Validation study of earnings data in the SIPP - do older workers have larger measurement error?

Authors: Peter T. Gottschalk, Minh Huynh

Chestnut Hill, Mass.: Center for Retirement Research at Boston College, 2005

VALIDATION STUDY OF EARNINGS DATA IN THE SIPP--
DO OLDER WORKERS HAVE LARGER MEASUREMENT ERROR?

Peter Gottschalk and Minh Huynh*

CRR WP 2005-07
Released: May 2005
Draft Submitted: April 2005

Center for Retirement Research at Boston College
550 Fulton Hall
140 Commonwealth Ave.
Chestnut Hill, MA 02467
Tel: 617-552-1762   Fax: 617-552-0191
http://www.bc.edu/crr

*Peter Gottschalk is a Professor of Economics in the School of Arts and Sciences at Boston College. Minh Huynh is a Senior Analyst with the Social Security Administration. The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA) to the Center for Retirement Research at Boston College (CRR). The findings and conclusions are solely those of the authors and do not represent the views of SSA, any agency of the Federal Government, or Boston College.

About the Center for Retirement Research

The *Center for Retirement Research at Boston College*, part of a consortium that includes a parallel centers at the University of Michigan and the National Bureau of Economic Research, was established in 1998 through a grant from the Social Security Administration. The goals of the Center are to promote research on retirement issues, to transmit new findings to the policy community and the public, to help train new scholars, and to broaden access to valuable data sources. Through these initiatives, the Center hopes to forge a strong link between the academic and policy communities around an issue of critical importance to the nation's future.

Center for Retirement Research at Boston College
550 Fulton Hall
140 Commonwealth Ave.
Chestnut Hill, MA 02467
phone: 617-552-1762 fax: 617-552-0191
e-mail: crr@bc.edu
http://www.bc.edu/crr

Affiliated Institutions:

American Enterprise Institute
The Brookings Institution
Center for Strategic and International Studies
Massachusetts Institute of Technology
Syracuse University
Urban Institute

Abstract:

The Survey of Income and Program Participation (SIPP) is a potentially useful data set to study earnings and retirement dynamics of older workers. Respondents' self-reported work and earnings in the SIPP are, however, likely to be measured with error, and this measurement error may be particularly large for older respondents who work non-standard hours. We explore the extent of measurement error by comparing SIPP employment and earnings data to administrative records contained in the matched Detail Earning Record.

I. Introduction

The Survey of Income and Program Participation (SIPP) is a potentially useful data set to contrast the earnings dynamics of older workers with their younger counterparts. For example, this data set contains a rich set of variables measuring earnings and employers on a monthly basis. While researchers have shown that other SIPP measures such as self-reported transfer program benefits are fairly accurate when compared to administrative sources, the accuracy of SIPP individuals' earnings histories has received little attention.[1]

Our work on measurement error in the SIPP focuses on the extent of measurement error in self-reported annual earnings for the elderly and the non-elderly. We are primarily interested in the effect of measurement error on the mean and dispersion of the marginal distributions of earnings for people of different ages and on the correlation in earnings across years, which is a measure of earnings mobility. Measurement error may lead to bias estimates of all the moments of these distributions, but this is not a necessary consequence of measurement error. For example, if measurement is random then it will have no effect on the mean of the earnings distributions but will increase earnings dispersion. If, however, measurement error is negatively correlated with earnings, then measurement error may actually reduce earnings dispersion, giving a false impression of inequality.[2] Likewise if measurement error is positively correlated across time, then mobility will be overstated.[3]

---

[1] The two exceptions are Coder (1992), who focuses on married couples, and Abowd and Stinson (2005) who study the accuracy of the earnings attached to jobs rather than individual earnings. We contrast our work with theirs thorough out the paper since it is the closest to our work. For an example of a study of the accuracy of SIPP program benefit data, see Huynh, Rupp and Sears (2002) http://www.census.gov/dusd/MAB/wp238.pdf
For a preliminary comparison between SIPP and IRS earnings data, see Hendrick,. King and Bienias (2002) http://www.census.gov/dusd/MAB/wp211.pdf .

[2] Let E(t) be measured earnings in period t and E*(t) be actual earnings where E(t)= E*(t)+m(t) so m(t) is measurement earnings. Then the variance of measured earnings is

While our primary focus is on labor force and earnings dynamics as respondents near retirement age, we also present evidence for the wider non-retired population. This allows us to contrast the role of measurement error across age groups.

