

**AN ANALYSIS OF SAMPLE SELECTION AND THE RELIABILITY OF USING
SHORT-TERM EARNINGS AVERAGES IN SIPP-SSA MATCHED DATA**

by

Jonathan Davis *
Federal Reserve Bank of Chicago

and

Bhashkar Mazumder *
Federal Reserve Bank of Chicago

CES 11-39

December, 2011

The research program of the Center for Economic Studies (CES) produces a wide range of economic analyses to improve the statistical programs of the U.S. Census Bureau. Many of these analyses take the form of CES research papers. The papers have not undergone the review accorded Census Bureau publications and no endorsement should be inferred. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Republication in whole or part must be cleared with the authors.

To obtain information about the series, see www.census.gov/ces or contact Cheryl Grim, Editor, Discussion Papers, U.S. Census Bureau, Center for Economic Studies, 4600 Silver Hill Road, Washington, DC 20233, CES.Papers.List@census.gov.

Abstract

In this paper, we document the extent to which the sample of the Survey of Income and Program Participation that is matched to the Social Security Administration's administrative earnings records is nationally representative. We conclude that the match bias is small, so selection is not a serious concern. The matched sample over-represents individuals who are wealthy, who have financial assets or who have received a government-transfer and under-represents individuals who attrited from the SIPP. We use this matched sample to examine the relationship between short-term averages of earnings from the SIPP earnings and average lifetime earnings from the administrative records. Our estimates suggest that using short averages of earnings may understate the effects of permanent income on particular outcomes of interest.

*Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. We thank Nathan Chan and Tony Zhang for excellent research assistance.

Introduction

The data linking the Census Bureau's Survey of Income and Program Participation (SIPP) to earnings records maintained by the Social Security Administration (SSA) offers researchers the potential to conduct innovative and wide-ranging research. The combination of rich survey data from the SIPP and a long panel of administrative earnings derived through tax records facilitate research on topics that cannot otherwise be addressed or, only in a more limited fashion. For example, the SIPP-SSA matched data has been used to study the effects of the Earned Income Tax Credit (Dahl, Deleire and Schwabel, 2009), the effects of parent income on childhood wellbeing (Mazumder and Davis, 2011), the effects of earnings volatility on health insurance coverage (Rutledge, 2011), the rate of immigrant earnings assimilation (Lubotsky, 2007) and intergenerational earnings mobility (Mazumder, 2005).

Although there are considerable advantages to this data, there are some potential issues that arise with its use. We focus on one of the more salient concerns which is the degree to which the matched sample is nationally representative. Although, the SIPP is a nationally representative survey, not all individuals are successfully matched to the SSA administrative data raising concerns about possible selection in the ensuing matched sample. Selection may also arise in the use of other similar matched data sources such as links between the Health and Retirement Survey (HRS) and SSA earnings data (e.g. Haider and Solon, 2006) and the Continuous Work History Sample (CWHS) (e.g. Kopczuk, Saez and Song, 2010). The Census Bureau, SSA and IRS have also been developing the SIPP/SSA Public use file (PUF) which aims to integrate survey data and administrative earnings records in a form that would be usable by the research community while maintaining the confidentiality of survey respondents. Therefore, an analysis of selection in the SIPP-SSA data potentially has broader applicability.

This paper addresses this by using several of the matched SIPP-SSA samples and systematically examining sources of potential bias due to matching. We document specific characteristics that are over-represented in the matched sample. Among these are individuals with financial assets and those who have received government transfers. Overall, though, the sizes of the effects are not especially large, so we do not think that selection is an especially serious concern. Nevertheless, we caution that researchers should consider the nature of their specific application.

We also explore the quality of short-term survey data on earnings contained in the SIPP relative to measures of lifetime earnings using the SSA data by estimating reliability ratios. The results are in line with previous work suggesting that the use of short-term time averages of earnings from survey data may significantly understate the effects of permanent income on particular outcomes of interest.

The rest of the paper proceeds as follows. We begin by providing some background on SSA's Master Earnings File (MEF). We then describe the process used in creating the matched data. We next describe the actual matched samples we use and provide some basic descriptives of the match rates. We then use cross-tabulations and regression analysis to more systematically analyze the degree of sample selection. Finally we present estimates of the reliability of using short-term time averages of earnings from the SIPP instead of long-term averages from the SSA administrative records.

Background on the MEF

The Master Earnings File (MEF) is used by SSA to track the earnings of individuals that may receive SSA benefits. The file has two main components: the Summary Earnings Records (SER) and Detailed Earnings Records (DER). The SER Earnings measure is annual FICA covered earnings from 1951 until 2007. SER earnings are censored at the FICA taxable

maximum (\$106,800 in 2010). The DER extract includes earnings information from 1978 until 2007 that is taken directly from the individual's W-2 tax form, so DER earnings should not be censored at the FICA.

Olsen and Hudson (2009) provide a detailed discussion of the history and mechanics of the MEF and make clear that the earnings data is not error free. In particular, SSA must match all reported earnings to its Numerical Identification (NUMIDENT) file. Apparently, 90% of all earnings are matched using name and social security number without issue and another 6% are matched using additional matching algorithms. Therefore, 4% of reported earnings do not actually make it into the MEF. Instead, these earnings are placed in the Earnings Suspense File until they can be successfully matched to an individual. Moreover, it is possible that some of the "successful" matches are spurious.

Olsen and Hudson also highlight the fact that not all workers are covered by FICA. For example, civilian federal employees hired before January 1, 1984, railroad workers and some self-employed workers with low earnings are not covered. Such workers who show up in the MEF will show zero earnings in the SER and DER files.¹

The Process of Matching

We begin by reviewing the mechanics of how individuals are matched between the SIPP and SSA files. Respondents are typically asked for their social security number (SSN) the first time they are interviewed. By providing a social security number, even an incorrect one, respondents give the Census Bureau consent to match a respondent's SIPP record to the

¹ A less known issue is that there are a small number of observations in the DER that have two implied decimal places. We located these observations by checking to see if DER earnings were exactly 100 times SER earnings. Since SER earnings are capped at the FICA taxable maximum and DER earnings are not, there are many situations where DER earnings are legitimately higher than SER earnings. Consequently, it may be difficult to identify which observations incorrectly have 2 implied decimal places. A conservative solution would be to drop any DER values that are more than 100 times greater than the SER earnings in a particular year.

MEF. Respondents who refuse to provide their social security numbers are not matched to the MEF.

The Census Bureau first attempts to validate matches by confirming that the name, address, birth date, gender and race in the SIPP matches those same fields in SSA's NUMIDENT file (Abowd, Stinson and Benedetto 2006). If the reported SSN appears to be incorrect, Census then tries to match the person using demographic characteristics. Vilhuber and Pedace (2003) review the rates of different SSN match outcomes for the 1990 through 1993 panels. They show that 6 to 9 percent of respondents refused to provide a SSN and were not matched. Close to 60 percent of respondents provided a valid SSN and slightly more than 80 percent of respondents were eventually matched to a MEF record.

Therefore, there are three distinct reasons why a SIPP respondent might not be successfully matched to the MEF: (i) they refused to provide a SSN (ii) they do not have a SSN (iii) they provided an invalid SSN and Census was not able to match them successfully.

Data

We use the 1984, 1990, 1991, 1992 and 1993 SIPP full panel files and the 1996, 2001 and 2004 SIPP core wave files matched to extracts from the SSA's Master Earnings File (MEF).² All of the analysis was conducted inside a Census Research Data Center (RDC) and the extracts we were provided did not contain SSN but instead appended the SIPP public use identifier to the administrative records.³ We confine our analysis to only the subsample of SIPP respondents who could have been matched to the earnings files.⁴

² We did not have access to the 2008 SIPP panel. The Census Bureau did not obtain social security numbers from respondents in the 1985 to 1989 SIPP panels.

³ The project was authorized by SSA, IRS and Census and the authors obtained special sworn status from the Census Bureau.

⁴ This was determined by an analysis of the match rate by the wave of entry into the survey. For example, since no one who entered the 1984 panel in waves 8 or 9 was matched, we inferred that there was no attempt to match individuals in these waves so we excluded them from the analysis. In all, we dropped individuals who entered in waves 8 or 9 of the

We made some further restrictions to the sample matched to the DER. In order to appear in the DER file, an individual must have had W-2 earnings sometime between 1978 and 2007. As a result, individuals who were never of working age during this period are unlikely to appear in the DER file. In contrast individuals of any age may potentially have a valid SSN. To address this issue, we restricted our SIPP sample to individuals born between 1918 and 1988 (older than 19 in 2007 and younger than 60 in 1978) when working with the DER data. The earnings measure that we focus on is annual earnings from wages, tips and other compensation as reported in Box 1 of the W-2 tax form.

