NBER WORKING PAPER SERIES

THE EFFECT OF FAMILY BACKGROUND ON ECONOMIC STATUS: A LONGITUDINAL ANALYSIS OF SIBLING CORRELATIONS

Gary Solon

Mary Corcoran

Roger Gordon

Deborah Laren

Working Paper No. 2282

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1987

This research was supported by a grant from the Ford Foundation to the Institute of Public Policy Studies at The University of Michigan. The authors are grateful for advice from John Bound, Charles Brown, Leslie Kish, James Lepkowski, and seminar participants at Northwestern University and the Universities of Michigan, Toronto, Chicago, and Western Ontario. The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #2282
June 1987

The Effect of Family Background on Economic Status: A Longitudinal Analysis of Sibling Correlations

ABSTRACT

Numerous previous studies have used sibling correlations to measure the importance of family background as a determinant of economic status. These studies, however, have been biased by several flaws: failure to separate permanent from transitory status variation (including that from measurement error), failure to account for life-cycle stage, and overly homogeneous samples. This paper presents a methodology to address these problems and applies it to longitudinal data from the Panel Study of Income Dynamics. Our main conclusion is that family background appears to exert greater influence on economic status than has been indicated by earlier research.

Gary Solon
Department of Economics
The University of Michigan
Ann Arbor, MI 48109-1220
(313) 763-1306

Mary Corcoran
Department of Political Science
The University of Michigan
Ann Arbor, MI 48109-1220
(313) 764-9517

Roger Gordon
Department of Economics
The University of Michigan
Ann Arbor, MI 48109-1220
(313) 764-6769

Deborah Laren
Institute for Social Research
The University of Michigan
Ann Arbor, MI 48109-2321
(313) 936-0098

The Effect of Family Background on Economic Status: A Longitudinal Analysis of Sibling Correlations

1. Introduction

The influence of family background on economic status has persistently interested social scientists and others concerned with social policy. This interest has stemmed largely from a belief that income inequality attributable to family background violates equal opportunity norms and warrants government intervention. Michael Harrington's influential book *The Other America*, for example, based its call for antipoverty efforts on just such a premise:

> ... the real explanation of why the poor are where they are is that they made the mistake of being born to the wrong parents, in the wrong section of the country, in the wrong industry, or in the wrong racial or ethnic group. Once that mistake has been made, they could have been paragons of will and morality, but most of them would never even have had a chance to get out of the other America.[1]

Similar (though less colorful) language has appeared in scholarly studies. The foreword to John Brittain's *The Inheritance of Economic Status*, for instance, says his findings "constitute impressive evidence of the effect of family background on a man's chances for economic success; they imply that inherited advantage, over which the individual has no control, is important; and they bolster the case for a public policy aimed at redistribution of income and wealth."[2]

One device social scientists have used for measuring the importance of family background is to estimate sibling correlations in economic status. The idea is that, if family background matters very much, siblings will show a strong resemblance in economic status; if it matters hardly at all, they will show little more resemblance than would randomly selected unrelated individuals. In Section 2 of this paper, we formalize this idea with a statistical model which also shows that previous sibling-based estimates of family background's influence on income variables have been biased (probably downward in most cases) by a series of shortcomings: measurement error, failure to distinguish permanent and transitory income, failure to adjust for life-cycle stage, and overly homogeneous samples. We then present a simple methodology for addressing these problems with longitudinal data and apply it to data from the Panel Study of Income Dynamics. In Section 3, we develop a more sophisticated extension of the methodology and report the corresponding results. We summarize our findings in Section 4.

2. A Simple Model

We begin with a simple statistical framework for analyzing sibling correlations. Let

$$(1) \qquad y_{ijt} = \beta' X_{ijt} + \epsilon_{ijt}$$

where y_{ijt} is some economic status measure in year t for the j^{th} sibling in the i^{th} family, X_{ijt} is a vector of explanatory variables chosen to account for life-cycle stage and general time effects (such as price inflation, business-cycle conditions, and cohort effects), and β is the associated vector of regression parameters. The error term ϵ_{ijt}, which then represents economic status after abstracting from life-cycle and time factors, is the object of our analysis.

We view ϵ_{ijt} as the sum of three components:

$$(2) \qquad \epsilon_{ijt} = a_i + u_{ij} + v_{ijt}$$

where a_i is a permanent component common to the siblings in the i^{th} family, u_{ij} is an individual-specific permanent component *not* shared by siblings in the same family, and v_{ijt} is a transitory component generated by some combination of measurement error and true transitory fluctuation in ϵ_{ijt}. With general time effects already accounted for, v_{ijt} is assumed to be cross-sectionally independent,[3] though it may be serially correlated. The three error components are to be interpreted so that they are orthogonal by construction. For example, the permanent component of an individual's status is conceptually partitioned into the part a_i that is perfectly correlated among siblings in the i^{th} family and the part u_{ij} that is perfectly uncorrelated.

