FORTUNATE SONS: NEW ESTIMATES OF INTERGENERATIONAL MOBILITY IN THE UNITED STATES USING SOCIAL SECURITY EARNINGS DATA*

Bhashkar Mazumder*

Abstract—Previous studies, relying on short-term averages of fathers' earnings, have estimated the intergenerational elasticity (IGE) in earnings to be approximately 0.4. Due to persistent transitory fluctuations, these estimates have been biased down by approximately 30% or more. Using administrative data containing the earnings histories of parents and children, the IGE is estimated to be around 0.6. This suggests that the United States is substantially less mobile than previous research indicated. Estimates of intergenerational mobility are significantly lower for families with little or no wealth, offering empirical support for theoretical models that predict differences due to borrowing constraints.

I. Introduction

PERHAPS one of the most widely held views about the U.S. is that it is a highly mobile society where individuals succeed or fail irrespective of their economic circumstances at birth. The stories written by Horatio Alger in the nineteenth century of individuals rising from rags to riches in a generation present the most idealized version of this belief of rapid upward mobility. The first economic studies of intergenerational mobility appeared to confirm the general view of the United States as extremely mobile. These studies found that the intergenerational elasticity (IGE) in earnings between fathers and sons in the United States and other industrialized countries was less than 0.2, prompting Gary Becker and Nigel Tomes to conclude in a 1986 paper that "Almost all earnings advantages or disadvantages of ancestors are wiped out in three generations."

During the 1990s, however, several economic studies (for example, Solon, 1992; Zimmerman, 1992), using better data, substantially revised this view of rapid mobility in the United States. These studies estimated that the IGE in earnings in the United States is at least 0.4, a figure twice as high as what researchers had previously thought and suggestive of a far less mobile society than was earlier believed.

In this article, I present new evidence that suggests that the IGE is actually close to 0.6. How much difference does it really make if the IGE is 0.6 rather than 0.4? To illustrate the implications in practical terms, let us assume (as Becker and Tomes do) that earnings across generations follow a simple autoregressive process. Now consider a family of four with two children whose income is right at the poverty threshold, roughly 75% below the national average.[1] If the IGE is 0.6, then on average, it will take the descendants of the family 5 to 6 generations (125 to 150 years) before their income would be within 5% of the national average. In contrast, if the parameter is 0.4, the comparable time will be just 3 generations (75 years). Clearly, an IGE of 0.6 paints a radically different picture of mobility in American society and renders stories of rags to riches in a generation more fiction than fact.

Assuming that the parameter represents a *causal* relationship, an IGE of 0.6 also has striking implications about the long-term effects of public policies.[2] For example, Krueger (1995) shows that the effect of the racial segregation of schools on blacks born in the 1920s was to reduce their earnings by 21%. An IGE of 0.6 implies that the offspring of this cohort will have earnings 12% lower due to the segregation of their parents' schools and would account for more than 40% of the black-white earnings gap for this second generation.

How is it that previous studies of intergenerational mobility have underestimated this parameter? The recent studies on the United States come from just two surveys, the Panel Study on Income Dynamics (PSID) and the National Longitudinal Surveys (NLS), both of which have some important shortcomings. The intergenerational samples that can be constructed are relatively small, and there is also considerable attrition in these data sets. This has made direct measurement of permanent earnings impossible. Therefore, researchers have had to rely on alternative strategies to estimate the IGE. The main strategy has been to use short-term averages of earnings as a proxy for fathers' permanent earnings. The use of short-term averages is problematic for two reasons. First, a long literature on earnings dynamics has shown that transitory shocks to earnings are persistent, so that averages of earnings taken over 4 or 5 years still provide rather poor measures of permanent economic status. Second, previous studies have been susceptible to downward-biased estimates in that some fathers' earnings are observed at ages when measurement error is especially high.

Received for publication August 13, 2003. Revision accepted for publication July 6, 2004.

* Federal Reserve Bank of Chicago

I thank David Card, Gary Solon, David Levine, Ken Chay, John DiNardo, Michael Reich, Nada Eissa, Mike Clune, and seminar participants at Berkeley, Illinois, the Chicago Fed, Cornell, UMass, BLS, Census, and The Santa Fe Institute for their helpful advice and comments. I greatly appreciated the help of Andrew Hildreth, Julia Lane, and especially Susan Grad in gaining access to the data. This research was conducted while I was an employee of the Social Security Administration. The help of the staff at SSA, especially Minh Hyunh, is also gratefully acknowledged. The views presented here do not reflect the views of the Federal Reserve System.

[1] The mean income for all families in the U.S. in 2001 was approximately $67,000. See table FINC-01, "Selected Characteristics of Families, by Total Money Income in 2001" at the Census Bureau Web site, census.gov. The poverty threshold is just under $18,000. See Proctor and Dalaker (2002, p. 5).

[2] Whether part of the IGE is causal is an open question. See Solon (1999) for a discussion.

The Review of Economics and Statistics, May 2005, 87(2): 235–255

Incorporating these fairly noncontroversial ideas into the analysis of intergenerational mobility has substantial implications for estimates of the IGE. This is shown both analytically and empirically. Using a simple simulation of the earnings process, I demonstrate that using even 5-year averages of fathers' earnings yields estimates that are biased down by approximately 30%. This is further reinforced by the results from Mazumder (2001), who estimates a highly structured earnings dynamics model on a sample of over 20,000 U.S. men over fifteen years that includes life cycle effects. Using the same methodological approach as Baker and Solon's (2003) study of Canadian men, Mazumder estimates that the bias from using a 5-year average of earnings is greater then 30%. Taking into account these measurement problems strongly suggests that the IGE is closer to 0.6.

I also use a new data source, the 1984 Survey of Income and Program Participation (SIPP) matched to the Social Security Administration's Summary Earnings Records (SER) to produce new empirical estimates of the IGE. Although the data set has some limitations, such as high rates of censored earnings in some years, it provides the long-term earnings histories for both parents and children without any problem of sample attrition. In addition, I use significantly larger samples and richer measures of income and wealth for the parents than in previous studies on U.S. data.

The main result of the study is that I estimate that the IGE between fathers and sons is 0.6 or higher. I attribute the higher estimate to the availability of many more years of earnings data on fathers. This eliminates the substantial downward bias stemming from persistent transitory shocks and also corrects for the age-related errors-in-variables bias by including more data on fathers' earnings during their prime working years. Indeed, the results when fathers' permanent earnings are measured using shorter time horizons closely track the findings from previous research. However, given some of the drawbacks of using social security earnings data, it would still be useful for future research to verify these findings. At a minimum, the results strongly suggest that previous estimates should be considered to be at the low end of the possible range of values for the IGE.

The results should not be viewed as completely surprising. Previous studies (Solon, 1992; Mulligan, 1997) using the PSID that have used instrumental variables (IV) have estimated the IGE to be in the range 0.5 to 0.6. Although researchers have expressed concern that these IV estimates are upward biased, for reasons I discuss in this paper the IV estimates are plausibly consistent and perhaps even downward biased.

This study also provides some other contributions to the literature on intergenerational mobility. I find that the IGE between fathers and daughters is nearly identical to that found between fathers and sons. Because there have been only a few studies that have examined daughters, it has not been entirely clear whether there are significant differences in mobility between sons and daughters. I also find that family income may be a better measure of parents' permanent economic status than fathers' earnings when only a few years of data are available. In fact, the elasticity between family income and children's earnings is higher than the elasticity between fathers' earnings and children's earnings. This is also consistent with the results reported by Chadwick and Solon (2002). This provides additional evidence that the importance of family resources on children's economic opportunities may not be accurately measured by using short-term averages of only fathers' earnings.

I also present new evidence that is consistent with theoretical models that emphasize borrowing constraints as a source of intergenerational inequality (Becker & Tomes, 1986; Mulligan, 1997). Using detailed information on wealth from the SIPP, the IGE is estimated to be significantly higher for families with low net worth, a group that is more likely to be borrowing-constrained. In contrast, the IGE is negligible for those in the top quartile of net worth. These results suggest that education policies that target borrowing-constrained families may foster greater intergenerational mobility. However, further research in this area is needed to more conclusively demonstrate this link.

The paper proceeds as follows: Section II describes the measurement issues involved in studies of intergenerational income mobility. In particular, this section demonstrates how averages of fathers' earning taken over short time spans can substantially underestimate the IGE. In Section III the construction of the matched data set is explained and a number of strategies are outlined to deal with some shortcomings in the data. Section IV presents the methodology used in the study and describes the main results. In addition, a variety of alternative approaches are presented that deal with possible criticisms of the research. Section V presents extensions of the research. This includes an analysis of the effects of family income on children's earnings and how borrowing constraints might influence the intergenerational transmission of inequality. Section VI concludes.

II. Measurement Issues

A. Background

A long tradition beginning with Sir Francis Galton in 1877 has examined the rate of regression to the mean of different characteristics across generations. The seminal work of Becker and Tomes (1979) was among the first attempts by economists to formalize the "heritability" of income. They developed a model in which parents choose between consumption and investment in their children's human capital to maximize utility. Under a set of simplifying assumptions they derived a straightforward result that son's income is a linear function of father's income—suggesting a similar statistical approach to that of the Galton regression model.

Empirical studies undertaken by economists have typically used the following regression model to measure the intergenerational elasticity between fathers and sons:

$$y_{1i} = \alpha + \rho y_{0i} + \beta_1 Age_{0i} + \beta_2 Age_{0i}^2 + \beta_3 Age_{1i} + \beta_4 Age_{1i}^2 + \varepsilon_i. \tag{1}$$

Here y_{1i} represents a measure of economic status such as the log of annual earnings of the son in family *i,* and y_{0i} is the corresponding measure for the father. The only additional right-side variables that are generally included are age and age squared, in order to take account of the effects of the lifetime profile of earnings for both the father and son. Other covariates are generally not included in equation (1), because the goal is to obtain a summary measure of all the factors related to income that are transmitted over generations. Therefore, ρ should not be given a causal interpretation. Ordinary least squares (OLS) is generally used to estimate the equation. The coefficient of interest, of course, is ρ, which measures the IGE.[3]

As might be expected, the first data sets used to create intergenerational samples used obscure samples with limited economic information. For example, Behrman and Taubman (1985) used a sample of white male twins who served in the armed forces. Sewell and Hauser (1975) used a sample of high school seniors in Wisconsin who were no longer in school 7 years later. These studies and others typically used only single-year measures of fathers' income or earnings and found the intergenerational coefficient to be less than 0.2. These results prompted Becker and Tomes (1986) to conclude that intergenerational mobility was quite rapid.

Bowles (1972) and Solon (1989, 1992) demonstrate that the use of single-year measures of economic status as a proxy for permanent status will sharply underestimate the IGE, due to transitory fluctuations to earnings. Solon (1992) also points out how the use of homogeneous samples will reduce the signal in the data because the variation in permanent status will be relatively low, leading to even greater downward bias.

Several studies in the early 1990s (such as Altonji & Dunn, 1991; Solon, 1992; Zimmerman, 1992) used either the PSID or the NLS—longitudinal data sets that were nationally representative and allowed for multiple-year measurements—to address these problems.[4] These studies typically used several empirical strategies. I begin by discussing the simplest and most commonly used method, which is to average fathers' earnings over multiple years. Using this approach, the estimates in nearly all cases were significantly higher than the coefficient 0.2 from the early literature and instead pointed to an IGE around 0.4 or higher.

Unfortunately, due to heavy attrition in the PSID and NLS, these researchers were forced to rely on short windows to measure fathers' earnings in order to maintain a reasonably large sample that was not highly influenced by selection issues. For example, Solon (1992), using the PSID, has only 290 father-son pairs in the samples that use 5-year averages of fathers' earnings. Zimmerman (1992) using the NLS has only 192 father-son pairs when using a 4-year average. Researchers understood that using short-term averages should reduce but not entirely eliminate the bias and that an estimate based on a time average should be treated as lower bound. The question then is how much bias remains? If there is a substantial *persistence* in transitory fluctuations, then even averaging earnings over four or five years may still provide a poor measure of permanent economic status and result in substantial attenuation bias when estimating the IGE. In fact, many studies have used error component models of earnings or wages and found that the transitory component is highly serially correlated (for example, Lillard & Willis, 1978; Card, 1994; Hyslop 2001). This implies that a significant fraction of a transitory shock to fathers' earnings at, say, age 30 might still be felt at age 35 but would have largely died out by age 40.

B. *Incorporating Autocorrelated Transitory Shocks*

Both Solon and Zimmerman consider more complicated error structures for the earnings process but do not calibrate the empirical effect of measuring father's earnings with an average taken over only a few years. To fully gauge the effects of persistent transitory fluctuations on estimates of the IGE, I revisit the standard statistical framework used by most studies:

$$y_{0is} = y_{0i} + w_{0is}, \tag{2}$$

$$y_{1it} = y_{1i} + w_{1it}, \tag{3}$$

$$w_{0is} = \delta w_{0\,i\,s-1} + \xi_{is}, \tag{4}$$

$$y_{1i} = \rho y_{0i} + \varepsilon. \tag{5}$$

In this setup, y_{0is} represents the father's log earnings in year *s,* and y_{1it} is the earnings of his son in year *t*.[5] Equation (2) breaks down the father's earnings in a particular year into

[3] In a bivariate regression, if the variance in log earnings is the same for both generations, then ρ is also the intergenerational correlation (IGC) which provides a useful measure of positional mobility. In contrast, the IGE provides a measure of the degree to which earnings regress toward the mean. In practice, the two measures are roughly comparable (Solon, 1992). With current intergenerational samples, the IGC is more susceptible to bias than the IGE, because it is typically more difficult to proxy for *children's* permanent status.

[4] Solon (1999) identifies fifteen different studies using these surveys.

