Statistical Examination of Rounding Tendencies in the Consumer Expenditure Interview Survey

Taylor Wilson[1] and Safia Abdirizak[2]
[1]U.S. Bureau of Labor Statistics, Washington, DC 20212
[2] U.S. Bureau of Labor Statistics, Washington, DC 20212

Abstract
Based on the results from the 2015 Consumer Expenditure Field Staff Survey Analysis Report, 98.4% of Field Representatives stated that record use improves the accuracy of the interview. The primary purpose of this project is to conduct research regarding the use of records as it relates to the rounding effect of recall interviews. This paper tests two hypotheses, [1] the use of records reduces the rounding effect and, as a result, increases data accuracy and quality and [2] the rounded expenditure amounts and non-rounded expenditure are significantly different. The hypotheses are important in aiding large survey programs in understanding how respondent rounding will affect the underlying data quality of the survey responses.

We employ a unique approach to isolate values which are most likely heaped—a coarse data property in which the respondents tend to converge their answers on 'round numbers.' This method examines the frequency of numbers represented in a domain of discrete values and examines the likelihood of observing any given value versus the rest of the values in that domain. The values which are statistically over represented by pre-defined threshold are heaped and thus more likely to be rounded. By identifying those values which have the highest probability of being a rounded value, we make an assumption that those values have actually been rounded. The findings will be of interest to survey methodologists and practitioners working in large scale survey operations with recall survey components and specific data quality goals.

Key Words: Record Use, Data Quality, Rounding, Reporting Behavior

All views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Bureau of Labor Statistics

1. Background

Defining and identifying rounding in large scale interviewer-administered surveys has been a key focus for researchers in assessing how rounding affects the data quality of the results. The Consumer Expenditure Survey (CE) is a nationwide household survey conducted by the U.S. Bureau of Labor Statistics (BLS) to find out how Americans spend their money. Its primary customer is the Consumer Price Index, which uses CE data for its expenditure weights. Other customers include the Department of Defense which uses CE data to help determine cost-of-living allowances for military personnel, and the Department of Agriculture which uses CE data to calculate the annual cost of raising a child.

The CE consists of two sub-surveys, an Interview survey (CEQ) and a Diary survey (CED), collected for the BLS by the U.S. Census Bureau. The purpose of the Diary survey is to collect detailed expenditure data on small, frequently purchased items such as food and

apparel. The purpose of the Interview survey is to collect detailed expenditure data on large items such as property, automobiles, and major appliances, as well as on recurring expenses such as rent, utilities, and insurance premiums. The data from the two surveys are then combined to provide a complete picture of consumer expenditures in the United States.

In recent years CE has engaged an effort to redesign the survey. That redesign plan, the "Gemini Project", introduces a records based interview to comply with specific recommendations by researchers to optimize the use of records in the CE. The stated goal of records according to the Gemini Redesign Proposal Report is to acquire precise and accurate information. Recent CE research by Elkin, Kopp, McBride, and Tan (2015), suggest that incidences of rounding are an indication of poor reporting quality. As expected, the findings from this report also show that recall data had a higher percent of rounded values. Similarly, in efforts to evaluate data quality, Murphy (2015) conducted a split sample experiment in the Community Advantage Panel Survey to explore the effects of record use. The sample consisted half of telephone respondents and half of in-person household interview respondents who were both encouraged to check records. As a result, the act of checking records was associated with lower levels of rounding, which is consistent with literature related to the use of records in interviewer and self-administered surveys.

Survey methodologists have conducted research to understand the mental process of answering survey questions, similar to the CEQ. The survey response model found in the second edition of Groves (2009) textbook on survey methodology describes the cognitive process of how respondents formulate responses. The first stage of the response model begins with "comprehension" in which the respondent interprets the question. The second stage, "retrieval", is when respondents recall the necessary information needed to answer the question. The third stage of the process "judgment" is a combination of estimating and summarizing the information recalled during the retrieval stage. Finally, in the "reporting" stage, respondents formulate their final response. During the reporting stage, respondents may adjust their response based on their ability to retrieve stored information during the recall stage.