We then view σ_ϵ^2, the stationary population variance of ϵ_{ijt}, as the sum of the variances of its three components:

(3) $$\sigma_\epsilon^2 = \sigma_a^2 + \sigma_u^2 + \sigma_v^2.$$

This variance decomposition is reminiscent of the numerous studies in the earnings function literature, such as Lillard and Willis (1978), which have broken earnings variance into permanent and transitory components. The present analysis further decomposes the permanent component into sibling and family components. The population variance in *permanent* status then is $\sigma_a^2 + \sigma_u^2$, and the correlation of permanent status among siblings is

(4) $$\phi = \sigma_a^2/(\sigma_a^2+\sigma_u^2).$$

This correlation ϕ tells, in a sense, what proportion of the population variance in permanent status is attributable to family background. It does so by describing what proportion of the variance is common to siblings. The particular sense in which ϕ summarizes family background's importance is broad in certain respects and narrow in others. On one hand, *anything* shared by siblings — not just parental characteristics, but also community characteristics such as school quality and status of neighbors — is included in the measured family background effect. On the other hand, some factors often thought of as family background factors are left out. For example, those genetic traits *not* shared by siblings are excluded. Also excluded are family or community factors that differ among siblings because the siblings are raised at different times or because the parents choose to treat the siblings differently.[4]

Nevertheless, sibling correlations have considerable appeal as omnibus measures of the importance of family background, and numerous researchers have used them as such. Corcoran, Jencks, and Olneck (1976), for example, estimated brother correlations for the natural logarithm of earnings and obtained values of .13, .21, and .22 in three different data sets. Similar analyses by other researchers usually have produced correlations in the .10–.35 range.[5]

Previous estimates, however, have been biased by several flaws. First, most studies have relied on only one observation of annual earnings (or some other status measure) per person. As a result, their status measures contain transitory components (including measurement error) along with the permanent components. Consequently, the brother correlations in these studies are effectively estimates of

$$(5) \qquad \gamma = \sigma_a^2/(\sigma_a^2+\sigma_u^2+\sigma_v^2)$$

rather than ϕ. As a comparison of (4) and (5) makes clear, the failure to distinguish transitory and permanent status tends to obscure the importance of family background in permanent status. To put it another way, by using only single-year earnings data, previous studies have tended to underestimate the brother correlation in permanent status by a factor of

$$(6) \qquad \gamma/\phi = (\sigma_a^2+\sigma_u^2)/(\sigma_a^2+\sigma_u^2+\sigma_v^2).$$

Second, many previous studies have relied not on representative national samples, but on remarkably homogeneous samples. For example, Olneck's (1977) study is based on brothers from Kalamazoo, and Kearl and Pope (1986) use data on Mormon brothers in 19th-century Utah. Studies that draw their samples from such homogeneous sets of families will tend to produce small estimates of σ_a^2 and therefore will again tend to underestimate the true ϕ for the overall U.S. population.

Third, most previous studies have not adjusted their status measures for stage of life-cycle.[6] Insofar as the siblings in a sample tend to be closer together in age than are randomly selected sample members, this failure will tend to overestimate sibling resemblance in permanent status. Since the age range in most samples has been fairly narrow, though, this bias probably is minor in most cases. We therefore expect that the biases toward underestimation typically dominate.

The remainder of this section outlines a strategy for addressing these estimation problems with longitudinal data from the Panel Study of Income Dynamics (PSID). We focus on a sample of "splitoffs" — children in the original 1968 PSID families who have since become heads of households (or spouses of heads) in their own families. This sample, of course, contains multiple siblings from the same families. The longitudinal nature of our data on splitoffs' adult status enables us to address the first estimation problem of separately identifying permanent and transitory variation. The national representativeness of the PSID alleviates the second problem (overly homogeneous samples), and regression adjustments described below address the third (life-cycle stage).

The general procedure is as follows. First we use least squares methods to estimate equation (1) where X_{ijt} is a vector of age and year variables. We then use the regression residuals

$$e_{ijt} = y_{ijt} - \hat{\beta}X_{ijt} \tag{7}$$

as consistent estimates of ϵ_{ijt}, the variable of interest. Next we apply analysis-of-variance methods to e_{ijt} to estimate the three variance components: σ_a^2, σ_u^2, and σ_v^2. Finally, substitution of these estimates into (4), (5), and (6) yields estimates of ϕ, γ, and γ/ϕ.

If our sample contained an equal number of siblings J per family and an equal number of annual observations T per sibling, if the transitory component v_{ijt} were serially uncorrelated, and if the PSID were a simple random sample, we could estimate the variance components with the classical analysis-of-variance formulas:

$$(8) \qquad \hat{\sigma}_v^2 = \sum_i \sum_j \sum_t (e_{ijt} - \bar{e}_{ij})^2 / [NJ(T-1)]$$

where N is the number of families in the sample and $\bar{e}_{ij} = \sum_t e_{ijt}/T$;

$$(9) \qquad \hat{\sigma}_u^2 = \sum_i \sum_j (\bar{e}_{ij} - \bar{e}_i)^2 / [N(J-1)] - (\hat{\sigma}_v^2/T)$$

where $\bar{e}_i = \sum_j \sum_t e_{ijt}/JT$; and

$$(10) \qquad \hat{\sigma}_a^2 = \sum_i (\bar{e}_i - \bar{e})^2/(N-1) - (\hat{\sigma}_u^2/J) - (\hat{\sigma}_v^2/JT)$$

where $\bar{e} = \sum_i \sum_j \sum_t e_{ijt}/NJT$. In effect, this procedure identifies the transitory variance on the basis of observed year-to-year variation in the same individual's status, then identifies the permanent variance among siblings in the same family on the basis of observed within-family variation, and finally identifies the permanent variance across families on the basis of observed between-family variation.

In fact, however, none of the required conditions for the classical formulas is met. Our sample contains different numbers of siblings J_i per family and different numbers of observations T_{ij} per person; numerous previous longitudinal studies have found evidence of serial correlation in the transitory component of earnings; and the PSID is a stratified multistage survey. We will defer treatment of the latter two problems to the next section, but we will close this section by illustrating our approach with an analysis that does treat the first problem.

