Federal Reserve Bank of Chicago

Earnings Mobility in the US: A New Look at Intergenerational Inequality

Bhashkar Mazumder

WP 2001-18

Earnings Mobility in the US: A New Look at Intergenerational Inequality

Bhashkar Mazumder
Federal Reserve Bank of Chicago

December, 2001

Abstract

This study uses a new data set that contains the Social Security earnings histories of parents and children in the 1984 Survey of Income and Program Participation, to measure the intergenerational elasticity in earnings in the United States. Earlier studies that found an intergenerational elasticity of 0.4 have typically used only up to five-year averages of fathers' earnings to measure fathers' permanent earnings. However, dynamic earnings models that allow for serial correlation in transitory shocks to earnings imply that using such a short time span may lead to estimates that are biased down by nearly 30 percent. Indeed, by using many more years of fathers' earnings than earlier studies, the intergenerational elasticity between fathers and sons is estimated to be around 0.6 implying significantly less mobility in the U.S. than previous research indicated. The elasticity in earnings between fathers and daughters is of a similar magnitude. The evidence also suggests that family income has an even larger effect than fathers' earnings on children's future labor market success. The elasticity of earnings is higher for families with low net worth, offering some empirical support for theoretical models that predict differences due to borrowing constraints. Some evidence of a higher elasticity among blacks is found but the results are not conclusive.

I am grateful to David Card, 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 Susan Grad in helping me gain access to the data. This research was conducted while the author 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.

I. Introduction

How economically mobile is America? Do all individuals have the same opportunity to achieve success in the United States labor market irrespective of their economic circumstances at birth? Is there an economic underclass that is essentially trapped in poverty for generations? The answers to these questions undoubtedly have a bearing on whether America should be viewed as an equal opportunity society and whether additional policies are needed to address long-term inequities. Despite the obvious importance of economic mobility as a basis for public policy, economists have only recently begun to gain access to the data and develop the tools that might allow for a clearer understanding of the dynamics of inequality among families over generations.

In recent years a growing body of research has used the regression coefficient relating a son's log earnings to his father's, as a summary measure of the degree of intergenerational mobility in society.[1] A high intergenerational elasticity is indicative of a rigid society, since it implies that an individual's position in the earnings distribution is largely a reflection of his parents' position in the previous generation. In contrast, a low intergenerational elasticity suggests a relatively mobile society in which an individual's lifetime income is largely independent of his or her parent's economic standing. In fact, one minus the regression coefficient provides a measure of the degree to which earnings "regress" towards the mean.

One useful way to illustrate the significance of this measure is to imagine what it implies about the evolution of the black-white wage gap in the United States under a set of simplifying assumptions. An intergenerational elasticity of 0.2, for instance, implies that only 20 percent of any earnings gap between groups would remain after a generation (say 25 years).[2] Using this logic, the black-white weekly wage differential that stood at about 25 percent for men of age 25

[1]Recent studies include Solon (1992), Zimmerman (1992), Altonji and Dunn (1991), Peters (1992), Shea (2000), Mulligan (1997) and Corak and Heisz (1999). A full survey can be found in Solon (1999).
[2]This example also assumes a common intergenerational elasticity for both groups and no other group-specific effects. For example, a number of factors such as skill-biased technical change or declining unionism could affect each group differently and temporarily widen the gap further.

to 40 in 1980[3] would be reduced to just 5 percent by 2005 for similarly aged men if all other shocks were ignored. If instead, the intergenerational coefficient was 0.6, then the black-white wage gap would still be a sizable 15 percent in 2005.

Another way to highlight the importance of the intergenerational elasticity is to consider its potential implications on the long-term effects of public policy. If the intergenerational elasticity is sizable and if it represents a causal link that can be exploited by policy makers, then actions taken to improve the fortunes of individuals in one generation might have a large effect on future generations as well.

The results from several studies from the 1990s (e.g. Solon 1992, Zimmerman 1992) have pointed to an intergenerational elasticity in the U.S. of about 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.[4] All of the recent studies on the U.S., however, come from just two surveys, the Panel Study on Income Dynamics (PSID) and the National Longitudinal Surveys (NLS), both of which have relatively small sample sizes, and suffer from considerable attrition when constructing intergenerational samples.[5] In addition, because of data limitations, researchers using these data must estimate fathers' permanent earnings using only a few years of earnings. Using a proxy for permanent earnings based on a short-term average, however, is likely to be flawed since many studies on earnings dynamics have shown that transitory shocks to earnings are highly serially correlated. In fact, using parameter estimates derived from previous studies on earnings dynamics, it is apparent that even five-year averages of earnings yield estimates of the intergenerational elasticity that are biased down by close to 30 percent. This implies that the true intergenerational elasticity may be closer to 0.6.

[3] See Smith and Welch (1989).
[4] Solon (1999) presents a summary of findings of other studies with similar results.
[5] For example, in the samples that use five year averages of fathers' earnings, Solon (1992) using the PSID has only 290 father-son pairs. Zimmerman (1992) using the NLS has only 192 when using a four year average.

This analysis uses a new data source, the 1984 Survey of Income and Program Participation (SIPP) matched to Social Security Administration's Summary Earnings Records (SER) to produce new estimates of the transmission of earnings inequality across generations. Although this data set has some drawbacks, it provides the long-term earnings histories for both parents and children without any problem of sample attrition. In addition, the data provides significantly larger samples and richer measures of income and wealth for the parents.

The key result of this study is that the intergenerational elasticity in earnings between fathers and sons is estimated to be 0.6 or higher, a figure substantially above previous estimates and indicative of a relatively immobile society. The higher estimate is largely attributed to the availability of many more years of earnings data on fathers which eliminates the substantial downward bias stemming from transitory shocks to earnings that exists in previous studies. Indeed, the results when fathers' permanent earnings are based on shorter time horizons closely track the findings from previous research.

This study also generates a number of new findings concerning the persistence of earnings across generations. The intergenerational elasticity between fathers and daughters is similar to that found between fathers and sons. The father-daughter relationship has received scant attention in most of the existing literature on intergenerational mobility.[6] Using data on both parents and using measures of non-earnings income, leads to higher estimates of the intergenerational elasticity. This provides further evidence that previous estimates of intergenerational mobility that were based on short-term averages of fathers earnings may have understated the degree of intergenerational persistence in economic status.

This study also presents evidence that is consistent with theoretical models that emphasize borrowing constraints as a source of intergenerational inequality (Becker and Tomes 1986, Mulligan 1997). Using detailed information on wealth from the SIPP, the intergenerational

[6] This is probably due, in part, to the fact that marriage (coupled with higher average earnings for men) weakens the reliability of daughters' earnings or income as a measure of economic standing.

elasticity is estimated to be significantly higher for families with low net worth and is negligible for those in the top quartile of net worth. These results suggest that policies that target borrowing constrained families may play an important role in reducing inequality over the long-term. The estimates in this study also show a higher intergenerational elasticity among black families than white families, particularly when both parents' earnings are included. The findings, however, are not precise enough to justify a strong conclusion.

A methodological contribution of this study is that careful attention is paid to sample selection rules to address Couch and Lillard's (1998) criticism that past studies may have inappropriately dropped observations if sons or fathers report zero earnings. In this study various exclusion rules are used to analyze the effects of including years of zero earnings for both fathers and their children, and the results are not highly sensitive to these variations.

The paper proceeds as follows: Section II describes the measurement issues involved in studies of intergenerational income mobility. In particular, this section demonstrates how the measures of permanent income in the existing literature that use averages over just five years can substantially underestimate the intergenerational elasticity. In Section III the construction of the matched dataset 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, how borrowing constraints might influence the intergenerational transmission of inequality and results concerning differences in mobility by race. Section VI concludes.

II. Measurement Issues

There is a long tradition dating back to Sir Francis Galton in 1877 that has examined the rate of regression to the mean of different characteristics across generations. Sociologists were the first to apply this type of model in analyzing the transmission of inequality across generations by measuring the correlation of various measures of economic status across generations.[7] The first major economic model to analyze the inheritability of income across generations was by Becker and Tomes (1979). They proposed a utility maximizing framework in which parents choose between current consumption and investment in their children's human capital. 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 as the Galton regression model. A major contribution of their model was their emphasis on human capital as a primary channel by which income inequality is transmitted. On the other hand, as Goldberger (1989) pointed out, it is not clear that the human capital model offers any more empirical content compared to earlier "mechanical" approaches to studying income transmission. More recently, Mulligan (1999) has found only mixed evidence in support of the human capital model versus the Galton model.[8]

In any case, empirical studies undertaken by economists have typically used the following regression model to measure the intergenerational elasticity between fathers and sons:

$$y_{1i} = \alpha + \rho y_{0i} + \beta_1 Age_{0i} + \beta_2 Age^2_{0i} + \beta_3 Age_{1i} + \beta_4 Age^2_{1i} + \varepsilon_i \quad (1)$$

Here y_{1i} represents a measure of economic status such as the log of annual earnings of the son in family i, while y_{0i} is the corresponding measure for the father. The only additional right hand side variables that are generally included are age and age squared, in order to account for the effects of

[7] An early example is Duncan (1961).
[8] In section V some results are presented that are consistent with the human captial model.

the lifetime profile of earnings for both the father and son.[9] Ordinary Least Squares (OLS) is generally used to estimate the equation. The coefficient of interest, of course, is r, which measures the intergenerational elasticity.[10]

As might be expected, the earliest datasets that contained detailed intergenerational information on income used relatively obscure samples.[11] These studies used only single-year measures of fathers' income or earnings and found the intergenerational correlation to be less than 0.2. On the basis of these results and other international studies, Gary Becker in his 1988 address to the American Economics Association, asserted that "In all these countries, low earnings as well as high earnings are not strongly transmitted from fathers to sons…".[12]

As carefully documented by Solon (1989, 1992), there are several problems with using only single-year measures of economic status as a proxy for permanent status that will have the effect of understating the true parameter estimate.[13] These are illustrated in the following statistical framework:

$$y_{0is} = y_{0i} + w_{0is} + v_{0is} \qquad (2)$$

$$y_{1it} = y_{1i} + w_{1it} + v_{1it} \qquad (3)$$

$$y_{1i} = r y_{0i} + e \qquad (4)$$

---

[9] Other covariates have generally not been included in these studies since the goal is to obtain a summary measure of all the factors related to income that are transmitted over generations. Therefore, r should not be given a *causal* interpretation.

[10] If earnings are age adjusted and the variance in log earnings is the same for both generations then r is also the intergenerational correlation. The intergenerational correlation has been emphasized in the sociology literature on intergenerational mobility. The two measures are roughly comparable even if the *variance* in permanent earnings differs across generations as shown by Solon (1992). The intergenerational correlation is more susceptible to bias from mis-measurement of *children's* earnings compared to the regression coefficient. Bowles and Gintis (2001) have also argued that the regression coefficient is a preferred measure since it does not confound changes in cross-sectional inequality with the association in earnings across generations.

[11] 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 seven years later.

[12] See Becker (1988).

[13] Bowles (1972) first pointed out some of the problems with using single year measures of income as a proxy for permanent income.