This paper approaches detection of inaccurate data from the perspective of respondent rounding as opposed to strictly fabricated responses. To detect possible fabricated data, Swanson, Cho, and Eltinge (2003) provide insight on methods by applying Benford's Law. Benford's law suggests that leading (or left-most) digits follow a certain distribution. Thus by identifying the distribution of the leading digits, one can detect if the data may be fraudulent or error-prone if the leading digits do not follow the expected distribution evident in Benford's law, shown in Swanson, Cho, and Eltinge (2003). Further discussion of detecting possible inaccurate data in surveys can be found in Manski (2010). Manski notes that responses typically follow a certain pattern. Five percent intervals are more common (i.e., 5, 10, …, 90, 95), with responses more bunched at 50 percent than at adjacent round values (40, 45, 55, 60).

The motivation behind the subsequent research is to identify the extent of rounded responses in the CEQ and understand how respondents who consult records may differ from those who do not. In addition, the findings from the analyses will inform which sections of the CEQ will benefit from encouraging record use. This will lead to a more efficient use of resources and help to minimize the burden to respondents by requesting the use of additional records in the interview.

2. Data

For the purpose of our research, we utilize the CEQ because we are interested in examining the impact of rounding and record use in this sub-survey. The CEQ is designed to obtain data on the types of expenditures respondents can recall for a period of three months or longer. In contrast, the CED includes records of daily expenditures collected over a two-week period. Most of the expenditures recorded are for "small-ticket" items, such as detailed groceries (lettuce, potatoes, dishwashing soap), for which long recall is not expected. Therefore, there is little opportunity to assess rounding for expenditures with wide-ranging costs.

Record use is collected as part of a post-interview questionnaire that the interviewer fills out following each interview. The interviewer can report one of four categories related to the frequency of record-use (1 "Always or almost always", 2 "Most of the time", 3 "Occasionally", 4 "Never or almost never"). For the purposes of this analysis, households that are recorded as a one or two are called record users and a three or four are called non-record users. To examine the effects of rounding in the CEQ, we explore a wide range of expenditures to evaluate the differences in the prevalence of rounding. Thus, the CEQ is an ideal source of data because the survey collects large expenditures, such as property and vehicles, and those that occur on a regular basis, such as rent or utilities.

In order to evaluate the effects of rounding in the CEQ, we utilized pre-processed data to look at raw respondent reported expenditure values. CE only collects integer responses in the CEQ, so a minor amount of rounding is already imposed by the instrument itself. Pre-processed data are data that CE receives from the Census Bureau before any CE processing has been done. The primary limitation of the data with respect to this analysis is that the overall record use question is based on the interviewer's recollection and is asked at the end of the interview. For this analysis, a year of data are constructed from the monthly files between November 2015 and November 2016 using Stata 15/SE.

3. Methodology

3.1 Defining a Rounded Value

In order to provide a reasonable and workable definition of a rounded value, it is important to consider the trade-off between arbitrary definitions of roundedness and the analytic power those definitions provide. The more relaxed the definition of roundedness, the more biased any analysis will be in its conclusions. This is primarily due to the increased likelihood of a false positive that arises from these arbitrary definitions. Consider the application of a loose roundedness rule to a dataset with discrete numeric values. For example, every value evenly divisible by five will be called a rounded value. This rule will cover many numbers that are not empirically being rounded to by respondents of the survey—even if that same rule does capture the values which are being rounded to. This addition of noise could result in inconclusive, insignificant, or incorrect results simply because the definition of roundedness was not well specified. Considering the purpose of this study is to examine the effects of records on respondent rounding, properly defining the dependent variable (i.e. roundedness) in the data is paramount to achieving meaningful results. As such, the analysis attempts to move away from *a priori* definitions of rounding toward an empirical *posteriori* approach.

Based on a review of literature related to rounding in large-scale surveys, rounded values are typically defined as any value with 00, 25, 50, or 75 as the trailing digits for

expenditures. This is the "round number" approach which has some foundation in psychological studies regarding the way individuals think about dividing large values into smaller parts[1]. Another way to approach the rounded value definition is by simply dividing all expenditure values by 5 or 10 and any number that is evenly divisible will be considered rounded. Both of these methods are highly arbitrary and will likely contain false positives or will miss values to which respondents are rounding.

