

Assessing Nonresponse Bias in the Consumer Expenditure Interview Survey

Susan L. King* Boriana Chopova Jennifer Edgar Jeffrey M. Gonzalez
Dave E. McGrath Lucilla Tan

Abstract

The Consumer Expenditure Interview Survey (CE) is a nationwide survey conducted by the U.S. Bureau of Labor Statistics to estimate the expenditures made by American households. The response rate for the survey has varied between 74.5 and 78.6 percent over the past six years. In 2006, the Office of Management and Budget (OMB) issued a directive for any household survey with a response rate below 80 percent to conduct a study determining whether nonresponse introduces bias into the survey estimates. This paper is a synthesis of four studies undertaken to respond to OMB's directive. The four studies are: a comparison of response rates between subgroups of the survey's sample; a comparison of respondent demographic characteristics between the CE and the American Community Survey, an analysis of nonresponse bias using 'harder-to-contact' respondents as proxies for nonrespondents; an analysis of nonresponse bias using intermittent respondents and attriters as proxies for nonrespondents. Collectively, the studies show no meaningful bias in the survey's estimates even though the nonresponse is not missing completely at random.

Key Words: Continuum of resistance, intermittent respondent, missing data, panel survey, proxy nonrespondent

1. Introduction

The Office of Management and Budget (OMB) is concerned with the decreasing unit response rates in household and establishment surveys. To address this concern, OMB (2006) issued new standards and guidelines for federal statistical surveys. These guidelines require that, for any survey with a response rate below 80 percent, the data are analyzed to determine whether the missing values are missing completely at random (MCAR), and that an estimate of the nonresponse bias be provided.

The Consumer Expenditure Interview Survey (CE) is a nationwide household survey conducted by the U.S. Bureau of Labor Statistics (BLS) to estimate expenditures made by American households, and it is one federal statistical survey with a response rate below 80 percent. The response rate for the CE ranged from 74.5 percent to 78.6 percent between years 2002 and 2007. The design includes a rotating panel in which approximately 14,000 households are visited each quarter of the year, and each is contacted for an interview every three months for five consecutive quarters. Expenditure information from the first interview is excluded from published estimates and is only used for inventory and bounding purposes. This addresses a common problem in panel surveys in which respondents erroneously recall and report events (in this case, expenditures) to have occurred more recently than they actually did. Only expenditure information from the second through fifth interviews is used in the published estimates.

To evaluate nonresponse bias in the CE, four studies were completed: a comparison of response rates among subgroups of the survey's sample; a comparison of socio-demographic characteristics to the American Community Survey (ACS); an analysis of nonresponse bias using 'harder-to-contact' respondents as proxies for nonrespondents; and, an analysis of nonresponse bias using intermittent respondents and attriters as proxies for nonrespondents. These studies were designed to answer the questions: (1) Are the data in the CE MCAR? (2) What are the demographic characteristics of the nonrespondents? and, (3) What is the level of nonresponse bias in the CE?

*King.Susan@bls.gov - U.S Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Washington, DC 20212

2. Methodology: Common Approaches across Studies

2.1 Data

For comparability of results, the four studies used a common analysis file, based on data collected for interview waves 1, 2, and 5 from April 2005 through June 2006. This time period was selected because the Contact History Instrument (CHI) was first available in April 2005. CHI data were used in the identification of proxy nonrespondents in one of the studies. The unit of analysis in these studies is a consumer unit (CU), which in most cases is a household. The common data file consists of one record per wave per CU for waves 1, 2, and 5. For each record, there are CU-level variables as well as variables for the survey respondent.

2.2 Weighting

The CE sample design is a nationwide probability sample of addresses. Most addresses consist of one CU, but some addresses have more than one. Each interviewed CU represents itself as well as other CUs; therefore, each interviewed CU must be weighted properly to account for all CUs in the target population. The U.S. Bureau of the Census selects the sample and provides the base weights, which are the inverse of a CU's probability of selection. Each CU in a Primary Sampling Unit (PSU)¹ has the same base weight. BLS makes three types of adjustments to the base weights: an adjustment if the interviewer finds multiple housing units where only a single housing unit was expected; a noninterview adjustment; and, a calibration adjustment. These weighting adjustments are made to each CU. The noninterview adjustment accounts for nonresponse by increasing the weight of the respondents in socio-demographic classes that are thought to be associated with nonresponse. Calibration adjusts the weights to Census population controls in order to account for frame undercoverage. All studies presented here use base weights, but additional analyses using the nonresponse and calibration weights are provided in the study comparing CE respondents to ACS respondents.

