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

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

$$URR = \frac{R}{E + U} \qquad (1)$$

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

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

Where
$w_i =$ the unbiased sampling weight for the ith tabulation unit

$y_i =$ the quantity of a key variable for the ith 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

units. The relatively high URR for the certainty units does not necessarily translate to a similar URR at the survey level (shown in green) because of the scarcity of certainty units in the sample.
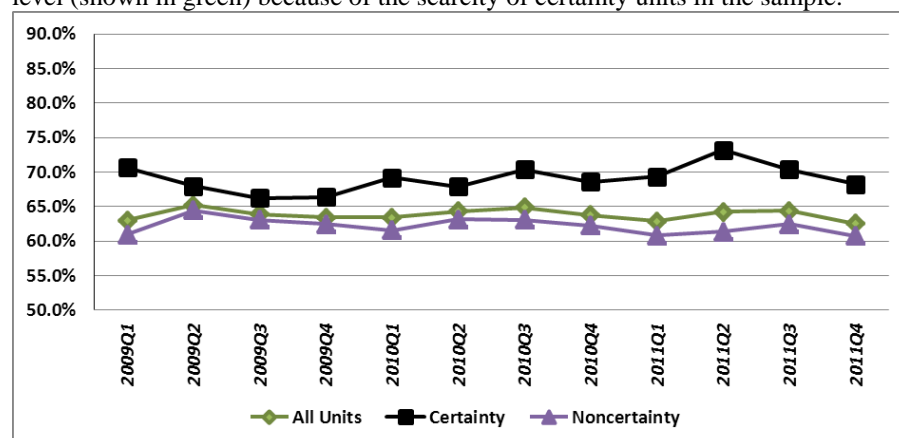

Figure 1: QSS URR by Size-of-Unit (2009Q1 – 2011Q4)

Figure 2 plots the QRR for revenue by size-of-unit category (certainty/noncertainty/all). In 9 of the 12 statistical periods, the *noncertainty* units have a higher QRR than the certainty units. That said, the corresponding measures are very close for most of the studied time-series, with absolute differences of less than three-percent from 2010Q1 through 2011Q3. In the absence of any other information, we suspect that this is in part a function of the analyst review procedures. Analyst review-and-contact may focus on the large certainty units, and large companies are more likely to have alternative sources of data available for replacement values (e.g. 10-Q filing). Because there are very few sources of alternative data that could be used for verification or substitution with the smaller units, a processed value from a noncertainty unit is unlikely to be examined by an analyst. Note that the QRR and the URR for revenue are generally very close, as the response status is usually determined by reported revenue.
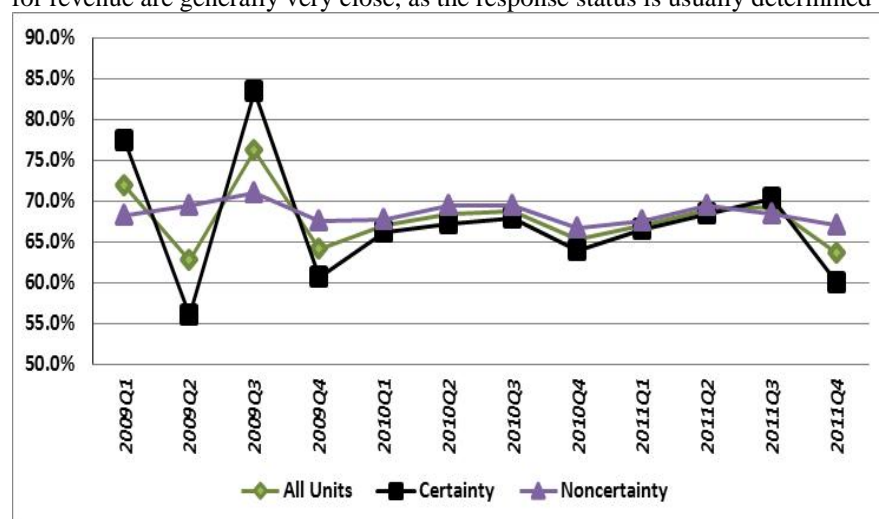

Figure 2: QSS QRR for Revenue by Size-of-Unit (2009Q1 – 2011Q4)

3.2.2 Response to the QSS by Data Collection Mode

Figure 3 plots URR by data collection mode. The primary data collection modes for the QSS are internet (red), form/questionnaire (blue), and fax (orange). We grouped the remaining modes into "other" (green). Within each statistical period, each data point represents the collection mode's *contribution* to the program-level URR for revenue (reported in Figure 1, green data points). Since the first quarter of 2010, internet collection has steadily increased, while collection by form has steadily decreased. Collection by fax has been steadily declining since the fourth quarter of 2010. The increase in the internet collection rate from 2011Q1 to 2011Q2 corresponds with a change in the QSS processing. Prior to 2011Q2, the first follow-up procedure for delinquent cases was to fax an

