

Swedish Institute for Social Research (SOFI)

Stockholm University

WORKING PAPER 3/2015

**INTERGENERATIONAL PERSISTENCE IN LATENT SOCIOECONOMIC
STATUS: EVIDENCE FROM SWEDEN**

by

Martin Nybom and Kelly Vosters

Intergenerational Persistence in Latent Socioeconomic Status: Evidence from Sweden

Martin Nybom

Kelly Vosters*

Abstract: In recent work, Gregory Clark and coauthors argue that rates of social mobility are constant across countries and generally much lower than traditionally estimated. The main explanation is that traditional estimates of intergenerational persistence are heavily attenuated because they use only one proxy measure (e.g., earnings) of *underlying* status. We examine this hypothesis within a suitable latent-variable framework, incorporating multiple proxy measures into a single “least-attenuated” estimate of persistence in latent status. With rich administrative data for Sweden, we exploit detailed proxy measures to test this proposition, and also conduct a Sweden-U.S. comparison. We find no evidence of substantial bias in prior estimates, or that the Sweden-U.S. difference in persistence is smaller than found in previous research. We further explore the concept of *family* status by incorporating mothers, thereby also contributing to the literature on intergenerational transmission for women. We find that while mothers’ income is a poor proxy for status, incorporating information on mothers’ occupation improves the ability to capture transmission from mothers to both sons *and* daughters.

* Nybom: Swedish Institute for Social Research (SOFI), Stockholm University, and IFAU; Vosters: Department of Economics, Michigan State University. We thank Gary Solon as well as participants at the 2015 UM-MSU-UWO Labor Day Conference for helpful comments and discussions. Nybom acknowledges support from Handelsbanken (the Wallander scholarship); Vosters from the Institute of Education Sciences Grant R305B090011 to Michigan State University. The opinions expressed are those of the authors and do not represent the views of the U.S. Department of Education.

I. Introduction

Researchers and policymakers have long shown a great deal of interest in understanding the degree of socioeconomic mobility within and across societies, resulting in a large body of economic research examining the extent to which income differences are passed on from parents to their children. One of this literature's most notable results is that intergenerational mobility in the Nordic countries is substantially higher than in countries such as the United States. However, recent work by Gregory Clark and coauthors has led to surprisingly contrary conclusions, suggesting that the "true" rate of mobility is generally very low and also steady across time and countries with vastly different social and economic contexts, including Sweden and the United States (Clark 2014, p.107).

The descriptive literature on intergenerational income mobility generally estimates an equation resembling a basic AR(1) process:

$$(1) \quad y_{it+1} = \beta y_{it} + \varepsilon_i,$$

where y_{it+1} is offspring log income of family i , y_{it} is parental (typically fathers') log income, and ε_i an idiosyncratic error; β is then interpreted as the intergenerational elasticity.¹ This process is not necessarily taken literally, nor is the estimate believed to be causal, but instead the goal is to obtain a summary statistic describing how differences in economic status persist from one generation to the next. For Sweden, the estimated persistence in income is around 0.2-0.3, compared to 0.4-0.6 in the U.S. (see Solon, 1999, Björklund & Jäntti, 2009, Black & Devereux,

¹ This parameter thus measures persistence, whereas $1-\beta$ is a measure of mobility. For this equation and all those that follow, variables are considered in deviation-from-mean form, allowing intercepts to be suppressed.

2011). The greater mobility in the Nordic countries is often attributed to policy differences, such as more redistributive tax structures, which facilitate public human-capital investments in terms of subsidized pre-school and college education.² Others point out that characteristics of the labor market also matter, such as differences in the returns to skills and the intergenerational transmission of employers (Björklund et al., forthcoming; Corak & Piraino, 2011). However, Clark (2014, p.5) follows the former argument, boldly interpreting the low and constant rates of mobility as evidence of large policy failure.

The creative methods underlying this recent work exploit the information content of *uncommon* surnames in lieu of actual intergenerational family links, and the results paint an extraordinarily different picture of mobility for Sweden as well as for other countries.³ The persistence rate for underlying status is estimated to be as high as 0.7-0.8 across a wide range of societies and time periods, leading to the conclusion that for Sweden, "The implied social mobility rates are as low as those of modern England or the United States" (Clark 2014, p. 41).

These claims are quite controversial, with important implications for the interplay between policy and socioeconomic mobility. They also clearly contradict conclusions from prior intergenerational studies. Acknowledging this incongruity, Clark and coauthors suggest that conventional methods have been limited in their measures of socioeconomic status. The main argument is that families have a *general social status* that underlies imperfect status measures such as income, education, or occupation, and these measures are linked to this underlying and unobserved latent factor with substantial random components. Formalized into a simple model,

² Public investments in children's human capital is put forth as one of the key determinants of the size of the reduced-form intergenerational income relationship in Solon's (2004) log-linear version of Becker and Tomes' (1979) model of parental investments.

³ Guëll et al. (2015) show that rare surnames do contain such information, developing a different method using the joint distribution of surnames and economic status to examine intergenerational transmission of status in Spain.

social mobility is reduced to a universal *law of mobility*, $x_{it+1} = bx_{it} + e_{it}$, where x_{it} is the underlying status of family i in generation t (Clark, 2014). A single measure such as income is then assumed to be related to status with some additive random noise, $y_{it} = x_{it} + u_{it}$, whereby substituting this into the conventionally estimated equation (1) leads to the classical errors-in-variables attenuation bias. Averaging within a surname, z , then reveals true status, $\bar{y}_{zt} = \bar{x}_{zt}$, as \bar{u}_{zt} is approximately zero for large enough surname groups. For data without surnames, Clark & Cummins (2015) propose that if the information from multiple measures—for example, income, education, and occupation—were combined for an individual, then conventionally estimated persistence would rise.

Applying an approach proposed by Lubotsky & Wittenberg (2006) to optimally aggregate information from multiple measures, Vosters (2015) tests this proposition using data from the Panel Study of Income Dynamics (PSID). The estimated persistence rates remain just under 0.5 and are insignificantly different from conventional estimates, even after accounting for multiple partial measures of underlying status. While this study shows that this approach does not substantially raise estimated persistence for the U.S., the question remains as to what information could be extracted from multiple status measures in a country with a more redistributive welfare state, such as Sweden. In fact, according to Clark's hypothesis, this approach should have a *greater* impact on estimated persistence in settings where persistence is conventionally estimated to be quite low.

Therefore, we follow the above approach, performing similar tests to look for any evidence of this asserted attenuation bias in conventional estimates for Sweden. We first provide estimates using measures constructed to take full advantage of the rich Swedish data. We construct nearly career-long income measures, which mitigates biases stemming from transitory

fluctuations (Mazumder, 2005) and life-cycle effects (Haider & Solon, 2006; Nybom & Stuhler, *forthcoming*) in short-run incomes. Our data also have more detailed occupation categories available, allowing us to better examine the degree to which information on status can be extracted from an individual's occupation. Moreover, the small sample size in Vosters (2015) yields low statistical precision and only very large attenuation biases can be formally rejected. In contrast, our sample consists of more than 167,000 parent-child pairs, which provides much greater statistical power. We also examine the claim that persistence is uniform across countries. For this, we provide estimates using variables constructed similarly to those based on the PSID, facilitating a test of whether persistence in underlying status is indeed of equal magnitude in Sweden and the U.S. In doing so, we also indirectly address implications of various data limitations with the U.S. data, such as inaccurate measurement of long-run income and occupations. As such, not only do we obtain results comparable to those for the U.S. to evaluate the applicability of the *simple law of mobility* across countries, we also obtain more robust results on the magnitude of the hypothesized attenuation bias in the Swedish estimates.

We find no evidence to support the *simple law of mobility*, as persistence estimates remain around 0.25-0.30 even after multiple measures are combined. Further, our comparison with the U.S. confirms the prior perception that mobility is indeed substantially higher in Sweden. These results are robust across a variety of specifications and methods for constructing the measures, and the country difference in persistence appears even greater when using measures constructed to mimic those used for the comparable U.S. study. Our findings thus support those in Vosters (2015), as well as the discussion in Chetty et al. (2014), suggesting that

the very low mobility rates provided by the surname approach strongly underestimate the degree of mobility in the population as a whole.⁴

Although much of our evidence reaffirms results from existing literature rather than lending support to Clark's conclusions, we do find that the latent status framework of the *simple law* can be empirically relevant for certain groups (e.g., mothers). Motivated primarily by the concept of *family* status, we add the analogous status measures for mothers as we incorporate those for fathers. We find mothers' occupation to be the most important addition, though producing only a nominal rise in persistence. Further exploring this result, we examine persistence with respect to mothers' status alone, finding that both the estimates of mother-son and mother-daughter relationships increase substantially when multiple measures are used, though rising from low levels compared to those typically found for father-child relationships. Still, these results highlight the unintentional implications of this framework for measurement issues specific to women, showing that combining multiple proxy measures can provide more informative estimates in cases where appropriate income data is not available.

