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

By STEVEN HAIDER AND GARY SOLON*

In the year 2003 alone, the *American Economic Review* published 14 articles reporting regression analyses involving individual or family income variables, and the May *Papers and 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 Franco Modigliani and Richard Brumberg (1954) and Milton Friedman (1957), which analyzed the properties of consumption functions estimated with current rather than permanent income variables as the regressors. Another instance is the literature (e.g., Lee A. Lillard, 1977) suggesting that inequality as measured in cross sections of annual earnings overstates the inequality in lifetime earnings. A recent off-shoot of that literature—exemplified by Peter Gottschalk and Robert Moffitt (1994), Haider (2001), and Michael 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 dataset, 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 Administration (SSA) earnings histories of the members of the Health and Retirement Study (HRS) sample. Despite some limitations discussed in Section II, these data provide nearly career-long earnings histories, which are based on relatively accurate administrative data and pertain to a broadly representative national sample. In Section I, we develop simple models to illustrate some important aspects of the association between current and

[1] See Bhashkar Mazumder (2001) for a relatively sophisticated recent example.

lifetime earnings and to demonstrate the implications for errors-in-variables biases in applied econometric research. In Section II, we describe the dataset and our econometric methods. In Section III, we present our evidence on the connections between annual and lifetime earnings. Section IV summarizes our findings and illustrates their usefulness with a brief application to intergenerational earnings mobility.

I. 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:

$$(1) \qquad y_{it} = y_i + v_{it},$$

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 v_{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 one. 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 $\text{Var}(y_i)/[\text{Var}(y_i) + \text{Var}(v_{it})]$.

These oft-used results no longer apply if the textbook errors-in-variables model incorrectly characterizes the relationship between current and lifetime income. In Section IA, 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 one. In Section IB, 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 changes over the life cycle. Anders Björklund (1993), the closest predecessor to our study, uses Swedish income tax data from 1951 to 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 (1993, p. 381), "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 Björklund 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 association between current and lifetime earnings that do have direct implications for errors-in-variables biases.

Another indication of life-cycle-related departures from the textbook errors-in-variables model, noted by Stephen Jenkins (1987) and Nathan D. Grawe (forthcoming), involves the estimation of intergenerational mobility models such as the regression of sons' log lifetime earnings on fathers' log lifetime earnings. If sons' 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 sons' earnings were observed. On the other hand, if the slope coefficient in the regression of sons' log annual earnings on sons' log lifetime earnings deviates from one 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 (e.g., Robert T. Reville, 1995) 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. As an extremely simple example, suppose that y_{it}, the log real earnings of worker i in year t of his career, follows

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

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 Jacob Mincer (1974), Baker (1997), Haider (2001), and Baker and Solon (2003), among others. For simplicity, assume zero covariance between α_i and γ_i, infinite lifetimes, and a constant real interest rate $r > \gamma_i$. Then the present discounted value of lifetime earnings is

$$(3) \qquad V_i = \sum_{s=0}^{\infty} \exp(\alpha_i + \gamma_i s)(1 + r)^{-s}$$

$$\cong \exp(\alpha_i)[(1 + r)/(r - \gamma_i)],$$

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

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

It follows that the slope coefficient in the re-

gression of current log earnings on the log of the present value of lifetime earnings is

$$(5) \qquad \lambda_t \equiv \frac{\mathrm{Cov}(\log V_i, y_{it})}{\mathrm{Var}(\log V_i)} \cong \frac{\sigma_\alpha^2 + (t\sigma_\gamma^2/r)}{\sigma_\alpha^2 + (\sigma_\gamma^2/r^2)}.$$

The main thing to note about this result is that, contrary to the textbook errors-in-variables model, λ_t generally does not equal one. Instead, it starts at a value less than one at the outset of the career and then increases monotonically over the life cycle. It reaches one when $t = 1/r$ and then exceeds one afterward. 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 may 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 (5) 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 one 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 through most of the life cycle. 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
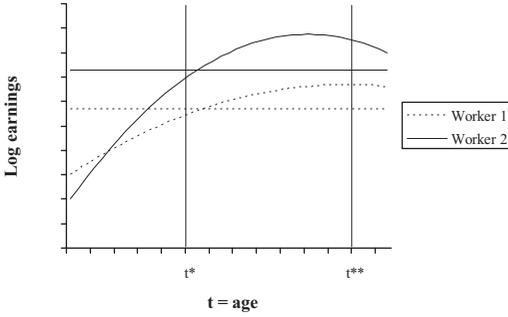
FIGURE 1. ILLUSTRATIVE EXAMPLE OF LOG ANNUAL
EARNINGS AND LOG ANNUITIZED LIFETIME EARNINGS

Note: For each worker, the upward-sloping line depicts log
annual earnings by age, and the horizontal line depicts log
annuitized lifetime earnings.

