

ADEP WORKING PAPER SERIES

**The Summary Earnings Record and Detailed Earnings
Record Extracts**

Katie R. Genadek
U.S. Census Bureau

Charles Hokayem
U.S. Census Bureau

Philip Pendergast
U.S. Census Bureau

Working Paper 2021-05

November 2021

Associate Directorate for Economic Programs
U.S. Census Bureau
Washington DC 20233

Disclaimer: Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. Results were approved for release by the Census Bureau's Disclosure Review Board, authorization number CBDRB-FY21-ERD002-026.

The Summary Earnings Record and Detailed Earnings Record Extracts

Katie R. Genadek, U.S. Census Bureau, katie.r.genadek@census.gov

Charles Hokayem, U.S. Census Bureau, charles.hokayem@census.gov

Philip Pendergast, U.S. Census Bureau, philip.m.pendergast@census.gov

ADEP Working Paper 2021-05

November 2021

Abstract

The Census Bureau and the Social Security Administration (SSA) have a long history of linking respondents in Census Bureau surveys at the individual level to earnings data from the SSA for research purposes. Yet, these data have been underutilized and are often difficult to understand due to disparate and limited documentation. This paper describes the history and contents of the restricted-use Summary Earnings Record (SER) data and the Detailed Earnings Records (DER) data extracts that can be linked to the Survey of Income and Program Participation (SIPP) and the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC). These linked files provide an exciting opportunity for researchers to study individual earnings and related outcomes such as poverty, inequality, and the earnings distribution. Further, these data can be used for research in measurement error, non-response, and imputation. We discuss the linkage of the datasets and provide detailed yearly observation and linkage statistics. The data availability and access for researchers are also described.

Keywords: Earnings Records, Census data, Administrative records, CPS ASEC, SIPP

JEL Classification Codes: J30, C55

The U.S. Census Bureau collaborates with the Social Security Administration (SSA) to make data from the SSA's Summary Earnings Record (SER) and Detailed Earnings Record (DER) linked at the individual-level to the Census Bureau's Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC) and Survey of Income and Program Participation (SIPP) available in a restricted research environment. While there has been a long, yet limited, history of researchers using the SSA SER and DER and this linked data, there is little documentation of the data files. These data provide opportunities for researchers to study individual earnings and related outcomes such as poverty, inequality, and more broadly, the earnings distribution. Analyses of reported earnings in one year can be supplemented with administrative individual earnings histories that date back to 1951 (SER) and 1978 (DER). The linkage between administrative earnings records to the CPS ASEC and SIPP also allows for the study of long-standing topics in the survey literature such as measurement error, non-response, and imputation. Moreover, the earnings histories can be used for longitudinal analyses of earnings. Without comprehensive documentation of the files, they are difficult to use, and the resulting analyses can be easily misinterpreted. In this paper we describe the SSA data in detail, discuss how these data are linked to the CPS ASEC and SIPP, and provide guidance on using the earnings data with the CPS ASEC and SIPP.

SSA SER-DER Background

The Social Security Act was enacted in 1935 to provide monthly benefits to qualified individuals aged 65 or older based on their pre-retirement wages. To accurately calculate benefit amounts for its beneficiaries, the SSA's predecessor established Social Security Numbers (SSNs) that allowed employers to uniquely identify and accurately report the qualifying earnings associated with each worker. Earnings data up to the taxable maximum of \$3,000 were collected for all qualified individuals starting in 1937. However, in 1939 tax provisions were taken out of the Social Security Act and the Federal Insurance Contributions Act (FICA) was enacted. These FICA taxes continue to be withheld from earnings up to a yearly maximum earnings amount, fostering an arrangement between the Internal Revenue Service (IRS), employers, and the SSA to track workers' contributions to the Social Security program. It is through these mechanisms that the SSA obtains the data ultimately used to create the individual-level yearly earnings records for nearly all workers in the United States. The Summary Earnings Record (SER) and

Detailed Earnings Record (DER) are files created from these records containing yearly earnings information for those with those with an SSN.

Composition of the SER and DER

Both the SER and DER are summary files created from the SSA's Master Earnings File (MEF). The MEF data are presently derived on a weekly basis from Internal Revenue Service Form W-2, quarterly earnings records, and annual income tax forms for the purpose of calculating benefit amounts. Because these earnings data are comprised of IRS tax information, they are subject to protection by Title 26 U.S.C (Internal Revenue Code, 1986). In total, the MEF contains data on regular wages and salaries, tips, self-employment income, and deferred compensation, with records of some types of earnings going back to 1937 for the past and present working population of the United States.

