How Good Are ASEC Earnings Data? A Comparison to SSA Detailed Earning Records[1]

Joan Turek, Kendall Swenson and Bula Ghose, Department of Health and Human Services
Fritz Scheuren and Daniel Lee, NORC University of Chicago

Abstract

On-going research has highlighted systematic differences over time in Current Population Survey (CPS), Annual Social and Demographic Supplement (ASEC), poverty rates among persons reporting all of their income items; those reporting some information on income but missing one or more amounts (item imputes); and, those who were imputed all ASEC supplement items including their income (whole imputes). [2] This research, however, cannot determine to what degree Census Bureau methods for imputing income bias reported poverty rates or, more generally, income estimates across the entire income distribution. These questions can be only addressed by matching survey data with administrative data which provide independent estimates of income[3]. With funding from the Office of the Assistant Secretary from Planning and Evaluation (ASPE), Department of Health and Human Services, the Census Bureau is matching ASEC respondents' information to the Social Security Administration's Detailed/Summary Earnings Records (DER). Amounts shown in these independent records will be compared to both reported and imputed ASEC responses. The match is conducted using 2005 income data, a year in which earnings accounted for over 82% of the total income reported on the ASEC.

Research

During the past four decades, many changes have occurred in our society and in the degree to which eligible ASEC respondents are willing to report their cash income, which is needed to construct income and poverty figures. Imputation is the most common method to address bias when respondents fail to answer any or all of the income questions on surveys[4].

How has the growing incidence of income imputation impacted income estimates and poverty rates? Income estimates are used to measure many critical phenomena including economic well-being, economic growth, the allocation of federal Funds[5], inequalities in the distribution of income,[6] and to make international comparisons. Imputation can change the allocation of income both among demographic groups and across the entire income distribution which could lead to misrepresenting many important phenomena if imputation deviates significantly from what the participants would have responded, if they had reported their incomes.

It also is important to understand how imputation affects poverty estimates, particularly given the central role that mitigating poverty has played in Federal income transfer and health programs. Is it possible that the growing non-response rate during this period may have altered historical trends or the characteristics of the poor? Could the growth in imputation also lead to misrepresenting the size, share, and composition of the poverty population if imputations by the Census Bureau significantly deviate from what the participants would have responded if they had reported their incomes?

This paper's primary focus is based on research from a joint study with the Census Bureau and the U.S. Department of Health and Human Services. The project is analyzing a dataset that consists of 2006 ASEC survey data (2005 annual income data) that has been merged with records from the Social Security Administration's 2005 Detailed Earnings Records (DER) file.[7] It examines the effects of substituting DER earnings records for ASEC reported earnings on income estimates and on the number of persons in poverty. Earnings are the sum of wages, salaries and farm and nonfarm self-employment and account for over 82% of the total income reported on the ASEC in income year 2005.

While we would like to compare ASEC income estimates to an independent source, the DER, for the entire period from 1977, such a task is not feasible. We set the stage for analysis of the matched calendar year 2005 data by showing trends in imputation rates for all persons with positive income from 1977 to the present.[8] Trends in imputation on total income and on poverty rates are presented. The trends in poverty rates are shown only from 1987 to the present because it is difficult to differentiate between the two types of imputes, item and whole, using the pre-1987 ASEC public use files. We distinguish between item and whole imputes in this analysis because our previous work suggests that poverty rates vary by imputation type and they may bias the data in different ways.[9] The population of interest for most of the analysis is all persons with positive income who are 15 and older and subgroups thereof.[10] Analysis is also conducted for selected demographic groups.

When looking at matched results for persons in poverty, those with negative income are not excluded. This differs from other tabulations presented in this report where persons with negative income are excluded, just as they were in our previous work.[11] Persons who are in poverty are identified using the information on poverty status on the ASEC. The portion of the poverty population that has no income, including dependents, is not included in our estimates. We are interested in the role of imputation and how it can be improved. Imputation is conducted at the person level. An equally important analysis would be to evaluate the overall effects of imputation on the entire poverty population. Those findings will be presented in a future paper.

ASEC Imputation

Missing data are always a factor to some degree on surveys. When missing data are accounted for through imputation or by some other means, usually there is an implicit assumption that data are missing at random after controlling for other variables. However, evidence indicates that missing ASEC income data may not be completely random. If these other variables are not properly accounted for, bias can result.[12]

The Census Bureau started regularly imputing for missing CPS income in 1962. Since then, the same basic strategy, “hot deck” imputation, has been employed, although the exact process employed has been revised several times[13]. With this procedure, non-respondents are assigned income amounts reported by respondents with similar characteristics. The process is conducted at the person level for each income source identified. A complex set of demographic, economic and social characteristics is used in identifying similar person-level respondents.[14] As discussed later, different types of missing income data are treated differently.

The current processing system was introduced in the 1988-89 period. With the previous processing system implemented in the mid-1970s, 11 income categories were reported on the public use ASEC. The new processing system expanded the editing and imputation process to reflect the substantial expansion in the number of income sources collected on the March 1979 questionnaire to permit imputation of all supplemental noninterviews from one source, and to retain all reported data during the nonrespondent/respondent match process.[15] Questionnaire revisions in March 1979 allowed respondents to report over 50 sources of income and the recording of 27 different amounts. With minor changes, this processing system is still being used.[16]

There are many statistical goals in imputation.[17] A particularly important one is determining the extent to which imputation does not create bias in ASEC survey estimates due to missing data. This goal is met to the degree that patterns of nonresponse are correctly identified and corrected for. In the ASEC there are two basic types of imputation for missing data:

- Item Imputes. Sample persons or other household members fail to respond to a specific question on the ASEC and data for that particular “item” is imputed. Responses to more than one item may be imputed.

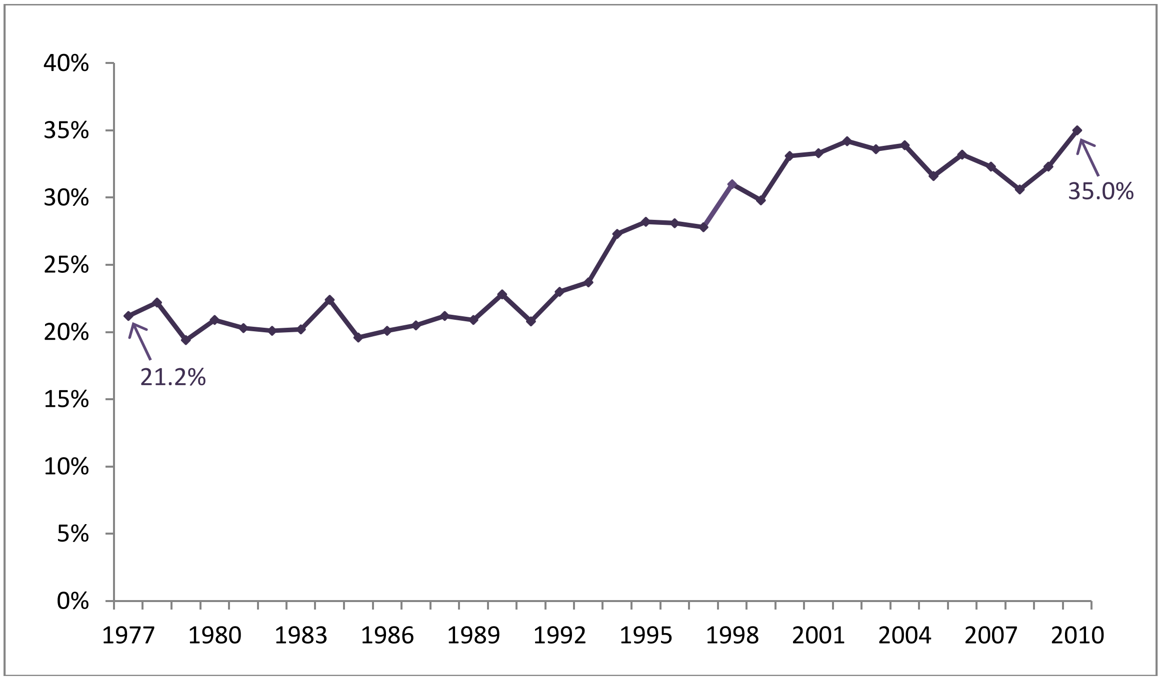
- Whole Imputes. Sample persons respond to the basic monthly survey questions but refuse to respond to the ASEC supplement. In this case, the “whole” or entire supplement has to be imputed.

Item imputes are based on responses to both the monthly survey and the ASEC supplement, while whole imputes are based only on the monthly survey.[18] This distinction is important because in the case of item imputes, many more variables are available to find a good match to impute any missing data.[19] For whole impute cases, where the whole supplement is missing, there are fewer variables to match on and thus we believe the chances are greater that the bias is increased from missing data.[20] In either case, after imputation a complete data set is created.