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, 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. For at least some of the life cycle beyond that age, the gap in current earnings tends to *overstate* the gap in lifetime earnings.

B. Implications for Errors-in-Variables Biases

Suppose we wish to estimate the regression model

$$(6) \qquad y_i = \boldsymbol{\beta}'\mathbf{X}_i + \varepsilon_i,$$

where the error term ε_i is uncorrelated with the regressor vector $\mathbf{X}_i$. Starting with the case of left-side measurement error, suppose that y_i is the log of lifetime earnings, which is not observed and hence is proxied by 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

$$(7) \qquad y_{it} = \lambda_t y_i + v_{it},$$

where λ_t, the slope coefficient in the linear projection of y_{it} on y_i, need not equal one and may vary over the life cycle. By construction, v_{it} is uncorrelated with y_i, and we will continue to maintain the textbook model's assumption that it also is uncorrelated with each separate determinant of y_i ($\mathbf{X}_i$ and ε_i).[2] Then, if OLS is applied to the regression of y_{it} on $\mathbf{X}_i$,

$$(8) \qquad y_{it} = \lambda_t \boldsymbol{\beta}'\mathbf{X}_i + (\lambda_t\varepsilon_i + v_{it}),$$

the probability limit of the estimated coefficient vector for $\mathbf{X}_i$ is $\lambda_t\boldsymbol{\beta}$ instead of $\boldsymbol{\beta}$. In the textbook case where $\lambda_t = 1$, measurement error in the dependent variable does not result in inconsistent estimation of $\boldsymbol{\beta}$. More generally, however, the OLS estimator is inconsistent, and the inconsistency varies as a function of the age at which annual earnings are observed.

Moving on to the case of right-side measurement error, suppose that the log of lifetime earnings is one element x_i in the regressor vector $\mathbf{X}_i$. Because x_i is not observed, it is proxied by x_{it}, log annual earnings at age t. Analogously to equation (7) for y_{it}, we express the linear projection of x_{it} on x_i as

$$(9) \qquad x_{it} = \lambda_t x_i + v_{it},$$

where v_{it} again is assumed to be uncorrelated with $\mathbf{X}_i$ and ε_i. If x_i is the only element in $\mathbf{X}_i$ and OLS is applied to the linear regression of y_i on x_{it}, the probability limit of the estimated slope coefficient is

$$(10) \qquad \text{plim } \hat{\boldsymbol{\beta}} = \frac{\text{Cov}(x_{it}, y_i)}{\text{Var}(x_{it})} = \theta_t\boldsymbol{\beta},$$

where

$$(11) \qquad \theta_t \equiv \frac{\text{Cov}(x_{it}, x_i)}{\text{Var}(x_{it})} = \frac{\lambda_t\text{Var}(x_i)}{\lambda_t^2\text{Var}(x_i) + \text{Var}(v_{it})}.$$

[2] When this assumption fails, as it sometimes does, neither the textbook analysis nor our extension is applicable. When $\text{Var}(v_{it}) = 0$, equation (7) specializes to the rescaling of variables often discussed in introductory econometrics textbooks (e.g., section 2.4 of Jeffrey M. Wooldridge, 2006). See section 4 of Joshua D. Angrist and Alan B. Krueger (1999) for an excellent overview of errors in variables, including nonclassical measurement error.

