The Mis-Measurement of Permanent Earnings: New evidence from Social Security Earnings Data

By: Bhashkar Mazumder

Federal Reserve Bank of Chicago

WP 2001-24

The Mis-Measurement of Permanent Earnings:
New evidence from Social Security Earnings Data

Bhashkar Mazumder
Federal Reserve Bank of Chicago
October, 2001

Abstract

This study investigates the reliability of using short-term averages of earnings as a proxy for permanent earnings in empirical research. An earnings dynamics model is estimated on a large sample of men covering the period from 1983 to 1997 following the cohort-based methodology of Baker and Solon (1999). The analysis uses a unique dataset that matches men in the 1984, 1990 and 1996 Surveys of Income and Program Participation (SIPP) to the Social Security Administartion's Summary Earnings Records (SER). The results confirm that using a short-term average of earnings can lead to spurious estimates of the effect of lifetime earnings on a particular outcome. In addition, the transitory variance appears to vary considerably over the lifecycle. The share of earnings variance due to transitory factors is higher among blacks and the persistence of transitory shocks appears to be greater for this group as well. Finally, the transitory variance appears to be a more important factor in explaining the overall earnings variance of college educated men than those without college.

I am grateful to David Card and David Levine for their helpful advice and comments. I greatly appreciated the help of Andrew Hildreth, Julia Lane and Susan Grad in helping me gain access to the data. The views presented here do not reflect the views of the Federal Reserve System.

I. Introduction

In recent years, numerous studies on a wide variety of topics have tried to incorporate measures of lifetime earnings or income into empirical economic research. Typically, researchers have used nationally representative longitudinal household survey data such as the Panel Study of Income Dynamics (PSID) and the National Longitudinal Survey (NLS) in order to calculate the earnings stream of individuals and families over time. Unfortunately, for many purposes, the longitudinal samples that can be constructed from these sources are relatively small, due in part to attrition. As a result, researchers have often been forced to rely on relatively short windows of time over which to measure lifetime economic status. This is particularly true in studies on intergenerational earnings mobility (e.g. Solon 1992; Zimmerman 1992; Mulligan 1997), where acquiring meaningful data for family members in *two* generations is an important limitation. Therefore, these studies have typically used only up to five years of data on fathers' earnings or income to proxy for lifetime earnings.

When researchers observe only a few years of an individual's lifetime earnings, exactly how good a reading does it provide of permanent economic status? While at first glance this might appear to be a somewhat narrow technical question, recent studies have found that the degree of the mis-measurement of lifetime earnings may have important ramifications on our understanding of long-term earnings dynamics and inequality. In the case of intergenerational earnings mobility, for example, due to the high persistence of transitory fluctuations in earnings, the bias from using even a five-year average of earnings may be quite substantial –in the order of 30 percent (Mazumder 2001). In addition, there is strong reason to suspect, based on both theory and empirical evidence that the variance in the transitory component of earnings changes considerably over the course of the lifecycle. Using an extraordinarily large database of Canadian taxpayers and a highly structured earnings dynamics model, Baker and Solon (1999) found that the transitory variance follows a U-shaped pattern over the lifecycle. This implies that the *age* at which fathers' earnings are measured may have a sizable effect on estimates of the intergenerational elasticity of earnings. Grawe (2000), building on the findings of Jenkins (1987) has estimated that lifecycle bias in intergenerational mobility studies may be of the order of 25 percent or more. Age-related heteroscedasticity in earnings is obviously an important issue for many other areas of empirical economic research as well. For example, recent studies that have tried to identify the causal effect of family income on children's outcomes, typically have ignored the age at which parent income is measured (e.g. Cameron and Heckman 2001; Duncan and Brooks-Gunn 1997; Mayer 1997 and Blau 1999).

Due to data limitations, however, a detailed empirical analysis of the variance of earnings comparable to Baker and Solon's work has not yet been attempted for the US. Such an analysis would provide greater insight into the nature and extent of bias in existing studies that use short-term proxies for permanent earnings. In addition, other dimensions of earnings instability have largely not been explored. It would be useful, for example, to document whether there are important differences in earnings dynamics by race, sex, income or education.[1] Recent studies that have failed to uncover differences in intergenerational mobility by groups with differential access to credit, (e.g. Mulligan 1997) might not have properly accounted for differences in transitory variance among these groups. Other research has also found that there has been a sharp rise in the variance of transitory earnings in the US during the 1980s (Gottschalk and Moffitt 1994; Moffitt and Gottschalk 1995; Gittleman and Joyce 1996; Haider 1998), and that this increase is an important factor in explaining growing inequality. It would obviously be important to know whether this trend has continued into the 1990s.

Following Baker and Solon, this study uses a new source of data that is able to identify a very rich model of earnings dynamics for the U.S. Specifically, the analysis uses the 1984, 1990 and 1996 Survey of Income and Program Participation (SIPP) matched to Social Security earnings records (SER). By pooling several SIPPs, a very large sample size is generated,

[1] An exception is Gittleman and Joyce (1996) but their analysis was limited to short-term changes in inequality.

allowing for a highly detailed analysis. The match to individuals' social security earnings histories, prevents attrition from dissipating the sample size. The methodology uses minimum distance techniques and a cohort-based estimation strategy that has not yet employed on U.S. data. Among other things, the study documents the lifecycle patterns in the variance of earnings as well as recent trends in the decomposition of earnings inequality between its permanent and transitory components. In addition, differences in earnings dynamics among subgroups of the population are also explored. The results of this analysis are then used to gauge the degree of bias in studies that use short-term proxies in place of true permanent earnings.

The paper proceeds as follows: section II briefly describes the measurement issues involved in using short-term averages of earnings to proxy for permanent earnings in regression models. Section III discusses the data and methodology used in the analysis. Section IV presents the results and their implications. Section V concludes.

II. Measurement Issues

In many economic studies, permanent earnings is used as a right hand side variable in a regression either as a control or as an explanatory variable. The problem with using short-term averages of earnings as a proxy for permanent earnings can be easily understood in the following simplified framework.

$$y_i = bx_{pi} + e \tag{1}$$

$$x_{it} = x_{pi} + w_{it} + v_{it} \tag{2}$$

$$w_{it} = r\, w_{it-1} + x_{it} \tag{3}$$

Here y_i is the dependent variable. In studies of intergenerational transmission of earnings this would be a measure of the child's earnings as an adult. x_{pi} represents permanent earnings and has typically referred to fathers' earnings in intergenerational studies. Of course, what is actually used in empirical work is not x_{pi}, but instead a proxy based on averaging annual earnings, x_{it}, over T number of years. As shown in (2), x_{it} can be decomposed into a permanent component, x_{pi}, a transitory component; w_{it}, and a measurement error term, v_{it}. In addition, w_{it} , follows a first order autoregressive process with r representing the autocorrelation coefficient. If , x_{it} is averaged over T years, the estimate of b derived from OLS will be biased down by an attenuation factor, l_t , as shown below in (4).

$$\hat{b}_{OLS} = b\, l_t \tag{4}$$

where,

$$l_T = \frac{s_x^2}{s_x^2 + \frac{1}{T} a\, s_w^2 + \frac{1}{T} s_v^2},$$

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

$$\text{var}(x_p) = \sigma_x^2, \ \text{var}(w) = \sigma_w^2 \ \text{and var}(v) = \sigma_v^2$$

Essentially, (4) demonstrates that as more years of data are used, the attenuation *factor* will rise and the attenuation *bias* will decline, since the transitory shocks and measurement errors are averaged away. On the other hand, the larger ρ is, the larger α will be, which will, to some extent, offset the benefits of averaging earnings. Using some estimates of the key parameters from previous earnings dynamics models, it can be shown that when using a five-year average of earnings, a reasonable value for λ_5 is about 0.67.[2] This suggests that the resulting estimate of β may be biased down by 33 percent. In studies of intergenerational earnings mobility, where consensus estimates have pointed to an elasticity of 0.4 (Solon, 1999), the actual value may be closer to 0.6.

