Bridging a Survey Redesign Using Multiple Imputation: An Application to the 2014 CPS ASEC

Jonathan Rothbaum[1]

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. In 2014, the CPS ASEC questionnaire was redesigned to improve data quality and to reduce misreporting, item nonresponse, and errors resulting from respondent fatigue. The sample was split into two groups, with nearly 70% receiving the traditional instrument and 30% receiving the redesigned instrument. Due to the relatively small redesign sample, analyses of changes in income and poverty between this and future years may lack sufficient power, especially for subgroups. The traditional sample is treated as if the responses were missing for income sources targeted by the redesign, and multiple imputation is used to generate plausible responses. A flexible imputation technique is used to place individuals into strata along two dimensions: 1) their probability of income recipiency and 2) their expected income conditional on recipiency for each income source. By matching on these two dimensions, this approach combines the ideas of propensity score matching and predictive means matching. In this article, this approach is implemented, the matching models are evaluated using diagnostics, and the results are analyzed.

Key words: Multiple imputation; survey redesign; bridge; CPS ASEC.

1. Introduction

The Current Population Survey Annual Social and Economic Supplement (CPS ASEC) is among the most widely used surveys conducted by the U.S. Census Bureau. CPS ASEC data are used to calculate measurements of national income and the official poverty rate. Rothbaum (2015) shows that the CPS ASEC suffers from underreporting of certain income types, including property income (especially interest and dividends), retirement income, and income from means-tested government transfer programs. Meyer et al. (2009) also show underreporting of participation in means-tested government programs.

To address this underreporting, the U.S. Census Bureau, in consultation with the private sector, implemented a redesign of the survey (see Czajka and Denmead (2008) and Hicks and Kerwin (2011) for results of that consultation). In 2014, approximately 30% of the CPS ASEC sample received the redesigned survey instrument, and approximately 70%

received the unchanged traditional instrument (in use since 1994). Assignment into the two groups was random at the household level. For more details about the redesign and the content tests, see Semega and Welniak (2013, 2015).

A major focus of the redesign was to improve reporting of property income, especially income earned from assets in the form of interest or dividends. In addition, since 1980, the nature of retirement savings has shifted from defined benefit to defined contribution plans. From 1980 to 2008, the share of private wage and salary workers with defined benefit plans fell from 38% to 20%. The share of private workers with defined contribution plans grew from eight percent to 31% over the same period (Butrica et al. 2009). Therefore, the survey was redesigned to improve reporting of retirement income, which has also historically been underreported (Czajka and Denmead 2008).

The redesigned instrument is being used for the full sample, starting with the 2015 CPS ASEC. However, in order to make apples-to-apples comparisons between the results in 2014 and 2015 and beyond, only 30% of the 2014 sample can be used. This significantly reduces the power of the comparisons that can be made, for example of median income or poverty rates, and is especially relevant for subgroups.

While the survey redesign significantly increased recipiency and aggregates for many income types, the majority of income (by US dollars) was not affected. For example, earnings comprised 75.9% of all income in the redesign sample, and there were no statistically significant differences in the number of earners or mean earnings across the two instruments. Although we do not observe what respondents to the traditional instrument would have said to the redesigned questions, we do have a considerable amount of information about them that is unaffected by the redesign.

This suggests treating the problem as one of missing data – as if the recipients of the traditional instrument did not respond to the redesigned income questions. The "missing" responses to the redesigned questions are multiply imputed for individuals in the traditional sample. This article adds to the literature using multiple imputation to bridge a survey or data classification change. Clogg et al. (1991) used multiple imputation to impute industry and occupation codes across a change in the coding scheme between 1970 and 1980 census data. Schenker and Parker (2003) imputed single-race reporting for multiple-race respondents after a change from single- to multi-race reporting in government survey data.

An approach developed by Bondarenko and Raghunathan (2007) is applied to impute these missing responses in the traditional sample. This approach combines the ideas of propensity score matching and predictive means matching. By matching donors to recipients within propensity score/predictive mean cells, this approach is similar to the hot deck procedure used in normal CPS ASEC processing. That makes it appealing for use in this case, as the completed data from the 2014 CPS ASEC can be used to make comparisons with data in subsequent years where all imputation of missing values is done using the hot deck.

The Bondarenko and Raghunathan technique is used to create an "Income-Consistent" full file that uses all of the CPS ASEC sample with imputed income in the affected categories for respondents to the traditional instrument. It is called the Income-Consistent file, as the responses for all individuals are consistent with the questions in the redesign survey instrument.

The article is organized as follows. In Section 2, the CPS ASEC and the survey redesign are described. Section 3 discusses the imputation methodology. Section 4 discusses

diagnostic results to evaluate the models used. Section 5 contains results relating to income and poverty, measurement using the imputed data. Section 6 contains a conclusion.

2. Data and Survey Redesign

The CPS ASEC is among the most widely used surveys conducted by the U.S. Census Bureau. The CPS ASEC uses a stratified random sample to survey about 100,000 households each year and includes questions on income and health insurance coverage.

The 2014 survey redesign included a number of changes. First, the survey was redesigned to specifically ask if anyone in the household has a pension, and separately if anyone has a retirement account (401(k), 403(b), IRA, or other account designed specifically for retirement savings). The traditional instrument includes one broad question on the receipt of pension and retirement income. The redesigned instrument also asks individuals over 70 years old about required distributions from retirement accounts. To ensure that the distribution is correctly identified as income, a follow-up question asks if the required distribution was "rolled over" or reinvested in another account. The traditional ASEC instrument makes no distinction between investment income received in retirement accounts or separately from them. This more detailed set of questions can improve misreporting of income and cue respondents to decrease underreporting.

Several additional changes were also made to the survey. Prior to the redesign, only households that reported less than USD 75,000 in combined family income were asked questions about means-tested transfer programs such as Temporary Assistance to Needy Families (TANF). Semega and Welniak (2015) cite evidence from the American Community Survey (ACS) that some screened households were likely to be recipients of these transfers making it inappropriate to remove them using the income screener. To prevent respondent fatigue from affecting answers to the income recipiency questions, the recipiency questions were separated from the amount questions as part of a "dual-pass" approach. Respondents were asked first about all sources of income received and then later were asked about amounts for only the received sources. In addition, the order of the income questions was changed based on respondent characteristics to match those sources most likely to be received. If a respondent was unsure of the income generated from assets, the value of the assets was collected. The questions on disability were clarified to eliminate confusion between disability income from Social Security and Supplemental Security Income (SSI).

