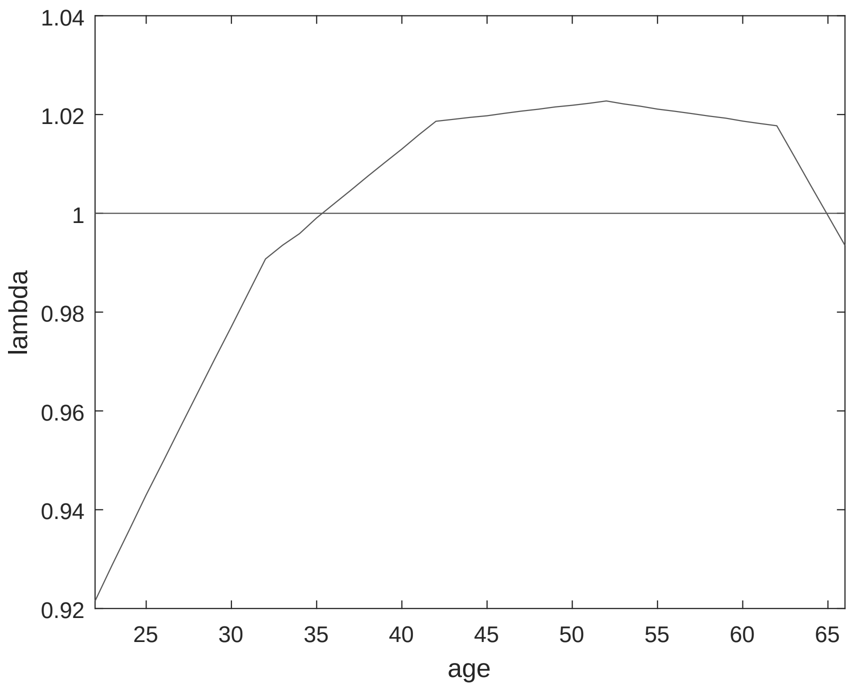
The data I generated with the model in Section 2 shows the typical relationship between annual income and average lifetime income. Annual income at younger ages lies below the lifetime average, while annual income at older ages lies above the lifetime average. Haider and Solon (2006) (H&S) suggested 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.

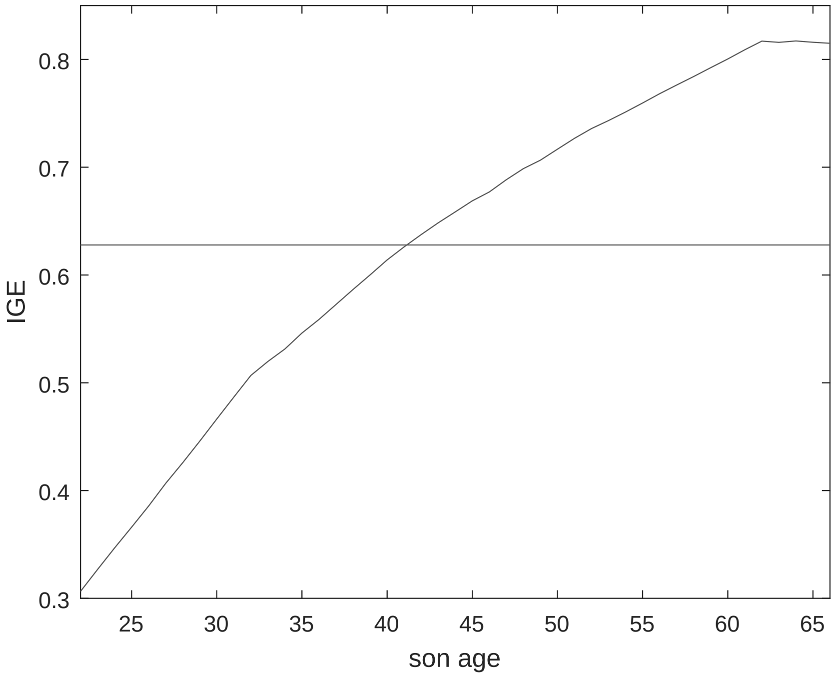
The bias of the IGE that is due to using an incorrect proxy for son average lifetime income is in the literature typically referred to as left-hand side measurement error. Using single annual income observations as a proxy for average lifetime income will lead to estimates of the IGE that are negatively 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.