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

Two further results about right-side measurement error are worth noting. First, if x_i is just one element in the regressor vector $\mathbf{X}_i$, the attenuation factor for its estimated coefficient is the same as the last expression in equation (11), except with $\text{Var}(x_i)$ replaced by the residual variance from the auxiliary regression of x_i on the other regressors in $\mathbf{X}_i$. Second, if the measurement error in x_{it} as a proxy for x_i is treated with an instrumental variable (IV) correlated with x_i but uncorrelated with ε_i or v_{it}, the probability limit of the conventional IV estimator of the coefficient of x_i is the coefficient divided by λ_t.[3]

The results presented in this subsection deliver two key messages. First, with plausible departures from the textbook errors-in-variables assumptions, the familiar textbook results about OLS and IV 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 inconsistency in OLS estimation as well as in IV estimation. Second, some of the estimation inconsistencies 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. In Section III, we will estimate those two parameters and examine how they vary over the life cycle.

II. 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 Björklund'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 SSA, the University of Michigan's Survey Research Center asked the participants in its HRS to permit access to their Social Security earnings histories for 1951–1991.[4] The HRS sample is a national probability sample of Americans born between 1931 and 1941, and about three-fourths 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 who granted access appear to be surprisingly representative of the complete sample. The earnings data supplied by the SSA 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 for every year over the major portion of their careers.[5] The other main strength of our dataset is that the Social Security

---

[3] The inconsistency of conventional IV estimation in the presence of nonclassical measurement error has been discussed previously by Thomas J. Kane et al. (1999), John Bound and Solon (1999), and Bonggeun Kim and Solon (2005).

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

[5] This sample is restricted to workers with positive earnings in at least ten years during the 1951–1991 period. This criterion, which excludes only 33 individuals, is less restrictive than the usual practice in survey-based earnings dynamics studies of requiring positive earnings in *every* year (e.g., John M. Abowd and David 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.

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 (CPS) used Social Security earnings data as the "true" values against which the CPS 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 SSA (1999, Table 3, B2), the percentage of earnings covered by Social Security has exceeded 80 percent ever since the coverage extensions effected by 1957, and exceeded 85 percent over most of our sample period. Between 1951 and 1956, however, this percentage ranged between 66 percent and 79 percent. Accordingly, in addition to our analyses 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 percentage in the sample with zero earnings, the taxable limit, and the percentage 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 when the taxable limit is low relative to the general earnings distribution. The worst year is 1965, when 62 percent of the sample is right-censored. Afterward, the degree of censorship lessens as the taxable limit is progressively increased. By 1991, only 9 percent 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 Björklund's (1993) approach of directly summarizing the observed joint distribution of

TABLE 1—DESCRIPTIVE STATISTICS FOR NOMINAL ANNUAL EARNINGS COVERED BY SOCIAL SECURITY

Year	Median	Percent with zero earnings	Taxable limit	Percent at taxable limit
1951	200	37.1	3,600	1.1
1952	200	41.0	3,600	2.9
1953	100	46.0	3,600	6.5
1954	200	43.5	3,600	10.1
1955	1,300	29.1	4,200	12.1
1956	2,200	21.8	4,200	20.0
1957	3,000	10.6	4,200	29.6
1958	3,100	11.9	4,200	35.1
1959	3,800	10.1	4,800	32.6
1960	4,100	11.2	4,800	39.6
1961	4,200	11.9	4,800	44.2
1962	4,700	11.0	4,800	49.5
1963	4,800	10.5	4,800	53.8
1964	4,800	9.7	4,800	56.4
1965	4,800	10.4	4,800	62.2
1966	6,600	8.5	6,600	50.3
1967	6,600	9.0	6,600	53.5
1968	7,500	7.7	7,800	47.7
1969	7,800	9.1	7,800	53.0
1970	7,800	9.0	7,800	56.5
1971	7,800	9.5	7,800	58.6
1972	9,000	10.6	9,000	55.3
1973	10,700	10.1	10,800	49.6
1974	11,400	10.4	13,200	39.5
1975	11,700	11.4	14,100	37.1
1976	12,600	11.3	15,300	37.1
1977	13,400	11.9	16,500	36.3
1978	15,000	12.2	17,700	40.0
1979	15,800	13.2	22,900	26.7
1980	16,500	13.5	25,900	22.5
1981	17,700	14.5	29,700	18.8
1982	17,800	17.1	32,400	17.2
1983	17,800	18.0	35,700	14.7
1984	18,900	19.4	37,800	15.1
1985	19,900	19.5	39,600	14.3
1986	19,100	20.1	42,000	13.4
1987	19,800	21.7	43,800	12.7
1988	20,000	22.2	45,000	13.6
1989	18,700	23.6	48,000	12.2
1990	18,200	23.6	51,300	10.4
1991	15,900	27.3	53,400	9.1

