Journal of Official Statistics, Vol. 28, No. 2, 2012, pp. 173–198

Sources of Error in Survey and Administrative Data: The Importance of Reporting Procedures

Jeffrey A. Groen[1]

With increased attention to administrative data for statistical purposes, analyses of the quality of administrative data and comparisons to survey data are greatly needed. This article presents a methodology for identifying sources of error in administrative and survey data and for identifying sources of differences between administrative and survey estimates. The first part of the methodology is a statistical decomposition of the difference between administrative and survey estimates. The second part investigates the causes of measurement error and the factors associated with differences between what the same respondents report in the survey and the administrative records. I illustrate this methodology using a case study of the monthly employment figures gathered from U.S. businesses. This analysis demonstrates that both administrative data and survey data may contain errors and that reporting procedures are likely to differ between the two types of data. The article also identifies practical ways to assess data quality.

Key words: Measurement error; data collection; establishment survey; Current Employment Statistics (CES) survey; Quarterly Census of Employment and Wages (QCEW).

1. Introduction

Administrative data offer several advantages over survey data, including larger sample size, lower cost, and lower respondent burden. As a result, national statistical agencies are trying to make better use of administrative data (Eurostat 2003; Prevost and Leggieri 1999; Wallgren and Wallgren 2007). Administrative data also have some disadvantages, including inappropriate constructs, matching problems, and other measurement issues. As a result, analyses of the quality of administrative data and reasons for differences between administrative data and survey data are greatly needed.

Researchers and practitioners have taken several approaches to the comparison of administrative and survey data. The most common approach is to assume that the administrative data represent the truth (e.g., Bound and Krueger 1991; Kreuter et al. 2010; Pischke 1995). This assumption is convenient, but for several reasons it may not be appropriate. Administrative data may contain measurement error because of problems

[1] U.S. Bureau of Labor Statistics, 2 Massachusetts Avenue NE, Room 4945, Washington, DC 20212, U.S.A. Email: groen.jeffrey@bls.gov

Acknowledgments: This article has benefitted from comments of participants at the American Economic Association meetings, the Society of Government Economists conference, and the UNECE/OECD/Eurostat meeting on business registers. I am grateful to Greg Erkens, Jason Faberman, Kristin Fairman, Scott Fricker, Julie Gershunskaya, Larry Huff, Chris Manning, Polly Phipps, Anne Polivka, Jim Spletzer, and Michele Walker for useful comments and helpful discussions. I also thank an Associate Editor and five anonymous referees for comments that greatly improved the article. The views expressed in this article are solely those of the author and do not reflect the views of the U.S. Bureau of Labor Statistics.

with matching individual observations across multiple sources (Kapteyn and Ypma 2007), because of reporting errors (Abowd and Stinson 2011), because of errors in constructing the statistical units of interest (Zhang 2011), because the construct behind a particular measure in administrative data differs from the construct that analysts require (Davies and Fisher 2008; Johnson and Moore 2008), or for other reasons.

Whether the administrative data can be assumed to be correct depends, of course, on the particular data source and proposed statistical use. It is well accepted that survey data are subject to measurement error and other types of error (Groves et al. 2009), but the notion that administrative data may contain errors is less widely recognized. Although some national statistical agencies have developed tools for assessing the quality of administrative data (e.g., Berka et al. 2010; Daas et al. 2009; Laitila et al. 2011), a conceptual framework has yet to develop that is as comprehensive and widely accepted as the "total survey error" paradigm for evaluating survey data (Groves et al. 2009).

This article presents a methodology for identifying sources of error in administrative and survey data and for identifying sources of differences between administrative and survey estimates. This methodology can be applied in contexts where the administrative data provide the sampling frame for the survey and the data item of interest is collected in both sources. The first part of this methodology is a statistical decomposition of the difference between administrative and survey estimates. The decomposition involves allocating the overall difference into four sources, three of which are familiar errors in survey-based estimates: coverage error, sampling error, and nonresponse error. The fourth source is differences between what the same respondents report in the survey and what they report in the administrative records. This source, which I refer to as "reporting differences" between survey and administrative data, is suggestive of the extent to which measurement error is present in the administrative data, survey data, or both.

The second part of the methodology investigates the causes of measurement error and the factors associated with reporting differences between survey and administrative data. This part requires matching the survey data to the administrative data at the micro level, computing measures of reporting differences for the matched sample, and relating these differences to characteristics of the reporting units, measures of how the data are collected and processed for each source, and the procedures used by reporting units to compile the data item that is reported in the survey and the administrative records. Some pieces of this information may be obtained from the internal records of data producers, but other pieces may require a special data collection such as a response analysis survey (RAS). Once this methodology is applied in a particular context, specific recommendations can be developed to reduce particular sources of error in both survey and administrative data and to reduce the magnitude of differences between survey and administrative estimates.

I illustrate this methodology using a case study of the monthly employment figures gathered from U.S. business establishments. The administrative data come from the Quarterly Census of Employment and Wages (QCEW), which is based on mandatory quarterly Unemployment Insurance (UI) reports. The survey data come from the Current Employment Statistics (CES) survey. Statistics based on these sets of data are produced by the U.S. Bureau of Labor Statistics (BLS) and used extensively by economists, policymakers, researchers, financial analysts, and government agencies as indicators of current economic conditions and as measures of labor-market activity.

The next section of the article describes the methodology and data. The first part of the section presents background information on the QCEW and CES and compares the respective estimates at the aggregate level. The section then develops the decomposition for the difference between administrative (QCEW) and survey (CES) estimates of employment along with a related decomposition for differences in employment growth. The remainder of Section 2 explains the method for exploring causes of reporting differences at the micro level in two matched samples of QCEW-CES data. The first matched sample is a large sample of CES respondents matched to their QCEW data. With this sample, I use regression analysis to identify whether reporting differences between QCEW and CES are related to particular establishment characteristics, imputation in the QCEW, and the timing of data collection. The second matched sample comes from a RAS in which a sample of CES respondents was contacted to gather information about the methods and sources used to compile employment data for QCEW and CES. With this sample, I use descriptive statistics to compare the reporting procedures used by respondents to provide data to both sources. I also use regression analysis to relate differences in reporting procedures to differences between the QCEW and CES data reported by establishments.

The results of the three pieces of empirical analysis are presented in Section 3. Section 4 discusses some practical implications of the comparative analysis for understanding the sources of micro-level reporting differences between QCEW and CES data and suggesting ways that BLS can reduce the magnitude of differences in employment estimates between the two sources. The section also discusses implications of the analysis for comparisons of administrative and survey data in other contexts, including sources of error in both types of data and practical ways to assess the quality of administrative data.

2. Methodology and Data

2.1. Background on QCEW and CES

2.1.1. QCEW Data

The QCEW is a quarterly census of all U.S. business establishments subject to UI taxes; in 2010, it covered approximately 9 million establishments nationwide and 98 percent of U.S. employment. Each quarter, businesses are required to submit quarterly contribution reports, either electronically or by mail, to state agencies responsible for administering UI programs. In their quarterly reports, businesses report the total wages paid to covered workers during the quarter (which is the tax base) and employment for each of the three months in the quarter.[2] Monthly employment is defined as the number of covered employees who worked or received pay during the pay period that includes the twelfth day of the month.

[2] Quarterly contribution report forms are specific to each state and are available from the state agencies responsible for administering UI programs; links to the websites for those agencies are available on the website of the U.S. Employment and Training Administration at https://ows.doleta.gov/unemploy/agencies.asp.

State agencies transmit the microdata to BLS, and QCEW estimates of employment and wages are released approximately seven months after the end of each quarter. Owing to the universal coverage of the QCEW data, QCEW estimates are produced at a fine level of detail by geography and industry. In cases where employers submit a quarterly report but do not provide employment data for all three months (a form of item nonresponse), the QCEW program imputes the employment data using historical QCEW data for the particular establishment. For the private sector in 2008, QCEW imputations represented about 5 percent of units and 3 percent of employment. The QCEW serves as the sampling frame for the CES and other BLS establishment-based surveys.

2.1.2. CES Data

The CES survey is a large-scale monthly survey of employment, hours, and earnings at nonagricultural establishments. In 2009, the set of respondents included about 140,000 businesses and government agencies nationwide that together covered approximately 410,000 individual worksites and about one-third of employment in the survey universe. Employment is defined in the CES as the number of employees who worked or received pay during the pay period that includes the twelfth day of the month–the same reference period as the QCEW.