II. Methods

We follow the well-developed procedures used in most previous validation studies in Bound, Brown and Mathiowetz (2001). The standard procedure in these studies is to identify a source of earnings data that is assumed to be free of error. The best know, validation study of the PSID by Bound, Brown, Duncan and Rodgers (1994) uses payroll records from a large manufacturing firm, but most other studies use Social Security or IRS records. Since Social Security records collect information only on earnings in covered sectors and only up to the FICA max, this source of administrative data has serious limitations. We, therefore, match SIPP respondents with earnings from IRS W-2 forms, which we assume to be free of reporting error.

Once a comparison data set has been identified, it is necessary to match respondents in the survey with their records in the administrative data. Our focus on earnings of individuals allows us to use Social Security numbers as unique identifiers to match them with their IRS records. Since Social security numbers are missing for some respondents these records cannot be matched and are, therefore, dropped.

given by $var(E(t))=var(E^*(t))+2(cov(E^*(t),m(t))+var(m(t))$. If measurement errors is negatively correlated with earnings then $cov(E^*(t),m(t))<0$ and if $2|cov(E^*(t),m(t))|>var(m(t))$ then the variance of measured earnings will be smaller than the variance of actual earnings, $var(E(t))<var(E^*(t))$.

[3] Let the covariance of earnings across time be a measure of mobility. The covariance of measured earnings is given by $cov[E(t), E(s)]= cov[E^*(t), E^*(s)]+ cov[E^*(t), m(s)]+ cov[m(t), E^*(s)]+ cov[m(t), m(s)]$ so the covariance in measured earnings will be larger than the covariance in actual earnings if $cov[E^*(t), m(s)]+ cov[m(t), E^*(s)]+ cov[m(t), m(s)]>o$. Any or all of these terms may be positive.

Since Abowd and Stinson (2004) also use IRS data to access the importance of measurement error in the SIPP, it is important to contrast our methods. There are two major differences between our studies. The first is that their objective is to compare the administrative and respondent report of the earnings received on jobs that are reported in SIPP. Their focus on the earnings of jobs is more ambitious than our focus on annual earnings of individuals since they need to match both individuals on the basis of Social Security records and employers on the basis of Employer Identification Numbers (EIN). As Abowd and Stinson describe, matching employers requires that many cases with valid Social Security numbers be dropped because these individuals hold jobs with employers that cannot be accurately matched to EIN's. As we will show, this additional sample restriction leads to somewhat different samples.

The second major difference between their study and almost all other studies, including our own, is that they depart from the usual assumption that it is only the self-reported earnings that are reported with error. Abowd and Stinson argue that the administrative source may itself depart from the earnings measure assumed in the survey. For example, earnings reported on W-2 forms do not include pre-tax health care premiums paid by the employee or contributions to 401(k) plans that come out of earnings. These exclusions will lead to a difference between W-2 earnings and self reported earnings if respondents to the survey accurately report their pre-deduction earnings. Abowd and Stinson, therefore, develop a procedure that allows for potential measurement error in both data sources.

III. Data

Our SSIP analysis file includes respondents 25 and over who are not in school. We use the 1996 SIPP panel, which includes earnings data covering a 36 month period starting in April 1996. As in previous SIPP panels, respondents were interviewed every four months. At each interview, respondents were asked to report their earnings for the previous four months, with detailed information on up to two jobs.[4] Respondents were interviewed four months later, yielding a total of 36 monthly observations on earnings.

Since we compare SIPP earnings with annual earnings from administrative records, which are based on annual earnings as reported on W-2 forms, our SIPP analysis file is limited to respondents with valid earnings data in all twelve month of the calendar year.[5] Since SIPP respondents are interviewed every four months and since the first interview is not necessarily in May (which would cover January through April) and the last interview can be earlier than July (which is the earliest interview that would cover August through December) some observations are dropped because they are in a year that is not fully covered in the 1996 SIPP panel for that individual.[6]

We compare measures of annual earnings from the SIPP, with its counterpart constructed from the Detail Earning Records (DER) which come from the detail segment of the Social Security Administration's Master Earnings File.[7] Since the DER contains earnings information from W-2s for all jobs held by the individual, it does not suffer from the standard limitations of FICA records which are top-coded at the FICA max and exclude persons in sectors not covered by the FICA tax, such as state and local government workers. In order to construct an earnings variable that is as close as possible to a SIPP earnings variable, we exclude self-employment earnings and deferred earnings.[8] Since SIPP data is top-coded, we impose a similar top-code to the DER.[9]

---

[4] Earnings of individuals under 15 years of age are not recorded in the SIPP and negative reported earnings are censored at zero.

