A Nonresponse Bias Study of the Consumer Expenditure Survey for the Ten-Year Period 2010-2019

May 9, 2022

Barry Steinberg, Sharon Krieger, Brett McBride, Brian Nix,
Michael Sverchkov, Daniel Yang

Consumer Expenditure Surveys Program Report Series

U.S. BUREAU OF LABOR STATISTICS • bls.gov | @BLS_gov

Table of Contents

1. Introduction

The Consumer Expenditure (CE) Surveys are nationwide household surveys sponsored by the U.S. Bureau of Labor Statistics to determine how U.S. consumers spend their money. There are two distinct CE Surveys, a quarterly Interview Survey and a two-week Diary Survey. The Interview Survey provides detailed information on large expenditures such as property, automobiles, and major appliances, as well as on recurring expenditures such as rent, utilities, and insurance premiums. By contrast, the Diary Survey provides detailed information on the expenditures of small, frequently purchased items such as food and apparel. The data from the two surveys are then integrated to provide a complete picture of consumer expenditures in the United States.

Over the past ten years (2010-2019), the response rates for both surveys decreased by about 20 percentage points, from 73 percent to 53 percent in the Interview Survey, and from 71 percent to 52 percent in the Diary Survey. These decreases are a concern because respondents and nonrespondents may have different expenditure patterns, and if so, there may be a bias in the survey estimates, with decreasing response rates increasing the amount of bias.

The Office of Management and Budget (OMB) which makes and enforces rules governing the collection of all data in federally sponsored surveys, encourages all federal surveys to study their nonresponse bias and requires such a study of all federal surveys whose response rates are below 80 percent. Both the CE Interview and Diary Surveys have response rates below 80 percent, so a nonresponse bias study is required of them. OMB's directive requires an analysis of nonresponse to determine whether the unreported data are "missing completely at random" (MCAR; see "**Studies to determine MCAR status,"** later in this text, for a definition), and another analysis to estimate the amount of nonresponse bias in the survey's estimates. Both analyses are summarized in this report.[1]

The analyses in this report show lower-income households are over-represented and higher-income households are under-represented in both CE Surveys. The analyses also show response rates are higher for rural households than for urban households in both CE Surveys; higher for homeowner households than for renter households in the Diary Survey; and lower for homeowner households than for renter households in the Interview Survey. Moreover, some of these relationships are changing over time. Taken together, all of these findings indicate that the unreported data in the two CE surveys are not MCAR. These over- and under-representations generally lead to nonresponse bias estimates of 0.0 percent to -2.0 percent in the Interview Survey and 0.0 percent to +3.5 percent in the Diary Survey.

2. Background and Approach

As mentioned above, over the ten-year period 2010-2019, the response rate for the Interview Survey decreased from 73.4 percent to 53.7 percent, and the response rate for the Diary Survey decreased from 71.5 percent to 52.8 percent. (See Table 1.) This decrease is a concern because it may affect the accuracy of CE's expenditure estimates if the respondents and nonrespondents have different expenditure patterns.

---

[1] OMB's standards and guidelines for statistical surveys were originally published in 2006 and have been modified periodically since then. The requirement that federal surveys with response rates below 80 percent conduct a nonresponse bias study can be found in the September 2006 version (guideline 3.2.9 on page 16) and the October 2016 version (question 66 on page 60), and guidelines for conducting a nonresponse bias study can be found in the September 2006 version (guideline 3.2.9 on page 16) and the October 2016 version (question 71 on page 64).

This report examines the possibility of respondents and nonrespondents having different expenditure patterns by examining the surveys' nonrespondents and determining whether their unreported data are MCAR. The research presented in this report updates and expands the research performed in 2008 by CE program staff, which found that expenditure estimates from the Interview Survey did not have a significant amount of nonresponse bias even though the respondents and nonrespondents had different characteristics, and even though the unreported data were not MCAR.[2]

Table 1. Unweighted Response Rates for the CE Interview and Diary Surveys, 2010-2019

	CE Interview Survey				CE Diary Survey			
Collection Year	Total Eligible Cases	Type A[3] Noninterviews	Complete Interviews	Response Rate	Total Eligible Cases	Type A Noninterviews	Complete Interviews	Response Rate
2010	38,718	10,289	28,429	73.4%	19,988	5,692	14,296	71.5%
2011	38,348	11,358	26,990	70.4%	19,823	5,898	13,925	70.2%
2012	38,835	11,842	26,993	69.5%	20,298	6,537	13,761	67.8%
2013	39,142	13,034	26,108	66.7%	20,296	7,961	12,335	60.8%
2014	39,003	13,095	25,908	66.4%	20,476	7,170	13,306	65.0%
2015	36,692	13,118	23,574	64.2%	20,517	8,676	11,841	57.7%
2016	40,375	14,934	25,441	63.0%	20,391	8,839	11,552	56.7%
2017	40,193	15,714	24,479	60.9%	20,110	8,452	11,658	58.0%
2018	40,366	17,207	23,159	57.4%	20,133	9,054	11,079	55.0%
2019	40,389	18,688	21,701	53.7%	20,244	9,562	10,682	52.8%

This report contains a summary of four studies undertaken with more recent data to determine whether CE's data are MCAR, and four studies to estimate the amount of nonresponse bias in CE's expenditure estimates. The four MCAR studies are:

- Study 1: A comparison of CE's respondent demographic characteristics to those of the American Community Survey (ACS).
- Study 2: A comparison of response rates between subgroups of CE's sample.
- Study 3: A linear regression analysis of CE's response rate trends and demographic characteristic trends over the ten-year period 2010-2019.
- Study 4: A logistic regression analysis of CE's response rates using socio-demographic variables that are available for both respondents and nonrespondents.

And the four nonresponse bias studies are:

[2] See the August 2008 study "Assessing Nonresponse Bias in the CE Interview Survey: A Summary of Four Studies," by Boriana Chopova, Jennifer Edgar, Jeffrey Gonzalez, Susan King, Dave McGrath, and Lucilla Tan.

[3] A Type A interview occurs when the survey's field representative finds an occupied housing unit but is unable to contact an eligible household member or is unable to convince a reluctant household member to participate in the survey.

- Study 1: A comparison of expenditure estimates from the survey's respondents weighted two different ways – with unadjusted base weights and with base weights adjusted to account for nonresponse.
- Study 2: A comparison of expenditure estimates identical to Study 1 except the base weights are adjusted in a different way.
- Study 3: A comparison of expenditure estimates identical to Studies 1 and 2 except the base weights are adjusted in a third way.
- Study 4: A comparison of expenditure estimates between two different subsets of CE's respondents, "pseudo respondents" and "pseudo nonrespondents." (See Section 6.4, "Method 4" for definitions.),

All four MCAR studies conclude that the nonrespondents' unreported data in both surveys are not MCAR. Study 1 and Study 3 show the distributions of various socio-demographic characteristics differ between the CE Surveys and the ACS, and the relationships between some of them are changing over time. Study 2 and Study 3 show the response rates among various subgroups of the CE's sample differ from each other and the relationships between some of them are changing over time as well. And Study 4 shows CE's response rate is affected by the demographic composition of the survey's sample. All four of these studies show different subgroups of the survey's sample respond to the CE Surveys at different rates, which means their patterns of missingness are not MCAR.

All four nonresponse bias studies conclude that nonresponse bias is small. Three of the four studies compare the average annual expenditure per household computed with base weights to the average annual expenditure per household computed with base weights adjusted for nonresponse. Study 1 uses the nonresponse adjustment used in production which is a traditional cell adjustment method, while Study 2 and Study 3 use the weight of each household adjusted according to its probability of responding to the survey using a logistic regression model. And Study 4 compares the average annual expenditure per household for respondents that are easy to contact (require few contact attempts) to that of all respondents (both those that are easy to contact and those that are hard to contact). All four of these studies show nonresponse bias for the total expenditures summary variable ranges from approximately 0.0 percent to -2.0 percent in the Interview Survey, and from 0.0 percent to +3.5 percent in the Diary Survey, which means CE's nonresponse bias is small.

The report will describe each study and provide tables and graphs highlighting the magnitude of the bias and the trends over the ten-year period.

Studies to determine MCAR status

To determine whether the missing values in the two CE Surveys are "missing completely at random" (MCAR), the four studies described above were performed. Details of each are below. But before proceeding, the term "MCAR" needs to be defined. The generally accepted definition comes from Roderick Little and Donald Rubin (2002). According to them, data are MCAR if the mechanism that produces the missing values is unrelated to the values of the data themselves and independent of any other characteristics as well.[4] The question of whether the data are MCAR is important because nonresponse bias is often associated with the data not being MCAR.

[4] For more details, see Roderick J.A. Little and Donald B. Rubin, "*Statistical Analysis with Missing Data,*" 2002, second edition.

In practical terms, this definition means CE's data are MCAR if the difference in amount spent between survey's respondents and nonrespondents for the same set of goods and services is not statistically significant (i.e., the pattern of missingness is independent of the data's actual values), and if every demographic subgroup of the survey's sample has, statistically speaking, the same response rate (i.e., the pattern of missingness is independent of any other characteristics). Since the expenditures of nonrespondents are unknown, examining the response rates among different demographic subgroups is one of the primary methods of determining whether a survey's data are MCAR, and it is one of the methods used in this report. [5]

The first study examines whether the data are MCAR by comparing the distribution of socio-demographic characteristics of the survey's respondents to those of a recent census or a "gold standard" survey. The American Community Survey (ACS) can be thought of as a "gold standard" survey because it has a large sample size, a high response rate, a high coverage rate, and a reputation for accuracy. [6] Any differences between the current survey and the gold standard survey suggest that the two surveys have different response mechanisms, and since the gold standard survey is presumed to have a response mechanism closer to that of an MCAR process, the other survey is presumed to have a response mechanism further from that of an MCAR process, which means its missingness pattern is probably not independent of the data themselves.

The second study examines whether the data are MCAR by comparing the survey's respondents and nonrespondents on several socio-demographic variables that are available for both groups. Any differences between them indicate that the pattern of missingness is not independent of other variables, and therefore the missing data are not MCAR. Despite the limited number of variables that can be used in this analysis, it is another standard method of figuring out whether the data are MCAR, and it is used in this study.