2.1. Results of the Redesign

In 2014, the CPS ASEC sample was randomly divided into two groups at the household level, with 31% (30,000 housing units) receiving the redesigned instrument and about 69% (68,000 housing units) receiving the traditional instrument. Within each sample, individual observations were weighted to national population controls, as is standard with the CPS ASEC. Both surveys were conducted primarily by home visit (with some by phone) by trained field representatives. Even the interviewer was not aware of the selection of a given household into the traditional or redesign sample until they began the survey.

Semega and Welniak (2015) compared income aggregates between the two samples. Table 1 shows a subset of their results for median income, updated to reflect recent edits of the redesign sample file. Household median income was USD 51,939 in the traditional sample and USD 53,585 in the redesign, a difference of 3.2%. When decomposed by race, the only statistically significant differences are for whites (and non-Hispanic whites).

Table 2 shows income statistics for total income and various income sources collected in the CPS ASEC. For each source, Semega and Welniak report the number of recipients in the population, the mean income earned by those recipients, and the aggregate value of that income estimated, using the traditional and redesign samples separately. For example, for total income, the number of income recipients estimated using the traditional sample is 218.7 million compared with 222.0 from the redesigned sample, a statistically significant difference of 1.5%. The estimated difference in mean total income is 2.6% (USD 41,319 in the traditional vs. USD 42,394 in the redesign), and the estimated difference in aggregate total income is 4.2% (USD 9.04 trillion in the traditional compared with USD 9.41 trillion in the redesign), both statistically significant. At the 90% confidence level, there are a number of income sources that have statistically significant differences in the number of recipients, mean income, or aggregate income. The sources with statistically significant differences in aggregate income include farm self-employment income (-42.1%), public assistance (28.8%), veterans' benefits (-23.1%), disability benefits (36.4%), retirement income (21.9%), interest (113.0%), and dividends (-20.1%).

Mitchell and Renwick (2015) study the effects of the redesign on poverty rates. While they find no statistically significant difference in the overall poverty rate, they do find differences for child and elderly poverty in the redesigned sample. In both cases, they suggest that differences in the sample populations may explain the increase in poverty in the redesigned sample. For child poverty, they show that the redesigned sample has a higher share of children living with female householders than the traditional sample (single-mother families). They also find that means-tested program recipiency was higher in the redesigned sample.

These potential differences in sample characteristics support the approach taken in this article. Because the changes in the questionnaire are treated as a problem of missing information, any differences in the samples can be controlled for as a part of the imputation modeling, and the combined sample should better reflect the intended full CPS ASEC sample.

2.2. *Selection of Income Sources to be Imputed*

Taking these analyses together, the redesign increased aggregate income, increased income recipiency and reporting in a number of income categories. However, some of the differences, especially in income types with no or little change in the questionnaire, may be due to random variation or differences in the samples. This is supported by differences in poverty that Mitchell and Renwick attribute to sample differences.

Because of this evidence of sample differences, the analysis focuses on those income types which were targeted by the questionnaire redesign. This eliminates farm self-employment, and veterans' benefits.

The income types that are sufficiently different between the two surveys and were specifically targeted by the questionnaire redesign include: 1) retirement income,

Table 1. Comparison of traditional and redesign samples: median income.

| | Traditional | | | | Redesign | | | | Percentage change $\left(\frac{R-T}{T}\right)$ | |
| | Number (thousands) | | Median income (USD) | | Number (thousands) | | Median income (USD) | | | |
Characteristic	Estimate	90% CI	Estimate	90% CI	Estimate	90% CI	Estimate	90% CI	Estimate	90% CI
All households	122,952	723	51,939	455	123,931	942	53,585	1,076	3.2*	2.08
Race[1] and Hispanic origin of householder										
White	97,774	605	55,257	699	98,807	756	56,745	850	2.7*	1.81
White, not Hispanic	83,641	544	58,270	1,066	84,432	732	60,329	876	3.5*	2.04
Black	16,108	262	34,598	1,198	16,009	355	35,324	1,410	2.1	5.13
Asian	5,759	151	67,065	2,830	5,818	215	72,383	5,531	7.9	7.92
Hispanic (any race)	15,811	210	40,963	908	16,088	354	39,687	1,953	-3.1	5.00
Earnings of full-time year-round workers										
Men with earnings	60,769	600	50,033	404	61,240	787	50,015	935	-0.4	1.23
Women with earnings	45,068	510	39,414	596	44,629	659	38,792	1,145	-0.9	3.19

Notes: Income in 2013 dollars. Households and people as of March of the following year. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see www.census.gov/prod/techdoc/cps/cpsmar14.pdf. * indicates statistically different from zero at the 90-percent confidence level. A 90-percent confidence interval (CI) is a measure of an estimate's variability. The larger the CI is in relation to the size of the estimate, the less reliable the estimate. CIs shown in this table are based on standard errors calculated using replicate weights. For more information, see "Standard Errors and Their Use" at <www.census.gov/hhes/www/p60_245sa.pdf>.

[1] Federal surveys now give respondents the option of reporting more than one race. Therefore, two basic ways of defining a race group are possible. A group such as Asian may be defined as those who reported Asian and no other race (the race-alone or single-race concept) or as those who reported Asian regardless of whether they also reported another race (the race-alone-or-in-combination concept). This table shows data using the first approach (race alone). The use of the single-race population does not imply that it is the preferred method of presenting or analyzing data. The U.S. Census Bureau uses a variety of approaches. Information on people who reported more than one race, such as White and American Indian and Alaska Native or Asian and Black or African American, is available from Census 2010 through American FactFinder. About 2.9 percent of people reported more than one race in Census 2010. Data for American Indians and Alaska Natives, Native Hawaiians and Other Pacific Islanders, and those reporting two or more races are not shown separately in this table.

Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplement.