[5] For simplicity, earnings are assumed to be measured as deviations from the sample mean and are adjusted for age and age squared.

two components: y_{0i}, a permanent component that reflects the true long-term earnings capacity, and w_{0is}, a component that captures any transitory shocks to earnings. Equation (4) models the transitory component as a first-order autoregressive process with parameter δ.[6] As before, equation (5) is the key equation relating father's *permanent* earnings to son's *permanent* earnings.

Of course, in practice researchers do not actually estimate equation (5) but instead use a T-year average of father's earnings as a proxy for y_{0i}. In such cases, the estimate $\hat{\rho}$ of ρ will be biased toward 0 by an attenuation factor λ_T that will vary depending on how many years are averaged:

$$\operatorname{plim} \hat{\rho} = \rho\lambda_T, \tag{6}$$

where

$$\lambda_T = \frac{\sigma^2_{y0}}{\sigma^2_{y0} + \frac{1}{T}\alpha\sigma^2_{w0}}$$

and

$$\alpha = 1 + 2\delta\,\frac{T - \frac{1-\delta^T}{1-\delta}}{T(1-\delta)}.$$

The attenuation factor, which is sometimes called the "reliability ratio," gives an estimate of how much signal is provided by the measure relative to the total variance (signal plus noise). As equation (6) shows, in the absence of serial correlation in transitory fluctuations (that is, $\delta = 0$), the coefficient $\alpha = 1$, and it is clear that averaging substantially lowers the noise relative to the signal, as Solon argued. With serial correlation, however, λ_T becomes a fairly complicated function of δ and it is no longer obvious whether short-term averaging (say, $T = 5$) will significantly reduce the noise. One simple way to gauge the potential effects on the attenuation bias is to conduct some simulations using plausible values for δ and the other parameters.[7]

Dividing the numerator and denominator of equation (6) by the variance in single-year earnings, σ^2_{yt} and then using estimates for δ, $\sigma^2_{y0}/\sigma^2_{yt}$, and $\sigma^2_{w0}/\sigma^2_{yt}$ enables one to calculate λ_T. Table 1 presents the results of a simulation that calculates the reliability ratio of a multiyear average as a proxy for the permanent component of earnings. A review of a few studies that decompose the variance of a single year of earnings or wages into permanent versus transitory factors finds that only about half the variance is due to the permanent component.[8] For this exercise, δ is allowed to vary from 0 to 0.7. As table 1 shows, using a plausible value of 0.5 for δ, the estimate for the reliability ratio when using a 5-year average of fathers' earnings is 0.69. This suggests that estimates of ρ of 0.4 based on 5-year averages are biased down by approximately 30% and that the true IGE may be approximately 0.6.[9] The profile of the reliability ratio as averages are taken over progressively more years is presented graphically in figure 1 and shows a concave pattern. Here it is clear that an average over something on the order of 20 to 25 years is needed to obtain a reliability ratio close to 0.9.

TABLE 1.—SIMULATION RESULTS ON ATTENUATION BIAS WHEN USING MULTIYEAR AVERAGES

No. of Years Averaged	Attenuation coefficient if $\sigma^2_{y0}/\sigma^2_{yt} = 0.5$, $\sigma^2_{w0}/\sigma^2_{yt} = 0.5$			
	$\delta = 0$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 0.7$
1	0.50	0.50	**0.50**	0.50
2	0.67	0.61	**0.57**	0.54
3	0.75	0.67	**0.62**	0.57
4	0.80	0.72	**0.66**	0.60
5	0.83	0.76	**0.69**	0.62
6	0.86	0.78	**0.72**	0.64
7	0.88	0.81	**0.74**	0.66
8	0.89	0.82	**0.76**	0.68
9	0.90	0.84	**0.78**	0.69
10	0.91	0.85	**0.79**	0.71
11	0.92	0.86	**0.81**	0.72
12	0.92	0.87	**0.82**	0.73
13	0.93	0.88	**0.83**	0.74
14	0.93	0.89	**0.84**	0.75
15	0.94	0.89	**0.85**	0.76
16	0.94	0.90	**0.85**	0.77
17	0.94	0.90	**0.86**	0.78
18	0.95	0.91	**0.87**	0.79
19	0.95	0.91	**0.87**	0.80
20	0.95	0.92	**0.88**	0.80
21	0.95	0.92	**0.88**	0.81
22	0.96	0.92	**0.89**	0.82
23	0.96	0.93	**0.89**	0.82
24	0.96	0.93	**0.89**	0.83
25	0.96	0.93	**0.90**	0.83
26	0.96	0.93	**0.90**	0.84
27	0.96	0.94	**0.90**	0.84
28	0.97	0.94	**0.91**	0.85
29	0.97	0.94	**0.91**	0.85
30	0.97	0.94	**0.91**	0.85

Note: Simulation is based on equation six. The permanent component share of the variance in single-year earnings is assumed to be 0.5. The autocorrelation coefficient is either 0, 0.3, 0.5, or 0.7. The preferred set of assumptions is shown in bold.

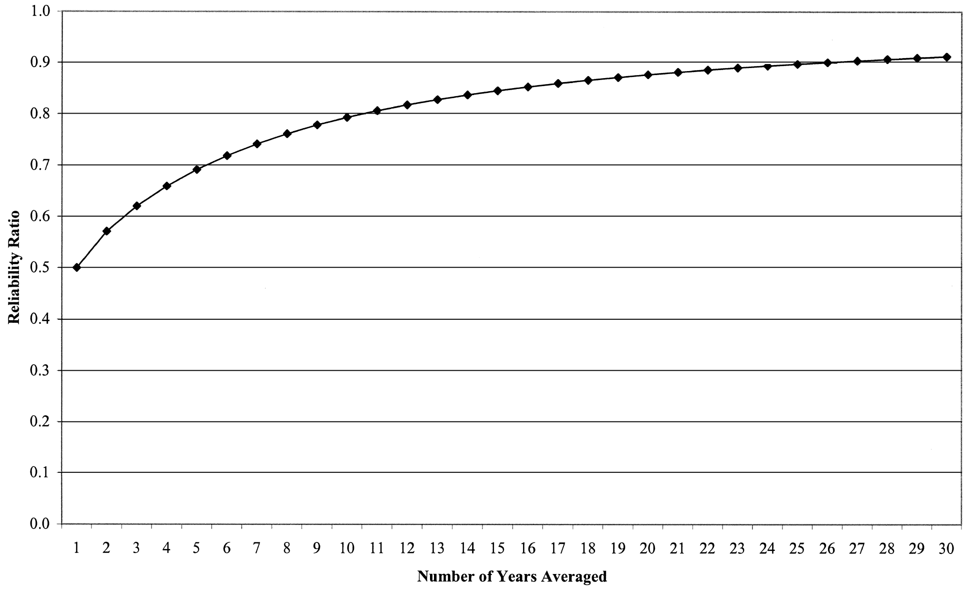
[6] The analogous equation for the transitory shocks to son's earnings is not shown. In the regression context, mismeasurement of the dependent variable will not result in bias unless it is correlated with the right-side variable.

[7] A possible problem arises if δ becomes very large. In that case transitory fluctuations for finite-lived individuals are effectively "permanent," for they may become a sizable part of an individual's actual lifetime income stream. An earlier version of this paper addresses this issue using a more complicated model but still finds similar results. I thank the referee for suggesting a more simple and transparent approach.

[8] Solon et al. (1991) find the permanent share to be approximately 0.55 (table 3) and argue that the upper bound is 0.7. Their analysis, however, strongly suggests that the estimates of 0.7 from studies in the 1970s and 1980s did not adequately adjust for age. Mazumder (2001), using social security earnings data, finds the permanent share to be 0.5. Card (1994) and Hyslop (2001), using the PSID, decompose the variance in hourly wages and find the permanent share to be 0.5.

[9] Taking Solon's (1992) estimate of 0.413 when using five-year averages and using a correction factor of 1/0.69 implies the IGE would be 0.6.

FIGURE 1.—RELIABILITY RATIO AS T INCREASES



An obvious criticism of this exercise is that the earnings process might be considerably more complicated than what is assumed here. For example, allowing for heterogeneity in experience-earnings profiles or including a random-walk component might conceivably alter the results. It could be that including these permanent components in an earnings-dynamics model will change the estimates of the persistence in the transitory component and instead assign a greater share of the overall earnings variance to the permanent component.

Until recently, there have been insufficient longitudinal data to settle these issues. However, a recent study of earnings dynamics among Canadian men by Baker and Solon (2003) uses an extraordinarily large administrative data set and estimates a highly structured model which includes a random-walk component, heterogeneous experience earnings profiles, and a transitory component that varies over the life cycle. Using the same methodology with administrative data in the United States, Mazumder (2001) has shown that such a detailed model has implications on the degree of attenuation bias when short-term averages of earnings are used as proxies for lifetime economic status that are almost identical to those of the simulation results described here.[10]

C. *Age-Related Errors-in-Variables Bias*

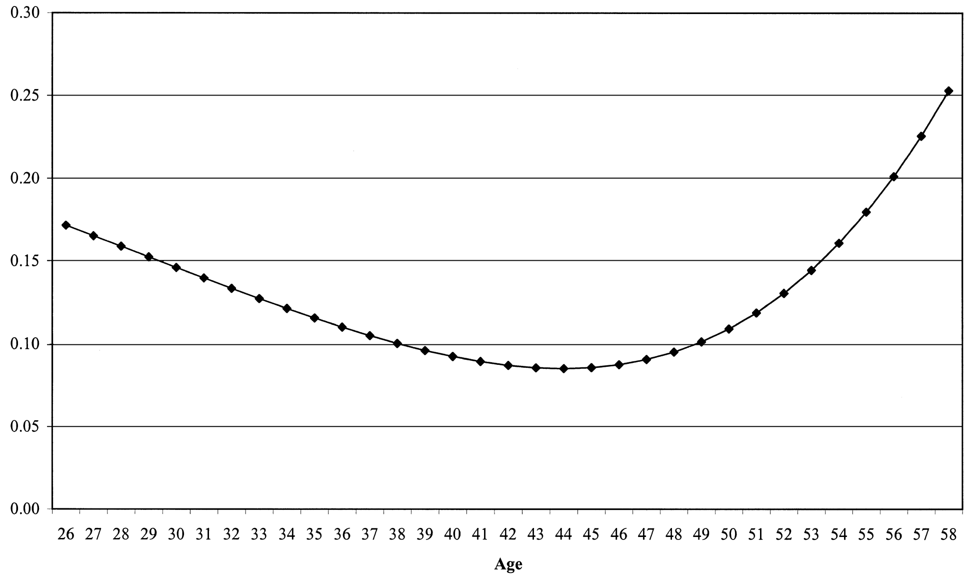
Estimates of the IGE may also be sensitive to the age at which father's earnings are measured. If the variance of the transitory component of earnings changes considerably over the course of the life cycle, then short-term averages of earnings taken at a time when earnings are noisy may lead to further bias. Like Baker and Solon (2003), Mazumder (2001) finds that innovations to the transitory component of earnings follow a U-shaped pattern over the life cycle, as shown in figure 2. This suggests that measures of fathers' earnings around age 40 produce less attenuation bias than those taken at age 30, and substantially less than those taken at age 55.

Recall that the simulation results in figure 1 assume that the transitory variance is constant over the life cycle. Given the pronounced life cycle pattern in earnings, it is worth considering how this finding might affect the results derived from the intergenerational samples that have been used in previous studies. The fact that the transitory noise is so high for older men is particularly troublesome in the case of the NLS, where fathers come from the older cohort of men who were between the ages of 45 and 59 at the start of the survey in 1966. This implies that the NLS studies that average father's earnings over four or five years are actually using income data when the mean age of fathers is in the mid-fifties. Indeed, the fact that several studies using the NLS have estimated the IGE to be substantially less than 0.4 (for example, Altonji & Dunn, 1991; Peters, 1992; Osbourne, 2001) may be due to this problem.

On the other hand, the higher reliability of earnings data for fathers in the middle of their life cycle makes the PSID samples advantageous. For example the mean age of fathers in Solon (1992) is 42, when earnings are likely to be a more reliable proxy for permanent earnings. Still, the fact that father's age in Solon's PSID sample ranges widely, from 27 to 68, suggests that many observations still contain considerable noise. Clearly, access to earnings during the middle

[10] Mazumder (2001) estimates the model on a sample of over 23,000 U.S. men over the period covering 1983 to 1997. The implied attenuation coefficient when earnings are averaged over five years for men born in 1943–1944 (aged 39 to 40 in 1983) is 0.68. This is the average of column 5 in the top panel of table 5 of Mazumder (2001).

FIGURE 2.—LIFE CYCLE PATTERN OF VARIANCE OF TRANSITORY INNOVATION



of the life cycle for *all* fathers would be even better. The administrative data used in this study can potentially address this problem by providing access to a wider range of years of father's earnings.

D. *Life Cycle Bias*

A few studies in the literature (Jenkins, 1987; Grawe, 2003; Haider & Solon, 2004) have made an entirely different point about the importance of age in estimating the IGE that is distinct from the problem of age-related errors in variables. Haider and Solon (2004) show that as a result of heterogeneous age-earnings profiles, the association between current and lifetime income is not constant over the life cycle and doesn't obey the standard errors-in-variables framework employed here and elsewhere in the literature. This issue of life cycle bias implies that the age of both fathers and sons is important to producing consistent estimates of the IGE. Haider and Solon use the entire lifetime social security earnings records of a subset of individuals in the Health and Retirement Survey (HRS) and simulate how the bias in the IGE changes with the age at which son's *and* father's earnings are measured.

Despite their alternative formulation, they reach roughly similar conclusions to those of Mazumder (2001) about the extent of attenuation bias over the life cycle arising solely from fathers' earnings. In fact, they find that the reliability ratio for using, five-year averages peaks at approximately 0.6 for men in their early forties, which is actually lower than what Mazumder (2001) finds. However, they also find that the age at which son's earnings are measured is a second important source of bias. Specifically, they show that the bias is negative for younger sons, positive for older sons, and close to 0 when sons are at age 40. This implies that estimates of the IGE will rise with the age of the sons when they are in their twenties and thirties—an empirical finding first noted by Reville (1995). Although I do not deal with the life cycle bias stemming from sons (or daughters) in this study, I show that the bias from this source is roughly comparable with that in previous studies. In any case, because the samples used here measure children's earnings before age 40, the results of this study may still underestimate the IGE.

