

**April 20, 2000**

**Brother Correlations in Earnings in Denmark, Finland, Norway and Sweden  
Compared to the United States\***

Anders Björklund  
SOFI, Stockholm University  
S-106 91 Stockholm  
SWEDEN  
Email: [anders@sofi.su.se](mailto:anders@sofi.su.se)

Tor Eriksson  
Aarhus School of Business  
DK-8210 Århus V  
DENMARK  
Email: [tor@asb.dk](mailto:tor@asb.dk)

Markus Jäntti  
Åbo Akademi  
FIN 20500 Åbo  
FINLAND  
Email: [mjantti@abo.fi](mailto:mjantti@abo.fi)

Oddbjörn Raaum  
Frisch Centre  
N-0349 Oslo  
NORWAY  
Email: [oddbjorn.raaum@frisch.uio.no](mailto:oddbjorn.raaum@frisch.uio.no)

Eva Österbacka  
Åbo Akademi  
FIN 20500 Åbo  
FINLAND  
Email: [eva.osterbacka@abo.fi](mailto:eva.osterbacka@abo.fi)

\* Comments from seminars at Aarhus School of Business, IZA in Bonn and SOFI at Stockholm University are gratefully acknowledged. We thank NOS-S for financial support. The Swedish data collection was also supported by HFR and SFR. We thank Tom Erik Aabø for preparing the Norwegian data, and Esben Agerbo for computational assistance with the Danish data.

**Abstract:** The correlation in economic status among siblings is a useful “omnibus measure” of the overall impact of family and community factors on adult economic status. In this study we compare brother correlations in long-run (permanent) earnings between the United States, on one hand, and the Nordic countries (Denmark, Finland, Norway and Sweden) on the other. Our base case results, based on very similar sample criteria and definitions for all countries, show that this correlation is above 0.40 in the United States and in the range 0.15-0.28 in the Nordic countries. Even though these results turn out to be somewhat sensitive to some assumptions that have to be made, we conclude that the family and community factors are more important determinants of long-run earnings in the United States than in the Nordic countries.

## 1. Introduction

Social scientists from several academic disciplines have long been interested in the association between family background and economic and social status during adulthood. This interest has stemmed largely from the view that inequality attributable to family background violates equal opportunity norms and is a pervasive motive for policy intervention. For this reason, we want to learn about the degree to which family background is related to outcomes during adulthood, whether the connection has changed over time, whether it is larger in some societies than in others, what the causal mechanisms are, and what policies affect the relationship.

In this study, we examine the impact of family and community background on economic status during adulthood by using sibling correlations. We measure outcome using permanent earnings, that is, annual earnings purged of its transitory component. A sibling correlation is a useful “omnibus” measure of the overall impact of family and community background. It can be interpreted as the proportion of the variance in the outcome variable that is attributable to factors that siblings share.<sup>1</sup> Siblings who have grown up together share the same family and community background. This is one reason why a sibling correlation is a broad measure. Strikingly, a sibling correlation is a broader measure than the seemingly more direct association between parents’ and childrens’ outcomes, the reason being that the sibling correlation captures the impact of both observable *and* unobservable parental characteristics.

---

<sup>1</sup> See Solon (1999) for an excellent survey of the empirical literature on intergenerational mobility using earnings and income as outcome measures.

Using data from the Panel Study of Income Dynamics (PSID), Solon, Corcoran, Gordon & Laren (1991) estimated brother correlations in long-run earnings to be around 0.45. In another study, Altonji & Dunn (1991) estimated brother correlations in long-run earnings using the National Longitudinal Survey (NLS) to be 0.37. So in the United States between one third and one half of the variance in men's long-run earnings seems to be attributable to family and community factors. Our aim is to carry out a cross-country comparison of brother correlations in long-run earnings. We start by updating the estimates of the brother correlation in earnings reported by Solon *et al.*; we observe men at a slightly older age and observe earnings over a longer period of time. Our goal is to get comparable estimates from our own countries, namely Denmark, Finland, Norway and Sweden. To achieve this, we use register information for each of our countries to construct large data sets of siblings. We use, more or less, the same sample criteria for all countries and estimate the same parameters for all five countries.

We believe that the U.S.-Nordic comparison is an interesting one. First, it is well known that the countries represent polar cases in comparisons of earnings and income inequality among developed countries. In general, the United States comes out as the most unequal and the Nordic countries as the most equal ones in these respects, so it is interesting to see how such countries compare in terms of a measure of equality of opportunity. Further, all Nordic countries have partly motivated their large public sectors by the desire to reduce inequality of opportunity. Universal access to public health care is one obvious example. The ambition to provide free education of equal quality in the public schools is another. That college education is offered free of tuition is a third example.

Our major finding is that the brother correlation in long-run earnings is higher in the United States than in the four Nordic countries. Our estimates cluster between 0.40 and 0.45 for the United States and in the range 0.14-0.26 for the Nordic countries. Statistical tests suggest that equal correlation in the United States and the Nordic countries can be rejected at conventional levels of significance. We also carry out a number of sensitivity tests to check whether some assumptions regarding sample restrictions and variable definitions affect the results. We do find that estimated brother correlations are sensitive in seemingly innocuous choices. Nevertheless, our overall conclusion is that it is more

likely that the U.S. brother correlation in long-run earnings exceeds those in the Nordic countries than the other way around.