Prior to posting income to an employee's earnings record, SSA verifies the name and SSN on all W-2 Forms or self-employment income reports (1040 SE) to match existing information in its Numerical Identification (Numident) file. Numident records are established when individuals apply for an SSN and contain information on each person's name, SSN, sex, race, birth date, and place of birth. Each year, the SSA processes approximately 245 million employee wage reports from 6.9 million employers and directly posts these earnings to the MEF for all employees with a name and SSN match in the Numident (Olsen and Hudson 2009). About 90% of wage reports are posted according to this procedure, but in cases where there is a mismatch on either identifier, a probabilistic match is generated. Between these two methods, 96% of wage items are posted to the MEF (Olsen and Hudson 2009). Remaining cases are placed in a separate file, the Earnings Suspense File (ESF), which can be resolved and posted to the MEF at a later date either through direct correspondence with employers or when workers apply for benefits and provide evidence of the wages missing from their earnings record. The SSA notes, however, that researchers using the MEF should be aware of possible incomplete or extraneous earnings for some workers, and that no indicator is provided to warn of erroneous earnings records (Olsen and Hudson 2009).

The MEF is stored in a compressed record structure, with each record divided into nine sections or segments (ten total historically) that each contain different types of information. The full list of segments are:

- 1) Header Segment – Contains SSN, name, sex, race, date of birth, and date of death from the Numident as well as a count of the number of entries found on the remainder of the record.
- 2) Summary Segment (Historical)– Annual FICA earnings from 1937 to present, along with annual information on Medicare Qualified Government Wages (MQGE/MQFE earnings), quarters of coverage, self-employment earnings, and military earnings indicators.
- 3) Summary Segment (Current) – Annual FICA earnings from 1951 to present, MQGE/MQFE earnings 1983 to present, and military credits 1968 to 2002, as well as quarters of coverage patterns 1951 to 1977, aggregate earnings 1937 to 1950, and indicators for disability, railroad earnings, and personal characteristics (sex, race, date of birth).
- 4) Posted Earnings Detail Segment – W-2 level data for all FICA and MQGE earnings since 1978, with “posted” referring to FICA-eligible income that is summarized in segments 2 and 3.
- 5) Unposted Detail Segment – Detailed non-FICA earnings, such as railroad wages, other non-covered earnings, deferred contributions, and Health Savings Account (HSA) contributions, which are not summarized in the summary segments.
- 6) Discrepancy Detail Segment – Indicates that earnings discrepancies have been identified and are being investigated.
- 7) Employer Special Payment Segment – Employer payments not considered to be detail items, such as sick pay, that are now housed in segment 5. This segment is not currently maintained.
- 8) Railroad Main Segment – Data from the Railroad Retirement Master File.
- 9) Railroad Film Segment – Data from the Railroad Retirement microfilm.
- 10) Old Start Quarterly Segment – Quarterly summaries, taken from microfilm, of earnings for the period 1937-1955.

The SER and DER are comprised of extracts from the MEF of these segments. A detailed examination of the data suggests that the SER consists of segments 2, 3, 8 and/or 9 and 10, whereas the DER appears to contain data from segments 4 and 5 (which contains data from 7). Generally, this means that the SER contains summary-level information of total FICA-eligible earnings with quarterly flags for different earnings types, whereas the DER contains more detail

on annual earnings, broken out into different sources of income, including deferred contributions and sick pay, which are not included in the SER.

Universe of the SER and DER

While the SER and the DER presently encompass nearly all types of workers, this has not always been the case. Under the original act, only workers in commerce and industries aside from railroads contributed to the Social Security program, and were included in the Social Security program. In 1951, coverage expanded to include most nonfarm self-employed workers, regularly employed farm and domestic workers, U.S. citizens employed abroad by U.S. employers, and workers in Puerto Rico and the U.S. Virgin Islands. Additional types of self-employed workers were added in 1955 (self-employed farm workers, self-employed professionals, and home workers) and 1966 (self-employed physicians, as well as interns and employees with tip income). Active duty military were added in 1956, with other types of Federal employees following in 1983 and 1984, state and local government employees in 1991, and police and firefighters in 1994. Generally, the SSA makes distinctions between three significant time periods related to these and other changes in Congressional law and technology: 1937-1950, when coverage is most limited and contain only aggregate, covered earnings; 1951-1977, when greater coverage and quarterly earnings data were first provided by employers; and 1978-present, when wage and salary earnings come from Form W-2 or 1040 Self Employment (SE) and nearly all workers and earnings are covered (Olsen and Hudson 2009). An important distinction between what is available in the SER versus the DER relates to which time periods are represented, with the SER containing data from all periods, while the DER covers only the most recent period (1978-present).

Data Available in the SER and DER Extracts

While the SER and DER ultimately contain similar information, the data files do differ in structure and specific content. The SER consists of two files: the adjusted file and the non-adjusted file. The adjusted SER file is topcoded at annual FICA limits, while the non-adjusted SER file is not, and the DER consists of just one file without topcoding. The SER is organized by worker and only contains one observation for the worker representing aggregated annual earnings, while the DER is organized by worker, job, and year, so a worker can appear multiple times per year in the DER if they have multiple jobs. Both SER files and the DER are bottom

coded at \$0, and only include positive self-employment earnings because individuals do not make self-employment tax contributions if they have losses (Nicholas and Wiseman 2009).