Finally, Solon (1992) shows that the attenuation bias is exacerbated if the sample is more homogeneous than the population. This is likely to be the case in the PSID and NLS, due to the high rates of sample attrition.[11] This is another potential advantage to using administrative data as compared to survey data.

E. *Alternative Empirical Approaches*

Several studies have also used IV as a second way to address measurement error. Solon (1992), for example, uses fathers' years of education to instrument for fathers' log earnings and estimates the IGE to be 0.53 using the PSID. Inasmuch as it is possible that fathers' education has an independent, and presumably positive, effect on sons' earnings, the IV estimate may not be consistent, but it arguably serves as an upper bound (Solon, 1992). However, in estimations using the National Longitudinal Survey of Youth (NLSY), I find that the effect of father's education on son's earnings is indistinguishable from 0 once family income is

[11] Solon (1992), for example, uses less than 60% of the original cohort of sons and acknowledges evidence of greater homogeneity in the resulting sample (for example, only 6% of the sample is black).

averaged over 3 or 4 years.[12] The same result has been found by Corcoran et al. (1992), using the PSID and other studies cited in that article. This suggests that 0.53 might actually be a consistent estimate for the IGE. Because it is estimated with a standard error of 0.14, one could not reject that the IGE is 0.6. Mulligan (1997) also uses the PSID and estimates the IGE in wages using education, sex, occupation, industry, and county dummies as instruments. Mulligan reports estimates in the range 0.5 to 0.6. Finally, Gaviria (2002) uses IV on the PSID and estimates the IGE to be 0.592 when including both sons and daughters and the oversample of poor families.

There is also reason to believe that the IV estimates may be biased down. Haider and Solon (2004) demonstrate that in the presence of life cycle bias, the IV estimator is no longer consistent and the bias will depend on the ages of the sample of fathers and sons used. In the usual case of sons under age 40 and fathers over age 40, the IV estimates are likely to be downward biased.[13]

Zimmerman (1992), using the NLS, instruments for father's log earnings with a Duncan index of socioeconomic status. The estimates are widely dispersed from 0.27 to 0.68 with large standard errors and have a central tendency of around 0.4. Zimmerman also uses the forward quasi difference of father's log earnings as an instrument to remove the transitory component with roughly similar results. Because Zimmerman's NLS sample uses a sample of older fathers, the life cycle bias may play a role in depressing the IV estimates as Haider and Solon's analysis suggests.

A third approach imposes a parametric structure on the earnings process as a way to deal with transitory fluctuations. Altonji and Dunn (1991) assume that earnings shocks follow an MA(2) process and estimate the IGE at 0.39. Zimmerman (1992) assumes that earnings shocks follow an AR(1) process and estimates the IGE at 0.40. However, as Solon (1992) points out, uncertainty at the time of those studies regarding how to correctly model the earnings process, combined with data limitations, suggests that this is not the best approach and may not lead to correct estimates.[14] In addition, both studies use the NLS, where the use especially, of older fathers may be problematic. As noted earlier, Mazumder (2001), using Baker and Solon's (2003) highly structured parametric model of earnings, finds much larger biases from using short-term averages than previous studies.

[12] The NLSY sample used for the estimation is described later, in section IV.

[13] Haider and Solon (2004) show that in the presence of nonclassical measurement error but with other standard assumptions in place, the probability limit of ρ using IV is equal to $(\lambda_t/\lambda_s)\rho$. Here, λ_t represents the slope coefficient from regressing sons' earnings in year t on the present discounted value of sons' lifetime earnings, and λ_s is the corresponding measure for fathers. With classical measurement error the λ's are equal to 1 and the IV estimator is consistent. However, Haider and Solon show that λ is less than 1 for those under 40 and greater than 1 for those over 40.

[14] See footnote 16 in Solon (1992).

III. Data Issues

A. *Overview of SIPP and SER*

This analysis uses the 1984 SIPP matched to Social Security Administration's (SSA) SERs.[15] The 1984 SIPP was a nationally representative longitudinal survey, which started with over 50,000 individuals in nearly 20,000 households. Interviews took place every 4 months and resulted in highly detailed data on employment, income, and government program participation for a $2\frac{1}{2}$-year period.[16] Respondents were asked to provide the social security numbers (SSNs) of their family members, and an attempt was made to match individuals in the SIPP to their SERs. The SER file contains the individual's annual taxable earnings from 1951 to 1998, but little other useful economic or demographic information.

B. *Matching Issues*

This matched file allows for intergenerational analysis of families where children were living with their parents between June 1983 and June 1984 and where the parents and children had SSNs that were provided to interviewers.[17] The universe selected for analysis is children born between 1963 and 1968 who were coresident with either or both parents during the first wave of the 1984 SIPP (June–September 1983).[18] The age range was limited to those 15 or older in 1983, because of the poor match rate for younger children.[19] This lower bound on age also ensures that the sons and daughters are at least 27 years old when their earnings are

[15] These data are not publicly available, and access to them requires a special arrangement with the Census Bureau or the Social Security Administration.

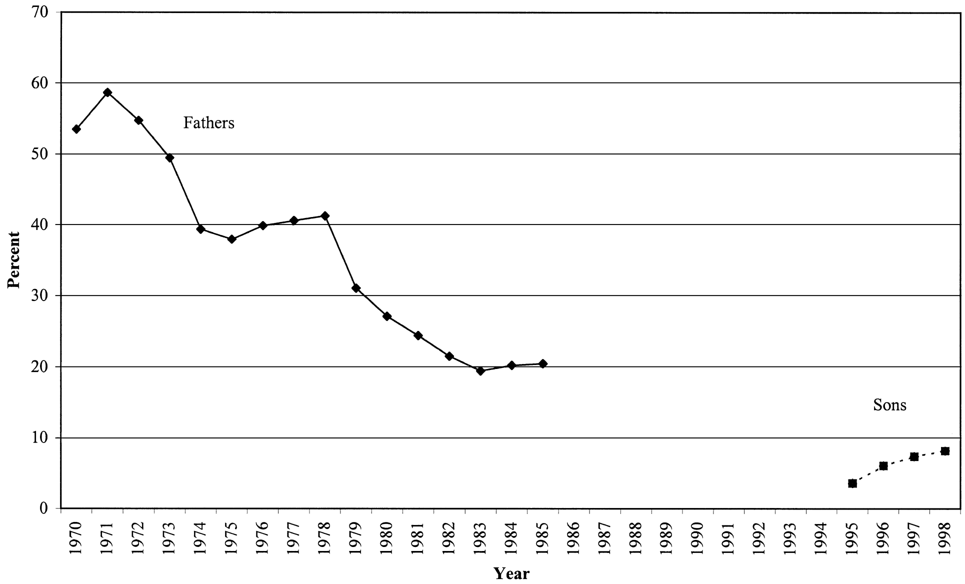
[16] Each interview wave also contains a variety of topical modules that provide rich information on employment history and wealth among other things.

[17] An intergenerational sample can also be created that links only the children to their SER data and uses parent information from the SIPP. Also, most children are directly linked only to their mother (when one exists in the household). Links to the fathers are via links from the mother to the mother's spouse. Following previous studies, the analysis here is focused on the broad effects of family, not just genetic influences.

[18] This includes students living away from home while in college.

[19] In the early 1980s social security numbers (SSNs) were not universal among children. Nonetheless, for those 15 to 20 close to 90% are matched. For those 17 and older, more than 93% are matched. An econometric analysis of the match, for both the 15-to-20 age group and the full sample of the 1984 SIPP, was conducted using an extensive set of demographic, economic, financial, and government program participation variables. For 15- to 20-year-olds, very few of the variables were statistically significant. The key factors that were statistically significant were having worked, receiving food stamps, being in the armed forces, and not being Mexican. A few states like Florida and Texas had significantly higher probabilities of a match that were statistically significant, and a few states like North Carolina and Missouri had lower match rates. The analysis for the sample of individuals of all ages found that the key factors associated with the match included employment, having a bank account, and receiving various forms of government assistance. This suggests that the samples using parent's SER earnings may slightly over-represent both poor and rich households. In section IV sensitivity tests are conducted which show that the match does not appear to affect the main results.

FIGURE 3.—PERCENTAGE OF SAMPLE TOPCODED



observed in the years 1995 through 1998. The sample was also restricted to those who were aged 20 or under in 1983 to ensure that the sample did not over-represent those who stayed at home until a late age.[20] The possible selection biases that could result from these rules are addressed in section IV.

There are a total of 4,072 child-parent pairs in which both the child and at least one parent are successfully matched to the SER file, representing an overall match rate of 87%.[21] In 3,158 cases, sons or daughter and their fathers are both successfully matched to their earnings records. Of these, 1,663 represent father-son pairs, and 1,495 father-daughter pairs.

An alternative to matching the parents to the SER data is to use their income or earnings measures directly from the SIPP for 1984 and 1985. A drawback to this approach is that because of attrition, budget cutbacks, and nonresponse to earnings questions, there is a much smaller sample with complete SIPP earnings data—only 912 father-son pairs and 809 father-daughter pairs.

C. SER Data Problems

In this study, the use of SER data introduces three key concerns. The first is that instances of zero annual earnings may either reflect nonworking or be due to employment in a job that is not covered by social security.[22] Although approximately 90% of jobs in the United States are now covered, in the early 1980s the figure was slightly lower. A second problem is that because earnings are taxed for social security only up to the taxable maximum for the year, the SER file *topcodes* earnings at this cutoff. This is further compounded by the fact that there have been large changes in the real value of the taxable maximum over the last 40 years, resulting in large changes in the fraction of the sample who are topcoded, as shown in figure 3. Finally, even among those with positive earnings, many individuals have both covered and noncovered earnings.[23] In particular, there is a large fraction of people who report dramatically higher earnings in the SIPP than are actually taxed for social security purposes. Each of these three issues affects both the dependent variable (children's adult earnings) and the independent variable (parents' earnings). Because of the differences in available information for the children and for their parents, each of these issues is addressed separately for the two groups.

D. Data Solutions: Children's Earnings

Identifying whether instances of zero covered earnings among the sons and daughters actually reflect nonwork is challenging, for survey data is not available. Approximately 12% of the sons and 21% of daughters had zero covered

[20] Solon (1992) used 18 as an upper age cutoff. Restricting the age to 18 has little effect on the results.

[21] The match rates within the pairs are as follows: fathers alone are matched at a 93.5% rate, mothers alone are matched at a 93.2% rate, sons alone are matched at an 88.8% rate and daughters alone are matched at an 88.2% rate.

[22] Many federal, state, and local government workers are not covered by social security. In addition, workers in the underground economy and workers in certain occupations are paid outside the tax system.

[23] This may be due to having more than one job, tax avoidance, or a desire to maintain social security eligibility if one's main job is not covered.

earnings in 1996. To address this, the 1996 SIPP-SER, which matches a completely different set of individuals interviewed in the late 1990s to their social security earnings, was used to distinguish noncovered employment from nonwork and impute the earnings data accordingly. The imputation procedure is described in detail in the appendix.

The problem of topcoding is not severe for the children: only 6% of the sons and 2% of the daughters were topcoded in 1996. Two approaches are used to address this problem. First, tobit models rather than OLS are estimated, as is discussed in the next section. Second, the 1996 SIPP-SER is used to impute earnings for those topcoded in 1995 through 1998, as described in the appendix. Results using both approaches will be shown in section IV.

The implications of the fact that some children will have both covered and noncovered earnings on the analysis are not entirely clear. Essentially, it means that for a fraction of the children, observed earnings from the SER will under-represent actual earnings. To the extent that this measurement error in the dependent variable is random, it will not bias the IGE. On the other hand, if this error is correlated with fathers' earnings, then the results will be biased. It is not obvious why sons or daughters whose SER earnings under-represent their true earnings would tend to have fathers with lower average earnings. This might arise if both generations' earnings are measured using data from the SER file and if noncovered status is correlated across generations. In this case the measurement error in children's earnings may be correlated with measured fathers' earnings. If this correlation is large enough, it might result in larger coefficients when SER data are used to measure fathers' earnings than when SIPP earnings are used. It turns out the opposite is true, as is shown in section IV. In any event, there is no simple way to solve the measurement error problem for the dependent variable, given the lack of direct survey data on the children in their adult years.

E. *Data Solutions: Parent Earnings*

The data problems are considerably easier to deal with for the parents, because of the rich information available in the SIPP. For 1984 and 1985 there is very detailed information on labor force status and pay in each month, and therefore it is quite easy to identify whether individuals with zero SER earnings also did not work during the year. For earlier years, a SIPP topical module on labor force history is used to classify fathers with zero earnings, in each year going back to 1979, as either noncovered or nonworkers.[24] Those classified as noncovered are either dropped from the analysis, in which case the estimates of the IGE only apply to the vast majority of workers in the covered sector, or their earnings are imputed using SIPP earnings.

[24] The analysis stops at 1979 because of evidence of recall problems for earlier years.

Topcoding is far more severe for the fathers than the children, because the taxable maximum affected a larger share of the sample in earlier decades, as figure 3 shows. To address this, fathers are divided into six groups by race and education level, and annual earnings are imputed based on the full sample of the 1984 SIPP-SER or from the March Supplements to the Current Population Survey (CPS). This is described in more detail in the appendix. It should be noted that this procedure takes the analysis a step toward the kind of two-stage estimation procedure used by some researchers (such as Bjorklund & Jantti, 1997) who have used demographic variables such as education to predict fathers' earnings in a first stage. As noted in section II, despite concerns that using father's education as an instrument may produce upward bias, there is reasonable evidence that this approach produces consistent estimates of the IGE.

