University of Wisconsin-Madison

IRP Discussion Papers

Gary Solon

INTERGENERATIONAL INCOME MOBILITY IN THE UNITED STATES

DP #894-89

ABSTRACT

Social scientists and policy analysts have long expressed concern about the extent of intergenerational income mobility in the United States. Nevertheless, remarkably little empirical evidence is available on this issue. The few existing estimates of the intergenerational correlation in income have been biased downward by a combination of measurement error and unrepresentative samples. This paper presents new estimates based on intergenerational data on a sample of 348 father-son pairs drawn from the Panel Study of Income Dynamics. The results imply that the correlation of fathers' income with that of their sons is at least .4, indicating dramatically less mobility than the figure of about .2 suggested by earlier research.

INTERGENERATIONAL INCOME MOBILITY IN THE UNITED STATES

The degree to which income status is transmitted from one generation to the next has persistently interested social scientists and others concerned with social policy. This interest has stemmed largely from a belief that intergenerationally transmitted income inequality 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. [1962, p. 21]

The recent literature on the "underclass" also has emphasized the extent to which income status, especially poverty, is passed from generation to generation. Auletta (1982, p. 268), for instance, has written, "Today, perhaps for the first time, America has a sizable, and so far intractable, intergenerational underclass." In a similar vein, Kilson (1981, p. 58) has argued that "those blacks who have come out of the 1960s and 1970s poverty ridden are more likely to pass on this awful plight to their offspring – offspring who, owing to inadequate schools, poor school performance, excessively high unemployment, low skills, and attendant social pathologies, have little opportunity to put the poverty of their parents behind them."

Given the widespread concern about intergenerational mobility, it is astonishing how few attempts have been made to measure the simple intergenerational correlation of income in the United States. A recent paper by Behrman and Taubman (1985), which estimated the father-son correlation in the natural logarithm of annual earnings, noted the existence of only one other published study, that of Sewell and Hauser (1975).[1] Although several studies have been conducted in other countries, these, of course, are of no help for ascertaining the degree of intergenerational mobility in the U.S.[2]

In stark contrast to the above quotations that stress the importance of intergenerational transmission, the statistical literature on the subject has found strikingly small intergenerational income correlations. Behrman and Taubman (1985, p. 147) estimated the father-son correlation in log earnings to be .2 or less and concluded, "The members of this sample come from a highly mobile society." Likewise, Sewell and Hauser (1975, p. 72) estimated only a .18 correlation between sons' earnings and parents' income. Based on a survey of European as well as U.S. studies, Becker and Tomes (1986, p. 51) concluded, "Regression to the mean in earnings in rich countries appears to be rapid." Becker's presidential address to the American Economic Association (1988, p. 10) similarly concluded, "In all these countries, low earnings as well as high earnings are not strongly transmitted from fathers to sons...."

The obvious question is: Are the policy-oriented writings that have emphasized intergenerational transmission unfounded, or is there something wrong with the statistical evidence? Section I of this paper demonstrates that previous estimates of

---

[1]Numerous studies, such as Duncan, Featherman, and Duncan (1972), have estimated intergenerational correlations in measures of occupational prestige. Such estimates typically are larger than the existing ones for income. It has been unclear whether the estimates for occupational status measures are higher because such measures are better indicators of long-run income than are the available income variables or because fathers and sons tend to be in similar occupational categories even when their long-run incomes are very different. Another study, by Treiman and Hauser (1977), imputed intergenerational income correlations in the absence of parental income data by imposing strong assumptions about the relationships among various economic status measures for parents and children. Treiman and Hauser repeatedly stressed the obvious desirability of obtaining parental income data to enable direct estimation of intergenerational income mobility. Still other studies, such as Solon et al. (1987) and Corcoran et al. (1989), have investigated the effects of family background on economic status by measuring sibling correlations in economic status or by estimating regression relationships between income variables and sets of background characteristics. Such studies, however, do not directly address the simple intergenerational correlation in income.

[2]See Becker and Tomes (1986) for an international survey.

intergenerational income correlations have been biased downward by a combination of measurement error and unrepresentative samples. Sections II, III, and IV describe a new analysis based on intergenerational data from the Panel Study of Income Dynamics. The results contain strong evidence that, in the United States, the father-son correlations in long-run earnings, hourly wages, and family income are about .4 or even higher. These results depict a much less mobile society than previous studies have portrayed. Section V summarizes and discusses the findings.