Table 2. *Comparison of traditional and redesign samples: income recipiency, mean income, and aggregate income by source.*

Type of income	Traditional Number (thousands)		Traditional Mean income (USD)		Traditional Aggregate income (thousands)		Redesign Number (thousands)		Redesign Mean income (USD)		Redesign Aggregate income (thousands)		Percent change $(\frac{K-T}{T})$ Number		Percent change Mean income		Percent change Aggregate income	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Total income	218,662	311	41,319	279	9,035,004	60,841	222,003	418	42,394	345	9,411,655	79,158	1.5*	0.2	2.6*	1.1	4.2*	1.1
Earnings	158,081	489	44,416	334	7,021,280	58,106	158,655	638	44,999	438	7,139,254	74,929	0.4	0.5	1.3	1.2	1.7	1.3
Wages and salary	148,752	492	44,931	336	6,683,647	55,005	149,546	684	45,695	457	6,833,462	75,382	0.5	0.5	1.7	1.2	2.2	1.4
Nonfarm self-emp	8,702	190	35,145	1,215	305,825	12,033	8,508	298	33,777	1,411	287,376	15,300	−2.2	4.1	−3.9	5.6	−6.0	6.2
Farm self-emp	627	59	50,728	6,870	31,808	5,359	601	73	30,662	4,448	18,416	3,278	−4.2	13.7	−39.6*	12.4	−42.1*	14.3
Unemployment	6,818	165	5,841	151	39,825	1,362	6,435	233	5,870	167	37,768	1,806	−5.6	4.1	0.5	4.1	−5.2	5.6
Workers' comp	1,186	60	9,224	566	10,940	930	952	77	10,156	851	9,671	1,135	−19.7*	7.3	10.1	10.5	−11.6	11.4
Social security	48,370	332	13,979	55	676,178	5,142	49,055	418	14,052	84	689,325	6,782	1.4	1.0	0.5	0.7	1.9	1.2
SSI	6,053	176	7,782	105	47,104	1,459	6,642	230	7,728	148	51,333	2,021	9.7*	5.0	−0.7	2.3	9.0	5.5
Public assistance	1,775	81	3,195	149	5,671	340	2,189	124	3,337	160	7,305	508	23.3*	8.8	4.4	6.5	28.8*	10.9
Veterans' benefits	3,517	127	14,640	424	51,493	2,503	3,296	164	12,021	584	39,619	2,757	−6.3	5.1	−17.9*	4.8	−23.1	6.1
Survivors' benefits	3,033	110	12,972	559	39,340	2,260	2,970	156	14,526	995	43,139	3,502	−2.1	6.5	12.0	8.6	9.7	11.0
Disability benefits	1,771	76	15,543	736	27,524	1,850	3,099	163	12,110	635	37,535	2,654	75.0*	11.8	−22.1*	5.6	36.4*	12.8
Retirement income	18,871	251	20,034	307	378,054	7,865	20,698	372	22,262	449	460,784	12,668	9.7*	2.4	11.1*	3.0	21.9*	4.2
Interest	86,142	588	2,120	68	182,619	5,963	123,772	887	3,142	107	388,943	13,403	43.7*	1.4	48.2*	6.8	113.0*	9.7
Dividends	33,243	432	4,424	170	147,050	6,225	31,804	568	3,693	211	117,454	6,954	−4.3*	1.9	−16.5*	5.8	−20.1*	5.6

*Statistically different from zero at the 90-percent confidence level. Standard errors are calculated using replicate weights. For more information, see "Standard Errors and Their Use" at <www.census.gov/hhes/www/p60_245sa.pdf>.

Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplement.

Absolute difference in aggregate income |redesign-traditional|
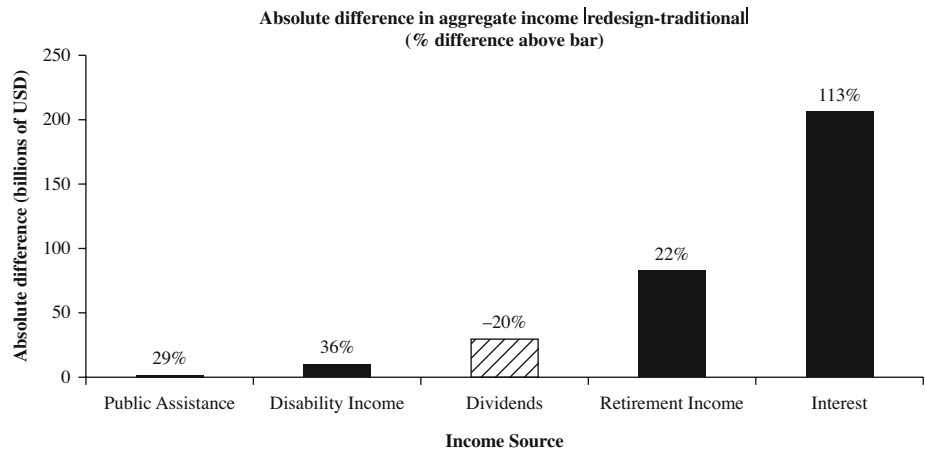**(% difference above bar)**

Fig. 1. Aggregate income differences between the traditional and redesign samples. Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplement.

2) interest, and 3) dividends. These three sources had the largest difference in estimated aggregate income of the types affected by the redesign. Fig. 1 shows changes in aggregate income for all income sources with a statistically significant difference in aggregate income between the traditional and redesign samples. For interest income, the number of recipients increased by 37.6 million and aggregate income increased by USD 206.3 billion. For retirement income, the number of recipients increased by 1.8 million and aggregate income increased by USD 82.7 billion. For dividend income, the number of recipients decreased by 1.4 million and aggregate income decreased by USD 29.6 billion.

3. Imputation Methodology

3.1. Hot Deck Imputation

As a part of the standard processing of the CPS ASEC, when an individual does not respond to a particular question, missing values are imputed using a hot deck procedure. In the hot deck, individuals are divided into cells based on the characteristics specified in the hot deck model. Within each cell, individuals without missing information (donors) are randomly selected and their income is assigned to the individuals with missing information (recipients). Donors and recipients in each cell must match on every variable in the hot deck model. If there are no donors in a given recipient's cell, the hot deck model is amended to reduce the number of categories for some variables (for example from nine age groupings to six) and to reduce the number of variables in the model.

The different hot deck models used in the CPS ASEC are called match levels. The first match level includes the largest number of variables and categories within each variable. If no matches are found at the first level, an attempt to match recipients and donors is made using the model at the second match level. This continues until a match level is reached for a given recipient in which at least one donor is present in the same cell. For missing earnings in the longest job, at the first match level there are 16 variables in the model and 621 billion possible cells; at the second match level there are 14 variables and 17 billion

possible cells; at the third match level there are eleven variables and 3.8 million possible cells, and by the sixth match level there are four variables and 96 possible cells. In the traditional sample for those observations missing earnings from the longest job only, 4.4% matched at the first level, 13.0% matched at the second level, 51.5% matched at the third level, and 6.4% matched at the sixth level. The variables and number of categories at each match level are available in Supplemental data online (URL: http://dx.doi.org/10.1515/JOS-2017-0010), Table 1.

As these numbers make clear, the number of variables that can be included in a hot deck model is limited by the size of the sample. While this is clearly a constraint even in the full CPS ASEC sample of about 200,000 individuals, the constraint is even more binding when imputing income from the redesign sample of about 60,000 individuals. If retirement, interest, and dividend income in the traditional sample were imputed using the hot deck model, it would not be possible to incorporate many variables in the model that are potentially correlated with each income type. This would limit the ability of the imputation to accurately match similar individuals as donors and recipients and reduce the quality of the matches.

