Life-Cycle Variation in the Association between Current and Lifetime Earnings

Steven Haider, Michigan State University

Gary Solon, University of Michigan

March 2004

The authors gratefully acknowledge grant support from the National Institute on Aging (2-P01 AG 10179). They also are grateful for advice from John Bound, Charlie Brown, Art Goldberger, Nathan Grawe, Jacob Klerman, Matthew Shapiro, Mel Stephens, Bob Willis, Jeff Wooldridge, and seminar participants at the University of Michigan, American University, Harvard's Kennedy School, Ohio State University, the University of Toronto, and Western Michigan University.

Life-Cycle Variation in the Association between Current and Lifetime Earnings

Abstract

Researchers in a variety of important economic literatures have assumed that current income variables as proxies for lifetime income variables follow the textbook errors-in-variables model. In an analysis of Social Security records containing nearly career-long earnings histories for the Health and Retirement Study sample, we find that the relationship between current and lifetime earnings departs substantially from the textbook model in ways that vary systematically over the life cycle. Our results can enable more appropriate analysis of and correction for errors-in-variables bias in any research that uses current earnings to proxy for lifetime earnings.

Life-Cycle Variation in the Association between Current and Lifetime Earnings

I. Introduction

In the year 2003 alone, the *American Economic Review*'s refereed issues contained 14 articles reporting regression analyses involving individual or family income variables, and the May *Proceedings* issue contained almost that many again. In some cases, the income variables were dependent variables; in others, they were regressors used to explain dependent variables ranging from child health in the United States to borrowing and lending behavior in Ghana. Without exception, the measured income variables were short-term values even though, in most cases, it appeared that the relevant economic construct was a longer-term value.

Many influential economic studies have recognized that the use of current income as a proxy for long-run income can generate important errors-in-variables biases. Perhaps the most famous examples are the seminal studies by Modigliani and Brumberg (1954) and Friedman (1957), which analyzed the properties of consumption functions estimated with current rather than permanent income variables as the regressors. Another instance is the rich literature – exemplified by Lillard (1977) and Shorrocks (1981) – suggesting that inequality as measured in cross-sections of annual earnings overstates the inequality in lifetime earnings. A recent offshoot of that literature – exemplified by Gottschalk and Moffitt (1994), Haider (2001), and Baker and Solon (2003) – has attempted to partition the upward trend in earnings inequality into persistent and transitory components. Still another recent example is the burgeoning literature on intergenerational income mobility (surveyed in Solon, 1999), which has found that the association between parents' and children's long-run income is susceptible to dramatic underestimation when current income variables are used as proxies for long-run income.

Nevertheless, applied researchers often ignore the distinction between current and long-run income. Most researchers who do attend to the issue assume the textbook errors-in-variables model and impute the noise-to-signal ratio by estimating restrictive models of income dynamics on the basis of short panels of income data spanning only a segment of the life cycle.[1] In this paper, we reconsider the appropriateness of the textbook errors-in-variables model, and we find that it does not accurately characterize current earnings as a proxy for lifetime earnings. Thanks to a remarkable new data set, we are able to generate detailed evidence on the association between current and lifetime earnings, including its evolution over the life cycle, without having to resort to an arbitrary specification of the earnings dynamics process.

Our empirical analysis uses the 1951-1991 Social Security earnings histories of the members of the Health and Retirement Study sample. Despite some limitations discussed in section III, these data provide nearly career-long earnings histories, based on relatively accurate administrative data, for a broadly representative national sample. In section II, we develop simple models to illustrate some important aspects of the association between current and lifetime earnings and to demonstrate the implications for errors-in-variables biases in applied econometric research. In section III, we describe the data set and our econometric methods. In section IV, we present our evidence on the connections between annual and lifetime earnings, and we apply the results to

[1] A relatively sophisticated recent example is Mazumder (2001). Like us, Mazumder uses Social Security earnings histories, but he uses no more than 15 years from any worker's history and therefore has to rely on parametric earnings dynamics models to infer the connection between current and lifetime earnings. Another important contrast with our work is that Mazumder's study, like most others, does not address the inappropriateness of the textbook errors-in-variables model.

interpreting evidence on intergenerational earnings mobility. Section V summarizes the main findings.

II. Models

Following Friedman (1957), most analyses of current income variables as proxies for unobserved lifetime income variables have adopted the textbook errors-in-variables model

(1) $$y_{it} = y_i + u_{it}$$

where y_{it} is a current income variable, such as log annual earnings, observed for individual i in period t; y_i is a long-run income variable, such as the log of the present discounted value of lifetime earnings; and u_{it}, the measurement error in y_{it} as a proxy for y_i, is assumed to be uncorrelated with y_i (and each of its determinants). Often, the current income variable y_{it} has been adjusted for stage of life cycle with a regression on a polynomial in age or experience or by subtracting out the cohort mean. Throughout this section, we will suppress intercepts by expressing all variables as deviations from their population means.

The textbook errors-in-variables model in equation (1) is effectively a regression model that assumes the slope coefficient in the regression of y_{it} on y_i equals 1. One familiar implication of that restriction is that, if y_{it} proxies for y_i as the dependent variable in a linear regression equation, ordinary least squares (OLS) estimation of that regression equation consistently estimates the equation's slope coefficients. Another well-known implication is that, if y_{it} proxies for y_i as the sole explanatory variable in a simple regression equation, the probability limit of the OLS estimator of the equation's slope coefficient equals the true coefficient times an attenuation factor equal to $Var(y_i)/[Var(y_i)+Var(u_{it})]$.

These oft-used results no longer apply if the textbook errors-in-variables model incorrectly characterizes the relationship between current and lifetime income. In part A of this section, we explain our reasons for suspecting that the slope coefficients in regressions of current income variables on lifetime variables vary systematically over the life cycle and do not generally equal 1.[2] In part B, we show how such departures from the textbook model alter the standard results on errors-in-variables bias.

A. Life-cycle variation

Several fragments of evidence suggest that the association between current and lifetime income variables varies over the life cycle. Bjorklund (1993), the closest predecessor to our study, uses Swedish income tax data from 1951-1989 to conduct a direct comparison of current and lifetime income. He finds a strong life-cycle pattern in the correlation between current and lifetime income. In his words, "the correlations are quite low – and in some cases even negative – up to around 25 years of age and are rather high after 35 years of age. In general the correlations are around 0.8 after the age of 35." Unfortunately, the correlations in income levels reported by Bjorklund do not map directly into magnitudes of errors-in-variables biases in the sorts of regression estimation that economists commonly do. In the next subsection, we develop measures of

[2] Another literature on departures from the textbook errors-in-variables model is the work by Bound and Krueger (1991) and others about the measurement error in survey reports of current earnings as measures of true current earnings. In contrast, our work is about the measurement error in true current earnings as measures of lifetime earnings. In section IV, we discuss the relevance of their results for the interpretation of ours.

association between current and lifetime earnings that do have direct implications for errors-in-variables biases.

Another indication of life-cycled-related departures from the textbook errors-in-variables model, noted by Jenkins (1987) and Grawe (2003), involves the estimation of intergenerational mobility models such as the regression of son's log lifetime earnings on father's log lifetime earnings. If son's log annual earnings as a proxy for the dependent variable obeyed the textbook errors-in-variables model, the estimated intergenerational elasticity would have the same probability limit regardless of the age at which the son's earnings were observed. On the other hand, if the slope coefficient in the regression of son's log annual earnings on son's log lifetime earnings deviates from 1 in a way that evolves over the life cycle, then analyses observing sons' earnings at different ages will yield systematically different elasticity estimates. Solon's (1999) survey of the intergenerational mobility literature reveals precisely such a pattern – the studies that estimate the smallest elasticities tend to be those that observe sons' earnings early in their careers. Correspondingly, several studies (Reville, 1995; Chadwick and Solon, 2002; Abul Naga, 2002; Dunn, 2003; Grawe, 2003) that have explicitly investigated the effects of varying the ages at which sons' earnings are observed have found that the estimated intergenerational elasticities increase substantially as the sons' earnings are observed further into their careers.