Trends in Income Imputation Rates for Persons with Positive Income

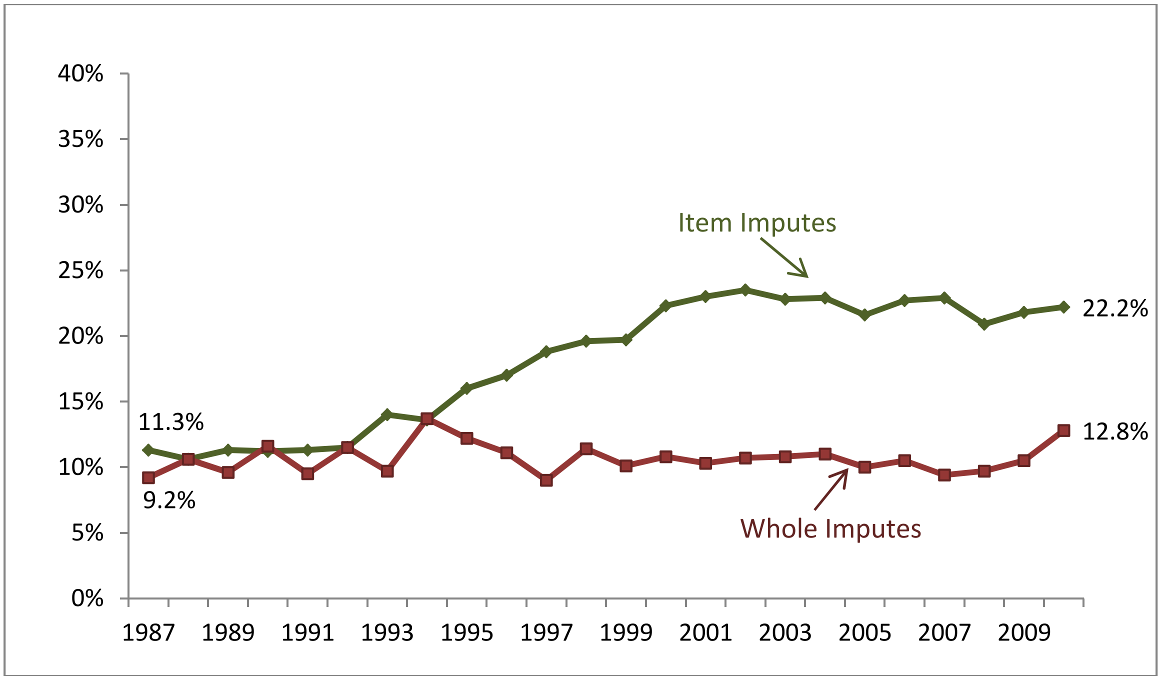
The imputation of income on the ASEC has increased substantially since the late 1970s. Between 1977 and 2010, the percent of dollar income received by persons with positive income that is imputed rose from 21.2% to 35.0 % as seen in Chart 1. This constitutes a 65% increase in the total amount of dollar income that is imputed. After remaining fairly stable in the 1980's, the amount of total income imputed increased rapidly during the 1990's. It has remained fairly stable since then reaching a high of 35.0% in 2010 after a short period of decline.

Chart 1 Percent of Total Dollar Income Imputed for Persons with Positive Income by Year



Trends for the two types of imputation are very different. Chart 2 shows trends in item and whole imputation for persons with positive income from 1987 to 2010. As can be seen, whole imputes have remained fairly stable over this period, with values of 9.2 % in income year 1987 and 12.8% in 2010. At their peak in 1994, whole imputes accounted for 13.7% of total dollar income. During this same time period, item imputes doubled from 11.3% to 22.2%. While persons with item imputes continued to respond to the ASEC, they were less willing to report answers to all of the income questions. Trends in income and poverty since 1987 appear to be more heavily influenced by changes in item imputes than whole imputes.

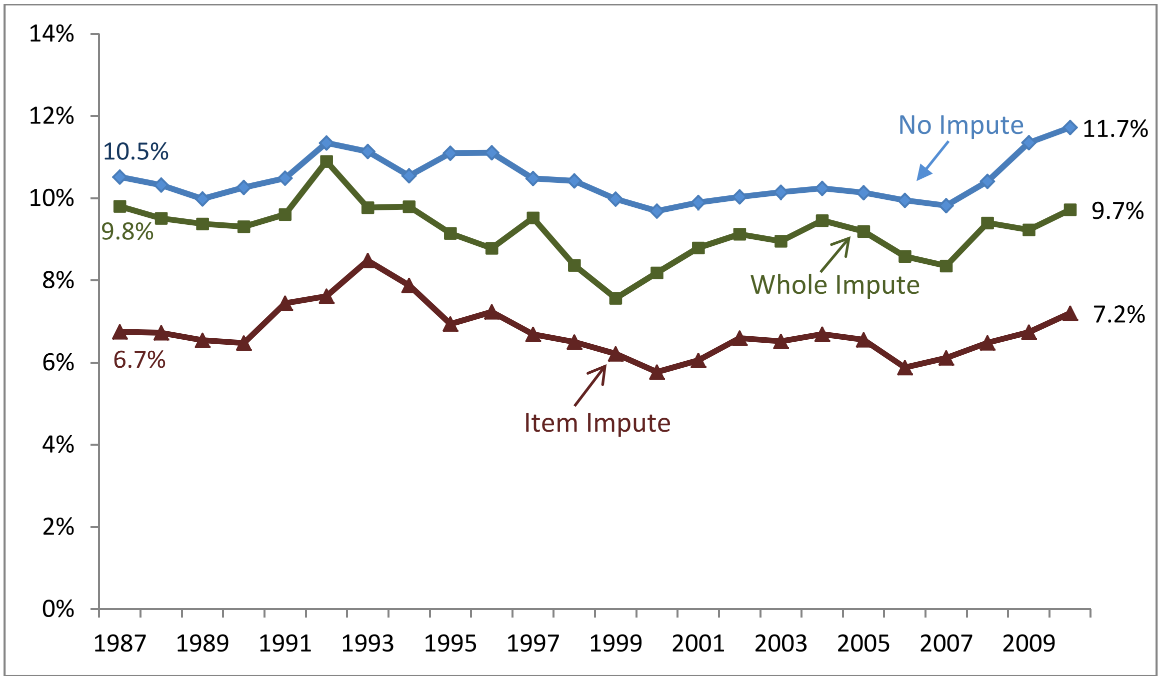
Chart 2 Percent of Total Dollar Income for Persons with Positive Income by Imputation Type and Year



Trends in Poverty Rates for Persons with Positive Income

Chart 3 shows trends in poverty rates separately for persons with positive income for item, whole or no imputes for the time period 1987-2010. As shown, the poverty rates trend lines for item, whole and no imputes are generally parallel over time, increasing and decreasing with similarity across the years.   However, poverty rates in any single year differ by imputation status. Persons with no imputes have the highest poverty rate, those with whole imputes are next and those with item imputes have the lowest poverty rate.

Chart 3 Percent of Persons with Positive Income that are Poor by Type of Imputation and Year



Including persons with whole imputes somewhat reduces the overall poverty rate for persons with positive income. This likely occurs because there is no information about the previous year's employment status of these persons and these persons are drawn from everyone in the imputation pool. That is, if full-year workers are more likely to need imputation, there is no way to reflect this in undertaking whole imputes. In contrast, poverty rates for persons with item imputes are much lower. Item imputes benefit from the availability of information reported by respondents on annual labor force participation and other activity such as presence of in-kind benefits, health insurance and country of birth.

Table 1 below estimates poverty rates for persons with positive income in 2005 using differing scenarios. First, the poverty rate is presented for all persons with positive income. Poverty rates are then provided for three separate populations: people with no imputes; people with item imputes, and people with whole imputes.[21] Combinations of persons with no imputes and either whole or item imputes are then presented. Finally, the table presents the resultant poverty rates as if there had been no item imputes at all, first for all persons with positive income and then only for those persons with item imputes. All item imputed income is set equal to zero, but the respondents remain in the denominator of the poverty rate calculation.

As can be seen, poverty rates for all persons with positive income would increase dramatically from 8.6% to 25.6% if there was no item imputed income. For only those persons with item imputes, zeroing out their imputed income would increase poverty rates from 6.5% to 48.7%. These large increases indicate that item imputes may exert a strong influence on the characteristics of persons in poverty. When whole imputes are excluded, the poverty rate increases somewhat to 9.2%. Further examination indicates that a major portion of the imputations that have been zeroed out are for persons with jobs—a not surprising finding, since earnings accounted for over 82% of total income in income year 2005.

