Supplementary material to:

The intergenerational transmission of cognitive and non-cognitive abilities

Erik Grönqvist, Björn Öckert and Jonas Vlachos

This online appendix has four sections. In Appendix A, we derive the formal conditions for the two-stage least squares with missing endogenous variables. Appendix B discusses the genetic inheritance of complex traits, such as cognitive abilities and personality. We also simulate the potential bias when using the uncle instrument under different assumptions about genetic inheritance. Appendix C provides a description of international adoptions in Sweden. In Appendix D we present additional results.

Appendix A. Generating missing regressors

This appendix lays out the formal conditions for estimating two-stage least squares with missing endogenous variables. We address the situation in which the missing regressor is unavailable in other samples, but an observed variable of the same *type* can be used to generate the missing regressor. In particular, the observed and unobserved variables are two different *alterations* of the same underlying variable. For example, different training programs (types) for both certified and uncertified teachers (alterations) in a study of school performance, where program participation is only observed for certified teachers. Another example would be pharmaceutical drugs (types) that are administered in both tablet and fluid form (alterations) in a study of infant health, but with consumption data available only for tablets. In our case, both mothers' and fathers' abilities are considered when estimating intergenerational correlations, but only the fathers' abilities are observed.

For each observation we consider the following structural equation:

$$y_i = \boldsymbol{x}_i\boldsymbol{\beta} + \varepsilon_i \,, \tag{1}$$

where $\boldsymbol{x}_i = [\boldsymbol{g}_{1i}\, \boldsymbol{h}_{2i}]$ is a $1 \times (k+l)$ vector of endogenous variables; $\boldsymbol{g}_{1i} = [g_{11i}\, g_{12i} \ldots g_{1ki}]$ is a $1 \times k$ vector of variables of alteration 1; and $\boldsymbol{h}_{2i} = [h_{21i}\, h_{22i} \ldots h_{2li}]$ is a $1 \times l$ vector of alteration 2 variables. $\boldsymbol{\beta} = [\boldsymbol{\beta}_1\, \boldsymbol{\beta}_2]'$ is the $(k+l) \times 1$ parameter vector of interest, and ε_i is an error term. We have the following first-stages:

$$\boldsymbol{g}_{1i} = \boldsymbol{z}_i\boldsymbol{\gamma}_1 + \epsilon_{1i} \tag{2}$$

$$\boldsymbol{h}_{2i} = \boldsymbol{z}_i\boldsymbol{\gamma}_2 + \epsilon_{2i}, \tag{3}$$

where $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are $k \times (k+l)$ and $l \times (k+l)$ matrices of first-stage parameters for the endogenous variables of alteration 1 and 2. $\boldsymbol{z}_i = [\boldsymbol{u}_{1i}\, \boldsymbol{v}_{2i}]$ is the $1 \times (k+l)$ vector of instruments. For each endogenous regressor in $\boldsymbol{g}_{1i}$ and $\boldsymbol{h}_{2i}$, there exists a primary instrument, designated to isolate the exogenous variation in the regressor; $\boldsymbol{u}_{1i} = [\boldsymbol{u}_{11i}\, \boldsymbol{u}_{12i} \ldots \boldsymbol{u}_{1ki}]$ includes k instruments for $\boldsymbol{g}_{1i}$, and $\boldsymbol{v}_{2i} = [\boldsymbol{v}_{21i}\, \boldsymbol{v}_{22i} \ldots \boldsymbol{v}_{2li}]$ contains l instruments for $\boldsymbol{h}_{2i}$.

Let $\boldsymbol{X}$ and $\boldsymbol{Z}$ denote data matrices of dimension $n \times (k + l)$, and let $\mathbf{Z}$ have the standard properties: $\text{plim}(\boldsymbol{Z}'\boldsymbol{\varepsilon}/n) = 0$ and $\text{plim}(\boldsymbol{Z}'\boldsymbol{X}/n) = \boldsymbol{\Sigma}_{ZX}$, where $\boldsymbol{\Sigma}_{ZX}$ is a finite matrix of full rank. When the regressors of both alteration 1 and 2 are observed, the first stage parameter vectors are estimated as $\widehat{\boldsymbol{\gamma}}^{FS} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}$, and the 2SLS estimator is given by:

$$\widehat{\boldsymbol{\beta}}^{2SLS} = \left(\widehat{\boldsymbol{X}}^{FS\prime}\widehat{\boldsymbol{X}}^{FS}\right)^{-1}\widehat{\boldsymbol{X}}^{FS\prime}\boldsymbol{Y}, \tag{4}$$

where $\widehat{\boldsymbol{X}}^{FS} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}$. This is a consistent estimator of $\boldsymbol{\beta}$ under the standard instrumental variable assumptions:

$$\text{plim}\ \widehat{\boldsymbol{\beta}}^{2SLS} = \text{plim}\left(\widehat{\boldsymbol{X}}^{FS\prime}\widehat{\boldsymbol{X}}^{FS}/n\right)^{-1}\widehat{\boldsymbol{X}}^{FS\prime}\boldsymbol{Y}/n,$$

$$= \text{plim}\,[(\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}')(\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X})/n]^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}/n$$

$$= \text{plim}\,[\boldsymbol{X}'\boldsymbol{P}_Z{}'\boldsymbol{P}_Z\boldsymbol{X}/n]^{-1}\boldsymbol{X}'\boldsymbol{P}_Z'\boldsymbol{Y}/n$$

$$= \text{plim}\,[\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{X}/n]^{-1}\boldsymbol{X}'\boldsymbol{P}_Z'\boldsymbol{X}\boldsymbol{\beta}/n + \text{plim}\,[\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{X}/n]^{-1}\boldsymbol{X}'\boldsymbol{P}_Z'\varepsilon/n = \boldsymbol{\beta}, \tag{5}$$

where $\boldsymbol{P}_Z$ is the projection matrix.