For the analysis of the CEQ, "heaped" values in the dataset are examined as a measure of coarseness. From Hao Wang's 2009 paper "Statistical Methods for Heaped Data", "Heaping is a common type of measurement error emerging when data are collected with various degrees of coarseness. We say that a dataset is heaped when it contains a mixture of exact and rounded-off values."[2] These spikes in frequency at certain numbers in the datasets strongly suggest heaping. It is important to also be able to call an individual number heaped as opposed to calling the entire dataset heaped in order to determine roundedness for a given number. The relationship between heaping and roundedness is ambiguous at best. Many of these frequency spikes are made up of a combination of factual expenditure amounts and rounded ones. These actual expenditure amounts are the effect of what we call "natural prices" of particular items. A simple example would be to consider an item, purchasable for 50 dollars per unit in a market of similar items that vary in price (e.g. 52 dollars, 48 dollars, etc.). Respondents display a propensity to round these varied prices near the "round number" to 50. Some of the 50 dollar responses will actually be 50 dollar purchases despite the over-representation of 50 in the dataset due to this rounding propensity. When markets have round natural prices, the argument that heaping implies roundedness becomes weaker and reports for each value on the domain can be hypothetically split into a proportion of rounded values and natural prices. The assumption is that values with a more pronounced spike (i.e. sufficiently higher than the typical spike) have a higher proportion of rounded values in them relative to their neighbors. We call these highly spiked values rounded by assuming the proportion of natural prices are comparatively low. The real challenge is deciding when a frequency spike at a certain number can be considered more severely heaped relative to its neighbors.

In order to accomplish this, it is necessary to characterize these spikes in a way which is objectively comparable. For this we use a method we call the "average fall approach." The approach begins by assigning the number of times an expenditure was reported to be a dollar amount d, denoted as n_d. Next the approach examines every possible integer along a domain and identifies the difference in the frequency of that integer from its left and right neighbor, D_L and D_R respectively, represented in equation (1) and (2).

$$D_L = n_d - n_{d-1} \quad (1)$$

$$D_R = n_d - n_{d+1} \quad (2)$$

Lastly, we define the threshold for which the joint distances are great enough that they can be considered different enough from the mean drop over the whole distribution to identify a value as heaped. Almost all expenditure distributions are right skewed or log-normal. This is due to a lower bound at zero for their domains and the relative infrequency of large individual expenditures in the data for most expenditure types. Because of this inherent

[1] See Groves (2009)

[2] For more information see, Wang (2009)

skew, the empirical rule whereby 95% of all observations fall within approximately two standard deviations of the mean does not automatically apply. The justification for selecting two standard deviations as the threshold in spite of the skews comes from Chebyshev's Inequality, represented here as equation (3).

$$Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{3}$$

This states that the probability of an observation less the mean value of a distribution being greater than k standard deviations is less than or equal to the reciprocal of the square of the number of standard deviations. Taking $k = 2$, the probability that an observation is outside two standard deviations from the mean is at most 0.25. This implies that at least 75% of the observations lie within two standard deviations of the mean for a wide class of probability density functions. The more normal the distribution, the more observations two standard deviations will capture--up to 95% of the distribution. This implies that a sensitivity can be chosen that guarantees a set percentage of observations will be covered. Since most expenditure distributions fall somewhere between normal and a right skew, two standard deviations are selected as the threshold for analysis guaranteeing a coverage from 75% to 95%. The condition for an expenditure value being identified as heaped and therefore rounded is specified in equation (4), where μ is the mean of the average drop along a specified domain.

$$D_L, D_R \geq \mu + 2\sigma \tag{4}$$

3.2 Order of Magnitude Effects

The definition of a rounded value specified here can be considered for specific orders of magnitude. Many expenditure value domains cover multiple orders of magnitude. For example, the clothing and accessories expenditure type has many expenditures valued between 1 and 100 dollars as well as many expenditures valued at over 1,000 dollars. However, given the right skew of the distributions, the frequency of the reported expenditures naturally approach zero because it becomes very unlikely to observe very large expenditure amounts in any frequency. This implies that using the method specified in the previous section, frequency differences in the higher orders will tend to be smaller and thus reduce the mean distance of the drops in the whole domain. Additionally, the relatively large frequency spikes in the smaller orders will obscure rounding behaviors in the larger domains where frequency spikes are less likely to be identified as being significantly different. Doing a simple logistic regression involving each order of magnitude as a predictor for observing a rounded value, while not previously restricting the sample to each order when doing the identification, will demonstrate this behavior. Table I reports the odds ratios from a logistic analysis of the rounded value identification procedure across the whole domain for Clothing and Accessories.

