.

Effects of Imputation on CPS Income and Poverty Series: 1981 – 2007[1]

Joan Turek[1], Fritz Scheuren[2], Brian Sinclair-James[1]
Bula Ghose[1], Sameer Desale[3]

[1]Department of Health and Human Services, 200 Independence Ave SW, Washington, DC 20201
[2]NORC University of Chicago, 1155 E 60th St, Chicago, IL 60637
[3]Synectics for Management Decisions, Inc., 1901 N Moore St, Arlington, VA 22209

Abstract

The Current Population Survey (CPS) has been the source of the official U.S. poverty estimates since their inception. Since then, many changes have occurred in society and in the willingness of CPS respondents to report the cash income used to construct these poverty figures. Improvements have also occurred in the CPS data collection instruments. We believe the series of annual CPS cross-sections provide a meaningful measure of changes in poverty. The current paper examines trends in poverty rates by type of imputation from 1981 until 2007 focusing particularly on how poverty series for reporters and for those with item and whole imputes have trended over this period. Differences for blacks and whites are also presented. We have undertaken a detailed reconstruction of imputation and poverty series for the period 1976 to 2007 and this paper provides a beginning of the analysis of these data.

Key Words: imputation, income, poverty, Current Population Survey, methodology

1. Introduction

The Annual Social and Economic Supplement (ASEC) to the Current Population Survey (CPS) is the source of official Federal statistics on poverty. U.S. poverty estimates based on the poverty thresholds originally developed by Molly Orshansky of the Social Security Administration in 1963-1964[2] are published annually.

Since the mid-1960's, many changes have occurred in our society and in the degree to which eligible CPS respondents are willing to report the cash income needed to construct these poverty figures. Many improvements have occurred in the CPS as well. Still, the series of annual cross-sectional poverty estimates are believed to provide a meaningful measure of change in U.S. poverty over time. A systematic review of this conjecture, though, seemed due and this paper begins that process.

Why do we feel such a systematic review is in order? The principal reason is the more than doubling in CPS income "missingness"[3] since 1981, and the consequent increase in CPS imputation. How has this added imputation impacted poverty rates during the period? We felt obligated to find out particularly given the central role that mitigating poverty has played in Federal income transfer and health programs. While the Census Bureau hot deck imputation procedures have known strengths, it is still possible that the

.

growing non-response during this period may have altered historical trends in the characteristics of the poor.

How do we go about tackling such a big undertaking? Well, we began by re-examining the CPS public use files and documenting changes in the survey processing done over the years. Assembling the available documentation and improving its accessibility, readability and interpretability were and are still a main project goal. Recent advances in the creation of metadata systems at the Census Bureau and elsewhere motivate us. We want, when done, to link up the early work of Molly Orshansky with the work now continuing and to be confident that the long historical series is a meaningful yardstick of the state of America's poor.

Now, frankly we expect to be at this big task for some time to come. Still, even during these "early days," we think that you may find some of our efforts and conjectures informative. Perhaps the natural place to begin is with how poverty is defined in the CPS. We then go on to discuss the treatment of CPS income non-response. A "peek" at the extensive data we have developed covering this time period is presented next. Finally trends in poverty rates classified by imputation type and race are examined. More about the impact of income non-response on the characteristics of the poor will be covered in later papers but we preview that here too. Some technical details are covered in the extensive footnotes and appendix.

2. Defining Poverty and the Official Measure

During the early 1960's, public policy began to reflect a growing belief that poverty resulted more from societal failure rather than primarily from the unwillingness of the poor to be gainfully employed. Michael Harrington's book, *The Other America: Poverty in the United States*, influenced both the Kennedy Administration and the Johnson Administration's War on Poverty[4].

