The Role of CPS Non-Response on Trends in Poverty and Inequality*

Charles Hokayem, U.S. Census Bureau

James P. Ziliak, Department of Economics and Center for Poverty Research, University of Kentucky

Christopher Bollinger, Department of Economics, University of Kentucky

Preliminary Draft, October 2012

Abstract: The Current Population Survey Annual Social and Economic Supplement (CPS ASEC) serves as the data source for official income, poverty, and inequality statistics in the United States. There is a concern that the rise in non-response to earnings questions could deteriorate data quality and distort estimates of these important metrics. We use a dataset of internal CPS ASEC records matched to Social Security Detailed Earnings Records (DER) to study the impact of earnings non-response on estimates of poverty and inequality over the time period 1997-2008. Our analysis does not treat the administrative data as the "truth"; instead, we rely on information from both administrative and survey data. Substituting administrative earnings data for earnings imputed in the CPS ASEC produces overall poverty rates that are higher than the official poverty rate but not as high as poverty rates produced from completely dropping imputed earners. Completely dropping imputed earners gives the highest poverty rates for adults, seniors, Whites, Blacks, men, women, and those with a high school education or less. Completely dropping imputed earners also gives the highest percentile ratio series (90/10,90/50, and 50/10), while replacing CPS earnings with DER earnings for only imputed earners produces the lowest series.

Key Words: CPS ASEC, poverty measurement, inequality, hot deck imputation, non-response bias, earnings, measurement error
JEL Codes: I32 (Measurement and Analysis of Poverty); J31 (Wage Level and Structure)

---

*Contact: Charles Hokayem, U.S.Census Bureau, SEHSD, HQ-7H176, 4600 Silver Hill Rd, Washington, DC 20233-8500. E-mail: charles.hokayem@census.gov Phone: 301-763-5330. James P. Ziliak, University of Kentucky, Lexington, KY 40506-0034; Email: jziliak@uky.edu; Phone: 859-257–2776. Christopher Bollinger, University of Kentucky, Lexington, KY 40506-0034; Email: crboll@uky.edu; Phone: 859-257-9524. We would like to thank Martha Stinson, Graton Gathright, and Gary Benedetto for their help in understanding the Detailed Earnings Record files. We also thank Ed Welniak, Trudi Renwick, Chuck Nelson, Adam Bee, and session participants at the 2012 Joint Statistical Meetings for helpful comments.

1. Introduction

The accurate measurement of the income distribution and poverty statistics is vital to assessing economic growth, characterizing income inequality, and gauging the effectiveness of the federal safety net. The Current Population Survey Annual Social and Economic Supplement (CPS ASEC) serves as the official source of income and poverty statistics for the United States. CPS ASEC respondents may be reluctant to answer income questions out of concern for response confidentiality, or they may just have insufficient knowledge of the answer (Groves 2001). As seen in Figure 1, the non-response rate for ASEC earnings in the population has risen dramatically since the early 1990's. Among the non-institutional population ages 16 and older, the imputation rate has reached over 10 percent (the line with squares), while among the subsample of wage, salary, and self-employed workers (the diamonds) it reached a peak of just over 20 percent in the early 2000's.[1] Rates of non-response for other earnings (e.g., self-employment) trended upward in the 1990's, but they only contribute 1-2 percentage points per year, implying most is due to wage and salary workers. Because earnings accounts for over 80 percent of total income, failure to accurately measure it may significantly bias estimates of poverty and inequality.

This paper assesses the extent of the bias in poverty rates and inequality measures caused by earnings non-response and the hot deck procedure. One method of addressing non-response simply deletes missing data and uses sampling weights to calculate population statistics of interest (Bollinger and Hirsch 2006; Ziliak 2006). An alternative method to address non-response fills in missing data using a matching procedure that relies on matching observations with missing data to observations with complete data based on socioeconomic characteristics

[1] This non-response rate is based on the earnings flag. If we also include those observations that have the entire ASEC imputed, the non-response rate averages just over 30 percent in the past decade.

(Little and Rubin 2002). This second procedure, referred to as a cell "hot deck," offers the advantage of retaining more observations in the final data set than simply deleting any observation with missing data; however, the hot deck procedure may bias estimates of population statistics if the conditional missing at random assumption does not hold (Bollinger and Hirsch forthcoming). Hirsch and Shumacher (2004) and Bollinger and Hirsch (2006) study the hot deck procedure in a related survey, the CPS Outgoing Rotation Group, and show the hot deck procedure causes earnings regression parameters to be biased. Given the bias in regression parameters there is a possibility the hot deck procedure could bias estimates of statistics derived from income such as poverty rates and inequality measures.

We assemble a dataset of internal CPS ASEC records matched to Social Security Detailed Earnings Records (DER) to study the impact of earnings non-response on estimates of poverty and inequality. The CPS ASEC-DER matched data file covers CPS ASEC years 1998-2009, allowing for the systematic study of long term trends in income imputation, poverty rates, and inequality. We present an analysis of the bias in poverty rates and inequality statistics (Gini coefficient, 90/10 ratio, 90/50 ratio, and 50/10 ratio) by comparing current Census practice of retaining imputed earnings with four alternatives: (1) dropping observations with imputed earnings; (2) dropping observations with imputed earnings and reweighting with inverse probability weights; (3) replacing ASEC earnings with DER earnings for all persons with a DER match regardless of imputation status and use ASEC earnings for persons without a DER match; and (4) replacing ASEC earnings with DER earnings only for those persons with imputed earnings and a DER match and use ASEC earnings for persons without a DER match. Our analysis does not treat the administrative data as correct or the "truth"; instead, we rely on information from survey and administrative sources.

Our results show that substituting administrative earnings data for earnings imputed in the CPS ASEC produces overall poverty rates that are higher than the official poverty rate but not as high as poverty rates produced from completely dropping imputed earners, suggesting survey non-response is more prevalent among higher earners. Moreover, completely dropping imputed earners also gives the highest percentile ratio series (90/10, 90/50, and 50/10), while replacing CPS earnings with DER Earnings for only imputed earners produces the lowest series.

2. Literature Review

Several papers examine the effect of measurement error and income imputation on poverty and inequality. Chesher and Schluter (2002) provide a theoretical treatment of measurement error on various measures of welfare. Their derivations allow a study of the sensitivity of income inequality and poverty measures to the amount of measurement error variance in the income distribution. Their simulations comparing income distributions with and without measurement error show measurement error can upwardly bias poverty rates and Gini coefficients. Poverty rates and Gini coefficients measured in surveys may overstate poverty and income inequality. Chesher and Schluter apply their method to measuring the degree of this bias to regional poverty and inequality in Indonesia.

Nicholas and Wiseman (2009) merge administrative data from the Social Security Administration (SSA) with the CPS ASEC 2003 to study poverty among the entire U.S. population and among the elderly for calendar year 2002. Their analysis uses several SSA files for earnings, Old-Age, Survivors, and Disability Insurance (OASDI) payments (social security), and SSI payments. Wage and salary earnings come from Summary Earnings Record (SER) and Detailed Earnings Record (DER) files; Social Security benefits come from the Payment History Update System (PHUS) file; and SSI payments come from the Supplemental Security Record file. Using administrative records for SSI payments corrects for underreporting of this benefit in the CPS ASEC. Their analysis substitutes administrative earnings for CPS earnings and self-employment income when available, leaving all other sources of income from the CPS. Nicholas and Wiseman develop measures of income that vary on the availability of administrative and CPS data and employ a reweighting adjustment for CPS observations unmatched to the administrative data. Their results confirm that the CPS substantially understates SSI receipt. They find that using administrative data reduces official poverty rates for the entire national population and for the SSI recipient population. The poverty rate for the entire U.S. population falls from 12.1 percent to between 9.3 percent and 11.8 percent while the SSI poverty rate falls from 44.3 percent to between 39.0 and 40.9 percent. Using a relative measure of poverty, half of equivalence-adjusted median income, has a smaller effect on poverty rates.

