

Sibling similarities and economic inequality in the US

Bhashkar Mazumder

Received: 22 April 2005 / Accepted: 17 November 2006 / Published online: 12 January 2007
© Springer-Verlag 2007

Abstract I use a new methodological approach and larger US samples than previous studies and estimate that the sibling correlation across a range of economic outcomes is around 0.5. This suggests that half of economic inequality in the US can be attributed to family and community influences. A comparison with noneconomic outcomes suggests that individual choices rather than a simple mechanical relationship governs the intergenerational transmission of income. A decomposition of the sibling correlation suggests that the acquisition of human capital is an important channel through which family background affects future success but that noncognitive factors also play a role.

Keywords Sibling correlation · Intergenerational mobility

JEL Classification J0 · D3 · J62

1 Introduction

A well-known feature of the US economy is its high level of income inequality. Recent research has also documented that intergenerational income mobility in the US is quite low compared to other industrialized countries (Corak 2006) suggesting that family resources during childhood may play an important role in determining future economic success. Few studies, however, have analyzed the contribution of family background as either a direct or indirect influence in explaining income inequality. For example, there might be important characteristics such as cognitive skills, looks, or social contacts that are transmitted by parents through either “nature” or “nurture” that are rewarded by the labor market.

Responsible editor: Christian Dustmann

B. Mazumder (✉)

Federal Reserve Bank of Chicago, 230 South LaSalle Street, Chicago, IL 60604-1413, USA
e-mail: bmazumder@frbchi.org

A useful way to measure the importance of family background is to examine the sibling correlation in economic outcomes. The sibling correlation provides a broad measure of the overall importance of a wide variety of factors common to the family, ranging from parental involvement to school and neighborhood quality.

Only a few previous studies have used large national samples to examine this measure for the US with estimates in the 0.3 to 0.45 range. I contribute to this literature in several ways. I employ variance component models using restricted maximum likelihood (REML) that have desirable statistical properties—most importantly, consistency—lacking in previous analyses. I also use much larger US samples from the National Longitudinal Survey of Youth (NLSY) containing many more siblings than previous work for a more recent set of cohorts. I present results on a variety of outcomes (earnings, family income, wages, and hours worked) for both men and women. I also contrast the correlation in economic outcomes to the correlation in a broad range of noneconomic outcomes.

A key finding is that the sibling correlation in economic outcomes in the US is around 0.5—suggesting that half of economic inequality may be attributed to family background. It also implies that intergenerational economic mobility is quite low. Solon et al. (1991) conduct a simulation to show that such a high sibling correlation implies that less than 7% of individuals who are near the bottom of the distribution of the family component of permanent income are likely to have their *own* earnings or wages surpass the median.

In addition, the sibling correlation in human capital (years of schooling and Armed Forces Qualifying Test [AFQT] test scores) is found to be around 0.6, which is higher than most previous estimates. In contrast, the sibling correlation for a variety of nonhuman capital measures is found to be much lower. It may be especially surprising to note that even measures of physical attributes such as height and weight, which presumably have a strong genetic component, are not as highly associated between brothers as is the permanent component of wages. This strongly suggests that there are factors related to individual or family decision making that lead to a high degree of similarity in the economic fortunes of siblings rather than some simple mechanical relationship. The fact that there is a reasonably strong evidence that the sibling correlation is lower in Nordic countries (Björklund et al. 2002) also suggests that economic mobility may be influenced by institutional differences or policy interventions.

This study also attempts to identify some of the underlying channels by which family and community affects future economic outcomes. Solon (1999) writes, “The mystery of what underlies the considerable resemblance between brothers in their long-run earnings remains a fascinating puzzle and should be a priority for continuing research.” A decomposition analysis finds that human capital acquisition can explain at most about 50% of the sibling correlation in earnings and wages. Other characteristics including “noncognitive” factors emphasized in some recent economic studies (e.g., Osbourne-Groves (2005), Heckman and Rubinstein (2001), and Dunifon et al. (2001)) are also found to be important. These results provide some initial clues for future research.

2 Background and previous US studies

A long literature in sociology and economics has tried to estimate the importance of family background on children's future economic success. Early studies often encountered important data limitations such as small intergenerational samples and missing variables on key family background characteristics. Beginning in the 1970s, researchers began to examine the sibling correlation as an alternative approach to measuring the importance of family background (e.g., Corcoran et al. 1976).

Conceptually, the sibling correlation in economic outcomes provides a summary statistic that captures all of the effects of sharing a common family as well as any other shared factors (e.g., common neighborhoods, school quality). Conversely, many aspects of family background including genetic traits and sibling-specific parental behaviors will not be captured. If the similarity in, say, wages between siblings is not much different compared to randomly chosen individuals, then we would expect a small correlation. If, however, a large fraction of the variance in wages is due to factors common to growing up in the same family environment, then the correlation might be sizable. In that sense, the sibling correlation tells us how much of inequality is due to differences between families.

The earliest studies in the US were conducted before large nationally representative data became available, and they typically used only a *single year* of earnings for each sibling. The results varied widely but were centered at around 0.25 (Solon 1999). Only a few studies have produced estimates of the sibling correlation in *permanent* economic status in the US. Solon et al. (1991) estimate the brother correlation in the permanent component of log annual earnings to be 0.34 when using the nationally representative portion of the Panel Study of Income Dynamics (PSID). The estimate rises to 0.45 when they include the oversample of poor families in the PSID. Björklund et al. (2002) focus on men and use the nationally representative portion of the PSID over a longer time period and produce estimates between 0.42 and 0.45. Both studies use analysis of variance (ANOVA) to produce estimates of variance components which, in turn, are used to assemble estimates of the sibling correlation in the permanent component of log earnings.