Efforts to create public programs helping the poor required some method of estimating who is poor and needing assistance. Molly Orshansky's poverty thresholds, which she characterized in her 1965 Social Security Bulletin article as income inadequacy thresholds, provided this measure. They were based on the Department of Agriculture's economy food plan and were adjusted for family size, farm/nonfarm status, sex of the family head, number of children and by age for families of size one and two. Food costs were multiplied by three to reflect other needs.

These thresholds, adjusted by the Consumer Price Index since 1969, were, with minor modification, used until 1981, to develop the annual Official Poverty Thresholds applied to the CPS income supplement each year to estimate US poverty. In 1981, the farm/nonfarm and female/male headed distinctions were eliminated and family size extended to nine or more persons. Several other efforts to revise the poverty thresholds have been proposed since then but have not been implemented[5].

.

3. Perspective on CPS Income "Missingness"

Missing data are always present 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 CPS income missingness may not be completely random. If these other variables are not properly accounted for, bias can result[6].

There are many statistical goals in imputation. The one we find particularly important is the extent that imputation reduces the missingness bias in CPS survey estimates. This goal is met to the degree that patterns of non-response are correctly identified and corrected for[7]. In the CPS there are two basic patterns of missingness – where some but not all of the CPS income questions are missing (Item Imputes) and where the entire CPS supplement is missing (Whole Imputes).

• Item Imputes. Sample persons or other household members who fail to respond to a specific question and "item" imputation is performed.

• Whole Imputes. Sample persons who only responded to the basic labor force questions. In this case, the "whole" or entire supplement has to be imputed.

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. For Whole Impute cases, where the whole supplement is missing, there are fewer variables to match on and thus we believe the chances of finding a way to reduce the missingness bias are likely to be less. In either case, after imputation a complete data set is created.

The Census Bureau started regularly imputing for missing CPS income in 1962. Since then, the same basis strategy, "hot deck" imputation, has been employed. 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[8]. Different types of missingness are treated differently. Item (partial) imputes are based on responses both to the entire monthly survey and the ASEC supplement, while whole imputes are only based on the monthly survey[9].

During the time period being considered, calendar years 1981 to 2007, the treatment of income changed significantly. A new processing system was introduced in the 1988-89 period. This revision expanded the editing and imputation process to reflect a substantial expansion in the number of income sources collected on the March questionnaire; to impute all supplemental non-interviews from one source, and to retain all reported data during the non-respondent/respondent match process.

The processing system in use before that time was implemented in 1976 and did not reflect the extensive changes to the income questionnaire in March of 1979[10]. The new questionnaire identifies over 50 sources of income and the recording of 27 different amounts. Prior to 1979, 11 income categories were collected. Due to resource constraints, the expanded income categories were collapsed back to the 11 categories for processing. The loss of information impaired the potential for doing a complete job of imputation.

.

For example, dividends, rents and royalties are asked separately on the new questionnaire. When collapsed, a yes to any one of these income sources would result in this category being treated as fully reported. With minor change, the processing system implemented in 1989 is still being used[11].