While this exercise suggests that there is likely to be significant downward bias, there are dimensions to this problem that have been simplified. For example, these calculations have assumed that the variance in the transitory component of earnings represents about 30 percent of the total variance in annual earnings and is constant over time and over the lifecycle. A number of studies (Gottschalk and Moffitt 1994; Moffitt and Gottschalk 1995; Gittleman and Joyce 1996, Haider 1998), have documented that transitory earnings variance has increased during the 1980s and has been an important cause of rising inequality in the US. In addition, Baker and Solon, using Canadian data, have documented that the transitory variance follows a U-shaped pattern over the lifecycle. This suggests that results from studies of the intergenerational earnings elasticity may be sensitive to the age composition of the particular sample. This point was first emphasized by Jenkins (1987) and refined by Grawe (2001) who has argued that this is a key factor in explaining some of the differences in results obtained by various studies. Gittleman and Joyce found evidence of greater earnings instability among blacks and those who are less

[2] Specifically, estimates are needed for ρ, the share of the variance in annual earnings accounted for by permanent earnings, transitory earnings and measurement error. This exercise uses parameter estimates drawn from Card (1994) and Hyslop (2001). See Mazumder (2001) for a more detailed description of the calculations.

educated, suggesting even greater downward bias in estimates for these groups. This suggests that attempts to uncover differences in the effects of family income among subgroups of the population might be biased away from finding large differences, even if they might actually exist for reasons such as borrowing constraints. Gittleman and Joyce's results, however, were based on a series of two-year panels. Finally, the model has also omitted a random-walk component to individual earnings which several empirical studies have found to be important (e.g. MacCurdy 1982; Abowd and Card 1989; Moffitt and Gottschalk 1995; Baker and Solon 1999).

III. Data and Methodology

This analysis pools data from the 1984, 1990 and 1996 Surveys of Income and Program Participation (SIPP). In each SIPP year, individuals are matched to their social security earnings records (SER) from 1951 to 1998. The relevant period of analysis for this study is 1983 through 1997. The time period is restricted to begin in 1983 because in earlier years, the maximum level of earnings that were taxed for Social Security was quite low in real terms. This is important because the SER data is censored at the taxable maximum. In section IV the issues that might arise from this topcoding are addressed. The match rate from the SIPPs to the SER is quite high at over 90 percent and does not appear to present any selection issues.[3]

The methodology closely follows that used by Baker and Solon, for ease of comparison. The sample is restricted to cohorts of men who were between the ages of 24 and 59 for at least nine years from 1982 to 1998. The sample is further divided into two-year birth cohorts beginning with 1931/32 and ending with 1965/66 yielding a total of 18 cohorts. Only men with positive earnings in all of the years that they meet the age requirement are included, thereby restricting the sample to those with a high attachment to the workforce.[4] In order to account for any peculiarities in the earnings variance arising from those entering or exiting the workforce, the first and last years of earnings are excluded. The actual working sample, therefore, consists of men aged between 25 and 58 and spans the years from 1983 and 1997. For 10 of the cohorts, starting with 1939/1940 and up to 1956/57, earnings information is used for all 15 years, allowing for estimates of the autocovariance at lags from 1 to 14 years. On the other hand, the 1931/32 cohort and the 1965/66 cohorts are restricted to earnings information for just 7 years with a maximum autocovariance lag of only 6 years. Some key features of the data are described in Table 1.

---

[3] See Mazumder (2001) for more detail.

[4] This is a standard practice in the literature and suggests that, if anything, the selection rule results in understating the degree of instability in earnings

The most obvious drawback with the selection rule is that it eliminates all men who do not have at least nine consecutive years of positive earnings, thereby excluding those with long spells of unemployment.[5] On the other hand, dropping men with years of zero earnings is preferable with this particular dataset because the SER data records a zero either because the individual did not work or because he did not work in a job covered by the Social Security system. Roughly 60 percent of the recorded zeroes for men are due to non-covered status.[6] Baker and Solon found that without the selection rule of consecutive years of positive earnings, the variances would be considerably larger but that the year-to-year movements would be unaffected. Therefore, for the purpose of identifying trends and *relative* differences in variance parameters, this selection rule is probably not a problem.

On important advance compared to previous research on earnings dynamics in the US is the sample size. This study uses a total of 23837 men with an average of 1324 in each *cohort*. This compares to samples of 2730 used by Gottschalk and Moffitt (1994) and 534 used by Baker (1997). It is also nearly three quarters the size of the working sample of 32,105 used by Baker and Solon.

The analysis in this paper estimates two sets of models. First, a fairly straightforward model of earnings dynamics is presented that allows estimation of a few parameters that would allow for a simple assessment of the bias from averaging earnings over a short span as was done in section II. At this point the estimation simply uses the empirical moments from the variance-covariance matrix of log earnings residuals for the entire sample following most of the earlier literature. The actual moments used for estimation are shown in Appendix Table A1. Like most previous studies, the sample is also restricted to be a balanced panel.[7] In the second set of results

[5] Since annual earnings are used, an individual would have to have no earnings for an entire calendar year to be excluded.
[6] See Mazumder (2001).
[7] This is equivalent to simply using the ten cohorts that are in the sample for all 15 years. Some experimentation was done using an unbalanced panel, but for the most part, results were not very different.

a more complex model is estimated that tries to identify, among other things, lifecycle and time effects and exploits a cohort-based estimation strategy. In this latter case, the set of empirical moments from the variance-covariance matrix of *each* of the cohorts is stacked and used for estimation of the model. For the ten cohorts with earnings data for all 15 years, a total of 120 moments ($15 \times 16 \div 2$) are calculated. For the other cohorts, obviously, fewer moments are available for the analysis although within each cohort, there is a balanced panel. In total, there are 1660 empirical moments that are used for estimation in the second set of results.

An advantage of using a revolving balanced panel design is that it largely removes the direct link between time and age so that both year and age effects can be separately identified. There is still some aging of the panel over time, however, that is most pronounced during the early 1980s and late 1990s (see Table 2).

