An Analysis of the Mixed Collection Modes for Business Surveys at the U.S. Census Bureau

Broderick Oliver[1] and Katherine Jenny Thompson
U.S. Census Bureau
4600 Silver Hill Road, Room 7K031M, Washington, DC 20233

Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference

Abstract

Business surveys conducted by the U.S. Census Bureau generally make initial contact with the respondents by mail and provide a variety of response options: mail, secure internet collection, fax, and telephone. There are many hypotheses about the merits of each collection method. Telephone collection can be more expensive for the data collection agency than mail, internet, and fax. However, telephone and internet collection can provide better quality data than mail and fax collection because of the opportunity for checking data during the data collection process. Using historic data from two ongoing programs, we investigate data quality as a function of data collection mode using various quality measures, including the unit response rate and the quantity response rate.

1. Introduction

The term "quality" as used by statisticians to describe data does not have a single accepted meaning. Statistical agencies and organizations often define data quality by various dimensions. Eurostat (2003) for example, identifies six dimensions to data quality: relevance, accuracy, timeliness, accessibility and clarity, comparability, and coherence. Of these dimensions, accuracy is the most important because "without accuracy, other quality features are irrelevant" (Biemer and Lyberg, 2003, p. 14). However, accuracy, which is usually defined in terms of total survey error (sampling error + nonsampling error), is difficult, if not impossible to measure (See Biemer and Lyberg, pp. 34-43). Therefore, in this paper, when we use the term "data quality" (or "quality"), we mean reporting accuracy. Moreover, we assume that the values of the edited data are correct.

The mode of data collection (i.e., the medium used to obtain a survey unit's responses to survey questions) has an effect on the quality of the reported data. Different modes (1) provide access to different types of people; (2) attract different types of respondents; and (3) elicit different responses (Jäckle et al., 2009). Business surveys (programs) conducted by the Economic Directorate of the U.S. Census Bureau are predominately self-administered and employ a large variety of data collection techniques, including mail-based methods, telephone-based methods, computer-based methods, electronic-based methods, and administrative records-based methods (Nicholls et al., 2000). Hence, businesses have several response options.

There are many hypotheses about the merits of each data collection method. Telephone collection can be more expensive for the data collection agency than mail, internet, and fax collection. However, telephone and internet collection can provide better quality data than mail and fax collection because of the ability for real-time editing during the data collection process (Biemer and Lyberg, 2003). Willimack and Nichols (2010) provide evidence that electronic data collection instruments can improve data quality by aiding in data retrieval and collection process, especially if "they are designed around spreadsheet applications."

Here, we explicitly link quality to post-data collection processing procedures, specifically editing, imputation, and analyst review. The first two procedures are applied automatically to the collected data and are designed to obtain accurate tabulations from all eligible units. Analyst review is performed selectively. In this framework, we assume that as the percentage of retained "reported data" increases, the quality of the collected data likewise increases. We study the quality of survey data as a function of collection mode through various measures that assess how much of the data reported by respondents are retained after all data processing has been completed. Equally important, we examine interactions between unit size and collection mode.

[1] Any views expressed in this paper are those of the authors and not necessarily those of the U.S. Census Bureau.

In general, analyst procedures in business surveys are designed to improve the quality of the key estimates of totals. This is usually best accomplished by unit nonresponse follow-up of the large cases expected to contribute substantially to the estimate, followed by intensive analyst research for auxiliary data sources such as publicly available financial reports to replace imputed values with equivalent data (Thompson and Oliver, 2012). For the smaller cases that may only be reviewed by machine checks, the effectiveness of the data collection mode (instrument) for obtaining accurate reported data is especially important.

The Census Bureau is trying to increase the usage of internet collection over mail-out for its economic programs. For the two studied surveys, internet collection has been increasing. Given the skewed populations and the resultant focus on obtaining reported data from the largest units, the first research question that we examine is whether the increased use of internet collection is an across-the-board phenomenon or confined to the larger units. If the latter is true, then protocols or collection instruments designed for small businesses should be examined. In terms of data quality, we specifically examine whether internet collection appears to be improving data quality in terms of preserved reported data, again assessing whether the "improvement" – if it exists -- is limited to the large cases.