This analysis focuses on men's annual earnings. The sample contains male splitoffs from the Survey Research Center component of the PSID[7] who were between the ages of 10 and 17 at the outset of the survey in 1968. Our data consist of annual earnings observations over the 1975–82 period, and we restrict our sample observations to positive earnings reported when the men were at least 25 years old. These data were collected in interviews from 1976 through 1983 and pertain to the preceding calendar years. Men that were under age 10 in 1968 are excluded from our sample because they had not reached age 25 by 1983. Men that were over age 17 in 1968 and entered the PSID because they were still living in their parents' households are excluded to avoid overrepresenting men that left home at late ages. The resulting sample contains a total of 1854 observations on 433 men from 342 families.

Our first step was to perform ordinary least squares estimation of equation (1) where the dependent variable y_{ijt} was the natural logarithm of annual nominal earnings and X_{ijt} was a vector of year and age dummy variables.[8] We then subjected the resulting residuals e_{ijt} to an analysis of variance. Our estimators of the variance components, listed in Appendix A, extend the classical formulas (8)-(10) to the case of varying J_i and T_{ij}.

The resulting estimates of the respective variance components are $\hat{\sigma}_a^2 = .119$, $\hat{\sigma}_u^2 = .229$, and $\hat{\sigma}_v^2 = .237$.[9] Substitution of $\hat{\sigma}_a^2$ and $\hat{\sigma}_u^2$ into (4) estimates the brother correlation in permanent log earnings to be $\hat{\varphi} = .342$. As expected, the use of longitudinal data to estimate the brother correlation in *permanent* status produces an estimate higher than most in the previous literature. This point is highlighted by substituting all three estimated variance components into (5) to obtain $\hat{\gamma} = .203$. This amounts to an estimate of the brother correlation in a single year's observation of status and indeed does correspond more closely to the previous estimates that relied on single-year data. The ratio of $\hat{\gamma}$ to $\hat{\varphi}$ suggests that using only single-year data underestimates the brother correlation in permanent status by a factor of .595.[10]

3. A More General Model

The results above are based on a procedure that assumes simple random sampling and serial noncorrelation of the transitory component v_{ijt}. With respect to the latter assumption, several previous studies — such as Lillard and Willis (1978), MaCurdy (1982), and Abowd and Card (1986) — have found evidence of substantial serial correlation in the transitory component of earnings. We therefore modify our model to allow v_{ijt} to follow the first-order autoregressive process

$$(11) \qquad v_{ijt} = \rho v_{ij,t-1} + z_{ijt}$$

where z_{ijt} is serially uncorrelated and $\sigma_v^2 = \sigma_z^2/(1-\rho^2)$. Substituting (11) into (2) and then subtracting ρ times the once-lagged version of (2) yields

$$(12) \qquad \epsilon_{ijt} - \rho\epsilon_{ij,t-1} = (1-\rho)a_i + (1-\rho)u_{ij} + z_{ijt}.$$

If a consistent estimate of ρ is available, one can apply the same analysis-of-variance procedures described in Section 2 to $e_{ijt} - \hat{\rho}e_{ij,t-1}$ to obtain estimates of the three variance components for (12): $(1-\rho)^2\sigma_a^2$, $(1-\rho)^2\sigma_u^2$, and σ_z^2. It is then straightforward to solve for the implied values of σ_a^2, σ_u^2, and σ_v^2.

To follow this procedure, one needs an estimate of ρ. A simple method is to regress the differenced residuals $e_{ijt} - e_{ij,t-1}$ on their lagged values $e_{ij,t-1} - e_{ij,t-2}$. As shown in Solon (1984), the resulting coefficient estimate r consistently estimates the first-order autocorrelation of $\epsilon_{ijt} - \epsilon_{ij,t-1}$, which is $(\rho-1)/2$. Inverting the relationship between r and ρ, one can then estimate ρ with $\hat{\rho} = 1 + 2r$.

The assumption of simple random sampling raises two other issues. First, the PSID sample actually was generated by a multistage survey design that randomly selected geographic clusters of households. This departure from simple random sampling causes bias in the variance component estimators described above, but a simple extension of the analysis in Section 9.3 of Cochran (1963) shows that, with the large number of clusters in the PSID, this bias becomes negligible. As will be discussed later, however, the nonindependence of observations in the same cluster is one of several complicating factors for the estimation of standard errors. Second, the full PSID sample contains both the Survey Research Center (SRC) sample used in Section 2 and an additional sample of low-income families drawn from the 1967 Survey of Economic Opportunity (SEO). The SEO sample was added in a deliberate effort to oversample the low-income population. Excluding SEO families from our analysis in Section 2 resulted in an unnecessarily small sample size, and we will include them in this section's analysis. Their inclusion, however, raises a new problem. With oversampling of low-income families, the sample distribution of economic status becomes systematically different from the population distribution, and therefore our variance component estimators based on observed sample variances become inconsistent. To correct this problem, we henceforth apply our estimation procedures to data weighted by the inverse of the probability of selection into the sample.[11] The details of the estimation procedures are given in Appendix A.