Table 1: Impact of Imputation on Poverty Rates for Persons with Positive Income in 2005

All Persons with Positive Income	8.59%
No Imputes Only	10.07%
Item Imputes Only	6.48%
Whole Imputes Only	9.14%
Whole Imputes Plus No Imputes	9.91%
Item Imputes Plus No Imputes	8.53%
All Persons with Item Imputes Equal Zero	25.60%
Item Imputes Only Equal Zero	48.66%

Effects of Imputation on Persons with Positive Income: Selected Demographic Groups

How imputation affects demographic groups that are important for policy reasons is of major importance. For this reason, estimates were developed for selected demographic characteristics at five year intervals from 1987 through 2007. They are presented in Appendix A classified by age, race, gender, family composition, and education. Similar poverty patterns emerged for all of these groups. That is, persons with no imputes tend to have the higher poverty rates while those with item imputes have the lowest poverty rates. Poverty rates for persons in categories likely to be eligible for federal programs tended to be higher than for other persons. At the same time, the significantly higher contribution of item imputation to lowering poverty rates stands out. Getting item imputation "right" appears critical to efforts toward "correctly" identifying the poor. For this reason, limited attention is paid to these findings and our analysis is directed to examination of results from the Census match project.

The 2005 Calendar Year Census Match Project

Because of the growing magnitude of income imputation and the potential for serious impacts on measures of the poor, the Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services, and the Census Bureau are matching 2006 ASEC and 2005 DER records in order to analyze how well earnings data obtained from each set of records match with each other. In calendar year 2005, 82% of the total reported income received by persons with positive income on the ASEC was from earnings.[22] The ASEC/DER merger yielded a matched data set where 92.4% of the earnings records on the ASEC were matched to earnings records on the DER (unit count).[23] As noted earlier, the matched records cover all persons with earnings and not just those with positive income[24]. When DER earnings are substituted for ASEC earnings, poverty is recalculated using the Official Poverty Thresholds based on family income and composition.[25] To be included, at least one DER matched person had to be in the family --- 82% of all persons in poverty (unit count) met this requirement. Income data was top-coded at $200,000 per person to reduce the effect of outliers.[26]

Neither data source can be viewed as the gold standard. While the DER provides accurate estimates of Social Security and Medicare covered earnings, there are other sources of earnings which are reported in the ASEC but not the DER or which are not provided to Census[27]. Sources of income other than earnings are not considered. In addition, the findings for all persons are not presented by demographic groups at this time. The need to meet confidentiality requirements for making the results public is more complicated. That is, we cannot report results when there are fewer than three unweighted cases per cell, so we are presenting less detailed income intervals[28].

It is important for policy makers to understand how the estimates of earnings obtained by the ASEC influence their analysis. It also is important to understand the degree to which these estimates might be biased. Findings for both all persons and for those in poverty are examined by looking at the degree of agreement between the ASEC and DER[29].

Findings for All Persons with Earnings

ASEC and DER earnings were cross classified by ten thousand dollar intervals. These intervals were selected in order to better meet disclosure requirements of at least three unweighted cases per cell. For presentation purposes, this information was converted to diagonals to show how often the two data sources agreed or disagreed on earnings within the same size class. As seen in Table 2, the category "same" shows the percent of all persons on the diagonal within $10,000 of each other. Diagonals on either side of the same category show how far earnings from the two sources differ. For example, the first size classes (+1 or – 1) on either side of "same" show persons who earnings on the ASEC are within plus or minus $20,000 of their earnings as reported on the DER. [30] When moving outward from the "same" diagonal, the divergence in earnings increases by $10,000 for each diagonal that is farther out. If the percent of persons who are above or below the diagonal are about equal and similar in characteristics, then the two measures are expected to have the same expected value and are unbiased[31].

Table 2: Schematic for Diagonal Charts

CPS-ASEC Reported Earnings	Matched DERS Earnings											
	$1 to $9999	$10,000 to $19,999	$20,000 to $29,999	$30,000 to $39,999	$40,000 to $49,999	$50,000 to $59,999	$60,000 to $69,999	$70,000 to $79,999	$80,000 to $89,999	$90,000 to $99,999	$100,000 or more	Sum of Diagonals
$1 to $999	Same	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	
$10,000 to $19,999	+1	Same	-1	-2	-3	-4	-5	-6	-7	-8	-9	Sum -10
$20,000 to $29,999	+2	+1	Same	-1	-2	-3	-4	-5	-6	-7	-8	Sum -9
$30,000 to $39,999	+3	+2	+1	Same	-1	-2	-3	-4	-5	-6	-7	Sum -8
$40,000 to $49,999	+4	+3	+2	+1	Same	-1	-2	-3	-4	-5	-6	Sum -7
$50,000 to $59,999	+5	+4	+3	+2	+1	Same	-1	-2	-3	-4	-5	Sum -6
$60,000 to $69,999	+6	+5	+4	+3	+2	+1	Same	-1	-2	-3	-4	Sum -5
$70,000 to $79,999	+7	+6	+5	+4	+3	+2	+1	Same	-1	-2	-3	Sum -4
$80,000 to $89,999	+8	+7	+6	+5	+4	+3	+2	+1	Same	-1	-2	Sum -3
$90,000 to $99,999	+9	+8	+7	+6	+5	+4	+3	+2	+1	Same	-1	Sum -2
$100,000 or more	+10	+9	+8	+7	+6	+5	+4	+3	+2	+1	Same	Sum -1
Sum of Diagonals		Sum +10	Sum +9	Sum +8	Sum +7	Sum +6	Sum +5	Sum +4	Sum +3	Sum +2	Sum +1	Sum Same

Charts 4 through 6 show the distribution of ASEC earnings category differences from DER earnings categories by $10,000 income intervals in calendar year 2005 for all persons with earnings, reporters (no imputes) and persons with imputed earnings. The latter includes both item and whole imputes. For all three charts there are more instances where the ASEC is higher than the DER, suggesting a slight tendency for the average respondent to report more on the ASEC than is reported by the DER records. Overall, the agreement between the two data sources is reasonably high for all persons and for those with no imputes. There is a steep drop-off and low fraction of the population in all but the same class in all instances.

As seen in Chart 4, over 50% of all persons were in the same category on both the ASEC and the DER; while 79% of all persons have earnings that are within one $10,000 interval of each other. About 5% of the persons had DER earnings that deviated by 4 or more $10,000 intervals from the reported values. Sixty one percent of persons with no imputes have DER earnings within the same $10,000 interval as reported on the ASEC; while, 88% of all persons with no imputes are within one $10,000 interval of each other. Only 27% of those with item and whole imputes are in the same $10,000 interval; while 57% were within one $10,000 interval of each other.

The picture is somewhat muddied, given the differences in methods for producing these two types of imputation, and further breakouts are needed by type of imputation to fully evaluate what is occurring. While imputation coarsened the relationship; it did not end it. Slightly over one fourth of the imputations matched the DER closely. However, in all instances, deviations are fairly evenly distributed around the same interval so in the aggregate the positive and negative deviations balance each other out. As seen later, the results are more accurate in the aggregate than at the unit level. Since the histograms are not symmetric but skewed, with more ASEC records showing slightly higher earnings; one might conclude that there is bias. It is likely however, that definitional differences in income reporting are responsible for much of the skewness[32].

Chart 4: Distribution of ASEC and DER Earnings Class Differences
All: 2005

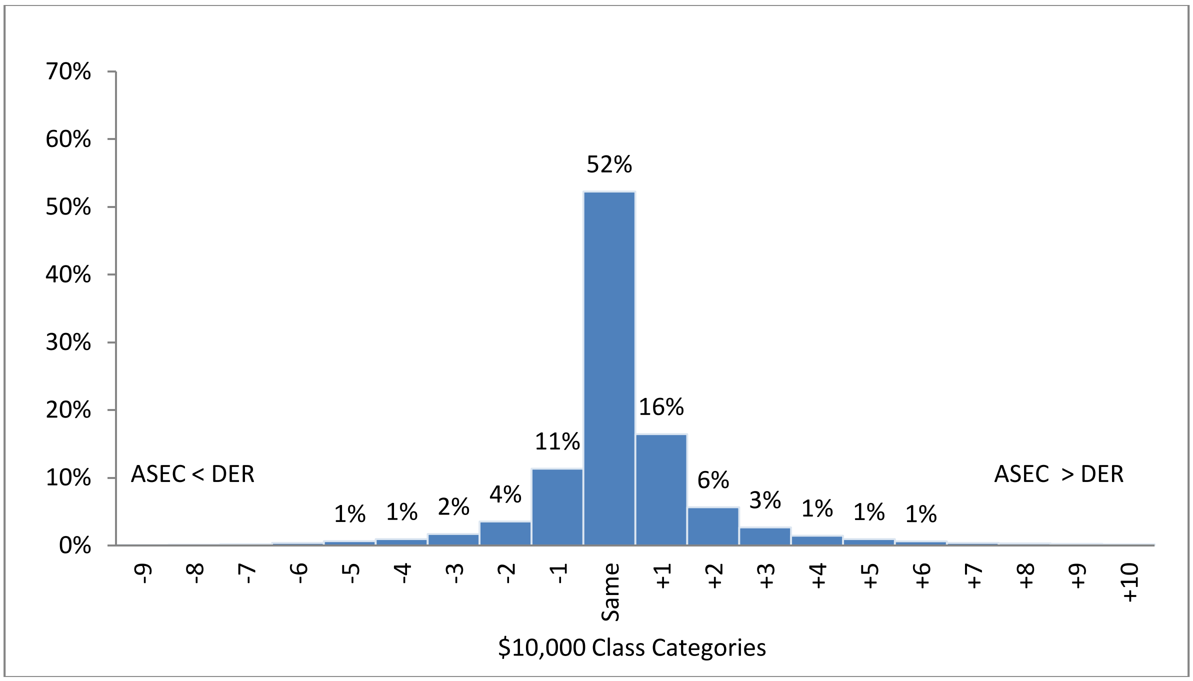


Chart 5: Distribution of ASEC and DER Earnings Class Differences
2005: No Imputes