While the March 1989 CPS file was the first to reflect the new processing system, the March 1988 CPS file was reprocessed to provide estimates for income year 1987, permitting evaluation of the effects of this change. Thus, we present two estimates for income year 1987. While our estimates of total imputation rates compared to those in Technical Paper 54 for 1983[12] were highly similar, there were significant differences in the split between item and whole imputes. For this reason, separately identifying item and whole imputes is limited to estimates from income year 1988 forward[13].
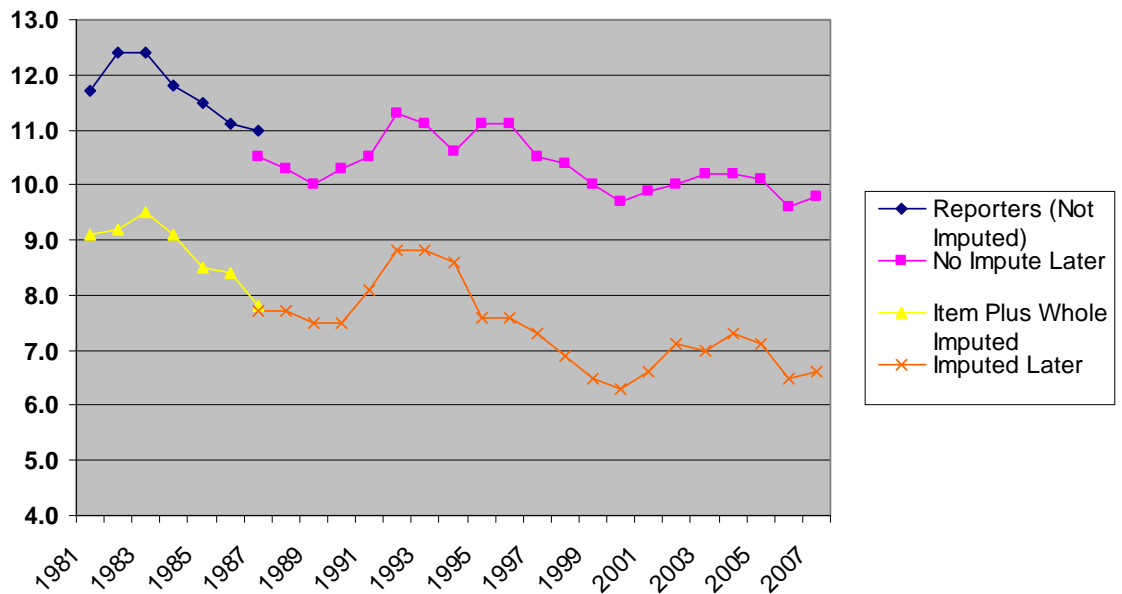
4. Trends in Poverty Rates

Figure 1, Poverty Rates: Persons with Positive Income, shows trends in poverty rates for persons with positive income from calendar year 1981 through calendar year 2007. During this time period, poverty rates for persons with positive income trended downward, falling from 10.94 percent in 1981 to 8.26 percent in 2007, but reached their highest level of 11.56 percent in 1983.
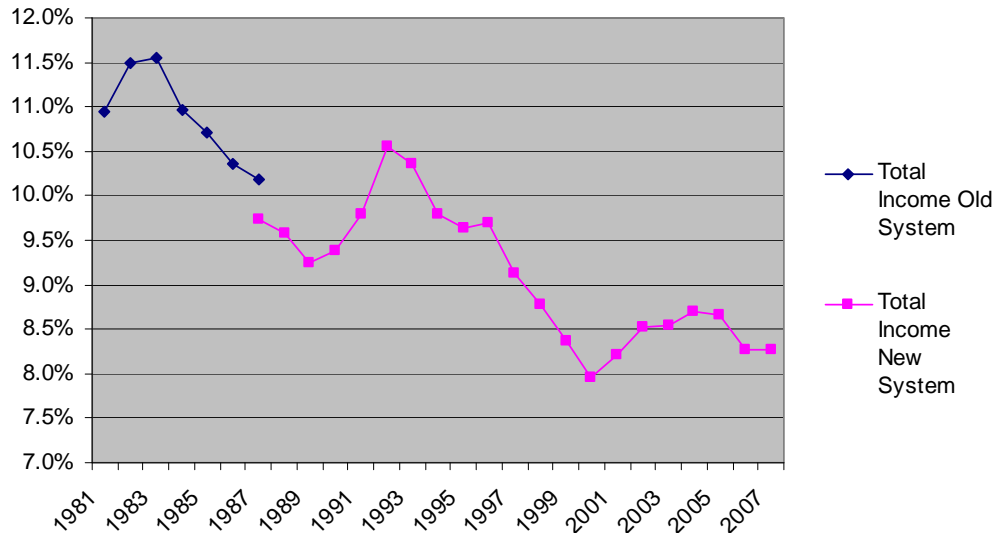
**Figure 1:** Poverty Rates: Persons with Positive Income

Figure 2, Poverty Rates, Reporters and Imputers Compared", shows poverty rates separately for reporters (persons with no imputation needed) and for imputers (persons with item and whole imputes). Over most of the time period, trends in poverty rates for reporters and imputers paralleled each other; however, poverty rates for imputers tended to be 2.5% to 3% lower than for reporters from year to year.