The SER annual earnings variables cover tax years 1951-present and are more useful than the agriculture and self-employment earnings variables. The agriculture and self-employment earnings variables include coded entries, but do not provide dollar amounts. The SER files have other individual variables such as date of birth, race, and sex. Additional variables provide the pattern of wage quarters credited toward Social Security in a specified year. For the period 1951-1977, the pattern for wage and salary workers is four digits with a 1 signifying a credit in the quarter and a zero for no credit in a quarter. For example, a pattern of 1101 would indicate credits earned in the first, second, and fourth quarter of the year. When a person reaches the annual taxable maximum in a quarter, subsequent quarters would be zero. If earnings are at the annual taxable maximum, a pattern of 1100 would indicate the maximum was reached in the first half of the year. The patterns begin including reports for agricultural and self-employed earnings, as well as wage and salary earnings, in 1978. Because agricultural quarters, self-employment quarters, and wage and salary quarters beginning in 1978 are based on annual earnings, these pattern variables cannot be used to identify patterns of employment and estimates of total earnings for persons at the taxable maximum.

The DER variables cover standard boxes on the yearly W-2 forms from 1978 forward. These boxes broadly cover wage and salary earnings, deferred contributions to 401(k), 403(b), 408(k), 457(b), and 501(c) retirement and trust plans as well as HSAs. Self-employment earnings variables come from 1040-SE forms. The DER also includes the Employer Identification Number (EIN) available on W-2 forms. This allows for linkages to other business and economic data that include EIN. The DER variables do not cover taxes withheld or taxes paid.

The earnings in the DER are not topcoded at FICA limits, and in theory, could be aggregated across jobs and earnings categories to match earnings reported in the non-adjusted SER. In many cases the combined DER earnings match or are similar to the SER earnings, yet they are often not the same. For the SER and DER data linked to the CPS ASEC, we find that the wage and salary earnings from the DER correspond exactly to SER earnings at the individual level for 86.5% of individuals with earnings. Combining wage and salary earnings, self-employment earnings, and deferred contributions in the DER results in an exact match to SER earnings for

about 12% of the remaining sample. For the final 1.5% of the sample of respondents with earnings, it is not clear from the data how the DER earnings correspond to the SER earnings.

SIPP and CPS ASEC Background

SIPP

The Survey of Income and Program Participation (SIPP) is a nationally representative survey that has been collected by the Census Bureau since 1983.¹ Households are surveyed regularly over 2.5 – 4 years in order to understand household income, family, and public program participation dynamics longitudinally. While the survey has changed over time, starting in 2014 SIPP went to an annual data collection rather than sub-annual. The SIPP remains the premier survey data for understanding households and government program participation. The SIPP includes between 15,000-53,000 designated households in each panel of the survey.

CPS

The Current Population Survey (CPS) is a household survey sponsored by the U.S. Census Bureau and the Bureau of Labor Statistics.² A monthly survey, the CPS is the primary source of information on employment status for the United States. The CPS originated as the Monthly Report of Unemployment in 1940, and it continues as a monthly survey with supplemental data collection in various months. The Annual Social and Economic Supplement (ASEC), formerly known as the March Demographic Supplement, is the largest supplemental CPS survey, with microdata available back to 1962.

The CPS ASEC includes detailed information on households and individuals within the household. In addition to the basic employment data collection, the supplement also includes extensive information on program participation, income from the prior year, education, health insurance and migration. It is the source of the Census Bureau's official income, inequality, poverty, and health insurance statistics. In 1962-1976, between 20,000 and 50,000 households were included in the sample each year, while in 1976-2000 about 70,000 households across the U.S. were surveyed each year. In 2001, the ASEC was expanded and redesigned and has since included about 100,000 households each year.

¹ For detailed information see: [Survey of Income and Program Participation \(SIPP\) \(census.gov\)](https://www.census.gov/sipp/). There is additional information on the SIPP found in the [Source and Accuracy Statements \(census.gov\)](https://www.census.gov/sipp/source-accuracy-statements/) and the [Nonresponse Bias Reports \(census.gov\)](https://www.census.gov/sipp/nonresponse-bias-reports/).

² For detailed information see [Current Population Surveys \(CPS\) \(census.gov\)](https://www.census.gov/cps/).