The methodology for the first approach follows previous studies (e.g. Abowd and Card 1989; Baker 1997) that first "demean" log earnings by using a regression that also adjusts for age, with a quartic function, and year effects, by using dummy variables. The resulting deviations of log earnings, y_{it}, are then modeled as follows:

$$(5) \qquad y_{it} = \alpha_i + \varepsilon_{it} + \zeta_{it}$$

where

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + u_{it},$$

$$\text{var}(\alpha_i) = \sigma^2_{\alpha}\ , \ \text{var}(\zeta_{it}) = \sigma^2_{\zeta}, \ \text{var}(u_{it}) = \sigma^2_{ut}\ ,$$

$$\text{cov}(\zeta_{it}, \zeta_{is}) = 0,\ t \neq s;\ \text{cov}(\zeta_{it}, \alpha_i) = \text{cov}(\zeta_{it}, u_i) = \text{cov}(u_{it}, \alpha_i) = 0$$

In this setup, α_i is the permanent component of earnings that varies across individuals; ε_{it} is a transitory component which follows a first-order autoregressive process with a time-varying variance and ζ_{it} is a white noise component. The final term, ζ_{it}, represents measurement error and

There are also some complexities involved in calculating standard errors that are avoided by using a balanced panel.

is identified by virtue of the fact that it is only contained in the variances not the covariances.[8] This implies, for example, that the variance of earnings in 1990 is as follows:

$$\text{(6)} \qquad \text{var}(y_{1990}) = s^2_a + r^2 s^2_{e1989} + s^2_{e1990} + s^2_z$$

One point that is immediately evident is that the transitory variance follows a recursive structure. Following previous research, it is assumed that the initial variance, σ^2_{e1983} captures all accumulated shocks up to 1983, and that this variance will be common to all individuals in the sample. Using this simple framework, all 120 moments from the variance-covariance matrix of the entire working sample are calculated and then used to estimate the parameters of the model using equally weighted minimum distance (EWMD) –which is equivalent to nonlinear least squares.[9] This approach is then used to estimate the parameters separately for various subgroups of the population.

In the second part of the study, the earnings model closely follows Baker and Solon and borrows their notation. Instead of using a regression to calculate age- and year-adjusted deviations, a more general procedure is now used:

$$\text{(7)} \qquad Y_{ibt} = m_{bt} + y_{ibt}$$

Let Y_{ibt} represent the log earnings of individual i, in birth cohort b, in year t; m_{bt} be the cohort-specific mean for that year; and y_{ibt} be the individual-specific deviation from that mean. Then y_{ibt} can simply be calculated by subtracting the sample average log earnings for cohort b in year t from the observed earnings, Y_{ibt}. The deviations from the mean are then modeled as follows:

$$\text{(8)} \qquad y_{ibt} = p_t[a_{ib} + b_{ib}(t-b-26) + u_{ibt}] + e_{ibt}$$

where

[8] Although one might expect that there would not be substantial measurement error in administrative data one problem that does arise with social security earnings data is that only covered earnings are reported when in fact, individuals may have non-covered earnings as well. See Mazumder (2001) for a discussion of this problem. In addition, there could be mismeasurement due to employer recording errors or changes in the timing of paychecks and bonuses.

[9] See the appendix of Abowd and Card (1989) for a description of the technique. Evidence from Altonji and Segal (1996) and Clark (1996) suggest that using the theoretically derived optimal weighting matrix

$$(9) \qquad u_{ibt} = u_{ibt-1} + r_{ibt}$$

$$(10) \qquad e_{ibt} = re_{ibt-1} + l_t v_{ibt}$$

and

$$(11) \qquad \text{var}(v_{ibt}) = g_0 + g_1\,(t\text{-}b\text{-}26) + g_2\,(t\text{-}b\text{-}26)^2 + g_3(t\text{-}b\text{-}26)^3 + g_4\,(t\text{-}b\text{-}26)^4$$

Deviations in log earnings have both a permanent and transitory component. The expression in brackets shown in the right hand side of (8) breaks down the permanent component into three parts. As before, a_{ib}, with variance $s^2{}_a$, is a fixed effect that varies across individuals. The b_{ib} term, with variance $s^2{}_b$, corresponds to Baker's (1997) finding of heterogeneity in the growth rate of earnings over time and captures the deviation of the individual's idiosyncratic growth rate from his cohort's, after age 26. In addition, since human capital theory typically implies a tradeoff between initial earnings and future earnings growth, s_{ab} is included to represent the covariance between a and b and is expected to be negative.[10] The u_{ibt} term is a random walk component as shown in (9). Here r_{ibt} is "white noise" with variance $s^2{}_r$. Various studies (MacCurdy 1982; Abowd and Card 1989; Moffitt and Gottschalk 1995; Baker and Solon 1999) have found evidence of a random walk component to earnings. As Baker and Solon point out, the random-walk component can be separately identified from the idiosyncratic growth parameter since the former implies a linear growth pattern in log earnings variance, whereas the latter implies a quadratic pattern. In order to capture changes over time in the permanent effect, a factor loading term represented by p_t is also included.

---

can produce serious bias in finite samples. Recent studies therefore, have used the identity matrix as the weighting matrix.

[10] Some researchers such as Mincer (1991) have found that this correlation is positive. This may be explained by omitted variable bias or an alternative theoretical model such as efficiency wages.

The last term in (8), e_{ibt}, is the transitory component which is modeled in (10) as following a first order autoregressive process with a factor loading, l_t, on innovations.[11] The factor loading terms simply capture changes over time in the relative importance of the components. These are standardized to equal 1 in the base year, 1983. Thereafter, increases in p_t relative to increases in l_t reflect movements in the permanent component relative to the transitory component. Finally, (11) models the innovations in the transitory component as following a quartic in experience since age 26, to capture life-cycle effects.

For clarity, an example of an element of the implied variance-covariance matrix is shown below. The variance of log earnings for the 1949/50 cohort in 1990 is presented in (12):

$$\text{var}(y_{49/50,\,1990}) = p^2_{1990}(s^2_a + 15^2 s^2_b + 30 s^2_{ab} + 15\sigma^2_r) + r^2\text{var}(\varepsilon_{i,49/50,1990}) + l^2_{1990}(g_0 + 15g_1 + 15^2 g_2 + 15^3 g_3 + 15^4 g_4) \tag{12}$$

The multiples of 15 in the expression arise from the fact that the cohort born in 1949 is 41 years old in 1990, and therefore has 15 years of experience since age 26. As in the first stage, the transitory variance has a recursive structure that must be traced back to the cohorts' initial entry into the sample.[12]

[11] In this formulation there is no term that corresponds to measurement error. This was omitted to maintain comparability with Baker and Solon.

[12] Following Baker and Solon, the initial transitory variance for each cohort is separately estimated since it makes no sense to impose a common initial variance when the model is designed to differentiate life-cycle effects.

IV. Results

The results for the first stage analysis are shown in Table 3. Here each row represents a different sample used for estimating the autocovariance of earnings. The first row utilizes the full sample and provides a set of baseline estimates. The estimate for s^2_a, the fixed effect, is 0.2, which is similar to previous estimates found by Baker (1997) and Card (1994). If the predicted values for the various parameters are used to decompose the cross-sectional variance of log earnings residuals over the fifteen year period into its permanent and transitory components, the implied permanent share of the cross-sectional variance is only 43 percent. This compares to 50 percent found by both Card (1994) and Hyslop (2001) using the PSID. The share of variance due to transitory earnings is 44 percent, similar to the estimated permanent share but sharply higher than the estimates of 30 percent in the aforementioned studies. In accordance with earlier results, however, r is estimated to be close to 0.8. The estimate of s^2_z, representing measurement error, is 0.06 which implies that the share of variance due to this component is 13 percent —a figure similar to what is found with survey data.[13] Incorporating these baseline estimates into the exercise discussed in Section II, confirms the notion that the use of short-term averages may lead to substantial underestimates of the effects of permanent earnings in regression models. For example, the parameter estimates suggest that a five-year average of earnings would lead to an attenuation factor of 0.56 —implying that coefficient estimates are biased down by more than 40 percent.