5

entire paper form. From 2011Q2 onward, the QSS changed the procedure and now only fax delinquent cases a *username* and *password* instead of a full form.
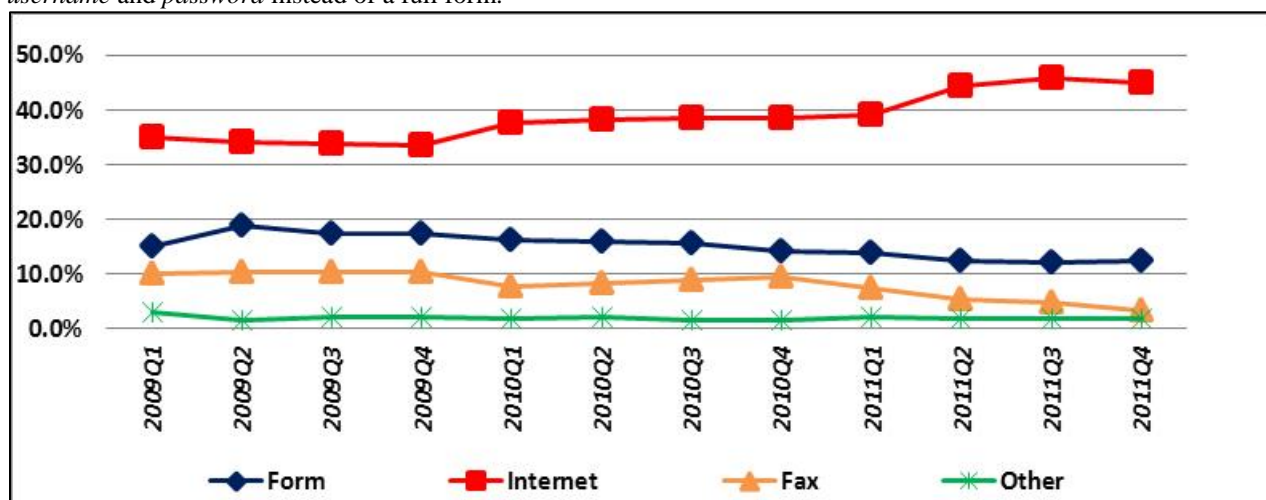
Figure 4 plots the QRR for revenue by mode of data collection. Within each statistical period, each data point represents the collection mode's *contribution* to the program-level QRR for revenue (reported in Figure 2, green data points). As with the URR, the majority of the QRR is derived from internet collection. After 2009Q2, the QRR percentage from internet collection is consistently much higher than the percentages obtained from the other collection methods and is quite stable. This provides evidence of the relationship between internet collection and quality in terms of obtaining usable reported data.

This result is not surprising. Over the past decade, the Census Bureau has made internet collection increasingly available and has performed numerous usability studies on the electronic collections instruments (Anderson and Harley, 2005). Although the data collection instruments are primarily designed for the larger businesses, they are available to all sampled units. Moreover, program managers are encouraged to promote internet collection as much as possible.
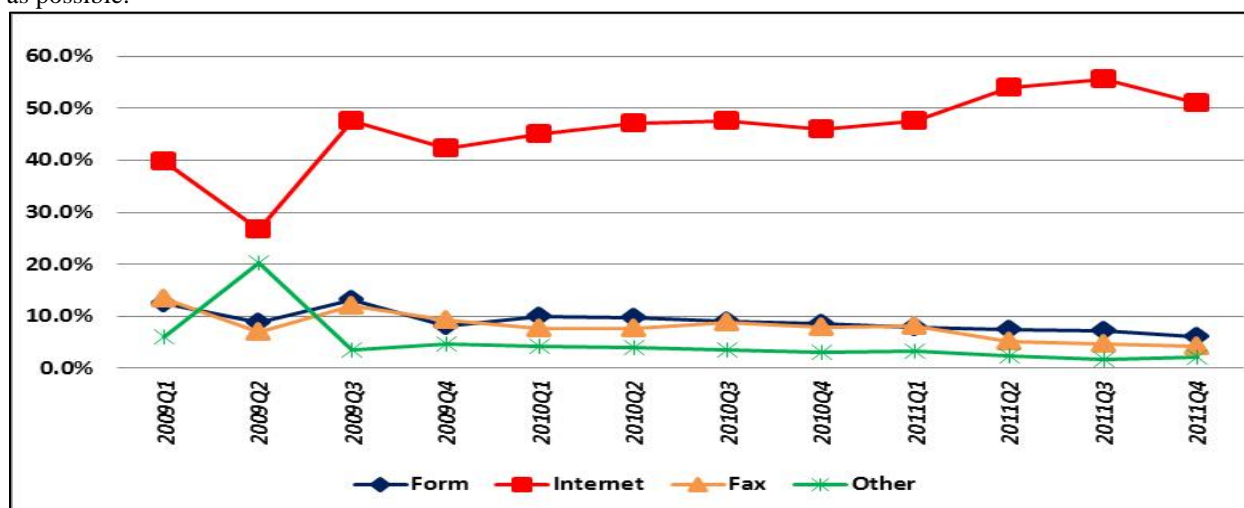
3.2.3 Examining Distribution of Response and Quality by Data Collection Mode

Figure 1 demonstrates that the larger units respond at a higher rate than the smaller units do, as is typical in a business survey. However, the reporting patterns are very consistent regardless of unit size, which may be at least in part due to the simplicity of the data collection instrument (questionnaire). After an introductory period, internet

6

collection has been consistently the most utilized, and values obtained by the internet collection contribute most to QRR for revenue (as demonstrated in Figures 3 and 4, respectively). On the surface, there is no obvious relationship between unit size and collection mode, and barely any relationship between unit size and quality. Here, we examine the three components jointly. Figure 5 examines the QRR for revenue by size-of-unit (certainty versus noncertainty) and mode of data collection (internet versus non-internet {form, fax, other}). Since 2010Q1 until 2011Q3, the data collected by internet consistently accounts for the largest share of the QRR value, regardless of unit size category (certainty/noncertainty). Of course, fewer cases are reporting by form, so this result is not unexpected.
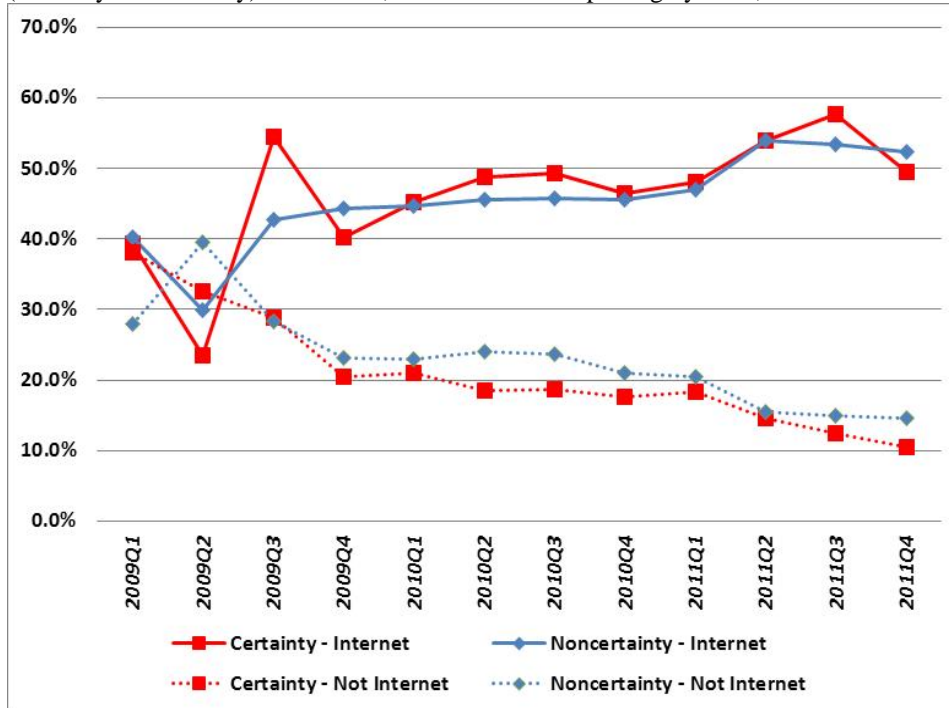