Figure 2: Estimates of lambda over the sons' life cycle



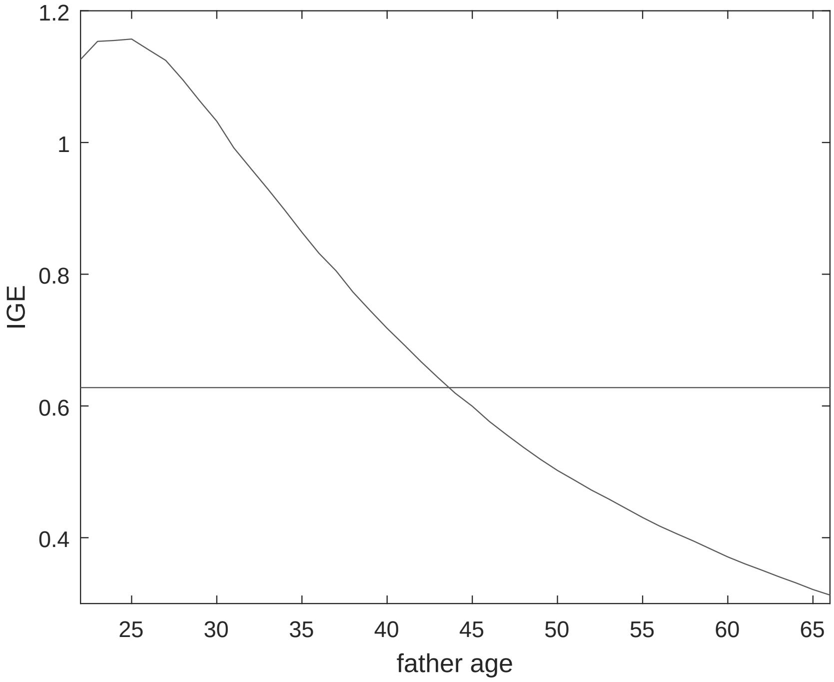
Note: This figure shows the slope coefficients from regressing 45 annual income observations of the sons on their average lifetime income. At age 35 and age 65, annual incomes are on average closest to average lifetime income (where the slope coefficient equals one).

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.

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.

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 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 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.

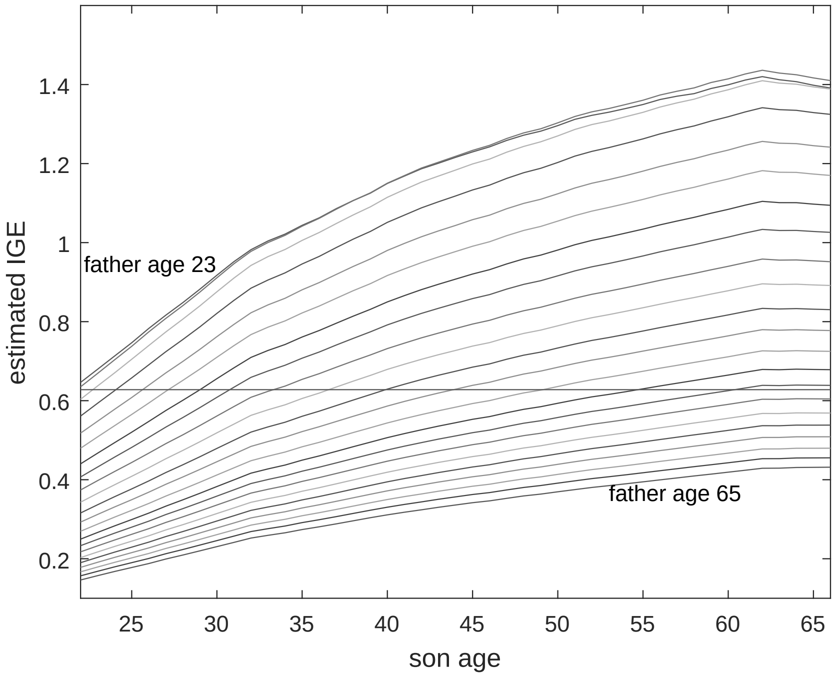
3.2 The effect of missing data on the bias of the IGE estimator