2.3 Nonresponse Bias Formula

For the estimates of nonresponse bias in the two proxy nonrespondent studies, *relative* nonresponse bias was computed instead of an estimate of *exact* nonresponse bias. The reason is that the dollar amounts vary substantially across expenditure categories; thus, making comparisons among them difficult. Relative nonresponse bias is a more appropriate statistic for comparisons across categories. The following formula was used to compute relative nonresponse bias, denoted as $RelBias(\bar{Z}_R)$, in the base-weighted sample mean:

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

where:

- $B(\bar{Z}_R)$ is the *absolute* nonresponse bias in the base-weighted respondent sample mean;
- \bar{Z}_R is the base-weighted *respondent* mean of expenditures;
- \bar{Z}_T is the base-weighted *all CU* mean of expenditures;
- \bar{Z}_{NR} is the base-weighted *proxy nonrespondent* mean of expenditures;
- N_{NR} is the base-weighted number of *proxy nonrespondent* CUs; and,
- N is the base-weighted *total* number of CUs.

¹In the CE a PSU is a geographic entity that consists of several counties. The average number of counties in a PSU is about five.

2.4 Variance Estimation and Significance Testing

Estimates of means, frequencies, and variances were made using procedures designed for complex sample surveys. Means and frequencies were calculated using SAS[®]9: PROC SURVEYMEANS and SURVEYFREQ. For comparisons on categorical variables, the test statistic used was the adjusted Rao-Scott chi-square (SAS Institute Inc., 2004). For two-way comparisons, the null hypothesis tested was “no association between response status and subgroup”. For the comparison of CE sample characteristics to the ACS, the one-way comparison null hypothesis, assuming that the ACS distribution is the truth, is that both distributions are similar. Since the variance for relative nonresponse cannot be expressed in closed-form, the random groups method (Wolter, 1985) was used to estimate the variance of the relative nonresponse bias for each expenditure category mean. 95% confidence intervals for the relative nonresponse bias estimates were calculated in a standard manner but using the random groups variance estimate.

3. Individual Studies

In this section, the basic methods and the key findings of each study are highlighted. Detailed results (e.g., estimates and significance tests) from most studies are omitted here due to space constraints, but they are available upon request from the authors. Conclusions about CE respondents drawn collectively from all four studies are presented in Table 3 and estimates of relative nonresponse bias in Table 2.

3.1 Comparisons of Response Rates Across Subgroups

This study examined the response rates among subgroups that could be identified for both respondents and nonrespondents. The goal was to determine whether the survey’s respondents and nonrespondents had the same distribution on various characteristics. The subgroups analyzed were: region of the country (Northeast, Midwest, South, West), urbanicity (urban, rural), type of PSU, housing tenure (owner, renter, other), and housing values for owners and renters².

Base-weighted response rates were calculated for each subgroup separately for waves 1, 2, and 5. Base-weighted response rates answer the question: “What percent of the survey’s target population do the respondents represent?” These response rates were computed as the sum of base-weighted interviewed units divided by the sum of base-weighted interviewed units plus the units with Type A noninterviews. Type A noninterviews occur when no interview is completed at an occupied eligible housing unit³.

The base-weighted response rates suggest that response rates differ within all of the subgroups examined. In particular, statistically significant differences ($p < 0.05$) were found in the following pairwise comparisons within the subgroups:

- across regions, CUs in the Northeast and West have lower response rates than those in the Midwest and South;
- across types of PSU, CUs in metropolitan Core Based Statistical Areas (CBSAs) with a population of more than 2 million people have lower response rates than those in other types of PSUs;
- renters in the third and fourth quartiles of housing values have lower response rates than renters in the lower quartiles in the Unit and Area frames⁴, with a similar trend among homeowners; and,
- CUs in urban areas have lower response rates than those in rural areas.

²The information on housing values is from the 2000 decennial census instead of the CE. This means that the information is available for every CU, including both respondents and nonrespondents, but is slightly out-of-date.

³When possible, neighbors or field interviewers provide demographic information for nonrespondents. However, the quality of this data has not been evaluated.

⁴The unit frame has complete addresses, whereas the area frame has incomplete addresses. The unit frame covers most of the population.

Although there is evidence of an association between housing tenure and survey participation, there were no statistically significant differences between the pairwise comparisons of owners and renters participation in waves 2 and 5 (respondents who do not own or rent their homes had significantly higher response rates, but the number of these ‘other’ respondents is very small and the difference is not thought to be substantively meaningful).