The third study looks at ten-year trends in response rate and demographic characteristic "relativities" using simple linear regressions to determine whether the relationships of the response rates to each other and the demographic characteristics to the ACS (the gold standard survey) are changing over time. In the case of response rates, the relativity of interest is the ratio of response rate for a subgroup in CE to the overall response rate. For example, the ratio of the response rate in the Northeast region to the response rate for the whole country. In the case of demographic characteristics, the ratio of interest is the proportion of CE's respondents in a specific demographic subgroup to the proportion of the population in the same demographic subgroup according to the ACS.

---

[5] For those who like formal logic, the following may be helpful. Start by recalling that these two statements are logically equivalent: "MCAR $\Leftrightarrow$ A *and* B" and "~MCAR $\Leftrightarrow$ ~A *or* ~B." The key difference between these two statements is that one has the word, "*and*" which means two things have to be demonstrated, while the other has the word "*or*" which means only one thing has to be demonstrated. Thus, two things need to be demonstrated to show the data are MCAR (the pattern of missing-ness is independent of the data's actual values *and* the values of any other variables), but only one thing needs to be demonstrated to show the data are not MCAR (the pattern of missing-ness is not independent of the data's actual values *or* the values of any other variables). Thus, we only need to show that the pattern of missing-ness is not independent of the values of any other variables to show that the data are not MCAR. We do not need to show anything about the unobservable expenditures of the survey's nonrespondents, hence demonstrating that the data are *not* MCAR is easier than demonstrating that they are MCAR.

[6] The ACS Survey is sent to over 3.5 million housing units per year, which is a large sample size, and in 2016 its response rate (94.7 percent and its coverage rate (91.9 percent) were both high.

Finally, the fourth study uses logistic regressions to examine whether the surveys' response rates are associated with certain socio-demographic characteristics. A logistic regression is a model of the outcomes of a binary process, such as whether a sample household participates in the CE Survey. It has a specific algebraic form that ensures its numeric values are between 0 and 1, which makes it suitable for modeling probabilities.

3. Methodology

3.1. Data

For comparability of results, the analyses for the Interview and Diary Surveys both used the same ten years of data, January 2010 through December 2019. The unit of analysis in these studies was generally the consumer unit (CU), but a mixture of information about the CU and individual members was used for the analyses comparing CE's demographic characteristics to those of the ACS. CUs are similar, but not identical to, households.[7] Nevertheless, most households comprise only one CU, so the terms are used interchangeably herein.

3.2. Sample Design and Weighting

The CE Surveys' sample design is a nationwide probability sample of addresses. That means a random sample of addresses is selected to represent the addresses of all CUs in the nation. Most addresses have only one CU living therein, hence the terms "address" and "CU" are used interchangeably in this report. The CE Surveys actually have a two-stage sample design in which a random sample of geographic areas called Primary Sampling Units (PSUs) is selected for the survey, and then a random sample of CUs is selected within those PSUs to represent them. The Bureau of Labor Statistics (BLS) selects the sample of PSUs and the U.S. Census Bureau selects the sample of CUs.

Each interviewed CU represents itself plus a number of other CUs that were not interviewed for the survey. Therefore, each interviewed CU must be weighted to properly account for the proportion of the population it represents. The weighting process starts with a "base weight," which is the number of CUs in the nation the CU selected for the sample represents. It is equal to the inverse of the CU's probability of selection, and since every CU in a PSU has the same probability of selection every CU selected in a PSU has the same base weight.

Then, BLS makes three types of adjustments to the base weights: an adjustment in the rare situation where a field representative finds multiple housing units where only a single housing unit was expected; a nonresponse adjustment to account for CUs that were selected for the survey but did not participate in it; and a calibration adjustment to account for nonresidential and other out-of-scope addresses in the sampling frame, as well as sampling frame under-coverage.[8] These weight adjustments are made to each

[7] A CU is a group of people living together in a housing unit who are related by blood, marriage, adoption, or some other legal arrangement; who are unrelated but pool their incomes to make joint expenditure decisions; or is a person living alone or sharing a housing unit with other people but who is financially independent of the other people. A household consists of one or more people who live in the same dwelling and may consist of a single family or another group of people.

[8] Since invalid addresses are available for selection, this is accounted for during the calibration adjustment process.

individual CU that participated in the survey, hence each respondent CU has its own unique "nonresponse-adjusted weight" and "final calibration weight." All of the studies in this report use base weights, except for the study comparing CE respondents to external data which uses final calibration weights.

3.3. Significance tests for one-way and two-way socio-demographic comparisons

Respondents and nonrespondents were compared on several categorical socio-demographic characteristics to ascertain whether the two groups had the same distribution of characteristics, and whether those characteristics were correlated with the likelihood of responding to the survey. For these comparisons, the Rao-Scott chi-square statistic was used, which is a design-adjusted version of the Pearson chi-square statistic involving differences between observed and expected frequencies. For one-way comparisons, the null hypothesis was that the respondents in the CE Surveys and the ACS had the same distribution of characteristics. And for two-way comparisons, the null hypothesis was that the response status (interview or noninterview) of CUs in the CE Survey was independent of their socio-demographic characteristics.

Ten years of data were analyzed in this study (2010-2019), with a separate analysis done for each year. The Rao-Scott chi-square statistic was generated for each comparison, with one statistic generated for each year, and the results of those ten yearly analyses were summarized by counting the number of times statistically significant results were obtained. For one-way comparisons, a difference in distribution was considered to be "strongly significant" if statistically significant differences ($p<0.05$) were found in 5 or more years analyzed, "moderately significant" for significant differences in 3 or 4 years, and "not significant" otherwise. For example, the difference between CE's and ACS's "education" distributions were statistically significant in 8 of the 10 years for the Interview Survey and in all 10 years for the Diary Survey, so both cases were considered to be "strongly significant" (see Appendix B).

For two-way comparisons, the scoring system was similar to the one-way comparisons. For each comparison, a net difference was calculated as the number of years the first subgroup listed had a statistically significantly higher response rate than the second subgroup listed ($p<0.05$) minus the number of years it had a statistically significantly lower response rate ($p<0.05$). In other words, for each year, if the first subgroup listed had a statistically significantly higher response rate than the second subgroup listed, then it was given a score of "+1"; if it had a statistically significantly lower response rate, then it was given a score of "–1"; and if there was no statistically significant difference, then it was given a score of "0." Then the ten scores for the ten years were summed, giving an overall score between –10 and +10. The difference between the two subgroups was then categorized as "strongly significant" if the overall score was greater than or equal to +5 or less than or equal to –5; "moderately significant" if it was equal to ±3 or ±4; and "not significant" if it was equal to ±2, ±1, or 0. Table 2 shows an example comparing the response rates for the South region to the West region of the country:

Table 2. Example comparing the response rates for the South region to the West region, 2010-2019

Category	# Years	Total Score = # Years * Score
South's response rate is significantly higher than the West's response rate	3 years	3*(+1) =+3
South's response rate is significantly lower than the West's	0 years	0*(–1) = 0
No significant difference between the South's and West's response	7 years	7*(0) = 0
Overall Score		+3

The overall score was +3, which was in the 3 or 4 range, hence the South's response rates were higher than the West's response rates, and the difference was "moderately significant."

3.4. Significance tests for linear regression analysis

For tests of significance pertaining to response rate subgroups, the ratio of each demographic subgroup's response rate to the overall response rate was calculated each year for the ten years analyzed. Then an ordinary least squares (OLS) regression line $Y = \beta_0 + \beta_1 X$ was fit to the data where the x-variable was the year in which the data were collected, and the y-variable was the ratio of response rates for that year. After fitting the line, a t-test was performed to determine whether its slope differed from zero. That is, the two-sided hypothesis test of the slope was this:

H_0: $\beta_1 = 0$

H_a: $\beta_1 \neq 0$

A level of significance of α=0.05 was used, so if $p<0.05$ for the t-test on the slope coefficient, then the ratio of response rates is linearly increasing over time if the slope coefficient is positive, and linearly decreasing over time if the slope is negative.

4. Individual studies to determine MCAR

4.1. Comparison of CE respondents to external data

As mentioned above, a common approach to analyzing the effect of nonresponse on a survey's estimates is to compare the distribution of socio-demographic characteristics of the survey's respondents to that of a recent census or other "gold standard" survey (Groves, 2006). If they are the same then the nonrespondents are likely MCAR, but if they are different then the nonrespondents are most likely not MCAR.

Appendix A for the Interview Survey and Appendix D for the Diary Survey show a 2019 comparison of the distribution of selected socio-demographic characteristics between the CE and the ACS. Calibration-weighted respondents for CE are used in the comparisons between the CE survey and the ACS. The characteristics compared are sex, age, race, education, CU size, housing tenure, number of rooms in a housing unit, owner-occupied housing value, monthly rent, and CU income. Housing information about the number of rooms in a housing unit, the housing unit's market value, and the housing unit's rental

value are available from the Interview Survey only. Tables for all years were produced but showing one year provides information to get a sense for the work that was done.

Comparing the distribution for a particular characteristic in the CE data to its distribution in the ACS data falls into the framework of a one-way chi-square goodness-of-fit test. The Rao-Scott chi-square statistic described above is used to find out whether a characteristic's distribution in the CE Survey and the ACS are the same or different.[9] For both surveys, statistically significant differences ($p < 0.05$) were found for many of the socio-demographic characteristics. Table 3 below summarizes these results.

Table 3. Summary of Comparison of Socio-demographic Characteristics in the CE Survey and ACS

	Calibration-weighted CE respondents versus ACS	
	CE Interview Survey	CE Diary Survey
"Strongly Significant" Differences	Gender Race Education CU size No. rooms in housing unit Owner-occupied housing value Monthly rent CU income	 Race Education CU income
"Moderately Significant" Differences		Gender CU size
Not Significant		

It should be pointed out that there are factors beyond the characteristics of the respondents in these two surveys that make differences likely to be statistically significant. First, the CE and the ACS differ in both their data collection modes and question wording. And second, for some of the CU-level variables examined, the definitional difference between CUs in the CE Surveys and households in the ACS may impact the results even though most of the time they are the same. As a result, the strength of the comparison of CE data with ACS data is limited by the extent to which the survey designs are truly comparable.