Linkage of SIPP/CPS ASEC to the SER/DER Extracts

The Census Bureau assigns unique, anonymous identifiers to respondents in surveys based on available Personally Identifiable Information (PII) using the Census Bureau's record linkage software system, the Person Identification Verification System (PVS). The unique identifiers, called Protected Identification Keys (PIKs), are assigned based on a probabilistic matching of available PII in a survey or administrative record (which can include name, SSN, address, date of birth, and gender) to a reference file built from administrative and survey data with PII and PIKs (Wagner and Layne 2014). Researchers only have access to data without PII, but they can link respondents across surveys and administrative data using PIKs, often using a crosswalk that only includes the PIK and the survey respondent-level identifiers. For the CPS ASEC and SIPP surveys, PIKs are assigned primarily based on the name, address, date of birth, and gender of respondents. Not all respondents are assigned PIKs due to missing or inaccurate information in the survey, respondents not being in the reference file, or respondents not being uniquely linked to the reference file. Thus, the availability of PIKs is non-random, yet the error in PIK assignment is minimal (Layne et al. 2014, Rastogi et al. 2012). Tables 2 and 3 show the rate of PIK assignment in the CPS ASEC and SIPP by year, respectively. As PII is only available in some years of the CPS ASEC, the PIK Crosswalks are limited to 1973, 1991, 1994, and 1996-present, while the PIK crosswalks for the SIPP files include 1984, 1990-1993, 1996, 2001, 2004, 2008, 2014, and 2018.³

For linking the CPS ASEC/SIPP to the SER and DER data, the Census Bureau refines the PIK crosswalks, then sends the SSNs associated with PIKs in the CPS and SIPP samples securely to SSA each year.⁴ SSA extracts full SER and DER histories for each SSN sent to them at the time the extract is taken and returns data files to the Census Bureau. Thus, earnings records are available before and after a respondent is observed in the SIPP or CPS ASEC, and they are updated with each additional extract. For example, an individual sampled in the 2010 ASEC will have earnings records before and after 2010 going back to 1978 for the DER and back to 1937 for the SER. The Census Bureau removes the SSNs and makes the SER and DER extracts available with PIK crosswalks linking the SER and DER to a PIK, and thus to the SIPP and CPS

³ The Census Bureau has developed PIK crosswalks for CPS ASEC files for 1979, 1981-1989, 1992, 1993, 1995. These years may be included in the SER and DER extracts in the future.

⁴ The original PIK crosswalks and refined crosswalks, known as the "consolidated crosswalks" are both available to researchers through a Federal Statistical Research Data Center (FSRDC).

ASEC data. While only respondents with earnings are included in the DER extract, the SER extract includes nearly all respondents with an SSN, so many of the SER entries have no earnings.

The linkage rates between the SIPP and the SER/DER using the SIPP crosswalk from 2019 are presented in Table 1. We consider a link to the DER to occur when the SIPP respondent's PIK is found in any year of the DER. Because the SER extract includes entries for nearly all respondents with PIKs, we consider a link occurring when a SIPP respondent has non-zero earnings in any year of the SER. In early years of the SIPP, very few respondents were assigned a PIK, but comparing the percent of total observations with a PIK and those linked to the DER and SER reveals that many of those with a PIK were linked to the DER and to a SER earnings record. In the 1984-1993 files, less than 20% of respondents in each SIPP year have a PIK, compared to 83.2% of respondents in 1996. While in 2001 only 50.8% of respondents have PIKs, 96.7% (44,500/46,000) of the respondents with PIKs link to earnings at some point in the DER and the SER. PIK rates increase dramatically starting in 2008, and over 90% of respondents have PIKs assigned after 2010. However, the rate at which these respondents link to a DER and SER records with non-zero earnings falls over time. This is in part a result of the increase in PIK rates among younger respondents who have not yet entered the workforce, but this also may reflect changes in the economy and those with reported earnings. In the most recent SIPP file, 2018, 72.7% of respondents were linked to the DER and the SER and displayed earnings in at least one year.

Table 2 shows the linkage rates for the CPS ASEC data to the SER and DER data.⁵ Similar to the older SIPP data, prior to 2001 the percentage of total records with PIKs and the percentage of total records linked to a DER record or a SER record with earnings are similar. However, the overall PIK rates are higher in the older CPS ASEC data than in the SIPP data. In 1973, 1991, 1994, 1998, 1999, and 2000, only respondents age 15 and over were eligible to be assigned a PIK, so between 52.7%-67.7% of respondents in each year have a PIK. For the 1996, 1997, and 2001-present CPS ASEC samples, all respondents were eligible to be assigned a PIK, so PIK rates are between 69.7% and 89.6%. Thus, as expected, a smaller percentage of

⁵ We use the same linkage definitions with the CPS ASEC and the SIPP. Thus, we consider a link to the DER to occur when the CPS ASEC respondent's PIK is found in any year of the DER, and we consider a link occurring when a CPS ASEC respondent has non-zero earnings in any year of the SER.

respondents with a PIK are linked to an earnings record in these more recent samples when compared to the older samples. Since 2000, linkage to an earnings record peaked in 2006 and has declined in the samples since that time. This is a byproduct of using the most recent extract of the SER and DER, where younger people in the later CPS ASEC sample years will not have entered the labor force yet. But in future SER and DER extracts, these respondents will most likely have an earnings record.