Case 1: Homogeneous first stages

Consider the case where $\boldsymbol{g}_{1i}$ is observed but $\boldsymbol{h}_{2i}$ is not. When the first stages are homogeneous across the alterations, $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2$, the unobserved regressors of alteration 2, $\boldsymbol{h}_{2i}$, can be generated using the observed variables of alteration 1. Let $\boldsymbol{h}_{1i} = [h_{11i}\ h_{12i}\ \ldots h_{1li}]$ be a $1 \times l$ vector of regressors of alteration 1 that corresponds to the unobserved regressors of alteration 2. That is, if $\boldsymbol{h}_{2i}$ contains the mother's cognitive skills, $\boldsymbol{h}_{1i}$ contains the father's cognitive skills. For each observation we then have the following *auxiliary stage*:

$$\boldsymbol{h}_{1i} = \boldsymbol{m}_i\boldsymbol{\theta}_1 + w_{1i}, \tag{6}$$

where $\boldsymbol{\theta}_1$ is a $l \times (k + l)$ matrix of auxiliary stage parameters, and $\boldsymbol{m}_i = [\boldsymbol{u}_{2i}\ \boldsymbol{v}_{1i}]$ is a $1 \times (k + l)$ vector of instrumental variables; $\boldsymbol{v}_{1i} = [v_{11i}\ v_{12i}\ \ldots v_{1li}]$ is a $1 \times l$ vector of primary instruments for the regressors in $\boldsymbol{h}_{1i}$, while $\boldsymbol{u}_{2i} = [u_{21i}\ u_{22i}\ \ldots u_{2ki}]$ is a $1 \times k$ vector of instruments for the unobserved regressors in $\boldsymbol{g}_{2i} = [g_{21i}\ g_{22i}\ \ldots g_{2ki}]$. $\boldsymbol{g}_{2i}$ is defined as a

$1 \times k$ vector of variables of alteration 2 that correspond to the variables of alteration 1 in $\boldsymbol{g}_{1i}$. $\boldsymbol{u}_{2i}$ is included for completeness. In our case, $\boldsymbol{v}_{1i}$ could be the paternal uncles' cognitive skills, and $\boldsymbol{u}_{i2}$ the maternal uncles' non-cognitive skills. w_{1i} is a disturbance term.

Let $\boldsymbol{M}$ and $\boldsymbol{H}_1$ be data matrices of dimensions $n \times (k+l)$ and $n \times l$, where $\boldsymbol{M}$ is assumed to have the following properties: $\text{plim}(\boldsymbol{M}'\boldsymbol{\varepsilon}/n) = 0$ and $\text{plim}(\boldsymbol{M}'\boldsymbol{H}_1/n) = \boldsymbol{\Sigma}_{MH_1}$, with $\boldsymbol{\Sigma}_{MH_1}$ being bounded and of full column rank. The auxiliary stage parameters are then estimated as follows:

$$\widehat{\boldsymbol{\theta}}_1^{AS1} = (\boldsymbol{M}'\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{H}_1. \tag{7}$$

Under the assumption that the first stage relation is homogeneous between alterations 1 and 2, the auxiliary stage parameters (7) can be used in place for the parameters in the unobserved first stage (3). The identifying assumption can be written as:

$$\text{plim}\ \widehat{\boldsymbol{\theta}}_1^{AS1} = \text{plim}\ \widehat{\boldsymbol{\gamma}}_2^{FS}. \tag{IA.1}$$

For later purposes, we denote the elements in the o:th row and p:th column of $\widehat{\boldsymbol{\theta}}_1^{AS1}$ by $\hat{\theta}_{1op}^{AS1}$. Under (IA.1), the prediction of the unobserved data can be written as $\widehat{\boldsymbol{X}}^{AS1} = \boldsymbol{Z}\widehat{\boldsymbol{\gamma}}^{AS1}$, where $\widehat{\boldsymbol{\gamma}}^{AS1} = [\widehat{\boldsymbol{\gamma}}_1^{FS}\ \ \widehat{\boldsymbol{\theta}}_1^{AS1}]$ is a $(k+l) \times (k+l)$ matrix of first stage and auxiliary stage parameters. The structural parameters in (1) can then be estimated as:

$$\widehat{\boldsymbol{\beta}}^{ASLS1} = \left(\widehat{\boldsymbol{X}}^{AS1\prime}\widehat{\boldsymbol{X}}^{AS1}\right)^{-1}\widehat{\boldsymbol{X}}^{AS1\prime}\boldsymbol{Y} = (\widehat{\boldsymbol{\gamma}}^{AS1\prime}\boldsymbol{Z}'\boldsymbol{Z}\widehat{\boldsymbol{\gamma}}^{AS1})^{-1}\widehat{\boldsymbol{\gamma}}^{AS1\prime}\boldsymbol{Z}'\boldsymbol{Y}. \tag{8}$$

Assumption (IA.1) implies $\text{plim}\widehat{\boldsymbol{X}}^{AS1} = \text{plim}\widehat{\boldsymbol{X}}^{FS}$, which yields the standard 2SLS estimator:

$$\text{plim}\ \widehat{\boldsymbol{\beta}}^{ASLS1} = \text{plim}\left(\widehat{\boldsymbol{X}}^{AS1\prime}\widehat{\boldsymbol{X}}^{AS1}/n\right)^{-1}\widehat{\boldsymbol{X}}^{AS1\prime}\boldsymbol{Y}/n = \text{plim}\left(\widehat{\boldsymbol{X}}^{FS\prime}\widehat{\boldsymbol{X}}^{FS}/n\right)^{-1}\widehat{\boldsymbol{X}}^{FS\prime}\boldsymbol{Y}/n = \boldsymbol{\beta}, \tag{9}$$

where the last equality follows from equation (5). Hence under the assumption that the first stage is homogeneous, $\widehat{\boldsymbol{\beta}}^{ASLS1}$ is a consistent estimator for $\boldsymbol{\beta}$.