There are also two studies that use the NLS original cohort of young men who are tracked from 1966 to 1981. Altonji and Dunn (1991) estimate the correlation in the permanent component of a variety of economic outcomes using both a time-averaging approach and a method-of-moments estimator. Their estimates of the brother correlation in log annual earnings are 0.32 and 0.37. For log hourly wages, their estimates are 0.33 and 0.42. Ashenfelter and Zimmerman (1997) report a brother correlation coefficient of 0.31 in log annual wages averaged over 1978 and 1981. Both studies only include individuals from multiple sibling families and use the oversample of black households without sampling weights. These results appear to support Solon's (1999) conjecture that the brother correlation in permanent status in the US is about 0.4.

Only a few studies have investigated which factors drive the sibling correlation in economic outcomes. Altonji and Dunn (2000) find evidence of linkages between family members (including siblings) in unobserved preferences for work hours using a factor model on US data. Solon et al. (2000) find that little of the sibling

correlation in years of schooling in the US can be explained by neighborhood effects. Rauum et al. (2006) reach a similar conclusion for Norway. Using a Swedish dataset containing specific information on various sibling types, Björklund et al. (2005) decompose the sibling correlation in earnings into genetic and environmental components. The results are sensitive to the specification but suggest that there is a large genetic component.

3 Statistical models and estimation

The following statistical framework based on Solon et al. (1991) is used to measure the sibling correlation in economic outcomes. Each economic outcome (e.g., wages, income) is denoted by y_{ijt} , where i indexes families, j indexes siblings and t indexes years. Outcomes are then modeled as follows:

$$y_{ijt} = \beta X_{ijt} + \varepsilon_{ijt} \quad (1)$$

Here, the vector, X_{ijt} , contains age and year dummies to account for lifecycle effects and year effects such as business cycle conditions. These are treated as fixed effects. The residual, ε_{ijt} , which is purged of these effects, is then decomposed as follows:

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

The three terms on the right hand side of (2) are treated as random effects that are assumed to be independent of each other.¹ The first term, a_i , is the permanent component that is common to all siblings in family i . The second term, u_{ij} , is the permanent component that is individual-specific. v_{ijt} represents the transitory component that reflects noise due to either temporary shocks to earnings or measurement error in the survey.² The variance of age-adjusted earnings, ε_{ijt} , then is simply:

$$\sigma_\varepsilon^2 = \sigma_a^2 + \sigma_u^2 + \sigma_v^2 \quad (3)$$

The first term, σ_a^2 , captures the variance in permanent economic outcomes that is due to differences *between* families, whereas the second term, σ_u^2 , captures the variance in permanent economic outcomes *within* families. These two components are then used to calculate the correlation in permanent outcomes between siblings, ρ , which is the focus of this analysis

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} \quad (4)$$

¹ The assumption that a_i and u_{ij} are uncorrelated is purely for analytical convenience and allows one, conceptually, to divide the permanent component into a part that is perfectly correlated among siblings and a part that is perfectly uncorrelated among siblings. For the assumption that a_i and v_{ijt} are uncorrelated, I find (as did Solon et al. 1991) that there is little or no cross-sectional correlation in the transitory component.

² Allowing for serial correlation in the transitory component has little effect on the results. Readers who are interested in a comparison of estimators for this case may wish to consult Mazumder (2004), which is an earlier version of this paper.

This is also equivalent to the fraction of the overall variance of the permanent components that is due to shared family and community background.

Solon et al. (1991) and (Björklund et al. 2002) use a two-step approach to estimate the variance components in this “mixed model” (mixed because it contains both fixed and random effects). First, they use a regression to estimate (1) and to produce the residuals. Then they use classical ANOVA formulas on the residuals that are adjusted for the fact that the data are “unbalanced” (the number of siblings varies by family, and the number of available years varies by sibling). The adjusted ANOVA formulas may be found in the Appendix of Solon et al. (1991).

Although ANOVA estimators of variance components have some desirable statistical properties for balanced data, virtually none of these properties transfer over to the case of unbalanced data. For this reason, the preferred approach is to use “REML”, which has a number of advantages such as consistency, asymptotic normality, and a known asymptotic sampling dispersion matrix. REML requires an assumption that the data are normally distributed. For many of the outcomes considered here (e.g., log wages, height), this is not likely to be a major factor. For other outcomes such as education, the assumption of normality may be more suspect. Until recent years, computational limitations also made practical implementation of maximum likelihood more difficult.

REML partials out the fixed effects and maximizes the likelihood of the residuals containing the random effects variance–covariance structure. Searle et al. (1992) conclude after an extensive review of approaches for estimating variance components that “It is our considered opinion that for unbalanced data each of ML and REML are to be preferred over any ANOVA method.” REML also appears to be the preferred estimator among quantitative geneticists (Meyer and Hill 1991; Visscher 1998).

A comforting feature of REML is that it produces identical results to ANOVA when the data are balanced. I will present some results using both approaches to show that although the results differ they are not dramatically different. Another nice feature of REML is that it directly produces standard errors of the variance components (standard errors for ρ are calculated by the bivariate delta method).

