Nonresponse Adjustment in the Current Employment Statistics Survey[1]

Kennon R. Copeland October 2003
U.S. Bureau of Labor Statistics
2 Massachusetts Avenue, N.E. Washington, DC 20212 (Copeland.Kennon@bls.gov)

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, initial estimates are generated three to four weeks after the survey reference period. Final estimates, incorporating late reports received after production of the preliminary estimates, are released two months later. Benchmark estimates, incorporating administrative population data from the BLS' ES-202 program for March of the prior year, are released annually with the data for May.

Nonresponse potentially introduces bias into survey estimates, if respondents differ from nonrespondents relative to the variables of interest, and also reduces the effective sample size of a survey, thereby increasing variances for survey estimates. Estimation methods are developed so as to account for nonresponse and lessen its impact on bias and variance. These methods, however, assume nonresponse is ignorable within defined estimation cells and, hence, do not distinguish among various patterns of nonresponse. Nonresponse is used here to encompass both nonreporting and late reporting. Late reporting is temporal nonresponse, as the data become available at a subsequent point in time.

This paper continues research presented in Copeland (2003). Section II presents a brief overview of the CES survey design, Section III defines and profiles CES survey reporting patterns, Section IV examines effects of late reporting and nonresponse on CES survey estimates, and Section V discusses alternative nonresponse and late reporting adjustment models.

II. CES Survey Design

The BLS recently completed a major redesign of the CES survey (Werking, 1997; Bureau of Labor Statistics, 2003), 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, with strata defined by state, industry, and employment size. Sampling rates for each stratum are determined through optimum allocation to minimize variance of total employment level.

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 prior month's second closing estimate from the current month's first closing estimate.

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 estimator for total employment (Bureau of Labor Statistics, 2001) takes the following form for month t and closing $k (=1,2,3)$

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

where

$c (=1,...,C)$ 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, as of closing k, reported data for both months t and $t-1$

w_i is the sampling weight for sample establishment i

Y_{ti} is the total employment reported for month t by sample establishment i

---

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

$\hat{Y}_{c(t-1)(k+1)}$ represents the prior month, $t-1$, link-relative estimate for estimation cell c based upon data reported as of closing $k+1$ (with a maximum value of three) for month $t-1$ (which corresponds to closing k for month t)

The link-relative, $LR_{ct,(t-1)}$, is thus a growth rate estimate for the period $t-1$ to t. 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. An implicit assumption of the estimator is that, within an estimation cell, establishments not reporting data (nonsampled, nonreporting, late reporting) for both months t and $t-1$ are assumed to have the same growth rate as for those establishments reporting data, i.e., the sampling, response, and reporting timeliness mechanisms are ignorable (Rubin, 1976).

III. 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 by survey period. Surveys that publish revised estimates offer yet another dimension to the response mechanism, that being timeliness of reporting.

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). While this categorization describes general patterns of nonresponse, the types are not mutually exclusive. A sample unit that stops reporting (attrition) could have had gaps in reporting for time periods prior to attrition (reentry) and may not have reported initially (late entry).

Little and Su (1989) identified two patterns of panel survey nonresponse – monotone (the only type of nonresponse is attrition) and haphazard (nonresponse is either late entry or reentry or both). Although a fully monotone pattern of nonresponse is unlikely, the actual pattern may be approximately monotone (e.g., dropouts in clinical trials). Haphazard nonresponse encompasses a wide variety of nonresponse patterns (i.e., Little and David's late entry and reentry panel nonresponse types, as well as any combination of the three nonresponse types).

These two taxonomies could be refined to reflect more completely the nature of reporting patterns. The Little and David taxonomy does not address mixtures of patterns, while the haphazard pattern of Little and Su does not provide distinctions among haphazard patterns. Clarifying these distinctions could prove useful, as distributional properties could differ among patterns. In addition, complete nonresponse (sample unit never reports) should be added to the list of nonresponse types and complete response (sample unit always reports) should be added so all sample units are encompassed by the classification. As a result, it may be more appropriate to talk in terms of panel survey reporting patterns rather than nonresponse patterns.

There are five basic types of reporting patterns.

Basic Reporting Patterns

Complete Response – unit reports every time period

Complete Nonresponse – unit does not report for any time period

Attrition – unit stops reporting after a given time period

Late Entry – unit begins reporting after the first time period

Episodic Nonresponse – unit experiences a mixture of reporting and nonreporting

An expanded and refined set of reporting patterns for panel surveys can be defined by mixtures of the basic reporting patterns. This taxonomy for reporting patterns, along with illustrations, is provided in Attachment 1. Note that classification of a sample unit in terms of a reporting pattern is temporary, unless the survey has ended and there will be no further time periods for which data will be collected. Reporting patterns defined by only one basic pattern may be thought of as first order reporting patterns, while other reporting patterns (based upon a combination of basic patterns) may be thought of as interaction reporting patterns.

