REPORTING PATTERNS IN THE CURRENT EMPLOYMENT STATISTICS SURVEY

Kennon R. Copeland
U.S. Bureau of Labor Statistics, 2 Massachusetts Ave, N.E., Room 1950, Washington, DC 20212

KEY WORDS: Late Reports; Attrition; Nonresponse; Panel Survey; Establishment Survey; Estimate Revisions

I. Introduction

The Bureau of Labor Statistics' (BLS) Current Employment Statistics (CES) survey collects employment, hours, and earnings data monthly from a sample of over 300,000 U.S. establishments. To provide timely information, preliminary estimates are generated three to four weeks after the survey reference period. Final estimates are released two months later, incorporating late reports received after production of the preliminary estimates. Benchmark estimates are released annually with the data for May, based upon population employment data from BLS' ES-202 program for March of the prior year.

Nonresponse and late reporting reduce the effective sample size of a survey, thereby increasing variances for survey estimates, and potentially introduce bias into survey estimates, if respondents differ from nonrespondents relative to the variables of interest. Estimation methods are developed so as to account for nonresponse and late reporting and lessen their impact on bias and variance. These methods, however, assume the response mechanism is ignorable within defined estimation cells and, hence, do not distinguish among various patterns of nonresponse.

Within this paper, CES reporting patterns will be profiled relative to establishment characteristics, effects of late reporting and nonresponse on CES survey estimates will be explored, and implications of the findings for alternative nonresponse and late reporting adjustment models will be discussed.

II. CES Survey Design

The BLS recently completed a major redesign of the CES survey (Werking, 1997), moving the survey from its historical quota sample design to a probability sampling basis. The new sample design is a stratified, simple random sample of establishments from the BLS's Longitudinal Data Base (LDB), with strata defined by state, industry, and employment size. Sampling rates for each stratum are determined through optimum allocation.

Data must be reported within a two to three week period for inclusion in the initial published estimates (referred to as first closing estimates) for the month. As additional responses are received after this first closing of the collection period, the estimates for a given month are revised twice (referred to as second and third closing estimates) to incorporate data from late reporters. The first closing estimate of month-to-month change is derived by subtracting the second closing estimate for the prior month from the first closing estimate for the current month.

Estimates are generated through use of a weighted link-relative estimator, which uses a weighted sample trend within an estimation cell, based upon common reporters between the prior and current months, to move forward the prior month's estimate for that cell. The current CES estimator for total employment (Bureau of Labor Statistics, 2001) takes the following form for month t and closing $k\ (=1,2,3)$

$$\hat{Y}^{LR}_{t|k} = \sum_{c=1}^{C}\left[\frac{\sum_{i\in s_{ct,(t-1)|k}} w_i Y_{ti}}{\sum_{i\in s_{ct,(t-1)|k}} w_i Y_{(t-1)i}}\hat{Y}^{LR}_{c(t-1)|k}\right] = \sum_{c=1}^{C}\left[LR_{ct,(t-1)|k}\hat{Y}^{LR}_{c(t-1)|k}\right]$$

where

- $c\ (=1,\ldots,C)$ refers to estimation cell (defined by industry and, for selected industries, region)
- $s_{ct,(t-1)\}k}$ represents the set of sample establishments in estimation cell c that reported data for both months t and $t-1$, as of closing k
- w_i is the sampling weight for sample establishment i
- Y_{ti} is the total employment reported for month t by sample establishment i
- $\hat{Y}^{LR}_{c(t-1)|k}$ represents the prior month, $t-1$, link-relative estimate for estimation cell c based upon data reported as of closing k

The link-relative, $LR_{ct,(t-1)}$, is thus a type of growth rate estimate for the period $t-1$ to t. As can be seen from the estimator, differences in estimates between closings will be due solely to the inclusion of late responses, while differences between estimated and benchmark values will be due to the combined effects of sampling, nonresponse, late reporting (if comparing first or second closing estimates), and measurement error.

III. CES Reporting Patterns

Survey nonresponse is frequently classified on the basis of reason for nonresponse. Panel surveys add another dimension to the response mechanism, that being response status at different points in time. Little and David (1983) distinguished three types of panel survey nonresponse – attrition (sample unit stops reporting), late entry (sample unit does not report initially), and reentry (sample unit has a gap in reporting).