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 IB, 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; 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 one, 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.[6]

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 4,000 random draws from the estimated joint distribution of the 41 years of annual earnings.[7] Then, for each of the 4,000 simulated earnings histories, we calculate the present discounted value of lifetime earnings. In the main version of the simulation, we perform the discounting by (a) using the personal consumption expenditures deflator to convert each year's nominal earnings to a real value and (b) assuming a constant real interest rate of 0.02. In the end, we have a simulated sample of 4,000 observations for which we observe the present discounted value of lifetime earnings as well as each year's earnings.

Finally, for this sample of 4,000 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

---

[6] 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.

[7] To implement the simulation, we need the estimated autocovariance matrix to be positive semidefinite (as the true one must be). Our procedure for imposing the restriction of positive semidefiniteness is described on the *AER* Web site (www.e-aer.org/data/sept06/20040284_data.zip).

TABLE 2—ESTIMATED AUTOCORRELATIONS IN LOG ANNUAL EARNINGS, 1975–1984

Year	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
1975	1	0.92 (0.01)	0.82 (0.01)	0.79 (0.01)	0.77 (0.01)	0.74 (0.01)	0.69 (0.02)	0.65 (0.02)	0.62 (0.02)	0.59 (0.02)
1976		1	0.88 (0.01)	0.83 (0.01)	0.79 (0.01)	0.77 (0.01)	0.73 (0.02)	0.69 (0.02)	0.66 (0.02)	0.64 (0.02)
1977			1	0.91 (0.01)	0.84 (0.01)	0.77 (0.01)	0.77 (0.01)	0.70 (0.02)	0.68 (0.02)	0.69 (0.02)
1978				1	0.88 (0.01)	0.80 (0.01)	0.78 (0.01)	0.72 (0.02)	0.70 (0.02)	0.69 (0.02)
1979					1	0.87 (0.01)	0.84 (0.01)	0.78 (0.01)	0.73 (0.02)	0.71 (0.02)
1980						1	0.89 (0.01)	0.83 (0.01)	0.76 (0.01)	0.73 (0.02)
1981							1	0.87 (0.01)	0.82 (0.01)	0.78 (0.01)
1982								1	0.89 (0.01)	0.82 (0.01)
1983									1	0.87 (0.01)
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	Bohlmark and Lindquist (2005), Table 2	Our W-2 data for 1980–1991	Our Social Security data for 1980–1991
1	0.89	0.82	0.84	0.89	0.91
2	0.82	0.75	0.79	0.83	0.85
3	0.78	0.71	0.80	0.79	0.81
4	0.75	0.69	0.73	0.74	0.77
5	0.72	0.65	0.72	0.71	0.73
6	0.69	0.62	0.69	0.67	0.70

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.

III. Empirical Results

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. The full matrix is available on the *AER* Web site (www.e-aer.org/data/sept06/20040284_data.zip). Table 2 shows the estimated autocorrelations

for 1975–1984, a period when our cohort born 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 0.89, the second-order autocorrelations average to 0.82, the third-order autocorrelations average to 0.78, and so forth. Table 3 also displays estimated autocorrelations from two other studies of administrative data. The most comparable results reported in Baker and Solon's (2003) study of Canadian income tax data are the autocorrelations over the 1985–1992 period for the cohort born 1942–1943. Their average autocorrelations, shown in the third column, are fairly similar to ours, but somewhat lower. As shown in the fourth column, the estimates

from Anders Bohlmark and Matthew J. Lindquist (2005), a replication of our study based on Swedish income tax data, are closer to ours than to Baker and Solon's. Note that this resemblance between other studies' estimates and ours occurs even though the other studies use uncensored data and therefore can estimate the autocorrelations directly without imposing distributional assumptions.[8] We find it reassuring that, despite the omission of earnings not covered by Social Security and the imputation of right-censored values, our autocorrelation estimates are similar to those from other datasets. Most of these estimated autocorrelations are somewhat higher than those reported by Baker (1997) and Haider (2001) in their analyses of the Panel Study of Income Dynamics, but the survey-based estimates may be biased downward by reporting error.

