

Panel Conditioning in the U.S. Consumer Expenditure Survey

Stephanie Eckman¹ and Ruben Bach²

The U.S. Consumer Expenditure Interview Survey asks many filter questions to identify the items that households purchase. Each reported purchase triggers follow-up questions about the amount spent and other details. We test the hypothesis that respondents learn how the questionnaire is structured and underreport purchases in later waves to reduce the length of the interview. We analyze data from 10,416 four-wave respondents over two years of data collection. We find no evidence of decreasing data quality over time; instead, panel respondents tend to give higher quality responses in later waves. The results also hold for a larger set of two-wave respondents.

Key words: Measurement error; panel conditioning; consumer expenditure.

1. Introduction

The Consumer Expenditure Interview Survey (CE) collects data from US households about expenditures and income. CE data contribute to the calculation of the Consumer Price Index, which determines adjustments to social security and other federal payments, and are also analyzed by economists and policymakers to understand consumption patterns. The CE has a rotating panel design: each month, some households are interviewed for the first time, others for a second time, and so on. Respondents are interviewed up to four times over ten months (National Research Council 2013; U.S. Bureau of Labor Statistics 2016). In each survey wave, respondents report purchases that their household has made in the last three months. Every reported purchase triggers follow-up questions about, for example, the month of purchase and the amount spent. This structure may allow respondents to learn that reporting fewer purchases makes the survey shorter. The strategy of giving false answers to filter questions to reduce the length of surveys is called motivated misreporting. This effect has been detected across survey modes, countries and topics (Kreuter et al. 2011; Eckman et al. 2014; Bach et al. 2020).

In a panel survey, respondents may remember from one wave to the next how the filter questions work and underreport more in later waves. In this situation, motivated

¹ RTI International, Statistical Methods, 701 13th St NW, Suite 750, Washington D.C., 20005 U.S.A. Email: steph.eckman@gmail.com

² University of Mannheim, Mannheim, Germany. Email: r.bach@uni-mannheim.de

Acknowledgments: The research reported here was supported by Eckman's ASA-NSF-BLS Fellowship. The authors thank Janel Brattland, Laura Erhard, Parvati Krishnamurty, Adam Safir, Lucilla Tan, and Erica Yu for discussion on the ideas behind this research, help with data processing, and comments on draft versions of the article.

misreporting becomes a form of changes-in-reporting panel conditioning: participating in a panel survey changes the answers that respondents give (Yan and Eckman 2012; Sun et al. 2018). A well known example of this phenomenon occurs in the eight wave Current Population Survey (CPS), which measures labor force participation and unemployment. Each report of unemployment triggers follow-up questions, and respondents in later waves tend to underreport unemployment (Hansen et al. 1955; Bailar 1975; Solon 1986; Halpern-Manners and Warren 2012; Hirsch and Winters 2016; Krueger et al. 2017). Halpern-Manners and Warren (2012) present evidence that respondents do so to avoid the follow-up questions that a report of unemployment triggers.

The CE interview is longer than the CPS interview, with many more filter and follow-up questions. For this reason, we hypothesize that motivated misreporting in the CE could also worsen with additional waves. The National Research Council worried about this possibility as well in its evaluation of the CE (National Research Council 2013; 84–85):

It seems likely that respondents learn quickly in the first interview, and are reminded in each successive one, that the interview will last longer if they answer “yes” to these (filter) questions. . . . Fifty percent of field representatives said (this) happened frequently or very frequently.

We hypothesize that repeated filter questions over waves also affect responses to the follow-up questions. For the purchases respondents do report in later interviews, they might round the amount spent or not report the amount, because recalling or looking up these details requires additional effort.

We test for motivated misreporting and panel conditioning in Waves 2, 3, and 4 of the CE using data from four-wave respondents between October 2016 and September 2018. This article expands upon preliminary results reported in a short research note in the economics literature, which used one year of data Bach et al. (2020). Section 2 reviews relevant previous research. Section 3 describes the CE in more detail and presents our approach to estimating the size of the panel conditioning effect. The results of the analyses are given in Section 4 and discussed in Section 5.

2. Relevant Literature

In this section, we review several streams of literature relevant for this study and describe how our study fills a gap in the existing research.

2.1. Motivated Misreporting

Experimental manipulations of filter question wording have shown that respondents take shortcuts in a survey if given the chance. When the structure of the filter questions makes it obvious that each “yes” response triggers one or more follow-up questions, respondents adapt their reporting behavior and answer “yes” less frequently.

Evidence for motivated misreporting (also called “fatigue bias” (Lehnen and Reiss 1978) has been found in many survey modes: web (Bach et al. 2020), mobile web (Daikeler et al. 2020), telephone (Kreuter et al. 2011; Bach et al. 2020; Eckman et al. 2014) and face-to-face (Bach et al. 2020). The effect occurs across countries (Bach et al. 2020) and across topics,

such as: household purchases (Kreuter et al. 2011; Eckman et al. 2014), mental health (Jensen et al. 1999; Kessler et al. 1998; Duan et al. 2007), crime victimization (Lehnen and Reiss 1978), and labor market behavior (Eckman et al. 2014; Eckman and Kreuter 2018). Several studies also show that misreporting extends to the follow-up questions (Kreuter et al. 2011; Eckman and Kreuter 2018; Daikeler et al. 2020). Overall, there is clear evidence from methodologically sound (experimental) studies that, on average, respondents do misreport within a single survey when the structure of the interview allows it.

2.2. Panel Conditioning

Motivated misreporting can be especially harmful to the quality of survey responses in longitudinal surveys: respondents may recall how the filter questions work and use this knowledge to avoid follow-up questions in later waves. Motivated misreporting is one way in which respondents in longitudinal surveys may become *worse-reporters* over time. It is also possible, however, that respondents become *better-reporters* in later waves, due to increasing motivation or trust (Waterton and Lievesley 1989). Both the worse- and better-reporters hypotheses are forms of panel conditioning in which participating in a survey changes the answers respondents give.