Notwithstanding the strong tradition of assuming that current income variables as proxies for lifetime income variables follow the textbook errors-in-variables model, indications that this assumption is false should not be surprising. Any realistic model of income evolution over the life cycle would contradict the traditional assumption. Here is an extremely simple example.

Suppose that y_{it}, the log real earnings of worker i in year t of his career, follows

(2) $$y_{it} = \alpha_i + \gamma_i t$$

where initial log earnings α_i varies across the population with variance σ_α^2 and the earnings growth rate γ_i varies across the population with variance σ_γ^2. Heterogeneity in earnings growth is a natural consequence of heterogeneity in human capital investment, and its empirical importance has been documented by Mincer (1974), Baker (1997), Haider (2001), and Baker and Solon (2003) among others. Purely for the sake of simplicity, assume that $Cov(\alpha_i, \gamma_i) = 0$. Then, assuming infinite lifetimes and a constant real interest rate $r > \gamma_i$, the present discounted value of lifetime earnings is

(3) $$\begin{aligned} V_i &= \sum_{s=0}^{\infty} \exp(\alpha_i + \gamma_i s)(1+r)^{-s} \\ &\cong \sum_{s=0}^{\infty} \exp(\alpha_i)[(1+\gamma_i)/(1+r)]^s \\ &= \exp(\alpha_i)[(1+r)/(r-\gamma_i)], \end{aligned}$$

and the log of the present value of lifetime earnings is thus

(4) $$\log V_i \cong \alpha_i + r - \log r + \frac{\gamma_i}{r}.$$

What then is the value of the slope coefficient in the regression of current log earnings on the log of the present value of lifetime earnings? Since

(5) $$Var(\log V_i) \cong \sigma_\alpha^2 + (\sigma_\gamma^2 / r^2)$$

and

(6) $$Cov(\log V_i, y_{it}) \cong \sigma_\alpha^2 + (t\sigma_\gamma^2 / r),$$

it follows that the slope coefficient is

(7) $$\lambda_t \equiv \frac{Cov(\log V_i, y_{it})}{Var(\log V_i)} \cong \frac{\sigma_\alpha^2 + (t\sigma_\gamma^2 / r)}{\sigma_\alpha^2 + (\sigma_\gamma^2 / r^2)}.$$

The main thing to note about this result is that, contrary to the textbook errors-in-variables model, λ_t generally does not equal 1. Instead, it starts at a value less than 1 at the outset of the career and then increases monotonically over the life cycle. It reaches 1 when $t = 1/r$ and then exceeds 1 afterwards. The intuition is that the workers with high lifetime earnings tend to be those with high earnings growth rates. Consequently, when comparing the current earnings of those with high and low lifetime earnings, an early-career comparison tends to understate their gap in lifetime earnings, and a late-career comparison tends to overstate it. Note that the common practice of adjusting current earnings for the central tendency of earnings growth over the life cycle does not undo this result. The result is due to heterogeneous variation *around* the central tendency.

Of course, the exact result in equation (7) is particular to the very simple assumptions of the model. A more realistic model would incorporate many additional features including transitory earnings fluctuations, nonzero covariance between initial earnings and earnings growth, nonlinear growth, and shocks with permanent effects. While these features would lead to a more complex relationship between λ_t and t, they clearly would not overturn the main qualitative results – that λ_t does not generally equal 1 and should be expected to vary over the life cycle.

Figure 1 provides a pictorial version of the argument. The figure contains the life-cycle log earnings trajectories of workers 1 and 2, with worker 2 attaining higher lifetime earnings. Both trajectories display the familiar concave shape documented and analyzed by Mincer (1974), and worker 2 experiences more rapid earnings growth. The horizontal lines depict the log of the annuitized value of each worker's present discounted value of lifetime earnings. The difference between the two workers' log lifetime earnings therefore is simply the vertical distance between the two horizontal lines. But how well is that difference estimated if it is proxied by the difference in log earnings at a particular age? If the worker with higher lifetime earnings has a steeper earnings trajectory, then the current earnings gap between the two workers early in their careers tends to understate their gap in lifetime earnings (and could even have the opposite sign). As the workers mature, this downward bias becomes less severe until age t^*, when the vertical distance between the current earnings trajectories equals the distance between the horizontal lines. That is the age at which the textbook errors-in-variables model is correct. Beyond that age, the gap in current earnings tends to *overstate* the gap in lifetime earnings.

In the next subsection, we will explain the econometric implications of such a departure from the textbook errors-in-variables model. Then, in the remainder of the paper, we will use the Social Security earnings histories for the Health and Retirement Study sample to ascertain what the empirical relationship between current and lifetime earnings actually is.

B. Errors-in-variables biases

Suppose we wish to estimate the simple regression model

(8) $$y_i = \beta X_i + \varepsilon_i$$

where the error term ε_i is uncorrelated with the regressor X_i. Either y_i or X_i can be a log lifetime earnings variable. In the case of an intergenerational mobility regression, both are: y_i for the sons and X_i for the fathers.

Start with the case in which only y_i is log lifetime earnings and, in the absence of a direct measure, we proxy for it with y_{it}, log annual earnings at age t. In accordance with the discussion in the preceding subsection, we do not assume the textbook errors-in-variables model in equation (1). Instead, we generalize that model to

$$(9) \qquad y_{it} = \lambda_t y_i + u_{it},$$

where λ_t, the slope coefficient in the linear projection of y_{it} on y_i, need not equal 1 and may vary over the life cycle. Equation (9) is not a structural model of either earnings dynamics or measurement error. Rather, it is just the linear projection of log earnings at age t on log lifetime earnings, which is the relevant statistical relationship for the ensuing analysis of errors-in-variables bias.

If OLS is applied to the regression of y_{it} on X_i, the probability limit of the estimated slope coefficient $\hat{\beta}$ is

$$(10) \qquad \text{plim}\,\hat{\beta} = \frac{Cov(X_i, y_{it})}{Var(X_i)} = \lambda_t \beta .$$

In the textbook case where $\lambda_t = 1$, measurement error in the dependent variable does not result in inconsistent estimation of β. More generally, however, the OLS estimator is inconsistent, and the inconsistency varies as a function of the age at which annual earnings are observed.

Next, consider the case in which the dependent variable y_i is observed (or is measured with error of the textbook variety), but the regressor X_i is unobserved log lifetime earnings, which is proxied for with log annual earnings at age s. As in equation (9) for y_{it}, we express the linear projection of X_{is} on X_i as

$$X_{is} = \lambda_s X_i + v_{is}. \tag{11}$$

If OLS is applied to the linear regression of y_i on X_{is}, the probability limit of the estimated slope coefficient is

$$\begin{aligned} \operatorname{plim}\hat{\beta} &= \frac{Cov(X_{is}, y_i)}{Var(X_{is})} \\ &= \frac{Cov(X_{is}, X_i)}{Var(X_{is})}\beta \\ &= \theta_s \beta \end{aligned} \tag{12}$$

where

$$\theta_s \equiv \frac{Cov(X_{is}, X_i)}{Var(X_{is})} = \frac{\lambda_s Var(X_i)}{\lambda_s^2 Var(X_i) + Var(v_{is})}. \tag{13}$$

The inconsistency factor θ_s, sometimes referred to as the "reliability ratio," is most simply interpreted as the slope coefficient in the "reverse regression" of X_i on X_{is}. In the textbook case where $\lambda_s = 1$, this factor simplifies to the familiar attenuation factor $Var(X_i)/[Var(X_i) + Var(v_{is})]$. More generally, the factor θ_s also depends on the value of λ_s. Indeed, with $\lambda_s < 1$ and sufficiently small $Var(v_{is})/Var(X_i)$, θ_s can exceed 1 so that the errors-in-variables bias is an amplification bias rather than an attenuation bias.