Previous comparative research on the impact of family background has mainly focused on parent-child relationships, and most often some measure of correlation between outcomes of fathers and sons. Although a brother correlation is a broader measure of the total impact of family and community background than a father-son correlation, we note that some recent studies have estimated lower father-son earnings correlations for Finland and Sweden than for the United States.<sup>2</sup> Our results reinforce these findings.

We continue the paper in section 2 by describing our data sets. We explain the model and the estimation technique in section 3. Section 4 gives the empirical results, and we conclude in section 5 by summarizing and discussing possible explanations to our results.

## 2. Data

In defining siblings and in choosing outcome variables for the United States, we closely follow Solon *et al.* (1991). Hence, we define as (social) siblings those children, aged 17 years or younger, who lived in the same PSID household in 1968. We also require that the person is the household head, or the spouse, in the outcome years between 1977 to 1993. Solon *et al.* covered the years 1975 to 1982, so they used a younger sample covered during a shorter period of time.

The *Danish* data set is constructed by merging two longitudinal databases. One is a representative 5 per cent sample of the population aged 15 to 74 in the period 1980-93, which contains detailed information about the individuals' labor market status and earnings for each year (for further information, see <http://www.cls.dk>). The other is called the fertility database and provides detailed demographic information, but also information about other individual characteristics and earnings (see below), about all Danes borne since 1942.

The current sample set up in the following way. The point of departure is the 5 per cent sample. By use of the unique personal identification number, the persons' biological

---

<sup>2</sup> See Björklund & Jäntti (1997) and Jäntti & Österbacka (1995). Björklund & Jäntti(2000) offer a survey of comparative research on intergenerational mobility.

parents and siblings are found in the fertility database, from which also information about some of their background characteristics and their earnings is obtained. As a consequence of the age restriction in the fertility database, only individuals below the age 52 in 1993 can be used. The earnings information comes from tax records, is annual and covers the years 1980-93. All earnings exceeding 5 Danish crowns are recorded as positive earnings.

The *Finnish* data stem from the census in 1970. Persons aged 17 years or younger who lived in the same census household are considered siblings. We use tax register based measures of annual earnings from 1985, 1990 and 1995. See also Jäntti & Österbacka (1995).

The *Norwegian* data are constructed from a complete register of all residents in Norway by 1 January 1993, administered by Statistics Norway. The register is, however, restricted to individuals with parents alive and living in Norway in 1993. For each such individual, the register identifies the biological mother and father. These links enable us to define various biological sibling relations, but the present data are for siblings with the same parents (whole siblings). Because we only have information on biological siblings for Norway and Denmark, we need to assess whether this leads to different results. Therefore, we estimate correlations for both social and biological siblings for the one country for which this is feasible, namely Sweden.

Annual earnings in 1992-1995 are collected from the registers of Statistics Norway. These registers are based on reports from employers, various public offices and tax declarations. Earnings include wages and salaries, earnings from self-employment, and some sick-leave payments.

The *Swedish* data set is entirely based on registers held by Statistics Sweden. The starting point consists of simple random samples from three disjoint populations of persons who lived in Sweden in 1992 and were born between 1951 and 1964. The largest sample (n=100,000) consists of persons who were born in Sweden and were not adopted by neither parent. A second sample (n=3,000) consists of persons who were born in Sweden and were adopted by both parents. A third sample (n=5,000) consists of persons who were born abroad but moved to Sweden before the age of 18. Persons born abroad and adopted by Swedish parents were very few until 1964 and are not parts of the sample we use.

The siblings of these persons are located in two types of registers held by Statistics Sweden. First, “the second-generation register” was used to locate biological whole siblings, biological half siblings on mother’s side (common mother), and biological half siblings on father’s side (common father). Second, we located the households in which the sampled individuals were living as a child (0-17 years of age) in the censuses of 1960, 1965, 1970, 1975 and 1980. We identified other children (same age) in these households and considered them as social siblings. Of course, most of these siblings are also biological. In the final step we added (among others) annual earnings in 1987, 1990, 1993 and 1996 from registers based on employers’ reports for tax purposes.

The earnings data differ between the countries in some respects. First, the PSID questionnaire imposes an upper limit that was \$99,999 from 1977-82, \$999,999 from 1983-91, and \$9,999,999 from 1992-93. No Nordic country applied an upper limit.<sup>3</sup> To achieve a higher degree of comparability, we therefore censor the top one percent of all annual earnings observations to the value of the 99<sup>th</sup> percentile in the earnings distribution. Second, there is no lower earnings limit in any country. Nonetheless, we do believe that there is a difference between the countries that require some treatment of the lowest earnings observations. The Nordic data sets include earnings observations as low as DK5 ( $\approx$  \$0.8) for Denmark, FIM100 ( $\approx$  \$20) for Finland, NOK100 ( $\approx$  \$15) for Norway, and SEK100 ( $\approx$  \$15) for Sweden. Although a rule about a lower earnings limit is not applied by the PSID, it is not likely that respondents report such low annual earnings in the interviews. Inspection of the data revealed that the lowest U.S. earnings observations are considerably higher than the dollar value of the Nordic lowest observations. We therefore decided to truncate the lowest earnings observations to \$100 in 1990 prices. Hence, earnings observations lower than that are treated as missing observations. We apply some sensitivity tests to see if the results are affected by these choices.