Case 2: Heterogeneous first stages

The assumption of homogeneous first stages can be too strong. In our case, it may seem unlikely that the auxiliary stage parameters (the brother-brother correlations) are equal to the

unobserved first stage parameters (the brother-sister correlations). We therefore propose using proxy variables for the unobserved regressors to adjust for the heterogeneity in the unobserved first stage. To this end we utilize $\widetilde{\boldsymbol{h}}_{1i}$ and $\widetilde{\boldsymbol{h}}_{2i}$ which are $1\times l$ vectors of proxies for $\boldsymbol{h}_{1i}$ and $\boldsymbol{h}_{2i}$, respectively. In our case, the proxy variables are compulsory school grades available for boys and girls. The relationship between the proxy variables and their corresponding instruments are given by a second auxiliary stage:

$$\widetilde{\boldsymbol{h}}_{1i} = \boldsymbol{m}_i\boldsymbol{\Pi}_1 + e_{1i} \tag{10}$$

$$\widetilde{\boldsymbol{h}}_{2i} = \boldsymbol{z}_i\boldsymbol{\Pi}_2 + e_{2i}, \tag{11}$$

where $\boldsymbol{\Pi}_1$ and $\boldsymbol{\Pi}_2$ are $l \times (k + l)$ matrices of second auxiliary stage parameters, while e_{1i} and e_{2i} are error terms. The second auxiliary stage parameters can be estimated as follows:

$$\widehat{\boldsymbol{\Pi}}_1^{AS2} = (\boldsymbol{M}'\boldsymbol{M})^{-1}\boldsymbol{M}'\widetilde{\boldsymbol{H}}_1 \tag{12}$$

$$\widehat{\boldsymbol{\Pi}}_2^{AS2} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\widetilde{\boldsymbol{H}}_2 \tag{13}$$

where $\widetilde{\boldsymbol{H}}_1$ and $\widetilde{\boldsymbol{H}}_2$ denote data matrices of dimension $n \times l$. Assuming that the relative impact of the instruments in the second auxiliary stages for alterations 1 and 2 can be used to infer the relative effects of the first stages for different alterations, we can use the ratio between the second auxiliary stage parameters to adjust the parameters in the first auxiliary stage. Formally, we define $\widehat{\boldsymbol{\theta}}_1^{AS2}$, which is a $l \times (k + l)$ matrix of adjusted first auxiliary stage parameters, where the elements are of the form $\hat{\theta}_{1op}^{AS2} = (\widehat{\Pi}_{2op}^{AS2}/\widehat{\Pi}_{1op}^{AS2})\hat{\theta}_{1op}^{AS1}$. That is, we scale the first auxiliary stage parameters with the ratio between the second auxiliary stage parameters. Hence, we can define the following identifying assumption:

$$\text{plim}\ \widehat{\boldsymbol{\theta}}_1^{AS2} = \text{plim}\ \widehat{\boldsymbol{\gamma}}_2^{FS}. \tag{IA.2}$$

Under (IA.2), the predicted observations for the unobserved regressors are given by $\widehat{\boldsymbol{X}}^{AS2} = \boldsymbol{Z}\widehat{\boldsymbol{\gamma}}^{AS2}$, where $\widehat{\boldsymbol{\gamma}}^{AS2} = [\widehat{\boldsymbol{\gamma}}_1^{FS}\ \ \widehat{\boldsymbol{\theta}}_1^{AS2}]$ is a $(k + l) \times (k + l)$ matrix of first stage and adjusted auxiliary stage parameters. The structural parameters in (1) can then be estimated as:

$$\widehat{\boldsymbol{\beta}}^{ASLS2} = \left(\widehat{\boldsymbol{X}}^{AS2\prime}\widehat{\boldsymbol{X}}^{AS2}\right)^{-1}\widehat{\boldsymbol{X}}^{AS2\prime}\boldsymbol{Y} = (\widehat{\boldsymbol{\gamma}}^{AS2\prime}\boldsymbol{Z}'\boldsymbol{Z}\widehat{\boldsymbol{\gamma}}^{AS2})^{-1}\widehat{\boldsymbol{\gamma}}^{AS2\prime}\boldsymbol{Z}'\boldsymbol{Y}. \quad (14)$$

Assumption (IA.2) implies that $\text{plim}\widehat{\boldsymbol{X}}^{AS1} = \text{plim}\widehat{\boldsymbol{X}}^{FS}$, and $\widehat{\boldsymbol{\beta}}^{ASLS2}$ is a consistent estimator for $\boldsymbol{\beta}$ by equation (5).

The credibility of IA.2

To see when assumption (IA.2) is reasonable, consider the following data-generating process:

$$\tilde{h}_{1ji} = h_{1ji}\kappa_{1j} + n_{1ji} \quad (15)$$

$$\tilde{h}_{2ji} = h_{2ji}\kappa_{2j} + n_{2ji}, \quad (16)$$

where $\tilde{h}_{1ji}$ and $\tilde{h}_{2ji}$ are 1×1 proxy variables for regressor j, while κ_{1j} and κ_{2j} are scalars describing the proxy relation, and $n_{1ji}$and n_{2ji} are error terms. Assuming that $E\left(\boldsymbol{M}'n_{1ji}\right) = 0$ and $E\left(\boldsymbol{Z}'n_{2ji}\right) = 0$, the parameters for regressor j in the second auxiliary stage is given by:

$$\widehat{\Pi}_{1j}^{AS2} = (\boldsymbol{M}'\boldsymbol{M})^{-1}\boldsymbol{M}'\widetilde{\boldsymbol{H}}_{1j} = (\boldsymbol{M}'\boldsymbol{M})^{-1}\boldsymbol{M}'\left(\boldsymbol{H}_{1j}\kappa_{1j} + n_{1j}\right) = (\boldsymbol{M}'\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{H}_{1j}\kappa_{1j} = \hat{\theta}_{1j}^{AS1}\kappa_{1j} \quad (17)$$

