

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

Steven Haider, Michigan State University

Gary Solon, University of Michigan

September 2003

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. Contrary to that assumption, our analysis of Social Security earnings histories for the Health and Retirement Study sample shows that the slope coefficient in the regression of log annual earnings on log lifetime earnings differs from 1 and varies 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.

The authors gratefully acknowledge grant support from the National Institute on Aging (2-P01 AG 10179) and thank Jeff Wooldridge for his expert advice on the issue discussed in footnote 6.

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

I. Introduction

The importance of the relationship between current and lifetime income variables has been a recurring theme in many areas of economics. Ever since Friedman's (1957) seminal analysis, research on consumption behavior has stressed the distinction between current and permanent income. Another example is the rich literature – exemplified by Lillard (1977) and Shorrocks (1981) – suggesting that earnings inequality as measured in cross-sections of annual earnings overstates the degree of 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. Another recent example is the burgeoning literature on intergenerational income mobility (surveyed in Solon, 1999), which has found that estimation of intergenerational income elasticities is highly sensitive to the length of time and the life-cycle stage over which the income variables are measured.

Despite the pervasive importance of the connection between current and lifetime income variables, surprisingly little direct evidence is available on the statistical relationship between the two and how it varies over the life cycle. Most existing analyses, including those cited above, combine short panels of income data with highly restrictive models of income dynamics in heroic efforts to impute lifetime income from data spanning only a segment of the life cycle.¹

¹ A 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

In this paper, we reexamine the association between current and lifetime earnings using 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 new evidence on the connections between annual and lifetime earnings, and we apply the results to 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

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

where y_{it} is a current income variable, such as log annual earnings, observed for individual i in period t ; y_i is a lifetime 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). Throughout

dynamics models to infer the connection between current and lifetime earnings. Another important contrast with our work is that Mazumder does not address the inappropriateness of the textbook errors-in-variables model.

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.

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 III, we discuss the relevance of their results for the interpretation of ours.

A. Life-cycle variation

Several fragments of evidence suggest life-cycle-related patterns in the association between current and lifetime income variables. 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 that the correlation between current and lifetime income varies dramatically over the life cycle. 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 association between current and lifetime earnings that do have direct implications for errors-in-variables biases.

As noted by Jenkins (1987) and Grawe (2003), life-cycle-related departures from the textbook errors-in-variables model have particular implications for the estimation of intergenerational mobility models such as the regression of son’s log lifetime earnings on father’s log lifetime earnings. For example, 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 intergenerational mobility analyses that use son’s log current earnings to proxy for his log lifetime earnings as the dependent variable will yield systematically different elasticity estimates as a function of the age at which the sons are observed. 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; 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

$$(2) \quad y_{it} = \mathbf{a}_i + \mathbf{g}_i t$$

where initial log earnings \mathbf{a}_i varies across the population with variance \mathbf{s}_a^2 and the earnings growth rate \mathbf{g}_i varies across the population with variance \mathbf{s}_g^2 . Purely for the sake of simplicity, assume that $Cov(\mathbf{a}_i, \mathbf{g}_i) = 0$. Then, assuming infinite lifetimes and a constant real interest rate $r > \mathbf{g}_i$, the present discounted value of lifetime earnings is

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

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

$$(4) \quad \log V_i \cong \mathbf{a}_i + r - \log r + \frac{\mathbf{g}_i}{r}.$$

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

$$(5) \quad \text{Var}(\log V_i) \cong \mathbf{s}_a^2 + (\mathbf{s}_g^2 / r^2)$$

and

$$(6) \quad \text{Cov}(\log V_i, y_{it}) \cong \mathbf{s}_a^2 + (t\mathbf{s}_g^2 / r),$$

it follows that the slope coefficient is

$$(7) \quad \mathbf{I}_t \equiv \frac{\text{Cov}(\log V_i, y_{it})}{\text{Var}(\log V_i)} \cong \frac{\mathbf{s}_a^2 + (t\mathbf{s}_g^2 / r)}{\mathbf{s}_a^2 + (\mathbf{s}_g^2 / r^2)}.$$

The main thing to note about this result is that, contrary to the textbook errors-in-variables model, \mathbf{I}_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, earnings early in the career are an attenuated indicator of lifetime earnings, and earnings late in the career may be an amplified indicator.