Both in terms of individuals and families, the sample sizes for the Nordic countries are much larger than the U.S. one. Another advantage of the Nordic data sets is

---

<sup>3</sup> Finland is an exception, though. In 1985 and 1990, earnings above FIM 310,000 were censored and given the mean value of all the earnings above the limit. In 1985, no one in the sample was affected by the upper limit. In 1990, 268 individuals were affected, i.e., around 1.1 percent of the sample. In 1995, there was no upper limit.

that they do not suffer from the non-response problem that plagues all survey-based data. The smaller U.S. sample in terms of individuals and families is, however, partly compensated by a longer time series of earnings observations. We give details about sample sizes in Table 1 below.

### 3. Models and estimation

In estimating a sibling correlation, we closely follow the previous literature. Let

$$y_{ijt} = X'_{ijt} \beta + \varepsilon_{ijt}, \quad (1)$$

where  $y_{ijt}$  denotes the logarithm of annual earnings in year  $t$  for the  $j$ th sibling in family  $i$ ;  $X_{ijt}$  is a vector of exogenous variables that account for lifecycle stage and time effects with  $\beta$  as the associated vector of coefficients;  $\varepsilon_{ijt}$  is an error term that represents earnings net of lifecycle and general time factors. Because the error term captures the factors that influence the long-run components of earnings, it is the main object of the analysis.

The error term has three components

$$\varepsilon_{ijt} = a_i + u_{ij} + v_{ijt} \quad (2)$$

where  $a_i$  is a permanent component common to all siblings of family  $i$ ;  $u_{ij}$  is an individual-specific permanent component; and  $v_{ijt}$  is a transitory component. In an extended model we allow the transitory component to follow an AR(1) process, i.e.

$$v_{ijt} = \lambda v_{ijt-1} + z_{ijt}. \quad (3)$$

We assume that the three error components are orthogonal. This assumption implies that the individual-specific permanent component is not shared by the siblings of the

same family, but is purely individual. The assumption also implies that the variance of the error term in (1),  $\sigma_\varepsilon^2$ , becomes:

$$\sigma_\varepsilon^2 = \sigma_a^2 + \sigma_u^2 + \sigma_v^2 \quad (4)$$

In this framework, the covariance of a pair of randomly drawn siblings' earnings (purged of lifecycle and time effects) is

$$\text{Cov}(\varepsilon_{ijt}, \varepsilon_{ij't}) = \text{Cov}(\varepsilon_{ij}, \varepsilon_{ij'}) = \sigma_a^2, \quad (5)$$

and the correlation of long-run earnings among siblings is

$$\rho = \sigma_a^2 / (\sigma_a^2 + \sigma_u^2). \quad (6)$$

This expression shows that the sibling correlation has an appealing interpretation within the framework of this model, namely as the proportion of the population variance in long-run earnings that is due to factors shared by siblings. Such factors are to be found both within the family and in the surrounding neighborhood of the family. Our goal is to produce comparable estimates of  $\rho$  for the five countries.

The estimation technique we use is also quite similar to previous studies. In the first step, we estimate equation (1) by OLS. We include a cubic in age and dummies for each outcome year (except one) among the X-variables, and we use real earnings with the national consumer price indexes as deflator. In the second step, we compute the residuals from (1) to estimate the variance of the three error components in (2). These components give us the information needed to estimate the sibling correlation  $\rho$ .<sup>4</sup> We follow the estimation procedure described in detail by Solon *et al.* The standard error of the sibling correlation is obtained by so called "naïve" bootstrapping from the original sample of

---

<sup>4</sup> Note that we do not directly estimate correlation coefficients, but components of earnings variation that within the framework of the model can be interpreted as a correlation coefficient.

families.<sup>5</sup> A copy of a SAS macro for estimation of the variances of the three error components and their sampling distributions are available from the authors upon request.<sup>6</sup>

#### 4. Results

As already mentioned, our sibling definition, choice of age limits and time periods were guided by the U.S. data and the study by Solon *et al.* We also had to make a number of decisions regarding the specific samples to use. For example, on one hand we can only use persons with observed earnings in at least three consecutive years in order to estimate the AR(1) structure of the earnings process. On the other hand, persons with observed earnings in only one year are useful in estimating equation (1) and the composite variance. Further, only a person with a brother in the sample is useful in estimating the variance within a family, whereas singletons are useful in estimating the variance among families. We start by presenting results for a base case in which we require that we observe positive earnings in only one year and where we include singletons. To illustrate how sensitive the results are to these choices, we also present results using other sample restrictions.

*Table 1* contains an overview of the definitions and sample restrictions that we use. The country samples are very close in terms of years of birth and age during outcome years. But there are also important differences. The Danish and Norwegian data sets use biological siblings. The Finnish and Swedish data sets only have earnings for every five and every three years respectively, whereas the other countries have earnings data for each year during a sequence of years. For Norway we only have data for four years, in contrast to 14 years of Danish earnings data and 17 years of U.S. earnings data. These differences motivate some sensitivity analyses that we report below.

It is also interesting to note the much larger sample sizes for the Nordic countries. Nonetheless, our U.S. sample is larger than in Solon *et al.* In their analysis of a sample similar to our base case, they use 1,854 annual earnings observations, 433 individuals (men) and 342 families. Their earnings data covered 1975-82. In our base case sample we have 1,674 individuals from 993 families, and because we observe earnings during 1977-

---

<sup>5</sup> Hinkley (1988) explains bootstrapping techniques. We found that the standard errors stabilized around 100 replications, so we used this number of replications in the bootstrapping.

<sup>6</sup> Please address requests to Markus Jäntti.