The problem of measurement error due to fathers with both covered and noncovered earnings is handled through the use of the "class of worker" variable in the 1984 SIPP. This variable identifies those who worked for the government or who were self-employed at any point that they were in the SIPP. These two categories comprise the vast majority of workers who have some noncovered earnings, and so removing these fathers from the analysis should correct for this problem. In addition to removing downward bias due to measurement error, this procedure has the additional advantage of reducing the possible bias arising from the joint mismeasurement of fathers' and children's earnings that might bias estimates upward as was discussed earlier. The drawback of this approach is that it reduces the sample size by roughly a third and it changes the interpretation of the IGE as applying only to workers in the private sector who are not self-employed.

IV. Methodology and Main Results

A. *SIPP Results*

This study begins by estimating the IGE using the SIPP earnings data for fathers. Although the SIPP is limited to just 2 years of earnings and necessitates a smaller sample, it serves as a useful benchmark for the main analysis that uses the SER data. It also serves as a point of comparison with other studies that use survey data from the PSID or NLS. The econometric approach follows the recent literature and estimates the following equation:

$$y_{1i} = \alpha + \rho y_{0i} + \beta_1 Age_{0i} + \beta_2 Age_{0i}^2 + \beta_3 Age_{1i} + \beta_4 Age_{1i}^2 + \varepsilon. \tag{7}$$

Specifically, y_{0i}, the father's earnings, is the log of the average annual earnings of fathers over 1984 and 1985. This includes earnings from up to two jobs and two businesses. In all aspects of this analysis, earnings are converted to 1998 dollars using the Consumer Price Index for all urban consumers. Only those fathers with earnings that are not imputed by the Census Bureau due to nonresponse are included. The father's age, Age_{0i}, and age squared, Age_{0i}^2, are measured in 1984. The son's or daughter's earnings, y_{1i}, is the log of average annual earnings over the years 1995 to 1998, with the added requirement that they must have been classified as working in at least one of the four years.[25] In the estimates that pool sons and daughters, a female dummy variable is included to allow for the large earnings gap between men and women. These years are chosen so the children are no younger than 27 in any of the years that their earnings are measured, thereby giving a more reasonable picture of lifetime earnings.[26] In cases of zero earnings, the procedure described in section III is used to identify and then impute the earnings of noncovered and nonworkers. The children's age measures, Age_{1i} and Age_{1i}^2, use their age in 1998. Table 2 presents the key sample statistics. Unlike some previous studies, if more than one child is matched to a father, all father-child cases are used and the standard errors are corrected for within-family correlation.[27]

The model is estimated in two ways to deal with the issue of topcoded earnings of the sons and daughters. One way is to simply use OLS, but adjust the dependent variable using the imputed earnings calculated from the 1996 SIPP-SER when sons or daughters have been topcoded. The second way is to estimate a tobit model with an individual specific right-censoring point, as follows:

$$y_{1i}^* = \alpha + \rho y_{0i} + \beta_1 Age_{0i} + \beta_2 Age_{0i}^2 + \beta_3 Age_{1i} + \beta_4 Age_{1i}^2 + \varepsilon_i, \tag{8}$$

$$y_{1i} = y_{1i}^* \quad \text{if} \quad y_{1it} < top_t \quad \forall\, t, \tag{9}$$

$$y_{1i} = k_i \quad \text{if} \quad y_{1it} \geq top_t \quad \text{in some } t. \tag{10}$$

Here y_{1i} is the observed level of permanent earnings, which is equal to the actual permanent earnings level y_{1i}^* only if annual earnings each year is below top_t, the taxable maximum earnings in each year. If earnings are topcoded in any one year, then the actual permanent earnings are treated as right-censored at the observed point k_i. A problem with this approach is that it treats individuals the same regardless of the number of times they were censored over the four years. One solution would be to estimate a tobit model for each year using a standard human capital earnings function and then average the predicted earnings over the four years for the censored observations. Given the lack of survey data for the sons and daughters as adults, this was not possible. The disturbance term is assumed to be normally distributed, and maximum likelihood estimation is used to estimate the intergenerational elasticity.[28] An alternative approach that avoids making strong distributional assumptions is to use a technique such as Powell's (1984) censored least absolute deviations estimator (CLAD). However, in order to maintain comparability with the vast majority of the previous literature which has focused almost exclusively on the OLS coefficient as the key parameter, this approach was not chosen.

In the first set of results, shown in table 3, three different sample selection rules are used. First, fathers who do not have positive earnings in both 1984 and 1985 are dropped (row 1). This has been the common practice in previous research. Given that we have only two years of earnings, allowing zero earnings in any year is likely to add considerable noise. The other two exclusion rules drop fathers who

TABLE 2.—SUMMARY STATISTICS FOR FATHERS AND CHILDREN

Variable	N	Mean	S.D.	Minimum	Maximum
Samples Using 1984 SIPP for Fathers' Earnings					
Father's age in 1984	796	46.9	6.2	28	71
Log average father's earnings 1984–1985	796	10.4	0.8	6.1	11.9
Son's age in 1998	796	32.4	1.7	30	35
Log average son's earnings 1995–1998	796	10.0	1.2	2.5	11.1
Daughter's age in 1998	719	32.5	1.7	30	35
Log average daughter's earnings 1995–1998	719	9.1	1.7	4.1	11.1
Samples Using SER for Fathers' Earnings					
Father's age in 1984	1,262	47.1	6.3	27	69
Log average father's earnings 1984–1985	1,262	10.5	0.9	4.0	11.5
Log average father's earnings 1982–1985	1,218	10.6	0.7	6.5	11.5
Log average father's earnings 1979–1985	1,160	10.7	0.6	7.3	11.5
Log average father's earnings 1976–1985	1,111	10.7	0.5	7.7	11.3
Log average father's earnings 1970–1985	1,063	10.7	0.4	8.1	11.2
Son's age in 1998	1,262	32.4	1.7	30	35
Log average son's earnings 1995–1998	1,262	10.0	1.2	2.5	11.1
Daughter's age in 1998	1,178	32.5	1.7	30	35
Log average daughter's earnings 1995–1998	1,178	9.1	1.8	3.1	11.1

Note: All earnings are converted to 1998 dollars using the CPI. Children's earnings are imputed for those predicted to be noncovered or nonworkers as described in text. The SIPP sample pertains to those shown in row 1 of table 3. Fathers in SIPP sample must be present for all of 1984 and 1985 and have no instances of nonresponse to earnings questions. The samples for the SER pertains to those shown in row 1 of table 4. Fathers' age in the SER sample is for the sample used when earnings are averaged over 1984–1985. SER earnings of those topcoded are imputed as described in text. For both SIPP and SER samples, father statistics correspond to the relevant father-*son* samples.

[25] Averages are taken over several years for the children to gain greater efficiency as well as to address the criticism by Couch and Lillard (1998) that Solon (1992) and Zimmerman (1992) both omit years of zero earnings among the children in their work. The requirement of one year of work is to avoid having the correlation between fathers' earnings and children's labor force participation overly influence the results. In section IV, the results are found to be fairly insensitive to this restriction.

[26] As noted earlier, studies with young samples have a lower IGE because of life cycle bias (Haider & Solon, 2004). The average age of the children is 31, which is similar to the averages 29.6 reported by Solon (1992) and 33.8 reported by Zimmerman (1992).

[27] The effects of restricting the sample to the oldest child in a family is shown later in the section.

[28] The "intreg" command in STATA is used, which allows for a variable censoring point for each observation and for clustered standard errors.

TABLE 3.—INTERGENERATIONAL ELASTICITIES USING SIPP FOR FATHERS' EARNINGS

	Elasticity (Standard Error) N					
	Sons		Daughters		Pooled	
Fathers	Tobit	OLS	Tobit	OLS	Tobit	OLS
Log avg. 1984–1985	0.384	0.342	0.360	0.341	0.369	0.332
father earnings >0	(0.091)	(0.085)	(0.106)	(0.103)	(0.069)	(0.066)
each year	796	796	719	719	1515	1515
Log avg. 1984–1985	0.337	0.293	0.367	0.346	0.344	0.304
father earnings >1,000	(0.080)	(0.072)	(0.117)	(0.112)	(0.070)	(0.065)
each year	788	788	713	713	1501	1501
Log avg. 1984–1985	0.349	0.292	0.361	0.337	0.355	0.305
father earnings >3,000	(0.078)	(0.070)	(0.128)	(0.122)	(0.076)	(0.070)
each year	767	767	702	702	1,469	1,469

Dependent variable is log average earnings, 1995–1998.

Note: For the dependent variable, probit models based on the 1996 SIPP matched to SER were used to determine if zero earnings reflected noncoverage or nonworker status and were imputed accordingly. In the case of OLS specification, topcoded children are imputed based on the earnings distribution in 1996 SIPP-SER. Fathers must have been present for all interview months and have no cases of nonresponse to earnings questions. Standard errors are adjusted for within family correlation when more than one sibling is present.

have earnings below either $1,000 (row 2) or $3,000 (row 3) in either year.

Without using any earnings cutoff, the father-son elasticity, which has been the focal point of the literature, is estimated at 0.342 using OLS and a bit higher at 0.384 using the tobit specification. The IGE between fathers and daughters is also quite similar. The tobit estimate is 0.360, which is only slightly higher than the OLS estimate of 0.341. The difference between OLS and tobit should be quite small, for only about 2% of the daughters are topcoded. The results for the daughters might be biased upward if low average earnings are due to a high incidence of nonworking among daughters, which in turn is correlated with fathers' earnings.[29] This could arise, for example, if women who choose not to have children or delay childbearing come from higher-income families.

Using an earnings cutoff does not appear to change the results appreciably. In these cases, the father-son earnings elasticity appears to drop slightly while the father-daughter elasticity remains remarkably stable. A reasonable summary of Table 3 is that the IGE is approximately 0.35 and is not significantly different between sons and daughters.

It should be kept in mind that these results are based only on two-year averages of fathers' earnings. The comparable result from Solon (1992) is 0.385, and from Zimmerman (1992) is 0.481.[30] Zimmerman's sample selection rules, however, focus on parents and children with high labor force attachment, so it is not clear how comparable the results are. Interestingly, Couch and Lillard (1998), using selection rules similar to those employed by Zimmerman on the same data, find the IGE to be 0.37 when using a four-year average. As noted in section II, other studies using the NLS to estimate the IGE using time averages have found lower estimates. Overall, the findings in this study using the two-year averages from the SIPP are roughly comparable to those obtained using the PSID and NLS.

The estimates of the father-daughter elasticity are of some interest, because only a few studies have examined earnings mobility among daughters. To some extent this is due to the difficulties in using earnings as a proxy for daughter's economic status given the more varied labor force participation patterns among women. Previous estimates of the IGE in earnings among daughters using the NLS are only in the 0.1 to 0.2 range but are roughly similar for sons.[31] Studies using the PSID obtain estimates in the vicinity of 0.4 and for the most part also find little difference between sons and daughters.[32]

B. SER Results

The main analysis uses the SER earnings data for the fathers. This not only significantly enlarges the sample, but it also allows for averaging fathers' earnings over many more years. The longer time averages should eliminate most of the attenuation bias stemming from serially correlated transitory fluctuations in earnings. Once again the IGE is estimated separately for sons and daughters and also with both groups pooled. In this exercise all the results are based on the tobit specification using the same dependent variable as in the prior analysis. Fathers' earnings are progressively averaged over more years, beginning with the two-year average of 1984 and 1985, as was done with the SIPP earnings. Additional estimates are based on averages of four years, seven years, ten years, and sixteen years ending in 1985.

Table 4 presents results using the SER data. Two broad categories of selection rules on fathers' earnings are used in

[29] Recall that daughters' average earnings over 1995–1998 may include some years of very low earnings due to nonwork.

[30] This estimate for Solon is the average of the results found in table 2, column 2, of Solon (1992). The estimate for Zimmerman is from table 6, column 2, of Zimmerman (1992).

[31] Altonji and Dunn (1991) obtain a coefficient of 0.22; Peters (1992) estimates the IGE to be only 0.11. See Solon (1999) for a full review of the literature.

[32] These include Minicozzi (1997) and Shea (2000). Chadwick and Solon (2002) report modest differences between sons and daughters but point out that the differences are not always statistically significant.

TABLE 4.—INTERGENERATIONAL ELASTICITIES USING SER FOR FATHERS' EARNINGS

Fathers	Elasticity (Standard Error) N														
	Sons					Daughters					Pooled				
Log Avg. Earn.	84–85	82–85	79–85	76–85	70–85	84–85	82–85	79–85	76–85	70–85	84–85	82–85	79–85	76–85	70–85
Father Earnings Must Be Positive Each Year															
Drop	0.253	0.349	0.445	0.553	0.613	0.363	0.425	0.489	0.557	0.570	0.308	0.388	0.470	0.559	0.600
noncovered	(0.043)	(0.059)	(0.079)	(0.099)	(0.096)	(0.065)	(0.087)	(0.110)	(0.140)	(0.159)	(0.039)	(0.052)	(0.067)	(0.084)	(0.093)
fathers	1262	1218	1160	1111	1063	1178	1124	1070	1031	982	2440	2342	2230	2142	2045
Impute	0.289	0.313	0.376	—	—	0.350	0.395	0.422	—	—	0.322	0.358	0.404	—	—
noncovered	(0.050)	(0.052)	(0.062)			(0.062)	(0.081)	(0.096)			(0.039)	(0.048)	(0.056)		
fathers	1485	1462	1433			1360	1339	1310			2845	2801	2743		
Drop	0.273	0.419	0.474	0.533	0.652	0.526	0.563	0.635	0.750	0.754	0.393	0.487	0.553	0.643	0.707
government &	(0.060)	(0.082)	(0.096)	(0.111)	(0.135)	(0.089)	(0.137)	(0.150)	(0.173)	(0.192)	(0.057)	(0.077)	(0.086)	(0.100)	(0.118)
self-employed	844	825	801	779	746	782	758	736	719	690	1626	1583	1537	1498	1436
Allow Some Years of Zero Father Earnings*															
Drop	0.234	0.334	0.434	—	—	0.312	0.423	0.506	—	—	0.269	0.377	0.472	—	—
noncovered	(0.043)	(0.057)	(0.069)			(0.060)	(0.065)	(0.091)			(0.034)	(0.043)	(0.056)		
fathers	1295	1268	1227			1201	1168	1127			2496	2436	2354		
Impute	0.238	0.342	0.403	—	—	0.295	0.384	0.474	—	—	0.266	0.365	0.441	—	—
noncovered	(0.042)	(0.057)	(0.059)			(0.055)	(0.061)	(0.080)			(0.033)	(0.042)	(0.049)		
fathers	1534	1550	1571			1394	1406	1424			2928	2956	2995		
Drop	0.242	0.355	0.441	0.523	0.575	0.400	0.504	0.600	0.731	0.847	0.304	0.422	0.570	0.622	0.703
government &	(0.059)	(0.080)	(0.084)	(0.101)	(0.109)	(0.084)	(0.083)	(0.113)	(0.130)	(0.145)	(0.046)	(0.061)	(0.073)	(0.081)	(0.087)
self-employed	874	869	862	895	917	803	794	785	825	831	1677	1663	1647	1720	1748

Dependent variable is children's log average earnings, 1995–1998. All results use tobit specification.