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 \mathbf{I}_t and t , they

clearly would not overturn the main qualitative results – that I_t does not generally equal 1 and should be expected to vary over the life cycle.

In the next subsection, we will explain the econometric implications of this 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

$$(8) \quad y_i = \mathbf{b}X_i + \mathbf{e}_i$$

where the error term \mathbf{e}_i is uncorrelated with the regressor X_i . Either y_i or X_i is 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

$$(9) \quad y_{it} = \mathbf{I}_t y_i + u_{it}$$

where \mathbf{I}_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.

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

$$(10) \quad \text{plim } \hat{\mathbf{b}} = \frac{\text{Cov}(X_i, y_{it})}{\text{Var}(X_i)} = \mathbf{I}_t \mathbf{b} .$$

In the textbook case where $\mathbf{I}_t = 1$, measurement error in the dependent variable does not result in inconsistent estimation of \mathbf{b} . 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

$$(11) \quad X_{is} = \mathbf{I}_s X_i + v_{is} .$$

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

$$(12) \quad \begin{aligned} \text{plim } \hat{\mathbf{b}} &= \frac{\text{Cov}(X_{is}, y_i)}{\text{Var}(X_{is})} \\ &= \frac{\mathbf{b} \text{Cov}(X_{is}, X_i)}{\text{Var}(X_{is})} \\ &= \mathbf{q}_s \mathbf{b} \end{aligned}$$

where

$$(13) \quad \mathbf{q}_s \equiv \frac{\text{Cov}(X_{is}, X_i)}{\text{Var}(X_{is})} = \frac{\mathbf{I}_s \text{Var}(X_i)}{\mathbf{I}_s^2 \text{Var}(X_i) + \text{Var}(v_{is})} .$$

In the textbook case where $I_s = 1$, the probability limit of $\hat{\mathbf{b}}$ is \mathbf{b} times the familiar attenuation factor $Var(X_i)/[Var(X_i) + Var(v_{is})]$. More generally, the factor \mathbf{q}_s also depends on the value of I_s . Sometimes referred to as the “reliability ratio,” \mathbf{q}_s is most simply interpreted as the slope coefficient in the “reverse regression” of X_i on X_{is} .

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 $\hat{\mathbf{b}}$ is

$$(14) \quad \text{plim } \hat{\mathbf{b}} = \mathbf{I} \mathbf{q}_s \mathbf{b}.$$

The key message of these results is that 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.³ In section IV, we will estimate those two parameters and examine how they vary over the life cycle.

³ 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 \mathbf{e}_i , u_{it} , or v_{is} . It is easy to show that the probability limit of the instrumental variable estimator is simply $(\mathbf{I}_t / \mathbf{I}_s) \mathbf{b}$. 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 (2003).

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 1951-1991.⁴ The HRS sample is a national probability sample of Americans born between 1931 and 1941, and about $\frac{3}{4}$ 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 $\frac{3}{4}$ that granted access appears 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

⁴ 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.⁵ 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

⁵ 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. 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.

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 that 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

The main assumption in our treatment of the censorship of the Social Security earnings data is the traditional Tobit assumption that the uncensored version of our dependent variable, log annual earnings, is normally distributed. Given the normality assumption, the joint distribution of all 41 annual earnings variables from 1951 to 1991 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 treated as left-censored.

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.

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 \mathbf{l}_t , the slope coefficient in the regression of log (uncensored) earnings in year t on the log of the present value of lifetime earnings. The other is \mathbf{q}_t , the slope coefficient in the reverse regression of log lifetime earnings on log (uncensored) earnings in year t . We are unable to estimate these coefficients directly because we cannot observe the exact value of log earnings in the cases where earnings are right-censored and furthermore, in those cases, we also cannot compute the present value of lifetime earnings.