Like Nicholas and Wiseman (2009) Turek et al. (2012) use administrative data from the Social Security Administration to study poverty with a focus on the effects of income imputation in the CPS on poverty. Turek et al. merge earnings information from the Detailed Earnings Record file with the CPS ASEC 2006 (calendar year 2005) to examine the effect of substituting DER earnings for reported CPS earnings on income estimates and number of persons in poverty. Their analysis separates individuals by CPS imputation status: no imputes, item imputes, and whole imputes. Item imputes are individuals who respond to the CPS ASEC supplement but need specific income questions imputed. Whole imputes are individuals who refuse to respond to the CPS ASEC supplement and need the entire supplement, including all income questions, imputed. After substituting DER earnings for CPS earnings, an overwhelming majority of individuals do not change poverty status. The poverty status for 93.7 percent of all individuals does not change. This result holds by all three imputation types: no imputes (94.4 percent), item imputes (92.8 percent), and whole imputes (89.2 percent).

Recent research on income inequality emphasizes the roles of measurement error in CPS-ASEC implicit hourly wage rates (Lemieux 2006) and the effect of top-coded incomes on the top 1% of the distribution (Piketty and Saez 2003; Burkhauser et al. 2012). Except for Piketty and Saez who use IRS tax return data, the latter papers (and indeed most of the inequality literature) rely on the CPS-ASEC (or the Outgoing Rotation Group) for their analyses. Lemieux (2006) and Autor et al. (2008) eliminate imputed earnings from their analyses, but do not address the broader issue of earnings non-response. Piketty and Saez (2003) find that growth in wage income at the top is fueling the growth in overall incomes among the upper 1% in tax return data, and Burkhauser et al. (2012) largely confirm this finding in the CPS-ASEC using the internal files at Census (but not matched to DER records) and adjusting for top-coding.

This paper differs from the previous literature on poverty and inequality in several ways. First, the analysis assembles a data set matched to administrative records covering a long time period, 1997-2008. Second, the analysis examines trends in non-response and imputation and their impact on poverty rates and inequality measures. This paper is the first to examine both poverty and inequality using administrative data. Third, while previous analyses study different components of income, this analysis focuses on just earnings imputation since earnings account for over 80 percent of income.

3. The Current Population Survey Hot Deck Imputation Procedure

The Census Bureau has used a hot deck procedure for imputing missing income since 1962. The current system has been in place with few changes since 1989 (Welniak 1990). The CPS ASEC uses a variation of the cell hot deck procedure to impute missing income and earnings data. The cell hot deck procedure assigns individuals with missing income values that come from individuals with similar characteristics. The hot deck procedure for the CPS ASEC earnings variables relies on a sequential match procedure. First, individuals with missing data are divided into one of 12 allocation groups defined by the pattern of non-response. Welniak (1990) lists the 12 allocation groups and non-response patterns. Examples include a group that is only missing earnings from longest job or a group that is missing both longest job and earnings from longest job. Second, an observation in each allocation group is matched to another observation with complete data based on a large set of socioeconomic variables, the match variables.[2] If no match is found based on the large set of match variables, then a match variable is dropped and variable definitions are collapsed to be less restrictive. This process of sequentially dropping a variable and collapsing variable definitions is repeated until a match is found. When a match is found, the missing income amount is substituted with the reported income amount from the first available matched record. The missing income amount does not come from an average of the available matched records.

For example, suppose the set of match variables consists of gender, race, education, age, and region where education is defined by less than high school, high school, some college, and college or more. If no match is found using this set of match variables, then the race variable could be dropped and education could be redefined by collapsing education categories to high school or less, some college, and college or more. If no match exists, then region could be dropped to obtain a match. This process of dropping and redefining match variables continues until the only match variable remaining is gender. This sequential match procedure always ensures a match.

[2] The set of match variables includes gender, race, age, relationship to householder, years of school completed, marital status, presence of children, labor force status of spouse, weeks worked, hours worked, occupation, class of worker, other earnings receipt, type of residence, region, transfer payments receipt, and person status.

4. Data

The data used for the analysis come from the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) for survey years 1998-2009 (reporting income for 1997-2008). The Census internal CPS ASEC is matched to the Social Security Administration's Detailed Earnings Record (DER) file. The Detailed Earnings Record file is an extract of Social Security Administration's Master Earning File (MEF) and includes data on total earnings, including wages and salaries and income from self-employment subject to Federal Insurance Contributions Act (FICA) and/or Self-Employment Contributions Act (SECA) taxation. Since individuals do not make SECA contributions if they lose money in self-employment, only positive self-employment earnings are reported in the DER file (Nicholas and Wiseman 2009). The DER file contains all earnings reported on a worker's W-2 forms. Figure 2 provides a sample W-2 form with the circled boxes we use in the analysis. These earnings are not capped at the FICA contribution amounts and include earnings not covered by Old Age Survivor's Disability Insurance (OASDI) but subject to Medicare tax. The DER earnings are also not capped by Census as are ASEC earnings, thus mitigating top code issues that plague inequality analyses. The DER file also contains deferred wages such as contributions to 401(k), 403(b), 408(k), 457(b), 501(c), and HSA plans. The DER file is not a comprehensive source of gross compensation. Abowd and Stinson (forthcoming) describe parts of gross compensation that may not appear in the DER file such as pre-tax health insurance premiums and education benefits. Workers in the DER file are uniquely identified by a Protected Identification Key (PIK) assigned by the Census Bureau. The PIK is a confidentiality-protected version of the Social Security Number.

The Census Bureau's Center for Administrative Records Research and Applications (CARRA) matches the DER file to the CPS ASEC. Since the CPS does not currently ask respondents for a Social Security Number, CARRA uses its own record linkage software system, the Person Validation System, to assign a Social Security Number.[3] This assignment relies on a probabilistic matching model based on name, address, date of birth, and gender (NORC 2011). The Social Security Number is then converted to a Protected Identification Key. The Social Security Number from the DER file received from SSA is also converted to a Protected Identification Key. The CPS ASEC and DER files are matched based on the Protected Identification Key and do not contain the Social Security Number.

5. Analysis

A worker can appear multiple times per year in the DER file if they have several jobs. The DER file is collapsed into one earnings observation per worker per year by aggregating total compensation (Box 1 of W-2), SSA covered self-employment earnings (SEI-FICA), and Medicare covered self-employment earnings (SEI-MEDICARE) across all employers. DER earnings are defined as the sum of total compensation plus the maximum of SSA covered self-employment income or Medicare covered self-employment:

$$\text{DER Earnings} = (\text{Box 1 of W-2}) + \max(\text{SEI-FICA}, \text{SEI-MEDICARE})$$

In this way DER Earnings is most compatible with the CPS earnings. CPS earnings (PEARNVAL) cover earnings from all wage and salary jobs (WSAL-VAL), business self-employment (SEMP-VAL), and farm self-employment (FRSE-VAL). The CPS total personal income variable (PTOTVAL) used to determine poverty consists of adding a person's total earnings (PEARNVAL) to a person's total other income (POTHVAL):

$$\text{PTOTVAL}=\text{PEARNVAL}+\text{POTHVAL}$$

[3] The final year the CPS collected respondent Social Security Number is CPS survey year 2005 (calendar year 2004). Beginning with survey year 2006 (calendar year 2005), all respondents were assigned a Social Security Number using the Person Validation System.