In this setup, y_{0is} represents the father's log earnings in year s, while y_{1it} is the earnings of his son in year t.[14] Equation 2 breaks down the father's earnings in a particular year into three components: y_{0i}, a permanent component that reflects the true long-term earnings capacity; w_{0is}, a component that captures any transitory shocks that might affect that particular year's earnings; and finally, v_{0is}, a term that captures any errors due simply to mismeasurement such as an inaccurate report of earnings.[15] Equation 3 is the analogous decomposition for the son.

Equation 4 is the relationship of interest between the father's *permanent* earnings and the son's *permanent* earnings. In actuality, researchers with access to only one year's measure of the father and son's earnings will not be able to estimate (3) but instead, will regress the father's *measured* earnings from a single year on the son's *measured* earnings also from a single year. If we assume that the transitory shocks and the measurement error are independent of the true permanent earnings, then the estimate of r, $\hat{r}$, will be biased towards zero by an attenuation factor. It is easily shown that:

$$\text{(5)} \qquad \operatorname{plim} \hat{r} = r\, l,$$

$$\text{where } l = \left(\frac{s_{y0}^{2}}{s_{y0}^{2} + s_{w0}^{2} + s_{v0}^{2}} \right),$$

is an "attenuation" coefficient" arising from the mismeasurement of father's permanent income.[16] The first source of downward bias is generated through s^2_{w0} term, the variance of transitory fluctuations. Second, there is bias due to measurement error in the father's earnings, which is captured by s^2_{v0}, the variance of the measurement error term. Finally, many of the studies use relatively homogeneous samples of fathers, which has the effect of reducing the "signal" in the

[14] For simplicity, earnings are assumed to be measured as deviations from the mean and are adjusted for age and age squared.
[15] For the moment, both the transitory component and the measurement error component are viewed as white noise.
[16] In the regression context, any errors in measuring the son's permanent earnings may lead to less precise estimates but should not lead to biased coefficients. Errors in fathers' earnings, in contrast, will bias the coefficient.

data because s^2_{y0}, is relatively low. Unless the use of a homogeneous sample also happens to reduces the noise, the downward bias will be exacerbated. The severity of these biases may be quite substantial. By some estimates, the transitory component and measurement error term account for about half of the total variance in a single year's earnings.[17]

Several studies in the early 1990s used either the Panel Study of Income Dynamics (PSID) or the National Longitudinal Surveys (NLS) —longitudinal datasets that were nationally representative and allowed for multiple year measurements—to address these problems.[18] By averaging the father's earnings over several years they were able to reduce the bias from transitory income shocks and measurement error.[19] The results in nearly all cases were significantly higher than the 0.2 coefficient from the early literature and instead pointed to an intergenerational elasticity of around 0.4.

These studies, however, overlooked the fact that averaging earnings over a short time span might still result in considerable attenuation bias if there is persistence in transitory fluctuations. In fact, it is well established from many error-component models of long-term earnings profiles that the transitory component of income is highly serially correlated.[20] The implications of this finding on past econometric results that used multiyear averages can best be seen by extending the statistical framework to incorporate serial correlation in the transitory component.[21] Specifically if we model w_{0is}, the transitory component of earnings as a stationary autoregressive process,

$$w_{0is} = d w_{0is-1} + x_{is} \quad (6)$$

[17] See Card (1994) and Hyslop (2001). Solon's (1992) survey of several studies suggested that noise may account for about 30 percent of the variance in single-year earnings.
[18] Solon (1999) identifies 15 different studies using these surveys. Probably the most widely cited are Solon (1992) and Zimmerman (1992).
[19] Additional techniques such as instrumental variables were also used, though these estimates often introduced a positive bias and could only provide an upper bound estimate.
[20] Some examples are Lillard and Willis (1978), MacCurdy (1982), Card (1994) and Hyslop (2001).
[21] While both Solon (1992) and Zimmerman (1992) present formulas on the bias when incorporating serial correlation in the transitory component, they do not pursue the implications of this on their results.

where d represents the autoregressive parameter, then the attenuation coefficient when averaging over T years, l_T, can be expressed as follows:

$$(7) \qquad l_T = \frac{s^2_{y0}}{s^2_{y0} + \frac{1}{T} a\, s^2_w + \frac{1}{T} s^2_v}$$

$$\text{where, } a = 1 + 2d \left\{ \frac{T - \left[\frac{\left(1 - d^T\right)}{\left(1 - d\right)} \right]}{T\left(1 - d\right)} \right\}$$

In the absence of serial correlation in transitory fluctuations (i.e. $d = 0$), the coefficient $\alpha = 1$ in equation (7), and it is clear that averaging lowers the noise relative to the signal. With serial correlation, however, the a term creates an offsetting factor. Indeed, the larger d is, holding the other parameters constant, the larger the overall attenuation bias will be.[22] In order to get a sense of the possible implications, some simulations using plausible values for d, and for the fraction of total variance in one year's earnings that is due to transitory factors, permanent factors and measurement error were undertaken.[23] Using one set of estimates for these parameters from a recent study by Hyslop (2001), the attenuation coefficient when averaging earnings over five years was found to be 0.66.[24]

A limitation with this approach, however, is that a very high persistence in transitory shocks might effectively be considered "permanent" if the effects do not die off over the course of an individual's working life. This problem can be addressed by assuming that the typical father will work for 45 years (from age 20 to age 65) and that what is really of interest for the

---

[22] It should be noted that s^2_w and d are related by $s^2_w = \frac{s^2_x}{1 - d^2}$. In the simulations that follow, assumptions are made regarding s^2_w and d and s^2_x is assumed to adjust to satisfy this relationship.

[23] If the numerator and denominator of (7) are divided by s^2_{yt}, then these are all the parameters that are required to simulate the attenuation coefficients for any given value of T.

analysis is the fathers' average earnings over this 45 year period. If we use the same assumptions on the other parameters as before, then the the attenuation coefficient increases to 0.74.[25] A full set of simulation results is shown in Table 1 where the value of *d* is either 0.5 or 0.8 under three different assumptions about the breakdown in the variance of single-year earnings. These results suggest that estimates of the intergenerational elasticity of 0.4 using five-year averages may still be biased down by about 25 to 30 percent.

As mentioned earlier this bias may be further compounded if the sample is more homogenous which is a distinct possibilty in the PSID and NLS due to the high rates of sample attrition. Solon (1992) for example, uses less than 60 percent of the original cohort of sons and acknowledges evidence of greater homogeneity in the resulting sample.[26]

Recent research has also found that estimates of the intergenerational elasticity may be sensitive to "lifecycle biases".[27] If the variance of the transitory component of earnings changes considerably over the course of the lifecycle, then short-term averages of earnings taken at a time when earnings are considerably noisy may lead to further bias. Indeed, since researchers don't have earnings information from *before* the starting point of longitudinal surveys, the father's age at the time earnings are measured may be quite high.[28] Several studies have found that the transitory component of earnings follows a "U-shaped" pattern over the lifecycle.[29] This suggests that measures of earnings around age 40 may have less attenuation bias than those taken at age 30 or 50.

---

[24] This assumes that *d* = 0.8, that share of the variance in earnings accounted by permanent factors is 0.5, by transitory factors is 0.3, and by measurement error is 0.2. These are also precisely the same estimates found by Card (1994) and Mazumder (2001).

[25] The procedure is described in detail in the appendix.

[26] See Solon (1992) p 398.

[27] See Jensen (1987) and Grawe (2000).

[28] The average age of fathers is 42 for Solon (1992) and is 50 for Zimmerman (1992).

[29] See Gordon (1984), Baker and Solon (1999) and Mazumder (2001).

III. Data Issues

Overview of SIPP and SER

This analysis uses the 1984 Survey of Income and Program Participation (SIPP) matched to Social Security Administration's (SSA) summary earnings records (SER). The 1984 SIPP was a nationally representative longitudinal survey, which started with over 50,000 individuals in nearly 20,000 households.[30] Interviews took place every four months and resulted in highly detailed data on employment, income and government program participation.[31] The survey began in October 1983 and continued until July 1986 covering the period from June 1983 to June 1986. Respondents were asked to provide the social security number of their family members and the SSA subsequently attempted to match individuals who entered the SIPP in one of the first three waves to their SER via their social security numbers. The resulting file contains the individual's SIPP identifiers along with annual taxable earnings from 1951 to 1998.[32]

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 children had social security numbers that were provided to SIPP interviewers.[33] That information would allow

[30] Unlike later SIPPs, there was no oversample of low income households.
[31] There are also a variety of topical modules in each interview wave that provide rich information.
[32] An additional set of variables is also available in the file including date of birth, sex, race, self-employment status, agricultural status, military status and a number of variables related to social security coverage. It should also be noted that the SER file can only be used to gather information on earnings and not other forms of income (e.g., asset income and transfers) that are available in the SIPP and may have been used in previous studies on income mobility.
[33] There are some difficulties in matching children to their fathers using the 1984 SIPP. An explicit description of family relationships does not take place until the eighth wave (January to March 1986) at which point the sample size significantly declines due to budget cuts and attrition. Therefore, in order to use a representative sample, the sons and daughters are matched to their father in the first wave using a roundabout procedure. Although children are directly linked only to their mother (when one exists in the household), they can be linked through the mother to the mother's spouse. This is consistent with the existing literature which has largely not been concerned about whether the matches are to the biological

researchers to link data on parents contained in the 1984 SIPP to their children's earnings as adults up to 1998. In order to go a step further and also access the full social security earnings history for the parents, it is necessary that the parents also provided social security numbers. Therefore an analysis of both children's and parents' full history of earnings requires that *both* be successfully matched to their SER earnings.

The universe selected for analysis in this study includes children born between 1963 and 1968 who were coresident with either or both parents or living away from home while at college during the first wave of the 1984 SIPP (June-September 1983). The age range was limited to those 15 or older in 1983 because of the poor match rate for younger children.[34] This lower bound on age also ensures that the sons and daughters are at least 27 years old when their earnings are observed in the years 1995 through 1998. The sample was also restricted to those who were age 20 or under in 1983 to ensure that the sample did not over-represent those who stayed at home until a late age.[35] The possible selection biases that could result from these rules are addressed in Section IV.

There are a total of 4072 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 percent.[36] In 3158 cases, sons or daughter and their fathers are both successfully matched to their earnings records. Of these, 1663 represent father-son pairs while 1495 cases are father-daughter pairs. An

---

father, arguing that what is being investigated is the broad effect of family background and not just genetic influences.

[34] In the early 1980s social security numbers were not nearly as universal among children as they are today. The key factors that determined whether someone would have a social security number were if he or she worked, had a bank account, owned stocks or received any form of government assistance. An econometric analysis of whether a 15 to 20 year old in the SIPP was matched to the SER showed all of these factors to be significant. This suggests that the sample used here over-represents both poor and rich households. Weighting the sample by the inverse of the probability of being matched has a minor effect on the results as shown in section IV.

[35] Earlier studies such as Solon (1992) and Zimmerman (1992) have used 18 years of age as an upper age cutoff for kids living at home. In the SIPP, however, sons and daughters living away while attending college were considered living at home and are included. The percent of 19 and 20 year olds still living at home or at college in the 1984 SIPP is over 70 percent.

alternative approach is to use SIPP income or earnings data from 1984 and 1985 for the parents instead of matching them to the SER file. A major drawback, however, 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.