Table 1 shows the sample sizes of the relevant SIPP universe for matching to the SER and DER files along with the match rates. These rates are shown separately for each SIPP panel. The match rates are roughly in the low 80 percent range for the 1984 and the 1990-1993 panels. However, match rates fell sharply in subsequent SIPP panel years. Czajka, Mabli and Cody (2008) have previously documented this result for the SIPP-SER match and argued that the lower match rates did not increase selection bias. We will return to this issue and extend the analysis to the SIPP-DER match.

Comparing the match rates between the SIPP to the SER and the SIPP to the DER, we find that the match rates to the DER are in all cases lower. Indeed, it appears that the match to the DER is always a subset of cases of successful matches to the SER. It is worth reiterating that we find lower match rates to the DER extract even after removing those individuals who are not of working age, suggesting that the DER match rates would be even lower if we compared the full samples (unconditional on age). The difference in match rates

1984 SIPP, wave 9 of the 2001 SIPP or waves 9 through 12 of the 2004 SIPP. Individuals who entered in wave 10 of the 1992 SIPP were not included in the analysis because we weighted our results using the interview weights from the core wave files, which we merged in to the full panel files, and these weights were not available for wave 10 of the 1992 panel. Additionally, respondents did not consent to being matched until wave 2 of the 2001 Panel, so anyone who attrited after wave 1 of the 2001 SIPP panel was dropped.

between the SER and DER appears to have narrowed to just 1 percentage point for the 2001 and 2004 panels.

Since the SIPP is a panel where households are tracked over time and interviewed many times, a certain amount of attrition is inevitable. Certain individuals may be particularly difficult to track or may not be as strongly tied to the formal economy, therefore, attrition status might conceivably serve as a useful proxy for individuals who might also be difficult to match to administrative records. Table 2 shows the match rates stratified by attrition status for each SIPP panel in our analysis. For the 1984 panel there is only about a 2 percentage point difference in the match rate between those who attrite and those who do not. The gap increases to between 7 and 12 percentage points for the 1990 through 2001 SIPP panels and is especially severe at over 30 percent for the 2004 panel. The patterns are very similar between matches to the SER and DER. Given that the average match rate across all panels between the SIPP and the SER is just over 75%, the effects due to attrition are clearly quantitatively meaningful.

Figure 1 shows match rates by age⁵ for each panel for the SER and DER. The most striking patterns are the low match rates for young children in the 1984 panel. This pattern gradually disappears for the SER in later SIPP panels. This is likely due the fact that SSA created a program in 1987 that issued SSNs to newborn infants when their birth was registered with the state. Notably, this program was implemented nationwide by 1989 (SSN Chronology). By construction, the low match rates for young children in the DER are not present in later SIPP panels since we dropped anyone born after 1988. Limiting the age range to 25-55 year olds, the match rates look much more stable within each panel.

⁵ We defined age as the individual's age at the end of the SIPP panel year. For example, if a respondent was born in 1970 and was part of the 1996 SIPP panel, we would list their age as 26. We coded children born after the start of the SIPP as 0 for this exercise because of disclosure concerns.

Analysis of Selection in the Matched Sample

Mean Comparisons

We begin by using a simple analysis of differences in mean characteristics among the matched sample versus the universe in order to get a general idea of what factors might matter most. We then turn to a more sophisticated multivariate approach where we can account for all the factors at once. In Table 3 we present a set of summary statistics for our full SIPP sample, the subsample that was matched to the SER (“Matched to SER”), the SIPP subsample of respondents born between 1918 and 1988 (“DER Full Sample”) and the subsample of the previous sample that was matched to the DER (“Matched to DER”).

The top panel of Table 3 shows summary statistics when we pool all seven SIPP panels. We chose to examine a range of characteristics that we hypothesized were likely to influence the likelihood of either having a valid SSN or refusing to provide one to the interviewer. In part, this was based on a previous analysis of the match rate for the 1984 SIPP-SER described in Mazumder (2005) who argued that that households of both lower and higher socioeconomic status may be over-represented. In addition to showing the raw sample means we also show the percent difference in these means between the matched samples and the relevant universes from which they are derived.

In percentage terms, the largest difference in means for the SER match is whether one was ever in the armed forces. For example, in the full SIPP sample, approximately 10 percent had served in the military. Among those matched to the SER the rate was around 12 percent. The percentage difference (using data that is not rounded) is 13.4 percent. Other important characteristics are whether one has a bank account (8.2 percent) or another form of financial asset (5.7 percent). This suggests that match rates may be higher for wealthier individuals. Indeed we find that mean personal income is 4.5 percent higher in the matched SER sample. The differences in mean characteristics between the universe

and matched sample are generally similar with respect to the to the DER data. Differences in mean personal income, however, are a bit larger at 5.9 percent. There were also bigger differences in the rate of government aid receipt in the case of the DER match (-8.0 percent) compared to the SER match (-1.1. percent).

The remaining panels in Table 3 show the same comparisons by SIPP panel. For simplicity we group the 1990-1993 panels which were very similar in their execution. In the 1984 panel the SER matched sample is 16% older, has 18% more years of education, has 20% higher personal income and was 21% more likely to be in the military. The matched sample means of average family income, the share of respondents who are white, male, ever married, employed, home owners or the share of respondents who have a bank account, who have another financial asset or who received any government aid are all within 5% of the means of the full sample. Moving to the DER, the patterns look broadly similar to the SIPP-SER match, except that the employment rate of the matched SIPP-DER sample is 5% higher than the universe, and the receipt of government aid is 16% lower in the matched sample.

Moving to the third panel where we examine characteristics in the 1990-1993 SIPP panels, we find fewer cases of large differences between the matched samples and the universe compared to the 1984 panel. In particular, mean differences are never more than 7 percent higher in the matched samples. The SIPP-SER matched sample has 5% higher personal income, was 5% more likely to have a financial asset other than a bank account and was 6% more likely to have been in the armed forces. The SIPP-DER matched sample had 7% higher personal income, was 5% more likely to have a financial asset other than a bank account, was 7% more likely to have ever been in the armed forces and was 7% less likely to have received any government aid.

When we move to the 1996 panel we begin to see much sharper differences in coverage by financial asset ownership. For the 1996 panel, the sample matched to the SER was 7% more likely to have a bank account and was 8% more likely to have another financial asset than the full SIPP sample. The sample matched to the DER had 6% higher personal income, was 7% more likely to have a bank account, was 9% more likely to have another financial asset, was 7% more likely to have served in the armed forces and was 6% less likely to have received government aid. For the most part, the measures where we see significant differences between the full and matched samples remain the same in 2001 for both the SER and DER even though the match rates were significantly lower.

In the 2004 panel, like the 1996 and 2001 panels, the respondents in the SER matched sample had higher personal incomes were more likely to have had a bank account, to have had another financial asset and to have ever been in the military. But unlike the earlier panels, the matched sample also had 7% higher family income and had almost 5% more white respondents. Additionally, the SER matched sample had 8% more homeowners than the full sample and was almost 6% less likely to have received government aid. Similarly, the respondents in the DER matched sample had higher personal and family income, were more likely to be home owners, to have a bank account or another financial asset or to have ever been in the armed forces and were less likely to have received any government aid.

Although we are primarily interested in differences between the matched subsample and the SIPP sample that we attempted to match—the full SIPP sample for the SER and the sample born between 1918 and 1988 for the DER—it is worth noting that the sample born between 1918 and 1988 is 8% older on average, had 14% higher personal income, were 10% more likely to be employed, were 11% more likely to have a bank account, were 13% more likely to have another financial, were 10% more likely to have

ever served in the armed forces and were 15% less likely to have received government aid. As one might expect, the differences between the full SIPP sample and the universe used to match to the DER, change over time. In the 1984 panel we only drop older individuals because the panel did not extend to 1988 whereas in the 2004 panel we drop all respondents who are under 19 and older than 86, so most of the dropped individuals are younger rather than older.

Regression Analysis

In Tables 4a and 4b we extend the means comparisons to a multivariate setting. We estimate linear probability models using Ordinary Least Squares (OLS). The dependent variable in the regression is an indicator that equals 1 if the respondent was matched and 0 otherwise. We now enter all of the characteristics simultaneously to estimate the effects of particular characteristics conditional on all of the others. As before, we chose covariates that are potentially correlated with either having a valid SSN or the likelihood of refusing to provide one. The covariates include age, gender, race, ethnicity, marital status, education, employment status, veteran status, personal and family income as well as whether someone is a home owner, has a bank account or another financial asset or has ever received WIC, AFDC, Food Stamps, General Assistance, Foster Care, Medicare, Medicaid, Other Welfare, Social Security Benefits, Railroad Retirement, Veterans Assistance, Rent Assistance or Subsidized Housing.⁶ We also include indicator variables for the entry wave.