The analysis calculates poverty and inequality measures, considering four alternatives of handling earnings observations. In constructing the four alternatives, our approach relies on information from both administrative and survey data. We do not treat the administrative as the "truth," recognizing the advantage of survey data which can collect income not reported to employers, income from tips, and "under the table" income. The first two alternatives use unmatched internal CPS ASEC data and differ by either retaining all observations or dropping imputed earners. The last two alternatives use matched internal CPS-DER data and replace the portion of total personal income due to earnings (PEARNVAL) with DER Earnings while keeping income from other sources the same (no change in POTHVAL). These alternatives use information from both data sources. Replacing earnings income differs by imputation status and the availability of DER Earnings. An individual's imputation status is determined by having either wages and salary from longest job imputed (I-ERNVAL) or wages and salary from other jobs imputed (I-WSVAL). The four alternatives are listed below:

1. Retain all observations (Internal CPS)

2. Drop imputed earners (Internal CPS-No Imputed Earners)

3. Replace CPS Earnings (PEARNVAL) with DER Earnings **for all persons with a DER match regardless of imputation status** and use CPS Earnings for persons without a DER match (CPS-DER Match Method 1)

4. Replace CPS Earnings (PERNVAL) with DER Earnings **for ONLY those persons with imputed earnings and a DER match** and use CPS Earnings for persons without a DER match. Use imputed earnings for persons with no CPS Earnings and no DER match (CPS-DER Match Method 2)

We compare poverty and inequality measures computed under each alternative. We consider four standard measures of family inequality: Gini coefficient, the 90/10 percentile ratio, the 90/50 percentile ratio, and the 50/10 percentile ratio.

6. Results

Table 1 shows the results of matching internal CPS ASEC and DER files for CPS survey years 1998-2009.[4] The table displays the person count based on the CPS ASEC person file, the number of earners, the number of matched records, and the match rate among earners. The match rate is defined as the number of earners matched to a DER record divided by the total number of earners. The match rates range from 66 percent to 85 percent. The table also shows the imputation rate among earners. The rate of imputed earnings begins at 16 percent for 1998, rises to 21 percent for 2003-2005, and falls to 19 percent for 2009. The lower panel of the table shows how match rates differ by imputation status among earners. Individuals with no imputed earnings are more likely to have a matched DER record. All counts and rates are unweighted. Figure 3 plots the overall match rate for earners and the match rate for earners by imputation status.

Table 2 shows the effect of imputation and replacing CPS earnings with DER Earnings on the official poverty rate. Poverty rates are weighted using the March supplement person weight. Column 1 shows the official poverty rate over the time period while column 3 shows the official poverty rate after dropping individuals with imputed earnings. This comparison gives a sense of the bias introduced by the imputation process. Columns 5 and 6 give the difference from the official poverty rate and a test for equality to the official poverty rate at the 10 percent

[4] The matched data for CPS survey year 2001 do not include the SCHIP sample expansion. Matched data for survey years after 2001 include the SCHIP sample expansion.

level of significance. Excluding imputed earners from the poverty calculation raises the poverty rate across all years by an average of 0.7 percentage points (Internal CPS-No Imputed Earners). This translates into 2-3 million additional persons in poverty in an average year. Column 7 shows the poverty rate after replacing CPS earnings with DER earnings for all persons regardless of imputation status (CPS-DER Match Method 1). Comparing this poverty series to the official poverty series for all years excluding 1999-2000 and 2008 still raises the rate but by a smaller amount (average of 0.3 percentage points).

Column 11 shows the poverty rate after replacing CPS earnings with DER earnings for only those persons with imputed earnings (CPS-DER Match Method 2). Again, the poverty series is higher than the official poverty series, but only for 2001-2008, by an average of 0.3 percentage points. The earlier years, 1997-2000, are not statistically different at the 10 percent level of significance. Figure 4 plots each series and shows the effects of dropping imputed earners and replacing CPS earnings with DER earnings by each CPS-DER Match Method. Figure 4 illustrates how substituting DER Earnings produces a poverty rate series that falls between the Internal CPS poverty rate and Internal CPS-No Imputed Earners poverty rate. Figure 4 clearly shows excluding imputed earners produces the highest poverty series, suggesting higher earners are more likely to not respond and require imputation.

Tables 3-6 repeat the analysis but for various subgroups of the population: age (Table 3), race (Table 4), gender (Table 5), and education (Table 6). Future versions of the paper will include standard errors and statistical testing of comparisons. Child poverty rates under each alternative for 1997-2000 are closely aligned but begin to diverge after 2000. Using DER Earnings for all persons (CPS-DER Match Method 1) produces the highest child poverty rate for 2000-2006. Imputations have the strongest effect for adults and seniors. Dropping imputed earners produces the highest poverty rate for these two groups. For seniors, using DER Earnings for all persons produces the lowest poverty rate. Dropping imputed earners also produces the highest poverty rate for Whites and Blacks (Table 4). Poverty rates for men and women after dropping imputed earners exceed the official rate by an average of 0.8 and 0.7 percentage points, respectively (Table 5). Poverty rates for individuals with less than a high school education or a high education also exceed the official rate by an average of 0.70 and 0.73 percentage points, respectively.

Figures 5-8 show the analysis for four standard measures of inequality for families: Gini coefficient (Figure 5), the 90/10 percentile ratio (Figure 6), the 90/50 percentile ratio (Figure 7), and the 50/10 percentile ratio (Figure 8). Future versions of the paper will include standard errors and statistical testing of comparisons. For consistency with the analysis of poverty, we restrict the sample to individuals in the poverty universe. Gini coefficients based on both Internal CPS alternatives are smaller than Gini coefficients based on both CPS-DER Match Methods. This is not surprising given the untopcoded DER Earnings data. The Gini coefficients based on the DER Earnings peak at the end of the dot-com bubble in 2000, showing the largest inequality during the sample time period. Generally, each percentile ratio measure exhibits a similar rank ordering among the alternative methods and similar trends in the time series. Dropping imputed earners produces the highest ratio series, followed by using DER Earnings for only imputed earners (CPS-DER Match Method 2), followed by only using the internal CPS. Replacing CPS earnings with DER Earnings all persons with a match (CPS-DER Match Method 1) produces the lowest series. Like the poverty series, excluding imputed earners produces the highest series, suggesting non-response may be more prevalent among higher earners.

7. Conclusion and Future Work

This paper uses a unique dataset of administrative earnings data matched to internal CPS ASEC to study the effects of earnings imputation on poverty and inequality measurement. Our analysis recalculates the official poverty rate and inequality statistics based on different assumptions on the availability of DER earnings and imputation status of survey respondents. In this way, we allow for both sources of earnings to contribute to total income and do not take the survey response or administrative record as the "truth." Substituting administrative earnings data for earnings imputed in the CPS ASEC produces overall poverty rates that are higher than the official poverty rate but not as high as poverty rates produced from completely dropping imputed earners. Completely dropping imputed earners gives the highest poverty rates for adults, seniors, Whites, Blacks, men, women, and those with a high school education or less. Likewise, completely dropping imputed earners also gives the highest percentile ratio series (90/10, 90/50, and 50/10) while replacing CPS earnings with DER Earnings for only imputed earners produces the lowest series.

Future work will include an examination of whole ASEC supplement imputations on poverty and inequality. Whole ASEC supplement imputations are for individuals who refuse to respond to the CPS ASEC supplement and need the entire supplement, including all income questions, imputed. Over the last decade about 10 percent of ASEC supplements were imputed. Future work will also closely examine non-response and measurement error. To address non-response we will explore the possibility of estimating the probability of non-response and reweighting the respondent sample by adjusting the survey weights. If non-response is non-random with respect to observables, (i.e., non-ignorable), we will explore an alternative approach and estimate earnings regressions for the respondents with sample selection corrections for non-response. The resulting model could be used to predict earnings for non-respondents--conditional on non-response--which correct for any non-random non-response. The advantage of these approaches, provided the non-response mechanism remains stable across time, is that these methods can be implemented by users who do not have access to the DER data and can be carried forward regardless of the availability of future CPS-DER matches.

Roemer (2002), Kapteyn and Ypma (2007) and Abowd and Stinson (forthcoming) provide a modeling approach to address measurement error that relies on both administrative and survey data. These approaches recognize that treating the administrative data as the “truth” may miss the advantage of survey data which, unlike administrative data, can collect “under the table” earnings. We will modify the framework of these papers and model the income measures from each data source as both containing measurement error and represent some underlying true amount of earnings.

References

Abowd, John and Martha Stinson. Forthcoming. "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data." *Review of Economics and Statistics*.