Note: For the dependent variable, probit models based on the 1996 SIPP matched to SER were used to determine if zero earnings reflected noncoverage or nonworker status and were imputed accordingly. For fathers, earnings for those identified as noncovered are either dropped or imputed for the years 1979–1985 as indicated. For the years before 1979, no adjustment is attempted. Earnings for topcoded fathers are imputed using March CPS data for 1970 to 1980 and using 1984 SIPP for 1981 to 1985. Standard errors are adjusted for within family correlation when more than one sibling is present.

*Required years of positive earnings are: 1 for 2-year averages; 2 for 4-year averages; 3 for 7-year averages; 7 for 10-year averages; and 11 for 16-year averages.

this analysis. In the top panel of the table, fathers' earnings must be positive in each year. In the lower panel, some years of zero earnings are allowed. Within each panel, there are three additional selection rules: noncovered fathers are dropped; noncovered fathers' earnings are imputed; and government and self-employed fathers and noncovered fathers are dropped. In the first set of results in the top panel (row 1 of table 4), it is not necessary to actually identify covered status, because all fathers with years of zero earnings are dropped. Therefore, it is possible to construct averages that include years prior to 1979. Under the second rule (estimates in row 2), in contrast, averages can only be constructed going back to 1979, because it is difficult to identify covered status in prior years. Under the third rule (row 3), those identified as government or self-employed workers at any time during the 1984 SIPP survey period are dropped.

The results from using the two-year average with SER data are clearly lower than what was found using the SIPP. The highest coefficient is 0.289 when noncovered fathers are dropped from the analysis. The fact that many fathers have noncovered earnings (in addition to covered earnings) that are not captured in the SER data is the obvious explanation for the greater attenuation using the SER data. In fact, when noncovered fathers are dropped and earnings are required to be at least $3,000 in each year, thereby eliminating many of those whose covered earnings severely misrepresent their true earnings, the estimated coefficient rises to 0.334 (not shown), which is comparable to the SIPP results from table 3. This suggests that the results based on the SER may, in fact, be biased down by even more than would be the case with comparable survey data. It also suggests that the possibility of upward bias from correlated measurement error between fathers and children when using SER data is more than offset by the overall attenuation bias. Otherwise the estimates using the SER would have been higher than those found when using the SIPP. It is also apparent from table 4 that on the whole, the IGE is only slightly lower when the imputed noncovered fathers are added to the sample.

The most striking finding is that the IGE rises dramatically as the fathers' earnings are increasingly averaged over more years. Indeed, the estimated father-son elasticity is 0.613 when the fathers' earnings are averaged over 16 years. The father-daughter elasticity is a bit lower at 0.570. When the sample of fathers is restricted to private-sector, non-self-employed workers, however, the father-daughter elasticity is estimated at 0.754. Such a high degree of transmission is rather surprising and may be due to the possible positive correlation between fathers' earnings and daughters' labor force participation among this group, as discussed earlier.

C. Does Excluding Years of Nonemployment Matter?

Couch and Lillard (1998) argue that the results of Solon (1992) and Zimmerman (1992) are sensitive to the inclusion

TABLE 5.—INTERGENERATIONAL ELASTICITIES USING BALANCED SAMPLES

Fathers	Elasticity (Standard Error) N				
Log Avg. Earn. over	84–85	82–85	79–85	76–85	70–85
Sons	0.298	0.401	0.462	0.516	0.613
	(0.063)	(0.071)	(0.079)	(0.085)	(0.096)
	1063	1063	1063	1063	1063
Daughters	0.387	0.410	0.490	0.516	0.570
	(0.103)	(0.121)	(0.139)	(0.149)	(0.159)
	982	982	982	982	982
Pooled	0.315	0.395	0.440	0.517	0.600
	(0.061)	(0.067)	(0.076)	(0.086)	(0.093)
	2045	2045	2045	2045	2045

Dependent variable is children's log Average earnings, 1995–1998. All results use tobit specification.
Note: For the dependent variable, probit models based on the 1996 SIPP matched to SER were used to determine if zero earnings reflected noncoverage or nonworker status and were imputed accordingly. For fathers, SER earnings for those identified as noncovered are dropped. Earnings for those topcoded are imputed using March CPS data for 1970–1980 and using 1984 SIPP for 1981–1984. Fathers must have positive earnings in all years from 1970 to 1985.

TABLE 6.—EFFECTS OF TOPCODED FATHERS ON INTERGENERATIONAL ELASTICITIES

Fathers	Elasticity (Standard Error) N				
Log Avg. Earn. over	84–85	82–85	79–85	76–85	70–85
Positive earnings	0.312	0.385	0.472	0.570	0.624
each year	(0.041)	(0.056)	(0.071)	(0.088)	(0.099)
	2440	2342	2230	2142	2045
Positive earnings	0.245	0.317	0.439	0.451	0.295
each year	(0.049)	(0.074)	(0.121)	(0.182)	(0.237)
drop topcoded fathers	1,713	1,530	1,144	784	343

Dependent variable is children's log average earnings, 1995–1998 (pooled, sons and daughters). All results use tobit specification.
Note: For the dependent variable, probit models based on the 1996 SIPP matched to SER were used to determine if zero earnings reflected noncoverage or nonworker status and were imputed accordingly. For fathers, SER earnings for those identified as noncovered are dropped. In row 1, earnings for those topcoded are imputed using March CPS data for 1970–1980 and using 1984 SIPP for 1981–1984. In row 2 fathers topcoded in any year over the relevant period are dropped.

of years of zero earnings. The estimates in the lower panel of table 4 suggest that the results with this sample are not affected by the inclusion of years of zero earnings for the fathers. For example, when averaging earnings from 1979 to 1985, allowing as many as four years of zero earnings to be averaged in has almost no effect. When noncovered fathers are dropped, the father-son elasticity estimate falls slightly, from 0.445 to 0.434. However, when noncovered fathers are imputed, the coefficient actually rises, from 0.376 to 0.403. In other sensitivity tests the results are found not to depend on the number of years of zero earnings that are included in the father's average.[33]

D. Additional Robustness Checks

One concern might be that the large changes in the IGE are due to changes in the sample composition as fathers' earnings are averaged over more years. In table 5 I address this by using the same sample to estimate the IGE for all of the time spans. Specifically, I reestimate row 1 from table 4 (which drops noncovered fathers) for each of the three groups, using a fixed sample that contains the full sample of those used in sixteen-year averages. The estimates for sons are a bit higher for the two- and four-year averages than they were in table 4, but are fairly close for the longer-term averages. The differences among daughters do not fit any clear pattern. Overall, it is clear that it is the time averaging and not sample composition that drives the results.

A possible problem when using the SER data for fathers' earnings is topcoding of the independent variable. In the absence of any correction, this would result in an upward bias in the IGE. If the true model is in fact a linear relationship, then imputing the topcoded fathers with the mean level of earnings for those topcoded (by race-education groups), in principle, should correct this problem. A simple way to check the robustness of the results of this procedure is to simply drop the topcoded fathers. However, if there are nonlinearities in the IGE, then dropping topcoded fathers may also induce a selection effect. If, for example, the IGE is higher at the top of the income distribution, then dropping the topcoded fathers may in fact lower the estimates due to this selection problem.[34] In addition, because multiyear averages are used, it is not clear how to handle a father who is, say, topcoded in only one of sixteen years. Nevertheless, to make some attempt to address this issue, fathers who are topcoded in any year over the relevant time horizon are dropped.

Table 6 presents the results of the exercise. Sons and daughters are pooled, in order to try to keep the sample as large as possible. For the most part it appears that dropping these fathers lowers the estimates of the elasticity. When using the seven-year average, however, the results are still quite similar. Including topcoded fathers and using the imputation procedure results in an estimate of 0.472, whereas dropping these fathers results in an estimate of 0.439. As noted earlier this could simply be due to a selection effect. This comparison suggests that topcoding does not dramatically affect the results. Averages of fathers' earnings using years before 1979 are probably meaningless for this exercise, because so many of the observations are topcoded at least once (see figure 3) and therefore dropped. For example, the average over 1970 to 1985 has a sample too small to even estimate the coefficient precisely.

E. The Role of Persistent Transitory Earnings and Age-Related Bias

I now examine the extent to which the results are driven by the age of the fathers in the sample, as opposed to the longer-time averages. As discussed earlier, studies have shown that the reliability ratio of fathers' earnings peaks when fathers are around age 40 (Mazumder, 2001; Haider & Solon, 2004). The mean age of fathers in the sample is 47 in 1984. This suggests that using time averages that begin in

[33] These results are available upon request.

[34] Corak and Heisz (1999) in a study of Canada find that the IGE rises sharply at the upper tail of the fathers' earnings distribution.

progressively earlier years and end in 1985 should increase the reliability of the estimates. In fact, this is one of the points of the study. Longer-time averages may produce better estimates of the IGE in part, because they cover the prime working years of fathers. In contrast, a limitation of existing longitudinal studies using the PSID and NLS is that they are unable to measure fathers' earnings in the years *preceding* the beginning of the survey. The key question is whether all of the rise in the estimates is due to the changing age composition of the sample.

One way to check the importance of age is to take time averages of the same length as before (2, 4, 7, 10 and 16 years) but vary the actual years that are covered. I do this in two ways. First, I have all the time averages begin in 1970, when the mean age of fathers is approximately 35, and second, I center all the time averages at the middle of the full 16-year period (1977–1978), when the mean age of fathers is approximately 41. The results when using time averages covering 1970–1971, 1970–1973, 1970–1976, 1970–1979, and 1970–1985 are shown in the top panel of table 7. The first thing to notice is that, as in table 4, the estimates rise as more years are included in the average. However, as suspected, using the earlier time span raises the estimates for the short-term averages. For example, for sons, the IGE for a four-year average beginning in 1970 is 0.48, compared to 0.35 for a four-year average beginning in 1982 (table 4). This probably reflects the fact that many fewer fathers are observed when they are in their fifties, a period when their earnings are especially noisy measures of permanent status.[35]

In the bottom panel of table 7 the averages are taken over the time periods covering 1977–1978, 1976–1979, 1975–1981, 1973–1982, and 1970–1985. For sons, the short-term averages also produce higher estimates of the IGE than the averages ending in 1985, but not as high as the estimates using averages beginning in 1970. This could simply be due to sampling error, but might be explained by the fact that the centered averages contain many more fathers whose earnings are observed in their fifties than the time averages beginning in 1970. On the other hand, using these years result in much higher estimates for daughters. In any case, the central point of the paper, that increasing the length of the time averages reduces the noise from transitory fluctuations, is not affected by this exercise.

Though it appears that the use of short-term averages leads to attenuated estimates of the IGE, it is interesting to compare the empirical results from the various time averages in this section with the simulation exercise undertaken in section II. Figure 4 presents this comparison, assuming that the true IGE is 0.7 and using the pooled estimates from row 1 of table 4 and rows 3 and 6 of table 7. The simulated

[35] Recall the asymmetrical life cycle pattern of transitory innovations shown in figure 2.

TABLE 7.—INTERGENERATIONAL ELASTICITIES OVER DIFFERENT TIME PERIODS

Using Averages Beginning in 1970					
Fathers'	Elasticity (Standard Error) *N*				
Log Avg. Earn. over	70–71	70–73	70–76	70–79	70–85
Sons	0.324	0.483	0.539	0.582	0.613
	(0.068)	(0.090)	(0.102)	(0.103)	(0.096)
	1,453	1,407	1,331	1,252	1,063
Daughters	0.307	0.474	0.521	0.503	0.570
	(0.088)	(0.126)	(0.147)	(0.152)	(0.159)
	1,302	1,263	1,199	1,140	982
Pooled	0.321	0.485	0.537	0.552	0.600
	(0.060)	(0.081)	(0.093)	(0.094)	(0.093)
	2755	2670	2530	2392	2045

Using Averages Centered in 1977–1978					
Fathers'	Elasticity (Standard Error) *N*				
Log Avg. Earn. over	77–78	76–79	75–81	73–82	70–85
Sons	0.296	0.398	0.548	0.532	0.613
	(0.060)	(0.077)	(0.099)	(0.089)	(0.096)
	1393	1331	1237	1185	1063
Daughters	0.452	0.475	0.496	0.545	0.570
	(0.101)	(0.120)	(0.131)	(0.140)	(0.159)
	1276	1223	1143	1103	982
Pooled	0.371	0.441	0.528	0.545	0.600
	(0.058)	(0.070)	(0.082)	(0.084)	(0.093)
	2669	2554	2380	2288	2045

Dependent variable is children's log average earnings, 1995–1998. All results use tobit specification.
Note: For the dependent variable, probit models based on the 1996 SIPP matched to SER were used to determine if zero earnings reflected noncoverage or nonworker status and were imputed accordingly. For fathers, SER earnings for those identified as noncovered are dropped. Earnings for those topcoded are imputed using March CPS data for 1970–1980 and using 1984 SIPP for 1981–1984. Fathers must have positive earnings in all years from 1970 to 1985.