In Table 4a, we begin with our analysis of the SIPP-SER sample. Column (1) presents coefficients for the sample that pools all 7 SIPP panels while column (2) through (6) show the results for specific SIPP panel years (or groups). When we pool the samples

⁶ Since respondents were generally asked for their SSN during their first interview, we coded these measures to reflect the first four reference months that the respondent was in the panel. For respondents who entered the 2001 panel in wave 1, we coded these to reflect their wave 2 responses, since this is when they were asked their SSN.

we also add indicator variables for each of the SIPP panels (or groups). For the pooled sample we explain about 15 percent of the variance. It is worth noting that a comparison of the R^2 across the columns indicates that our models are predicting a much greater share of the variance for the 1984 panel than in the other panels. This is entirely a consequence of the low match for children less than 15 in the 1984 panel. If we estimate the regressions using only respondents who are 15 or older (not shown) the R^2 are comparable across all of the panels.

Looking at the individual coefficients in column (1) of Table 4a the coefficients with the largest magnitude (in absolute value), are the indicators for the wave of entry (after wave 1). For example, the coefficient on entering in Wave 8, which was the last wave of the 1990, 1991, 2001 and 2004 panels was -0.33. This means that, all else equal, individuals who entered in the 8th wave of a SIPP panel were 33 percentage points less likely to be matched to the SER earnings data than a respondent who entered in wave 1. To put this in perspective, since the average match rate across all panels is approximately 75%, a 33 percentage point reduction implies an effect size of 44 percent evaluated at the mean. Other than the entry wave, the only other covariates that are associated with a 10 percentage point or larger drop in the probability of being matched are indicators for panel year. This suggests that for the most part, bias due to sample selection does not appear to be an especially serious concern.

Other covariates with potentially important, albeit smaller effects are receipt of: food stamps, rent assistance, Medicaid or Social Security. These increase the probability that a respondent was matched by between 3 and 7 percentage points. Therefore in contrast to the simple means comparison, the multivariate analysis shows that receipt of government aid raises the probability of a match and suggests that recipients of government transfer programs are more likely to be represented in the sample.

Relative to those in the fifth decile of the family income distribution, being in the bottom decile, reduces the probability that someone was matched by 5 percentage points. Therefore, conditional on controlling for government program participation, those of low socioeconomic status are less likely to be matched. We also find that being Hispanic, Black or another race other than White reduces the match rate by 5, 4 and 3 percentage points, respectively.

In contrast, to the simple means comparisons, we now find that financial asset ownership matters much less. For example, having a bank account or having another financial asset increases the match rate by between 3 and 4 percentage points. Similarly, having served in the military now only has a 3 percentage point effect on the likelihood of being matched.

The sign of the coefficients are generally stable across SIPP panels but the magnitudes do vary considerably. For example, being a home owner has no greater than a 2 percentage point effect on match rates prior to the 2004 panel, but increases match rates by 12 percentage points in the 2004 panel. This could be related to the many “subprime” mortgages that enabled some households to enter the housing market during the residential real estate boom. Many government assistance programs, like AFDC/TANF, WIC, Food Stamps and Social Security, had large positive effects on match rates in the 1984 panel and much smaller positive effects in later panels. The coefficient on AFDC/TANF benefits bucked the trend of the direction of coefficients remaining constant across panels, declining from a 6 percentage point increase in the 1984 panel to a 5 percentage point decrease in the 2001 panel.

Although the linear probability model estimates are broadly similar for the SER and DER there are some important differences. Moving to Table 4b, we show that being employed increased a respondent’s chance of being matched by 7 percentage points for the

DER. The corresponding effect for the SER was only 2 percentage points. The magnitude of this effect was primarily driven by a 15 percentage point increase in the probability of being matched for respondent's who were employed in the 1984 SIPP, but the coefficients are larger for the DER across all panels. Being more educated also seems to increase the probability of being matched with the DER more than with the SER, depending on the particular level, respondents with more than a high school diploma were 3 to 4 percentage points more likely to have been matched.

As with the SER results, receiving government assistance increased the probability that a respondent was matched, and being Black, Hispanic or another race other than White decreased the probability that a respondent was matched. Having a bank account or another financial asset or having been in the armed forces increase the probability a respondent was matched by 3 to 4 percentage points and being a home owner significantly increased the probability of being matched in the 2004 panel.

Reliability of Short-term Earnings Averages in the SIPP

One of the key advantages of using the matched SIPP-SSA data is the potential to construct measures of lifetime earnings instead of being limited to a short-run measure of earnings such as that available in the SIPP. One way to demonstrate the value of having access to the long-term earnings history is to consider the case where a true regression model should have the average lifetime earnings as an explanatory variable but where the researcher is forced to use a short-term measure of earnings as a proxy.

The intergenerational mobility literature for example is typically concerned with persistence in lifetime earnings (e.g. Mazumder, 2005). One might also be interested in the income gradient in childhood health or well-being (Mazumder and Davis, forthcoming). In these cases, the regression coefficient obtained from using a short-term measure of earnings would be biased towards zero relative to the true coefficient. The amount of bias

can be inferred by estimating a regression of the lifetime earnings measure on the short-term measure. The coefficient can be interpreted as the reliability ratio. A reliability ratio of 0.8 for example would suggest that the effect of lifetime earnings on an outcome would be biased down by about 20 percent if one used a short-term average as a proxy, and that the true coefficient could be obtained by multiplying the regression coefficient by a factor of 1.25.

In Table 5 we estimate the reliability ratio of using a two year average of earnings from the SIPP instead of a 25 year average of earnings from administrative records. Specifically, we use a sample of men born between 1948-1953 who had positive earnings in each of the first two calendar years of their SIPP panel. We selected this sample because they were between 25 and 57 during the 1978 to 2002 period, so their lifetime earnings should be well represented by the SER and DER data. Panel A of Table 5 runs a regression of the log of the 25-year average of SER earnings over 1978 to 2002 on the log of the 2-year average in the SIPP. Panel B shows the analogous estimates using the DER earnings and Panel C uses the maximum of SER or DER earnings. The last sample is worth considering since the DER sample is smaller than the SER sample and one might want to consider combining the samples to maximize cases.

Starting with Panel A we show that for the full sample, the reliability ratio with respect to SER earnings is 0.42 suggesting substantial bias from using short-term earnings. When we examine this by age category, we find that the reliability ratio peaks for 36-40 year olds and declines as the sample gets older. The reliability ratio for Whites is nearly identical to the ratio for the full sample, but that the reliability ratio for Blacks is only 0.35. Interestingly, the ratio is higher for respondents who had a high school degree or less than for those with some college.

Turning to Panel B, for the full sample, the reliability ratio is 0.58, almost 40% higher than the reliability ratio calculated using SER earnings. The DER results display a broadly similar lifecycle pattern to the SER results although the reliability ratio peaks with 41-45 year olds instead of 36-40 year olds. Indeed the reliability ratio is close to 70 percent for those in their early 40s. For the DER earnings, we find that the reliability ratios are substantially higher for Whites (0.60) than for Blacks (0.44) and increase sharply by education level. In Panel C where we use the maximum of SER and DER earnings we find reliability ratios that are roughly in between the results found for Panels A and B.

Conclusion

The matched sample generated by linking the Census Bureau's Survey of Income and Program Participation (SIPP) to the Social Security Administration's (SSA) administrative earnings records has opened up new avenues of research opportunities. We examine the degree to which this matched sample is nationally representative. We generally find that selection is not a large concern and that individuals with financial assets and who participate in government transfer programs are somewhat over-represented.

A key advantage of the matched data is the potential to construct measures of lifetime earnings. We show that the use of such data offers the potential for more accurately identifying the associations between lifetime earnings and many other outcomes of interest.

References

- Abowd, John, Martha Stinson, and Gary Benedetto (2006). "Final Report to the Social Security Administration on the SIPP/SSA/IRS Public Use File Project". November 5, 2006. <http://www.census.gov/sipp/FinalReporttoSocialSecurityAdministration.pdf>
- Czajka, John L., James Mabli and Scott Cody (2008). "Sample Loss and Survey Bias in Estimates of Social Security Beneficiaries: A Tale of Two Surveys". February 2008. <http://www.mathematica-mpr.com/publications/pdfs/samplelosssurvey.pdf>
- Dahl, Molly, Thomas Deleire and Jonathan Schwabish (2009). "Stepping Stone or Dead End: The Effect of the EITC on Earnings Growth". *National Tax Journal* 62:329-346.
- Haider, Steven and Gary Solon (2006). "Life-Cycle Variation in the Association between Current and Lifetime Earnings." *American Economic Review*, 96(4):1308-1320.
- Lubotsky, Darren (2007). "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings," *Journal of Political Economy*, 115 (5) 820-867.
- Mazumder, Bhashkar (2005). "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data." *The Review of Economics and Statistics*, MIT Press, vol 87(2), pp. 235-255.
- Mazumder, Bhashkar and Jonathan Davis (forthcoming). "Parental Income and Children's Well-being and Future Success: An Analysis of the SIPP matched to SSA Earnings Data" *Economic Inquiry*.
- Kopczuk Wojciech, Emmanuel Saez and Jae Song (2010) "Earnings Inequality and Mobility in the United States: Evidence From Social Security Data Since 1937" *Quarterly Journal of Economics* 125(1), 91-128.
- Olsen, Anya and Russell Hudson. "Social Security Administration's Master Earnings File: Background Information". *Social Security Bulletin*, Vol. 69, No. 3, 2009.
- Rutledge, Matthew (2011). "Long-Run Earnings Volatility and Health Insurance Coverage: Evidence from the SIPP Gold Standard File" Center for Economic Studies Working Paper 11-35.
- "Social Security Number Chronology". Last Updated 11/09/05. <http://www.socialsecurity.gov/history/ssn/ssnchron.html>
- W-2 Tax Form. <http://www.irs.gov/pub/irs-pdf/fw2.pdf>
- Vilhuber, Lars and Robert Pedace. "Longitudinal analysis of SSN response on SIPP 1990-1993 panels". July 7, 2003.