Autor, David, Lawrence Katz, and Melissa Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *Review of Economics and Statistics* 90: 300-323.

Burkhauser, Richard V., Shuaizhang Feng, Stephen P. Jenkins, and Jeff Larrimore. 2012. "Recent Trends in Top Income Shares in the USA: Reconciling Estimates from March CPS and IRS Tax Return Data." *Review of Economics and Statistics* 94(2), pp. 371-388.

Bollinger, Christopher and Barry Hirsch. 2006. "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching." *Journal of Labor Economics*, 24(3), pp. 483-519.

Chesher, Andrew and Christian Schluter. 2002. "Welfare Measurement and Measurement Error." *Review of Economic Studies*, 69, pp. 357-378.

Groves, Robert. 2001. Survey Nonresponse. New York: Wiley.

Hirsch, Barry and Edward Schumacher. 2004. "Match Bias in Wage Gap Estimates Due to Earnings Imputation." *Journal of Labor Economics*, 22(3), pp. 689-722.

Kapteyn, Arie and Jelmer Ypma. 2007. "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data." *Journal of Labor Economics* 25:513-551.

Lemieux, Thomas. 2006. "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *American Economic Review* 96: 461–498.

Little, Roderick and Donald Rubin. 2002. Statistical Analysis With Missing Data. New York: Wiley.

Nicholas, Joyce and Michael Wiseman. 2009. "Elderly Poverty and Supplemental Security Income" *Social Security Bulletin*, 69(1), pp. 45-73.

NORC at the University of Chicago. 2011. "Assessment of the US Census Bureau's Person Identification Validation System." Final Report presented to the US Census Bureau.

Piketty, Thomas and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118: 1–39.

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, US Census Bureau.

Turek, Joan, Kendall Swenson, Bula Ghose, Fritz Scheuren, and Daniel Lee. 2012. “How Good Are ASEC Earnings Data? A Comparison to SSA Detailed Earnings Records” Presented at 2012 Federal Committee On Research Methodology Research Conference.

Welniak, Edward J. 1990. “Effects of the March Current Population Survey’s New Processing System on Estimates of Income and Poverty.” US Census Bureau, Washington, DC, 1990.

Ziliak, James P. 2006. “Understanding Poverty Rates and Gaps: Concepts, Trends, and Challenges.” *Foundations and Trends in Microeconomics* 1: 127–199.

Table 1: Match Rate and Imputation Rate

Calandar Year	Person Record Count	Total Earners	Total Matched Records	Match Rate (Earners)	Imputation Rate (Earners)
1997	131,617	69,573	53,005	71%	16%
1998	132,324	70,218	49,474	66%	18%
1999	133,710	71,783	50,661	66%	17%
2000	218,269	69,040	51,311	68%	20%
2001	217,219	113,577	89,543	73%	20%
2002	216,424	111,698	84,692	70%	21%
2003	213,241	109,672	74,585	62%	21%
2004	210,648	108,120	71,632	61%	21%
2005	208,562	107,532	100,013	85%	19%
2006	206,639	106,738	99,633	85%	20%
2007	206,404	107,038	99,217	84%	20%
2008	207,921	107,134	98,764	84%	19%

Match Rate by Imputation Status

Calendar Year	Matched Earner **and Imputed**	Total Imputed Earners	Match Rate (Imputed Earners)	Matched Earner **and Not Imputed**	Total Earners Not Imputed	Match Rate (Not Imputed Earners)
1997	6,111	11,329	54%	43,193	58,244	74%
1998	5,873	12,363	48%	40,173	57,855	69%
1999	6,079	12,492	49%	41,071	59,291	69%
2000	7,255	13,771	53%	39,916	55,269	72%
2001	12,983	22,534	58%	69,388	91,043	76%
2002	12,510	23,097	54%	65,235	88,601	74%
2003	10,340	22,649	46%	58,111	87,023	67%
2004	10,057	22,296	45%	55,812	85,824	65%
2005	15,631	20,016	78%	75,786	87,516	87%
2006	16,145	20,853	77%	74,524	85,885	87%
2007	16,201	21,174	77%	74,100	85,864	86%
2008	15,086	20,014	75%	74,981	87,120	86%

Sources: U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <www.census.gov/apsd/techdoc/cps/cpsmar12.pdf>.

Social Security Administration, Detailed Earnings Record, 1997-2008.

Table 2: Poverty Rates Based on Alternative Treatment of Imputed Earnings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Calendar Year	Internal CPS (Official Rate)	Std. Error	Internal CPS-No Imputed Earners	Std. Error	Difference from Official Rate	Test for Equality	CPS-DER Match (Method 1)	Std. Error	Difference from Official Rate	Test for Equality	CPS-DER Match (Method 2)	Std. Error	Difference from Official Rate	Test for Equality
1997	13.3	0.211	13.8	0.225	0.6	*	13.4	0.212	0.2	*	13.4	0.212	0.1	
1998	12.7	0.206	13.3	0.222	0.6	*	12.9	0.208	0.2	*	12.9	0.207	0.2	
1999	11.9	0.199	12.4	0.213	0.5	*	11.9	0.199	0.1		11.9	0.199	0.1	
2000	11.3	0.193	12.0	0.211	0.7	*	11.3	0.194	0.0		11.4	0.194	0.1	
2001	11.7	0.139	12.5	0.152	0.8	*	12.1	0.141	0.4	*	12.0	0.141	0.3	*
2002	12.1	0.140	12.9	0.154	0.8	*	12.4	0.142	0.3	*	12.4	0.142	0.3	*
2003	12.5	0.142	13.3	0.154	0.8	*	12.7	0.143	0.3	*	12.7	0.143	0.2	*
2004	12.7	0.142	13.6	0.155	0.8	*	13.0	0.143	0.3	*	12.9	0.143	0.2	*
2005	12.6	0.141	13.3	0.152	0.7	*	13.0	0.143	0.4	*	12.9	0.142	0.3	*
2006	12.3	0.139	13.1	0.151	0.8	*	12.7	0.141	0.4	*	12.7	0.141	0.5	*
2007	12.5	0.139	13.3	0.152	0.9	*	13.0	0.141	0.5	*	13.0	0.141	0.5	*
2008	13.2	0.142	14.0	0.154	0.8	*	13.2	0.142	0.0		13.6	0.143	0.3	*

Standard errors are estimated using generalized variance parameters. *p<0.10

Sources: U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <www.census.gov/apsd/techdoc/cps/cpsmar12.pdf>.

Social Security Administration, Detailed Earnings Record, 1997-2008.

Table 3: Poverty Rate By Age Based on Alternative Treatment of Imputed Earnings

Children (<18)

Calandar Year	Internal CPS	Internal CPS-No Imputed Earners	CPS-DER Match (Method 1)	CPS-DER Match (Method 2)
1997	19.9	19.9	20.1	20.0
1998	18.9	19.0	19.3	19.0
1999	17.1	17.2	17.3	17.2
2000	16.2	16.3	16.2	16.2
2001	16.3	16.4	17.1	16.7
2002	16.7	16.8	17.2	17.1
2003	17.6	17.7	18.0	17.9
2004	17.8	17.9	18.4	18.0
2005	17.6	17.7	18.5	18.0
2006	17.4	17.5	18.1	17.9
2007	18.0	18.1	19.0	18.8
2008	19.0	19.1	19.1	19.5

Adults (18-64)

Calandar Year	Internal CPS	Internal CPS-No Imputed Earners	CPS-DER Match (Method 1)	CPS-DER Match (Method 2)
1997	10.9	11.5	11.2	11.1
1998	10.5	11.1	10.7	10.7
1999	10.1	10.6	10.2	10.2
2000	9.5	10.2	9.6	9.6
2001	10.1	11.0	10.4	10.3
2002	10.6	11.5	10.8	10.8
2003	10.8	11.7	11.2	11.0
2004	11.3	12.2	11.5	11.4
2005	11.1	11.9	11.5	11.3
2006	10.8	11.9	11.3	11.3
2007	10.9	11.9	11.3	11.4
2008	11.7	12.6	11.7	12.0