Finally, consider the case (such as the intergenerational mobility model) where both y_i and X_i are log lifetime earnings variables proxied respectively by y_{it} and X_{is}.

If u_{it} and v_{is} are uncorrelated, it is straightforward to combine the results above to show that, if OLS is applied to the regression of y_{it} on X_{is}, the probability limit of the estimated slope coefficient $\hat{\beta}$ is[3]

(14) $\quad \text{plim}\,\hat{\beta} = \lambda_t \theta_s \beta$.

These results deliver two key messages. First, with plausible departures from the textbook errors-in-variables model, the familiar textbook results about OLS estimation are overturned. Measurement error in the dependent variable is not innocuous for consistency, and measurement error in the explanatory variable can induce either amplification or attenuation bias. Second, the estimation biases from using log annual earnings as a proxy for log lifetime earnings can be summarized with just two simple parameters: the slope coefficient in the "forward regression" of log annual earnings on log lifetime earnings and the slope coefficient in the "reverse regression" of log lifetime earnings on log annual earnings.[4] In section IV, we will estimate those two parameters and examine how they vary over the life cycle.

---

[3] We are using the t subscript to denote the age at which the left-side proxy is observed and s to denote the age at which the right-side proxy is observed. In the special case in which $t = s$ and the λ and θ parameter vectors are the same for both y and X (e.g., the same for both generations in an intergenerational mobility regression), $\lambda_t \theta_t$ is the squared correlation between log earnings at age t and log lifetime earnings, which means it is the R^2 for both the "forward" and "reverse" regressions between log current earnings and log lifetime earnings. In this special case, correlations such as those reported by Bjorklund (1993) are sufficient to characterize the errors-in-variables bias. More generally, however, identifying the errors-in-variables bias requires separate information on λ_t and θ_s.

[4] Of course, the analysis here can be extended as needed to other practically relevant estimation situations. One such situation is the case of multiple regressors. Another is where u_{it} and v_{is} are correlated. Another is where the measurement error in the regressor is treated with an instrumental variable correlated with X_i, but not with ε_i, u_{it}, or v_{is}. It is easy to show that the probability limit of the instrumental variable estimator is simply $(\lambda_t / \lambda_s)\beta$. The inconsistency of conventional instrumental variable estimation in the presence of non-classical measurement error has been discussed previously by Kane, Rouse, and Staiger (1999), Bound and Solon (1999), and Kim and Solon (forthcoming).

III. Data and Methods

A. Data

Most U.S. studies of the relationship between current and lifetime income variables have been based on longitudinal survey data from only a limited portion of the respondents' careers. In contrast, like Bjorklund's (1993) study of Swedish income tax data, our study is based on nearly career-long earnings histories. This information is now available for a U.S. sample because, in accordance with an agreement with the Social Security Administration, the University of Michigan's Survey Research Center asked the participants in its Health and Retirement Study (HRS) to permit access to their Social Security earnings histories for 1951-1991.[5] The HRS sample is a national probability sample of Americans born between 1931 and 1941, and about ¾ of the respondents agreed to permit access to their Social Security earnings histories. As shown in Haider and Solon (2000), in terms of observable characteristics, the respondents that granted access appear to be surprisingly representative of the complete sample. The earnings data supplied by the Social Security Administration round the earnings observations to the nearest hundred dollars, with a distinction made between zero and positive amounts less than $50.

Our analysis is for male HRS respondents born between 1931 and 1933, who were about 19 years old at the beginning of the 1951-1991 earnings period and about 59 at the end. Thus, for the 821 men in our analysis, we have annual earnings information

[5] Because of the highly confidential nature of the data, the earnings histories are not part of the HRS public release data sets, but are provided only through special permission from the Survey Research Center. For information on accessing "HRS restricted data," see the HRS website http://hrsonline.isr.umich.edu. For more general information on the HRS, see the website or Juster and Suzman (1995).

for every year over the major portion of their careers.[6] The other main strength of our data set is that the Social Security earnings histories tend to be more accurate than the survey reports of earnings used in most previous research. Indeed, Bound and Krueger's (1991) influential study of errors in earnings reports in the Current Population Survey used Social Security earnings data as the "true" values against which the Current Population Survey measures were compared.

The strengths of the Social Security earnings data are accompanied by two serious limitations. First, the earnings data pertain only to jobs covered by Social Security. According to Social Security Administration (1999, table 3.B2), the percentage of earnings covered by Social Security has exceeded 80% ever since the coverage extensions effected by 1957 and exceeded 85% over most of our sample period. Between 1951 and 1956, however, this percentage ranged between 66 and 79%. Accordingly, in addition to our main analysis for 1951-1991, we also will report results for 1957-1991.

Second, the Social Security earnings in our data are measured only up to the maximum amount subject to Social Security tax. In some years, the proportion of observations that are "right-censored" is quite large. For the 821 men in our sample, table 1 displays the median observed earnings, the number in the sample with zero earnings, the taxable limit, and the number with earnings at the taxable limit for each year from 1951 to 1991. The table shows that, in the early years, very few sample members are earning enough to approach the taxable limit. As their earnings grow over their careers, however, the taxable limit becomes more constraining, especially in the years

---

[6] This sample is restricted to workers with positive earnings in at least 10 years during 1951-1991. This criterion is less restrictive than the usual practice in survey-based earnings dynamics studies of requiring positive earnings in *every* year (e.g., Abowd and Card, 1989; Baker, 1997). Within this sample, our main analysis includes years of zero earnings, but we also will report results from an analysis based only on the positive earnings observations.

when the taxable limit is low relative to the general earnings distribution. The worst year is 1965, when 62% of the sample is right-censored. Afterwards, the degree of censorship lessens as the taxable limit is progressively increased. By 1991, only 9% of the sample is right-censored. Although some previous studies of current and lifetime earnings have used annual earnings data with less severe right-censorship, their observation of earnings usually has been limited to relatively short segments of the life cycle. In effect, they have used restrictive models of earnings dynamics to impute missing earnings data over most years of their sample members' careers.

If not for the right-censorship, we would follow Bjorklund's (1993) approach of directly summarizing the observed joint distribution of annual and lifetime earnings. Because of the right-censorship, however, we are forced instead to estimate the joint distribution in a way that imputes the censored right tails of the annual earnings distributions. We describe our methods in the next subsection.

B. Econometric methods

As explained above in section II.B, our ultimate goal is to summarize the association between annual and lifetime earnings in terms of two types of parameters. One is λ_t, the slope coefficient in the regression of log earnings in year t on the log of the present value of lifetime earnings. The other is θ_t, the slope coefficient in the reverse regression of log lifetime earnings on log earnings in year t. If we had complete data, we would estimate these parameters simply by applying least squares to the forward and reverse regressions of the relevant variables.

Because of the censorship of the Social Security earnings data at the taxable limit, however, we cannot observe the exact value of annual earnings in the cases where earnings are right-censored and furthermore, in those cases, we also cannot compute the present value of lifetime earnings. We therefore apply a three-step procedure for estimating the λ and θ coefficients. First, we use a limited-dependent-variable model to estimate the joint distribution of uncensored annual earnings in the 41 years from 1951 through 1991. Second, drawing from that estimated joint distribution, we generate a simulated sample of uncensored earnings histories, for which we can calculate the present discounted value of lifetime earnings. Third, using the uncensored earnings data for that sample, we apply least squares to the forward and reverse regressions to obtain our estimates of the λ and θ parameters.