We introduce our notation and metrics in Section 2. In Section 3, we apply these metrics to historic data from two ongoing programs: the Quarterly Services Survey (QSS) and the Annual Capital Expenditures Survey (ACES). Both surveys use the Standard Economic Processing System (StEPS) developed at the U.S. Census Bureau for economic surveys (Sigman, 2001), which allows us to profit from the system's standard data flagging rules, respondent definitions, and existing metrics.

2. Definitions and Metrics

2.1. Definitions

For many economic programs, there is a need to distinguish between the survey (sampling) unit, the reporting unit, and the tabulation unit. A *survey unit* is an entity selected from the underlying statistical population of similarly constructed units (i.e. from the frame). A *reporting unit* is an entity from which data are collected. Reporting units are the vehicle for obtaining data and may or may not correspond to a survey unit for several reasons. For example, a company may request several forms (one per establishment) or a group of sampled establishments may request to report on a single consolidated form. Thus, the survey unit(s) establishes reporting unit(s) for their convenience. Lastly, a *tabulation unit* houses the data used in estimation. In the case of multiple reporting unit forms for the same unit, form data are consolidated to create a tabulation unit. In the case of a single reporting unit providing data for several establishments or other categories (e.g., industry), the reporting unit data may be split among the different categories. Note that the original definition of the survey unit may change in composition over time (perhaps due to mergers, acquisitions, or divestitures), and the associated reporting and tabulation units may likewise change.

Following the U.S. Census Bureau Quality Standards (Methodology and Statistics Council, 2012), a **respondent** is an eligible reporting unit for which

- an attempt was made to collect data
- the unit belongs to the target population
- *the unit provided sufficient data to be classified as a response*

Surveys establish their rules for respondent definitions by establishing required data items and conditions for requirement. A survey can elect to have one or more required data items and may require all required items provide sufficient data or that some combination of items provide sufficient data. For example, a survey could require sufficient data for two items, for one of two items, or for one of two items depending on the unit's classification (e.g., one item if unit sells material on the internet, a different item otherwise). These definitions are established before data collection begins and are not expected to change.

In the Economic Directorate, data are sufficient if processed data fields contain **reported data**. The respondent directly provides the values in the data collection period. Examples of reported data include data values that are reported by the respondent on the form and that pass data edits, data values that are reported by the respondent to an analyst, and data values that are reported on the form and receive minor corrections from an analyst (e.g., divide original value by 1,000).

Standard 3.3 of the Office of Management and Budget Statistical Standards (Federal Register Notice, 2006) states that "Agencies must add codes to collected data to identify aspects of data quality from the collection (e.g., missing data) in order to allow users to appropriately analyze the data. Codes added to convert information collected as text into a form that permits immediate analysis must use standardized codes, when available, to enhance comparability." StEPS identifies changes to collected data using two flags: a flag that identifies the module where the data change occurred (e.g., review and correction, simple imputation, imputation) and a flag that identifies the source of the final item value (e.g., wrong units, summing error, instructed by respondent to use company web site).

Although StEPS provides tools for correctly flagging each data item, the integrity checks on these flags can be easily disabled, especially on the source flag. Moreover, StEPS allows users to subjectively determine whether to flag a value as "Analyst Corrected" (treat as reported) or "Analyst Imputed" (treat as imputed). The metrics described assume that the standard flagging rules have been properly and consistently applied. Our applications do incorporate known exceptions provided by subject matter experts, which we discuss in Section 3.

2.2. Quality Measures

We use four quality measures to link quality to data collection mode.

The **Unit Response Rate (URR)** is the unweighted proportion of responding reporting units. In each statistical period, StEPS automatically computes the URR as

$$URR = \frac{R}{E + U} \tag{1}$$

Where
R = count of reporting units that were eligible for data collection in the statistical period and classified as a respondent.
E = count of reporting units that were eligible for data collection in the statistical period.
U = count of reporting units in the statistical period whose eligibility for reporting could not be determined.

The numerator of the URR can be further cross-classified to study contributions from mutually exclusive domains such as sampling unit type (certainty or noncertainty) or data collection mode.

Business populations are highly skewed, and large units consequently are included in samples with certainty or with very high probability (small design weights). Computing an unweighted rate reduces the influence of small units on the program level URR. However, we are interested in the interaction between response, size of the unit, and data collection mode. For this study, we use certainty status (certainty units versus noncertainty units) as a proxy for unit size. We examine interaction between data quality (in terms of retained reported data), unit size, and data collection mode.