I. Biases in Previous Studies

Previous estimates of intergenerational income mobility have been based on error-ridden data on unrepresentative samples. To explore the likely effects of these problems, consider the following model. Let y_{1i} represent long-run economic status (e.g., the "permanent" component of log annual earnings) for a son in family i, and let y_{0i} be the same variable for his father. Let both variables be measured as deviations from generation means. Let ρ denote the true population correlation between y_{0i} and y_{1i}, and assume for now that the population variance in y is the same σ_y^2 in either generation. Then, if y_{0i} and y_{1i} were directly observed for a random sample of families, one could estimate ρ by applying least squares to the regression equation

$$(1) \qquad y_{1i} = \rho y_{0i} + \epsilon_i \quad .$$

The intergenerational correlation ρ could be consistently estimated by either $\hat{\rho}$, the estimated slope coefficient, or R, the square root of the R^2 statistic.[3]

This is essentially the estimation approach used in previous studies of intergenerational income mobility (both in the U.S. and in other countries), with two crucial exceptions. First, lacking direct measures of long-run status, these studies instead

[3]If the variance in y differs between generations ($\sigma_{y0}^2 \neq \sigma_{y1}^2$), then the estimated slope coefficient estimates $\rho\sigma_{y1}/\sigma_{y0}$ rather then the correlation ρ itself. The empirical relevance of intergenerational change in the variance of long-run economic status is discussed in Section V.

have used short-run proxies, such as single-year earnings or income measures. Second, they typically have used data from peculiarly homogeneous samples, rather than random samples. As discussed in detail in Solon (1989), both factors generate downward biases in the estimated intergenerational correlations.

The first bias can be simply characterized by assuming that the short-run proxy for son's long-run status is his measured status in period t,

(2) $$y_{1it} = y_{1i} + v_{1it},$$

where v_{1it} is a transitory fluctuation around long-run status due to both actual transitory movement and random measurement error.[4] Similarly, the proxy for father's status is his measured status in period s,

(3) $$y_{0is} = y_{0i} + v_{0is}.$$

Let σ^2_{v0} and σ^2_{v1} denote the population variances of v for each generation, and assume v_{0is} and v_{1it} are uncorrelated with each other and with y_{0i} and y_{1i}. Then, when previous studies have applied least squares to equation (1) with y_{0is} and y_{1it} in place of y_{0i} and y_{1i}, the resulting estimates have been subject to errors-in-variables biases. In particular, the probability limit of the estimated slope coefficient $\hat{\rho}$ is

(4) $$\text{plim } \hat{\rho} = \rho\sigma^2_y/(\sigma^2_y + \sigma^2_{v0}) < \rho,$$

and the probability limit of R is

(5) $$\text{plim } R = \rho\sigma^2_y/\sqrt{(\sigma^2_y+\sigma^2_{v0})(\sigma^2_y+\sigma^2_{v1})} < \rho.$$

Whether this tendency to underestimate ρ is practically important depends on whether the variances of the transitory fluctuations are substantial relative to the variance

---

[4]For simplicity, this formulation abstracts from life-cycle profiles in income variables. Such profiles are incorporated into the analysis in Section III. For evidence that the measurement error aspect of v_{1it} is empirically important, see Duncan and Hill (1985) and Bound and Krueger (1989).

in long-run status. Information on this point for the U.S. is available from several longitudinal studies of earnings and wage rates, which have decomposed the population variance in annual measures of these variables into permanent and transitory components.[5] The results of these studies suggest that, in an intergenerational analysis based on only single-year data, such as Behrman and Taubman (1985), errors-in-variables bias alone could be expected to depress estimates of ρ by more than 30 percent. In a study such as Sewell and Hauser (1975), which used status measures averaged over a few years, this bias would be reduced, but not eliminated.

The second source of bias is unrepresentatively homogeneous samples. Behrman and Taubman's fathers sample, for instance, was drawn from a sample of white male twins born between 1917 and 1927. To remain in the sample, both twins had to have served in the armed forces, and both had to have survived until and cooperated with a succession of surveys. One would expect this sample to be more homogeneous than a random cohort sample, and Solon (1989) summarizes some corroborating evidence. Most other intergenerational studies, both in the U.S. and abroad, also have relied on homogeneous samples. Sewell and Hauser's study, for example, was based on a sons sample of Wisconsin high school seniors that graduated in 1957 and were no longer in school in 1964.

To focus on the bias from homogeneity, assume for now that permanent status can be directly observed, but that the fathers sample, as in the Behrman-Taubman study, is selected from a relatively homogeneous subpopulation whose variance in permanent status is $s_{y0}^2 < \sigma_y^2$. In that case, if one applies least squares to equation (1), the probability limit of R is[6]

$$\text{(6)} \qquad \text{plim } R = \rho/\sqrt{1+(1-\rho^2)[(\sigma_y^2/s_{y0}^2)-1]} < \rho.$$

[5]See Lillard and Willis (1978), MacDonald and Robinson (1985), and Solon et al. (1987).

[6]See Solon (1989) for the derivation.

The reason for the downward inconsistency is that the small sample dispersion in the regressor y_{0i} depresses the R^2 statistic. A similar result of downward inconsistency applies to the case where R is based on a homogeneous sons sample, as in the Sewell-Hauser study.