SER Data Problems

In this study, the use of SER data introduces three key concerns. The first is that although instances of zero annual earnings may reflect non-working, they could also be due to employment in a job that is not covered by social security.[37] Although about 90 percent of jobs in the U.S. are now covered, in the early 1980s the figure was somewhat lower. However, if even 10 percent of the sample is incorrectly classified as zeroes, this presents a significant problem if regression results are sensitive to sample selection rules around zero earnings. A second problem is that because earnings are only taxed for Social Security 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 forty years resulting in large changes in the fraction of the sample who are topcoded, as shown in Figure 1.

Finally, even among those with positive earnings, a large number of individuals have both covered and non-covered earnings.[38] This is illustrated in Figure 2 which uses the full sample of adults in the 1984 SIPP-SER and plots SER earnings on the x-axis and SIPP earnings on the y-axis. If there was random reporting error, the graph would show a random scattering of points around the forty-five degree line. Instead, there is a large fraction of people who report dramatically higher earnings in the SIPP than are actually taxed for social security purposes.

---

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

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

Each of these three issues may present econometric problems not only because they affect the dependent variable, children's adult earnings, but because they also affect the *independent variable*, parents' earnings. Because of the differences in available information for the children compared to their parents, each of these issues are addressed separately for the two groups.

Data Solutions: Children's Earnings

Distinguishing instances of zero covered earnings among the sons and daughters that are due to lack of work rather than resulting from employment in the non-covered sector is perhaps the most difficult issue of all. The problem is that there are no available data on hours worked for the children in the sample as of the late 1990s. Approximately 12 percent of the sons and 21 percent of daughters had zero covered earnings in 1996.

What turns out to be useful, however is the use of another confidential dataset, the 1996 SIPP-SER, which matches a completely different set of individuals to their social security earnings. In particular, this matched dataset contains detailed earnings from the SIPP for the years 1996 and 1997 along with the social security earnings histories over the years 1951 to 1998. This allows one to identify individuals who had zero social security earnings but reported positive earnings in the SIPP in 1996 or 1997. Focusing exclusively on the 1963-1968 cohort, a series of models were estimated that would allow for classifying men and women as either "non-working" or "non-covered" in each year.[39] These models are then applied to the children in the 1984 SIPP to classify them into these groups for the years 1995 through 1998. Those identified as non-covered are then either dropped from the analysis or their earnings are imputed using the mean level of log earnings for the analogous group from the 1996 SIPP. Similarly, those identified as non-employed are assigned the mean level of log earnings for that group. Based on the within

---

[38] 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.

sample forecasting results, the procedure appears to do a remarkably good job in classifying men into the two groups. The details of the methodology and the statistical results are shown in the Appendix.

The second problem, topcoding at the taxable maximum, is much easier to handle for the children than for the parents. For example, only 6 percent of the sons and 2 percent of the daughters were topcoded in 1996. There are two approaches that are used to address this problem. The first is to estimate tobit models rather than Ordinary Least Squares (OLS) as will be discussed in the next section. The second approach is to use the 1996 SIPP-SER, once again, to impute earnings for those topcoded in 1995 through 1998.[40] Results using both approaches will be shown in section IV.

The implications of the fact that some children will have *both* covered and non-covered earnings 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 intergenerational elasticity coefficient although it will enlarge the standard errors. On the other hand, if this error is correlated with *fathers' earnings*, then the results would 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.[41] In any event, there is no simple way to solve the measurement

[39] Non-covered are defined as those with zero covered earnings but who worked in each month of the year that they are surveyed by the SIPP. Non-workers are classified as those with zero covered earnings who worked between 0 and 2 months. See the Appendix for an explanation of this categorization.
[40] Specifically, the mean value of SIPP earnings of those in the cohort with SIPP earnings above the social security taxable maximum was calculated for 1996 and 1997. There was no significant difference between the imputed values for men and 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.
[41] One way that this could arise is if fathers who have some non-covered earnings, typically have lower total earnings, and also have children who are more likely to have non-covered earnings and hence, lower observed earnings. In this case the bias would be upwards. This can be seen in the following example: If $y_{1i} = y_{1i}^* + t$, where y_{1i}^* is the actual child's earnings and t is the measurement error, then, plim of r "hat" $= r + (\mathrm{Cov}(t, y_{0i})/\mathrm{Var}(y_{0i}))$. If errors are larger in magnitude (more negative) at low values of fathers' earnings, y_{0i}, then $\mathrm{Cov}(t, y_{0i})$ will be positive. While there is evidence that there is a positive intergenerational correlation in self-employment status (Dunn & Holtz-Eakin, 1996) it is not clear that this translates into a sizable correlation in overall non-covered status. In addition, there is no clear evidence

error problem for the dependent variable given the lack of direct survey data on the children in their adult years.

Data Solutions: Parent Earnings

The problems with using SER data are considerably easier to deal with for the parents because of the rich information available in the 1984 SIPP. In addition, it is also possible to use the parents' earnings data directly from the SIPP for the years 1984 and 1985 as an additional check on the results obtained using the SER, keeping in mind of course, the limitations of using just 2 years of data. While this strategy results in smaller samples, they are still significantly higher than those using the PSID and NLS.

The SIPP survey questions are particularly useful with respect to the first problem, that zero earnings might reflect non-covered status. 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 who had zero SER earnings also reported no paid weeks of employment for the full year.[42] For earlier years, a topical module from the second wave on labor force history is used to classify fathers with zero earnings in each year as either non-covered or non-workers.[43] Those classified as non-covered may either be dropped from the analysis or have their earnings imputed using the SIPP earnings.

---

that the distribution of earnings among the self-employed is different from the overall population. A second possibility is that the same form of measurement error exists for both children's earnings and fathers' earnings. This might be the case if both generations' earnings are measured using data from the SER file and if non-covered 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 is used to measure fathers' earnings than when SIPP earnings are used. It turns out the opposite is true as is shown in section IV.

[42] For individuals who are not in the SIPP for the full year of 1984 or 1985, the criteria are modified based on whatever survey information is available in order to classify zero earners.

[43] Specifically, the questionnaire asks individuals a series of questions about recent employment experiences such as tenure and time between jobs, that enables one to construct instances of year long unemployment spells, reasonably well. Because of evidence of poor recall into the distant past, the process is only used to classify non-workers going back to 1979.

The issue of topcoding is far more severe for the fathers since the taxable maximum affected a higher share of the sample in earlier decades. Specifically, for the sample of fathers, the topcode rate was above 50 percent during the early 1970s falling to about 20 percent by the mid-1980s. The approach taken to correct for this is to divide the fathers into 6 groups by race and education level and to impute annual earnings based on information from the full sample of the 1984 SIPP-SER or from the March Supplements to the Current Population Survey (CPS). The procedure is described in greater detail in the Appendix.

The problem of measurement error due to fathers with both covered and non-covered 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 non-covered earnings. 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, as was discussed earlier. One drawback of this approach is that it reduces the sample size by roughly a third. In addition, because there is no information on class of worker for years before the 1984 SIPP, the classification based on 1984 and 1985 must be imposed when averaging fathers' earnings over many years.

IV. Methodology and Main Results

SIPP Results

This study begins by estimating the intergenerational elasticity in earnings between fathers and their children using the SIPP earnings data for fathers. Although the SIPP is limited to just two years of earnings and necessitates a smaller sample, it serves as a useful benchmark for the main analysis that uses the SER data. The econometric approach follows the recent literature and estimates the following equation:

$$(8) \qquad y_{1i} = a + r y_{0i} + b_1 Age_{0i} + b_2 Age^2_{0i} + b_3 Age_{1i} + b_4 Age^2_{1i} + e$$

Specifically, y_{0i}, the father's earnings, will be 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 CPI.[44] 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^2_{0i}, 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. These years are chosen so the kids are no younger than 27 in any of the years that their earnings are measured, thereby giving a more reasonable picture of lifetime earnings.[45] Each year's earnings for the sons and daughters are first adjusted using the procedure described in section III to identify and then impute the earnings of non-covered and non-workers. The children's age measures, Age_{1i} and Age^2_{1i}, use their age in 1998. Table 2 presents the key sample statistics. Unlike some previous

[44] Specifically this is the Bureau of Labor Statistics headline series (BLS code "CUUR0000SA0") for all urban consumers.

[45] Solon (1999) has argued that studies with young samples have found lower correlations because of mean reversion in the transitory income component, i.e. those with higher permanent income have lower transitory incomes at a young age, thereby inducing an attenuation bias. The average age of the kids in this study is 31 which is similar to the average age of 29.6 reported by Solon (1992) and 33.8 reported by Zimmerman (1992). Averages are taken over several years for the children to address the criticism by Couch and Lillard (1998) that Solon and Zimmerman both omit years of zero earnings among the children in their work.

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.[46]

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 set up a tobit model with an individual specific right-censoring point, as follows:

$$(9) \qquad y^*_{1i} = \rho y_{0i} + \beta_1 Age_{0i} + \beta_2 Age^2_{0i} + \beta_3 Age_{1i} + \beta_4 Age^2_{1i} + \varepsilon_i$$

$$(10) \qquad y_{1i} = y^*_{1i} \text{ if } y_{1it} < top_t \ \forall \ t$$

$$(11) \qquad y_{1i} = k_i, \text{ if } y_{1it} \geq top_t \text{ in some } t$$

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.[47] The disturbance term is assumed to be normally distributed and maximum likelihood estimation is used to estimate the intergenerational elasticity.[48] In the case of the daughters, there is likely to be little difference between the OLS and tobit estimates since few women in the sample are censored at the taxable maximum.

In the first set of results, three different sample selection rules are used. First, fathers who do not have positive earnings in both 1984 and 1985 are dropped. This has been the common practice in previous research. Given that there are only two years of earnings, allowing

[46] The effects of restricting the sample to only the oldest child in a family is shown later in the section.
[47] 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. Ideally, one would want 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.
[48] The "intreg" command in STATA is used which allows for a variable censoring point for each observation and for clustered standard errors.

zero earnings in any year is likely to add considerable noise. The other two exclusion rules drop fathers who have earnings below a cutoff point in either year. The cutoffs used are $1000 and $3000 in 1998 dollars.

The results are shown in Table 3. 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 elasticity 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 since only about 2 percent of the daughters are topcoded. The results for the daughters might be biased upwards if the high incidence of non-working among daughters is due to other factors such as child-bearing which in turn, is correlated with parent earnings. 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 intergenerational elasticity is about 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[49]. Couch and Lillard (1998) using selection rules similar to those employed by Zimmerman on the same data, find the elasticity to be 0.37 when using a four-year average. This suggests that simply using the two-year averages from the SIPP gives results similar to those obtained using the PSID and NLS. At a minimum, this adds further confirmation to the argument that the early studies that found elasticities of 0.2 or less did not accurately reflect the degree of earnings mobility in the U.S.

---

[49] 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).