The key assumption in our first step is that the uncensored values of log annual earnings over the 41 years from 1951 to 1991 follow a multivariate normal distribution. Given this variant of the traditional Tobit assumption for limited dependent variables, the joint distribution of the 41 annual earnings variables can be fully characterized by the mean and variance of log earnings for each year and the cross-year autocorrelations of log earnings for every pair of years.

To estimate the year-specific means and variances for our sample cohort born in 1931-1933, we simply apply the conventional cross-sectional Tobit maximum likelihood estimator separately for each year from 1951 to 1991. The only regressor in each year's equation is 1, the coefficient of which is the intercept. The estimated intercept is our estimate of the cohort's mean log earnings in that year. In our main analysis, we use a two-limit Tobit model. The right-censorship threshold is the Social Security taxable limit for that year. The left-censorship threshold is \$50. Observations of zero earnings and of positive earnings less than \$50 are both included as observations left-censored at \$50.[7]

Having estimated each year's mean and variance in the cross-sectional Tobits, we still need to estimate the autocorrelations between years. To obtain those estimates, we apply the conventional bivariate Tobit maximum likelihood estimator separately for each of the $41 \times 40/2 = 820$ distinct pairs of years in our 1951-1991 period. With those autocorrelations estimated along with the mean and variance for every year, we have an estimated version of the entire joint distribution of uncensored annual earnings over all 41 years.

In the second step of our procedure, we use our estimated joint distribution of uncensored earnings for 1951-1991 to perform the following simulation. First, we take 2,000 random draws from the estimated joint distribution of the 41 years of annual earnings.[8] Then, for each of the 2,000 simulated earnings histories, we calculate the present discounted value of lifetime earnings. In the main version of the simulation, we perform the discounting by (1) using the personal consumption expenditures deflator to

---

[7] In the simulation described below, our treatment of zero-earnings observations as left-censored observations from a lognormal distribution causes our simulated observations to include no zeros, but instead small annual earnings values less than \$50. The purpose of the simulation is to generate observations for the present discounted value of lifetime earnings. For that purpose, the difference between annual earnings of zero or a few dollars is of practically no consequence.

[8] To implement the simulation, we need the estimated autocovariance matrix to be positive semi-definite (as the true one must be). Our element-by-element method of estimation does not guarantee that the initial estimate of the autocovariance matrix is positive semi-definite, and indeed it is not. Our procedure for imposing the restriction of positive semi-definiteness begins by characterizing the autocovariance matrix Ω in terms of the Cholesky decomposition $\Omega = TT'$ where T is lower triangular. The matrix Ω is positive semi-definite if and only if the diagonal elements of T are non-negative. We therefore choose $\hat{T}$ to minimize the distance between $\hat{\Omega}$ and our initial estimate of the autocovariance matrix subject to the constraint that the diagonal elements of $\hat{T}$ are non-negative. We measure distance as the sum of squared deviations between the distinct elements in $\hat{\Omega}$ and the corresponding elements in the initial estimate of the autocovariance matrix. Because our initial estimate of the autocovariance matrix is nearly positive semi-definite, the elements in $\hat{\Omega}$ differ only slightly from those in the original estimate. We are very grateful to Jeff Wooldridge for his help in devising this method.

convert each year's nominal earnings to a real value and (2) assuming a constant real interest rate of 0.02. In the end, we have a simulated sample of 2,000 observations for which we observe the present discounted value of lifetime earnings as well as each year's earnings.

Finally, for this sample of 2,000 individuals, we apply OLS to the regression of each year's log annual earnings on the log of the present value of lifetime earnings, and thereby produce a $\hat{\lambda}_t$ for each year from 1951 to 1991. Similarly, we obtain a $\hat{\theta}_t$ for each year by applying OLS to the reverse regression of the log of the present value of lifetime earnings on each year's log annual earnings. Plotting each of these coefficient estimates over time depicts the life-cycle trajectory of the association between current and lifetime earnings in a way that translates directly into implications for errors-in-variables biases.

IV. Empirical Results

A. Main estimates

In the first step of our estimation procedure, the Tobit analysis described above results in a 41×41 estimated autocovariance matrix for log annual earnings from 1951 to 1991. For the sake of brevity, we will not display the entire matrix (available on request from the authors) but will report some illustrative portions. Table 2 shows the estimated autocorrelations for 1975-1984, a period when our cohort born in 1931-1933 is between the ages of about 43 and 52. As shown in the second column of table 3, the first-order autocorrelations over this period average to 0.80, the second-order autocorrelations average to 0.72, the third-order autocorrelations average to 0.69, and so forth. This pattern is similar to Baker and Solon's (2003) report that, in their Canadian income tax data, "we find autocorrelations of around 0.8 at the first order, followed by gradual declines at higher orders." The third column of table 3 shows the average autocorrelations over 1976-1992 that Baker and Solon report for the cohort born in 1942-1943, and they are strikingly similar to our results in the second column. Note that this resemblance occurs even though Baker and Solon use uncensored data and therefore can estimate the autocorrelations directly without imposing distributional assumptions.

These autocorrelation patterns from administrative earnings data are, in turn, similar to those reported by Baker (1997) and Haider (2001) in their analyses of the Panel Study of Income Dynamics (PSID). Their estimates are summarized in the fourth and fifth columns of table 3. At first, the similarity is surprising because one might expect that reporting errors in survey data on earnings would add noise that is less serially correlated than true earnings and, hence, would reduce the measured autocorrelations. Pischke's (1995) analysis of the PSID Validation Study, however, finds a tendency for survey respondents to underreport the magnitude of transitory earnings fluctuations. He concludes that, because this mean-reversion in the reporting of transitory earnings partly offsets the contribution of measurement noise to apparent transitory variation, "autocorrelations in the changes of earnings can be estimated relatively accurately despite the presence of measurement error." Bound et al. (2001) conjecture that this is why Baker and Solon's results from administrative data resemble those from surveys.

We find it somewhat reassuring that, despite the omission of earnings not covered by Social Security and the imputation of right-censored values, our autocorrelation estimates are quite similar to those from other data sets. Another relevant comparison is to an alternative earnings variable available for our sample for 1980-1991, the last 12 years of our sample period. For those years, in addition to the Social Security earnings data, the Survey Research Center also has obtained earnings data from employers' W-2 reports to the Internal Revenue Service. Unlike the Social Security data, the W-2 variable includes earnings not covered by Social Security, and it is right-censored (for confidentiality reasons) at $125,000, which is far less constraining than the Social Security taxable limits listed in table 1. On the other hand, the W-2 variable leaves out self-employment earnings and earnings allocated to 401(k) pensions. As shown in the sixth column of table 3, when we use the W-2 data to reestimate our Tobits for 1980-1991, the first-order autocorrelations average to 0.83, the second-order autocorrelations average to 0.77, the third-order autocorrelations average to 0.72, and so forth. As shown in the last column, the corresponding average autocorrelations for the Social Security earnings variable over the same period are 0.81, 0.74, and 0.71. Once again, the idiosyncrasies of the alternative earnings measures do not appear to generate major discrepancies in the estimated persistence of earnings.[9]

While the results described so far give a good sense of the earnings autocorrelation pattern for mature workers, the pattern is quite different in our sample's early years. Table 4 shows the estimated autocorrelations for 1951-1960, when our sample is between the ages of about 19 and 28. One obvious and unsurprising pattern is that the autocorrelations are much lower in this period, when many members of our

[9] Perhaps the similarity of the autocorrelation estimates should not be a surprise. If one thinks of the log of covered earnings as the sum of log total earnings and the log of the proportion covered, one would expect the autocorrelation of log covered earnings to be approximately a weighted average of the autocorrelations for log total earnings and log coverage. Presumably, both of these autocorrelations are highly positive. If they are not very different from each other, then "averaging in" the coverage autocorrelation will not produce a large bias in estimating the earnings autocorrelation.

sample have not yet settled into their career paths. The table also suggests a discrete increase in autocorrelations for pairs of years no earlier than 1957, when Social Security's coverage was increased. In the next subsection, we will describe an alternative analysis that excludes the years prior to 1957.