The other rows of Table 3 compare differences by subgroups of the population. The estimates for whites are virtually identical to that of the overall population while blacks have a higher share of transitory variance at nearly 50 percent. In addition, the estimates of r are also a bit higher for blacks at 0.81 compared to 0.78 for whites, a difference that is statistically significant. These parameters suggest that five-year averages of earnings for blacks may result in

[13] See Bound and Krueger (1991).

coefficient estimates that are biased down by more than 50 percent. One implication of this finding is that tests designed to uncover differences in parameter estimates by race may be biased away from finding differences.

One concern is that the selection criteria rule, by screening out those with very long spells of unemployment, effectively eliminates some of the racial differences in earnings dynamics that might otherwise be captured. To see what might happen if the selection rule was eased, the same model was also run allowing for some years of zero earnings.[14] The results of this exercise are shown in the last two rows of Table 3. The overall variance in log earnings is about two and a half times larger using this less restrictive selection rule for both blacks and whites. As would be expected, the transitory share is now sharply higher, reaching nearly two thirds in the case of blacks. In addition, r is now estimated to be 0.83 for whites and 0.86 for blacks. Still, the *difference* in the transitory share of the overall earnings variance between blacks and whites is not appreciably different using this selection rule, suggesting that the requirement of positive earnings in all years may not be distorting the comparison.

The fourth and fifth rows of Table 3 compare differences by educational attainment. Here the sample is split between those who have completed college and those who haven't. The overall variance of earnings is sharply lower among college educated men. However, those with college actually appear to have greater earnings instability, as the transitory share of the total variance is higher for college graduates. Gittleman and Joyce, in contrast, found that those at the lowest educational level have a higher share of transitory variance.[15] One potential explanation is that the analysis might conflate age effects with education level. In other words, older workers may have more earnings stability than younger workers due to their position in the lifecycle

[14] Specifically, individuals with positive earnings in approximately two-thirds of the years that they match the age criteria are included. Since the analysis uses log earnings, zero earnings are recoded as $1000 or approximately 6.9 on the log scale. The results are not very sensitive to the imputation value chosen.

[15] It should be noted that in their second set of results (Table 2 on p.192), the coefficient of variation actually rises with education after 12 years of schooling, a result that is consistent with what is presented here.

despite having less education. The model in the second stage analysis will be able to separate these effects.

One possible problem with this analysis is that it uses social security earnings data that is censored at the social security taxable maximum. This might be especially important given the fact that what is being estimated is the *variance* of earnings which may be more sensitive to changes in the tails of the distribution. One comforting fact, however, is that the rate of topcoding, has not changed very much over the period under examination, ranging from a low of 11.6 percent in 1983 to a high of 17.6 percent in 1997.[16] To try to gauge the sensitivity of the results to topcoding, another estimation was attempted that imputed censored observations with random draws from the upper tail of an estimated distribution.[17] The results for the full sample are shown in the last row of Table 3. The only meaningful difference is that r is now estimated to be close to 0.85. The implied share of the permanent component is exactly the same as before while the share of earnings variance is now slightly higher than before.

The parameter estimates from the second part of the study are shown in Table 4. The first panel of the table shows the estimates of the various parameters that comprise the permanent component. Each pair of columns refers to a different sample used for analysis. Recall that the model allows for heterogeneity across individuals in the intercept and the idiosyncratic growth rate, a random walk component and year-specific factor loadings. For the full sample, the estimate for the fixed effect, 0.17, is only slightly lower than what was found in the earlier results. It also slightly larger than the 0.13 in the baseline estimates reported by Baker and Solon for Canada. Unlike the findings from previous studies (Baker 1997, Haider 1998 and Baker and

---

[16] It should be noted that the rising share of topcoded observation is due to the increasing age of the sample over time. The rate of topcoding for those aged 35-45 in each year actually falls from 22.9 percent in 1983 to 16.5 percent in 1997.

[17] The procedure was to assume that the true distribution of log earnings was normally distributed. The mean of the *truncated* earnings distribution in each year along with the rate of topcoding was then used to infer the parameters of the complete distribution (see Greene, p.951). The complete distribution was then truncated from below at the censoring point and then used to randomly draw observations for the purpose of imputing censored observations.

Solon, 1999), there does not appear to be any significant heterogeneity in the growth rate component. Also the estimated covariance between the intercept and growth rate terms is positive. Conforming to previous research, however, there is a positive and significant random walk component. Still the economic significance of the parameter is not clear. For example, for fifty year olds in 1990, the estimated contributed of the random walk component to the cohorts' earnings variance is just 4.9%. The year specific loading factors on the permanent component appear to rise over time but not in a smooth manner. They surge in the early 1990s but begin to stabilize thereafter.

The transitory component is modeled as first order autogressive process with age-varying parameters and different initial variances for each cohort. The estimate for r is 0.67, which is considerably smaller than what was estimated in the bare-boned first stage analysis but still suggests a significant degree of persistence in transitory shocks. The analogous estimate for r in Canada as found by Baker and Solon was just 0.54. The implied effects of this degree of persistence on short-term averages of earnings will be discussed below. There does not appear to be any pattern in the initial variances of each cohort. Given that the initial age at which each cohort is first observed (see Table 1) is monotonically declining, one might have expected a U-shaped pattern to be apparent in the initial variances given the findings of Baker and Solon.

One somewhat surprising result is that the year-specific weights on the transitory component, as reflected by the l_t's, decline through the 1980s and then reverse course and sharply increase during the 1990s. While a higher transitory variance might be expected during the recession years of the early 1990s, it is clear that there has been a secular increase in earnings instability that has persisted through at least 1997.

One clear result is that the age parameters on the transitory variance do, indeed, imply a U-shaped pattern as shown in Figure 1. This pattern supports the notion that regressions using short-term averages of earnings may highly be sensitive to the age-composition of the sample.

Theoretical models that emphasize learning or matching can easily explain the falling transitory variance in earnings as individuals enter their thirties and forties but it is not so clear what accounts for the substantial increase in earnings instability as workers enter their fifties. Clearly this result offers interesting fodder for future research.

The remaining columns in Table 4 estimate the same model by level of education. While it would be ideal to have used the cohort-based approach to measure differences by race, initial results indicated that the small sample of blacks led to unstable results.[18] The comparison between education groups appears to be more meaningful since the sample sizes are reasonably large for both groups. The fixed effect appears to be significantly higher for the no college group compared to those with a college education. The random walk term, in contrast, is much higher for the college sample. The loading factor on the permanent component has remained above 1 for the no college group since 1988 but has been below 1 for the college group. The transitory component appears to be more persistent for college graduates as the estimate for r is slightly higher for this sample. The initial variances for the cohorts are higher for the college group in two-thirds of the cases. The yearly loading factors on the transitory variance for the two groups also show distinct patterns. For college graduates the loading factor has remained above 1 for nearly the entire period with a more pronounced trend up in the 1990s. In contrast for the no college group, the loading factor had declined on average, through the 1980s, but has rebounded in the 1990s.