93, we get as many as 12,712 annual earnings observations.<sup>7</sup> The prospects for getting better precision in the U.S. estimates than in the previous studies are therefore quite good. The actual sample used is, of course, also affected by the requirement of one positive (> \$100) earnings observation. In the first row of Table 3, we show the sample sizes (in terms of individuals, families and singletons), when only the age and time limits and the overall family restrictions are imposed. In the second row, we show the sample sizes that we get when we impose the additional requirement of one positive earnings. The loss of observations due to this additional requirement is highest for the Norwegian sample, which is reduced by only around 2.5 percent.

In Table 2a we present results for our base case sample that includes all individuals who have positive earnings in at least one year. The transitory error in (2) is assumed to be serially uncorrelated. We find a substantial difference between the United States on one hand, and the Nordic countries on the other. The U.S. brother correlation is estimated to 0.429, whereas the Nordic estimates cluster between 0.138-0.264. The standard errors of the estimated correlations are quite low compared to these differences, so the t-values reported in the last row lead us to reject the hypotheses of equal correlations in the Nordic countries and in the United States.<sup>8</sup> More specifically, we can reject both that the correlation in each single Nordic country is equal to the U.S. correlation and that the average correlation in the Nordic countries equals the U.S. one.

---

<sup>7</sup> We have not used the oversample of the low-income population in PSID's Survey of Economic Opportunity component. Neither did Solon *et al.* in their estimations that are similar to our base case. In their more general analysis, they added this sample and could use 2,656 annual observations on 738 individuals from 582 families.

<sup>8</sup> The same conclusion follows from the confidence intervals that are simulated by the bootstrapping technique.

Table 1. Data and sample characteristics

	Denmark	Finland	Norway	Sweden	United States
1. Sibling definition	Biological siblings from fertility data base	Social siblings in 1970 census	Biological whole siblings from population registers	Social siblings in 1965 census	Social siblings in a 1968 PSID-household.
2. Years of birth.	1951-1968	1953-1965	1950-1970	1948-1965	1951-1967
3. Earnings definition	Annual earnings from registers	Annual earnings From registers	Annual earnings From registers	Annual earnings from registers	Wages and salaries plus the estimated part of farm and business income from interviews
4. Years of outcome.	1980-93	1985, 1990, 1995	1992-1995	1987, 1990, 1993, 1996	1977-93
5. Age interval during outcome years	25-42	25-42	25-42	25-42	25-42
6. Sample size (for base case with (1) positive earnings in at least one year, and (2) top censoring to the 99 <sup>th</sup> percentile). <i>Nobs</i> : no. of positive earnings in any year. <i>Nind</i> : no. of individuals. <i>Nfam</i> : no. of families.	<i>Nobs</i> : 561,386 <i>Nind</i> : 53,911 <i>Nfam</i> : 35,570 <i>Nsingletons</i> : 22,008	<i>Nobs</i> : 62,191 <i>Nind</i> : 24,294 <i>Nfam</i> : 15,703 <i>Nsingletons</i> : 9,565	<i>Nobs</i> : 962,740 <i>Nind</i> : 301,752 <i>Nfam</i> : 198,383 <i>Nsingletons</i> : 118,786	<i>Nobs</i> : 383,152 <i>Nind</i> : 112,966 <i>Nfam</i> : 72,973 <i>Nsingletons</i> : 43,183	<i>Nobs</i> : 12,712 <i>Nind</i> : 1,674 <i>Nfam</i> : 993 <i>Nsingletons</i> : 547
7. Other sample restriction	Resident in Denmark in all outcome years	Resident in Finland in all census years	Resident in Norway by Jan. 1 1993; parents alive and living in Norway in 1993	Resident in Sweden in all outcome years	Household head or spouse in each outcome year

*Notes*: 1. In the base case all earnings observations below \$100 in 1990's dollar are truncated to zero (treated as missing), and the top one percent annual earnings observations are censored to the 99<sup>th</sup> percentile earnings observation in each annual distribution. 2. In the Swedish sample, the "main persons" of the original sample of individuals are 0-14 years old in 1965 (born 1951 to 1965), but brothers are allowed to be 17 years old (born in 1948 or later)

Next, we compare the estimates with those from a model with serially correlated errors. This model's requirement of earnings observations in three consecutive years causes sample reductions for all countries—c.f. rows 2 and 3 in table 3. The largest relative sample size reduction occurs in the Finnish data, in which earnings only are observed in 1985, 1990 and 1995. For this reason, we also estimated the previous model without serially correlated errors on these samples. The results are shown in table 2c. It turns out that the U.S. correlation estimate is practically unaffected by the reduction of the sample. Both the Danish and the Swedish estimates increase by 0.024, the Norwegian by 0.055, and the Finnish by as much as 0.096. The Finnish estimate is no longer significantly different from the U.S. one.

By comparing the estimated brother correlations in tables 2b and 2c, we see how sensitive the results are to extending the model with an AR(1) structure of the transitory error component. Again, the results for the United States and Sweden are not much affected, whereas those for Finland and Norway are affected to a considerable extent. The estimate for Finland goes up from 0.360 to 0.519. It even exceeds the estimate for the United States, but is not significantly different. And the estimate for Norway increases from 0.193 to 0.357. Denmark is a middle case with a decrease from 0.240 to 0.203, almost identical to the estimate in the base case.

With this model and this sample, only the Swedish and the Danish estimates are significantly different from the U.S. one. The average Nordic estimate is not significantly different from the U.S. estimate either.

*Table 2a.* Estimated brother correlations, components of earnings inequality, and tests of equal correlation among countries. Base case. Estimated standard errors within parentheses. Model without autocorrelated errors.