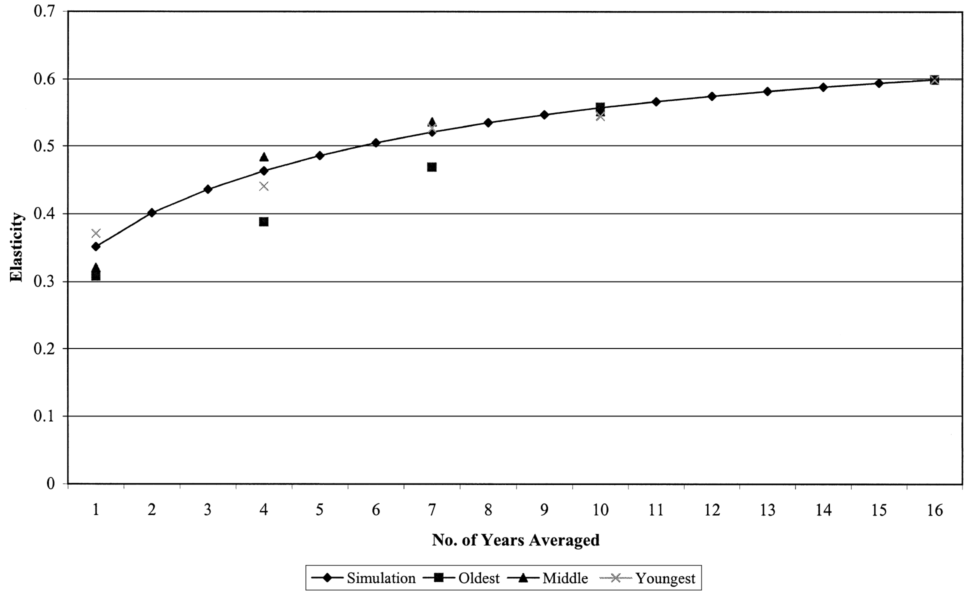
coefficient increases as more years are used, but at a decreasing rate. The results from using the time averages of the fathers ending in 1985 (labeled "oldest"), in contrast, show a much more linear pattern. The results when using the time averages centered in 1977–1978 ("middle") and the time averages beginning in 1970 ("youngest") exhibit much more of the expected concave pattern and may provide more accurate estimates.

Overall, this comparison suggests that it is largely the correction for persistent transitory earnings fluctuations that explains the strikingly high estimates in this study. Still, because of the issues involved in using social security earnings data, the analysis of other matched data sets containing administrative data that may become available in the future may help to settle these measurement issues more conclusively.

F. Comparison with NLSY Estimates

Given the high estimates of the IGE in this sample, it is natural to question whether there is anything about this sample that differs from those used in the earlier studies. One clear difference is that the cohort examined here is more recent, and it might be the case that the IGE is higher

FIGURE 4.—SIMULATION AND ACTUAL ESTIMATES FROM AVERAGING FATHERS' EARNINGS



for this cohort.[36] This is probably not the case, for the estimates taken over shorter time periods appear to match the results from previous studies. To confirm this, a separate analysis was undertaken using the NLSY, 1979 cohort.

The NLSY sample was chosen as a point of comparison because the cohort includes those born between 1957 and 1965, which overlaps with the 1963–1968 cohort used here. The sample is restricted to those born between 1961 and 1965 who lived with their father in 1979, 1980, and 1981. Due to data limitations, the log of a three-year average of family income over the period 1978–1980 is used as the right-side variable. Sons' earnings are measured in 1993 when they were between the ages of 28 and 32. The estimated IGE is 0.448 suggesting that the results from the SIPP-SER sample are consistent with other data covering the same time period.

G. *Other Sample Selection Issues*

I first examine whether the results are sensitive to the requirement that children be living at home with their parents and that the children be matched to their SER data. Inasmuch as the age range of the sample is 15 to 20 and the match rate is approximately 90%, these are not likely to have much effect on the results. Nonetheless, I address this by separately estimating the probability of living at home and the probability of being matched to one's social security records. I then reestimate the IGE using the reciprocals of the predicted probabilities as weights.[37] Table 8 shows the results. The first row presents the results from the bottom row of table 3. The second row weights the observations. The overall elasticity when sons and daughters are pooled is nearly identical at 0.355, but rises slightly for sons and falls slightly for daughters.

Other variations are also attempted in table 8. Restricting the sample to the oldest child in each family has a small but insignificant effect on sons and virtually no effect on daughters. Dropping those aged 19 or 20 in 1983 lowers the elasticity for sons to 0.283. The difference is still within the sampling error but might indicate some effect. The result is consistent with the observation by Reville (1995) and Haider and Solon (2004) that using the earnings of sons when they are observed at a younger age can bias the results downward.[38] The final two rows of table 8 use different sample selection rules on children. Dropping those children identified as noncovered rather than imputing their earnings has almost no effect for sons but a significant positive effect for daughters. Finally, it might be the case that outliers due

[36] For example, Solon (1992) uses those born between 1951 and 1959, and Zimmerman uses those born between 1942 and 1952.

[37] To estimate the probability of living at home, the entire cohort was divided into 24 groups by year of birth, sex, and race. The rate of living at home was then calculated for each group. For estimating whether individuals would be matched to their SER data, a probit was estimated using a wide range of demographic, economic, financial, and government-program-participation variables. The reciprocal of the predicted probability of living at home and being matched to the SER produces the final weight. This approach assumes that there is "selection on observables," in the parlance of Fitzgerald, Gottschalk, and Moffitt (1998)—in other words, that there are no omitted variables that both influence the probability of selection and influence the dependent variable (children's earnings) conditional on the regressors. This appears to be a reasonable assumption for the SER match, for the possession of an SSN is almost wholly a function of observable characteristics that are measured in the SIPP (such as having worked). This is evident when examining the history of the use of the SSN (see http://www.ssa.gov/history/ssn/ssnchron.html).

[38] See footnote 26. It is probably not because older children are more similar to their parents because they lived at home at a late age; for many of those aged 19 or 20 are actually attending college.

TABLE 8.—THE EFFECTS OF SAMPLE SELECTION USING THE SIPP FOR FATHERS' EARNINGS

	Elasticity (Standard Error) N		
Fathers	Sons	Daughters	Pooled
	0.349	0.361	0.355
Log avg. 84–85	(0.078)	(0.128)	(0.076)
father earnings >3,000 each year	767	702	1469
Weighted for	0.375	0.339	0.356
match likelihood	(0.086)	(0.128)	(0.080)
& prob. living at home	767	702	1469
Eldest children only	0.386	0.357	0.378
	(0.095)	(0.147)	(0.087)
	548	506	1054
Aged 15–18 only	0.283	0.400	0.344
	(0.085)	(0.155)	(0.090)
	542	486	1028
Noncovered	0.362	0.473	0.406
children are dropped	(0.094)	(0.113)	(0.072)
	644	498	1142
Require 2 years of positive	0.358	0.363	0.360
children's earnings	(0.080)	(0.130)	(0.078)
	736	687	1423

Dependent variable is log average earnings, 1995–1998. All results use tobit specification.
Note: For the dependent variable, probit models based on the 1996 SIPP matched to SER were used to determine if zero earnings reflected noncoverage or nonworker status and were imputed accordingly (except where otherwise indicated). Fathers' earnings from 1984 SIPP required that the father be present for all interview months and have no cases of nonresponse to earnings questions. Standard errors are adjusted for within-family correlation when more than one sibling is present.

to extremely low values of children's average earnings have influenced the parameter results. To correct for this possibility, children who are identified as nonworkers in more than two of the four years are dropped. This appears to have no effect.

V. Extensions

A. *Family Income*

The analysis so far has focused on intergenerational mobility with respect to labor market earnings. An important question, however, is whether this is the only measure that deserves attention and how the results might differ on examining other outcomes.[39] The main reason to focus on earnings as opposed to say, income or wealth is that conceptually, earning capacity (skills, effort, and the like) cannot be transferred from parents to children, say, the way a house or a financial asset can. In that sense, mobility with respect to earnings provides a better measure of societal opportunities related to individual merit. In addition, from a theoretical perspective, economists have had little to say about the intergenerational effects of unearned income. Nonetheless, from the point of view of measuring the effect of family resources on children's economic opportunities, there may be reasons to prefer alternate measures of parents' economic status, especially when limited years of data on earnings are available.

An interesting finding in some previous studies that has received little attention is that family income is more highly associated across generations than is earnings.[40] This result is consistent with Becker and Tomes's (1986) intergenerational permanent income model where earnings mobility is dictated by regression to the mean in ability, whereas income mobility may be much slower due to the additional possibility of direct financial asset transfers from parents to children.

With the exception of Chadwick and Solon (2002), previous studies have not analyzed the IGE between family income and children's *earnings.* Because this measure looks at children's earnings as an outcome it provides a valid reading of children's economic opportunity as opposed to directly inherited economic status. At the same time, it uses a potentially less error-prone measure of parent permanent status as an explanatory variable. So, though the elasticity of family income on children's earnings is not a measure that comes out of the standard theoretical models, it may still shed light on the magnitude of the effect of family resources on children's future earning opportunities, which is presumably a question of great interest.

Family income provides a broader measure that includes not only the mother's earnings, but also forms of nonearnings income, in the analysis. In the permanent income model, the desire to smooth consumption over time suggests that families would use nonearnings forms of income to offset transitory shocks to earnings. Using family income to estimate the IGE, therefore, presents an alternative approach to overcoming the measurement problems associated with transitory fluctuations to earnings. This might be particularly true at the low end of the parents' earnings distribution, where individuals may receive nonearnings income (such as family transfers or unemployment insurance) at times when they receive virtually no earnings. So we would expect to estimate a higher IGE when family income is used rather than earnings.

Although the SER do not have data on other forms of income, the SIPP provides a highly detailed breakdown of more than thirty different possible sources of income for the parents. Table 9 provides the results of an analysis that substitutes income for earnings in the model and also looks separately at two-parent families, single-mother families, and both types of families pooled together.[41] In all cases, only parents whose income measure exceeds $3,000 in 1998 dollars in 1984 and 1985 are included. Using fathers'

[39] Mulligan (1997), for example, has analyzed mobility with respect to consumption, which for economists is arguably the most important measure, because consumption typically enters the utility function. Unfortunately, due to lack of good data, consumption has been close to impossible to use with intergenerational samples.

[40] These include Mulligan (1997), Shea (2000), Solon (1992), Altonji and Dunn (1991), Corak and Heisz (1999), and Peters (1992).

[41] Single-mother families are simply those where there is no spouse identified for the mother. Obviously, this will miss unmarried couples and other living arrangements where there might be additional sources of income.

TABLE 9.—INTERGENERATIONAL ELASTICITY OF PARENTS' INCOME ON CHILDREN'S EARNINGS

	Elasticity (Standard Error) *N*		
	Sons	Daughters	Pooled
Father earnings	0.349	0.361	0.355
Log avg. 84–85	(0.078)	(0.128)	(0.076)
	767	702	1469
Father income	0.518	0.496	0.507
Log avg. 84–85	(0.102)	(0.119)	(0.084)
	871	773	1644
Two-parent earnings	0.385	0.491	0.442
Log avg. 84–85	(0.075)	(0.118)	(0.069)
	776	719	1495
Two-parent income	0.553	0.708	0.632
Log avg. 84–85	(0.103)	(0.118)	(0.082)
	842	768	1610
Single-mother earnings	0.215	0.357	0.283
Log avg. 84–85	(0.170)	(0.306)	(0.168)
	161	145	306
Single-mother income	0.362	0.287	0.318
Log avg. 84–85	(0.151)	(0.183)	(0.118)
	231	219	450
All family earnings	0.322	0.502	0.412
Log avg. 84–85	(0.060)	(0.098)	(0.055)
	959	879	1838
All family income	0.478	0.558	0.522
Log avg. 84–85	(0.067)	(0.080)	(0.054)
	1105	1006	2111

Dependent variable is log average earnings, 1995–1998. All results use tobit specification.

Note: Probit models based on the 1996 SIPP matched to SER were used to determine if children's zero earnings reflected noncoverage or nonworker status and were imputed accordingly. For SIPP parent measures, parent must be present for all interview months and have no cases of nonresponse to earnings questions. All parent measures require earnings greater than $3,000 in 1998 dollars in 1984 and 1985. Standard errors are adjusted for within family correlation when more than one sibling is present.

income rather than earnings raises the IGE quite a bit. For sons the estimate increases from 0.349 to 0.518.

Using income rather than earnings also appears to raise the IGE sharply when two parents are used and when only single mothers are examined. Adding mothers to the analysis also appears to raise the IGE, particularly for daughters. For example, looking at both parents' income instead of just the fathers' income raises the IGE with daughters' earnings from 0.496 to 0.708. The comparable increase for sons is from 0.518 to 0.553. Looking at single mothers only, the estimated IGEs are dramatically lower and, in most cases statistically insignificant. This is no doubt due, in part, to poor classification of families and resulting mismeasurement.

The results are consistent with the hypothesis that family income may be a less error-prone measure of permanent economic status than earnings, especially when limited years of data are available. However, there also appears to be a sample selection effect. If the IGE is higher at the low end of the distribution, and if more fathers are dropped from the earnings analysis because of exclusion rules on earnings, then including these individuals by using income rather than earnings might raise the elasticity. In fact, if the same sample that is used to estimate the IGE with fathers' earnings in row 1 is also used to estimate the IGE with fathers' income, then the latter estimate falls from 0.518 to 0.385 (not shown). In any case, it appears that using family income rather than fathers' earnings may provide a more accurate measure of the importance of family resources for children's economic opportunities, especially when only a few years of data are available.