Figure 2: Poverty Rates, Reporters and Imputers Compared

A more detailed look at trends in poverty rates separately for persons with item, whole or no imputes (reporters) is presented in Figure 3, Poverty Trends by Types of Missingness 1987-2007, for the time period 1987-2007 under the new processing system. The pattern of item, whole and no imputes are parallel over time. As seen in Chart III, 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.

Whole imputes do not reduce poverty much because there is no information about the previous year's labor force status of these persons and they are drawn from everyone in the imputation pool. If full-year workers are more likely to need imputation, there is no way to reflect this in the whole imputation system. It can be reflected in the item imputation system, as more characteristics are available for use in doing the imputation. As seen in the appendix, zeroing out item imputes for persons with positive income in 2007 dramatically increased poverty rates.
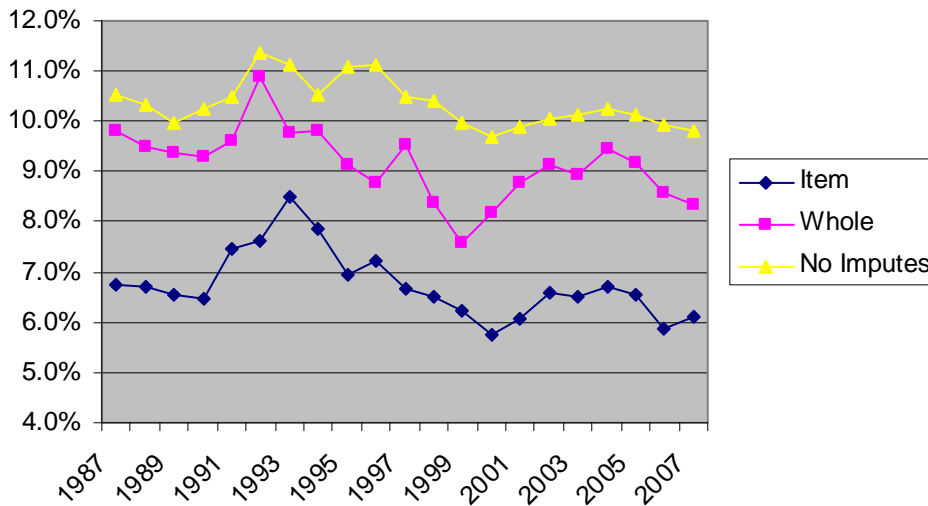
Figure 3: Poverty Trends by Types of Missingness 1987-2007