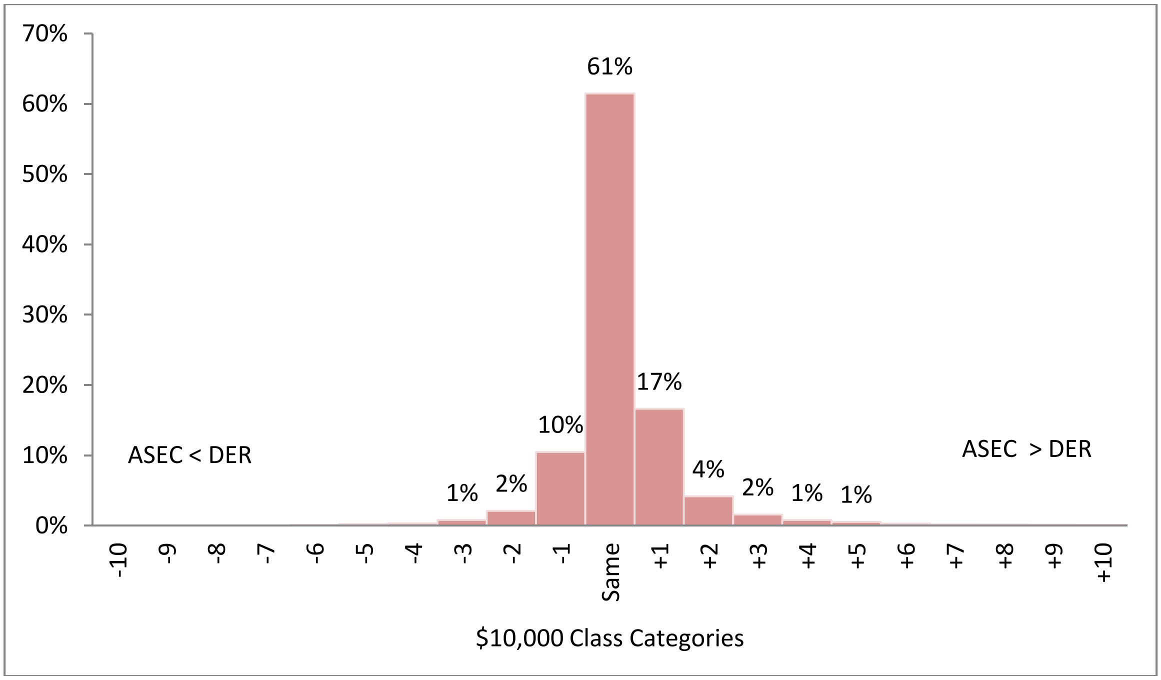
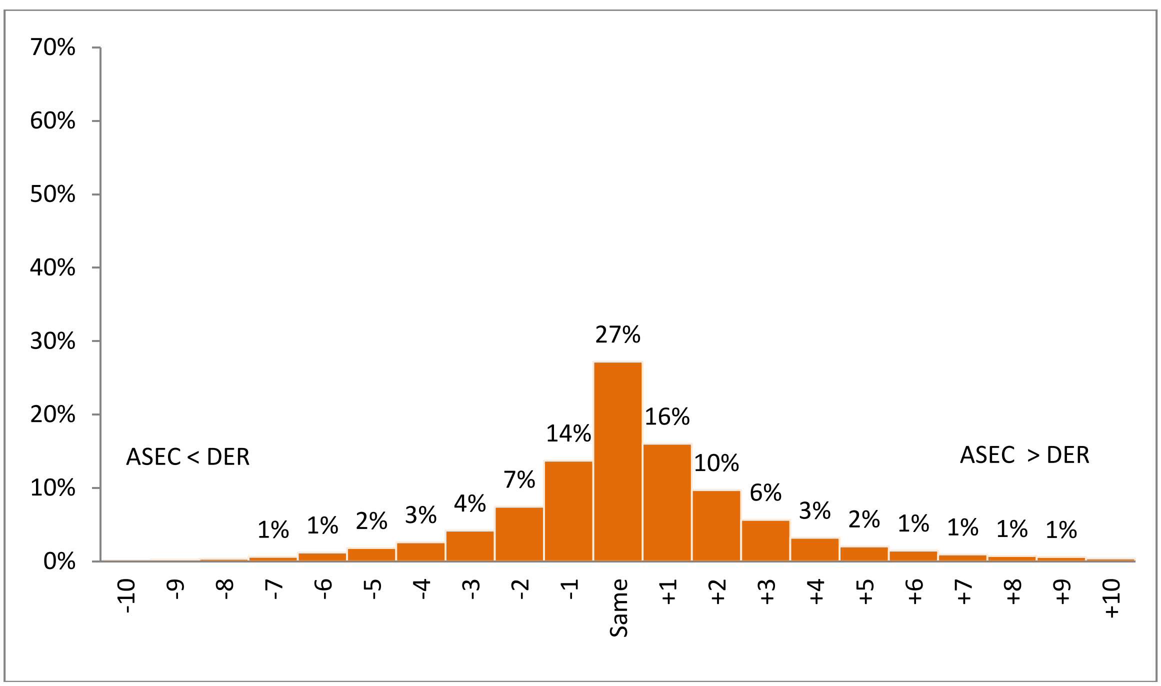


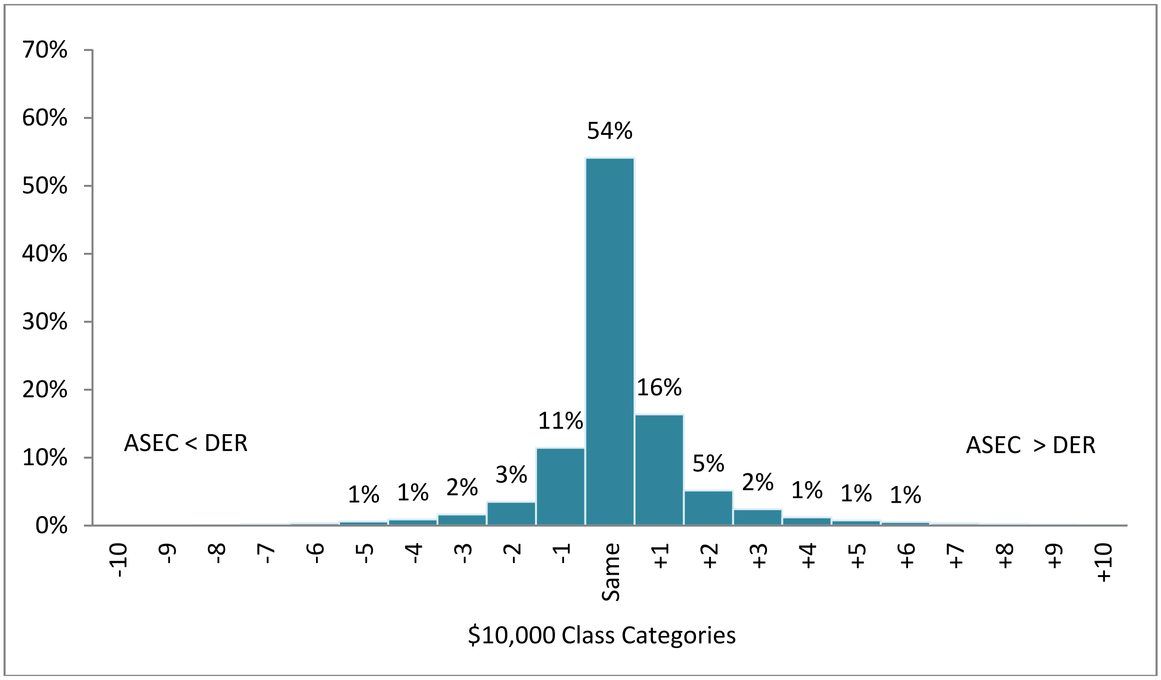
Chart 6: Distribution of ASEC Earnings Class Differences
2005: Whole and Item Imputes



For Wages Only

Self-employment earnings are less well reported on the ASEC than wages[33]. What would our charts look like if we looked only at wages which accounted for 76% of income and 93% of earnings? As seen in Chart 7, there was a slight increase in the percent of all persons who were in the same category when wages are considered on both the ASEC and the DER from 52% to 54%. Eighty one percent of persons with wages were within plus or minus $20,000 of each other. Overall, wages dominated the earnings estimates.