Several studies find evidence for the worse-respondent hypothesis: for example, in reports of expenditures for home alteration and repair jobs (Neter and Waksberg 1964), functional limitations among the elderly (Mathiowetz and Lair 1994), and use of personal hygiene products (Nancarrow and Cartwright 2007). Likewise, Schonlau and Toepoel (2015) find that straightlining occurs more often among experienced respondents. Moreover, as mentioned above, CPS respondents underreport unemployment after their first interview (Hansen et al. 1955; Bailar 1975; Solon 1986; Halpern-Manners and Warren 2012; Hirsch and Winters 2016; Krueger et al. 2017). Other studies support the better-respondent hypothesis. Over waves, respondents respond more honestly (Struminskaya 2016), more accurately (Angel et al. 2017), and are less likely to give socially desirable answers (Waterton and Lievesley 1989).

However, other scholars find no evidence of either decreasing or increasing data quality over the waves of a panel survey. These studies investigate reports of health care utilization and expenditures (Cohen and Burt 1985), responses to knowledge questions (Struminskaya 2016) and attitudinal questions (Sun et al. 2018). An experimental test of the motivated misreporting effect over waves finds no evidence that respondents underreport more in the second wave in a monthly online panel (Bach and Eckman 2018).

The inconclusive findings regarding motivated misreporting in panel surveys (i.e., panel conditioning) are likely the result of the different methods used to measure panel conditioning. Identifying panel conditioning effects is difficult, in part because of the need to disentangle bias caused by panel attrition from bias caused by panel conditioning (Bach 2021). Unfortunately, most of the studies mentioned above do not disentangle these effects (for exceptions, see Halpern-Manners and Warren (2012); Struminskaya (2016); and Bach and Eckman (2018)). However, elimination of these confounding sources of error is essential if we wish to attribute changes in reporting behavior to repeated panel survey participation. Before we describe our approach to estimating panel conditioning, we briefly review previous research on motivated misreporting and panel conditioning in the CE.

2.3. Misreporting in the CE

A handful of studies have investigated motivated misreporting and panel conditioning in the CE. For a review of other forms of measurement error in the CE, see [Fricker et al. \(2015\)](#).

Previous research is mixed about whether motivated misreporting occurs in the first wave of the CE. Comparison of CE purchase reports to external measures of expenditures reveals that CE data are generally of high quality. Small and irregular purchases (such as clothing) and items that may be susceptible to bias due to social desirability (such as alcohol and tobacco) are most likely to be underreported ([Bee et al. 2015](#)). [McBride \(2013\)](#) investigates whether reports of household purchases decline over the course of the CE interview using data collected in May 2011, but does not find strong evidence for his hypothesis. [Eckman \(2020\)](#) estimates that respondents underreport purchases by about 5 percentage points in Wave 1 of the CE. However, [Bosley et al. \(1999\)](#) find no evidence of motivated misreporting in a small study modelled after the CE.

Regarding reporting *over waves* in the CE, previous studies find small to no evidence supporting decreasing quality ([Silberstein and Jacobs 1989](#); [Yan and Copeland 2010](#); [Shields and To 2005](#)). These studies account for potentially confounding effects due to attrition by restricting their analysis samples to all-wave respondents. Silberstein and [Jacobs \(1989, 296\)](#) find no changes across waves in more than half of the expenditure categories. In a follow-up paper, [Silberstein \(1990\)](#) detects more reports in Wave 1 relative to later waves, which can be interpreted as evidence of motivated misreporting panel conditioning. However, she interprets the finding as overreporting in Wave 1 due to telescoping. [Yan and Copeland \(2010\)](#) study CE reports from one quarter in 2008 and detect no panel conditioning effects in the mean number of expenditure types reported or in total expenditures. [Shields and To \(2005\)](#) analyze expenditures for trips and vacations reported by respondents between April 2001 and March 2002 and report small to no evidence for underreporting of expenditures in these categories. These studies do not clearly support either the worse- or better-reporters hypothesis.

Although these studies exploit the rotating structure of the CE to eliminate confounding sources of error (attrition), their findings are limited to mean or total expenses reported ([Silberstein and Jacobs 1989](#); [Silberstein 1990](#)), short periods of data collection ([Yan and Copeland 2010](#)) or a small subset of purchases ([Shields and To 2005](#)). For these reasons, we have conducted a new study of motivated misreporting over waves in the CE, using a diverse set of purchases and a large sample.

3. Consumer Expenditure Interview Survey

The CE uses face-to-face and telephone interviewing to administer a 60 minute interview about household purchases ([U.S. Bureau of Labor Statistics 2016](#)). Any adults in the household can respond to the survey, and households remain in the survey for four waves (ten months). Data collection is continuous: each month some households are participating for the first time and others for the second, third or fourth time ([U.S. Bureau of Labor Statistics 2016](#)).

We utilize the CE's rotating design to investigate whether respondents in later waves change their response behavior. We think of the respondents in a given month as having different exposure to treatment, the survey interview, and estimate the effects of increasing exposure. For a discussion of panel conditioning as a treatment effect, see [Bach and Eckman \(2019\)](#).

3.1. Data Preparation

If we simply compare the responses collected from Wave 1 respondents to those collected from Wave 2, 3 and 4 respondents in the same months, we risk confounding changes in reports with changes in who responds in each wave. For example, households with stable employment may be more likely to stay in the sample for four waves, but these households may also have higher incomes and thus different spending habits. We want to estimate the difference in reports over waves after controlling for differences in the characteristics of the respondents over waves. To do so, we used the approach of [Halpern-Manners and Warren \(2012\)](#) and [Halpern-Manners et al. \(2016\)](#).

Table 1 depicts the design. The different samples are shown down the columns. Sample A responded for the first time in January 2016 and for the second time in April 2016. In October 2016, when Sample A was responding for the fourth time, Sample D responded for the 3rd time, Sample G for the second time, and Sample J for the first time. For analysis, we use responses only from October 2016–September 2018 (shown in bold in the table) from four-wave respondents. The data from the prior months is used only to determine which cases in Samples A–I were four-time respondents. Similarly, data from months after September 2018 are needed only to identify which respondents in Samples AE–AM completed four waves of the survey. Consider Sample J, which responded for the first time in October 2016. Not all data collected from Sample J cases in that month are included in the analysis data set. We peek ahead to January, April, and July 2017 to see which cases in Sample J responded four times. We then include in the analysis data set responses in October 2017 from only those Sample J cases that went on to respond four times. The analysis data set contains 10,416 cases and 30,487 interviews. In each month, our data set contains responses from respondents in each wave, which eliminates confounding with time and seasonal patterns in purchases.