Our paper thus extends the literature on the measurement of intergenerational mobility. To date, research has mostly focused on the measurement of specific status indicators, with the approximation of lifetime (or permanent) income being the prime example. Inspired by the work by Gregory Clark and others, we complement this research by providing new evidence on whether such status indicators themselves, even when accurately measured, suffice to capture a broader concept of socioeconomic status. Our findings imply that for men detailed measures of long-run income are indeed good proxies for latent status. In contrast, for women combining individual income information with occupation improves the measurement of status substantially.

⁴ See also Braun & Stuhler (2015), who discuss the surname approach in the context of mobility across multiple generations.

We also add to the large literature on cross-national mobility differences (e.g., Solon, 2002). The finding based on surnames data, that social mobility is constant across countries, is put into question by our results; these show a Sweden-U.S. country differential that is in line with previous income-based evidence.

The rest of the paper unfolds as follows. Section 2 describes the data, before Section 3 discusses our empirical approach. Section 4 presents our main findings. Section 5 presents our extension to both parents, and the intergenerational associations related to mothers and daughters. The final section offers some concluding remarks.

II. Data

A. Sources and Sample Selection

We use administrative data from various sources, which have been merged by Statistics Sweden using unique personal identifiers. A multigenerational register links children to their biological parents; censuses provide data on parents' occupation and education; income tax declarations for both parents and offspring provide data on total *individual* income. Our main sample is based on a random draw of 35 percent of all children born in Sweden 1951-1961 and their biological parents.⁵ We restrict our analysis to these cohorts for a couple of reasons. Given the available income data, we can observe long-run prime-age incomes of both these offspring cohorts and their parents. Moreover, these are the cohorts used in Vosters (2015), so this selection further facilitates the comparison between our estimates for Sweden with those for the U.S.

B. Construction of Status Measures

⁵ We exclude those with parents who were more than 40 years old at the birth of the child.

For annual income, we use administrative data covering the years 1968-2007. The data are based on individual income-tax declarations and we define the income measure separately for fathers and mothers. Our measure includes income before taxes from all sources except means-tested benefits and universal child benefits. These data come with a number of advantages: they are almost entirely free from attrition and reporting error, pertain to all jobs, and are not censored.⁶ For parents, we approximate log lifetime income by the average of log annual income over ages 30-60. For offspring, we construct measures of log lifetime income as the average of log income over ages 27-46. We require parents and offspring to have at least five non-missing annual income measures.⁷ Throughout all analyses we control flexibly for parental and offspring birth year using cohort dummies.

While the tests we conduct focus on the measurement of “status”, our income measure is particularly important because it also minimizes the potential for two well-known sources of bias in estimated intergenerational associations; bias arising from transitory shocks to income and from life-cycle effects. By using a long-run average of annual income observations, potential attenuation biases from transitory fluctuations are greatly reduced (Mazumder, 2005). In our sample, 88 percent of sons have 20 non-missing log income observations and 88 percent of fathers have at least 10 non-missing log income observations. Further, we measure income as long-run averages during mid-life in order to minimize so-called life-cycle bias (Nybom &

⁶ In contrast to many other administrative data sources, our data are not censored (nor truncated) in the top of the income distribution. Further, the Swedish system provides strong incentives to declare some taxable income since doing so is a requirement for eligibility to most social insurance programs. Hence, we expect very little missing data in the low end of the distribution.

⁷ Missing income is rare in our sample, and such occurrences could be due to quite different reasons; individuals could be living abroad, they could fail to file their tax declaration, or it might arise due to coding errors.

Stuhler, *forthcoming*). While there are slightly fewer mid-life income observations for fathers, 91 percent of them have at least one annual income observation from before age 50.

We use occupation data from national censuses conducted every five years between 1960 and 1990. The occupational classification employed in the censuses builds on the Nordic Occupational Classification (NYK), which is based on the International Standard Classification of Occupations (ISCO). The NYK categorizes occupations according to the end result of the tasks and duties undertaken in the job. Hence, level of education and professional status are typically not considered in the categorization (Statistics Sweden, 2004). The classification has a hierarchical structure, allowing for analyses at different aggregation levels. Three-digit codes denote unique occupations, two-digit codes denote minor occupation groups and one-digit codes denote major occupation groups. To fully exploit the available information, we use the unique occupation indicators in our main analyses, but also test the sensitivity of our results to using the broader classification levels.

We define a parent's occupation as the occupation he or she had in the 1970 census. Fathers in our sample are, on average, about 44 years old in 1970, so this census provides a good prime-age occupation measure. If occupation is missing in this census, however, we use the corresponding data from the 1975 or the 1980 censuses.⁸ For those with occupation still not coded, we include indicators for missing and undefined in our main specifications to flexibly account for these special cases. Including missing and undefined as separate categories, the resulting sample holds 270 unique occupations classified into 61 minor occupation groups, or 12 major occupation groups. To demonstrate the nature of the classification, the major occupation

⁸ Incorporating the later censuses is primarily beneficial in obtaining more accurate information on occupation for mothers, who are more likely to have missing data in 1970. Very few fathers have missing occupation in 1970.

groups are: 1) Professional work (arts and sciences); 2) Managerial work; 3) Clerical work; 4) Wholesale, retail, and commerce; 5) Agriculture, forestry, hunting, and fishing; 6) Mining and quarrying; 7) Transportation and communication; 8) Manufacturing; 9) Services; 10) Military/Armed forces; 11) Undefined; 12) Missing.

For parental education, we use data on final education in 1970 according to the data from Statistics Sweden's education register, which is based on a standard conversion translating each level into years of education. The measures of parental education reflect their highest educational attainment, with the levels including: less than nine years of primary school, nine years of primary school, two-year secondary school, three-year secondary school, less than three years of post-secondary school, three years or more of post-secondary school, and graduate school. We also perform a set of robustness tests in which we control for education more flexibly. First, we again use the above measure but now by including a dummy variable for each of the different levels. Second, we also exploit more detailed information on educational attainment from the same data source. In doing so, we include a large set of dummies reflecting length and *type* of education, distinguishing between various tracks within high school as well as a large number of different academic and vocational post-secondary educational categories.

Because we are exploring implications of aggregating the information on parental income, education, and occupation, we only include parent-child pairs for which the parents have non-missing information on all of these measures and the child has the requisite non-missing income measures. Table 1 provides descriptives for our resulting main sample of 167,552 sons matched to 153,920 fathers.

C. Alternative Measures for U.S. Comparison

We also construct alternative measures to facilitate a Sweden-U.S. comparison based on comparable findings in Vosters (2015). The analysis by Vosters is based on data from the nationally representative part of the Panel Study of Income Dynamics (PSID), which began with a sample of about 3,000 families in 1968. Importantly, the PSID includes family links and follows original sample members and their children over time. Fathers are identified as the male head of the household in which the child resided at the time of the initial survey, which does not necessarily represent a biological link. Thus, our Swedish sample differs slightly in that we use biological rather than cohabitating fathers.⁹

To enable a credible cross-country comparison, we construct alternative measures for Sweden that are analogous to those from the PSID. For offspring income, we use the log of annual income in 1991. Fathers' income is defined as the average of log income in 1968-72.¹⁰ Our education measure is very similar, reflecting the highest level of attainment. For occupation, we use the major groups described above, which differ slightly but not much from the seven groups used in the PSID (see Vosters, 2015). To better match the last "residual" category in the PSID, we add missing and undefined occupations to our military/armed services category, resulting in 10 major categories for the Swedish sample. In the U.S. data, education and occupation are from the 1968-1969 surveys, while our corresponding data are from 1970. Although there are minor differences in some variable definitions across the two countries, they are marginal at most and should have very little effect on our results. The sample with non-missing data on these measures includes 146,783 sons matched to 135,020 fathers.

⁹ For approximately 95 percent of the sons in the Vosters (2015) PSID sample, the identified cohabitating father is in fact the biological father, so this difference is minor.