Taken together, these results appear to bolster the earlier finding that the permanent component accounts for a larger share of the earnings variance of non-college educated men than it does for college educated men. Figure 2 plots the share of earnings variance due to the permanent component for those aged 40 in each year by education group. Figure 2 makes it clear that the share of the earnings variance due to the permanent component was far higher for high-

[18] Specifically some of the estimated initial variances were negative and the loading factors were extremely volatile. Of some interest is that r was estimated to be about 0.95 indicating substantial persistence.

school educated men from 1985 until 1992. Since that time, however, both education groups appear to have roughly similar shares of the permanent component. Put another way, rising earnings instability affected college educated workers far more than high school educated workers in the 1980s, but since the early 1990s it has become an equally important component of the overall earnings variance of both education groups.

The key motivation in estimating this model, however, was to answer the question of how reliable a measure of permanent earnings is a short-term average of earnings. In order to assess this empirically, the following exercise was performed. As before let x_{it}, the earnings of individual i in period t, be decomposed into a permanent component a_{it} and transitory component, e_{it}.[19]

$$(13) \qquad x_{it} = a_{it} + e_{it}$$

Essentially, (13) simply summarizes the more complicated expression for the permanent component shown in (8), with a_{it}. Now define permanent earnings, x_{pi} as the average of the permanent component of earnings, over the fifteen years spanning 1983 through 1997.

$$(14) \qquad x_{pi} = \frac{1}{15}\sum_{t=1983}^{1997} a_{it}$$

With this setup we can then easily use the second stage baseline results to derive the attenuation factors that would arise from using any multiyear average of earnings during the 1983-1997 time span as a proxy for x_{pi} in a regression. In general, if an n year average of earnings beginning in year s, $\bar{x}_{n,s}$, is used as a proxy for x_{pi}, the attenuation factor $\lambda_{n,s}$ is the following:

$$(15) \qquad l_{n,s} = \frac{\text{cov}(\bar{x}_{n,s}, x_p)}{\text{var}(\bar{x}_{n,s})}$$

[19] For simplicity, the notation drops the indexing of cohorts.

To calculate the numerator and denominator of this expression would simply involve summing various estimated variances and covariances involving both the permanent and transitory components.[20] This yields a huge number of estimated attenuation factors, since there are unique results for each potential time span over which an average is taken and since there are distinct results for each cohort. As an example, the attenuation factors for the cohorts born in 1943/44 and 1953/54 are shown in Table 5. Clearly, looking across the rows, as averages are taken over progressively more years the attenuation bias declines as expected. This pattern is robust across all cohorts. The results here support the findings in Mazumder (2001), however, that even a five-year average still results in an estimate that is biased down by about 30 percent.[21] What is quite important, however, is the *age* at which the average is taken. For example, in some cases, the attenuation coefficient from using a single year of earnings, is actually higher than it is when averages are taken over as many as five years. Given these results, it is clear that extraordinary care must be taken in interpreting results about parameters dealing with permanent earnings when only short periods are available for measurement. Clearly some corrections should be undertaken to appropriately adjust the sample composition or to reweight observations, however, such techniques are beyond the scope of this paper.

It is also important to note that there maybe many instances where researchers should not be concerned about the earnings stream over one's *entire lifetime*. For example, the dependent variable of interest may be a child's outcome that is related only to parent's income at the time of

---

[20] For example, if a five-year average of earnings covering the years 1983-1987, $\bar{x}_{5,83}$, was used as a proxy for xp_i the resulting attenuation factor, $\lambda_{5,83}$ is as follows:

$$l_{5,83} = \frac{\frac{1}{5}\sum_{s=83}^{87}\frac{1}{15}\sum_{t=83}^{97}\mathrm{cov}(a_{is},a_{it})}{\frac{1}{25}(\sum_{s=83}^{87}\mathrm{var}(a_{is})+\sum_{s=83}^{87}\mathrm{var}(e_{is})+2\sum_{s=83}^{87}\sum_{t=s+1}^{87}\mathrm{cov}(a_{is},a_{it})+2\sum_{s=83}^{87}\sum_{t=s+1}^{87}\mathrm{cov}(e_{is},e_{it})}$$

[21] The average of column 5 in Table 5 is 0.68.

child rearing.[22] In that case transitory fluctuations that occur at the time that parents are actively raising children may actually matter. If such fluctuations are highly persistent, then a short-term average may actually be preferable to a long-term average. The attenuation coefficients presented here, obviously should not be applied in those instances.

[22] For example certain health outcomes at young ages may be especially sensitive to current income, especially among groups that are borrowing constrained.

V. Conclusion

As researchers increasingly exploit longitudinal data to study the effects of income on various outcomes, it is essential that the pitfalls of using short-term proxies as a substitute for their permanent counterparts, are well understood. This study adds to our knowledge by using a new longitudinal data source that contains a large panel of years and a sizable sample to estimate a detailed earnings dynamics model. The results of this analysis bolster the findings of other studies that have found that the variance of the transitory component of earnings accounts for a large share of the overall observed variance in earnings. In addition, the variance in transitory earnings appears to follow a pronounced U-shaped pattern over the lifecycle as was found by Baker and Solon (1999) and has risen considerably over time.

The estimates of the model suggest that a short-term average of earnings, when used as a substitute for permanent earnings in a regression model, can lead to serious bias. In particular, even a five year average of earnings can lead to a parameter estimate that is biased down by 30 percent. Of particular importance is the age over which earnings are averaged. In fact, earnings taken from just one or two years at the right stage of the lifecycle may yield less biased estimates than those stemming from a five or six-year average. Future research should pay careful attention to the age composition of the sample and try to develop methods to overcome this bias.

Attempts to uncover differences in earnings dynamics among subgroups of the population lead to some interesting results. The share of earnings variance due to transitory factors is higher among blacks and the persistence of transitory shocks appears to be greater for this group as well. This suggests that comparisons of the effect of income on outcomes by race may lead to failure to detect differences when in fact, such differences may well exist. A somewhat surprising result is that the transitory variance appears to be a more important factor in explaining the overall earnings variance of college educated men than those without college. However, in recent years, both groups appear to be equally impacted by growing earnings instability.

References

Abowd, John M and David Card (1989), "On the Covariance Structure of Earnings and Hours Changes" *Econometrica*, 57(2) 411-445

Altonji, Joseph G. and Lewis M. Segal, "Small Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business and Economic Statistics*, 14() 353-366.

Baker, Michael (1997), "Growth-Rate Heterogeneity and the Covariance Structure of Life Cycle Earnings", Journal of Labor Economics, 15(2) 338-375

Baker, Michael and Gary Solon (1999), "Earnings Dynamics and Inequality Among Canadian Men, 1976-1992: Evidence from Longitudinal Tax Records" *NBER Working Paper* 7370, NBER, Cambridge Mass.

Blau, David (1999), "The Effect of Income on Child Development", *Review of Economics and Statistics*, 81(2) 261-276.