To test the robustness of our results, we repeat all analyses on two-wave respondents. We repeat all of the above data preparation steps to extract Wave 1 and Wave 2 responses collected October 2016–September 2018 from households that completed two or more waves of the survey. This data set contains 14,442 cases and 27,852 interviews.

The data preparation approach described above should remove attrition bias. To test that the data set is balanced across waves, we fit two multinomial regression models on the four-wave data set. The dependent variable in each is wave, taking values 1, 2, 3 and 4. The independent variables are housing unit, household and respondent characteristics. The first model ($n = 29,886$) drops cases with missing values in the independent variables. The second ($n = 30,487$) uses a data set where the missing values are filled via multiple imputation. Significant coefficients in the models would suggest that respondents in a wave differed from respondents in Wave 1. None of the variables is significantly correlated with wave at the 5% level. These results reassure us that our approach has eliminated confounding on these variables, suggesting that there are no meaningful differences (in these variables) between Wave 1 respondents and Wave 2, 3 and 4 respondents, aside from their exposure to the survey. Therefore, differences in reports between respondents with varying levels of panel experience should be due only to experience with the survey itself. See Appendix (Subsection 6.1) for more information on these models.

Table 1. Example to illustrate case selection method.

Month ^a	Sample																					
	A	B	C	D	E	F	G	H	I	J	...	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	
Jan 2016	1																					
Feb 2016		1																				
Mar 2016			1																			
Apr 2016	2		1																			
May 2016		2		1																		
Jun 2016			2		1																	
Jul 2016	3		2			1																
Aug 2016		3		2			1															
Sep 2016			3		2			1														
Oct 2016	4		3		2		2		1													
Nov 2016	4		4		3		2															
Dec 2016		4		4		3		2														
Jan 2017				4		3		2														
...																						
Jun 2018									3						2		1					
Jul 2018										3				2		1						
Aug 2018											3			2		1						
Sep 2018									4				3		2		1					
Oct 2017																						
Nov 2017																						
Dec 2017																						
Jan 2018																						
Feb 2018																						
Mar 2018																						
Apr 2018																						

Only data from bolded rows (October 2016–September 2018) appear in analysis data set

3.2. Analysis

After creating the analysis data set, we compare the means of four outcome variables across the waves: the number of purchases reported, the average amount spent, the number of reports where the amount is rounded, and the number of reports for which the amount spent is missing. We include expenditures from six sections of the CE: clothing, memberships, utilities, vehicle licensing expenses, vehicle operating expenses, and miscellaneous. These sections have high rates of purchases per item and are asked in the same way in each wave. If motivated misreporting panel conditioning is present, we should detect it in these sections. [Table 2](#) gives descriptive statistics for the four outcome variables for the four-wave and two-wave data sets.

Rounding is defined as a substantially higher frequency of a value relative to the frequency of neighboring values. To flag rounded values, we first divide the reported amounts (a) into five ranges (r): $a < \text{USD } 10$; $\text{USD } 10 \leq a < \text{USD } 100$; $\text{USD } 100 \leq a < \text{USD } 1,000$; $\text{USD } 1,000 \leq a < \text{USD } 10,000$; $a \geq \text{USD } 10,000$. (All amounts collected in the survey are rounded to the nearest whole dollar.) We next calculate the frequency of each reported amount (n_a) and order those frequencies by the amount. We then take the first differences – the difference in frequencies between two consecutive amounts: $d_a = n_a - n_{a+1}$. We find the average and standard deviation of these first differences within each range: $\bar{d}_r, \sigma_r(d)$. If the frequency of a given amount is more than two standard deviations larger than the average frequency, then that amount is flagged as rounded; that is: $I(d_a > \bar{d}_r + 2 \times \sigma_r(d))$ ([Wilson and Abdirizak 2017](#)).

Motivated misreporting predicts fewer purchases over waves, as hypothesized by the National Research Council report quoted above. It also predicts greater average amounts in later waves: respondents who want to reduce the burden of the survey might fail to report small (in price) purchases, increasing the average reported amount. Higher average amounts would indicate that only large purchases are reported and small purchases ignored. Following the motivated misreporting literature, we also expect to see more rounded amounts and more missing values in Waves 2, 3 and 4 relative to Wave 1 as respondents put less effort into the retrieval process. If we have eliminated attrition bias from the data set, simple mean comparisons are sufficient to test for changes in responses across waves ([Halpern-Manners and Warren 2012](#)).

All analyses were done in R 4.0.2 ([R Core Team 2020](#)); Appendix (Subsection 6.2) gives the packages used. The analyses are unweighted and do not account for the clustering in the sample, because the goal is not to make inference to the population but rather to investigate the response behavior of the panel respondents.

4. Results

For each case-wave in the four wave analysis data set, we calculate the number of purchases reported, the average amount of each purchase, the number of rounded amounts, and the number of missing amounts. [Figure 1](#) shows the mean of each measure by wave (indicated by a dot) as well as the 66% and 95% confidence intervals on the estimate of the mean.

Looking at the upper left graph, the average number of purchases reported by respondents increases from Wave 1 to Wave 2 and decreases again in Waves 3 and 4, but

Table 2. Descriptive statistics for outcomes on four-wave respondent data set.

