Correlates of Data Quality in the Consumer Expenditure Quarterly Interview Survey

October 2009

Jennifer Edgar[1] and Jeffrey M. Gonzalez[2]
[1]US Bureau of Labor Statistics, Washington, DC, Edgar.Jennifer@bls.gov
[2]US Bureau of Labor Statistics, Washington, DC, Gonzalez.Jeffrey@bls.gov

Abstract

The Consumer Expenditure Quarterly Interview Survey (CEQ) is an ongoing panel survey which collects detailed expenditure information from a national sample of households. High data quality is essential to accurately reflect the spending habits of American consumers. This study examines CEQ data quality in terms of the editing required during the data processing phase. Editing methods include imputation of item missing data and allocation of expenses reported in aggregate across different items. We explore the relationship between this measure of data quality and a variety of covariates, including respondent characteristics (e.g., educational attainment), household characteristics (e.g., relationship of persons living within the household), and interview characteristics (e.g., use of interviewing aids). We find that many of these factors are significantly associated with data quality.

Key Words: Data quality, data processing, imputation, measurement error, paradata, processing error

1. Introduction

Survey methodologists generally agree that data quality is one of the most important elements to consider when evaluating the success of a survey. How they define data quality, however, is not uniformly agreed upon. For each study making an assessment of the quality of a survey, there may be a different definition of data quality. Furthermore, publicly and/or easily accessible assessments of survey quality usually only consist of "low hanging fruit" for example, overall response rates and standard errors of key survey estimates. On occasion, research offices associated with the survey program office might go a few steps beyond that and publish academic-type reports on other aspects of survey data quality.

The Consumer Expenditure Survey program is no exception. There are few systematic attempts made within the program office to study data quality. In fact, CEQ data quality assessments are often limited to quarterly response rates, standard errors on survey estimates and overall post-survey processing editing rates. It is rarely the case that any of these assessments relate data quality to the reasons that data quality might be compromised or measurement errors are thought to arise (e.g., mode of data collection, the interview, the survey instrument, and the respondent). However, data quality assessments of other surveys have taken this next step. For instance, researchers have explored the impact on various factors on data, including interviewer effects (Kennickell, 2002), questionnaire design (Hess, et al., 2001), and reluctant reporters (Yan, et al., 2004), to name a few.

This study offers an additional perspective on the relationship between data quality and various interview and respondent characteristics. It identifies a set of characteristics of the sample unit, respondent, and interview that might be related to inadequate response. We believe that this is an important step in assuring an on-going effort to track data quality. In addition, by knowing which factors lead to high quality data, the CE survey program office can make informed decisions about the best ways to improve the overall quality of the existing survey.

1.1 The Consumer Expenditure Quarterly Interview Survey (CEQ)

The CEQ collects information on the spending habits of American Consumers. The survey design is a rotating panel survey and is conducted quarterly over the course of thirteen months. Each respondent is asked a series of questions on a variety of expenditures. For the most part, the expenditures captured in the CEQ are those that can be expected to be recalled for a period of three months or longer and tend to be relatively large purchases, such as for automobiles, and recurrent expenses, such as utility bills. However, a few smaller expenditures, such as those for clothing, or less frequently occurring expenses, such as those for household appliances (e.g., toaster ovens), are collected. These data are important because they provide the basis for revising the weights and associated pricing samples for the Consumer Price Index (CPI), one of the nation's leading economic indicators. They also allow us to get a picture of a household's spending pattern (BLS *Handbook of Methods*, 2007).

US Census Bureau field staff capture the data using a computer-assisted personal interviewing (CAPI) survey instrument. Once the data are collected, they are transmitted to the national office for processing. This processing involves a series of reviews and edits. There are three major types of data adjustment routines that are performed during data processing aimed at improving the estimates derived from the CEQ. They are *imputation*, *allocation*, and *time adjustment*. Data imputation routines account for missing or invalid entries. Missing or invalid attributes or expenditures are imputed, or filled in with plausible values based on some model. Allocation routines are applied when respondents provide insufficient detail to meet tabulation requirements. For example, an allocation routine would be invoked when a respondent reports an expenditure amount for an aggregate group of items (e.g., spent $100 on clothing), but cannot provide the costs of the specific items (e.g., bought one $50 pair of pants and two $25 shirts). The allocation routine would distribute the total amount of the expenditures for the combined report among the components of the report. Finally, time adjustment routines are used to classify expenditures by month, prior to aggregation of the data to calendar-year expenditures. Time adjustment will not be discussed further in this paper.