CES data are collected using computer-assisted telephone interviews (CATI) and various self-reporting modes, including touchtone data entry, direct electronic file transmission, the Internet, mail, and fax.[3] With direct electronic transmission, large firms provide data on all of their establishments that are participating in the CES to a central collection facility. Firms that report their CES data in this way usually also report their QCEW data to the same facility, though not necessarily at the same time.

Approximately 40,000 new sample units are enrolled in the CES sample each year. New units generally report using CATI for a number of months and are then converted to one of the automated modes if they are willing to use them. The sample rotation plan allows most firms to report for four years and then be rotated out of the sample for a similar period. The sample design is a stratified, simple random sample of worksites, clustered by UI account number. The sample strata are based on state, industry, and employment-size class. The sampling frame and the CES sample are updated annually with current data from the QCEW.

CES estimates are released approximately three weeks after the reference period. This set of estimates is considered preliminary because some establishments do not respond by the primary deadline for data receipts. In order to incorporate additional sample received after this deadline, each sample-based estimate undergoes two monthly revisions before being finalized (Copeland and Valliant 2007).

2.1.3. CES Estimation

Once a year, the CES obtains a total employment figure (or "benchmark") as of March for each estimating cell (defined by industry) from universe employment counts derived mainly from the QCEW. The CES estimate for March is determined by this benchmark,

[3] CES report forms are available on the BLS website at http://www.bls.gov/ces/idcfcesforms.htm.

and employment estimates for subsequent months are computed using a ratio derived from CES respondents who provided data in both the current month and the previous month. The numerator of this ratio is the weighted employment in the current month for all such respondents in the estimating cell, and the denominator is the weighted employment in the previous month for these respondents. The employment estimate for a given month is computed by multiplying this ratio by the estimate for the prior month and then adding an estimate of residual net birth-death employment. Because the QCEW benchmark data are available only on a lagged basis, the initial published CES estimates for March of year t are computed from the QCEW benchmark figure for March of year $t - 1$ and the CES monthly growth-rate estimates and birth-death factors for each of the 12 months from April of year $t - 1$ through March of year t.

The birth-death factor helps account for the failure of the CES sample to capture employment changes due to business births and deaths (Mueller 2006). Although these changes are not captured, the death-related employment losses tend to offset the birth-related employment gains. As a result, what is needed is an estimate of the difference between birth-related employment gains and death-related employment losses. This factor is estimated using a time-series model based on historical QCEW data.

When new benchmark figures are determined, CES estimates for the previous 21 months are subject to revision. Benchmark revisions can have adverse effects on users of CES data. Many users have made business or policy decisions based on the initial estimates, and they must reconsider these decisions in light of the revised data. Benchmark revisions are typically on the order of 0.1 to 0.2 percent of total nonfarm employment, but revisions for two recent years were relatively large (0.6 percent for March 2006 and -0.7 percent for March 2009).

Reducing differences between QCEW and CES estimates is important because these differences contribute to benchmark revisions. The magnitude of benchmark revisions is determined largely by differences between CES and QCEW in over-the-year employment growth. BLS has considered converting benchmarking from an annual process to a quarterly process. Whether such a change would reduce the absolute value of total revisions depends in large part on differences in seasonal patterns between CES and QCEW (Battista et al. 2009). At issue is the difference between the CES and QCEW measures of employment change at monthly and quarterly frequencies.

2.1.4. Comparing QCEW and CES Estimates at the Aggregate Level

To give a sense of how QCEW and CES compare at the aggregate level, I compare the respective estimates of total employment using monthly data from March 2003 to March 2007. This period covers four "benchmark years" – 13-month periods that run from March of one year to March of the following year. At the beginning of each benchmark year, the QCEW and CES estimates are identical because CES estimates are benchmarked as of March. The CES estimates presented here are the final sample-based estimates for that month; they are not adjusted to reflect benchmarking in later years. Both sets of estimates refer to total private nonfarm employment, and they are not seasonally adjusted.

In the analysis in Section 2.2 and Section 2.3, I compare the QCEW and CES in terms of the level of employment at a point in time and in terms of the change in employment at monthly and quarterly frequencies. An important aspect of employment change over these

frequencies is the seasonal pattern of employment. As shown in Figure 1, the seasonality of employment is quantitatively important and consistent over time in both series. The seasonal pattern consists of a large increase during the second quarter, little change during the third quarter, an increase during the fourth quarter (more for QCEW than CES), and a large decrease from December to January. This seasonal pattern exists for other macroeconomic quantity variables (such as output), though the seasonal movements are smaller in employment than in output (Barsky and Miron 1989). Seasonal movements are thought to be caused by holiday spending, weather, summer vacations, and the opening and closing of schools (Miron 1996; Rydzewski et al. 1993).

Figure 2 plots the difference between the QCEW and CES estimates for each month. Over this period the difference between the estimates was as large as 1 million. In each 13-month period, the difference was largest in December. From September through January, there is a consistent pattern across the years in the difference between the estimates. The difference decreased from September to October, increased from October to December (as QCEW increased while CES was roughly constant on average), and decreased from December to January (as both QCEW and CES decreased but QCEW decreased by more).

Users of these data are typically more interested in the magnitude and direction of employment change over time than the level of employment at a particular point in time. Table 1 shows the average percentage change in employment from month to month. The period with the largest difference in growth rates between QCEW and CES is December to

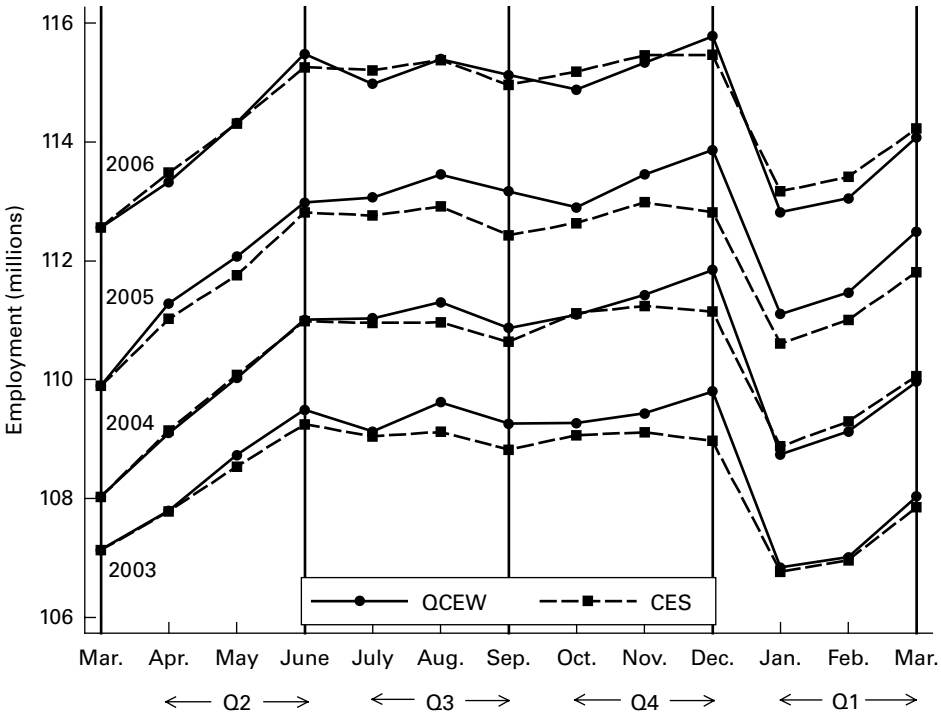

Fig. 1. QCEW and CES Employment Estimates, 2003–2006. Notes: Employment is total private nonfarm employment. The series start in March of each year from 2003 to 2006 (after the benchmark) and end a year later, in March 2004 to March 2007 (before the benchmark). Vertical lines identify the third month of each quarter.

