

Achieving timeliness in the BLS CES program

Abstract

This document contains a description of methods used relevant to the *STES timeliness framework* within the United States Bureau of Labor Statistics (BLS) Current Employment Statistics (CES) program. These methods assist the BLS in achieving very timely monthly estimates for the following statistics:

Indicator	Common periodicity
Total employment (not from household labor force survey)	Monthly
Average weekly hours of production/nonsupervisory workers	Monthly
Average weekly earnings of production/nonsupervisory workers	Monthly
Average hourly earnings of production/nonsupervisory workers	Monthly
Average weekly overtime of production/nonsupervisory workers in manufacturing	Monthly
Women worker employment	Monthly
Production/nonsupervisory worker employment	Monthly

For more information on this program in general, please contact Patricia Getz at: Getz_P@bls.gov

Frame selection:

The Longitudinal Data Base (LDB) is the business sample frame and research database for BLS. The data are collected under the ES-202 Program, which is a Federal/State Cooperative Program conducted in cooperation with the State Employment Security Agencies (SESA). The LDB longitudinally links records across time and is updated on a quarterly basis.

The sampling frame, and the CES sample itself, are updated twice a year with new quarters of UI (Unemployment Insurance tax accounts)-based universe data. This helps keep the sample up-to-date by adding new firm births and deleting business deaths. In addition, the design specifies an annual update process, which includes sample frame maintenance and the redrawing of the entire sample for the first quarter of each year. Frame maintenance provides for the updating of industry, size class, and metropolitan area designations and for the merging of semi-annual birth samples into the overall frame.

Larry Huff (202) 691-6362

Questionnaire design:

The design of the CES reporting form (BLS form 790 series) is particularly important in maintaining continuity and consistency in reporting from month to month. The use of a single form for the entire year allows the respondent to compare the latest data submitted to the data submitted in prior. The CES form is one page, front and back. Respondents enter the monthly

employment and payroll figures on the front side. Definitions are available on the back of the form to assist the respondents, if necessary.

A CES reporting form is provided to all CES respondents except those that report via electronic file. For units that report via Computer-Assisted Telephone Interviews (CATI), Touchtone Data Entry (TDE), or the Internet, the form provides a convenient means to record their data each month. For units that report via mail, the form is submitted each month by the respondent, edited by the State agency, and returned to the respondent for use again the following month.

The CES reporting form has been used since 1930, but there have been substantial changes in its design and in the data collected over this time span. Six variations of the basic CES form are currently used. The variations tailor the data items, concepts, and definitions for each major industry sector. Separate forms are used for natural resources and mining, construction, manufacturing, service-providing industries, public administration, and educational services.

Rick Rosen (202) 691-6524

Sample design and selection:

Industries are classified in accordance with the 2002 North American Industrial Classification System (NAICS). Industrial classification refers to the grouping of reporting establishments into industries based on their major activity. Based on a description provided by the employer on a questionnaire, State Employment Security Agencies assign an industrial code to each establishment as an administrative byproduct of the UI reporting system. All data for an establishment making more than one product or engaging in more than one activity are classified under the industry of the primary product or activity, based on the information reported.

The design is a stratified, simple random sample of worksites, clustered by UI account number. The sample strata, or sub-populations, are defined by state, industry, and employment size, yielding a state-based design. Sampling rates for each stratum are determined through optimum allocation, which distributes a fixed number of sample units across a set of strata to minimize the overall variance or sampling error on the primary estimate of interest, the Statewide total nonfarm employment level.

Each year, new sample units are enrolled into the CES survey to add new firm births, realign the sample distribution with the universe distribution, and to rotate a portion of the sample. Approximately 40,000 new sample units are enrolled each year.

All firms with 1,000 employees or more and a sample of firms with less than 1,000 employees are asked to participate in the survey. In 2003, the CES sample consisted of about 160,000 businesses and government agencies that represented approximately 400,000 individual worksites drawn from a UI sampling frame. The sample rotation plan allows most firms with employment less than 1,000 to report for 4 years and then rotate out of the sample for a similar period.

Larry Huff (202) 691-6362

Reference period & due dates:

The reference week for the survey is the pay period that includes the 12th of the month, and the National News Release is usually the first Friday of the following month. Usually, the first deadline for data transmission to BLS is the second Friday after the survey week. There is a secondary transmission the Monday after the primary transmission. When the primary transmission is on Thursday, the secondary transmission is on the next business day that is not a Federal holiday. The second deadline for data transmission to BLS is usually the fifth Friday after the survey week. The third deadline for data transmission to BLS is usually the eighth Friday after the survey week. Until the third deadline, data received by the BLS is processed. The longer time period allows more sample to be used in estimation for subsequent press releases.