We apply these procedures first to men's log annual earnings. The sample selection criteria are the same as in Section 2 except that we now use both the SRC and SEO components of the PSID. The initial (unweighted) sample size is 3533 annual observations on 855 individuals from 651 families, but the "ρ-differencing" procedure eliminates one observation per individual and also eliminates individuals who start with only one observation. The sample size for the differenced data is 2656 observations on 738 individuals from 583 families. Slightly over half these observations come from the SRC sample. Of the 738 sample individuals, 448 are "singletons," i.e., either they have no brothers or their brothers do not meet the sample selection criteria. The other 290 sample individuals consist of 116 brother pairs, 18 triples, and 1 quadruple.

The first column of Table I shows the estimation results. The first-order serial correlation ρ of the transitory component is estimated at .479, a finding similar to those in previous longitudinal studies of earnings. The analysis of variance produces the variance component estimates $\hat{\sigma}_a^2 = .148$, $\hat{\sigma}_u^2 = .183$, and $\hat{\sigma}_v^2 = .267$. Substitution of $\hat{\sigma}_a^2$ and $\hat{\sigma}_u^2$ into (4) yields an estimated brother correlation in permanent status of $\hat{\phi} = .448$. This estimate, even higher than the one in Section 2, suggests a distinctly larger influence of family background than implied by most earlier siblings studies. Again, it is interesting to substitute all three estimated variance components into (5) to estimate the brother correlation in single-year (rather than permanent) status. The resulting $\hat{\gamma} = .248$ is fairly typical of the estimates in previous studies based on single-year data. The ratio $\hat{\gamma}/\hat{\phi} = .554$ implies that measuring the brother correlation in single-year status underestimates the correlation in permanent status by almost half.

Our estimate of γ is based on a complex estimation procedure that uses multiple years of panel data. It is worth asking what we would obtain if, like previous studies, we actually used only single-year data to estimate σ_a^2, the sum $\sigma_u^2 + \sigma_v^2$, and the ratio $\gamma = \sigma_a^2/(\sigma_a^2+\sigma_u^2+\sigma_v^2)$. We have performed such estimation[12] for each of the years 1979 through 1982, and the resulting estimates of γ range from .150 for 1981 to .511 for 1979. The surprising instability of the estimates across years led us to examine the data more closely. What we discovered is that the variance component estimators are highly sensitive to outliers.[13] In particular, the estimates are greatly influenced by the log earnings of the few individuals who earned less than several hundred dollars in a year.[14] This problem becomes particularly important when the outliers happen to be singletons. A brother outlier inflates the estimates of both σ_a^2 and $\sigma_u^2 + \sigma_v^2$ so that the estimate of the ratio γ is not dramatically affected. A singleton outlier, however, inflates only $\hat{\sigma}_a^2$ (singletons provide no information on within-family variation and therefore do not appear in the estimator for $\sigma_u^2 + \sigma_v^2$) and hence substantially inflates $\hat{\gamma}$ as well. This suggests the possible advisability of excluding singletons (as most previous studies have done). With singletons excluded, our single-year estimates of γ do indeed compress to a narrower range — from .235 for 1981 to .342 for 1982.[15] These estimates tend toward the higher end of the range from previous studies, but remain well below $\hat{\phi} = .448$, our estimated brother correlation in *permanent* status.

Beyond its magnitude relative to earlier estimates, our estimate of ϕ is difficult to interpret in practical terms. We therefore present, in the first column of results in Table II, answers to the question: If an individual's family background component is in some specified decile of the family background distribution, in what percentile rank of the permanent log earnings distribution is the individual's expected permanent log earnings? The figures in the table are based on our variance component estimates and the additional (and questionable[16]) assumption that the two permanent components a_i and u_{ij} are both normally distributed. The results suggest a startlingly important influence of family background. For example, an individual whose family background is in the bottom decile has expected permanent log earnings in only the 12^{th} percentile of the permanent log earnings distribution.

Although these findings have substantively important implications, their reliability depends on the precision of our estimation. The complexity of our estimation procedure combined with the complexity of the PSID sample design has prevented us from analytically deriving standard errors for our estimators. A correct analytical solution would have to take account of the imprecision in our estimation of the original regression adjustment coefficients β and the serial correlation parameter ρ as well as the imprecision from the analysis-of-variance procedure itself. It also would have to account for the apparent nonnormality of ϵ_{it}, the nonindependence of observations from different families in the same geographic cluster, and the weighting necessitated by the oversampling of the low-income population.[17] To bypass these difficult problems, we instead have used a nonparametric "balanced half-sample replication" approach to produce standard error estimates. This approach — described in detail in Kish and Frankel (1970), McCarthy (1969), and Wolter (1985) — repeatedly applies the entire estimation procedure to a succession of strategically chosen half-samples. Each estimator's observed variance across the half-sample replications is then used to infer an estimate of the variance of that estimator as applied to the full sample. Our application of this procedure is described more fully in Appendix B.

As shown in Table I, some of the standard error estimates obtained in this manner turn out to be quite large. Most importantly, the estimated brother correlation in permanent status ($\hat{\phi}$) has an estimated standard error of .295. This is much larger than the standard error estimates reported for $\hat{\gamma}$ in previous studies. even those studies based on brothers samples no larger than our own. Our large standard errors seem to arise from two factors. The first is simply that we have tackled a more difficult estimation problem than that undertaken by earlier studies — we have attempted separate identification of the permanent and transitory individual-specific components in order to estimate a brother correlation for unobserved permanent status rather than for observed single-year status. Given the obvious difficulty of distinguishing a permanent component from a serially correlated transitory component on the basis of only a few years of panel data, the imprecision of our estimation should not be surprising. The second factor is the previously mentioned sensitivity of the estimation to singleton outliers. In an effort to treat this problem, we have repeated our analysis with singletons excluded. The resulting point estimates are fairly similar to those in Table I ($\hat{\sigma}_a^2 = .114$, $\hat{\sigma}_u^2 = .198$, $\hat{\sigma}_v^2 = .248$, $\hat{\phi} = .367$, $\hat{\gamma} = .205$, $\hat{\gamma}/\hat{\phi} = .557$), but the estimated standard errors become even larger. The reason is that the robustness gain from excluding singletons is more than offset by the efficiency loss from deleting most of the individuals in our sample. To paraphrase MacDonald and Robinson (1985), the exclusion of singletons throws the sample size out with the bathwater.