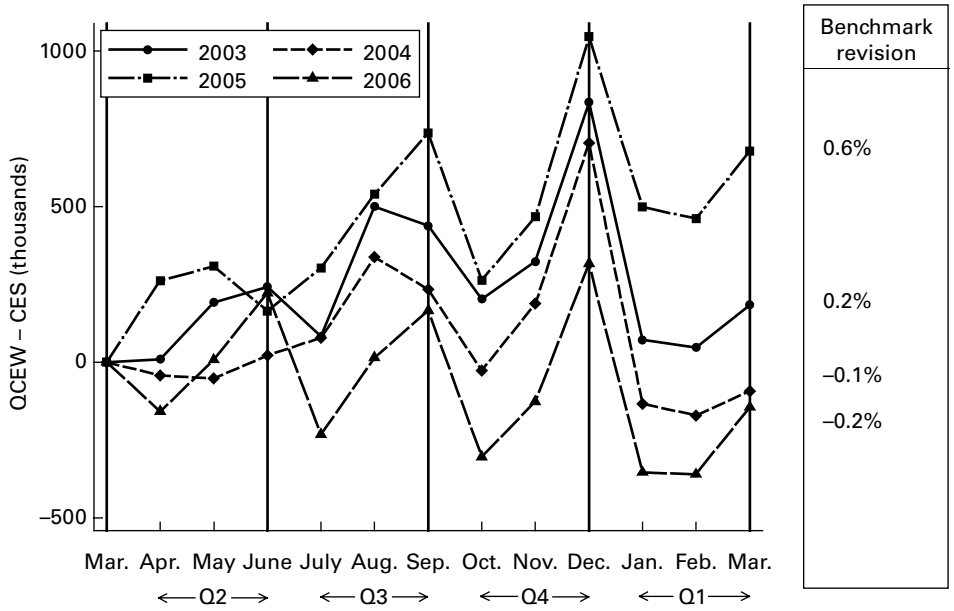

Fig. 2. Difference between QCEW and CES Employment Estimates, 2003–2006. Notes: Employment is total private nonfarm employment. The series labels correspond to the year at the beginning of a given 13-month period; for instance, the series labeled "2003" begins in March 2003 (after the benchmark) and ends in March 2004 (before the benchmark). Vertical lines identify the third month of each quarter.

January. QCEW employment fell by 2.62 percent on average, while CES employment fell by 2 percent. The series also show large differences from November to December, with QCEW employment rising by 0.37 percent and CES employment falling by 0.09 percent.

The end of the year also stands out when viewing the data on a quarterly basis (using the estimates for the third month of each quarter). QCEW estimates showed a greater buildup of employment at the end of the calendar year and a larger drop of employment moving into the following year. This pattern of differences in growth rates explains why switching from an annual benchmarking process to a quarterly one would increase the absolute value of total revisions (Battista et al. 2009). Over the first three quarterly periods (covering March–December) shown in Table 1, QCEW grew faster than CES. But in the other period (December–March), QCEW grew less than CES. With annual benchmarking these changes offset to some extent, but with quarterly benchmarking they would not.

2.2. Decomposition of Differences in Employment and Employment Growth

Much attention has been paid by researchers and practitioners to sources of error in surveys. The "total survey error" paradigm (Groves et al. 2009) identifies multiple sources of error in surveys: measurement error, processing error, coverage error, sampling error, nonresponse error, and adjustment error. Administrative data may also have some of these errors. Although administrative data do not typically involve sampling error, these data may contain measurement error, processing error, coverage error, nonresponse error, and adjustment error. Nonresponse error may exist in administrative data due to missing data.

Table 1. Changes in QCEW and CES employment estimates, 2003–2006

Period	Percentage Change		
	QCEW	CES	Diff.
Monthly			
Mar. to Apr.	0.89	0.87	0.02
Apr. to May	0.83	0.74	0.09
May to June	0.85	0.81	0.04
June to July	−0.17	−0.08	−0.09
July to Aug.	0.35	0.09	0.26
Aug. to Sep.	−0.30	−0.34	0.04
Sep. to Oct.	−0.06	0.26	−0.32
Oct. to Nov.	0.33	0.17	0.16
Nov. to Dec.	0.37	−0.09	0.46
Dec. to Jan.	−2.62	−2.00	−0.61
Jan. to Feb.	0.26	0.29	−0.02
Feb. to Mar.	0.89	0.74	0.15
Quarterly			
Mar. to Jun.	2.59	2.44	0.15
Jun. to Sep.	−0.12	−0.33	0.21
Sep. to Dec.	0.64	0.34	0.30
Dec. to Mar.	−1.49	−0.99	−0.50

Imputation aims to reduce nonresponse error, but imputation procedures may create adjustment error.

2.2.1. Sources of Error in QCEW and CES

Both the QCEW and CES are subject to measurement error. The two data sources have the same definition of employment: the number of employees who worked or received pay during the pay period that includes the twelfth day of the month. When reporting their data, employers may introduce errors by deviating from this definition. For instance, they may include in their employment counts workers who were on layoff or on unpaid leave for the entire pay period. Or they may include in their counts all employees who worked any time during the month, rather than limiting their counts to those who worked during the pay period that includes the twelfth.

These reasons apply equally, at least in principle, to the QCEW and CES. There is an additional reason to suspect measurement error in monthly employment data in the QCEW: data for all three months of a quarter are reported simultaneously. This raises the possibility of a seam effect: that month-to-month changes in employment are larger for the seam months (across quarters) than for adjacent months off the seam (within a quarter). Seam effects are common in panel surveys that interview respondents every three or four months but ask respondents to provide data for each month within the reference period (e.g., Kalton and Miller 1991; Rips et al. 2003). In part due to concern about seam effects, BLS publications from the database of longitudinal QCEW data use employment in the third month of a quarter but do not use employment for the first and second months (Pivetz et al. 2001).

Errors may also arise in the QCEW and CES due to nonresponse. The nonresponse error in the CES is typical of surveys: when some of the sampled units do not respond to the

survey, it increases the variance of the estimates. In addition, if nonrespondents are systematically different than respondents, nonresponse will contribute bias to the employment estimates. The average CES response rate by the final deadline for sample-based estimates from March 2003 to March 2008 was about 55 percent (Huff and Gershunskaya 2009). In the QCEW, item nonresponse occurs when employers submit a quarterly contribution report but do not provide employment counts for all three months. In these cases, the employment data are imputed, and the imputed values may differ from the true values. As a result, errors related to nonresponse in the QCEW are manifest as adjustment errors.

Sampling error and coverage error are present for the CES but not the QCEW. Sampling error exists because estimates are based on only part of the sampling frame; because the CES is a probability sample, sampling error contributes only to the variance of the estimates. Coverage error exists in the CES because its sample is drawn only once a year and therefore does not capture employment changes resulting from business births (openings) and deaths (closings) in real time. The estimation procedure used to adjust CES estimates for births and deaths reduces coverage error but does not eliminate it. The QCEW has no coverage error because it captures births and deaths within its normal reporting time frame.

2.2.2. Derivation of Decomposition

I now develop a decomposition of the difference between QCEW and CES estimates of employment. The decomposition is useful for identifying some sources of error in the CES and for identifying the reasons for QCEW-CES differences in employment and employment growth. As a result, the findings from the decomposition can guide program improvements and identify ways for BLS to reduce differences between QCEW and CES estimates. The decomposition builds on the approach used in Gershunskaya et al. (2002) and Huff and Gershunskaya (2009). Earlier work treated the administrative data from the QCEW as the truth, but my approach allows for errors in the administrative data. The key insight is that because the QCEW serves as the sampling frame for the CES, CES-type estimates can be constructed for the CES frame and sample using employment values from the QCEW. The method presented here can be applied to comparisons of survey and administrative data in other contexts where the administrative data provide the sampling frame for the survey and the data item of interest is collected in both sources.

Inputs for the decomposition are six research series of estimates of total private nonfarm employment (not seasonally adjusted) for each month from March 2003 to March 2007. The first series is the usual QCEW estimate, $\hat{E}^{QCEW}$. The other series are CES-type estimates, computed using CES sampling weights and the CES estimation procedure; the series vary in the sample used to compute the estimates, the employment values used for that sample, and whether the birth-death adjustment factors are used. In the notation used for these series, the superscript indicates the set of establishments used to compute the estimates ("r" for CES respondents, "s" for CES sample, or "f" for CES frame). The first subscript indicates the source of the employment values used to construct the estimate ("Q" for QCEW or "C" for CES). The second subscript indicates whether the birth-death adjustment factors are used ("y" for yes or "n" for no).

$\hat{E}^r_{C,y}$ is the usual CES estimate; it is based on CES respondents, uses CES employment values, and uses the birth-death adjustment. $\hat{E}^r_{Q,n}$ and $\hat{E}^r_{C,n}$ are also based on CES respondents but do not use the birth-death adjustment; $\hat{E}^r_{Q,n}$ uses QCEW employment values and $\hat{E}^r_{C,n}$ uses CES employment values. The final two series are based on either the CES sample ($\hat{E}^s_{Q,n}$) or the CES frame ($\hat{E}^f_{Q,n}$); these series use QCEW employment values but not the birth-death adjustment.