To further investigate the CPS ASEC linked to the administrative earnings records, we focus on all CPS ASEC respondents with reported earnings in the prior year. We identify earners in the CPS ASEC as any respondents reporting any earnings in the previous year. We link these respondents to the appropriate earnings year in the DER, and count this as a match if they also have earnings according to the DER in that year. This exercise represents a common use of these data, comparing self-reported survey earnings in the CPS ASEC to the administrative earnings in the SSA files. The results of this linkage are shown in Table 3. As shown in the table, on average, 93.8% of earners with a PIK assigned in the CPS ASEC in a given year have corresponding earnings in the DER. Thus, this group has individual administrative earnings that are directly comparable to the earnings measures in the CPS ASEC. We also look at those in the CPS ASEC that do not report any earnings in the previous year, and we match these respondents to the DER for that earnings year. The percent of these respondents that we are able to match to the DER range from 7.5% in 2005 to 16.7 in 2000, with an outlier in 1991 of 37.5%. To examine how the linkage varies across the earnings distribution, we identified earners in the CPS ASEC across all survey years with PIK crosswalks, and adjusted earnings using the Consumer Price Index (CPI), and categorized the earners by percentile of earnings as reported in the CPS ASEC. As shown in Table 4, the linkage to the DER is not constant throughout the earnings distribution. For the bottom earnings percentiles (0-5%), 62.2% of the CPS ASEC earners have earnings in the DER. However, for the top percentiles (95%-100%), 83% of the earners have earnings in the DER.

As mentioned earlier, PIKs are probabilistically assigned, and based on data available in various administrative records, the assignment of PIKs is non-random in these samples. Table 5 shows the demographic and geographic breakdown of the 2014 SIPP and CPS ASEC samples to further understand the biases that may be present in the CPS ASEC and SIPP samples linked to the DER earnings files. For each demographic characteristic, Columns 1 and 4 show the mean

estimate from the unweighted CPS ASEC and SIPP samples, followed by the estimates from the weighted samples in Columns 2 and 5, and estimates from the sample linked to the DER in Columns 3 and 6. The SIPP sample that is linked to the earnings data is less Hispanic and has more U.S. born individuals than the weighted SIPP. More of the linked respondents are living in the south and in urban areas than the weighted SIPP as well. Similarly, the CPS ASEC sample in 2014 that is linked to the DER is less Hispanic than the weighted CPS ASEC, and the linked respondents are more likely to be born in the U.S. than the weighted estimates. Unlike the linked SIPP-DER sample, the CPS ASEC-DER linked sample is less likely to be living in the south but still more likely to be living in a rural area than the weighted CPS ASEC sample.

Previous Research using CPS ASEC/SIPP linked to SER/DER

The first linkage of the CPS ASEC data to the SER/DER data used the 1973 CPS ASEC. This work is documented in Kliss and Scheiren’s “The 1973 CPS-IRS-SSA exact match study,” which describes in detail the linkage and includes many references to research and reports using these data (Kliss and Scheiren 1978). More recently, the SER and DER data linked to the CPS and SIPP have been used to compare survey and administrative data (Roemer 2002; Abowd and Stinson 2013; O’Hara et al 2017; Davies and Fisher 2019; Medalia et al 2019) and to understand survey non-response and improve survey imputation (Chenevert et al. 2015; Hokayem et al. 2015; Bollinger et al. 2019; Klee et al. 2019; Hokayem et al. 2020; Ziliak et al. 2020; Eggleston et al. 2020). These data have also been used to answer various research questions related to employment, earnings, and poverty (Joyce and Wiseman 2009; Dahl et al. 2011; Bound et al. 2014; Tamborini et al. 2015; Klee and Warren 2017).

Data Access

The CPS ASEC, SIPP, and their associated SER and DER extracts are available to researchers on approved projects in a restricted research environment through the Federal Statistical Research Data Centers (FSRDC). The FSRDC network currently includes 32 physical research centers at universities and research institutions. Researchers can apply to use these data through the standard Census Bureau FSRDC application process, which starts by contacting the closest FSRDC.⁶ In addition to using the CPS ASEC and SIPP data linked to the SER/DER extracts, The CPS ASEC and SIPP data can be linked to any available Census Bureau data set with an associated PIK crosswalk.

⁶ FSRDC Locations: <https://www.census.gov/about/adrm/fsrdc/locations.html>

The CPS ASEC and SIPP samples linked to the SSA earnings data have a near 50-year history of enhancing our understanding of wages and income in the United States. With the detailed history provided in this paper, alongside observations counts, linkage information, and sample statistics, researchers are now able to fully understand the data available and propose research using these files. The SER and DER extracts contain one of the longest panels of earnings in the U.S., and the ongoing large survey samples from the CPS ASEC and SIPP provide rich additional information that increase the potential power in using these data going forward. Researchers now have an opportunity to study earnings over time and related outcomes as well as survey measurement error and non-response.