$$\widehat{\Pi}_{2j}^{AS2} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\widetilde{\boldsymbol{H}}_{2j} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\left(\boldsymbol{H}_{2j}\kappa_{2j} + n_{2j}\right) = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{H}_{2j}\kappa_{2j} = \hat{\gamma}_{2j}^{FS}\kappa_{2j}, \quad (18)$$

where $\widetilde{\boldsymbol{H}}_{1j}$ and $\widetilde{\boldsymbol{H}}_{2j}$ are $n \times 1$ vectors of observations for $\tilde{h}_{1ji}$ and $\tilde{h}_{2ji}$. Now, consider the elements of $\hat{\theta}_1^{AS2}$ and substitute in elements from equations (17) and (18)

$$\hat{\theta}_{1op}^{AS2} = \left(\widehat{\Pi}_{2op}^{AS2}/\widehat{\Pi}_{1op}^{AS2}\right)\hat{\theta}_{1op}^{AS1} = \left(\hat{\gamma}_{2op}^{FS}\kappa_{2op}/\hat{\theta}_{1op}^{AS1}\kappa_{1op}\right)\hat{\theta}_{1op}^{AS1} = \hat{\gamma}_{2op}^{FS}\left(\kappa_{2op}/\kappa_{1op}\right) \quad (19)$$

Hence, if $\text{plim}\kappa_{1op} = \text{plim}\kappa_{2op}$ then $\text{plim}\hat{\theta}_{1op}^{AS2}$=$\text{plim}\hat{\gamma}_{2op}^{FS}$. For (IA.2) to hold, the proxy relation in equations (15) and (16) must be homogeneous; i.e., the proxy variables must be equally good measures for the regressors of alteration 1 and 2. In our case this translates to GPA's being an equally good proxy of skills for both genders.

Standard errors

The sampling errors of the auxiliary stage estimates must be taken into account when calculating standard errors. Therefore, we chose to bootstrap the standard errors. Our findings suggest that the bootstrapped standard errors are not much larger than when ignoring the

sampling variability. We believe this is due to our large samples in combination with the strong first stage and auxiliary stage relations.

Appendix B. Simulating the Validity of the Uncle Instrument

Using the uncle's abilities as instruments for parental abilities relies on the assumption that there are no direct links between uncles' and nephews' traits (over and above the relation between uncles' and parents' abilities). In principle, uncles may be related to their nephews either through genetic similarity or environmental influence. This appendix assesses the possibility of a direct genetic link between uncles' and nephews' abilities. We discuss the recent evidence for a shared genetic expression and simulate the validity of the uncle instrument under different assumptions of genetic transmission over generations.

To understand how uncles and nephews are genetically related, it is necessary to introduce some basic concepts in genetics. Humans receive one complete copy (allele) of their genetic material from each parent. They can either have two copies of the same allele or one copy each of two different alleles. The allele that an individual has at a given location on the chromosome (locus) is called a genotype. Children receive exactly half of each parent's genotypes. This implies that genes do not skip over generations and that uncles cannot share genotypes with their nephews that are unshared by the parents. Thus, the validity of uncle instrument is not threatened by any direct genetic link between uncles and nephews.

The visible expression of the genotype is called an individual's phenotype. Complex phenotypes, including stature, intelligence and personality, are controlled by many genes and by environmental factors. The simplest model of how genes combine to determine the phenotype assumes that genes interact additively with each other. However, alleles may interact both within loci (dominance) and between loci (epistasis). When inheritance is dominant, only one allele is required for the trait to be observed, and a dominant allele will mask a recessive allele. Likewise, when phenotypes are formed by the interaction of alleles at different loci (epistasis), the alleles of one gene may conceal the expression of alleles at other genes. If all alleles are not expressed to the same extent in the phenotype, inheritance is said

to be non-additive. In principle, there may then be a direct link between uncles' and nephews' phenotypes, even though there is no connection between their genotypes. In practice, the validity of the uncle instrument essentially depends on how much of the genetic variation in humans that are transmitted additively and how much is due to dominance or epistasis.

There has been a long controversy in genetics concerning the relative proportion of additive and non-additive genetic variation for human complex traits. Recent evidence, however, suggests that most of the genetic variation is additive. Although the prevalence of both dominance and epistasis have been documented in genome-wide association studies of human traits (e.g., Hemani, et al, 2014), non-additive effects seem to contribute to only a small fraction of the genetic variation in the population. In a survey of the literature, Wei et al. (2014) concluded that large interaction terms between pairwise SNPs (epistasis) are very unlikely. Further, Zhu et al. (2015) find that dominance variation at common SNPs explains only a small part of phenotypic variation in humans. Consistent with this, twin studies typically find little support for substantial non-additive effects in the heritability of human traits (see, e.g., the survey by Polderman et al, 2015). Thus, the available empirical evidence suggests that most genetic variation is additive, which implies that a direct phenotypic link between uncles and nephews poses less of a threat against the validity of the uncle instrument.

Although the direct phenotypic link between uncles and nephews seems weak, it is still useful to see how different degree of non-additive transmission may affect the validity of the uncle instrument. Therefore, we will simulate how the uncle instrument performs under different assumptions of genetic heritability. It is then important to note that the share of non-additive genetic variation in a population is determined by a number of factors. First, the genetic variability depends on the degree of dominance or epistasis at the gene level. Thus, only in the presence of significant interaction effects at individual loci may there be large non-additive genetic variance in the population. Second, the share of non-additive variation

depends on the distribution of allele frequencies. If dominant alleles are rare, the genetic variation may be predominately additive at the population level, even if there are non-additive effects at the gene level (Hill et al., 2008).

