

Examining the relationship between the accuracy of self-reported data and the availability of respondent financial records

Emily Geisen, Charles Strohm, M. Christopher Stringer, Brandon Kopp, Ashley Richards

Abstract

Use of respondent records is a method for supplementing survey data. Records can improve data quality when data collected by self-report are subject to recall or other measurement errors. One limitation of using respondent records is that respondents often cannot provide records for all survey items of interest. If the accuracy of self-reports for items with records is different from the accuracy of self-reports for items without records, the estimates based on records are subject to nonresponse bias. However, if accuracy of self-reports is not related to availability of records, then even a limited set of records can improve data quality and also provide information about the direction and type of measurement error.

We explored the relationship between accuracy of self-reports and record availability using data from the Consumer Expenditure Records Study (CE Records Study). The CE Records Study is a non-probability feasibility study that examined the accuracy of self-reported data for various types of household expenditures. In the first interview, participants provided self-reports about the cost of household expenditures from the previous three months. In a follow-up interview, participants provided records (e.g., receipts, bank statements, bills) for all expenditures asked about. By comparing self-reports and records, we were able to evaluate the accuracy of self-reports.

Records were available for 36% of the 3,039 expenditures reported. Of the expenditures with records, 46% had incorrect self-reports. On average, expenditure amounts were under- or overestimated by 29.9%. Several factors such as the date, frequency, and cost of expenditures were associated with the availability of records. Using these factors and respondent demographics, we developed a propensity model to determine the likelihood that a record is available for a given expenditure. We then determined if expenditures with higher propensities for having records were more or less likely to be accurate than those with lower propensities. Finally, we evaluated the implications this has on reducing measurement error.

1. Introduction

Self-reported data are subject to recall error, which can affect the overall quality of the data collected. A review of the literature on recall error shows that both event characteristics and respondent characteristics are affect the quality of retrospective reporting. Events that are frequent, recent, salient, and distinctive are more likely to be reported accurately (Ebbinghaus, 1964; Sudman, Bradburn, and Schwarz, 1996; Tourangeau, Rips, & Rasinski, 2000).

In addition, respondent characteristics affect the accuracy of self-reports. Older respondents and those with less education tend to have greater recall error. In addition, some studies find that women provide higher-quality reports than do men (Auriat 1993; Smith and Thomas 2003).

In many areas of government statistics, administrative records are being explored as a means of reducing measurement error associated with recall failure and limiting respondent burden. However, administrative records for certain types of information, such as consumer expenditures (e.g., loyalty card information, store-based credit cards) are not easily accessed due to the wide range of vendors (e.g. national chains, small outlets, banks, and utility companies) and the proprietary nature of private business records. One solution proposed by other researchers (Guðnason, 2004) has been to ask respondents to save and provide their paper or electronic records themselves. A major limitation to this approach is that respondent records can be incomplete and not available for all survey items of interest.

The focus of this article is to explore the impact an incomplete set of records has on the ability to reduce measurement error, primarily due to recall failure, in the U.S. Consumer Expenditure Quarterly Interview Survey (CEQ). The CEQ is an ongoing monthly survey of the buying habits of American consumers, which is conducted by the U.S. Census Bureau for the U.S. Bureau of Labor Statistics (BLS). Data from the CEQ are used to calculate the weight of index items in the Consumer Price Index (CPI), one of the nation's leading economic indicators.

Past research suggests that respondents have difficulty recalling and reporting information about expenditures (Ahmed, Brzozowski, & Crossley, 2006; Tucker, 1985). The range of items and the depth of information asked about in the CEQ can make it a difficult questionnaire to answer and the interview experience has been described by some as burdensome (Fricker, Gonzalez, & Tan, 2011). These factors along with the factors affecting recall error can lead to measurement error. While measurement error is often cited as a problem in the CEQ (Miller & Downes-Le Guin, 1989; Steinberg, Chopova, Tan, Ogden, & Bahizi, 2006), little work has been done to evaluate the extent and direction of that error and whether there are other ways of collecting expenditure information to limit error.

If we assume that records are the gold standard, even a limited number of records can improve data quality. For items with records, we can also estimate the extent of measurement error for certain estimates. However, what can we determine about the items without records? What impact does this have on our ability to assess and reduce measurement error for the survey overall?