These reporting patterns are of interest in that distributional properties may differ among the patterns, thereby affecting the assumptions of the underlying estimation model.

Table 1 presents the CES survey reporting pattern distribution for four industries for the period Jan '01 - Jun '02, excluding Complete Nonresponse. Complete Response accounts for 47% - 57% of the sample. Other first order reporting patterns (Strict Attrition, Strict Late Entry, Strict Episodic Nonresponse) account for 34% - 41% of the sample. Attrition (classified based on observing reporting patterns through Dec '02) occurred for 12% to 16% of the sample, while some type of episodic nonresponse occurred for 24% - 30% of the sample. (Note: Late entry could not be distinguished from initiation of new sample units (which was carried out on a flow basis); thus, some establishments classified as late entry may actually belong to the next higher level. In addition, some establishments classified as attrition may have become out of business.)

Table 1
Reporting Pattern Distributions
Selected Industries, Jan '01 - Jun '02

	Manufacturing NAICS 31xx-33xx	Wholesale Trade NAICS 42xx-43xx	Mining NAICS 1133, 21xx	Construction NAICS 23xx
Complete Response	57.0%	49.8%	51.2%	47.4%
Strict Attrition	9.3%	11.8%	10.6%	9.4%
Strict Late Entry	7.0%	10.5%	10.2%	9.9%
Strict Late Entry Attrition	2.1%	3.5%	3.2%	3.5%
Attrition with Episodic Nonresponse	4.1%	4.6%	4.2%	5.4%
Late Entry with Episodic Nonresponse	2.0%	2.8%	2.0%	2.5%
Late Entry Attrition with Episodic Nonresponse	0.6%	0.6%	0.6%	0.4%
Strict Episodic Nonresponse	17.8%	16.5%	17.9%	21.5%

For complete nonresponse and attrition, reason for nonresponse (e.g., refusal, noncontact) could be of interest as it may provide an indication of the establishment's situation (noncontacts may indicate out of business). Reason for episodic nonresponse (e.g., on vacation, data unavailable) may also be of interest for understanding and addressing nonresponse.

For a survey such as the CES survey, in which estimates are revised several times, late reporting adds another dimension to reporting patterns, as illustrated in Attachment 2. Late reporting units are treated differently than on-time reporters, thus affecting the eligible sample for the first closing link-relative estimator. The key aspect of timeliness of reporting relative to the current form of the link-relative estimator is the fact that late reporting establishments that reported (whether on-time or late) the prior month are not utilized in the link-relative estimation for the current month. This serves as a complement to the key aspect of nonreporting relative to the current form of the link-relative estimator being the fact that on-time reporters for the current month are not utilized in the link-relative estimation if they did not report in the prior month.

For the CES survey, timeliness of reporting is an issue for most sample establishments, although not on a continual basis. Table 2 presents top-level distribution of frequency of first-closing reporting for establishments in the Complete Response reporting pattern, for the eighteen-month period Jan '01 – Jun '02. The proportion of establishments in the Complete Response reporting pattern that reported on-time every month ranged from 23% to 29% at the industry level, while the proportion of establishments that reported late every month ranged from 1% to 12%. Further classification of timeliness of reporting patterns could take into account shorter time frames (perhaps most recent six months), relationship with length of reporting period (e.g., only late when reporting period less than 11 days), and characteristics of the sample establishments (e.g., length of pay period).

Table 2
Timeliness of Reporting Pattern Distributions
Selected Industries, Jan '01 - Jun '02
Sample Reporting all Eighteen Months

	Manufacturing NAICS 31xx-33xx	Wholesale Trade NAICS 42xx-43xx	Mining NAICS 1133, 21xx	Construction NAICS 23xx
Every Month by First Closing	27.7%	22.7%	22.8%	29.1%
12 - 17 Months by First Closing	55.4%	53.6%	51.3%	60.3%
6 - 11 Months by First Closing	10.6%	8.3%	16.2%	8.1%
1 - 5 Months by First Closing	4.0%	3.1%	7.6%	2.0%
No Month by First Closing	2.2%	12.3%	2.1%	0.6%

Copeland (2003) provides additional tabulations of CES survey reporting patterns. The distributional information on reporting patterns and timeliness of reporting indicates: 1) when viewed across long time periods, incomplete data or late reported data occurs for a large proportion of sample establishments; 2) prior CES survey reported employment information is available for many nonreporting/late reporting sample establishments.