An alternative classification, appropriate for the CES survey, reflects the current month's reporting status, timing of

Any opinions expressed in this paper are those of the author and do not constitute policy of the Bureau of Labor Statistics

reporting for current month reporters, and prior reporting patterns for current month nonreporters. Following is a set of classifications for the CES survey to reflect current month reporting status (Figure 1 illustrates the nature of data reporting for these classifications, with month T classification determined following subsequent months of data collection):

Reporters

Early Reporters

First Closing Reporters – sample establishments reporting data for the month prior to the cutoff date for processing first closing estimates

Late Reporters

Second Closing Reporters – sample establishments reporting data for the month after the cutoff date for processing first closing estimates, but prior to the cutoff date for processing second closing estimates

Third Closing Reporters – sample establishments reporting data for the month after the cutoff date for processing second closing estimates, but prior to the cutoff date for processing third closing estimates

Nonreporters

Complete Nonreporters – sample establishments not reporting for any month

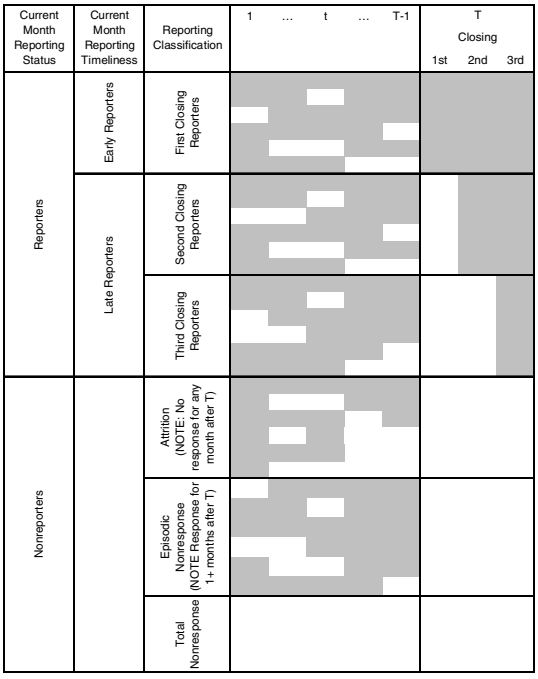
Attritors – sample establishments reporting data for at least one month, but which no longer report data

Episodic Nonreporters – sample establishments not reporting data for the month, but which do report for a subsequent month

Figure 1

CES Reporting Patterns - Illustration for Month T

(Shaded area represents data reported for month, by closing within month)



All three nonreporter types (complete nonreporters, attritors, and episodic nonreporters) impact the overall accuracy of the CES estimates, regardless of closing. Late reporters (second closing reporters, third closing reporters) affect the accuracy of preliminary estimates only. The extent of the impact late reporters have on the preliminary estimates can be assessed by examining the direction and magnitude of revisions between first and third closing estimates.

IV. CES Survey Nonresponse Profile

The focus of this profile is on the dynamic portion of the CES survey nonresponse—late reporting, attrition, and episodic nonreporting. Complete nonreporters, while contributing to the overall nonresponse impact, are less tractable in terms of a nonresponse adjustment strategy due to the lack of any reported data.

A. Late Reporting

A portion of the nonresponse present in first closing estimates is temporal. For a variety of reasons, some sample establishments are unable to respond within the narrow timeframe required for publication of first closing results, but do provide data for the survey month at a later point in time (Rosen, et al, 1991). Calendar effects appear to play a role in late reporting. For the CES survey, the time available for data collection depends upon the day of the week the 12th of the month falls on; the shorter the data collection period, the greater the likelihood for late reporting. In addition, as data are to be reported for the pay period containing the 12th day of the month, the length of a sample establishment's pay period could affect availability of the information to be reported

While the data for these late reporters are utilized in second and third closing estimates (depending upon when they report), any differences between their month-to-month trends and that assumed by the weighted link-relative estimator will drive the direction and magnitude of revisions to the first closing estimates.