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

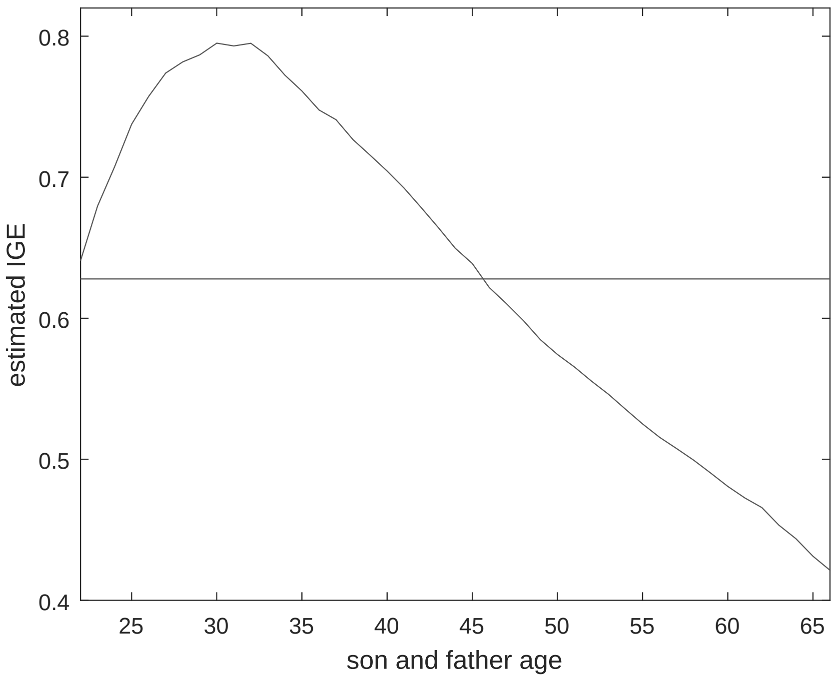
I start by demonstrating the effect of using just one single income observation for each generation as a proxy for lifetime income, based on my simulated father and son sample with complete income histories. Figure 5 shows the estimated IGE for all possible combinations of father and son individual income 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

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.

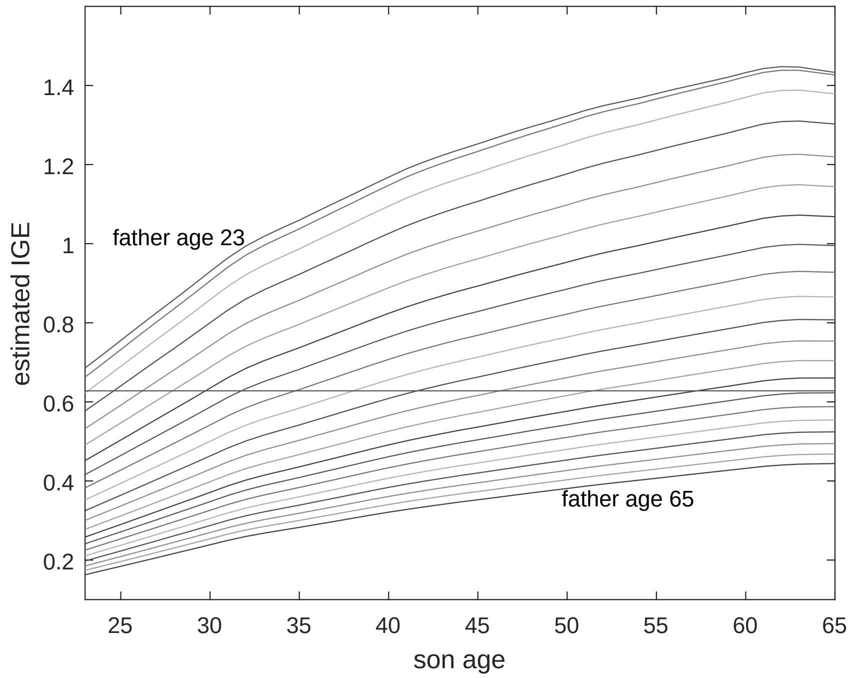
estimated IGEs over the full range of son ages, while holding father age constant at 65. The figure shows that, in principle, there are many different combinations of ages to observe income for the two generations that will give the correct IGE (wherever the reference line crosses the estimated plots). 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 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.