Another relevant comparison is to an alternative earnings variable available for our sample for 1980–1991. 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 fifth column of Table 3, when we use the W-2 data to reestimate our Tobits for 1980–1991, the first-order autocorrelations average to 0.89, the second-order autocorrelations average to 0.83, and the third-order autocorrelations average to 0.79. As shown in the last column, the corresponding average autocorrelations for the Social Security earnings variable over the same period are 0.91, 0.85, and 0.81. The idiosyncrasies of the alternative earnings measures do not appear to generate major discrepancies in the estimated persistence of earnings.[9]

FIGURE 2. MAIN ESTIMATES OF λ_t AND θ_t

Note: The solid lines graph the parameter estimates, and the dotted lines are 1.96 estimated standard errors above and below the solid lines.

In the second step of our estimation procedure, we perform the simulation in which we take 4,000 draws from the estimated joint distribution of the 41 years of annual earnings. Then, using the resulting sample of 4,000 uncensored earnings histories, our third step summarizes the connection between annual and lifetime earnings by estimating the forward and reverse regressions between the logs of annual and lifetime earnings.

The top portion of Figure 2 plots our estimates of λ_t, the slope coefficient in the regression of log annual earnings at time t on the log

---

[8] Also like these other studies, we find that earnings autocorrelations are somewhat smaller early in the life cycle.

[9] Perhaps the similarity of the autocorrelation estimates should not be a surprise. If one thinks of the log of covered
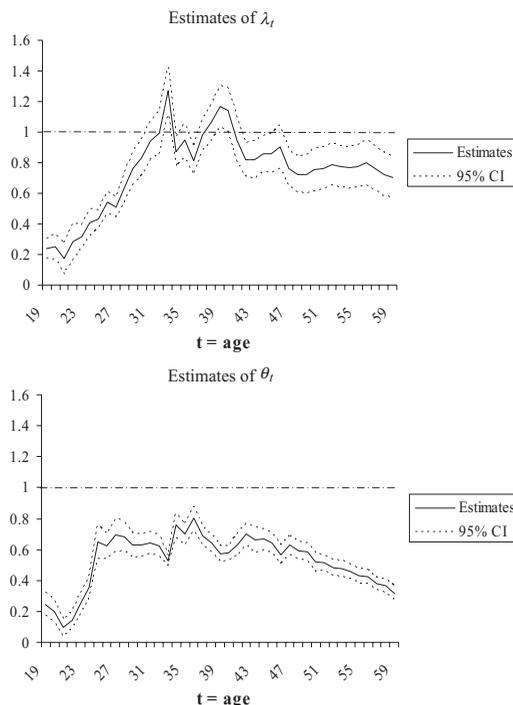
earnings as the sum of log total earnings and the log of the proportion covered, one would expect the autocorrelation of log covered earnings to be approximately a weighted average of the autocorrelations for log total earnings and log coverage. Presumably, both of these autocorrelations are highly positive. 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.

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.[10] In contrast to the textbook assumption that λ_t equals 1 throughout the life cycle, $\hat{\lambda}_t$ begins at 0.24 at age 19, increases steadily until it rises to about 1 at age 32, and then declines some in the late forties. 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 can induce an errors-in-variables bias. Most importantly, using current earnings in the twenties causes a large attenuation bias. A constructive implication is that the bias is small if one uses current earnings between the early thirties and the mid-forties, when the textbook assumption that $\lambda_t = 1$ is reasonably accurate.

The lower portion of Figure 2 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 only about 0.2, increases to a fairly flat peak averaging about 0.65 between the late twenties and mid-forties, and then decreases. Our discussion in Section IB 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 that using current earnings to proxy for lifetime earnings as a regressor induces an 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.