SER Results

The second stage of the study uses the SER earnings data for the fathers. This not only significantly enlarges the sample, since SIPP nonresponse and attrition is eliminated, but it also allows for averaging fathers' earnings over many more years. This longer time period should largely eliminate the problem of attenuation bias stemming from measurement error and transitory fluctuations in earnings. Once again the earnings elasticity 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 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. In all cases the averages are taken over the range of years ending in 1985.

Table 4 presents results using the SER data. There are two broad categories of selection rules on fathers' earnings that are used in 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: non-covered fathers are dropped; non-covered fathers' earnings are imputed; and government and self-employed fathers and non-covered 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, since *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 since 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 non-covered fathers are dropped from the analysis. The fact that many fathers have non-covered 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 non-covered fathers are dropped and earnings are required to be at least $3000 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.

Another finding that is readily apparent from Table 4 is that the estimated elasticity is only slightly lower when the imputed non-covered fathers are added to the sample. In fact, when fathers' earnings are averaged over short time horizons the results are sometimes larger with this adjustment.

The most striking finding is that the elasticity 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 .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 correlation between fathers' earnings and daughters' labor force participation.

Does Excluding Years of Non-Employment Matter?

The estimates in the lower panel of Table 4 also suggest that the results are not sensitive to the inclusion of years of zero earnings. 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 non-covered fathers are dropped, the father-son elasticity estimate falls slightly from 0.445 to 0.434. However, when non-covered fathers are imputed, the coefficient actually *rises*, from 0.376 to 0.403. While the choice of how many years of zero earnings to include is somewhat arbitrary, as long as one positive year of earnings is required, the estimated elasticity is raised substantially from the results that allow for zero earnings in all years.[50] To illustrate this, Appendix Table A2 shows the effects of varying the number of years of zero fathers' earnings that are included, on the father-son intergenerational elasticity. It seems reasonable to conclude that the results are not very sensitive to this variation.[51] Given that children who are not working are also not excluded from the analysis, the criticism by Couch and Lillard (1998) that high estimates of the intergenerational elasticity are based on exclusion rules are not supported by this dataset.

The Effects of Topcoding

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 elasticity coefficient. Imputing the topcoded fathers with the mean level of earnings for those topcoded, ideally, should correct this problem.[52] A simple way to check the robustness of the results of this procedure is to simply drop the topcoded fathers. Table 5 presents the results of this exercise when fathers who are topcoded in *any year* over the relevant time horizon are

[50] In cases where fathers' earnings are zero in *all* years, obviously, the log-log specification is untenable. Following Couch and Lillard (1998), cases with *zero average* earnings are recoded as $1 so that the log would be zero rather than negative infinity. It is not clear that this is a reasonable approach since zeroes on a *log*-scale may significantly alter the results due to the leverage of such observations. In other specifications (not shown) recoding zero earnings in a range from $500 to $3000 (6.2 to 8.0 on the log scale) substantially raises the coefficient compared to what is shown in Table A2.

[51] It should be noted that the results from the last row of Table 4 which average over 10 years and 16 years, include years of zero earnings that are due to *non-covered status.* In these cases, more restrictive rules are used. It was decided that fathers must have positive earnings in about 70 percent of the years in these cases. This was chosen because under this rule, the results for a seven-year average when non-covered zeroes are included is similar to the results when only zeroes due to non-working are allowed.

[52] Of course, this assumes that true statistical model is a linear relationship between fathers' earnings children's earnings, which itself, is the subject of inquiry in Section V.

dropped from the sample. The results are shown for sons and daughters 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 results in an estimate of 0.472 while dropping these observations fathers results in an estimate of 0.439. Averaging fathers' earning using years before 1979 is particularly troublesome because the taxable maximum in real terms was so much lower during that time. As a result so many of the observations are topcoded, and hence, dropped, that it is not clear that the results are meaningful. In fact, the average over 1970 to 1985 has a sample too small to even precisely estimate the coefficient.

The Role of Persistent Transitory Earnings

Given the high estimated elasticity in this sample, a natural question is what explains the difference between the results presented here and the earlier literature? Since estimates taken over shorter time periods match the results from previous studies it does not appear that there is anything especially different about the sample, data or cohort that was chosen. The analysis from section II suggests that it might be the case that short-term fluctuations in earnings are highly persistent and are not adequately “averaged away”, especially when averages are taken over short time periods. The simulation exercise presented in section II suggested that the intergenerational elasticity calculated with a five-year average may be biased down by twenty five to thirty percent under plausible assumptions. Taking this approach a step further, the entire "path" of the attenuation factor can be plotted as the average of fathers earnings are taken over more years. This can then be compared to the empirical results in this study under various assumptions on the true intergenerational elasticity. Figure 3 shows this comparison using assumptions based on results found by Card (1994) and Hyslop (2001) and assuming that the true intergenerational elasticity is 0.6 as the results of Table 4 indicate.[53]

The simulated attenuation bias declines but a slowing rate as more years are used in the averaging process. The results from Table 4, in contrast, show a more linear increase in the estimated coefficient. In fact, the results when fathers' earnings are averaged over 16 years, appear to be somewhat higher than what would be predicted using the simulated model when the true intergenerational elasticity is assumed to be 0.6. A likely explanation is that transitory fluctuations vary over an individual's lifespan, whereas in this simulation they are treated as constant. Gordon (1984) and Baker and Solon (1999) for example, have shown that the transitory variance follows a "U-shape" over an individual's lifetime. If this is indeed the case, then the attenuation *factor*, l^*_t, should be somewhat higher than the simple simulation predicts when the fathers' earnings are averaged using years when their age is closer to forty. The longer term averages may result in higher estimates because they average in years when there is less transitory noise.

In the final analysis, this exercise is merely suggestive. The estimates of the intergenerational elasticity are subject to sampling error and we certainly do not know the "true" parameter values of the statistical model. Analysis of other matched datasets that may become available in thecoming years and which do not suffer from some of the problems present in this study may help to resolve these measurement issues. Still, highly serially correlated transitory earnings along with lifecycle bias appears to be a reasonable explanation for why the results from five-year averages might be so different from that found using a sixteen-year average.

Other Sample Selection Issues

---

[53] The procedure for the simulation is described in detail in the Appendix.

There are some issues related to the construction of the matched dataset that can potentially bias the results. First, children must have been coresident with their parents or living away at college at the beginning of the survey. Second, in order to have been matched, they must also have a social security number that was provided to the interviewer. To handle the first problem, the sample of *all* individuals born between 1963 and 1968 in the 1984 SIPP were divided into 24 groups by year of birth, sex and race. The rate of "living at home" was calculated for each group. The inverse of these rates could then be used to weight the children in the intergenerational samples used in this study. For the problem of matches based on social security numbers, a probit analysis was done to predict the likelihood that individuals from the cohort in the SIPP would be matched to their fathers. The inverse of the predicted probabilities can also be used to weight the father-son pairs in the analysis. Table 6 shows the effects of incorporating these weights on the estimated elasticities using the SIPP-based sample of fathers. The first row simply presents the earlier estimated results from the bottom row of Table 3. The second row weights the observations by the inverse of the probability that they are both living at home and have provided a social security number. The overall elasticity when sons and daughters are pooled is identical at 0.365 but rises slightly for sons and falls slightly for daughters.

Other variations are also attempted in Table 6. Restricting the sample to only 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 to 0.283. The difference is still within the sampling error but might indicate some effect. The result is consistent with the observation by Solon (1999) that using the earnings of sons when they are observed at a younger age can bias the results downwards. It is probably not due to the fact that older kids living at home are more similar to their parents since many of those aged 19 or 20 are actually attending college. The final two rows of Table 6 use different sample selection rules on children. Dropping those children identified as non-covered rather than imputing them has almost no effect for sons but a significant positive effect on daughters. Finally, it might be the case that outliers due to extremely low values of children's average earnings have influenced the parameter results. The approach used to correct for this is possibility is to drop those children who are identified as non-workers in more than two of the four years. This rule appears to have no effect.[54]

[54] There is still some possibility that children with positive, but very low covered earnings, come from families whose fathers, on average, have lower earnings, introducing an upward bias as described in section II. Unfortunately there does not appear to be any way to correct for this possibility with the available data.

V. Further Extensions

Family Income

An interesting finding in some previous intergenerational studies is that family income is more highly correlated across generations than is fathers' earnings.[55] Most of these studies, however, have not discussed this result in much detail.[56] While this study is limited to the use of earnings as an outcome for children it can examine the effects of other measures of parental economic status. The use of family income provides a broader measure that includes not only the mother but also incorporates other forms of *non-earnings* income into the analysis. Although the SER data does not have data on other forms of income, the SIPP is particularly useful because it provides a very detailed breakdown of sources of income that can be used for the parents. Table 7 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.[57] In all cases, only parents whose income measure exceeds $3000 in 1998 dollars in 1984 and 1985 are included. Using fathers' income rather than earnings raises the intergenerational elasticity quite a bit. For sons the estimate increases from 0.349 to 0.518. Using income rather than earnings also appears to raise the elasticity sharply when two parents are used and if only single mothers are examined. Adding mothers to the analysis also appears to raise the elasticity, particularly for daughters. For example, looking at both parents' income instead of just the fathers' income, raises the elasticity with daughters earnings from 0.496 to

---

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

[56] The exception is Mulligan (1997) who argues that this result makes perfect sense in a standard intergenerational permanent income model. In such a model under certain assumptions earnings mobility is dictated by regression to the mean in ability which might be relatively rapid. Income mobility, however, might be much slower because of financial asset transfers from parents to children irrespective of investment in children's human capital. That analysis, however, does not explain differences between the effects of parents' earnings as compared to parents' income on children's *earnings.*

[57] 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.

0.708. The comparable increase for sons is from 0.518 to 0.553. Looking at single mothers only, the estimated elasticities are dramatically lower, and in most cases, statistically insignificant. This is no doubt due to poor classification of families and therefore, significant mis-measurement.

What might explain the higher results from parental income? For one thing, income may be a less noisy measure of economic status than earnings. This is likely to be particularly true at the low end of the parents' earnings distribution where individuals may receive income at times when they receive virtually no earnings due to unemployment, e.g. unemployment insurance or workers compensation. This may result in a higher estimated elasticity when parents' income is used rather than earnings because of a smaller attenuation bias due to measurement error or transitory shocks. In addition, there appears to be a sample selection effect. If the intergenerational elasticity 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 elasticity with fathers earnings in row 1 is also used to estimate the elasticity with fathers' income, then the latter estimate falls from 0.518 to 0.385 (not shown). In any case, it appears that using income rather than earnings for parents may give a more accurate reading of intergenerational mobility, especially when only a few years of parents earnings data are available.

Borrowing Constraints

Theoretical models of intergenerational mobility have emphasized borrowing constraints as a key factor in the transmission of earnings inequality. Becker and Tomes (1986) and Mulligan (1997) have argued that if parents can borrow from their children's future earnings, then all parents will invest the optimal amount in their children's human capital. If earnings are determined by human capital, and human capital is a function of ability, then the intergenerational elasticity in earnings will only be positive if earnings and ability are correlated and will depend on the rate at which ability regresses to the mean. With borrowing constraints, however, parents with low income and able children will not invest the optimal amount in their children's education inducing a higher intergenerational elasticity in earnings. Mulligan (1997) has attempted to test this hypothesis using the PSID and by splitting the sample by those who expect to receive an inheritance. He found no significant difference in elasticities between the two groups. 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 intervivos transfers. The model focuses solely on an intergenerational budget constraint and does not analyze parents' potential inability to borrow from *their own* future income.