3.2. Model-Based Matching Imputation

Instead, a more flexible technique is implemented to impute the missing responses to the redesigned questions in the traditional sample for the research file. The approach, developed by Bondarenko and Raghunathan (2007), hereafter BR, matches donors and recipients using summaries of covariates estimated by logistic and ordinary least squares regression modelling.

The primary reason the BR approach was chosen in this research is its similarity to the hot deck model. As in the hot deck model, individuals are matched based on similarities in observable characteristics. In the hot deck model, matching is directly based on the characteristics. In the BR approach, the matching is based on the predicted probability of recipiency and expected income conditional on recipiency, both of which can be estimated from observable characteristics. This is advantageous, as the imputed data must be comparable to data from subsequent years where all missing data are imputed using the hot deck model. However, by efficiently summarizing the model covariates in two statistics, recipiency and expected income, the BR approach allows for the inclusion of many more variables in the imputation model.

Next, the BR method is described, with slight modification for this application. Suppose that the dataset has P variables of observable characteristics, X_p, $p = 1, 2, \ldots, P$ and $X = (X_1, \ldots, X_P)$ and Q income types where Y_q, $q = 1, 2, \ldots, Q$, represents the income value and R_q represents recipiency status ($R_q \in \{0, 1\}$). There are two groups in the sample, one for which the income types q are observed (group O) and one for which income types q are unobserved (group M) so that each vector can be partitioned among O and M as $X_p = \left(X_p^O, X_p^M \right)$, $Y_q = \left(Y_q^O, Y_q^M \right)$, and $R_q = \left(R_q^O, R_q^M \right)$. Because missingness is complete for all Y_q^M, income can be imputed sequentially without iteration. Therefore, $<q$ is defined as the set of incomes with indices less than q so that $Y_{<q} = \left(Y_1, \ldots, Y_{q-1} \right)$ and $R_{<q} = \left(R_1, \ldots, R_{q-1} \right)$ and $Y_{<1}$ and $R_{<1}$ are empty sets. Two efficient summaries are constructed of the income variables through two regression predictions:

1. Probability of recipiency: $\hat{R}_q = \Pr\left(R_q = 1 | Y_{<q}, R_{<q}, X\right)$ estimated using a logistic regression model. This is an efficient summary of R_q that can be used to balance income recipients and nonrecipients.
2. $\hat{R}_q$ is stratified into K equal size strata, where $k = 1, \ldots, K$.
3. Predicted value of income conditional on recipiency within each stratum k: $\hat{Y}_q = E\left(Y_q | R_q = 1, Y_{<q}, R_{<q}, X\right)$ is estimated using an OLS regression model on all individuals in stratum k. Then, individuals are subdivided in stratum k into J equal sized substrata, where $j = 1, \ldots, J$. This creates $K \times J$ equal size strata.

Within each stratum k, j there are n individuals with observed income and recipiency, and m individuals with missing income and recipiency for income type q. A sample size m is drawn from the observed set of n individuals as the imputed values by Approximate Bayesian Bootstrap (ABB). This step is repeated for each stratum k, j and income type q and then sequentially for all $q = 1, \ldots, Q$. This entire process is repeated independently to obtain multiple imputations.

This approach relies on the same assumption that underlies matching models (see, for example, Rosenbaum and Rubin 1983). The traditional sample is M in this exercise, as the responses to the redesigned questions are missing for all individuals in the traditional sample. The redesign sample is O as the responses are observed. For this approach to be valid, it must be assumed that inclusion in the traditional sample can be controlled for by the variables in the imputation model (unconfoundedness). Specifically, given the probability of missingness $P(M)$, it is assumed that $P\left(M | X, Y_q^M, Y_q^O\right) = P(M | X)$ where X can be summarized by $\hat{R}_q$ and $\hat{Y}_q$. Although the random selection into the redesign sample implies missingness should be completely at random (MCAR), Mitchell and Renwick (2015) suggest that sample differences do exist on observable characteristics. Therefore, it is conservatively only assumed that the responses are missing at random (MAR).

There are a number of challenges to implementing BR method in the CPS ASEC. First, many income types do not follow a normal distribution or any simple transformation of a normal distribution. Because the missing income sources are modeled with continuous covariates, some distributional assumptions must be made about the relationships between them. Second, predictors (X) must be selected for the modelling of each income variable from a very large set of possible covariates ($>1{,}200$) in the CPS ASEC.

As shown in Hokayem et al. (2015), the distribution of income is rarely normally distributed. Simple transformation (such as log) and more flexible ones such as Tukey's gh distribution (He and Raghunathan 2006) also can fail to convert the distribution to normal. Therefore, an empirical normal transformation proposed by Woodcock and Benedetto (2009) is used to convert all income values to normal distributions (this includes income and other continuous variables in X as well) prior to imputation.

The most significant challenge to applying the BR method to the CPS ASEC was to select the models for each imputed variable. In order to avoid omitted variable bias in the imputation model, as many potential predictors as possible should be included. However, if too many variables are included, overfitting the model is a risk. The list of potential predictors used includes all unchanged income information (imputation flag, recipiency, value), spouse/partner earnings, race (separate dummy for each), gender, age (including dummies for each age between 62 and 70), weeks worked last year, hours worked per

week, as well as the hot deck categories for relationship to householder, education level, marital status, presence of children, occupation (22 categories), type of residence, Census region, recipiency of means-tested government transfers. A large set of interaction terms are included in the list of predictors, including for major income types (earnings, spouse earnings, etc.), education, weeks and hours worked, race and age, and means-tested transfers. In all, over 1,200 potential predictors and interaction terms can be included in the BR models. A list of the modelling variables is available in Supplemental data, Table 2 (URL: http://dx.doi.org/10.1515/JOS-2017-0010).

The parameters of two models are estimated: (1) $R_q = F(X_q \beta_R + v_q)$ using logistic regression and (2) $Y_q = X_q \beta_Y + e_q$ by OLS. However, with more than 1,000 possible covariates, all possible covariates cannot be included in X and some values in β must be set to 0 in each regression.