Seniors (65+)

Calandar Year	Internal CPS	Internal CPS-No Imputed Earners	CPS-DER Match (Method 1)	CPS-DER Match (Method 2)
1997	10.5	10.8	10.1	10.5
1998	10.5	10.7	10.1	10.5
1999	9.7	10.0	9.4	9.6
2000	10.2	10.6	9.9	10.2
2001	10.1	10.5	9.7	10.2
2002	10.4	10.8	10.2	10.5
2003	10.2	10.6	9.9	10.2
2004	9.8	10.2	9.5	9.9
2005	10.1	10.5	9.6	10.2
2006	9.4	9.8	9.0	9.4
2007	9.7	10.1	9.1	9.7
2008	9.7	10.1	9.1	9.7

Note: CPS-DER Match (Method 1) uses DER earnings for all persons. CPS-DER Match (Method 2) uses DER earnings for persons with imputed earnings. Sources: U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <www.census.gov/apsd/techdoc/cps/cpsmar12.pdf>.
Social Security Administration, Detailed Earnings Record, 1997-2008.

Table 4: Poverty Rate By Race Based on Alternative Treatment of Imputed Earnings

White

Calandar Year	Internal CPS	Internal CPS-No Imputed Earners	CPS-DER Match (Method 1)	CPS-DER Match (Method 2)
1997	11.0	11.4	11.4	11.1
1998	10.5	11.0	10.6	10.7
1999	9.8	10.2	9.9	9.8
2000	9.5	10.0	9.6	9.6
2001	9.9	10.5	10.3	10.1
2002	10.3	10.9	10.6	10.5
2003	10.6	11.2	10.9	10.7
2004	10.9	11.6	11.2	11.1
2005	10.7	11.2	11.2	10.9
2006	10.4	11.1	11.0	10.8
2007	10.6	11.3	11.2	11.0
2008	11.4	12.0	11.5	11.7

Black

Calandar Year	Internal CPS	Internal CPS-No Imputed Earners	CPS-DER Match (Method 1)	CPS-DER Match (Method 2)
1997	26.5	28.2	25.6	26.6
1998	26.1	27.6	26.5	26.3
1999	23.6	25.1	23.6	23.8
2000	22.1	23.9	21.8	21.9
2001	22.7	24.4	22.8	23.2
2002	24.0	25.8	23.5	24.2
2003	24.3	26.3	23.9	24.6
2004	24.7	26.5	24.8	25.1
2005	24.8	26.6	24.6	25.0
2006	24.2	26.1	23.8	24.8
2007	24.5	26.7	24.5	25.6
2008	24.6	26.5	23.8	25.0

Other

Calandar Year	Internal CPS	Internal CPS-No Imputed Earners	CPS-DER Match (Method 1)	CPS-DER Match (Method 2)
1997	16.1	16.9	16.3	16.2
1998	14.5	15.2	15.7	14.6
1999	14.5	15.2	14.8	15.1
2000	13.7	14.6	13.6	14.0
2001	12.8	13.4	13.2	13.1
2002	12.2	12.9	12.7	12.7
2003	13.5	14.5	13.8	14.1
2004	12.0	12.9	12.4	12.4
2005	13.0	13.8	13.2	13.2
2006	12.8	13.9	13.1	13.6
2007	12.1	13.0	12.7	12.7
2008	14.3	15.3	13.6	14.3

Note: CPS-DER Match (Method 1) uses DER earnings for all persons. CPS-DER Match (Method 2) uses DER earnings for persons with imputed earnings. Sources: U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <www.census.gov/apsd/techdoc/cps/cpsmar12.pdf>.
Social Security Administration, Detailed Earnings Record, 1997-2008.

Table 5: Poverty Rate By Gender Based on Alternative Treatment of Imputed Earnings

Men

Calandar Year	Internal CPS	Internal CPS-No Imputed Earners	CPS-DER Match (Method 1)	CPS-DER Match (Method 2)
1997	11.6	12.2	11.9	11.7
1998	11.1	11.7	11.4	11.3
1999	10.4	11.0	10.6	10.5
2000	10.0	10.7	10.2	10.1
2001	10.4	11.2	10.9	10.7
2002	10.9	11.7	11.2	11.1
2003	11.2	12.0	11.6	11.4
2004	11.5	12.3	11.9	11.7
2005	11.1	11.8	11.7	11.4
2006	11.0	11.9	11.6	11.5
2007	11.1	12.0	11.7	11.7
2008	12.0	12.8	12.2	12.3

Women

Calandar Year	Internal CPS	Internal CPS-No Imputed Earners	CPS-DER Match (Method 1)	CPS-DER Match (Method 2)
1997	14.9	15.4	14.9	14.9
1998	14.3	14.8	14.3	14.4
1999	13.2	13.7	13.2	13.3
2000	12.6	13.2	12.4	12.6
2001	12.9	13.6	13.2	13.2
2002	13.3	14.0	13.5	13.6
2003	13.7	14.4	13.9	13.9
2004	13.9	14.7	14.1	14.1
2005	14.1	14.7	14.3	14.3
2006	13.6	14.3	13.8	14.0
2007	13.8	14.6	14.2	14.3
2008	14.4	15.2	14.2	14.8

Note: CPS-DER Match (Method 1) uses DER earnings for all persons. CPS-DER Match (Method 2) uses DER earnings for persons with imputed earnings. Sources: U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <www.census.gov/apsd/techdoc/cps/cpsmar12.pdf>.
Social Security Administration, Detailed Earnings Record, 1997-2008.

Table 6: Poverty Rate By Education Based on Alternative Treatment of Imputed Earnings

Less than High School

Calandar Year	Internal CPS	Internal CPS-No Imputed Earners	CPS-DER Match (Method 1)	CPS-DER Match (Method 2)
1997	23.2	23.7	22.2	23.2
1998	21.8	22.5	21.3	22.1
1999	21.1	21.5	20.4	21.2
2000	20.1	20.8	19.4	20.1
2001	20.4	21.3	20.2	20.8
2002	21.0	21.7	20.8	21.3
2003	21.3	22.1	20.9	21.6
2004	21.8	22.7	21.3	21.9
2005	21.6	22.3	21.5	21.9
2006	20.9	21.7	20.4	21.4
2007	22.4	23.1	21.6	22.9
2008	23.5	24.1	22.0	23.9

High School

Calandar Year	Internal CPS	Internal CPS-No Imputed Earners	CPS-DER Match (Method 1)	CPS-DER Match (Method 2)
1997	10.8	11.2	11.2	11.0
1998	11.0	11.5	11.0	11.1
1999	10.3	10.8	10.6	10.5
2000	9.9	10.5	10.1	10.0
2001	10.6	11.3	10.7	10.8
2002	11.1	12.0	11.3	11.3
2003	11.3	12.1	11.7	11.5
2004	11.9	12.7	12.1	12.0
2005	11.9	12.6	12.3	12.1
2006	11.7	12.7	12.3	12.2
2007	12.0	12.9	12.3	12.4
2008	12.6	13.5	12.8	13.0