¹⁰ Since our income data start in 1968, this measure is marginally different from the U.S. data that are based on earnings in 1967-71.

We provide descriptive statistics for both the full sample and this restricted U.S. comparison sample in Table 1. The samples are very similar across all observables. Sons' average income is slightly higher than that for fathers; in logarithmic form, these averages are 12.22-12.29.¹¹ Fathers' average education of just over 9 years, as well as the distribution among the various levels of attainment, is nearly identical across samples, as are the proportions in each occupation category. Professional work and manufacturing comprise much of the sample of fathers, with 19 and 38 percent in the respective categories.

3. Empirical Approach

Our empirical approach is designed to test the hypothesis that estimates of intergenerational persistence in socioeconomic status approach 0.7-0.8 as we add the proposed partial measures. We then proceed to contrast our results with comparable estimates for the U.S. to also test the claim that persistence in latent status is the same across countries. We begin by obtaining a baseline estimate of persistence by estimating the usual intergenerational income equation above in (1). To gauge the degree of attenuation bias in this estimate, we then add the additional measures of parental status to this equation. Although this provides insights into the sensitivity of conventional estimates to accounting for other status measures, it does not provide a *single* estimate of persistence in underlying status that *combines* information from all measures.

Our preferred method, proposed by Lubotsky & Wittenberg (2006), estimates the persistence in *latent status*, aggregating the information in the included proxy measures. To better illustrate our methodological approach, we first present the hypothesis in a simple latent variables framework, writing measurement equations for each of the partial measures, y_{jit} , of the form:

¹¹ Income is provided in 2005 Swedish kronor (SEK).

$$(2) \quad y_{jit} = \rho_j x_{it}^* + u_{jit},$$

where j indexes the measure, i indexes family, and t generation. We generalize the measurement equations from the *simple law* to allow for slope coefficients, ρ_j . Our main empirical specifications include equations for y_{1it} for parental (e.g., fathers') income, y_{2it} for parental education, and $y_{3it} - y_{kit}$ for the $k-2$ parental occupation indicators. x_{it}^* is the unobserved latent status and the u_{jit} are the measurement errors. The so-called structural equation can then be written:

$$(3) \quad y_{it+1} = \beta x_{it}^* + \varepsilon_{it},$$

where β is the measure of intergenerational persistence in underlying latent status. This notation shows that we do not explicitly address offsprings' latent status with multiple partial measures.¹²

However, the outcome variable we do use—a twenty-year average of annual incomes during mid-life—is likely one of the best *single* measures of socioeconomic status available. Further, the simple conditions underlying the *simple law* of mobility rely on the assumptions of a classical errors-in-variables model, under which measurement error on the left-hand side is innocuous.

¹² To assess sensitivity to this choice, we performed two different tests. First, we created omnibus measures of status for fathers *and* sons (applying fathers' weights to sons' measures) and obtained an estimate of 0.237, which is nearly identical to the comparable estimate (0.238) using only fathers' measures. Second, we used average log incomes across same-sex siblings as measure of offspring status (excluding those without same-sex siblings from the sample). While baseline estimates and thus the scaling differ in the latter case, the estimated decrease in attenuation bias is very similar.

Under the classical assumptions, the measurement errors (u_{jit}) are all uncorrelated, and the coefficients ρ_j are equal to one. In this simple case, there are several econometric methods available. For example, instrumental variables (IV) estimation using one measure to instrument for another is common solution. We provide one such estimate, using father's education to instrument for father's income, which under the proposed *law* should estimate persistence levels in the 0.7-0.8 range.¹³ Other possible approaches include the MIMIC (multiple indicators, multiple causes) or LISREL frameworks (see, e.g., Jöreskog & Goldberger, 1975, and Bollen, 1989). More recently, Black & Smith (2006) propose a GMM estimator with potential efficiency gains. However, each of these approaches relies critically on the assumption of uncorrelated measurement errors, and we find this restriction to be particularly concerning in the setting considered here.¹⁴ First, the nature of the suggested measures (income, education, and occupation) makes the case of zero correlation among measurement errors unlikely. Second, the anecdotal examples used to motivate the concept of underlying latent status directly imply correlation among the measurement errors.¹⁵ Further, our main purpose is not to point identify β , rather we seek to test whether attenuation bias *decreases* as multiple proxies for latent status are taken into account. The LW approach is in this respect superior, allowing us to compare different

¹³ Note that in this particular IV setup, consistency requires only the coefficient in the income measurement equation to equal one (and the measurement errors still being uncorrelated), which is not problematic as this is the normalization we adopt for our preferred approach described below. This normalization simply sets the scale of latent status to be on that of fathers' income.

¹⁴ If the measurement errors were positively correlated, Black & Smith (2006) point out that the IV estimate from using one measure to instrument for the other provides a benchmark for a lower bound. In our case though, the measurement errors may be negatively correlated, which would leave the IV estimate biased upward.

¹⁵ Clark (2014, p.11) refers to education being a poor measure of status for Bill Gates (who presumably has high status), as he is a college dropout but has incredibly high income. Conversely, the other example posits that income would be a poor measure of status for a philosophy professor, whose education would be a more appropriate measure. These scenarios imply a negative correlation among the measurement errors for income and education.

lower bounds without making restrictive assumptions on cross-correlations of the measurement errors.

In addition to relaxing the assumption of zero correlation among the measurement errors (u_{jit}), we also allow the coefficients, ρ_j , in the measurement equations to be mostly unrestricted (subject to a normalization discussed below). The approach from Lubotsky & Wittenberg (2006; henceforth, LW) is ideally suited for this scenario, as it actually exploits the correlation in the measurement errors and estimates the coefficients in the measurement equations. In fact, the LW approach incorporates the information from all included measures of status in an optimal fashion, producing the estimate of persistence with the least attenuation bias. The LW estimator can be written as:

$$(4) \quad \beta_{LW} = \hat{\rho}_1 \hat{\phi}_1 + \hat{\rho}_2 \hat{\phi}_2 + \dots + \hat{\rho}_J \hat{\phi}_J,$$

where the $\hat{\rho}_j$'s are estimates of the slope coefficients in the measurement equations, and the $\hat{\phi}_j$'s are obtained from an auxiliary OLS regression described below. Hence, actual computation entails a multistep process.

The first step of the LW approach is to obtain the auxiliary OLS coefficient estimates of ϕ_j from regressing the dependent variable on all measures of status:

$$(5) \quad y_{it+1} = \phi_1 y_{1it} + \phi_2 y_{2it} + \dots + \phi_J y_{Jit} + v_i.$$

To identify the coefficients in the measurement equations, we need a normalization assumption on one of the ρ_j 's. We normalize $\rho_1 = 1$, which simply sets the scale of the latent status to be on

that of fathers' log income.¹⁶ This implies the following formula to obtain estimates of the ρ_j 's:

$$(8) \quad \rho_j = \frac{\text{cov}(y_{it+1}, y_{jit})}{\text{cov}(y_{it+1}, y_{1it})}$$

Estimating this ratio can be done in a single step via IV estimation, with y_{jit} as the outcome variable and using y_{it+1} to instrument for y_{1it} . We obtain standard errors for the β_{LW} estimate using a block bootstrap (100 replications) to account for within-family correlation. While not identifying β itself, this estimator provides an estimate of β with the least attenuation bias based on the joint information in the proxy measures of status. If the *simple law of mobility* does hold, we should see estimated persistence levels rising as we add these measures of status.

In addition to the proclaimed elevated persistence (i.e., lower mobility), the other controversial aspect of the *simple law* is the assertion that rates of mobility are constant across countries. To facilitate a cross-country comparison between Sweden and the U.S., we estimate analogous specifications using a Swedish sample with the measures constructed similarly to those used for the U.S. by Vosters (2015). From this we can also examine the consequences of various data limitations within the Swedish setting, thus providing indirect evidence on whether the U.S. estimates would change if based on richer data. We also conduct various robustness checks with other constructs of the income, education, and occupation measures.

Further, because the hypothesized simple law relies on the social status of *families*, we extend our analysis to other family members, by adding the analogous measures for mothers, as

¹⁶ This normalization hence allows the LW estimate to be directly comparable to the conventionally estimated intergenerational income elasticity. In fact, in the case where income is the only status measure used, it is easily seen from equations (4) and (5) that the LW estimate is identical to this conventional estimate.

well as estimating specifications with only mothers' measures. This exercise provides some suggestive evidence not only on the role of mothers but also on new methods for measuring mothers' status. In addition, given the paucity of evidence on intergenerational persistence for daughters, we also extend our analysis to daughters.