A late reporting rate may be calculated as

$$LR\% = \frac{n_{R2} + n_{R3}}{n_{R1} + n_{R2} + n_{R3}} \times 100\%$$

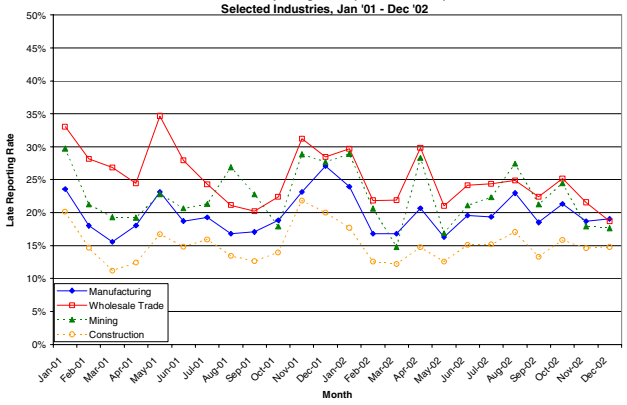
where n_{Ra} represents the number of sample establishments within the response status group corresponding to closing a

Graph 1 presents, for selected industries, late reporting rates for the period January 2001 through June 2002. Graph 2 provides similar information, weighted by employment.

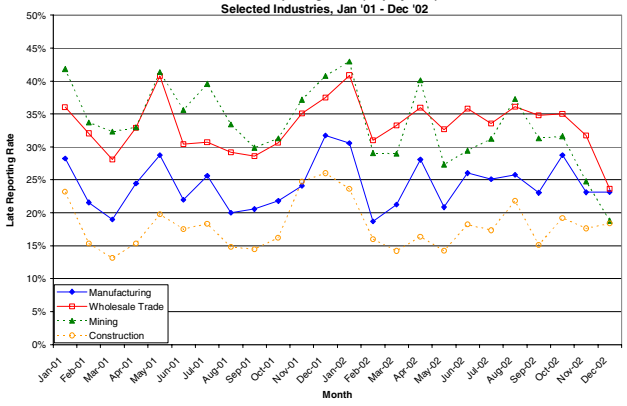
These graphs show late reporting rates range between 11%-35% of the sample establishments and between 13%-43% of the employment that will report by third closing. This percentage varies across both time and industry. The results also suggest late reporters tend to be larger establishments, as the late reporting rate weighted for employment is greater than

that for establishments (anywhere from 1 to 19 percentage points).

Graph 1
CES Late Reporting Rates (establishments)
Selected Industries, Jan '01 - Dec '02



Graph 2
CES Late Reporting Rates (employment)
Selected Industries, Jan '01 - Dec '02



As suspected, the variability across time appears to be partly driven by calendar effects, especially when the data reporting period is small. Months with the shortest data reporting period (9 or 10 days) accounted for eight of the ten largest late reporting rates, with Nov '01 (which contains a holiday during the data collection period) accounting for the other two largest late reporting rates. The converse was not the case, however, as months with the longest reporting period (14 or 15 days) only accounted for three of the ten smallest late reporting rates. The correlation between length of reporting period and late reporting rate ranged from -0.30 to -0.53 for establishments and from -0.33 to -0.71 for employment.

This finding suggests a fixed-effects linear model, based upon data reporting period (e.g., short, not short) along with some fixed effect for industry segment, could be used to describe the late reporting rate.

B. Attrition

A second portion of nonresponse in a panel survey is due to sample establishments that stop reporting as of some point in time. Rosen, et al (1993) classified attrition for the CES survey as: establishment went out of business; establishment overtly refused to continue participation; and establishment simply ceased reporting. Reasons for refusal and ceasing reporting include fatigue and, for establishment surveys, change in contact person within the establishment, with the result that a new decision is made relative to survey participation.

Data for attritors are not utilized in the weighted link-relative estimator, with the implicit assumption being that the growth rate from month $t-1$ to t is the same for attritors as for available reporters within estimation cell. To the extent this assumption fails to hold, the accuracy of the CES survey estimates will be adversely affected.