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

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

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

Clark, Todd (1996), "Small-Sample Properties of Estimators of Nonlinear Models of Covariance Structure," *Journal of Business and Economic Statistics*, 14() 367-373.

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

Duncan, Greg J. and Jeanne Brooks-Gunn (1997), eds. "Consequences of Growing up Poor". Russell Sage Foundation, New York

Gittleman, Maury and Mary Joyce (1996), "Earnings Mobility and Long-Run Inequality: An Analysis Using Matched CPS Data", Industrial Relations, 35(2) 180-196.

Gottschalk, Peter and Robert A. Moffitt (1994), "The Growth of Earnings Instability in the U.S. Labor Market," Brookings Papers on Economic Activity, 217-272

Grawe, Nathan D. (2000), "Lifecycle Bias in the Estimation of Intergenerational Income Persistence". Manuscript, University of Chicago.

Haider, Steven J. (1998), “Earnings Instability and Earnings Inequality of Males in the United States 1967-1991,” Chapter 2 in Econometric Studies of Long-Run Earnings Inequality, PhD. Dissertation, University of Michigan.

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

Jenkins, Stephen (1987), “Snapshots Versus Movies: ‘Lifecycle Biases’ and the Estimation of Intergenerational Earnings Inheritance”, *European Economic Review* 31:1149-1158

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

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

Mayer, Susan (1997), “What Money Can’t Buy: Family Income and Children’s Life Chances”. Harvard University Press, Cambridge Mass.

Mazumder, Bhashkar (2001), “Earnings Mobility in the U.S.” A New Look at Intergenerational Inequality.” *Federal Reserve Bank of Chicago Working Paper 2001-18*.

Mincer, Jacob (1991), “Human Capital, Technology, and the Wage Structure: What do Time Series Show?,” *NBER Working Paper* 3581.

Moffitt, Robert A and Peter Gottschalk, “Trends in the Autocovariance Structure of Earnings in the U.S. 1969-1987,” Working Paper No. 335, Department of Economics, Johns Hopkins University.

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

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

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

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

Table 1: Sample Size by Cohort

Birth Year	Sample Size	Years Observed	Initial Age	Final Age
1931/32	728	1983-1989	52	58
1933/34	735	1983-1991	50	58
1935/36	795	1983-1993	48	58
1937/38	802	1983-1995	46	58
1939/40	806	1983-1997	44	58
1941/42	1028	1983-1997	42	56
1943/44	1047	1983-1997	40	54
1945/46	1196	1983-1997	38	52
1947/48	1436	1983-1997	36	50
1949/50	1490	1983-1997	34	48
1951/52	1537	1983-1997	32	46
1953/54	1580	1983-1997	30	44
1955/56	1647	1983-1997	28	42
1957/58	1626	1983-1997	26	40
1959/60	1732	1985-1997	26	38
1961/62	1869	1987-1997	26	36
1963/64	1923	1989-1997	26	34
1965/66	1860	1991-1997	26	32
Total	23837			

Note : Age refers to age of older members in 2-year birth cohorts

Table 2: Sample Statistics by Year

Year	N	Mean Log Earnings	Std. Dev. Log Earnings	Min Age	Max Age	Mean Age
1983	16453	10.273	0.767	25	52	36.1
1984	16453	10.348	0.699	26	53	37.1
1985	18185	10.344	0.695	25	54	36.9
1986	18185	10.386	0.691	26	55	37.9
1987	20054	10.367	0.702	25	56	37.7
1988	20054	10.402	0.660	26	57	38.7
1989	21977	10.371	0.678	25	58	38.4
1990	21249	10.370	0.678	26	57	38.8
1991	23109	10.275	0.715	25	58	38.6
1992	22374	10.244	0.713	26	57	39
1993	22374	10.324	0.753	27	58	40
1994	21579	10.367	0.713	28	57	40.4
1995	21579	10.373	0.727	29	58	41.4
1996	20777	10.376	0.733	30	57	41.8
1997	20777	10.394	0.757	31	58	42.8

Note: Earnings are measured in 1998 dollars deflated by using the CPI-U-X1 series.

Table 3: Results from simple earnings dynamics specification

	s^2_a (fixed effect)	r	s^2_{ut}** (trans. Var.)	s^2_z (meas. Error)	*F-stat*	*N**	*Implied shares of cross-sectional variance due to…* *permanent component*	*transitory component*	*measurement error*
Whole sample	0.20	0.78	0.09	0.06	163.29	13393	0.43	0.44	0.13
	(0.006)	(0.010)		(0.003)					
Whites	0.19	0.78	0.09	0.06	156.46	12121	0.43	0.44	0.13
	(0.005)	(0.010)		(0.003)					
Blacks	0.24	0.81	0.13	0.09	167.62	962	0.37	0.49	0.14
	(0.037)	(0.035)		(0.011)					
College	0.14	0.72	0.09	0.04	165.14	4312	0.41	0.49	0.10
	(0.007)	(0.018)		(0.005)					
No College	0.20	0.75	0.11	0.06	200.52	9081	0.42	0.45	0.13
	(0.007)	(0.012)		(0.003)					
Whites less restricted	0.34	0.83	0.25	0.16	238.23	17492	0.29	0.57	0.14
	(0.012)	(0.006)		(0.004)					
Blacks less restricted	0.34	0.86	0.34	0.21	299.22	1568	0.22	0.64	0.14
	(0.064)	(0.018)		(0.016)					
Full Sample Imputed	0.23	0.84	0.08	0.09	160.05	13393	0.43	0.41	0.16
	(0.007)	(0.009)		(0.003)					

*Here, N refers to the number of observations used to calculate the variance-covariance matrix.

**This is the average of the transitory variance over the fifteen years

Notes: All specifications use a total of 120 moments, those used for row 1 are shown in Appendix Table A1.
Less restricted sample allows some years of zero earnings. Standard errors are shown in parenthesis