Table 1: Match Rates by SIPP Panel

Survey	SIPP		SER	SIPP		DER
	N	N	Match Rate	N	N	Match Rate
1984	62183	48915	0.79	56243	44638	0.74
1990	67114	58792	0.86	60320	54152	0.83
1991	42992	36597	0.83	38242	33423	0.80
1992	59696	51843	0.84	52332	46407	0.81
1993	60009	51228	0.83	51875	45468	0.80
1996	113345	92031	0.74	93883	77455	0.71
2001	81540	48688	0.54	63845	39158	0.53
2004	125705	95656	0.69	95467	74210	0.68

The SIPP sample sizes and match rates reflect the sample from the SIPP that could have been matched with the SER or DER. Individuals who entered the SIPP in waves in which no one was matched to the SIPP were dropped from the analysis. The DER SIPP sample only includes individuals born between 1918 and 1988. Our sample included Waves 1 through 7 of the 1984 SIPP, waves 1 through 8 of the 1990, 1991, 2001 and 2004 SIPP, waves 1 through 9 of the 1992 and 1993 SIPP and waves 1 through 12 of the 1996 SIPP.

Table 2: Match Rates by Attrition Status

Panel	SER			DER		
	In Full Panel	Attriter	Difference	In Full Panel	Attriter	Difference
1984	0.795	0.776	-0.019	0.744	0.723	-0.021
1990	0.881	0.787	-0.094	0.854	0.741	-0.113
1991	0.847	0.752	-0.095	0.820	0.715	-0.105
1992	0.870	0.801	-0.068	0.844	0.767	-0.077
1993	0.858	0.749	-0.109	0.832	0.710	-0.122
1996	0.768	0.683	-0.086	0.747	0.656	-0.091
2001	0.559	0.442	-0.118	0.554	0.430	-0.125
2004	0.772	0.467	-0.305	0.767	0.457	-0.310

Differences are all significant at the 1% level.

Table 3: Summary Statistics by Match Status

Pooled										
Measures	Full Sample		Matched to SER		% Diff	DER Full Sample		Matched to DER		% Diff
In MEF	612584	0.74	454664	1.00		512207	0.73	370277	1.00	
Age	612572	33.25	454664	34.71	4.4%	512207	35.88	370277	36.17	0.8%
Education	534817	10.97	405249	11.15	1.6%	496740	11.35	360233	11.63	2.4%
Family Income	582083	3698.77	440722	3733.94	1.0%	492184	3726.28	362115	3819.60	2.5%
Person Income	582083	1316.99	440722	1375.93	4.5%	492184	1501.49	362115	1590.60	5.9%
White Non-Hispanic	612584	0.73	454664	0.75	3.2%	512207	0.74	370277	0.76	3.2%
Male	612584	0.49	454664	0.49	-0.6%	512207	0.49	370277	0.50	1.2%
Ever Married	433852	0.75	335229	0.77	2.2%	412193	0.74	304256	0.75	1.8%
Employed	474224	0.63	364628	0.63	0.8%	426855	0.69	316639	0.72	3.7%
Home Owner	595919	0.68	444528	0.70	2.5%	498430	0.68	362489	0.70	2.2%
Bank Account	539514	0.50	404034	0.54	8.2%	454939	0.55	332771	0.59	6.8%
Other Financial Asset	539514	0.28	404034	0.30	5.7%	454939	0.32	332771	0.33	4.8%
Ever in Armed Forces	549367	0.10	414500	0.12	13.4%	472112	0.12	346204	0.13	13.4%
Any Government Aid	521561	0.32	398731	0.31	-1.1%	437598	0.27	326403	0.25	-8.0%

1984										
Measures	Full Sample		Matched to SER		% Diff	DER Full Sample		Matched to DER		% Diff
In MEF	62183	0.79	48368	1.00		56243	0.74	40824	1.00	
Age	62171	33.09	48368	38.44	16.1%	56243	28.68	40824	32.84	14.5%
Education	62183	9.04	48368	10.70	18.3%	56243	8.93	40824	10.84	21.3%
Family Income	62183	2205.17	48368	2262.68	2.6%	56243	2278.29	40824	2385.16	4.7%
Person Income	62183	741.67	48368	893.07	20.4%	56243	733.74	40824	928.23	26.5%
White Non-Hispanic	62183	0.79	48368	0.81	1.8%	56243	0.78	40824	0.80	2.0%
Male	62183	0.49	48368	0.49	-0.9%	56243	0.50	40824	0.51	1.8%
Ever Married	44122	0.76	40939	0.76	0.8%	38234	0.73	33538	0.73	-0.2%
Employed	44835	0.65	41680	0.65	0.0%	39087	0.72	34436	0.76	4.6%
Home Owner	57995	0.69	46233	0.69	0.4%	52146	0.68	38807	0.68	0.1%
Bank Account	45980	0.62	42510	0.63	1.7%	40098	0.60	35104	0.61	2.5%
Other Financial Asset	45980	0.22	42510	0.23	2.7%	40098	0.21	35104	0.22	2.4%
Ever in Armed Forces	62183	0.13	48368	0.15	20.7%	56243	0.12	40824	0.16	25.7%
Any Government Aid	46787	0.33	42477	0.31	-4.9%	40909	0.24	35063	0.20	-16.2%

1990-1993										
Measures	Full Sample		Matched to SER		% Diff	DER Full Sample		Matched to DER		% Diff
In MEF	229811	0.84	193872	1.00		202769	0.81	165116	1.00	
Age	229811	33.38	193872	34.44	3.2%	202769	33.43	165116	33.59	0.5%
Education	229811	9.44	193872	9.65	2.2%	202769	10.14	165116	10.35	2.1%
Family Income	229811	3045.21	193872	3108.12	2.1%	202769	3163.39	165116	3263.62	3.2%
Person Income	229811	1076.57	193872	1131.66	5.1%	202769	1160.87	165116	1242.05	7.0%
White Non-Hispanic	229811	0.75	193872	0.77	2.1%	202769	0.75	165116	0.77	2.3%
Male	229811	0.49	193872	0.49	-0.1%	202769	0.50	165116	0.51	1.6%
Ever Married	164824	0.74	142501	0.76	2.1%	153529	0.73	126847	0.74	1.6%
Employed	167482	0.65	145598	0.65	-0.7%	156509	0.70	130311	0.72	3.0%
Home Owner	217334	0.69	185871	0.69	0.7%	193089	0.68	159345	0.69	0.6%
Bank Account	172944	0.60	149100	0.62	2.7%	161646	0.60	133330	0.62	3.4%
Other Financial Asset	172944	0.22	149100	0.23	4.5%	161646	0.21	133330	0.22	4.8%
Ever in Armed Forces	196133	0.17	167188	0.18	5.8%	181735	0.17	148748	0.18	6.7%
Any Government Aid	176235	0.35	153848	0.35	-1.3%	159017	0.28	133419	0.26	-7.3%
1990	229811	0.24	193872	0.25		202769	0.25	165116	0.25	
1991	229811	0.25	193872	0.24		202769	0.25	165116	0.24	
1992	229811	0.25	193872	0.25		202769	0.25	165116	0.25	
1993	229811	0.26	193872	0.25		202769	0.25	165116	0.25	

Table 3 continued: Summary Statistics by Match Status

1996										
Measures	Full Sample		Matched to SER		% Diff	DER Full Sample		Matched to DER		% Diff
In MEF	113345	0.74	82702	1.00		93883	0.71	66098	1.00	
Age	113345	33.03	82702	33.62	1.8%	93883	36.88	66098	36.88	0.0%
Education	84616	12.72	62171	12.80	0.6%	81331	12.79	57491	12.97	1.4%
Family Income	100489	3834.19	75373	3902.64	1.8%	85227	3932.17	61771	4055.04	3.1%
Person Income	100489	1359.25	75373	1407.10	3.5%	85227	1554.07	61771	1640.76	5.6%
White Non-Hispanic	113345	0.72	82702	0.74	2.8%	93883	0.73	66098	0.75	3.2%
Male	113345	0.49	82702	0.49	-0.3%	93883	0.50	66098	0.50	1.3%
Ever Married	80100	0.74	58982	0.76	2.2%	76815	0.73	54359	0.75	2.0%
Employed	82066	0.68	60887	0.68	0.1%	78890	0.71	56350	0.73	3.1%
Home Owner	113345	0.65	82702	0.67	2.2%	93883	0.67	66098	0.68	2.3%
Bank Account	113345	0.41	82702	0.44	6.5%	93883	0.48	66098	0.51	7.0%
Other Financial Asset	113345	0.29	82702	0.32	8.0%	93883	0.35	66098	0.38	9.3%
Ever in Armed Forces	101018	0.07	75696	0.07	4.4%	85756	0.08	62090	0.08	7.0%
Any Government Aid	103695	0.29	77268	0.29	0.2%	86983	0.25	62617	0.24	-6.1%