Figure 5: QSS QRR for Revenue (Reported Units Only) by Size-of-Unit and Collapsed Data Collection Mode (2009Q1 – 2011Q4)

Lastly, we examine the proportion of the responding units that retain reported data for revenue, by unit size. Figure 6 plots the SDI for revenue by size-of-unit for the five most recent quarters.
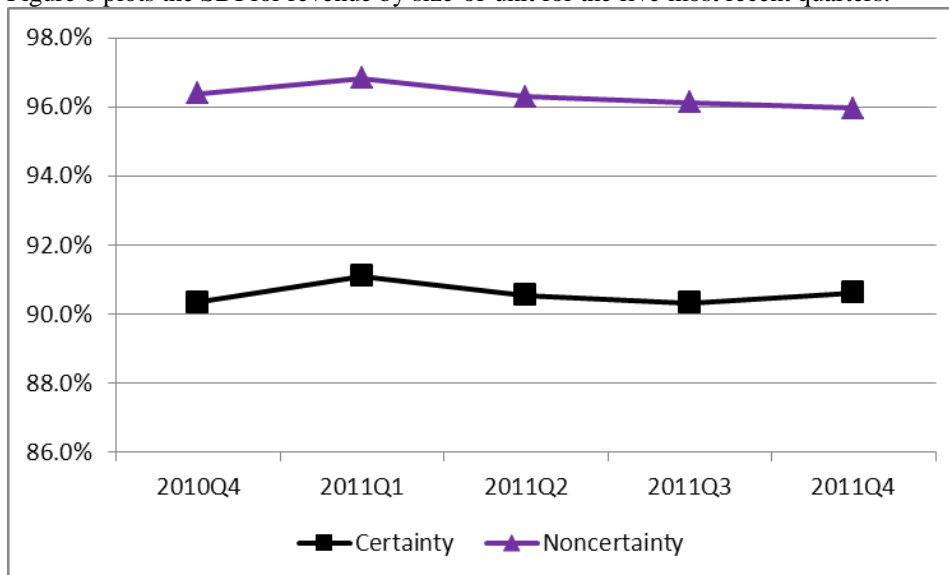

Figure 6: QSS SDI for Revenue by Size-of-Unit (2010Q4 – 2011Q4)

7

Regardless of unit size (certainty versus noncertainty), the percentage of retained reported data (i.e., reported value = edited value) is extremely high. The SDI values for the noncertainty units are consistently higher than the SDI for the certainty cases. Table 1 presents the SDI for revenue cross-classified by unit size and collection mode from the five most recent quarters studied.

Generally, the revenue values reported to the QSS are retained after processing, regardless of unit size or collection mode. Recall that the QSS questionnaire is quite straightforward and collects very few items. Hence, it is not a very complex form, and there is very little reporting burden associated with the survey. That said the certainty units have lower retention rates across the board. This is consistent with earlier-presented results showing that the QRR is actually *lower* for certainty units than for noncertainty units. That is, analysts appear to alter or use alternate source data at a higher rate for certainty cases than for noncertainty cases, which makes sense, as large companies are more likely to have alternate source data.

Table 1: QSS SDI for Revenue by Size-of-Unit and Data Collection Mode

Collection Mode	Unit Size	2010Q4	2011Q1	2011Q2	2011Q3	2011Q4
Form	*Certainty*	91.1%	93.0%	91.2%	87.8%	92.8%
	Noncertainty	96.9%	97.9%	97.2%	96.5%	96.4%
Internet	*Certainty*	89.9%	90.7%	90.2%	90.5%	90.5%
	Noncertainty	96.0%	96.3%	95.8%	96.0%	95.9%
Fax	*Certainty*	91.7%	91.3%	91.4%	90.7%	89.1%
	Noncertainty	97.2%	97.4%	96.8%	96.4%	95.3%
Other	*Certainty*	92.2%	91.7%	95.1%	93.8%	90.4%
	Noncertainty	94.3%	96.1%	97.9%	95.9%	95.9%

4. Case Study: Annual Capital Expenditures Survey (ACES)

4.1 Background

The Annual Capital Expenditures Survey (ACES) provides data on capital spending for new and used structures and equipment by U.S. nonfarm businesses with and without employees. Respondents report capital expenditures for the calendar year in all subsidiaries and divisions for all operations within the United States. The ACES selects a new sample each year. For more details on the ACES design and estimation procedures, see http://www.census.gov/econ/aces/.