Ideally, we would like to base the simulation on empirical estimates of both the degree of gene interaction and the allele frequencies in humans. Unfortunately, genome-wide studies have not been able to provide credible evidence regarding degree of dominance and epistasis in the formation of complex human traits (see surveys by Henn, et al., 2015 and Wei et al., 2014). A typical finding is that most gene action is additive. In addition, the incidence of gene interaction is low, and the magnitudes of the effects are small. Due to the lack of solid empirical evidence, we will study the validity of the uncle instrument under different distributional assumptions about gene interaction, but where gene transmission is expected to be additive on average. In particular, we let the rate of gene interaction follow one of three distinct distributions: (1) Normal (most gene action is additive), (2) Uniform (different degrees of non-additive interactions are equally likely), and (3) U-shaped distribution (non-additive gene action is frequent). We believe that normally distributed gene interactions are the most realistic assumption, but it is instructive to see how sensitive the validity of the uncle instrument is to more extreme (unlikely) assumptions about dominance and epistasis.

Unlike the degree of gene interaction, there is much more empirical evidence about allele frequencies in populations. In population genetics, allele frequencies indicate the genetic diversity of a population. Selection and genetic drift increase the frequencies of rare alleles, whereas migration and admixture reduce the genetic diversity. Genome-wide studies show that the genetic variability in humans is relatively low and that the majority of human genomic variable sites are rare (e.g., Elhaik, 2012, Gravel et al, 2011 and Henn et al., 2014). Allele frequencies can, thus, best be described by a U-shaped distribution, with high frequencies of alleles in the tails. The low incidence of rare alleles may explain why genetic

variation tends to be predominantly additive in the population, even if there are non-additive effects at the gene level (Hill et al., 2008). Based on the available empirical evidence, we will assume that allele frequency follows a U-shaped distribution in our simulations.

With these basic conditions established, we set up a multi-generational model in which individuals' skills are determined by (the weighted sum of) several genes. The weights are assumed to be normally distributed. Each gene has two alleles, which can take two values; a (a recessive allele) or A (a dominant allele).[1] Genes are assumed to interact pairwise with each other, with positive (synergistic) interaction effects.[2] As discussed above, we assume that the allele frequencies in the population follow a U-shaped distribution, and we let the degree of dominance and epistasis vary.

The simulation model consists of three generations; grandparents, parents and children. In the first step, grandparents' genotypes are generated based on the initial allele frequency distribution, and phenotypes (skills) are formed under different assumptions of the degree of dominance and epistasis. Paternal and maternal grandparents are assumed to mate partly based on observed skills. In the next step, mothers' and fathers' genotypes are generated by randomly drawing one allele per gene from each parent. The genotypes of paternal uncles are generated analogously. The genotypes of fathers, mothers and uncles are then transformed into phenotypes under the same distributional assumptions as in the grandparents' generation, and mothers and fathers are assumed to sort positively on skills. In the final step, children's genotypes are determined by randomly drawing alleles from mothers and fathers, and genes are transformed into phenotypes as for previous generations.

[1] We assume the following genetic values: AA = 1; aa = 0; aA and Aa = 1 – h, where h is the degree of dominance. Under additive inheritance (h = 0.5), Aa and aA equal 0.5.

[2] The positive epistatic effects are set to the weighted average of the genes' phenotypes, where the weights are equal to the degree of epistasis. However, if one gene has two recessive alleles, the other gene solely determines the combined genetic expression.

The simulations are based on 5,000 individuals for each biological relation in every generation. This means that the population consists of 20,000 individuals in the first generation (paternal grandmothers, paternal grandfathers, maternal grandmothers and maternal grandfathers), 15,000 in the second generation (mothers, fathers and paternal uncles) and 5,000 in the last generation (children). For each repetition, we estimate the (true) correlation between fathers' and sons' skills with OLS and IV (where fathers' skills are instrumented with the uncles' skills). For completeness, we also present first-stage and reduced form estimates. Since skills are assumed to be measured without error, the potential bias arising from a direct phenotypic link between uncles and nephews is given by the difference between the OLS- and IV-estimates. This exercise was repeated 300 times, and we present the average and the standard deviation of the point estimates.

Table B1 presents the results from the simulations. The first column shows our preferred specification, in which the degrees of dominance and epistasis are both normally distributed. In this case, the OLS- and IV-estimates are very similar, and we cannot reject that the uncle instrument is valid. An increasing degree of gene-gene interaction in columns (2) and (3) does not change this conclusion. This is consistent with the evidence provided in Mäki-Tanila and Hill (2014). They showed that gene interactions typically only make small contribution to non-additive genetic variation and that the effects of epistasis decline rapidly with the number of genes involved. Because genes are only allowed to interact pairwise in our simulation, it is likely that epistasis would be even less important in a more general model.

In the remaining columns in Table B1, we let the degree of dominance increase substantially. Columns (4) – (6) show the results when different degrees of non-additive interactions are equally likely (uniformly distributed). This increases the potential bias of the IV-estimator slightly. However, the magnitude is generally small (less than 2.5 % of the true estimate) and not significantly different from zero. In the last three columns, we let the gene

transmission be mainly dominant. This leads to a small increase in the bias of the estimates. Even in the extreme case, the potential bias is relatively small (less than 6.5 % of the true estimate). However, it is important to stress that the degree of dominance is likely to be much lower. Thus, we conclude that a direct phenotypic relationship between uncles and nephews is unlikely to invalidate the uncle instrument.

Table B1 Simulation of intergenerational correlations in abilities under different distributional assumptions about allele transmission

Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
OLS	0.648	0.650	0.649	0.632	0.628	0.622	0.592	0.592	0.583
	(0.013)	(0.011)	(0.011)	(0.021)	(0.020)	(0.020)	(0.032)	(0.029)	(0.028)
First stage	0.488	0.489	0.487	0.478	0.480	0.476	0.460	0.462	0.457
	(0.015)	(0.012)	(0.012)	(0.016)	(0.015)	(0.015)	(0.019)	(0.020)	(0.016)
Reduced form	0.321	0.323	0.320	0.310	0.309	0.304	0.284	0.287	0.283
	(0.013)	(0.014)	(0.015)	(0.019)	(0.017)	(0.017)	(0.022)	(0.022)	(0.019)
IV	0.659	0.660	0.657	0.648	0.643	0.639	0.618	0.621	0.619
	(0.020)	(0.021)	(0.022)	(0.027)	(0.024)	(0.027)	(0.033)	(0.029)	(0.029)
Bias	0.011	0.009	0.008	0.016	0.015	0.016	0.026	0.029	0.036
	(0.020)	(0.020)	(0.020)	(0.021)	(0.022)	(0.024)	(0.024)	(0.025)	(0.023)
Distributional assumptions:									
Allele dominance	Normal	Normal	Normal	Uniform	Uniform	Uniform	U-shaped	U-shaped	U-shaped
Allele epistasis	Normal	Uniform	U-shaped	Normal	Uniform	U-shaped	Normal	Uniform	U-shaped