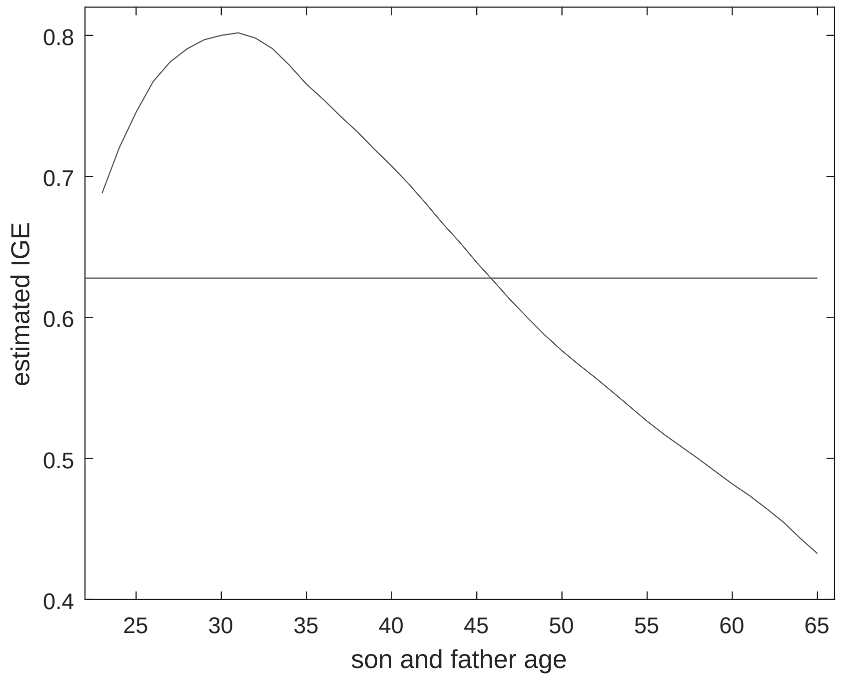
In order to test for the theoretical prediction by Solon (1992) that averaging over many parent income observations decreases attenuation bias, I compute three-, five-, eleven-, and twenty one-year averages for the fathers over all possible ages and then estimate the IGE (using the true average lifetime 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

Figure 7: IGE estimates using three-year averages



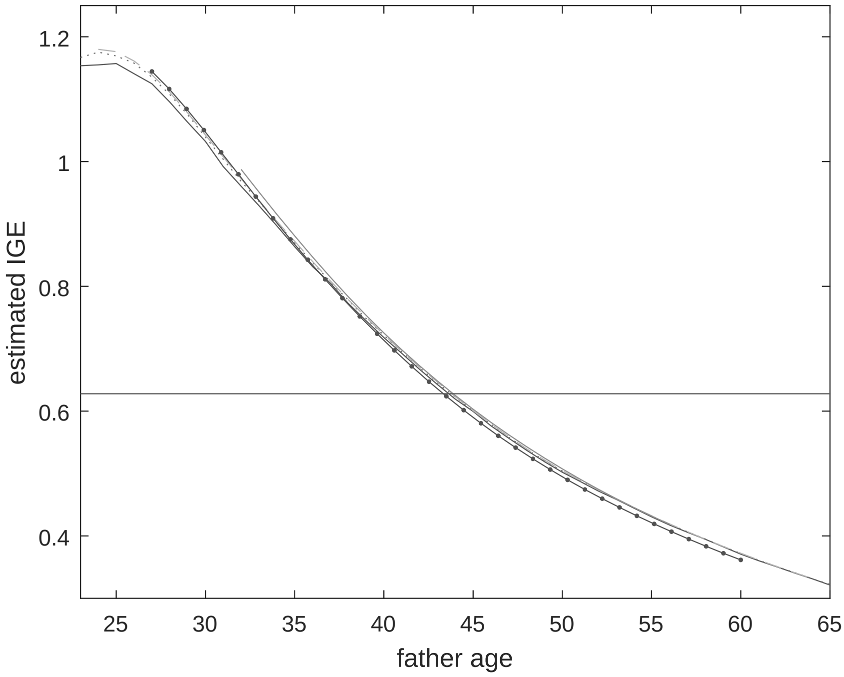
Note: This figure shows half of the IGE estimates using all possible combinations of three-year averages of annual incomes for both generations. The horizontal line indicates the true IGE. Each line shows the IGE estimates over the life cycle of the son, holding constant the age around which father income is observed.

Figure 8: IGE estimates using three-year averages at the same age



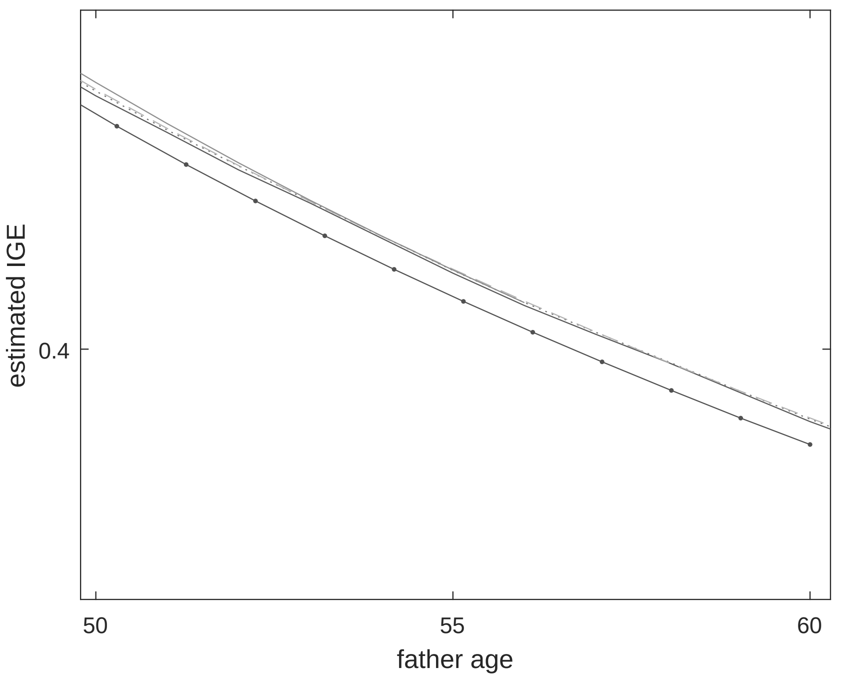
Note: This figure shows the IGE estimates using three-year averages of annual income observations for both generations at the same age. Using the average around age 46 for both father and son results in an unbiased IGE estimate.