1.2 Theory on Data Quality

There are two main paradigms for survey data quality. The first is usually referred to as the Total Survey Error (TSE) paradigm (Groves, et al., 2004). This perspective on data quality focuses on how at each stage of the survey process errors, either systematic or variable, can arise. The TSE paradigm generally consists of the following errors or error sources: coverage error, sampling error, nonresponse error, construct validity, measurement error, processing error, and post-survey adjustment error. Most relevant to the research presented here is measurement error. Measurement error is said to occur when there is a mismatch between the response provided by the respondent and its corresponding "true" value. The reason for this error in measurement may be a consequence of the mode of data collection, the interviewer, the survey instrument (including the question wording), and/or the respondent. Definitions of each of the other sources of error can be found in Groves et al. (2004). In sum, the primary focus of this paradigm tends to be the accuracy of the survey estimates.

The second paradigm, often referred to as the Total Quality Management (TQM) paradigm, includes accuracy (and all the types of errors that comprise of accuracy) as a single dimension of data quality (Biemer and Lyberg, 2003). In addition, this paradigm draws on other dimensions of quality by incorporating a user's perspective. These dimensions include relevance, timeliness, coherence, interpretability, and accessibility (Brackstone, 1999). While it is important to recognize that accuracy is not the only important dimension or component of data quality, it is the primary focus of this research. We believe that assessing components of the accuracy dimension of data quality is an important first step in providing an overall comprehensive picture of data quality as it relates to the CEQ.

A complete assessment of survey data quality should include both quantitative and qualitative statements on each component of data quality from both frameworks; however, that type of assessment is beyond the scope of this research. We believe that this report would be subsumed in a more comprehensive evaluation of CEQ data quality, and acknowledge that examining other aspects of data quality is an essential next step in this research.

2. Methods

2.1 Operational Definition of Data Quality

As mentioned in the introduction, after the data are collected, they undergo a series of processing and edits. For the purposes of this study, data quality is defined, or operationalized, as whether an interview needed post-data collection processing in the form of imputation or allocation. Other metrics for data quality were considered, such as interview length, number of reports, interview mode, and use of respondent aids, but we determined that those are potential *indicators* of data quality, rather than *measures* of data quality. In other words, indicators are generally known to have an effect on the quality of the survey data, but by themselves do not measure data quality.

2.2 Data Creation

We analyzed CEQ data collected during the time period April 2006 to March 2008 yielding a total of 85,440 completed interviews (observations). Each observation was treated as an independent interview despite some consumer units[1] (CU) providing responses for up to four interviews. We excluded data from interview one from our

[1] A consumer is the unit for which expenditure reports are collected. It is defined as: "(1) all members of a particular housing unit who are related by blood, marriage, adoption, or some other legal arrangement, such as foster children; (2) a person living alone or sharing a household with others, or living as a roomer in a private home, lodging house, or in permanent living quarters, in a hotel or motel, but who is financially independent; or, (3) two or more unrelated persons living together who pool their income to make joint expenditure decisions. Students living in university-sponsored housing are also included in the sample as separate consumer units." (Bureau of Labor Statistics, US Department of Labor, *Handbook of Methods*, Chapter 16, April 2007 edition, Consumer Expenditures and Income)

analyses as these data are primarily used for inventory and bounding purposes and are not subjected to the same processing and review that data from interviews two through five are.

We identified thirty-six expenditure variables that represent the majority of expenses captured in the CEQ. These expenses vary in their frequency, salience, and amount and it should be noted that a CU may report expenses for all, some, or none of these expenditure. For each of these expenditure variables, a CU may have multiple reports for expenses. For instance, clothing expenses for persons age two or older are contained in a single variable, but each report is listed as a separate record (or row) in the data set; thus, all of these records are summarized so that in the final data set each record (row) represents one CU. However, to get an accurate assessment of whether any report by a CU was edited, each individual report was investigated. Once each report was assessed, we classified any CU as needing editing (i.e., providing data of poor quality) if any of its reports were edited during the processing and review stage of data collection.