4. Empirical Results

A. Main Results

We first examine the conventionally estimated intergenerational persistence of income in Sweden, and whether adding additional partial measures affects the estimated coefficient on log income. In these and all other estimations, we control flexibly for cohorts of each generation using birth-year dummies. For the results presented in Table 2, we use the long-run average of sons' log income as the dependent variable, and fathers' measures of status include the long-run average of fathers' log income, educational attainment, and unique occupation indicators. The first set of results in Table 2 provides OLS estimates (omitting those for the 269 occupation indicators for brevity), with columns [1]-[4] progressively adding measures of fathers' status. Note that these estimates also correspond to the OLS components (ϕ_j) of the LW estimate obtained in the auxiliary regression.

The baseline OLS estimate of equation (1) is 0.23. This estimate of the intergenerational income elasticity is of similar magnitude to previous estimates for Sweden. Moving to column [2], fathers' educational attainment is added to the regression, but the coefficient estimate on fathers' income remains nearly identical. When instead fathers' occupation indicators are added to the regression in column [3], the coefficient on income does fluctuate some, falling to 0.21. This estimate is hardly affected by the inclusion of education, as shown in column [4], indicating

that while there is some sensitivity to the addition of the occupation measure, we see very little sensitivity to the addition of educational attainment.

With two noisy measures of status, and assumptions of classical measurement error, IV estimation provides a consistent estimate of intergenerational persistence in underlying status. Considering this scenario, the next rows of Table 2 include first and second stage results when instrumenting for fathers' income using fathers' education. This estimate of persistence is 0.24, similar to conventionally estimated persistence for Sweden. However, it is important to recognize the possibility that the assumptions for consistency may be violated. In particular, the measurement error in income as a measure of social status may be correlated with the measurement error in educational attainment, leaving the direction of bias unknown without further information on the nature of the correlation.

The final estimation approach, proposed by Lubotsky & Wittenberg (2006), exploits such violations by using the information on the relationships among the measurement errors and providing the greatest lower bound on persistence in underlying status. The LW estimate in column [1] is identical to the OLS estimate (by construction). However, as we incorporate more measures of status, this approach provides a single estimate from an optimal aggregation of the information from all measures. Given that the OLS estimates shown in the top of Table 2 are underlying components of the LW coefficient estimate, it is unsurprising that adding education does not change the LW estimate, as shown in column [2]. Similarly, given the sensitivity of the OLS coefficients to adding the occupation indicators, the increased persistence with the inclusion of occupation in column [3] is somewhat expected. However, the nominal rise from the conventional estimate of 0.23 to 0.26 when all suggested partial measures are included (column [4]) does not support the hypothesis of substantial attenuation bias in prior estimates. This pattern of results is similar to that found for the U.S. (Vosters, 2015), exhibiting minimal increases in

persistence when more partial measures of status are included, despite claims of elevated persistence in all countries. However, an important difference here is the statistical certainty. Due to our much smaller standard errors, we can reject even moderate drops in attenuation bias (for this specific model).

B. A Comparison of Sweden and the United States

Next we turn to directly address the hypothesis that persistence in status is in fact constant across countries. Our main results (provided again in Table 3) show that the persistence estimates for Sweden remain in the previously cited range of 0.20-0.30. Further, these estimates are substantially lower than the U.S. estimates of 0.44-0.47 found by Vosters (2015), illustrating a meaningful distinction in persistence between the two countries. However, the Swedish measures are constructed differently (e.g., the long-run income measure and the unique occupation indicators). While the measurement differences would likely bias the U.S. estimates towards the Swedish estimates, we carefully construct our measures to mimic those used by Vosters to allow for a more sound comparison. Using the Swedish data, we also indirectly gauge what the estimated persistence in status might look like in the U.S. if based on richer data.

With the five-year average of log income and broad occupation categories constructed to match those for the U.S., we find that estimated persistence in Sweden is lower at 0.19-0.22. To check whether this might be due to sample composition differences between our main sample and this smaller sample, we also analyze the same sample using our original measure constructs, and find estimates (0.23-0.26) nearly identical to our main results. Sample composition does not appear to be driving the differences. While our results do show that the U.S. estimates may be somewhat attenuated, possibly by some 10-20 percent, we can also see that the increase in estimates as additional measures are added does not change regardless of how measures are

constructed; in no cases do the estimates rise substantially when additional measures of status are included. Further, the estimates for Sweden remain in the approximate range previously asserted in the literature, albeit at the low end around 0.20, and clearly differ from the estimates for the U.S.¹⁷ Thus, our results fail to support either aspect of the *simple law of mobility*.

C. Robustness of Main Results

Next we examine the sensitivity of our main results to various modifications to the measures of status. For our measure of income, we did see some sensitivity to the number of yearly income observations included in the average, as the five-year average used for the Sweden-U.S. comparison produced lower estimates than our longer-term measure. Another more arbitrary aspect of our measure construction is the choice to use the average of the annual log earnings rather than the log of average annual earnings. We provide estimates based on this alternative income measure construction in Table 4. While these estimates are slightly higher than our main results, they still remain in the typical range of estimates for Sweden. Moreover, the general pattern of the estimates as additional measures of status are added remains unchanged.

The other adjustments to the income measure, as well as the education and occupation measures, are motivated in part by our chosen empirical approach. For example, our long-term income measure gives equal weight to each annual measure from age 30 to 60, while each annual measure entering separately would allow the LW method to optimally choose these weights, which may vary over the life cycle. However, since the LW method also excludes any observations with a missing covariate and several of the fathers in our sample have incomplete

¹⁷ That the estimate for this specification is slightly lower than previous ones in the literature is not unexpected. While previous estimates have been based on long-run income measures and an optimal use of existing data, our goal here is to use data constructs comparable to the U.S.

income histories, we estimate specifications that include annual log income from age 40 to 50, to reduce the data requirements while still focusing on income measures during mid-life. The persistence estimates are higher, ranging between 0.25-0.30, but are based on a much smaller and presumably more homogeneous sample of fathers that have log income observed in each of these eleven consecutive years. For comparison, we also estimate persistence based on the average of these annual log incomes, finding persistence estimates to be slightly lower (0.24-0.30), suggesting only trivial gains from allowing the LW method to determine the weights on the separate annual income measures. For another point of reference, the corresponding estimates using our baseline income measure for this sample are 0.28-0.33, which are even higher. So it appears that this sample exhibits more persistence than the full sample, but we also see that our longer-term average is serving as a better proxy for status than using the more flexible annual income measures when limited to fewer years.

We also adjust the education and occupation measures. For our main specifications, educational attainment enters under the assumption of a linear relationship in years of schooling. We relax this by using indicators for each level of highest attainment. Even with this flexible approach, education does not appear to provide substantial information on status (conditional on income), with estimates increasing by less than 0.01. We also estimate specifications indicating the *type* of education along with each level, again with increases of less than 0.01 in the estimate. Our main specification used the most flexible measure available for occupation. However, these detailed occupation indicators can be grouped into minor or major occupation groups (similar to those used for our U.S. comparison), resulting in estimates of 0.25 and 0.24, respectively. We thus see some numerical sensitivity of the estimates in this regard, though not to an extent that would affect the conclusions reached with our main analysis. In our main analysis, we included observations with occupation missing or undefined, accounted for using separate category

indicators. When excluding these two groups, the baseline estimates increase by around 0.05. However, this modest numerical change has little effect on our main conclusions regarding the level of persistence in Sweden (or the comparison to the U.S.).

In general, our robustness checks in Table 4 show that while there is some sensitivity of the estimates to how the partial measures are constructed, none of the changes are meaningful qualitatively. In particular, they do not change our conclusion that estimated persistence is not converging to 0.7-0.8 as additional measures are included, nor the conclusion of higher mobility in Sweden relative to the U.S.