In general, response rate differences among subgroups suggest that the data are not MCAR because the respondent and nonrespondent CUs are not simple cross sections of the original sample.

3.2 Comparisons of the CE Respondents to the ACS

Another approach to analyzing nonresponse bias is to compare the distribution of socio-demographic characteristics of respondents to that of a recent census or other ‘gold standard’ survey (Groves, 2006). The ‘gold standard’ survey chosen for this study was the 2005 ACS. The ACS satisfied three important criteria: (1) its estimates are considered to be very accurate; (2) it has key socio-demographic variables available; and, (3) it was conducted in a time period very close to that which was used to analyze the CE. The ACS is a mandatory survey with a response rate of 97.3% and a coverage rate of 95.1% (Census Bureau, 2006).

CE and ACS respondents were compared on the following characteristics: gender, age, race, educational attainment, household size, tenure, the number of rooms in the dwelling unit, housing value, rent, and CU income. Statistically significant differences ($p < 0.05^5$) were found between the two distributions for all comparisons and all types of weighting with only two exceptions, calibration-weighted age and housing. The majority of the statistically significant differences were smaller than six percentage points. However, larger differences were found for race and rent. The ACS reported 74.9% of the population as white, whereas the base, noninterview, and calibration weighted percentage of whites was 82.3%, 83.3%, and 81.4%, respectively. From the ACS, 20.5% of the population had a monthly rent less than \$500, whereas the interview survey reported 38.7%, 38.5%, and 38.4%. This indicates that the CE data are probably not MCAR.

There are several factors (e.g., the extremely large sample size) beyond the characteristics of the respondents that make differences likely to be statistically significant. For example, differences in proportions of subgroups of less than 1.5% between the ACS and CE distributions for the characteristics of gender, education attainment, and CU size were found to be statistically significant. In addition, the CE and the ACS collect data differently, the two surveys use different data collection modes, and the wording of the questions are different. As a result, the strength of the comparison to the ACS is limited by the extent to which the survey designs are truly comparable.

In short, the first study found that the data are not MCAR, and this study provided further evidence to substantiate that conclusion.

3.3 ‘Harder-to-Contact’ Respondents as Proxies for Nonrespondents

The third study uses ‘harder-to-contact’ respondents as proxies for nonrespondents. It draws on a theory known as the ‘continuum of resistance’ to identify a subset of 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 and those requiring the most effort should be similar to actual nonrespondents (Groves, 2006).

Using CHI data, a respondent was classified as ‘harder-to-contact’ when over 45 percent of their contact attempts resulted in noncontacts. This cut-off was selected to yield a response rate slightly under 80 percent, which is similar to the CE’s actual response rate during the study period. Also, the noncontact rate was chosen as the classification variable because this measure standardizes the amount of effort exerted by an interviewer to make contact across all sample units.

As an example, consider the contact history of a CU that had 6 contact attempts (Table 1). In this example, of the 6 contact attempts that were made, 2 resulted in contacts and

⁵All differences were significant at the $\alpha = 0.001$ level.

Table 1: Classification of contact attempts

Contact attempt	Classification
1. No one home	Noncontact
2. No one home	Noncontact
3. Got answering machine/service	Noncontact
4. No one home	Noncontact
5. Respondent too busy, appointment set	Contact
6. Complete case - ready to transmit	Contact

4 noncontacts, for a 67% noncontact rate. Since the noncontact rate is greater than 45%, this particular CU was classified as ‘harder-to-contact.’

In this study respondents and proxy nonrespondents were compared at each wave on the following socio-demographic characteristics: gender, marital status, race/ethnicity, age, educational attainment, household tenure, Census Region (Northeast, Midwest, South, West), urbanicity (urban, rural) and CU size. Statistically significant differences between the two groups ($p < 0.05$) across the three waves were found for respondent age, marital status, CU size, and census region. Other statistically significant differences occurred at individual waves for race, educational attainment, and housing tenure. Collectively, these differences suggest, once again, that the CE data may not be MCAR.

This study also estimated the relative nonresponse bias for total expenditures and for 13 expenditure categories. Table 2 shows the relative nonresponse bias estimates along with their 95% confidence intervals for waves 2 and 5. In wave 2, estimates of relative nonresponse bias for total expenditures [-0.14%, 95% CI: (-1.4%, 1.12%)] and 11 expenditure categories were not significantly different from zero. Two relative nonresponse bias estimates that were significantly different from zero were Health expenditures [3.68%, 95% CI: (1.80%, 5.56%)] and Reading materials expenditures [3.82%, 95% CI: (0.51%, 7.13%)]. Similar results were found for wave 5. It is worth noting that Health and Reading materials expenditures represent only 6% and 0.3%, respectively, of total spending; thus, the impact of any nonresponse bias from them is probably very small.