Interestingly, in the case where the sample selection is on fathers, the estimated regression coefficient $\hat{\rho}$, unlike R, would consistently estimate ρ if long-run status were directly observed. But, with short-run proxies in place of long-run status, sample homogeneity aggravates the errors-in-variables bias in $\hat{\rho}$. This occurs because, with a homogeneous fathers sample, the small sample dispersion in father's long-run status reduces the "signal-to-noise ratio" in father's measured status. In mathematical terms, the factor in equation (4) declines from $\sigma_y^2/(\sigma_y^2+\sigma_{v0}^2)$ to $s_{y0}^2/(s_{y0}^2+\sigma_{v0}^2)$.[7]

Because the crucial quantities are variances in permanent status, which is not directly observed, it is difficult to ascertain the severity of sample homogeneity in previous studies or its impact on their estimates of intergenerational mobility. Nevertheless, the sample selection criteria appear strikingly prone to produce homogeneous samples, and it is quite conceivable that such samples combined with substantial error in measuring long-run status could produce extreme biases in the estimation of intergenerational income correlations. It therefore seems worthwhile to conduct a new analysis designed to be less susceptible to the biases of earlier studies.

II. Data Description

The new analysis uses intergenerational data from the Panel Study of Income Dynamics (PSID), a nationally representative longitudinal survey of about 5,000 families that The University of Michigan's Survey Research Center has conducted annually since

---

[7]This assumes the sample is homogeneous with respect to permanent status, but not with respect to transitory fluctuations in status. This seems a reasonable characterization of the Behrman-Taubman sample.

1968.[8] Because the survey has followed children from the original PSID families as they have grown into adulthood and formed their own households, it is now possible to relate the children's income status as adults to the status of their parents, as annually reported by the parents themselves since the outset of the survey.[9] The PSID data are especially well-suited for reducing the biases of earlier research. First, because the data come from a national probability sample, they avoid the homogeneity of the samples used in previous studies. Second, the longitudinal nature of the data makes it possible to explore the empirical importance of using short-run versus long-run status measures.

This study focuses mainly on father-son correlations in earnings, hourly wage rates, and family income. The main sample is comprised of 348 father-son pairs from the Survey Research Center component of the PSID.[10] The sons in the sample are children from the original 1968 PSID households who, in the 1985 survey, reported positive annual earnings for 1984.[11] The sons sample is restricted to those born between 1951 and 1959. Sons born before 1951, who were older than 17 at the 1968 interview, are excluded to avoid overrepresenting sons that left home at late ages. The 1959 restriction assures that the sons' 1984 status measures are observed at ages of at least 25. Where more than one

---

[8]See Survey Research Center (1988) for documentation. The PSID data used in this study come from the 1985 cross-year family-individual response-nonresponse file.

[9]Self-reported parental income is more accurate than the retrospective child-reported measures of parental status sometimes used in intergenerational research. For a detailed discussion, see Massagli and Hauser (1983). A drawback of using the intergenerational span of the PSID is that the available sample is subject to considerable attrition. For a lucid discussion of attrition in the PSID, see Becketti et al. (1988), who conclude that "attrition has not substantially reduced the representativeness of the PSID." Other recent efforts to exploit the intergenerational span of the PSID include Altonji (1988), Behrman and Taubman (1987), Corcoran et al. (1989), and Hill and Duncan (1987).

[10]The Survey of Economic Opportunity (SEO) component, designed to overrepresent the low-income population, is excluded from the analysis. The families in the SEO component were selected on the basis of their low 1966 incomes. Because the transitory term v_{0is} in parental income is serially correlated, including the SEO component would generate a nonrandom sample of v_{0is} in the 1967–71 parental income data used in this study.

[11]Those with earnings imputed by "major assignments" are excluded from the sample.

son from the same family meets all the above restrictions, only the oldest is retained in the main sample.

The "fathers" in the sample are the male heads of the households the sons inhabited in 1968. In some cases, these "fathers" are not the sons' natural fathers. Such cases are retained in the sample because the object of this study is not to measure genetic transmission, but to measure the correlation between son's economic status as an adult and the status of the household in which he grew up.

Table 1 presents some summary statistics on the age and annual earnings of the main sample's fathers and sons. Despite the study's preference for older sons, the sample mean age for sons in 1984 is still slightly less than 30, while the sample mean for fathers in 1967 is 42. Because the sons are observed at an early stage of the life cycle, their mean earnings are lower, and the standard deviation of the natural logarithm of their earnings is higher.

III. Econometric Models

The models estimated in this study extend the model in Section I to incorporate age profiles in earnings, wages, and income. For any of these status variables measured in year t, the model for son's status in equation (2) is extended to

$$y_{1it} = y_{1i} + \alpha_1 + \beta_1 A_{1it} + \gamma_1 A_{1it}^2 + v_{1it} \tag{7}$$

where A_{1it} is the age of the son from family i in year t. Similarly, the model in equation (3) for father's status in year s is extended to

$$y_{0is} = y_{0i} + \alpha_0 + \beta_0 A_{0is} + \gamma_0 A_{0is}^2 + v_{0is} \tag{8}$$

where A_{0is} is the father's age in year s.[12] Solving equations (7) and (8) for y_{0i} and y_{1i} and substituting the results into equation (1) yield

[12]These models, which account for life-cycle stage with individual-invariant age coefficients, assume that different individuals do not have systematically different age