5. Extension to Mothers and Daughters

Our results thus far have focused on male lineages, as is common in the intergenerational literature (including Vosters' and Clark's work). However, the *simple law* is described to pertain to underlying latent *family* status. To more appropriately address the concept of *family* status, we perform tests analogous to those above but including mothers' income, education, and occupation in addition to the same measures for fathers. This extension is warranted both by the specific hypothesis we are testing, but also by the dearth of evidence pertaining to mothers. To supplement the limited evidence in the literature, we also estimate persistence based on only mothers' status, and then attempt to disentangle contributions of status measures separately for mothers and fathers, in determining their child's later socioeconomic status. Since intergenerational associations for daughters are also much less common in the literature—

especially mother-daughter associations in individual income—we conduct all of these tests for daughters as well.¹⁸

Similar to our main analysis, we begin by obtaining a baseline estimate via OLS and further augment the regression with additional measures. The first set of results in Table 5 replicates the main analysis for fathers and sons, only now restricting the sample to sons matched to both a father and mother, to facilitate comparisons with the different parent-offspring samples considered here.¹⁹ For sons, the coefficient on fathers' income is not affected by the addition of education, while for daughters, adding fathers' education does seem to matter. The estimates for both daughters and sons are somewhat sensitive to accounting for fathers' occupation. When we add the corresponding measures for mothers to each of these specifications, the changes in the coefficient estimates are negligible for both sons and daughters (comparing the first panel to the second). The last set of results is for specifications using only mothers' measures. As the coefficient on mothers' income is very low, these estimates illustrate why mothers are generally not considered in studies of intergenerational *income* persistence. While today Sweden indeed has a high rate of female labor force participation, it was much lower for the cohorts of mothers in our sample (born before 1940), and thus individual income is a very noisy indicator of socioeconomic status.

In Table 6 we present the LW results, which aggregate information from additional status measures for each of the different parent-child samples. For fathers and sons, the results are

¹⁸ Chadwick & Solon (2002) for the U.S. along with Rauum et al. (2007) for several different countries look at intergenerational income associations for daughters, circumventing the labor force participation issues by using a family income measure. Altonji & Dunn (1991) comprehensively looks at associations in family income and individual income, for all parent-child pairs, using U.S. data.

¹⁹ Descriptives for these samples can be found in Appendix Table A1. OLS and LW results for the full mother-offspring and father-offspring samples can be found in Appendix Tables A2 and A3, respectively.

nearly identical to the main results from the full sample, with persistence estimates ranging 0.23-0.26. For daughters, the intergenerational persistence in status with regard to their fathers is slightly lower, with estimates ranging 0.15-0.19. An important difference is that fathers' education does matter for the association in status with daughters, while it did not for sons. Fathers' occupation is similarly important for persistence in status with daughters and sons. The results for mothers are more striking, showing that mothers' occupation is crucial for measuring mothers' status. This holds especially when considering intergenerational associations with sons, as the estimated persistence rises from 0.03 to 0.24. For daughters, the corresponding increase is from 0.06 to 0.13. These estimates are similar to the mother-son association found for the U.S. by Altonji & Dunn (1991), though they obtained a larger mother-daughter estimate. Clearly income is a very poor measure of status for mothers, and this is further confirmed by the results in Table 6; what was not obvious in the OLS results in Table 5 is the substantial impact of accounting for mothers' occupation, which is made apparent by the LW method's aggregation of all information contained in mothers' income, education, and occupation. Education is also salient to mothers' status, as shown in columns [6] and [8], though less so than occupation.

Next we include mothers' and fathers' measures jointly, to consider how persistence might change if we take more literally the concept of *family* status. When we compare these estimates to those accounting for only fathers' status (i.e., estimates reflecting the same information as most of the literature), we see that mothers' occupation does contain additional information on family status with respect to intergenerational transmission to sons, and even more so for daughters. Further, mothers' income seems salient to transmission of family status for daughters, a result consistent with Altonji & Dunn's (2000) finding that factors underlying earnings had stronger intergenerational associations along gender lines.

To further assess the relative importance of mothers' and fathers' status measures, we also attempt to separate the relative contributions of each parent to the intergenerational persistence estimate. Decomposing the estimate into portions due to mothers' and fathers' status, we see in the bottom portion of Table 6 that the vast majority of the persistence for sons is coming from fathers' measures, with only 4-5 percent from mothers in the income and income/education specifications.²⁰ Mothers' occupation appears more important though, shifting more weight to mothers so they account for 15-16 percent. For daughters, the role of mothers' status is more substantial, accounting for 32-43 percent of estimated persistence in underlying status.

Whether mothers should contribute (conditional on fathers) to intergenerational persistence in family status is an empirical question. Theoretically, one could posit several stories. For example, if we believe there to be substantial positive assortative matching on latent status in the marriage market, then we might expect mothers' or fathers' status measures to serve as equally suitable measures of family status. Indeed, for the sample of sons, LW estimates from specifications including all measures for fathers are very similar to those including all measures for mothers (0.26 and 0.25, respectively). However, this is not the case when occupation indicators are omitted; nor does it hold as convincingly for the sample of daughters (with estimates of 0.19 and 0.14). While previous studies have found evidence of positive assortative matching in both Sweden (Hirvonen, 2008; Nakosteen et al., 2004) and the U.S. (Chadwick & Solon, 2002), this does not seem to explain our results here. In auxiliary correlational analyses, we find the mother-father correlation in educational attainment to be the highest at 0.55, but the correlation in long-run income is low (0.06). The correlations between mothers' and fathers'

²⁰ The decomposition is done by separating the sum $\beta_{LW} = \hat{\rho}_1 \hat{\phi}_1 + \hat{\rho}_2 \hat{\phi}_2 + \dots + \hat{\rho}_J \hat{\phi}_J$ into the sum of elements from fathers' measures and the sum of elements from mothers' measures.

estimated latent status is also low (0.08), which is not surprising given that income both weights heavily into the status measures and exhibits low parental matching.

More likely, our results are explained by the well-known issues with using mothers' income, some of which we mentioned above. For education, it is less clear what the explanation is; educational attainment is both believed to suffer less from measurement problems and exhibit smaller male-female differences than what is the case for income. However, we do see that combining information from income *and* education can mitigate these measurement issues, and adding occupation is especially helpful. So Clark & Cummins' (2015) proposition that persistence estimates will rise when combining information from multiple measures seems to have some merit for capturing intergenerational associations with mothers. Each of our noisy measures contributes to measuring mothers' status, however not to the extent of raising persistence estimates to the levels proposed in the *simple law*.

6. Conclusions

Clark's work shifts the focus to be on *underlying socioeconomic status*, which is described to be a slightly different—presumably more general—concept relative to the purely economic ones economists have thus far considered. While it is not entirely clear to what extent these concepts should differ, Clark's work is painting an entirely different landscape for intergenerational persistence, provoking a new set of studies (such as this one) testing the surname results and associated hypotheses. Very few of these papers are confirming the results found with the surnames approach or the proposed reasoning for the contradictory results, as in the present paper (e.g., Chetty et al., 2014; Braun & Stuhler, 2015; Vosters, 2015).

We tested two facets of the hypothesized simple law of mobility, failing to find evidence to support either claim. We first looked for evidence of substantially increased intergenerational

persistence in underlying social status in Sweden when information from several partial measures of parental status was combined. Incorporating information on educational attainment has almost no effect on the conventionally estimated persistence rate of 0.23. When occupation is included, the estimate increases slightly to 0.26, but does not come close to the hypothesized “true” persistence rate of about 0.7-0.8. We then investigated the claimed uniform persistence across countries, by comparing our Swedish estimates with those for the U.S. (presented in Vosters, 2015). Even after harmonizing our sample and variables as to mimic those used in the U.S. study, our estimates still differ substantially, with the U.S. estimates of persistence being more than twice as large as the Swedish. Our analysis thus confirms the previously established higher levels of intergenerational mobility for Sweden relative to the US.

Prior studies, such as Goldberger (1989), also recognized that non-income measures may be important in measuring persistence in socioeconomic status.²¹ However, Clark’s theory formalizes this notion with a very simple measurement error framework and proposes an easily testable hypothesis. So while Clark is not the first to emphasize the importance of non-income measures, the exercise of considering a more general latent status has also prompted various extensions to the literature. For example, ours along with Vosters (2015) is one of the first studies to aggregate information from different dimensions of status into a single measure of persistence. While sociologists and economists have included, say, income and education in the same regression, these have been attempts to identify mechanisms, or simply reactions to data limitations, rather than for the purposes of obtaining one aggregate persistence estimate.