There are several advantages that this dataset can bring to this question. First, with a larger sample it is possible to split the sample along some dimension that directly reflects fathers' ability to access capital, and still estimate the parameters reasonably well.[58] Second, the topical module from wave 4 of the 1984 SIPP can be used to gather more detailed information on household balance sheets to more accurately classify families by their ability to invest fully in their children's human capital. It was decided to use net worth to classify fathers as either borrowing constrained or not borrowing constrained. This measure captures the ability of individuals to borrow against their current wealth or to draw down assets in order to finance human capital acquisition. One problem with this approach, of course, is that the measure is from 1984, when kids are aged 16 to 21 while the relevant period to measure borrowing constraints is arguably at an earlier point in the child's educational career. In addition, since net worth and income are highly correlated, any nonconstancies in the intergenerational income elasticity may also be reflected in differences in r by levels of net worth that may or may not be due to borrowing constraints.

[58] Another approach is to include nonlinearities in fathers' earnings. Experimentation with this approach did not yield any statistically significant results.

Table 8 shows the results of this exercise. First, using the SIPP for parents' earnings, and dividing the sample by the median level of net worth (about $65,000 in 1984 dollars) the results point to a sharp difference between those below the median and those above. The elasticity is 0.422 for those with lower than median net worth but only 0.140 for those above the median level. While the difference is large, one could not reject the null hypothesis of equality at the 5 percent 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. In fact for the top quartile, there appears to be zero elasticity. Indeed, the permanent income model would predict this result if income is uncorrelated with ability. Similar attempts were less conclusive using SER data for fathers' earnings as the bottom half of Table 8 shows. A possible explanation for this result is that the high topcoding rate of fathers compresses the fathers earnings distribution and given the strong correlation between net worth and earnings, the full variation in the intergenerational elasticity is also compressed.

Differences by Race

One of the key comparisons that has not been explored in previous studies is whether there are significant differences in mobility between blacks and whites. A higher elasticity among blacks might suggest that even if overall mobility is high, economic progress for blacks might be more difficult for other reasons such as borrowing constraints, neighborhood effects or discrimination. Once again, obtaining reasonable sample sizes for such a comparison has been virtually impossible in previous datasets. Table 9 shows the difference in estimates for blacks and whites. Using the seven-year average of fathers' earnings from the SER, the elasticity among blacks (0.487) was found to be nearly twice as high as the elasticity among whites (0.271) but the difference was not significantly greater than zero at the five percent level.[59] In order to keep the sample size as large as possible, the SER results imputed non-covered fathers and used all fathers with positive earnings in *any* year.[60] Additional results were attempted using the SIPP for parent earnings. The comparison of fathers' earnings by race yields a very similar result to what was found using the SER. The difference in elasticities is estimated at 0.222 which is nearly identical to the 0.216 obtained using the SER sample, but in this case the smaller sample leads to a far less precise estimate. The comparison of fathers' income elasticities leads to a larger difference, though it is still estimated imprecisely. Looking at combined two parent earnings and income, however, leads to incredibly large estimates for blacks that exceed 1. If taken seriously, it implies no regression to the mean. The difference in estimates when using two parents is on the border of significance at the five percent level. The results are similar, though less precise, when the samples include only low net worth families (not shown) suggesting that the racial difference is not simply due to borrowing constraints.

One difficulty in these comparisons lies in family composition. A much higher percentage of black families are headed by single mothers where the estimated elasticities are substantially lower (see Table 7). The small sample size of single mother families, however, does not permit a breakdown by race. While further research is clearly needed, the results presented here are suggestive of less mobility among blacks. Some plausible explanations for the higher persistence might lie in employment discrimination, borrowing constraints, neighborhood effects, inferior schools or disparities in home ownership.

---

[59] The seven-year average was used because that is the longest average over which there is still a classification of social security coverage status among fathers. This allows inclusion of zero earning years that reflect non-employment but not non-coverage.

[60] Using more restrictive exclusion rules raises the estimated correlation for whites slightly and lowers the correlation for blacks slightly. The difference remains large but insignificant.

VI. Conclusion

The 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 dataset 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 father 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 highly persistent. Short-term proxies for permanent income may also be susceptible to lifecycle bias due to the fact that the variance of the transitory component of earnings varies considerably by age. Overall, the results point toward an intergenerational elasticity of about 0.6. If accurate, this suggests that many well-documented wage gaps may persist for several generations.

The results appear to be fairly robust to sample selection rules, the match process, and to 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 and lifecycle bias and have significantly smaller samples. What may be required in the future is access to other administrative datasets that overcome these data problems.

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 0.5 to 0.6 range.

While this study provides new descriptive evidence of the extent of mobility in the U.S. 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? While far from conclusive, new evidence is provided suggesting that intergenerational inequality may be related, in part, to access to capital. This offers a potential avenue by which greater mobility may be fostered through public policy.

Some suggestive evidence also points to less mobility among blacks, a minority group that has struggled to achieve full economic parity many decades after the end of slavery. This suggests that the black-white wage gap may take considerably longer to equalize than discrepancies among other groups.

References

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

Atkinson, A.B., A.K. Maynard and C.G. Trinder (1983), Parents and Children: Incomes in Two Generations (Heinemann, London).

Baker, Michael and Gary Solon (1999), "Earnings Dynamics and Inequality Among Canadian Men, 1976-1992: Evidence From Longitudinal Tax Records," NBER Working Paper 5622, National Bureau of Economic Research.

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

Becker, Gary S. and Nigel Tomes (1986), "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics*, 4:S1-S39.

Becker, Gary S. (1988), "Family Economics and Macro Behavior," *American Economic Review*, 78:1-13

Bowles, Samuel (1972). "Schooling and Inequality from Generation to Generation." Journal of Political Economy 80: S219-251.

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

Card, David (1994), "Intertemporal Labor Supply: An Assessment", in Christopher A. Sims (ed.) *Advances in Econometrics, Sixth World Congress*, Vol. 2, Cambridge University Press. Cambridge.

Chay, Kenneth Y. (1995), "Evaluating the Impact of the 1964 Civil Rights Act on the Economic Status of Black Men Using Censored Longitudinal Earnings Data," October 1995, mimeo.

Chay, Kenneth Y. and Bo E. Honoré (1998), "Estimation of Semiparametric Censored Regression Models: An Application to Changes in Black-White Earnings Inequality During the 1960s," *Journal of Human Resources* 33(1):4-38.

Corak, Miles and Andrew Heisz. (1999). "The Intergenerational Earnings and Income Mobility of Canadian men: Evidence from Longitudinal Income Tax Data." *Journal of Human Resources* 34(3):504-533.

Couch, Kenneth A. and Dean R. Lillard (1998). "Sample Selection Rules and the Intergenerational Correlation of Earnings." *Labour Economics* 5: 313-329.

Dearden, Lorraine, Stephen Machin, and Howard Reed (1997), "Intergenerational Mobility in Britain," *Economic Journal*, 107:47-66.

Duncan, Greg, Johanne Boisjoly and Timothy Smeeding (1996), "Economic Mobility of Young Workers in the 1970s and 1980s," *Demography*, 33:497-509.

Dunn, Thomas and Douglas Holtz-Eakin (1996), "Financial Capital, Human Capital and the Transition to Self-Employment: Evidence from Intergeneration Links", NBER Working Paper 5622, National Bureau of Economic Research

Goldberger, Arthur S. (1989), "Economic and Mechanical Models of Intergenerational Transmission", *American Economic Review*, 79:504-513.

Gordon, Roger H.. "Differences in Earnings and Ability," Garland, New York.

Hyslop, Dean (2001), "Rising U.S. Earnings Inequality and Family Labor Supply: The Covariance Structure of Intrafamily Earnings," *American Economic Review,* 91:755-777.

Krueger, Alan (1995), "The Legacy of Separate and Unequal Schooling" Paper presented at conference of Southern Economic Association, November 18-20, New Orleans, LA.

Levine, David I. (2000). "Choosing the Right Parents: Changes in the Intergenerational Transmission of Inequality Between the 1970s and the Early 1990s", manuscript, University of California, Berkeley.

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

Mazumder, Bhashkar (2001), "Earnings Mobility in the U.S.,: A New Look at Intergenerational Inequality", PhD. Dissertation, University of California, Berkeley.

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

Minicozzi, Alexandra L. (1997), "Nonparametric analysis of intergenerational income mobility", PhD dissertation (University of Wisconsin).

Mulligan, Casey B. (1997), Parental Priorities and Economic Inequality. University of Chicago Press, Chicago.

Peters, Elizabeth H. (1992). "Patterns of Intergenerational Mobility in Income and Earnings," *Review of Economics and Statistics*, 74:456-466.

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

Smith James P. and Finis R. Welch (1986), "Closing the Gap: Forty Years of Economic Progress for Blacks.", The Rand Corporation, Santa Monica CA.

Solon, Gary (1989), "Biases in the Estimation of Intergenerational Earnings Correlations" *Review of Economics and Statistics*, 71:172-174

Solon, Gary (1992), "Intergenerational Income Mobility in the United States," *American Economic Review*, 82:393-408

Solon, Gary (1994) "Comments on 'Sample Selection Rules and the Intergenerational Correlation of Earnings: A Comment on Solon and Zimmerman'". Unpublished manuscript..

Solon, Gary (1999), "Intergenerational Mobility in the Labor Market," *Handbook of Labor Economics,* Elsevier

Zimmerman, David J. (1992), "Regression Toward Mediocrity in Economic Stature," *American Economic Review*, 82:409-429

Table 1: Simulation Results on Attenuation Bias when Using Multiyear Averages

Attenuation Coefficient if….