Some College

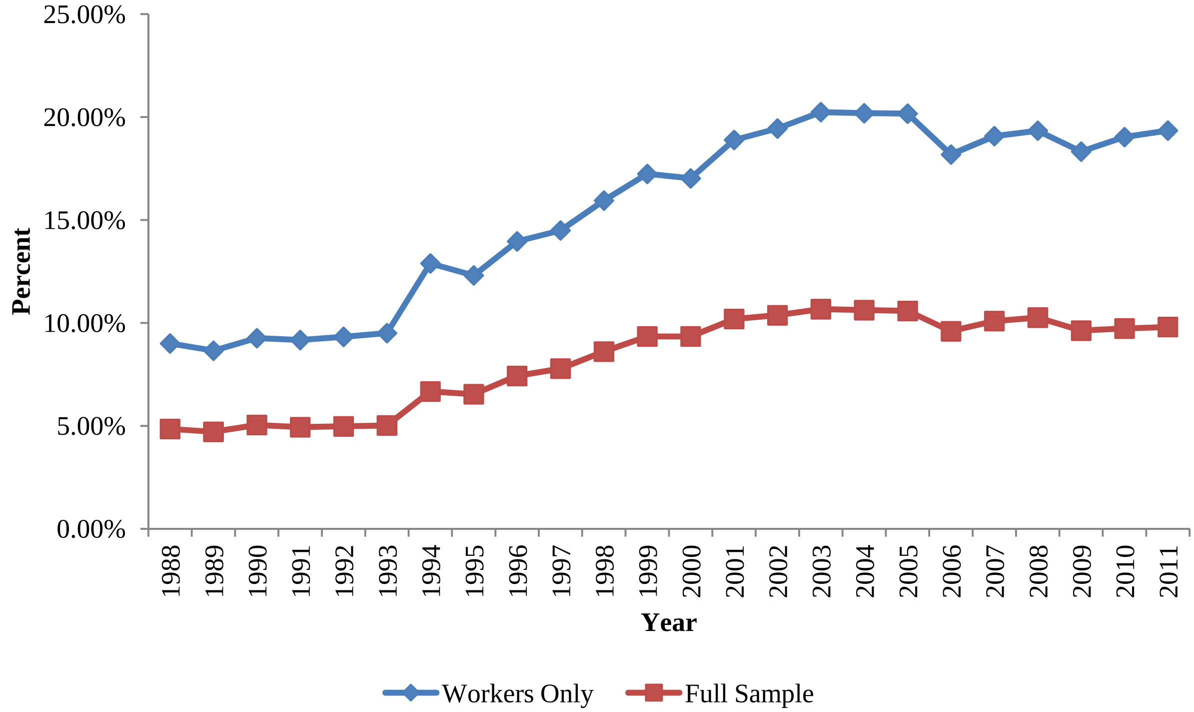
Calandar Year	Internal CPS	Internal CPS-No Imputed Earners	CPS-DER Match (Method 1)	CPS-DER Match (Method 2)
1997	8.1	8.6	8.8	8.4
1998	8.2	8.4	8.8	8.3
1999	7.9	8.1	8.1	8.0
2000	7.6	8.0	7.7	7.6
2001	8.4	9.0	8.8	8.6
2002	8.9	9.4	9.3	9.2
2003	9.3	10.0	9.9	9.7
2004	9.6	10.2	10.1	9.8
2005	9.7	10.2	9.8	9.8
2006	9.5	10.2	10.3	9.9
2007	9.1	9.7	10.2	9.7
2008	10.5	11.1	11.0	10.9

College Plus

Calandar Year	Internal CPS	Internal CPS-No Imputed Earners	CPS-DER Match (Method 1)	CPS-DER Match (Method 2)
1997	4.0	4.0	4.3	4.0
1998	3.8	3.9	4.1	3.9
1999	3.7	3.9	3.8	3.7
2000	3.8	4.1	4.1	3.9
2001	4.1	4.4	4.5	4.2
2002	4.4	4.8	4.6	4.6
2003	4.8	5.1	5.1	4.9
2004	4.7	5.0	5.0	4.9
2005	4.7	4.9	5.1	4.9
2006	4.6	5.0	4.9	4.9
2007	4.4	4.8	4.8	4.7
2008	4.7	5.1	4.9	4.9

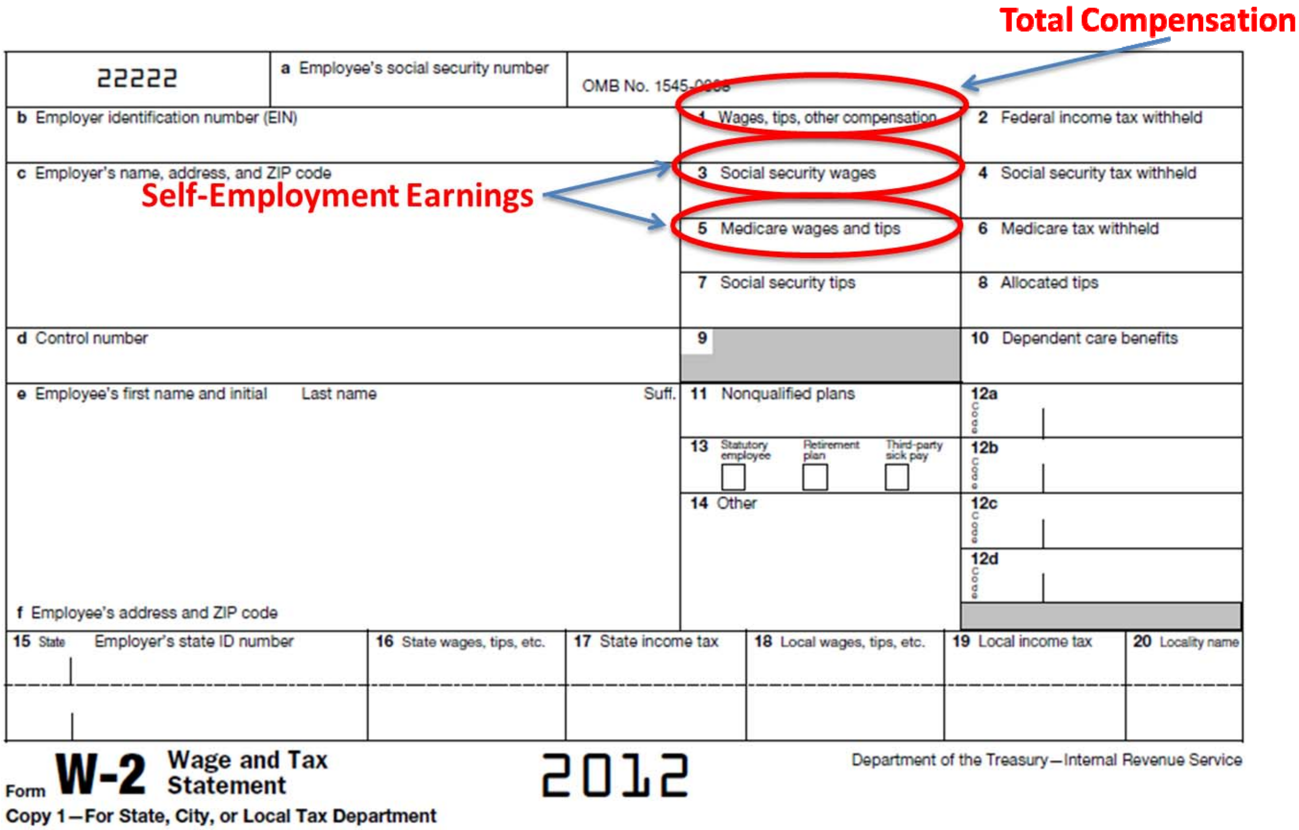
Note: CPS-DER Match (Method 1) uses DER earnings for all persons. CPS-DER Match (Method 2) uses DER earnings for persons with imputed earnings. Sources: U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <www.census.gov/apsd/techdoc/cps/cpsmar12.pdf>.
Social Security Administration, Detailed Earnings Record, 1997-2008.

Figure 1: Imputation Rates for Earnings in the CPS-ASEC



Sources: U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <www.census.gov/apsd/techdoc/cps/cpsmar12.pdf>.

Figure 2: Sample W-2 Form


Total Compensation

Self-Employment Earnings

22222

a Employee's social security number

OMB No. 1545-0008

b Employer identification number (EIN)

1 Wages, tips, other compensation

2 Federal income tax withheld

c Employer's name, address, and ZIP code

3 Social security wages

4 Social security tax withheld

5 Medicare wages and tips

6 Medicare tax withheld

7 Social security tips

8 Allocated tips

d Control number

9

10 Dependent care benefits

e Employee's first name and initial Last name Suff.