A cumulative attrition rate through month T may be calculated as

$$Att\%_{1,T} = \frac{\sum_{t=1}^{T} n_{Att,t}}{n_{Act,1}} \times 100\%$$

where

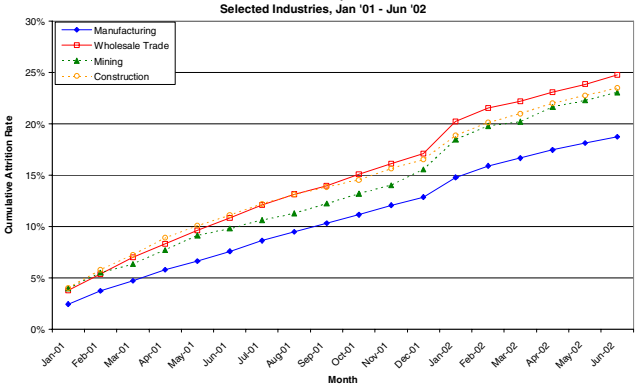
$n_{Att,t}$ is the number of sample establishments becoming attritors effective month t

$n_{Act,1}$ is the number of active sample establishments as of month 1

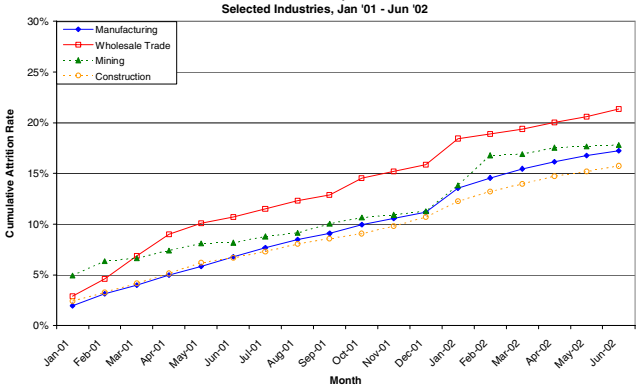
Graph 3 presents cumulative attrition rates by major industry segment for the period Jan 2001 through Jun 2002, relative to active sample establishments as of Dec 2000. Graph 4 provides similar information, based upon employment rather than establishments.

These graphs suggest cumulative attrition rates at the establishment level were slightly less for Manufacturing, while cumulative attrition rates weighted by employment tended to be slightly greater for Wholesale Trade. These data also provide an indication that Attritors tend to be smaller establishments, as the cumulative attrition rate is greater for establishments than for employment.

Graph 3
Cumulative Attrition Rate (establishments)
Relative to Active Sample units Dec '00
Selected Industries, Jan '01 - Jun '02



Graph 4
Cumulative Attrition Rate (employment)
Relative to Active Sample units Dec '00
Selected Industries, Jan '01 - Jun '02

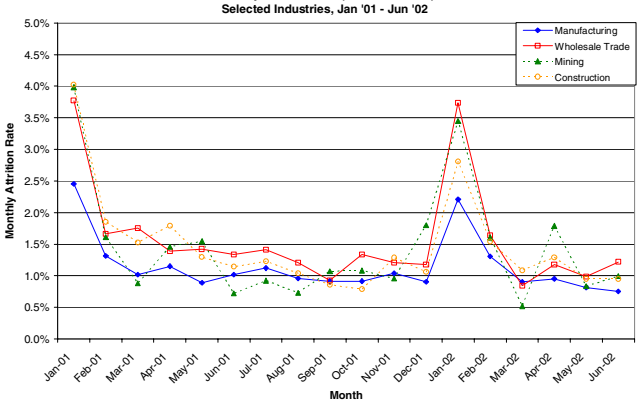


A monthly attrition rate for month T may be calculated as

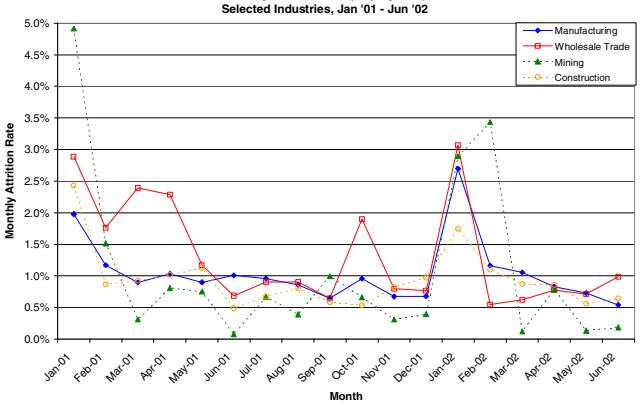
$$Att\%_T = \frac{n_{Att,T}}{n_{Act,1} - \sum_{t=1}^{T-1} n_{Att,t}} \times 100\%$$

Graphs 5 and 6 present monthly attrition rates for the period Jan 2001 through Jun 2002, based on establishments and employment, respectively.