2001										
Measures	Full Sample		Matched to SER		% Diff	DER Full Sample		Matched to DER		% Diff
In MEF	81540	0.54	42989	1.00		63845	0.53	33029	1.00	
Age	81540	34.85	42989	36.04	3.4%	63845	41.12	33029	41.44	0.8%
Education	62121	12.94	33204	13.05	0.8%	60930	12.97	31592	13.16	1.5%
Family Income	74263	4606.32	41654	4662.56	1.2%	58898	4671.80	32257	4788.87	2.5%
Person Income	74263	1666.98	41654	1749.19	4.9%	58898	2045.59	32257	2172.84	6.2%
White Non-Hispanic	81540	0.70	42989	0.72	2.8%	63845	0.71	33029	0.74	3.2%
Male	81540	0.49	42989	0.49	-1.0%	63845	0.49	33029	0.49	0.4%
Ever Married	58868	0.74	31599	0.76	2.6%	57677	0.74	30023	0.75	2.4%
Employed	60337	0.68	32909	0.69	1.4%	59177	0.70	31325	0.72	3.9%
Home Owner	81540	0.69	42989	0.70	2.1%	63845	0.70	33029	0.72	2.5%
Bank Account	81540	0.43	42989	0.48	11.4%	63845	0.53	33029	0.59	10.6%
Other Financial Asset	81540	0.34	42989	0.38	12.3%	63845	0.42	33029	0.47	12.1%
Ever in Armed Forces	74526	0.06	41706	0.06	6.4%	59161	0.07	32308	0.07	7.9%
Any Government Aid	76334	0.29	42101	0.29	-0.5%	60045	0.27	32451	0.25	-5.4%

2004										
Measures	Full Sample		Matched to SER		% Diff	DER Full Sample		Matched to DER		% Diff
In MEF	125705	0.69	86733	1.00		95467	0.68	65210	1.00	
Age	125705	32.25	86733	33.63	4.3%	95467	40.81	65210	41.42	1.5%
Education	96086	13.08	67634	13.32	1.9%	95467	13.10	65210	13.44	2.5%
Family Income	115337	5103.90	81455	5466.11	7.1%	89047	5100.17	62147	5501.91	7.9%
Person Income	115337	1844.02	81455	2024.34	9.8%	89047	2351.85	62147	2599.68	10.5%
White Non-Hispanic	125705	0.67	86733	0.70	4.5%	95467	0.70	65210	0.73	4.5%
Male	125705	0.49	86733	0.48	-1.9%	95467	0.49	65210	0.48	-0.6%
Ever Married	85938	0.77	61208	0.80	3.5%	85938	0.77	59489	0.80	3.2%
Employed	119504	0.52	83554	0.54	3.3%	93192	0.66	64217	0.70	4.7%
Home Owner	125705	0.68	86733	0.74	7.8%	95467	0.70	65210	0.75	7.5%
Bank Account	125705	0.43	86733	0.48	11.5%	95467	0.56	65210	0.62	10.5%
Other Financial Asset	125705	0.35	86733	0.39	14.1%	95467	0.45	65210	0.51	13.8%
Ever in Armed Forces	115507	0.05	81542	0.05	5.7%	89217	0.06	62234	0.07	7.0%
Any Government Aid	118510	0.31	83037	0.29	-5.6%	90644	0.29	62853	0.26	-8.5%

Table 4a: Regression Analysis of SIPP Matched to the Summary Earnings Record

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled	1984	1990-1993	1996	2001	2004
Not a HS Graduate	0.0272 [0.0018]***	-0.0069 [0.0032]**	0.0318 [0.0025]***	0.0367 [0.0045]***	0.0065 [0.0060]	-0.0332 [0.0051]***
Some College	0.0296 [0.0017]***	0.0184 [0.0033]***	0.0382 [0.0024]***	0.0377 [0.0040]***	0.0028 [0.0053]	0.0222 [0.0042]***
BA	0.0195 [0.0022]***	0.0007 [0.0046]	0.0273 [0.0031]***	0.0239 [0.0052]***	-0.0089 [0.0068]	0.0176 [0.0054]***
Graduate Education	0.0281 [0.0025]***	0.0048 [0.0048]	0.0406 [0.0033]***	0.0401 [0.0069]***	-0.0022 [0.0089]	0.0082 [0.0067]
Black	-0.0407 [0.0019]***	-0.0292 [0.0045]***	-0.0488 [0.0029]***	-0.0558 [0.0044]***	-0.0278 [0.0056]***	-0.0288 [0.0049]***
Other Race	-0.0328 [0.0029]***	-0.008 [0.0090]	-0.0412 [0.0048]***	-0.0423 [0.0068]***	-0.0405 [0.0079]***	-0.0129 [0.0055]**
Hispanic	-0.0476 [0.0021]***	-0.0167 [0.0064]***	-0.0382 [0.0032]***	-0.06 [0.0050]***	-0.0564 [0.0059]***	-0.0658 [0.0052]***
Male	-0.0013 [0.0012]	0.0035 [0.0031]	0.0049 [0.0018]***	0.001 [0.0029]	-0.0052 [0.0037]	-0.0226 [0.0030]***
Widowed or Divorced	0.0141 [0.0019]***	-0.0088 [0.0038]**	0.0232 [0.0027]***	0.012 [0.0046]***	0.0046 [0.0062]	-0.0013 [0.0049]
Separated	0.0233 [0.0041]***	-0.0138 [0.0086]	0.035 [0.0059]***	0.036 [0.0096]***	0.0196 [0.0130]	-0.0145 [0.0116]
Never Married	-0.0125 [0.0020]***	-0.0086 [0.0039]**	-0.0182 [0.0029]***	-0.0209 [0.0049]***	-0.0017 [0.0063]	-0.0138 [0.0052]***
Had Job	0.019 [0.0020]***	0.0543 [0.0040]***	0.0095 [0.0029]***	0.0048 [0.0050]	0.0359 [0.0068]***	0.0122 [0.0053]**
Armed Forces	0.0265 [0.0018]***	0.0094 [0.0037]**	0.0216 [0.0023]***	0.03 [0.0058]***	0.0233 [0.0087]***	-0.0023 [0.0070]
1st Decile of Family Income Distribution	-0.0534 [0.0032]***	-0.0247 [0.0073]***	-0.038 [0.0050]***	-0.0487 [0.0071]***	-0.0504 [0.0091]***	-0.0914 [0.0076]***
2nd Decile of Family Income Distribution	-0.0238 [0.0028]***	0.0076 [0.0061]	-0.0192 [0.0042]***	-0.0175 [0.0066]***	-0.033 [0.0085]***	-0.0278 [0.0072]***
3rd Decile of Family Income Distribution	-0.0077 [0.0026]***	0.0122 [0.0056]**	-0.0031 [0.0037]	0.002 [0.0062]	-0.032 [0.0082]***	-0.0019 [0.0069]
4th Decile of Family Income Distribution	-0.0007 [0.0025]	0.0091 [0.0055]*	0.0068 [0.0035]*	-0.0078 [0.0062]	-0.0219 [0.0081]***	0.0041 [0.0067]
6th Decile of Family Income Distribution	0.0024 [0.0024]	0.002 [0.0052]	0.006 [0.0034]*	-0.0016 [0.0060]	-0.0113 [0.0079]	0.0095 [0.0064]
7th Decile of Family Income Distribution	0.0024 [0.0025]	-0.009 [0.0056]	0.0066 [0.0034]*	0.0071 [0.0060]	-0.0373 [0.0080]***	0.0291 [0.0063]***
8th Decile of Family Income Distribution	0.0069 [0.0025]***	0.0015 [0.0054]	0.0084 [0.0034]**	0.0126 [0.0060]**	-0.0338 [0.0080]***	0.0401 [0.0063]***
9th Decile of Family Income Distribution	0.0057 [0.0025]**	-0.0001 [0.0056]	0.0064 [0.0035]*	0.0196 [0.0061]***	-0.0164 [0.0082]**	0.0157 [0.0066]**
10th Decile of Family Income Distribution	-0.0029 [0.0026]	0.0005 [0.0058]	-0.0014 [0.0038]	0.0027 [0.0064]	-0.0454 [0.0086]***	0.0242 [0.0067]***
No Personal Income	-0.0256 [0.0035]***	-0.003 [0.0075]	-0.031 [0.0054]***	-0.0471 [0.0074]***	-0.0181 [0.0093]*	-0.0292 [0.0084]***
4th Decile of Personal Income Distribution	0.0081 [0.0030]***	0.0274 [0.0059]***	0.0029 [0.0046]	0.0009 [0.0067]	0.0218 [0.0089]**	0.0098 [0.0073]
5th Decile of Personal Income Distribution	0.008 [0.0026]***	0.0046 [0.0048]	0.0081 [0.0037]**	0.0019 [0.0064]	0.0123 [0.0087]	0.025 [0.0069]***
6th Decile of Personal Income Distribution	0.00 [0.0025]	-0.0057 [0.0046]	0.0031 [0.0035]	-0.0008 [0.0063]	0.0053 [0.0083]	-0.0032 [0.0068]
8th Decile of Personal Income Distribution	0.0012 [0.0024]	0.0106 [0.0043]**	-0.0017 [0.0033]	-0.0004 [0.0061]	0.0168 [0.0082]**	0.0057 [0.0066]
9th Decile of Personal Income Distribution	0.0014 [0.0024]	0.0088 [0.0046]*	0.0037 [0.0034]	-0.0024 [0.0062]	-0.0016 [0.0085]	0.0148 [0.0066]**
10th Decile of Personal Income Distribution	-0.0055 [0.0027]**	0.0022 [0.0052]	-0.0018 [0.0038]	-0.0161 [0.0069]**	0.0131 [0.0093]	0.0062 [0.0071]