Nevertheless, we still can estimate \mathbf{l}_t and \mathbf{q}_t by using our estimated joint distribution of uncensored earnings for 1951-1991 to perform the following simulation. First, we take 2,000 random draws from our estimated joint distribution of the 41 years of annual earnings.⁶ Then, for each of the 2,000 simulated earnings histories, we calculate

⁶ To implement the simulation, we need the estimated autocovariance matrix to be positive 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 definite, and indeed it is not. Our procedure for imposing

the present discounted value of lifetime earnings. We perform the discounting 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. 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{I}_t for each year from 1951 to 1991. Similarly, we obtain a \hat{q}_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. Estimates

The Tobit estimation procedure 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

the restriction of positive definiteness begins by characterizing the autocovariance matrix Ω in terms of the Cholesky decomposition $\Omega = TT'$ where T is lower triangular. Ω is positive definite if and only if the diagonal elements of T are positive. We therefore choose T to minimize the distance between TT' and our initial estimate of the autocovariance matrix subject to the constraint that the diagonal elements of T are positive. We measure distance as the sum of squared deviations between the distinct elements in TT' and the corresponding elements in the initial estimate of the autocovariance matrix. Because our initial estimate of the autocovariance matrix is nearly positive definite, the elements in TT' , our positive definite estimate of the autocovariance matrix, differ only slightly from those in the original estimate.

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. The first-order autocorrelations in this table 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 (1993) 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."

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). At first, this 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.

While the results in table 2 give a good sense of the autocorrelation pattern once our sample has matured, the pattern is quite different in the sample's early years. Table 3 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 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. Later in this section, we will describe an alternative analysis that excludes the years prior to 1957.

Next, 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 4 shows the coefficient of variation 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 1 plots our estimates of I_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 I_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.

The estimated life-cycle trajectory shown in figure 1 is similar to the one predicted by the simple model in section II.A. In contrast to the textbook assumption that I_t equals 1 throughout the life cycle, \hat{I}_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 does induce 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 $I_t = 1$ is approximately correct.

Figure 2 shows the estimated life-cycle trajectory of the reliability ratio q_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{q}_t begins at 0.09 at age 19, increases to a fairly flat peak between 0.4 and 0.5 between ages 33 and 56, and then decreases. These results underscore the point, emphasized by Friedman (1957) and countless subsequent authors, that using current earnings to proxy for lifetime earnings as a regressor induces a large errors-in-variables 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.

We intend to perform a series of robustness checks for the next draft of this paper. For now, we have one to report. 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. 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 the supplementary 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, as shown in figures 3 and 4, the new estimates of I_t and q_t also remain similar to those in figures 1 and 2.

B. Application to intergenerational earnings mobility

We can make the lessons of these 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 mentioned in section II.A, the intergenerational mobility literature has exhibited a systematic tendency to estimate lower intergenerational elasticity estimates 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{I}_t in figures 1 and 3. 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 from Solon (1989) on 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 q_t depicted in figures 2 and 4, 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. To explore the extent to which such averaging corrects the bias, in figure 5 we repeat the analysis in figure 1 except that the new estimates of q_t are for five-year 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 5 is higher than in figure 1. Nevertheless, although the estimates of q_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) 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 \mathbf{l}_t and \mathbf{q}_t can be used to correct for errors-in-variables bias. For example, suppose an intergenerational study observes sons at an age t when $\hat{\mathbf{l}}_t = 0.8$ and fathers at an age s when $\hat{\mathbf{q}}_s = 0.6$. Then, using equation (14), one can correct the errors-in-variables bias in $\hat{\mathbf{b}}$ 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 \mathbf{l}_t from 1 and have obtained their estimates of \mathbf{q}_s by combining longitudinal earnings data from limited segments of the life cycle with tightly specified parametric models of earnings dynamics.