2.3 Covariate Selection

To address our primary research objective, the identification of factors associated with data quality, the characteristics of interest were classified into three categories – respondent-level, CU-level, and interview-level. The respondent-level covariates included a categorical version of age (under 35, 35 to 54, and 55 and older), education level (less than high school, high school graduate, some college or Associate's degree, and Bachelor's degree or higher), an indicator for Hispanic origin, primary language of the respondent (English or other), gender (male or female), and an indicator for whether or not the respondent had previously refused to complete the current interview. The CU-level characteristics included family type, housing tenure (owner or renter), and Census region (Northeast, South, Midwest, and West). It is important to note that family type is a covariate that describes the relationships among the persons living together in a particular CU. In our analyses, this covariate had four levels – husband and wife only, a husband and/or wife with children, single consumers, and all other types. The interview-level covariates included an indicator for receipt of the advance letter, an indicator for bill usage during the reporting of utility expenditures, general record usage during the interview (mostly to always and occasionally to never), an indicator for whether the information book was used, and two mode (personal visit or telephone) variables – one for the mode in which the interview was completed and the second indicating the mode most often used during the collection process.

2.4 Inferential and Graphical Methods

The characteristics within each of these categories were used to build separate logistic regression models in which each set of covariates was used to predict whether or not any of the reports given in a completed interview would need editing. Our hypothesis was that the prevalence of editing would significantly vary by these characteristics. Since this was an exploratory analysis, we identified factors commonly used in statistical assessments of data quality to include in each model, but did not hypothesize about the nature or direction of these relationships. The primary statistic used in these analyses was the adjusted odds ratio. The odds ratio is one commonly used statistic to assess the risk of a particular outcome occurring, and in our case the outcome of interest is poor data quality, as measured by the presence of edited reports in a completed interview.

To develop the final first-order logistic regression models for each set of covariates, we used a step-wise regression procedure. We used step-wise regression for two primary reasons: (1) this research is meant to extract unknown information from the data (i.e., we are looking for things that may or may not have strong theoretical motivation); and, (2) this type of procedure holds advantages over other selection procedures in that a regressor can either enter or leave the model at each stage.

After final first-order logistic regression models were produced, we reconsidered each of the three models by including all of the first order terms as well as each covariate's interaction with the amount of expenditure reports made. The rationale for including this interaction was that the relationship between the specific characteristic and whether or not the completed interview needed editing may be impacted by the amount of expenditure reports a respondent provides. As in the first set of logistic regression models, we used a step-wise regression procedure to determine the final set of covariates in each model.

We produced a series of odds ratio plots for each of the characteristics. On the vertical axis is the odds ratio and on the horizontal axis is the count of expenditures reported. The solid black line represents the odds ratio for two levels of characteristic as a function of the count of expenditures reported during the interview. The dashed red lines on each side of the black line represent lower and upper 95% confidence bands. The dashed blue horizontal line is a reference line for a null association. Finally, the three dashed green horizontal lines (from left to right) represent the 25th, 50th, and 75th percentiles of the count of expenditures reported.

The lines displayed in the graphs for the odds ratios and lower and upper 95% confidence bands as a function of the count of expenditures reported were produced using the following procedure:

1. From each of the final regression models we estimated the odds ratio between two levels of a characteristic (as a function of the count of expenditures reported), which we denote as, $\hat{\theta}_{A,B}$, to be $e^{\hat{\beta}_1+\hat{\beta}_2 X}$. The parameter estimate $\hat{\beta}_1$ represents the estimated log odds ratio between two levels of the characteristic of interest, controlling for other factors in the model and $\hat{\beta}_2$ represents the estimated effect due to the interaction with the count of expenditures reported. Finally, X denotes the count of expenditures reported.
2. Next, to find the estimated standard error associated with the estimated log odds ratio while accounting for the interaction term, we used the following[2]:
   a. $v\left(\hat{\beta}_1+\hat{\beta}_2 X\right)=v\left(\hat{\beta}_1\right)+X^2 v\left(\hat{\beta}_2\right)+2X\operatorname{cov}\left(\hat{\beta}_1,\hat{\beta}_2\right)$
   b. $se\left(\hat{\beta}_1+\hat{\beta}_2 X\right)=v\left(\hat{\beta}_1+\hat{\beta}_2 X\right)^{1/2}$.
3. Thus, the lower and upper 95% confidence bands for $\hat{\theta}_{A,B}$ become $e^{\hat{\beta}_1+\hat{\beta}_2 X\pm 1.96 se\left(\hat{\beta}_1+\hat{\beta}_2 X\right)}$.

3. Results

[2] Estimates of the individual variance and covariance terms can be obtained directly from the SAS output.