Table 4a continued: Regression Analysis of SIPP Matched to the Summary Earnings Record

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled	1984	1990-1993	1996	2001	2004
Home Owner	0.0268 [0.0014]***	0.0118 [0.0032]***	0.0145 [0.0021]***	0.0093 [0.0033]***	0.0002 [0.0044]	0.1189 [0.0038]***
Has a Bank Account	0.0351 [0.0015]***	0.0262 [0.0030]***	0.0241 [0.0022]***	0.0424 [0.0038]***	0.0594 [0.0051]***	0.0494 [0.0041]***
Has Other Financial Asset	0.0276 [0.0015]***	0.0124 [0.0026]***	0.0096 [0.0022]***	0.0358 [0.0038]***	0.0327 [0.0052]***	0.0373 [0.0041]***
Received WIC	0.0155 [0.0053]***	0.0999 [0.0151]***	0.0227 [0.0076]***	0.021 [0.0114]*	0.0163 [0.0150]	-0.01 [0.0131]
Received AFDC/TANF Benefits	0.0033 [0.0044]	0.0646 [0.0113]***	-0.0068 [0.0057]	-0.0336 [0.0129]***	-0.0515 [0.0241]**	-0.0577 [0.0200]***
Received Food Stamps	0.0698 [0.0030]***	0.1409 [0.0068]***	0.0423 [0.0042]***	0.0692 [0.0086]***	0.0837 [0.0130]***	0.0706 [0.0102]***
Received Medicare	0.007 [0.0048]	-0.0111 [0.0106]	0.0369 [0.0070]***	-0.0233 [0.0134]*	0.0038 [0.0178]	-0.0088 [0.0123]
Received Medicaid	0.0345 [0.0025]***	0.0503 [0.0084]***	0.0295 [0.0044]***	0.0514 [0.0052]***	0.0546 [0.0064]***	0.013 [0.0052]**
Received Social Security	0.0297 [0.0036]***	0.0781 [0.0073]***	0.0112 [0.0048]**	0.0345 [0.0110]***	0.0049 [0.0155]	0.0226 [0.0115]**
Received Rent Assistance	0.0533 [0.0029]***	0.0299 [0.0079]***	0.0396 [0.0042]***	0.0295 [0.0067]***	0.0875 [0.0093]***	0.1024 [0.0081]***
Received Energy Assistance	0.0254 [0.0033]***	0.0155 [0.0105]	0.0254 [0.0043]***	0.0412 [0.0080]***	0.0773 [0.0128]***	0.0083 [0.0091]
0-5	-0.0104 [0.0072]	-0.4728 [0.0267]***	0.0051 [0.0148]	0.1935 [0.0145]***	0.0718 [0.0167]***	0.0248 [0.0161]
10-15	0.016 [0.0074]**	-0.3637 [0.0279]***	0.0276 [0.0148]*	0.1744 [0.0153]***	0.0483 [0.0179]***	0.0153 [0.0162]
16-20	0.0276 [0.0064]***	-0.2686 [0.0270]***	0.0275 [0.0140]**	0.1485 [0.0143]***	0.0408 [0.0164]**	0.0202 [0.0144]
20-25	0.0014 [0.0034]	-0.0141 [0.0071]**	0.0011 [0.0052]	0.0353 [0.0082]***	0.0267 [0.0108]**	-0.0354 [0.0088]***
26-30	0.0104 [0.0029]***	0.0112 [0.0057]*	0.0125 [0.0042]***	0.006 [0.0073]	0.0151 [0.0098]	-0.0391 [0.0081]***
31-35	0.0048 [0.0027]*	0.0056 [0.0050]	0.0058 [0.0037]	-0.0025 [0.0068]	0.0122 [0.0095]	-0.0248 [0.0077]***
36-40	-0.0009 [0.0026]	-0.0025 [0.0051]	0.0023 [0.0036]	-0.0055 [0.0064]	-0.0019 [0.0091]	-0.0023 [0.0071]
41-45	0.0011 [0.0027]	-0.0096 [0.0056]*	-0.0003 [0.0037]	-0.0016 [0.0066]	0.0112 [0.0091]	0.0122 [0.0071]*
46-50	-0.0075 [0.0028]***	-0.0015 [0.0057]	-0.0091 [0.0040]**	-0.031 [0.0069]***	-0.0024 [0.0094]	0.0113 [0.0072]
51-55	-0.0031 [0.0030]	-0.0004 [0.0060]	-0.0113 [0.0044]***	-0.0123 [0.0075]*	0.0014 [0.0099]	0.0189 [0.0076]**
56-60	-0.0023 [0.0031]	0.0131 [0.0056]**	-0.0157 [0.0045]***	-0.0206 [0.0081]**	0.0176 [0.0107]*	0.02 [0.0082]**
61-65	-0.0184 [0.0035]***	0.0024 [0.0062]	-0.0219 [0.0050]***	-0.0269 [0.0093]***	-0.003 [0.0126]	0.019 [0.0108]*
66-70	-0.0256 [0.0049]***	-0.0045 [0.0102]	-0.0387 [0.0072]***	-0.01 [0.0117]	-0.0211 [0.0164]	0.0189 [0.0126]
71-75	-0.0262 [0.0051]***	0.0013 [0.0104]	-0.0379 [0.0075]***	-0.0184 [0.0124]	-0.0096 [0.0169]	-0.0048 [0.0131]
76-80	-0.0293 [0.0054]***	0.0035 [0.0110]	-0.0401 [0.0080]***	-0.0184 [0.0132]	0.0058 [0.0178]	-0.0506 [0.0145]***
81+	-0.0476 [0.0056]***	-0.0018 [0.0116]	-0.0471 [0.0080]***	-0.0229 [0.0135]*	-0.0123 [0.0178]	-0.1422 [0.0157]***

Table 4a continued: Regression Analysis of SIPP Matched to the Summary Earnings Record

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled	1984	1990-1993	1996	2001	2004
Entered in Wave 2	-0.1681 [0.0060]***	-0.2059 [0.0107]***	-0.1097 [0.0087]***	-0.1139 [0.0110]***	-0.3342 [0.0124]***	-0.0998 [0.0127]***
Entered in Wave 3	-0.2335 [0.0060]***	-0.2322 [0.0099]***	-0.1793 [0.0089]***	-0.1609 [0.0125]***	-0.3558 [0.0123]***	-0.0824 [0.0137]***
Entered in Wave 4	-0.2635 [0.0061]***	-0.3244 [0.0102]***	-0.1949 [0.0096]***	-0.2141 [0.0136]***	-0.4182 [0.0127]***	-0.0966 [0.0143]***
Entered in Wave 5	-0.2696 [0.0064]***	-0.2612 [0.0111]***	-0.2413 [0.0102]***	-0.2247 [0.0137]***	-0.4068 [0.0131]***	-0.0656 [0.0140]***
Entered in Wave 6	-0.2839 [0.0064]***	-0.2343 [0.0147]***	-0.2647 [0.0096]***	-0.204 [0.0150]***	-0.4431 [0.0119]***	-0.0888 [0.0137]***
Entered in Wave 7	-0.3272 [0.0066]***	-0.2586 [0.0143]***	-0.2855 [0.0105]***	-0.233 [0.0152]***	-0.4651 [0.0121]***	-0.2864 [0.0157]***
Entered in Wave 8	-0.3344 [0.0072]***		-0.2845 [0.0110]***	-0.2706 [0.0151]***	-0.5305 [0.0109]***	-0.2427 [0.0166]***
Entered in Wave 9	-0.2534 [0.0110]***		-0.2635 [0.0141]***	-0.2423 [0.0165]***		
Entered in Wave 10	-0.2276 [0.0167]***			-0.2642 [0.0173]***		
Entered in Wave 11	-0.2381 [0.0164]***			-0.2723 [0.0169]***		
Entered in Wave 12	-0.263 [0.0174]***			-0.2958 [0.0180]***		
1990	0.1038 [0.0021]***					
1991	0.1661 [0.0022]***		-0.029 [0.0023]***			
1992	0.1374 [0.0024]***		-0.0177 [0.0021]***			
1993	0.1055 [0.0022]***		-0.0274 [0.0021]***			
1996	0.0945 [0.0022]***					
2001	-0.2103 [0.0022]***					
2004	-0.0527 [0.0021]***					
Observations	612584	62183	229811	113345	81540	125705
R-squared	0.15	0.47	0.09	0.09	0.12	0.11

