

**Influences of Nature and Nurture on Earnings Variation:
A Report on a Study of Various Sibling Types in Sweden**

Anders Björklund
Swedish Institute for Social Research
Stockholm University
Sweden
anders.bjorklund@sofi.su.se

Markus Jäntti
Department of Economics and Statistics
Abo Akademi University
Finland
markus.jantti@abo.fi

Gary Solon
Department of Economics
University of Michigan
USA
gsolon@umich.edu

April 2003

Prepared for *Unequal Chances: Family Background and Economic Success*, edited by Samuel Bowles, Herbert Gintis, and Melissa Osborne, New York: Russell Sage Foundation, forthcoming.

We are grateful to Nancy Pedersen at the Swedish Twin Registry for making the twins data available for this study. Ake Nilsson at Statistics Sweden provided indispensable guidance in constructing the siblings data set. Financial support from the Joint Committee of the Nordic Social Science Research Councils, the Swedish Council for Research in the Humanities and Social Sciences, and the Swedish Council for Social Research is gratefully acknowledged. We have received useful comments from John Bound, Arthur Goldberger, an anonymous reviewer, participants at the Russell Sage Foundation conference held at the Santa Fe Institute in October 2001, and seminar participants at Southampton University, Nuffield College, the Stockholm School of Economics, and SOFI at Stockholm University.

Influences of Nature and Nurture on Earnings Variation: A Report on a Study of Various Sibling Types in Sweden

1. Introduction

Understanding the sources of earnings inequality is a central topic in labor economics. Indeed, accounting for the rise in earnings inequality that has occurred in most developed countries over the last quarter-century probably has been the field's most active research area in recent years (Katz and Autor 1999). Another active area of inequality research has focused on the role of family and community origins. One line of this research has used sibling correlations to measure the proportion of earnings variation that can be attributed to the family and community background factors that siblings have in common. The basic idea is that, if family and community origins account for a large portion of earnings inequality, siblings will show a strong resemblance in earnings; if family and community background matters hardly at all, siblings will show little more resemblance than would randomly selected unrelated individuals.

Most of the empirical evidence on sibling correlations in earnings (reviewed in Solon 1999) pertains to brothers in the United States. A reasonable summary of that evidence is that the correlation among U.S. brothers in the permanent component of their logarithmic earnings may be around 0.4. Bjorklund et al. (2002) present evidence that brother correlations in long-run earnings are lower in Sweden, Denmark, Finland, and Norway. Another strand of research has focused on twins. Most of the studies of monozygotic twins, both in the United States and elsewhere, have estimated earnings correlations of around 0.6. Estimates for dizygotic twins have tended to be lower than those for monozygotic twins, but higher than those for non-twin brothers.

While these brother correlations are far less than 1, they are large enough to suggest a substantial role for family and community origins in accounting for earnings inequality. Once that is recognized, it is natural to ask which specific background factors make a difference. One

aspect of that question is whether the brother resemblance in earnings stems from similarities between brothers in their genetic endowments or from similarities in their family and community environments. That question has been hotly debated – most recently in response to Herrnstein and Murray’s *The Bell Curve* (1994) – with respect to scores on intelligence tests. Much less attention has been devoted to nature vs. nurture as sources of earnings inequality. The main exception is the work during the 1970’s by Paul Taubman and colleagues (Taubman 1976, Behrman et al. 1977). Using the National Research Council sample of monozygotic and dizygotic twins, Taubman et al. attempted to disentangle the roles of nature and nurture on the assumption that the greater correlation typically observed for monozygotic twins occurs mainly because monozygotic twins (so-called “identical” twins, who come from one fertilized egg that splits in two) have identical genes, whereas the genes of dizygotic twins (“fraternal” twins, who come from two different eggs fertilized by different sperm) are correlated only in the same way as the genes of non-twin siblings.

In this chapter, we use an extraordinary Swedish data set on various types of sibling pairs to reconsider the extent to which sibling correlations in earnings stem from genetic and environmental sources. Following the suggestion of Feldman et al. (2000), we make use of a wide variety of sibling types, differing in both their genetic connectedness and the extent to which they were reared together. Contrasting sibling correlations across a wider variety of sibling types provides additional leverage for disentangling nature and nurture effects, and for examining the sensitivity of results to alternative modeling assumptions.

Although we believe it is worthwhile to use sibling comparisons to generate new clues about the sources of earnings inequality, we stress that the policy implications are far less clear than is sometimes supposed. The notion that environmentally-induced inequality is easily susceptible to policy remediation, and that genetically-based inequality is not, is a *non sequitur*. To borrow an example from Goldberger (1979), a finding of a large genetic role in poor eyesight

would in no way indicate that remediation with eyeglasses is ineffectual. Whether earnings status is determined largely by nature or nurture, any proposed policy ought to be evaluated on the basis of the particular policy's benefits and costs.

In the next section, we outline some simple models of the dependence of sibling correlations on variation in genetic and environmental factors. In the process, we highlight how results from the previous literature and from our own analysis might be sensitive to arbitrary modeling assumptions. In Section 3, we describe our data on Swedish siblings. In Section 4, we use those data to estimate alternative models, and we discuss what the results do (and do not) reveal about nature and nurture as sources of earnings variation. Section 5 summarizes and concludes.

2. Models

A simple model of genetic and environmental influences on an earnings measure Y is

$$(1) \quad Y = gG + sS + uU$$

where G represents the genetic factor; S is an environmental factor that may be at least somewhat shared between siblings; U is an environmental factor totally idiosyncratic to the individual (i.e., not shared at all between siblings); Y , G , S , and U are all standardized to have mean 0 and variance 1; and g , s , and u are parameters that will be related to the relative importance of genes, shared environment, and non-shared environment in accounting for variation in Y . G , S , and U are all "latent variables," i.e., they are not directly observed.