[5] Zero earnings are treated as valid values that indicate that the individual was not working.

[6] Respondents were divided into four rotation groups. Each rotation group was interviewed in a different month. Interviews covered the previous four months. Therefore, earlier rotation groups covered earlier time periods than later rotation groups.

[7] Earnings in both data sets are deflated by the CPI series CWUR0000SA0 using 1999 as the base year.

[8] Deferred earnings include contributions to retirement plans that are not taxable.

[9] We impose a $150,000 top-code on annual earnings in the DER and recode these topcoded values to the mean earnings of top-coded respondents classified by sex, race and labor force status. This differs from SIPP which top-codes monthly values that would

The data in the DER are matched to SIPP respondents on the basis of their self-reported Social Security numbers. Respondents who fail to give their Social Security numbers or give invalid Social Security numbers cannot be matched and are, therefore, dropped.

While this matching procedure allows us to compare reported earnings in SIPP with earnings in fairly inclusive administrative records, SIPP earnings may differ from DER earnings for several reasons. First SIPP respondents are only asked to report earnings on up to two jobs in any month. If the respondent held more than two jobs, either simultaneously or in succession, then the earnings in the additional jobs will be missed in the SIPP but not in the DER. While this is not likely to be a serious problem for prime aged individuals, it may affect some older workers who hold a series of part time jobs.

More importantly, SIPP fills in missing values wherever possible. Two methods are used, edits and imputations. Edits fill in missing values based on values in previous waves for the same respondent. Alternatively SIPP imputes missing values using a hot-deck procedure in which earnings are imputed from the observed values for similar respondents.[10] These two methods clearly introduce measurement error in individual earnings, even if they are unbias estimates.[11]

IV. Results

This section presents our results in four sections. The first section provides evidence that the sub-sample of the SIPP that we use in the following sections is representative of the full SIPP sample. We then turn to a comparison of the moments of the joint distribution of SIPP and DER log earnings. This is followed by an assessment of the impact of measurement error on parameter estimates and general conclusions based on our findings.

---

exceed $150,000 on an annual basis. See SIPP 2001 User Guide Appendix B Top-coding for detail.

[10] See SIPP User Guide pages 4-8 to 4-16 for detail.

Matched and Unmatched Cases

Since we can only compare annual earnings in the SIPP and the DER for a subset of observations that are matched to the DER, this opens the possibility that the comparison between reported and actual earnings that we document in the following section cannot be generalized to the full SIPP sample. While we cannot observe the measurement error for the unmatched cases, this section explores whether these unmatched cases differ systematically from the matched cases in terms of observed SIPP labor market characteristics. If measurement error is a function of these observed SIPP characteristics and if unmatched cases differ systematically on the basis of these characteristics, then this implies that the unmatched cases would have different measurement error than the matched cases.[12] This would limit our ability to generalize the findings in the following sections to the full SIPP.

Table 1 shows the number of person years cross-classified by whether the respondent was matched to the DER and whether the yearly observation had reported earnings in all twelve months of the calendar year. Of the 184,989 person years in the full sample, 36,137 had to be dropped because earnings did not cover the full calendar year. Of the remaining 148,852 cases, 23,329 cases were dropped because they were not matched to the DER. The overall match rate of 85 percent was not only high but also similar for observations that had valid earnings and those that did not. This resulted in an initial analysis sample of 125,523 person years.[13]

We now turn to a comparison of observed labor market characteristics of persons in the SIPP classified according to whether they were matched to the DER. If the matched and unmatched cases do not differ systematically on the basis of observed SIPP characteristics, then this indicates that there is no systematic bias on the basis of selection on observables. Since retirement research focuses specifically on the elderly, Tables 2-5

---

[11] Unbias estimates assign the correct value on average, but not necessarily the correct value for each individual.

[12] See Moffitt, Fitzgerald and Gottschalk (1999) for a discussion of selection on observables.

present match rates for observations classified by age of the respondent. For these and all following tables, we limit the sample to observations with valid earnings in all 12 months.

Table 2 shows that the SIPP sample includes a substantial number of matched observations for the elderly. The sample includes 23,744 person years for respondents 70 and over. This is not only large but the match rate is very close to the overall match rate of 84 percent, yielding a sample of 19,919 matched observations for respondents 70 and over. The fact that the match rate does not vary with age indicates that the selection effects of matching are likely to be similar for the elderly and the non-elderly.