Notes: The regressions also included indicators for all missing values, receiving general assistance, foster care, other welfare, railroad retirement, veteran's assistance and state of residence. Robust standard errors in brackets.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4b: Regression Analysis of SIPP Matched to the Detailed Earnings Record

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled	1984	1990-1993	1996	2001	2004
Not a HS Graduate	0.0036 [0.0021]*	-0.0277 [0.0046]***	0.0129 [0.0030]***	0.0199 [0.0049]***	0.0006 [0.0062]	-0.0511 [0.0058]***
Some College	0.0345 [0.0018]***	0.0212 [0.0041]***	0.0439 [0.0027]***	0.0423 [0.0042]***	0.0091 [0.0054]*	0.0303 [0.0043]***
BA	0.0252 [0.0023]***	0.0047 [0.0055]	0.034 [0.0034]***	0.0304 [0.0054]***	-0.0011 [0.0069]	0.025 [0.0055]***
Graduate Education	0.041 [0.0026]***	0.0125 [0.0058]**	0.0543 [0.0036]***	0.052 [0.0072]***	0.0085 [0.0090]	0.207 [0.0067]***
Black	-0.034 [0.0022]***	-0.0275 [0.0052]***	-0.0452 [0.0033]***	-0.0475 [0.0050]***	-0.0133 [0.0065]**	-0.0136 [0.0057]**
Other Race	-0.0467 [0.0033]***	-0.0099 [0.0102]	-0.0564 [0.0053]***	-0.063 [0.0076]***	-0.0529 [0.0091]***	-0.0221 [0.0065]***
Hispanic	-0.0526 [0.0024]***	-0.0231 [0.0073]***	-0.0348 [0.0036]***	-0.0673 [0.0057]***	-0.0711 [0.0069]***	-0.0788 [0.0064]***
Male	0.0116 [0.0014]***	0.0113 [0.0035]***	0.0154 [0.0020]***	0.018 [0.0033]***	0.0057 [0.0043]	-0.0144 [0.0035]***
Widowed or Divorced	0.0199 [0.0022]***	-0.0017 [0.0059]	0.0359 [0.0032]***	0.0182 [0.0050]***	0.006 [0.0064]	-0.0123 [0.0051]**
Separated	0.0253 [0.0044]***	0.0027 [0.0104]	0.037 [0.0065]***	0.0412 [0.0100]***	0.0218 [0.0131]*	-0.0169 [0.0117]
Never Married	-0.0219 [0.0021]***	-0.0148 [0.0046]***	-0.0257 [0.0031]***	-0.0272 [0.0051]***	-0.0085 [0.0064]	-0.022 [0.0053]***
Had Job	0.0713 [0.0023]***	0.1553 [0.0061]***	0.0623 [0.0034]***	0.0548 [0.0055]***	0.0612 [0.0073]***	0.0346 [0.0055]***
Armed Forces	0.0392 [0.0020]***	0.0476 [0.0046]***	0.0457 [0.0026]***	0.0258 [0.0059]***	0.02 [0.0088]**	-0.0079 [0.0071]
1st Decile of Family Income Distribution	-0.0335 [0.0037]***	-0.0205 [0.0086]**	-0.0162 [0.0056]***	-0.0428 [0.0083]***	-0.0454 [0.0107]***	-0.0671 [0.0091]***
2nd Decile of Family Income Distribution	-0.0219 [0.0033]***	-0.0026 [0.0080]	-0.0208 [0.0050]***	-0.0169 [0.0078]**	-0.0245 [0.0098]**	-0.0256 [0.0084]***
3rd Decile of Family Income Distribution	-0.0097 [0.0031]***	0.0077 [0.0073]	-0.0073 [0.0044]*	-0.0034 [0.0073]	-0.0304 [0.0095]***	-0.0089 [0.0079]
4th Decile of Family Income Distribution	-0.0007 [0.0029]	0.0079 [0.0068]	0.0081 [0.0041]**	-0.0166 [0.0071]**	-0.0156 [0.0093]*	-0.0089 [0.0077]
6th Decile of Family Income Distribution	0.0011 [0.0027]	-0.002 [0.0063]	0.0067 [0.0038]*	-0.0036 [0.0068]	-0.0122 [0.0091]	-0.0051 [0.0073]
7th Decile of Family Income Distribution	0.0007 [0.0028]	-0.011 [0.0066]*	0.0059 [0.0038]	0.0083 [0.0067]	-0.0342 [0.0091]***	0.0131 [0.0073]*
8th Decile of Family Income Distribution	0.0034 [0.0028]	-0.0023 [0.0064]	0.007 [0.0038]*	0.0066 [0.0068]	-0.0288 [0.0092]***	0.207 [0.0073]***
9th Decile of Family Income Distribution	-0.0008 [0.0028]	-0.007 [0.0066]	0.0027 [0.0039]	0.0158 [0.0068]**	-0.0204 [0.0093]**	-0.005 [0.0076]
10th Decile of Family Income Distribution	-0.011 [0.0030]***	-0.0061 [0.0068]	-0.006 [0.0042]	-0.0029 [0.0072]	-0.0478 [0.0098]***	-0.003 [0.0078]
1st "Decile" of Personal Income Distribution	-0.0253 [0.0039]***	0.0025 [0.0101]	-0.0474 [0.0060]***	-0.022 [0.0089]**	-0.0103 [0.0115]	-0.0606 [0.0096]***
4th Decile of Personal Income Distribution	-0.0149 [0.0033]***	-0.0003 [0.0083]	-0.0285 [0.0052]***	-0.0215 [0.0074]***	0.0083 [0.0096]	-0.0175 [0.0077]**
5th Decile of Personal Income Distribution	-0.0339 [0.0030]***	-0.0368 [0.0068]***	-0.046 [0.0044]***	-0.042 [0.0073]***	-0.0149 [0.0092]	-0.0177 [0.0073]**
6th Decile of Personal Income Distribution	-0.02 [0.0029]***	-0.0497 [0.0069]***	-0.0199 [0.0042]***	-0.013 [0.0069]*	0.0044 [0.0087]	-0.0135 [0.0070]*
8th Decile of Personal Income Distribution	0.0001 [0.0026]	0.0153 [0.0056]***	-0.0052 [0.0037]	-0.0078 [0.0065]	0.0118 [0.0085]	0.0037 [0.0067]
9th Decile of Personal Income Distribution	-0.002 [0.0026]	0.0083 [0.0058]	-0.003 [0.0037]	-0.0165 [0.0066]**	-0.0072 [0.0088]	0.0096 [0.0068]
10th Decile of Personal Income Distribution	-0.0112 [0.0029]***	0.0012 [0.0064]	-0.0121 [0.0041]***	-0.0329 [0.0072]***	0.0036 [0.0097]	0.0032 [0.0074]

Table 4b continued: Regression Analysis of SIPP Matched to the Detailed Earnings Record