The ACES collection instrument is quite complex. Figure 7 presents the extracted total capital expenditures collection complex from the ACE-1 short form (the same data collection matrix is used for all ACES cases). Respondents are instructed to complete the matrix of capital expenditures information completely and to ensure that all rows and columns are fully additive. Businesses that operate in more than one industry are later asked to further break down the aggregate company level totals by industry. The more detailed breakdown values must add to the totals reported earlier in Item 2 (see Figure 7)[2] and to a separately collected value of total capital expenditures -- reported in Item 1 of the questionnaire. In addition, businesses must provide some explanation for all amounts listed in the "Other" category of Item 2. Unlike the QSS, administrative data are not available for the ACES sample units, although some of the larger companies may publish their annual capital expenditures, which the analysts may use for validation.

---

[2] In addition, ACES has a more detailed data collection every five years – 2003, 2008, and 2012 were such years.

Figure 7: Capital Expenditures Collection for ACES

Currently, the most prevalent forms for collecting respondent data are forms (questionnaires), the internet (via Centurion), and to a lesser extent, fax. For our analysis of the ACES, we grouped all other data collection modes (analyst phone contact, clerk phone contact, respondent phones in, and touchtone data entry) into a single "other" category. We use historic data from the 2002 through 2011 data collections, but note that ACES introduced internet collection in 2006.

ACES classifies a unit as a respondent if it reported a valid value for at least one of the seven "non-other" cells provided in Figure 7 (all items in columns (1), (2), and (4), to be specific). We computed all measures for three of these items, but only present results for total capital expenditures, as the results obtained for capital expenditures on structures and capital expenditures on equipment were essentially the same. Since the ACES respondents have to provide at least one valid value of a capital expenditures item, there is a possibility that the QRR for some items is low. The WVRR, which uses the business register payroll value on the frame, will better capture the impact of a given unit in terms of size on the weighted volume measures.

4.2 Results

In this section, we present the results obtained when we applied our quality measures discussed in Section 2 to ACES data from the 2002 through 2011 data collections.

4.2.1 Response to the ACES by Size-of-Unit

Figure 8 plots the URR for ACES for all eligible units (green), for the certainty units (black), and for the non-certainty units (purple). The unit response rates for the certainty cases are quite stable. In this program, each subject matter expert is assigned to review and monitor a group of industries. Analysts often have frequent personal contact with the largest companies in their assigned industries and are instructed to focus on obtaining their reported data and making certain that their reported data are consistent. However, since 2008, the rates for the noncertainty units have been declining. However, this survey does have a high percentage of noncertainty units, so the program level URR (shown in green) is affected by their decline in response.
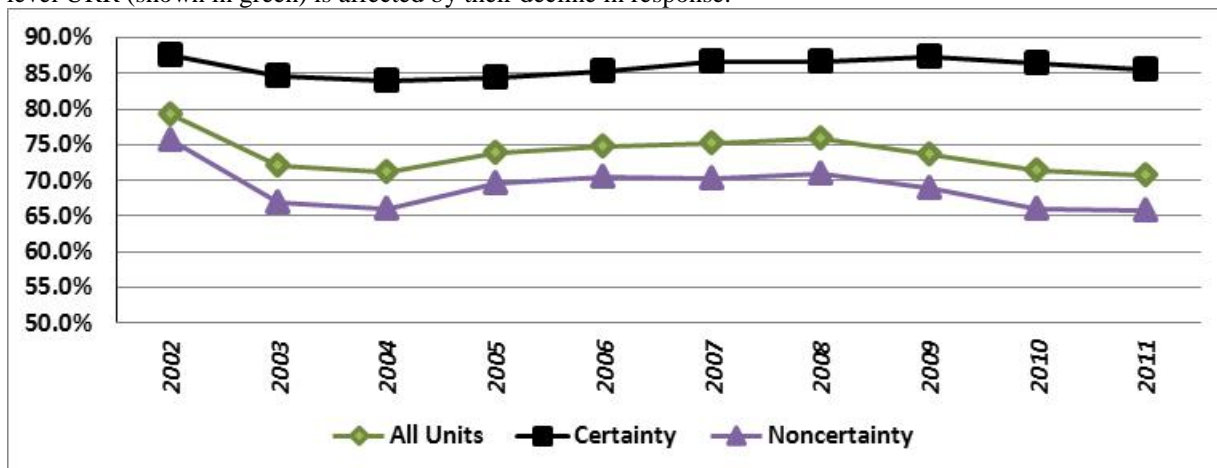

Figure 8: ACES URR by Size-of-Unit (2002 – 2011)

9

If the larger cases represent the majority of a weighted tabulation, then the QRR for the tabulation item will be larger than the URR. In Figure 9, we present the QRR for total capital expenditures, using data from the 2007 through 2011 collections[3]. The high percentage of retained reported data from the certainty units indeed contributes to the very high survey level values of the QRR (shown in green). The lower percentage of reported data retained by the noncertainty units does affect the QRR, but the impact is fairly muted.
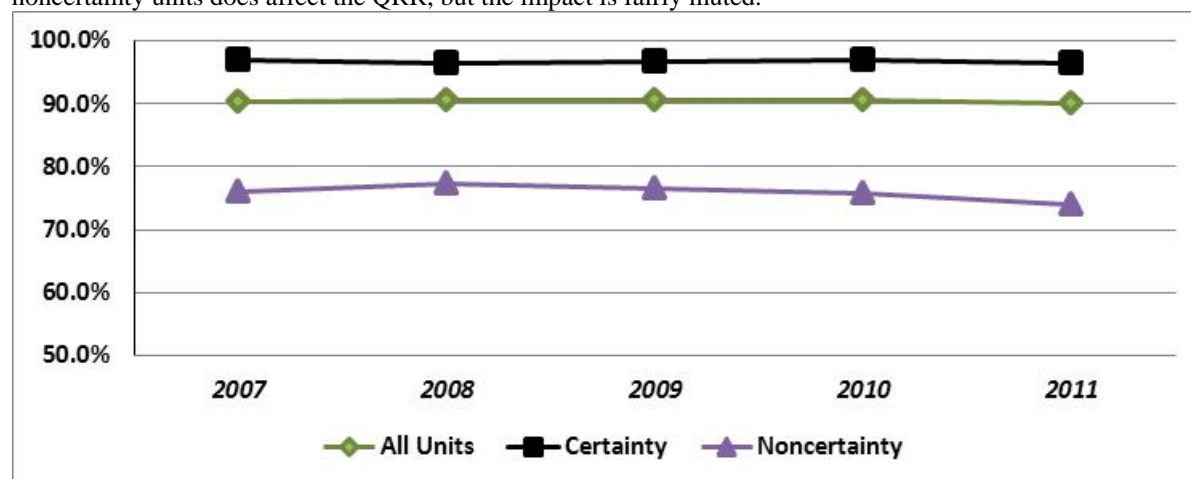