	Denmark	Finland	Norway	Sweden	United States
$\rho$ (sibling correlation)	0.230 (0.011)	0.264 (0.027)	0.138 (0.006)	0.250 (0.012)	0.429 (0.043)
$\sigma_a^2$ (family c.)	0.039	0.077	0.062	0.059	0.165
$\sigma_u^2$ (indiv. c.)	0.129	0.216	0.390	0.176	0.219
$\sigma_v^2$ (transit. c.)	0.148	0.425	0.228	0.328	0.245
$\sigma_a^2 + \sigma_u^2$	0.168	0.293	0.452	0.235	0.384
$\sigma_a^2 + \sigma_u^2 + \sigma_v^2$	0.316	0.718	0.680	0.563	0.629
T-value	U.S. = D: 4.52	U.S. = F: 3.25	U.S. = N: 6.77	U.S. = S: 3.98	U.S. = Nordic: 5.03

*Table 2b.* Model with autocorrelated errors.

	Denmark	Finland	Norway	Sweden	United States
$\rho$ (sibling correlation)	0.290 (0.021)	0.519 (0.116)	0.357 (0.024)	0.283 (0.017)	0.443 (0.059)
$\sigma_a^2$ (family c.)	0.031	0.107	0.073	0.070	0.160
$\sigma_u^2$ (indiv. c.)	0.075	0.099	0.132	0.177	0.202
$\sigma_v^2$ (transit. c.)	0.140	0.484	0.332	0.344	0.228
$\lambda$ (serial corr.)	0.593	0.725	0.546	0.540	0.370
$\sigma_a^2 + \sigma_u^2$	0.106	0.207	0.205	0.247	0.362
$\sigma_a^2 + \sigma_u^2 + \sigma_v^2$	0.246	0.691	0.537	0.591	0.590
T-value	U.S. = D: 2.44	U.S. = F: -0.53	U.S. = N: 1.35	U.S. = S: 2.61	U.S. = Nordic: 1.22

Table 2c. Model without autocorrelated errors.

	Denmark	Finland	Norway	Sweden	United States
$\rho$ (sibling correlation)	0.254 (0.014)	0.360 (0.032)	0.193 (0.008)	0.274 (0.016)	0.427 (0.048)
$\sigma_a^2$ (family c.)	0.032	0.073	0.066	0.055	0.155
$\sigma_u^2$ (indiv. c.)	0.093	0.130	0.277	0.146	0.208
$\sigma_v^2$ (transit. c.)	0.127	0.387	0.201	0.314	0.240
$\sigma_a^2 + \sigma_u^2$	0.125	0.203	0.343	0.201	0.363
$\sigma_a^2 + \sigma_u^2 + \sigma_v^2$	0.252	0.590	0.544	0.515	0.603
T-value	U.S. = D: 3.46	U.S. = F: 1.16	U.S. = N: 4.81	U.S. = S: 3.02	Nordic = U.S.: 3.27

Notes: In Table 2a we use the base case sample in which each individual must have positive earnings (at least \$100) in at least one year. In Table 2b and 2c each individual must have positive earnings in three consecutive years. In testing whether the Nordic correlations are equal to the U.S. correlation, we use the unweighted average of the estimates for the Nordic countries.

The estimated autocorrelation coefficients (the  $\lambda$ 's) are in the range 0.370 (United States) to 0.725 (Finland), and strongly significantly different from zero. It could also be noted that our measures of the composite earnings variance in the last row of tables 2a-2c are not unambiguously lower in the Nordic countries than in the United States. This could indicate a data problem because so many studies have shown that earnings inequality is much higher in the United States. Most likely, the reason to this result is that we have applied a quite low earnings limit in our base case, namely \$100. When we instead apply the limit \$1000 (see below), the composite earnings variance is markedly higher in the US than in the Nordic countries.<sup>9</sup>

<sup>9</sup> In Table 4 below we report estimates of brother correlations with \$1000 as the lower earnings limit. With this restriction, the composite earnings variance is 0.50 for the United States, 0.20 for Denmark, 0.43 for Finland, 0.38 for Norway, and 0.33 for Sweden. Note also that there is a difference between conventional estimates of overall earnings inequality and our composite variance since the latter is life-cycle adjusted.

Table 3. Sample sizes for alternative earnings limits.

		Denmark	Finland	Norway	Sweden	U.S.
Population, i.e. those fulfilling 2, 5 and 7 in Table 1.	Nind:	56,317	24,437	309,261	114,502	1,691
	Nfam:	36,989	15,739	201,688	73,578	1,003
	Nsingletons:	22,860	9,560	119,257	43,209	556
Case 1a. Base case: Earnings $\geq$ \$100 in at least one year and censoring at 99 <sup>th</sup> percentile.	Nind:	53,911	24,294	301,192	112,966	1,674
	Nfam:	35,570	15,703	198,114	72,973	993
	Nsingletons:	22,008	9,565	118,715	43,183	547
Case 1b. Earnings $\geq$ \$100 in at least three consecutive years and censoring at 99 <sup>th</sup> percentile.	Nind:	40,326	15,033	233,584	97,106	1,394
	Nfam:	27,989	11,868	159,238	67,274	865
	Nsingletons:	18,529	9,202	100,470	43,773	513
Case 2. Truncation at \$1000 and censoring at the 99 <sup>th</sup> percentile. Positive earnings in at least one year.	Nind:	52,380	24,080	297,401	112,653	1,670
	Nfam:	34,706	15,615	196,386	72,849	991
	Nsingletons:	21,598	9,552	118,414	43,177	547
Case 3. Truncation at \$100 and no censoring. Positive earnings in at least one year.	Nind:	53,911	24,294	301,192	112,966	1,674
	Nfam:	35,570	15,703	198,114	72,973	993
	Nsingletons:	22,008	9,565	118,715	43,183	547
Case 4. $Y_5 < \Sigma Y(it)/T(I) < Y_{95}$ . Positive earnings in at least one year.	Nind:	38,367	21,852	271,073	105,982	1,510
	Nfam:	26,732	14,582	184,955	70,039	929
	Nsingletons:	17,797	9,242	117,268	42,842	535