References

- Abowd, John M. and Martha H. Stinson. (2013). "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data." *Review of Economics and Statistics*, 95 (December): 1451-1467.
- Bollinger, Christopher R., Barry Hirsch, Charles Hokayem, and James P. Ziliak. (2019). "Trouble in the Tails? What We Know About Earnings Nonresponse Thirty Years After Lillard, Smith, and Welch." *Journal of Political Economy*, 129(5):2143-2185.
- Bound, John, Stephan Lindner and Timothy Waidmann. (2014). "Reconciling Findings on the Employment Effect of Disability Insurance." *IZA Journal of Labor Policy* 3, 11.
- Chenevert, Rebecca L., Mark A. Klee and Kelly R. Wilkin. (2015). "Do Imputed Earnings earn Their Keep? Evaluating SIPP Earnings and Nonresponse with Administrative Records." SIPP Working Paper Number 275.
- Davies, Paul S. and T. Lynn Fisher. (2009). "Measurement Issues Associated with Using Survey Data Matched with Administrative Data from the Social Security Administration." *Social Security Bulletin*, 69(2).
- Dahl, Molly, Thomas DeLeire, and Jonathan Schwabish. (2011). "Estimates of Year-to-Year Variability in Worker Earnings and in Household Incomes from Administrative, Survey, and Matched Data." *Journal of Human Resources*, 46(4): 750-774
- Eggleston, Jonathan, Mark A. Klee, and Robert Munk. (2020) "Improving Self-Employment Imputations with Administrative Data and Model Based Imputations: An Analysis Using the Survey of Income and Program Participation". Working Paper, U.S. Census Bureau.
- Hokayem, Charles, Christopher Bollinger, and James P. Ziliak. (2015). "The Role of CPS Nonresponse in the Measurement of Poverty." *Journal of the American Statistical Association* 110 (September): 935-045.
- Hokayem, Charles, Trivellore Raghunathan, and Jonathan Rothbaum. "Match Bias or Nonignorable Nonresponse? Improved Imputation and Administrative Data in the CPS ASEC." *Journal of Survey Statistics and Methodology*, Forthcoming.
- Internal Revenue Code, 26 U.S.C. § 1 et seq. (1986).
- Klee, Mark A. and Lewis H. Warren. (2017) "What Are We Searching for? Estimating the Returns to Job Search." SEHSD Working Paper, U.S. Census Bureau.
- Klee, Mark A., Rebecca L. Chenevert, and Kelly R. Wilkin. (2019). "Revisiting the Shape of Earnings Nonresponse" *Economics Letters*, 184.
- Kilss, Beth, and Frederick J. Scheuren. (1978) "The 1973 CPS-IRS-SSA exact match study." *Social Security Bulletin* 41(10): 14-22.
- Layne, Mary., Wagner, Deborah., and Rothhaas, Cynthia. (2014). "Estimating record linkage false match rate for the Person Identification Validation System." Center for Administrative Records Research and Applications Working Paper, No. 2014-02.
- Medalia, Carla, Bruce Meyer, Amy O'Hara and Derek Wu. (2019). "Linking Survey and Administrative Data to Measure Income, Inequality, and Mobility." *International Journal of Population Data Science*, 4(1), 4-12.

- Nicholas, Joyce and Michael Wiseman. (2009). "Elderly Poverty and Supplemental Security Income." *Social Security Bulletin*, 69 (May): 45-73.
- O'Hara, Amy, Rachel M. Shattuck and Robert M. George. (2017). "Linking Federal Surveys with Administrative Data to Improve Research on Families." *The ANNALS of the American Academy of Political and Social Science*, 669(1), 63-74.
- Olsen, Anya and Russel Hudson (2009). "Social Security Administration's Master Earnings File: Background Information" *Social Security Bulletin*, 69(3): 29-45.
- Roemer, Mark. (2002). "Using Administrative Earnings Records to Assess Wage Data Quality in the Current Population Survey and the Survey of Income and Program Participation." *Longitudinal Employer-Household Dynamics Program Technical Paper No. TP-2002-22*, U.S. Census Bureau.
- Rastogi, Sonya, Amy O'Hara, James Noon, Ellen A. Zapata, Cindy Espinoza, Leah B. Marshall, Teresa A. Schellhamer, and J. David Brown. (2012). "2010 Census Match Study." *Center for Administrative Records Research and Applications Report*, U.S. Census Bureau.
- Tamborini, Christopher R., Changhwan Kim and Arthur Sakamoto. (2015). "Education and Lifetime Earnings in the United States." *Demography*, 52, 1383-1407.
- Wagner, Deborah., and Layne, Mary. (2014). "The person identification validation system (PVS): applying the Center for Administrative Records Research and Applications' (CARRA) record linkage software." *Center for Economic Studies*, U.S. Census Bureau, No. 2014-01.
- Ziliak, James P., Charles Hokayem, and Christopher Bollinger (2020, May). "Trends in Earnings Volatility using Linked and Administrative Survey Data." *University of Kentucky Center for Poverty Research Discussion Paper Series*, DP2020-01.