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)

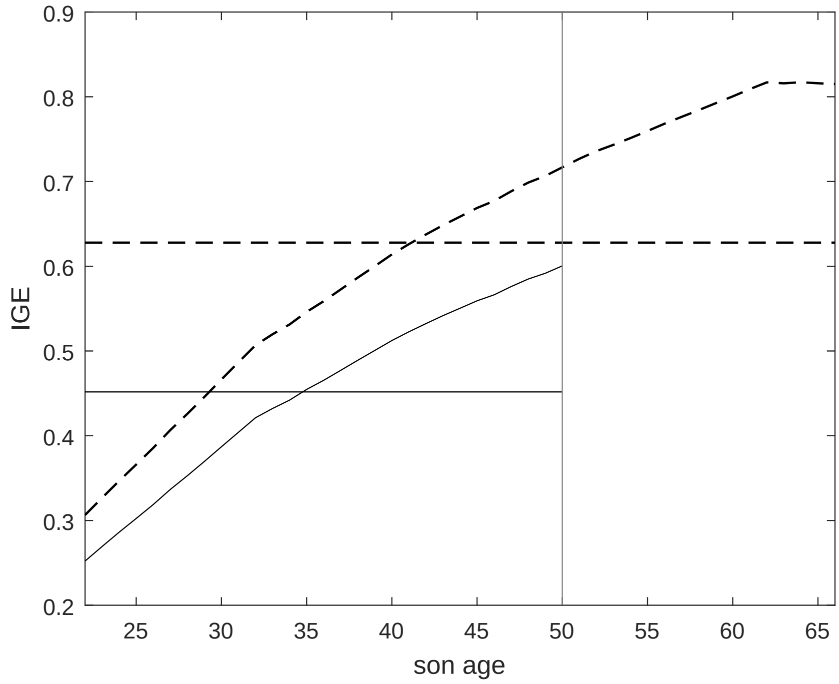


Note: This figure shows the bias of the IGE when using the true son average lifetime income, and 3-, 5-, 11-, and 21-year averages of father income over all possible ages.

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.

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

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).

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 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

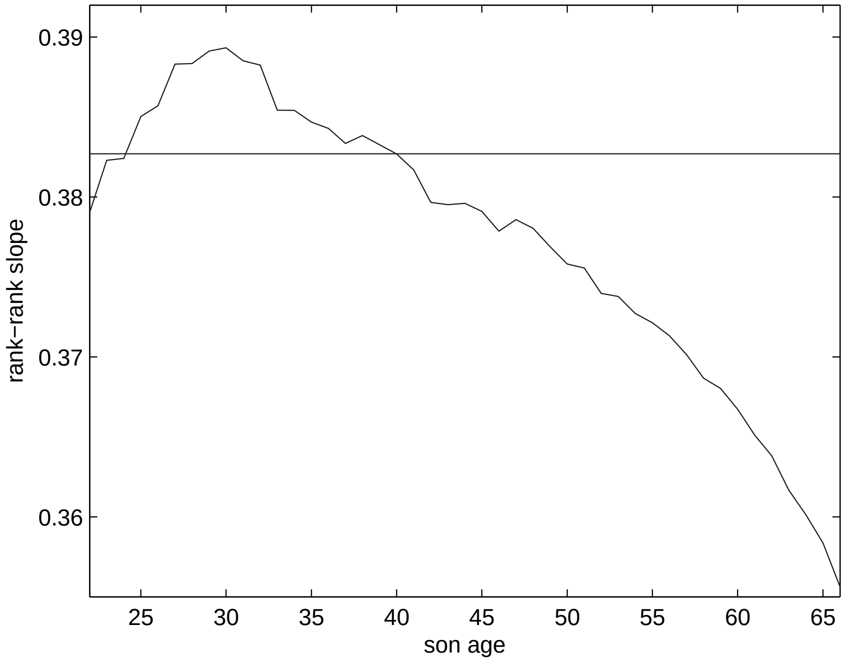
One relatively new development in the literature of intergenerational income associations has been to use percentile ranks instead of log incomes. The relationship between the ranks of parent and child have, for US data, been shown to be well represented by a linear relationship (Chetty et al., 2014a), 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:

$$R_m^s = \beta_0^{rank} + \beta_{rank} R_m^f + \varepsilon_m^{rank} \tag{12}$$

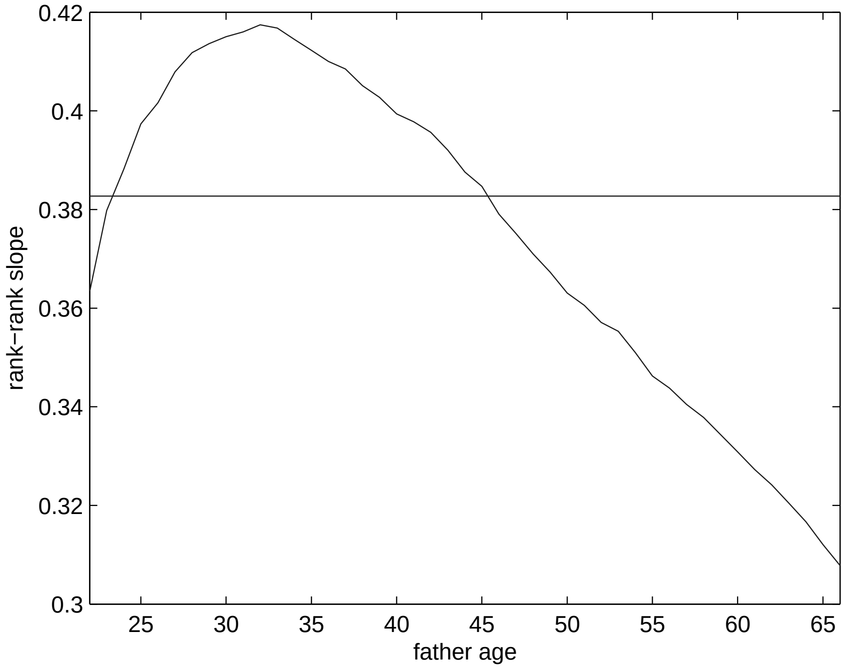
where R_m^s and R_m^f are the average lifetime income ranks for sons and fathers, respectively. All ranks are normalized to lie between 0 and 100.

Figure 12: Rank-rank slope estimates over the life cycle of the sons



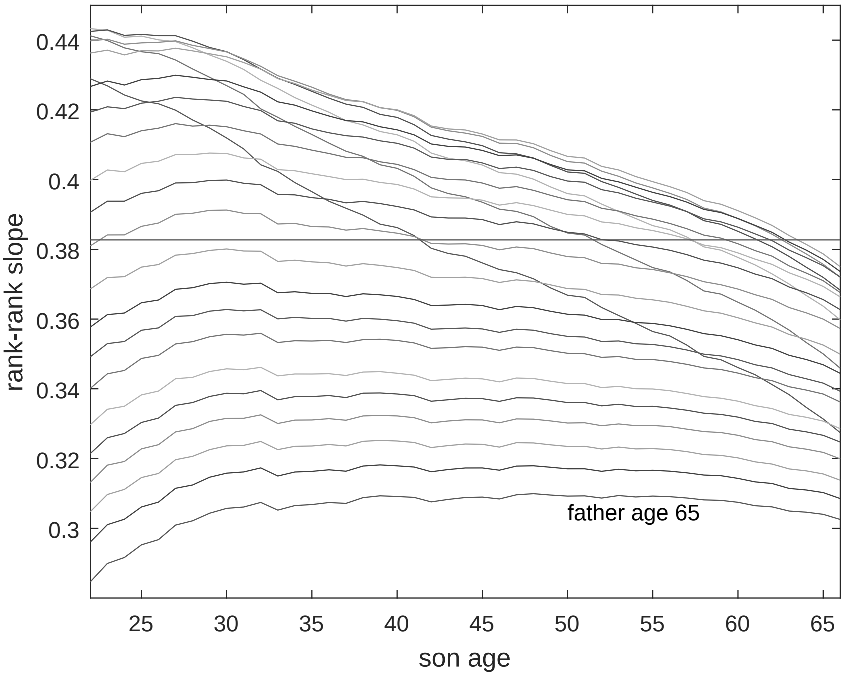
Note: This figure shows the rank-rank slope estimates from regressing 45 annual income rank observations of the son on the average lifetime income rank of the father. The estimates are equal to the true value of the rank-rank slope at around ages 24 and 40.