We need to caution readers, however, about several problems in applying our estimates of \mathbf{l}_t and \mathbf{q}_s to other sets of earnings data. To begin with, the life-cycle trajectories for our cohort born in 1931-1933 may differ from those for other cohorts. 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 above, 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

Researchers have long 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 Annual Earnings (N=821)

Year	Median	Number with Zero Earnings	Taxable Limit	Number at Taxable Limit
1951	200	305	3600	9
1952	200	337	3600	24
1953	100	378	3600	53
1954	200	357	3600	83
1955	1300	239	4200	99
1956	2200	179	4200	164
1957	3000	87	4200	243
1958	3100	98	4200	288
1959	3800	83	4800	268
1960	4100	92	4800	325
1961	4200	98	4800	363
1962	4700	90	4800	406
1963	4800	86	4800	442
1964	4800	80	4800	463
1965	4800	85	4800	511
1966	6600	70	6600	413
1967	6600	74	6600	439
1968	7500	63	7800	392
1969	7800	75	7800	435
1970	7800	74	7800	464
1971	7800	78	7800	481
1972	9000	87	9000	454
1973	10700	83	10800	407
1974	11400	85	13200	324
1975	11700	94	14100	305
1976	12600	93	15300	305
1977	13400	98	16500	298
1978	15000	100	17700	328
1979	15800	108	22900	219
1980	16500	111	25900	185
1981	17700	119	29700	154
1982	17800	140	32400	141
1983	17800	148	35700	121
1984	18900	159	37800	124
1985	19900	160	39600	117
1986	19100	165	42000	110
1987	19800	178	43800	104
1988	20000	182	45000	112
1989	18700	194	48000	100
1990	18200	194	51300	85
1991	15900	224	53400	75

Table 2. Estimated Autocorrelations in Log Annual Earnings, 1975-1984
(Estimated Standard Errors in Parentheses)

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

Table 3. Estimated Autocorrelations in Log Annual Earnings, 1951-1960
(Estimated Standard Errors in Parentheses)

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

Table 4. Coefficients of Variation for Earnings Variables from Simulations

Year	Coefficient of Variation
1951	1.33
1952	1.21
1953	1.27
1954	1.57
1955	1.17
1956	1.05
1957	.94
1958	.98
1959	1.05
1960	.99
1961	.96
1962	.94
1963	.98
1964	.89
1965	.94
1966	1.07
1967	.97
1968	1.13
1969	.93
1970	.92
1971	1.15
1972	1.05
1973	.98
1974	1.20
1975	1.33
1976	1.27
1977	1.24
1978	1.43
1979	1.07
1980	1.49
1981	1.40
1982	1.34
1983	1.46
1984	1.26
1985	1.43
1986	1.69
1987	1.22
1988	1.09
1989	1.23
1990	1.22
1991	1.35
Lifetime	.67