Stepwise model selection is used to determine which values in β in each regression to set to 0. It was chosen as a pragmatic tool to efficiently capture the correlations between covariates X_q and dependent variables R_q and Y_q. However, there is uncertainty about which are the correct items in each β that should be set to zero which must be accounted for in order for the imputation to be proper. If the model variables were known with certainty (known nonzero items in β), after regressing Y on $X\beta$ parameter uncertainty could be accounted for using the variance-covariance matrix of β. However, in this case, both parameter and model uncertainty are present. In order to approximate both sources of uncertainty and have proper imputation variance estimates, for each income type, all regressions for each income type q are run on an approximate Bayes Bootstrap (ABB) sample.

In summary, the imputation steps to create the Income-Consistent file are:

1. Normal transformation – transform all income value variables to normal distribution with empirical normal transformation.
2. BR Imputation – sequentially impute interest, dividends, and retirement income from the redesign (donors/observed) to the traditional sample (recipients/missing). For each income type:
   a. Select a random sample by ABB.
   b. Predict probability of income recipiency using logistic regression on the redesign ABB sample with stepwise model selection to choose list of predictors. Only those individuals with non-imputed values of recipiency are included in the regression.
   c. Stratify the sample into K equal-sized groups based on probability of income recipiency in the original sample.
   d. Within each stratum k, predict expected income conditional on recipiency using OLS regression on the redesign ABB sample that is within the probability of recipiency bounds of that stratum.
   e. Stratify subsample k into J equal sized substrata based on the expected income of the original sample.
   f. Within each substratum j, select a random sample of m donors from the redesign sample (where m is the number of recipients with missing responses in stratum k, j) using ABB. Each donor receives all income, source, and value variables from the recipient. By donating source information (i.e., whether the retirement income is

from a 401k, IRA, or others), this implies the additional assumption that for each income type q and source S_q, $P\left(M|X, Y_q^M, Y_q^O, S_q^M, S_q^O\right) = P(M|X)$, where X can be summarized by $\hat{R}_q$ and $\hat{Y}_q$.

    g. Repeat for each stratum k, j until all missing observations for income type q have been imputed.

3. Transform to original scale – return all variables to their original scales.

4. Repeat the entire process to create ten implicates.

These steps are done after processing and allocation of the survey data. This means that hot deck imputed values in the redesign file can be used as part of the imputation process. However, all modelling and prediction is done only on actual responses with allocated values excluded from the modelling step.

Since all of the interest, dividends, and retirement income are missing for all observations in the traditional sample, the order of imputation should not matter. Consider, for example, the case where interest is imputed first and dividends second. In that case, the imputation for interest should capture the relationship between interest and all other variables in $X_{<q}$. In the imputation for dividends, information on interest is included in $X_{<q}$, which should capture the relationship between dividends and interest as well as dividends and all other variables in X. For both missing income types, the imputation captures the conditional relationship between the other type and the variables in X. The same is true if dividends are imputed first and interest second. As a result, the variables are imputed by frequency of recipiency, from most common to least common: 1) interest, 2) dividends, and 3) retirement income. Note that this invariance to imputation order is only true if missingness for each imputed variable implies missingness for all others.

4. Diagnostic Results

One way of evaluating the imputation model is to construct an R^2 from the set of regressions on the ABB sample. For the logistic regressions, the Tjur-R^2 (Tjur 2009) is used, which is calculated by comparing the average predicted probability of recipiency for those who did and did not receive income of that type, or

$$R_{\text{Tjur}}^2 = E\left(\hat{R}_q|R = 1, Y_{<q}, R_{<q}, X\right) - E\left(\hat{R}_q|R = 0, Y_{<q}, R_{<q}, X\right).$$

The Tjur-R^2 is bounded between 0 and 1.

For the OLS regressions, the R^2 used is the squared correlation between the transformed income and the predicted income from the strata regressions, shown in Table 3. The average Tjur-R^2 for interest, dividends, and retirement are 0.35, 0.30, and 0.39 respectively. The OLS R^2 values for interest, dividends, and retirement income are 0.12, 0.10, and 0.15 respectively.

The relatively low R^2 are in part due to the fact that predictions are made on ABB samples, not the original one. The regression R^2 are much higher, but they reflect the match between the predictions and the bootstrapped sample, which will by definition be higher than for the original sample, which was not used for the prediction.

After imputing interest, dividend, and retirement income responses for the traditional sample, the two samples are combined to create the Income-Consistent file, as all responses

Table 3.   Model diagnostics – effective R^2 of recipiency and value regressions.

Variable	Recipiency			Value		
	Average	Min	Max	Average	Min	Max
Interest	0.35 (0.01)	0.33	0.36	0.12 (0.03)	0.05	0.15
Dividends	0.30 (0.01)	0.28	0.32	0.10 (0.03)	0.03	0.13
Retirement	0.39 (0.02)	0.36	0.43	0.15 (0.05)	0.03	0.19

The R^2 are calculated by taking the predicted recipiency and values conditional on recipiency from the prediction models used to define the donor/recipient cells and calculating the Tjur R^2 for recipiency and squared correlation for the value. The average, minimum, and maximum effective R^2 across the 10 implicates are reported with standard deviations in parentheses.

are now consistent with the redesigned income questionnaire. Estimates are calculated for the number of recipients and mean income in the traditional, redesign, each of the Income-Consistent implicates as well as the multiple imputation estimates, shown in Table 4. Standard errors in each file or implicate are calculated using replicate weights. Throughout the article, multiple imputation standard errors for the Income-Consistent file are calculated using the multiple imputation variance formula in Rubin and Schenker (1986). There are no statistically significant differences when comparing recipiency or mean income between the redesign and Income-Consistent files for any of the three income sources.

Another statistic that can be used to evaluate the value of applying the imputation to create the Income-Consistent file is the estimated rate of missing information, denoted by γ (Rubin 1987). Very high values of γ (for example, 0.7) would imply that there is little additional benefit to using the traditional sample with imputed interest, dividend, and retirement income. As the relevance of the missing interest, dividend, and retirement income may differ for different statistics, for each parameter of interest, a separate γ can be computed.

I estimated γ for the recipiency and mean income statistics in Table 4. Recall that approximately 30% of the Income-Consistent sample comes from the redesign sample and is the same across all ten implicates. The rate of missing information varies across the income types from 0.09 (interest recipiency) to 0.53 (mean retirement income).

I also calculated γ for household median income and poverty of 0.15 and 0.08 respectively. Both of these are low values, which indicates that a considerable amount of information in estimating median income and poverty is contributed by the other variables in the traditional sample that were not imputed, as they were not affected by the redesign. For family and household income statistics, these low γ values also validate the general approach of combining the samples to take full advantage of the information available in the questions unaffected by the redesign.