In this section, we present descriptive statistics of the sample used in our analyses and abbreviated results and findings from the step-wise logistic regressions. The remaining results are available upon request from the authors.

3.1 Overall Descriptive Statistics

In Table 1, we present an overall description (based on the covariates identified in Section 2.3) of the sample we used in our analysis. To highlight a few of the findings, we see that a majority of respondents tend to be non-Hispanic, English speaking females who were not converted refusers. Most respondents acknowledged receiving the advance letter, but failed to use utility bills when reporting utility expenditures. Also, about two-thirds of the completed interviews were conducted by a personal visit interview and over eighty percent of the completed interviews contained reports that were edited.

Table 1: Overall Descriptive Statistics

Respondent-Level Characteristics			
Age (%)		*Educational Attainment (%)*	
Under 35	22.85	Less than high school	15.43
35-54	40.78	High school graduate	25.35
55 and older	36.36	Some college	30.69
Sex (%)		Bachelor's or higher	28.53
Male	46.84	*Language (%)*	
Female	53.16	English	94.16
Hispanic Origin (%)		Other	5.84
Hispanic	12.21	*Converted Refusal (%)*	
Non-Hispanic	87.79	Yes	11.79
		No	88.21
CU-Level Characteristics			
Census Region (%)		*Family Type (%)*	
Northeast	21.24	Husband and wife only	21.45
Midwest	22.11	Husband and/or wife w/children	31.83
West	25.41	Other types	18.34
South	31.22	Single consumers	28.38
Household Tenure (%)			
Owner	67.06		
Renter	32.94		
Interview Level			
Advance Letter (%)		*Record Usage (%)*	
Yes	94.63	Mostly to always	40.82
No	5.37	Occasionally to never	59.18
Utility Bill Usage		*Mode Most Used (%)*	
Yes	26.21	Personal visit	64.28
No	73.79	Telephone	35.72
Infobook Usage		*Mode (Final) (%)*	
Yes	46.33	Personal visit	63.31
No	53.67	Telephone	36.69
Data Quality			
Proportion of Edited Values (%)			
No editing performed	18.43		
Any editing	81.57		

3.2 First-Order Logistic Regression Models

3.2.1 Respondent-Level Characteristics Model

The first logistic regression model considered all of the first-order respondent-level characteristics. All covariates were retained in the regression model. Table 2 displays the adjusted odds ratios from this model as well as their lower and upper 95% confidence limits.

Table 2: Adjusted Odds Ratios from the Respondent-Level Characteristics Model

Comparison	Estimate	95% LCL	95% UCL
Age			
35 – 54 vs. Under 35	1.24	1.20	1.28
55+ vs. Under 35	0.88	0.85	0.90
Education			
High school vs. Less than high school	1.52	1.42	1.62
Some college vs. Less than high school	1.55	1.48	1.62
BS or higher vs. Less than high school	2.41	2.25	2.58
Hispanic: No vs. Yes	0.90	0.86	0.93
Language: English vs. Other	1.11	1.05	1.17
Converted refusal: No vs. Yes	0.90	0.87	0.93
Sex: Male vs. Female	0.96	0.94	0.98

The first logistic regression model considered all of the first-order respondent-level characteristics. All covariates were retained in the regression model. The results indicate non-Hispanic respondents have lower odds of needing editing than their Hispanic counterparts, adjusted for all other characteristics in the model. A similar trend was found for each of non-converted refusers, males, and respondents who completed the interview in a language other than English. There also seemed to be a slight gradient with respect to the effect of education on editing. In other words, as a respondent’s level of education increased from not completing high school, he/she had higher odds of needing editing, adjusted for all other factors in the model. Finally, when respondents age 35 to 54 were compared to those under 35, they had higher odds of needing editing while the reverse trend was observed for those aged 55 and over, adjusted for all other covariates in the model.

3.2.2 CU-Level Characteristics Model

The second logistic regression considered all first-order CU-level characteristics. All covariates were retained in the regression model. The adjusted odds ratios for all statistically significant associations and their 95% confidence intervals from this model are presented in Table 3.