Notes: The table shows simulated intergenerational correlations under different distributional assumptions of allele transmission. Abilities are assumed to be determined by the weighted sum of 200 different genes, with normally distributed weights. Each gene has two alleles (one allele from each parent) which can take two different values (A, a). Alleles are assumed to interact both within loci (dominance) and between loci (epistasis). The degrees of dominance and epistasis are determined by the distributions indicated in the table. Abilities are assumed to be unaffected by environmental factors, and partners are sort partly on observed abilities (assortative mating). All abilities are assumed to be measured without error, and the variables have been standardized. In the IV-estimates, fathers' abilities are instrumented with uncles' abilities. The bias term is calculated as the difference between the IV- and OLS-estimates. The simulations are based on 5,000 observations for each biological relation and generation. Mean values and standard deviations of the point estimates from 300 repetitions are reported.

References for Appendix B

Elhaik, E. (2012), "Empirical Distributions of FST from Large-Scale Human Polymorphism Data", *PLoS One,* Vol. 7, Issue 11.

Gravel, S, B. Henn, R. Gutenkunst, A. Indap, G. Marth, A. Clark, F. Yu, R. Gibbs, The 1000 Genomes Project, and C. Bustamant (2011), "Demographic History and Rare Allele Sharing Among Human Populations", *Proceedings of the National Academy of Sciences of the United States of America* (PNAS), vol. 108 no. 29, pp. 11983–11988.

Hemani, G., K. Shakhbazov, H.-J. Westra, T. Esko, A. Henders, A. McRae, J. Yang, G. Gibson, N. Martin, A. Metspalu, L. Franke, G. Montgomery, P. Visscher, and J. Powell (2014), "Detection and Replication of Epistasis Influencing Transcription in Humans", *Nature*, Vol. 508, pp. 249-256.

Henn, B., L. Botigué, C. Bustamante, A. Clark and S. Gravel (2015), "Estimating the Mutation Load in Human Genomes", *Nature Reviews Genetics*, Vol. 16, pp. 333–343.

Hill, W., M. Goddard, and P. Visscher (2008), "Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits", *PloS Genetics*, Vol. 4, Issue 2, pp. 1-9.

Mäki-Tanila, A., and Hill, G. (2014), "Influence of Gene Interaction on Complex Trait Variation with Multilocus Models", *Genetics*, 198, 355–367.

Polderman, T., B. Benyamin, C. de Leeuw, P. Sullivan, A. van Bochoven, P. Visscher, and D. Posthuma (2015), "Meta-analysis of the Heritability of Human Traits Based on Fifty Years of Twin Studies", *Nature Genetics*, 47, pp. 702–709.

Zhu, Z., A. Bakshi, A. Vinkhuyzen, G. Hemani, S. Hong Lee, I. Nolte, J. van Vliet-Ostaptchouk, H. Snieder, The LifeLines Cohort Study, T. Esko, L. Milani, R. Mägi, A. Metspalu, W. Hill, B. Weir, M. Goddard, P. Visscher, and J. Yang (2015), "Dominance Genetic Variation Contributes Little to the Missing Heritability for Human Complex Traits", *The American Journal of Human Genetics*, Vol. 96, pp. 377–385.

Appendix C. Foreign adoptions to Sweden

During the relevant time period, children were mainly adopted to Sweden from Chile, Colombia, India, Sri Lanka, South Korea and Thailand. The process for foreign adoptions in Sweden has changed slightly over time, but the general features of the process in the 1980s can be described as follows. To become eligible for adoption, parents were screened and approved by the social authorities. The approval stated that the parents could adopt a child within a specified age range. Parents then send their application (including information on their education, social and economic conditions) to a specific country, where the local authorities allocated a child to them. Although this process is a black-box, the scope for matching children to parents based on abilities is limited. Information on the children was essentially limited to birth weight and growth indices 6-9 months prior to adoption, and the background information was also limited, as most children were either left for adoption anonymously or given up by young mothers. It should be noted that the adopting parents could not make specific requests about the child; for example, they could not request a specific sex. When adopting parents were notified about their child, they typically received a picture and some basic health information (i.e., early growth indices). Given the limited scope for matching parents and children based on abilities, and given the limited information that parents had to base a decision to decline a child, the allocation can plausibly be viewed as exogenous. It should also be stressed that the cost for parents of declining a child is potentially high. This description of the adoption process is based on SOU (2003) and on a telephone interview on November 30, 2015, with Birgitta Lönnemar at Adoptionscentrum in Stockholm.

Reference for Appendix C

SOU 2003:49, “Adoption – till vilket pris?”, Fritzes förlag, Stockholm.

Appendix D. Supplementary material

Table D1 Correlation between non-cognitive ability at age 18 and measures of non-cognitive ability at age 13

	Cohort:	
	1953	1967-72
Survey questions for pupils in 6th grade		
"Do you think you do well in school?"		0.092
(Self-confidence)		(0.013)
"Are you often worried about things happening in school?"		-0.022
(Anxiety)		(0.012)
"Do you give up if you get a difficult task in school?"		-0.062
(Persistence)		(0.012)
"Do you like working together with your classmates?"		0.022
(Cooperativeness)		(0.012)
"Do you sometimes bring schoolmates home?"		0.048
(Sociability)		(0.012)
"How often do you spend time with schoolmates after school?"	0.071	
(Peer interaction)	(0.015)	
"Do you often think about other things when you are in school?"		-0.040
(Concentration)		(0.012)
"Do you get disappointed if you get bad results on a test?"		0.084
(Expectations)		(0.012)
"What occupation do you aim for?"	0.103	0.081
(Aspirations)	(0.018)	(0.011)
School administrative data in 6th grade		
Absenteeism		-0.044
		(0.012)
School grades in non-academic subjects	0.131	
	(0.017)	
R^2	0.068	0.115
Number of observations	3,949	6,632

Notes: All variables are standardized by cohort. All models controls for cognitive ability at age 13. Aspiration is measured as the number of years of schooling required for the occupation that the pupil aspires to.