Figure 9: ACES QRR for Total Capital Expenditures (2007 -2011)

Recall that to be classified as a responding unit in the ACES, reported data must be retained in one of the nine items collected in the capital expenditures matrix shown in Figure 7. By examining the WVRR, we remove potential confounding related to a specific item, allowing us to assess the relationship between response and unit size. Table 2 presents the ACES WVRR and the URR by unit size.

For the certainty units, the contact strategies are working very well overall, with almost all of the certainty units providing valid reported data on at least one item (see WVRR values in the "Certainty Units" column). For the noncertainty units, there may be some room for improvement. Even though more reported data are captured in terms of contribution to weighted estimates than would be demonstrated by the URR, the WVRR measures are far from 100%.

Table 2: ACES WVRR versus URR by Size-of-Unit (2007 – 2011)

	All Units		Certainty Units		Noncertainty Units	
Year	WVRR	URR	WVRR	URR	WVRR	URR
2007	88.5%	75.2%	95.6%	86.6%	79.6%	70.2%
2008	88.9%	75.8%	95.4%	86.6%	80.7%	70.9%
2009	88.6%	73.6%	95.4%	87.3%	79.9%	68.8%
2010	88.6%	71.4%	95.2%	86.5%	79.9%	66.0%
2011	87.7%	70.7%	94.7%	85.6%	78.4%	65.7%

4.2.2 Response to ACES by Data Collection Mode

Figure 10 plots the unit response rates by mode of data collection. The primary data collection modes for the ACES are internet (red), form (i.e., questionnaire) (blue), and fax (orange). Within each statistical period, each data point represents the collection mode's *contribution* to the program-level URR for revenue (reported in Figure 8, green data points). Note that ACES did not introduce internet collection into this survey until 2006. However, its use has been climbing steadily while other modes of data collection have been simultaneously declining.

---

[3] We exclude the earlier data to allow comparability with analysis by collection mode, as internet collection was not introduced until 2006. We exclude the 2006 data collection to provide a small burn-in period.
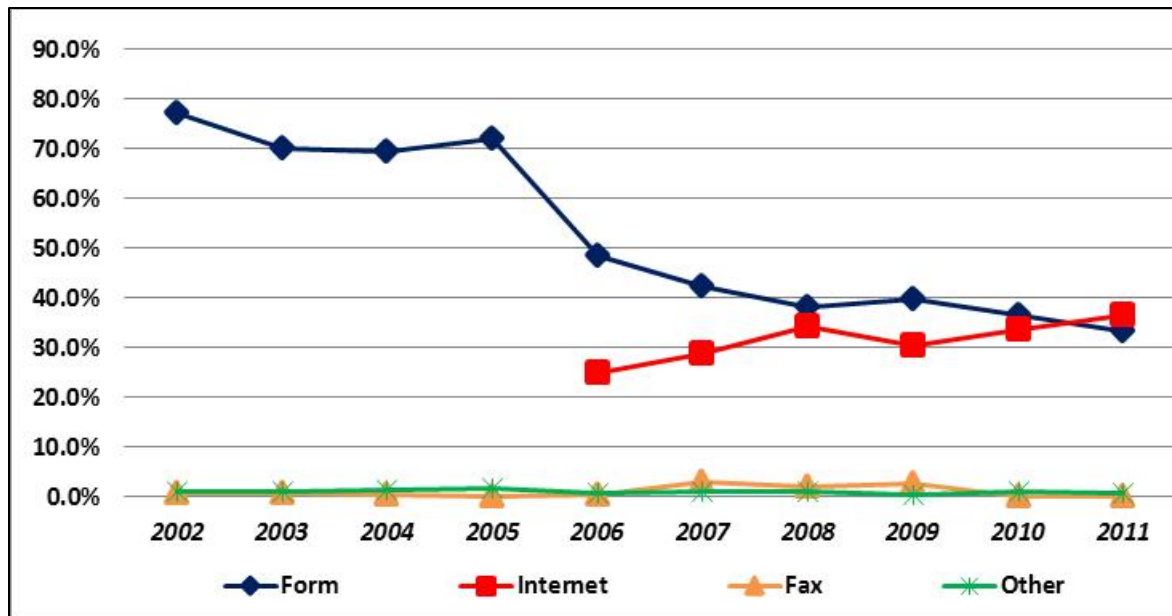
Figure 10: ACES URR by Data Collection Mode (2002 – 2011)