IV. Impact of Late Reporting and Nonresponse

The impact of nonresponse in the CES survey may be examined directly for late reporting, and indirectly for both late reporting and nonresponse. Comparisons of first and third closing estimates provide a direct indication of the impact of late reporting, as the only difference between the two estimates is the inclusion of late reporters into the sample. The relative difference between first and third closing estimates for month t,

$$\text{RelDiff}_{t(1,3)} = \frac{\hat{Y}_{t|1}^{LR} - \hat{Y}_{t|3}^{LR}}{\hat{Y}_{t|3}^{LR}} \times 100\%$$

and the difference between first and third closing estimates of the month-to-month change from month $t-1$ to month t,

$$\text{Diff}_{t(1,3)} = \left[\hat{Y}_{t|1}^{LR} - \hat{Y}_{(t-1)|2}^{LR}\right] - \left[\hat{Y}_{t|3}^{LR} - \hat{Y}_{(t-1)|3}^{LR}\right] = \Delta_{t,(t-1)|1}^{LR} - \Delta_{t,(t-1)|3}^{LR}$$

provide measures of the extent to which growth rates for late reporters differed from those for early reporters. Large differences provide an indication of nonignorability of late reporting.

Figure 1 presents relative differences between first and third closing non-seasonally adjusted estimates of monthly employment for the period May '01 – Feb '02 and May '02 – Feb '03 for the four industries for which probability sampling had been implemented as of May 2001. March and April were excluded from this graph due to the nature of CES survey processing, with annual benchmark data incorporated with the publication of first closing estimates for May (and thus second closing for April and third closing for March), and thus negating the ability to measure solely late reporting impact for these months. Although the larger industries have experienced fairly small revisions (absolute relative differences less than 0.25%), the revisions for Mining have been much greater, with the absolute relative difference as high as 1.1% in Feb '03.

Figure 1
First Closing Revision, Relative to Third Closing Estimate
Selected Industries: May '01 - Feb '02, May '02 - Feb '03
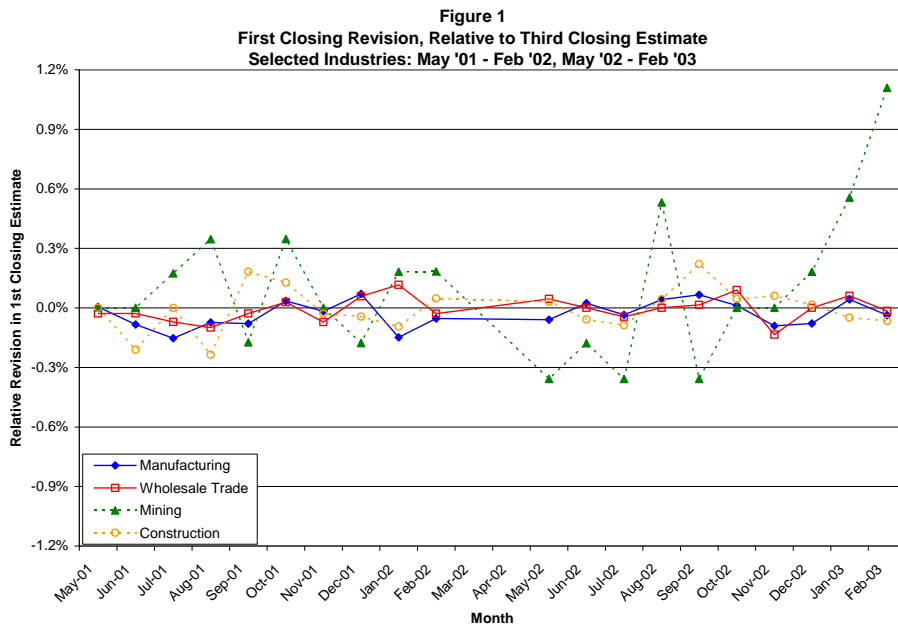
Revisions in the monthly employment estimates and in the estimates of month-to-month change in employment can also be compared with the month-to-month change in employment, which is a primary measure for assessing the employment data. Revisions that are large relative to the estimated change could serve to decrease the utility of the preliminary reports. Table 3 compares the magnitudes of the revisions in monthly and month-to-month change in employment to the first closing estimate of month-to-month change in employment for the period May '02 – Feb '03.