The difference between the QCEW and CES estimates for a given month can be decomposed into four components:

$$\hat{E}^{QCEW} - \hat{E}^r_{C,y} = \left[\left(\hat{E}^{QCEW} - \hat{E}^f_{Q,n}\right) - \left(\hat{E}^r_{C,y} - \hat{E}^r_{C,n}\right)\right] + \left(\hat{E}^f_{Q,n} - \hat{E}^s_{Q,n}\right) + \left(\hat{E}^s_{Q,n} - \hat{E}^r_{Q,n}\right) + \left(\hat{E}^r_{Q,n} - \hat{E}^r_{C,n}\right). \quad (1)$$

The term in brackets consists of two parts. The first part, $\hat{E}^{QCEW} - \hat{E}^f_{Q,n}$, reflects the coverage of the CES frame (which is fixed at the time the CES sample is drawn) relative to the QCEW universe (which is dynamic). The second part, $\hat{E}^r_{C,y} - \hat{E}^r_{C,n}$, reflects the cumulative birth-death adjustment (from the most recent benchmark to the current month) that is made to correct for the lack of coverage of births and deaths in the CES; this part can be taken as an estimate of the first part. The difference between these parts represents coverage error in the CES.

The second term on the right side of Equation (1), $\hat{E}^f_{Q,n} - \hat{E}^s_{Q,n}$, represents CES sampling error. The third term, $\hat{E}^s_{Q,n} - \hat{E}^r_{Q,n}$, represents CES nonresponse error. The fourth term, $\hat{E}^r_{Q,n} - \hat{E}^r_{C,n}$, reflects reporting differences between QCEW and CES – that is, differences between what the same respondents report in the survey and what they report in the administrative records. Conceptually, the term "reporting differences" refers to differences between the information about a single unit (a business in this case; a household or individual in other cases) available in survey and administrative sources.

In this case, reporting differences at the micro level can arise because of differences in the QCEW and CES definitions of employment (due to employment not covered by UI) or because establishments report different employment values to QCEW and CES even though the definitions are identical. Reporting differences can also reflect issues in linking CES and QCEW data at the establishment level. The linking was based on an exact match using state, UI account number, and reporting-unit number (an establishment identifier for account numbers with multiple establishments). The linking process is difficult because of mergers, acquisitions, and the opening of new establishments and because firms may combine employment information for multiple establishments in their reports for the CES and QCEW. It is likely that linking issues contribute primarily to reporting differences, but they may also contribute to the measures of CES coverage error and CES nonresponse error.

In order to simplify the notation, it is useful to rewrite Equation (1) as

$$Q_t - C_t = CV_t + SE_t + NR_t + RP_t, \quad (2)$$

where Q_t and C_t are the QCEW and CES estimates for month t, CV_t is coverage error, SE_t is sampling error, NR_t is nonresponse error, and RP_t is reporting differences between QCEW and CES. The shares of the overall difference between QCEW and CES estimates

of employment for a given month can be obtained by dividing each term in Equation (2) by $Q_t - C_t$.

Beyond the difference in QCEW and CES employment levels, also of interest is the difference between QCEW and CES measures of employment growth over time. The difference between monthly employment growth in the QCEW and the CES is

$$\Delta Q_t - \Delta C_t = (Q_t - Q_{t-1}) - (C_t - C_{t-1}) = (Q_t - C_t) - (Q_{t-1} - C_{t-1})$$
$$= \Delta CV_t + \Delta SE_t + \Delta NR_t + \Delta RP_t. \quad (3)$$

The shares of the difference in employment growth between QCEW and CES for a pair of consecutive months can be obtained by dividing each term in Equation (3) by $\Delta Q_t - \Delta C_t$.

For these decompositions, a given share may be positive or negative, but the four shares must sum to 100. For example, suppose that for a given month the overall difference ($Q_t - C_t$) is 10,000 and the components are $CV_t = 6{,}000$, $SE_t = 2{,}000$, $NR_t = -3{,}000$, and $RP_t = 5{,}000$. Then the respective shares are 60 percent for coverage error, 20 percent for sampling error, -30 percent for nonresponse error, and 50 percent for reporting differences. The shares for coverage error, sampling error, and reporting differences are positive because their components have the same sign as the overall difference; as such, these components are interpreted as explaining the overall difference. By contrast, the nonresponse share is negative because the nonresponse error has the opposite sign as the overall difference.

For the decomposition of the difference in employment levels, I first compute shares for each month during the period and then compute a weighted average of the shares across months, weighting by the absolute value of the difference in each month, $|Q_t - C_t|$. I follow a similar procedure for the decomposition of differences in employment growth at monthly and quarterly frequencies, except that the weight used in averaging the shares across periods is the absolute value of the difference in employment growth over each period, $|\Delta Q_t - \Delta C_t|$. In each case, the weight is the total amount to be explained.

2.3. Exploring Reporting Differences Between QCEW and CES

As shown and discussed in Section 3.1, the results of the decomposition reveal that reporting differences between QCEW and CES are responsible for a large share of both the difference in employment at a point in time and the difference in employment change from month to month. Therefore, the next stage of the analysis focuses on differences between QCEW and CES in the reported data on the level of employment and the change in employment. Specifically, I am interested in identifying factors that are related to such differences. I consider a variety of factors including (1) establishment characteristics such as pay frequency; (2) aspects related to how the data are collected and processed by BLS; and (3) the procedures used by businesses to compile employment data for the two programs. The analysis is based on two samples of matched QCEW-CES microdata.

2.3.1. Influence of Establishment Characteristics, Imputation, and Timing of Data Collection

The first dataset used to analyze reporting differences is a large sample of CES respondents matched to their QCEW data. The sample was constructed by taking all CES

respondents from January 2006 to March 2007 and attempting to match them to their QCEW data for that period. Of the 367,155 respondents in the private, nonfarm sector, an exact match (based on state, UI account number, and reporting-unit number) was obtained for 242,110 respondents. Data for some multi-establishment firms could not be matched because the establishment-level information was consolidated into an aggregate report for one or both programs and the level of aggregation was not the same in the QCEW and CES. However, the CES respondents in the matched and unmatched samples are broadly similar in terms of average employment, the distribution by employment-size class, and the distribution by industry.

The dataset for the matched sample provides information on employment at establishment i in month t according to the QCEW (Q_{it}) and the CES (C_{it}). The reported employment data are compared in terms of the absolute value of the difference in employment in month t ($|Q_{it} - C_{it}|$) and the absolute value of the difference in employment change from month t' to month t ($|\Delta Q_{it} - \Delta C_{it}|$, where $\Delta Q_{it} = Q_{it} - Q_{it'}$ and $\Delta C_{it} = C_{it} - C_{it'}$). Linear regressions are estimated in which the dependent variable is either (1) the difference in employment in a given month, (2) the difference in employment change over consecutive months, or (3) the difference in employment change over a quarter (using data for the third month of consecutive quarters).

The regressions comparing employment levels are estimated using data from March 2006 to February 2007, and those comparing employment change are estimated using data from March 2006 to March 2007. The sample used for each regression is constructed by pooling information across the multiple periods relevant for the dependent variable. The explanatory variables of interest are an indicator for whether QCEW data are imputed, pay frequency, the mode by which CES data are collected, and the timing of the CES report. The remaining explanatory variables are establishment size, industry, geographical division, and an indicator for whether the establishment is part of a multi-establishment firm.

The information on QCEW imputation and CES collection/timing are specific to the time period of the dependent variable. When the dependent variable refers to employment change over a monthly or quarterly period, the imputation variable for the period is an indicator for QCEW employment being imputed in either month t' or month t. The closing code (which identifies the timing of the CES report) for the period is defined as the greater of the codes in month t' and month t. If the collection mode is the same in the beginning and ending months, the collection mode for the period is defined as this common mode; otherwise the mode for the period is a residual category labeled "mixed."

2.3.2. Influence of Reporting Procedures

To examine the influence on QCEW-CES reporting differences of characteristics related to how establishments derive the employment counts for the QCEW and CES, I use data collected from a RAS that was conducted by phone in 2008. A sample of CES respondents was contacted to gather information about the methods and sources used to compile employment data for the two reports. Since the early 1980s, BLS has used the RAS method to investigate data quality in establishment surveys (Goldenberg et al. 1993). For a RAS, a respondent is contacted after survey completion and is asked a series of standardized questions on record-keeping practices, records availability and use, understanding of

survey instructions and definitions, discrepancies between survey definitions and answers, and other data-quality issues (Phipps et al. 1995).

