Appendix

A1 Derivations for Section 2.1

We can solve for the model's parameters using its implied covariances and variances listed in equations (2a)-(2g). One should note that, as in Atkinson and Jenkins (1984), this simultaneous-equations model is overidentified – we have five unknowns and seven equations – and we thus have multiple solutions to the system of equations. We proceed by combining equations (2e) and (2f) which gives:

$$\gamma = \frac{\sigma_{y_t\ 1 y_t} \quad \rho\lambda\sigma_{y_t\ 1 e_t\ 1}}{\sigma_{y_t\ 1 y_t\ 1}} = \frac{\sigma_{y_t\ 1 y_t} \quad \rho\sigma_{y_t\ 1 e_t}}{\sigma_{y_t\ 1 y_t\ 1}}. \tag{A1}$$

Using equations (2a) and(2c) gives:

$$\rho = \frac{\sigma_{e_t y_t} \quad \gamma\lambda\sigma_{y_t\ 1 e_t\ 1}}{\sigma_{e_t e_t}} = \frac{\sigma_{e_t y_t} \quad \gamma\sigma_{y_t\ 1 e_t}}{\sigma_{e_t e_t}}, \tag{A2}$$

where we again used equation (2f) in the second step. We can then use these last two expressions (equations (A1) and (A2)) to write γ and ρ in terms of observable population moments:

$$\gamma = \frac{\sigma_{y_t\ 1 y_t}\sigma_{e_t e_t} \quad \sigma_{y_t e_t}\sigma_{y_t\ 1 e_t}}{\sigma_{y_t\ 1 y_t\ 1}\sigma_{e_t e_t} \quad \sigma^2_{y_t\ 1 e_t}} \tag{A3}$$

$$\rho = \frac{\sigma_{y_t\ 1 y_t\ 1}\sigma_{e_t y_t} \quad \sigma_{y_t\ 1 e_t}\sigma_{y_t\ 1 y_t}}{\sigma_{y_t\ 1 y_t\ 1}\sigma_{e_t e_t} \quad \sigma^2_{e_t y_t\ 1}}. \tag{A4}$$

So far these expressions are identical to those in Atkinson and Jenkins (1984), despite our slightly different and simpler model. In addition, we get our corresponding expression for λ directly from equation (2f). One can then express these in terms of correlations rather than covariances:

$$\lambda = \frac{\sigma_{y_t\ 1 e_t}}{\sigma_{e_t\ 1 y_t\ 1}} = \frac{r_{y_t\ 1 e_t}}{r_{e_t\ 1 y_t\ 1}} \left(\frac{\sigma_{e_t}}{\sigma_{e_t\ 1}}\right) \tag{A5}$$

$$\rho = \frac{\sigma_{y_t\ 1 y_t\ 1}\sigma_{e_t y_t} \quad \sigma_{y_t\ 1 e_t}\sigma_{y_t\ 1 y_t}}{\sigma_{y_t\ 1 y_t\ 1}\sigma_{e_t e_t} \quad \sigma^2_{e_t y_t\ 1}} = \frac{r_{y_t e_t} \quad r_{y_t y_t\ 1} r_{y_t\ 1 e_t}}{1 \quad r^2_{y_t\ 1 e_t}} \left(\frac{\sigma_{y_t}}{\sigma_{e_t}}\right) \tag{A6}$$

$$\gamma = \frac{\sigma_{y_t\ 1 y_t}\sigma_{e_t e_t} \quad \sigma_{y_t e_t}\sigma_{y_t\ 1 e_t}}{\sigma_{y_t\ 1 y_t\ 1}\sigma_{e_t e_t} \quad \sigma^2_{e_t y_t\ 1}} = \frac{r_{y_t y_t\ 1} \quad r_{y_t e_t} r_{y_t\ 1 e_t}}{1 \quad r^2_{y_t\ 1 e_t}} \left(\frac{\sigma_{y_t}}{\sigma_{y_t\ 1}}\right), \tag{A7}$$

where we used the notation r_{XY} for the correlation between two variables X and Y, and σ_X for the standard deviation of X. Our focus is on the case in which parental income y_{t-1} is unobserved. We restrict our attention to γ and ρ again and note that these depend on three moments involving y_{t-1}. But since this model (as the one in Atkinson and Jenkins (1984)) is overidentified, one can use the previously unused correlation in equation (2g) to substitute out $r_{y_t y_t\ 1}$:

$$\rho = \frac{r_{y_t e_t} r_{e_t \ 1 y_t \ 1} \quad r_{y_t e_t \ 1} r_{e_t y_t \ 1}}{r_{e_t \ 1 y_t \ 1} \quad r_{y_t \ 1 e_t} r_{e_t \ 1 e_t}} \left(\frac{\sigma_{y_t}}{\sigma_{e_t}} \right) \tag{A8}$$

$$\gamma = \frac{r_{y_t e_t \ 1} \quad r_{y_t e_t} r_{e_t \ 1 e_t}}{r_{e_t \ 1 y_t \ 1} \quad r_{y_t \ 1 e_t} r_{e_t \ 1 e_t}} \left(\frac{\sigma_{y_t}}{\sigma_{y_t \ 1}} \right) \tag{A9}$$