Table 1
Sample Characteristics

	Mean	Standard deviation	Minimum	Maximum
Son's age in 1984	29.6	2.4	25.0	33.0
Son's earnings in 1984	22,479	15,019	19	147,656
Son's log earnings in 1984	9.75	.94	2.94	11.90
Father's age in 1967	42.0	7.7	27.0	68.0
Father's earnings in 1967*	29,304	20,015	405	202,500
Father's log earnings in 1967*	10.10	.69	6.00	12.22

*The sample statistics for father's 1967 earnings are in 1984 dollars and pertain to the sample of 322 fathers analyzed in the first row and column of Table 2.

$$(9) \quad y_{1it} = (\alpha_1 - \rho\alpha_0) + \rho y_{0is} + \beta_1 A_{1it} + \gamma_1 A_{1it}^2 - \rho\beta_0 A_{0is} - \rho\gamma_1 A_{0is}^2 + \epsilon_i + v_{1it} - \rho v_{0is}.$$

Equation (9) expresses son's observed status in year t as a regression function of father's observed status in year s and age controls for both father and son. If equation (9) is estimated by least squares, the resulting $\hat{\rho}$ is subject to an errors-in-variables bias because of the correlation between v_{0is} and y_{0is}. In fact, if, in addition to the assumptions in Section I, the age variables are assumed to be uncorrelated with long-run status and the v's, $\hat{\rho}$ continues to be downward inconsistent by a factor of $\sigma_y^2/(\sigma_y^2 + \sigma_{v0}^2)$. This inconsistency should be less severe with the PSID data than with more homogeneous samples, but still could be quite substantial.

The analysis in Section IV pursues two strategies for treating the errors-in-variables bias. One approach is to average father's status in equation (8) over T years, so that equation (9) is modified to

$$(10) \quad y_{1it} = (\alpha_1 - \rho\alpha_0) + \rho\bar{y}_{0i} + \beta_1 A_{1it} + \gamma_1 A_{1it}^2 - \rho\beta_0 \bar{A}_{0i} - \rho\gamma_0 \overline{A_{0i}^2} + \epsilon_i + v_{1it} - \rho\bar{v}_{0i}$$

where, for any variable z_{0is}, $\bar{z}_{0i} = \sum_{j=s}^{s+T} z_{0ij}/T$. If equation (10) is estimated by least squares, the resulting $\hat{\rho}$ is still downward inconsistent, but the magnitude of the inconsistency is reduced because the averaging across years decreases the variance of the "noise" relative to the "signal."

The second approach is to apply instrumental-variable estimation to equation (9), with father's years of education as the instrument for father's single-year status. This is a somewhat odd context for instrumental-variable estimation because father's education probably would not be excluded from a structural model for son's economic status. As will

profiles for earnings, wages, or income. This assumption is supported by several recent longitudinal analyses of earnings and wage rates. Results in Abowd and Card (1989), Solon and Barsky (1989), and Topel (1987), as well as results for the present study's sample, indicate that the serial correlation in earnings or wage growth is essentially zero at lags longer than two years, which would not be the case in the presence of substantial individual-specific profiles.

be discussed later, however, under plausible assumptions, the inconsistency of this instrumental-variable estimator of ρ is in an upward direction. The probability limits of the two proposed estimators therefore bracket the true value of ρ.

IV. Empirical Results

The first part of this section presents estimates of ρ based on ordinary least squares (OLS) estimation of equations (9) and (10). The second part presents instrumental-variable (IV) estimates. The third discusses the implications of the estimates for the degree of intergenerational income mobility in the U.S.

A. OLS Results

Tables 2 and 3 display estimates of ρ from OLS estimation of equations (9) and (10), where y_{1it} is the natural logarithm of the son's annual earnings in 1984 and y_{0is} is the natural logarithm of the father's annual earnings in year s. Results are reported for each of s = 1967, 1968, ..., 1971. All earnings variables, as well as the wage and family income variables considered later, are in 1984 dollars as measured by the Consumer Price Index. The results in Table 2 are based on different sample sizes (shown in brackets) because the number of missing observations varies with s. In particular, father's earnings might be missing in year s because of the father's attrition from the sample, because his earnings were not reported, or, in a few instances, because he had zero earnings.[13]

The estimates of ρ in the first column of Table 2 come from OLS estimation of equation (9), that is, from regressions involving single-year measures of father's log earnings. These estimates, which are expected to suffer from substantial errors-in-variables bias, range from .25 when father's 1971 log earnings are the regressor to .39

[13]In addition, for comparability with the later IV estimates, two observations are excluded because father's years of education were not reported. Retaining those observations has virtually no effect on the OLS results.

Table 2
OLS Estimates of ρ from Log Earnings Data

First year in average of father's log earnings	Measure of father's log earnings				
	Single-year measure	Two-year average	Three-year average	Four-year average	Five-year average
1967	.386 (.079) [322]				
		.425 (.090) [313]			
1968	.271 (.074) [326]		.408 (.087) [309]		
		.365 (.081) [317]		.413 (.088) [301]	
1969	.326 (.073) [320]		.369 (.083) [309]		.413 (.093) [290]
		.342 (.078) [312]		.357 (.088) [298]	
1970	.285 (.073) [318]		.336 (.084) [301]		
		.290 (.082) [303]			
1971	.247 (.073) [307]				

Standard error estimates are in parentheses, and sample sizes are in brackets.