With U defined as perfectly idiosyncratic to the individual, it is natural to assume that U is uncorrelated with both G and S . With one exception, the models to follow also will make the simplifying assumption that G and S are uncorrelated. In that case, g^2 is the fraction of the

variance in Y that is due to genetic variation, and $1 - g^2 = s^2 + u^2$ is the fraction due to the combination of shared and non-shared environment.

When this model is combined with sufficiently strong assumptions about the extent to which G and S are correlated between different sibling types, it becomes possible to use data on sibling correlations in Y to infer how much of earnings inequality can be ascribed to each source. We will illustrate with a simplified caricature of the Taubman et al. analysis of monozygotic and dizygotic twins. Let Y and Y' denote the earnings measures for the two members of a twin pair. In addition to the assumption that G , S , and U are mutually uncorrelated *within* individuals, assume that they are uncorrelated *between* twins – e.g., one twin's G is uncorrelated with the other twin's S' . Let $Corr(G, G') = 1$ for monozygotic twins, who have identical genes, and assume that $Corr(G, G')$ is 0.5 for dizygotic twins, whose genetic resemblance is the same as for non-twin full siblings. (See Otto, Christiansen, and Feldman 1995 for a thorough discussion of genetic resemblance among various family members.) Finally, assume that $Corr(S, S') = 1$ for both monozygotic and dizygotic twins. This last assumption implies that the twin correlation in environmental influences on earnings is $s^2 / (s^2 + u^2)$ for both twin types, so the assumption is that monozygotic twins experience no more (and no less) similarity in environment than dizygotic twins do.

It follows from these strong assumptions that $Corr(Y, Y')$ is $g^2 + s^2$ for monozygotic twins and $0.5g^2 + s^2$ for dizygotic twins. Therefore, g^2 can be estimated by doubling the difference between the earnings correlations observed for monozygotic and dizygotic twins, and s^2 can be estimated by subtracting the estimated g^2 from the earnings correlation observed for monozygotic twins. For example, in the Taubman et al. sample, the earnings correlation was 0.54 for monozygotic twins and 0.30 for dizygotic twins. Processing these correlation estimates through the model described above leads to an estimated g^2 of 0.48 and an estimated s^2 of

0.06. These estimates imply that 48 percent of earnings inequality stems from genetic variation, that 0.48 of the 0.54 earnings correlation for monozygotic twins is due to their identical genes, and that 0.24 of the 0.30 earnings correlation for dizygotic twins is due to their similar genes.

While this example illustrates the possibility of disentangling nature and nurture effects by contrasting the earnings correlations of different sibling types, it also can be used to illustrate the possibility that inferences may be very sensitive to the modeling assumptions. Following Goldberger (1979), replace the assumptions above with the assumptions that $g = 0$ (i.e., genetic variation is of absolutely no consequence in determining earnings) and that $Corr(S, S')$ for dizygotic twins is a fraction ρ of the corresponding correlation for monozygotic twins (i.e., dizygotic twins may be treated less similarly than monozygotic twins). Processing the same observed earnings correlations through this alternative model leads to an estimated g^2 of 0 (by assumption), an estimated s^2 of 0.54, and an estimated ρ of $0.30/0.54 = 0.56$. These estimates imply that 0 percent of earnings inequality stems from genetic variation and that the entirety of the 0.54 and 0.30 earnings correlations for monozygotic and dizygotic twins is due to similarity in environment. Like the preceding model, this one delivers an exact fit to the observed correlations, so the two models (and their dramatically different implications) cannot be distinguished empirically on the basis of comparing correlations for only these two types of sibling pairs.

The empirical strategy in our study is to use a wider variety of sibling types to enable estimation of more general (i.e., less restrictive) models and to examine the sensitivity of our results to variation in modeling assumptions. As described in more detail in the next section, we use observed sibling correlations in earnings for nine types of sibling pairs: monozygotic twins reared together, monozygotic twins reared apart, dizygotic twins reared together, dizygotic twins reared apart, non-twin full siblings reared together, non-twin full siblings reared apart, half-

siblings reared together, half-siblings reared apart, and adoptive siblings. As with the simple models above, we can process the observed correlations through a set of assumptions about the relative genetic and environmental connectedness of the various sibling types to obtain estimates of parameters related to the importance of genes and environment in accounting for earnings variation. The advantage of working with a wider variety of sibling types is that we will be able to use models that invoke somewhat less restrictive assumptions, and we will have greater latitude for checking the robustness of our results to alternative assumptions.

In our work so far, we have estimated four models. “Model 1” is essentially the extension of the first model above to the setting with nine different types of sibling pairs. In this model, $Corr(G, G') = 1$ for monozygotic twins (reared together or apart), and we continue to assume that $Corr(G, G') = 0.5$ for dizygotic twins (reared together or apart) as well as for non-twin full siblings (reared together or apart). We also assume that $Corr(G, G') = 0.25$ for half-siblings (reared together or apart) and $Corr(G, G') = 0$ for adoptive siblings. We continue to assume that $Corr(S, S') = 1$ for all types of sibling pairs reared together, and we assume that $Corr(S, S') = 0$ for sibling pairs reared apart. In effect, the model assumes that all types of sibling pairs reared together experience the same degree of environmental similarity and that the environments of siblings reared apart are absolutely unrelated. Although this model is similar to models frequently used in twins-based research on intelligence and personality traits, we recognize that the assumptions are terribly restrictive, and we do not mean them to be taken very seriously. Rather, we will use this model as a point of departure for considering several less restrictive models.

Model 2 dispenses with the simplifying assumption that G and S are uncorrelated. Instead, recognizing the possibility that those with genes conducive to high earnings also tend to have advantaged environments, this model treats $Corr(G, S')$ as a quantity to be estimated. In

particular, model 2 adds two new parameters that allow for three distinct nonzero correlations between G and S' : one for biological siblings reared together, one for siblings reared apart, and one for adoptive siblings.