If respondents tend to have records for items they recall easily and accurately, the use of records will not substantially improve data quality. In fact, the effort required to gather records may actually have a negative impact on data quality by affecting response rates and increasing respondent burden. On the other hand, if respondents tend to have records for items they recall incorrectly, the use of records can greatly improve data quality. And finally, if there is no relationship between accuracy of self-reports and availability of records, then even a limited set of records can provide detailed information about the direction and type of measurement error in the survey overall. Using data from the Consumer Expenditures Record Study (CE Records Study), we explore the relationship between the availability of respondent records and the accuracy of self-reports.

2. Methods

Data

The CE Records Study was designed to investigate the availability of respondent financial records, to measure the direction and magnitude of measurement error in the CEQ, and to explore respondent burden in completing the CEQ. The U.S. Census Bureau contracted with RTI International to conduct the CE Records Study.

RTI International collected data from 115 adults in two interviews, spaced four to seven days apart. Respondents were recruited through convenience sampling methods in Raleigh-Durham, North Carolina and Washington, D.C. metropolitan area. Respondents were provided a \$40 incentive to complete the first interview and \$60 to complete the second interview. Data were collected via computer-assisted personal interviewing (CAPI) in respondents' homes.

During the first interview, respondents completed an abbreviated version of the typical CEQ instrument: respondents provided self-reported data (item purchased, date of purchase, and cost) for eight expenditure categories (housing, utilities, appliances, furniture, clothing, health insurance, subscriptions/memberships, and miscellaneous).¹ For each expenditure type, respondents reported whether they made an expenditure in the past three months and if so, the amount of the purchase. Respondents provided reports about expenditures they made and expenditures by others in their household. Adjustments were made to the CEQ instrument to allow for one-time administration (instead of inclusion in a panel) and to limit the instrument to the expenditure categories described above. With these exceptions, the question wording, interviewer instructions, and other data collection procedures in the CE Record Study were identical to the procedures used by the U.S. Census Bureau during standard CEQ interviews.

At the end of the first interview, respondents were asked to collect records for the expenditures asked about in the survey. We defined "records" broadly to include receipts, bills, bank and credit card statements, and respondents' notes or electronic budgets. Records could be hardcopy or electronic. Four to seven days after the first interview, during the second interview, the respondent provided the collected records to the interviewer. If a record was not available for an expenditure reported in the first interview, respondents were asked why the record was unavailable. When records were available, the interviewer recorded the information on the record such as the date and the amount of the expenditure. If the amount on the record did not match the self-report from the first interview, interviewers asked respondents to provide a reason, if applicable, for the discrepancy. In the vast majority of discrepancies, respondents indicated that the record amount was valid, but their self-report was incorrect. In this paper, we assume that any difference between the record and the self-report can be considered measurement error.

Analytic Approach

The analysis proceeded in four steps. First (Step 1), we determined if there was potential for bias in the availability of a record for a reported expenditure item. We examined both respondent characteristics and expenditure (i.e., event) characteristics to determine if

¹ The interview also collected self-reports about gross household income. We do not analyze these reports, however, because the paper focuses on expenditures rather than income or money received.

certain factors affected the likelihood that a record was provided for an expenditure. Next (Step 2) we assessed the extent of measurement error in the amount of reported expenditures. Then (Step 3), we examined how measurement error is associated with the propensity for having a record. To do this, we divided the predicted propensities for having a record into five strata based on the quartile scores. Finally (Step 4), we examined changes in measurement error for the amount of the reported expenditure for the lower propensity strata compared to the higher propensity strata.

3. Findings

Respondents reported a total of 3,039 expenditures in Interview 1. Records were available for 1,082 or 36% of the expenditures reported.

Factors affecting likelihood of having records:

First we explored what factors, if any, affected the likelihood that a record could be provided for a given expenditure. Logistic regression using the SAS Genmod procedure allowed us to explore both respondent characteristics and expenditure characteristics by accounting for the clustering of expenditures within respondents (i.e., respondents reported multiple expenditures). Table 1 provides the mean value for each characteristic for expenditures with records and those without, the odds ratio, and the significance-level for each factor. Reference categories are provided in parentheses.