To understand how different observable characteristics (e.g., parent income, schooling) influence the sibling correlation in economic outcomes, I calculate an estimate of the contribution of various factors. I add the relevant variables to the vector X in (1) and treat them as additional fixed effects in the REML framework. The inclusion of additional fixed effects should sop up some of the residual variation in the outcome variable and produce lower estimates of the family component (σ_a^{2*}) than what was found without their inclusion (σ_a^2). I then take the reduction in the variance of the family component ($\sigma_a^2 - \sigma_a^{2*}$) as an estimate of the amount of the overall variance of the family component that can be attributed to the specific factor (s) in question. This provides an upper-bound estimate of the causal effect because it includes all omitted factors that are also correlated with the included fixed effects. For example, the reduction in (σ_a^2) due to the inclusion of years of schooling would be comprised of both the direct effects of schooling as well as any omitted factors (e.g., perseverance) that also contribute to the outcome variable and are correlated with years of schooling. The change in the variance of the family component divided

by the overall variance of the permanent component tells us what fraction of the overall sibling correlation is due to the factor(s) in question. Implementing this approach for a wide variety of possible explanatory variables, either by including them one at a time or all at once, should tell us something about which measures are critical to explaining the correlation in economic outcomes.

I also investigate the sibling correlation in several noneconomic outcomes where I do not need to examine multiple measurements at different points in time. This requires a far simpler model where I simply drop v_{ijt} from (2) and then use REML to simply calculate the two variance components (σ_a^2) and (σ_u^2) to calculate ρ . In an earlier version of this paper (Mazumder 2004), I compared the REML estimates to a different approach employed by Solon et al. (2000) and found that the results were not very different.

4 Data

The analysis uses the NLSY79 which followed individuals between the ages of 14 and 21 on December 31, 1978 every year from 1979 through 1994 and then every other year. For economic outcomes, I use the NLSY data through the 1998 survey. The survey includes an oversample of black, Hispanic, and (nonblack, non-Hispanic) disadvantaged families. However, the NLSY identifies a nationally representative cross-section of families. I make use of both the full samples by using survey year weights and also use the nationally representative sample without weights.

I use men and women between the ages of 14 and 22 in the initial survey in 1979. I examine four economic outcomes: log annual earnings, log annual family income, log hourly wages, and log annual hours. The outcome variable must be observed and positive at least once when they are at least 26 years old and not enrolled in school. Earnings include wage and salary income as well as business income. I also imposed the following sample restrictions for each outcome; earnings and family income had to be at least \$500 in 1979 dollars; wages had to be at least \$0.50 and no greater than \$100; and annual hours had to be at least 100. The NLSY identifies up to five siblings for each individual. Siblings are identified based on variables that attempt to identify precise family relationships between household members. I do not include nonbiological siblings (e.g., stepbrothers, brothers-in-law, foster brothers, adopted brothers). Only siblings born within the 8-year cohort window are tracked. This restriction on the difference in ages between siblings is very similar to previous studies.³

I also examine a variety of noneconomic outcomes. For education, I measure years of completed schooling by age 26. For test scores, I use the AFQT. The AFQT is part of the Armed Services Vocational Aptitude Battery, a battery of ten tests given

³ Solon et al. (1991) implicitly have a restriction of an 8-year sibling age difference as the cohorts must be born between 1951 and 1958. Similarly, Altonji and Dunn (1991) use cohorts born between 1942 and 1952 who, in practice, are observed no more than eight times when they are at least 24 years old making it relatively rare to observe siblings spaced more than 8 years apart. In principle, (Björklund et al. 2002) may observe more widely spaced siblings; however, they do not report the age distribution within families.

to applicants to the US military. The AFQT score is based on four of the tests that focus on reading skills and numeracy and is a primary criterion for enlistment eligibility. I use the percentile ranking for the renormed (1989 version) score. The test was administered to nearly all respondents in the NLSY in 1980 to provide new norms for the test based on a nationally representative sample. The AFQT is not meant to be viewed as a measure of general intelligence or IQ. For illegal drug use, I use a 1988 survey question asking how many times the respondent has used marijuana or hashish in their lifetime. The responses are presented in five categorical groups. I use the type of residence variable from all survey years starting in 1983 to determine if individuals were ever in jail at the time of the interview. For women, I examine the age of first pregnancy (for the subset of women who were ever pregnant). I look at three measures of physical attributes, height (in 1985), weight at age 28 or 29, and body mass index (BMI). Unfortunately, height is only asked in 1985 so I am unable to hold the age at which height is measured fixed. However, the youngest respondents who are born in 1964 will be 21 by 1985. Note that height is also used to calculate BMI.

Finally, I look at two attitudinal measures from the psychology literature. The first measure I use is the “Rotter scale,” which measures the degree to which individuals feel they have control over their lives. See Osbourne (2000) for a discussion of this measure. The second measure is a self-esteem scale, which combines responses to ten questions designed to determine respondent’s views of self-worth. In all the samples that examine noneconomic outcomes, I require that observations do not have missing data on the relevant outcome.

In most of the analysis, the samples include siblings as well as non-siblings or “singletons” in estimating population variances. I do this to increase efficiency and to maintain comparability with Solon et al. (1991) who had too small a sample of siblings in the PSID to confine the analysis only to multiple sibling families. Solon et al. speculate that including singletons in the analysis may lead to an overestimate of ρ if outliers tend to be more common among singletons than siblings. This is because although singletons’ earnings are used to calculate σ_a^2 , the variance of the family component used in both the numerator and denominator of ρ , they are not included in σ_u^2 , the variance of the individual component, which is only in the denominator of ρ . In the results that follow in the next section, I conduct several robustness checks that include using a sample of only siblings. A set of summary statistics is provided in Table 1.