5.   Income and Poverty Statistics

To further assess the Income-Consistent file, median income and poverty statistics are calculated. These statistics are in the annual Income and Poverty reports published by the

Table 4. Recipiency and mean Income in traditional, redesign, and income-consistent files.

	Retirement		Interest		Dividends	
	Recipiency (Thousands)	Mean (USD)	Recipiency (Thousands)	Mean (USD)	Recipiency (Thousands)	Mean (USD)
Traditional	18,871	20,034	86,142	2,120	33,243	4,424
	(251)	(307)	(588)	(68)	(432)	(170)
Redesign	20,698	22,262	1,23,772	3,142	31,804	3,693
	(372)	(449)	(887)	(107)	(568)	(211)
Income-Consistent Implicate #						
1	20,709	22,406	1,25,594	3,033	32,095	3,584
	(265)	(271)	(1,866)	(75)	(587)	(172)
2	20,777	22,195	1,24,925	3,160	32,045	3,669
	(292)	(377)	(1,404)	(75)	(411)	(141)
3	20,511	22,054	1,25,339	3,168	31,236	3,706
	(241)	(251)	(1,834)	(87)	(386)	(109)
4	20,395	22,383	1,25,813	3,083	31,657	3,734
	(225)	(257)	(2,152)	(54)	(368)	(96)
5	20,826	22,667	1,24,748	3,115	32,167	3,633
	(455)	(291)	(1,228)	(55)	(616)	(122)
6	20,945	22,371	1,24,858	3,087	31,357	3,923
	(428)	(287)	(1,282)	(49)	(366)	(133)
7	20,891	22,826	1,24,804	3,082	31,660	3,706
	(520)	(377)	(1,191)	(51)	(295)	(135)
8	20,659	22,252	1,25,513	3,195	31,827	3,629
	(248)	(267)	(2,129)	(75)	(404)	(129)
9	20,788	22,019	1,25,805	3,037	31,741	3,597
	(327)	(353)	(2,026)	(74)	(302)	(165)
10	20,793	21,867	1,25,996	3,134	31,733	3,711
	(418)	(393)	(2,189)	(58)	(312)	(116)
Income-Consistent Multiple Imputation	20,729	22,304	1,25,340	3,109	31,752	3,689
	(398)	(442)	(1,841)	(88)	(525)	(168)
Rate of Missing Information (γ)	0.23	0.53	0.09	0.47	0.41	0.41

Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplement. Standard errors (in parenthesis) are calculated using replicate weights. For more information, see "Standard Errors and Their Use" at <www.census.gov/hhes/www/p60_245sa.pdf>. Standard errors for the Income-Consistent file are calculated using the multiple imputation formula in Rubin and Schenker (1986).

Census Bureau from the CPS ASEC (available at http://www.census.gov/topics/income-poverty/income.html). Table 5 shows the median income statistics (Table 1 from the annual report) comparison between the Income-Consistent full sample and the traditional and redesign sample. Compared with the redesign sample, the only statistically significant differences are for median income of nonfamily households with a female householder (3.6% greater) and households headed by individuals without a disability (2.3% greater). At the 90% confidence interval, fewer than ten percent of the tested statistics are significantly different. For the comparison with the traditional sample, nearly all of the median income comparisons are statistically significant.

Table 6 compares poverty estimates in each of the traditional and redesign sample to the Income-Consistent file. The headline poverty number for all individuals is not statistically significantly different between the Income-Consistent file and either sample. However, for the traditional sample, poverty is lower in the Income-Consistent file for blacks (1.0%),

Table 5. Comparison of traditional, redesign, and income-consistent files: household median income by selected characteristics.

Characteristic	Traditional (T)			Redesign (R)		Income-Consistent Multiple Imputation (IC)		(IC-T)/T	(IC-R)/R
	Number (thousands)	Median (USD)	SE	Median (USD)	SE	Median (USD)	SE	% Diff	% Diff
All households	1,22,952	51,939	455	53,585	1,076	53,499	687	* 3.00	−0.16
Family households	81,192	65,587	643	66,923	872	67,301	673	* 2.61	0.56
Married-couple families	59,669	76,509	674	78,897	1,359	79,275	946	* 3.61	0.48
Female householder, no husband present	15,193	35,154	832	35,412	1,512	35,896	793	* 2.11	1.37
Male householder, no wife present	6,330	50,625	1,503	52,480	2,730	52,248	1,528	* 3.21	−0.44
Nonfamily households	41,760	31,178	518	31,480	951	31,977	539	* 2.56	1.58
Female householder	22,266	26,425	795	26,238	1,019	27,186	766	* 2.88	3.61 *
Male householder	19,494	36,876	937	39,379	1,674	38,242	1,094	* 3.70	−2.89
White	97,774	55,257	699	56,745	850	56,708	606	* 2.63	−0.06
White, not Hispanic	83,641	58,270	1,006	60,329	876	60,225	657	* 3.36	−0.17
Black	16,108	34,598	1,198	35,324	1,410	35,429	948	2.40	0.29
Asian	5,759	67,065	2,830	72,383	5,531	71,743	2,564	* 6.98	−0.88
Hispanic (any race)	15,811	40,963	908	39,687	1,954	41,341	894	0.92	4.17
Under 65 years	94,223	58,448	958	60,265	771	60,528	499	* 3.56	0.44
15 to 24 years	6,323	34,311	1,808	33,791	3,156	34,845	1,610	1.55	3.12
25 to 34 years	20,008	52,702	1,489	52,416	2,098	53,592	1,366	1.69	2.24
35 to 44 years	21,046	64,973	1,620	67,594	1,976	66,985	1,198	* 3.10	−0.90
45 to 54 years	23,809	67,141	1,265	70,598	2,114	70,671	1,214	* 5.26	0.10
55 to 64 years	23,036	57,538	1,662	60,481	1,835	60,735	1,503	* 5.56	0.42
65 years and older	28,729	35,611	722	37,297	1,283	36,352	808	2.08	−2.53
Native born	1,05,328	52,779	754	55,087	940	54,615	737	* 3.48	−0.86

Table 5. Continued.