Model 1's assumptions that the genetic correlation is 0.5 for dizygotic twins and non-twin full siblings, 0.25 for half-siblings, and 0 for adoptive siblings are frequently used in siblings research, but they are highly questionable. The assumptions of 0.5 and 0.25 correlations fail in the presence of assortative mating, genetic dominance, or non-additivities in the effects of different genes. The assumption of a zero correlation for adoptive siblings fails in the presence of non-randomness in adoption placements. Therefore, in model 3, we replace these three assumed values of 0.5, 0.25, and 0 with three parameters to be estimated.

Finally, we need to loosen up the restrictions that $Corr(S, S') = 1$ for all sibling pairs reared together and $Corr(S, S') = 0$ for all those reared apart. As emphasized by Goldberger (1979), the implication that the sibling correlation in environmental influences on earnings is the same $s^2 / (s^2 + u^2)$ for all types reared together is hard to believe. This assumption is false, for example, if monozygotic twins are treated more similarly than dizygotic twins, or if twins are treated more similarly than non-twins. Kamin and Goldberger (2002) caution that the assumption of zero environmental correlation for siblings reared apart also is questionable. Twins classified as reared apart, for example, shared the same womb; after birth, they may have been reared together for some time before they were separated; and, even after they were separated, they may have experienced correlated environments, an obvious example being if they were reared by relatives. Therefore, in model 4, instead of imposing correlations of 1 for all pairs reared together and 0 for all those reared apart, we normalize $Corr(S, S')$ to 1 for monozygotic twins reared together, and we introduce three new parameters to represent this

correlation for other sibling types: one for dizygotic twins reared together, one for non-twins reared together, and one for siblings reared apart.

Of course, each of these embellished models is still a very stylized model of how nature and nurture affect earnings. In principle, we would like to estimate more general and realistic models, for example, by combining the features of models 2, 3, and 4. Unfortunately, simultaneously incorporating the features of even any two of these extended models results in under-identification; that is, the combined model does not lead to well-defined parameter estimates because very different sets of parameter values produce the exact same fit to the data. Nevertheless, as we will see in Section 4, the estimates from models 1-4 do shed some light on the range of conclusions about nature vs. nurture that can be supported under various assumptions. First, however, we will describe the extraordinary data set on which our estimates are based.

3. Data

Our samples of twins and non-twins come from different sources. Our twins sample comes from the middle cohort of the Swedish Twin Registry, developed and administered at the Karolinska Institutet in Stockholm. The starting point for this sample is the population of all twins born in Sweden between 1926 and 1967, in all 54,890 pairs. Out of these, all 17,992 same-sex twin pairs born between 1926 and 1958 who were alive and living in Sweden in 1970 were sent a questionnaire in 1972 (Medlund et al. 1977). Responses were received from both members of 13,664 pairs.

One of the variables elicited in the questionnaire was the twins' own report of their zygosity. More objective information on zygosity is available for a small sub-sample included in SATSA (the Swedish Adoption/Twin Study of Aging). In that project, 351 twin pairs reared apart and a control sample of 407 twin pairs reared together have been subjected to intensive

study, including a medical determination of zygosity based on blood samples. Whenever possible, we classify twins as monozygotic or dizygotic on the basis of the SATSA information, but, for the majority of our twins sample that is not in the SATSA sub-sample, we must rely on self-reports of zygosity. Fortunately, cross-tabulations of the two zygosity measures for the SATSA sub-sample have shown that the self-reports conform remarkably well with the determinations from the blood tests (Pedersen et al. 1991).

The 1972 questionnaire also asked, “How long did you live with your twin partner?” We follow previous studies of the Swedish twins in categorizing the pairs as reared together or apart on the basis of whether they were separated before age 10. That some of the twins “reared apart” did live together for a substantial period before age 10 is part of the motivation for our model 4, which estimates the correlation in shared environment for siblings “reared apart” instead of assuming that the correlation is zero.

Our data on non-twin siblings stem from two simple random samples of the Swedish population drawn by Statistics Sweden. The first is a sample of 100,000 persons born in Sweden between 1951 and 1964. A further requirement is not to be adopted by the parents. The second sample consists of 3,000 persons born in Sweden between 1951 and 1964 who were adopted by both parents. A further requirement in both samples is that the persons lived in Sweden in 1993. Because these data are collected from population registers, there is no non-response.

For the members of these two samples, Statistics Sweden identified their siblings according to several definitions. First, full siblings were identified as siblings with the same two biological parents. From the resulting sample of full-sibling pairs, we drew our sample of non-twin full siblings as those who were not born in the same year and the same month (or adjacent months). Second, half-siblings were identified, and a distinction was made between half-siblings on the mother’s side and on the father’s side. Because our preliminary analyses showed similar results for both types of half-siblings, we pool them together in the current analysis. Third,

siblings related by adoption were identified by means of an adoption code recorded in a special population register held by Statistics Sweden. In our present analysis, we use only pairs containing a biological child and an adopted child. We suspect that the higher correlations observed for pairs of adopted children might reflect biological relationships between the two adopted children. Finally, Statistics Sweden also recorded whether the siblings lived together in the censuses of 1960, 1965, 1970, 1975, and 1980. We classify non-twin siblings as reared apart if they never lived in the same household in any of the censuses.

We confine our present analysis to brother-brother pairs and sister-sister pairs. When more than two same-sex siblings from the same family meet our sample restrictions, we use all available pairs. For example, a family with three brothers contributes three brother pairs to our sample. Our standard error estimates do not account for the non-independence of the overlapping sibling pairs from the same family.

What distinguishes our study from the many previous studies of Swedish twins is our focus on earnings as the variable of interest. We obtained special permission from Statistics Sweden to access longitudinal data on annual labor earnings for the members of our siblings samples. The earnings data are for 1987, 1990, and 1993. They come from compulsory reports by employers to tax authorities and should be very reliable. Earnings of the self-employed are included.