Table 3
First Closing Revisions Relative to First Closing Month-to-Month Employment Change
Selected Industries: May '02 - Feb '03
(Numbers in thousands)

	Manufacturing				Wholesale Trade				Mining				Construction			
	Employment (1st Closing)	Revision in 1st Closing Employment	Employment Change from Prior Month (1st Closing)	Revision in 1st Closing Employment Change	Employment (1st Closing)	Revision in 1st Closing Employment	Employment Change from Prior Month (1st Closing)	Revision in 1st Closing Employment Change	Employment (1st Closing)	Revision in 1st Closing Employment	Employment Change from Prior Month (1st Closing)	Revision in 1st Closing Employment Change	Employment (1st Closing)	Revision in 1st Closing Employment	Employment Change from Prior Month (1st Closing)	Revision in 1st Closing Employment Change
May-02	16,769	-10	24	-9	6,682	3	19	4	561	-2	4	-2	6,595	2	196	1
Jun-02	16,838	4	78	5	6,713	0	27	1	562	-1	3	-1	6,794	-4	200	-7
Jul-02	16,755	-6	-88	-5	6,716	-3	3	-3	561	-2	-1	-1	6,857	-6	61	0
Aug-02	16,784	7	30	12	6,698	0	-15	0	562	3	3	3	6,864	3	13	3
Sep-02	16,709	11	-70	-1	6,672	1	-27	2	561	-2	-4	-2	6,785	15	-78	11
Oct-02	16,643	2	-74	-1	6,667	6	-6	6	560	0	0	1	6,752	3	-50	5
Nov-02	16,575	-15	-73	-12	6,662	-9	-11	-9	554	0	-6	0	6,645	4	-111	5
Dec-02	16,487	-13	-72	-14	6,646	0	-7	0	550	1	-4	1	6,448	1	-196	-4
Jan-03	16,341	7	-136	10	6,585	4	-62	5	537	3	-14	3	6,128	-3	-323	-1
Feb-03	16,293	-6	-58	-3	6,584	-1	-4	-2	535	6	-3	4	6,065	-4	-66	2

Although revisions for several months are larger than the first closing estimate of month-to-month employment change, the changes in these situations are small. For months with larger employment changes, revisions are not of the magnitude of the change, but could nonetheless be viewed as substantial (five of eighteen first closing changes of at least 50,000 saw a revision in the first closing estimated employment level that was 10%+ of the magnitude of the first closing estimated change, while four saw a 10%+ revision in the magnitude of the change). Viewed from this perspective, late reporting could be considered to have an impact on the accuracy of the first closing estimates.

A direct indication of the appropriateness of the assumption of ignorability for late reporting can be made by comparing growth rates for late reporters with those for preliminary reporters. Table 4 provides comparisons of current month-to-month growth rates for current month first closing reporters and late reporters, for those sample establishments for which data were available for the prior month.

Table 4
Month-to-Month Growth Rates by Current Month Timeliness Status
Selected Industries, Jan '02 - Dec '02
Establishments with 50+ Employment

	Manufacturing		Wholesale Trade		Mining		Construction	
Month	1st Closing	2nd/3rd Closing	1st Closing	2nd/3rd Closing	1st Closing	2nd/3rd Closing	1st Closing	2nd/3rd Closing
Jan '02	-0.9%	-1.5%	-0.9%	-0.1%	-1.6%	-0.6%	-4.4%	-4.8%
Feb ' 02	-0.4%	-0.7%	-0.8%	-0.7%	-0.8%	-1.0%	-0.2%	-0.9%
Mar ' 02	-0.4%	-0.5%	-0.4%	-0.3%	-0.4%	0.0%	1.0%	1.7%
Apr ' 02	-0.2%	-0.3%	0.2%	0.2%	1.2%	2.8%	2.5%	1.5%
May ' 02	0.0%	-0.4%	0.0%	-0.3%	1.4%	1.0%	1.8%	0.5%
Jun '02	0.5%	0.3%	0.1%	0.8%	0.8%	-0.5%	2.1%	1.0%
Jul '02	-0.5%	-0.4%	-0.1%	-0.1%	-0.1%	-0.3%	0.9%	1.4%
Aug '02	0.1%	-0.3%	-0.3%	-0.3%	-0.1%	2.5%	0.5%	1.5%
Sep '02	-0.7%	-0.5%	0.2%	-0.6%	-0.9%	-0.8%	-1.0%	-0.1%
Oct '02	-0.7%	0.1%	-0.1%	-0.3%	-0.2%	0.0%	-0.7%	-0.4%
Nov '02	-0.4%	-0.6%	-0.4%	-0.2%	-1.3%	-0.6%	-2.1%	-2.3%
Dec '02	-0.2%	-0.3%	0.1%	-0.5%	0.6%	0.0%	-2.6%	-2.1%