The sample for this RAS was not designed to be representative of all CES respondents; as a result, the findings of any analysis for this sample do not necessarily generalize to the entire population. The RAS targeted respondents that exhibited specific types of reporting differences based on previously reported data for January 2006–March 2007. The types of differences are based on measures of employment change over various periods of time, as specified by the groups that are defined in Appendix Table A1. The proportion of the sample allocated to each group was based on the perceived importance of each group to overall QCEW-CES differences. Although an establishment could fall into more than one group, for sampling purposes each establishment was assigned to a single group based on an order of precedence. A small control group was created to represent establishments whose monthly QCEW and CES employment data were identical during this period. The analysis reported here is based on unweighted data.

The sample consisted of 3,002 actively reporting establishments of various sizes and industries. In order for an establishment to be eligible for sampling, its QCEW data must have been reported (not imputed) for all months from January 2006 to March 2007. A handful of industries were excluded from the sample frame because of scope differences between QCEW and CES or because of potential complications with collecting RAS data: educational services, government, hospitals, and professional employer organizations. In addition, establishments that used direct electronic transmission to report their CES data were excluded because they had been promised exclusive contact through the central collection facility.

The RAS questionnaire was divided into two sections with similar questions: one section focused on the monthly CES report, the other on the Quarterly Contribution Report (QCR) – the tax form that is the source of QCEW data. Initial contact was made with the CES respondent of record, who was then asked to complete the CES section. If the CES respondent was also familiar with the QCR, he or she was also asked the questions regarding the QCR. If another person at the establishment was responsible for the QCR, the interviewer contacted that individual and attempted to complete the QCR section. If an outside organization (such as an accounting firm or payroll processor) was responsible for the QCR, the interviewer did not contact that organization.

Among the sampled establishments, 63 percent answered one or both parts of the questionnaire, while 27 percent refused to answer either part and 10 percent were unable to be contacted. Nonresponse was relatively high for the QCR section, which limits the number of cases for which comparisons can be made between the procedures used to compile QCEW and CES data. A major factor in this nonresponse was that an outside organization was responsible for preparing the QCR figures. Among establishments that completed one or both parts of the questionnaire, 56 percent completed only the CES section, 44 percent completed both sections, and less than 1 percent completed only the QCR section.

I use regression analysis of the RAS data to relate differences in procedures to differences between the QCEW and CES data reported by establishments. I estimate a set of linear regression models of the form: $\Pr(D_i = 1) = \alpha + \beta R_i + \theta Z_i + \varepsilon_i$, where $D_i = 1$ if establishment i exhibits reporting differences between QCEW and CES (i.e., it was *not*

in the control group) and $D_i = 0$ if the establishment had no such differences (i.e., it was in the control group). R_i is a variable constructed from the RAS responses; in many cases, it is an indicator for whether an establishment follows a different procedure for QCEW than for CES. Estimates of β are interpreted as follows: if $\beta > 0$, then R contributes to reporting differences between QCEW and CES. Z_i is a vector of control variables, including establishment size and industry.

3. Results

3.1. Decomposition of Differences in Employment and Employment Growth

Table 2 contains the results of the decompositions described in Section 2.2. The most important sources of the difference in employment levels between QCEW and CES are coverage error (59 percent) and reporting differences (48 percent). Sampling error accounts for 10 percent of the difference and nonresponse accounts for − 17 percent. The importance of coverage error suggests that improvements to the CES birth-death procedures would reduce differences between CES and QCEW estimates of monthly employment.

Regarding the differences in monthly employment growth between QCEW and CES, 75 percent are due to reporting differences, 10 percent to nonresponse error, 10 percent to coverage error, and 4 percent to sampling error. On a quarterly basis, the share of differences in employment growth that is due to reporting differences (27 percent) is much lower than on a monthly basis. The largest shares of the differences in growth at a quarterly frequency are due to coverage error (41 percent) and nonresponse error (32 percent). That the portion due to reporting differences is lower at the quarterly frequency might arise from seam effects in the QCEW because the quarterly changes are constructed using data from only the third month of each quarter. In Section 3.2, I examine seam effects directly using matched QCEW-CES microdata.

Table 2. Decomposition of QCEW-CES differences, 2003–2006

Source	Levels	Monthly changes	Quarterly changes
Coverage error	59	10	41
	(34, 103)	(− 86, 56)	(− 1, 99)
	[0, 12]	[2, 10]	[1, 3]
Sampling error	10	4	0
	(− 41, 67)	(− 19, 114)	(− 10, 67)
	[4, 8]	[5, 7]	[2, 2]
Nonresponse error	− 17	10	32
	(− 213, 20)	(− 103, 79)	(− 213, 85)
	[10, 2]	[6, 6]	[1, 3]
Reporting differences	48	75	27
	(1, 164)	(− 2, 163)	(− 24, 164)
	[0, 12]	[1, 11]	[2, 2]

Notes: Numbers in the first row for each source are the overall average shares. Numbers in parentheses are the smallest and largest period-specific average shares. Numbers in brackets are the number of negative and positive period-specific average shares. Each period-specific average refers to a particular month or quarter and is based on four shares, one for each year from 2003 to 2006.

3.2. Influence of Establishment Characteristics, Imputation, and Timing of Data Collection

Given the importance of reporting differences between QCEW and CES in the differences in estimates of employment and employment growth, I now explore factors that are related to reporting differences. Coefficient estimates for the regressions on the first matched sample, as described in Section 2.3.1, are reported in Table 3. In general, the regression results indicate that factors related to employment differences at a point in time are also related to differences in employment change. In particular, imputation in the QCEW is strongly associated with both larger differences between QCEW and CES employment at a point in time and larger differences in employment change on a monthly or quarterly basis. For example, when QCEW data are imputed, the difference between QCEW and CES employment in a given month is 6.3 workers greater than when QCEW data are not imputed.

All else equal, establishments with weekly payrolls exhibit QCEW-CES differences that are larger than those for establishments with less-frequent payrolls (bi-weekly or semi-monthly). A potential explanation for this pattern is that establishments with weekly payrolls are more likely to use different reference periods for QCEW and CES. With a weekly payroll, an establishment has at least four pay periods in a given month. With a semi-monthly payroll, by contrast, there is only two pay periods in each month and therefore a greater likelihood that establishments will use the same reference period for QCEW and CES.

Differences between QCEW and CES are also associated with the timing of the receipt of CES data for a given reference month. A majority of the survey responses (about 75 percent for this sample) are received by the primary deadline for data receipts, which is the last Friday of the reference month and referred to as "first closing." Second closing is three weeks after first closing, and third closing (the final deadline for sample-based estimates) is three weeks after second closing. All else equal, establishments reporting by first closing have smaller differences (in both employment levels and employment change) than establishments reporting later. This pattern might reflect differences in the underlying characteristics of establishments that report before or after first closing, such as how organized their record-keeping is or the complexity of their operations.

The mode by which CES data are collected is predictive of QCEW-CES differences in employment and employment change. Compared with CATI, differences are somewhat smaller for touchtone data entry and much smaller for electronic file transmission. The estimated coefficients on the mode variables could represent causal effects of mode, but because modes are not randomly assigned these coefficients could also reflect the underlying characteristics of establishments that use each mode.

Another potential reason for reporting differences between QCEW and CES microdata on monthly employment is differences in the timing of data collection. While CES data are collected monthly, QCEW data are collected quarterly. To test for seam effects, I compute the absolute value of the percentage change in employment over each pair of consecutive months from January 2006 to January 2007. Then I average these changes across establishments separately for month pairs that are across quarters and those that are within a quarter.