Table 3: Adjusted Odds Ratios from the CU-Level Characteristics Model

Comparison	Estimate	95% LCL	95% UCL
Family type			
Husband and wife only vs. Single consumers	0.92	0.89	0.96
Husband and/or wife with children vs. Single consumers	1.40	1.35	1.45
Other types vs. Single consumers	1.43	1.38	1.49
Housing tenure: Owner vs. Renter	1.61	1.58	1.65
Census region			
Midwest vs. West	1.18	1.14	1.22
Northeast vs. West	0.93	0.90	0.96
South vs. West	0.81	0.79	0.83

From the model, family type (a classification of the relationship among persons within the CU) appears to be a strong predictor of whether or not a completed interview will need editing. More specifically, when compared to single consumers, a CU with only a husband and wife had lower odds of needing editing, after controlling for other covariates in the model. The opposite association was observed the two other types of CUs when compared to single consumers. Another finding from this model was that CU members living in an owned housing unit had higher odds of needing editing than their renting counterparts, controlling for all other factors in the model. Finally, CUs located in the Northeast or South Census regions had lower odds of needing editing than those located in the West, while the opposite trend was observed for those in the Midwest, adjusting for other factors in the model.

3.2.3 Interview-Level Characteristics Model

The third logistic regression considered all first-order interview-level characteristics. Of all the characteristics considered only two were retained in the final regression model after the stepwise selection process. They were the final data collection mode and an indicator for the use of records during the interview process. The estimated adjusted odds ratios for these two factors are displayed in Table 4.

Table 4: Adjusted Odds Ratios from the Interview Characteristics Model

Comparison	Estimate	95% LCL	95% UCL
Final data collection mode: Personal visit vs. Telephone	0.84	0.82	0.86
Record usage: Mostly to always vs. Occasionally to never	0.89	0.88	0.91

The results of this model suggest that interviews completed via a personal visit have lower odds of needing editing when compared to interviews completed via the telephone, controlling for record usage. Similarly, completed interviews in which records were mostly to always used had lower odds of needing editing than completed interviews in which records were only occasionally or never used, after adjusting for final data collection mode.

3.3 Interactions with Expenditure Reports

3.3.1 Respondent-Level Characteristics with Expenditure Count Interaction Model

The fourth model we considered included the first-order terms for all of the respondent-level characteristics as well as each characteristic's interaction with the amount of expenditures reported. With the exception of the respondent's gender, all characteristics as well as their interaction terms were retained in the model. Only the first-order term for the respondent's gender remained in the final model. Since logistic regression parameter estimates are difficult to interpret in the presence of interactions, we graphically summarize a few of the key findings below.

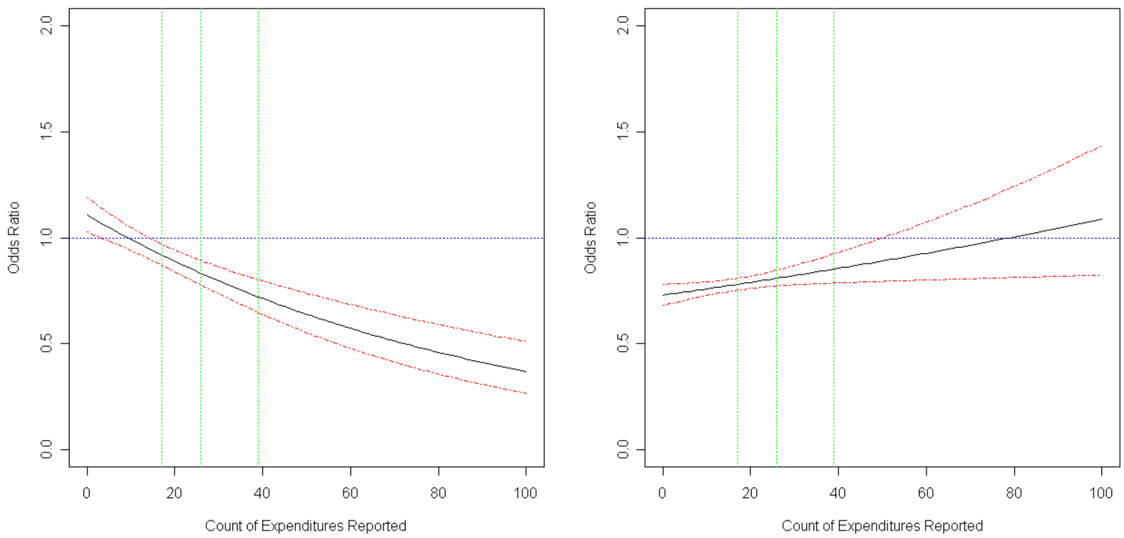

Figure 1: Hispanic origin: No vs. Yes (left); Converted refusal: No vs. Yes (right)

The following trends were observed when comparing non-Hispanic respondents to Hispanic respondents. In general, as the count of expenditure items reported increased, the odds for needing editing among the non-Hispanic respondents decreased. Among respondents in the first quartile of expenditure reporting, there was little evidence, if any, of a statistically significant association. Beyond the first quartile of expenditure reporting; however, non-Hispanic respondents had lower odds of needing editing than Hispanic respondents. This association was statistically significant and the odds ratio decreased as the level of expenditure reporting increased.