Variable	N						
	Respondents	Respondent-waves	Mean	Median	Std dev	Min	Max
Four-wave respondents							
Number of purchases	10,416	30,487 ^a	9.28	8	6.84	0	74
Average amount	10,347 ^b	30,114 ^b	186.04	150	159.19	0 ^c	5111.11
Number of rounded amounts	10,347 ^b	30,114 ^b	2.57	2	2.38	0	26
Number of missing amounts	10,347 ^b	30,114 ^b	0.10	0	0.45	0	14
Two-wave respondents							
Number of purchases	14,442	27,852 ^a	9.03	7	7.09	0	90
Average amount	14,323 ^b	27,432 ^b	190.30	150	141.8	0 ^d	15724.67
Number of rounded amounts	14,323 ^b	27,432 ^b	2.66	2	2.58	0	26
Number of missing amounts	14,323 ^b	27,432 ^b	0.12	0	0.23	0	9

^a Number of Respondent-Waves is less than 4 (or 2) times the number of respondents because some waves fall outside October 2016–September 2018 window

^b Excludes respondents (respondent-waves) with no reported purchases

^c In 10 respondent-waves, all reported amounts were USD0

^d In 19 respondent-waves, all reported amounts were USD0

no wave is significantly different from Wave 1. This pattern does not match our expectations that the average number of reported purchases would decline over waves. In the upper right graph, we see lower average amounts reported in Waves 2, 3 and 4, though only Waves 3 and 4 differ significantly from Wave 1 ($t = -3.41$, $t = -2.52$, respectively). We interpret lower average amounts as an indicator of better data quality – respondents take the time to report small purchases that they could underreport if they wanted to make the survey shorter. In the bottom row of Figure 1, there are fewer rounded and missing amounts in later waves than in the first wave; the means in all waves are significantly lower than Wave 1 at any conventional threshold ($t < -5.5$ for all comparisons of later waves with Wave 1). These results are clear evidence that four-wave respondents become better reporters after the first interview.

The results are similar when we broaden the analysis sample to include those who responded in two (or more) waves (Figure 2). Again the number of reported purchases and the average amount spent are not significantly different across waves. The decrease in the number of rounded and missing amounts is highly significant ($t < -7.8$).

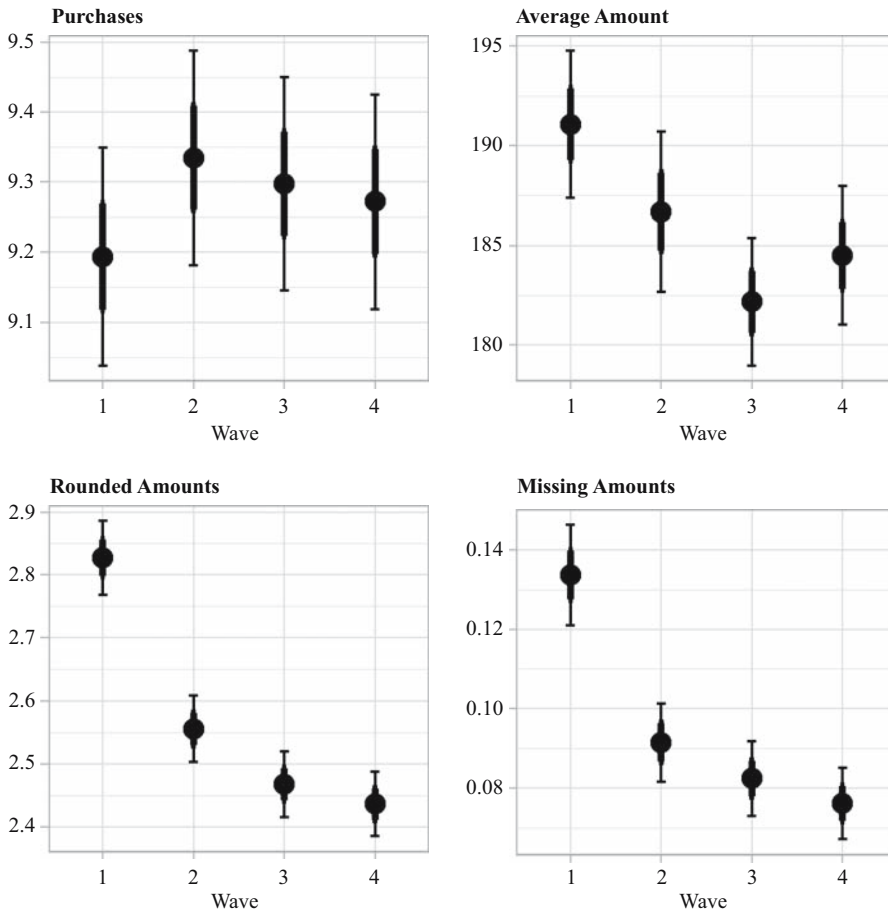


Fig. 1. Estimates of means of outcome variables among four-wave respondents, by wave. Note: Figures show mean and 66% (thick lines) and 95% (thin lines) confidence intervals.