Characteristic	Traditional (T)			Redesign (R)		Income-Consistent Multiple Imputation (IC)		(IC-T)/T	(IC-R)/R
	Number (thousands)	Median (USD)	SE	Median (USD)	SE	Median (USD)	SE	% Diff	% Diff
Foreign born	17,624	46,939	1,037	46,795	1,563	48,156	1,259	* 2.59	2.91
Naturalized citizen	9,491	54,974	2,898	56,354	3,098	57,406	1,947	* 4.42	1.87
Not a citizen	8,133	40,578	1,113	40,185	1,944	41,020	950	1.09	2.08
Households with householders aged 18 to 64	94,024	58,492	955	60,310	742	60,566	501	* 3.55	0.42
With disability	8,794	25,421	1,260	25,337	1,746	26,476	1,152	* 4.15	4.50
Without disability	84,784	61,979	564	62,487	1,021	63,924	892	* 3.14	* 2.30
Northeast	22,053	56,775	1,426	56,868	2,563	58,121	1,852	2.37	2.20
Midwest	27,214	52,082	1,160	53,426	2,102	53,207	1,427	* 2.16	−0.41
South	46,499	48,128	1,104	49,854	1,335	50,136	809	* 4.17	0.57
West	27,186	56,181	1,190	59,525	2,067	57,775	1,336	* 2.84	−2.94

The 2014 CPS ASEC included redesigned questions for income and health insurance coverage. All of the approximately 98,000 addresses were eligible to receive the redesigned set of health insurance coverage questions. The redesigned income questions were implemented to a subsample of these 98,000 addresses using a probability split panel design. Approximately 68,000 addresses were eligible to receive a set of income questions similar to those used in the 2013 CPS ASEC and the remaining 30,000 addresses were eligible to receive the redesigned income questions. Income in 2013 US dollars. Households and people as of March of the following year. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see http://www2.census.gov/programs-surveys/cps/techdocs/cpsmar14.pdf. Standard errors calculated using replicate weights for the traditional and redesigned samples, and using replicate weights for each Income-Consistent implicate and the multiple imputation variance formula from Rubin and Schenker (1986) to combine estimates across implicates for the Income-Consistent estimates. * if statistically significant difference at the 90% confidence level.

Table 6. Comparison of traditional, redesign, and income-consistent files: poverty by selected characteristics.

Characteristic	Total Population	Traditional				Redesign				Income-Consistent Multiple Imputation				IC-T	IC-R
		Number (thousands)	SE	% in Pov	SE	Number (thousands)	SE	% in Pov	SE	Number (thousands)	SE	% in Pov	SE	% in Pov	% in Pov
PEOPLE															
Total	3,12,970	45,318	616	14.5	0.2	46,269	896	14.8	0.3	45,257	549	14.5	0.2	0.0	−0.3
Family Status															
In families	2,54,990	31,530	514	12.4	0.2	32,786	833	12.8	0.3	31,668	480	12.4	0.2	0.0	−0.4
Householder	81,217	9,130	150	11.2	0.2	9,645	256	11.7	0.3	9,171	145	11.3	0.2	0.0 *	−0.5
Related children under 18	72,573	14,142	271	19.5	0.4	15,116	440	20.9	0.6	14,492	256	20.0	0.4 *	0.5 *	−0.9
Related children under 6	23,585	5,231	137	22.2	0.6	5,590	207	23.7	0.9	5,339	122	22.6	0.5	0.5	−1.0
In unrelated subfamilies	1,413	608	69	43.0	3.8	776	134	47.7	5.1	622	62	42.5	3.2	−0.5	−5.2
Reference person	595	246	29	41.3	3.9	291	52	44.0	5.0	243	24	40.3	3.0	−1.0	−3.6
Children under 18	714	340	42	47.7	4.1	448	79	53.1	5.7	358	40	47.5	3.5	−0.2	−5.6
Unrelated individual	56,564	13,181	252	23.3	0.4	12,707	352	22.9	0.5	12,967	225	23.0	0.3	−0.3	0.0
Race and Hispanic Origin															
White alone	2,43,080	29,936	496	12.3	0.2	31,287	652	12.9	0.3	30,235	415	12.4	0.2	0.1 *	−0.4
White alone, not Hispanic	1,95,170	18,796	439	9.6	0.2	19,552	495	10.0	0.3	19,102	349	9.8	0.2	0.1	−0.3
Black alone	40,615	11,041	308	27.2	0.8	10,186	384	25.2	0.9	10,615	270	26.2	0.7 *	−1.0 *	1.0
Asian alone	17,063	1,785	107	10.5	0.6	2,255	201	13.1	1.2	1,881	100	11.1	0.6	0.6 *	−2.0
Hispanic (of any race)	54,145	12,744	312	23.5	0.6	13,356	487	24.7	0.9	12,750	299	23.5	0.6	0.0	−1.1
Sex															
Male	1,53,360	20,119	345	13.1	0.2	20,294	468	13.2	0.3	20,101	291	13.1	0.2	0.0	−0.1
Female	1,59,600	25,199	348	15.8	0.2	25,975	548	16.3	0.3	25,156	327	15.8	0.2	0.0 *	−0.5
Age															
Under 18 years	73,625	14,659	277	19.9	0.4	15,801	441	21.5	0.6	15,039	261	20.5	0.4 *	0.6 *	−1.1
18 to 64 years	1,94,830	26,429	394	13.6	0.2	25,899	533	13.3	0.3	26,000	358	13.3	0.2	−0.2	0.0
65 years and over	44,508	4,231	138	9.5	0.3	4,569	174	10.2	0.4	4,218	135	9.5	0.3	0.0 *	−0.7

Table 6.   *Continued.*

Characteristic	Total Population (thousands)	Traditional				Redesign				Income-Consistent Multiple Imputation				IC-T	IC-R
		Number (thousands)	SE	% in Pov	SE	Number (thousands)	SE	% in Pov	SE	Number (thousands)	SE	% in Pov	SE	% in Pov	% in Pov
Nativity															
Native	2,71,970	37,921	573	13.9	0.2	38,831	790	14.3	0.3	38,064	492	14.0	0.2	0.0	−0.3
Foreign born	40,997	7,397	227	18.0	0.5	7,438	338	18.3	0.7	7,193	201	17.7	0.4	−0.4	−0.6
Naturalized citizen	19,147	2,425	105	12.7	0.5	2,132	151	11.1	0.8	2,245	89	11.7	0.4 *	−0.9	0.7
Not a citizen	21,850	4,972	189	22.8	0.7	5,306	303	24.8	1.1	4,948	174	22.9	0.7	0.1 *	−1.9

The 2014 CPS ASEC included redesigned questions for income and health insurance coverage. All of the approximately 98,000 addresses were eligible to receive the redesigned set of health insurance coverage questions. The redesigned income questions were implemented to a subsample of these 98,000 addresses using a probability split panel design. Approximately 68,000 addresses were eligible to receive a set of income questions similar to those used in the 2013 CPS ASEC and the remaining 30,000 addresses were eligible to receive the redesigned income questions. * if statistically significant difference at the 90% confidence level. Income in 2013 US dollars. Households and people as of March of the following year. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see ftp://ftp2.census.gov/programs-surveys/cps/techdocs/cps-mar13.pdf. Standard errors calculated using replicate weights for the traditional and redesigned samples, and using replicate weights for each Income-Consistent implicate and the multiple imputation variance formula from Rubin and Schenker (1986) to combine estimates across implicates for the Income-Consistent estimates.

naturalized citizens (-0.9%), residents of principal cities (0.5%), and workers (0.2%), and is higher for children (0.6%). For the redesign sample, unlike median income, there are significant differences: lower in poverty in the Income-Consistent file for children (1.1%) and those aged 65 and older (0.7%).