Table D2 OLS- and IV-estimates of intergenerational correlation in abilities – different ability measures

Son's ability:	Cognitive ability	Logical ability	Verbal ability	Spatial ability	Technical ability
			OLS		
Father's ability	0.350	0.292	0.306	0.205	0.281
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
			IV		
Father's ability	0.479	0.417	0.490	0.467	0.483
	(0.009)	(0.009)	(0.011)	(0.017)	(0.012)
Statistic:					
Reliability ratio	0.731	0.700	0.624	0.439	0.582
	(0.016)	(0.018)	(0.016)	(0.018)	(0.017)
Son's ability:	Non-cognitive ability	Social maturity	Intensity	Psychological energy	Emotional stability
			OLS		
Father's ability	0.210	0.156	0.155	0.159	0.136
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
			IV		
Father's ability	0.422	0.424	0.345	0.414	0.387
	(0.015)	(0.020)	(0.019)	(0.021)	(0.021)
Statistic:					
Reliability ratio	0.498	0.368	0.449	0.384	0.351
	(0.021)	(0.021)	(0.029)	(0.023)	(0.023)
N	50,171	50,171	50,171	50,171	50,171

Notes: All estimates come from separate regressions using the same ability measure for father and son. The ability measures have been standardized. Standard errors adjusted for clustering on the father are in parentheses. In the IV-specification, the father's ability has been instrumented with the uncle's ability. The reliability ratios have been calculated by dividing the OLS-estimates by the IV-estimates, and the standard errors have been computed by the delta method.

Table D3 Sibling and self correlations

Dependent variable:	Brother's cognitive ability at age 13	Sister's cognitive ability at age 13	Boy's own GPA at age 16	Girl's own GPA at age 16	Brother's GPA at age 16	Sister's GPA at age 16
Independent variable:						
Brother's cognitive ability at age 18	0.418	0.381	.	.	0.363	0.333
	(0.011)	(0.011)	.	.	(0.011)	(0.011)
Own cognitive ability at age 13	.	.	0.581	0.572	.	.
	.	.	(0.009)	(0.009)	.	.
Statistic:						
Relative correlation	0.910		0.986		0.917	
	(0.034)		(0.022)		(0.039)	
n	7,833	7,604	7,833	7,604	7,833	7,604

Notes: All estimates come from separate regressions. The ability measures have been standardized. The standard error for the relative sibling correlation has been calculated by the Delta method.

Table D4 Correlation matrix for different ability measures

Dependent variable:	Cognitive ability	Logical ability	Verbal ability	Spatial ability	Technical ability	Non-cognitive ability	Social maturity	Intensity	Psychological energy	Emotional stability
Cognitive ability	1.000									
Logical ability	0.898	1.000								
Verbal ability	0.812	0.738	1.000							
Spatial ability	0.843	0.696	0.592	1.000						
Technical ability	0.861	0.713	0.672	0.732	1.000					
Non-cognitive ability	0.390	0.384	0.345	0.313	0.346	1.000				
Social maturity	0.346	0.335	0.314	0.278	0.301	0.765	1.000			
Intensity	0.225	0.234	0.191	0.176	0.209	0.801	0.559	1.000		
Psychological energy	0.305	0.299	0.274	0.245	0.266	0.732	0.745	0.605	1.000	
Emotional stability	0.324	0.312	0.286	0.264	0.289	0.788	0.748	0.574	0.707	1.000

Notes: The table shows the correlation between the variables listed in the rows and columns. The correlations are based on the entire population of men born 1950-1987 (n=1,194,820).

Table D5 Father-son and mother-son correlations – different ability measures

Son's ability:	Cognitive ability	Logical ability	Verbal ability	Spatial ability	Technical ability
Father's ability	0.332	0.322	0.354	0.320	0.351
	(0.022)	(0.019)	(0.025)	(0.036)	(0.025)
Mother's ability	0.424	0.353	0.404	0.470	0.451
	(0.020)	(0.021)	(0.023)	(0.034)	(0.021)
n	25,252	25,252	25,252	25,252	25,252

Son's ability:	Non-cognitive ability	Social maturity	Intensity	Psychological energy	Emotional stability
Father's ability	0.325	0.294	0.304	0.287	0.276
	(0.030)	(0.049)	(0.030)	(0.043)	(0.045)
Mother's ability	0.285	0.332	0.262	0.325	0.283
	(0.031)	(0.048)	(0.034)	(0.046)	(0.046)
n	25,252	25,252	25,252	25,252	25,252

Notes: All estimates come from regressions using the same ability measures for the father and mother as for the son. The ability measures have been standardized. The father's abilities have been predicted by using the paternal uncle's abilities. The mother's abilities have been predicted by using the maternal uncle's abilities and the relative sibling correlations in Table 7. Bootstrapped standard errors (200 reps) accounting for clusters on the father are in parentheses.