5. Defining Poverty and the Official Measure

How imputation affects demographic groups is of major policy importance. For this reason, tabulations were developed for selected demographic characteristics at five year intervals from 1987 through 2007. A detailed analysis of these tabulations is beyond the scope of the current paper; however, some preliminary analysis for Blacks and Whites is presented.
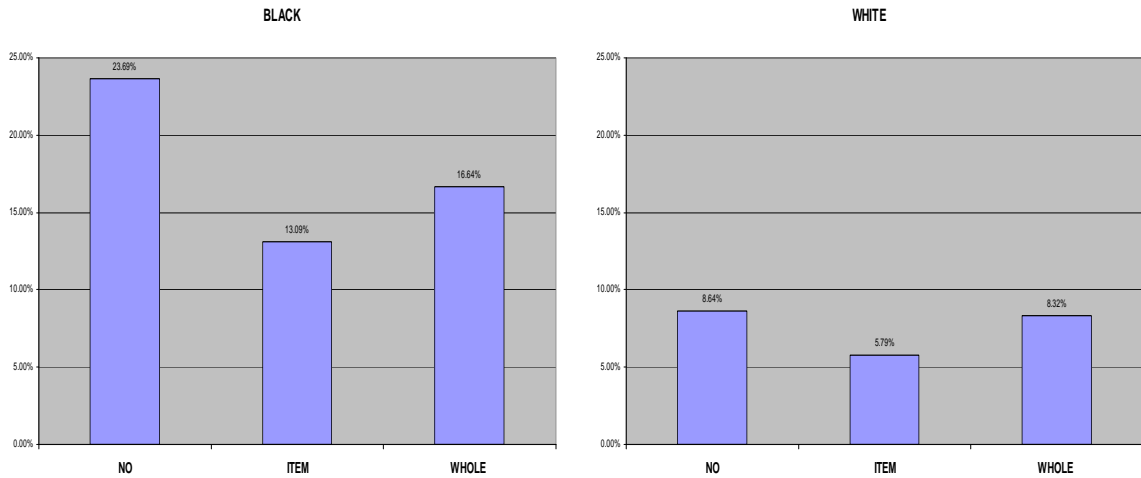
Figure 4, Black and White CPS Poor by Type of Imputation: 1987-2007, provides average estimates of item, whole and no imputation for Blacks and Whites during this period. Averages are constructed by summing each imputation category across selected years and then dividing by the number of years (five). While whole imputes are about the same percent for Blacks and Whites, 11% vs. 10%, there is significantly more item nonresponse for Whites than for Blacks, 24% vs. 18%. The higher item nonresponse is consistent with the more complex affairs and higher overall income levels of Whites.
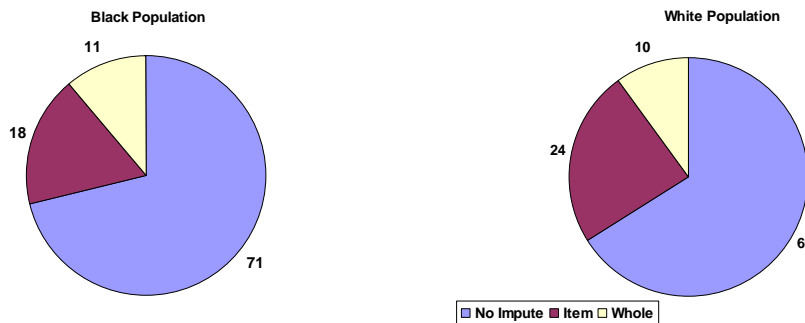
Figure 4: Percent Black and White CPS Poor by Type of Imputation: 1987-2007

When poverty rates are compared in Figure 5, Black and White Poverty Rates by Type of Imputation, Black poverty rates for persons with no imputes are more than double those of whites, 23.7% vs. 8.6%. The pattern of poverty rates for Blacks and Whites is similar

.

to that in Figure 3 for all persons with positive income. Poverty rates for Blacks for the sum of no imputes plus item imputes are considerably higher (21.6%) than for Whites (16.6%).

Figure 5: Black and White Poverty Rates by Type of Imputation

6. Interpretation of Time Series

The three series we have been looking at tell a similar story but they are open to two further perspectives: Why were the two series with missing data parallel? Was this because only some of the known bias in the missingness[14] (e.g., Scheuren et al. 1980) was "corrected" by the imputation? We frankly do not know.

There have been recent CPS efforts to compare survey and administrative data[15] (e.g., O'Hare 2004) but these do not address poverty directly and so are only partially of value in firming up or rebutting the working hypotheses we began this exercise with. We did not expect that poverty rates for the whole impute series and for the combined (item plus no impute) series would be so similar. For the time period 1987 to 2007, the ratio of poverty rates for whole imputes to item plus no imputes varied from of low of .9 in 1999 to a high of 1.1 in 2004. But this makes sense, given how weak the variables are to match whole imputes to the rest of the sample. Maybe another approach to handling the whole imputes should be tried to see what the two distributions might be? This is implicit in our discussion of next steps.