11 Nonqualified plans

12a Code

13 Statutory employee Retirement plan Third-party sick pay

12b Code

14 Other

12c Code

12d Code

f Employee's address and ZIP code

15 State Employer's state ID number

16 State wages, tips, etc.

17 State income tax

18 Local wages, tips, etc.

19 Local income tax

20 Locality name

Form **W-2** **Wage and Tax Statement**

2012

Department of the Treasury—Internal Revenue Service

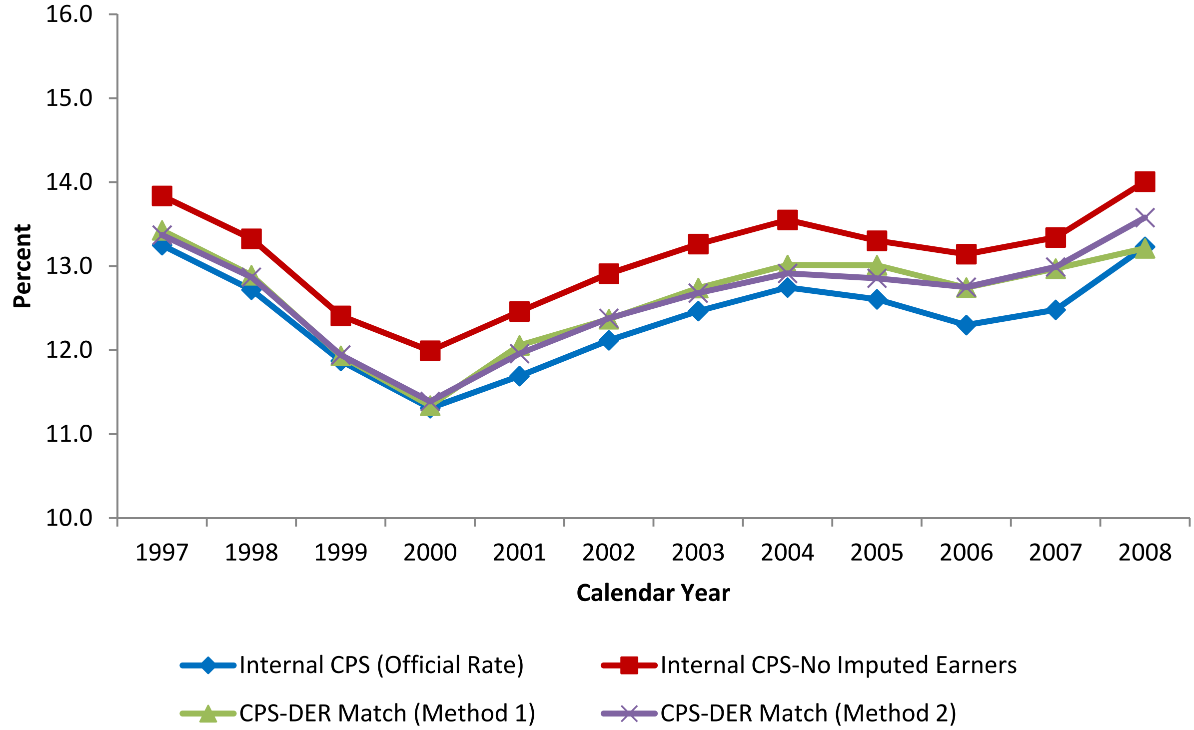
Copy 1—For State, City, or Local Tax Department

Figure 3: Match Rate By Imputation Status

Sources: U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <www.census.gov/apsd/techdoc/cps/cpsmar12.pdf>.

Social Security Administration, Detailed Earnings Record, 1997-2008.

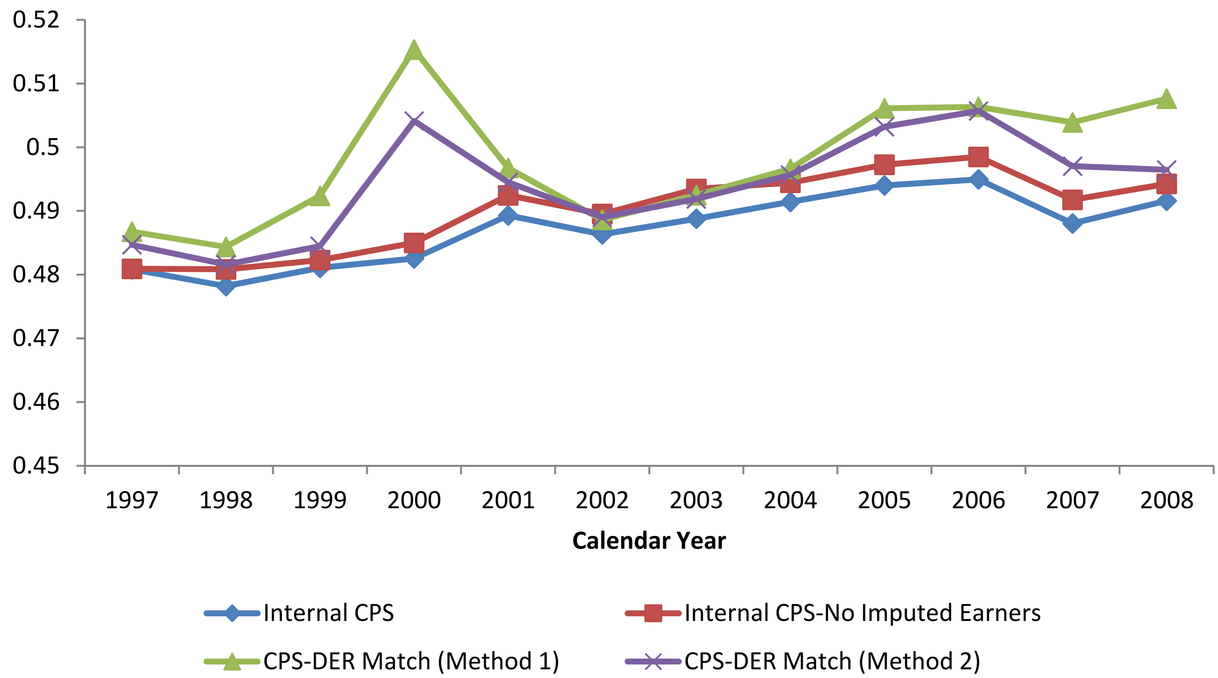
Figure 4: Trends in Poverty Rates Based on Alternative Treatment of Imputed Earnings



Sources: U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <www.census.gov/apsd/techdoc/cps/cpsmar12.pdf>.

Social Security Administration, Detailed Earnings Record, 1997-2008.