Table D6 Descriptive statistics for different samples

	(1)	(2)	(3)	(4)	(5)
Son's variables:					
Cognitive ability at age 13	-0.035	0.051	0.083	0.012	.
	(0.978)	(0.948)	(0.885)	(1.021)	.
Compulsory school GPA at age 16	-0.042	0.071	0.057	0.058	0.257
	(0.975)	(0.940)	(0.943)	(0.959)	(0.992)
Cognitive ability at age 18	.	0.092	0.094	0.097	0.279
	.	(0.930)	(0.935)	(0.958)	(0.933)
Non-cognitive ability at age 18	.	0.060	0.049	0.040	0.187
	.	(0.970)	(0.970)	(0.978)	(0.966)
Years of schooling at age 30-40	12.620	12.841	12.787	12.956	12.529
	(2.791)	(2.608)	(2.57)	(2.672)	(2.699)
Earnings at age 30-40 (SEK 1000s)	274.723	288.192	284.670	298.818	265.831
	(171.075)	(162.723)	(160.697)	(186.567)	(134.166)
Father's variables:					
Cognitive ability at age 13	0.002	0.014	-0.011	-0.048	.
	(0.974)	(0.970)	(0.990)	(1.007)	.
Cognitive ability at age 18	.	-0.004	-0.030	-0.043	0.024
	.	(0.957)	(0.972)	(0.962)	(1.014)
Non-cognitive ability at age 18	.	0.085	0.072	0.065	0.169
	.	(0.975)	(0.977)	(0.982)	(0.983)
Years of schooling at age 40-50	8.266	8.263	8.347	8.158	8.827
	(1.901)	(1.920)	(2.021)	(1.890)	(2.386)
Earnings at age 40-50 (SEK 1000s)	306.949	313.974	313.564	301.590	363.698
	(166.689)	(167.369)	(172.083)	(158.553)	(340.911)
Sample:					
Sons and fathers with draft records		X	X	X	X
Paternal uncles with draft records			X	X	
Maternal uncles with draft records				X	
Dead or emigrated paternal uncles					X
Number of sons	362,297	160,219	50,171	25,252	287
Number of fathers	269,604	128,890	40,242	19,625	257

Notes: The descriptive statistics are restricted to sons and fathers born 1950-1987. Standard deviations are in parentheses. All ability measures have been standardized by year in the entire population. The cognitive skill measure at age 13 is only available for a random sample of 10 % of men born 1953, 1967, 1972, 1977 (5 % sample) or 1982.

Table D7 Addressing sample selection due to military draft: Intergenerational correlations in cognitive abilities at age 13 and 18 for different samples

Dependent variable:	Son's cognitive ability at age 18	Son's cognitive ability at age 18	Son's cognitive ability at age 18	Son's cognitive ability at age 13	Son's cognitive ability at age 13	Son's cognitive ability at age 18
	OLS					
Independent variable:						
Father's cognitive ability at age 13	0.331	0.328	.	.	.	.
	(0.020)	(0.021)	.	.	.	.
Father's cognitive ability at age 18	.	.	0.335	0.387	0.356	0.347
	.	.	(0.021)	(0.020)	(0.023)	(0.023)
	IV					
Father's cognitive ability at age 13	0.460	0.455	.	.	.	.
	(0.028)	(0.029)	.	.	.	.
Father's cognitive ability at age 18	.	.	0.446	0.529	0.487	0.474
	.	.	(0.029)	(0.028)	(0.032)	(0.031)
Sample restriction:						
Sons with valid draft records	X	X	X		X	X
Fathers with valid draft records		X	X	X	X	X
N	2,580	2,294	2,294	2,210	1,526	1,526

Notes: All estimates come from separate regressions. The ability measures have been standardized. In column (2) in the lower panel, the father's ability at age 13 has been instrumented with the father's ability at age 18. The reliability ratio obtained by dividing the OLS-estimate with the IV-estimate in column (2) is used to correct the estimate for measurement error in column (1) in the lower panel. In column (3) in the lower panel the father's ability at age 18 has been instrumented with the father's ability at age 13. In column (4) – (6) in the lower panel, the father's abilities have been corrected for measurement errors using the reliability ratios in Table 3. Standard errors adjusted for clustering on the adoptive father are in parentheses.

Table D8 Addressing sample selection due to uncles: Intergenerational correlations in cognitive and non-cognitive abilities for different samples

Dependent variable:	Son's cognitive ability	Son's cognitive ability	Son's cognitive ability	Son's non-cognitive ability	Son's non-cognitive ability	Son's non-cognitive ability
	OLS					
Independent variable:						
Father's cognitive ability	0.344	0.350	0.369	.	.	.
	(0.002)	(0.004)	(0.021)	.	.	.
Father's non-cognitive ability	.	.	.	0.208	0.210	0.221
	.	.	.	(0.003)	(0.005)	(0.023)
	IV					
Father's cognitive ability	0.471	0.478	0.505	.	.	.
	(0.003)	(0.006)	(0.029)	.	.	.
Father's non-cognitive ability	.	.	.	0.419	0.422	0.445
	.	.	.	(0.005)	(0.009)	(0.046)
Sample restriction:						
Sons and fathers with valid draft records	X	X	X	X	X	X
Paternal uncles with valid draft records		X	X		X	X
Maternal uncles with valid draft records			X			X
N	160,219	50,171	14,799	160,219	50,171	14,799

Notes: All estimates come from separate regressions. The sample is restricted to sons and fathers with valid draft records. The ability measures have been standardized. In the lower panel, the father's abilities have been corrected for measurement errors using the reliability ratios in Table 3. Standard errors adjusted for clustering on the adoptive father are in parentheses.

Table D9 Descriptive statistics of adoptee sample

	Sons	Fathers	Paternal uncles	Maternal uncles
Variables:				
Year of birth	1980.73	1953.05	1954.79	1955.39
	(4.34)	(1.99)	(3.55)	(3.86)
Age at draft	18.25	18.72	18.60	18.54
	(0.43)	(0.69)	(0.73)	(0.71)
Cognitive ability at age 18	-0.35	0.26	0.14	0.06
	(0.90)	(0.93)	(0.93)	(0.98)
Non-cognitive ability at age 18	-0.09	0.28	0.16	0.12
	(0.96)	(0.93)	(0.93)	(1.00)
N	1,519	567	677	833

Notes: Standard deviations are in parentheses. The cognitive and non-cognitive ability measures have been standardized by year of draft in the entire population.