Table 1. Odds Ratios from Logistic Analysis of Whole Sample Rounding Identification Procedure

	Odds Ratio	Standard Error	Z - Statistic
Order One (n x 10^1)	11.14	0.91	29.67***
Order Two (n x 10^2)	7.26	0.52	27.89***

Source: CE Microdata – Clothing and Accessory Section

Compared to the third order expenditure values ($n \times 10^3$), the odds of observing a rounded value in the first order is 11 times more likely and the odds of observing a rounded value in the second order is 7 times more likely. To assess the real incidence of rounding behavior in all domains, it is necessary to restrict the domain to each order of magnitude in which expenditures occur. Repeating the procedure separately for each order of magnitude will appropriately classify values as rounded within its local area. This will dramatically increase the number of values identified as being rounded, especially in the higher valued domains where there is relative scarcity of reported expenditures and smaller variance in the drops. We can see this by repeating the logistic analysis of these split samples presented in table II.

Table 2. Odds Ratios from Logistic Analysis of Split Sample Rounding Identification Procedure

	Odds Ratio	Standard Error	Z - Statistic
Order One ($n \times 10^1$)	0.47	0.03	-10.49***
Order Two ($n \times 10^2$)	0.59	0.03	-9.85***

Source: CE Microdata – Clothing and Accessory Record Type

The odds ratios show that it is half as likely to observe a rounded value in order one relative to order three and sixth tenths as likely to observe a rounded value in order two relative to order three. To understand this result, consider that the largest frequency records are in orders one and two for the Clothing and Accessories record type. Because of this, the average fall in the distribution is much larger in order one and two than it is in the higher domains making it easier for a value to be considered rounded in the higher domains. When considering the whole sample, it was virtually impossible for a third order value to be considered rounded because of the influence of first order frequencies. After identifying values in comparison to their peers, the results are much more indicative of respondent behavior. There is no optimal odds ratio between the orders that should be expected. As long as the expenditures are being properly identified relative to their peers, the odds ratios can be any non-negative number. It may be possible in some expenditure sections for rounding to not occur in one order and occur frequently in another. In this case, the odds ratios would only show the relative difference in the frequency of the identified values. Figure I shows the behavior of one of these expenditure distributions as it approaches the higher orders. Evaluating expenditure values relative to their peers captures the smaller spikes at the right hand side of the frequency distribution.