7. Next Steps and Future Directions

This is the first paper to undertake analysis of the complex data set developed for analyzing trends in imputation from calendar 1976 to 2007. A variety of future efforts are planned. These include a more complete analyses of imputation trends since 1976, analysis of demographic characteristics beyond race that target persons of policy interest, further examination of bias adjustment due to imputation, and further exploration of the relationship between bias reduction and observed differences in poverty rates by imputation status.

.

8. Appendix

That item imputes have a greater effect than whole imputes in reducing poverty is shown in Table 1 below where poverty rates are presented for the population of persons in 2007 whose positive item imputes have been set equal to zero. As can be seen, poverty rates increase dramatically when item imputes are set equal to zero. Poverty rates increase from 8.2% to 35.1% for all persons with positive income and from 6.1% to 51.7% for persons whose positive item imputes have been set equal to zero[16]. When whole imputes are excluded, the poverty rate is 8.3%. By comparison, the poverty rate for persons with whole imputes was 8.4%. 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 account for around 80% of total income.

Table 1: Comparison of Poverty Rates in CY 2007: With and Without Item Imputes
Population: All Persons with Positive Income (including item imputes)

Poverty Rate	All Persons Positive Income	All Persons, Item Imputes Set to Zero	Item Imputes	Item Impute Set to Zero
Under 100%	8.2%	35.1%	6.1%	51.7%
100 to 200%	16.5%	13.2%	13.4%	7.9%
GT 200%	75.3%	51.7%	80.6%	40.5%

Acknowledgements

References

1. The views are those expressed by the authors and are not the official position of any of their organizations.
2. Social Security Bulletin, January 1965
3. "Missingness" is used as a term of art in this paper to mean a failure by a survey respondent to fully report their income. It is defined more fully in the section on handling survey non-response.
4. Michael Harrington, *The Other America: Poverty in the United States* (New York: Macmillan, 1962.
5. Gordon M. Fisher, *The Development and History of the U.S. Poverty Thresholds --- A Brief Overview*, U.S. Department of Health and Human Services, http://aspe.hhs.gov/poverty/papers/hptgssiv.htm)

.

6. H. Lock Oh, Social Security Administration, Fritz Scheuren, Internal Revenue Service, Hal Nisselson, WESTAT, *Differential Bias Impacts of Alternative Census Bureau Hot Deck Procedures for Imputing Missing CPS Income Data*, Papers and Proceedings, Joint Statistical Meetings, 1980.

7. U.S. Census Bureau, Survey of Program Dynamics, Data Editing and Imputation Goals, http://www.census.gov/spd/goals.html)

8. *Technical Paper 54: Income Nonresponses: March 1983 CPS*, September 1985, U.S. Department of Commerce, Bureau of the Census, Washington D.C and Edward J Welniak, Effects of the March Current Population Survey's New Processing System on Estimates of Income and Poverty, Bureau of the Census, Washington D.C. 20233.

9. Noninterview nonresponses, 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.

10. The new questionnaire was fully adopted in 1980 with minor modification.

11. Op.cit, Welniak. This discussion relies heavily on his paper.

12. Op.cit. Based on date in Table A, p4.

13. Minor differences occurred since we excluded persons with negative income from our analysis this failure to exclude all persons in families containing a person with negative income has a slight, virtually insignificant impact on poverty rates. Poverty rate changes for 1983 and 1997 showed a less than one percent difference and do not materially affect the results presented in this paper.

14. Op.cit

15. Amy O'Hare, *Allocated Values in Linked Files*, US Bureau of the Census, presented at the Federal Committee on Statistical Methodology.

16. The population being considered is all persons who have positive income including item imputes.