B. *Borrowing Constraints*

The evidence presented thus far provides powerful descriptive evidence about the nature of inequality in the United States, but in some ways it raises as many questions as it answers. Should anything be done to promote greater economic mobility? and if so, what? If we start with the assumption that we ought to consider policies to foster greater intergenerational mobility, the research thus far, provides little in the way of policy recommendations. Bowles and Gintis (2002), for example, attempt to decompose the intergenerational correlation using variables such as schooling and measures of cognitive skills such as IQ, but find that at most this accounts for 60% of the parameter value. Other researchers have argued that there is no causal connection between the financial resources of parents and the long-term outcomes of children.[42]

The natural starting point for economists to gain insight into this issue is the human capital framework advanced by Becker and Tomes (1979, 1986) and extended by Mulligan (1997). I briefly summarize the main implications of the model for empirical estimates of the IGE. If parents can freely borrow from their children's future earnings, then all parents will invest the privately optimal amount in their children's human capital. In this scenario, Mulligan shows that because earnings are determined by human capital, and human capital in turn is a function of ability, then the IGE in earnings will only be positive if earnings and ability are positively correlated and will depend on the rate at which ability regresses to the mean. In the presence of intergenerational borrowing constraints, however, parents with high-ability children but insufficient funds will not invest the privately optimal amount in their children's education, thereby inducing a sizable IGE in earnings.

One testable prediction of Becker and Tomes's model is that mobility should be greater among families that make intergenerational income transfers—an indicator that they are not borrowing-constrained. Mulligan in fact attempts to test this hypothesis using the PSID by splitting the sample by those who expect to receive an inheritance. He finds no significant difference in intergenerational mobility between the two groups and concludes that "borrowing constraints do not appear to be an important determinant of

[42] See for example, Mayer (1997) and Shea (2000). See Solon (1999) for a review of these studies.

intergenerational mobility."[43] One problem with this approach is that it does not directly measure parents' ability to finance schooling for their children at the time that such an investment is made. Mulligan's measure also does not capture inter vivos transfers. Finally, the model focuses solely on an intergenerational budget constraint and does not analyze parents' potential inability to borrow from their own future income, which may be an important issue in its own right.

Gaviria (2002) addresses some of these problems. He also uses the PSID, but categorizes the nonconstrained as those who have actually reported receiving large financial transfers or whose parents have a high net worth. Gaviria also uses a split-sample estimation approach and finds some evidence that intergenerational mobility is in fact lower among borrowing-constrained families. However, the differences are not large, and the samples are too small to find differences in the IGE at the 5% significance level.

The SIPP-SER data can bring several clear advantages to this question. First, with a larger sample than the PSID it is easier to detect differences among subgroups when using a sample-splitting strategy. Second, the highly detailed wealth data available in the SIPP make it possible to measure borrowing constraints more directly through net worth. Net worth measures the ability of parents to borrow against their current wealth or to draw down assets in order to finance human capital acquisition for their children. Third, the data on wealth are available for 1984, when the children are between the ages of 16 and 21, and at a time when critical decisions regarding college attendance or continuation are being made.

Table 10 shows the results of this exercise. First, on using the SIPP sample, pooling both sons and daughters together, and splitting the sample by the median level of net worth (approximately $65,000 in 1984 dollars) the results point to a sharp difference between those below the median and those above. The IGE is 0.458 for those with lower than median net worth, but only 0.274 for those above the median level. Though the difference looks large, one could not reject the null hypothesis of equality at the 5% significance level. The second set of results compares those at or below the first quartile of net worth with those at the top quartile. In this case the difference is even more dramatic and is statistically significant at the 5% level. In fact, for the top quartile the IGE appears to be close to 0. Indeed, the permanent income model would predict this result if income is uncorrelated with ability.

Similar attempts were slightly less conclusive using SER data for fathers' earnings, as the bottom half of table 10 shows. Whereas estimates for the low end of the net-worth distribution were similar to that found using the SIPP, the estimates for those with high net worth were significantly

TABLE 10.—INTERGENERATIONAL ELASTICITY BY LEVEL OF NET WORTH

	Elasticity (Standard Error) N				
	Overall	High Net Worth	Low Net Worth	Diff.	t-Stat.
		SIPP Results			
Father earnings	0.369	0.274	0.458	0.184	0.855
Log avg. 84–85	(0.069)	(0.184)	(0.112)	(0.215)	
Low is ≤median	1514	757	757		
High is >median					
Father earnings		−0.044	0.465	0.508	2.795
Log avg. 84–85		(0.135)	(0.122)	(0.182)	
Low is ≤25th percentile		380	379		
High is ≥75th percentile					
		SER Results			
Father earnings	0.480	0.304	0.465	0.160	1.130
Log Avg. 79–85	(0.068)	(0.110)	(0.090)	(0.142)	
Low is ≤median	2,186	1,093	1,093		
High is ≥median					
Father earnings		0.205	0.515	0.310	1.799
Log Avg. 79–85		(0.113)	(0.130)	(0.172)	
Low is ≤25th percentile		547	547		
High is ≥75th percentile					

Dependent variable is log average earnings, 1995–1998. All results use tobit specification. Sons and daughters pooled.

Note: For the dependent variable, probit models based on the 1996 SIPP matched to SER were used to determine if zero earnings reflected noncoverage or nonworker status and were imputed accordingly. Fathers must have positive earnings in each year. When fathers' earnings are from the 1984 SIPP, they must be present for all interview months and have no cases of nonresponse to earnings or income questions. Only those fathers successfully matched to their wave 4 questionnaire are kept in all samples. Standard errors are adjusted for within-family correlation when more than one sibling is present.

higher than that found in the SIPP. A possible explanation for this result is that despite the imputation procedure, the topcoding of fathers' earnings resulted in a more compressed earnings distribution among those with high net worth. In any case, the difference in point estimates when comparing the top and bottom quartiles (0.310) is still quite large and significant at the 10% level.

Although the fact that net worth is measured during college-going years has been presented as a strength, it may actually be a limitation if the relevant period for measuring borrowing constraints should include *earlier* points in the child's educational career. Cameron and Heckman (2001) and Carneiro and Heckman (2003) have argued that much of the inequity in human capital development arises early in life. In this case, empirical evidence of differences in intergenerational mobility by wealth at such a late age might actually be viewed as even stronger evidence in favor of borrowing constraints.

One might argue that because net worth and earnings are highly correlated, any nonlinearities in the IGE may also be reflected in differences in ρ by levels of net worth that may or may not be due to borrowing constraints. To address this issue, the same comparisons in table 10 were also done using quartiles of the earnings distribution rather than net worth as sample cutoff points. The results (not shown) indicated that there are very modest differences between the high and low ends of the earnings distribution. In no case were the differences remotely close to significant, and in

[43] See Mulligan (1997, p. 247).

one case, the sign was actually the opposite of what was expected.[44]

Ultimately the key problem for this type of analysis is that it cannot control for other unobservable factors. For example, it might be the case that families that tend to save also tend to invest in their children's human capital, or that differences in preferences or altruism might account for the observed difference in mobility by net worth. Future research will need to develop more credible research designs that can address this formidable problem. Therefore, though the results here are consistent with the borrowing-constraints hypothesis, they are not conclusive. Still, the findings here run counter to Mulligan's conclusion and deserve attention, particularly given the possibility that easing credit constraints might be a policy lever to influence intergenerational mobility.

VI. Conclusion

This study uses a new, nationally representative intergenerational sample and finds strong evidence that there is far less intergenerational mobility in the United States than was previously thought. The unique advantage of this data set is the availability of long-term earnings histories of fathers. It appears that it is precisely this characteristic of the data which results in the higher estimates. Indeed, estimates based on short-term averages of fathers' earnings closely track the existing literature. Averages of fathers' earnings taken over long periods of time, however, appear to be less sensitive to transitory fluctuations that many studies have shown are persistent. Short-term proxies for fathers' permanent income may also be susceptible to bias because the variance of the transitory component of earnings varies considerably by age.

Overall, the results point toward an intergenerational elasticity of approximately 0.6. This suggests that existing inequalities may persist for many more decades than previously thought. If this parameter reflects a causal relationship, it may have striking implications for understanding the long-term effects of policies, particularly those that either caused, or attempt to redress, historical inequities such as the black-white earnings gap.

The results appear to be fairly robust to sample selection rules, the match process, and the problems that are inherent in the use of social security earnings data. Ideally, future research should attempt to verify the results here using long-term measures of permanent earnings from other sources that do not require the kind of imputations that were necessary in this study. It may be difficult, however, given that existing public-use longitudinal data sets suffer from attrition, age-related measurement error bias, and significantly smaller samples. What may be required in the future is access to other administrative data sets that overcome these data problems. At a minimum, these results suggest that previous estimates of the IGE are likely to be on the low end of the range of possible values.

The use of highly detailed survey data on income from the SIPP from just two years also appears to bolster the main findings. The elasticity of parent income on children's future earnings is estimated to be in the range 0.5 to 0.6. This is due in part to the fact that family income appears to provide a less error-prone measure of permanent economic status—consistent with the permanent income model.

This study provides new descriptive evidence of the extent of intergenerational mobility in the United States, but there is still a tremendous amount that is not understood about how the transmission process works. To what extent is the high estimate of the intergenerational elasticity truly a reflection of the importance of financial resources as opposed to less tangible characteristics that cannot be influenced by public policy? Though not conclusive, new evidence is provided suggesting that intergenerational inequality may be related to access to credit. This points to an important causal channel through which greater mobility may be fostered through public policy.

An obvious candidate for policymakers to consider is whether greater educational attainment can be promoted among poorer households. Some research has suggested that government financial aid promotes college enrollment (for example, Dynarski, 2000; Kane, 2003). Others however, are skeptical about the influence of government programs directed at college enrollment (for example, Cameron & Heckman, 2001) and argue that differences in educational achievement are larger for younger children. An important area for further research is the extent to which borrowing constraints might impede parental investment in their children at earlier ages.

[44] These results are available on request.

REFERENCES

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

Baker, Michael, and Gary Solon, "Earnings Dynamics and Inequality among Canadian Men, 1976–1992: Evidence from Longitudinal Tax Records," *Journal of Labor Economics* 21:2 (2003), 289–321.

Becker, Gary S., and Nigel Tomes, "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy* 87 (1979), 1153–1189.

———, "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics* 4 (1986), S1–S39.

Behrman, Jere R., and Paul Taubman, "Intergenerational Earnings Mobility in the United States: Some Estimates and a Test of Becker's Intergenerational Endowments Model," this REVIEW, 67 (1985), 144–151.

Bjorklund, Anders, and Markus Jantti, "Intergenerational Income Mobility in Sweden Compared to the United States," *American Economic Review,* 87:5 (1997), 1009–1018.

Bowles, Samuel, "Schooling and Inequality from Generation to Generation," *Journal of Political Economy* 80 (1972), S219–S251.

Bowles, Samuel, and Herbert Gintis, "The Inheritance of Inequality," *Journal of Economic Perspectives* 16 (2002), 3–30.

Cameron, Stephen V., and James J. Heckman, "The Dynamics of Educational Attainment for Black, Hispanic and White Males," *Journal of Political Economy* 109:3 (2001), 455–499.