With earnings as our variable of interest, we wish to observe our sample members when they are of working age. The participants in the twins survey were born between 1926 and 1958, so they were between the ages of 29 and 61 in 1987. We restrict our sample of non-twin siblings to those born between 1943 and 1965, so they were between 22 and 44 in 1987. (Note that, although our non-twins sample began with individuals born between 1951 and 1964, their siblings could have been born outside that range.) We include individuals only if they had positive earnings in at least one of the years 1987, 1990, and 1993.

Ideally, we would like to use a long-run measure of earnings that is not greatly influenced by transitory fluctuations in a single year's earnings and that is adjusted for stage of life cycle. For smoothing out transitory fluctuations, we are fortunate to have access to up to three years of earnings data over a span of seven years. We also have performed a first-stage regression adjustment to account for stage of life cycle. In particular, for a more inclusive preliminary sample (e.g., including individuals in sister-brother pairs), we pooled all observations of positive earnings in 1987, 1990, and 1993. Then, separately for women and men, we applied least squares to the regression of the natural logarithm of annual earnings on year dummy variables and a cubic in age. Then, taking the residual from that regression as an age-adjusted measure of log annual earnings, we smoothed out transitory earnings variation by averaging each individual's residualized log earnings variable over all of his/her available observations from the three years. The resulting multi-year average of age-adjusted log earnings is the earnings measure Y for which we measure sibling correlations, which in turn are used to decompose earnings inequality into genetic and environmental components.

The first column of Table 1 restates the nine types of sibling pairs used in our analysis, and the second column shows our sample's number of pairs of each type. Our sample sizes for monozygotic and dizygotic twins reared apart are very small, so our estimates of the sibling correlations for those sibling types will be very unreliable. Accordingly, the model-fitting method we will present in the next section will be designed to give relatively little weight to imprecise correlation estimates based on small samples. The sample sizes for our other seven types of sibling pairs, however, are quite substantial, especially relative to the tiny sample sizes often used in siblings research.

The third column of Table 1 shows the sample estimates of the sibling correlation $Corr(Y, Y')$ for each type of sibling pair. Letting N denote the sample's number of pairs of a

particular type, we calculate our estimate of $Corr(Y, Y')$ as the ratio of the siblings' sample covariance in Y to the sample variance:

$$(2) \quad \hat{C} = \frac{\left(\sum_{i=1}^N Y_i Y'_i / N \right) - \bar{Y}^2}{\left[\sum_{i=1}^N (Y_i^2 + Y_i'^2) / (2N) \right] - \bar{Y}^2}$$

where

$$(3) \quad \bar{Y} = \sum_{i=1}^N (Y_i + Y'_i) / (2N).$$

In parentheses below each estimated sibling correlation is the associated standard error estimate, computed as

$$(4) \quad \hat{\sigma} = \sqrt{(1 - \hat{C}^2) / N}.$$

For our very large samples (over 40,000 each) of non-twin full siblings reared together, we estimate the sibling correlation in our earnings measure Y to be 0.17 for brothers and 0.13 for sisters. While significantly positive, these estimates accord with the finding of Bjorklund et al. (2002) that sibling correlations in earnings are lower in Sweden than in the United States. Our measured sibling correlations for monozygotic twins reared together are 0.36 for brothers and 0.31 for sisters. As expected, these estimates are larger than for the other types of sibling pairs, but again they are distinctly lower than most of the corresponding estimates for the United States. The correlation estimates for dizygotic twins reared together are similar to those for non-twin full siblings, while those for half-siblings and adoptive siblings are somewhat lower. At first glance, the estimates for monozygotic twins reared apart seem surprisingly low, but, as they are based on samples of only slightly more than 40 pairs, they are extremely unreliable.

Just from eyeballing these numbers, one might believe one sees some patterns consistent with a genetic role in earnings inequality. For example, the measured earnings correlations for monozygotic twins reared together are greater than those for dizygotic twins and non-twin full

siblings, which in turn exceed the correlations for half-siblings and adoptive siblings. Possible signs of a contribution of shared environment to the sibling correlations are that the correlation estimates for adoptive siblings are significantly positive (though not very large) and that the measured correlations for particular genetic types of sibling pairs are often (though not always) greater for those reared together than for those reared apart. That none of the observed earnings correlations exceeds 0.36 suggests a predominant role for non-shared environment. In the next section, we proceed to a more formal analysis of these patterns.

4. Results from Model Estimation

We begin by estimating model 1 as described in Section 2. Recall that this simple model makes very strong assumptions: that all types of sibling pairs reared together experience the same degree of environmental similarity; that those reared apart experience *no* similarity in environment; that the genetic and environmental factors are uncorrelated with each other; and that the genetic similarity of dizygotic twins and non-twin full siblings is half that of monozygotic twins, the genetic similarity for half-siblings is a quarter of that for monozygotic twins, and adoptive siblings have no genetic resemblance at all. In terms of equation (1), these assumptions imply that the earnings correlation is $g^2 + s^2$ for monozygotic twins reared together, $0.5g^2 + s^2$ for dizygotic twins and non-twin full siblings reared together, $0.25g^2 + s^2$ for half-siblings reared together, s^2 for adoptive siblings, g^2 for monozygotic twins reared apart, $0.5g^2$ for dizygotic twins and non-twin full siblings reared apart, and $0.25g^2$ for half-siblings reared apart.