Coupled with our method for obtaining an aggregate estimate, Clark’s theory regarding latent *status* unintentionally inspires another important contribution to both the measurement and

²¹ Sociologists also consider non-income measures, instead often focusing on social “class” and various measures of occupational prestige.

intergenerational literatures, enabling further examination of intergenerational associations related to mothers. Studies rarely consider status transmission from mothers to children, or even fathers to daughters, due to data limitations stemming from lower labor force participation rates for females. In the context of Clark's work, despite the underlying theory being presented in the realm of male lineages, the latent variable approach might be more relevant for females. Hence, we first extended our analysis to more carefully consider the concept of *family* status by accounting for mothers' in addition to fathers' status measures, which had very little impact on estimated persistence, especially for sons, with persistence rising to only 0.28. Although beyond the scope of the *simple law*, we do find the framework to be more relevant to females, especially mothers. We show that a modified version of the measurement error framework presented by Clark proves useful in estimating intergenerational associations between mothers and their offspring. In contexts where income is a very noisy measure of socioeconomic status, as is often the case for mothers, supplementing this information with additional noisy measures can make an important difference, as shown by our analyses incorporating mothers' education and occupation.²² In fact, for daughters the intergenerational persistence estimates accounting for all measures are only slightly lower for mothers relative to fathers. While these results warrant future research for a better understanding of these transmission channels, our results here illustrate what information might be gained by considering other estimation approaches, and other measures of status.

²² If the available income data is of low quality (or observed only as short snapshots), the same approach could also be potentially useful when studying men.

References

Altonji, J. G., & Dunn, T. A. (1991). Relationships among the family incomes and labor market outcomes of relatives (No. 3724). National Bureau of Economic Research.

Altonji, J. G., & Dunn, T. A. (2000). An intergenerational model of wages, hours, and earnings. *Journal of Human Resources*: 221-258.

Becker, G.S., & Tomes, N. (1979) An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy*, 87: 1153-1189.

Björklund, A., & Jäntti, M. (2009). Intergenerational Income Mobility and the Role of Family Background, in (W. Salverda, B. Nolan and T. Smeeding eds.) Oxford Handbook of Economic Inequality, Oxford University Press.

Björklund, A., Jäntti, M., & Nybom, M. (forthcoming). The contribution of early-life vs. labour-market factors to intergenerational income persistence: a comparison of the UK and Sweden. *Economic Journal*.

Black, D. A., & Smith, J. A. (2006). Estimating the returns to college quality with multiple proxies for quality. *Journal of Labor Economics*, 24(3): 701-728.

Black, S. E. & Devereux, P. J. (2011). Recent developments in intergenerational mobility, in (O. Ashenfelter and D. Card, eds.), *Handbook of Labor Economics*, Vol. 4: 1487-1541, Amsterdam: Elsevier.

Bollen, K. A. (1989). *Structural Equations With Latent Variables*. Wiley, New York, NY.

Braun, S., & Stuhler, J. (2015). *The Transmission of Inequality Across Multiple Generations: Testing Recent Theories with Evidence from Germany*. Mimeo.

Chadwick, L., & Solon, G. (2002). Intergenerational income mobility among daughters. *American Economic Review*, 335-344.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States. *Quarterly Journal of Economics*, 129(4), 1553-1623.

Clark, G. (2014). *The son also rises: surnames and the history of social mobility*. Princeton University Press.

Clark, G. & Cummins, N. (2015). Intergenerational Wealth Mobility in England, 1858–2012: Surnames and Social Mobility. *The Economic Journal*, 125(582): 61-85.

Corak, M. & Piraino, P. (2011). The Intergenerational Transmission of Employers. *Journal of Labor Economics*, 29(1): 37-68.

Güell, M., Rodríguez Mora, J. V. & Telmer, C. (2015). The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating. *Review of Economic Studies*, 82(2): 693-735.

Goldberger, A.S. (1989). Economic and mechanical models of intergenerational transmission. *American Economic Review*, 79(3): 504-513.

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

Hirvonen, L. H. (2008). Intergenerational earnings mobility among daughters and sons: Evidence from Sweden and a comparison with the United States. *American Journal of Economics and Sociology*, 67(5): 777-826.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351a): 631-639.

Lubotsky, D. & Wittenberg, M. (2006). Interpretation of regressions with multiple proxies. *The Review of Economics and Statistics*, 88(3): 549-562.

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

Nakosteen, R. A., Westerlund, O., & Zimmer, M. A. (2004). Marital Matching and Earnings: Evidence from the Unmarried Population in Sweden. *Journal of Human Resources*, 39(4): 1033-1044.

Nybohm & Stuhler (*forthcoming*). Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation. *Journal of Human Resources*.

Raaum, O., Bratsberg, B., Røed, K., Österbacka, E., Eriksson, T., Jäntti, M., & Naylor, R.A. (2007). Marital Sorting, Household Labor Supply, and Intergenerational Earnings Mobility Across Countries. *The BE Journal of Economic Analysis & Policy*, 7(2).

Solon, G. (1999). 'Intergenerational mobility in the labor market', in (O. Ashenfelter and D. Card, eds.), *Handbook of Labor Economics*, Vol. 3A: 1761-1800, Amsterdam: North-Holland.

Solon, G. (2002). Cross-country differences in intergenerational earnings mobility. *The Journal of Economic Perspectives*, 16(3): 59-66.

Solon, G. (2004). A Model of Intergenerational Mobility Variation over Time and Place. In M. Corak, *Generational Income Mobility in North America and Europe* (pp. 38-47). Cambridge University Press.

Statistics Sweden (2004). Jämförelse mellan yrkesuppgifter i FoB och yrkesregistret.

URL: <http://www.scb.se/sv/Hitta-statistik/Statistik-efter-amne/Arbetsmarknad/Sysselsattning-forvarvsarbete-och-arbetstider/Yrkesregistret-med-yrkesstatistik/59068/Jamforelser-mot-aldre-yrkesstatistik/> (Accessed: 2015-05-28)

Vosters, K. (2015). Is the Simple Law of Mobility Really a Law? Testing Clark's Hypothesis.
Unpublished manuscript.

Table 1 - Summary statistics for full sample and U.S. comparison sample

Variable	Full sample		U.S. comparison sample	
	mean	std dev	mean	std dev
Offspring				
Year of birth	1956	3.14	1956	3.12
Average income, age 27-46	250,584	198,680	253,318	201,623
Average log income, age 27-46	12.22	0.53	12.25	0.50
Non-missing incomes, age 27-46	19.52	1.81	19.66	1.41
Log income in 1991			12.23	0.70
Years of education	11.79	2.40	11.83	2.41
Number of offspring (N)	167,552		146,783	
Fathers				
Age when offspring born	30.05	5.13	30.22	5.06
Year of birth	1926	6.53	1925	6.29
Average income, age 30-60	245,013	166,478	249,144	164,416
Average log income, age 30-60	12.26	0.48	12.29	0.45
Non-missing incomes, age 30-60	17.78	6.56	17.71	6.39
Average log income 1968-72			12.28	0.48
Years of education	9.14	2.88	9.15	2.91
Educational attainment (years)				
< 9 years of primary school	0.58	0.49	0.58	0.49
9 years of primary school	0.04	0.21	0.04	0.20
2-year secondary school	0.18	0.38	0.17	0.38
3-year secondary school	0.11	0.31	0.11	0.31
< 3 years of post-secondary school	0.03	0.17	0.03	0.17
3+ years of post-secondary school	0.06	0.23	0.06	0.24
Graduate school	0.01	0.08	0.01	0.08
Occupation category				
1. Professional work (arts & sciences)	0.19	0.40	0.20	0.40
2. Managerial work	0.04	0.21	0.05	0.21
3. Clerical work	0.04	0.19	0.04	0.19
4. Wholesale, retail, & commerce	0.08	0.28	0.08	0.27
5. Agriculture, forestry, hunting, & fishing	0.10	0.30	0.10	0.30
6. Mining & quarrying	0.01	0.08	0.01	0.08
7. Transportation & communication	0.09	0.29	0.09	0.29
8. Manufacturing	0.38	0.48	0.38	0.49
9. Services	0.04	0.20	0.04	0.19
10. Military / armed forces	0.01	0.10	0.01	0.10
Undefined	<0.00	0.01	0.00	0.01
Missing	0.02	0.13	0.01	0.10
Number of fathers (N)	153,920		135,020	

Notes: The main sample is the full sample used for our main analysis as well as robustness checks. The U.S. comparison sample is the subset that has the income measures needed to compute the PSID comparable income measures (average of log income in years 1968-72 for fathers and, for sons, log income in 1991).