Table 3
OLS Estimates of ρ from Log Earnings Data for "Balanced" Sample (N = 290)

First year in average of father's log earnings	Measure of father's log earnings				
	Single-year measure	Two-year average	Three-year average	Four-year average	Five-year average
1967	.369 (.094)				
		.409 (.093)			
1968	.396 (.087)		.431 (.093)		
		.422 (.088)		.420 (.094)	
1969	.406 (.085)		.405 (.090)		.413 (.093)
		.382 (.089)		.397 (.090)	
1970	.309 (.087)		.374 (.088)		
		.324 (.086)			
1971	.285 (078)				

Standard error estimates are in parentheses.

when his 1967 earnings are used.[14] These estimates differ because of both the change in regressors and the change in sample composition. To hold the latter constant, equation (9) is reestimated for the 290 cases in which father's earnings are available for all the years 1967-71. The resulting estimates of ρ, shown in the first column of Table 3, range from .28 for s=1971 to .41 for s=1969. Once sample composition is held constant, the estimates for 1967-69 are fairly similar, but those for 1970-71 are noticeably smaller. Part of the explanation, especially for 1971, seems to be that the increased variance in father's log annual earnings in recession years worsens the errors-in-variables bias.

Despite the wide variation in results, *all* the estimates are distinctly above .2, the value described by both Behrman and Taubman (1985) and Becker and Tomes (1986) as an upper bound on the intergenerational correlation in log earnings. Even though the present estimates are biased downward by the use of single-year measures of father's earnings, they apparently are less biased than previous estimates based on samples more homogeneous than the PSID. A simple exercise to illustrate the importance of the homogeneity issue is to imitate Sewell and Hauser's exclusion of sons that are not high school graduates. When the analyses reported in the first column of Table 2 are repeated with the sons samples restricted to those with at least 12 years of education, the estimated ρ for s=1967 falls from .39 to .26 (with estimated standard error .08) with sample size 285. Similarly, $\hat{\rho}$ declines from .27 to .20 (.07) for s=1968, from .33 to .22 (.08) for s=1969, from .29 to .17 (.08) for s=1970, and from .25 to .18 (.08) for s=1971.

[14]Some additional results, not reported in the table, involve exclusion of outlier observations. With s=1967, reestimation excluding sons and fathers whose annual earnings were less than $1,000 reduces the sample size to 311 and gives a $\hat{\rho}$ of .358 (with estimated standard error .064). Excluding fathers whose age in 1967 was less than 30 or greater than 59 leads to $\hat{\rho}$ = .412 (.085) with sample size 308. Imposing both restrictions simultaneously leads to $\hat{\rho}$ = .374 (.066) with sample size 298. Another experiment involves including multiple sons from the same family. In the main analyses, when more than one son meets the other sample selection criteria, only the oldest is used because his 1984 earnings are likely to be a more accurate indicator of his long-run status and because the correlation of error terms across sons from the same family would complicate the correct estimation of standard errors. It is nevertheless worth reporting that, if multiple sons are included, the OLS results for s=1967 are $\hat{\rho}$=.348 (.066) with a sample size of 428 sons.

Next, to reduce the errors-in-variables bias, OLS is applied to equation (10), that is, to regressions in which father's log earnings are averaged over multiple years. The results are displayed in the remaining columns of Tables 2 and 3. For example, the entries in the second column of Table 2 indicate that $\hat{\rho}$ equals .425 when the regressor is father's log earnings averaged over 1967 and 1968, .365 when the average is over 1968 and 1969, and so forth. Table 3 gives the corresponding results for the "balanced" sample of 290 cases in which father's earnings are available for all years. As expected, the general pattern in both tables is that $\hat{\rho}$ tends to get larger as father's log earnings are averaged over more years. Most of the estimates based on at least three years are in the neighborhood of .4, double the upper limit claimed in previous studies. Furthermore, even these estimates presumably are subject to at least minor downward biases.

To supplement the results on intergenerational earnings correlations, Table 4 presents results in which the economic status measures for fathers and sons are the logarithms of their hourly wage rates, their family incomes, and their family incomes relative to the official federal poverty threshold. The hourly wage is measured as the ratio of annual earnings to annual hours of work. Division of family income by the relevant poverty standard is a crude effort to adjust family income for family size and composition.

The first column of Table 4 reports OLS estimates of ρ based on single-year measures of father's status with $s = 1967$. The estimated ρ of .39 for log earnings is copied over from Table 2, while ρ is estimated at .29 for the log wage and .48 for both family income variables. That the smallest estimate appears for the hourly wage is unsurprising given Duncan and Hill's (1985) finding that measurement error in both earnings and hours of work causes the ratio of the two to be especially noisy. Even though all these OLS estimates are biased downward by their reliance on single-year measures, they are strikingly large relative to previous studies' estimates of intergenerational correlations.

Although this study focuses mainly on father-son correlations, it is reasonable to ask how the results would be affected by inclusion of sons from mother-headed families.