In the second step of our estimation procedure, we perform the simulation in which we take 2,000 draws from the estimated joint distribution of the 41 years of annual earnings. For the resulting simulated population of 2,000 observations, table 5 shows the coefficient of variation (i.e., the ratio of the standard deviation to the mean) for annual earnings in each year from 1951 to 1991 as well as for the present discounted value of lifetime earnings. The results echo two familiar findings in the earnings dynamics literature. First, as found by Baker and Solon (2003) and others, relative inequality in annual earnings declines in the early years of the life cycle and then goes back up. Second, as found by Lillard (1977), Shorrocks (1981), Bjorklund (1993), and others, inequality in cross-sections of annual earnings overstates the inequality in lifetime earnings.

Finally, we summarize the connection between annual and lifetime earnings by estimating the forward and reverse regressions between the logs of annual and lifetime earnings. Figure 2 plots our estimates of λ_t, the slope coefficient in the regression of log annual earnings at time t on the log of the present value of lifetime earnings. To focus on the life-cycle variation in λ_t, we express t on the horizontal axis as year minus 1932, which gives the approximate age of our 1931-1933 cohort in each year. In addition to plotting our point estimates of λ_t at each age, figure 2 also shows approximate 95% confidence intervals around the estimates. These are constructed as 1.96 estimated standard errors above and below the point estimates, with the standard errors estimated by a bootstrap procedure.[10]

The estimated life-cycle trajectory shown in figure 2 is similar to the one predicted by the simple model in section II.A. In contrast to the textbook assumption that λ_t equals 1 throughout the life cycle, $\hat{\lambda}_t$ begins at 0.29 at age 19, increases steadily until it crosses 1 at age 42, and exceeds 1 afterwards, peaking at 1.39 at age 48. The main implication is that, contrary to the textbook errors-in-variables model, using log current earnings to proxy for log lifetime earnings as the dependent variable induces an errors-in-variables bias. Using log current earnings at early ages causes a large attenuation bias, and using log current earnings late in the life cycle causes an amplification bias. A constructive implication is that the bias is small if one uses log current earnings around age 40, when the textbook assumption that $\lambda_t = 1$ is approximately correct.

Figure 3 shows the estimated life-cycle trajectory of the reliability ratio θ_t, the relevant parameter for assessing errors-in-variables bias from using log annual earnings to proxy for log lifetime earnings as the explanatory variable in a simple regression. $\hat{\theta}_t$ begins at 0.09 at age 19, increases to a fairly flat peak between 0.4 and 0.5 between ages

[10]Plotting the 95th and 5th percentiles from the bootstrap replications yields slightly tighter confidence intervals. To construct the confidence intervals for the estimates of λ_t and θ_t in figures 2 and 3, we conduct 80 iterations of choosing new samples of size 821 by sampling with replacement from our original sample of 821 individuals. For each of the bootstrap samples, we perform our entire sequential estimation procedure to generate estimates of λ_t and θ_t for each t. For each parameter estimate plotted in figures 2 and 3, we estimate the standard error with the sample standard deviation of that parameter estimate across the bootstrap replications. The only departure from the estimation procedure in our main analysis is that we use a different method for imposing positive semi-definiteness of the autocovariance matrix. Instead of using the method described in footnote 8, we perform a spectral decomposition on the estimated covariance matrix, set the negative eigenvalues to zero, and then re-multiply the various elements together. This change greatly reduces the computational time, and we have verified that the resulting positive semi-definite matrix is very similar to what would be obtained using the previous method. Furthermore, to the extent that a "closer" positive semi-definite matrix would exist, this simplification can be interpreted as introducing noise into our bootstrap procedure, which probably would produce overly large confidence intervals.

33 and 56, and then decreases. Our discussion in section II.B showed that theoretically the errors-in-variables bias could be either an attenuation bias or an amplification bias. Our empirical results, however, confirm the conventional presumption of Friedman (1957) and many subsequent authors that using current earnings to proxy for lifetime earnings as a regressor induces a large attenuation bias. The bias is especially large if current earnings are measured early in the life cycle. There is a wide age range in mid-career when the errors-in-variables bias stays about the same, but it remains quite substantial even in that range.

B. Robustness checks

To check the robustness of our main results, we have carried out a series of sensitivity analyses, the results of which are displayed in figures 4 and 5. The first supplementary analysis is motivated by the question of how to treat years of zero earnings. Our main results are based on two-limit Tobit estimates that retain observations of zero earnings in the analysis. Most previous analyses of earnings dynamics, however, have excluded observations of zero earnings (e.g., Abowd and Card, 1989; Baker, 1997). We therefore supplement our main analysis with another that excludes the zeros, codes positive earnings less than \$50 as \$25, and estimates one-limit Tobits with only right-censorship. As shown in table 1, zero earnings are especially prevalent in the early years of our sample, both because many of our sample members are not yet working for pay and because the Social Security system's coverage is less extensive before 1957. We therefore conduct this analysis only for 1957-1991. Excluding the zeros changes the estimates of the variances and autocovariances in log annual earnings, but because those changes are roughly proportional, the estimated autocorrelations are similar to those in the main analysis. Accordingly, the new estimates of λ_t and θ_t, denoted by the lines with triangles in figures 4 and 5, are quite similar to the estimates from our main analysis repeated from figures 2 and 3.

A second supplementary analysis is designed to probe further into the sensitivity of our results to our Tobit-based imputation of right-censored earnings. Our reanalysis with the 1980-1991 W-2 data explored that issue by easing the right-censorship. Another approach is to check the consequences of *increasing* the severity of the right-censorship in a way that imposes more uniform censorship across the years. This is particularly useful for checking that our estimated life-cycle patterns in λ_t and θ_t are not just artifacts of temporal variation in the degree of censorship. Accordingly, we reapply our entire estimation procedure after right-censoring the upper 40% of the earnings distribution in every year from 1957 to 1991 that really has less than 40% censorship (i.e., the 22 years in the intervals 1957-1960 and 1974-1991). The results, shown as the dotted lines in figures 4 and 5, are again quite similar to the main results. Obviously, these results based on throwing out a lot of information should not be viewed as preferred estimates, but the robustness of the results to a substantial change in the frequency of imputations suggests that the results are not driven by the imputations.

Third, we have checked the sensitivity of our results to our choice of interest rate series. In our main simulation, we calculated the present discounted value of lifetime earnings by (1) using the personal consumption expenditures deflator to convert each year's nominal earnings to a real value and (2) assuming a constant real interest rate of 0.02. We also have tried discounting nominal earnings by a nominal interest rate series, the annual one-year T-note interest rates.[11] The results, shown as the lines with asterisks in figures 4 and 5, are quite similar to those based on our original interest rate series.

Fourth, we have checked whether our results are affected by the Health and Retirement Study's oversampling of blacks, Hispanics, and residents of Florida. To do so, we have reestimated the joint distribution of earnings with a Tobit quasi-maximum likelihood procedure that weights each observation's contribution to the likelihood function by its inverse probability of selection into the sample. The resulting Tobit estimates are very similar to those from our original unweighted analysis, and consequently the new estimates of λ_t and θ_t, shown as the dashed lines in figures 4 and 5, are again very similar to the main estimates.

To summarize, all of these analyses tell the same story. Contrary to the textbook errors-in-variables model usually assumed in applied research, the slope coefficient in the regression of log current earnings on log annual earnings varies systematically over the life cycle and is not generally equal to 1. Any study that uses current earnings to proxy for long-run earnings should consider the implications of this pattern for errors-in-variables bias, which were detailed above in section II.B. The next subsection illustrates with an application to the empirical literature on intergenerational earnings mobility.