Chart 7: Distribution of ASEC and DER Wage Class Differences
2005: All



All things considered, it would have been nice if there had been a closer correspondence between reported earnings on the ASEC and DER. The results, however, do not invalidate the process. Clearly, imputation adds value by increasing the income available. It does not appear to bias the results. Although there are many cases where the dollar values for individuals did not closely match, aggregate statistics developed from these two data sources appear to be highly similar to each other. The following sections provide comparisons of aggregate income and poverty measures.

Aggregates

Table 3 shows the relationship between the mean earnings obtained on the ASEC and on the DER when earnings on the two data sets are classified by the ratio of ASEC to DER earnings. As can be seen, differences in earnings reported by individual persons wash out in the aggregate. That is, total, not imputed and imputed earnings all have mean ratios that are within 82% or more even when they differ by 25 to < 50%. Even though imputed earnings are not as good a match at the person level, total mean imputed earnings on the ASEC for all persons are larger than total mean earnings on the DER by only 1.9%. Thus, in the aggregate, the two data sources will have very similar estimates of total earnings.

Table 3: Ratio of Mean Earnings, ASEC/DER (Unit Counts)
Percent of Persons

ASEC/DER	Total	Not imputed	Imputed
ASEC earnings < 5% different from DER	99.9%	99.9%	99.9%
5% to < 10% different	98.9%	98.9%	99.4%
10% to < 15% different	97.1%	97.0%	97.6%
15% to < 20%different	95.4%	95.4%	95.5%
20% to < 25% different	92.8%	92.8%	92.5%
25% < 50% different	86.3%	89.3%	821%
50% < 75% to < 100% different	71.7%	87.7%	62.6%
75% to , 100% different	73.0%	110.9%	54.1%
100% or more different	370.0%	357.9%	379.4%
Total	100.7%	100.3%	101.9%

For the findings to be useful for policy purposes, much more information is needed. Item and whole imputed earnings need to be considered separately. More complete information on wages also would be interesting. What is happening to important demographic groups? How good are the data for high income persons? These, and other, questions are still to be addressed. However, findings are available for one important group --- those in poverty—both for everyone and by major demographic group.

Findings for Persons in Poverty

Overall, the hot deck procedures used by Census in undertaking imputations appear to be performing well for the poverty population. Table 4 shows the percent of all persons with positive income who changed poverty status when DER earnings are substituted for ASEC earnings in income year 2005. Results are also presented for age, race, gender, family composition, education, family size and employment status in Appendix B.

Perhaps the most important finding from this exercise is that when DER earnings are substituted for ASEC earnings, the great majority of persons do not change poverty status. As seen in Table 2, the poverty status of over 90 percent of the persons with positive income, who either have no imputes or item imputes, did not change. Slightly fewer than 9 out of 10 persons with whole imputes did not have their poverty status change when DER earnings were substituted for ASEC earnings. This finding remains for most of the demographic groups examined as well. Where exceptions occur, they tend to be for lower income persons who are black, single parents or persons who are not high school graduates (less than high school and some high school). These are all important policy groups and attempts should be made to determine why this is the case. Even for these groups, over 85% of persons with positive income have no change in poverty status, with the exception of single parents where no change in poverty was in the low eighties. Irrespective of the group examined, the composition of poverty for the total is similar to the composition for no imputes. What does this mean? Do people in poverty do a fairly good job reporting their earnings in the ASEC? Can the same be said for item imputes (it appears so) and whole imputes (not as good)?

The differences in income between the DER and ASEC are roughly symmetrical with slightly more persons entering poverty rather than leaving poverty. The percentages entering and leaving poverty are relatively small and most are not statistically significant. The changes do not dramatically alter our understanding of the size and composition of the poor, as measured by the federal government. Only five of the net poverty changes presented in Table 4 and in Appendix B were statistically significant at the 95 percent confidence level.

Table 5: Change in Poverty Status When DER Earnings Substituted for ASEC Earnings: Calendar Year 2005

	No Imputes	Item Imputes	Whole Imputes	Total
No change in poverty Status	94.4%	92.8%	89.2%	93.7%
Change from Poor to Non-poor	2.8%	3.0%	5.3%	3.0%
Change for Non-poor to Poor	2.9%	4.3%	5.5%	3.3%
Net Change Non-poor to Poor	0.1%	1.3%	0.3%	0.2%

Note: the gross change rate in this table is 100.0% minus 93.7% (no change in poverty status).

Future Directions

Several reports have already presented findings for portions of the complex data set developed for analyzing trends in imputation.[34] Further analyses will look more in depth at wages in addition to earnings as well as at important policy groups. Nonmatches will also be examined to see if they differ from matches. ASEC data is also being matched to Supplemental Security Income and OASDI records. The data set is so rich that it can support a wide range of analyses for many years.

Acknowledgements

The authors thank Charles T. Nelson, Assistant Division Chief for Economic Characteristics, and Edward J. Welniak Jr., Chief of the Income Statistics Branch, Bureau of the Census, for their extensive assistance in developing the comprehensive data series, including the matched data, used in this report, as well as for reviewing this report and its supporting materials.

Appendix A: Poverty Rates by Selected Demographic Characteristics

Poverty rates are presented In Appendices A and B for the following demographic groupings for everyone and for persons in poverty at five year intervals and for the five year average:

- **Family Composition:** Single parent families with children are important recipients of Federal aid, such as welfare programs, and are about four times more likely to be poor than married-parent families. They have the highest poverty rates of any of the demographic groups studied.

- **Age:** The three selected age groupings , persons aged 18 to 44, aged 45 to 64, and 65 and older, can be viewed as separating persons in the earlier stages of their careers and more likely to have children under 18 from those who are farther advanced and more likely to have college aged children. Persons 18 to 44 are more likely to have children under 18 and to be eligible for welfare programs. Many persons 65 and older are likely to be retired, although this may be changing as health improves and economic circumstances worsen. Persons aged 18 to 44 and 65 and older have similar poverty rates while persons 45 to 64 have the lowest poverty rates. Persons 65 and older have the highest poverty rate. Their higher poverty rate may reflect underreporting of asset income on the ASEC and their higher percentage of income from unearned sources.

- **Race:** Two race categories are presented, Whites and Blacks. Finer distinctions identifying additional race categories (e.g., Asians) and separately identifying Hispanics could not be constructed covering the entire time period covered by the analysis. Large disparities exist by race. The poverty rate of Blacks with no imputes is more than double that for Whites. That is, they are 23.7% vs. 8.6%, respectively.

- **Gender:** Women also have considerably higher poverty rates than men. Females are about 50 percent more likely to be poor than males. Again, the biggest differences are between persons with no or whole imputes.

- **Education:** Poverty rates vary inversely with education level. Overall, poverty rates for persons without a high school education are substantially higher. In fact, persons with less than a high school education have poverty rates about 8 times higher than for college graduates. The educational categories are “less than a high school education” (less HS); “some high school, but no degree” (Some HS); “high school graduates” (HS Grad), some college, (Some Col), and college graduate with a BA or advanced degree (Col Grd).

Poverty Rates by Selected Demographic Characteristics
Family Composition

Non Single parents	1987	1992	1997	2002	2007	5 Yr. Average
No	8.60%	9.29%	8.46%	8.44%	8.20%	8.62%
Item	5.98%	6.69%	5.85%	5.96%	5.45%	5.89%
whole	8.64%	9.29%	8.18%	8.18%	7.07%	8.27%
Total pop	8.09%	8.76%	7.55%	7.38%	6.91%	7.70%
Single Parents						
No	42.05%	41.83%	37.17%	31.28%	31.78%	36.95%
Item	29.41%	31.85%	25.06%	20.10%	20.57%	23.43%
whole	33.38%	39.32%	35.44%	25.12%	29.16%	32.07%
Total pop	39.78%	40.26%	34.09%	27.06%	28.30%	33.40%

Age

18 to 45	1987	1992	1997	2002	2007	5 Yr. Average
No	10.46%	11.89%	11.44%	10.94%	11.06%	11.17%
Item	7.07%	8.13%	7.49%	7.25%	6.97%	7.33%
whole	10.24%	11.62%	10.50%	10.47%	9.61%	10.51%
Total pop	9.87%	11.21%	10.17%	9.50%	9.52%	10.06%
45 to 64						
No	8.34%	8.36%	7.89%	7.79%	8.09%	8.10%
Item	4.97%	5.37%	4.79%	5.07%	4.69%	4.92%
whole	7.78%	7.55%	7.56%	6.49%	6.80%	7.15%
Total pop	7.51%	7.57%	6.72%	6.44%	6.51%	6.87%
65 and older						
No	13.70%	15.39%	11.19%	11.36%	9.46%	12.31%
Item	7.98%	9.24%	7.42%	7.78%	7.18%	7.75%
whole	11.56%	13.67%	10.08%	10.36%	7.61%	10.83%
Total pop	12.19%	12.45%	9.60%	9.58%	8.28%	10.31%