Graph 5
Monthly Attrition Rate (establishments)
Selected Industries, Jan '01 - Jun '02



Graph 6
Monthly Attrition Rate (employment)
Selected Industries, Jan '01 - Jun '02



These graphs show attrition rates higher in January (2.2% - 4.0% for establishments and 1.7% - 4.9% for employment) than for the remaining months (0.5% - 1.9% for establishments and 0.1% - 3.4% for employment). Attrition rates are more variable for employment, especially for Mining.

The larger January attrition rate is likely due to the data collection process, in which establishments are mailed a calendar year log form in January. It is reasonable to assume some establishments opt out of continued reporting when they receive the new log form. There appears to be a carry-over of this attrition effect in February. The monthly attrition rate for establishments is generally higher than that for employment, and is likely due to operational efforts aimed at ensuring continued participation of larger establishments, so as to control the impact on survey estimates.

C. Episodic Nonreporting

Episodic nonreporting represents sample establishments that do not report for a given month, but do report for a subsequent month. Gaps could be due to a variety of factors, such as change in data reporters, and seasonal closings. Episodic nonreporting can only be distinguished from attrition post hoc. CES guidelines treat reporting gaps of six months as attrition.

Episodic nonresponse may be viewed relative to the total sample size, with a within-month episodic nonresponse rate calculated as

$$ENR\%_T = \frac{n_{ENR,T}}{n_{Act,T}} \times 100\%$$

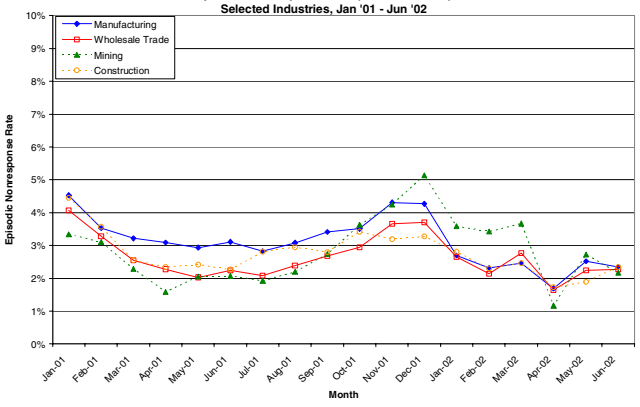
where

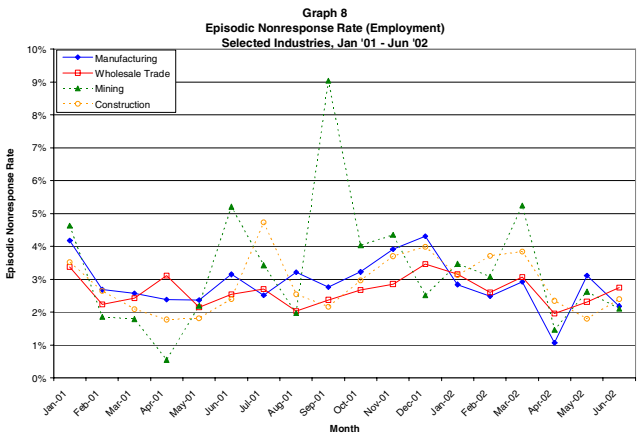
$n_{ENR,T}$ is the number of sample establishments that are episodic nonreporters in month T

$n_{Act,T}$ is the number of active sample establishments as of month T

Graphs 7 and 8 present monthly episodic nonresponse rates for the period Jan '01 – Jun '02, based on establishments and employment, respectively.