Figure 1. Estimates of I_t , Based on Two-Limit Tobits

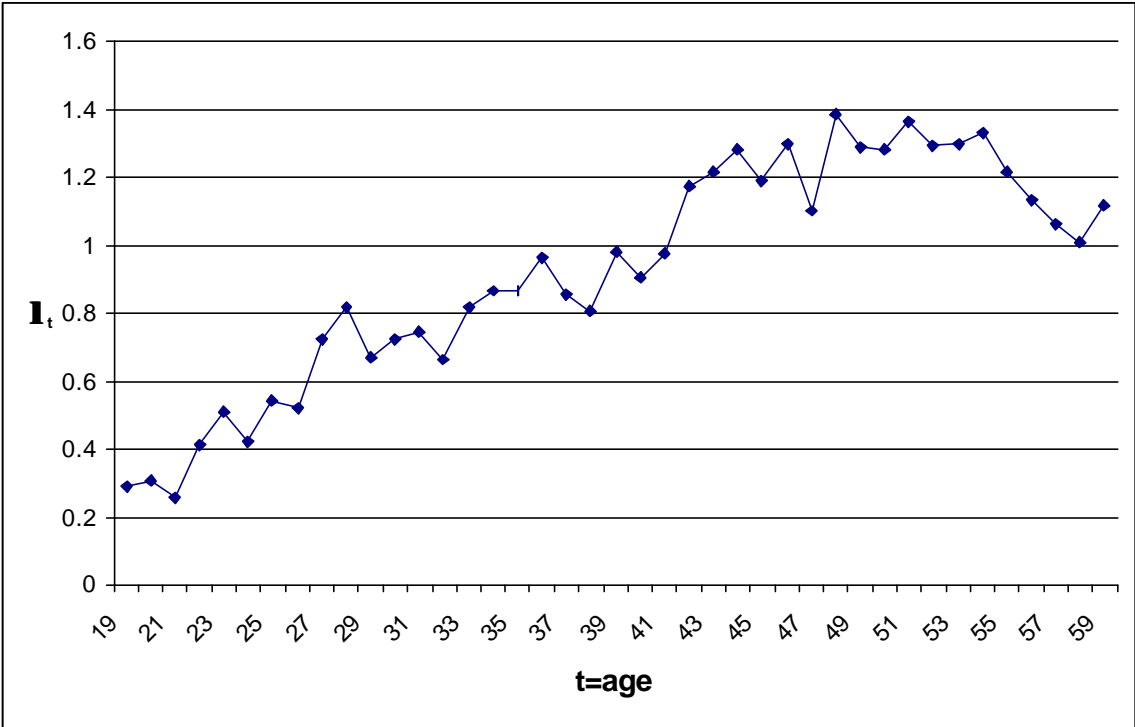


Figure 2. Estimates of q_t Based on Two-Limit Tobits

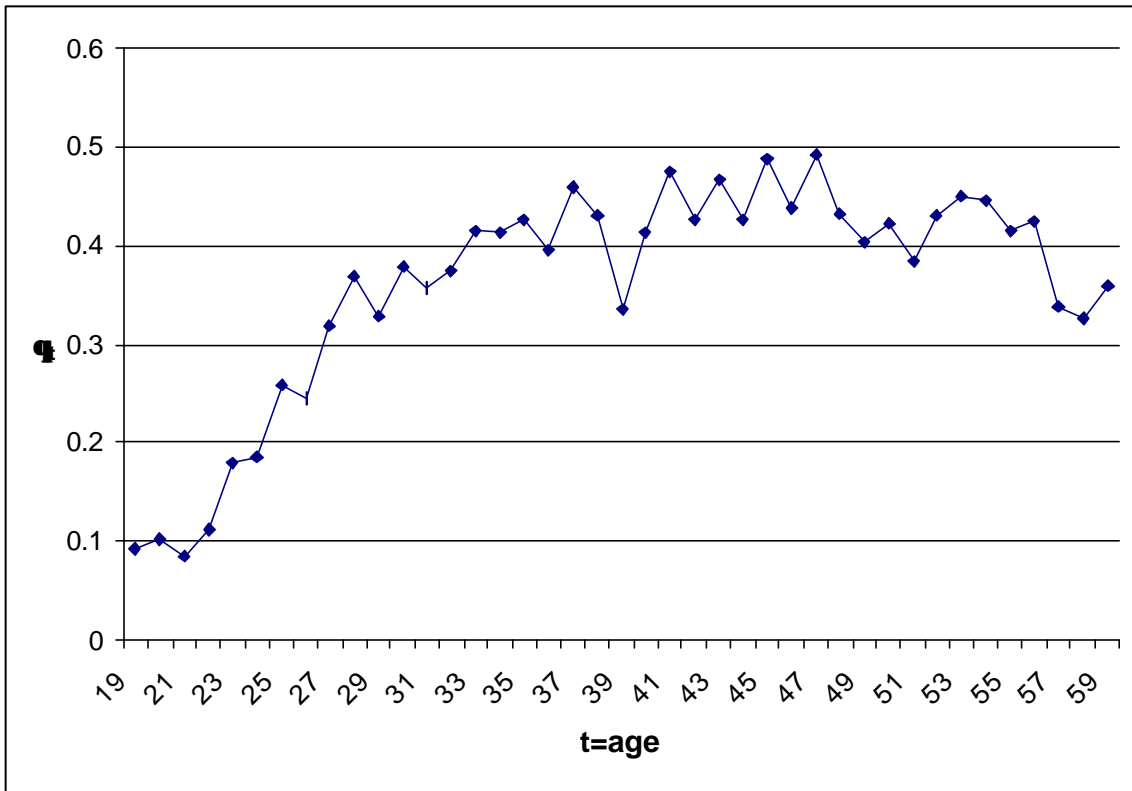


Figure 3. Estimates of I_t Based on One-Limit Tobits