Further analysis was done to observe trends over time for the CE data compared to the ACS data by using linear regression analysis over the ten-year period from 2010 to 2019. The goal of this analysis is to determine whether the CE and ACS have the same distributions of socio-demographic characteristics, and

[9] For these comparisons, the Rao-Scott Chi-square test with the BRR variance method is used to reflect CE's sample design. The chi-square test is included for all socio-demographic characteristics listed, except for age and housing tenure. The reason for excluding comparisons for these two variables is that they are used in calibration, meaning their replicate weights and final weight in the BRR procedure create design correction factors that are zero or very close to zero. This causes the resulting test statistics to become extremely large and their associated p-values to become extremely small. Therefore, the comparison of CE's distribution to ACS's distribution for these two variables is not practical.

if they are different whether they are moving towards each other or moving away from each other over time. In other words, whether the CE/ACS ratio is moving towards 1.00 or moving away from 1.00. The ten yearly CE/ACS relativities over the ten-year period 2010-2019 are plotted and analyzed to determine whether their relationships are changing or holding steady over time.

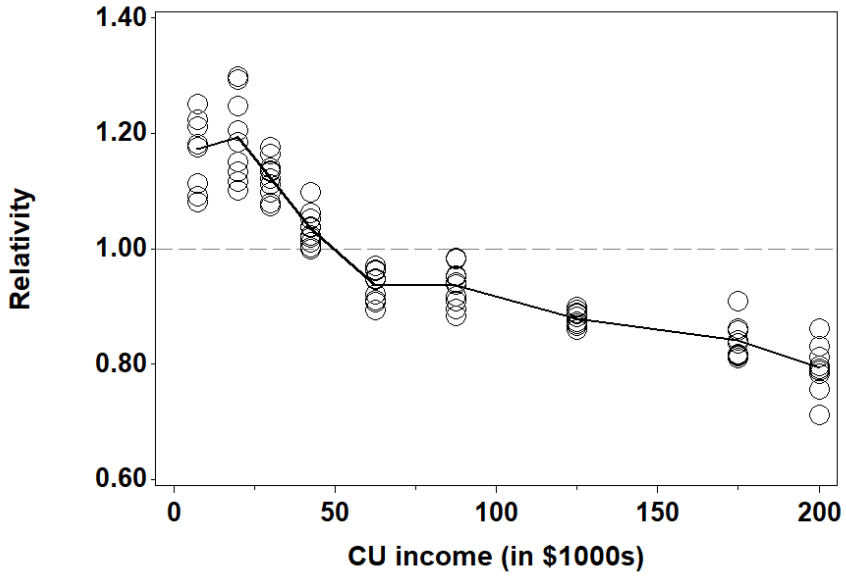
CE-to-ACS comparison for the Interview Survey

As mentioned above, several of the characteristics have different distributions in the two surveys. Some of the differences are rather small and statistically significant but a few of them have noticeable patterns in which some socio-demographic subgroups are systematically over-represented or under-represented relative to the ACS. Characteristics with noticeable patterns include the market value of owner-occupied housing units, the monthly rent of rental housing units, and especially CU income (Figure 1). The graphs below show the patterns of over-representation or under-representation for CU income. The graphs show the socio-demographic subgroups along the horizontal axis, then above them there are ten circles showing the CE/ACS relativities for those subgroups for each of the ten years, and a solid line connecting the average value of the CE/ACS relativities to show the patterns.

Figure 1 shows the patterns of over-representation and under-representation for CUs' by ranges of annual incomes (less than $0 to $200,000+) in the Interview Survey.[10] The graph shows that CUs with incomes below $50,000 are over-represented in the CE Interview Survey relative to the ACS, while CUs with incomes of $50,000 or higher are under-represented. Specifically, CUs with incomes below $50,000 are over-represented by 5 to 20 percent, while CUs with incomes of at least $50,000 are under-represented by 5 to 20 percent. Furthermore, for CUs with higher incomes (i.e., at least $50,000), the under-representation grows with their incomes, so that, for example, the $100,000-$149,999 subgroup is under-represented in the CE Survey by about 10 percent, the $150,000-$199,999 subgroup is under-represented by about 15 percent, and the $200,000+ subgroup is under-represented by about 20 percent. For the Diary Survey, the patterns of over-representation and under-representation are similar to what is shown for the Interview Survey in Figure 1.

[10] The lowest income group ("Less than $15,000") includes CUs with negative incomes. [Negative incomes can occur when they have large capital losses, such as when they sell a house for less than its purchase price.] WORD DOES NOT ALLOW COMMENTS IN FNs, so note the "Negative incomes" sentence is incorrect, plus the addition of new text here: Negative incomes can occur when consumer units incur losses via self-employment or rental property income. In addition to the "Less than $15,000" range, the other ranges are: $15,000 – $X; $X+1 to $Y; …, and $200,000+. ALTERNATIVE: Since the authors use microdata, it is plausible that they excluded the negative incomes. If that is the case, I suggest: In standard publications, the lowest income group ("Less than $15,000") includes CUs with negative incomes. (Negative incomes can occur when consumer units incur losses via self-employment or rental property income.) However, we have excluded negative incomes from this analysis, so that the minimum possible value is $0. In addition to the "$0 to $15,000" range, the other ranges we examine are: $15,000 – $X; $X+1 to $Y; …, and $200,000+

Figure 1. CE-to-ACS Relativities for CU Income Subgroups in the Interview Survey, 2010-2019



These patterns for income are a problem because incomes are correlated with expenditures Therefore, their patterns suggest that CUs with higher expenditures are under-represented in CE Surveys, which may lead to bias.

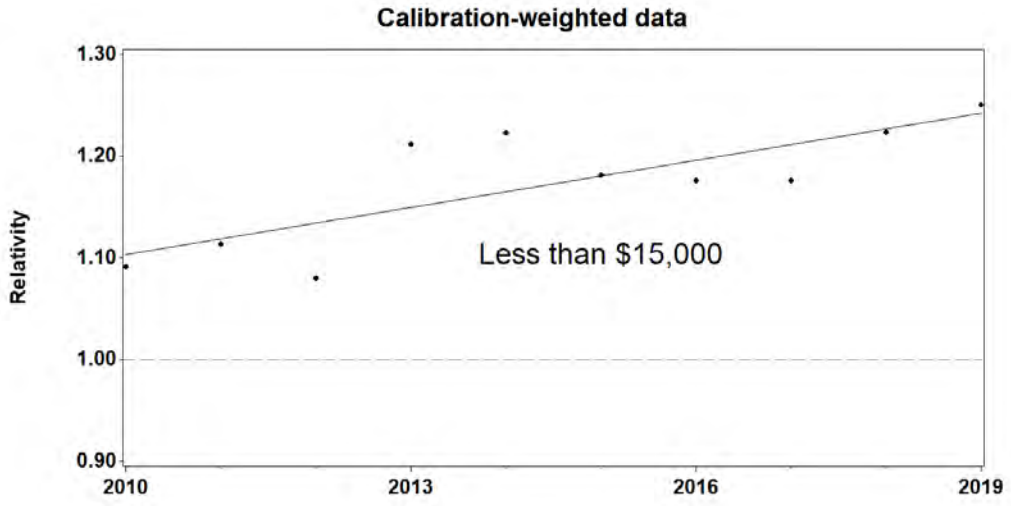
Regression Analysis for the Interview Survey

In the previous two sections, we looked at the distributions of various socio-demographic characteristics among CE's respondents relative to the ACS. The next two sections look at how those distributions changed over the ten-year period. The graphs show the ten-year period (2010-2019) on the horizontal axis, and the yearly "relativities" of selected socio-demographic characteristics on the vertical axis. Each graph also has a linear regression line showing how the relativities changed over time. The regression lines are trend lines as opposed to a formal time series analysis.

Appendix C shows the following socio-demographic characteristics have statistically significant trends (based on the coefficient for year) for one or more subgroups: age, race, education, CU size, monthly rent, owner-occupied housing value, and CU income.

CU Income. For the Interview Survey, the subgroups with incomes less than $15,000, $15,000 to $24,999, and $25,000 to $34,999 have regression lines with statistically significant slopes. Their p-values were p=0.007, p<0.001, and p=0.008, respectively, for the calibration-weighted data. All three subgroups have regression lines with CE/ACS relativities that start between 1.09 and 1.12 and increase to between 1.18 and 1.30. That means CUs in these subgroups were over-represented by 9 percent to 12 percent relative to the ACS at the beginning of the ten-year period (i.e., in 2010) and they were over-represented by 18 percent to 30 percent at the end of the ten-year period (i.e., in 2019). Since movement towards 1.00 is desirable, these subgroups are moving in the wrong direction. The less than $15,000 subgroup is shown in Figure 2.

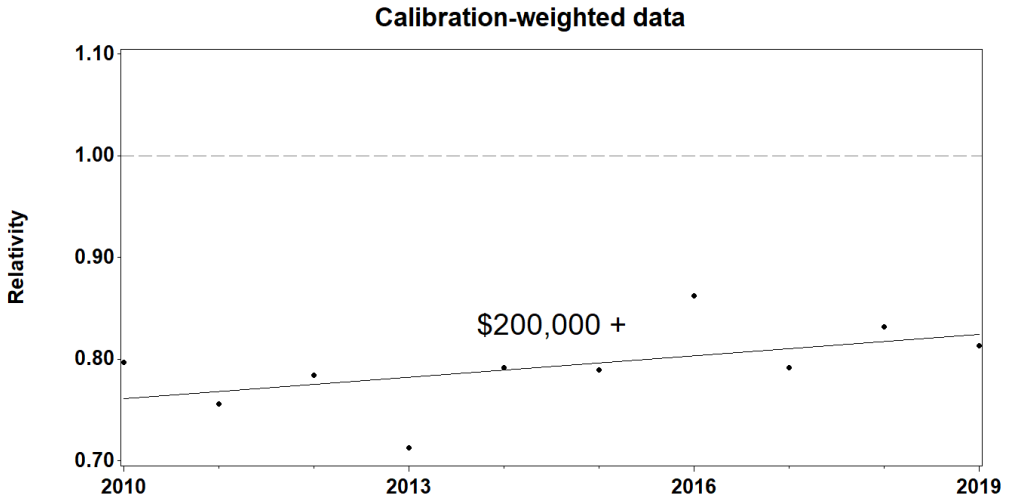
Figure 2. Lowest CU income group over-represented in the CE Interview Survey



Also, for the Interview Survey, the subgroups with incomes between $150,000 to $199,999 and $200,000+ have regression lines with slightly positive slopes, but they are not statistically significant. Their regression lines are well below the preferred ratio of 1.00 throughout the ten-year period. The CUs in these subgroups were under-represented by about 10 to 20 percent relative to the ACS throughout the ten-year period. These results are consistent with other recent research findings that show high-income CUs are under-represented in the CE Interview Survey and that CE's weighting procedures do not fix the problem.[11] The $200,000+ subgroup is shown in Figure 3.