Figure 11 plots the QRR for total capital expenditures by data collection mode, starting with the 2007 data collection. Within each statistical period, each data point represents the collection mode's *contribution* to the program-level QRR for revenue (reported in Figure 9, green data points). Data reported via the internet is accounting for an increasing percentage of the retained reported data for total capital expenditures.
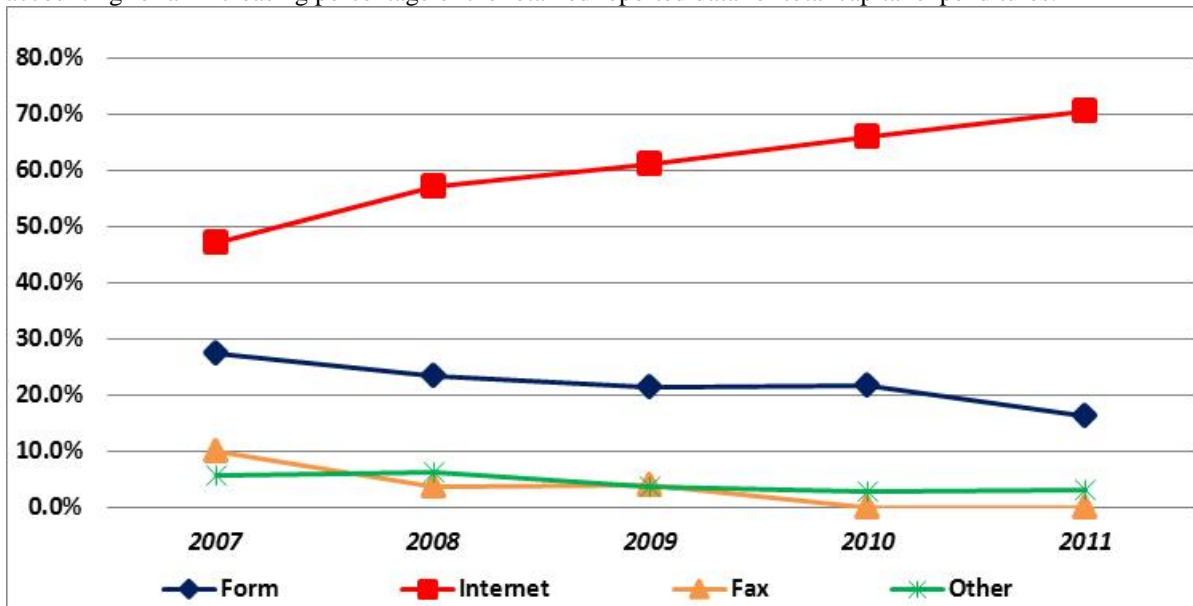

Figure 11: ACES QRR for Total Capital Expenditures by Collection Mode (2007 – 2011)

4.2.3. Examining Distribution of Response and Quality by Data Collection Mode

The earlier sections show that in ACES, the larger units (certainty units) consistently have higher reporting rates than the noncertainty units do and that internet collection has become the most utilized method. Here, we examine the quality of response for total capital expenditures by unit size (certainty versus noncertainty) and data collection mode (internet versus non-internet {form, fax, other}). See Figure 12 below.

11

Total capital expenditures collected by internet account for a larger share of the QRR for both the certainty and noncertainty units (except in 2007 for the latter) than the non-internet modes. Moreover, with internet collection, the percentage of retained reported data is increasing, regardless of unit size. The difference in mode usage is very apparent for the certainty cases. In contrast, for the noncertainty units, the percentages of retained reported data do not really differ by collection mode. Given the increased across-the board usage of the internet, this lack of a difference is indicative of a potential area of improvement for ACES. It is possible that the internet form could be made more user friendly for small companies or perhaps be better advertised.
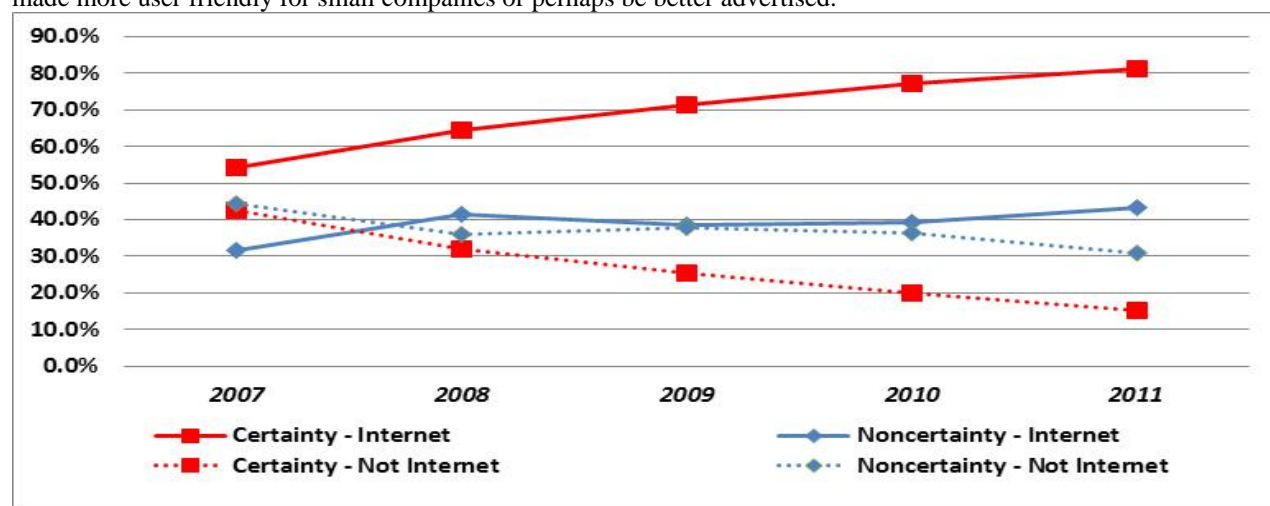

Figure 12: ACES QRR for Total Capital Expenditures by Size-of-Unit and Data Collection Mode (2007 – 2011)

Table 3 presents the WVRR by mode for the ACES. For the internet collection, the WVRR is always larger than the URR, again demonstrating the impact of the larger cases' usage of internet reporting. The differences between the WVRR and URR for the non-internet collection modes are less pronounced. For form collection, the WVRR values are lower than the corresponding URR values. This indicates that smaller units may not be taking full advantage of the internet collection.