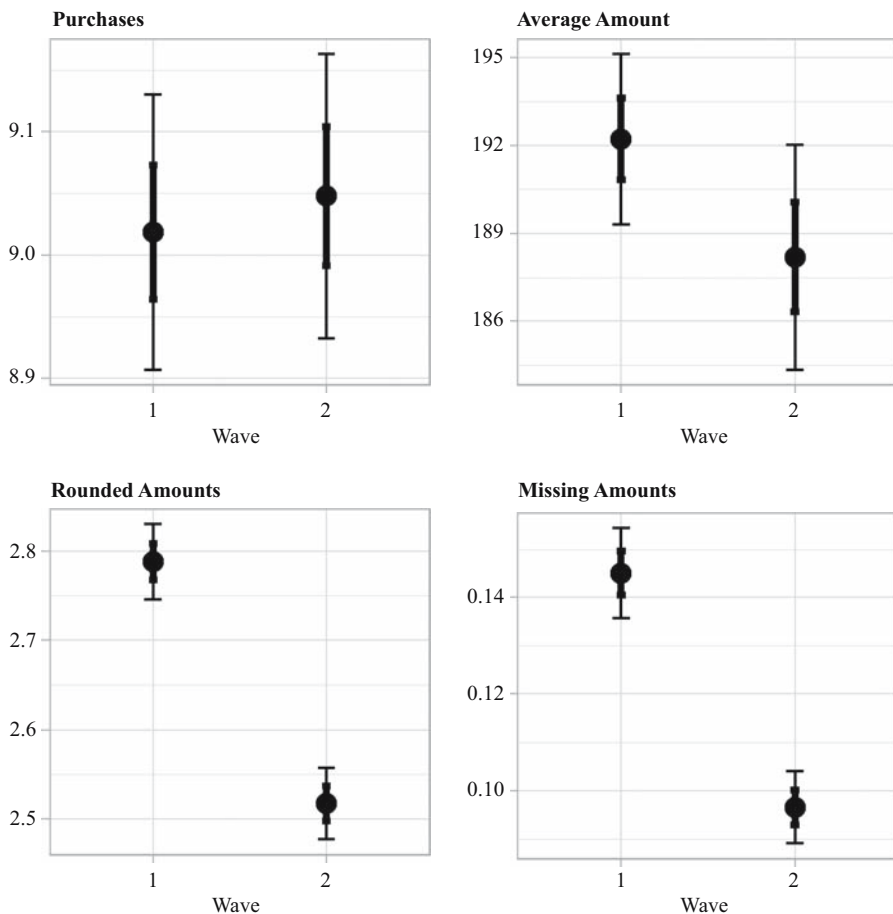


Fig. 2. Estimates of means of outcome variables among two-wave respondents, by wave. Note: Figures show mean and 66% (thick lines) and 95% (thin lines) confidence intervals.

We find no evidence that respondents on average underreport purchases or introduce more measurement error in later waves of the CE. In fact, there is strong evidence that they become better reporters over the four waves.

5. Discussion

We have tested the hypothesis that CE respondents underreport purchases in later waves to reduce the length and burden of the survey. To estimate the effects of panel conditioning, we took advantage of the rotating structure of the CE and limited analysis to respondents who completed four interviews. This approach to measuring panel conditioning should eliminate confounding with attrition. Our results show that panel conditioning occurs in the CE; respondents change the way they respond over time as a result of participating in the study. However, the direction of the changes is different than we hypothesized. Contrary to expectations, respondents become better reporters in later waves. Four-wave respondents do not appear to underreport purchases in later waves to avoid follow-up questions. In later waves, they are less likely to report rounded and missing amounts.

These results are not good news for all users of CE data, however. Economists often use the longitudinal CE data to understand how household purchases change over time in response to policy or income changes (Parker et al. 2014). If measurement error were constant over waves, it would fall out of estimates of changes and not bias results. However, our results show that measurement error is not constant. Thus, the panel conditioning we detect in this study can introduce bias into longitudinal estimates. We recommend that researchers carefully think through how changing measurement error over waves may affect the results of their analyses.

The results presented here help clear up the contradictory findings in previous investigations of motivated misreporting in the CE. Eckman (2020) detects underreporting in Wave 1, but Bosley et al. (1999) does not. However, the latter study used a panel design and tested for underreporting only in the second wave. The results in this article suggest that behavior in the second wave is different than it is in the first.

However, if respondents underreport in Wave 1, why does the effect not grow in later waves? A possible explanation is that respondents become more engaged with the interviewing task over time (Waterton and Lievesley 1989; Yan and Eckman 2012). After the first interview, respondents may realize how important it is that they give accurate reports of their spending. Previous research has shown that respondents make more use of records in later waves (Edgar 2010) and record use is associated with fewer rounded and missing responses (Edgar and Gonzalez 2009).

A second explanation is that motivated misreporting occurs in the CE through attrition rather than through measurement error. That is, respondents who found the survey repetitive or burdensome may simply choose not to participate in additional waves. If so, only those who did not find the filter questions burdensome would go on to the Waves 2, 3 and 4, explaining why we do not find motivated misreporting in these waves.

We can find some support for the hypothesis that motivated misreporting in the CE leads to attrition rather than measurement error by comparing the CE to the CPS. Both are longitudinal surveys collected by U.S. Census Bureau interviewers and selected from the same household frame. But, as discussed above, CPS respondents engage in motivated misreporting in later waves, while CE respondents do not. What might explain the different response behavior in the CE and CPS? An important difference between the two surveys is their response rates. In September 2018, the last month in our analysis data set, the CE response rate was 56.5%; in the same month, the CPS response rate was 85.1% (U.S. Bureau of Labor Statistics 2020). It may be that CE respondents who would underreport in later waves attrit instead, whereas CPS respondents do take part in later waves and underreport unemployment. However, Bach et al. (2020) finds little evidence that reluctant respondent engage in more motivated misreporting than others. We hope future research will investigate this hypothesis further.

6. Appendix

6.1. A Balance Across Waves Among Four-Wave Respondents

The approach we use to study panel conditioning restricts analysis to data collected from four-wave respondents between October 2016 and September 2018. In this data set, we compare the responses given by respondents at different waves. This approach relies upon