5 Sibling correlations in economic and noneconomic outcomes

5.1 Economic outcomes

I start by contrasting a few selected results with those obtained by using ANOVA formulas. In Table 2, I present estimates of the variance components and the sibling correlation in annual earnings and hourly wages separately for brothers and sisters. The ‘Brothers’ part shows the results for men, and the ‘Sisters’ part shows the results for women. The estimated brother correlations are 0.492 and 0.536 for earnings and

Table 1 Selected summary statistics for men^a

Year	Number of individuals	Weighted mean age ^b	Log earnings		Log fam. inc.		Log wage		Log annual hours	
			Weighted mean	SD	Weighted mean	SD	Weighted mean	SD	Weighted mean	SD
1983	192	26.0	9.31	0.75	10.14	0.64	1.86	0.52	7.40	0.56
1984	610	26.2	9.57	0.74	10.35	0.60	2.03	0.55	7.61	0.37
1985	1,103	26.7	9.61	0.72	10.22	0.86	2.03	0.59	7.61	0.36
1986	1,601	27.2	9.69	0.71	9.94	0.78	2.10	0.57	7.64	0.36
1987	2,224	27.7	9.76	0.67	10.01	0.76	2.13	0.56	7.65	0.37
1988	2,803	28.2	9.75	0.76	10.01	0.83	2.15	0.63	7.65	0.38
1989	3,405	28.7	9.77	0.71	10.02	0.79	2.15	0.59	7.66	0.37
1990	3,516	29.2	9.74	0.73	10.00	0.80	2.14	0.62	7.64	0.40
1991	3,793	29.8	9.79	0.90	10.04	0.96	2.15	0.64	7.63	0.40
1992	3,674	30.8	9.80	0.72	10.06	0.81	2.16	0.60	7.66	0.39
1993	3,687	31.8	9.80	0.76	10.08	0.82	2.16	0.61	7.67	0.38
1995	3,594	33.8	9.89	0.75	10.18	0.83	2.21	0.65	7.70	0.39
1997	3,365	35.8	9.97	0.75	10.24	0.83	2.28	0.66	7.71	0.39

^a Log earnings, log family income and log wages are expressed in 1983 dollars. The statistics for the outcomes pertain to those used in column 1 of Table 3.

^b These pertain to the earnings sample.

Table 2 Sibling correlations in earnings and wages, REML vs ANOVA^a

	REML		ANOVA	
	Earnings	Wages	Earnings	Wages
Brothers				
Family component	0.194 (0.013)	0.114 (0.008)	0.157 (0.015)	0.103 (0.011)
Individual component	0.201 (0.012)	0.099 (0.006)	0.208 (0.015)	0.116 (0.010)
Transitory component	0.206 (0.002)	0.161 (0.001)	0.215 (0.006)	0.169 (0.004)
Correlation	0.492 (0.024)	0.536 (0.022)	0.430 (0.039)	0.470 (0.047)
Observations	33,567	31,528	33,567	31,528
Individuals	5,213	5,102	5,213	5,102
Families	4,303	4,215	4,303	4,215
Singletons	3,542	3,471	3,542	3,471
Sisters				
Family component	0.164 (0.020)	0.071 (0.009)	0.149 (0.021)	0.075 (0.012)
Individual component	0.318 (0.020)	0.127 (0.009)	0.312 (0.021)	0.131 (0.010)
Transitory component	0.295 (0.003)	0.180 (0.002)	0.306 (0.008)	0.187 (0.006)
Correlation	0.340 (0.037)	0.360 (0.033)	0.323 (0.044)	0.363 (0.054)
Observations	28,177	27,263	28,177	27,263
Individuals	4,994	4,956	4,994	4,956
Families	4,267	4,245	4,267	4,245
Singletons	3,648	3,635	3,648	3,635

^a Standard errors use the delta method for REML and are bootstrapped for ANOVA. The sample includes the oversample of minority and poorer households using sampling weights.

wages, respectively. Although these results are similar to the higher set of estimates in Solon et al. (1991), they are more robust due to the significantly larger number of observations used. For example, for men, Solon et al. use around 2,500 observations, 750 individuals, and 600 families, whereas this analysis uses 30,000 observations, 5,000 individuals, and 4,000 families.

The results are higher for both outcomes when using REML, although the differences are not statistically significant. As REML is known to produce consistent estimates, and as the NLSY samples are quite a bit larger, these results suggest that the consensus view of a sibling correlation in log earnings of around 0.4 expressed in Solon (1999) perhaps ought to be revised up to near 0.5. The results also suggest that the correlation in log wages appears to be higher than 0.5, which is also consistent with Solon et al.’s (1991) finding. In any case, these results reinforce the main point in Solon et al. (1991) that accounting for the transitory variance is critical. If I were to use data from just a single year, the implied correlation in brothers’ wages would be just 0.306.

To ensure that the difference in estimates by statistical method was not due to a dataset effect, I also conducted an analysis using the PSID. I first attempted to replicate the baseline results in Solon et al. (1991) using ANOVA. Compared to their

estimate of a sibling correlation in log annual earnings of 0.342, I obtain an estimate of 0.310. When I use REML instead, my estimate increases to 0.385. The increase of 0.075 is roughly comparable to the magnitude of the increases I find for men in Table 1 when moving from ANOVA to REML (0.062 for earnings and 0.066 for wages). Therefore, I conclude that the higher estimates from REML are not a result of using any particular dataset.