Table 2 - OLS, IV, and LW estimates for full sample (fathers and sons)

	[1]	[2]	[3]	[4]
OLS estimates				
Fathers' log average income	0.231 <i>0.003</i>	0.230 <i>0.004</i>	0.208 <i>0.004</i>	0.207 <i>0.004</i>
Fathers' years of education		0.000 <i>0.001</i>		0.000 <i>0.001</i>
Fathers' unique occupation (indicators)			x	x
IV estimates				
First stage (educ. IV for log income)	0.083 <i>0.000</i>			
Second stage	0.235 <i>0.006</i>			
LW estimates of the IGE				
B	0.231 <i>0.004</i>	0.231 <i>0.004</i>	0.260 <i>0.004</i>	0.260 <i>0.004</i>
Observations (N)	167,552	167,552	167,552	167,552

Notes: All specifications use the average of sons' log income as the dependent variable and include birth-year dummies of fathers and sons as controls. The noisy measures of status for fathers included in each model are: [1] income; [2] income and education; [3] income and occupation; [4] income, education and occupation. Because the occupation measure is 270 unique occupation categories, the OLS coefficients and standard errors for occupations are omitted from the table. OLS and IV standard errors are clustered by family and LW standard errors are computed using a block bootstrap to account for within-family correlation (100 repetitions).

Table 3 - Comparison of LW estimates for Sweden with those for the U.S.

	N	[1]	[2]	[3]	[4]
Sweden estimates					
Main results (full sample)	167,552	0.231 <i>0.004</i>	0.231 <i>0.004</i>	0.260 <i>0.004</i>	0.260 <i>0.004</i>
Main specifications for restricted sample used in U.S. comparable specification	146,783	0.231 <i>0.003</i>	0.231 <i>0.003</i>	0.262 <i>0.004</i>	0.262 <i>0.004</i>
Sweden estimates using U.S. (PSID) comparable specification	146,783	0.194 <i>0.004</i>	0.194 <i>0.004</i>	0.215 <i>0.005</i>	0.215 <i>0.005</i>
U.S. estimates (from Vosters, 2014)	415	0.439 <i>0.075</i>	0.445 <i>0.072</i>	0.465 <i>0.080</i>	0.473 <i>0.080</i>

Notes: The noisy measures of status for fathers included in each model are: [1] income; [2] income and education; [3] income and occupation; [4] income, education and occupation. The main specifications use the average of sons' log income (age 27-46) as the dependent variable, average of log income (age 30-60) for father's income, unique occupation indicators for fathers' occupation, and include birth-year dummies of fathers and sons as controls. The PSID-comparable measures are: sons' log income in 1991; father's average log income 1968-1972; indicators for fathers' major occupation category. All specifications use years of education as the measure of educational attainment. LW standard errors are computed using a block bootstrap to account for within-family correlation (100 repetitions).

Table 4 - Robustness of LW estimates to construction of status measures (fathers and sons)

	N	[1]	[2]	[3]	[4]
Main results	167,552	0.231 <i>0.004</i>	0.231 <i>0.004</i>	0.260 <i>0.004</i>	0.260 <i>0.004</i>
Adjusting the occupation measure					
Indicators for minor occupation (2-digit)	167,552	0.231 <i>0.004</i>	0.231 <i>0.004</i>	0.247 <i>0.004</i>	0.247 <i>0.004</i>
Indicators for major occupation	167,552	0.231 <i>0.004</i>	0.231 <i>0.004</i>	0.238 <i>0.004</i>	0.238 <i>0.004</i>
Excluding “undefined” and missing	164,678	0.235 <i>0.003</i>	0.235 <i>0.003</i>	0.265 <i>0.004</i>	0.265 <i>0.004</i>
Adjusting the education measure					
Indicators for each education level	167,552	0.231 <i>0.004</i>	0.233 <i>0.004</i>	0.260 <i>0.004</i>	0.261 <i>0.004</i>
Indicators for level/type of attainment	167,552	0.231 <i>0.004</i>	0.241 <i>0.004</i>	0.260 <i>0.004</i>	0.265 <i>0.004</i>
Adjusting the income measure					
Log (average annual income)	167,550	0.270 <i>0.003</i>	0.274 <i>0.003</i>	0.296 <i>0.003</i>	0.297 <i>0.003</i>
Separate log annual income measures, age 40-50	57,728	0.247 <i>0.008</i>	0.247 <i>0.008</i>	0.304 <i>0.009</i>	0.304 <i>0.010</i>
<i>Using average of log annual income, age 40-50</i>	57,728	0.241 <i>0.008</i>	0.241 <i>0.008</i>	0.298 <i>0.009</i>	0.298 <i>0.009</i>
<i>Main specification using this restricted sample</i>	57,728	0.279 <i>0.007</i>	0.279 <i>0.007</i>	0.333 <i>0.009</i>	0.333 <i>0.009</i>

Notes: All specifications use the average of sons' log income as the dependent variable and include birth-year dummies of fathers and sons as controls. The noisy measures of status for fathers included in each model are: [1] income; [2] income and education; [3] income and occupation; [4] income, education and occupation. LW standard errors are computed using a block bootstrap to account for within-family correlation (100 repetitions).

Table 5 - OLS estimates from extensions with mothers' measures of status, for all child-parent samples

	Sons				Daughters			
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Fathers' measures								
Log average income	0.231 <i>0.003</i>	0.229 <i>0.004</i>	0.208 <i>0.004</i>	0.207 <i>0.005</i>	0.152 <i>0.003</i>	0.128 <i>0.004</i>	0.130 <i>0.004</i>	0.123 <i>0.004</i>
Education		0.001 <i>0.001</i>		0.001 <i>0.001</i>		0.008 <i>0.001</i>		0.006 <i>0.001</i>
Fathers' & Mothers' measures								
Fathers' log avg. income	0.225 <i>0.003</i>	0.227 <i>0.004</i>	0.203 <i>0.005</i>	0.203 <i>0.005</i>	0.142 <i>0.003</i>	0.125 <i>0.004</i>	0.125 <i>0.004</i>	0.120 <i>0.004</i>
Mothers' log avg. income	0.024 <i>0.002</i>	0.023 <i>0.002</i>	0.010 <i>0.002</i>	0.009 <i>0.002</i>	0.059 <i>0.002</i>	0.053 <i>0.002</i>	0.048 <i>0.003</i>	0.046 <i>0.003</i>
Fathers' education		-0.001 <i>0.001</i>		-0.001 <i>0.001</i>		0.003 <i>0.001</i>		0.002 <i>0.001</i>
Mothers' education		0.002 <i>0.001</i>		0.003 <i>0.001</i>		0.005 <i>0.001</i>		0.005 <i>0.001</i>
Mothers' measures								
Log average income	0.034 <i>0.002</i>	0.021 <i>0.002</i>	0.002 <i>0.002</i>	-0.001 <i>0.002</i>	0.064 <i>0.002</i>	0.052 <i>0.002</i>	0.043 <i>0.003</i>	0.040 <i>0.003</i>
Education		0.016 <i>0.001</i>		0.011 <i>0.001</i>		0.015 <i>0.001</i>		0.012 <i>0.001</i>
Observations (N)	152,486	152,486	152,486	152,486	145,256	145,256	145,256	145,256

Notes: All specifications use the average of sons' or daughters' log income as the dependent variable and include birth-year dummies of included parents and offspring as controls. The noisy measures of status for parents included in each model are: [1], [5] income; [2], [6] income and education; [3], [7] income and occupation; [4], [8] income, education and occupation. Because the occupation measure is 270 unique occupation categories, the OLS coefficients and standard errors for occupations are omitted from the table. Standard errors are clustered by family to account for within-family correlation.