The following trends were observed when comparing converted refusal respondents to non-converted refusal respondents. When comparing non-converted refusers to converted refusals, the odds ratio increased as the count of expenditures reported increased. The odds of expenditure reports needing editing was lower for non-converted refusers than it was for converted refusers. This trend was statistically significant for respondents reporting fewer than about fifty items. For respondents with more than fifty expenditure reports, there was no evidence of an association between converted refusal status and expenditure reports needing editing.

3.3.2 CU-Level Characteristics with Expenditure Count Interaction Model

The fifth model that we considered included all first-order terms for the CU-level characteristics as well as each term's interaction with the count of expenditures. After the stepwise logistic regression was performed, all covariates and their interactions remained in the model. An interesting finding regarding family type is presented in Figure 2.

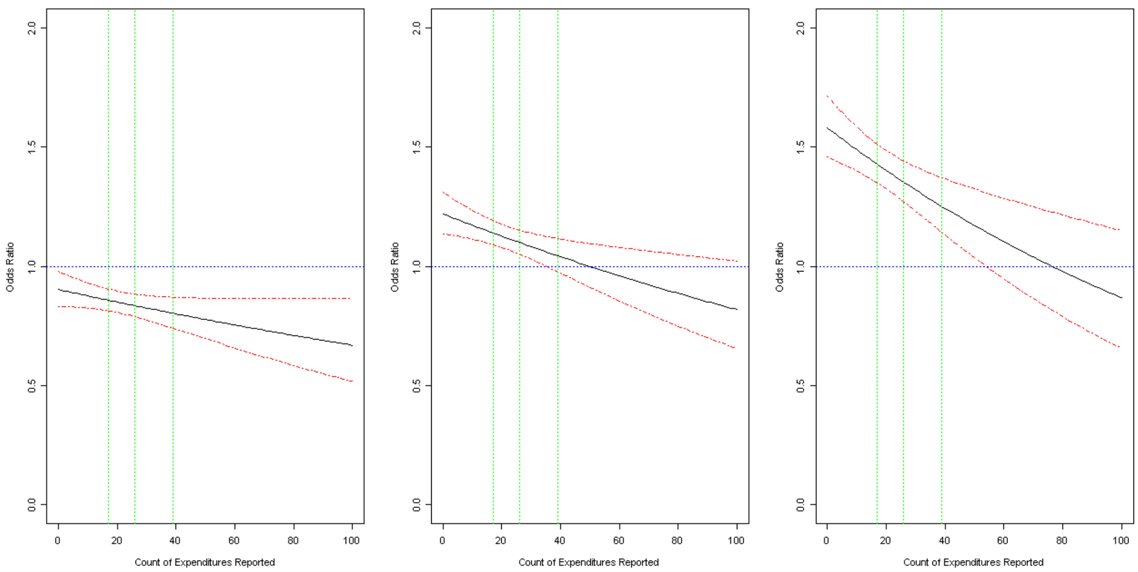

Figure 2: Family type (from left to right): Husband and wife only vs. Single consumers; Husband and/or wife with children vs. Single consumers; Other types vs. Single consumers

The following trends were observed for family type. When respondents from husband and wife CUs were compared to single CU respondents, the odds of needing editing on their reports decreased as the count of expenditure reports increased. Across all levels of expenditure reporting, the odds of needing editing was lower (and statistically significant) for respondents from husband and wife only CUs when compared to single CU respondents.

When comparing respondents from husband and/or wife with children CUs to single CU respondents, we again observed a decreasing odds ratio as the count of expenditure reports increased. However, for respondents in below the third quartile of expenditure reporting, the odds of needing editing was higher for respondents in husband and/or wife with children CUs when compared to respondents in single-person CUs. Beyond the third quartile, there was no evidence of a statistically significant association.

When comparing all other types of CUs to single-person CUs, the odds ratio for needing editing on the expenditure records reported decreased as the number of expenditure items reported increased. For respondents reporting fewer than fifty expenditure items, the odds of needing editing was higher for other types of CUs than it was for single-person CUs.