Table 4: Results from cohort-based earnings dynamics model

	Full sample		*College*		*No College*	
	estimate	s.e	estimate	s.e	estimate	s.e
			permanent component			
s^2_a	0.170	(0.005)	0.118	(0.007)	0.172	(0.006)
s^2_b	1.01E-07	(1.56E-07)	-2.88E-06	(6.40E-07)	-4.89E-06	(6.75E-07)
s_{ab}	2.62E-04	(2.63E-05)	5.90E-04	(8.09E-05)	3.32E-04	(4.36E-05)
s^2_r	6.52E-04	(6.36E-05)	1.49E-03	(2.12E-04)	8.17E-04	(1.07E-04)
			yearly loading factors on permanent component			
p84	0.992	(0.017)	0.940	(0.031)	0.982	(0.022)
p85	0.995	(0.019)	0.902	(0.036)	0.988	(0.023)
p86	1.044	(0.019)	0.903	(0.036)	1.036	(0.024)
p87	1.068	(0.020)	0.970	(0.037)	1.042	(0.025)
p88	1.017	(0.020)	0.901	(0.036)	0.994	(0.024)
p89	1.063	(0.020)	0.928	(0.036)	1.039	(0.025)
p90	1.117	(0.020)	0.996	(0.037)	1.082	(0.025)
p91	1.175	(0.021)	1.046	(0.038)	1.135	(0.026)
p92	1.174	(0.021)	0.984	(0.037)	1.145	(0.026)
p93	1.148	(0.021)	0.998	(0.037)	1.095	(0.025)
p94	1.112	(0.020)	1.050	(0.037)	1.048	(0.024)
p95	1.092	(0.020)	0.983	(0.035)	1.038	(0.024)
p96	1.112	(0.019)	0.949	(0.034)	1.072	(0.024)
p97	1.151	(0.019)	0.978	(0.034)	1.108	(0.023)
			transitory component			
r	0.670	(0.006)	0.694	(0.009)	0.675	(0.008)
			cohort specific initial variances			
$\sigma^2_{31/32}$	0.296	(0.022)	0.411	(0.031)	0.218	(0.027)
$\sigma^2_{33/34}$	0.574	(0.022)	0.571	(0.036)	0.533	(0.029)
$\sigma^2_{35/36}$	0.209	(0.023)	-0.051	(0.031)	0.242	(0.029)
$\sigma^2_{37/38}$	0.257	(0.021)	0.337	(0.035)	0.209	(0.029)
$\sigma^2_{39/40}$	0.271	(0.021)	0.191	(0.034)	0.265	(0.026)
$\sigma^2_{41/42}$	0.275	(0.025)	0.172	(0.032)	0.299	(0.026)
$\sigma^2_{43/44}$	0.320	(0.024)	0.091	(0.032)	0.395	(0.028)
$\sigma^2_{45/46}$	0.256	(0.022)	0.260	(0.029)	0.257	(0.025)
$\sigma^2_{47/48}$	0.262	(0.022)	0.405	(0.033)	0.213	(0.026)
$\sigma^2_{49/50}$	0.241	(0.022)	0.171	(0.029)	0.268	(0.030)
$\sigma^2_{51/52}$	0.437	(0.023)	0.326	(0.031)	0.489	(0.029)
$\sigma^2_{53/54}$	0.325	(0.023)	0.401	(0.033)	0.294	(0.026)
$\sigma^2_{55/56}$	0.393	(0.025)	0.348	(0.034)	0.399	(0.027)
$\sigma^2_{57/58}$	0.481	(0.024)	0.635	(0.034)	0.435	(0.029)
$\sigma^2_{59/60}$	0.204	(0.025)	0.409	(0.036)	0.186	(0.029)
$\sigma^2_{61/62}$	0.289	(0.023)	0.419	(0.033)	0.266	(0.028)
$\sigma^2_{63/64}$	0.322	(0.025)	0.505	(0.032)	0.318	(0.027)
$\sigma^2_{65/66}$	0.293	(0.025)	0.342	(0.035)	0.315	(0.028)

Table 4: Results from cohort-based earnings dynamics model (cont.)

	Full sample		*College*		*No College*	
	estimate	s.e	estimate	s.e	estimate	s.e
			yearly loading factors on transitory component			
l_{84}	1.000	---	1.000	---	1.000	---
l_{85}	0.995	(0.051)	1.377	(0.105)	0.897	(0.062)
l_{86}	0.953	(0.047)		(0.099)	0.936	(0.053)
l_{87}	0.875	(0.048)	1.079	(0.098)	0.840	(0.056)
l_{88}	0.918	(0.044)	1.023	(0.091)	0.887	(0.053)
l_{89}	0.801	(0.048)	0.996	(0.092)	0.748	(0.056)
l_{90}	0.875	(0.047)	1.041	(0.090)	0.832	(0.057)
l_{91}	0.940	(0.044)	1.041	(0.092)	0.907	(0.051)
l_{92}	1.103	(0.043)	1.218	(0.093)	1.076	(0.052)
l_{93}	1.149	(0.042)	1.193	(0.089)	1.148	(0.050)
l_{94}	1.106	(0.044)	1.063	(0.094)	1.099	(0.052)
l_{95}	1.252	(0.042)	1.327	(0.088)	1.225	(0.051)
l_{96}	1.237	(0.048)	1.358	(0.091)	1.191	(0.054)
l_{97}	1.244	(0.044)	1.219	(0.094)	1.256	(0.060)
			parameters on quartic on age			
g_0	0.172	(0.009)	0.101	(0.011)	0.181	(0.012)
g_1	-6.33E-03	(4.10E-04)	-3.64E-03	(4.48E-04)	-6.22E-03	(4.92E-04)
g_3	-1.54E-05	(2.32E-06)	-5.81E-08	(2.90E-07)	1.35E-06	(5.18E-07)
g_4	1.02E-06	(1.95E-07)	3.95E-07	(1.78E-07)	8.41E-07	(1.81E-07)
g_5	2.54E-07	(1.65E-08)	1.70E-07	(1.94E-08)	2.45E-07	(1.98E-08)

Table 5: Implied Attenuation Factors for 1943/44, 1953/54 Cohorts

starting year	*starting age*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
		1943/44 Cohort, Average earnings over…years														
1983	40	0.40	0.47	0.53	0.59	0.63	0.68	0.71	0.74	0.76	0.77	0.79	0.79	0.80	0.80	0.80
1984	41	0.49	0.55	0.60	0.65	0.69	0.72	0.75	0.77	0.78	0.79	0.81	0.80	0.80	0.80	
1985	42	0.54	0.60	0.65	0.69	0.72	0.75	0.76	0.78	0.78	0.79	0.80	0.79	0.79		
1986	43	0.58	0.64	0.68	0.72	0.74	0.76	0.77	0.77	0.78	0.78	0.79	0.77			
1987	44	0.62	0.67	0.71	0.73	0.75	0.75	0.76	0.76	0.76	0.76	0.77				
1988	45	0.63	0.69	0.71	0.73	0.74	0.74	0.74	0.74	0.74	0.74					
1989	46	0.66	0.69	0.71	0.71	0.71	0.72	0.72	0.72	0.71						
1990	47	0.64	0.66	0.67	0.68	0.68	0.68	0.68	0.68							
1991	48	0.60	0.62	0.63	0.64	0.64	0.65	0.65								
1992	49	0.54	0.57	0.59	0.60	0.60	0.61									
1993	50	0.50	0.53	0.55	0.56	0.57										
1994	51	0.48	0.50	0.51	0.52											
1995	52	0.42	0.41	0.47												
1996	53	0.39	0.38													
1997	54	0.36														
		1953/54 Cohort, Average earnings over…years														
1983	30	0.38	0.44	0.49	0.53	0.57	0.61	0.64	0.67	0.70	0.72	0.73	0.75	0.76	0.77	0.78
1984	31	0.42	0.47	0.52	0.56	0.60	0.64	0.67	0.69	0.71	0.73	0.76	0.76	0.77	0.77	
1985	32	0.44	0.50	0.55	0.59	0.63	0.66	0.69	0.71	0.72	0.74	0.76	0.76	0.77		
1986	33	0.47	0.53	0.58	0.62	0.65	0.68	0.70	0.72	0.73	0.74	0.76	0.76			
1987	34	0.52	0.57	0.61	0.65	0.67	0.69	0.71	0.72	0.73	0.74	0.76				
1988	35	0.53	0.59	0.63	0.66	0.68	0.69	0.71	0.72	0.73	0.73					
1989	36	0.58	0.62	0.65	0.66	0.68	0.69	0.70	0.71	0.72						
1990	37	0.58	0.62	0.63	0.65	0.66	0.67	0.69	0.69							
1991	38	0.57	0.59	0.61	0.63	0.64	0.66	0.67								
1992	39	0.53	0.56	0.59	0.61	0.63	0.64									
1993	40	0.50	0.55	0.58	0.60	0.62										
1994	41	0.51	0.54	0.57	0.59											
1995	42	0.48	0.48	0.55												
1996	43	0.47	0.48													
1997	44	0.47														