[11] John Sabelhaus, David Johnson, Stephen Ash, David Swanson, Thesia Garner, and Steve Henderson, *Is the Consumer Expenditure Survey Representative by Income?*, (NBER Working Paper No. 19589, October 2013).

Figure 3. Highest CU income group under-represented in the CE Interview Survey



Summary. The graphs in this section highlight two points, one being that the CE Interview Survey and the ACS have different distributions for several socio-demographic characteristics, and the other being that the relationships between some of those distributions are changing over time. Assuming ACS's distributions are more accurate than CE's distributions, both of these findings/results suggest that the CE Interview Survey's data are not MCAR. Furthermore, the difference in the distribution of CU incomes from the ACS distribution is growing over time for low-income CUs. Low-income CUs are over-represented in the CE Interview Survey relative to the ACS and their over-representation is growing over time, and high-income CUs are under-represented in the CE Interview Survey relative to the ACS and their under-representation is relatively unchanged over time. This is a concern since CE is a survey about spending, and the over-representation of low-income/under representation of high-income CUs may result in CE's expenditure estimates being under-estimated, and with the under-estimation growing over time. This will be discussed later in the report.

Just like in the Interview Survey, Appendix E shows the results of the regression analysis from 2010 to 2019 for all subgroups in the Diary Survey, with statistically significant test results highlighted in gray. The results relative to statistically significant trends for CU income in the Diary Survey are fairly consistent with those of the Interview Survey.

4.2. Comparison of response rates across subgroups: General information

This study examined the response rates by socio-demographic characteristics among subgroups for which such characteristics could be identified for both respondents and nonrespondents. Any differences between them indicate that the pattern of missingness is not independent of other variables, and therefore the missing data are not MCAR. As mentioned above, such comparisons are usually limited in scope because little is known about the nonrespondents since they do not respond to the survey. Consequently, the variables examined for them are often limited to a small number of variables on the sampling frame and maybe a few other variables that data collectors are able to collect for every sample unit regardless of their participation in the survey. For this reason, the subgroups analyzed were limited to region of the

country (Northeast, Midwest, South, West), "urbanicity" (urban, rural), PSU size class, housing tenure (owner or renter), and housing values for owners and renters.[12]

Base-weighted response rates were calculated for these subgroups separately for both the Diary Survey and the four waves of the Interview Survey.[13] As a reminder, base weights are the inverse of a sample address's probability of selection. Base-weighted response rates answer the question "What percent of the survey's target population do the respondents represent?" Base-weighted response rates are defined as the sum of base-weighted interviewed units divided by the sum of base-weighted interviewed units plus the base-weighted Type A noninterviews units. Type A noninterviews occur when no interview is completed at an occupied eligible housing unit.

Base-weighted response rate = $\frac{\sum_{i \in I} w_i}{\sum_{i \in I} w_i + \sum_{i \in A} w_i}$ where:

- w_i = base weight for the i^{th} CU.
- "I" is the set of all CUs that completed an interview; and
- "A" is the set of all CUs that are Type A noninterviews.

4.2.1. Interview Survey: Comparison of response rates across subgroups

Interview Survey response rates were examined across socio-demographic subgroups for the ten-year period (2010-2019) and their results are summarized in Appendix F, Appendix G, and Appendix H. Appendix F shows response rates for each subgroup and the nation by wave. Only results for 2019 are shown to keep the report more condensed. Appendix G summarizes the test results from the Rao Scott chi-square tests for each of the subgroup comparisons only for Wave 4, again to keep the report more condensed. And Appendix H shows response rate relativities for each of the subgroups relative to the nation by year and wave. Examples of these subgroup comparisons include the Northeast vs. Nation, Midwest vs. Nation, South vs. Nation, and West vs. Nation. The response rate relativities are then used to create the regression lines that demonstrate the statistical significance of the slope. For each of the four interview waves, all possible pairs of subgroups within the six categories were examined over the ten-year period.

Using a level of significance α=0.05, a t-test of slope coefficients from linear regression is used to find whether the slope of the ten-point regression line (each point represents one year) differs from zero. For example, if the slope is 0.0038 (i.e., the response rate relativity increases 0.0038 per year) and the

[12] The information on housing values is from the 2000 decennial census for CUs that were in the sample in 2010 through 2014, and it is from the 2010 decennial census for CUs that were in the sample in 2015 through 2019. This means the information is available for every CU, both respondents and nonrespondents, but it is out-of-date.

[13] If a household is selected to participate in the Interview Survey, that address will be scheduled to have four interviews, one every three months, over the course of a ten-month period. The interview number (from 1 to 4), also called the wave, will be used to identify which visit in the sequence it represents. For example, if the household is scheduled for interviews in February, May, August, and November then the February interview would be Wave 1, the May interview would be Wave 2, the August interview would be Wave 3, and the November interview would be Wave 4.

standard error of the slope is 0.0016, then it has a t-statistic of 2.38 (= (0.0038 – 0.0000)/0.0016), which means the slope is statistically different from zero at α=0.05 level of significance.

The two-way comparisons show that there are many statistically significant differences in response rates for every subgroup and since there is not a trend for convergence for the overwhelming majority of these comparisons, this strongly demonstrates that the data are not Missing Completely at Random for the Interview Survey.

The Diary Survey response rates analyses were examined across socio-demographic subgroups in a similar fashion to the Interview Survey and their results are summarized in Appendix I, Appendix J, and Appendix K. Much like the Interview Survey, response rate differences within the subgroups suggest that the data are not MCAR because the respondent and nonrespondent CUs are not simple cross sections of the original sample.

4.3 Models for determining MCAR

In this study a model of the probability of a sample household participating in the CE Survey is developed. It is a logistic regression model where the independent variables are chosen from a list of geographic and housing characteristics that are available for every household on the sampling frame, and the dependent variable is the household's probability of participating in the survey. The purpose of the model is to determine whether there is a relationship between these geographic and housing characteristics and a household's probability of participating in the survey. If there is a relationship, then the nonrespondents' missing data is not MCAR.

The variables were chosen for the model with a forward stepwise selection process, and then after the main effect variables were chosen the interaction terms were evaluated and chosen. Separate models were generated for the Interview and Diary Surveys using all ten years of data. The two models were nearly identical to each other, with region of the country (Northeast, Midwest, South, West), urbanicity (urban or rural), and household tenure (homeowner or renter) as the main effect variables, and many of their two- and three-way combinations as the interaction terms. The small p-values on the individual variables in the models and the high overall goodness-of-fit statistics on the models indicate that there is a relationship between the variables and a household's probability of participating in the survey. That means the nonrespondents' missing data is not MCAR.

5. Calculating Relative Nonresponse Bias

5.1. OMB nonresponse bias equation

To estimate nonresponse bias, OMB (2006) provided a specific formula for computing the nonresponse bias of the respondent sample mean. This is given by:

$$B(\bar{y}_R) = \bar{y}_R - \bar{y}_T = \left(\frac{n_{NR}}{n}\right)(\bar{y}_R - \bar{y}_{NR})$$

where:

- $B(\bar{y}_R)$ is the nonresponse bias of the respondent sample mean;
- $\bar{y}_R$ is the mean based only on respondent cases;
- $\bar{y}_T$ is the mean based on all sample cases;
- n_{NR} is the number of nonrespondent cases in the sample; and
- n is the number of cases in the sample; and
- $\bar{y}_{NR}$ is the mean based only on nonrespondent cases.

Slight modifications to the nonresponse bias formula were necessary because relevant data (e.g., expenditures) were not available for the CE nonrespondents. After the modifications (described later in Section 6.1 through 6.4) were made, the application of the formula to CE expenditure data becomes:

$$B(\bar{Z}_R) = \bar{Z}_R - \bar{Z}_T = \left(\frac{N_{NR}}{N}\right)(\bar{Z}_R - \bar{Z}_{NR})$$

where:

- $B(\bar{Z}_R)$ is the nonresponse bias in the base-weighted respondent sample mean.
- $\bar{Z}_R$ is the base-weighted mean of expenditures for all respondent CUs (this estimate *excludes* pseudo nonrespondent CUs, defined below, from the calculation);
- $\bar{Z}_T$ is the base-weighted mean of expenditures for all CUs (this estimate includes all CUs, respondents *and* pseudo nonrespondents);
- N_{NR} is the base-weighted number of pseudo nonrespondent CUs;
- N is the base-weighted number of CUs; and
- $\bar{Z}_{NR}$ is the base-weighted mean of expenditures for all pseudo nonrespondent CUs;

Pseudo nonrespondents are respondents with low contact rates. "Harder-to-contact" respondents are considered as proxy nonrespondents. It draws on a theory known as the 'continuum of resistance' to identify appropriate respondents to serve as proxy nonrespondents. This theory suggests that sample units can be ordered across a continuum by the amount of interviewer effort exerted in order to obtain a completed interview (Groves, 2006). Their difficulty in being contacted makes them similar to nonrespondents in terms of their low probability of participating in the survey, and it is assumed that they are similar to nonrespondents in other ways as well, such as in the expenditures they make.

For the estimates of nonresponse bias in the pseudo nonrespondent study, we computed *relative* nonresponse bias, instead of the *absolute* nonresponse bias described in the formula above. The reason is that the dollar amounts vary substantially across expenditure categories, making comparisons difficult. For example, a nominal difference in dollars could be considered large for a lower expenditure variable but small for a high expenditure variable. Therefore, a relative bias percentage is a more appropriate statistic for comparisons across categories. The relative nonresponse bias is a percentage calculated by dividing the nonresponse bias by the adjusted base-weighted mean expenditures of all CUs and is shown below:

$$Relative\ Bias\ (\bar{Z}_R) = \frac{(\bar{Z}_R - \bar{Z}_T)}{\bar{Z}_T} \times 100\%$$

As a final point of clarification, the above formula was applied separately for each method. For Method 1, $\bar{Z}_R$ represents the base-weighted respondent mean and $\bar{Z}_T$ represents the final calibration weighted respondent mean, for Methods 2 and 3, $\bar{Z}_R$ represents the base-weighted respondent mean and $\bar{Z}_T$ represents the Propensity-weighted respondent mean (see sections 6.2 and 6.3), and for Method 4, $\bar{Z}_R$ represents the base-weighted "pseudo respondent" mean and $\bar{Z}_T$ represents the all base-weighted respondents' mean.