To check the robustness of our main results, we have carried out a series of sensitivity analyses, the results of which are displayed in Figure 3. The first is motivated by the question of how to treat years of zero earnings. Our main results are based on two-
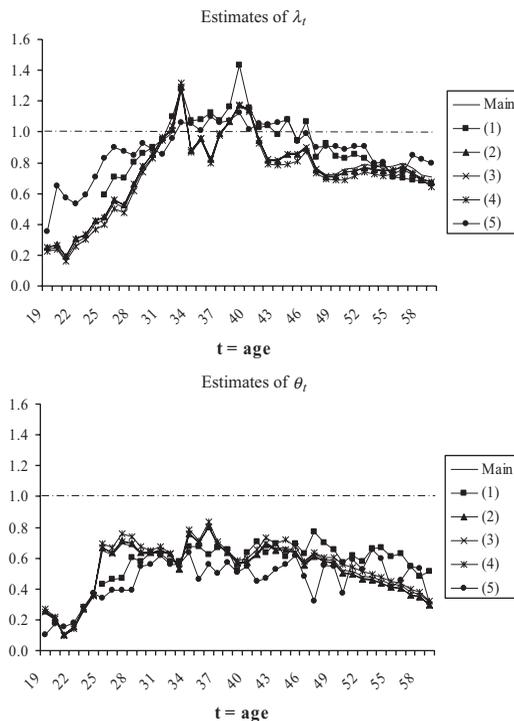

FIGURE 3. ALTERNATIVE ESTIMATES OF λ_t AND θ_t

Notes: The plotted estimates are from six 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 using 0.04 real interest rate.
(3)—same as main, but discounting with one-year T-note interest rates.
(4)—same as main, but weighting by inverse probabilities of selection.
(5)—Swedish estimates from Bohlmark and Lindquist (2005).

limit Tobit estimates that retain observations of zero earnings in the analysis. Because most previous analyses of earnings dynamics, however, have excluded observations of zero earnings, we 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

---

[10] The point estimates plotted in Figure 2 and the associated standard error estimates are tabulated in our electronic Appendix, which also describes our bootstrap procedure for estimating the standard errors.

of the variances and autocovariances in log annual earnings, but because those changes are roughly proportional, the estimated auto-correlations are similar to those in the main analysis. Accordingly, the new estimates of λ_t and θ_t shown in Figure 3 are similar to the estimates from our main analysis repeated from Figure 2.

Our second and third robustness checks explore 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 (a) using the personal consumption expenditures deflator to convert each year's nominal earnings to a real value and (b) assuming a constant real interest rate of 0.02. Our third supplementary analysis uses a real interest rate of 0.04, and our fourth discounts nominal earnings by a nominal interest rate series, the annual one-year T-note interest rates.[11] The results shown in Figure 3 are quite similar to those based on our original interest rate series.

Fourth, we have checked whether our results are affected by the HRS oversampling of blacks, Hispanics, and residents of Florida. To do so, we have reestimated the joint distribution of earnings with a Tobit quasi-maximum like-lihood procedure that weights each observa-tion's contribution to the likelihood function by its inverse probability of selection into the sam-ple. The resulting Tobit estimates are very sim-ilar to those from our original unweighted analysis, and consequently the new estimates of λ_t and θ_t are again very similar to the main estimates.

Finally, Figure 3 includes the results of Bohl-mark and Lindquist's (2005) replication of our main analysis based on Swedish income tax data. This comparison is particularly interesting because Bohlmark and Lindquist's data are largely free of the censorship and coverage is-sues that afflict our U.S. Social Security earn-ings data. As a result, Bohlmark and Lindquist estimate λ_t and θ_t directly with the forward and reverse regressions involving log current and lifetime earnings, without having to resort to

our more complex estimation procedure based on the multivariate normality assumption. Their estimates of λ_t in the twenties are somewhat higher than ours, but still much less than one. In general, the patterns of the Swedish and U.S. results are strikingly similar.

IV. Summary and Discussion

All of our analyses tell the same story: con-trary to the textbook errors-in-variables model usually assumed in applied research, the slope coefficient in the regression of log current earn-ings on log annual earnings varies systemati-cally over the life cycle and is not generally equal to one. We can illustrate the usefulness of our results by applying them to the intergenera-tional mobility regression in which sons' log of lifetime earnings is the dependent variable and fathers' log of lifetime earnings is the explana-tory variable. As summarized in Solon (1999), most recent research in that literature has de-voted considerable attention to the right-side measurement error from using short-run proxies for fathers' lifetime earnings. Our estimates of θ_t shown in Figures 2 and 3 confirm the litera-ture's presumption that right-side measurement error causes an attenuation inconsistency in OLS estimation of the intergenerational elasticity.[12]