Table 3: WVRR and URR for ACES by Data Collection Mode

Year	Form		Internet		Fax		Other	
	WVRR	URR	WVRR	URR	WVRR	URR	WVRR	URR
2007	35.4%	42.3%	41.3%	28.8%	8.0%	3.0%	3.7%	1.1%
2008	30.7%	38.1%	49.4%	34.4%	4.6%	2.1%	4.2%	1.2%
2009	28.5%	39.7%	52.2%	30.6%	4.7%	2.8%	3.1%	0.6%
2010	27.3%	36.7%	58.4%	33.6%	0.2%	0.2%	2.6%	0.9%
2011	23.0%	33.2%	61.8%	36.6%	0.1%	0.1%	2.9%	0.8%

Finally, we examine the proportion of responding units that retain reported data (i.e., reported equals adjusted) by certainty status and size of unit. Figure 13 plots the SDI for total capital expenditures by certainty/noncertainty status. Except for the 2008 collection, the corresponding SDI values are very close, and in both cases, the reported item values are consistently retained. A subject matter expert hypothesizes that the lower SDI for the certainty firms could have been caused by the expansion of data items in that survey year. As with the QSS, this is less likely a function of the high quality of data reported by the smaller units than a consequence of the analyst review procedures.
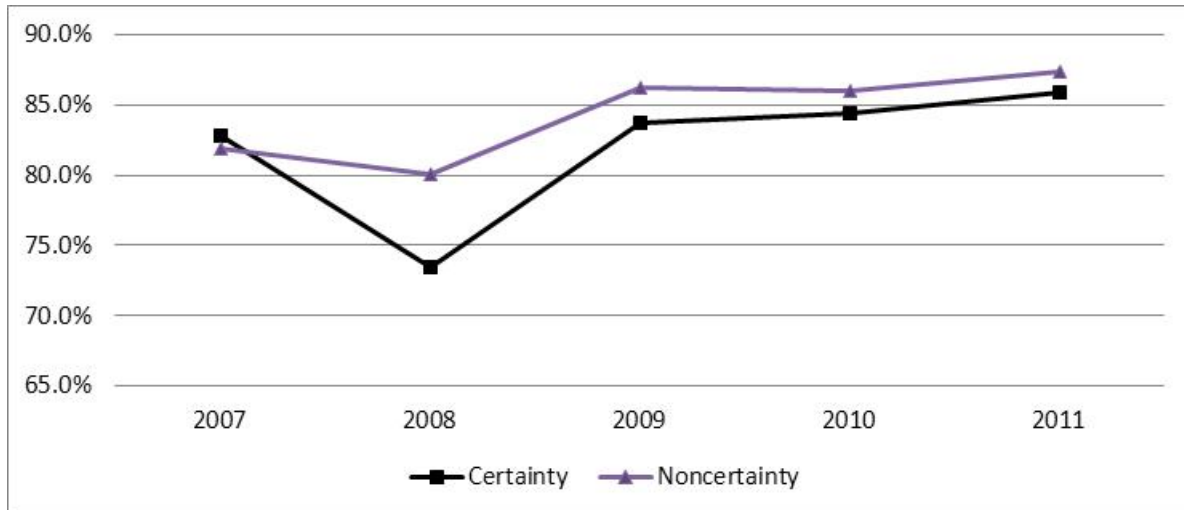
Figure 13: SDI for Total Capital Expenditures by Size-of-Unit (2007-2011)

Figure 14 examines the SDI by data collection mode. The percentage of retained reported data is consistently quite high for the three most heavily utilized data collection modes. The highest percentages of retained reported data are consistently obtained through internet collection. That is good news, in the sense that the internet collection is retaining the same level of quality or higher than other methods.
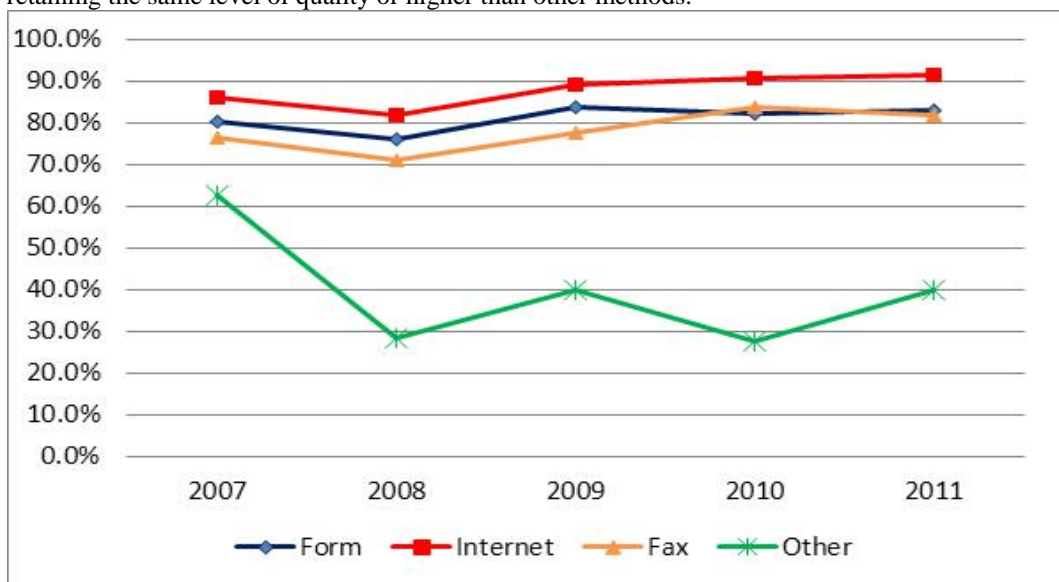

Figure 14: ACES SDI for Total Capital Expenditures by Data Collection Mode (2007-2011)

5. Discussion

In this paper, we apply several quality measures to historic data from two separate programs to evaluate the effects of mixed data collection modes on data quality. Both programs are business surveys, characterized by highly skewed populations. Both programs are sample surveys, with small units having large sampling weights. There are, however, some key differences. First, the revenue data collected by the QSS are expected to be non-zero for the majority of the sampled units, and corresponding values are often available from alternative source data. With the ACES, not all companies capitalize their expenditures, many small companies report legitimate zero values, and administrative data are not available. Second, the data collection instrument for the QSS is straightforward, with few items collected. The ACES' data collection instruments require complex additive matrices. In addition, the item definitions (data concepts) for the ACES are quite complex.