6. Description of the Four Methods Used in Calculation of Relative Bias

The following four methods were used to estimate the amount of nonresponse bias in the CE Surveys. The exact amount of nonresponse bias is unknown, so four methods of estimating it were developed to generate a range of plausible values for the true but unknown amount but taken together they give a good idea of the amount of nonresponse bias in the CE Surveys.

6.1. Method 1

Method 1 calculates nonresponse bias as the difference between the weighted estimate of the population mean prior to any nonresponse adjustment and the weighted estimate of the mean after nonresponse adjustment. "Base" weights were used for the weights prior to nonresponse adjustment, and "final" weights were used for the weights after nonresponse adjustment. The final weights include adjustments for both nonresponse and calibration, but the nonresponse adjustment is its largest component. This estimate of nonresponse bias assumes the nonresponding CUs are MAR, and the nonresponse adjustment factor in the final weights is a reasonable estimate of the inverse of each CU's response probability.

This method uses the general bias formula, $B(\hat{\bar{y}}) = E(\hat{\bar{y}}) - \bar{y}$, where $\hat{\bar{y}} = \frac{\sum_{i \in R} w_i^B y_i}{\sum_{i \in R} w_i^B}$ is a weighted estimate of the population mean expenditure ignoring nonresponse, w_i^B is the base-weight of the i-th CU, and R denotes the set of all respondents. $E(\hat{\bar{y}})$ can be estimated by $\hat{\bar{y}}$, and $\bar{y}$, can be estimated by an unbiased (or consistent) estimate that takes into account nonparticipation and more specifically, nonresponse. Assuming that the final calibration weighted estimate accounts for nonresponse, $\bar{y}$ can be estimated by $\hat{\bar{y}} = \frac{\sum_{i \in R} w_i^{F21} y_i}{\sum_{i \in R} w_i^{F21}}$, where w_i^{F21} is the final calibration weight. This estimate assumes that response is MAR and response correction to the final calibration weight is a reasonable estimate to the inverse of the response probability. Therefore, the nonresponse bias can be estimated by $\hat{B}^{F21}(\hat{\bar{y}}) = \frac{\sum_{i \in R} w_i^B y_i}{\sum_{i \in R} w_i^B} - \frac{\sum_{i \in R} w_i^{F21} y_i}{\sum_{i \in R} w_i^{F21}}$ where $Relative\ Bias\ (\bar{Z}_R) = \frac{(\bar{Z}_R - \bar{Z}_F)}{\bar{Z}_F} \times 100\%$ where $\bar{Z}_R$ is the base-weighted mean and $\bar{Z}_F$ is the final calibration weighted mean.

6.2. Method 2

Method 2 calculates the difference between the weighted estimates of the population mean before any nonresponse adjustments minus the propensity-weighted estimate. This propensity-weighted estimate is developed using a logistic regression model that contains socio-demographic variables. Initial research

showed that the surveys' response rates are affected by certain socio-demographic variables. Those variables were household tenure, urbanicity, region of residence, and many of their two-way interaction terms. Further research showed that CU size was also a good variable to use and was therefore added to the current model.

The selected Interview Survey model was:

$\boldsymbol{Ln}\,(\boldsymbol{p}/(\mathbf{1}-\boldsymbol{p}))$ = $\boldsymbol{\beta_0}$ + $\boldsymbol{\beta_1}$**I**(Rural) + $\boldsymbol{\beta_2}$**I**(Renter) + $\boldsymbol{\beta_3}$**I**(Tenure Other) + $\boldsymbol{\beta_4}$**I**(Midwest) + $\boldsymbol{\beta_5}$**I**(South) + $\boldsymbol{\beta_6}$**I**(West) + $\boldsymbol{\beta_7}$**I**(CU Size 1) + $\boldsymbol{\beta_8}$**I**(CU Size 2) + $\boldsymbol{\beta_9}$**I**(CU Size 3 or 4) + $\boldsymbol{\beta_{10}}$(Percentage of Noncontacts) + $\boldsymbol{\beta_{11}}$**I**(Rural $*$ Midwest) + $\boldsymbol{\beta_{12}}$**I**(Rural $*$ South) + $\boldsymbol{\beta_{13}}$**I**(Rural $*$ West)+ $\boldsymbol{\beta_{14}}$**I**(Renter $*$ Midwest) + $\boldsymbol{\beta_{15}}$**I**(Renter $*$ South) + $\boldsymbol{\beta_{16}}$**I**(Renter $*$ West) + $\boldsymbol{\beta_{17}}$**I**(Tenure Other $*$ Midwest) + $\boldsymbol{\beta_{18}}$**I**(Tenure Other $*$ South) + $\boldsymbol{\beta_{19}}$**I**(Tenure Other $*$ West) + $\boldsymbol{\beta_{20}}$**I**(CU Size 1*Rural) + $\boldsymbol{\beta_{21}}$**I**(CU Size 2*Rural) + $\boldsymbol{\beta_{22}}$**I**(CU Size 3 or 4 * Rural) + $\boldsymbol{\beta_{23}}$**I**(CU Size 1 * Renter) + $\boldsymbol{\beta_{24}}$**I**(CU Size 1 * Tenure Other) + $\boldsymbol{\beta_{25}}$**I**(CU Size 2 * Renter) + $\boldsymbol{\beta_{26}}$**I**(CU Size 2 * Tenure Other) + $\boldsymbol{\beta_{27}}$**I**(CU Size 3 or 4 * Renter) + $\boldsymbol{\beta_{28}}$**I**(CU Size 3 or 4 * Tenure Other) + $\boldsymbol{\beta_{29}}$**I**(CU Size 1 * Midwest) + $\boldsymbol{\beta_{30}}$**I**(CU Size 1 * South) + $\boldsymbol{\beta_{31}}$**I**(CU Size 1 * West) + $\boldsymbol{\beta_{32}}$**I**(CU Size 2 * Midwest) + $\boldsymbol{\beta_{33}}$**I**(CU Size 2 * South) + $\boldsymbol{\beta_{34}}$**I**(CU Size 2 * West) + $\boldsymbol{\beta_{35}}$**I**(CU Size 3 or 4 * Midwest) + $\boldsymbol{\beta_{36}}$**I**(CU Size 3 or 4 * South) + $\boldsymbol{\beta_{37}}$**I**(CU Size 3 or 4 * West) where $\boldsymbol{p}$ is the probability of response.

The Interview Survey and Diary Survey have similar models with the only difference being the exclusion of a few interaction terms not being used in the Diary Survey model because they were not statistically significant.

As a reminder, logistic regression is a model of the probability of outcomes of a binary process, such as whether a sample household participates in the CE Surveys.

To estimate nonresponse bias, Method 2 estimates $\bar{y}$ using the estimate of the CU's probability of responding:

$$\hat{B}^{Prop}(\hat{\bar{y}}) = \frac{\sum_{i\in R} w_i^B y_i}{\sum_{i\in R} w_i^B} - \frac{\sum_{i\in R}(w_i^B/p_i)y_i}{\sum_{i\in R}(w_i^B/p_i)} \quad \text{where } p_i = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n+\varepsilon)}}$$

is the estimate of the CU's response probability (propensity score), assuming that the non-response is MAR and $p_i$is a reasonable estimate of probability to respond. The resulting propensity scores will have a score between 0.0 and 1.0 and the reciprocal of this propensity will be multiplied by the current base-weight to get the adjusted base-weight. Relative bias can then be estimated by $(\bar{Z}_R) = \frac{(Z_R - Z_F)}{Z_F} \times 100\%$ where $\bar{Z}_R$ is the base-weighted mean and $\bar{Z}_p$ is the propensity adjusted base-weighted mean.

6.3. Method 3

Method 3 is nearly identical to Method 2 except that it contains a contact history variable (noncontact percentage) in the logistic regression model in addition to all of the socio-demographic variables discussed in Method 2. This contact history variable was calculated as the percent of noncontacts during the interview process and was found to have a strong relationship to propensity to respond. Having this variable in the model resulted in a much wider range of propensity scores than for those in Method 2. When the resulting propensity adjusted base-weighted means were calculated, there was noticeably more variance of the relative bias. Therefore, the model with this variable included was analyzed separately. Everything else pertaining to Method 2 described above also applied to Method 3.

6.4. Method 4

For Method 4, responders were divided into "pseudo responders" and "pseudo nonresponders" based on contact history. Responders who have high contact rates were treated as pseudo responders while those with low contact rates were treated as pseudo nonresponders since they were harder-to-contact. We assume that the pseudo nonresponders from the real respondent part of the sample behave like real nonrespondents regarding expenditure patterns. While not directly verifiable, this assumption is based on the theory known as the "continuum of resistance" to identify certain respondents to serve as pseudo nonrespondents. The theory suggests that sampling units can be ordered by the amount of interviewer effort needed to obtain a completed interview (Groves 2006) and was used in the previous nonresponse bias study.[14]

Using data collected in the Interview Survey Contact History Instrument (CHI), respondents were defined to be "harder to contact" when greater than 50 percent of the contact attempts resulted in noncontacts. The only exception was if there were two contact attempts resulting in one contact, these CUs were also considered "harder to contact" and were treated as pseudo nonresponders. This cut-off was selected to yield a response rate that coincided with the observed response rates during the 2010-2019 period covered by the data that ranged from the 53.7 percent 73.4 percent.

The formula used to calculate the relative bias was similar to those mentioned above except the numerator is the difference between the base-weighted mean of the pseudo respondents and the base-weighted mean of all respondents divided by the base-weighted mean of all respondents. It can be shown as follows:

$$Relative\ Bias\ (\bar{X}_R) = \frac{(\bar{X}_R - \bar{X}_T)}{\bar{X}_T} \times 100\%$$

Where:

$Relative\ Bias\ (\bar{X}_R)$ is the relative nonresponse bias % of the weighted sample mean,

$\bar{X}_R$ is the weighted mean of the pseudo respondent expenditures.

$\bar{X}_T$ is the weighted mean of all respondent expenditures,

6.5. Results Using the Four Methods to Quantify Relative Bias

Interview Survey Variables Analyzed

As has been discussed, four methods to estimate relative nonresponse bias were used in the analysis. The main variable used in the report is the summary variable, ZTOTALX4, that contains all CE Interview expenditures. This variable was analyzed by year to find out if relative bias has changed over time, and by

[14] See the August 2008 study "Assessing Nonresponse Bias in the CE Interview Survey: A Summary of Four Studies," by Boriana Chopova, Jennifer Edgar, Jeffrey Gonzalez, Susan King, Dave McGrath, and Lucilla Tan.