Tables 3 through 5 provide descriptive statistics on the proportion of the observations in which the respondents are employed, and for those who are employed their mean earnings and mean hours. These tables confirm that these measures of labor market activity are similar in the matched sample and cases that could not be matched. Table 3 shows that employment rates start high and are very similar for matched and unmatched cases, the employment rate for males 25 to 29 matched to the DER is .878. For unmatched cases the employment rate is .866. While the difference between employment rates of similar aged matched and unmatched females is a bit wider (.711 and .648), it narrows for females in the next highest age range.

As expected, employment rates decline substantially for workers nearing retirement age. This decline is similar in both matched and unmatched cases. For example, males 70 and over matched to the DER have an employment rate of 5.5 percent. The employment rate for unmatched males in this age category is also 5.5 percent. The comparable employment rates for females are 3.2 and 2.6 percent.

Mean earnings and mean hours worked for those employed are shown in Tables 4 and 5. Again, these data show the expected patterns for both matched and unmatched cases.

---

[13] The analysis sample includes 31,947 persons out of 41,316 persons with valid earnings in at least one year.

Mean earnings are again similar for matched and unmatched respondents, starting at $27,127 for matched males 25 to 29 and $25,273 for unmatched males in the same age range ($16,855 and $14,295 for females). Mean earnings for both matched and unmatched cases increase until respondents reach their late 40s and then decline. By the time respondents reach their 70s, those still working are earning considerably less than at their peak earnings years.

Table 5 shows that hours worked among employed respondents are again similar for matched and unmatched cases. Both series show the expected patterns in which average hours of employed males largely reflect full-time work until these males approach retirement age, at which point average hours drop to roughly half-time. Females work somewhat fewer hours when they are young but show the same reduction as they reach retirement age. Average hours for workers 70 and over are similar for matched and unmatched males (1,169 and 1,189 hours) and for females (1,026 and 1,049).

We conclude that matched and unmatched observations in the SIPP have similar reported employment rates, average earnings and average hours. Since we cannot observe earnings in the DER for the unmatched SIPP observations, we cannot rule out that the patterns in SIPP and DER earnings, which we explore in the following section for the matched sample, would also hold for the unmatched observations. However, the evidence provided in this section indicates that there is little selection on observables. We, of course cannot, rule out selection on unobservables.

Comparison of SIPP and DER

We now turn to a direct comparison of the SIPP with the DER for the matched sample. We start by comparing measures of central tendency and dispersion for the marginal distributions of reported earnings in the SIPP and the marginal distribution of earnings in the DER. In order to make our findings comparable to findings in other validation studies, we present evidence on the distribution of log earnings rather than absolute earnings. In the following section, we turn to the potential bias caused by measurement error in the SIPP.

Joint Distribution of SIPP and DER Earnings

We start by comparing measures of employment and measures of central tendency and dispersion in the marginal distributions of positive earnings in the SIPP and DER.[14] Table 6 shows employment rates in the two data sets for persons classified by age.[15] Both data sets show the expected lifetime pattern in employment. Employment rates start high and decline rapidly as persons near retirement age.

While the life-cycle patterns are similar in the two data sets, the DER gives consistently higher employment rates than the SIPP, a finding that is consistent with Abowd and Stinson (2004). This indicates that some SIPP respondents fail to report earnings that are captured in the DER. These respondents with missing SIPP earnings tend to have lower earnings in DER than respondents with observed earnings in both data sets.[16] A smaller number of respondents had positive SIPP earnings but no earnings in DER, possibly reflecting informal work arrangements. These workers had lower SIPP earnings than workers with earnings in both data sets. Comparing earnings for persons with reported earnings in both data sets, therefore, leaves out some workers at the bottom of the earnings distributions.

Tables 8 and 9 compare the mean and variance of log earnings in the SIPP and DER for persons with positive earnings in both data sets. Earnings in both data sets exhibit the standard rapid increase in mean earnings while respondents are in their 20s and 30s. This growth in mean earnings continues through the late 40s for both males and females. Earnings profiles then flatten and then decline rapidly as workers over 60 cut back on hours.

---

[14] Employment is defined as having positive earnings during the year.
[15] The columns for SIPP are identical to the columns for matched cases in Table 3.
[16] Mean log earnings in the DER are 10.43 for males in both data sets. Males with missing earnings in SIPP have a mean of 9.94 in DER. The corresponding means for females are 9.26 and 8.30.