Taken literally, the model delivers all too many ways of estimating the parameters. Consider the estimation of g^2 , which is the fraction of earnings inequality attributable to genetic variation. As noted in Section 2, one way of estimating g^2 from this model is to double the

difference between the measured correlations for monozygotic twins and dizygotic twins reared together. With our correlation estimates for brothers, that delivers an estimated g^2 of 0.39, which is *too* large for explaining even our very highest correlation estimate, even if environmental variation is of no consequence at all. Another way to estimate g^2 is by doubling the difference between the measured correlations for monozygotic and dizygotic twins reared apart. With our correlation estimates for either brothers or sisters, that delivers a ridiculous negative estimate of g^2 , but of course the estimate is very unreliable because of the small sample sizes for twins reared apart. Clearly, the statistical challenge is how best to combine the available evidence from the nine different types of sibling pairs.

We use the approach of “minimum distance estimation,” which chooses parameter estimates so as to match the implied population values of the sibling correlations as closely as possible to the values observed in our sample. In other words, we estimate g^2 and s^2 by taking the nine sample sibling correlations as nine observations of our dependent variable and then applying least squares to the regression of those observed correlations on the functions of g^2 and s^2 listed in the first paragraph of this section. To improve the precision of our parameter estimation, we use weighted least squares, with the nine observations’ contributions to the sum of squares weighted by the sample sizes on which they are based.¹ For example, in our model estimation for brothers, the observed correlation for non-twin full siblings reared together, which is based on almost 50,000 brother pairs, gets more than 1,000 times as much weight as the correlation estimate for monozygotic twins reared apart, which is based on only 45 pairs. This is appropriate because the latter correlation estimate is subject to vastly greater sampling error and

¹ In principle, we could improve the asymptotic efficiency of our estimation still further by using weights that incorporate the influence of the sibling correlations, as well as the sample sizes, on the error variance in our regression equation. We do not do so because, as explained by Altonji and Segal (1996) and Clark (1996), basing the weights on the estimated correlations themselves might induce a large finite-sample bias in our parameter estimation. We do, however, take account of how the error variance depends on the sibling correlations when we

therefore is vastly less informative. Thus, our estimation procedure will labor mightily to get a close fit to the sample correlation for full siblings reared together and will hardly bother at all with fitting the sample correlations for monozygotic and dizygotic twins reared apart. Our estimates of g^2 and s^2 will be based mainly on contrasts among the seven types of sibling pairs for which we do have sizeable samples and will depend very little on the other two.

Estimating model 1 with our brothers samples produces an estimated g^2 of 0.281 (with estimated standard error 0.080) and an estimated s^2 of 0.038 (0.037). The estimated value of $u^2 = 1 - g^2 - s^2$ therefore is 0.681, suggesting that non-shared environment is responsible for most earnings variation. The implied values of the population sibling correlations for the nine types of sibling pairs are listed in the fourth column of Table 1, and the fifth and sixth columns decompose the implied correlations into their genetic and environmental components. The results imply that genetic variation accounts for 28 percent of earnings inequality, 0.28 of the 0.32 earnings correlation fitted for monozygotic twins reared together, and 0.14 of the 0.18 correlation fitted for dizygotic twins and non-twin full brothers reared together. The results for sisters, shown in the lower panel of Table 1, are qualitatively similar to those for brothers.

Our weighted estimation scheme assures a good fit to the observed correlations for the huge samples of non-twin full siblings reared together, with the fitted value of 0.179 for brothers coming very close to the observed value of 0.174. Similarly, the fitted value of 0.131 for sisters only slightly over-predicts the observed value of 0.127. The model comes nowhere near fitting the observed values for monozygotic twins reared apart, but that is just as well because those strangely small sample correlations are based on tiny samples and are probably far from the true population values. For the other sibling types with sizeable samples, the fitted values from the model are fairly close to the observed values, though not close enough to satisfy a chi-square

perform our standard error estimation, which relies on standard results on the asymptotic covariance matrix for nonlinear least squares estimation.

goodness-of-fit test, which rejects the model at the 0.05 level for both brothers and sisters. For brothers, the most noticeable failures of the model are that it under-predicts the correlations for monozygotic twins reared together and adoptive siblings by 0.04, for half-siblings reared together by 0.03, and for full siblings reared apart by 0.02. Some similar discrepancies appear for the sisters sample.

If anything, it is surprising that the model fits as well as it does. As emphasized in Section 2, many of the model's assumptions are highly implausible, so we ought to consider several less restrictive models. Recall that, because one might expect that individuals with genes conducive to high earnings also tend to enjoy advantaged environments, model 2 loosens up the assumption of zero correlation between the genetic factor G and the shared-environment factor S . In particular, the new model allows three distinct non-zero values of $Corr(G, S')$, one each for biological siblings reared together, siblings reared apart, and adoptive siblings. To our surprise, for both brothers and sisters, these estimated cross-correlations turn out to be insignificantly *negative*.

As shown in Table 2, except for the exact fit enabled by a free parameter for adoptive siblings, the addition of parameters does not dramatically improve the fit to the data; indeed, for brothers, the chi-square statistic for improvement of fit is insignificant at the 0.05 level. Summarizing the implications for nature vs. nurture is complicated by the fundamental ambiguity about which should be credited for the correlation between the two. The decompositions in Table 2 provide a range between the two polar extremes of attributing all of the cross-term to either nature or nurture. For the most part, the qualitative implications are similar to those from model 1. For both brothers and sisters, the larger part of most of the fitted correlations is attributed to the siblings' genetic resemblance. Although the results of this exercise should not be read as proving the absence of a positive correlation between nature and nurture, it appears that the arbitrary practice of assuming zero correlation among the factors in

the variance decomposition (which is quite common in many applications, not just in siblings research) does not cost much in the present context.