Prior month-to-month growth rates for first closing and late reporters differ by more than one percentage point four times for Mining, three times for Construction, and no times for Manufacturing and Wholesale Trade. In one instance (for Mining) the differences are greater than two percentage points. Deriving a rough impact on the overall estimate (accounting for the relative sizes of the first closing and late reporting samples) showed the impact would generally be less than two-tenths of a percentage point, although in some cases the impact could be greater than seven-tenths of a percentage point.

Assessing the impact of nonresponse is more difficult, as reported survey data are not available. Although annual benchmark revisions provide an indication of overall error in the third closing estimates – the net result of sampling, nonresponse, and measurement errors – nonresponse error cannot be isolated from the benchmark results alone. One approach for assessing nonresponse effects would be to use ES-202 data at the establishment level as a proxy for CES survey data, and compare estimates derived including and excluding nonrespondents.

An indirect measure of the potential for a nonresponse impact on third closing estimates can also be obtained by comparing prior employment trends between reporters and nonreporters. The current estimator assumes nonresponse is ignorable within an estimation cell, and thus makes the implicit assumption of equality of month-to-month growth rates for respondents and nonrespondents within an estimation cell. Comparing immediately prior growth rates for current period reporters and nonreporters for which prior data are available provides an indirect indication of the appropriateness of the equivalence assumption. Table 5 provides comparisons of prior month-to-month growth rates for current month reporters and nonreporters, for those sample establishments for which data were available for the prior two months.

Table 5
Prior Month-to-Month Growth Rates by Current Month Reporting Status
Selected Industries, Jan '02 - Dec '02
Establishments with 50+ Employment

Month	Manufacturing		Wholesale Trade		Mining		Construction	
	Nonreporters	Reporters	Nonreporters	Reporters	Nonreporters	Reporters	Nonreporters	Reporters
Jan '02	-0.9%	-0.4%	-1.1%	-0.1%	-2.3%	-1.0%	-2.2%	-2.9%
Feb ' 02	-0.6%	-1.1%	-1.5%	-0.5%	-1.0%	-1.1%	-5.1%	-4.4%
Mar ' 02	-0.4%	-0.5%	0.1%	-0.8%	-3.3%	-0.7%	1.6%	-0.4%
Apr ' 02	0.2%	-0.5%	-0.2%	-0.5%	-0.3%	-0.5%	1.2%	1.0%
May ' 02	-0.2%	-0.2%	0.2%	0.2%	12.2%	0.8%	3.4%	2.2%
Jun '02	-0.2%	-0.1%	0.4%	-0.2%	0.8%	1.2%	0.3%	1.6%
Jul '02	0.4%	0.5%	0.6%	0.4%	0.8%	0.4%	2.2%	1.8%
Aug '02	-0.5%	-0.5%	0.3%	-0.1%	0.6%	-0.2%	2.2%	0.9%
Sep '02	-1.0%	-0.1%	-2.4%	-0.3%	-0.6%	0.9%	-0.5%	0.8%
Oct '02	-0.9%	-0.6%	-1.5%	0.0%	-1.4%	-0.8%	0.0%	-0.9%
Nov '02	0.2%	-0.5%	-1.3%	0.0%	-0.6%	-0.1%	-0.3%	-0.6%
Dec '02	-1.1%	-0.4%	-1.1%	-0.2%	-2.3%	-1.0%	-1.6%	-2.2%

Results show prior month-to-month growth rates for reporters and nonreporters differ by more than one percentage point five times each for Mining and Construction, three times for Wholesale Trade, an no times for Manufacturing. In three instances (twice for Mining and once for Manufacturing) the differences are greater than two percentage points. Deriving a rough impact on the overall estimate (accounting for the relative sizes of the reporting and nonreporting samples) showed the impact would generally be less than one-tenth of a percentage point, although in some cases the impact could be greater than eight-tenths of a percentage point.

Taken together, these results suggest the implicit assumption of equivalence between reporters and nonreporters and of first closing and late reporters may not hold, and that the lack of equivalence may impact both first and third closing estimates.