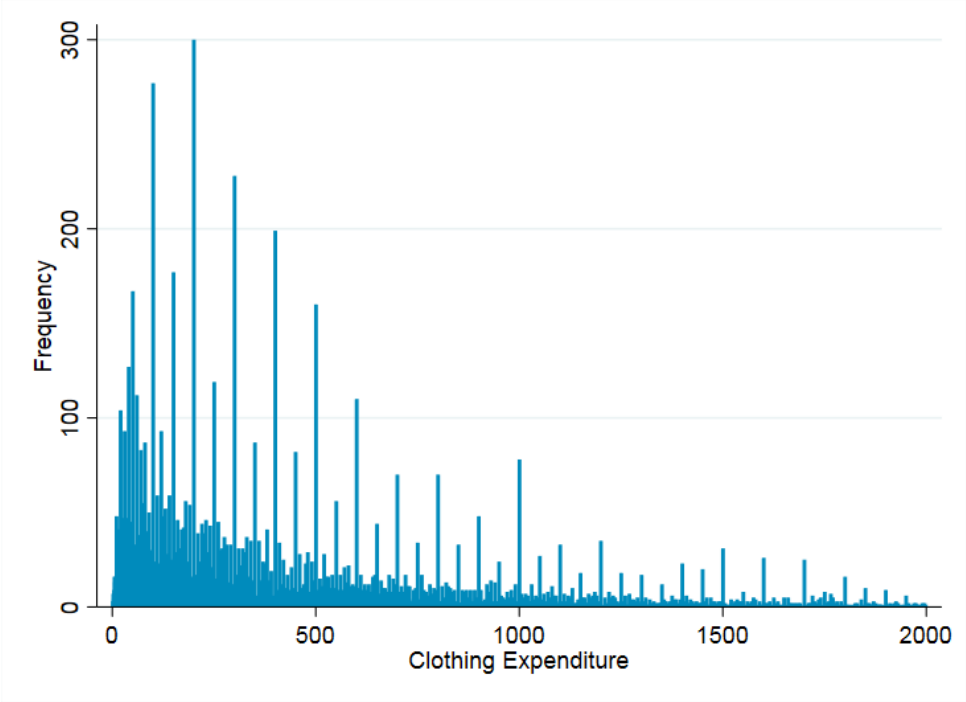

Figure 1. Histogram of Reported Clothing and Accessories Values

3.3 Odds Ratios for Evaluating the Impact of Records

To determine the odds of observing a rounded value in the records group versus the non-records group, odds ratios are computed from a logistic regression where records use is the independent variable and the identified rounded values are the dependent variable. Orders of magnitude are incorporated when identifying the rounded values in the procedure described in section 3.2. This ensures that the rounded values are appropriately identified relative to their peers as opposed to the whole sample. This allows us to predict the odds of observing a rounded value if a respondent utilized records. The resulting odds ratio will be in comparison to the baseline group of records equal to zero (non-record users). The null hypothesis is that records do not reduce the incidence of rounding in the data. The alternative hypothesis is that records reduce the incidence of rounding in the data. Therefore, we expect to observe an odds ratio that is less than one for all tested expenditure categories. Odds that are not significantly different from one imply that records do not impact rounding in the expenditure category in a statistically significant way.

3.4 Mann-Whitney u Test

The Mann-Whitney u test is used to test the significant differences in the expenditure distributions between the record using group and the non-record using group. This provides insight into whether or not the values reported are affected by the introduction of records. The primary advantage of this test over a standard t test is that the u test is non-parametric and does not require the data be normally distributed. The u test requires that two independent samples are compared, in this case record and non-record users, and that the data be ordinal. In order to compare the majority and non-majority record user groups, the response of the interviewer is considered. That is, the interviewer reports whether or not a household used records with varying degrees of frequency. That response is recoded into a 1 for respondents who used records 50 percent or more of the time and a 0 for respondents

who used records less than 50 percent of the time. The data are then ranked against each other by which value is greater and those identified ranks for each independent sample are added up. The u statistic based on the sum of ranks and the standard deviation are then computed from the two independent samples. From this it is possible to determine whether one distribution is significantly different from another. Procedural details are outlined in the 1947 paper by Mann and Whitney.[3] This test does not assess whether medians between the groups are significantly different. For that, a chi-squared statistic is computed for a nonparametric equality-of-medians hypothesis test. It is possible for one of these tests to be insignificant while the other is significant. Mann-Whitney u tests look at shape and location of the distribution while the equality of medians test examines differences in the median value of the expenditures.[4]

4. Results

The analysis of responses provides results consistent with other recall surveys in that there is an observed heaping pattern that occurs in all record types. Figure II shows examples from four of the observed record types. Value domains are restricted to 1000 for clarity of presentation.

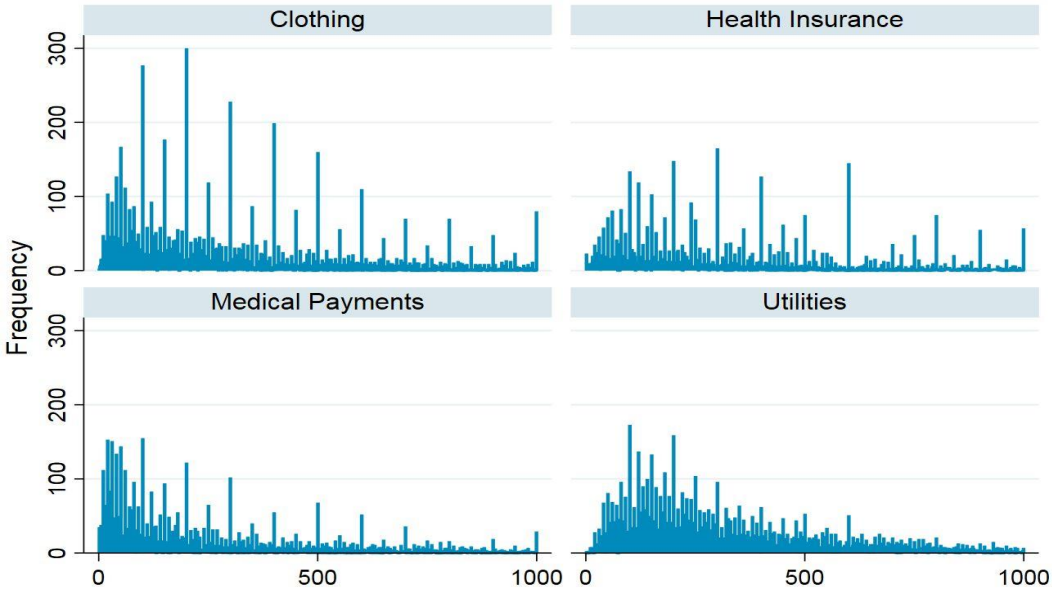

Figure 2. Frequency Histograms Demonstrating Heaping Patterns for Four Record Types