Table 3. Reporting differences and establishment characteristics

Variable	Mean	Levels	Monthly changes	Quarterly changes
QCEW data imputed	0.08	6.32**	2.49**	4.80**
		(0.54)	(0.20)	(0.32)
Pay frequency				
Weekly	0.45	—	—	—
Bi-weekly	0.28	−3.36**	−2.19**	−2.83**
		(0.31)	(0.11)	(0.18)
Semi-monthly	0.24	−0.95**	−0.85**	−1.00**
		(0.28)	(0.10)	(0.17)
Monthly	0.03	1.26	−0.14	−0.41
		(1.19)	(0.44)	(0.60)
CES collection mode				
CATI[a]	0.22	—	—	—
Electronic file transmission	0.36	−4.45**	−1.81**	−2.78**
		(0.28)	(0.11)	(0.20)
Touchtone data entry	0.19	−0.75**	−0.46**	−0.87**
		(0.31)	(0.11)	(0.20)
Fax, mail, or Internet	0.21	0.07	0.27**	0.20
		(0.36)	(0.13)	(0.24)
Other or mixed	0.01	−1.21	0.31	1.06**
		(1.69)	(0.26)	(0.45)
CES closing code				
1 (By 1st closing)	0.75	—	—	—
2 (Between 1st & 2nd closing)	0.19	1.06**	0.72**	0.38**
		(0.21)	(0.08)	(0.13)
3 (Between 2nd & 3rd closing)	0.03	2.09**	1.48**	2.79**
		(0.43)	(0.17)	(0.40)
4 (After 3rd closing)	0.04	0.39	1.32**	1.53**
		(0.37)	(0.14)	(0.26)
R^2		0.09	0.05	0.07
N (observations)		2,362,616	2,297,084	726,539
N (establishments)		241,312	239,965	227,666
Mean of dependent variable		6.38	3.15	4.50

[a] Computer-assisted telephone interview.

* $p < .10$; ** $p < .05$.

Notes: Regressions also include controls for establishment size (7 categories), industry (13 categories), geographical division (9 categories), and an indicator for being part of a multi-establishment firm. The unit of observation is an establishment in a particular period; the standard errors (in parentheses) account for multiple observations per establishment. The reported means of the explanatory variables are based on the sample used for the "levels" regression.

Source: Merged QCEW and CES data for March 2006–March 2007.

These averages are reported in Table 4 separately by establishment-size class. For the QCEW, the variation in monthly employment is larger across quarters than within a quarter. For the CES, by contrast, the variation across quarters is about the same as the variation within a quarter. Examined another way, the variation in monthly employment across quarters is larger in the QCEW than in the CES. These patterns are all consistent with seam effects in the QCEW.

Table 4. Testing for seam effects in monthly employment data

	QCEW			CES			QCEW – CES	
Size class	Across	Within	Diff.	Across	Within	Diff.	Across	Within
1 to 9	12.92	9.06	3.87	9.19	8.98	0.21	3.73	0.07
10 to 19	10.22	7.32	2.90	7.54	7.38	0.16	2.68	−0.06
20 to 49	8.67	6.52	2.15	7.10	6.84	0.26	1.57	−0.32
50 to 99	7.57	5.63	1.93	6.47	6.02	0.45	1.09	−0.39
100 to 249	5.91	4.42	1.49	5.20	4.92	0.28	0.71	−0.50
250 to 499	5.14	3.81	1.33	4.64	4.34	0.30	0.50	−0.53
500 or more	4.27	2.83	1.44	3.90	3.63	0.27	0.38	−0.80
Total	9.78	7.03	2.76	7.46	7.21	0.25	2.32	−0.19

Notes: The numbers in the table are averages of absolute percentage changes of employment in consecutive months. The absolute percentage change is defined as $|x_1 - x_2|/\bar{x}$, where x_1 is employment in the first month, x_2 is employment in the second month, and $\bar{x} = (x_1 + x_2)/2$. Due to rounding, the differences shown in the table may not exactly equal the difference between the corresponding values. Size class is based on average monthly employment at the establishment between January 2006 and January 2007.
Source: Merged QCEW and CES data for January 2006–January 2007.

3.3. Influence of Reporting Procedures

As described in Section 2.3.2, the RAS data provide the opportunity to examine the procedures that establishments use to compile QCEW and CES data. Some of the procedures used for the two sources are compared in Table 5. The sample of respondents used for a given comparison is establishments that provided information on the procedures used for both programs. Establishments were more likely to use an incorrect reference period for QCEW (48 percent) than for CES (15 percent). The correct reference period is the pay period of the twelfth day of the month. Incorrect reference periods recorded in the RAS interviews include the entire month (the most common response) and a pay period other than the one including the twelfth. Respondents were asked if their employment counts represented a count of checks issued rather than the number of persons receiving pay; roughly the same proportion of respondents (10 to 12 percent) used a count of checks for deriving QCEW and CES employment.

RAS interviewers presented respondents with an extensive list of 20 employee types, asking if the establishment had each type and, if so, whether that type was included in the CES or QCR employment figures. The list included such types as employees working in locations outside the state (should be excluded), trainees (included), employees on leave without pay (excluded), and employees on layoff or strike (excluded if they are away for the entire pay period). According to the responses, nearly half of respondents reported incorrectly including/excluding at least one type in their employment counts for both QCEW and CES.

The final set of variables in Table 5 relates to the data sources that an establishment uses to complete the CES and QCR reports, who prepared the reports, whether there were any changes to the data sources, and record clean-up procedures such as purging of employee records. In a relatively small share (8 percent) of these establishments, a different data source (such as payroll, memory, or a count of time cards) was used for the two reports.

Table 5. Procedures used to compile QCEW and CES data

Procedure/characteristic	Percent	N
Number of payrolls		1,835
Single	85.9	
Multiple	14.1	
Pay frequency (if single payroll)		1,577
Weekly	39.6	
Bi-weekly	38.9	
Semi-monthly	12.2	
Monthly	1.7	
Missing or other	7.6	
Reference period		691
Incorrect for QCEW	47.8	
Incorrect for CES	15.3	
Different for QCEW & CES	42.3	
Check counting		827
Used for QCEW	10.0	
Used for CES	12.1	
Different for QCEW & CES	10.8	
Treatment of employee types		711
Incorrect for QCEW	41.8	
Incorrect for CES	45.9	
Different for QCEW & CES	17.9	
People and data sources		
Different data source for QCEW & CES	8.2	803
Different people prepare QCEW & CES	58.5	1,819
Change in QCEW data source	3.6	882
Change in CES data source	4.3	1,777
Purging of employee records	34.5	1,505
QCR data source doesn't have monthly counts	18.0	750

Source: 2008 CES-QCEW Response Analysis Survey.

By contrast, in over half (59 percent) of these establishments, different people prepared the two reports. Taken together, the summary statistics in Table 5 support the view that employment data from both QCEW and CES involve measurement error, but the types and magnitude of errors are different in each source.

Table 6 reports the results of the regression analysis that relates differences in procedures to differences between the QCEW and CES data reported by establishments. The number and timing of payrolls are both associated with reporting differences. Establishments with multiple payrolls are more likely than those with single payrolls to exhibit reporting differences. Among establishments with a single payroll, those with monthly payrolls are less likely to exhibit reporting differences than those with weekly payrolls. This pattern is related to differences in the reference period used for QCEW and CES. Establishments with a monthly payroll have only one pay period to choose from and therefore use the correct pay period for both reports. By contrast, 43 percent of establishments with a weekly payroll use a different pay period for constructing the QCEW and CES counts.

Table 6. Reporting differences and reporting procedures

Variable	Coeff.	N	Mean
Multiple payrolls	0.039**	1,835	0.141
	(0.013)		
Payroll frequency			
Weekly	—	1,577	0.396
Bi-weekly	0.012	1,577	0.389
	(0.013)		
Semi-monthly	0.011	1,577	0.122
	(0.018)		
Monthly	−0.087**	1,577	0.017
	(0.041)		
Missing or other	0.022	1,577	0.076
	(0.021)		
Reference periods different	0.088**	691	0.423
	(0.019)		
Check counting different	0.050*	827	0.108
	(0.026)		
Employee types treated differently	0.044*	711	0.179
	(0.024)		
Different data source used for QCEW and CES	0.045	803	0.082
	(0.030)		
QCR data source doesn't have monthly counts	−0.002	750	0.180
	(0.022)		
Different people prepare QCEW and CES	0.032**	1,819	0.585
	(0.009)		
Change in CES data source	0.027	1,727	0.043
	(0.023)		
Change in QCEW data source	0.023	882	0.036
	(0.041)		
Purging of employee records	0.018	1,505	0.345
	(0.012)		

* $p < .10$; ** $p < .05$.

Notes: Each row comes from a separate regression, except for payroll frequency (for which the set of rows covers a single regression). The dependent variable (mean = 0.961) in each regression equals 1 if the establishment exhibits reporting differences between QCEW and CES and equals 0 otherwise. Regressions also include establishment size (6 categories), industry (7 categories), and an indicator for being part of a multi-establishment firm. Sample for pay-frequency regression is establishments with a single payroll. Standard errors are in parentheses.