Table 6 - LW estimates from extensions with mothers' measures of status, for all child-parent samples

	Sons				Daughters			
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Fathers	0.231 <i>0.003</i>	0.231 <i>0.003</i>	0.262 <i>0.004</i>	0.262 <i>0.004</i>	0.152 <i>0.003</i>	0.163 <i>0.003</i>	0.188 <i>0.004</i>	0.193 <i>0.004</i>
Mothers	0.034 <i>0.002</i>	0.096 <i>0.006</i>	0.235 <i>0.014</i>	0.252 <i>0.015</i>	0.064 <i>0.002</i>	0.096 <i>0.003</i>	0.132 <i>0.004</i>	0.142 <i>0.004</i>
Fathers & Mothers	0.234 <i>0.003</i>	0.235 <i>0.003</i>	0.283 <i>0.004</i>	0.283 <i>0.004</i>	0.209 <i>0.004</i>	0.218 <i>0.005</i>	0.273 <i>0.006</i>	0.276 <i>0.006</i>
Fathers' portion	0.225 96%	0.222 95%	0.242 85%	0.239 84%	0.142 68%	0.140 64%	0.160 59%	0.157 57%
Mothers' portion	0.009 4%	0.012 5%	0.041 15%	0.044 16%	0.066 32%	0.078 36%	0.113 41%	0.118 43%
Observations (N)	152,486	152,486	152,486	152,486	145,256	145,256	145,256	145,256

Notes: These estimation samples have non-missing data on all measures for mothers and fathers. All specifications use the average of sons' or daughters' log income as the dependent variable and include birth-year dummies of included parents and offspring as controls. The noisy measures of status for parents included in each model are: [1], [5] income; [2], [6] income and education; [3], [7] income and occupation; [4], [8] income, education and occupation. Because the occupation measure is 270 unique occupation categories, the OLS coefficients and standard errors for occupations are omitted from the table. Standard errors are computed using a block bootstrap to account for within-family correlation (100 repetitions).

Table A1 - Summary statistics for mothers & fathers (balanced samples)

Variable	mean	std dev	mean	std dev	mean	std dev	mean	std dev
Offspring	<i>Sons</i>				<i>Daughters</i>			
Year of birth	1956	3.13			1956	3.12		
Average income, age 27-46	252,496	203,293			169,955	78,675		
Average log income, age 27-46	12.23	0.53			11.84	0.54		
Non-missing incomes, age 27-46	19.52	1.79			19.57	1.66		
Years of education	11.81	2.39			12.21	2.30		
Number of offspring (N)	152,486				142,020			
Parents	<i>Fathers</i>		<i>Mothers</i>		<i>Fathers</i>		<i>Mothers</i>	
Age when offspring born	29.93	5.12			29.92	5.13		
Year of birth	1926	6.38	1929	6.16	1926	6.38	1929	6.16
Average income, age 30-60	245,518	161,788	121,359	65,144	245,259	182,143	121,419	64,393
Average log income, age 30-60	12.26	0.48	11.38	0.78	12.26	0.48	11.39	0.78
Non-missing incomes, age 30-60	18.22	6.45	19.49	6.60	18.22	6.45	19.53	6.61
Years of education	9.18	2.90	8.53	2.36	9.17	2.89	8.53	2.37
Educational attainment (years)								
< 9 years of primary school	0.57	0.49	0.63	0.48	0.57	0.49	0.63	0.48
9 years of primary school	0.04	0.21	0.11	0.31	0.04	0.20	0.10	0.31
2-year secondary school	0.18	0.38	0.18	0.38	0.18	0.39	0.18	0.38
3-year secondary school	0.11	0.31	0.02	0.14	0.11	0.31	0.02	0.14
< 3 years of post-secondary school	0.03	0.17	0.03	0.17	0.03	0.17	0.03	0.17
3+ years of post-secondary school	0.06	0.24	0.03	0.18	0.06	0.24	0.03	0.18
Graduate school	0.01	0.08	0.00	0.02	0.01	0.08	0.00	0.02
Occupation category								
1. Professional work	0.20	0.40	0.18	0.38	0.20	0.40	0.18	0.38
2. Managerial work	0.04	0.21	0.01	0.08	0.04	0.20	0.01	0.08
3. Clerical work	0.04	0.19	0.16	0.36	0.04	0.19	0.16	0.36
4. Wholesale, retail, & commerce	0.09	0.28	0.11	0.31	0.09	0.28	0.11	0.32
5. Agriculture, forestry, hunting, fishing	0.09	0.29	0.06	0.23	0.09	0.29	0.06	0.23
6. Mining & quarrying	0.01	0.08	0.00	0.02	0.01	0.08	0.00	0.02
7. Transportation & communication	0.09	0.29	0.04	0.19	0.09	0.29	0.04	0.19
8. Manufacturing	0.38	0.48	0.09	0.29	0.38	0.49	0.10	0.29
9. Services	0.04	0.20	0.26	0.44	0.04	0.20	0.26	0.44
10. Military / armed forces	0.01	0.10	0.00	0.00	0.01	0.10	0.00	0.00
Undefined	0.00	0.01	0.00	0.04	0.00	0.02	0.00	0.04
Missing	0.02	0.13	0.10	0.30	0.02	0.13	0.10	0.29
Number of parents (N)	140,052		140,234		133,884		134,108	

Table A2 - OLS estimates from extensions with mothers' measures of status, for all child-parent samples

	Sons				Daughters			
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Fathers' measures								
Log average income	0.231 <i>0.003</i>	0.230 <i>0.004</i>	0.208 <i>0.004</i>	0.207 <i>0.004</i>	0.153 <i>0.003</i>	0.129 <i>0.004</i>	0.130 <i>0.004</i>	0.122 <i>0.004</i>
Education		0.000 <i>0.001</i>		0.000 <i>0.001</i>		0.008 <i>0.001</i>		0.006 <i>0.001</i>
Observations (N)	167,552	167,552	167,552	167,552	159,172	159,172	159,172	159,172
Fathers' & Mothers' measures								
Fathers' log avg. income	0.225 <i>0.003</i>	0.227 <i>0.004</i>	0.203 <i>0.005</i>	0.203 <i>0.005</i>	0.142 <i>0.003</i>	0.125 <i>0.004</i>	0.125 <i>0.004</i>	0.120 <i>0.004</i>
Mothers' log avg. income	0.024 <i>0.002</i>	0.023 <i>0.002</i>	0.010 <i>0.002</i>	0.009 <i>0.002</i>	0.059 <i>0.002</i>	0.053 <i>0.002</i>	0.048 <i>0.003</i>	0.046 <i>0.003</i>
Fathers' education		-0.001 <i>0.001</i>		-0.001 <i>0.001</i>		0.003 <i>0.001</i>		0.002 <i>0.001</i>
Mothers' education		0.002 <i>0.001</i>		0.003 <i>0.001</i>		0.005 <i>0.001</i>		0.005 <i>0.001</i>
Observations (N)	152,486	152,486	152,486	152,486	145,256	145,256	145,256	145,256
Mothers' measures								
Log average income	0.032 <i>0.002</i>	0.019 <i>0.002</i>	-0.002 <i>0.002</i>	-0.005 <i>0.002</i>	0.062 <i>0.002</i>	0.049 <i>0.002</i>	0.039 <i>0.002</i>	0.036 <i>0.002</i>
Education		0.016 <i>0.001</i>		0.010 <i>0.001</i>		0.015 <i>0.001</i>		0.012 <i>0.001</i>
Observations (N)	173,608	173,608	173,608	173,608	165,161	165,161	165,161	165,161

Notes: All specifications use the average of sons' or daughters' log income as the dependent variable and include birth-year dummies of included parents and offspring as controls. The noisy measures of status for parents included in each model are: [1], [5] income; [2], [6] income and education; [3], [7] income and occupation; [4], [8] income, education and occupation. Because the occupation measure is 270 unique occupation categories, the OLS coefficients and standard errors for occupations are omitted from the table. Standard errors are clustered by family to account for within-family correlation.

Table A3 - LW estimates from extensions with mothers' measures of status, for all child-parent samples

	Sons				Daughters			
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Fathers	0.231 <i>0.004</i>	0.231 <i>0.004</i>	0.260 <i>0.004</i>	0.260 <i>0.004</i>	0.153 <i>0.004</i>	0.164 <i>0.004</i>	0.190 <i>0.005</i>	0.194 <i>0.005</i>
Observations (N)	167,552	167,552	167,552	167,552	159,172	159,172	159,172	159,172
Mothers	0.032 <i>0.003</i>	0.098 <i>0.007</i>	0.246 <i>0.012</i>	0.263 <i>0.012</i>	0.062 <i>0.003</i>	0.049 <i>0.004</i>	0.039 <i>0.006</i>	0.036 <i>0.007</i>
Observations (N)	173,608	173,608	173,608	173,608	165,161	165,161	165,161	165,161

Notes: All specifications use the average of sons' or daughters' log income as the dependent variable and include birth-year dummies of included parents and offspring as controls. The noisy measures of status for parents included in each model are: [1], [5] income; [2], [6] income and education; [3], [7] income and occupation; [4], [8] income, education and occupation. Because the occupation measure is 270 unique occupation categories, the OLS coefficients and standard errors for occupations are omitted from the table. Standard errors are computed using a block bootstrap to account for within-family correlation (100 repetitions).