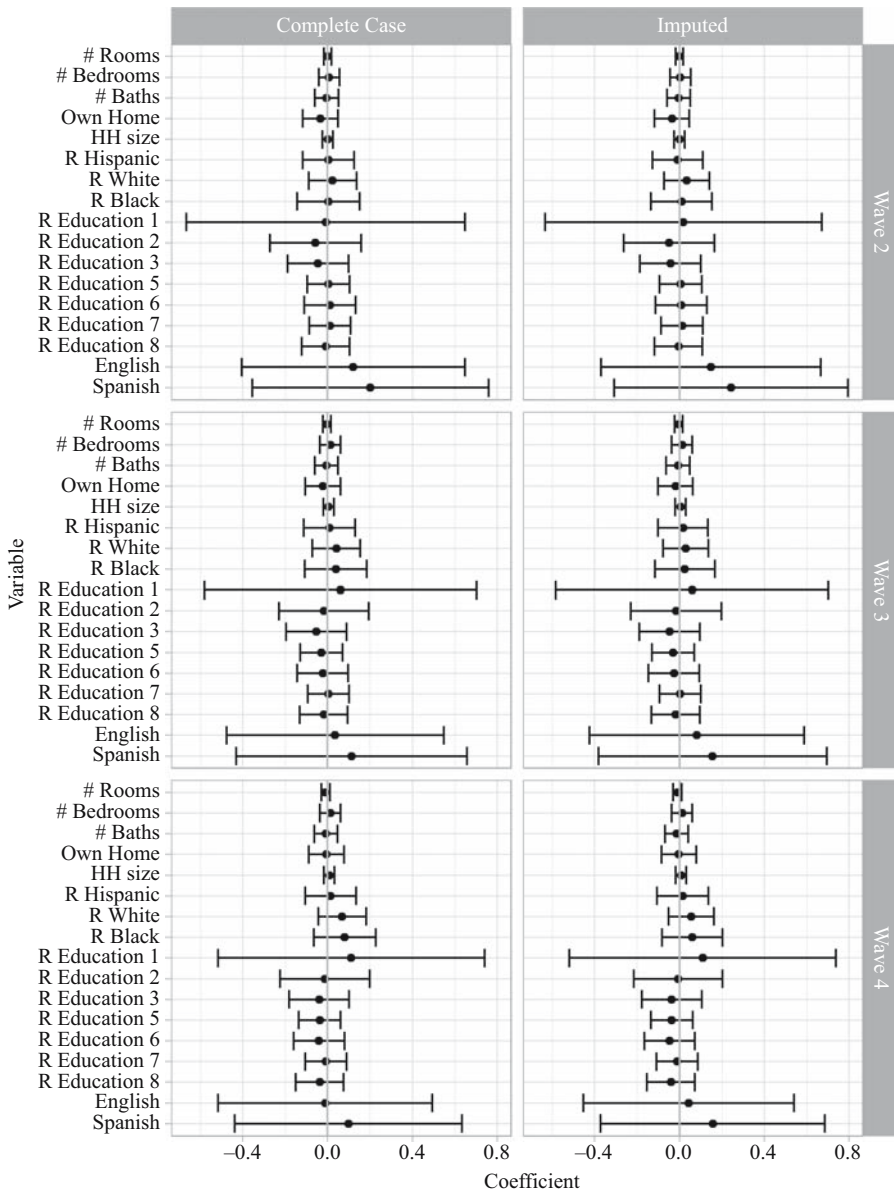


Fig. 3. Balance in household and respondent characteristics over waves coefficients capture deviations from baseline (Wave 1).

the assumption that there are no other differences in the respondents at each wave except their exposure to the survey.

We test this assumption with a multinomial model. The dependent variable in the model is the wave indicator. The independent variables are household and respondent characteristics collected in the CE. The household characteristics used are number of rooms, bedrooms and bathrooms in the housing unit, tenure, and the number of people in the household. The values of these variables are taken from the first-wave interview. The respondent characteristics used are race (white/African American/other), education (eight categories), and language of interview. Because the respondent can change over waves, the values of these variables are taken from each wave's interview. We do not expect that the respondent characteristics themselves suffer from panel conditioning, because they are collected early in the interview and do not trigger follow-up questions. About 2% of cases ($n = 601$) had missing values in one or more of the independent variables. We imputed values for the missing data using predictive mean matching (Little 1988) and created five imputed data sets.

Figure 3 displays the estimated coefficients and their standard errors from the two models. The right side shows the coefficients from the model run on the complete cases ($n = 29,886$). The left side shows the results from the model run on the imputed data set ($n = 30,487$). In each model, no coefficients are significantly different from zero at the 5% level, evidence that there are no substantial differences among the cases in our analysis data set by wave.

6.2. R Packages Used

The following R packages were used in analysis: tidyverse (Wickham et al. 2019), mice (Van Buuren and Groothuis-Oudshoorn 2011), nnet (Venables and Ripley 2002), table (Huntington-Klein 2020), Hmisc (Harrell Jr et al. 2020), jtools (Long 2020), skimr (Waring et al. 2020), and markdown (Allaire et al. 2019).

7. References

- Allaire, J.J., J. Horner, Y. Xie, V. Marti, and N. Porte. 2019. *markdown: Render Markdown with the C Library 'Sundown'*. R package version 1.1. Available at: <https://CRAN.R-project.org/package=markdown> (accessed November 2020).
- Angel, S., R. Heuberger, and N. Lamei. 2017. "Differences Between Household Income from Surveys and Registers and How These Affect the Poverty Headcount: Evidence from the Austrian SILC." *Social Indicators Research*, 138(2):575–603. DOI: <https://doi.org/10.1007/s11205-017-1672-7>.
- Bach, R.L. 2021. "A Methodological Framework for the Analysis of Panel Conditioning Effects." In *Measurement Error in Longitudinal Data*. In *Measurement Error in Longitudinal Data*, edited by A. Cernat and J. Sakshaug. Oxford University Press, Oxford, UK: 19–41.
- Bach, R.L., and S. Eckman. 2018. "Motivated Misreporting in Web Panels." *Journal of Survey Statistics and Methodology*, 6(3):418–430. DOI: <https://doi.org/10.1093/jsam/smx030>.