Note: In case 4 we start from the base case with truncation at \$100 and censoring at the 99<sup>th</sup> percentile and then define long-run earnings from this distribution. Then we eliminate those individuals with long-run earnings below the 5<sup>th</sup> and above the 95<sup>th</sup> percentiles in the distributions of their long-run earnings.

The results in tables 2a-2c place us in a dilemma. On one hand, we want to avoid the sample reduction caused by the requirement of three consecutive earnings observations. On the other hand, we do not want to use a model that is rejected by the data and that for two countries give markedly higher estimates of the parameter of interest, namely the brother correlation. We have decided, however, to put less emphasis on the results from the model with serially correlated errors. One reason is, of course, the fact that the sample restriction needed to estimate this model raises the estimates markedly for Finland and Norway; this was documented in Table 2a and 2c. Another reason is that only the estimates for Finland and Norway are sensitive to the error specification. We have also reason to believe that estimates from a model that distinguishes between several error components, including a transitory one with serial correlation, is hard to estimate on data with only three years of observations, like for Finland, and four consecutive years, like - for Norway. Solon et al. (1991) also report some problems in estimating a model with serial correlation. Therefore we use the model without serially correlated error terms in what follows.

For at least two different reasons, one could expect that the results also are sensitive to high and low earnings observations. First, one might expect that our treatment of the low and the high earnings observations adversely affect the cross-country comparison. Second, the true relationship between siblings' earnings could be stronger or weaker at the ends of the distributions. Even though we would not be able to distinguish between these two explanations to such a sensitivity in the results, it is useful to know whether the results are sensitive to the exclusion of high and low observations.

To examine how sensitive the results are to our treatment of low and high earnings, we experimented with alternative treatments. Further, we eliminated the individuals with long-run earnings (average earnings in the years in which they were observed) below the 5<sup>th</sup> and above the 95<sup>th</sup> percentile of the distribution of long-run earnings. The estimated sibling correlations from these sensitivity tests are presented in Table 4, and the corresponding sample sizes are shown in Table 3.

The first two rows repeat the estimates from Tables 2a and 2c. In the third row (case 2), we show the estimates that we obtained when we truncated at \$1000 instead of \$100. The correlation falls for the United States and increases for the Nordic countries.

But the magnitudes of these changes are so small that the overall conclusions are not affected.<sup>10</sup> In the next row (case 3), we show estimates based on a sample where we have not censored at the 99<sup>th</sup> percentile. The results are not affected much by this choice either.

In the last row (case 4), we eliminated the bottom and top five percentiles of the observations in the distribution of long-run earnings. In this case, the estimated correlations for Sweden and the United States are practically the same (around 0.34), the estimate for Finland is higher (0.452), and the estimate for Norway is lower (0.186). Overall, the Nordic-U.S. differences are smaller. It is possible that this sensitivity can be explained by measurement errors. A more likely explanation, however, is that the country differences are concentrated to the extreme parts of the distribution of long-run earnings.

The cross-country comparisons could also be sensitive to the fact that we have different periodicity and length in the earnings information for the four countries. In Table 5, we show results where we have re-estimated the Danish and U.S. parameters, using the same periodicity and length as in the Nordic countries. We do find a quite remarkable effect of the periodicity in the U.S. data. There is also an effect in the same direction for Denmark, but the magnitude is smaller. When we re-estimated the brother correlation for the United States with the same periodicity as in the Finnish and Swedish data, we get much higher U.S. estimates: 0.497 with the Swedish periodicity (three years) and 0.559 with the Finnish periodicity (five years). The use of only four years, as in the Norwegian data set, does not have a big impact though. We have no simple explanation for the sensitivity to the periodicity of the earnings observations. The results suggest, however, that the U.S. correlations could be higher—compared to Finland and Sweden—than what follows from the previous comparisons.

---

<sup>10</sup> Note, though, that as mentioned above, the composite variance is more affected for the Nordic countries than for the United States.

Table 4. Estimated brother correlations with alternative earnings limits. Estimated standard errors in parentheses and total variance in brackets.