Note: Attenuation factors are calculated according to formula shown in equation 15

Table A1: The Variance-Covariance Matrix of Log Earnings Residuals for the Full sample

	1983	*1984*	*1985*	*1986*	*1987*	*1988*	*1989*	*1990*	*1991*	*1992*	*1993*	*1994*	*1995*	*1996*	*1997*
1983	0.556	0.705	0.616	0.563	0.528	0.505	0.497	0.475	0.461	0.424	0.404	0.396	0.375	0.365	0.343
1984	0.361	0.472	0.750	0.638	0.588	0.559	0.551	0.522	0.505	0.460	0.447	0.441	0.420	0.408	0.384
1985	0.302	0.339	0.433	0.764	0.661	0.619	0.600	0.575	0.556	0.510	0.484	0.484	0.449	0.436	0.409
1986	0.279	0.292	0.334	0.442	0.770	0.669	0.651	0.614	0.578	0.528	0.509	0.507	0.474	0.455	0.423
1987	0.257	0.264	0.284	0.335	0.427	0.770	0.704	0.652	0.612	0.556	0.534	0.534	0.497	0.481	0.454
1988	0.229	0.234	0.248	0.271	0.306	0.370	0.791	0.703	0.647	0.587	0.561	0.558	0.525	0.502	0.471
1989	0.225	0.230	0.240	0.263	0.280	0.293	0.370	0.792	0.705	0.640	0.609	0.604	0.560	0.533	0.501
1990	0.223	0.225	0.238	0.257	0.268	0.269	0.303	0.396	0.789	0.678	0.645	0.634	0.586	0.555	0.517
1991	0.226	0.227	0.240	0.252	0.262	0.258	0.281	0.326	0.430	0.765	0.683	0.657	0.614	0.580	0.542
1992	0.221	0.220	0.234	0.245	0.254	0.249	0.272	0.298	0.351	0.488	0.771	0.692	0.633	0.582	0.543
1993	0.212	0.216	0.224	0.238	0.246	0.240	0.261	0.286	0.316	0.379	0.496	0.796	0.706	0.638	0.587
1994	0.199	0.204	0.215	0.227	0.235	0.229	0.247	0.269	0.291	0.326	0.378	0.455	0.815	0.719	0.661
1995	0.194	0.200	0.205	0.219	0.225	0.221	0.236	0.256	0.280	0.306	0.345	0.381	0.481	0.816	0.722
1996	0.195	0.200	0.205	0.216	0.225	0.218	0.232	0.250	0.272	0.290	0.321	0.347	0.405	0.511	0.820
1997	0.193	0.199	0.203	0.213	0.224	0.217	0.230	0.246	0.269	0.286	0.312	0.337	0.378	0.443	0.571

Note: Autocorrelations are presented above the diagonal

Figure 1: Lifecycle Pattern of Transitory Variance

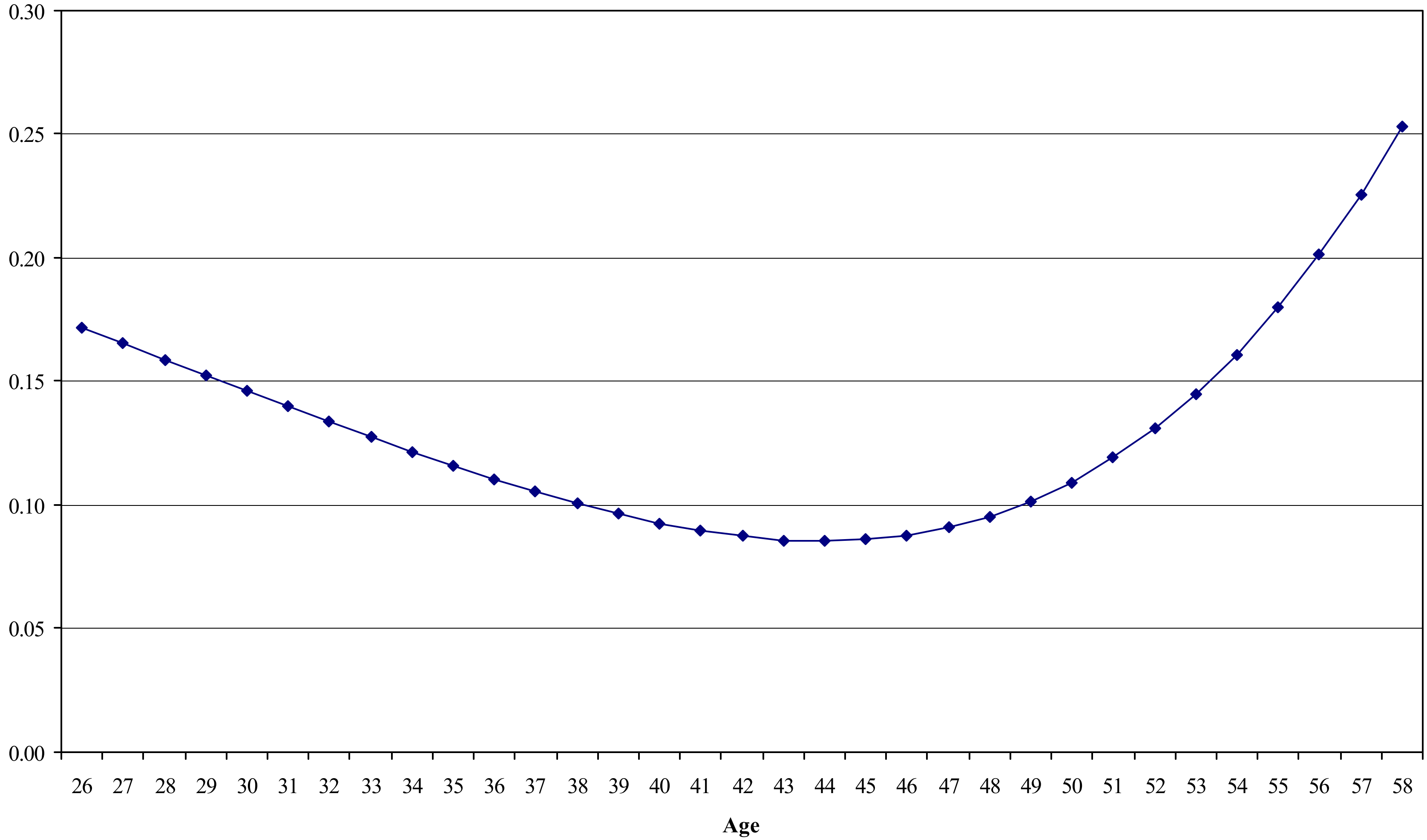

Figure 2: Permanent Component Share of Earnings Variance, 40 Year Olds