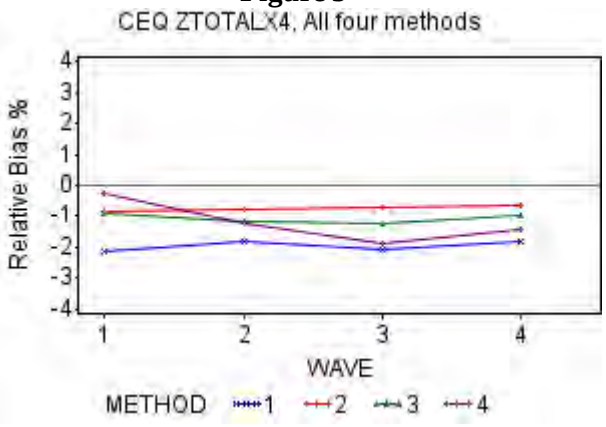
wave to find out if different interview waves contain more or less bias than others. As described above in the report, relative nonresponse bias point estimates are a percentage calculated by dividing the nonresponse bias by the adjusted base-weighted mean expenditures of all CUs.

Interview Survey Findings

Figures 4 and 5. Relative Bias by Year and Wave

Figure 4

Figure 5



ZTOTALX4 by Year

The graphs for Methods 1-4 in Figure 4 show negative relative bias for all three methods for years 2010-2014 while there are varying degrees of negative relative bias for the earlier years and in later years of the research period. In fact, starting in 2015, Method 2 and 3 for actually show a small positive bias. In general, the four methods have relative bias ranging from -1.8 percent to 0.8 percent. Over the ten-year period, relative bias shows a negative trend for Method 1, a bit of an upward trend for Method 2 and Method 3 and flat over the 2010-2019 period for Method 4.

The data points in Figure 4 also show the relative bias for all four methods by year do not differ that dramatically, especially in the early years 2010-2014. They are within one percentage point of each other for these early years of the research period but Methods 2 and 3 increase slightly starting in 2015. A slight negative relative bias implies that our responders spend a little less than our nonresponders and vice versa for positive relative bias. Method 4, which separates the responders into pseudo responders and pseudo nonresponders (described above), shows similar results.

Overall, Figure 4 shows that there is not total agreement regarding a recent relative bias trend where Method 2 has trended slightly positive where Method 3 has been flat while Method 1 has trended negative, but Method 4 has begun to trend higher. In addition, through 2014, Method 1 and Method 2 were similar but split beginning after 2014 with Method 1 now showing close to -1.5 percent while Method 2 trending very close to 0 percent.

ZTOTALX4 by Wave

The chart in Figure 5 clearly shows negative relative bias between -0.2 percent and -2.0 percent. There is some evidence that ZTOTALX4 expenditures for Wave 4 display slightly less negative relative bias than the other waves for most of the methods. Perhaps, this is due to an extra effort by the field reps to get the CUs to participate in the survey.

Interview Survey Summary

In summary, the four methods for the Interview Survey show the presence of slight negative relative bias over time and by wave. The level of negative relative bias varies by method and is generally in the range of -0.5 percent to -2.0 percent apart from 2015-2019 for Method 2 and Method 3. However, the relative bias does not appear to be strongly correlated to the decreasing response rates over the ten-year period.

Diary Survey Variables Analyzed

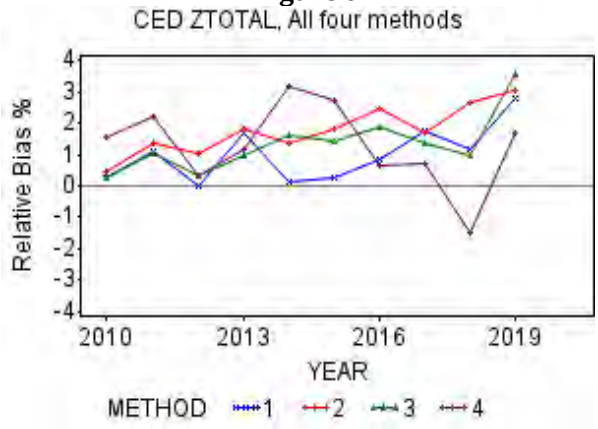
Similar to the Interview Survey, four methods to estimate relative bias were used in the Diary Survey analysis using the variable ZTOTAL (total expenditures). ZTOTAL was analyzed by year to find out if bias has changed over time.

The relative bias estimates for the Diary Survey using Methods 1-3 were calculated in the same manners as were their Interview Survey counterparts, with one difference: For the Diary Survey, a noncontact percentage greater than 40 percent was used as the cut-off to separate pseudo responders versus pseudo nonresponders. This cut-off was selected to yield a response rate between 52.8 percent and 71.5 percent, which corresponds to the Diary Survey's actual response rate during the research period.

Diary Survey Findings

Figure 6. Relative Bias by Year

Figure 6



ZTOTAL

In general, the four methods for ZTOTAL, the Diary Survey summary variable that contains all expenditures, show a non-negative relative bias (with one data point exception) in the range of 0.0 percent to 3.6 percent over the ten-year period. However, in 2019, all four Methods have slightly higher relative bias in the 1.7 percent to 3.6 percent range as shown in Figure 6 above. As a reminder, a slight positive relative bias implies that CE responders are spending a little more than the estimates for nonresponders, which is opposite to the results for most summary variables in the Interview Survey.

Diary Survey Summary

In summary, there is a slight level of positive relative bias for most of the summary variables in the Diary Survey using the all expenditures variable, ZTOTAL. When looking at the ZTOTAL graph, the four methods hint at a slight trend towards increasing positive bias but not as dramatically as the drop in Diary Survey response rates over the ten-year period. A strong correlation that would show increasing nonresponse relative bias, either positive or negative, related to declining response rates could be cause for concern because survey data may become inaccurate in representing expenditures and other items that CE publishes. A reminder that positive relative nonresponse bias is not a measure of respondents underreporting expenditures but instead compares the responders' actual reported expenditures to the four estimates of the nonresponders' reported expenditures.

7. Conclusion

In 2006, OMB issued a directive requiring any federal household survey with a response rate below 80 percent to perform a nonresponse analysis. Both the Interview and Diary Surveys have a response rate below 80 percent, so they are required to perform the nonresponse analysis. Each of the four studies in this report was designed to analyze nonresponse in the Interview and Diary Surveys by answering one or more of the following questions: (1) Are the data in the Interview and Diary Surveys MCAR? (2) What are the demographic characteristics of the nonrespondents and respondents? and (3) What additional information does the linear trend analysis provide regarding socio-demographic movement over the ten collection years.

The studies undertaken to find out whether missing data in the CE Surveys are MCAR were described in Section 3. Statistically significant differences were found by region of the country, PSU size class, urbanicity, and housing tenure for the Interview Survey and for all subgroups except Housing value for the Diary Survey. Likewise, the study comparing respondent demographic characteristics to the American Community Survey's population found statistically significant differences for most of the variables examined. Because statistically significant differences were found in each of these studies, we conclude that the data are not MCAR. No individual study was intended to provide a definitive answer to the questions raised in this research. However, all the studies conclude that the Interview and Diary Survey respondents and nonrespondents have different characteristics for many variables and the data are not MCAR.

The studies undertaken to estimate nonresponse bias in the CE Surveys were described in Section 5. The total expenditure summary variable for the Interview Survey, ZTOTALX4, was analyzed in detail to determine if there was a presence of relative nonresponse bias. Analysis of the Interview Survey presented robust graphic detail and tables of bias for ZTOTALX4 by year and wave. The results showed

a slight negative relative bias in a general -0.5 percent to -2.0 percent range over the ten-year period. This implies that the responders spent a little less than the nonresponders over the period and there was statistical evidence supporting this.

The Diary Survey total expenditures summary variable, ZTOTAL, was also analyzed in detail to determine if there was a presence of relative nonresponse bias. As opposed to the total expenditures variable in the Interview Survey, this variable showed a slight positive relative bias in a general 0.0 percent to 2 percent range over the ten-year period. This implies that the responders spent a little more than the nonresponders over the period.

None of the four methods was designed to exclusively find the exact level of relative bias but rather provide a range of estimates. Each method had its strengths and weaknesses and they differ enough to provide a realistic range of estimates for the analysis. Under the MAR[15] assumption, the conclusion is that the relative bias seems to be minor and not essentially important.

[15] Missing at Random means the propensity for a data point to be missing is not related to the missing data, but it is instead related to some of the observed data.

Appendix A

Interview Survey 2019 – Comparison of selected characteristics of Calibration-weighted CE respondents to the ACS

	ACS	CE
Gender (%) [1]		
Male	49.2	48.9
Female	50.8	51.1
Age (%)		
Under age 25	31.5	31.7
25-34	13.9	13.8
35-44	12.8	12.6
45-54	12.4	12.5
55-64	12.9	13.0
65-74	9.6	9.7
75 and over	6.9	6.7
Race (%) [1]		
White	72.0	78.4
Black	12.8	13.3
Other	15.2	8.4
Education [2] (%) [1]		
Less than high school	11.4	11.0
High school graduate	26.9	25.3
Some college/Assoc degree	28.6	27.9
College graduate	33.1	35.8
CU size (%) [1]		
1 person	28.3	30.1
2 persons	34.3	33.2
3 persons	15.3	14.3
4+ persons	22.1	22.4
Housing Tenure (%)		
Owner	64.1	63.6
Renter	35.9	36.4

	ACS	CE
No. of rooms in housing unit (%) [1]		
1	2.4	1.1
2	2.9	2.7
3-4	25.5	22.9
5-6	36.9	38.3
7-8	20.7	23.8
9 +	11.6	11.2
Owner-occupied housing value (%) [1]		
Less than $50,000	6.2	6.2
$50,000 to $99,999	10.2	9.7
$100,000 to $149,999	11.4	11.9
$150,000 to $199,999	13.2	13.8
$200,000 to $299,999	20.4	20.6
$300,000 to $499,999	21.4	21.6
$500,000 to $999,999	13.1	13.2
$1,000,000 +	3.9	3.0
Monthly rent (%) [1]		
Less than $500	14.1	15.3
$500 to $749	18.8	19.7
$750 to $999	18.8	19.7
$1,000 to $1,499	23.4	23.7
$1,500 to $1,999	11.4	10.5
$2000 +	8.7	8.7
No cash rent	4.8	2.4
CU income (%) [1]		
Less than $15,000	9.8	12.3
$15,000 to $24,999	8.3	10.7
$25,000 to $34,999	8.4	9.6
$35,000 to $49,999	11.9	13.1
$50,000 to $74,999	17.4	16.0
$75,000 to $99,999	12.8	11.5
$100,000 to $149,999	15.7	13.7
$150,000 to $199,999	7.2	6.2
$200,000 +	8.5	6.9

[1] Indicates a statistically significant difference ($p < 0.05$) between the Interview Survey and the ACS, using calibration-weighted CE respondents. For these comparisons, the Rao-Scott Chi-square test with the BRR variance method is used to reflect CE's sample design. The distributions for age and housing tenure are shaded in this table because their differences were not compared. The reason for excluding comparisons for these two variables is that they are used in calibration, meaning their replicate weights and final weight in the BRR procedure create design correction factors that are zero or very close to zero. This causes the resulting test statistics to become extremely large and their associated p-values to become extremely small. Therefore, the comparison of CE's distribution to ACS's distribution for these two variables is not practical.