Number of Years Averaged	$d = 0.5$	$d = 0.8$	$d = 0.5$	$d = 0.8$	$d = 0.5$	$d = 0.8$
1	0.641	0.670	0.519	**0.554**	0.526	0.572
2	0.733	0.735	0.630	**0.637**	0.619	0.629
3	0.783	0.767	0.693	**0.680**	0.677	0.662
4	0.817	0.790	0.737	**0.710**	0.720	0.687
5	0.843	0.808	0.772	**0.734**	0.754	0.709
6	0.863	0.823	0.799	**0.754**	0.782	0.728
7	0.879	0.837	0.821	**0.772**	0.806	0.746
8	0.892	0.849	0.840	**0.788**	0.826	0.762
9	0.904	0.859	0.856	**0.802**	0.843	0.777
10	0.913	0.869	0.869	**0.815**	0.857	0.792
11	0.921	0.878	0.881	**0.827**	0.870	0.805
12	0.928	0.887	0.891	**0.839**	0.882	0.817
13	0.935	0.895	0.900	**0.849**	0.892	0.829
14	0.940	0.902	0.908	**0.859**	0.900	0.840
15	0.945	0.908	0.915	**0.868**	0.908	0.850
16	0.949	0.915	0.922	**0.877**	0.916	0.860
17	0.953	0.921	0.927	**0.885**	0.922	0.869
18	0.957	0.926	0.933	**0.892**	0.928	0.877
19	0.960	0.931	0.937	**0.899**	0.934	0.886
20	0.963	0.936	0.942	**0.906**	0.939	0.893
21	0.966	0.940	0.946	**0.912**	0.943	0.901
22	0.968	0.945	0.950	**0.918**	0.947	0.908
23	0.970	0.948	0.953	**0.924**	0.951	0.914
24	0.973	0.952	0.956	**0.929**	0.955	0.920
25	0.974	0.956	0.959	**0.934**	0.958	0.926
26	0.976	0.959	0.962	**0.939**	0.961	0.932
27	0.978	0.962	0.964	**0.943**	0.964	0.937
28	0.980	0.965	0.967	**0.947**	0.967	0.942
29	0.981	0.968	0.969	**0.952**	0.970	0.947
30	0.982	0.971	0.971	**0.955**	0.972	0.952

Note: Simulation is based on equation 14 (See appendix). In the first pair of columns, the share of single year variance in earnings accounted for by permanent factors is 0.7. In the last two pairs of columns the share is assumed to be 0.5. Within each pair of columns, assumptions are made about the share of transitory variance in the variance of a single year of earnings and the auto correlation coefficient. The assumptions based on Hyslop (2001) are shown in bold.

Table 2: Summary Statistics for Fathers and Children

Samples using 1984 SIPP for fathers' earnings

Variable	*N*	*Mean*	*S.D.*	*Minimum*	*Maximum*
Father's Age in 1984	796	46.9	6.2	28	71
Log Average Father's Earnings 84-85	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 95-98	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 95-98	719	9.1	1.7	4.1	11.1

Samples using SER for fathers' earnings

Variable	*N*	*Mean*	*S.D.*	*Minimum*	*Maximum*
Father's Age in 1984	1262	47.1	6.3	27	69
Log Average Father's Earnings 84-85	1262	10.5	0.9	4.0	11.5
Log Average Father's Earnings 82-85	1218	10.6	0.7	6.5	11.5
Log Average Father's Earnings 79-85	1160	10.7	0.6	7.3	11.5
Log Average Father's Earnings 76-85	1111	10.7	0.5	7.7	11.3
Log Average Father's Earnings 70-85	1063	10.7	0.4	8.1	11.2
Son's Age in 1998	1262	32.4	1.7	30	35
Log Average Son's Earnings 95-98	1262	10.0	1.2	2.5	11.1
Daughter's Age in 1998	1178	32.5	1.7	30	35
Log Average Daughter's Earnings 95-98	1178	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 non-covered or non-workers as described in text. The SIPP sample pertains to those shown in row 1 of Table 2. 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 3. 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.

Table 3: Intergenerational Elasticities Using SIPP for Fathers' Earnings

elast. *(s.e.)* *N*	***Dependent Variable is Log Avg. Earnings, 1995-1998***					
	Sons		**Daughters**		**Pooled**	
Fathers	Tobit	OLS	Tobit	OLS	Tobit	OLS
Log Avg. 84-85	0.384	0.342	0.360	0.341	0.357	0.322
Father Earnings >0	(0.091)	(0.085)	(0.106)	(0.103)	(0.074)	(0.070)
each year	796	796	719	719	1515	1515
Log Avg. 84-85	0.337	0.293	0.367	0.346	0.339	0.300
Father Earnings >1000	(0.080)	(0.072)	(0.117)	(0.112)	(0.074)	(0.069)
each year	788	788	713	713	1501	1501
Log Avg. 84-85	0.349	0.292	0.361	0.337	0.365	0.315
Father Earnings >3000	(0.078)	(0.070)	(0.128)	(0.122)	(0.080)	(0.074)
each year	767	767	702	702	1469	1469

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 non-worker 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.

Table 4: Intergenerational Elasticities Using SER for Fathers' Earnings

elast. *(s.e.)* *N*	***Dependent Variable is Children's Log Avg Earnings, 1995-1998*** All results use tobit specification														
Fathers	**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	0.312	0.385	0.472	0.570	0.624
Non-Covered	(0.043)	(0.059)	(0.079)	(0.099)	(0.096)	(0.065)	(0.087)	(0.110)	(0.140)	(0.159)	(0.041)	(0.056)	(0.071)	(0.088)	(0.099)
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	--	--	0.323	0.358	0.406	--	--
Non-Covered	(0.050)	(0.052)	(0.062)			(0.062)	(0.081)	(0.096)			(0.041)	(0.051)	(0.059)		
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	0.394	0.487	0.557	0.659	0.727
Government &	(0.060)	(0.082)	(0.096)	(0.111)	(0.135)	(0.089)	(0.137)	(0.150)	(0.173)	(0.192)	(0.062)	(0.084)	(0.094)	(0.109)	(0.128)
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	--	--	0.264	0.372	0.474	--	--
Non-Covered	(0.043)	(0.057)	(0.069)			(0.060)	(0.065)	(0.091)			(0.037)	(0.046)	(0.059)		
Fathers	1295	1268	1227			1201	1168	1127			2496	2436	2354		
Impute	0.238	0.342	0.403	--	--	0.295	0.384	0.474	--	--	0.260	0.357	0.438	--	--
Non-Covered	(0.042)	(0.057)	(0.059)			(0.055)	(0.061)	(0.080)			(0.035)	(0.044)	(0.052)		
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	0.294	0.417	0.519	0.626	0.704
Government &	(0.059)	(0.080)	(0.084)	(0.101)	(0.109)	(0.084)	(0.083)	(0.113)	(0.130)	(0.145)	(0.051)	(0.064)	(0.072)	(0.086)	(0.094)
Self-Employed	874	869	862	895	917	803	794	785	825	831	1677	1663	1647	1720	1748

Note: See text for how children's and fathers' earnings are constructed. Standard errors are adjusted for multiple siblings. *Required years of pos. earnings are: 1 for 2-yr. averages; 2 for 4-yr. averages; 3 for 7-yr. averages; 7 for 10 yr.-averages and 11 for 16-yr. averages.

Table 5: Effects of Top-Coded Fathers on Intergenerational Elasticities

elast. *(s.e.)* *N*	***Dependent Variable is Children's Log Avg Earnings, 1995-1998*** All results use tobit specification				
Fathers			**Pooled (Sons & Daughters)**		
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 dads	1713	1530	1144	784	343

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 non-worker status and were imputed accordingly. For fathers, SER earnings for those identified as non-covered are dropped. In row 1, earnings for those topcoded are imputed using March CPS data for 1970-80 and using 1984 SIPP for 1981 to 1984. In row 2 fathers topcoded in any year over the relevant period are dropped.

Table 6: The Effects of Sample Selection Using the SIPP for Fathers' Earnings

elast. *(s.e.)* *N*	***Dependent Variable*** ***is Log Avg. Earnings, 1995-1998*** All results use tobit specification		
Fathers	**Sons**	**Daughters**	**Pooled**
	0.349	0.361	0.365
Log Avg. 84-85	(0.078)	(0.128)	(0.080)
Father Earnings >3000 each year	767	702	1469
Weighted for	0.375	0.339	0.365
Match Likelihood	(0.086)	(0.128)	(0.084)
& Prob Living at home	767	702	1469
Eldest Kids Only	0.386	0.357	0.358
	(0.095)	(0.147)	(0.092)
	548	506	1054
Aged 15 to 18 only	0.283	0.400	0.367
	(0.085)	(0.155)	(0.095)
	542	486	1028
Non-Covered	0.362	0.473	0.409
Children are Dropped	(0.094)	(0.113)	(0.074)
	644	498	1142
*Require 2 years of Positive**	0.358	0.363	0.369
Children's Earnings	(0.080)	(0.130)	(0.082)
	736	687	1423

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 non-worker staus 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.
*Really, this means children cannot be classified as non-workers in more than two years.

Table 7: Intergenerational Elasticity of Parents' Income on Children's Earnings

elast. *(s.e.)* *N*	***Dependent Variable*** ***is Log Avg. Earnings, 1995-1998*** All results use tobit specification		
	Sons	**Daughters**	**Pooled**
Father Earnings	0.349	0.361	0.365
Log Avg. 84-85	(0.078)	(0.128)	(0.080)
	767	702	1469
Father Income	0.518	0.496	0.499
Log Avg. 84-85	(0.102)	(0.119)	(0.088)
	871	773	1644
Two Parent Earnings	0.385	0.491	0.444
Log Avg. 84-85	(0.075)	(0.118)	(0.073)
	776	719	1495
Two Parent Income	0.553	0.708	0.635
Log Avg. 84-85	(0.103)	(0.118)	(0.086)
	842	768	1610
Single Mother Earnings	0.215	0.357	0.239
Log Avg. 84-85	(0.170)	(0.306)	(0.178)
	161	145	306
Single Mother Income	0.362	0.287	0.320
Log Avg. 84-85	(0.151)	(0.183)	(0.123)
	231	219	450
All Family Earnings	0.322	0.502	0.406
Log Avg. 84-85	(0.060)	(0.098)	(0.058)
	959	879	1838
All Family Income	0.478	0.558	0.523
Log Avg. 84-85	(0.067)	(0.080)	(0.056)
	1105	1006	2111

Note: Probit models based on the 1996 SIPP matched to SER were used to determine if children's zero earnings reflected noncoverage or non-worker 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 $3000 in 1998 dollars in 1984 and 1985. Standard errors are adjusted for within family correlation when more than one sibling is present.

Table 8: Intergenerational Elasticity by Level of Net Worth

elast. *(s.e.)* *N*	***Dependent Variable*** ***is Log Avg. Earnings, 1995-1998*** All results use tobit specification				
	Pooled (Sons and Daughters)				
	Overall	***High Net Worth***	***Low Net Worth***	***Diff.***	***t-stat***
	SIPP Results				
Father Earnings	0.358	0.146	0.412	0.265	1.729
Log Avg. 84-85	(0.074)	(0.108)	(0.109)	(0.153)	
Low is <=median	1514	757	757		
High is >median					
Father Earnings		-0.022	0.450	0.472	2.414
Log Avg. 84-85		(0.140)	(0.136)	(0.195)	
Low is <=25th percentile		374	379		
High is >=75th percentile					
	SER Results				
Father Earnings	0.482	0.286	0.467	0.181	1.212
Log Avg. 79-85	(0.072)	(0.114)	(0.097)	(0.149)	
Low is <=median	2186	1111	1075		
High is >median					
Father Earnings		0.193	0.471	0.278	1.467
Log Avg. 79-85		(0.124)	(0.143)	(0.189)	
Low is <=25th percentile		559	532		
High is >=75th percentile					

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 non-worker 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 succesfully matched to their wave 4 questionnaire are kept in the sample. Standard errors are adjusted for within family correlation when more than one sibling is present.