Table 4
OLS and IV Estimates of ρ for Various Single-Year Income Measures in 1967

Income measure	OLS	IV	Sample size
Log earnings	.386 (.079)	.526 (.135)	322
Log wage	.294 (.052)	.449 (.095)	316
Log family income	.483 (.069)	.530 (.123)	313
Log (family income/ poverty line)	.476 (.060)	.563 (.103)	313

Standard error estimates are in parentheses.

Doing so expands the sample size for the family income analyses from 313 to 340 and decreases $\hat{\rho}$ from .48 to .44 (.06) for both family income variables. Again, despite the errors-in-variables bias, these estimates are dramatically larger than those from previous studies.

B. IV Results

An alternative strategy for treating the errors-in-variables problem is to apply IV estimation to equation (9) with father's years of education as the instrument for y_{0is}. Because the PSID's 1968 information on education is in interval form, the instrument actually used is set at the midpoint of the reported interval except that fathers in the highest education category are assigned 18 years of education. Although this procedure inescapably produces measurement error in father's years of education, as long as the measurement error is uncorrelated with the error term in equation (9), the IV estimator remains consistent.

A more subtle issue is whether father's education can be a valid instrument when it belongs as a regressor in a structural model for son's income status. As detailed in the Appendix, this problem may cause inconsistency in the IV estimator, but, under plausible assumptions, the inconsistency is in an upward direction. If so, the probability limits of the OLS and IV estimators bracket the true ρ. If not, even the IV estimator may tend to underestimate ρ.

The second column of Table 4 presents IV estimates of ρ for $s=1967$. As expected, the IV estimates are larger than the OLS estimates, ranging from .45 for the log of the wage to .56 for the log of family income relative to the poverty line. While these estimates are likely to be upward-biased, in combination with the downward-biased OLS estimates, they strongly suggest that the intergenerational income correlation in the U.S. is around .4 or possibly higher.

C. Implications for Intergenerational Mobility

Contrary to previous studies' conclusion that the intergenerational income correlation in the U.S. is less than .2, this study's results suggest the correlation is at least .4, and the family income results cluster around .5. What do these different correlation estimates imply about the extent of intergenerational income mobility in the U.S.? To highlight the differences in implications, Table 5 displays the probability that a son's long-run status lies in each decile of the population distribution as a function of the percentile of his father's status. The first panel of the table shows the probabilities implied by $\rho=.2$, the second shows the probabilities implied by $\rho=.4$, and the third shows the probabilities implied by $\rho=.5$. The figures are based on the assumption that long-run status (e.g., the permanent component of log earnings) is normally distributed in each generation.

Inspection of the table emphasizes that a ρ of .4 or .5 implies a very different degree of intergenerational mobility than a ρ of .2. For example, if $\rho=.2$, a son whose father's status is at the fifth percentile has a .30 chance of remaining in the bottom quintile, a .37 chance of rising above the median, and a .12 chance of reaching the top quintile. But, if $\rho=.4$, he has a .42 chance of remaining in the bottom quintile, only a .24 chance of rising above the median, and only a .05 chance of reaching the top quintile. And, if $\rho=.5$, he has a .49 chance of remaining in the bottom quintile, only a .17 chance of rising above the median, and a mere .03 chance of reaching the top quintile. Clearly, the higher intergenerational correlations estimated in this study imply a dramatically less mobile society.

V. Summary and Discussion

Measurement error and homogeneous samples have caused previous studies to exaggerate the extent of intergenerational income mobility in the United States. This paper has presented a new analysis based on intergenerational data from the Panel Study of Income Dynamics and designed to be less susceptible to the biases in earlier studies. The results, which indicate that the intergenerational income correlation in the U.S. is at

Table 5
Probability That Son's Long-Run Status Is in Specified Decile Given Percentile of Father's Status

Father's percentile	Decile of son's status									
	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100
					$\rho = .2$					
5	.17	.13	.12	.11	.10	.09	.08	.08	.07	.05
15	.14	.12	.11	.11	.10	.10	.09	.09	.08	.06
25	.12	.11	.11	.11	.10	.10	.10	.09	.09	.07
35	.11	.11	.11	.10	.10	.10	.10	.10	.09	.08
45	.10	.10	.10	.10	.10	.10	.10	.10	.10	.09
55	.09	.10	.10	.10	.10	.10	.10	.10	.10	.10
65	.08	.09	.10	.10	.10	.10	.10	.11	.11	.11
75	.07	.09	.09	.10	.10	.10	.11	.11	.11	.12
85	.06	.08	.09	.09	.10	.10	.11	.11	.12	.14
95	.05	.07	.08	.08	.09	.10	.11	.12	.13	.17
					$\rho = .4$					
5	.25	.17	.14	.11	.09	.08	.06	.05	.03	.02
15	.17	.15	.13	.12	.10	.09	.08	.07	.05	.03
25	.13	.13	.12	.12	.11	.10	.09	.08	.07	.05
35	.11	.12	.12	.11	.11	.10	.10	.09	.08	.06
45	.09	.10	.11	.11	.11	.11	.10	.10	.09	.07
55	.07	.09	.10	.10	.11	.11	.11	.11	.10	.09
65	.06	.08	.09	.10	.10	.11	.11	.12	.12	.11
75	.05	.07	.08	.09	.10	.11	.12	.12	.13	.13
85	.03	.05	.07	.08	.09	.10	.12	.13	.15	.17
95	.02	.03	.05	.06	.08	.09	.11	.14	.17	.25