The **Quantity Response Rate (QRR)** is the weighted proportion of an estimate that uses reported data. Unlike the URR, each item has its own QRR, so there may be several QRR measures per survey. StEPS automatically computes this measure for a given item y as

$$QRR = \frac{\sum_{i \in S} w_i \cdot y_i \cdot R_{Ti}}{T} \tag{2}$$

Where

$w_i =$ the unbiased sampling weight for the i^{th} tabulation unit

$y_i =$ the quantity of a key variable for the i^{th} tabulation unit

$R_{Ti} =$ 1 if i^{th} tabulation unit retained its reported value for the item; 0 otherwise

$T =$ the estimated weighted total of y (includes imputed data and nonresponse weight adjustment)

Using weighted estimates incorporates the contribution of each tabulation unit to the estimated total. We use the QRR to examine the interaction between data quality (in terms of retained reported data), size of the unit, and data collection mode for each *key item*, considering the same cross-classifications as with the URR.

The **Weighted Volume Response Rate (WVRR)** uses the frame measure of size (MOS) instead of a survey characteristic (y) in (2). This eliminates several sources of confounding in our analysis. First, the denominator (T) is never adjusted for nonresponse, as frame MOS is available by definition for all tabulation units. Second, the denominator of the WVRR is essentially constant throughout the survey design (it may change slightly due to status changes in tabulation units). This facilitates comparisons between statistical periods on the same item, c.f. the QRR, whose denominator will change each statistical period. Finally, unit response to a program can be determined by a single item, by more than one item ("and" relationships), or by responding to one of a group of items ("or" relationships). In the latter case, the QRRs for the set of required items can vary quite a bit, making analysis difficult. We use the WVRR to examine data quality as a function of unit size and data collection mode.

The **Source of Data Item (SDI)** measures the proportion of responding units that retain their reported data (i.e., reported value equals edited value) for an item. Similar to (1), this proportion uses unweighted counts. For multi-mode data collection, within an item, we compare the SDI by mode to determine whether there are particular data collection modes that retain a higher proportion of the reported data. By cross classifying the SDI by unit size, we can explore the influence of the size on a particular collection mode as well. If the collection mode is effective, we expect the SDI to approach 100%.

3. Quarterly Services Survey (QSS)

3.1. Background

The Quarterly Services Survey (QSS) is a principal economic indicator series that produces quarterly estimates of total operating revenue and the percentage of revenue by class of customer (government, business, consumers, and individuals) for selected industries. The survey also produces estimates of total operating expenses from tax-exempt firms in industries that have a large not-for-profit component. The QSS sample is comprised of service businesses with paid employees that operate in the covered sectors. A new QSS sample is selected every five years, and the sample is updated quarterly to reflect births and to (temporarily) exclude out-of-scope and inactive cases. For details on the QSS design and the estimation methodology, see http://www.census.gov/services/qss/qsstechdoc.html.

Businesses selected for the QSS may respond through the internet, by mail, fax, or telephone. Currently, the prevalent forms of data collection are internet, form (paper questionnaire), and, to a lesser extent, fax. For our study, we examine these modes separately and group the remaining modes (analyst phone contact, clerk phone contact, respondent phones in, and touchtone data entry) into a single "other" category. The QSS collects revenue from all units and operating expenses from a subset of units, depending on the industry in which the unit operates.
For our study, we used QSS data from 2009 through 2011, with revenue as the studied item. Unfortunately, the measure of size variable (census equivalent receipts) was not available for all eligible units in our historic data sets. Consequently, we cannot compute the WVRR for the QSS data.

Since response status is entirely determined by reported revenue, the *URR* is approximately the same as the *QRR* when response is representative in the sense that small and large units are reporting at the same rate. Likewise, the *SDI* values will be highly correlated with the two response rate measures.

3.2 QSS Results

In this section, we present the results obtained when we applied our quality measures discussed in Section 2 to QSS data from the first quarter of 2009 (i.e., 2009Q1) to the last quarter of 2011 (i.e., 2011Q4).

3.2.1 Response to the QSS by Size-of-Unit

Figure 1 plots the URR for the QSS for all eligible units (green), for the certainty units (black), and for the non-certainty units (purple). Within size-of-unit category (certainty/noncertainty/all), the URR are extremely stable across time. However, the URR for the certainty units is consistently higher than the URR for the noncertainty