The results for the sister correlation in earnings and wages are somewhat lower at 0.340 and 0.360, respectively. This is not so surprising because women's labor force participation patterns during their 20s and 30s are very different from men and may produce much noisier estimates of long-term economic status for these outcomes. Indeed, both the permanent individual component and the transitory components are dramatically higher than for men. I will look at the correlation in family income, a broader measure of economic status, in the next set of results. The ANOVA results are a bit lower for women's earnings but are actually higher for women's wages.

In Table 3, I confine the results to men and only present REML estimates. I present the same estimates for earnings and wages as in Table 2, but I now add family income and annual hours as additional outcomes. In column 1, I employ the

Table 3 Brother correlation in economic outcomes^a

	REML estimates for men			
	Whole sample weighted include singletons	Rep. sample unweighted include singletons	Whole sample weighted no singletons	Rep. sample unweighted no singletons
Earnings				
Correlation	0.492	0.464	0.494	0.459
(SE)	(0.029)	(0.037)	(0.030)	(0.039)
Obs	33,567	19,518	10,723	6,516
Ind'ls	5,213	2,791	1,671	950
Families	4,303	2,279	761	438
Family income				
Correlation	0.466	0.423	0.468	0.418
(SE)	(0.030)	(0.039)	(0.031)	(0.040)
Obs	32,371	18,998	10,576	6,402
Ind'ls	5,164	2,796	1,680	953
Families	4,248	2,282	764	439
Wages				
Correlation	0.536	0.509	0.535	0.519
(SE)	(0.029)	(0.037)	(0.031)	(0.039)
Obs	31,528	18,642	10,087	6,198
Ind'ls	5,102	2,756	1,631	931
Families	4,215	2,255	744	430
Annual hours				
Correlation	0.392	0.398	0.374	0.371
(SE)	(0.043)	(0.051)	(0.043)	(0.050)
Obs	33,511	19,621	10,833	6,568
Ind'ls	5,179	2,785	1,660	942
Families	4,275	2,278	756	435

^a Estimates are produced using REML, standard errors use the delta method.

base specification, which uses the full sample of men including singletons, and weight the sample with survey year weights. The brother correlation in family income is estimated to be 0.466 or slightly lower than the correlation in earnings, 0.492. As family income includes spouse income, it may be that assortative mating acts to lower the correlation between brothers in this outcome. The correlation in annual hours is estimated to be just under 0.4, which is similar to the reported results in Solon et al. (1991) and the method of moment results in Altonji and Dunn (1991).

In Table 3, I also test whether the results are sensitive to the use of weights on the oversample of poorer families. In column 2, I use only the sample of families that are identified as nationally representative in 1979. This results in keeping a little more than half the observations. The estimates drop only slightly for three of the four outcomes and are virtually identical for the fourth.

I also experiment with confining the sample to only individuals from families with multiple siblings. In column 3, I show that compared to column 1, the results are virtually identical when using a sibling-only sample. This is true although only about a third of the total observations are used. In column 4, I combine both conditions (a nationally representative sample of only siblings). This strongly suggests that including singletons has little effect on the results. I have also experimented with changing the age cutoff and found that the results are not sensitive to these perturbations. For example, if I restrict the age of observation to be at least 28, the estimate of the brother correlation in wages falls slightly from 0.54 to 0.53. For those aged at least 30, the estimate falls to 0.51. One issue is that raising the age cutoff also necessitates dropping some of the sample years. For example, an age restriction of 34 would confine my sample to just 5 years, 1991–1993, 1995, and 1997.

In Table 4, I do the same analysis for women. The estimates for the correlation in family income among sisters are in indeed much higher than for the other outcomes. In fact, they are virtually identical to the correlation between brothers. For the nationally representative samples (columns 2 and 4), the correlation in family income is actually higher among sisters than among brothers (Table 3). Chadwick and Solon (2002) similarly find that the intergenerational elasticity in family income for daughters is close to that found for sons. The correlation in annual hours is very low but not surprising given the wide variance in labor force participation. Having established that, with respect to the most comprehensive measure of economic status, family income, the sibling correlation is essentially the same for men and women, I proceed with the most of remaining analysis focusing just on men.

5.2 Sibling correlations in noneconomic outcomes

Table 5 presents the results for a variety of noneconomic outcomes for brothers, sisters, and all siblings as appropriate. This serves as a useful point of contrast with the economic outcomes and is also interesting in its own right. The analysis includes singletons using only the nationally representative portion of the NLSY. Including the oversample of poorer and minority households without weights results in similar estimates. I found, however, that when I used weights on the noneconomic outcomes, it sometimes led to implausibly large estimates. Therefore, I chose to only use the representative sample without weights when studying noneconomic outcomes.

Table 4 Sister correlation in economic outcomes^a

	REML estimates for women			
	Whole sample weighted include singletons	Rep. sample unweighted include singletons	Whole sample weighted no singletons	Rep. sample unweighted no singletons
Earnings				
Correlation	0.340	0.301	0.333	0.287
(SE)	(0.040)	(0.050)	(0.041)	(0.050)
Obs	28,177	17,013	7,660	4,952
Ind'ls	4,994	2,781	1,346	816
Families	4,267	2,339	619	374
Family income				
Correlation	0.453	0.438	0.466	0.428
(SE)	(0.033)	(0.041)	(0.034)	(0.043)
Obs	31,038	19,290	8,669	5,749
Ind'ls	5,195	2,906	1,435	879
Families	4,420	2,431	660	404
Wages				
Correlation	0.360	0.318	0.357	0.301
(SE)	(0.042)	(0.053)	(0.043)	(0.053)
Obs	27,263	16,528	7,388	4,810
Ind'ls	4,956	2,767	1,321	806
Families	4,245	2,333	610	372
Annual Hours				
Correlation	0.149	0.149	0.140	0.142
(Se)	(0.048)	(0.059)	(0.048)	(0.057)
Obs	29,434	17,628	8,091	5,130
Ind'ls	5,068	2,817	1,383	833
Families	4,321	2,366	636	382

^a Estimates are produced using REML, standard errors use the delta method.