Card, David, "Intertemporal Labor Supply: An Assessment," in Christopher A. Sims (Ed.), *Advances in Econometrics, Sixth World Congress,* Vol. 2 (Cambridge: Cambridge University Press, 1994).
Carneiro, Pedro, and James J. Heckman, "Human Capital Policy," NBER working paper no. 9495 (2003).
Chadwick, Laura, and Gary Solon, "Intergenerational Income Mobility among Daughters," *American Economic Review* 92:1 (2002), 335–344.
Corcoran, Mary, Roger Gordon, Deborah Laren, and Gary Solon, "The Association between Men's Economic Status and Their Family and Community Origins," *Journal of Human Resources* 27:4 (1992), 575–601.
Corak, Miles, and Andrew Heisz, "The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data," *Journal of Human Resources* 34:3 (1999), 504–533.
Couch, Kenneth A., and Dean R. Lillard, "Sample Selection Rules and the Intergenerational Correlation of Earnings," *Labour Economics* 5 (1998), 313–329.
Dynarski, Susan, "Hope for Whom? Financial Aid for the Middle Class and Its Impact on College Attendance," *National Tax Journal* 53:3 (2000), 629–661.
Fitzgerald, John, Peter Gottschalk, and Robert A. Moffitt, "An Analysis of the Impact of Sample Attrition on the Second Generation of Respondents in the Michigan Panel Study of Income Dynamics," *Journal of Human Resources* 33 (1998), 300–344.
Gaviria, Alejandro, "Intergenerational Mobility, Sibling Inequality and Borrowing Constraints," *Economics of Education Review* 2 (2002), 331–340.
Grawe, Nathan D., "Lifecycle Bias in the Estimation of Intergenerational Income Persistence," University of Chicago manuscript (2003).
Haider, Steven J., and Gary Solon, "Life Cycle Variation in the Association between Current and Lifetime Earnings," University of Michigan manuscript (2004).
Hyslop, Dean, "Rising U.S. Earnings Inequality and Family Labor Supply: The Covariance Structure of Intrafamily Earnings," *American Economic Review* 91 (2001), 755–777.
Jenkins, Stephen, "Snapshots versus Movies: 'Lifecycle Biases' and the Estimation of Intergenerational Earnings Inheritance," *European Economic Review* 31 (1987), 1149–1158.
Kane, Thomas J., "A Quasi-Experimental Estimate of the Impact of Financial Aid on College Going," NBER working paper no. 9703 (2003).
Krueger, Alan, "The Legacy of Separate and Unequal Schooling," presentation at Southern Economic Association, New Orleans (November 18–20 1995).
Lillard, Lee A., and Robert J. Willis, "Dynamic Aspects of Earning Mobility," *Econometrica* 46 (1978), 985–1012.
Mayer, Susan, *What Money Can't Buy: Family Income and Children's Life Chances* (Cambridge MA: Harvard University Press, 1997).
Mazumder, Bhashkar, "The Mis-measurement of Permanent Earnings: New Evidence from Social Security Earnings Data," Federal Reserve Bank of Chicago working paper no. 2001-24 (2001).
Minicozzi, Alexandra L., "A Nonparametric Analysis of Intergenerational Income Mobility," University of Wisconsin PhD dissertation (1997).
Mulligan, Casey B., *Parental Priorities and Economic Inequality* (Chicago: University of Chicago Press, 1997).
Osbourne, Melissa A., "Personality and the Intergenerational Transmission of Earnings," Towson State University manuscript (2001).
Peters, Elizabeth H., "Patterns of Intergenerational Mobility in Income and Earnings," this REVIEW, 74 (1992), 456–466.
Powell, James L., "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics* 25:3 (1984), 303–325.
Proctor, Bernadette D., and Joseph Dalaker, *Current Population Reports, P60-219, Poverty in the United States: 2001* (Washington, DC: U.S. Government Printing Office, 2002).
Reville, Robert T., "Intergenerational and Life Cycle Variation in Measured Intergenerational Earnings Mobility," RAND Corp. manuscript (1995).
Sewell, William H., and Robert M. Hauser, *Education, Occupation and Earnings: Achievements in the Early Career* (New York: Academic Press, 1975).
Shea, John, "Does Parents' Money Matter," *Journal of Public Economics* 77:2 (2000), 155–184.
Solon, Gary, "Biases in the Estimation of Intergenerational Earnings Correlations," this *REVIEW,* 71 (1989), 172–174.
——— "Intergenerational Income Mobility in the United States," *American Economic Review* 82 (1992), 393–408
——— "Intergenerational Mobility in the Labor Market," in Orley C. Ashenfalter and David Card (Eds.), *Handbook of Labor Economics,* Vol. 3A, (Amsterdam: Elsevier North Holland, 1999).
Solon, Gary, Mary Corcoran, Roger Gordon, and Deborah Laren, "A Longitudinal Analysis of Sibling Correlations in Economic Status," *Journal of Human Resources* 26:3 (1991), 509–534.
Zimmerman, David J., "Regression toward Mediocrity in Economic Stature," *American Economic Review* 82 (1992), 409–429.

APPENDIX

1. Procedure for Assigning Covered Status among Children

The 1996 SIPP-SER was utilized to classify those born in 1963–1968 with zero earnings in 1996 or 1997 as either nonworkers or noncovered. In this subsample, approximately 57% of the men with zero SER earnings were employed for the full year and are classified as noncovered, while 32% worked for only 0 to 2 months of the year and are called nonworkers.[45] Those working in the noncovered sector are primarily government workers or self-employed. The comparable rates for the daughters were 21% and 71%, respectively. These numbers suggest two conclusions. First, most of those with zero SER earnings are either nonworkers or full-time workers in the noncovered sector. Only about 10% of zero earners fall in the gray area of having zero earnings and working part year. Second, the problem of noncovered workers is particularly important for men.

Because of the clear dichotomy among those with zero covered earnings, probit models were used to predict the probability that individuals with zero covered earnings will have actually worked a full year as a function of all available information contained in the 1996 SER file as well as any basic demographic information that can be determined by adolescence.[46] This function is then applied to the sample of sons and daughters from the 1984 SIPP-SER to obtain predicted probabilities for each individual that they were noncovered. A second set of probit models were also estimated to predict the likelihood that someone with zero covered earnings worked no more than two months of the year. The estimated function is then applied to the sons and daughters to obtain a second set of predicted probabilities. Each of these probit models was estimated separately for men and women, and for both 1996 and 1997.[47] The estimates from the probit models were then combined in order to classify each son or daughter either as a nonworker or as noncovered for each year.[48] Those identified as noncovered are then dropped from the analysis or else their earnings are imputed using the mean level of log earnings for the group from the 1996 SIPP. Similarly, those identified as nonemployed are assigned the mean level of log earnings for those who worked between 0 and 2 months.[49]

[45] The universe is restricted to those who remained in the 1996 SIPP through the end of 1996. Those who are considered employed for the whole year may have worked for 10 to 12 months. Because of the rotation-group structure of the SIPP, some individuals may have only joined the survey starting in February or March of 1996.

[46] For the most part, survey information from the 1996 SIPP is deliberately omitted from this analysis, because such information is obviously unavailable for the sample of sons and daughters from the 1984 SIPP. The exceptions are some basic demographic information and whether individuals ever attended a college. For the sons and daughters, data on whether they ever attended a college over the period of the 1984 SIPP (June 1983–June 1986) can then be exploited.

[47] These were the only years from the 1996 SIPP for which annual earnings were available at the time the research was permitted.

[48] Specifically, individuals are classified according to the category in which they have a higher predicted probability. This is equivalent to assigning them according to the sign of the difference in predicted probabilities. The results for 1996 are used to classify those with zero covered earnings in 1995, and similarly the 1997 results are applied to those with zero covered earnings in 1998.

[49] This strategy allows those children with zero SER earnings in all four years not to be entirely dropped from the analysis simply because of the log

The results of the two probit models for men in 1996 are shown in table A1.[50] Among the key variables that are significant are: having attended college; the number of years of zero earnings during the late 1990s; total lifetime covered earnings; annual earnings in specific years; a flag indicating an active earnings discrepancy;[51] being 29 years old; never having positive covered earnings; being Mexican and being self-employed interacted with 1995 earnings. The fit of these models is quite high as measured by the pseudo R^2. The within-sample forecasting record is also very impressive. For men in 1996, over 90% of the true classifications of noncovered and nonworkers were correctly predicted. In terms of the entire sample of the cohort of men in the 1996 SIPP-SER, this implies that less than 1% of the sample was incorrectly classified. The error in forecasting women's status is higher and implies that approximately 3% of the sample is incorrectly classified. Although it is impossible to know how well this model predicts the correct classification of earnings for the sons and daughters in the 1984 SIPP, the low forecast errors in the 1996 SIPP sample suggest that we can have high confidence in the results.

2. Procedure for Handling Topcoding among Children

The mean value of SIPP earnings of those in the 1963–1968 cohort with SIPP earnings above the social security taxable maximum was calculated for 1996 and 1997. This mean value was then assigned to the children with topcoded earnings. There was no significant difference between the imputed values for men and for women. The 1995 imputation value simply used the 1996 value converted to 1995 dollars using the CPI. Similarly, the 1998 value used the 1997 inflation-adjusted value.

3. Procedure for Handling Topcoding among Fathers

Fathers with topcoded earnings are divided into six race-education cells: by white and black, and by those with less than 16 years of schooling, exactly 16 years of schooling, and more than 16 years of schooling. For each year from 1981 to 1985 the *full sample* of the 1984 SIPP-SER data set is used to create imputed values for each group. Specifically, the mean value of SIPP earnings in 1984 for each topcoded group is calculated and used for imputation.[52]

For the years 1970 to 1980, the imputation values are derived from each year's March CPS instead of the 1984 SIPP. Given the well-documented change in the earnings distribution from the 1970s to the 1980s, it is clearly inappropriate to use the 1984 earnings distribution to calculate the imputed values during the 1970s. For these years, the actual taxable maximum published by the Social Security Administration is used as a cutoff point for the CPS analysis. The mean value of earnings above the taxable maximum for each group is used to impute earnings for those who were topcoded during these years.[53]

Table A1.—Probit Results on Predicting Noncovered versus Nonworker, Men in 1996 SIPP-SER

	Dependent Variable				
	Noncovered		Nonworker		
Independent Variable	*dF/dx* × 100	*z*-Stat.	*dF/dx* × 100	*z*-Stat.	Mean
Black*	6.72	0.47	0.35	0.05	0.13
College*	21.98	2.35	−10.73	−1.91	0.35
Years of 0 Earn. 81–90	−10.30	−2.13	1.97	0.75	5.22
Years of 0 Earn. 91–94	2.77	0.2	−4.12	−0.55	2.17
Years of 0 Earn. 95–98	48.64	2.19	−10.12	−1.21	3.16
Self-employed*	−79.99	−1.59	33.34	0.87	0.22
Agricultural*	15.51	0.92	−8.15	−0.92	0.14
Total quarters of coverage	−2.78	−1.77	0.95	1.15	24.83
Earnings 1981	−0.03	−2.19	0.02	2.31	556.23
Earnings 1982	−0.02	−2.14	0.00	0.81	749.30
Earnings 1983	−0.02	−2.38	0.01	2.39	1,254.62
Earnings 1984	−0.01	−1.68	0.01	1.68	1,703.05
Earnings 1985	−0.01	−1.78	0.01	1.94	2,318.21
Earnings 1986	−0.02	−2.12	0.01	2.2	3,208.63
Earnings 1987	−0.02	−2.62	0.01	2.71	3,981.88
Earnings 1988	−0.02	−2.36	0.01	2.15	4,162.30
Earnings 1989	−0.01	−1.9	0.01	1.92	5,328.86
Earnings 1990	−0.01	−1.97	0.01	2.33	6,030.20
Earnings 1991	−0.02	−2.76	0.01	2.54	5,964.49
Earnings 1992	−0.02	−2.2	0.01	2.45	5,756.44
Earnings 1993	−0.02	−2.11	0.01	2.05	5,153.96
Earnings 1994	−0.02	−2.24	0.01	2.35	3,450.91
Earnings 1995	−0.01	−1.31	0.01	1.72	2,099.37
Earnings 1997	−0.02	−2.35	0.01	2.15	2,532.74
Earnings 1998	−0.02	−2.32	0.01	2.46	5,006.78
Earnings discrepancy flag*	−90.26	−5.97	59.04	6.49	0.29
Military*	6.79	0.38	0.94	0.09	0.07
Age 29*	29.15	2.42	−8.75	−1.19	0.15
Age 30*	0.28	0.02	−7.82	−1.06	0.20
Age 31*	15.58	1.05	−7.49	−0.95	0.16
Age 32*	5.09	0.3	6.69	0.67	0.19
Age 33*	−28.45	−1.22	15.28	1.22	0.16
First year of earnings	−1.13	−0.63	−0.78	−0.68	1,722.25
Last year of earnings	7.58	3.1	−2.92	−2.52	1,732.53
Total earnings to date	0.02	2.35	−0.01	−2.38	60,245.30
Quarters of coverage 1990	0.92	0.2	−3.11	−1.14	2.02
Quarters of coverage 1991	−2.25	−0.37	−0.41	−0.13	2.02
Quarters of coverage 1992	9.97	1.43	−6.86	−1.87	1.81
Quarters of coverage 1993	−3.01	−0.55	1.58	0.52	1.48
Quarters of coverage 1994	−2.63	−0.41	−0.37	−0.11	1.23
Quarters of coverage 1995	4.68	0.6	2.72	0.67	0.72
Quarters of coverage 1997	7.94	0.94	0.12	0.03	0.93
Quarters of coverage 1998	6.86	1.11	−0.35	−0.1	1.19
Never covered earnings	100.00	2.46	−100.00	−2.49	0.13
Newly posted credit earn*	−14.11	−0.55	16.10	1	0.06
Mexican*	25.88	2.86	−11.44	−1.63	0.07
Mexican-American*	−20.86	−0.71	3.44	0.25	0.05
Hispanic*	−7.56	−0.31	−9.18	−1.21	0.05
Earnings 1998 × self-emp.	0.01	2.46	0.00	−1.79	981.06
0 earnings 1995 × self-emp.*	41.91	1.84	−18.20	−1.31	0.16
0 earnings 1997 × self-emp.*	−36.59	−0.98	13.84	0.57	0.14
0 earnings 1998 × self-emp.*	28.55	1.71	−3.45	−0.2	0.11
0 earnings 1995 × agr.*	−25.60	−0.84	31.93	1.32	0.10
No. of 0's 95–98 × 1995 earn.	−0.01	−3.27	0.00	2.23	3,760.91
Observations	258		258		
Pseudo R^2	0.60		0.53		

*Dummy variable; *dF/dx* shows the effect of a discrete change in the variable from 0 to 1.

Note: Sample is from 1996 SIPP matched to SER for cohort born in 1963 to 1968 with zero SER earnings. Sample is restricted to those who are interviewed for at least ten months of 1996 SIPP. "Noncovered" have zero SER earnings but at least 10 months of paid work. Unemployed have zero SER earnings and 0 to 2 months of paid work.

specification. In table 6, results are shown when the sample is restricted to children who are not classified as nonworkers in more than two years.

50 Results for women and for 1997 are available on request.

51 These are cases where individuals have contested what they believe to be inaccurate reports of their earnings with the SSA and where the dispute has not yet been resolved.

52 Only topcoded individuals for whom SIPP earnings in 1984 is greater than or equal to 1984 SER earnings and who are in the SIPP for all twelve months of 1984 are used in the calculations. For the years 1981 to 1983 and for 1985, calculating the imputations involves an added step. The percentile to use as a cutoff for calculating the imputed values for each year is determined by using the percentage topcoded in that year based on the SER data for all the sample members in the 1984 SIPP-SER data set (not just the fathers). For example, in 1980, 8.8% of those with positive earnings in the full sample of the 1984 SIPP-SER matched data set had topcoded earnings. The strategy then was to use the top 8.8% of the SIPP earnings distribution in 1984 to calculate the imputed values for each of the six groups for 1980. Of course, the 1984 dollar values were then converted to 1980 dollars using the CPI.

53 An attempt was also made to use information in the SER data file on the quarter of the year in which full coverage was achieved. For years before 1978 this variable could be used to estimate full year earnings for those topcoded. The results, however, were not significantly different using this strategy.