Graph 7
Episodic Nonresponse Rate (Establishments)
Selected Industries, Jan '01 - Jun '02





These results show episodic nonresponse rates ranging from 1.2% to 5.1% for establishments and, excluding Mining, from 1.1% - 4.7% for employment. Mining episodic nonresponse rates for employment were much more variable, ranging from 0.6% to 9.0%. Thus, for episodic nonresponse rates, there do not appear to be any differences due to employment size.

Table 1 presents the distribution of the maximum gap in nonreporting for episodic nonreporters in 2001.

Table 1
Nonreporting Gaps
Episodic Nonreporters in 2001

Longest gap for episodic nonreporters	Manufacturing	Wholesale Trade	Mining	Construction
1 month	42.7%	43.5%	49.3%	40.3%
2 months	21.2%	20.4%	19.7%	21.3%
3 months	17.3%	16.4%	13.5%	18.5%
4 months	11.0%	13.2%	8.8%	12.4%
5 months	5.9%	5.4%	6.2%	5.8%
6 months	1.9%	1.1%	2.6%	1.7%
7+ months	0.0%	0.0%	0.0%	0.0%

Between 40% and 49% of the episodic nonreporters experienced no more than a one month gap in nonreporting, while 18% - 20% experienced a gap of more than three months. Long gaps not leading to attrition may be a result of nonresponse conversion efforts undertaken for the CES survey.

Episodic nonreporting creates a carry-over effect in the use of a sample unit, due to the form of the CES estimator. A sample establishment that does not report for a given month will be left out of the calculation of the weighted link-relative not only for that month, but also for the succeeding month, as it will not be contained within the set of constant reporters.

D. Combined Nonresponse

The prior information about the components of nonresponse can be viewed as a whole across time. Such a picture can provide some insight into the nature of the problems faced in appropriately compensating for nonresponse.

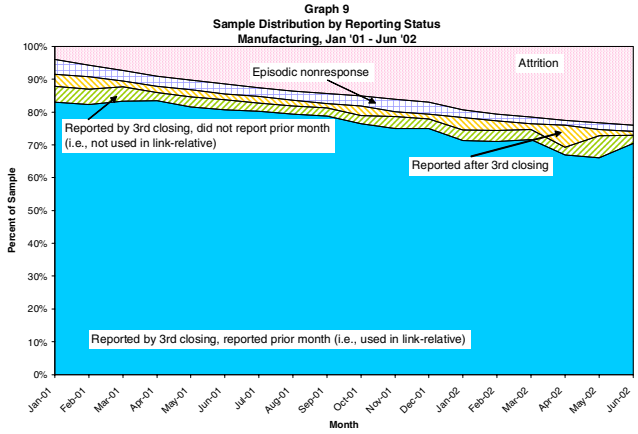
Table 2 presents information about the distribution of the reporting behavior in 2001 for the active sample as of December 2000.

Table 2
Reporting Behavior 2001
Active Sample Units as of Dec '00

	Manufacturing	Wholesale Trade	Mining	Construction
Respond all 12 months	74.5%	69.4%	70.3%	68.7%
Attritor during 12 months	11.0%	15.0%	13.6%	13.0%
Episodic NR during 12 months	14.5%	15.6%	16.1%	18.3%

Sample establishments are classified as reporting all 12 months (69% - 75%), becoming attritors from the sample during the year (11% - 15%), or being an episodic reporter in one or more month of the year (14% - 18%).

Although episodic reporting occurred for 14%+ of the sample in 2001, the impact on a monthly basis is somewhat less. Graph 9 shows the distribution of reporting status for Manufacturing from Jan '01 through Jun '02.