While lifetime patterns are similar in the two data sets, earnings are higher in the DER than the SIPP for males up through their late 50s. For older males and for females, there is no systematic difference between data sets. This indicates that males under 60 tend to understate their earnings in the SIPP.

Table 8 shows the dispersion in log earnings for persons with positive earnings within each age group. Not surprisingly, there is a sharp increase in dispersion in the older age groups as some respondents start cutting back on hours while others continue to work full-time. Once again, the two data sets show similar lifecycle patterns but the levels of dispersion are different in the two data sets, with the SIPP showing consistently lower dispersion of annual earnings than the DER.[17] This reflects longer tails at both ends of the DER distribution than the SIPP distribution.[18] The reason for these differences in dispersion across data sets is unclear.

Tables 9 and 10 show how the mean and variance of log earnings changed in these two data sets over time. This reflects both the effects of the aging of the panels and the pure effects of calendar time. Mean log earnings are constant for both males and females in the SIPP while the DER shows some growth in the mean. Neither data set shows a clear trend in dispersion.[19]

Having compared the marginal distributions of log earnings in these two data sets, we now turn to the correlation between DER and SIPP earnings. Table 11 presents the correlations for all reported earnings, whether imputed or not. Table 12 limits the sample to persons with non-imputed earnings. Both tables show correlations for respondents under 65 and for respondents 65 and over.

---

[17] While Abowd and Stinson (2004, Table 9) also find that the dispersion in the distribution of log earnings across jobs is lower in the SIPP than the DER, the difference in dispersion of log earnings across persons is even larger.

[18] The first percentiles of the DER and SIPP distributions are 7.29 and 7.80. The corresponding 99^{th} percentiles are 12.63 and 12.04.

[19] Abowd and Stinson (2004, Table 9) show growth in the means of the distributions of earnings by job in the two data sets. Dispersion in the DER is constant while it declines in the SIPP

Table 11 shows that the correlation between DER log earnings and SIPP log earnings for males less than 65 varies from .64 to .72 across years. For males over 65, the correlation starts higher than for the younger age group but declines over time, which may reflect the effects of non-random attrition through death and non-random exits from the labor market. Females exhibit similar patterns. The younger age group has correlations that vary between .68 and .72, while the older age group also shows correlations that decline over time. These correlations indicate that older SIPP respondents still working are not making larger reporting errors than younger workers.

Not surprisingly, Table 12 shows that these correlations increase when imputed earnings are dropped. The correlation between SIPP and DER non-imputed earnings varies from .72 to .79 for males under 65 and from .72 to .78 for females in the same age category. Older workers again have correlations that are similar to those of younger workers.

Since Abowd and Stinson (2004) restrict their analysis to carefully matched jobs that appear in both data sets, it is not surprising that they find somewhat higher correlations in SIPP and DER job earnings than we find in our measures of annual earnings of individuals that include the earnings of jobs in the DER but not the SIPP (which includes the earnings in only jobs in each month). The difference across these studies is, however, surprisingly small. While we find correlations around .75, the correlations they find range from .83 to .85.[20]

The correlations we find can also be contrasted with the results of the validation study of the PSID reported in Bound, Brown, Duncan, and Rogers (1994). This study compares earnings from the payroll records of a large manufacturing firm with reported earnings from a PSID questionnaire administered to 418 workers. This validation study finds that the correlation between PSID earnings and payroll records are .89 and .92 in the two

[20] See Table 12 of their paper. Coder (1992) finds a correlation of .83 between the reported SIPP earnings and IRS data but his sample is restricted to married couples who could be matched to a joint return.

years they studied.[21] As the authors acknowledge, these high correlations reflect the fact that hourly workers in the firm were unionized and that the firm had a highly compressed earnings distribution. This made it easier for respondents to report their earnings annually.

In summary, we find annual earnings as reported in the SIPP show less dispersion than annual earnings in the DER but that the correlation between these two measures is fairly high, especially when imputed earnings are dropped. Furthermore, these patterns are similar across age groups, indicating that any bias caused by measurement error is not likely to be worse for older workers nearing retirement. We now turn to the impact of measurement error in SIPP earnings on regression coefficients that use SIPP earnings as a dependent or independent variable.