V. Alternative Models

The working model that applies to the CES survey estimator of total employment is a proportional regression model, in which the current month's value is assumed proportional to the prior month's value (West, et al, 1989), with the proportionality factor, ρ, assumed to vary by estimation cell, $c \, (= 1, \ldots, C)$ (which are collapsed sample design cells), and month

$$Y_{tci} = \rho_{tc} Y_{(t-1)ci} + k_{tci}$$

$$k_{tci} \sim iid \, N\left(0, \sigma_k^2 Y_{(t-1)ci}\right)$$

This proportional regression model has appeal for use in establishment surveys, where inference is often made about the rate of change for the population. In that sense, it can be thought of as a longitudinal analogue to the mean imputation model. However, the weakness of this model lies in its assumption of ignorable nonresponse (for both nonreporters and late reporters) within an estimation cell, and its failure to utilize reported data for establishments other than those reporting for both months $t-1$ and t.

Review of data from late reporters and nonreporters provided in this paper suggests the ignorability assumption may not always be met. Thus, some refinement to the model may provide opportunities for improvement. Prior research into occurrence of

nonresponse and late reporting (Copeland, 2003) indicates roughly 5% of current month first closing reporters are not used in the CES estimator due to lack of data for the prior month, and roughly 15% - 25% of prior month first and second closing reporters are not used in the CES estimator due to lack of first closing data for the current month. Thus, approaches incorporating this current information may provide opportunities for improvement.

One useful path for further research would appear to be refinement of estimation cell definitions, in an attempt to define a more appropriate ignorable nonresponse model, using a response propensity approach as described in David, et al (1983). Factors for consideration beyond industry, based upon research to date, would include size of establishment and geography. These factors are utilized in the sample design, but not (with a few exceptions) in the definition of estimation cells. If the estimation cell refinements were to involve such auxiliary data from the frame, the current model could be continued.

A second path would be to explore other ignorable nonresponse models. Auxiliary information known for sample establishments, such as length of pay period and prior employment changes may be correlated with reporting status and/or growth, and therefore be useful in a refined model. However, these auxiliary variables cannot be incorporated into the current model, as they are not known for the population. One approach would result in imputing for missing values due to late reporters and nonrespondents under a model defined for the sample establishments, with the completed sample data file being used to create link-relative estimates at the cell level as in the current approach. The difference would be that all sample establishments, not just those reporting current and prior months' data, would be used in the estimator. This approach would thus allow use of auxiliary data known only for the sample.

A simple model for deriving imputed values would be the current proportionality model, with the proportionality factor, ρ, assumed to vary by imputation cell $g\ (=1,\ldots,G)$, defined by a combination of design and sample auxiliary variables

$$Y_{tgi} = \rho_{tg} Y_{(t-1)gi} + k_{tgi}$$
$$k_{tgi} \sim iid\ N\left(0, \sigma_k^2 Y_{(t-1)gi}\right)$$

The proportionality factors could be estimated from the reporting sample, then used to impute values (using either a deterministic or a stochastic approach) for late reporters and nonreporters. The completed data set (reported data for preliminary reporters, imputed values for late reporters and nonrespondents) could then be used in the current link-relative estimator to derive employment estimates.

A third path would be to address potential nonignorability of nonresponse. The impact assessment of CES nonreporting and late reporting suggests refinement of the model to allow proportionality factors to vary by reporting pattern within estimation cell. If the assumption of ignorable nonresponse is not used, the question becomes how to estimate the month-to-month employment growth rates (proportionality factors) for nonreporters and late reporters. In addition, if differing proportionality factors are assumed for nonreporters and late reporters, it is necessary to predict whether a nonrespondent at first closing will report late or will be a nonreporter.

Little (1993) proposed the use of pattern-mixture models for handling incomplete multivariate data, such as that arising from a panel survey. Under this approach, models are defined for reporters and nonreporters, with identifying restrictions established to allow for estimation of parameters that are non-estimable under the base models. For the CES survey, a simple pattern-mixture approach would be to extend the proportional regression model to allow different growth rates by reporting pattern in the current and prior months. Before specifying the model, it is helpful to define two reporting status indicator variables, $\left(X_{tci}^{LR}, X_{tci}^{NR}\right)$, for late reporting (LR) and nonreporting (NR). The pair $\left(X_{tci}^{LR}, X_{tci}^{NR}\right) = \{(0,0),(1,0),(0,1)\}$ defines the month t reporting status for the ith establishment in estimation cell c.