Table 1. SIPP Observations and Linkage to DER and SER

SIPP Year	Observations	Observations with PIK assigned	% of total observations	Observations linked to any year of DER	% of total observations	Observations linked to any year of SER with earnings	% of total observations
1984	64,500	11,000	17.1%	10,000	15.5%	10,500	16.3%
1990	69,500	12,500	18.0%	12,000	17.3%	12,500	18.0%
1991	44,500	8,000	18.0%	7,600	17.1%	7,800	17.5%
1992	62,500	11,000	17.6%	10,500	16.8%	11,000	17.6%
1993	62,500	11,000	17.6%	10,500	16.8%	11,000	17.6%
1996	95,500	79,500	83.2%	76,500	80.1%	77,500	81.2%
2001	90,500	46,000	50.8%	44,500	49.2%	44,500	49.2%
2004	111,000	86,000	77.5%	81,000	73.0%	79,500	71.6%
2008	106,000	94,500	89.2%	84,000	79.2%	80,000	75.5%
2014	73,000	66,500	91.1%	54,000	74.0%	52,000	71.2%
2018	64,000	60,000	93.8%	46,500	72.7%	46,500	72.7%

Source: 1984-2018 Survey of Income and Program Participation (SIPP) and Social Security Administration (SSA) Summary Earnings Record (SER) and Detailed Earnings Record (DER) extracts.

Notes: All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-026. The counts are rounded per Census Bureau Disclosure Review Board guidelines.

Table 2. CPS ASEC Observations and Linkage to DER and SER

ASEC Year	Observations	Observations with PIK assigned	% of total observations	Observations linked to any year of DER	% of total observations	Observations linked to any year of SER with earnings	% of total observations
1973	136,000	84,000	61.8%	67,000	49.3%	78,000	57.4%
1991	158,000	106,000	67.1%	99,000	62.7%	103,000	65.2%
1994	151,000	92,500	61.3%	87,500	57.9%	90,500	59.9%
1996	130,000	109,000	83.8%	105,000	80.8%	107,000	82.3%
1997	132,000	108,000	81.8%	104,000	78.8%	105,000	79.5%
1998	132,000	75,500	57.2%	72,500	54.9%	74,000	56.1%
1999	132,000	69,500	52.7%	67,000	50.8%	68,500	51.9%
2000	134,000	71,000	53.0%	68,500	51.1%	70,000	52.2%
2001	129,000	100,000	77.5%	93,000	72.1%	94,000	72.9%
2002	217,000	172,000	79.3%	165,000	76.0%	167,000	77.0%
2003	216,000	167,000	77.3%	158,000	73.1%	159,000	73.6%
2004	213,000	152,000	71.4%	142,000	66.7%	143,000	67.1%
2005	211,000	147,000	69.7%	134,000	63.5%	135,000	64.0%
2006	209,000	184,000	88.0%	168,000	80.4%	169,000	80.9%
2007	207,000	183,000	88.4%	164,000	79.2%	165,000	79.7%
2008	206,000	181,000	87.9%	160,000	77.7%	161,000	78.2%
2009	208,000	182,000	87.5%	158,000	76.0%	159,000	76.4%
2010	210,000	185,000	88.1%	156,000	74.3%	156,000	74.3%
2011	205,000	184,000	89.8%	152,000	74.1%	152,000	74.1%
2012	201,000	180,000	89.6%	146,000	72.6%	146,000	72.6%
2013	203,000	179,000	88.2%	144,000	70.9%	144,000	70.9%
2014	200,000	174,000	87.0%	139,000	69.5%	139,000	69.5%
2015	199,000	173,000	86.9%	136,000	68.3%	136,000	68.3%
2016	185,000	161,000	87.0%	124,000	67.0%	124,000	67.0%
2017	186,000	159,000	85.5%	121,000	65.1%	121,000	65.1%
2018	180,000	154,000	85.6%	115,000	63.9%	115,000	63.9%

Source: 1973-2018 Current Population Survey (CPS) Annual Social and Economic Supplements (ASEC) and Social Security Administration (SSA) Summary Earnings Record (SER) and Detailed Earnings Record (DER) extracts.

Notes: All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-026. The counts are rounded per Census Bureau Disclosure Review Board guidelines.