Impact of Measurement Error

The impact of measurement error on parameter estimates depends crucially on whether measurement error is random or whether it is correlated with the true level of earnings, which we take to be DER earnings. If measurement error is uncorrelated with earnings, then this classical measurement error will not affect regression coefficients as long as the earnings variable is the dependent variable. If the earnings variable is an independent variable, then even classical measurement error will lead to bias coefficient estimates. This bias carries over to other independent variables correlated with earnings. It is often asserted that classical measurement error will lead to downward bias in regression coefficients, but such bias cannot be signed in the multivariate case unless the regressors are orthogonal. As Bound, Brown and Mathiowetz (2001) show, it is impossible to sign the direction of the bias if more than one independent variable is measured with error.

If measurement error is not classical, then bias will result even if mismeasured earnings is used as a dependent variable. Furthermore, one can no longer sign the bias even if earnings is the only independent variable. If the measurement error is sufficiently

[21] Correlations reported in Bound, Brown and Mathiowetz (2001).

negatively correlated with true earnings, then this can even lead to upward bias in estimated coefficients.

Bound, Brown and Mathiowetz (2001) show that when mismeasured earnings, E, is used as the dependent variable, the bias in the regression coefficients depends on the correlation between actual earnings, E*, and the measurement error, m, (i.e. the bias depends on cov(E,m) where E=E*+m.) They, likewise, show that if earnings is a single mismeasured independent variable, then the bias depends on cov(E*,m). Specifically they show that the size of the bias can be obtained from the regression coefficient on E or on E* in a set of auxiliary regressions with m as the dependent variable. We denote these regression coefficients from the auxiliary regressions $\beta_{m,E}$ and β_{m,E^*}.

A central issue in determining the likely impact of measurement error on estimated regression coefficients that use log earnings as a dependent or independent variable is whether the measurement error is classical. This is obtained from the auxiliary regression of measurement error on DER earnings (which we use as our measure of E*.) If the coefficient on DER earnings is not zero, then the measurement error is not classical and $\beta_{m,DER}$ gives the bias in regression coefficients when SIPP log earnings is used as a dependent variable. Bias arising out of non-classical measurement error when SIPP log earnings is used as the independent variable in a bivariate regression is given by the coefficient on SIPP log earnings in the second auxiliary regression. We denote this coefficient as $\beta_{m,SIPP}$.

Table 13 shows the relevant regression coefficients from the two auxiliary regressions[22]. Row 1 shows our estimate of $\beta_{m,DER}$. This estimated coefficient of -.39 for males and -.37 for females indicates that measurement error is negatively correlated with DER earnings for both males and females. This negative relationship is especially strong when earnings are imputed.

[22] m is calculated by the difference between ln(SIPP earnings) and ln(DER earnings).

The negative correlation between DER earnings and measurement error is consistent with our finding that earnings inequality is considerably higher in the DER than in the SIPP data. This negative correlation implies that workers with low earnings tend to overstate their earnings, while respondents with high earnings tend to underreport their earnings in SIPP. This reduces inequality of reported earnings in the SIPP.

As Bound, Brown and Mathiowetz (2001) show, these coefficients indicate that measurement error in the SIPP tends to decrease regression coefficients when SIPP earnings are used as the dependent variable. This table shows that the non-classical nature of measurement error in SIPP earnings decreases regression coefficients by over 35 percent when SIPP earnings are the dependent variable. This proportional bias is reduced to 30 percent if imputed earnings are excluded.

The second row of coefficients gives the impact of the non-classical measurement error when SIPP earnings are used as an independent variable. Let $\beta_{Y,SIPP}$ be the estimated coefficient in a regression of Y on SIPP log earnings. If the measurement error were classical and SIPP earnings were the only independent variable, this would lead to a downward bias in the coefficient on earnings. Bound, Brown and Mathiowetz (2001) show that the proportional bias (i.e. $\beta_{Y,SIPP}$ / $\beta_{Y,DER}$) is given by (1- $\beta_{m,SIPP}$).[23] The larger the value of $\beta_{m,SIPP,}$ the smaller the bias.

Estimates of $\beta_{m,SIPP}$ are given in the second row in Table 13. These estimates are .18 for males and .21 for females. This implies that regression coefficient estimates using SIPP earnings as an independent variable would be roughly 80 percent as large as the coefficient estimates researchers would obtain if they had access to DER earnings data.[24]

[23] $\beta_{m,SIPP}$ would be equal to the ratio of measurement variance to total variance if the measurement error were classical (i.e. $\beta_{m,DER}$ =0).
[24] This holds only for bivariate regressions. Bound, Brown and Mathiowetz (2001) show how the bias depends on the full covariance of regressors in the multivariate context.