Another of the questionable assumptions in model 1 is that the genetic correlation is 0.5 for dizygotic twins and non-twin full siblings, 0.25 for half-siblings, and 0 for adoptive siblings. This set of restrictions, quite commonly imposed in siblings research, assumes away the effects of assortative mating, genetic dominance, non-additivities in influences of different genes, and non-randomness in adoption placements. Model 3 therefore replaces the previously assumed values of 0.5, 0.25, and 0 with three new parameters to be estimated. For brothers, the respective estimates turn out to be 0.43 (with estimated standard error 0.01), 0.25 (0.04), and 0.14 (0.09); for sisters, they are 0.39 (0.02), 0.26 (0.04), and 0.18 (0.10). Although the joint hypothesis that the 0.5, 0.25, and 0 values were correct is rejected at the 0.05 significance level for both brothers and sisters, the newly estimated values are not hugely different from the values previously assumed. Consequently, the qualitative implications from model 1 are preserved again. The results for brothers imply that genetic variation accounts for 32 percent of earnings inequality, 0.32 of the 0.36 earnings correlation fitted for monozygotic twins reared together, and 0.14 of the 0.18 correlation fitted for dizygotic twins and non-twin full siblings. The implications for sisters are similar.

Finally – and, we think, most importantly – we turn to the assumption that all siblings reared together experience the same degree of environmental similarity and that all siblings reared apart experience no environmental similarity at all. This assumption of model 1 ignores the likelihood that monozygotic twins are treated more similarly than other sibling pairs are, and it overlooks the many reasons that siblings classified as “reared apart” might have correlated environments (e.g., some of them did live together for a substantial time before being separated, or they were raised by relatives). As explained in Section 2, it is easy to see how misspecification of environmental effects could distort decompositions of genetic and

environmental influences. For example, the assumption that monozygotic and dizygotic twins differ only in their genetic resemblance, and not at all in their environmental similarity, forces the greater earnings correlation observed for monozygotic twins to be attributed to an important role for genetics.

Therefore, in model 4, we explore the implications of a more flexible specification of environmental effects. Normalizing $Corr(S, S')$ to 1 for monozygotic twins reared together, we estimate three new parameters to represent this correlation for dizygotic twins reared together, non-twins reared together, and siblings reared apart. For brothers, the respective estimates are 0.406 (with estimated standard error 0.147), 0.461 (0.216), and 0.209 (0.311); for sisters, they are 0.282 (0.131), 0.340 (0.074), and 0.254 (0.180). As suspected, the estimates for reared-together siblings other than monozygotic twins are significantly less than 1 (in both the statistical and substantive senses), and the estimates for siblings reared apart are also more than 0, although not significantly so in the statistical sense. For both brothers and sisters, chi-square tests of model 1 vs. model 4 show that model 4's improvement of fit is statistically significant. Furthermore, model 4 fits the data better than models 2 and 3. For both brothers and sisters, goodness-of-fit tests of model 4 fail to reject it at even the 0.20 significance level.

Table 4 displays model 4's implications for decomposing earnings variation between nature and nurture. This time the qualitative findings are somewhat altered. The fraction of men's earnings inequality attributable to genetic variation, which had been estimated at about 0.3 in previous models, is now estimated at 0.2. With previous models, about 80 percent or more of the earnings correlations for twin brothers and non-twin full brothers usually had been attributed to their genetic resemblance. Now the proportion from genetic resemblance is more like 60 percent, with shared environmental influences now playing a larger role than before. For sisters, the estimated fraction of earnings inequality attributable to genetic variation, previously between 0.2 and 0.3, now falls to 0.13. The estimated genetic share of the earnings correlations for twin

sisters and non-twin full sisters, which typically had been estimated at about 90 percent before, now has about a 50/50 split with shared environment.

The intuition for the change in results is straightforward. In model 1, which assumed that all siblings reared together experience the same environmental similarity, the only way to account for the large gap between the earnings correlation for monozygotic twins reared together and the much lower correlations for dizygotic twins and full siblings reared together was to estimate a large value for g^2 , the genetic component of earnings variation. For brothers, for example, explaining the gap between the 0.36 correlation measured for monozygotic twins reared together and the 0.17 correlations measured for both dizygotic twins and full siblings reared together would require a g^2 of almost 0.40. Even with no role whatsoever for shared environment, however, a g^2 of almost 0.40 would lead to over-prediction of the earnings correlations for every type of biological sibling. Furthermore, avoiding over-prediction of those correlations requires a small value of s^2 , the shared environment component, which leads to under-prediction of the correlation for adoptive siblings. The reason that model 4 fits the data better is that it interprets the much higher correlation observed for monozygotic twins reared together as reflecting greater similarity in environment in addition to greater genetic resemblance. Model 4 therefore can fit the relatively high correlation for monozygotic twins reared together without having to over-predict the correlations for other biological siblings. At the same time, by assigning a substantial role to shared environment, model 4 can explain the significantly positive correlation observed for adoptive siblings, and, by allowing for some environmental similarity among siblings reared apart, it can fit the reared-apart correlations without requiring quite so large a genetic component.

5. Conclusions

In this chapter, we have used new evidence from Sweden on earnings correlations among a variety of sibling types in an attempt to shed new light on nature vs. nurture as sources of earnings inequality. The richness of our data set has enabled us to explore the robustness of our results to variations in model specification. We have found that the results are indeed sensitive to flexibility in modeling the variation across types of sibling pairs in the similarity of their environments. Models that assume monozygotic twins experience the same environmental similarity as other sibling pairs and that assume no environmental resemblance among siblings reared apart tend to exaggerate the importance of nature relative to nurture. Even our smallest estimates of the genetic component of earnings variation, however, suggest that it accounts for about 20 percent of earnings inequality among men and more than 10 percent among women.

Although our results point to a significant role for genetic variation, perhaps the most striking finding is the most obvious one – about the importance of *non*-shared environment. The largest sibling correlation in earnings that we estimate is a 0.36 correlation for monozygotic twin brothers. Even though these brothers have identical genes and, according to our preferred model, experience even more similar environments than other sibling pairs do, an estimated 64 percent of their earnings variation is explained by neither genetic nor environmental resemblance. In other words, much and perhaps most of earnings variation in Sweden stems from environmental factors that are not shared even by monozygotic twins.