[2] Comparison for persons age 25 and older

Appendix B

Comparison of CE's and ACS's demographic distributions over the 10-year period 2010–2019: The number of years the Rao-Scott chi-square statistic showed a statistically significant difference between CE and ACS ($p<0.05$) for both the Interview and Diary Surveys

Interview Survey:

Demographic characteristic	Calibration-weighted CE respondents vs. ACS
Gender	8
Age	*n.a.*
Race	10
Education	8
CU size	10
Tenure	*n.a.*
# Rooms in housing unit	10
Owner-occupied housing value	7
Monthly rent	10
CU income	10

Diary Survey:

Demographic characteristic	Calibration-weighted CE respondents vs. ACS
Gender	4
Age	*n.a.*
Race	10
Education	10
CU size	4
Tenure	*n.a.*
CU Income	10

Appendix C

Interview Survey: Linear Regression of time (by data collection year) on CE/ACS ratio

	CE subgroup percentage / ACS subgroup percentage Calibration-weighted CE Respondents	
Subgroup*	P-value	Slope
Gender		
Male	0.698	Positive
Female	0.705	Negative
Age		
Under age 25	0.024	Negative
25-34	0.143	Negative
35-44	0.543	Positive
45-54	0.571	Positive
55-64	0.052	Positive
65-74	0.027	Positive
75 and over	0.001	Positive
Race		
White	0.020	Positive
Black	0.795	Positive
Other	0.871	Positive
Education		
Less than high school	0.002	Negative
High school graduate	0.033	Negative
Some college/Assoc degree	0.014	Positive
College graduate	0.003	Positive
CU size		
1 person	0.168	Negative
2 persons	0.011	Positive
3 persons	0.106	Negative
4+ persons	0.120	Negative
Housing tenure		
Owner	0.003	Negative
Renter	0.002	Positive
Number of Rooms		
1	0.934	Negative
2	0.818	Negative
3-4	0.402	Positive
5-6	0.556	Positive
7-8	0.378	Positive
9 +	0.082	Negative

* Shaded data in this table show the subgroups where the β_1 coefficient is significant, as well as the direction of the slope for the ten-year regression line.

Interview Survey: Linear Regression of time (by data collection year) on CE/ACS relativity-- Continued

	CE subgroup percentage / ACS subgroup percentage Calibration-weighted CE Respondents	
Subgroup*	P-value	Slope
Owner-occupied housing value		
Less than $50,000	0.041	Positive
$50,000 to $99,999	0.598	Negative
$100,000 to $149,999	0.232	Negative
$150,000 to $199,999	0.102	Negative
$200,000 to $299,999	0.218	Negative
$300,000 to $499,999	0.311	Positive
$500,000 to $999,999	0.012	Positive
$1,000,000 +	0.697	Negative
Monthly rent		
Less than $500	0.007	Negative
$500 to $749	0.226	Positive
$750 to $999	0.055	Positive
$1,000 to $1,499	0.065	Positive
$1,500 to $1,999	0.003	Positive
$2000 +	0.019	Positive
No cash rent	0.002	Negative
CU income		
Less than $15,000	0.007	Positive
$15,000 to $24,999	0.000	Positive
$25,000 to $34,999	0.008	Positive
$35,000 to $49,999	0.323	Positive
$50,000 to $74,999	0.011	Negative
$75,000 to $99,999	0.151	Negative
$100,000 to $149,999	0.161	Negative
$150,000 to $199,999	0.095	Positive
$ 200,000 +	0.116	Positive

* Shaded data in this table show the subgroups where the β_1 coefficient is significant, as well as the direction of the slope for the ten-year regression line.

Appendix D

Diary Survey 2019 – Comparison of selected characteristics of Calibration-weighted CE respondents to the ACS

	ACS	CE
Gender (%)		
Male	49.2	49.2
Female	50.8	50.8
Age (%)		
Under age 25	31.5	31.7
25-34	13.9	13.8
35-44	12.8	12.6
45-54	12.4	12.5
55-64	12.9	13.0
65-74	9.6	9.7
75 and over	6.9	6.7
Race (%) [1]		
White	72.0	78.1
Black	12.8	13.3
Other	15.2	8.6
Education [2] (%) [1]		
Less than high school	11.4	10.7
High school graduate	26.9	25.1
Some college/Assoc degree	28.6	28.3
College graduate	33.1	35.9
CU size (%)		
1 person	28.3	29.9
2 persons	34.3	33.0
3 persons	15.3	14.7
4+ persons	22.1	22.4
Housing tenure (%)		
Owner	64.1	63.6
Renter	35.9	36.4
CU income (%) [1]		
Less than $15,000	9.8	10.5
$15,000 to $24,999	8.3	10.2
$25,000 to $34,999	8.4	9.8
$35,000 to $49,999	11.9	13.1
$50,000 to $74,999	17.4	16.8
$75,000 to $99,999	12.8	11.8
$100,000 to $149,999	15.7	14.5
$150,000 to $199,999	7.2	6.0
$200,000 +	8.5	7.2

[1] Indicates a statistically significant difference ($p < 0.05$) between the Diary Survey and the ACS, using calibration-weighted CE respondents. For these comparisons, the Rao-Scott Chi-square test with the BRR variance method is used to reflect CE's sample design. The distributions for age and housing tenure are shaded in this table because their differences were not compared. The reason for excluding comparisons for these two variables is that they are used in calibration, meaning their replicate weights and final weight in the BRR procedure create design correction factors that are zero or very close to zero. This causes the resulting test statistics to become extremely large and their associated p-values to become extremely small. Therefore, the comparison of CE's distribution to ACS's distribution for these two variables is not practical.

[2] Comparison for persons age 25 and older

Appendix E

Diary Survey – Linear Regression of time (by data collection year) on CE/ACS relativity

	CE subgroup percentage / ACS subgroup percentage Calibration-weighted CE Respondents	
Subgroup*	P-value	Slope
Gender		
Male	0.794	Negative
Female	0.791	Positive
Age		
Under age 25	0.020	Negative
25-34	0.145	Negative
35-44	0.518	Positive
45-54	0.400	Positive
55-64	0.049	Positive
65-74	0.023	Positive
75 and over	0.001	Positive
Race		
White	0.075	Negative
Black	0.104	Positive
Other	0.000	Positive
Education		
Less than high school	0.260	Negative
High school graduate	0.000	Negative
Some college/Assoc degree	0.020	Positive
College graduate	0.253	Positive
CU size		
1 person	0.683	Negative
2 persons	0.913	Negative
3 persons	0.254	Positive
4+ persons	0.392	Negative
Housing tenure		
Owner	0.003	Negative
Renter	0.002	Positive
CU income		
Less than $15,000	0.021	Positive
$15,000 to $24,999	0.299	Positive
$25,000 to $34,999	0.044	Positive
$35,000 to $49,999	0.726	Negative
$50,000 to $74,999	0.159	Negative
$75,000 to $99,999	0.584	Negative
$100,000 to $149,999	0.758	Positive
$150,000 to $199,999	0.531	Positive
$200,000 +	0.044	Positive

* Shaded data in this table show the subgroups where the β_1 coefficient is significant, as well as the direction of the slope for the ten-year regression line.

Appendix F

Interview Survey 2019 – Subgroup response rates by wave

Subgroup	Wave 1		Wave 2		Wave 3		Wave 4	
	n	Weighted Response Rate %	n	Weighted Response Rate %	n	Weighted Response Rate %	n	Weighted Response Rate %
Overall	10,115	56.02	10,055	53.60	10,128	52.84	10,091	53.93
Region[1,2,3 4]								
Northeast	1,954	50.26	1,920	48.22	1,905	48.75	1,893	50.10
Midwest	2,155	58.22	2,160	53.91	2,185	51.61	2,208	52.77
South	3,382	57.00	3,397	54.85	3,448	54.68	3,440	55.09
West	2,624	56.86	2,578	55.55	2,590	54.23	2,550	56.24
PSU size class[1, 2, 3, 4]								
Self-representing	4,270	50.93	4,284	48.87	4,272	48.60	4,254	49.26
Non-self-representing	5,267	58.00	5,210	55.48	5,276	54.85	5,262	56.04
Rural	578	70.62	561	67.49	580	61.37	575	64.59
Housing value - Owners								
Quartile 1-2	3,112	58.84	3,061	55.15	3,085	54.08	3,074	55.04
Quartile 3-4	2,845	58.02	2,860	54.18	2,879	53.19	2,906	54.50
Housing value - Renters[3]								
Quartile 1-2	1,588	52.77	1,565	52.72	1,586	52.97	1,541	52.89
Quartile 3-4	1,615	53.02	1,603	51.72	1,610	50.36	1,616	52.11
Urbanicity[1, 2, 3, 4]								
Urban	8,391	54.61	8,334	52.19	8,367	51.89	8,358	52.99
Rural	1,724	62.59	1,721	60.17	1,761	57.17	1,733	58.28
Housing tenure[1, 2, 3, 4]								
Owner	6,474	56.66	6,520	52.71	6,612	52.13	6,582	52.88
Renter	3,595	54.70	3,481	54.95	3,446	53.74	3,443	55.73
Other	46	61.88	54	64.44	70	64.18	66	59.04

[1, 2, 3, 4] Indicates a statistically significant difference ($p < 0.05$) was found for at least one comparison using the computed Rao-Scott chi-square statistic for the test of no association between survey participation and subgroup in waves 1, 2, 3 and 4 respectively.