	Denmark	Finland	Norway	Sweden	United States
Case 1a. Base case: positive earnings in at least one year	0.230 (0.011) [0.316]	0.264 (0.027) [0.718]	0.138 (0.007) [0.680]	0.250 (0.012) [0.563]	0.429 (0.043) [0.647]
Case 1b. Positive earnings in at least three consecutive years	0.254 (0.014) [0.252]	0.360 (0.032) [0.590]	0.193 (0.007) [0.544]	0.274 (0.016) [0.515]	0.427 (0.048) [0.615]
Case 2. Truncation at \$1000 and censoring at the 99 <sup>th</sup> percentile. Positive earnings in at least one year.	0.246 (0.011) [0.202]	0.319 (0.019) [0.427]	0.165 (0.005) [0.382]	0.273 (0.008) [0.332]	0.392 (0.042) [0.501]
Case 3. Truncation at \$100 and no censoring. Positive earnings in at least one year.	0.213 (0.011) [0.579]	0.266 (0.023) [0.724]	0.139 (0.005) [0.679]	0.252 (0.013) [0.568]	0.432 (0.057) [0.641]
Case 4. $Y(5) <$ $\Sigma Y(it)/t(i) < Y(95)$	0.240 (0.016) [0.526]	0.452 (0.032) [0.552]	0.186 (0.006) [0.427]	0.332 (0.015) [0.471]	0.341 (0.053) [0.421]

Notes: In all cases we use the model without autocorrelation. Total variance means the total variance after adjustment for age and period effects, i.e.,  $\sigma_a^2 + \sigma_u^2 + \sigma_v^2$ .

Table 5. Estimated brother correlations for Denmark and the United States using different number of years. Estimated standard errors within parentheses. Base case.

Denmark			United States		
As Norway (four consecutive years)	As Sweden (four years, each of which are three years apart)	As Finland (three years, each of which are five years apart)	As Norway (four consecutive years)	As Sweden (four years, each of which are three years apart)	As Finland (three years, each of which are five years apart)
0.142 (0.037)	0.256 (0.019)	0.306 (0.022)	0.398 (0.055)	0.497 (0.048)	0.559 (0.063)

Notes: Model without autocorrelation. The last four years of the Danish and US data are used to mimic the Norwegian results. 1984, 1987, 1990 and 1993 are used to mimic the Swedish results. 1983, 1987 and 1993 are used to mimic the Finnish results.

Finally, we have checked whether the results are sensitive to the choice of age limits. In particular, it could be argued that annual earnings at the age of 25 to 29 years are not very informative of long-run earnings. Several previous studies have reported that annual earnings in the mid-20s are, at the most, weakly positively correlated with more permanent measures of earnings.<sup>11</sup> In the Nordic countries, youths typically complete education later than in the United States and most men do military service for one year. In Table 6, we report results where we have raised the lower age limit to 30 years. In general, the results are not affected by this change; the estimates are close to those reported in table 2a, and the tests yield the same conclusions.<sup>12</sup>

*Table 6.* Estimations as before but lower age limit 30 years in year of outcome. Base assumption about lower earnings limit. Standard errors within parentheses.

	Denmark	Finland	Norway	Sweden	United States
$\rho$ (sibling correlation)	0.219 (0.018)	0.223 (0.026)	0.142 (0.009)	0.225 (0.012)	0.448 (0.053)
$\sigma_a^2$ (family c.)	0.033	0.084	0.062	0.061	0.197
$\sigma_u^2$ (indiv. c.)	0.116	0.294	0.372	0.210	0.242
$\sigma_v^2$ (transit. c.)	0.086	0.402	0.183	0.296	0.223
$\sigma_a^2 + \sigma_u^2$	0.149	0.378	0.435	0.271	0.447
$\sigma_a^2 + \sigma_u^2 + \sigma_v^2$	0.235	0.781	0.618	0.567	0.680
<i>Nobs:</i>	349,486	45,118	557,786	320,070	7,989
<i>Nind:</i>	32,517	23,446	176,620	112,087	1,364
<i>Nfam:</i>	23,880	15,446	121,876	72,608	858
<i>Nsingletons:</i>	17,029	9,618	78,275	43,135	523
T-value	U.S. = D: 4.06	U.S. = F: 3.81	U.S. = N: 5.71	U.S. = S: 4.11	U.S. = Nordic: 4.59

Note: model without autocorrelation.

The sensitivity analyses that we have presented so far are all motivated by technical issues related to the nature of the data and the exact specification of the model.

<sup>11</sup> See Björklund (1993) for Sweden and, e.g., Gordon (1984) for the United States.

We conclude this section by reporting two additional analyses that are motivated by two conceptual issues involved in the interpretation of sibling correlations. In the Danish and Norwegian data sets, siblings are identified from registers of biological siblings, whereas we use social siblings for Finland, Sweden and the United States. We have no a priori reasons to believe that one sibling type shares more than the other of the family and community factors that influence earnings in adult life. On one hand, biological siblings have some common genetic traits. On the other hand, social siblings have more community factors in common. Furthermore, in our data sets that consist of people born in the 1950s and 1960s, we expect quite a big overlap between social and biological siblings. Nonetheless, we would like to know if the results are affected by the use of different sibling definitions.

Fortunately, the Swedish data set contains the information required to apply both sibling definitions. We used the same age and time restrictions as in the base case and estimated the counterpart to the correlation in the first row of Table 2a for Swedish biological siblings. The point estimate was 0.280 (standard error: 0.012) compared to 0.250 (standard error: 0.012) for social siblings.<sup>13</sup> This is a small difference and it does not give us any reason to believe that the cross-national differences between Denmark and Norway, on one hand, and the United States, on the other, are due to different sibling definitions.

Finally, we address an issue that often arises in comparisons of north European countries and the United States, namely that the United States has a more heterogeneous population. In one respect this is not the case, e.g. the fraction of foreign born is quite high in the Nordic countries. It is highest in Sweden, where around 11 percent of the population are either a foreign citizen or a foreign-born. Nonetheless, the United States has the mix of a black and a white population, and this racial background is definitely shared by brothers. In order to examine whether this mix of the U.S. population affects the overall brother correlation, we estimated the correlation for whites only.<sup>14</sup> The

---

<sup>12</sup> We could also for some countries remove the upper age limit, but the results were not affected by this change either.