Figure 13: Rank-rank slope estimates over the life cycle of the fathers



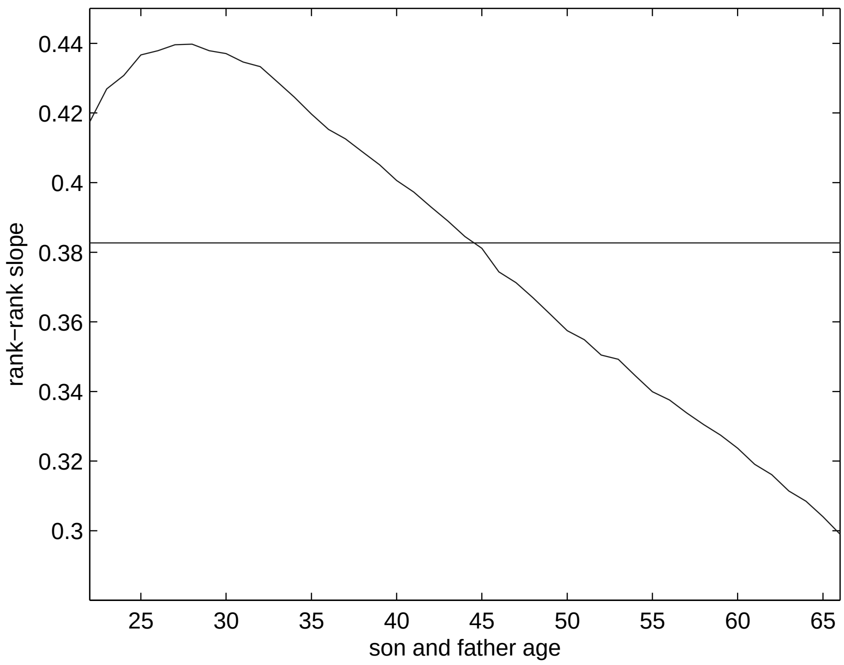
Note: This figure shows the rank-rank slope estimates from regressing the average lifetime income rank of the son on 45 annual income rank observations of the father. The estimates are equal to the true value of the rank-rank slope at around ages 24 and 45.

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.

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.

3.3 Reducing the bias in applied work

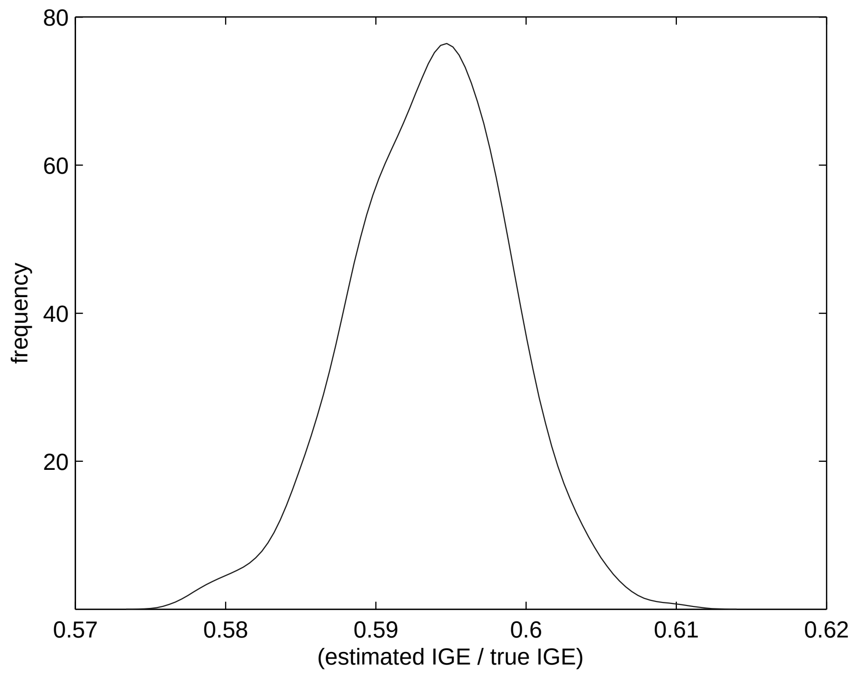
Next, I turn to the typical data situation a researcher faces. The base for most studies are (earned) income observations from either survey data or from register data, usually based on individuals' annual income tax reports. If the data is available for a short time span only and we want to study the 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)