Appendix G

Interview Survey Wave 4 comparison of subgroup response rates by year (2010-2019): Number of occurrences using Rao-Scott chi-square test (significance where $p < 0.05$)

Region	Northeast v. Midwest	Northeast v. South	Northeast v. West	Midwest v. South	Midwest v. West	South v. West
Higher	0	0	1	2	3	3
Lower	8	10	8	6	4	0
Not Significant	2	0	1	2	3	7
SCORE	-8	-10	-7	-4	-1	3

PSU size class	Self-representing v. Non-Self-representing	Self-representing v. Rural	Non-Self-representing v. Rural
Higher	0	0	0
Lower	9	9	5
Not Significant	1	1	5
SCORE	-9	-9	-5

Housing value - Owners	1st and 2nd Quartiles v. 3rd and 4th Quartiles
Higher	6
Lower	0
Not Significant	4
SCORE	6

Housing value - Renters	1st and 2nd Quartiles v. 3rd and 4th Quartiles
Higher	6
Lower	0
Not Significant	4
SCORE	6

Urbanicity	Urban v. Rural
Higher	0
Lower	8
Not Significant	2
SCORE	-8

Tenure	Owners v. Renters	Owners v. Other	Renters v. Others
Higher	0	0	0
Lower	10	9	4
Not Significant	0	1	6
SCORE	-10	-9	-4

Appendix H

Interview Survey: Relativity regression results for response rate comparison of subgroups

	Wave 1		Wave 2		Wave 3		Wave 4	
Subgroup*	P-value	Slope	P-value	Slope	P-value	Slope	P-value	Slope
Region								
Northeast	0.88873	Positive	0.61112	Positive	0.78111	Negative	0.51139	Negative
Midwest	0.37342	Positive	0.41514	Negative	0.22274	Negative	0.04188	Negative
South	0.28505	Negative	0.36411	Negative	0.98336	Positive	0.87270	Positive
West	0.52744	Positive	0.01883	Positive	0.05997	Positive	0.05477	Positive
PSU size class								
Self-representing	0.03539	Negative	0.14395	Negative	0.16954	Negative	0.07388	Negative
Non-self-representing	0.60718	Negative	0.64737	Negative	0.65847	Negative	0.99626	Negative
Rural	0.00004	Positive	0.00035	Positive	0.00183	Positive	0.00545	Positive
Housing – Owners								
Quartiles 1-2	0.10719	Positive	0.40687	Positive	0.40220	Positive	0.33878	Positive
Quartiles 3-4	0.00421	Positive	0.00640	Positive	0.02356	Positive	0.00118	Positive
Housing – Renters								
Quartiles 1-2	0.00014	Negative	0.00857	Negative	0.06381	Negative	0.03252	Negative
Quartiles 3-4	0.00382	Negative	0.08759	Negative	0.09115	Negative	0.02610	Negative
Urbanicity								
Urban	0.00095	Negative	0.10006	Negative	0.06758	Negative	0.10468	Negative
Rural	0.00027	Positive	0.05031	Positive	0.02326	Positive	0.07469	Positive
Housing Tenure								
Owner	0.00001	Positive	0.00212	Positive	0.06485	Positive	0.01560	Positive
Renter	0.00001	Negative	0.00168	Negative	0.02363	Negative	0.00778	Negative
Other	0.04018	Negative	0.48356	Negative	0.83441	Negative	0.67836	Negative

*****Shaded data in this table show the subgroups where the β_1 coefficient is significant, as well as the direction of the slope for the ten-year regression line.

Appendix I

Table I.1. Diary Survey: subgroup response rates for 2010-2014

Subgroup	2010		2011		2012		2013		2014	
	n	Response Rate %	n	Response Rate %	n	Response Rate %	n	Response Rate %	n	Response Rate %
Overall	19,988	71.92	19,823	70.29	20,298	67.74	20,296	60.65	20,476	64.80
Region [1]										
Northeast	4,146	68.23	4,042	68.14	4,079	66.35	4,056	58.35	4,084	64.87
Midwest	4,452	76.73	4,396	75.92	4,489	71.66	4,421	65.94	4,521	66.53
South	6,933	71.37	6,880	68.96	7,162	66.09	7,276	60.88	7,216	64.83
West	4,457	71.00	4,505	68.55	4,568	67.59	4,543	56.96	4,655	63.03
PSU size class [1]										
Self-representing	10,752	70.80	10,767	70.09	10,902	66.56	10,848	58.89	11,041	63.51
Non-self-representing	8,208	72.29	8,109	70.74	8,417	68.67	8,423	61.41	8,429	65.72
Rural	1,028	76.04	947	68.73	979	69.05	1,025	66.19	1,006	66.94
Housing Value - Owners										
Quartile 1-2	6,526	73.78	6,354	70.16	6,385	69.19	6,455	62.97	6,415	66.15
Quartile 3-4	5,296	71.75	5,208	71.96	5,361	69.70	5,263	61.03	5,316	66.90
Housing Value - Renters										
Quartile 1-2	2,877	70.36	2,834	68.49	2,903	64.83	2,921	58.69	2,999	61.56
Quartile 3-4	2,852	70.29	2,788	69.18	2,896	66.18	2,912	57.65	2,919	64.37
Urbanicity [1]										
Urban	16,521	71.58	16,508	69.68	16,844	66.91	16,757	59.69	17,050	63.82
Rural	3,467	73.28	3,315	72.88	3,454	71.24	3,539	64.55	3,426	69.02
Tenure [1]										
Owner	12,984	72.76	12,797	71.93	12,864	69.45	12,794	62.93	12,739	66.51
Renter	6,797	70.23	6,901	67.21	7,285	64.59	7,384	56.58	7,569	61.70
Other	207	73.22	125	68.43	149	71.79	118	60.97	168	74.05

[1] Indicates a significant difference ($p < 0.05$) was found for the computed Rao-Scott chi-square statistic for the test of no association between at least two subgroups for at least five of the ten years in the study.

Table I.2. Diary Survey: subgroup response rates for 2015–2019

Subgroup	2015		2016		2017		2018		2019	
	n	Response Rate %	n	Response Rate %	n	Response Rate %	n	Response Rate %	n	Response Rate %
Overall	20,517	57.73	20,391	56.75	20,100	58.36	20,124	55.42	20,238	53.27
Region [1]										
Northeast	3,817	58.50	3,855	59.48	3,808	55.93	3,802	51.92	3,688	47.01
Midwest	4,338	57.66	4,080	55.89	4,182	61.49	4,108	56.23	4,178	53.88
South	7,063	56.55	7,323	54.84	7,064	59.17	7,110	55.20	6,937	53.66
West	5,299	59.14	5,133	58.58	5,046	55.75	5,104	57.80	5,435	57.10
PSU size class [1]										
Self-representing	8,519	57.82	8,515	56.18	8,434	56.13	8,418	52.26	8,673	47.56
Non-self-representing	10,493	57.38	10,733	56.04	10,463	58.81	10,518	56.41	10,347	56.38
Rural	1,505	60.63	1,143	67.67	1,203	68.34	1,188	66.44	1,218	60.56
Housing Value - Owners										
Quartile 1-2	7,731	58.25	6,554	58.59	6,353	60.44	6,206	56.96	6,223	56.76
Quartile 3-4	4,741	59.88	5,836	60.40	5,843	61.74	5,775	60.94	5,743	55.98
Housing Value - Renters										
Quartile 1-2	3,320	55.64	3,210	53.88	3,164	52.11	3,131	50.23	3,102	48.60
Quartile 3-4	3,447	55.83	3,402	52.70	3,243	56.54	3,231	49.46	3,173	48.69
Urbanicity [1]										
Urban	16,851	57.26	17,006	55.46	16,596	57.03	16,657	54.44	16,743	51.79
Rural	3,666	60.01	3,385	63.07	3,504	64.46	3,467	60.03	3,495	60.03
Tenure [1]										
Owner	12,649	59.26	12,352	59.85	12,567	60.82	12,487	58.66	12,668	56.49
Renter	7,692	54.79	7,902	51.48	7,414	53.54	7,517	49.40	7,466	47.45
Other	176	73.97	137	72.43	119	81.52	120	75.62	104	66.63

[1] Indicates a significant difference ($p < 0.05$) was found for the computed Rao-Scott chi-square statistic for the test of no association between at least two subgroups for at least five of the ten years in the study.

Appendix J

Diary comparison of subgroup response rates by year (2010-2019):
Number of occurrences using Rao-Scott chi-square test (significance where $p < 0.05$)

Region	Northeast v. Midwest	Northeast v. South	Northeast v. West	Midwest v. South	Midwest v. West	South v. West
Higher	1	2	1	6	6	3
Lower	8	5	3	0	4	4
Not Significant	1	3	6	4	0	3
SCORE	-7	-3	-2	6	2	-1

PSU size class	Self-representing v. Non-Self-representing	Self-representing v. Rural	Non-Self-representing v. Rural
Higher	0	0	0
Lower	7	8	7
Not Significant	3	2	3
SCORE	-7	-8	-7

Housing value - Owners	1st and 2nd Quartiles v. 3rd and 4th Quartiles
Higher	2
Lower	5
Not Significant	3
SCORE	-3

Housing value - Renters	1st and 2nd Quartiles v. 3rd and 4th Quartiles
Higher	0
Lower	2
Not Significant	8
SCORE	-2

Urbanicity	Urban v. Rural
Higher	0
Lower	10
Not Significant	0
SCORE	-10

Tenure	Owners v. Renters	Owners v. Other	Renters v. Others
Higher	10	0	0
Lower	0	6	7
Not Significant	0	4	3
SCORE	10	-6	-7

Appendix K

Diary Survey: Relativity regression results for response rate comparison of subgroups

Subgroup*	P-value	Slope
Region		
Northeast	0.455	Negative
Midwest	0.032	Negative
South	0.317	Positive
West	0.072	Positive
PSU size class		
Self-Representing	0.024	Negative
Non-Self-Representing	0.225	Positive
Rural	0.007	Positive
Housing – Owners		
Quartiles 1-2	0.098	Positive
Quartiles 3-4	0.006	Positive
Housing – Renters		
Quartiles 1-2	0.001	Negative
Quartiles 3-4	0.009	Negative
Urbanicity		
Urban	0.002	Negative
Rural	0.001	Positive
Housing Tenure		
Owner	0.000	Positive
Renter	0.000	Negative
Other	0.001	Positive

*Shaded data in this table show the subgroups where the β_1 coefficient is significant, as well as the direction of the slope for the ten-year regression line.