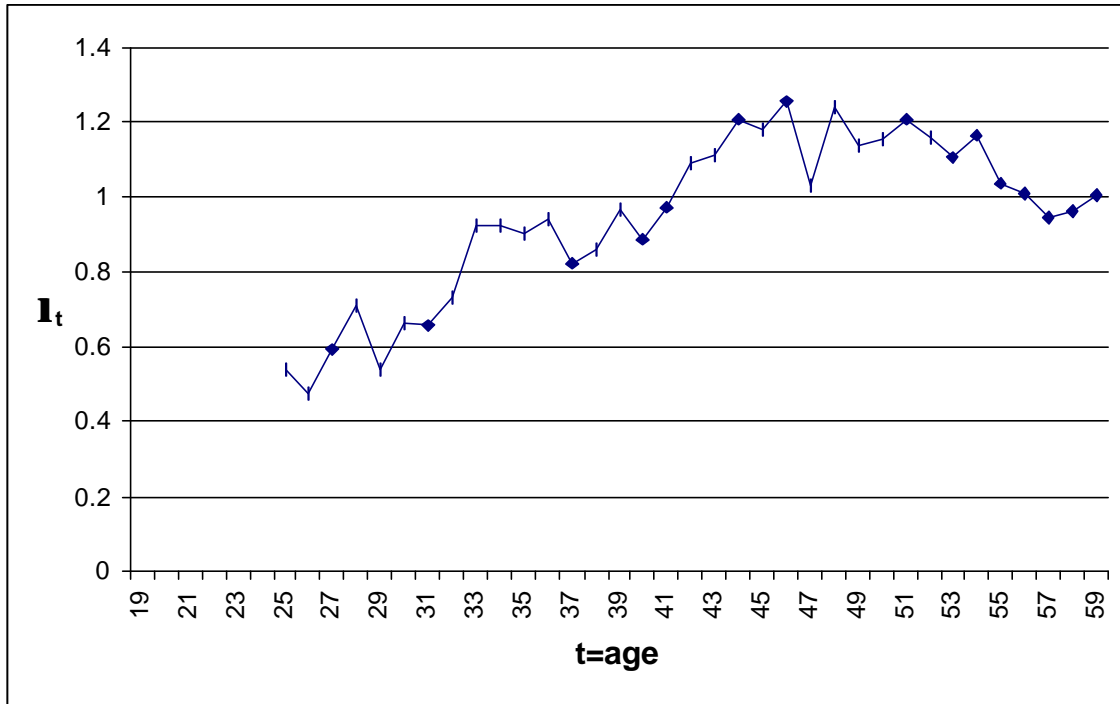


Figure 4. Estimates of q_t Based on One-Limit Tobits

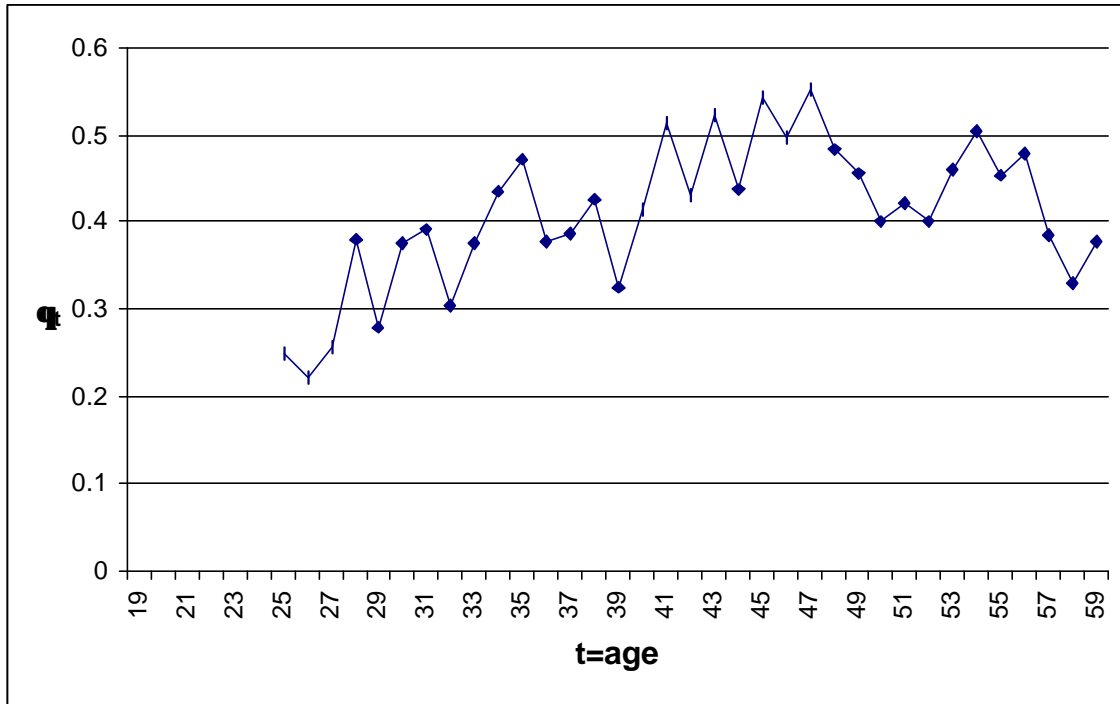
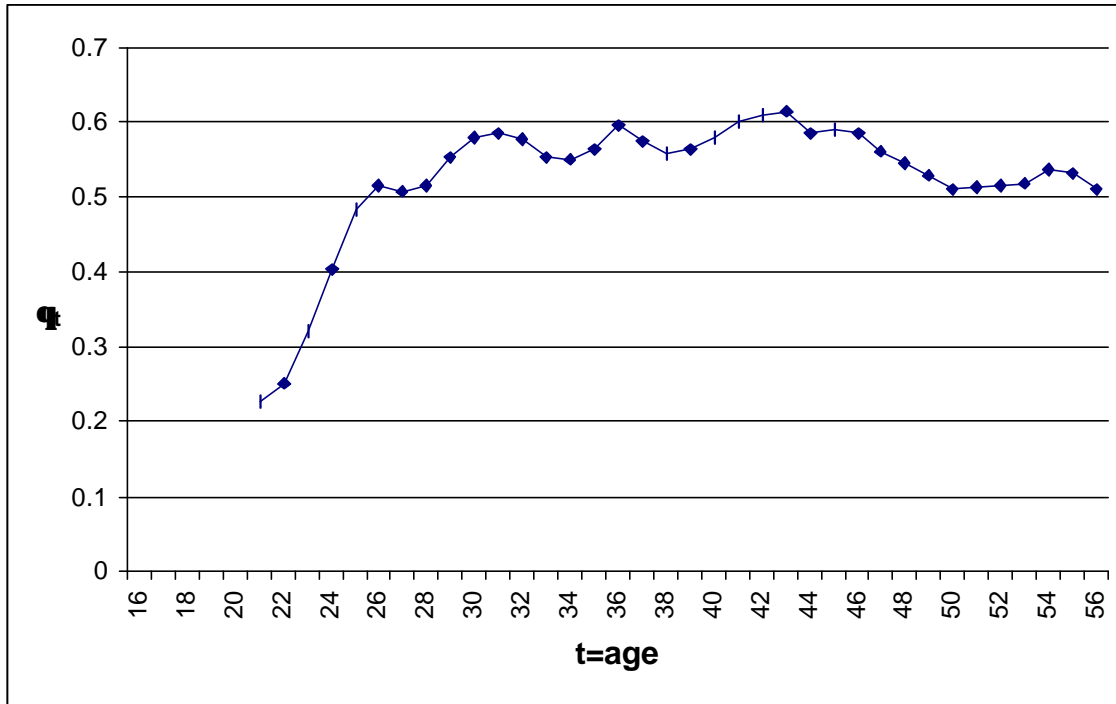


Figure 5. Estimates of q_t for Five-Year Averages Based on Two-Limit Tobits



References

- 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." Unpublished, University of Michigan, 2003.
- 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.
- 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. Klevmarcken 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.