Given the proposed definition of a rounded record defined in the above section, the ratio of rounded values to non-rounded values will indicate the relative frequency of this behavior in the data. Respondent rounding varies across record type in frequency. The use of records in all examined cases lowers the frequency of rounded values with the notable exception of Renter's Insurance, which is not significantly different between the groups. In certain cases, record use improves the data distributions, resulting in more varied

[3] For more information see, Mann, H. B. (1947)
[4] For more information on these tests, see:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1120984/

responses and less heaping. Evidenced by the statistically significant reduction in rounding in most cases. Most record types follow this pattern. In certain cases, the effect of record use is less pronounced. The subscriptions record type experiences a smaller reduction of rounded values when records are used in the interview. The survey does not capture whether or not a record was used on a particular response so it is possible that a records respondent is still rounding some of their expenditures for which they did not have a record. Table III shows these figures for a selection of record types in the Interview Survey.

Table 3. Comparison of CU rounded expenditures Non-Record vs. Record User

	Non-Record			Record			Odds Ratio
Expenditure Section	# Rounded	# Total	% Rounded	# Rounded	# Total	% Rounded	
Owned Housing (MRTPMTX)	1,405	3,561	39	312	1,205	26	0.54***
Utilities (UTLCHGX2)	1,740	8,938	19	223	2,609	9	0.39***
Rented Vehicles (RENTEXPX)	222	509	44	83	246	34	0.67***
Non-health Insurance (INSNEXXB)							
Life	614	2,126	29	167	857	19	0.61***
Automobile	1,384	4,924	28	319	1,093	23	0.79***
Homeowners'	497	1,536	32	374	1,526	25	0.68***
Renter's	267	710	38	66	168	39	0.95
Health Insurance (HHIRPMXB)	1,480	5,112	29	322	1,680	19	0.59***
Medical (MEDPMTX)	1,551	4,981	31	298	1,586	19	0.52***
Clothing (CLOTHXA)	2,110	6,721	31	397	1,958	20	0.57***
Subscriptions (SUBEXPX)	415	2,892	14	81	637	12	0.76*

Source: CEQ Survey Unbox Data (November 2015- November 2016)
*p<.05; **p<.01; ***p<0.001 | *Most Appropriate Order of Magnitude Selected*

While some of the heaping observed can be attributed to respondent rounding, a portion of the heaping is due to the underlying natural prices of the items that make up that category. Because of this, some of those values will align with the actual expenditure amount paid by the consumer for the item. In distributions where the typical pre-tax prices of items tend to align with what we define as a rounded value, the introduction of a record will not add any new information for the respondent. This typically occurs with large ticket items where companies apply marketing practices when pricing. If a company does the rounding for the

consumer (e.g. 500 dollar washing machine), then that record type will report a higher frequency of these expenditures.

The literature also suggests that recurring expenditures are easier to remember for recall surveys. When comparing aggregate CE data with administrative sources, categories like rent typically tend to match very well.[5] Despite categories like mortgage payments being recurring, they also are rarely round numbers to which a respondent would choose to round. This implies that the introduction of records will still provide new information (i.e. the actual amount paid) for the respondent and therefore a more accurate value. In the case of the subscriptions category which is typically composed of smaller expenses, the natural price effect is influencing the outcome whereby the expenses are both repeatable and similar to what a respondent may choose to round to anyway (e.g. 10 dollar Netflix subscription). Adding records may not add much information in this case, which is demonstrated by the comparatively small two percentage point difference between the record group and non-record group. Opposite to this, there are those expenditure categories that naturally have items priced very differently from the rounded value definition. These will typically be smaller expenditures where the prices are situated toward the lower end of the value distribution and are non-repeated (e.g. Clothing and Accessories). These expenditure categories are expected to be greatly influenced by the introduction of records. This is confirmed by percentage decreases and the significance of the odds ratios.

Generally, survey methodologists prefer decreasing the amount of heaping in the dataset yields higher quality data. Given the general finding from the literature that households tend to underreport the value of many of their expenditures[6,7], it is reasonable to expect that any significant differences in medians by the introduction of records would result in a higher reported value for record users. Table IV reports the median expenditures for each expenditure type by record and non-record users and records the difference between the values. The significance of these differences are tested in the following section.