Table 1. Factors Affecting Likelihood of Having Records

Respondent Characteristics	Record	No Record	Odds ratio	p-value
High education (vs low educ.)	56.19%	49.21%	9.85	0.02
High income (vs low income)	43.16%	34.70%	3.68	0.02
Non-Hispanic white	82.90%	61.68%	1.60	0.04
Male (vs female)	31.61%	37.51%	1.59	0.05
Owners (vs renters)	68.11%	54.06%	1.73	0.06
North Carolina (vs DC)	87.43%	90.70%	0.66	0.17
Household size	2.46	2.71	0.89	0.18
Age	46.48	44.26	1.01	0.32
Married (vs unmarried)	64.88%	53.24%	0.70	0.32
Working (vs not working)	46.83%	46.30%	0.86	0.48
Expenditure Characteristics	Record	No Record	Odds ratio	p-value
Housing expenditure (vs other)	9.33%	8.12%	7.84	< 0.0001
Phone expenditure (vs other)	20.05%	18.14%	8.78	< 0.0001
Utility expenditure (vs other)	21.07%	19.67%	7.47	< 0.0001
Appliance/furniture (vs other)	11.9%	20.2%	0.9	< 0.0001
Clothing expenditure (vs other)	20.7%	18.7%	1.19	0.46
Order reported	16.49	17.19	0.96	<0.01
Recent (vs not recent)	34.20%	21.82%	1.36	<0.01
Amount of expenditure (\$)	\$210.40	\$159.26	1.01	<0.01
QIC goodness of fit: 3379.64				

Both respondent and expenditure characteristics were significant predictors of having a record. Expenditures reported by respondents who were female, non-Hispanic white,

home-owners, had higher income, higher education, and were from the DC area (compared to North Carolina) were more likely to have a record. Expenditures that were reported earlier in the interview, expenditures that were purchased more recently, expenditures that were more expensive, and expenditures that were for housing, telephone, or utility purchases were also more likely to have a record. It is important to note that the housing, telephone, and utility purchases were frequent and recurring purchases. Respondents typically received and paid those bills each month.

These factors were selected because of their association with potential causes for recall error and because they were available for all reported expenditures.

Measurement Error

Measurement error was determined by the accuracy of the cost of expenditures reported by the respondents compared to the cost provided on the record for the same expenditure. Self-reports that matched the record within 10% were considered a match. Self-reports that were different from the record by more than 10% were considered not matching.

To determine whether records matched within 10%, we calculated the magnitude of misreporting for each expenditure. This was determined by the percent difference between the respondent's self-report of the expenditure amount and the amount shown on the record. We use the following formulas to calculate magnitude of difference:

$$\text{If record} > \text{self-report, magnitude of difference} = \frac{[(\text{record} - \text{self-report}) / \text{self-report}] \times 100\%}{}$$

$$\text{If record} < \text{self-report, magnitude of difference} = \frac{[(\text{self-report} - \text{record}) / \text{record}] \times 100\%}{}$$

$$\text{If record} = \text{report, magnitude of difference} = 0$$

We use this measure because it retains the relative difference between overestimates and underestimates. The following example illustrates the difference between our measure and a simple formula of self-report / record. If the self-report were \$100 and the record were \$150, a simple ratio of record to self-report would be 0.67 (100/150), showing a 33% underestimate. However, if the case was the opposite and the self-report were \$150 and the record were \$100, the ratio would be 1.5 (150/100), showing an overestimate of 50%. This is problematic because the average of the 33% underestimate and the 50% overestimate is 41.5% instead of 50%. Overestimates are always going to receive a higher weight than underestimates. The formulas above, take the difference between the two values and divides the difference by the smaller amount. Therefore, whether the self-report was an over-estimate or under-estimate, the ratio is the same. In the two examples sited above, the value for measurement error in both is 50% indicating that the self-report was misreported by 50% compared to the record value. In future work, we will investigate the direction of measurement error (i.e., whether the self-report is higher or lower than the record). We multiplied the values by 100% to facilitate interpretation of the results. For example, a measurement error value of 0.12 indicates that there was a 12% difference between the amount on the record and self-report (without regard to which was higher).

The magnitude of misreporting was 29.9% on average. To calculate the percent of self-report that matched the record within 10%, we coded expenditures with a magnitude of

difference between 0% and 10% as a match; values greater than 10% were considered not matches. The value of 10% was selected because it was determined by the Bureau of Labor Statistics to be a reasonable margin of error for reporting consumer expenditures. Of the expenditures with records, 54% of self-reports were matched within 10%.

Propensity Model

In this study, record propensity is the probability that a record was provided for an expenditure reported in Interview 1. A number of factors can affect the propensity that an expenditure will have a record. These include respondent characteristics – some respondents may be more or less likely to maintain records than other respondents. Other factors include expenditure (i.e., event) characteristics such as the cost and nature of the expenditure. Disentangling the effects between event and respondent characteristics would require multivariate modeling. In this paper, we used logistic regression, which predicts the availability of a record, to create propensity scores (i.e. the predicted probability that a given expenditure would have a record, regardless of whether a record was actually provided or not). Response propensity models are sometimes used to understand the risk of nonresponse bias in surveys. In this study, the likelihood that a reported expenditure has a record is similar to examining the likelihood that a sampled case completes a survey.