The correlation in years of schooling appears to be roughly similar for both brothers and sisters, and the correlation across all sibling pairs is about 0.6. This is slightly higher than the estimates in Solon et al. (2000). For AFQT scores, the estimates are also similar across genders and are even higher than the education estimates. These estimates are similar to those obtained by Oettinger (1999) who also uses the NLSY. I next focus on a few socioeconomic outcomes that have been commonly analyzed in studies of neighborhood/peer effects (e.g., Case and Katz 1991). I find that the correlation in drug use among all siblings is a bit below 0.3 with a slightly higher point estimate for sisters (0.37) than brothers (0.3), although the difference is not statistically significant. The fact that the overall correlation is lower than the correlation within gender type suggests that the correlation across siblings of different genders is lower. When I examine whether respondents were ever in jail, the estimate for sisters is zero. It is worth noting that as variance component models, by definition, are bounded at zero, REML cannot produce negative estimates. For brothers, the correlation estimates are around 0.25. This sharp distinction by sex is similar to that found by Duncan et al. (2001) in their analysis of measures of delinquency among teenagers. Finally, the correlation in age at pregnancy is about 0.3.

Table 5 Sibling correlation in selected noneconomic outcomes^a

Corr. (SE) <i>N</i>	All	Brothers	Sisters
Yrs of schooling	0.602 (0.014) 6,097	0.622 (0.023) 3,000	0.602 (0.026) 3,097
AFQT scores	0.619 (0.014) 5,751	0.620 (0.024) 2,811	0.618 (0.024) 2,940
Illegal drug use	0.268 (0.023) 5,413	0.300 (0.040) 2,639	0.371 (0.041) 2,774
Ever in jail	0.152 (0.021) 6,111	0.263 (0.035) 3,003	0.000 – 3,108
Age of pregnancy	–	–	0.305 (0.044) 2,495
Height	–	0.492 (0.032) 2,803	0.467 (0.035) 2,937
Weight	–	0.330 (0.039) 2,652	0.293 (0.039) 2,768
Body mass index	–	0.265 (0.044) 2,580	0.301 (0.039) 2,696
Rotter scale	0.094 (0.021) 6,053	0.067 (0.039) 2,974	0.112 (0.043) 3,079
Self-esteem	0.248 (0.021) 5,809	0.219 (0.042) 2,860	0.281 (0.038) 2,949

^a Estimates are produced using REML with the nationally representative subsample and do not use survey weights. Standard errors use the delta method.

I now turn to physical characteristics/health outcomes. The correlation in height between siblings is slightly below 0.5, whereas the estimates for weight are around 0.3 for brothers and sisters. The height correlation is similar to what has been found in previous studies (e.g., Duncan et al. 2001). For BMI, I find that the correlation among sisters (0.30) is slightly higher than among brothers (0.26). These estimates are similar to the estimates of sibling correlations in cholesterol levels and blood pressure (Lee et al. 2003). Finally, with the attitudinal variables, the correlation in the Rotter scale is only about 0.1 in all cases, whereas the correlation in self-esteem is in the 0.2 to 0.3 range. These are similar to estimates of “extraversion” and “emotional stability” reported in Loehlin and Rowe (1992).

Overall, it appears that the correlations in the human capital measures are actually the highest at around 0.6. Otherwise, the only variable with a sibling correlation comparable to the economic outcomes is height. Other outcomes that presumably have a large genetic component such as weight, BMI, and personality characteristics are considerably lower and correspond to findings in the existing literature. This strongly suggests that there are factors related to individual or family decision

making, particularly with respect to schooling, that lead to a high degree of similarity in the economic fortunes of siblings.

6 Contributions to the brother correlation in economic outcomes

I now examine the potential impact of various explanatory variables on the sibling correlation in economic outcomes among men. Table 6 presents estimates of the contribution of various variables using the methodology described in “[Background and previous US studies](#)”. A growing literature on intergenerational economic inequality (e.g., Solon 1999; Mazumder 2005) has emphasized the importance of parent or family income on children’s economic outcomes. Therefore, one obvious candidate for explaining the sibling correlation is parent income. With the NLSY79 sample, there is only information for a subset of individuals for just a few years on parent income. As Solon (1992) has shown, using income from just a single year is a poor proxy for permanent income and leads to downward-biased coefficients. Similarly, using just a two-year average of income from 1978–1979 will also likely to lead to biased estimates of the residuals (purged of parent income) and, therefore, biased estimates of the variance components. In any case, using this proxy for parent permanent income, I find that the variance in the family component in earnings residuals is reduced by 0.17, which explains about 36% of the sibling correlation in earnings. Interestingly, Solon (1999) shows that using the consensus estimates of the

Table 6 Contributions to the brother correlation in economic outcomes^a

Upper-bound estimates of the contribution to the correlation from...