One important parameter we do estimate with some precision is the ratio γ/ϕ, estimated at .554 with estimated standard error .082. Exact confidence intervals are precluded by our ignorance of the distribution of $\hat{\gamma}/\hat{\phi}$, but we can make conservative statements on the basis of the Chebyshev inequality.[18] According to the Chebyshev inequality, no random variable, regardless of its distribution, can deviate from its mean by more than k standard deviations with a probability greater than $1/k^2$. Thus, if .082 is a good estimate of the standard error, then .554 ± 3 (.082) is a conservative .89 confidence interval around $\hat{\gamma}/\hat{\phi}$. The upper end of this range, .800, still implies that studies using only single-year data are biased toward substantial underestimation of the brother correlation in permanent status. To put it another way, .554 is more than five estimated standard errors below unity. We expect therefore that a hypothesis that single-year analyses are only negligibly biased could be rejected at any conventional significance level.

Thus far we have focused on men's earnings, the variable considered in most earlier work. We also have applied our estimation procedure to several other variables related to economic status: log hourly wage, log annual hours of work, log family income, and log family income/needs for men as well as the same two family income variables for women. The hourly wage measure is the ratio of annual earnings to annual hours, and the income/needs ratio is family income divided by a needs standard related to the official poverty line for the individual's family size.[19] The results, shown in columns 2–7 of Table I, are qualitatively similar to those for men's earnings. The sibling correlations in permanent status are generally high compared to those in previous studies; they range from .276 for women's log family income to .534 for men's log wage. The computing expense of implementing the balanced half-sample replications prevented us from estimating standard errors, but our experience with men's earnings suggests that many of the standard errors are probably quite large.

Finally, columns 2–7 of Table II repeat the exercise of interpreting the estimated sibling correlations in terms of percentile ranks. Of course, the results differ across variables in accordance with the differences in estimated sibling correlations, but in all cases the estimated influence of family background remains strikingly large.

4. Summary

Previous studies of sibling correlations in economic status have been biased by several flaws: failure to separate permanent from transitory status variation (including that from measurement error), failure to account for life-cycle stage, and overly homogeneous samples. We have presented a methodology to address these problems and have applied it to longitudinal data from the Panel Study of Income Dynamics. The resulting point estimates of sibling correlations in permanent status suggest a greater sibling resemblance than indicated by earlier studies. Unfortunately, standard error estimates based on balanced half-sample replications reveal that these point estimates are very imprecise. Nevertheless, the evidence of substantial downward bias in the previous studies that used single-year data is overwhelming. Our main substantive conclusion, then, is that family background appears to exert greater influence on economic status than has been suggested by earlier research.

The policy implications of this finding are far from obvious. First, whether income inequality attributable to family background *can* be affected by policy intervention depends on just what it is about siblings' shared background that matters. To address this issue, we are beginning research on the relationship between the PSID splitoffs' status and a wide variety of characteristics of their parents and the communities in which the splitoffs grew up. A reliable assessment of which characteristics are important, however, will be difficult in the face of multicollinearity among the characteristics, measurement error, and omitted-variable problems. Second, even if the influential aspects of background can be identified, whether the resulting income inequality *ought* to be affected by policy intervention will depend inevitably on value judgments concerning what an equitable income distribution is and the extent to which efficiency losses should be suffered to achieve it.

Table I

Estimated Parameters (and Standard Errors) from Full PSID Sample

Parameters	Men					Women	
	Log Earnings	Log Wage	Log Hours	Log Income	Log Income/Needs	Log Income	Log Income/Needs
ρ	.479 (.095)	.372	.449	.329	.297	.417	.361
$(1-\rho)^2\sigma_a^2$	.040 (.027)	.039	.013	.046	.074	.027	.062
$(1-\rho)^2\sigma_u^2$	.050 (.046)	.034	.019	.088	.080	.071	.061
σ_z^2	.206 (.029)	.105	.137	.119	.109	.173	.150
σ_a^2	.148 (.103)	.100	.043	.102	.149	.079	.153
σ_u^2	.183 (.097)	.087	.062	.196	.161	.208	.149
σ_v^2	.267 (.051)	.122	.171	.133	.119	.210	.172
ϕ	.448 (.295)	.534	.410	.342	.481	.276	.507

(Table I continued)

Parameters	Men					Women	
	Log Earnings	Log Wage	Log Hours	Log Income	Log Income/Needs	Log Income	Log Income/Needs
γ	.248 (.160)	.323	.156	.237	.348	.159	.322
γ/ϕ	.554 (.082)	.605	.380	.691	.723	.577	.637
Sample Size:*							
Observations	2656	2656	2656	2764	2764	3272	3272
Individuals	738	738	738	752	752	875	875
Families	583	583	583	592	592	673	673

*These sample sizes pertain to the "ρ-differenced" observations.