Table 5 (continued)

Father's percentile	Decile of son's status									
	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100
					ρ = .5					
5	.30	.19	.14	.11	.08	.06	.05	.03	.02	.01
15	.19	.17	.14	.12	.11	.09	.07	.06	.04	.02
25	.14	.14	.13	.12	.11	.10	.09	.07	.06	.03
35	.10	.12	.12	.12	.12	.11	.10	.09	.07	.04
45	.08	.10	.11	.12	.12	.11	.11	.10	.09	.06
55	.06	.09	.10	.11	.11	.12	.12	.11	.10	.08
65	.04	.07	.09	.10	.11	.12	.12	.12	.12	.10
75	.03	.06	.07	.09	.10	.11	.12	.13	.14	.14
85	.02	.04	.06	.07	.09	.11	.12	.14	.17	.19
95	.01	.02	.03	.05	.06	.08	.11	.14	.19	.30

least .4 and possibly higher, portray a much less mobile society than has been described in earlier research.

One obvious limitation of the present study is its reliance on a single data set. Further research with other data, perhaps from the National Longitudinal Surveys of labor market experience, would be very worthwhile. Another limitation is that all of this study's analyses characterize the association between father's and son's incomes as a linear relationship. This overlooks the possibility that the strength of intergenerational transmission may be greater at one end of the income distribution than at the other. Although the limited size of this study's sample precludes a reliable investigation of such nonlinearities, a richer data set might enable exploration of this issue in the future.

Finally, all of this study's estimates have been based on the simplifying assumption that the variance in long-run status is the same in both generations ($\sigma^2_{y0} = \sigma^2_{y1}$). The numerous studies that have found increasing inequality in annual earnings over recent years[15] call this assumption into question, though whether inequality in *long-run* status has grown remains unclear. But, even if the variance in long-run status grew by as much as 20 percent from the fathers' generation to the sons' ($\sigma^2_{y1}/\sigma^2_{y0} = 1.2$), the estimates in this study would need to be divided through by only $\sqrt{1.2}$.[16] Thus, for example, an estimated intergenerational income correlation of .40 would be revised to .37. Clearly, even extreme adjustments for intergenerational change in inequality would leave intact this study's main finding that intergenerational income mobility in the U.S. is much weaker than previous estimates have suggested.

[15] See, for example, Grubb and Wilson (1989) and Dooley and Gottschalk (1984).

[16] See footnote 3.

Appendix

Suppose that son's long-run income status y_{1i} is determined by

$$\text{(A1)} \qquad y_{1i} = \beta_1 y_{0i} + \beta_2 E_i + u_i$$

where y_{0i} is father's long-run income status, E_i is father's years of education, and all variables are expressed as deviations from means. Equation (A1) differs from equation (1) in Section I in that it distinguishes separate effects of father's income and education. The object of this paper, however, is not to estimate β_1 and β_2, but to estimate ρ, the projection of y_{1i} on y_{0i} alone. The relationship between ρ and β_1 and β_2 follows the familiar omitted-variable formula:

$$\text{(A2)} \quad \rho = \beta_1 + \beta_2 \operatorname{Cov}(E_i, y_{0i})/\sigma_y^2$$
$$= \beta_1 + \beta_2 \lambda \sigma_E / \sigma_y$$

where λ is the correlation between E_i and y_{0i} and σ_E^2 is the variance of E_i.

The difficulty for consistent estimation of ρ is that neither y_{1i} nor y_{0i} is directly observed. Instead, they are proxied by the short-run measures $y_{1it} = y_{1i} + v_{1it}$ and $y_{0is} = y_{0i} + v_{0is}$. Under the assumptions described in Section I, if OLS is applied to the regression of y_{1it} on y_{0is}, the probability limit of the estimated coefficient is

$$\text{(A3)} \qquad \operatorname{plim} \hat{\rho}_{OLS} = [\sigma_y^2/(\sigma_y^2 + \sigma_{v0}^2)]\rho,$$

so that $\hat{\rho}_{OLS}$ is downward-inconsistent.

An alternative strategy is to estimate the regression of y_{1it} on y_{01s} by IV with father's education E_i as the instrument. Assuming that E_i is uncorrelated with v_{1it} and v_{0is}, the probability limit of the IV estimator is

$$
\begin{aligned}
\text{(A4)}\quad \text{plim } \hat{\rho}_{IV} &= \text{plim } [\Sigma E_i y_{1it}/\Sigma E_i y_{0is}] \\
&= \text{plim } [\Sigma E_i(\beta_1 y_{0is} + \beta_2 E_i + u_i + v_{1it} - \beta_1 v_{0is})/\Sigma E_i y_{0is}] \\
&= \beta_1 + \beta_2 \sigma_E^2/(\lambda \sigma_E \sigma_y) \\
&= \beta_1 + \beta_2 \sigma_E/(\lambda \sigma_y) \\
&= (\beta_1 + \beta_2 \lambda \sigma_E/\sigma_y) + \beta_2[(\sigma_E/\lambda\sigma_y) - (\lambda\sigma_E/\sigma_y)] \\
&= \rho + \beta_2 \sigma_E(1-\lambda^2)/(\lambda\sigma_y).
\end{aligned}
$$