Table 3. CPS ASEC and DER observations for direct earnings comparisons

CPS ASEC Survey Year	Earnings Year	Earners with PIK	Earners Linked to DER	Percent	Non- earners with PIK	Non- earners linked to DER	Percent
1991	1990	73,500	68,500	93.2%	32,000	12,000	37.5%
1994	1993	63,500	59,500	93.7%	29,000	4,200	14.5%
1996	1995	56,500	53,000	93.8%	52,500	4,100	7.8%
1997	1996	56,000	52,500	93.8%	52,000	4,000	7.7%
1998	1997	52,500	49,500	94.3%	23,000	3,700	16.1%
1999	1998	49,000	46,000	93.9%	21,000	3,400	16.2%
2000	1999	50,500	47,000	93.1%	21,000	3,500	16.7%
2001	2000	50,000	47,000	94.0%	47,000	4,100	8.7%
2002	2001	87,500	83,000	94.9%	84,500	7,200	8.5%
2003	2002	83,000	78,500	94.6%	84,000	7,000	8.3%
2004	2003	73,500	69,500	94.6%	79,000	6,200	7.8%
2005	2004	70,500	66,500	94.3%	77,000	5,800	7.5%
2006	2005	97,000	91,500	94.3%	87,500	8,600	9.8%
2007	2006	96,000	91,000	94.8%	87,000	9,000	10.3%
2008	2007	95,500	91,000	95.3%	85,500	9,000	10.5%
2009	2008	96,000	90,500	94.3%	86,500	8,800	10.2%
2010	2009	94,500	88,000	93.1%	90,500	9,000	9.9%
2011	2010	92,500	85,500	92.4%	92,000	9,400	10.2%
2012	2011	90,000	83,500	92.8%	90,000	9,200	10.2%
2013	2012	90,000	84,500	93.9%	88,500	9,100	10.3%
2014	2013	88,000	82,500	93.8%	86,000	9,100	10.6%
2015	2014	87,500	82,000	93.7%	85,500	9,300	10.9%
2016	2015	82,000	76,500	93.3%	79,000	8,400	10.6%
2017	2016	82,000	77,000	93.9%	77,500	8,300	10.7%
2018	2017	79,000	74,000	93.7%	75,000	8,300	11.1%

Source: 1973-2018 Current Population Survey (CPS) Annual Social and Economic Supplements (ASEC) and Social Security Administration (SSA) Summary Earnings Record (SER) and Detailed Earnings Record (DER) extracts.

Notes: All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-026. The counts are rounded per Census Bureau Disclosure Review Board guidelines.

Table 4. DER Linkage to CPS ASEC by
CPS ASEC Earnings Percentile

CPS ASEC Earnings Percentile	% Linked to DER
0-5	62.2
5-10	74.5
10-15	75.0
15-20	75.5
20-25	73.8
25-30	74.2
30-35	75.4
35-40	76.5
40-45	78.0
45-50	78.6
50-55	79.7
55-60	79.5
60-65	81.1
65-70	81.7
70-75	81.8
75-80	82.6
80-85	83.1
85-90	83.0
90-95	83.2
95-100	83.0

Source: 1973-2018 Current Population Survey (CPS) Annual Social and Economic Supplements (ASEC) and Social Security Administration (SSA) Summary Earnings Record (SER) and Detailed Earnings Record (DER) extracts.

Notes: All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-026. The counts are rounded per Census Bureau Disclosure Review Board guidelines.

Table 5. Characteristics of 2014 CPS ASEC and SIPP Samples, for Earners (Age 15+)

	CPS ASEC	Weighted CPS ASEC	CPS ASEC- Linked to DER	SIPP	Weighted SIPP	SIPP- linked to DER
Age	41.9	42.2	42.0	49.6	49.0	50.2
Male	0.52	0.53	0.51	0.48	0.49	0.48
Female	0.48	0.47	0.49	0.51	0.50	0.52
White, Non-Hispanic	0.64	0.66	0.67	0.78	0.79	0.79
Black, Non-Hispanic	0.10	0.11	0.10	0.15	0.13	0.15
Asian, Non-Hispanic	0.06	0.05	0.06	0.04	0.06	0.04
Other race, Non-Hispanic	0.03	0.02	0.03	0.03	0.03	0.03
Hispanic	0.17	0.16	0.15	0.13	0.13	0.12
US born	0.83	0.84	0.86	0.86	0.84	0.88
Foreign Born, Citizen	0.17	0.16	0.14	0.08	0.10	0.08
Foreign Born, Noncitizen	0.09	0.08	0.06	0.06	0.06	0.04
Northeast	0.19	0.18	0.19	0.13	0.18	0.13
South	0.31	0.36	0.31	0.43	0.37	0.43
Midwest	0.23	0.22	0.24	0.22	0.22	0.23
West	0.27	0.23	0.25	0.21	0.23	0.20
Urban	0.82	0.86	0.81	0.77	0.81	0.77
Rural	0.18	0.14	0.19	0.23	0.19	0.23
N (unweighted)	99,000	99,000	82,500	47,500	47,000	44,000

Source: 2014 Survey of Income and Program Participation (SIPP), 2014 Current Population Survey (CPS) Annual Social and Economic Supplements (ASEC) and Social Security Administration (SSA) Summary Earnings Record (SER) and Detailed Earnings Record (DER) extracts.

Notes: All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-026. The counts are rounded per Census Bureau Disclosure Review Board guidelines.