Table II

Percentile of Expected Permanent Economic Status Given Percentile of Family Background Component

Percentile of Family Background Component	Men					Women	
	Log Earnings	Log Wage	Log Hours	Log Income	Log Income/Needs	Log Income	Log Income/Needs
0 – 10	12	10	13	14	11	18	11
10 – 20	24	22	25	26	23	29	23
20 – 30	33	31	33	34	32	36	32
30 – 40	40	39	40	41	39	42	39
40 – 50	47	46	47	47	47	47	46
50 – 60	53	54	53	53	53	53	54
60 – 70	60	61	60	59	61	58	61
70 – 80	67	69	67	66	68	64	68
80 – 90	76	78	75	74	77	71	77
90 – 100	88	90	87	86	89	82	89

Appendix A

The estimators used in Section 2 are:

(13) $$\hat{\sigma}_v^2 = \sum_i \sum_j \sum_t (e_{ijt} - \bar{e}_{ij})^2 / (\sum_i \sum_j T_{ij} - \sum_i J_i)$$

where $\bar{e}_{ij} = \sum_t e_{ijt} / T_{ij}$;

(14) $$\hat{\sigma}_u^2 = \frac{\sum_i \sum_j T_{ij}(\bar{e}_{ij} - \bar{e}_i)^2 - (\sum_i J_i - N)\hat{\sigma}_v^2}{\sum_i \sum_j T_{ij} - \sum_i \left[\frac{\sum_j T_{ij}^2}{\sum_j T_{ij}}\right]}$$

where $\bar{e}_i = \sum_j \sum_t e_{ijt} / \sum_j T_{ij}$; and

(15) $$\hat{\sigma}_a^2 = \frac{\sum_i (\sum_j T_{ij})(\bar{e}_i - \bar{e})^2 - \left[\sum_i \left(\frac{\sum_j T_{ij}^2}{\sum_j T_{ij}}\right) - \frac{\sum_i \sum_j T_{ij}^2}{\sum_i \sum_j T_{ij}}\right]\hat{\sigma}_u^2 - (N-1)\hat{\sigma}_v^2}{\sum_i \sum_j T_{ij} - \frac{\sum_i (\sum_j T_{ij})^2}{\sum_i \sum_j T_{ij}}}$$

where $\bar{e} = \sum_i \sum_j \sum_t e_{ijt} / \sum_i \sum_j T_{ij}$.

For the analysis of data weighted by w_{ij}, the inverse of the individual's probability of selection into the sample,[20] the above equations can be modified as follows:

$$(16) \qquad \hat{\sigma}_v^2 = \sum_i \sum_j \sum_t w_{ij}(e_{ijt} - \bar{e}_{ij})^2 / \sum_i \sum_j w_{ij}(T_{ij} - 1);$$

$$(17) \qquad \hat{\sigma}_u^2 = \frac{\displaystyle\sum_i \sum_j w_{ij} T_{ij} (\bar{e}_{ij} - \bar{e}_i)^2 - \left[\sum_i \sum_j w_{ij} - \sum_i \left(\frac{\sum_j w_{ij}^2 T_{ij}}{\sum_j w_{ij} T_{ij}} \right) \right] \hat{\sigma}_v^2}{\displaystyle\sum_i \sum_j w_{ij} T_{ij} - \sum_i \left[\frac{\sum_j (w_{ij} T_{ij})^2}{\sum_j w_{ij} T_{ij}} \right]}$$

and, in equation (15) for $\hat{\sigma}_a^2$, T_{ij} is everywhere replaced by $w_{ij}T_{ij}$, and the coefficient of $\hat{\sigma}_v^2$ changes from N-1 to

$$\sum_i \left(\frac{\sum_j w_{ij}^2 T_{ij}}{\sum_j w_{ij} T_{ij}} \right) - \frac{\sum_i \sum_j w_{ij}^2 T_{ij}}{\sum_i \sum_j w_{ij} T_{ij}}.$$

In the analysis in Section 3, these weighted estimators are applied not to e_{ijt}, but to the "ρ-differenced" e_{ijt}. The sample counts T_{ij}, J_i, and N must then be modified to account for the loss of one observation per individual and the loss of individuals who had only one observation initially.

Appendix B

To facilitate half-sample replications, the Survey Research Center has characterized the PSID sample as consisting of two independent "primary selections" from each of 32 strata. The pair of selections in the k^{th} stratum might be, say, the PSID samples from the Milwaukee and Minneapolis areas. The coding of these pairs is described on pages 89-90 and 310-11 of Survey Research Center (1985). A half-sample comprised of only one selection from each of the 32 strata more or less duplicates the complex survey design of the PSID, but at only about half the size.

We used the 32×32 Hadamard matrix on page 325 of Wolter (1985) to select a set of 32 "balanced" half-samples. For any parameter μ, if $\hat{\mu}$ denotes the estimate from the full sample and $\hat{\mu}_k$ the estimate from the k^{th} half-sample, we estimate the variance of $\hat{\mu}$ with

$$(18) \qquad \widehat{\text{Var}}\,(\hat{\mu}) = \sum_{k=1}^{32} (\hat{\mu}_k - \hat{\mu})^2/32.$$