$$X_{tci}^{LR} = \begin{cases} 1 \text{ if } r_{tci|1} = 0 \text{ and } r_{tci|3} = 1 \text{ (LR for month } t) \\ 0 \text{ if } r_{tci|3} = 0 \text{ or } r_{tci|1} = 1 \text{ (not LR for month } t) \end{cases}$$

$$X_{tci}^{NR} = \begin{cases} 1 \text{ if } r_{tci|3} = 0 \text{ (NR for month } t) \\ 0 \text{ if } r_{tci|3} = 1 \text{ (not NR for month } t) \end{cases}$$

where $r_{tci|k}$ represents the month t reporting status, as of closing $k\ (=1,2,3)$, for establishment i in estimation cell c

A simple extension of the CES model can then be written as

$$Y_{tci} = \rho_{tci} Y_{(t-1)ci} + k_{tci}$$

$$\ln(\rho_{tci}) = \ln(\rho_{tci}) + \lambda_{0tc}^{LR} X_{tci}^{LR} + \lambda_{0tc}^{NR} X_{tci}^{NR} + \lambda_{1(t-1)c}^{LR} X_{(t-1)ci}^{LR} + \lambda_{1(t-1)c}^{NR} X_{(t-1)ci}^{NR} + e_{tci}$$

$$k_{tci} \overset{iid}{\sim} N(0, \sigma_k^2)$$

$$e_{tci} \overset{iid}{\sim} N(0, \sigma_e^2)$$

Note that the approach to assigning reporting status indicator variables also establishes underlying proportionality factors at the estimation cell level, ρ_{tci}, based upon consistent preliminary reporters (PR) (i.e., based upon zero values for LR and NR status in both months). This is analogous to the current approach of estimating the proportionality factor for an estimation cell based solely upon PR establishments which were either PR or LR in the prior month; however, the extended model recognizes the proportionality factor could vary for LR and NR establishments.

The coefficients for the reporting status indicators, $\lambda_{0tc}^{LR}, \lambda_{0tc}^{NR}, \lambda_{1(t-1)c}^{LR}, \lambda_{1(t-1)c}^{NR}$, are not estimable at first closing. Therefore, further assumptions are required for estimation—a set of identifying restrictions linking the nonestimable parameters to some set of estimable parameters. A simple set of identifying restrictions is that the expected value of the coefficients is invariant over time.

$$\lambda_{0tc}^{LR} = \lambda_{0c}^{LR} + \tau_{tc}^{LR}$$

$$\lambda_{0tc}^{NR} = \lambda_{0c}^{NR} + \tau_{tc}^{NR}$$

$$\lambda_{1(t-1)c}^{LR} = \lambda_{1c}^{LR} + \tau_{(t-1)c}^{LR}$$

$$\lambda_{1(t-1)c}^{NR} = \lambda_{1c}^{NR} + \tau_{(t-1)c}^{NR}$$

where the $\tau's$ are assumed normal with mean 0 and constant variance

This model also requires prediction of current month's reporting status for establishments, given reporting status is only known for first closing reporters. A simple logit predictive model for current month's reporting status, based on prior months' reporting status may be appropriate.

$$\log\left(\frac{\pi_{tci|R_{(t-1)ci}, R_{(t-2)ci}}^{LR}}{\pi_{tci|R_{(t-1)ci}, R_{(t-2)ci}}^{NR}}\right) = \alpha_c + \beta_{R_{(t-1)ci}} + \beta_{R_{(t-2)ci}}$$

where

$$\pi_{tci|R_{(t-1)ci}, R_{(t-2)ci}}^{LR} = P\left(X_{tci}^{LR} = 1 \mid r_{tci|1} = 0, R_{(t-1)ci}, R_{(t-2)ci}\right)$$

$$\pi_{tci|R_{(t-1)ci}, R_{(t-2)ci}}^{NR} = 1 - \pi_{tci|R_{(t-i)ci}, R_{(t-2)ci}}^{LR}$$

$$R_{tci} = \begin{cases} 0 \text{ if } r_{tci|1} = 1 \, (PR) \\ 1 \text{ if } r_{tci|1} = 0, r_{tci|3} = 1 \, (LR) \\ 2 \text{ if } r_{tci|3} = 0 \, (NR) \end{cases}$$

This approach can be used, as in the extended ignorable nonresponse approach described previously, to impute values for late reporters and nonrespondents and use the completed sample data in the weighted link-relative estimator. A more extensive change would involve imputing values for nonsampled establishments as well, with the resultant complete population file used to tabulate estimates.