Therefore, $\hat{\rho}_{IV}$ consistently estimates ρ only if $\beta_2 = 0$ (father's education does not influence son's status) or $|\lambda| = 1$ (father's education and income are perfectly correlated). Under the more plausible assumptions that $\beta_2 > 0$ (father's education positively influences son's status) and $0 < \lambda < 1$ (father's education and income are positively, but imperfectly correlated), ρ_{IV} is upward-inconsistent. If so, the probability limits of $\hat{\rho}_{OLS}$ and $\hat{\rho}_{IV}$ bracket the true ρ. It may be worth noting that these results are unaffected if the education instrument actually used is not the true E_i, but a proxy subject to classical measurement error.

References

Abowd, John M., and Card, David, "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, March 1989, 57, 411–45.

Altonji, Joseph G., "The Effects of Family Background and School Characteristics on Education and Labor Market Outcomes," unpublished, December 1988.

Auletta, Ken, *The Underclass*, New York: Random House, 1982.

Becker, Gary S., "Family Economics and Macro Behavior," *American Economic Review*, March 1988, 78, 1–13.

_______, and Tomes, Nigel, "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics*, July 1986, 4, S1–39.

Becketti, Sean, Gould, William, Lillard, Lee, and Welch, Finis, "The Panel Study of Income Dynamics after Fourteen Years: An Evaluation," *Journal of Labor Economics*, October 1988, 6, 472–92.

Behrman, Jere, and Taubman, Paul, "Intergenerational Earnings Mobility in the United States: Some Estimates and a Test of Becker's Intergenerational Endowments Model," *Review of Economics and Statistics*, February 1985, 67, 144–51.

_______, "The Intergenerational Correlation between Children's Adult Earnings and Their Parents' Income: Results from the Michigan Panel Survey of Income Dynamics," unpublished, May 1987.

Bound, John, and Krueger, Alan B., "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" Working Paper No. 2885, National Bureau of Economic Research, March 1989.

Corcoran, Mary, Gordon, Roger, Laren, Deborah, and Solon, Gary, "Effects of Family and Community Background on Men's Economic Status," Working Paper No. 2896, National Bureau of Economic Research, March 1989.

Dooley, Martin D., and Gottschalk, Peter, "Earnings Inequality among Males in the United States: Trends and the Effect of Labor Force Growth," *Journal of Political Economy*, February 1984, 92, 59–89.

Duncan, Greg J., and Hill, Daniel H., "An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data," *Journal of Labor Economics*, October 1985, 3, 508–32.

Duncan, Otis Dudley, Featherman, David L., and Duncan, Beverly, *Socioeconomic Background and Achievement*, New York: Seminar Press, 1972.

Grubb, W. Norton, and Wilson, Robert H., "Sources of Increasing Inequality in Wages and Salaries, 1960–80," *Monthly Labor Review*, April 1989, 112, 3–13.

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

Hill, Martha S., and Duncan, Greg J., "Parental Family Income and the Socioeconomic Attainment of Children," *Social Science Research*, March 1987, 16, 39–73.

Kilson, Martin, "Black Social Classes and Intergenerational Poverty," *The Public Interest*, Summer 1981, 64, 58–78.

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

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

Massagli, Michael P., and Hauser, Robert M., "Response Variability in Self- and Proxy Reports of Paternal and Filial Socioeconomic Characteristics," *American Journal of Sociology*, September 1983, 89, 420–31.

Sewell, William H., and Hauser, Robert M., *Education, Occupation, and Earnings: Achievement in the Early Career*, New York: Academic Press, 1975.

Solon, Gary, "Biases in the Estimation of Intergenerational Earnings Correlations," *Review of Economics and Statistics*, February 1989, 71, 172–74.

_______, and Barsky, Robert, "Real Wages over the Business Cycle," Working Paper No. 2888, National Bureau of Economic Research, March 1989.

_______, Corcoran, Mary, Gordon, Roger, and Laren, Deborah, "The Effect of Family Background on Economic Status: A Longitudinal Analysis of Sibling Correlations," Working Paper No. 2282, National Bureau of Economic Research, June 1987.

Survey Research Center, *A Panel Study of Income Dynamics: Procedures and Tape Codes, 1985 Interviewing Year – Wave XVIII, A Supplement*, Ann Arbor: Institute for Social Research, The University of Michigan, 1988.

Topel, Robert, "Wages Rise with Job Seniority," unpublished, November 1987.

Treiman, Donald J., and Hauser, Robert M., "Intergenerational Transmission of Income: An Exercise in Theory Construction," in Robert M. Hauser and David L. Featherman, *The Process of Stratification: Trends and Analyses*, New York: Academic Press, 1977, 271–302.