Table 1. Results from Model 1

Type of sibling pair	Number of pairs	Sibling correlation	Fitted value from model	Genetic component	Env. Component
Brothers					
MZ twins reared together	2,052	0.363 (0.021)	0.319	0.281 (0.080)	0.038 (0.037)
MZ twins reared apart	45	0.072 (0.149)	0.281	0.281	0
DZ twins reared together	3,269	0.166 (0.017)	0.179	0.141	0.038
DZ twins reared apart	41	0.165 (0.154)	0.141	0.141	0
Full siblings reared together	48,389	0.174 (0.004)	0.179	0.141	0.038
Full siblings reared apart	3,297	0.159 (0.017)	0.141	0.141	0
Half-siblings reared together	2,862	0.138 (0.018)	0.108	0.070	0.038
Half-siblings reared apart	4,782	0.068 (0.014)	0.070	0.070	0
Adoptive siblings	1,954	0.082 (0.023)	0.038	0	0.038
Sisters					
MZ twins reared together	2,395	0.309 (0.019)	0.254	0.245 (0.080)	0.009 (0.037)
MZ twins reared apart	41	-0.048 (0.156)	0.245	0.245	0
DZ twins reared together	3,474	0.116 (0.017)	0.131	0.123	0.009
DZ twins reared apart	64	0.177 (0.123)	0.123	0.123	0
Full siblings reared together	42,510	0.127 (0.005)	0.131	0.123	0.009
Full siblings reared apart	3,310	0.102 (0.017)	0.123	0.123	0
Half-siblings reared together	2,747	0.069 (0.019)	0.070	0.061	0.009
Half-siblings reared apart	4,321	0.086 (0.015)	0.061	0.061	0
Adoptive siblings	1,705	0.066 (0.024)	0.009	0	0.009

Numbers in parentheses are estimated standard errors. For men, the chi-square goodness-of-fit statistic is 15.6 with 7 degrees of freedom (p-value 0.030). For women, it is 22.9 (p-value 0.002).

Table 2. Results from Model 2

Type of sibling pair	Number of pairs	Sibling correlation	Fitted value from model	Genetic component	Env. component
Brothers					
MZ twins reared together	2,052	0.363 (0.021)	0.334	0.250 - 0.314	0.020 - 0.084
MZ twins reared apart	45	0.072 (0.149)	0.307	0.307 - 0.314	-0.007 - 0
DZ twins reared together	3,269	0.166 (0.017)	0.177	0.093 - 0.157	0.020 - 0.084
DZ twins reared apart	41	0.165 (0.154)	0.150	0.150 - 0.157	-0.007 - 0
Full siblings reared together	48,389	0.174 (0.004)	0.177	0.093 - 0.157	0.020 - 0.084
Full siblings reared apart	3,297	0.159 (0.017)	0.150	0.150 - 0.157	-0.007 - 0
Half-siblings reared together	2,862	0.138 (0.018)	0.098	0.015 - 0.079	0.020 - 0.084
Half-siblings reared apart	4,782	0.068 (0.014)	0.072	0.072 - 0.079	-0.007 - 0
Adoptive siblings	1,954	0.082 (0.023)	0.082	-0.002 - 0	0.082 - 0.084
Sisters					
MZ twins reared together	2,395	0.309 (0.019)	0.274	0.201 - 0.289	-0.015 - 0.073
MZ twins reared apart	41	-0.048 (0.156)	0.277	0.277 - 0.289	-0.012 - 0
DZ twins reared together	3,474	0.116 (0.017)	0.129	0.057 - 0.145	-0.015 - 0.073
DZ twins reared apart	64	0.177 (0.123)	0.132	0.132 - 0.145	-0.012 - 0
Full siblings reared together	42,510	0.127 (0.005)	0.129	0.057 - 0.145	-0.015 - 0.073
Full siblings reared apart	3,310	0.102 (0.017)	0.132	0.132 - 0.145	-0.012 - 0
Half-siblings reared together	2,747	0.069 (0.019)	0.057	-0.016 - 0.072	-0.015 - 0.073
Half-siblings reared apart	4,321	0.086 (0.015)	0.060	0.060 - 0.072	-0.012 - 0
Adoptive siblings	1,705	0.066 (0.024)	0.066	-0.007 - 0	0.066 - 0.073

Numbers in parentheses are estimated standard errors. For men, the chi-square goodness-of-fit statistic is 10.3 with 5 degrees of freedom (p-value 0.07). For women, it is 14.9 (p-value 0.01).

Table 3. Results from Model 3

Type of sibling pair	Number of pairs	Sibling correlation	Fitted value from model	Genetic component	Env. component
Brothers					
MZ twins reared together	2,052	0.363 (0.021)	0.357	0.320 (0.059)	0.037 (0.026)
MZ twins reared apart	45	0.072 (0.149)	0.320	0.320	0
DZ twins reared together	3,269	0.166 (0.017)	0.175	0.138	0.037
DZ twins reared apart	41	0.165 (0.154)	0.138	0.138	0
Full siblings reared together	48,389	0.174 (0.004)	0.175	0.138	0.037
Full siblings reared apart	3,297	0.159 (0.017)	0.138	0.138	0
Half-siblings reared together	2,862	0.138 (0.018)	0.118	0.080	0.037
Half-siblings reared apart	4,782	0.068 (0.014)	0.080	0.080	0
Adoptive siblings	1,954	0.082 (0.023)	0.082	0.044	0.037
Sisters					
MZ twins reared together	2,395	0.309 (0.019)	0.303	0.291 (0.064)	0.012 (0.027)
MZ twins reared apart	41	-0.048 (0.156)	0.291	0.291	0
DZ twins reared together	3,474	0.116 (0.017)	0.126	0.114	0.012
DZ twins reared apart	64	0.177 (0.123)	0.114	0.114	0
Full siblings reared together	42,510	0.127 (0.005)	0.126	0.114	0.012
Full siblings reared apart	3,310	0.102 (0.017)	0.114	0.114	0
Half-siblings reared together	2,747	0.069 (0.019)	0.086	0.074	0.012
Half-siblings reared apart	4,321	0.086 (0.015)	0.074	0.074	0
Adoptive siblings	1,705	0.066 (0.024)	0.066	0.054	0.012

Numbers in parentheses are estimated standard errors. For men, the chi-square goodness-of-fit statistic is 6.7 with 4 degrees of freedom (p-value 0.17). For women, it is 7.4 (p-value 0.12).