The literature, however, has given much less attention to the left-side measurement error from using short-run proxies for sons' lifetime earnings. Presumably, this neglect reflects an assumption by researchers that, in accordance with the textbook errors-in-variables model, left-side measurement error is innocuous for consistency. All our estimates of λ_t suggest that assumption would be fairly well founded if sons' earnings were measured between the early thirties and mid-forties. Many intergenerational mobility studies, however, have measured sons' earnings at earlier ages, and this has substan-

---

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

[12] Many researchers have attempted to reduce the atten-uation inconsistency by averaging fathers' log earnings over multiple years. In an analysis summarized in our electronic Appendix, we repeat our estimation of θ_t except that the new estimates are for five-year averages of log annual earnings, rather than for single years. Our results strongly support the conclusion of Mazumder (2001, 2005) that even estimates based on five-year averages of the earnings vari-able for fathers are subject to a substantial errors-in-variables bias.

tially affected the findings. Reville (1995), for example, estimates intergenerational 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 3. An important implication is that many estimates of the intergenerational earnings elasticity have been subject to substantial attenuation inconsistency from left-side measurement error, in addition to the well-known inconsistency from right-side measurement error.

Of course, interpreting evidence on intergenerational earnings mobility is just one example of how our results might be applied. We advise readers, however, to exercise due caution in importing our estimates of λ_t and θ_t to other earnings data. We already have mentioned issues of comparability between administrative and survey data. Furthermore, the life-cycle trajectories for our U.S. cohort born 1931–1933 may differ from those for other cohorts and other countries. In addition, 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. Nevertheless, taking account of our evidence on departures from the textbook errors-in-variables model should enable better-informed analyses of estimation biases in a wide variety of research that uses current earnings variables as proxies for long-run earnings.

REFERENCES

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

Angrist, Joshua D. and Krueger, Alan B. "Empirical Strategies in Labor Economics," in Orley C. Ashenfelter and David Card, eds., *Handbook of labor economics*. Vol. 3A. Amsterdam: Elsevier Science, North-Holland, 1999, pp. 1277–1366.

Baker, Michael. "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings." *Journal of Labor Economics*, 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*, 2003, *21*(2), pp. 289–321.

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

Bohlmark, Anders and Lindquist, Matthew J. "Life-Cycle Variations in the Association between Current and Lifetime Income: Country, Gender and Cohort Differences." Unpublished Paper, 2005.

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*, 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*, 1999, *18*(2), pp. 169–82.

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*, 1994, 2, pp. 217–54.

Grawe, Nathan D. "Lifecycle Bias in Estimates of Intergenerational Earnings Persistence." *Labour Economics* (forthcoming).

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

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

Jenkins, Stephen. "Snapshots versus Movies: 'Lifecycle Biases' and the Estimation of Intergenerational Earnings Inheritance." *European Economic Review*, 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."

National Bureau of Economic Research, Inc., NBER Working Papers: No. 7235, 1999.

Kim, Bonggeun and Solon, Gary. "Implications of Mean-Reverting Measurement Error for Longitudinal Studies of Wages and Employment." *Review of Economics and Statistics*, 2005, *87*(1), pp. 193–96.

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

Mazumder, Bhashkar. "The Mis-Measurement of Permanent Earnings: New Evidence from Social Security Earnings Data." Federal Reserve Bank of Chicago, Working Paper Series: 01-24, 2001.

Mazumder, Bhashkar. "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data." *Review of Economics and Statistics*, 2005, *87*(2), pp. 235–55.

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 Kenneth K. Kurihara, ed., *Post-Keynesian economics*. New Brunswick: Rutgers University Press, 1954, pp. 388–436.

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

Solon, Gary. "Intergenerational Income Mobility in the United States." *American Economic Review*, 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: Elsevier Science, North-Holland, 1999, pp. 1761–1800.

U.S. Social Security Administration. *Annual statistical supplement to the social security bulletin: 1999*. Washington, DC: U.S. Government Printing Office, 1999.

Wooldridge, Jeffrey M. *Introductory econometrics: A modern approach*. 3rd ed. Mason, OH: Thompson South-Western, 2006.