C. Application to intergenerational earnings mobility

We can make the lessons of our results more concrete by applying them to the intergenerational mobility regression in which son's log of lifetime earnings is the dependent variable and father's log of lifetime earnings is the explanatory variable. As

[11] This series is available only back to 1954. For 1951-1953, we added 0.003 to the interest rates for three-month T-bills. This adjustment was based on the relationship between the one-year and three-month rates observed for 1954-1959.

mentioned in section II.A, the intergenerational mobility literature has exhibited a systematic tendency to estimate lower intergenerational elasticities when the sons' earnings are measured early in the life cycle. Reville (1995), for example, estimates elasticities of about 0.25 when he measures the sons' earnings in their twenties, but his estimates start approaching 0.5 when he observes the sons well into their thirties. This is just the pattern one should expect from the trajectories of $\hat{\lambda}_t$ in figures 2 and 4. Our results indicate that measuring sons' earnings at around age 40 should avoid any serious errors-in-variables bias from mismeasurement of log lifetime earnings as the dependent variable. Our results also suggest that, once the longitudinal surveys on which the intergenerational studies are based have followed the sons well past age 40, using their current earnings later in the life cycle could cause an amplification bias in the estimation of the intergenerational elasticity of lifetime earnings.

Most intergenerational studies following Solon (1989) have emphasized that using one year of father's log annual earnings as a proxy for father's log lifetime earnings as the explanatory variable induces a large attenuation bias. This conclusion is strongly supported by the estimates of the reliability ratio θ_t depicted in figures 3 and 5, which never exceed 0.6 and are much lower early in the life cycle. In an effort to reduce this errors-in-variables bias, many researchers – such as Altonji and Dunn (1991), Solon (1992), and Zimmerman (1992) – have averaged father's log earnings over multiple years.[12] To explore the extent to which such averaging corrects the bias, in figure 6 we repeat the analysis in figure 3 except that the new estimates of θ_t are for five-year

[12] These same authors and others also have used instrumental variables estimation strategies to address the errors-in-variables issue. As noted above in footnote 4, our results are pertinent for interpreting IV estimates as well.

averages of log annual earnings, rather than for single years. For example, the observation plotted for age 30 is based on a five-year average for ages 28-32.

As expected, the trajectory in figure 6 is higher than in figure 3. Nevertheless, although the estimates of θ_t reach close to 0.6 over a wide age range from about 29 to 48, they never exceed 0.6 by much. This finding strongly supports the conclusion of Mazumder (2001, 2003) that even five-year averages of the earnings variable for fathers are subject to a large errors-in-variables bias.

Of course, the ideal solution would be to observe lifetime earnings for both generations. In the more usual case in which lifetime data are unavailable, our estimates of λ_t and θ_s can be used to correct for errors-in-variables bias. For example, suppose an intergenerational study observes sons at an age t when $\hat{\lambda}_t = 0.8$ (e.g., around age 30, as has been common in intergenerational mobility studies) and fathers at an age s when $\hat{\theta}_s = 0.6$. Then, using equation (14), one can correct the errors-in-variables bias in $\hat{\beta}$ by dividing the initial estimate through by $0.8 \times 0.6 = 0.48$. This is essentially one of the approaches already used by several intergenerational mobility researchers (Altonji and Dunn, 1991; Zimmerman, 1992; Mazumder, 2001) except that they typically have overlooked departures of λ_t from 1 and have obtained their estimates of θ_s by combining longitudinal earnings data from limited segments of the life cycle with tightly specified parametric models of earnings dynamics.[13]

[13] Zimmerman (1992) models log annual earnings as the sum of a permanent component and a transitory component that follows a stationary first-order autoregressive process. Altonji and Dunn (1991) assume the transitory component follows a second-order moving average process. Mazumder (2001) uses the more elaborate earnings dynamics model developed by Baker and Solon (2003).

We caution readers, however, about several problems in applying our estimates of λ_t and θ_s to other sets of earnings data. To begin with, the life-cycle trajectories for our U.S. cohort born in 1931-1933 may differ from those for other cohorts and other countries. Also, as emphasized in Solon (1992), sample selection criteria that affect the sample's dispersion in earnings also affect the measurement error properties of current earnings as proxies for lifetime earnings. Furthermore, patterns observed for administrative data on earnings covered by Social Security cannot be exactly applicable to other earnings data, such as survey reports in the Panel Study of Income Dynamics. As discussed in section IV.A, the mean-reversion in survey reports of transitory earnings variation makes this discrepancy less serious than it otherwise would be, but surely does not eliminate it.

V. Summary

Many researchers have recognized that current income variables are error-ridden proxies for lifetime income variables, but their analyses of the resulting biases typically have assumed the textbook errors-in-variables model, in which the regression of the error-ridden variable on the true one has a slope coefficient of 1. Using nearly career-long earnings histories from administrative Social Security data, we have documented the association between annual earnings at each age and the present discounted value of lifetime earnings. Our results show that this association departs substantially from the textbook model in ways that vary systematically over the life cycle. We have illustrated the use of our results in correcting for errors-in-variables biases in the estimation of intergenerational earnings elasticities, but the results are more broadly applicable to any setting in which annual earnings variables are used as proxies for lifetime earnings.

Table 1. Descriptive Statistics for Nominal Annual Earnings Covered by Social Security

Year	Median	Number with Zero Earnings	Taxable Limit	Number at Taxable Limit
1951	200	305	3,600	9
1952	200	337	3,600	24
1953	100	378	3,600	53
1954	200	357	3,600	83
1955	1,300	239	4,200	99
1956	2,200	179	4,200	164
1957	3,000	87	4,200	243
1958	3,100	98	4,200	288
1959	3,800	83	4,800	268
1960	4,100	92	4,800	325
1961	4,200	98	4,800	363
1962	4,700	90	4,800	406
1963	4,800	86	4,800	442
1964	4,800	80	4,800	463
1965	4,800	85	4,800	511
1966	6,600	70	6,600	413
1967	6,600	74	6,600	439
1968	7,500	63	7,800	392
1969	7,800	75	7,800	435
1970	7,800	74	7,800	464
1971	7,800	78	7,800	481
1972	9,000	87	9,000	454
1973	10,700	83	10,800	407
1974	11,400	85	13,200	324
1975	11,700	94	14,100	305
1976	12,600	93	15,300	305
1977	13,400	98	16,500	298
1978	15,000	100	17,700	328
1979	15,800	108	22,900	219
1980	16,500	111	25,900	185
1981	17,700	119	29,700	154
1982	17,800	140	32,400	141
1983	17,800	148	35,700	121
1984	18,900	159	37,800	124
1985	19,900	160	39,600	117
1986	19,100	165	42,000	110
1987	19,800	178	43,800	104
1988	20,000	182	45,000	112
1989	18,700	194	48,000	100
1990	18,200	194	51,300	85
1991	15,900	224	53,400	75

Note: These descriptive statistics are for our main sample of 821 men from the Health and Retirement Study.

Table 2. Estimated Autocorrelations in Log Annual Earnings, 1975-1984

Year	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
1975	1	.75	.70	.75	.61	.66	.64	.56	.51	.52
		(.02)	(.02)	(.02)	(.03)	(.02)	(.02)	(.03)	(.03)	(.03)
1976		1	.71	.67	.62	.62	.61	.52	.49	.44
			(.02)	(.02)	(.03)	(.02)	(.02)	(.03)	(.03)	(.04)
1977			1	.82	.73	.67	.64	.50	.52	.54
				(.01)	(.02)	(.02)	(.02)	(.03)	(.03)	(.03)
1978				1	.82	.69	.68	.60	.56	.67
					(.01)	(.02)	(.02)	(.03)	(.03)	(.02)
1979					1	.80	.78	.64	.65	.59
						(.01)	(.01)	(.02)	(.02)	(.03)
1980						1	.84	.80	.74	.72
							(.01)	(.01)	(.02)	(.02)
1981							1	.84	.68	.73
								(.01)	(.02)	(.02)
1982								1	.76	.73
									(.01)	(.02)
1983									1	.82
										(.01)
1984										1

Note: Numbers in parentheses are estimated standard errors.