Table 9: Intergenerational Elasticity by Race

elast. *(s.e.)* *N*	***Dependent Variable*** ***is Log Avg. Earnings, 1995-1998*** All results use tobit specification **Pooled (Sons and Daughters)**				
	Overall	***White***	***Black***	***Diff.***	***t-stat***
			SER Results		
Father Earnings	0.328	0.271	0.487	0.216	1.500
Log Avg. 79-85	(0.046)	(0.048)	(0.136)	(0.144)	
	3077	2726	255		
			SIPP Results		
Father Earnings	0.357	0.312	0.534	0.222	0.605
Log Avg. 84-85	(0.074)	(0.074)	(0.359)	(0.367)	
	1515	1362	108		
Father Income	0.364	0.254	0.620	0.366	1.265
Log Avg. 84-85	(0.084)	(0.082)	(0.277)	(0.289)	
	1690	1498	134		
Two Parent Earnings	0.464	0.343	1.013	0.670	1.942
Log Avg. 84-85	(0.082)	(0.066)	(0.339)	(0.345)	
	1573	1409	117		
Two Parent Income	0.518	0.381	1.109	0.728	1.713
Log Avg. 84-85	(0.109)	(0.100)	(0.413)	(0.425)	
	1674	1488	130		

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 non-worker status and were imputed accordingly. For SER results, fathers must have positive earnings in at least one year and fathers who are classified as non-covered are imputed. 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. They must also have positive earnings in each year. Standard errors are adjusted for within family correlation when more than one sibling is present.

Figure 1: Percent of Sample Topcoded

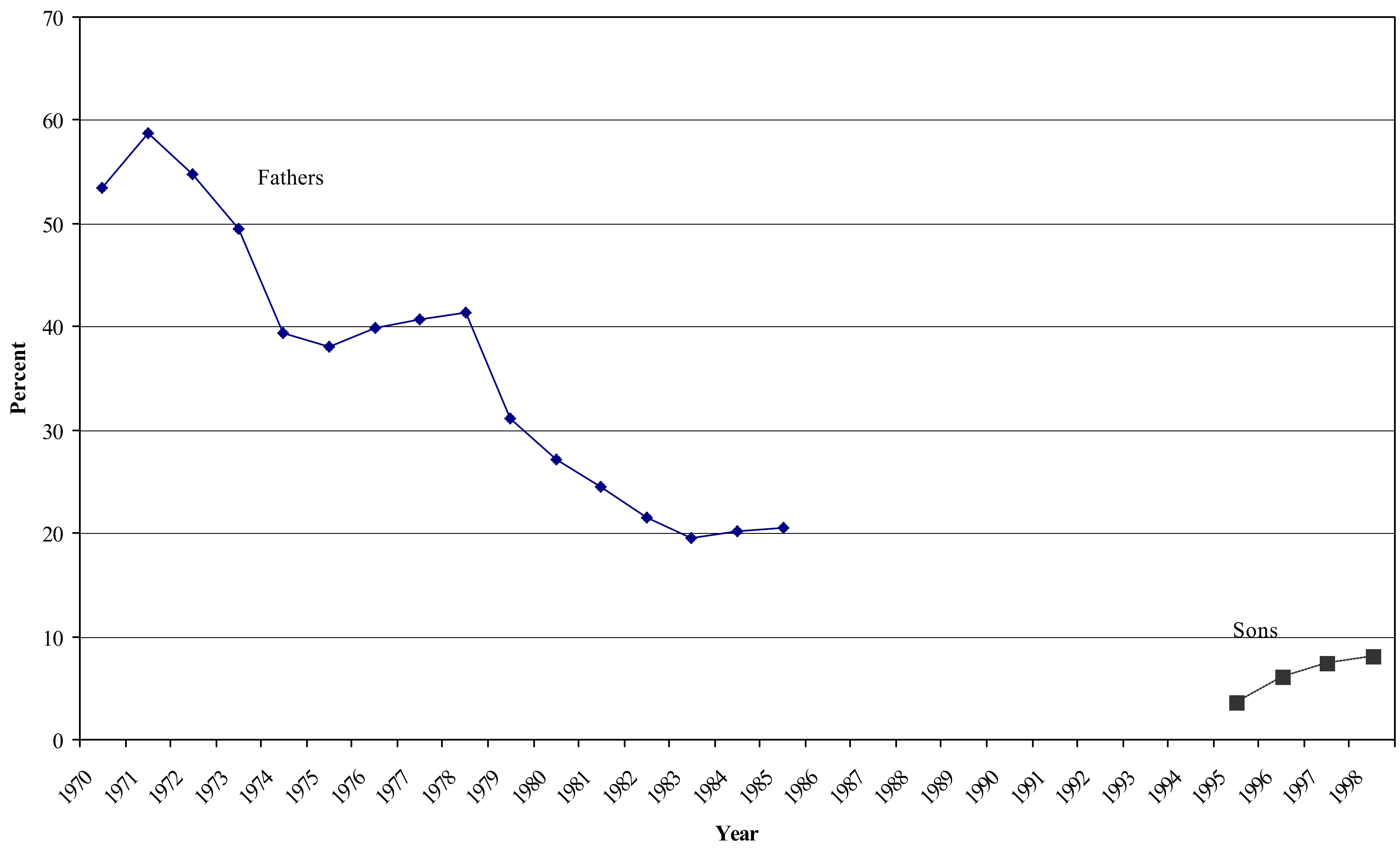


Figure 2: 1984 SIPP vs. SER Comparison

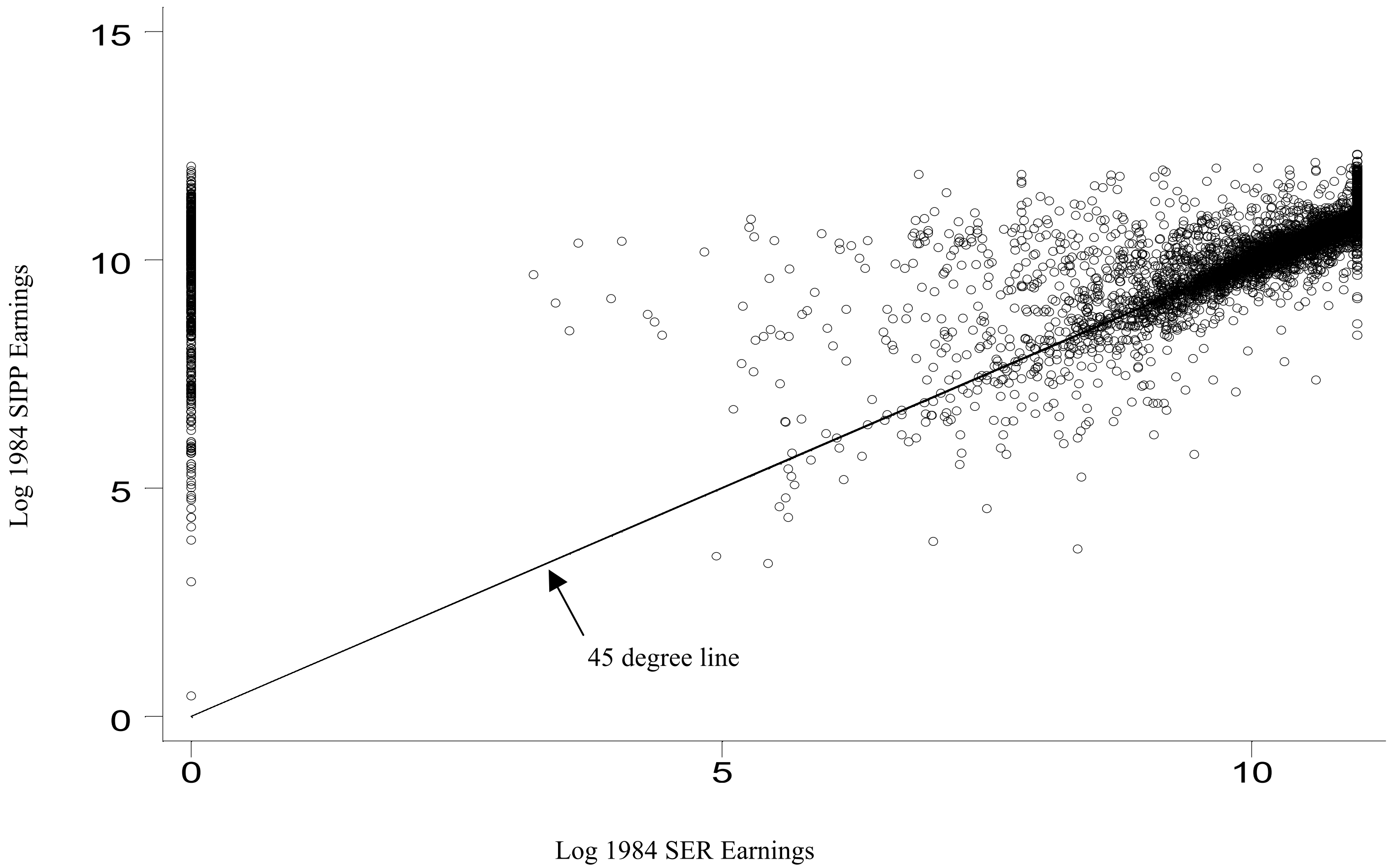
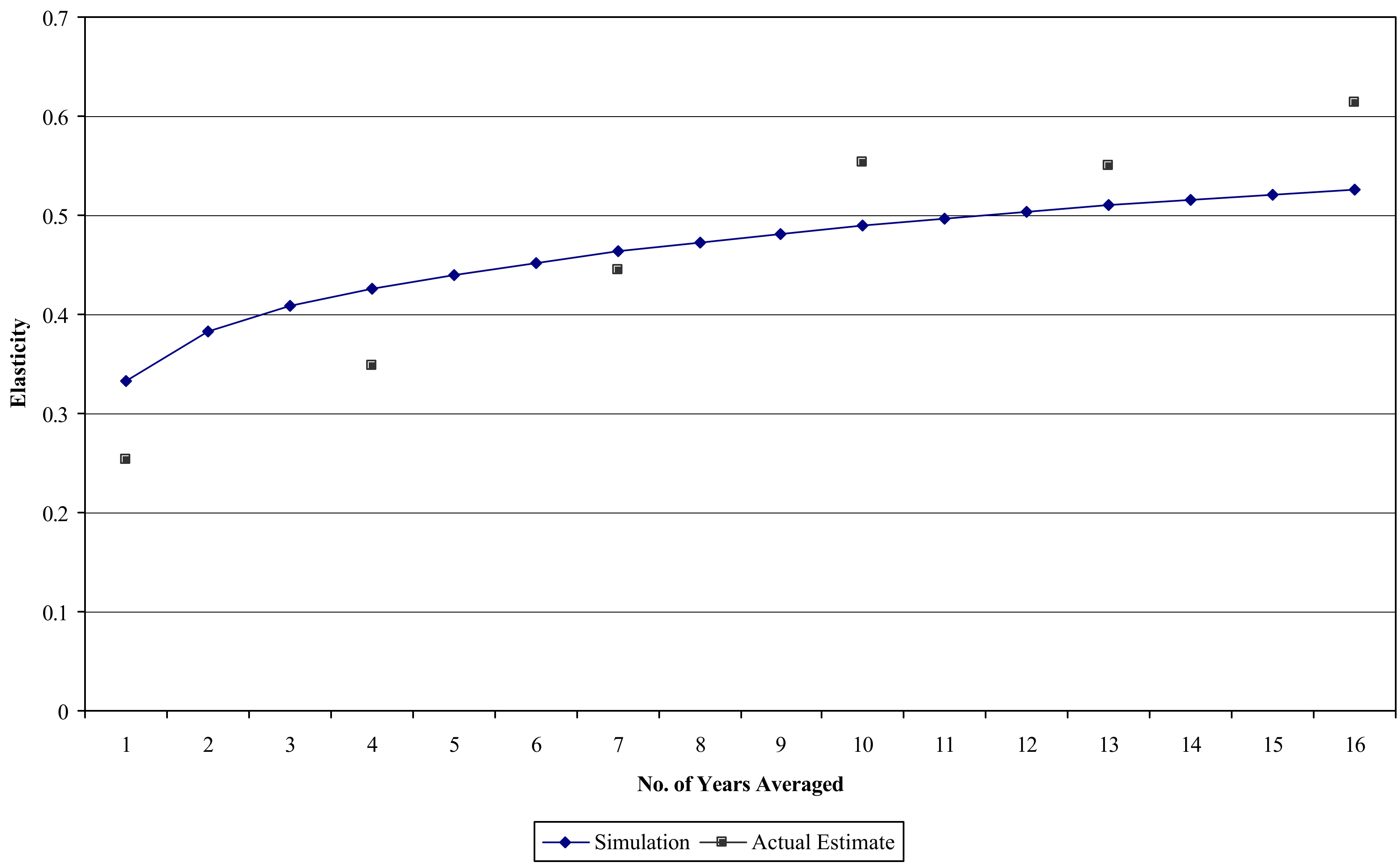