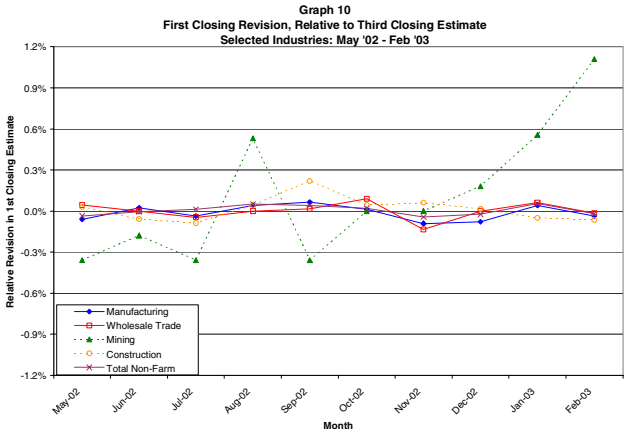
This graph shows episodic nonreporting accounted for less than 5% of the sample within a month. However, as stated earlier, episodic nonreporting also affects the usability of a subsequent month reporter, due to the need for two consecutive months of data for the weighted link-relative. As seen from the diagonally hatched portion of the bar, this carry-over affect resulted in an additional 2% - 7% of the sample being unusable for the weighted link-relative within a month. In addition, there are a percentage of the sample establishments (1% - 7%) that report too late for inclusion even in the third closing estimates.

V. Potential Impacts of CES Nonresponse

Nonresponse affects the accuracy of survey estimates through variance increase and potential for bias. Commonly, indirect indicators of the impact are used to assess the potential impact, as data for the nonrespondents are not known. However, the CES survey provides a more tangible indicator of the impact of nonresponse: revisions for late reporting.

A. Late Reporting Revisions

Graph 10 shows the relative difference between the first closing and third closing estimates for the period May 2001 through Feb 2002 for selected industries.



These results show revisions are less than 0.1% at the total non-farm level and generally less than 0.2% at the industry level; however, revisions for Mining were larger (as high as 1.1%).

The relationship between late reporting rates and first closing revisions was also examined. Table 3 presents the correlations between the late reporting rate and the size of the relative revision to the first closing estimates.

Table 3

Correlations between late reporting rates and first closing revision

	Correlation between late reporting rate and:	
Industry	**Relative revision**	**Absolute relative revision**
Manufacturing	0.50	-0.50
Wholesale Trade	0.21	0.05
Mining	0.47	0.35
Construction	-0.35	-0.34
Overall	0.09	0.18

The correlation of 0.18 found when combining all the observations across industry is in the direction expected. Results at the industry level, however, are mixed, suggesting there are other factors contributing to the magnitude of the first closing revision.

B. Benchmark Revisions

Table 4 provides information about the relative size of benchmark revisions, the revision between 3rd closing and ES-202, for Mar '01 and for the 10 year period ending in Mar '01.

Table 4

CES Historic Benchmark Revisions

	Benchmark Revision		
Industry	Mar '01	10 yr average	
		Actual	Absolute
Total	-0.1%	0.2%	0.3%
Manufacturing	0.1%	0.5%	0.6%
Wholesale Trade	-3.6%	-0.7%	1.2%
Mining	0.4%	0.7%	1.1%
Construction	-2.2%	0.1%	1.4%

These revisions are higher than the revisions seen between first and third closing. It must be noted, however, that these revisions incorporate the net effects of sampling, nonresponse, and measurement error, and thus cannot be used in isolation to assess the impact of nonresponse.

VI. Implications

This study provided a framework for viewing nonresponse and late reporting for the CES survey in terms of reporting patterns and timeliness. Such an approach provides a basis for considering alternative estimation approaches to compensate for nonresponse.

The findings of this study suggest further research is needed to provide an understanding of the relationship between reporting patterns and other establishment characteristics and data collection aspects. Such research should be carried out in an attempt to better explain the nature of nonreporting in the CES survey and suggest models that may prove useful in reducing the impact on nonresponse on the accuracy of the first closing estimates.

References

Bureau of Labor Statistics. (2001), Chapter 7, "Estimation," *Current Employment Statistics Manual*: U.S. Bureau of Labor Statistics, Washington, D.C.

Little, R.J.A. and David, M.H. (1983), "Weighting Adjustments for Non-response in Panel Surveys," Working paper: U.S. Bureau of the Census, Washington, D.C.

Rosen, R.J., Clayton, R.L., Rubino, T.R. (1991), "Controlling Nonresponse in an Establishment Survey," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 587-592.

Rosen, R.J., Clayton, R.L., Wolf, L.L. (1993), "Long Term Retention of Sample Members under Automated Self-Response Data Collection," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 748-752.

Werking, G.S. (1997), "Overview of the CES Redesign," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 512-516.