The non-classical nature of measurement error in earnings data is consistent with the findings in a number of earnings validation studies reviewed in Bound, Brown and Mathiowetz (2001, Table 1). Our estimates of $\beta_{m,DER}$ of roughly -.30 for non-imputed earnings are somewhat larger than the estimate of -.17 for the PSID found in Bound, Brown, Duncan, and Rogers (1994). However, our estimates of $\beta_{m,Sipp}$ of .18 for males and .21 for females are similar to the estimates of .24 in the PSID validation study.

As an example of the impact of measurement error on parameters of interest, Table 14 shows transition rates between 1996 quintiles and 1999 quintiles. While it is often assumed that measurement error increases mobility, this is not necessarily true if measurement error is negatively correlated with earnings, which we have shown to be the case in SIPP. As Table 14, indicates SIPP earnings show less mobility than DER earnings. The DER transition rates in the top panel show that for respondents starting in the lowest quintile, the probability of staying in that quintile is .74. When persons are classified by their SIPP earnings, the probability of staying in the lowest quintile increases to .83. The probability of staying in the top quintile also increases from .82 when DER earnings are used to .83 when SIPP earnings are used. When the sample is limited to respondents with non-imputed SIPP earnings in either year, these probabilities increase to .88 and .85.

Conclusions

This study has shown that lifecycle changes in the mean and variances of log earnings are similar in the SIPP and DER. We have also shown that while the correlation in log earnings in these two data sets is somewhat lower than correlations found in the PSID, these differences are consistent with the differences in samples. Furthermore, measurement error is no larger for older workers than for younger workers. This is important since it might be assumed that older workers, who are the subjects of interest in retirement research, are more likely to make errors in reporting annual earnings. While older workers are not more likely to have measurement error, workers of all ages with imputed earnings have larger measurement error. These workers should be dropped in order to reduce the role of measurement error.

Our study finds that measurement error in the SIPP is not classical measurement error. We find the same negative correlation between earnings and measurement error in the SIPP as has been found in the PSID. This negative correlation is one possible explanation for our finding that the variance of SIPP log earnings is smaller than the variance of DER earnings, since the negative correlation implies that workers with low earnings tend to overstate their earnings, while respondents with high earnings tend to underreport their earnings. This non-classical nature of the measurement error in SIPP tends to lessen the measurement error bias when earnings are used as an independent variable but tends to introduce bias when earnings are used as a dependent variable.

Reference

Abowd ,J, and M. Stinson "Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Survey and SSA Administrative Data" mimeo July 2004.

Bound, J. C. Brown, G. Duncan, and W. Rodgers "Evidence on the Validity of Cross-sectional and Longitudinal labor Market Data" Journal of Labor Economics 12(3) 1994.

Bound, J., C. Brown, and N. Mathiowetz. "Measurement Error in Survey Data" Handbook of Labor Economics volume 5, Chapter 59 J. Heckman and E. Leamer editors Amsterdam: Elsevier 2001.

Coder, J. "Using Administrative Record Information to Evaluate the Quality of Income Data Collected in the Survey of Income and Program Participation" Proceedings of Statistics Canada Symposium 92 (Statistics Canada, Ottawa) 1992.

Moffitt, R, J. Fitzgerald and P. Gottschalk "Sample Attrition in Panel Data" The Role of Selection on Observables" Annales: d'Economie et de Statistique, 55-56, 129-152. December 1999.

Hendrick, M, K King and J. Bienias "Research on Characteristics of Survey of Income and Program Participation Non-respondents Using IRS Data," US Census Bureau, SIPP Working Paper #211.

Huynh, M, K. Rupp and J. Sears "The Assessment of Survey of Income and Program Participation (SIPP) Benefit Data Using Longitudinal Administrative Records", US Census Bureau, SIPP Working Paper #238, June 2002.

Moffitt, R, J. Fitzgerald and P. Gottschalk "Sample Attrition in Panel Data" The Role of Selection on Observables" Annales: d'Economie et de Statistique, 55-56, 129-152. December 1999.