However, we have yet another unused equation in the case of our simplified model. We can therefore also substitute out $r_{y_t \ 1 e_t}$, using $\lambda = r_{y_t \ 1 e_t} = r_{e_t \ 1 e_t} r_{e_t \ 1 y_t \ 1}$, which we get from equations (2d) and (2f). We now have:

$$\rho = \frac{r_{y_t e_t} \quad r_{y_t e_t \ 1} r_{e_t \ 1 e_t}}{1 \quad r^2_{e_t \ 1 e_t}} \left(\frac{\sigma_{y_t}}{\sigma_{e_t}} \right) \tag{A10}$$

$$\gamma = \frac{r_{y_t e_t \ 1} \quad r_{y_t e_t} r_{e_t \ 1 e_t}}{r_{e_t \ 1 y_t \ 1}(1 \quad r^2_{e_t \ 1 e_t})} \left(\frac{\sigma_{y_t}}{\sigma_{y_t \ 1}} \right), \tag{A11}$$

where ρ does not depend on the unobserved y_{t-1} at all, while γ depends on two moments involving y_{t-1}. However, the counterpart of those moments can be observed for generation t and it is therefore possible to adopt a couple of steady-state assumptions to ensure identfication. In particular, we might be willing to assume that $r_{y_t,e_t} = r_{y_t \ 1, e_t \ 1} = r^*_{y,e}$ and $\sigma_{y_t} = \sigma_{y_t \ 1} = \sigma^*_y$, i.e. that the intragenerational income-education correlation and the standard deviation of income are both constant across generations. Under these assumptions, we can take as the steady-state estimator:

$$\gamma_{ss} = \frac{r_{y_t e_t \ 1} \quad r^*_{ye} r_{e_t \ 1 e_t}}{r^*_{ye}(1 \quad r^2_{e_t \ 1 e_t})}. \tag{A12}$$

Thus, under these steady-state assumptions we are capable of identifying both γ and ρ. Note that the identification of λ follows directly from equation (2d), such that

$$\lambda = \frac{\sigma_{e_t e_t \ 1}}{\sigma_{e_t \ 1 e_t \ 1}} = r_{e_t \ 1 e_t} \left(\frac{\sigma_{e_t}}{\sigma_{e_t \ 1}} \right) \tag{A13}$$

and, with the additional (but in this case unnecessary) steady-state assumption that $\sigma_{e_t} = \sigma_{e_t \ 1}$, this reduces to $\lambda = r_{e_t \ 1 e_t}$.

A2 Data Description

Our empirical illustrations are based on a 35 percent random sample of the Swedish population born between 1932 and 1967. Using information based on population registers, we can link these sampled individuals to their biological parents and children. We then individually match data on demographic characteristics based on bi-decennial censuses starting from 1960, as well as education and income data stemming from official registers. For males born 1951-1980 we also match data on cognitive and noncognitive test scores from the mandatory military enlistment held by the Swedish War Archives.

Educational registers were compiled in 1970, 1990 and about every third year thereafter, containing detailed information on each individual's highest educational attainment. We consider the highest schooling level recorded across these years, and translate it into years of education, with 7 years for the old compulsory school being the minimum, and 20 years for a doctoral degree the maximum. Education data in 1970 is available only for those born 1911 and later. As the data are collected from official registers there are no standard non-response problems.

The income data stem from official tax declaration files and are held by Statistics Sweden. We construct measures of long-run income based on age-specific averages of annual incomes, which are observed for the years 1968-2007. We use total (pre-tax) income, which is the sum of an individual's labor (and labor-related) earnings, early-age pensions, and net income from business and capital realizations. We adjust all incomes for inflation using the CPI. Incomes for parents are necessarily measured at a later age than incomes for their offspring, which may bias estimates. Specifically, we construct five-year averages of annual incomes measured between age 45 and 49 for fathers, and between age 30 and 34 for sons. To limit the influence of outliers, we winsorize these averages at the 1st and 99th percentile of their respective birth cohort.

To these data we add military enlistment test scores. Complete information from the draft is available for males who were drafted between 1969 and 2000. During these years, almost all males went through the draft procedure at age 18 or 19, and enlistment scores are available for 88-95 percent of each birth cohort. The data include an overall measure of cognitive skill and a corresponding measure of overall non-cognitive skill. The overall cognitive score is based on four sub-tests measuring: inductive skill (or reasoning); verbal comprehension; spatial ability; and technical understanding. Overall cognitive skill is reported on an integer Stanine scale, which varies from one to nine.[33] The evaluation of non-cognitive ability is based on a procedure that was adopted in 1969 and consists of a 25-minute interview with a certified psychologist. As a basis for the interview, the psychologist had information on the results and the cognitive tests, the results on various physical tests, school grades, and answers from a questionnaire about families, friends and hobbies. The interview as such was centered around a number of pre-specified behavioral topics. Based on the interview, the draftee gets an overall score on an integer Stanine scale. The overall score reflects social maturity, psychological energy (e.g., focus and perseverance), intensity (e.g., activation without external pressure), and emotional stability (e.g., tolerance to stress).

For the analysis in Section 2, we consider (log) incomes and years of education for males born 1942-1955 and their fathers (if born after 1911). For the analyses in Section 3 we consider males born between 1951 (the earliest cohort for which cognitive and noncognitive test scores are available) and 1975, as well as the fathers and paternal grandfathers of the 1971 cohort. For the analyses in Section 4 we consider males born between 1955 and 1975, as well as their fathers. To ensure that incomes are observed at the same age range, we consider father-son pairs in which the father was between 25 and 30 years old at the birth of the child. We verified

[33] The Stanines are (approximately) normally distributed with a mean of 5 and a standard deviation of 2.

that this restriction has only a small effect on our estimates in those cohorts in which the age restrictions were not necessary.