Race

White						
No	8.17%	9.14%	8.87%	8.61%	8.38%	8.64%
Item	5.91%	6.77%	5.85%	5.79%	5.21%	5.79%
whole	8.64%	9.34%	8.10%	8.23%	7.25%	8.32%
Item + no	7.68%	8.60%	7.75%	7.28%	7.03%	7.65%
Total pop	7.77%	8.68%	7.78%	7.38%	7.05%	7.71%
Black						
No	27.27%	28.21%	22.20%	20.21%	19.53%	23.69%
Item	14.21%	15.34%	12.69%	12.47%	12.70%	13.09%
whole	16.96%	18.85%	17.30%	15.62%	14.31%	16.63%
Item + no	24.96%	25.57%	18.93%	16.90%	16.80%	20.31%
R Total pop	24.00%	24.45%	18.73%	16.74%	16.48%	19.82%

Gender

	1987	1992	1997	2002	2007	5 Yr. Average
Males						
No	7.86%	8.75%	7.94%	8.12%	8.09%	8.16%
Item	5.73%	6.12%	5.21%	5.29%	4.77%	5.29%
whole	8.52%	8.17%	7.71%	7.11%	6.82%	7.62%
item plus no	7.40%	8.15%	6.93%	6.81%	6.69%	7.16%
Total pop	7.51%	8.15%	7.00%	6.84%	6.70%	7.21%
Females						
No	13.03%	13.80%	12.91%	11.87%	11.54%	12.70%
Item	7.74%	9.11%	8.15%	7.88%	7.44%	7.96%
whole	11.06%	13.56%	11.23%	11.09%	9.89%	11.40%
item plus no	11.92%	12.76%	11.19%	10.04%	9.81%	11.08%
Total pop	11.84%	12.85%	11.19%	10.15%	9.82%	11.11%

Education

	1987	1992	1997	2002	2007	5 Yr. Average
Less HS						
No	27.89%	28.89%	26.81%	25.66%	24.78%	27.14%
Item	15.26%	18.40%	18.20%	16.90%	17.25%	17.22%
whole	21.15%	21.89%	19.52%	16.36%	18.25%	19.75%
Total pop	25.21%	26.34%	23.95%	21.94%	22.05%	24.15%
Some HS						
No	18.23%	22.12%	20.63%	19.23%	21.01%	20.15%
Item	12.57%	14.31%	14.88%	13.39%	14.51%	13.96%
whole	15.10%	19.46%	16.42%	14.98%	16.54%	16.58%
Total pop	17.03%	20.48%	18.71%	16.54%	18.64%	18.24%
HSGrad						
No	8.88%	10.55%	10.07%	10.22%	10.89%	10.07%
Item	6.33%	8.04%	7.06%	7.40%	7.18%	7.22%
whole	8.65%	10.99%	9.79%	9.41%	8.63%	9.55%
Total pop	8.37%	10.10%	9.04%	8.99%	9.28%	9.16%
Some Col						
No	6.10%	7.31%	7.09%	7.13%	7.21%	6.99%
Item	4.75%	5.74%	4.95%	5.80%	5.48%	5.42%
whole	6.70%	7.75%	7.51%	8.54%	7.81%	7.77%
Total pop	5.88%	7.04%	6.38%	6.73%	6.61%	6.56%
Col Grd						
No	2.56%	2.81%	2.77%	3.03%	2.90%	2.82%
Item	2.84%	2.74%	2.50%	2.78%	2.46%	2.62%
whole	4.28%	3.36%	4.15%	4.31%	3.46%	3.88%
Total pop	2.76%	2.85%	2.78%	3.03%	2.75%	2.84%

Appendix B: Change in Poverty Status When DER Earnings Substituted for ASEC Earnings by Selected Demographic Groups, Calendar Year 2005[35]

For decades, the Census Bureau's P-60 income and poverty series has used a generalized variance estimation approach (described in Wolter 2007)[36]. That approach was changed a few years ago to a form of direct replication (also described in Wolter 2007). So we had two ways of doing the formal hypothesis testing for bias.

We judged, for the exercise here, that both methods would have been roughly equivalent. For reasons of convenience we chose the generalized variance approach. What were our steps? These are listed below. We began with 90% upper confidence bounds as reported in the P-60 for income year 2005 by race and ethnic origin. We interpolated, as would have been done historically, from them to the numbers shown in the basic tables displayed here. We looked only at percentages for the 2005 poverty rate, not at the population estimates. The task was to compare the differences between the percentage of non-poor who became poor versus the percentage that started out poor but became non-poor. Take for example the overall percents in the first column of Table 4. The percent that were non-poor and became poor was 2.9%, while the percent that started out non-poor but ended up poor was 2.8%. The net difference is thus 2.9% minus 2.8 =0.1%. To formally test the hypothesis that there was no bias, we need to see if this net difference was statistically significantly different from zero. In most cases, incidentally, except for a few very small subpopulations with large sampling uncertainty, the numerical net differences found were not substantively sizable enough to worry about.

How to proceed formally? To start with, the two percents are not independent; hence, the usual t-test of the difference between two proportions does not apply. To treat the percents as independent would yield confidence intervals that were too wide. What to do? Well this problem turns out to be akin to deciding which of two candidates has a statistically significant lead in an election. Conveniently, then, we can adjust the standard error of the level estimate, using Ansolabehere and Belin's[37] result as found in Chance (1998) by the square root of 3 that is by 1.73. This means that if the overall confidence interval at a 90% confidence value is, say, 0.2% , then the difference between the two percents, a measure of the bias, would need to be greater than about .4% for there to be a statistically significant difference at the 95% level.

Are we done now? Not quite. Now of course, for subpopulations that are smaller --Blacks, Women, Hispanics -- the percent differences would have to be much larger to achieve significance since the ASEC sample size is smaller than the DER. The factor would involve the ratio of the square roots of the total population to the subpopulation, say, Blacks. For the bulk of the cases this is enough, but there is still one more step, to account for the increase in the variance due to the hot deck imputation itself. We cannot estimate this directly. To do that would require we use multiple imputation (Rubin 1978)[38] and in the ASEC only one imputation was done.[39] In part, following Hansen, Hurwitz and Madow (1953)[40], we increased the item and whole impute estimates by 1.125.

While approximations are required here the calculations show that there is little or no bias in the ASEC poverty estimates as measured by substituting administrative earnings data for the earnings data obtained in the ASEC process. Replicate estimates would have been quite a challenge to carry out and they too have problems that make them hard to interpret. Bottom line, we came away convinced that the results were fit for our preliminary use here.

Change in Poverty Status When DER Earnings Substituted for ASEC Earnings by Selected Demographic Groups, Calendar Year 2005
Age

	Persons 18 to 44			
	No Imputes	Item Imputes	Whole Imputes	Total
No change in poverty Status	93.9%	91.1%	88.0%	92.9%
Change from Poor to Non-poor	3.0%	3.5%	5.8%	3.3%
Change for Non-poor to Poor	3.2%	5.4%	6.2%	3.8%
Net Change Non-poor to Poor	0.2%	1.9%	0.5%	0.5%
	Persons 45 to 64			
No change in poverty Status	96.7%	94.6%	92.7%	96.0%
Change from Poor to Non-poor	1.8%	2.4%	4.1%	2.1%
Change for Non-poor to Poor	1.5%	2.9%	3.2%	1.9%
Net Change Non-poor to Poor	-0.4	0.5%	-0.9%	-0.3%
	Persons 65 and Older			
No change in poverty Status	97.0%	96.5%	94.3%	96.7%
Change from Poor to Non-poor	2.3%	1.4%	4.0%	2.4%
Change for Non-poor to Poor	0.7%	2.0%	1.6%	0.9%
Net Change Non-poor to Poor	-1.7%	0.6%	-2.4%	-1.5%*

Note: the gross change rate in this table is 100.0% minus the no change in poverty status.

Race

	No Imputes	Item Imputes	Whole Imputes	Total
	White			
No change in poverty Status	95.1%	93.3%	89.9%	94.5%
Change from Poor to Non-poor	2.3%	2.8%	5.6%	2.6%
Change for Non-poor to Poor	2.6%	4.0%	4.6%	2.9%
Net Change Non-poor to Poor	0.4%	1.2%	-1.0%	0.3%
White, Non-Hispanic				
No change in poverty Status	96.1%	93.6%	90.9%	95.4%
Change from Poor to Non-poor	1.6%	2.5%	4.7%	1.9%
Change for Non-poor to Poor	2.3%	3.9%	4.5%	2.7%
Net Change Non-poor to Poor	0.7%	1.4%	-0.2%	0.8%
White, Hispanic				
No change in poverty Status	89.9%	90.6%	85.3%	89.5%
Change from Poor to Non-poor	6.0%	4.9%	9.6%	6.2%
Change for Non-poor to Poor	4.2%	4.5%	5.1%	4.3%
Net Change Non-poor to Poor	-1.8%	-0.4%	-4.5%*	-1.9%
	Black			
No change in poverty Status	89.8%	89.1%	85.8%	89.3%
Change from Poor to Non-poor	6.0%	4.6%	4.5%	5.6%
Change for Non-poor to Poor	4.2%	6.3%	9.7%	5.1%
Net Change Non-poor to Poor	-1.8%	1.7%	5.2%*	0.5%
	Other			
No change in poverty Status	93.1%	94.4%	89.5%	92.8%
Change from Poor to Non-poor	3.6%	2.1%	4.3%	3.5%
Change for Non-poor to Poor	3.3%	3.5%	6.3%	3.6%
Net Change Non-poor to Poor	-0.4%	1.4%	2.0%	0.1%

Note: the gross change rate in this table is 100.0% minus the no change in poverty status.