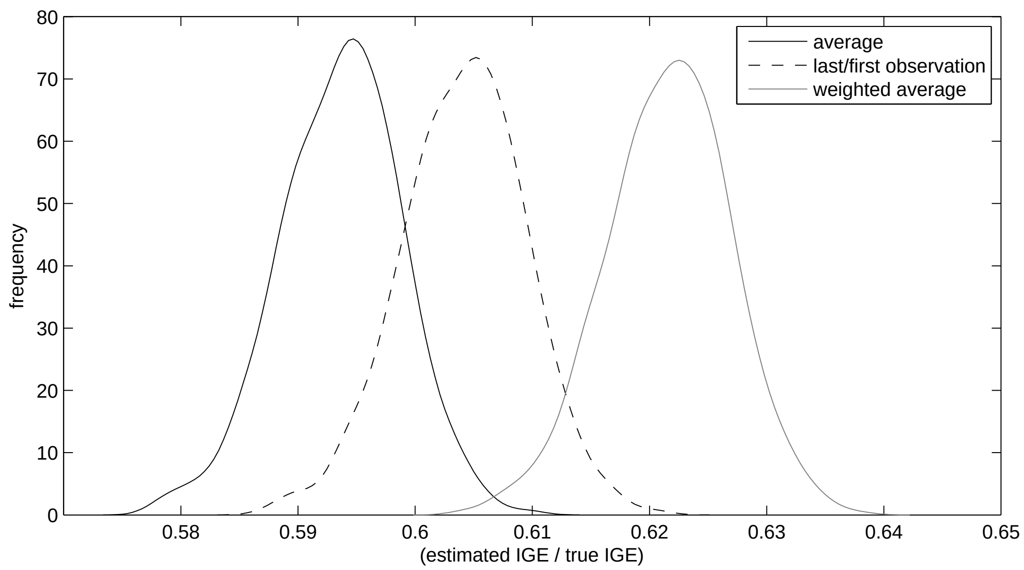
Note: This figure shows the distribution of the fraction of the true IGE that is estimated when using averaged log income at mixed parent ages around 50, and at son ages 31 and 32, to approximate life time income. The mean is 0.59, i.e. the estimated IGE is on average only 59 percent of the true IGE 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.

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

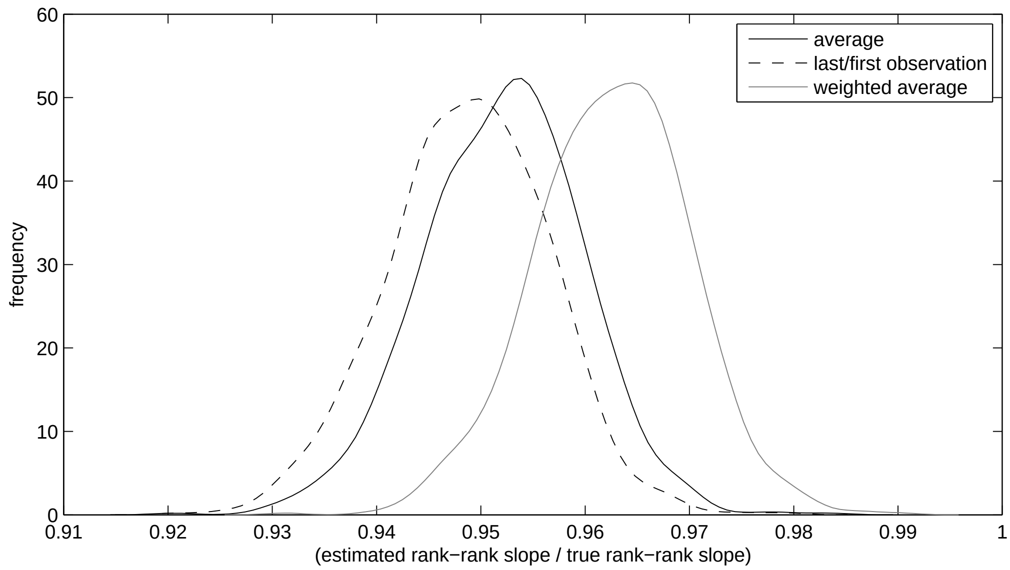
Figure 17: Monte Carlo simulation of typical data: three ways to use the data (log)



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.

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
Corr$(\alpha_f, \beta_f)^{e1}$	-0.25
Corr$(\alpha_f, \beta_f)^{e2}$	-0.7
Corr(α_f, α_s)	0.15
Corr(β_f, β_s)	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.*