- Bach, R.L., and S. Eckman. 2019. "Participating in a Panel Survey Changes Respondents' Labour Market Behaviour." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(1):263–281. DOI: <https://doi.org/10.1111/rssa.12367>.
- Bach, R., and S. Eckman. 2020. "Rotation Group Bias in Reporting of Household Purchases in the U.S. Consumer Expenditure Survey." *Economics Letters*. DOI: <https://doi.org/10.1016/j.econlet.2019.108889>.
- Bach, R.L., S. Eckman, and J. Daikeler. 2020. "Misreporting Among Reluctant Respondents." *Journal of Survey Statistics and Methodology*, 8(3):566–588. DOI: <https://doi.org/10.1093/jssam/smz013>.
- Bailar, B.A. 1975. "The Effects of Rotation Group Bias on Estimates from Panel Surveys." *Journal of the American Statistical Association*, 70(349):23. DOI: <https://doi.org/10.2307/2285370>.
- Bee, A., B.D. Meyer, and J.X. Sullivan. 2015. "The Validity of Consumption Data. Are the Consumer Expenditure Interview and Diary Surveys Informative?" In *Improving the Measurement of Consumer Expenditures*, edited by C.D. Carroll, T.F. Crossley, and J. Sabelhaus: 204–240. University of Chicago Press, Chicago.
- Bosley, J.M. Dashen, and J. Fox. 1999. "When Should We Ask Follow-up Questions About Items in Lists?" In *Proceedings of the Section on Survey Research Methods, American Statistical Association*. Available at: http://www.asasrms.org/Proceedings/papers/1999_131.pdf (accessed November 2020).
- Cohen, S.B., and V.L. Burt. 1985. "Data Collection Frequency Effect in the National Medical Care Expenditure Survey." *Journal of Economic and Social Measurement*, 13(2):125–151.
- Daikeler, J., R.L. Bach, H. Silber, and S. Eckman. 2020. "Motivated Misreporting in Smartphone Surveys." *Social Science Computer Review*. DOI: <https://doi.org/10.1177/0894439319900936>.
- Duan, N., M. Alegria, G. Canino, T. McGuire, and D. Takeuchi. 2007. "Survey Conditioning in Self-Reported Mental Health Service Use: Randomized Comparison of Alternative Instrument Formats." *Health Research and Educational Trust*, 42(2):890–907. DOI: <https://doi.org/10.1111/j.1475-6773.2006.00618.x>.
- Eckman, S. 2020. "Underreporting of Purchases in the U.S. Consumer Expenditure Survey." Available at: <https://osf.io/e3sgz/> (accessed November 2020).
- Eckman, S., and F. Kreuter. 2018. "Misreporting to Looping Questions in Surveys: Recall, Motivation and Burden." *Survey Research Methods*, 12(1):59–74. DOI: <https://doi.org/10.18148/srm/2018.v12i1.7168>.
- Eckman, S., F. Kreuter, A. Kirchner, A. Jäckle, R. Tourangeau, and S. Presser. 2014. "Assessing the Mechanisms of Misreporting to Filter Questions in Surveys." *Public Opinion Quarterly*, 78(3):721–733. DOI: <https://doi.org/10.1093/poq/nfu030>.
- Edgar, J. 2010. "Respondent Record Use in U.S. Consumer Expenditure Survey." In Presented at the American Association for Public Opinion Research Conference, May 13–16, 2010, Chicago, Illinois, U.S.A. Available at: https://www.bls.gov/cex/research_papers/pdf/Respondent-Record-Use-in-the-US-Consumer-Expenditure-Survey-Presentation.pdf (accessed November 2020).
- Edgar, J., and J. Gonzalez. 2009. "Correlates of Data Quality in the Consumer Expenditure Quarterly Interview Survey." In Proceedings of the Survey Research Methods Section

- of the American Statistical Association August 1–6, 2009, Washington D.C., U.S.A. Available at: <https://www.bls.gov/osmr/research-papers/2009/pdf/st090140.pdf> (accessed November 2020).
- Fricker, S, B. Kopp, L. Tan, and R. Tourangeau. 2015. “A Review of Measurement Error Assessment in a U.S. Household Consumer Expenditure Survey.” *Journal of Survey Statistics and Methodology*, 3(1):67–88. DOI: <https://doi.org/10.1093/jssam/smu025>.
- Halpern-Manners, A., and J.R. Warren. 2012. “Panel Conditioning in Longitudinal Studies: Evidence From Labor Force Items in the Current Population Survey.” *Demography*, 49(4):1499–1519. DOI: <https://doi.org/10.1007/s13524-012-0124-x>.
- Halpern-Manners, A., J.R. Warren, and F. Torche. 2016. “Panel Conditioning in the General Social Survey.” *Sociological Methods and Research*, 46(1): 103–124. DOI: <https://doi.org/10.1177/0049124114532445>.
- Hansen, M.H., W.N. Hurwitz, H. Nisselson, and J. Steinberg. 1955. “The Redesign of the Census Current Population Survey.” *Journal of the American Statistical Association*, 50(271):701–719. DOI: <https://doi.org/10.2307/2281161>.
- Harrell Jr, F.E. 2020. with contributions from C. Dupont, and many others. *Hmisc: Harrell Miscellaneous*. R package version 4.4-1. Available at: [https://CRAN.R-project.org/package = Hmisc](https://CRAN.R-project.org/package=Hmisc) (accessed November 2020).
- Hirsch, B.T., and J.V. Winters. 2016. “Rotation Group Bias in Measures of Multiple Job Holding.” *Economic Letters*, 147:160–163. DOI: <https://doi.org/10.1016/j.econlet.2016.08.039>.
- Huntington-Klein, N. 2020. *vtable: Variable Table for Variable Documentation*. R package version 1.2.1. Available at: [https://CRAN.R-project.org/package = vtable](https://CRAN.R-project.org/package=vtable) (accessed November 2020).
- Jensen, P.S., H.K. Watanabe, and J.E. Richters. 1999. “Who’s Up First? Testing for Order Effects in Structured Interviews Using a Counterbalanced Experimental Design.” *Journal of Abnormal Child Psychology*, 27:439–445. DOI: <https://doi.org/10.1023/A:1021927909027>.
- Kessler, R.C., H.-U. Wittchen, J.M. Abelson, K. McGonagle, N. Schwarz, K.S. Kendler, B. Knäuper, and S. Zhao. 1998. “Methodological Studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey (NCS).” *International Journal of Methods in Psychiatric Research*, 7(1):33–55. DOI: <https://doi.org/10.1002/mp.33>.
- Kreuter, F., S. McCulloch, S. Presser, and R. Tourangeau. 2011. “The Effects of Asking Filter Questions in Interleafed versus Grouped Format.” *Sociological Methods and Research*, 40(88):88–104. DOI: <https://doi.org/10.1177/0049124110392342>.
- Krueger, A.B., A. Mas, and X. Niu. 2017. “The Evolution of Rotation Group Bias: Will the Real Unemployment Rate Please Stand Up?” *Review of Economics and Statistics*, 99(2):258–264. DOI: https://doi.org/10.1162/REST_a_00630.
- Lehnen, R.G., and A.J. Reiss. 1978. “Response Effects in the National Crime Survey.” *Victimology*, 3:110–160.
- Little, R.J.A. 1988. “Missing-Data Adjustments in Large Surveys.” *Journal of Business and Economic Statistics*, 6(3):287. DOI: <https://doi.org/10.2307/1391878>.