Why is this a sensible estimator of Var $(\hat{\mu})$? Let $\hat{\mu}_{k'}$ denote the estimate of μ from the complement of the k^{th} half-sample, and suppose $\hat{\mu} = (\hat{\mu}_k + \hat{\mu}_{k'})/2$, as is exactly true if $\hat{\mu}$ is a linear estimator and is likely to be approximately true otherwise. Then, for any arbitrary half-sample k,

$$
\begin{aligned}
(19) \qquad E(\hat{\mu}_k - \hat{\mu})^2 &= E[\hat{\mu}_k - (\hat{\mu}_k + \hat{\mu}_{k'})/2]^2 \\
&= E[(\hat{\mu}_k - \hat{\mu}_{k'})/2]^2 \\
&= E(\hat{\mu}_k - \hat{\mu}_{k'})^2/4 \\
&= E[(\hat{\mu}_k - \mu) - (\hat{\mu}_{k'} - \mu)]^2/4 \\
&= 2\ \text{Var}\,(\hat{\mu}_k)/4 \\
&= \text{Var}\,(\hat{\mu}_k)/2
\end{aligned}
$$

$$= \text{Var}\ (\hat{\mu}).$$

Thus, for any particular half-sample k, the squared deviation of $\hat{\mu}_k$ from $\hat{\mu}$ is an approximately unbiased estimator of Var ($\hat{\mu}$). The point of taking 32 different half-samples and averaging the squared deviations of the $\hat{\mu}_k$ from $\hat{\mu}$ is to improve the precision of the variance estimator. The optimal method of choosing "balanced" half-sample replications is discussed in detail in McCarthy (1969) and Wolter (1985).

Footnotes

[1]Harrington (1962), p. 21.

[2]Brittain (1977), pp. vii-viii.

[3]In response to comments from Angelo Melino and Kevin Murphy, we have checked the assumption that the contemporaneous sibling correlation in the transitory component v_{ijt} is zero. Based on the relationship between brothers' earnings changes in 1981–82, we have estimated that correlation to be .128. Some rough calculations suggest that, if .128 is the true sibling correlation in v_{ijt}, then our subsequent estimation of the brother correlation in permanent status under the assumption of cross-sectional independence of v_{ijt} is biased upward, but by only a small magnitude. We also have checked the assumption that v_{ijt} is homoskedastic with respect to age and business-cycle stage, and have found no clear-cut evidence to the contrary.

[4]The parameter ρ can be affected in complex ways by parents' behavior regarding investment in their children. For example, if, as Griliches (1979) claims, parents invest more in their less able children, this "compensatory" behavior reduces σ_u^2; if, as Becker and Tomes (1976) claim, parents invest more in their abler children, this "reinforcing" behavior increases σ_u^2. These different investment strategies might affect σ_a^2 as well, but the directions of the effects on that variance component are less obvious.

[5]See Behrman, Taubman, and Wales (1977), Bound, Griliches, and Hall (1986), Brittain (1977), Corcoran and Jencks (1979), Griliches (1979), Kearl and Pope (1986), and Olneck (1977). The main outlier is the high correlation (.54) that Behrman, Taubman, and Wales found for monozygotic twins. It should not be surprising, though, that monozygotic twins show stronger resemblance than do other siblings.

[6]A related point, with similar implications, is that these studies also have failed to account for cohort effects.

[7]As will be discussed later, the PSID oversampled the low-income population by also including a sample component drawn from the Survey of Economic Opportunity.

[8]The dummy variables included one each for ages 26–32 (25 omitted) and one each for years 1976–82 (1975 omitted). Accounting for life-cycle stage with individual-invariant age coefficients assumes that different individuals do not have systematically different earnings-age profiles. Somewhat surprisingly, the longitudinal analyses of earnings by MaCurdy (1982) and Abowd and Card (1986) support this assumption. They find that the serial correlation in change of log earnings is essentially zero at lags longer than two years, which would not be the case in the presence of substantial individual-specific time trends. We find similar evidence in our data.

[9] Our estimate of σ_v^2 is larger than most in the existing literature. As noted by MacDonald and Robinson (1985), most previous studies have depressed their estimates of σ_v^2 by deleting outliers and all observations for individuals who had zero earnings in any year in the sample period. Our approach is to include all positive observations and recognize that our estimate of σ_v^2 is partly comprised of variance due to measurement error. Another possible factor behind our large $\hat{\sigma}_v^2$ is the young age range in our sample. Our sample members may display especially volatile earnings because many of them are not yet settled into their regular career paths. If the PSID continues well into the future, it would be worthwhile to conduct a later replication of our study once our sample has matured.

[10] Other estimates of the ratio in (6) can be inferred from results of previous longitudinal analyses of earnings. The results in the first row of Table 1 in Lillard and Willis (1978), for example, imply $\hat{\gamma}/\hat{\phi} = .73$. These results, however, do not adjust for stage of life-cycle and are based on a sample restricted to individuals with positive earnings in *all* years of the sample period. The results in the seventh row of Table 1 in MacDonald and Robinson (1985) imply $\hat{\gamma}/\hat{\phi} = .57$. This analysis pertains to hourly wages and adjusts for schooling as well as life-cycle and year effects, but is otherwise similar to ours in that it does not exclude outliers or individuals with zero earnings in any year.

[11] This technique has been suggested in related contexts by Hausman and Wise (1981) and Manski and Lerman (1977).

[12] The estimation formulas specialize those in Appendix A to the case where $T_{ij} = 1$ and the within-family variance components σ_u^2 and σ_v^2 are not separated.

[13] We are not the first to encounter this general problem. Section 8.1 of Huber (1981) comments on the special sensitivity of estimated second moments to outliers.

[14] It was tempting, of course, simply to delete the outliers. An examination of the outlier cases, however, did not give us any sound basis for concluding that they were any less valid than other observations.