Gender

	Male			
	No Imputes	Item Imputes	Whole Imputes	Total
No change in poverty Status	94.4%	93.3%	89.4%	93.8%
Change from Poor to Non-poor	2.7%	2.6%	5.0%	2.9%
Change for Non-poor to Poor	2.9%	4.2%	5.6%	3.3%
Net Change Non-poor to Poor	0.3%	1.6%	0.6%	0.5%
	Female			
No change in poverty Status	94.3%	92.2%	89.0%	93.7%
Change from Poor to Non-poor	2.9%	3.5%	5.6%	3.2%
Change for Non-poor to Poor	2.8%	4.4%	5.5%	3.2%
Net Change Non-poor to Poor	-0.1%	1.0%	-0.1%	0.0%

Note: the gross change rate in this table is 100.0% minus no change in poverty status.

Family Composition

	Single Parents			
	No Imputes	Item Imputes	Whole Imputes	Total
No change in poverty Status	88.1%	80.8%	76.3%	86.4%
Change from Poor to Non-poor	5.3%	6.3%	10.1%	5.8%
Change for Non-poor to Poor	6.6%	12.9%	13.7%	7.7%
Net Change Non-poor to Poor	1.4%	6.6%	3.6%	1.9%*
	Married Parents			
No change in poverty Status	95.1%	94.4%	92.1%	94.8%
Change from Poor to Non-poor	2.3%	2.8%	3.8%	2.5%
Change for Non-poor to Poor	2.6%	2.8%	4.1%	2.8%
Net Change Non-poor to Poor	0.3%	0.1%	0.3%	0.3%

Note: the gross change rate in this table is 100.0% no change in poverty status.

Education

	Less than High School			
	No Imputes	Item Imputes	Whole Imputes	Total
No change in poverty Status	92.1%	86.0%	84.9%	91.5%
Change from Poor to Non-poor	3.8%	6.4%	6.7%	4.1%
Change for Non-poor to Poor	4.1%	7.5%	8.4%	4.5%
Net Change Non-poor to Poor	0.3%	1.1%	1.7%	0.4%
	Some High School			
No change in poverty Status	91.9%	88.6%	86.2%	90.9%
Change from Poor to Non-poor	4.4%	4.2%	6.9%	4.6%
Change for Non-poor to Poor	3.8%	7.2%	6.9%	4.5%
Net Change Non-poor to Poor	-0.6%	3.0%	0.0%	-0.1%
	High School Graduate			
No change in poverty Status	94.2%	91.7%	89.2%	93.3%
Change from Poor to Non-poor	2.8%	3.4%	5.4%	3.2%
Change for Non-poor to Poor	2.9%	4.8%	5.3%	3.5%
Net Change Non-poor to Poor	0.1%	1.4%	-0.1%	0.3%
Some College				
No change in poverty Status	95.9%	92.4%	90.8%	94.9%
Change from Poor to Non-poor	2.1%	3.1%	4.6%	2.4%
Change for Non-poor to Poor	2.1%	4.5%	4.5%	2.6%
Net Change Non-poor to Poor	0.0%	1.4%	-0.1%	0.2%
	College Graduate			
No change in poverty Status	97.7%	96.2%	94.7%	97.2%
Change from Poor to Non-poor	1.1%	1.6%	2.8%	1.3%
Change for Non-poor to Poor	1.2%	2.2%	2.4%	1.5%
Net Change Non-poor to Poor	0.1%	0.6%	-0.4%	0.2%

Note: the gross change rate in this table is 100.0% minus no change in poverty status).

Employment Status

	Non-worker			
	No Imputes	Item Imputes	Whole Imputes	Total
No change in poverty Status	92.1%		85.6%	91.5%
Change from Poor to Non-poor	4.5%		8.5%	4.9%
Change for Non-poor to Poor	3.4%		5.9%	3.6%
Net Change Non-poor to Poor	-1.1%		-2.6%	-1.3%
Less than 35 hours per week				
No change in poverty Status	95.2%	90.1%	90.0%	93.7%
Change from Poor to Non-poor	2.0%	4.9%	6.0%%	2.9%
Change for Non-poor to Poor	2.8%	5.0%	4.0%	3.3%
Net Change Non-poor to Poor	0.8%	0.1%	-2.0%	0.4%
35 or more hours per week				
No change in poverty Status	96.6%	93.4%	92.0%	95.6%
Change from Poor to Non-poor	1.1%	2.5%	2.4%	1.5%
Change for Non-poor to Poor	2.3%	4.1%	5.6%	2.9%
Net Change Non-poor to Poor	1.2%	1.6%	3.2%	1.4%

Note: the gross change rate in this table is 100.0% minus the no change in poverty status

Family Size

	Unrelated Individual			
	No Imputes	Item Imputes	Whole Imputes	Total
No change in poverty Status	92.0%	84.1%	77.5%	89.5%
Change from Poor to Non-poor	3.4%	5.5%	10.3%	4.3%
Change for Non-poor to Poor	4.6%	10.5%	12.3%	6.2%
Net Change Non-poor to Poor	1.2%	5.1%*	2.0%	1.9%
Two Person Family				
No change in poverty Status	95.8%	95.1%	91.1%	95.3%
Change from Poor to Non-poor	2.2%	2.1%	5.0%	2.4%
Change for Non-poor to Poor	2.0%	2.8%	3.9%	2.3%
Net Change Non-poor to Poor	-0.2%	0.7%	-1.1%	-0.1%
Three person family				
No change in poverty Status	95.0%	95.0%	91.5%	94.7%
Change from Poor to Non-poor	2.4%	1.9%	3.8%	2.5%
Change for Non-poor to Poor	2.6%	3.1%	4.7%	2.8%
Net Change Non-poor to Poor	0.2%	1.1%	0.9%	0.3%
	Four person family			
No change in poverty Status	95.5%	95.0%	92.9%	95.3%
Change from Poor to Non-poor	2.1%	2.5%	3.5%	2.2%
Change for Non-poor to Poor	2.4%	2.5%	3.6%	2.5%
Net Change Non-poor to Poor	0.3%	0.0%	0.1%	.3%
	Five Person Family			
No change in poverty Status	93.7%	92.1%	88.9%	93.1%
Change from Poor to Non-poor	3.6%	3.6%	4.7%	3.7%
Change for Non-poor to Poor	2.8%	4.4%	6.5%	3.2%
Net Change Non-poor to Poor	-0.8%	0.8%	1.8%	-0.5%
Six person family				
No change in poverty Status	91.9%	92.3%	87.5%	91.5%
Change from Poor to	4.1%	3.9%	7.5%	4.4%

Non-poor				
Change for Non-poor to Poor	4.0%	3.7%	5.1%	4.1%
Net Change Non-poor to Poor	0.1%	-0.2%	-2.4%	-0.3%
Seven or more person family				
No change in poverty Status	88.3%	90.6%	85.2%	88.1%
Change from Poor to Non-poor	6.2%	5.2%	7.8%	6.3%
Change for Non-poor to Poor	5.5%	4.1%	7.1%	5.6%
Net Change Non-poor to Poor	-0.8%	-1.1%	-0.7%	-0.8%

Note: the gross change rate in this table is 100.0% minus the no change in poverty status.