Estimation for this model can be carried out through Bayesian analysis utilizing Markov Chains Monte Carlo (MCMC) methods. Evaluation of the performance of the model can be carried out in several ways – comparison with third closing estimates, and comparison with CES benchmark estimates. Based upon the results of a performance analysis of the alternative

model, refinements may be proposed and evaluated. An additional challenge will be model validation; however, this is no different than that faced with the current working model.

VI. Summary

The current CES survey estimator of total employment assumes ignorable nonresponse within an estimation cell. Findings suggest that assumption may not always hold and, as a result, employment estimates may be adversely impacted. Although the impact may be relatively small, it may be sizeable in terms of employment change. Further work is needed to evaluate estimation cell definitions and to incorporate information on standard errors of the estimates into the impact assessment.

Alternative models, incorporating auxiliary information from sample establishments under an ignorable nonresponse mechanism may provide an opportunity to account for variability not addressed by the current model. Further work is needed to explore the nature of the variability in employment data and to establish and evaluate the performance of appropriate working models addressing this variability.

A broad reporting pattern classification that accounts for both reporting status and timeliness of reporting may provide a structure for developing a pattern-mixture model to estimate growth rates without the assumption of ignorable nonresponse. Alternative models would seek to leverage prior information about nonreporters where available, thereby improving upon the current working model that only incorporates information about constant reporters. Further work is needed to explore application of the pattern-mixture model with CES survey data, establish a reasonable working model, identify appropriate identifying restrictions (especially where no prior information is available), and determine impact on estimator accuracy.

References

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

Bureau of Labor Statistics. (2003), "BLS Establishment Estimates Revised to Incorporate March 2002 Benchmarks," *Employment and Earnings*: U.S. Bureau of Labor Statistics, Washington, D.C.

Copeland, K.R. (2003), "Reporting Patterns in the Current Employment Statistics Survey," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (forthcoming).

David, M.H., Little, R., Samuhel, M., Triest, R. (1983), "Imputation Models Based On the Propensity to Respond," *Proceedings of the Section on Business and Economic Statistics, American Statistical Association*, 168-173.

Little, R.J.A. (1993), "Pattern-Mixture Models for Multivariate Incomplete Data," *Journal of the American Statistical Association,* **88**, 125-134.

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.

Little, R.J.A. and Su, H.-L. (1989), "Item Nonresponse in Panel Surveys," in Kasprzyk, D., Duncan, G., Kalton, G., and Singh, M.P. (eds.), *Panel Surveys* (pp. 403-423), New York: John Wiley and Sons, Inc.

Rubin, D.B. (1976), "Inference and Missing Data," *Biometrika*, **63**, 581-592.

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

West, S., Butani, S., Witt, M., and Adkins, C. (1989), "Alternative Imputation Methods for Employment Data," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 227-232.

Response Pattern Illustrations

Shaded area represents data reported for month

Response Pattern Classification	Response Pattern Description	Month								
		1	2	…	t1	…	t2	…	T-1	T
Total Response	Unit reports every time period									
Total Nonresponse	Unit does not report for any time period									
Strict Attrition	Unit reports for every time period until some point in time, after which it no longer reports									
Strict Late Entry	Unit does not report until some point in time subsequent to the first time period, after which it continues to report for every time period									
Strict Late Entry Attrition	Unit does not report until some point in time subsequent to the first time period, after which it continues to report for every time period until some point in time, after which it no longer reports									
Attrition with Episodic Nonresponse	Unit reports for the first time period, then experiences a mixture of reporting and nonreporting until some point in time, after which it no longer reports									
Late Entry with Episodic Nonresponse	Unit does not report until some point in time subsequent to the first time period, after which it experiences a mixture of reporting and nonreporting for succeeding time periods									
Late Entry Attrition with Episodic Nonresponse	Unit does not report until some point in time subsequent to the first time period, after which it experiences a mixture of reporting and nonreporting until some point in time, after which it no longer reports									
Strict Episodic Nonresponse	Unit reports for the first time period, and experiences a mixture of reporting and nonreporting for all subsequent time periods									

Timeliness Pattern Illustrations
Shaded area represents data reported on-time for month
Dotted area represents late reported data for month

Response Pattern Classification	Timeliness Classification	Month T-1	T	Use
Current, Prior Month Reporter	On-Time both months			Preliminary
	On-time current month only			Preliminary
	On-time prior month only			Final
	Late both months			Final
Current, Prior Month Nonreporter	N/a			No
Prior Month Only Reporter	On-time			No
	Late			No
Current Month Only Reporter	On-time			No
	Late			No