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled	1984	1990-1993	1996	2001	2004
Home Owner	0.0276 [0.0016]***	0.0204 [0.0039]***	0.0193 [0.0023]***	0.0053 [0.0037]	-0.0007 [0.0050]	0.1119 [0.0044]***
Has a Bank Account	0.0394 [0.0017]***	0.0302 [0.0038]***	0.0244 [0.0024]***	0.0472 [0.0041]***	0.0622 [0.0052]***	0.0522 [0.0042]***
Has Other Financial Asset	0.0283 [0.0017]***	0.0155 [0.0039]***	0.012 [0.0025]***	0.0441 [0.0040]***	0.0364 [0.0053]***	0.0415 [0.0042]***
Received WIC	0.0378 [0.0061]***	0.1025 [0.0155]***	0.0608 [0.0111]***	0.0398 [0.0118]***	0.0304 [0.0152]**	0.0018 [0.0131]
Received AFDC/TANF Benefits	0.066 [0.0053]***	0.1451 [0.0160]***	0.0733 [0.0075]***	0.0392 [0.0139]***	-0.0102 [0.0245]	-0.0471 [0.0203]**
Received Food Stamps	0.0686 [0.0037]***	0.1457 [0.0087]***	0.0517 [0.0052]***	0.0644 [0.0098]***	0.0788 [0.0137]***	0.0594 [0.0107]***
Received Medicare	0.0016 [0.0059]	-0.1712 [0.0212]***	-0.0005 [0.0095]	-0.0073 [0.0158]	0.0124 [0.0180]	0.0087 [0.0127]
Received Medicaid	-0.019 [0.0035]***	-0.0239 [0.0138]*	-0.0521 [0.0062]***	0.0083 [0.0067]	0.0127 [0.0086]	-0.0051 [0.0073]
Received Social Security	0.0467 [0.0044]***	0.0875 [0.0109]***	0.029 [0.0059]***	0.0504 [0.0129]***	0.0277 [0.0156]*	0.0372 [0.0118]***
Received Rent Assistance	0.0463 [0.0037]***	0.031 [0.0102]***	0.0339 [0.0052]***	0.0225 [0.0082]***	0.0828 [0.0115]***	0.0993 [0.0102]***
Received Energy Assistance	0.0214 [0.0042]***	0.0047 [0.0141]	0.0184 [0.0055]***	0.0329 [0.0100]***	0.0753 [0.0155]***	0.026 [0.0114]**
0-5	-0.1438 [0.0088]***	-0.418 [0.0277]***	-0.0297 [0.0154]*			
10-15	-0.0086 [0.0083]	-0.3089 [0.0288]***	0.0614 [0.0148]***	0.0364 [0.0179]**		
16-20	0.0282 [0.0069]***	-0.2138 [0.0279]***	0.0845 [0.0140]***	0.081 [0.0156]***	0.0201 [0.0189]	0.0432 [0.0147]***
20-25	0.0438 [0.0036]***	0.0531 [0.0081]***	0.0507 [0.0054]***	0.0633 [0.0085]***	0.0491 [0.0110]***	-0.0221 [0.0089]**
26-30	0.0247 [0.0030]***	0.0358 [0.0063]***	0.0316 [0.0044]***	0.0148 [0.0075]**	0.0219 [0.0099]**	-0.0377 [0.0081]***
31-35	0.0085 [0.0027]***	0.0142 [0.0055]***	0.0113 [0.0038]***	-0.0015 [0.0069]	0.0115 [0.0095]	-0.0273 [0.0077]***
36-40	-0.0063 [0.0027]**	-0.0126 [0.0058]**	-0.0069 [0.0037]*	-0.0065 [0.0065]	-0.0004 [0.0091]	-0.002 [0.0072]
41-45	-0.0107 [0.0027]***	-0.0354 [0.0066]***	-0.0179 [0.0039]***	-0.0068 [0.0067]	0.0107 [0.0091]	0.0104 [0.0071]
46-50	-0.0243 [0.0029]***	-0.0539 [0.0072]***	-0.0303 [0.0042]***	-0.0362 [0.0070]***	-0.005 [0.0094]	0.0095 [0.0073]
51-55	-0.0317 [0.0032]***	-0.0739 [0.0079]***	-0.0489 [0.0047]***	-0.025 [0.0077]***	-0.002 [0.0099]	0.0129 [0.0077]*
56-60	-0.0457 [0.0035]***	-0.0881 [0.0083]***	-0.0715 [0.0051]***	-0.0412 [0.0084]***	0.0078 [0.0107]	0.0084 [0.0084]
61-65	-0.0836 [0.0041]***	-0.108 [0.0097]***	-0.1059 [0.0059]***	-0.0734 [0.0098]***	-0.0299 [0.0127]**	-0.0139 [0.0111]
66-70	-0.109 [0.0060]***	-0.0163 [0.0247]	-0.1219 [0.0098]***	-0.0813 [0.0130]***	-0.0662 [0.0164]***	-0.0415 [0.0130]***
71-75	-0.1269 [0.0066]***		-0.1466 [0.0109]***	-0.1227 [0.0139]***	-0.0779 [0.0169]***	-0.0934 [0.0136]***
76-80	-0.1185 [0.0084]***			-0.1391 [0.0164]***	-0.0864 [0.0180]***	-0.1518 [0.0150]***
81+	-0.2065 [0.0111]***				-0.0785 [0.0224]***	-0.2724 [0.0155]***

Table 4b continued: Regression Analysis of SIPP Matched to the Detailed Earnings Record

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled	1984	1990-1993	1996	2001	2004
Entered in Wave 2	-0.2077 [0.0069]***	-0.209 [0.0112]***	-0.1812 [0.0109]***	-0.1287 [0.0136]***	-0.3437 [0.0143]***	-0.1246 [0.0166]***
Entered in Wave 3	-0.2647 [0.0069]***	-0.2377 [0.0103]***	-0.2461 [0.0108]***	-0.1978 [0.0152]***	-0.3686 [0.0139]***	-0.123 [0.0174]***
Entered in Wave 4	-0.2907 [0.0069]***	-0.3195 [0.0103]***	-0.2471 [0.0115]***	-0.2388 [0.0168]***	-0.427 [0.0141]***	-0.134 [0.0181]***
Entered in Wave 5	-0.2821 [0.0073]***	-0.2542 [0.0112]***	-0.2725 [0.0120]***	-0.2261 [0.0164]***	-0.3992 [0.0151]***	-0.0851 [0.0178]***
Entered in Wave 6	-0.2944 [0.0074]***	-0.2304 [0.0151]***	-0.3003 [0.0113]***	-0.2063 [0.0183]***	-0.4411 [0.0134]***	-0.0923 [0.0171]***
Entered in Wave 7	-0.3344 [0.0076]***	-0.2598 [0.0144]***	-0.3294 [0.0124]***	-0.2271 [0.0185]***	-0.4758 [0.0125]***	-0.2681 [0.0196]***
Entered in Wave 8	-0.3476 [0.0087]***		-0.3273 [0.0130]***	-0.2815 [0.0181]***	-0.5194 [0.0126]***	-0.1425 [0.0211]***
Entered in Wave 9	-0.3128 [0.0137]***		-0.3192 [0.0170]***	-0.2693 [0.0206]***		
Entered in Wave 10	-0.2645 [0.0209]***			-0.2628 [0.0214]***		
Entered in Wave 11	-0.305 [0.0205]***			-0.3047 [0.0209]***		
Entered in Wave 12	-0.3336 [0.0217]***			-0.3243 [0.0224]***		
1990	0.0801 [0.0024]***					
1991	0.1459 [0.0025]***		-0.0231 [0.0026]***			
1992	0.118 [0.0027]***		-0.0117 [0.0024]***			
1993	0.0967 [0.0025]***		-0.0216 [0.0024]***			
1996	0.085 [0.0025]***					
2001	-0.193 [0.0025]***					
2004	-0.0456 [0.0023]***					
Observations	512207	56243	202769	93883	63845	95467
R-squared	0.15	0.4	0.1	0.1	0.12	0.12

Notes: We also included indicators for all missing values, receiving general assistance, foster care, other welfare, railroad retirement, veteran's assistance and state of residence. Robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Relationship between SIPP, SER and DER Earnings Measures

Panel A				
SER (1978-2002) on SIPP				
	Reliability Ratio	S.E.	N	R-Squared
All	0.422	[0.013]***	9304	0.27
31-35	0.438	[0.047]***	797	0.35
36-40	0.451	[0.029]***	1554	0.31
41-45	0.448	[0.025]***	2822	0.27
46-49	0.409	[0.033]***	880	0.21
50-55	0.34	[0.032]***	1268	0.21
White	0.423	[0.014]***	8290	0.26
Black	0.347	[0.037]***	631	0.23
Less HS	0.421	[0.034]***	780	0.48
HS	0.426	[0.030]***	2454	0.26
College	0.395	[0.018]***	6069	0.19

Panel B				
DER (1978-2002) on SIPP				
	Reliability Ratio	S.E.	N	R-Squared
All	0.583	[0.018]***	6030	0.52
31-35	0.48	[0.059]***	493	0.4
36-40	0.638	[0.044]***	961	0.55
41-45	0.678	[0.028]***	1821	0.63
46-49	0.566	[0.050]***	589	0.59
50-55	0.45	[0.040]***	862	0.46
White	0.598	[0.019]***	5517	0.53
Black	0.442	[0.049]***	353	0.45
Less HS	0.395	[0.065]***	336	0.31
HS	0.534	[0.041]***	1570	0.49
College	0.583	[0.023]***	4123	0.49

Panel C				
Maximum of SER and DER (1978-2002) on SIPP				
	Reliability Ratio	S.E.	N	R-Squared
All	0.524	[0.014]***	9441	0.52
31-35	0.506	[0.051]***	807	0.49
36-40	0.547	[0.037]***	1570	0.54
41-45	0.555	[0.023]***	2870	0.56
46-49	0.535	[0.040]***	894	0.55
50-55	0.431	[0.037]***	1282	0.47
White	0.529	[0.015]***	8416	0.52
Black	0.436	[0.043]***	638	0.5
Less HS	0.424	[0.035]***	784	0.46
HS	0.48	[0.029]***	2481	0.47
College	0.518	[0.018]***	6175	0.48

Figure 1: Match Rates by Age