<sup>13</sup> The sample size in this estimation was in all dimensions very similar to the one for social siblings.

<sup>14</sup> In the PSID white is defined from the answer by the household head in the interviews during the outcome years.

estimated correlation was 0.321 (standard error: 0.053), i.e. 0.1 lower than for all. This is a considerable difference. More detailed analyses are needed to find out what family and community attributes shared by siblings that can explain this result. It could be, however, that the explanations are the same as those that make the U.S. correlation fall considerably when the top and bottom five percent of observations in the long-run earnings distribution were deleted.

We also divided the Swedish sample into those born abroad and those born in Sweden. The estimates were almost identical to those for the whole sample though; 0.250 for those born in Sweden and 0.242 for those born abroad with our base case assumptions. This result suggests that this kind of population heterogeneity in Sweden is not driving the results. Although, the construction of the Norwegian data set eliminates most immigrants from that sample, we are inclined to conclude that a larger part of immigrants in Sweden cannot explain the difference between Norway and Sweden either.

## **5. Conclusions**

We started by updating previous estimates of the brother correlation in long-run earnings on U.S. PSID data. Our point estimate, 0.429, is very close to the previous ones but thanks to a larger sample, the precision of our estimate is higher. Our goal was to get as comparable estimates as possible for the four Nordic countries, namely Denmark, Finland, Norway and Sweden. We find these countries interesting to compare to the United States, because of the large differences in earnings and income inequality as well as in public policy and labor market institutions. We have collected large samples of siblings making use of register information in the Nordic countries. Using similar sample criteria, we obtained estimates of 0.230 for Denmark, 0.264 for Finland, 0.138 for Norway and 0.250 for Sweden. The differences between the United States on one hand, and each Nordic country, on the other, were strongly statistically significant.

We found that this conclusion is sensitive to one choice regarding model specification. When we allowed the transitory error component to follow an AR(1) process, we obtained estimates for Finland and Norway that were close to, and not significantly different from, the U.S. one. We are, however, inclined to put less emphasis on this feature of our results than in others. There are three reasons for that. First, these

estimates are based on a more restrictive sample, and simply reducing the sample size to accommodate this sample requirement raised the estimates markedly, by 0.1 for Finland. Second, the Finnish data set has only three years of observations and the Norwegian one only four years. Variance estimates of complicated error components from so short panels are not likely to be reliably estimated. Third, another of our sensitivity analyses suggests, that the longer periodicity in the Finnish data (five years) and the Swedish data (three years) tends to yield much higher estimates than from data sets with earnings observations for each successive year, like in Denmark and the United States. For these three reasons, we think that the evidence still points to the United States having a higher correlation.

What could explain such U.S.-Nordic differences in earnings correlations among brothers? We would suggest future research along two alternative lines. First, it could be that family and community background factors are more important determinants of men's productive characteristics in the U.S. than in the Nordic countries. Studies of brother correlations in productivity indicators like health and education would be useful to explore this hypothesis. Of course, correlations, or similar relationships, among other relatives would also be informative.

Second, it could be that the relationship between earnings and productive characteristics is weaker in the Nordic countries than in the U.S. Some comparative studies of earnings inequality are quite suggestive. For example, the return to schooling is much lower in the Nordic countries, see e.g. Asplund & Pereira (1999) and Freeman and Katz (1995). The much more centralized wage setting in the Nordic countries is one candidate explanation to the lower earnings inequality along the schooling dimension.

We would guess that both hypotheses have some explanatory power. In pursuing research along these lines, it may also be fruitful to consider our finding that the similarity among U.S. brothers seems to be particularly strong in the very top and the very bottom of the long-run earnings distribution.

## References

Altonji Joseph G. & Thomas A. Dunn (1991), "Relationships among the Family Incomes and Labor Market Outcomes of Relatives", *Research in Labor Economics*, 12: 269-310.

Asplund Rita & Pedro Telhado Pereira (1999), *Returns to Human Capital in Europe*, ETLA: Helsinki.

Björklund Anders (1993), "A Comparison Between Actual Distributions of Annual and Lifetime Income: Sweden 1951-89", *Review of Income and Wealth*, 39(4): 377-386.

Björklund Anders & Markus Jäntti (1997), "Intergenerational Income Mobility in Sweden Compared to the United States", *American Economic Review*, 87: 1009-1018.

Björklund Anders & Markus Jäntti (2000), "Intergenerational Mobility of Economic Status in Comparative Perspective", *Nordic Journal of Political Economy*, forthcoming.

Freeman Richard B. & Lawrence F. Katz , eds., (1995), *Differences and Changes in Wage Structures*, Chicago: University of Chicago Press and NBER.

Gordon Roger H. (1984), *Differences in Earnings and Ability*, Garland: New York.

Hinkley David V. (1988), "Bootstrap methods", *Journal of Royal Statistical Society*, 50: 321-377.

Jäntti Markus & Eva Österbacka (1995), How Much of the Variance in Income can be Attributed to Family Background?, Åbo Akademi University.

Solon Gary (1999) "Intergenerational Mobility in the Labor Market", in Orley Ashenfelter & David Card (eds.), *Handbook of Labor Economics*, North-Holland.

Solon Gary, Mary Corcoran, Roger Gordon & Deborah Laren (1991), "A Longitudinal Analysis of Sibling Correlations in Economic Status", *Journal of Human Resources*, 26: 509-534.