3.3.3 Interview-Level Characteristics with Expenditure Count Interaction Model

The last model we considered was similar to the interview-level model but now included terms for each covariate's interaction with the amount of expenditures reported. With the exception of mode used most often during the data collection process, all first-order terms for the covariates were retained in the model. However, only the interaction terms for utility bill usage and record usage remained in the final model. The graphs of the odds ratios for these factors and their interaction with expenditure count are presented below in Figure 3.

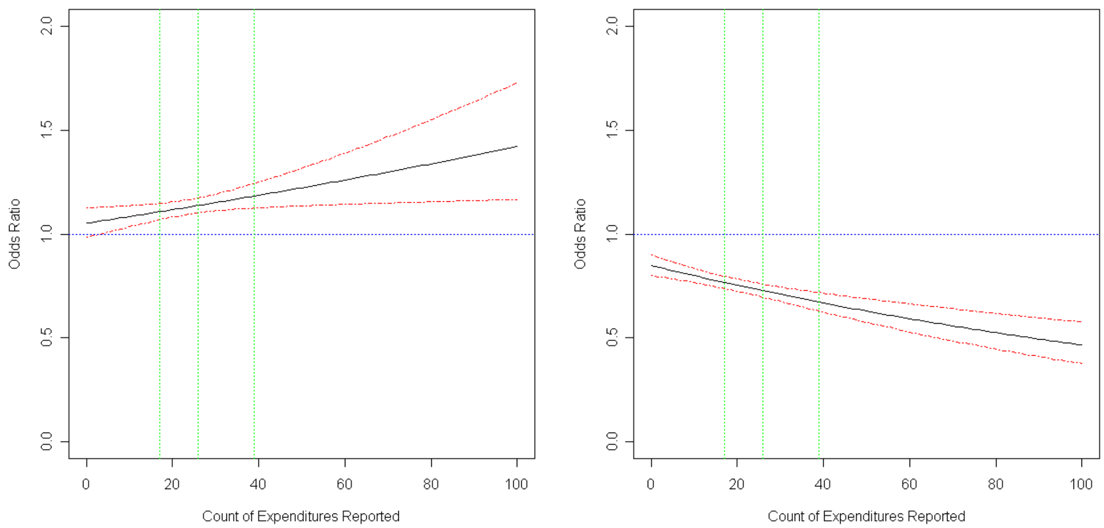

Figure 3: Utility bill usage: No vs. Yes (left); General record usage: Mostly to always vs. Occasionally to never (right)

The following trends were observed for utility bill usage. When comparing respondents who did not consult their utility bills during the interview to those respondents that did, the odds of needing editing on their reports increased as the count of expenditure reports increased. At all levels of expenditure reporting, the odds of needing editing was higher and statistically significant for those who did not consult their utility bills when compared to those that did.

The following trends were observed for general record usage. When comparing respondents who mostly to always consulted some kind of expenditure record to those that occasionally or never did, the odds of needing editing on their reports decreased as the count of expenditure reports increased. At all levels of expenditure reporting, the odds of needing editing was lower and statistically significant for those who mostly to always consulted their records when compared to those that occasionally to never did.

4. Discussion

Given our operational definition of data quality (whether an interview required post-collection data processing in the form of imputation and/or allocation), we found that several respondent, CU, and interviewer level characteristics were associated with poor data quality. In particular, the following respondent-level characteristics were significantly associated with data quality: age, education, Hispanic origin, language, sex, and converted refusal status (Table 2). For the CU-level characteristics, family type, housing tenure, and Census region were significantly associated with data quality (Table 3). Of the interview-level characteristics only final data collection mode and record usage were statistically significantly associated with data quality (Table 4).

The effect of these characteristics was significantly modified when an interaction with the count of expenditure reports was introduced into each of these models (Figures 1 – 3). On the one hand, some of these results are fairly intuitive. For example, respondents who use records (both general records and utility bill statements) to report expenditures during the interview seem to have better data quality than respondents who do not use such aids and this discrepancy gets larger as the number of expenditure reports increases. On the other hand, some of the findings are not readily explainable. For instance, why the discrepancy between Hispanic and non-Hispanic respondents with respect to their data quality is greater at higher levels of expenditure reporting is not immediately obvious. To explain these types of findings, more systematic assessments of data quality are necessary.