	Earnings (contrib)	Percent	Family income (contrib)	Percent	Wages (contrib)	Percent	Hours (contrib)	Percent
Parent income	0.169	36	0.178	41	0.145	27	0.077	21
Human capital	0.248	51	0.261	57	0.252	47	0.070	20
Education	0.215	44	0.223	48	0.209	39	0.062	16
AFQT scores	0.217	45	0.233	51	0.219	41	0.066	18
Physical char.	0.027	5	0.035	7	0.024	5	0.015	4
Height	0.024	5	0.031	7	0.023	4	0.013	3
Weight	0.006	1	0.011	2	0.002	0	0.007	2
BMI	0.001	0	0.009	2	0.004	1	0.000	0
Illegal behavior	0.108	22	0.111	24	0.043	8	0.112	2
Jail	0.112	23	0.122	26	0.042	8	0.112	2
Illegal drug use	0.001	0	0.006	1	0.005	1	0.003	1
Psychological char.	0.094	20	0.096	21	0.097	18	0.029	8
Rotter scale	0.050	10	0.054	12	0.058	11	0.009	2
Self-esteem	0.073	15	0.076	17	0.068	12	0.027	8
Occupation	0.289	59	0.288	62	0.291	54	0.204	52
All except occupation	0.311	65	0.310	73	0.294	56	0.140	45
All	0.382	80	0.360	84	0.380	71	0.230	73
	REML estimates using full samples and weights							
Table 3 estimate	0.492		0.466		0.536		0.392	

^a Estimates are produced using REML. Samples include the oversample of poor and minority households and use survey weights. Individuals with missing data on the relevant characteristic are dropped from the analysis.

intergenerational elasticity in earnings of 0.4, one would expect the contribution to be $(0.4)^2$ or 0.16. If, however, the intergenerational elasticity is actually closer to 0.6 (Mazumder 2005), then this is a vast underestimate of the true contribution of parent income. The contribution of parent income to the sibling correlation in family income is slightly higher at 41%. The contributions to the sibling correlation in wages and hours are 27 and 21%, respectively.

Economic models of wage determination strongly emphasize the importance of human capital. Consequently, I next examine how years of schooling and AFQT scores influence the sibling correlation in economic outcomes. Both measures fare almost equally well across the outcomes. For earnings, family income, and wages, each measure explains anywhere between 40 and 50% of the sibling correlation. Including both human capital measures in conjunction explains more than half of the sibling correlation in earnings and family income. These human capital measures, however, only explain about 20% of sibling correlation in annual hours.

Many studies have also emphasized the importance of physical characteristics such as height and appearance on wages, so it is interesting to understand the importance of these variables in explaining the sibling correlation in economic outcomes. Interestingly, physical characteristics only account for about 5% of the sibling correlation, most of which is due to the inclusion of height. Accounting for any time spent in jail explains more than 20% of the sibling correlation in earnings, family income, and hours but less than 10% of the sibling correlation in hourly wages. Illegal drug use, however, appears to explain virtually none of the sibling correlation. Both psychological measures make an important contribution to explaining the sibling correlation in earnings, family income, and wages. Combined, they account for about 20% of the sibling correlation in these measures. This result adds to the growing research that has found that noncognitive factors such as personality traits play an important role in the intergenerational persistence in economic status.

Finally, I also try to explain the importance of occupation by including three-digit occupation dummies. This is more controversial, as occupation is often viewed as an

Table 7 Conditional contributions to the brother correlation in wages^a

REML estimates using full samples and weights						
Upper-bound estimates of the contributions to the brother correlation in wages from...						
	Unconditional (%)	Conditional on parent inc. (%)	Conditional on human cap. (%)	Conditional on physical char. (%)	Conditional on illegal behav.	Conditional on psych. char.
Parent income	27	–	6	26	24	20
Human capital	47	25	–	42	41	31
Physical characteristics	5	1	0	–	4	2
Illegal behavior	8	2	0	8	–	4
Psychological characteristics	18	8	1	16	13	–

^a Estimates are produced using REML. Samples include the oversample of poor and minority households and use survey weights. Individuals with missing data on the relevant characteristic are dropped from the analysis.

outcome rather than a causal factor determining economic success. Still, occupation may act as a proxy for other forms of human capital or capture social connections that perpetuate intergenerational inequality. In any case, occupation appears to account for around 60% of the sibling correlation in earnings, family income, and wages.

Including all of the variables at the same time accounts for 80% or more of the brother correlation in earnings and family income and upwards of 70% of the brother correlation in wages and hours worked. Excluding the occupation dummies, 65% of the correlation in earnings between brothers and 73% of the correlation in family income can be accounted for by these variables. On the other hand, just under half of the sibling correlation in wages is still unexplained even after including all of these variables.

The fact that the variables are likely to be correlated with one another is a limitation of this decomposition approach. It would be interesting to know for example, the importance of parent income *controlling* for my measures of human capital. To get a sense of the conditional contributions to the brother correlation in wages, Table 7 shows how the contributions for each broad category of factors are affected if one controls for each of the other factors, one at a time. For example row 1 of Table 7 shows that the unconditional contribution of parent income (27%) is reduced to 6% if one controls for years of schooling and AFQT scores. These results continue to show that the human capital measures contribute the most followed by parent income and then psychological characteristics.