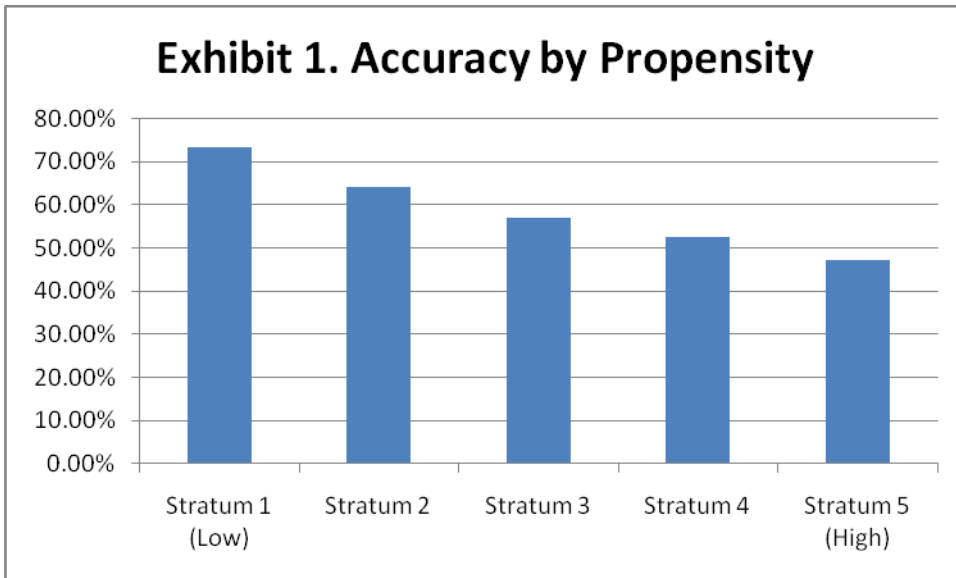
For the propensity model, exploratory models were computed using the respondent and expenditure characteristics described in Step 1. Other variations of the model presented in Step 1 (i.e., removing insignificant factors such as marital status) did not improve the goodness of fit of the model substantially. Therefore, we opted to use the full model presented in Step 1.

Next we divided the predicted propensity scores for likelihood of having a record into five strata of roughly equal-sizes, ordered from the lowest propensity to the highest propensity (Table 2). In a good propensity model, we would expect to see that the predicted propensities closely matched the actual proportion of expenditures with records. As we can see in Table 2, the predicted propensities are a good match for the actual proportion.

Stratum	Predicted Propensity	Actual proportion with record
Stratum 1 (Low propensity)	11.4%	10.0%
Stratum 2	22.9%	19.6%
Stratum 3	33.1%	35.8%
Stratum 4	45.8%	44.2%
Stratum 5 (High propensity)	68.3%	74.0%

Relationship between the likelihood of having a record and accuracy of self-reports

The final step was to compare differences in the accuracy of self-reports by the different propensity strata. Exhibit 1 shows the accuracy of self-reports by the propensity of having a record. Respondents were significantly more likely to have records for expenditures for which they were the least accurate ($F=5.80, df=4, p=.0001$). Similarly, respondents were less likely to have record for expenditure amounts they reported correctly.



4. Discussion and Conclusions

The analysis revealed that there are a variety of factors that affect the likelihood of providing a record for a reported expenditure. The respondent characteristics associated with having a record are being female, non-Hispanic white, a home-owner, having a higher income, higher education, and being from the DC area (compared to North Carolina). The expenditure characteristics associated with having a record were items that were earlier in the interview, more recent, more expensive, and for more expenditures that were frequent (e.g., housing, telephone, utilities). If the accuracy of self-reports for expenditures differs based on these factors, there is the potential for bias when evaluating the extent of measurement error in the survey.

Our analysis also showed that for expenditures with records, measurement error bias was a concern. Nearly half of the self-reports (46%) were inaccurate compared to the records, when using a margin of 10%. While we cannot determine exactly how much of the inaccuracy in reporting was due to respondent error, such as recall failure, versus other types of measurement error, we did ask respondents to explain why their self-reported amount was different than the record. The vast majority of the explanations provided by respondents were that they could not accurately recall the cost of the expenditure, indicating recall error. However, some explanations did show that there were other causes for the inaccuracies in the self-reports. For example, in the utility section, respondents frequently commented that they reported the amount they paid each month for the utility bill. The CEQ instrument, however, asks for the amount billed each month. Therefore interviewers recorded the amount associated with the bill date as opposed to the due date. The bill date for the records was frequently one month prior to the date the bill was actually paid. Therefore the discrepancies between self-reports and records were not due to recall error, but due to aspects of the survey instrument.