Table 10-- Variance of ln Earnings in SIPP and DER						
by Gender and Year						
	Males			Females		
	SIPP		DER	SIPP		DER
	All	non-imputed		All	non-imputed	
1996	0.580	0.558	0.721	0.634	0.659	0.836
1997	0.565	0.547	0.685	0.647	0.700	0.786
1998	0.551	0.542	0.756	0.678	0.742	0.803
1999	0.558	0.554	0.854	0.657	0.718	0.827

Table 11--Correlation of ln Earnings in SIPP and DER				
by Age Gender and Year				
	Males		Females	
	Age<65	Age>=65	Age<65	Age>=65
1996	0.696	0.774	0.683	0.673
1997	0.717	0.750	0.719	0.713
1998	0.710	0.618	0.704	0.673
1999	0.648	0.685	0.694	0.599

Table 12--Correlation of Non-imputed ln Earnings in SIPP and DER-- by Age Gender and Year				
	Males		Females	
	Age<65	Age>=65	Age<65	Age>=65
1996	0.767	0.832	0.717	0.711
1997	0.789	0.816	0.784	0.750
1998	0.787	0.741	0.747	0.676
1999	0.717	0.758	0.756	0.597

Table13-- Auxilliary Regression Coefficients						
	Males			Females		
	All	Non-imputed	Imputed	All	Non-imputed	Imputed
ln(DER)	-0.389	-0.301	-0.529	-0.367	-0.299	-0.516
	[0.003]	[0.004]	[0.006]	[0.004]	[0.004]	[0.007]
ln(SIPP)	0.182	0.131	0.278	0.212	0.191	0.271
	[0.004]	[0.005]	[0.009]	[0.005]	[0.005]	[0.010]
Note:Coefficients in auxilliary bivariate regressions with measurement error as the dependent variable						

Table14-- 1996 to 1999 Transition Probabilities in SIPP and DER					
All Respondents with Positive Earnings in Both Years					
DER			1999 Quintile		
1996 Quintile	1	2	3	4	5
1	*0.74*	0.20	0.04	0.01	0.00
2	0.15	*0.61*	0.20	0.04	0.01
3	0.06	0.15	*0.59*	0.18	0.03
4	0.03	0.03	0.15	*0.65*	0.14
5	0.02	0.01	0.02	0.13	*0.82*
SIPP			1999 Quintile		
1996 Quintile	1	2	3	4	5
1	*0.83*	*0.13*	*0.02*	*0.01*	*0.01*
2	*0.13*	*0.70*	*0.14*	*0.03*	*0.00*
3	*0.03*	*0.14*	*0.65*	*0.17*	*0.02*
4	*0.01*	*0.03*	*0.16*	*0.66*	*0.14*
5	*0.01*	*0.01*	*0.03*	*0.13*	*0.83*
Sipp non-imputed			1999 Quintile		
1996 Quintile	1	2	3	4	5
1	*0.88*	0.11	0.01	0.00	0.00
2	0.10	*0.76*	0.13	0.01	0.00
3	0.01	0.12	*0.71*	0.15	0.01
4	0.00	0.01	0.14	*0.71*	0.14
5	0.00	0.00	0.01	0.13	*0.85*
Note: Each observation is a person year					

Recent Working Papers from the Center for Retirement Research at Boston College

Investment Choice in the Swedish Premium Pension Plan
Mårten Palme, Annika Sundén, and Paul Söderlind, April 2005

Design and Implementation Issues in Swedish Individual Pension Accounts
R. Kent Weaver, March 2005

Lashed to the Mast?: The Politics of Notional Defined Contribution Pension Reforms
Sarah M. Brooks and R. Kent Weaver, January 2005

Understanding Expenditure Patterns in Retirement
Barbara A. Butrica, Joshua H. Goldwyn, and Richard W. Johnson, January 2005

Changes in the Distribution of Long-Run Earnings and Retirement Incomes—Have Recent Cohorts Fallen Behind?
Peter Gottschalk and Minh Huynh, January 2005

The Age Profile of Income and the Burden of Unfunded Transfers in Four Countries: Evidence from the Luxembourg Income Study
Gary Burtless, December 2004

Projecting Immigration: A Survey of the Current State of Practice and Theory
Neil Howe and Richard Jackson, December 2004

Nonearnings Income Migration in the United States: Anticipating the Geographical Impacts of Baby Boom Retirement
Peter B. Nelson, December 2004

Does Work Pay at Older Ages?
Barbara A. Butrica, Richard W. Johnson, Karen E. Smith, and Eugene Steuerle, November 2004

Poverty and Income Maintenance in Old Age: A Cross-National View of Low Income Older Women
Timothy M. Smeeding and Susanna Sandstrom, November 2004

All working papers are available on the Center for Retirement Research website (http://www.bc.edu/crr) and can be requested by e-mail (crr@bc.edu) or phone (617-552-1762).