Figure 3: Simulation and Actual Estimates from Averaging Fathers' Earnings



Appendix

Description of simulation to calculate attenutation coefficients shown in Table 1

In order to account for the fact that transitory shocks that persist over the course of an individual's working life are effectively "permanent", a formula analogous to (7) is derived for calculating attenuation coefficients when using multi-period averages of fathers' earnings. As before, the earnings process for fathers and sons is assumed to follow equations (2) and (3). It is now assumed, however, that instead of (4) the equation of interest is:

$$y_{1i} = r\bar{y}_{45,i} + e \tag{12}$$

where $\bar{y}_{45,i}$ is the average of fathers' earnings over the 45 years of his working life. Let *t* index the years from 1 to 45. If a *T* year average of earnings beginning in year *s,* $\bar{y}_{T,s}$, is used as a proxy for $\bar{y}_{45,i}$, the attenuation factor $l^*_{T,s}$ is the following:

$$l^*_{T,s} = \frac{\operatorname{cov}(\bar{y}_{T,s}, \bar{y}_{45})}{\operatorname{var}(\bar{y}_{T,s})} \tag{13}$$

While the denominator of this expression will be exactly the same as in (7), the numerator is more complicated. The covariance between any multiyear average of fathers' earnings and the entire lifetime average of earnings will depend not only the number of years that are averaged but also on exactly *which* years are used in the average.[1] Equation (14) provides the exact formula for calculating $l^*_{T,s}$.

$$l^*_{T,s} = \frac{s^2_{y0} + \frac{1}{T}\sum_{s}^{s+T-1}\frac{1}{45}\sum_{t=1}^{45} d^{|t-s|} s^2_w}{s^2_{y0} + \frac{1}{T} a s^2_w + \frac{1}{T} s^2_v} \tag{14}$$

---

[1] The autoregressive structure implies that the correlations between the transitory components will depend on the distance in time between the years used for the short-term average and the full 45 year average. For example, a five-year average taken at the very beginning or end of one's life will be less correlated with lifetime earnings than a five-year average taken during the middle of one's life.

where, as before, $a = 1 + 2d\left\{\dfrac{T - \left[\dfrac{(1-d^{T})}{(1-d)}\right]}{T(1-d)}\right\}$

Dividing the numerator and denominator by the variance in single year earnings, s^2_{yt} and then using estimates for d and the share of the variance of single year earnings accounted for by the permanent component, transitory component and measurement error, enables one to calculate $l^*_{T,s}$ for all possible values of T and s. In order to get a summary measure of the degree of attenuation bias that is only a function of T, we can simply average the $l^*_{T,s}$ over all possible s for a given value of T. Table 1 presents the values of l^*_T as averages are taken over progressively more years using three different sets of assumptions on the parameter values.

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 non-workers or non-covered. In this sub-sample, about 57 percent of the men with zero SER earnings were employed for the full year and are classified as non-covered while 32 percent worked for only zero to two months of the year and are called non-workers.[2] Those working in the non-covered sector are primarily government workers or self-employed. The comparable rates for the daughters were 21 percent and 71 percent, respectively. These numbers suggest two conclusions. First, most of those with zero SER earnings are either non-workers or full-time workers in the non-covered sector. Only about 10 percent of zero earners fall in the gray area of having zero earnings and working part-year. Second, the problem of non-covered workers is particularly important for men.

[2] 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.

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.[3] 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 non-covered. 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 were estimated separately for men and women, and for both 1996 and 1997.[4] The estimates from the probit models were then combined in order to classify each son or daughter as either a non-worker or as non-covered for each year.[5] Those identified as non-covered are then either dropped from the analysis or their earnings are imputed using the mean level of log earnings for the group from the 1996 SIPP. Similarly, those identified as non-employed may be assigned the mean level of log earnings for those who worked between zero to two months.[6]

The results of the two probit models for men in 1996 are shown in Appendix Table A1.[7] 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

---

[3] For the most part, survey information from the 1996 SIPP is deliberately omitted from this analysis since 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.
[4] These were the only years from the 1996 SIPP for which annual earnings were available at the time the research was permitted.
[5] Specifically, individuals are classified based on the category in which they have a higher predicted probability. This is equivalent to assigning them based on 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.
[6] This strategy allows those children with zero SER earnings in all four years not to be entirely dropped from the analysis
[7] Results for women and for 1997 are available on request.

years; a flag indicating an active earnings discrepancy;[8] 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 percent of the true classifications of non-covered and non-workers 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 percent of the sample was incorrectly classified. The error in forecasting women's status is higher and implies that about 3 percent of the sample is incorrectly classified. While 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 a high level of confidence in the results.

Procedure for handling topcoding among fathers

Fathers with topcoded earnings are divided into six race-education cells: by white or 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 dataset 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.[9]

For the years 1970 to 1980, the imputation values are derived from each year's March Current Population Survey (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

---

[8] These are cases where an individual has contested what they believe to be inaccurate reports of their earnings with SSA and where the dispute has not yet been resolved.

[9] 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 12 months of 1984 are used in the calculations. For the years 1981 to 1983, and 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 percent topcoded in that year based on the SER data for all the sample members in the 1984 SIPP-SER dataset (not just the fathers). For example, in 1980, 8.8 percent of those with positive earnings in the full sample of the 1984 SIPP-SER matched dataset, had topcoded earnings. The strategy then, was to use the top 8.8 percent of the SIPP earnings distribution in 1984 to calculate the imputed values for each of the 6 groups for 1980. Of course, the 1984 dollar values were then converted to 1980 dollars using the CPI.

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.[10]

[10] 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 no different using this strategy.

Table A1: Probit Results on Predicting Non-Covered vs. Non-Worker, 1996 SIPP-SER

Men in 1996	*Dependent Variable*				
	Non-Covered		**Non-Worker**		
Variable	*dF/dx times 100*	*z-stat*	*dF/dx times 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	1254.62
earnings 1984	-0.01	-1.68	0.01	1.68	1703.05
earnings 1985	-0.01	-1.78	0.01	1.94	2318.21
earnings 1986	-0.02	-2.12	0.01	2.2	3208.63
earnings 1987	-0.02	-2.62	0.01	2.71	3981.88
earnings 1988	-0.02	-2.36	0.01	2.15	4162.30
earnings 1989	-0.01	-1.9	0.01	1.92	5328.86
earnings 1990	-0.01	-1.97	0.01	2.33	6030.20
earnings 1991	-0.02	-2.76	0.01	2.54	5964.49
earnings 1992	-0.02	-2.2	0.01	2.45	5756.44
earnings 1993	-0.02	-2.11	0.01	2.05	5153.96
earnings 1994	-0.02	-2.24	0.01	2.35	3450.91
earnings 1995	-0.01	-1.31	0.01	1.72	2099.37
earnings 1997	-0.02	-2.35	0.01	2.15	2532.74
earnings 1998	-0.02	-2.32	0.01	2.46	5006.78
earnings discrepancy flag*	-90.26	-5.97	59.04	6.49	0.29
military*	6.79	0.38	0.94	0.09	0.07
age29*	29.15	2.42	-8.75	-1.19	0.15
age30*	0.28	0.02	-7.82	-1.06	0.20
age31*	15.58	1.05	-7.49	-0.95	0.16
age32*	5.09	0.3	6.69	0.67	0.19
age33*	-28.45	-1.22	15.28	1.22	0.16
first year of earnings	-1.13	-0.63	-0.78	-0.68	1722.25
last year of earnings	7.58	3.1	-2.92	-2.52	1732.53
total earnings to date	0.02	2.35	-0.01	-2.38	60245.30
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

Table A1: Probit Results on Predicting Non-Covered vs. Non-Worker, 1996 SIPP-SER (cont.)

Men in 1996 *Dependent Variable*

Variable	**Non-Covered** *dF/dx times 100*	*z-stat*	**Non-Worker** *dF/dx times 100*	*z-stat*	*Mean*
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 X self-emp.	0.01	2.46	0.00	-1.79	981.06
0 earnings 1995 X self-emp.*	41.91	1.84	-18.20	-1.31	0.16
0 earnings 1997 X self-emp.*	-36.59	-0.98	13.84	0.57	0.14
0 earnings 1998 X self-emp.*	28.55	1.71	-3.45	-0.2	0.11
0 earnings 1995 X agr.*	-25.60	-0.84	31.93	1.32	0.10
# of 0's 95-98 X 1995 earn.	-0.01	-3.27	0.00	2.23	3760.91
Observations	258		258		
Pseudo R squared	0.60		0.53		

* indicates a 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. "Not Covered" have zero SER earnings but at least 10 paid months of work. Unemployed have zerc SER earnings and between 0 to 2 months of paid work.

Table A2: The Effects of Varying the Exclusion Rule on Years of Fathers' Zero Earnings

elast. *(s.e.)* *N*	***Dependent Variable*** *Son's Log Avg Earnings, 1995-1998*							
	Require the following number of years of positive earnings…							
	0	1	2	3	4	5	6	7
Fathers'	0.160	0.290	0.336	0.434	0.468	0.462	0.440	0.445
Log Avg. Earnings	(0.034)	(0.062)	(0.058)	(0.069)	(0.074)	(0.076)	(0.075)	(0.079)
1979-1985	1299	1256	1245	1227	1212	1201	1181	1160

Note: For the dependent variable, probit models based on the 1996 SIPP matched to SER were used to determine if zero earnings reflected noncovered status and if so were imputed. For fathers, earnings for those identified as non-covered are dropped. Earnings for those topcoded are imputed using March CPS data for 1970-80 and using 1984 SIPP for 1981 to 1984.