3.4 Pattern of Participation while in the Sample

The fourth study is based on the premise that nonrespondents are similar to respondents who failed to complete the entire panel of interviews (Reyes-Morales, 2003; 2007). Attriters and intermittent respondents were classified as proxy nonrespondents, and then these proxy nonrespondents were compared to respondents on socio-demographic variables and expenditures.

The study was based on a single cohort of CUs who had their first interview in April-June 2005, their second interview in July-September 2005, and their last interview in April-June 2006. The cohort had 3,071 unique CUs out of which 2,468 were used in this study. The CUs were then divided into three groups according to their pattern of participation in waves 2 through 5⁶: *complete respondents* (CUs who participated in the survey in all contiguous waves of the survey period), *attriters* (CUs who participated in the first wave for which they were eligible, and possibly completed the second and third waves [if eligible] but then refused to participate in all subsequent waves for which they were eligible), and *intermittent respondents* (CUs who participated in at least one but not all waves for which they were eligible). These groups comprised 78.6%, 14.1%, and 7.3% of the cohort, respectively.

The following set of demographic characteristics were analyzed in this study: household tenure (owner, renter), marital status, gender, respondent age, race, Hispanic origin, CU size, educational attainment, region, and urbanicity (urban, rural). Missing demographic information for proxy nonrespondents was imputed using a method similar to the ‘last observation carried forward’, where any missing values for a particular CU were imputed by

⁶For any wave, only CUs who resided at addresses eligible for inclusion in the sample were considered for the construction of the 3 analysis groups.

copying the values recorded for that CU in a previous interview (Verbeke and Molenberghs, 2000). This assumes that the demographic characteristics do not change from one wave to the next. Statistically significant differences ($p < 0.05$) were found between intermittent respondents and complete respondents with respect to age and Hispanic origin; while attriters were found to differ from complete respondents only with respect to age.

For relative nonresponse bias computations, we combined attriters and intermittent respondents to form one group of proxy nonrespondents and averaged expenditures across waves 2 through 5 for each expenditure category (Table 2). Estimates of relative nonresponse bias for total expenditures [-0.54%, 95% CI: (-2.31%, 1.24%)] and 10 expenditure categories were not significantly different from zero. However, relative nonresponse bias was significantly different from zero in three expenditure categories: Entertainment [3.46%, 95% CI: (0.35%, 6.57%)]; Personal insurance [3.82%, 95% CI: (1.72%, 5.93%)]; and, Transportation [-5.65%, 95% CI: (-9.11%, -2.19%)]. These categories represent 5.1%, 10.3%, and 19.1%, respectively.

4. Conclusions

All of the studies found that the data are not MCAR. Significantly different response propensities were found for various demographic characteristics in the studies on subgroup comparisons, respondent characteristics compared to the ACS, as well as the two proxy nonrespondent studies. Because statistically significant differences were found in each of these studies, we concluded that the data are not MCAR. Any characteristic for which a statistically significant difference was observed suggests that the respondent sample disproportionately represents particular subgroups of the survey's target population. Common findings across the studies are summarized in Table 3 and indicate that blacks are underrepresented among the respondents while those age 65 and over tend to be overrepresented.

Both proxy nonrespondent studies found no evidence of relative nonresponse bias in total expenditures. Between them, these studies found significant relative nonresponse bias estimates in expenditures on health, reading materials, entertainment, personal insurance, and transportation. With the exception of transportation and personal insurance, the other expenditures comprised less than 10% of total expenditures. In addition, because each study identified different categories to have a significant relative nonresponse bias and since some bias could be expected to occur at random, we conclude that the expenditure estimates derived from the CE are not subject to high levels of nonresponse bias.

No study by itself provides a definitive answer to the questions raised in this research. Taken together, the four studies indicate that nonresponse bias is not a significant problem for CE estimates, even though respondents and nonrespondents tend to be demographically dissimilar and the data not missing at random. The findings contradict the commonly held belief that if a survey's missing data are not MCAR, then its estimates are subject to nonresponse bias. From the nonresponse bias equation provided by OMB (2006), a mean estimate's nonresponse bias disappears if there is complete response or if the mean expenditure is similar for respondents and nonrespondents. For the CE, the absence of meaningful bias in total expenditures in spite of nonresponse suggests that the bias in an underrepresented group (e.g., blacks) is offset by a similar bias in an overrepresented group (e.g., the over-65 age group).