Table 4. Medians by Group – Values in USD

Expenditure Section	N	Record Users	Non-Record Users	Difference
Owned Housing (MRTPMTX)	4,766	$2188	$2300	-$112
Utilities (UTLCHGX2)	11,547	325	277	48
Rented Vehicles (RENTEXPX)	755	235	238	-3
Non-health Insurance (INSNEXXB)				
Life	2,983	106	100	6
Automobile	6,017	228	240	-12
Homeowners'	3,062	182	190	-8
Renter's	878	35.5	40	-4.5

[5] The estimated aggregate rent in CE is 98 percent of the aggregate estimated in the 2015 National Accounts. https://www.bls.gov/cex/cepceconcordance.htm
[6] See Tucker et al (2005), https://www.bls.gov/cex/research_papers/pdf/st040070.pdf
[7] See McCully et al (2013), https://www.bls.gov/osmr/pdf/ec130020.pdf

Health Insurance (HHIRPMXB)	6,792	262	242.5	19.5
Medical (MEDPMTX)	6,567	225	170	55
Clothing (CLOTHXA)	8,679	271	240	31
Subscriptions (SUBEXPX)	3,529	40	36	4

Source: CEQ Unbox Data (November 2015- November 2016)
Most Appropriate Order of Magnitude Selected

Each record type was examined for equality of medians and similarity of distributions to determine what effect, if any, records have on the reported expenditure values. The Mann-Whitney u test is used to test the null hypothesis that values from each group are drawn from the same expenditure distribution. The median test is used to test the null hypothesis that the medians of both groups are equal. Unequal medians may suggest that record use may have an effect on the magnitude of reported expenditures. Both tests are run for all examined record types and the results are reported in table V.

Table 5. Mann-Whitney u Test and Equality-of-Medians Test for Each Expenditure

		Mann-Whitney u Test (Record v. Non-Record)		Equality of Medians (Record v. Non-Record)	
Expenditure Section	N	Z Statistic	P Value	Chi Squared	P Value
Owned Housing (MRTPMTX)	4,766	2.117	0.034*	2.981	0.084
Utilities (UTLCHGX2)	11,547	-8.257	0.000***	45.089	0.000***
Rented Vehicles (RENTEXPX)	755	0.737	0.461	0.099	0.753
Non-health Insurance (INSNEXXB)					
Life	2,983	-0.495	0.621	0.414	0.520
Automobile	6,017	1.749	0.080	2.573	0.109
Homeowners'	3,062	0.020	0.984	0.349	0.555
Renter's	878	0.600	0.549	0.766	0.382
Health Insurance (HHIRPMXB)	6,792	-2.468	0.014*	5.913	0.015*
Medical (MEDPMTX)	6,567	-7.050	0.000***	32.135	0.000***
Clothing (CLOTHXA)	8,679	-3.852	0.000***	12.098	0.001**
Subscriptions (SUBEXPX)	3,529	-2.403	0.016**	1.483	0.223

Source: CEQ Unbox Data (November 2015- November 2016)
*p<.05; **p<.01; ***p<0.001 | *Most Appropriate Order of Magnitude Selected*

Subscriptions and mortgage payments are the most interesting result from these tests. The results show that these expenditure types have medians that are not statistically different,

but distributions that are significantly different. Non-health insurance and rented vehicle expenditure types produce insignificant results in both tests despite having significant odds ratios. This shows that the incidence of rounding significantly decreased but the general shape of the distribution and the median was preserved. The remaining expenditure types have significant differences in the medians and in distributions, showing that the record use changed the shape of the data and also the median value in a significant way. Differences in the inherent qualities of a record user and a non-record user may also have an effect on their spending and reporting style. For all significant differences in medians, the use of records has generated higher values overall. This is also evident from the negative z-scores produced by the test. Negative z-scores indicate that the median of the record users was higher than the median of the non-record users.