Table 3. Average Estimated Autocorrelations from Various Studies

Order of Autocorrelation	Our Table 2	Baker and Solon (2003), Table 3	Baker (1997), Table 1	Haider (2001), Table 3	Our W-2 Data for 1980-1991	Our Social Security Data for 1980-1991
1	.80	.80	.81	.78	.83	.81
2	.72	.72	.74	.71	.77	.74
3	.69	.69	.70	.67	.72	.71
4	.64	.66	.67	.65	.68	.66
5	.58	.64	.64	.62	.64	.63
6	.59	.62	.62	.59	.61	.60

Table 4. Estimated Autocorrelations in Log Annual Earnings, 1951-1960

Year	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
1951	1	.47	.27	.20	.29	.09	.09	.04	.08	.06
		(.04)	(.06)	(.06)	(.05)	(.07)	(.06)	(.07)	(.06)	(.07)
1952		1	.38	.29	.40	.25	.10	.10	.19	.17
			(.05)	(.06)	(.05)	(.06)	(.06)	(.06)	(.06)	(.07)
1953			1	.56	.43	.34	.10	.13	.16	.19
				(.04)	(.05)	(.06)	(.07)	(.07)	(.07)	(.06)
1954				1	.50	.28	.18	.21	.21	.12
					(.04)	(.06)	(.06)	(.05)	(.05)	(.07)
1955					1	.51	.32	.21	.20	.16
						(.04)	(.05)	(.05)	(.06)	(.07)
1956						1	.33	.26	.27	.19
							(.04)	(.05)	(.04)	(.06)
1957							1	.60	.48	.32
								(.03)	(.04)	(.05)
1958								1	.64	.46
									(.02)	(.03)
1959									1	.67
										(.02)
1960										1

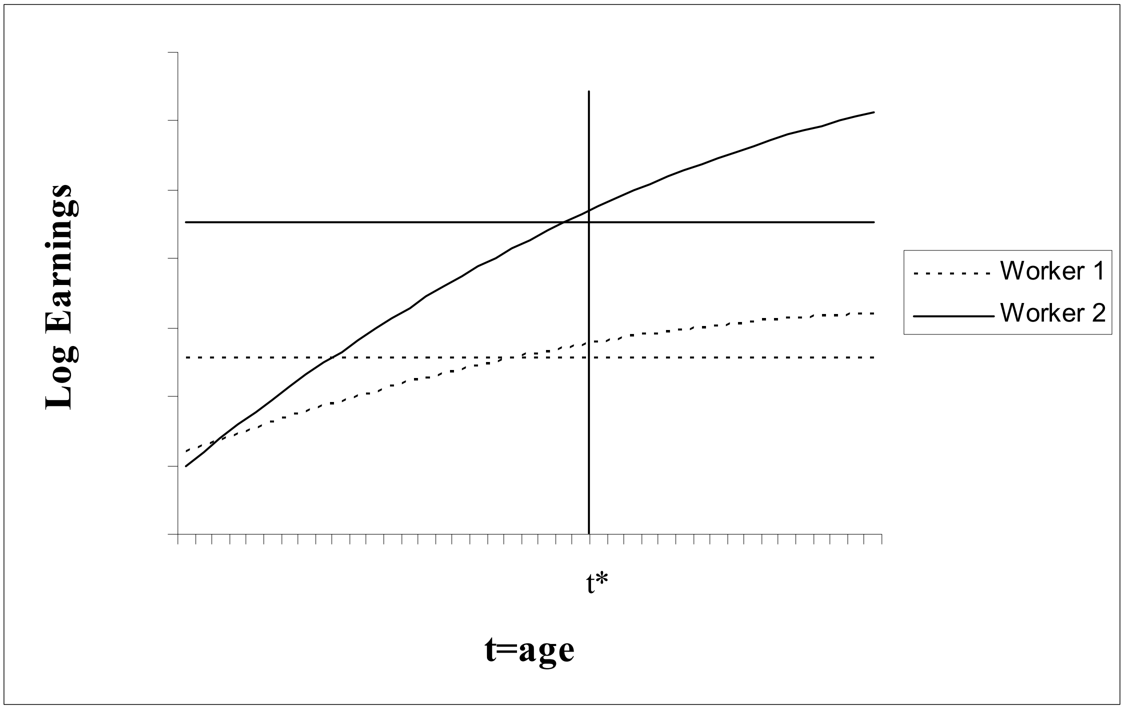
Note: Numbers in parentheses are estimated standard errors.

Table 5. Coefficients of Variation for Earnings Variables from Simulations

Year	Coefficient of Variation	Year	Coefficient of Variation
1951	1.33	1972	1.05
1952	1.21	1973	0.98
1953	1.27	1974	1.20
1954	1.57	1975	1.33
1955	1.17	1976	1.27
1956	1.05	1977	1.24
1957	0.94	1978	1.43
1958	0.98	1979	1.07
1959	1.05	1980	1.49
1960	0.99	1981	1.40
1961	0.96	1982	1.34
1962	0.94	1983	1.46
1963	0.98	1984	1.26
1964	0.89	1985	1.43
1965	0.94	1986	1.69
1966	1.07	1987	1.22
1967	0.97	1988	1.09
1968	1.13	1989	1.23
1969	0.93	1990	1.22
1970	0.92	1991	1.35
1971	1.33	Lifetime	0.67

.

Figure 1. Illustrative Example of Log Annual Earnings and Log Annuitized Lifetime Earnings



Notes: The dotted lines are for worker 1, and the solid lines are for worker 2. For each worker, the upward-sloping line depicts log annual earnings by age, and the horizontal line depicts log annuitized lifetime earnings.

Figure 2. Main Estimates of λ_t

Notes: The solid line graphs the parameter estimates, and the dotted lines are 1.96 estimated standard errors above and below the solid line. The standard error estimates come from the bootstrap procedure described in footnote 10.

Figure 3. Main Estimates of θ_t

Notes: The solid line graphs the parameter estimates, and the dotted lines are 1.96 estimated standard errors above and below the solid line. The standard error estimates come from the bootstrap procedure described in footnote 10.

Figure 4. Alternative Estimates of λ_t

Notes: The plotted estimates are from five different analyses:

Main – main estimates copied from figure 2

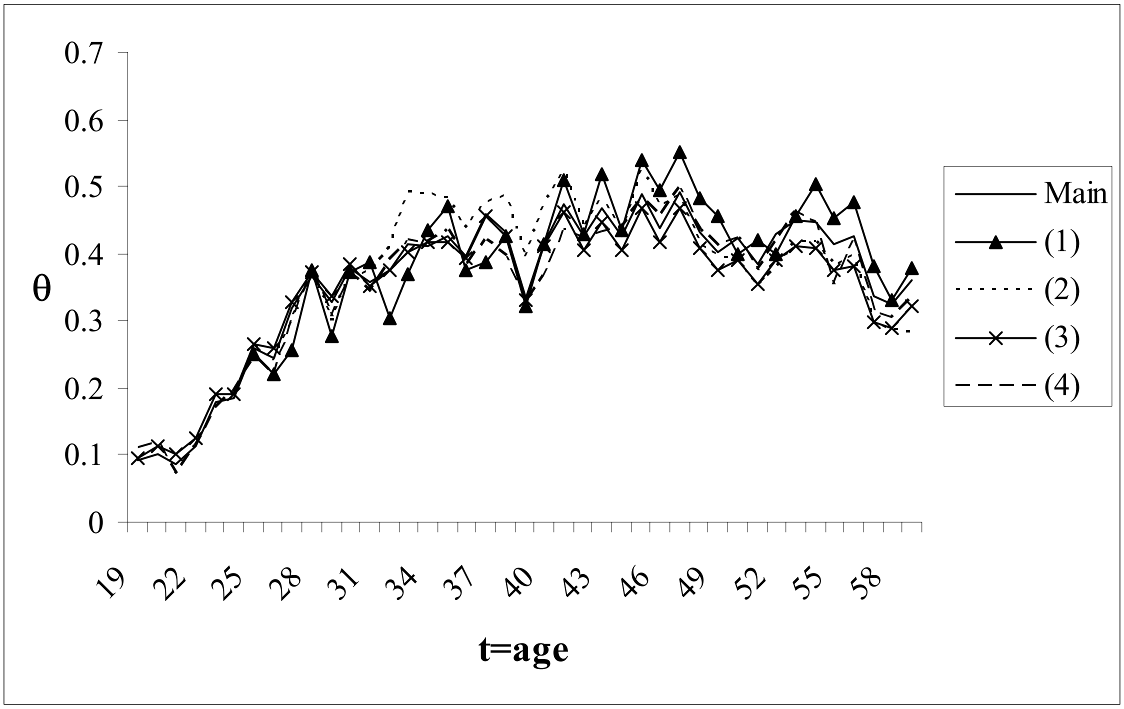
(1) – same as main, but dropping zeros and estimating one-limit Tobits

(2) – same as main, but censoring top 40% of earnings distribution

(3) – same as main, but discounting with one-year T-note interest rates

(4) – same as main, but weighting by inverse probabilities of selection

Figure 5. Alternative Estimates of θ_t



Notes: The plotted estimates are from five different analyses:

Main – main estimates copied from figure 3

(1) – same as main, but dropping zeros and estimating one-limit Tobits

(2) – same as main, but censoring top 40% of earnings distribution

(3) – same as main, but discounting with one-year T-note interest rates

(4) – same as main, but weighting by inverse probabilities of selection

Figure 6. Estimates of θ_t for Five-Year Averages

Notes: The solid line, copied from figure 3, plots the estimated reliability ratios for log earnings at each age. The dotted line plots the estimated reliability ratios for five-year averages of log earnings centered on each age.

References

Abowd, John M. and Card, David. "On the Covariance Structure of Earnings and Hours Changes." *Econometrica*, March 1989, 57(2), pp. 411-45.

Abul Naga, Ramses H. "Estimating the Intergenerational Correlation of Incomes: An Errors-in-Variables Framework." *Economica*, February 2002, 69(273), pp. 69-91.

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

Baker, Michael. "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings." *Journal of Labor Economics*, April 1997, 15(2), pp. 338-75.

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

Bjorklund, Anders. "A Comparison between Actual Distributions of Annual and Lifetime Income: Sweden 1951-89." *Review of Income and Wealth*, December 1993, 39(4), pp. 377-86.

Bound, John, Brown, Charles and Mathiowetz, Nancy. "Measurement Error in Survey Data," in James J. Heckman and Edward Leamer, eds., *Handbook of Econometrics*, Vol. 5. Amsterdam: North-Holland, 2001, pp. 3705-843.

Bound, John and Krueger, Alan B. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics*, January 1991, 9(1), pp. 1-24.

Bound, John and Solon, Gary. “Double Trouble: On the Value of Twins-Based Estimation of the Return to Schooling.” *Economics of Education Review*, April 1999, 18(2), pp. 169-82.

Chadwick, Laura and Solon, Gary. “Intergenerational Income Mobility among Daughters.” *American Economic Review*, March 2002, 92(1), pp. 335-44.

Dunn, Christopher. “Intergenerational Earnings Mobility in Brazil.” Unpublished, University of Michigan, 2003.

Friedman, Milton. *A theory of the consumption function*. Princeton: Princeton University Press, 1957.

Gottschalk, Peter and Moffitt, Robert. “The Growth of Earnings Instability in the U.S. Labor Market.” *Brookings Papers on Economic Activity*, 2:1994, pp. 217-54.

Grawe, Nathan D. “Life Cycle Bias in the Estimation of Intergenerational Earnings Persistence.” Working Paper No. 207, Analytical Studies Branch, Statistics Canada, 2003.

Haider, Steven J. “Earnings Instability and Earnings Inequality of Males in the United States: 1967-1991.” *Journal of Labor Economics*, October 2001, 19(4), pp. 799-836.

Haider, Steven and Solon, Gary. “Nonrandom Selection in the HRS Social Security Earnings Sample.” Working Paper No. 00-01, RAND Labor and Population Program, 2000.

Jenkins, Stephen. “Snapshots versus Movies: ‘Lifecycle Biases’ and the Estimation of Intergenerational Earnings Inheritance.” *European Economic Review*, July 1987, 31(5), pp. 1149-58.

Juster, F. Thomas and Suzman, Richard. "An Overview of the Health and Retirement Study." *Journal of Human Resources*, 1995, 30(Supplement), pp. S7-56.

Kane, Thomas J., Rouse, Cecilia Elena and Staiger, Douglas. "Estimating Returns to Schooling When Schooling Is Misreported." Working Paper No. 6721, National Bureau of Economic Research, 1999.

Kim, Bonggeun and Solon, Gary. "Implications of Mean-Reverting Measurement Error for Longitudinal Studies of Wages and Employment." *Review of Economics and Statistics*, forthcoming.

Lillard, Lee A. "Inequality: Earnings vs. Human Wealth." *American Economic Review*, March 1977, 67(2), pp. 42-53.

Mazumder, Bhashkar. "The Mis-Measurement of Permanent Earnings: New Evidence from Social Security Earnings Data." Working Paper No. 2001-24, Federal Reserve Bank of Chicago, 2001.

Mazumder, Bhashkar. "Fortunate Sons: New Estimates of Intergenerational Mobility in the U.S. Using Social Security Earnings Data." Unpublished, Federal Reserve Bank of Chicago, 2003.

Mincer, Jacob. *Schooling, experience, and earnings*. New York: National Bureau of Economic Research, 1974.

Modigliani, Franco and Brumberg, Richard. "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data," in K. K. Kurihara, ed., *Post-Keynesian economics*. New Brunswick: Rutgers University Press, 1954, pp. 88-436.

Pischke, Jorn-Steffen. "Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study." *Journal of Business and Economic Statistics*, July 1995, 13(3), pp. 305-14.

Reville, Robert T. "Intertemporal and Life Cycle Variation in Measured Intergenerational Earnings Mobility." Unpublished, RAND, 1995.

Shorrocks, A. F. "Income Stability in the United States," in N. A. Klevmarken and J. A. Lybeck, eds., *The statics and dynamics of income*. Clevedon: Tieto, 1981, pp. 175-94.

Social Security Administration. *Annual statistical supplement to the Social Security Bulletin*. Washington, DC: Social Security Administration, 1999.

Solon, Gary. "Biases in the Estimation of Intergenerational Earnings Correlations." *Review of Economics and Statistics*, February 1989, 71(1), pp. 172-4.

Solon, Gary. "Intergenerational Income Mobility in the United States." *American Economic Review*, June 1992, 82(3), pp. 393-408.

Solon, Gary. "Intergenerational Mobility in the Labor Market," in Orley C. Ashenfelter and David Card, eds., *Handbook of labor economics*, Vol. 3A. Amsterdam: North-Holland, 1999, pp. 1761-800.

Zimmerman, David J. "Regression toward Mediocrity in Economic Stature." *American Economic Review*, June 1992, 82(3), pp. 409-29.