5. Acknowledgments

The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the U.S. Bureau of Labor Statistics. The authors would like to thank Karen Goldberg and Dave Swanson for helpful discussions of nonresponse bias and earlier versions of this paper.

REFERENCES

- Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70, 646-675.

- Office of Management and Budget (OMB) (2006). Standards and guidelines for statistical surveys. Washington, D.C. Retrieved September 11, 2007 from www.whitehouse.gov/omb/infoereg/statpolicy/standards.pdf.
- Reyes-Morales, S.E. (2003). Characteristics of complete and intermittent responders in Consumer Expenditure Quarterly Interview Survey. Internal BLS Report.
- Reyes-Morales, S.E. (2007). Characteristics of complete and intermittent responders in Consumer Expenditure Quarterly Interview Survey: 2001-2006. Internal BLS Report.
- SAS Institute Inc. (2004). SAS/STAT user's guide, version 9.1. Cary, NC: SAS Institute Inc.
- Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.
- Wolter, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

Table 2: Estimates of relative bias in the CE for expenditure categories using proxy nonrespondents

Expenditure Category	Share of total expenditures (%)	'Harder-to-Contact' respondents as proxy nonrespondents						Pattern of Participation			
		Wave 2			Wave 5			Across waves			
		Relative Bias (%)	Lower 95% CI	Upper 95% CI	Relative Bias (%)	Lower 95% CI	Upper 95% CI	Relative Bias (%)	Lower 95% CI	Upper 95% CI	Upper 95% CI
Total Expenditures	0.81	-0.14	-1.40	-1.12	-0.10	-1.18	0.98	-0.54	-2.31	1.24	
Alcoholic beverages	2.88	-0.96	-3.74	-1.82	-2.82	-6.18	0.54	-2.37	-9.09	4.36	
Apparel and services	3.90	0.96	-3.48	5.40	0.37	-1.85	2.59	-0.01	-5.24	5.21	
Cash contributions	1.90	2.09	-2.60	6.78	0.73	-2.02	3.48	4.24	-0.98	9.46	
Education	5.10	-3.74	-10.83	3.35	1.86	-6.08	9.80	-2.27	-11.81	7.28	
Entertainment	13.40	0.49	-2.49	3.47	0.40	-2.43	3.23	3.46	0.35	6.57	
Food	6.00	0.19	-0.83	1.21	0.76	-0.06	1.58	0.09	-1.35	1.53	
Health	33.13	3.68	1.80	5.56	3.21	0.88	5.54	1.03	-2.76	4.82	
Housing	0.63	-0.28	-1.40	0.84	-0.89	-1.94	0.16	-0.56	-2.26	1.14	
Personal care	10.32	0.32	-2.24	2.88	0.59	-0.91	2.09	1.95	-0.74	4.65	
Personal insurance	0.29	-0.51	-2.72	1.70	-0.68	-2.36	1.00	3.82	1.72	5.93	
Reading materials	0.72	3.82	0.51	7.13	3.41	0.28	6.54	4.22	-1.16	9.60	
Tobacco	19.12	-0.27	-4.07	3.53	-1.51	-3.43	0.41	2.51	-1.79	6.81	
Transportation		-1.96	-5.04	1.12	-0.53	3.59	2.53	-5.65	-9.11	-2.19	

Table 3: Summary of significant differences in characteristics of respondents across studies

Characteristic	Response Rate Across Subgroups	Comparison to ACS	'Harder-to-Contact' as Nonrespondents	Pattern of Participation	Conclusion
Gender	-	males under-represented	inconclusive	inconclusive	inconclusive
Age	-	55+ over-represented	55+ over-represented	65+ over-represented	55+ over-represented
Race	-	blacks under-represented	blacks under-represented	blacks under-represented	blacks under-represented
Education level	-	HS graduate underrepresented	inconclusive	inconclusive	inconclusive
Housing tenure	renters over-represented	renters under-represented	renters under-represented	inconclusive	inconclusive
Marital status	-	-	not married underrepresented	inconclusive	inconclusive
Census region	Midwest, South overrepresented	-	West over-represented	inconclusive	inconclusive
Urbanicity	urban under-represented	-	inconclusive	inconclusive	inconclusive
CU size	-	2-4 person CUs underrepresented	2-, 4-person CUs overrepresented	inconclusive	inconclusive