Source: 2008 CES-QCEW Response Analysis Survey.

The regression results demonstrate that differences in the reference period used for QCEW and CES are directly related to reporting differences. Establishments using a different reference period are more likely than establishments using the same reference period to exhibit reporting differences. In addition, establishments that treated employee types incorrectly for one program but not the other are more likely to exhibit reporting differences than establishments following the same procedure for both programs. For reference period, employee types, check counting, and the people who prepare the two reports, the qualitative pattern is similar: establishments that use different procedures for QCEW and CES have larger reporting differences.

4. Discussion

4.1. Implications for Differences Between QCEW and CES

The empirical results in the previous section suggest that micro-level reporting differences between QCEW and CES are created by the interaction of the variation over time in actual (or "true") employment and differences in the procedures used in collecting QCEW and CES data. That relationship is captured by the following equation:

$$\boxed{\text{Differences in reported employment}} = \boxed{\text{Differences in procedures}} \times \boxed{\text{Variation in actual employment}} \tag{4}$$

In this equation, both factors on the right side have a positive relationship with differences in reported employment. However, if one factor is zero, the other factor has no effect on differences in reported employment. For example, differences in the reference period used will not create differences between the QCEW and CES measures of employment during periods in which actual employment is stable. Analogously, variation over time in actual employment will not create differences between the QCEW and CES measures of employment when establishments follow the same procedures for both programs.

This framework can explain why QCEW-CES differences are noticeably larger for December than other months. As suggested by my analysis of the RAS data, QCEW data are more likely than CES data to be based on a count of employees who worked anytime during the month, rather than the correct reference period of the twelfth of the month. By itself, this difference in reporting procedures could contribute to differences between QCEW and CES data in any month. But during months in which employment is changing rapidly, the difference in reference periods should translate into a very large difference in the reported data. This is precisely the case for December due to the expansion of employment during the month for the holiday season.

This analysis points to several ways that BLS can reduce the magnitude of differences in employment between QCEW and CES. Although these differences may be caused by several factors (namely, CES coverage error, CES sampling error, CES nonresponse, and reporting differences), it makes sense to focus on reporting differences given that these differences are responsible for a large share of the difference between QCEW and CES estimates of both the level of employment and the monthly change in employment. A way to address reporting differences while promoting correct measurement would be to highlight the definition of employment and to emphasize that in most cases the QCEW and CES definitions are identical. The CES program already does this to some extent within the normal framework of respondent contact. CES interviewers mention the QCEW to some respondents during the CES initiation process and discuss definitions with respondents if their first month of CES data differs greatly from their QCEW data. These efforts could be expanded by discussing the QCEW with more respondents during CES initiation or by reconciling CES and QCEW microdata more frequently. Definitions could be highlighted by redesigning the forms and instructions used to collect CES data.

It is not feasible or desirable to discuss reporting differences directly with QCEW respondents on an individual basis, because only a small percentage of the 9 million

establishments in the QCEW will ever be solicited for CES and because the QCEW source document is a tax form rather than a survey instrument. Still, BLS periodically reviews state QCR forms and works with states to standardize and improve the wording of key concepts on the forms. In addition, BLS stays in contact with payroll-processing and payroll-software firms to make sure they understand the employment definition on the QCR form and understand that in most cases it is the same as the definition used for CES; these efforts cover QCEW reporting for a very large number of establishments.

Beyond actions targeted to the particular employment values reported by establishments, BLS may be able to reduce differences between the QCEW and CES employment estimates by reducing the amount of imputation in the QCEW or improving the accuracy of the imputation that is done. Other promising steps include improving response rates in the CES and improving the CES birth-death procedure; BLS already has efforts in place in those areas and has done much work on them in the past (e.g., Cohen et al. 2006; Mueller 2006; Rosen et al. 2002).

4.2. Guidance for Comparisons of Administrative and Survey Data in Other Contexts

This article explores the problem of differences between administrative and survey data using as an example the monthly employment figures gathered from U.S. business establishments. A large share of the difference between administrative and survey estimates of employment and employment change is due to reporting differences – when survey respondents report different employment values to administrative and survey sources. One factor related to reporting differences is imputation in the administrative data that arises from the failure of some establishments to provide monthly employment data. Another factor contributing to reporting differences is the timing of data collection: the survey data are collected monthly, while the administrative data are collected quarterly. This produces a seam effect in the administrative data: the variation in monthly employment is larger across quarters than within a quarter.

Reporting differences are also related to differences in the reporting procedures: establishments that follow different procedures for reporting their data to administrative and survey sources have larger reporting differences. In other contexts, reporting procedures are likely to differ between survey and administrative data because these types of data are usually collected under different conditions. For example, survey data on individual earnings are collected during telephone or in-person interviews, whereas administrative data are collected from tax authorities based on reports from businesses and individuals (e.g., Abowd and Stinson 2011; Johnson and Moore 2008). Differences in reporting procedures can take place on a variety of dimensions, including collection mode, the timing of collection relative to the reference period, and the identity of respondents.

The analysis in this article suggests that data collected from survey and administrative sources may involve errors, but the types and magnitude of errors are different in each source. It is well accepted that survey data are subject to measurement error and other types of error (Groves et al. 2009), but the notion that administrative data may contain errors is less widely recognized. To be sure, in many contexts the quality of administrative data may exceed the quality of survey data. But analysts and national statistical agencies

should at least consider the possibility that administrative data may have errors, especially measurement error and adjustment error (related to imputation and other edits).

With the increased use of administrative data by national statistical agencies, an ongoing challenge for researchers is to extend "total survey error" concepts to administrative data. This article demonstrates that administrative data may contain measurement error and adjustment error; in addition, administrative data may be subject to processing error, coverage error, nonresponse error, and types of errors (such as matching errors) that are typically not present in survey data. Some work has already been done in this area (e.g., Bakker 2010; Davern, Roemer, et al. 2009), but more work is needed to conceptualize error components in administrative data and standardize terminology across contexts.

In evaluating the quality of administrative data, it is useful to ask whether the administrative data measure the construct that is appropriate for the statistic that is being considered and how it compares to the construct in survey data (e.g., Davern, Klerman, et al. 2009; Davies and Fisher 2008). If it is the appropriate construct, to what extent is the construct adhered to by respondents? The present analysis has documented that although the tax form that is the basis for the administrative data involves the same reference period as the corresponding survey instrument, in some cases businesses do not use this reference period when constructing their monthly employment counts for the tax form.

It is often difficult for a national statistical agency to completely assess the quality of administrative data because these data are often produced outside the agency, the agency may not have access to the microdata, and the agency may not be able to influence the collection procedures (Eurostat 2003). This article illustrates several ways to assess the quality of administrative data. One way is internal consistency checks that explore patterns in the administrative data. The test for seam effects takes this approach by comparing the magnitude of monthly employment changes within a quarter and across quarters.

Another way to assess the quality of administrative data is to match it at the micro level to existing survey data and compare the values of particular variables. Such a comparison will not directly identify errors in the administrative data because survey data may also have errors, but it can be a useful guide. A third way is to conduct a RAS, in which reporting units in the administrative data are asked questions about reporting procedures. Although such surveys are usually used to assess quality of survey data, this article demonstrates the value of a RAS for assessing the quality of administrative data. A RAS may be done for the administrative data alone or in combination with related survey data, as was done in this case.

When a statistical agency discovers measurement errors in administrative data, what can it do to reduce the extent of such errors? The agency could make changes to the procedures it uses to create statistics from the underlying microdata, such as the procedures related to imputation and data editing. But what if the errors relate to the construct or to the reporting procedures? In that case, the statistical agency must attempt to work with the administrative agency to make changes to the construct or reporting procedures in order to reduce measurement error while maintaining the validity of the administrative data for the administrative agency.