- Long, J.A. 2020. *jtools: Analysis and Presentation of Social Scientific Data*. R package version 2.1.0. Available at: [https://cran.r-project.org/package = jtools](https://cran.r-project.org/package=jtools) (accessed November 2020).
- Mathiowetz, N.A., and T.J. Lair. 1994. "Getting Better? Change or Error in the Measurement of Functional Limitations." *Journal of Economic and Social Measurement*, 20(3):237–262. DOI: <https://doi.org/10.3233/JEM-1994-20305>.
- McBride, B. 2013. "Examining Changes in Filter Question (FQ) Reporting in the Consumer Expenditure Quarterly Interview Survey." In Proceedings of the Survey Research Methods Section of the American Statistical Association. Boston, Massachusetts, U.S.A.: 4304–4316. Available at: http://www.asasrms.org/Proceedings/y2013/files/400269_500735.pdf (accessed November 2020).
- Nancarrow, C., and T. Cartwright. 2007. "Online access panels and tracking research: The conditioning issue." *International Journal of Market Research*, 49(5):573–594. DOI: <https://doi.org/10.1177/147078530704900505>.
- National Research Council. 2013. *Measuring What We Spend: Toward a New Consumer Expenditure Survey*. The National Academies Press, Washington, D.C.
- Neter, J., and J. Waksberg. 1964. "Conditioning effects from Repeated Household Interviews." *Journal of Marketing*, 28(2):51–56. DOI: <https://doi.org/10.1177/002224296402800211>.
- Parker, J.A., N.S. Souleles, and C.D. Carroll. 2014. "In benefits of panel data in consumer expenditure surveys." In *Improving the Measurement of Consumer Expenditures: 75–99*. University of Chicago Press. DOI: <https://doi.org/10.7208/chicago/9780226194714.003.0004>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/> (accessed November 2020).
- Schonlau, M., and V. Toepoel. 2015. "Straightlining in Web Survey Panels Over Time." *Survey Research Methods*, 9(2):125–137. DOI: <https://doi.org/10.18148/srm/2015.v9i2.6128>.
- Shields, J., and N. To. 2005. "Learning To Say No: Conditioned Underreporting in an Expenditure Survey." In Proceedings of the Survey Research Methods Section of the American Statistical Association. Available at: <http://www.asasrms.org/Proceedings/y2005/files/JSM2005-000432.pdf> (accessed November 2020).
- Silberstein, A.R. 1990. "First Wave Effects in the U.S. Consumer Expenditure Interview Survey." *Survey Methodology*, 16(2):293–304.
- Silberstein, A.R., and C.A. Jacobs. 1989. "Symptoms of Repeated Interview Effects in the Consumer Interview Survey." In *Panel Surveys*, edited by D. Kasprzyk, G.J. Duncan, G. Kalton, and M.P. Singh, 289–303. Wiley, New York.
- Solon, G. 1986. "Effects of Rotation Group Bias on Estimation of Unemployment." *Journal of Business and Economic Statistics*, 4(1):105–109. DOI: <https://doi.org/10.1080/07350015.1986.10509499>.
- Struminskaya, B. 2016. "Respondent Conditioning in Online Panel Surveys. Results of Two Field Experiments." *Social Science Computer Review*, 34(1):95–115. DOI: <https://doi.org/10.1177/0894439315574022>.

- Sun, H., R. Tourangeau, and S. Presser. 2018. "Panel Effects: Do the Reports of Panel Respondents Get Better or Worse over Time?" *Journal of Survey Statistics and Methodology*, 7(4):572–588. DOI: <https://doi.org/10.1093/jssam/smy021>.
- U.S. Bureau of Labor Statistics. 2016. *Consumer Expenditures and Income: Handbook of Methods*. Available at: <https://www.bls.gov/opub/hom/cex/pdf/cex.pdf> (accessed November 2020).
- U.S. Bureau of Labor Statistics. 2020. *Household Survey Response Rates*. Available at: <https://www.bls.gov/osmr/response-rates/household-survey-response-rates.htm> (accessed November 2020).
- Van Buuren, S., and K. Groothuis-Oudshoorn. 2011. "mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software*, 45(3):1–67. DOI: <https://doi.org/10.18637/jss.v045.i03>.
- Venables, W.N., and B.D. Ripley. 2002. *Modern Applied Statistics with S*. Springer, New York, fourth edition. Available at: <http://www.stats.ox.ac.uk/pub/MASS4> (accessed November 2020).
- Waring, E., M. Quinn, A. McNamara, E. Arino de la Rubia, H. Zhu, and S. Ellis. 2020. "skimr: Compact and Flexible Summaries of Data, 2020. R package version 2.1.2. Available at: <https://CRAN.R-project.org/package=skimr> (accessed November 2020).
- Waterton, J., and D. Lievesley. 1989. "Evidence of Conditioning Effects in the British Social Attitudes Panel." In *Panel Surveys*, edited by D. Kasprzyk, G.J. Duncan, G. Kalton, and M.P. Singh, 319–339. Wiley, New York.
- Wickham, H., M. Averick, J. Bryan, W. Chang, L. D'Agostino McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. Lin Pedersen, E. Miller, S. Milton Bache, K. Müller, J. Ooms, D. Robinson, D. Paige Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani. 2019. "Welcome to the Tidyverse." *Journal of Open Source Software*, 4(43): 1686. DOI: <https://doi.org/10.21105/joss.01686>.
- Wilson, T., and S. Abdirizak. 2017. "Statistical Examination of Rounding Tendencies in the Consumer Expenditure Interview Survey." In Proceedings of the American Statistical Association July 29–August 3, 2017, Baltimore, Maryland, U.S.A. 894–907. Available at: <https://www.bls.gov/osmr/research-papers/2017/pdf/st170160.pdf> (accessed November 2020).
- Yan, T., and K. Copeland. 2010. "Panel Conditioning in the Consumer Expenditure Quarterly Interview Survey." In Proceedings of the Survey Research Methods Section of the American Statistical Association, May 13–16, 2010, Chicago, Illinois, U.S.A. Available at: http://www.asarms.org/Proceedings/y2010/Files/307812_59394.pdf (accessed November 2020).
- Yan, T., and S. Eckman. 2012. "Panel Conditioning: Change in True Value versus Change in Self-Report." In Proceedings of the Survey Research Methods Section of the American Statistical Association, July 28–August 2, 2012, San Diego, California, U.S.A. Available at: http://www.asarms.org/Proceedings/y2012/Files/306203_76099.pdf (accessed November 2020).

Received May 2020

Revised September 2020

Accepted December 2020