Finally, the results showed that the accuracy of the self-reports did vary by the likelihood of having a record. The expenditures with the highest likelihood of having a record were the least accurate. The expenditures with the lowest likelihood of having a record were the most accurate. One possible explanation for this is that the expenditures people are

unlikely to have records for are smaller, every day type of purchases. These are expenditures that are relatively common and purchased frequently. Therefore respondents are more familiar with the cost of these expenditures and are able to report them more accurately.

These results are promising for several reasons. First, it means that in the CE Records Study, even an incomplete set of records can be used to reduce measurement error. Secondly, it suggests that we may want to use different data collection strategies for different items. For items that respondents recall accurately and are unlikely to have records for, a retrospective survey based on respondent self-report may work well. For other items that respondents do not recall accurately, but are likely to have records for, it may be worth pursuing records as the primary source rather than relying on respondent recall.

The results of this study are limited by the fact that the CE Records Study used a non-probability sample of only 115 respondents. Additional research is needed on a larger probability-based sample to draw conclusions about the use of records in the CEQ or other consumer studies. In addition, the results reported may be specific to consumer expenditures and financial records. Surveys with other types of recall or measurement error may produce different findings about the relationship between the accuracy of self-reports and the availability of records.

Our next steps are to continue exploring the use of records to supplement recall and reduce burden in the CEQ, consider having respondents collect records proactively, and explore the use of records as a primary source of data collection for certain expenditures.

5. References

- Ahmed, N., Brzozowski, M., & Crossley, T.F. (2006). Measurement errors in recall food consumption data. IFS Working Papers W06/21, Institute for Fiscal Studies. [Available at: <http://ideas.repec.org/p/ifs/ifsewp/06-21.html>]
- Auriat, N. (1993). "My Wife Knows Best: A Comparison of Event Dating Accuracy Between the Wife, the Husband, the Couple, and the Belgium Population Register." *Public Opinion Quarterly* 57: 165-90.
- Ebbinghaus, H. (1964). *Memory: A Contribution to Experimental Psychology* New York: Russell Sage Foundation.
- Fricker, S., Gonzalez, J.M., & Tan, L. (2011, May). *Are you burdened yet? Let's find out*. Paper presented at the meeting of the American Association for Public Opinion Research, Phoenix, AZ.
- Guðnason, Rósmundur (2004, June). *The receipts approach to the collection of household expenditure data*. Paper presented at the meeting of the SSHRC International Conference on Index Number Theory and the Measurement of Prices and Productivity, Vancouver, Canada. [Available at: <http://www.hagstofa.is/lisalib/getfile.aspx?itemid=862>]
- Hahn, R., Eaker, E., and Rolka, H. (1997). "Reliability of Reported Age at Menopause." *American Journal of Epidemiology*, 146, 9: 771-775.

- Miller, L. & Downes-Le Guin, T. (1989). *Improving comprehension and recall in the Consumer Expenditure Interview Survey: Discrepancies in comprehension and recall as a source of nonsampling error*. Proceedings of the Survey Research Methods Section, American Statistical Association. [Available at: http://www.amstat.org/sections/srms/proceedings/papers/1989_089.pdf]
- Smith, J. and Thomas, D. (2003). "Remembrances of Things Past: Test-Retest Reliability of Retrospective Migration Histories." *Journal of the Royal Statistical Society Series A* 166: 23-49.
- Steinberg, B., Chopova, B., Tan, L., Ogden, J., Bahizi, P. (2006). *Analysis of recall effect on the reporting of expenditures for the Consumer Expenditure Interview Survey*. BLS Working Paper.
- Sudman, S., Bradburn, N., and Schwarz, N. (1996). *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass.
- Tourangeau, R., Rips, L., & Rasinski, K. (2000). *The Psychology of Survey Response*. New York: Cambridge University Press.
- Tucker, C. (1985). *Identifying measurement error in consumer expenditure reports*. Proceedings of the Survey Research Methods Section, American Statistical Association. [Available at: http://www.amstat.org/sections/srms/proceedings/papers/1985_092.pdf]
- Verhaeghen, P., Marcoen, A., and Goosens, L. (1993). "Facts and Fiction about Memory Aging: A Quantitative Integration of Research Findings." *Journals of Gerontology: Psychological Sciences*, 48: 157-171.