7 Conclusion

This study uses an improved estimation approach on a much larger sample than previous studies to bolster the findings of previous economic research that has demonstrated that family and community influences account for a large portion of the variation in economic outcomes in the US. I find, for example, that the sibling correlation in men's permanent wages is greater than 0.5, suggesting a high degree of economic rigidity. Notably, this correlation is even higher than the correlation in height and weight, suggesting that the "inheritance" of economic inequality is particularly strong.

Using a decomposition analysis, I find that observable measures of family and individual characteristics can explain a large portion of the sibling correlation in earnings and family income. Economic models of wage determination typically emphasize human capital acquisition as a key variable, and it is clear that this is one important way by which families confer advantages to their offspring. Parent income also plays an important role. I also find that psychological characteristics play a smaller, although quantitatively important role. Although this analysis provide some initial clues for which mechanisms are important, future research using more convincing research designs are needed to better guide policymakers.

Acknowledgment I thank David I. Levine, Gary Solon, Dan Sullivan, Kristin Butcher, and three anonymous referees for helpful comments. I also greatly appreciate the outstanding research assistance provided by Sara Christopher. The views expressed here do not represent the views of the Federal Reserve System.

References

- Altonji JG, Dunn TA (1991) Relationships among the family incomes and labor market outcomes of relatives. In: Ehrenberg RG (ed) *Research in labor economics*, vol 12. JAI, Greenwich, CT, pp 269–310
- Altonji JG, Dunn TA (2000) An intergenerational model of wages, hours and earnings. *J Hum Resour* 35(2):221–257
- Ashenfelter O, Zimmerman D (1997) Estimates of the returns to schooling from sibling data: fathers, sons and brothers. *Rev Econ Stat* 79(1):1–9
- Björklund A, Eriksson T, Jantti M, Raaum O, Osterbacka E (2002). Brother correlations in earnings in Denmark, Finland, Norway and Sweden compared to the United States. *J Popul Econ* 15(4):757–772
- Björklund A, Jantti M, Solon G (2005) Influences of nature and nurture on earnings variation. In: Bowles S, Gintis H, Osbourne-Groves M (eds) *Unequal chances: family background and economic success*. Princeton University Press, Princeton, NJ, pp 145–164
- Case AC, Katz LF (1991) The company you keep: the effects of family and neighborhood on disadvantaged youths. NBER Working Paper 3705
- Chadwick L, Solon G (2002) Intergenerational income mobility among daughters. *Am Econ Rev* 92(1):335–344
- Corak M (2006) Do poor children become poor adults? Lessons for public policy from a cross country comparison of generational earnings mobility. *Res Econ Inequal* 13 (1):143–188
- Corcoran M, Jencks C, Olneck M (1976) The effects of family background on earnings. *Am Econ Rev* 66(2):430–435
- Duncan GJ, Boisjoly J, Harris KM (2001) Sibling, peer, neighbor, and schoolmate correlations as indicators of the importance of context for adolescent development. *Demography* 38(3):437–447
- Dunifon R, Duncan GJ, Brooks-Gunn J (2001) As ye sweep, so shall ye reap. *Am Econ Rev* 91(2):150–154
- Heckman JJ, Rubinstein Y (2001) The importance of noncognitive skills: lessons from the GED testing program. *Am Econ Rev* 91(2):145–149
- Lee KE, Klein B, Klein R (2003) Familial aggregation of components of the multiple metabolic syndrome in the Framingham Heart and Offspring Cohorts: genetic analysis workshop problem 1. *BMC Genet* 4 (Suppl 1):S94
- Loehlin JC, Rowe DC (1992) Genes, environment, and personality. In: Chapara G, Van Heck GL (eds) *Modern personality psychology: critical reviews and new directions*. Harvester Wheatsheaf, New York
- Mazumder B (2004) Sibling similarities, differences and economic inequality. Federal Reserve Bank of Chicago Working Paper 2004–13
- Mazumder B (2005) Fortunate sons: new estimates of intergenerational mobility in the U.S. using social security earnings data. *Rev Econ Stat* 87(2):235–255
- Meyer K, Hill WG (1991) Approximation of sampling variances and confidence intervals for maximum likelihood estimates of variance components. *J Anim Breed Genet* 109:264–280
- Oettinger GS (1999) Does the sibling correlation in economic status vary across family and sibling pair. Manuscript, University of Texas
- Osbourne M (2000) The power of personality: labor market rewards and transmission of earnings. Manuscript, University of Massachusetts
- Osbourne-Groves M (2005) Personality and the intergenerational transmission of earnings. In: Bowles S, Gintis H, Osbourne-Groves M (eds) *Unequal chances: family background and economic success*. Princeton University Press, Princeton NJ
- Rauum O, Salvanes KG, Sørensen EØ (2006) The neighborhood is not what it used to be. *Econ J* 116(1):200–222
- Searle SR, Casella G, McCulloch C (1992) *Variance components*. Wiley, New York
- Solon G (1992) Intergenerational income mobility in the United States. *Am Econ Rev* 82(3):393–408
- Solon G (1999) Intergenerational mobility in the labor market. In: Ashenfalter OC, Card D (eds) *Handbook of labor economics*, vol 3A. Elsevier, Amsterdam, North Holland
- Solon G, Corcoran M, Gordon R, Laren D (1991) A longitudinal analysis of sibling correlations in economic status. *J Hum Resour* 26(3):509–534
- Solon G, Page ME, Duncan GJ (2000) Correlations between neighboring children in their subsequent educational attainment. *Rev Econ Stat* 82(3):383–392
- Visscher PM (1998) On the sampling variance of intraclass correlations and genetic correlations. *Genetics* 149(3):1605–1614