References

[1] The views are those expressed by the authors and are not the official position of any of their organizations.
[2]Joan Turek, Fritz Scheuren, Brian Sinclair-James, Bula Ghose and Sameer Desale, *Effects of Imputation on CPS Income and Poverty Series";* Joint Statistical Meetings, Washington D.C.; August 4, 2009 and Joan Turek, Fritz Scheuren, Charles Nelson, Edward Welniak Jr., Brian Sinclair-James, and Bula Ghose; *Effects of Imputation on CPS Poverty Series: 1987- 2007;* Federal Committee on Statistical Methodology; Washington D.C., November 3, 2009; Joan Turek, Fritz Scheuren, Kendall Swenson, Bula Ghose, *Effects of Imputation on Trends and Demographic Characteristics,* 33rd Annual Fall Research Conference, Association of Public Policy Analysis & Management, Washington D.C., November 4, 2011; Joan Turek, Fritz Scheuren, *Imputation and Income Distribution Measures*, (forthcoming).
[3] Unfortunately, administrative records permitting evaluation of all income sources do not exist.
[4] Many improvements also have occurred in the ASEC during this time. These improvements include a large expansion of the questions on sources of income in the late 1980s and conversion from paper questionnaires to computer assisted interviewing techniques in the 1990s.
[5] Blumerman, L.M. and P.M. Vidal, 2009, *Uses of Population and Income Statistics in Federal Funds Distribution – with a Focus on Census Bureau Data*. Government Division Reports Series, Research Report "2009-1.
[6] Numerous studies have reported on the growing income disparity in the United States since the 1970s. By 2008, the trend towards greater inequality in the distribution of income had reached levels not seen since the Depression. See*:* Arthur E. Jones Jr. and Daniel H. Weinberg, *The Changing Shape of the Nation's Income Distribution: 1947-1998*, Current Population Reports, P60-204, U.S. Census Bureau, U.S. Department of Commerce, June 2000, p.11.; and Peter Whoriskey, *Income gap widens as executives prosper,* The Washington Post, Sunday, June 19, 2011, page A1.
[7] Marc Roemer, *Using Administrative Earning Records to Assess Wage Data Quality in the March Current Population Survey and the Survey of Income and Program Participation,* Technical Paper TP-2002-22, U.S. Census Bureau LEHD Program, November 2002, states on page 4: 'The Detailed Earnings Records are an extract from the Social Security Administration's (SSA) Master Earnings File of nearly all workers in the United States. The data are annual records of the full amount of wages (untopcoded) and self-employment paid to workers by each employer. An employer reports earnings to the SSA by Employer Identification Number (EIN) and the Social Security Number (SSN) of the worker. The SSA verifies the name and birth date of the record before entering the earnings data into the database. About 1 percent of the reported earnings records remain unverified and absent from the database supplying the DER…..Known reporting problems exist in the DER data, causing under coverage of workers in private households, agriculture and construction in particular. A special advantage of the DER data is that its worker specific." Some classes of workers, such as, Federal employees hired before 1982 and some state and local employees are excluded from the DER since they do not pay into the Social Security system.
[8] A detailed description of the development of this data set is presented in Joan Lee Turek Ph.D. and Bula Ghose, *Documentation for CPS Income Imputation Dataset,* 2008, Joan.Turek@hhs.gov.
[9] Joan Turek, Fritz Scheuren, Brian Sinclair-James, Bula Ghose and Sameer Desale, *Effects on Imputation on CPS Income and Poverty Series: 1981 – 2007*, Papers and Proceedings of the American Statistical Association, August 4, 2009, Washington D.C.; Joan Turek, Fritz Scheuren, Charles Nelson, Edward Welniak Jr., Brian Sinclair-James and Bula Ghose, *Effects of Imputation on CPS Poverty Series, 1987-2007,* Federal Committee on Statistical Methodology, November 2009, Washington D.C. and Joan Turek, Fritz Scheuren, Charles Nelson, Edward Welniak Jr., and Bula Ghose, *Effects of Income Imputation on Traditional Poverty Estimates, 1987-2007*, Association of Public Data Users, September 2010, Washington D.C.
[10] Op.cit., Turek and Ghose, p.4. In constructing estimates of persons with positive income, each component of income that was negative was set equal to zero and the person remained in the data set. Another approach would have been to exclude all persons in families which contained someone with negative income. The difference in approach has almost no effect on measured poverty rates.
[11] In identifying persons with negative income, we first zeroed out any source of negative income reported by a respondent and then re-estimated income. So only those persons whose total income remained negative when income was re-estimated were excluded. This was done because of our desire to support analysis by major sources of income.

[12]See for example, H. Lock Oh, Social Security Administration, Fritz Scheuren, Internal Revenue Service, Hal Nisselson, WESTAT, *Differential Bias Impacts of Alternative Census Bureau Procedures for Imputing Missing CPS Income Data,* Papers and Proceedings, Joint Statistical Meetings, 1980.
[13] Ford, Barry, *Hot Deck Imputation* in Theory of Incomplete Data, Volume II, National Academy of Science, 1983.
[14] *Technical Paper 54: Income Nonresponse: March 1983 CPS*, September 1985, U.S. Department of Commerce, Bureau of the Census, Washington D.C., 20233 and Edward J. Welniak Jr., *Effects of the March Current Population Survey's New Processing System on Estimates of Income and Poverty,* 1990, U.S. Department of Commerce, Bureau of the Census, Washington D.C. 20233.
[15] The new questionnaire was fully adopted in 1980 with minor modifications.
[16] Op. cit., Welniak. This discussion relies heavily on his paper.
[17] U.S. Census Bureau, Survey of Program Dynamics, *Data Editing and Imputation Goals,* http://www.census.gov/spd/goals.html.
[18] Non-interview non-responses, where both the income supplement and monthly labor force surveys are missing, are not covered in this paper, since they are handled by weight adjustments and not imputation.
[19] The ASEC collects detailed information on income by source; on employment, earnings, hours of work; and on a variety of demographic characteristics such as age, race, marital status, educational attainment, health insurance coverage and participation in transfer programs.
[20] For example, if a respondent fails to report his/her annual earnings but does report the number of weeks worked in the previous year, then this information can be used by Census in its hot deck procedures for imputing earned income in a manner that reflects annual employment. Holding all else constant, more earnings will be allocated to persons reporting more weeks of work than to those reporting fewer weeks of work. In contrast, for whole imputes, information about weeks worked is only available for one month.
[21] Before 1962, income data was only published for persons with no imputes.
[22] Nearly 75 percent of persons with positive income on the ASEC received earnings.
[23] Each DER record was assigned a personal identification key (PIK) which was used in matching with an ASEC person record. The 2006 ASEC contained a total of 208,562 person records of which 99,350 matched to a DER record of a person aged 15 or older (the universe for ASEC income). There were 107,532 ASEC person records for persons 15 and older with earnings. For detailed information on matched file, contact edward.j.welnick.jr@census.gov.
[24] The definition of earnings for the DER data is: max(Box1Comp+Deferred, Box3Wages) + max(Medicare_SE, FICA_SE). The ASEC variable used was PEARNVAL.
[25] There were 169,987 ASEC records where at least one family member age 15 and older matched to a DER record. These were used in the poverty tables.
[26] The presentation at the Federal Committee on Statistical Methodology's conference on January 12, 2012 presented unit counts for the basic matches.
[27] Anya Olsen and Russell Hudson, *Social Security Administration's Master Earnings File: Background Information,* Social Security Bulletin, vol. 69, No. 3, 2009, page 41, Footnote 6, states: "The major groups that are not covered include civilian federal employees hired before January 1, 1984; railroad workers, certain employees of state and local governments who are covered under their employers' retirement system; domestic and farm workers whose earnings do not meet certain minimum requirements; and persons with very low net earnings from self-employment." Employers pay some wages underground and employees in many service industries do not report all of their wages to Social Security. According to Martha Stinson of the Census, the DER file provided to Census also excludes health insurance premium amounts paid by employees that are excluded from taxable income. (E-mail on 1/30/3012).
[28] Analysis of persons who did not match with the DER is needed to complete the picture.
[29] There have been other efforts to compare survey and administrative data. See: Amy O'Hare, *Allocated Values in Linked Files,* U.S. Bureau of the Census, Federal Committee on Statistical Methodology, 2004 and Roemer, op.cit, footnote 7.
[30] See: Fritz J. Scheuren and H. Lock Oh, *A data analysis approach to fitting squared tables, Communications in Statistics – Theory and Methods,* 4: 7, 595 – 615.
[31] Measure proposed by Fritz Scheuren.
[32] Op.cit. footnote 27.

[33] Marc Roemer, Assessing the Quality of the March Current Population Survey and the Survey of Income and Program Participation, Income Surveys Branch, Housing and Household Economic Statistics Branch, U. S, Census bureau, June 16, 2000, http://www.census.gov/hhes/www/income/publications/assess1.pdf.

[34] Op.cit, This data set currently covers trends in imputation from income years 1976 through 20010. Detailed demographic data has also been developed at five year intervals starting in income year 1977 through income year 2007.

[35] Text by Fritz Scheuren. See: Oh, H. Lock, and Fritz Scheuren. "Weighting Adjustment for Unit Nonresponse." Incomplete Data in Sample Surveys. Vol. 2. New York: Academic, 1983 Variance Estimation. New York: Springer, 2007.

[36] Wolter, Kirk M., Introduction to Variance Estimation. New York: Springer, 2007.

[37] Ansolabehere, S., and T. Belin, "Poll Faulting." Chance 6.1 (1993).

[38] Rubin, D.B (1978) "Multiple Imputation in Sample Survey- A Phenomenological Bayesian Approach to Nonresponse," in Proceedings of the Survey Research Method Section, American Statistical Association, pp. 20-44.

[39] Shao, Jun and Qi Tang, "Random Group Variance Estimators for Survey Data with Random Hot Deck Imputation", Journal of Official Statistics, 27.3 (2011): 507-26. Print.

[40] Hansen, Morris H., William N. Hurwitz, and William G. Madow. Sample Survey Methods and Theory. Vol. 1. New York: John Wiley, 1953.