To summarize the results, the Income-Consistent file household median income estimates are more like the redesign file, but the poverty estimates lie between the two files. While the point estimate for poverty of 14.5% is not statistically significantly different from the point estimate for either file, it is much closer to the 14.5% estimate from the traditional file than the 14.8% estimate of the redesign file.

6. Conclusion

In this article, multiple imputation is applied to the problem of a split sample receiving different survey instruments in a bridge year. One possible way to use data from all survey respondents is shown, even though distinct sets of respondents answered different questions. This idea has an important potential benefit – by making use of all of the data during a bridge year, it potentially lowers the cost in terms of decreased statistical power of survey redesigns and bridges.

To address this problem of missing information during a survey bridge year, a semiparametric multiple imputation technique proposed by Bondarenko and Raghunathan is applied to the CPS ASEC 2014 redesign. The technique performs reasonably well and analysis of basic summary statistics shows how this technique affects important economic statistics that are widely reported on from the CPS ASEC.

For the 2014 CPS ASEC, this technique increases the potential sample that can be used to make comparisons to data from subsequent years, which uses the redesigned questionnaire for the entire sample. The larger sample facilitates analyses on subgroups, such as by state, where the redesign sample may lack the statistical power needed for comparisons. By combining the two samples, this technique may also address concerns about differences in sample composition raised in previous research by Mitchell and Renwick (2015).

7. References

Bondarenko, I. and T.E. Raghunathan. 2007. "Multiple Imputations Using Sequential Semi and Nonparametric Regressions." In Proceedings Section on Survey Research Methods: American Statistical Association, July 29, 2007. 3293–3300. Salt Lake City, UT. Available at: http://ww2.amstat.org/sections/srms/Proceedings/y2007/Files/JSM2007-000624.pdf (accessed November 2016).

Clogg, C.C., D.B. Rubin, N. Schenker, B. Schultz, and L. Weidman. 1991. "Multiple Imputation of Industry and Occupation Codes in Census Public-use Samples Using Bayesian Logistic Regression." *Journal of the American Statistical Association* 86: 68–78. Doi: http://dx.doi.org/10.1080/01621459.1991.10475005.

Czajka, J.L. and G. Denmead. 2008. "Income Data for Policy Analysis: A Comparative Assessment of Eight Surveys." Mathematica Reference No.: 6302-601. Available at:

https://www.mathematica-mpr.com/~/media/publications/PDFs/incomedata.pdf (accessed November 2016).

Hokayem, C., T.E. Raghunathan, and J. Rothbaum. 2015. "SRMI in the CPS ASEC." *Unpublished Manuscript*.

He, Y. and T.E. Raghunathan. 2006. "Tukey's gh Distribution for Multiple Imputation." *The American Statistician* 60: 251–256. Doi: http://dx.doi.org/10.1198/000313006X126819.

Hicks, W. and J. Kerwin. 2011. "Cognitive Testing of Potential Changes to the Annual Social and Economic Supplement of the Current Population Survey." Unpublished Westat report to the U.S. Census Bureau. July 25, 2011.

Butrica, B.A., I. Howard, K.E. Smith, and E.J. Toder. 2009. "The Disappearing Defined Benefit Pensions and Its Potential Impact on the Retirement Income of Baby Boomers." *Social Security Bulletin* 69: 1–28.

Meyer, B.D., W.K.C. Mok, and J.X. Sullivan. 2009. "The Under-Reporting of Transfers in Household Surveys: its Nature and Consequences." *National Bureau of Economic Research Working Paper #15181*.

Mitchell, J. and T. Renwick. 2015. "A Comparison of Official Poverty Estimates in the Redesigned Current Population Survey Annual Social and Economic Supplement." *U.S. Census Bureau SEHSD Working Paper #2014-35*.

Rosenbaum, P.R. and D.B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70: 41–55. Doi: http://dx.doi.org/10.1093/biomet/70.1.41.

Rothbaum, J. 2015. "Comparing Income Aggregates: How Do the CPS and ACS Match the National Income and Product Accounts, 2007–2012." *U.S. Census Bureau SEHSD Working Paper #2015-01*.

Rubin, D.B. 1987. *Multiple Imputation for Nonresponse in Surveys*. John Wiley and Sons. Doi: http://dx.doi.org/10.1002/9780470316696.

Rubin, D.B. and N. Schenker. 1986. "Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse." *Journal of the American Statistical Association* 81: 366–374. Doi: http://dx.doi.org/10.1080/01621459.1986.10478280.

Schenker, N. and J.D. Parker. 2003. "From Single-Race Reporting to Multiple-Race Reporting: Using Imputation Methods to Bridge the Transition." *Statistics in Medicine* 22: 1571–1587. Doi: http://dx.doi.org/10.1002/sim.1512.

Semega, J. and E. Welniak Jr. 2013. "Evaluating the 2013 CPS ASEC Income Redesign Content Test." Presented at the proceeding of the 2013 FCSM Conference. U.S. Census Bureau Income Statistics Working Paper. Available at: http://census.gov/library/working-papers/2013/demo/semega-01.html (accessed November 2016).

Semega, J. and E. Welniak Jr. 2015. "The Effects of the Changes to the Current Population Survey Annual Social and Economic Supplement on Estimates of Income." Presented at the Proceedings of the 2015 Allied Social Science Association (ASSA) Research Conference. U.S. Census Bureau Income Statistics Working Paper. Available at: http://www.census.gov/library/working-papers/2015/demo/cpsasec-red-income.html (accessed November 2016).

Tjur, T. 2009. "Coefficients of Determination in Logistic Regression Models—A New Proposal: The Coefficient of Discrimination." *The American Statistician* 63: 366–372. Doi: http://dx.doi.org/10.1198/tast.2009.08210.

Woodcock, S.D. and G. Benedetto. 2009. "Distribution-Preserving Statistical Disclosure Limitation." *Computational Statistics & Data Analysis* 53: 4228–4242. Doi: http://dx.doi.org/10.1016/j.csda.2009.05.020.