[15] As will be discussed later in this section, obtaining appropriate standard error estimates is a formidable task. We have applied the half-sample replication procedure described later to the $\hat{\gamma}$ for 1982 and estimated its standard error to be .081.

[16] The relative preponderance of low outliers indicates that ϵ_{ijt} is distributed asymmetrically and hence nonnormally. If the nonnormality of ϵ_{ijt} arises solely from

nonnormality of the transitory component v_{ijt}, however, Table II's normality assumption for a_i and u_{ij} is appropriate.

[17]See footnote 16 on the nonnormality issue. These same factors that complicate the estimation of standard errors also explain why we did not adopt a maximum likelihood approach for the estimation in this paper.

[18]See DeGroot (1975), p. 185.

[19]The needs standard is described on pages 115–16 of Survey Research Center (1985). The usual procedure of adjusting the standard for inflation is unnecessary in our study because we include year variables in X_{ijt} in equation (1).

[20]We use the weight described on page 459 of Survey Research Center (1985). Pages 6–19 of Survey Research Center (1979) give a detailed explanation of the construction of the weight.

References

Abowd, John M., and David Card. "On the Covariance Structure of Earnings and Hours Changes." Working Paper No. 1832, National Bureau of Economic Research, February 1986.

Becker, Gary S., and Nigel Tomes. "Child Endowments and the Quantity and Quality of Children." *Journal of Political Economy* 84 (August 1976): S143-62.

Behrman, Jere, Paul Taubman, and Terence Wales. "Controlling for and Measuring the Effects of Genetics and Family Environment in Equations for Schooling and Labor Market Success." *Kinometrics: Determinants of Socioeconomic Success within and between Families*. ed. by Paul Taubman. Amsterdam: North-Holland. 1977.

Bound, John, Zvi Griliches, and Bronwyn H. Hall. "Wages, Schooling and IQ of Brothers and Sisters: Do the Family Factors Differ?" *International Economic Review* 27 (February 1986): 77-105.

Brittain, John A. *The Inheritance of Economic Status*. Washington: Brookings Institution, 1977.

Cochran, William G. *Sampling Techniques*. New York: Wiley. 1963.

Corcoran, Mary, and Christopher Jencks. "The Effects of Family Background." Chapter 3 in Christopher Jencks et al., *Who Gets Ahead?* New York: Basic Books, 1979.

Corcoran, Mary, Christopher Jencks, and Michael Olneck. "The Effects of Family Background on Earnings." *American Economic Review* 66 (May 1976): 430-35.

DeGroot, Morris H. *Probability and Statistics*. Reading, MA: Addison-Wesley, 1975.

Griliches, Zvi. "Sibling Models and Data in Economics: Beginnings of a Survey." *Journal of Political Economy* 87 (October 1979): S37-64.

Harrington, Michael. *The Other America: Poverty in the United States*. Baltimore: Penguin, 1962.

Hausman, Jerry A., and David A. Wise. "Stratification on Endogenous Variables and Estimation: The Gary Income Maintenance Experiment." *Structural Analysis of Discrete Data with Econometric Applications*, ed. by Charles F. Manski and Daniel McFadden. Cambridge: MIT Press, 1981.

Huber, Peter J. *Robust Statistics*. New York: Wiley, 1981.

Kearl, J. R., and Clayne L. Pope. "Unobservable Family and Individual Contributions to the Distributions of Income and Wealth." *Journal of Labor Economics* 4 (July 1986): S48-79.

Kish, Leslie, and Martin R. Frankel. "Balanced Repeated Replications for Standard Errors." *Journal of the American Statistical Association* 65 (September 1970): 1071-94.

Lillard, Lee A., and Robert J. Willis. "Dynamic Aspects of Earning Mobility." *Econometrica* 46 (September 1978): 985-1012.

MacDonald, Glenn M., and Chris Robinson. "Cautionary Tales about Arbitrary Deletion of Observations; or, Throwing the Variance out with the Bathwater." *Journal of Labor Economics* 3 (April 1985): 124–52.

MaCurdy, Thomas E. "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis." *Journal of Econometrics* 18 (January 1982): 83–114.

Manski, Charles F., and Steven R. Lerman. "The Estimation of Choice Probabilities from Choice Based Samples." *Econometrica* 45 (November 1977): 1977–88.

McCarthy, Philip J. "Pseudo-Replication: Half Samples." *Review of the International Statistical Institute* 37 (1969): 239–64.

Olneck, Michael R. "On the Use of Sibling Data to Estimate the Effects of Family Background, Cognitive Skills, and Schooling: Results from the Kalamazoo Brothers Study." *Kinometrics: Determinants of Socioeconomic Success within and between Families*, ed. by Paul Taubman. Amsterdam: North-Holland, 1977.

Solon, Gary. "Estimating Autocorrelations in Fixed-Effects Models." Technical Working Paper No. 32, National Bureau of Economic Research, February 1984.

Survey Research Center. *A Panel Study of Income Dynamics: Procedures and Tape Codes, 1978 Interviewing Year — Wave XI, A Supplement.* Ann Arbor: Institute for Social Research, The University of Michigan, 1979.

Survey Research Center. *A Panel Study of Income Dynamics: Procedures and Tape Codes. 1983 Interviewing Year — Wave XVI, A Supplement.* Ann Arbor: Institute for Social Research, The University of Michigan, 1985.

Wolter, Kirk M. *Introduction to Variance Estimation.* New York: Springer-Verlag, 1985.