5. Conclusion

The use of the average fall approach to identify rounded records has enabled testing of whether record use significantly impacts the coarseness of data in the Consumer Expenditure Interview Survey. Eleven expenditures (e.g. Health Insurance, Subscriptions, etc.) were tested using this method. All expenditures displayed significant decreases in the likelihood of observing a rounded value when records were introduced, with the exception of renter's insurance. Further testing on the equality of medians showed that for all expenditure types where there was a significant difference in medians, records increased the median value of the expenditure. Changes in distributions were more common among expenditure categories. This is likely due to the redistribution of the heaped values to other values in the distribution. This lends credibility to the claim that rounding by the respondent or by a decision rule in processing could change the distributions in significant ways. Some expenditure types benefit more in terms of roundedness reduction from the use of records while others only have marginal benefits. It will be necessary to explore in future work whether the trade-off in increased respondent burden from record use for the observed changes in medians and distributions is worth it. Overall, the use of records appears to have a positive effect on the accuracy of the data and produces a higher quality product from the survey.

References

Bar, Haim, and Dean Lillard. "A Heap of Trouble? Accounting for Mismatch Bias in Retrospectively Collected Data on Smoking." Health, Econometrics and Data Group, 2010.

Battistin, Erich. "Errors in Survey Reports of Consumption Expenditures." 2003.

Dominitz, Jeff, and Charles F. Manski. "Perceptions of Economic Insecurity: Evidence from the Survey of Economic Expectations." Public Opinion Quarterly 61, no. 2 (1997)

Elkin, Ian, Brandon Kopp, Brett McBride, and Lucilla Tan. "Results from the Consumer Expenditure Surveys' Individual Diary Feasibility Test." 2015.

Fischhoff, Baruch, and Wändi Bruine De Bruin. "Fifty–Fifty=50%?" Journal of Behavioral Decision Making 12, no. 2 (1999)

Geisen, Emily, Ashley Richards, and Charles Strohm. "U.S. Consumer Expenditure Records Study." 2011.

Gideon, Michael, Joanne Hsu, and Brooke McFall. "Eliciting financial information on surveys: rounding in reported mortgage balances." FCSM, 2015.

Gideon, Michael, Brooke Helppie-Mcfall, and Joanne W. Hsu. "Heaping at Round Numbers on Financial Questions: The Role of Satisficing." Finance and Economics Discussion Series 2017, no. 006 (2017).

Groves, Robert M. Survey methodology. New York: J. Wiley, 2009.

Hart, Anna. "Mann-Whitney Test Is Not Just a Test of Medians: Differences in Spread Can Be Important." *BMJ : British Medical Journal* 323.7309 (2001): 391–393.

Lillard, Lee A. , and Robert J. Willis. "Cognition and Wealth: The Importance of Probabilistic Thinking." SSRN Electronic Journal.

Mann, H. B., and D. R. Whitney. "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other." *The Annals of Mathematical Statistics*, vol. 18, no. 1, 1947, pp. 50–60., doi:10.1214/aoms/1177730491.

Manski, Charles F. "Measuring Expectations." Econometrica 72, no. 5 (2004)

Manski, Charles F., and Francesca Molinari. "Rounding Probabilistic Expectations in Surveys." Journal of Business & Economic Statistics 28, no. 2 (2010)

McCully, Clinton, Thesia I. Garner, William Passero. "Understanding the Relationship: CE Survey and PCE." Bureau of Labor Statistics. 2013.

Moore, Jeffrey, Linda Stinson, and Edward Welniak. "Income Measurement Error in Surveys: A Review." 1999.

Murphy, J., J. Rosen, A. Richards, S. Riley, A. Peytchev, and M. Lindblad. "Financial Record Checking in Surveys: Do Prompts Improve Data Quality?" Field Methods 28, no. 3 (2015)

Pudney, Stephen. "Heaping and leaping: Survey response behaviour and the dynamics of self-reported consumption expenditure." Institute for Social and Economic Research University of Essex, 2008.

Swanson, David, Moon Jung Cho, and John Eltinge. "Detecting Possibly Fraudulent or Error-Prone Survey Data Using Benford's Law."

Tucker, Clyde, Paul Biemer and Brian Meekins (2005), "Estimating the Level of Underreporting of Expenditures among Expenditure Reporters: A Micro-Level Latent Class Analysis," American Statistical Association 2005 Proceedings of the Section on Survey Research Methods,, Washington, D.C., American Statistical Association.

Tversky, A., and D. Kahneman. "Judgment under Uncertainty: Heuristics and Biases." Science 185, no. 4157 (1974)

Wang, Hao, "Statistical methods for heaped data" (2009). *Dissertations available from ProQuest*. AAI3395726. http://repository.upenn.edu/dissertations/AAI3395726