13

Paper form collection has always been problematic, especially in mail-out surveys. Collection is even more challenging with large businesses, where the paper form may be completed in several different offices before being mailed, faxed, or keyed into the internet site (Snijkers et al, 2013). Almost 30 years ago, Granquist advocated, "moving editing closer to data collection" by developing electronic questionnaires that performed simple edit checks as data are entered by the respondent (Granquist, 1984). The Economic Directorate of the Census Bureau has been moving towards this goal by offering voluntary internet collection for many economic programs. With simple questionnaires, few edit checks are required, and electronic implementation is straightforward. With more complex questionnaires, form designers must balance usability with accuracy, which means that some edits may be sacrificed to avoid frustration on the respondents' part. Guidelines are available to help achieve this balance. However, pretesting is often confined to large businesses, and the difficulties experienced by a small business may not be addressed (see Nichols et al, 2005; Willimack and Nichols, 2010).

Our analyses demonstrated increased usage of internet collection for both programs. We assume that the implemented machine edits are very effective in locating erroneous data, and that the retained reported data are valid. However, we also note that alternative data sources are more often available for large businesses. As stated in the introduction, program managers strive to reduce the proportion of imputed values in the tabulated totals. Consequently, analysts will attempt to find an alternative value considered of equivalent quality to reported data for substitution over a machine-imputed value. As shown with the QSS results, less reported data might actually be retained for the larger cases because viable options for substitution exist; similarly, analysts may restore reported data for a smaller case that fails less important edits instead of accepting a machine-impute.

By funneling down from detailed response rates to proportions of retained reported data by item, we can assess whether a particular collection mode is improving (or not hurting) data quality. By incorporating unit size in our analyses, we can assess whether the improvements, if they exist, are found across-the-board or confined to a selected set of cases. When the questionnaire is not complex and the data concepts are straightforward, internet collection works quite well and does not seem to be affected by the size of unit. On the other hand, when these two conditions do not hold, although valid, voluntary internet collection is more confined to larger units. This can have quality implications, as the data collected via internet are validated by respondents during data entry, whereas editing is not performed until *after* data entry with paper form collection.

Business surveys are designed to produce reliable totals. As mentioned before, analyst review procedures are likewise designed to produce reliable totals, and generally focus on obtaining valid data for the largest units. There is a gap in the review procedures for small businesses. In part, this could be addressed by more adaptive collection procedures, such as subsampling small unit nonrespondents for follow-up, subsampling respondents for reinterview, or developing analyst review protocols that require checking a minimum number of small units. That said, such changes need to be researched and tested, and budget must be made available. However, internet collection is already available, and there is a missed opportunity to self-validate the data when internet collection is *not* used. When a subpopulation does not fully take advantage of a data collection mode that has such quality benefits, the administering organization should investigate. If the subpopulation is small businesses, this may represent a major paradigm shift. That said, the cost of investigating how to improve internet collection instrument usability – or increase desirability of internet usage – for a non-participating subpopulation could be greatly offset by the decrease in cost via internet collection. Ultimately, a survey whose collection instrument minimizes the probability of response error for all units maximizes a major component of quality. That is priceless.

References

Anderon, A.E. and Harley, Diane. (2007). "Using Software to Collect Data Electronically for the Economic Census." Paper presented at the ICES-III, June 18-21, Montreal, Canada. Retrieved from http://www.amstat.org/meetings/ices/2007/proceedings/ICES2007-000178.PDF.

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

Eurostat (2003). "Item 4.2: Methodological Documents: Definition of Quality in Statistics," Sixth Working Group Meeting of Assessment of Quality in Statistics. Luxembourg, October 2-3, 2003.

Federal Register Notice (2006). OMB Standards and Guidelines for Statistical Surveys. Washington, DC: Retrieved from http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf

Granquist, L. (1984). "On the Role of Editing." *Stastistisk Tidskrift*, **2**, pp. 105 – 118.

Jäckle, A., Roberts, C., and Lynn, P. (2008). "Assessing the Effect of Data Collection Mode on Measurement." Institute for Social and Economic Research (ISER) Working Paper. 2008-08. Colchester: University of Essex.

Methodology and Statistics Council (2012). *U.S. Census Bureau Statistical Standards*. Washington, DC: U.S. Census Bureau.

Nicholls, W.L., Mesenbourg, T.L., Andrews, S.H., De Leeuw, E. (2000). "Use of New Data Collection Methods in Establishment Surveys." *Proceedings of the Second International Conference on Establishment Surveys.* American Statistical Association.

Nichols, E.M., Murphy, E.D., Anderson, A.E, Willimack, D.K., and Sigman, R.S. (2005). Designing Interactive Edits for U.S. Electronic Economic Surveys and Censuses: Issues and Guidelines. *Internal Census Bureau report, available upon request.*

Snijkers, G., Haraldsen, G., Jones, J., & Willimack, D. K. (2013). Designing and Conducting Business Surveys. Hoboken, NJ: John Wiley & Sons, Inc.

Sigman, Richard (2001). "Editing and Imputation in a Standard Economic Processing System." *Proceedings of Statistics Canada Symposium*, Ottawa, Ontario: Statistics Canada.

Thompson, K.J. and Oliver, B. (2012). "Response Rates in Business Surveys: Going Beyond the Usual Performance Measure." *Journal of Official Statistics*, **28**, 221-237.

Willimack, D.K. and Nichols, E. (2010). "A Hybrid Response Process Model for Business Surveys." *Journal of Official Statistics*, **1**, 3-24.