Table 4. Results from Model 4

Type of sibling pair	Number of pairs	Sibling correlation	Fitted value from model	Genetic component	Env. Component
Brothers					
MZ twins reared together	2,052	0.363 (0.021)	0.363	0.199 (0.157)	0.164 (0.158)
MZ twins reared apart	45	0.072 (0.149)	0.233	0.199	0.034
DZ twins reared together	3,269	0.166 (0.017)	0.166	0.100	0.067
DZ twins reared apart	41	0.165 (0.154)	0.134	0.100	0.034
Full siblings reared together	48,389	0.174 (0.004)	0.175	0.100	0.076
Full siblings reared apart	3,297	0.159 (0.017)	0.134	0.100	0.034
Half-siblings reared together	2,862	0.138 (0.018)	0.125	0.050	0.076
Half-siblings reared apart	4,782	0.068 (0.014)	0.084	0.050	0.034
Adoptive siblings	1,954	0.082 (0.023)	0.076	0	0.076
Sisters					
MZ twins reared together	2,395	0.309 (0.019)	0.309	0.130 (0.085)	0.179 (0.044)
MZ twins reared apart	41	-0.048 (0.156)	0.176	0.130	0.045
DZ twins reared together	3,474	0.116 (0.017)	0.116	0.065	0.050
DZ twins reared apart	64	0.177 (0.123)	0.111	0.065	0.045
Full siblings reared together	42,510	0.127 (0.005)	0.126	0.065	0.061
Full siblings reared apart	3,310	0.102 (0.017)	0.111	0.065	0.045
Half-siblings reared together	2,747	0.069 (0.019)	0.093	0.033	0.061
Half-siblings reared apart	4,321	0.086 (0.015)	0.078	0.033	0.045
Adoptive siblings	1,705	0.066 (0.024)	0.061	0	0.061

Numbers in parentheses are estimated standard errors. For men, the chi-square goodness-of-fit statistic is 5.21 with 4 degrees of freedom (p-value 0.27). For women, it is 4.65 (p-value 0.34).

References

- Altonji, Joseph G., and Lewis M. Segal. 1996. "Small-Sample Bias in GMM Estimation of Covariance Structures." *Journal of Business and Economic Statistics* 14(3): 353–66.
- Behrman, Jere R., Paul Taubman, and Terence Wales. 1977. "Controlling for and Measuring the Effects of Genetics and Family Environment in Equations for Schooling and Labor Market Success." In *Kinometrics: Determinants of Socioeconomic Success within and between Families*, edited by Paul Taubman. Amsterdam: North-Holland.
- Bjorklund, Anders, Tor Eriksson, Markus Jantti, Oddbjorn Raaum, and Eva Osterbacka. 2002. "Brother Correlations in Earnings in Denmark, Finland, Norway, and Sweden Compared to the United States." *Journal of Population Economics* 15(4): 757-72.
- Clark, Todd E. 1996. "Small-Sample Properties of Estimators of Nonlinear Models of Covariance Structure." *Journal of Business and Economic Statistics* 14(3): 367-73.
- Feldman, Marcus W., Sarah P. Otto, and Freddy B. Christiansen. 2000. "Genes, Culture, and Inequality." In *Meritocracy and Economic Inequality*, edited by Kenneth Arrow, Samuel Bowles, and Steven Durlauf. Princeton: Princeton University Press.
- Goldberger, Arthur S. 1979. "Heritability." *Economica* 46(184): 327–47.
- Herrnstein, Richard J., and Charles Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Kamin, Leon J., and Arthur S. Goldberger. 2002. "Twin Studies in Behavioral Research: A Skeptical View." *Theoretical Population Biology* 61(1): 83-95.

Katz, Lawrence F., and David H. Autor. 1999. "Changes in the Wage Structure and Earnings Inequality." In *Handbook of Labor Economics*, vol. 3A, edited by Orley C. Ashenfelter and David Card. Amsterdam: North-Holland.

Medlund, P., R. Cederlöf, B. Floderus-Myrhed, L. Friberg, and S. Sörensen. 1977. "A New Swedish Twin Registry Containing Environmental and Medical Base Line Data from About 14,000 Same-Sexed Pairs Born 1926-58." *Acta Medica Scandinavica* 600(Suppl): 5-111.

Otto, Sarah P., Freddy B. Christiansen, and Marcus W. Feldman. 1995. "Genetic and Cultural Inheritance of Continuous Traits." Unpublished paper. Stanford University.

Pedersen, N. L., G. E. McClearn, R. Plomin, J. R. Nesselroade, S. Berg, and U. DeFaire. 1991. "The Swedish Adoption/Twin Study of Aging: An Update." *Acta Geneticae Medicae Gemellologiae* 40(1): 7-20.

Solon, Gary. 1999. "Intergenerational Mobility in the Labor Market." In *Handbook of Labor Economics*, vol. 3A, edited by Orley C. Ashenfelter and David Card. Amsterdam: North-Holland.

Taubman, Paul. 1976. "The Determinants of Earnings: Genetics, Family, and Other Environments; A Study of White Male Twins." *American Economic Review* 66(5): 858-70.