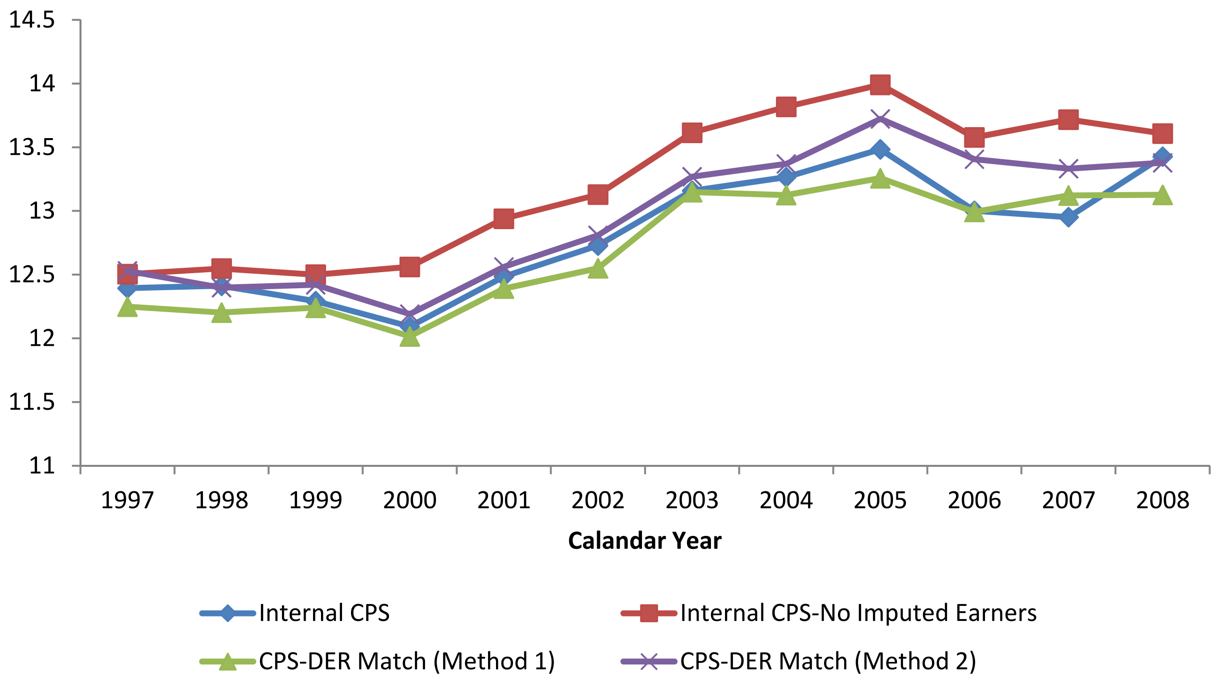
Figure 5: Trends in Family Income Gini Coefficients Based on Alternative Treatment of Imputed Earnings



Sources: U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <www.census.gov/apsd/techdoc/cps/cpsmar12.pdf>.

Social Security Administration, Detailed Earnings Record, 1997-2008.

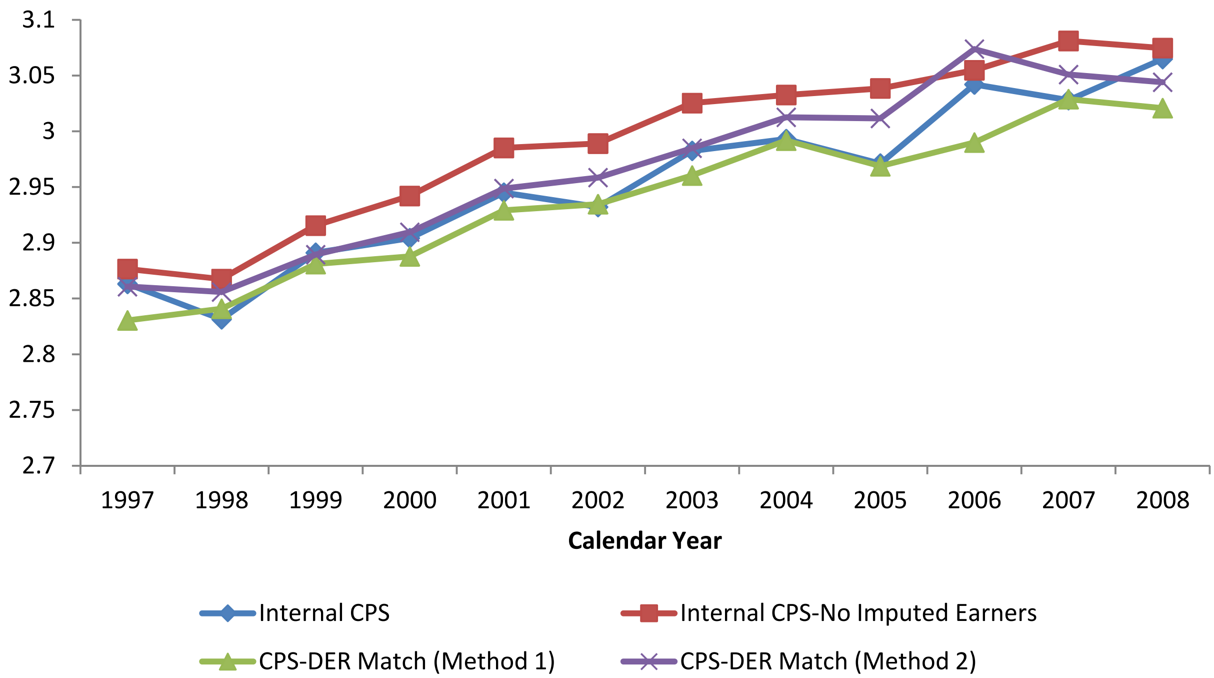
Figure 6: Trends in Family Income 90-10 Ratios Based on Alternative Treatment of Imputed Earnings



Sources: U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <www.census.gov/apsd/techdoc/cps/cpsmar12.pdf>.

Social Security Administration, Detailed Earnings Record, 1997-2008.

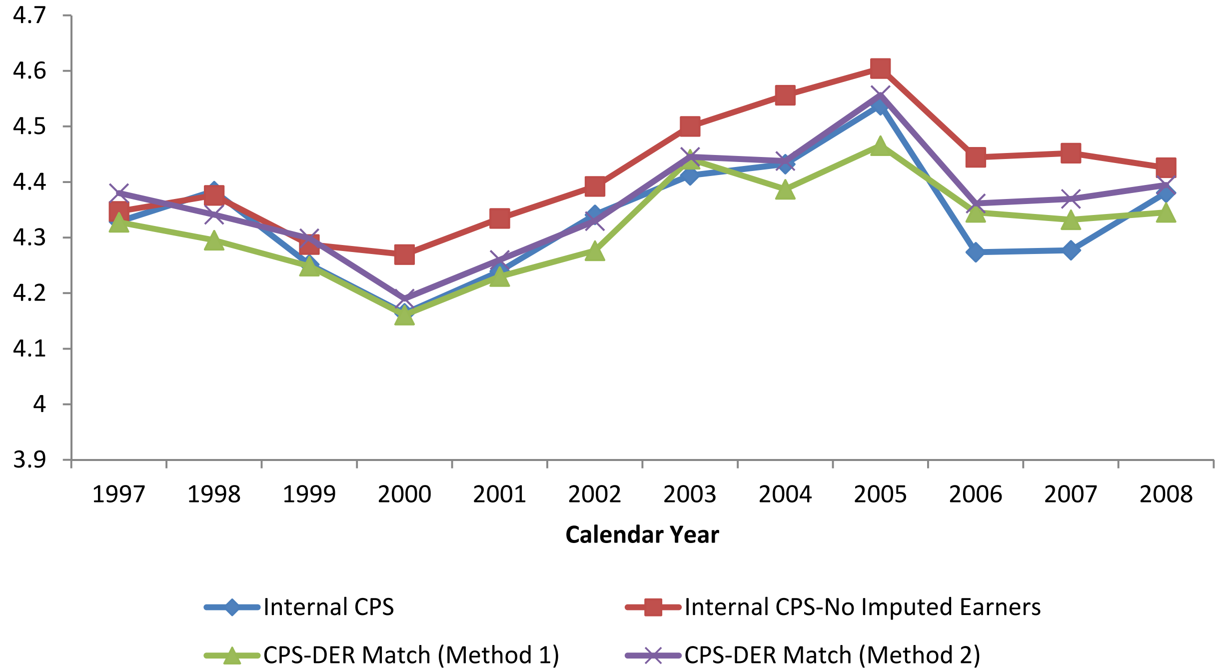
Figure 7: Trends in Family Income 90-50 Ratios Based on Alternative Treatment of Imputed Earnings



Sources: U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <www.census.gov/apsd/techdoc/cps/cpsmar12.pdf>.

Social Security Administration, Detailed Earnings Record, 1997-2008.

Figure 8: Trends in Family Income 50-10 Ratios Based on Alternative Treatment of Imputed Earnings



Sources: U.S. Census Bureau, Current Population Survey, 1998-2009 Annual Social and Economic Supplement. For information on sampling and nonsampling error, see <www.census.gov/apsd/techdoc/cps/cpsmar12.pdf>.

Social Security Administration, Detailed Earnings Record, 1997-2008.