Appendix

Table A1. Group definitions for the response analysis survey

Group	Definition	Sampling percent
Dec.-Jan.	Over-the-month change from December 2006 to January 2007 is different in QCEW and CES	40
June-Dec.-Jan.	Over-the-year buildup (from June to September to December 2006) is larger in QCEW than CES, and the drop from December 2006 to January 2007 is larger in QCEW than CES	4
March-March	Over-the-year change from March 2006 to March 2007 is different in QCEW and CES	23
Oct.-Dec.	Over-the-quarter change from October to December 2006 is different in QCEW and CES	20
Control	QCEW and CES employment data were identical (or nearly identical) for all months between January 2006 and March 2007	3.33
QCEW constant within	QCEW data are constant for all 3 months within a quarter while CES data are not; must see this pattern in at least 3 of the 5 quarters.	4.84[a]
QCEW constant across	QCEW data are constant for 2 months across quarters while CES data are not; must see this pattern in at least 3 of the 4 cross-quarter periods.	4.84[a]
QCEW stair-step	QCEW data exhibit stair-step pattern while CES data do not; must see this pattern in at least 3 of the 4 cross-quarter periods. A stair-step pattern is an increase (decrease) in employment over the months of a quarter followed by a decrease (increase) in the first month of the following quarter.	4.84[a]
CES constant within	CES data are constant for all 3 months within a quarter while QCEW data are not; must see this pattern in at least 3 of the 5 quarters.	4.84[b]
CES constant across	CES data are constant for 2 months across quarters while QCEW data are not; must see this pattern in at least 3 of the 4 cross-quarter periods.	4.84[b]

[a] The groups "QCEW constant within," "QCEW constant across," and "QCEW stair-step" were combined for sampling purposes, and the combined group was allocated 4.84 percent of the sample.

[b] The groups "CES constant within" and "CES constant across" were combined for sampling purposes, and the combined group was allocated 4.84 percent of the sample.

Notes: The definition of each of the first four groups involves computing a change (from month t' to month t) in CES employment ($\Delta C = C_t - C_{t'}$), the corresponding change in QCEW employment ($\Delta Q = Q_t - Q_{t'}$), and then computing the absolute value of the difference in these changes ($|\Delta Q - \Delta C|$). An establishment was then considered to be in the group if the difference exceeded a threshold that depended on employment-size class (based on the average of QCEW and CES employment in the base period). These thresholds were: 3 employees for size class 1–9, 7 employees for class 10–49, 10 employees for class 50–99, 15 employees for class 100–249, and 20 employees for class 250 or more.

Source: 2008 CES-QCEW Response Analysis Survey.

5. References

Abowd, J.M. and Stinson, M.H. (2011). Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Bureau Survey and SSA Administrative Data. U.S. Census Bureau Center for Economic Studies Paper No. CES-WP-11-20. Available at http://ssrn.com/abstract=1894690.

Bakker, B.F.M. (2010). Micro-Integration: State of the Art. Paper presented at Joint UNECE/Eurostat Expert Group Meeting on Register-Based Censuses, May 10–11, The Hague, The Netherlands.

Barsky, R.B. and Miron, J.A. (1989). The Seasonal Cycle and the Business Cycle. Journal of Political Economy, 97, 503–535.

Battista, V., Manning, C., and Robertson, K.W. (2009). Research on Quarterly Benchmarking for the Current Employment Statistics Survey. Proceedings of the American Statistical Association, Section on Government Statistics, 2866–2876.

Berka, C., Humer, S., Lenk, M., Moser, M., Rechta, H., and Schwerer, E. (2010). A Quality Framework for Statistics Based on Administrative Data Sources using the Example of the Austrian Census 2011. Austrian Journal of Statistics, 4, 299–308.

Bound, J. and Krueger, A.B. (1991). The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right? Journal of Labor Economics, 9, 1–24.

Cohen, S., McCarthy, D., Rosen, R., and Wiatrowski, W. (2006). Internet Collection at the Bureau of Labor Statistics: An Option to Report Data. Monthly Labor Review, 129(2), 47–57.

Copeland, K.R. and Valliant, R. (2007). Imputing for Late Reporting in the U.S. Current Employment Statistics Survey. Journal of Official Statistics, 23, 69–90.

Daas, P., Ossen, S., Vis-Visschers, R., and Arends-Toth, J. (2009). Checklist for the Quality Evaluation of Administrative Data Sources. Discussion Paper 09042, Statistics Netherlands.

Davern, M., Klerman, J.A., Baugh, D.K., Call, K.T., and Greenberg, G.D. (2009). An Examination of the Medicaid Undercount in the Current Population Survey: Preliminary Results from Record Linking. Health Services Research, 44, 965–987.

Davern, M., Roemer, M., and Thomas, W. (2009). Investing in a Data Quality Research Program for Administrative Data Linked to Survey Data for Policy Research Purposes is Essential. Paper presented at Research Conference of the Federal Committee on Statistical Methodology, November 2–4, Washington, DC.

Davies, P.S. and Fisher, T.L. (2008). Measurement Issues Associated with Using Survey Data Matched with Administrative Data from the Social Security Administration. Proceedings of the American Statistical Association, Section on Government Statistics, 147–157.

Eurostat. (2003). Quality Assessment of Administrative Data for Statistical Purposes. Paper presented at Meeting of Working Group "Assessment of Quality in Statistics," October 2–3, Luxembourg.

Gershunskaya, J., Eltinge, J.L., and Huff, L. (2002). Use of Auxiliary Information to Evaluate a Synthetic Estimator in the U.S. Current Employment Statistics Program. Proceedings of the American Statistical Association, Section on Survey Research Methods, 1149–1154.

Goldenberg, K.L., Butani, S., and Phipps, P.A. (1993). Response Analysis Surveys for Assessing Response Errors in Establishment Surveys. Proceedings of the International Conference on Establishment Surveys, American Statistical Association, 290–299.

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., and Tourangeau, R. (2009). Survey Methodology (2nd edn). Hoboken, NJ: John Wiley & Sons.

Huff, L. and Gershunskaya, J. (2009). Components of Error Analysis in the Current Employment Statistics Survey. Proceedings of the American Statistical Association, Section on Survey Research Methods, 2002–2009.

Johnson, B.W. and Moore, K. (2008). Differences in Income Estimates Derived from Survey and Tax Data. Proceedings of the American Statistical Association, Section on Survey Research Methods, 1495–1503.

Kalton, G. and Miller, M.E. (1991). The Seam Effect with Social Security Income in the Survey of Income and Program Participation. Journal of Official Statistics, 7, 235–245.

Kapteyn, A. and Ypma, J.Y. (2007). Measurement Error and Misclassification: A Comparison of Survey and Administrative Data. Journal of Labor Economics, 25, 513–551.

Kreuter, F., Muller, G., and Trappmann, M. (2010). Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data. Public Opinion Quarterly, 74, 880–906.

Laitila, T., Wallgren, A., and Wallgren, B. (2011). Quality Assessment of Administrative Data. Research and Development–Methodology Report 2011:2, Statistics Sweden. Available at http://www.scb.se/statistik/_publikationer/OV9999_2011A01_BR_X103BR1102.pdf.

Miron, J.A. (1996). The Economics of Seasonal Cycles. Cambridge, MA: MIT Press.

Mueller, K. (2006). Impact of Business Births and Deaths in the Payroll Survey. Monthly Labor Review, 129(5), 28–34.

Phipps, P.A., Butani, S.J., and Chun, Y.I. (1995). Research on Establishment-Survey Questionnaire Design. Journal of Business and Economic Statistics, 13, 337–346.

Pischke, J. (1995). Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study. Journal of Business and Economic Statistics, 13, 305–314.

Pivetz, T.R., Searson, M.A., and Spletzer, J.R. (2001). Measuring Job and Establishment Flows Using BLS Longitudinal Microdata. Monthly Labor Review, 124(4), 13–20.

Prevost, R. and Leggieri, C. (1999). Expansion of Administrative Record Uses at the Census Bureau: A Long-Range Research Plan. Paper presented at Research Conference of the Federal Committee on Statistical Methodology, November, Washington, DC.

Rips, L.J., Conrad, F.G., and Fricker, S.S. (2003). Straightening the Seam Effect in Panel Surveys. Public Opinion Quarterly, 67, 522–554.

Rosen, R., Hertwig, R., and Gomes, T. (2002). The Impact of Prompting on Response Rates: Experience with Touchtone Reporting in the CES Program. Proceedings of the American Statistical Association, Section on Survey Research Methods, 2959–2964.

Rydzewski, L.G., Deming, W.G., and Rones, P.L. (1993). Seasonal Employment Falls over Past Three Decades. Monthly Labor Review, 116(7), 3–14.

Wallgren, A. and Wallgren, B. (2007). Register-Based Statistics: Administrative Data for Statistical Purposes. Chichester, England: John Wiley & Sons.

Zhang, L.-C. (2011). A Unit-Error Theory for Register-Based Household Statistics. Journal of Official Statistics, 27, 415–432.

Received February 2011
Revised February 2012