There are some limitations to this research that should be addressed in subsequent analyses. As noted in the introduction, we only use one metric of data quality, whether an interview needed post-collection editing. A complete assessment of data quality should include other dimensions quality from both the TSE and TQM frameworks. Second, we assessed each covariate set in separate models. A natural next step would be to devise a unified model relating all of the various characteristics to data quality. Third, we dichotomized each interview into one of two categories, poor and high data quality, based on the entire set of expenditures inquired about during the CEQ. The entire set accounts for over seventy percent of a typical household's spending. Furthermore, expenditures collected in this interview range from frequently incurred expenses, such as those for housing and utilities, to less frequent expenses, such as those for major appliances and funerals. It is reasonable to assume that expenditure reporting accuracy would vary by the characteristics of the expenditure (e.g., frequency of occurrence, dollar amount), so a next step would be to look at individual expenditure categories or items and assess the quality of those reports. Finally, since the CEQ is a panel survey, it would be interesting to investigate whether the quality of the reports given by a particular CU increased or decreased as their tenure in the panel increased. Studies such as Shields and To (2005) have looked at the concept of conditioned underreporting, i.e., a respondent learning to say "no" to stem questions so that he/she would not get asked a series of follow-up questions, but their assessment was only limited to expenditures during trips.

We acknowledge that the survey community agrees that collecting high quality data is important, but definitions of data quality and conclusions from data quality studies are less likely to be agreed upon. Perhaps a common definition of data quality is not essential because each survey has different objectives and goals, but key elements such as those described in Section 1.2 should be incorporated into every survey practitioner's definition of quality. Another issue is that it is not always clear how to use and address the information learned from data quality studies. One potential use of the type of information presented here is to help shape field procedures. If we concretely identify characteristics of the respondent (or CU, interview) that are associated with poor data quality, then we can disseminate this information to our interviewers so that they can be better prepared to work more closely with these respondents in order to elicit complete and accurate data. As a final note, the CE program is currently investigating alternative approaches to collecting its data. Some of these approaches would potentially result in a massive redesign of the survey and survey procedures (Gonzalez and Eltinge, 2008). Therefore, it is imperative that we understand the survey's current issues related to data quality as these should be first addressed when moving forward with a redesign.

5. Acknowledgments

The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the U.S. Bureau of Labor Statistics. The authors would like to thank Maxine Denniston, John Eltinge, Scott Fricker, Karen Goldenberg, Bill Mockovak, Adam Safir, and Lucilla Tan for their useful discussion on the Consumer Expenditure Surveys, data quality, and other contributions to this research.

References

Biemer, P.P. and Lyberg, L.E. (2003). *Introduction to Survey Quality*, New York: Wiley.

Brackstone, G. (1999). "Managing Data Quality in a Statistical Agency," *Survey Methodology*, 25, 139-149.

Bureau of Labor Statistics, U.S. Department of Labor, *Handbook of Methods*, Chapter 16, April 2007 edition, Consumer Expenditures and Income. http://www.bls.gov/opub/hom/pdf/homch16.pdf (visited July 16, 2009).

Gonzalez, J. M. and Eltinge, J. L. (2008). Adaptive Matrix Sampling for the Consumer Expenditure Quarterly Interview Survey. In *JSM Proceedings*, Survey Methods Research Section. Alexandria, VA: American Statistical Association.

Groves, R.M., Fowler J., Couper M.P., Lepkowski J.M., Singer E., and Rourangeau R. (2004). *Survey Methodology*, New York: Wiley.

Groves, R.M. (1989).*Survey Errors and Survey Costs*, New York: Wiley.

Hess, J., Moore, J., Pascale, J., Rothgeb, J., and Keeley, C. (2001). The Effects of Person-Level versus Household-Level Questionnaire Design on Survey Estimates and Data Quality. *Public Opinion Quarterly, 65*(4), p 574-584

Kennickell, A. B. (2002). Interviewers and data quality: Evidence from the 2001 Survey of Consumer Finances. In Proceedings of the Annual Meetings of the American Statistical Association. New York.

Shields, J. and To, N. (2005). Learning to Say No: Conditioned Underreporting in an Expenditure Survey. *American Association for Public OpinionResearch - American Statistical Association, Proceedings of the Section on Survey Research Methods*, 3963-8.

Yan, T., Tourangeau, R., and Arens, Z. (2004), "When Less is More: Are Reluctant Respondents Poor Reporters?" In Proceedings of the Annual Meetings of the American Statistical Association. Toronto.