After the third deadline, data is no longer processed for the original survey reference week. Data transmission deadline sometimes vary due to holidays.

Rick Rosen (202) 691-6524

Data collection and validation:

Each month, the State agencies cooperating with BLS as well as BLS Data Collection Centers collect data on employment, payrolls, and paid hours from a sample of establishments. Data are collected through various automated collection modes and mail. Touch-tone data entry (TDE) serves as the primary type of electronic reporting, although a large number of reports are collected via direct electronic file transmission (EDI), and a small but growing number of reports are received via the Internet (Web). Additionally, many respondents report via Computer-Assisted Telephone Interviews (CATI) or FAX. Table 1 summarizes the distribution of sample by collection mode:

Collection Method	Percent of reports
TDE	30%
CATI	22%
EDI	19%
FAX	15%
Mail	13%
Web	1%

CES offers a multitude of reporting options to limit response burden. Whereas formerly the majority of respondents reported via mail, almost everyone now reports electronically. EDI centers lessen the burden for large companies with multiple establishments. Web reporting lessens the burden for establishments that continue to expand electronic mail communications. Respondent relationships are built by assisting new respondents with reporting (CATI) and afterwards transferring them to their preferred reporting method. CES has a Help Desk with a toll-free number and email to assist with respondent issues.

Depending on respondent reporting method, CES calls, mails, faxes, or emails an advance notice about an upcoming reporting deadline. In order to maintain good respondent relationship, CES carefully tracks who does not want to receive these notices. If a respondent fails to respond, CES calls, mails, faxes, or emails a Non Response Prompt (NRP) notice. CES also tracks the pay periods of establishments. If the pay period is not over by the reporting deadline (ex. monthly payrolls), CES will not send a NRP. This careful monitoring of payroll periods ensures a better respondent relationship because respondents are not bothered to report when they are unable to do so.

Sample data are tested to locate three types edit errors: *duplicate records*; *data and reporter identity problems*; and *internal inconsistencies*. Sample data which pass the edit tests are next put through a series of screening tests, which locate errors in the relationship between the present data and its corresponding preceding data. Screening tests check the reported data against a variety of plausible patterns in reported data over time. The purpose for screening is to find large month-to-month changes that require respondent verification. These screening tests are designed to closely imitate the decision process analysts use to determine the validity of the data; data which pass any one of these tests are accepted. Rejected data items are flagged for validation or correction. If any microdata fail all tests, then it is rejected by the system.

Rick Rosen (202) 691-6524

Organizational arrangement for input data processing:

The CES program is divided into four separate divisions:

(1) Data Collection - responsible for coordinating the flow of data to the CES program from approximately 300,000 reporters each month. There are several data collection sites responsible for collecting, editing, and conducting non-response prompting for CES sample units. These collection sites include three Data Collection Centers (DCCs), the Electronic Data Interchange (EDI) Center, the One-Point Touchtone Data Entry (TDE) Center, and each State. In addition, a small number of large reporters are collected by the Division of Federal/State Monthly Surveys (FSMS) in the BLS National Office. The CES program also has a contract with Westat, Inc., both to conduct non-response prompting for TDE units and to directly collect CES data for about 8,000 reports. DCC, EDI Center, and State employees are BLS employees; TDE Center and Westat employees are contract employees.

The Data Collection branch oversees the data collection activities of each of the data collection sites. This unit also works to improve the quality and timeliness of data collection.

Rick Rosen (202) 691-6524

(2) National Benchmarks - responsible for the annual benchmark adjustment of the national estimates. The unit also develops net birth/death factors and seasonal adjustment factors. Staff members work on a variety of technical research projects including examining disparities between the CES and ES-202 estimates, investigating data reporting problems, and modeling business births and deaths.

Kirk Mueller (202) 691-6554

(3) National Estimates - produces and analyzes the national monthly CES estimates of employment, hours, and earnings published in the Employment Situation news release. The unit also prepares the *Real Earnings* news release. Staff members prepare graphs, tables, and written and oral analyses for use by the Commissioner and other upper BLS staff in preparation for the *Commissioner's Statement to the Joint Economic Committee* and *Employment Situation* news release. Staff also works on projects to improve the quality of CES estimates and regularly contribute articles to Monthly Labor Review.

Angie Clinton (202) 691-6551

The three proceeding divisions use SAS-based Web programs to ensure program compatibility.

(4) The State and Area Estimates unit oversees the production and dissemination of estimates of State and metropolitan area employment, hours, and earnings produced by the State agencies. The unit develops seasonal adjustment factors and provides support and oversight for the Automated Current Employment Statistics system (ACES). The staff produces the employment portion of two news releases each month: *Regional and State Employment and Unemployment* and *Metropolitan Area Employment and Unemployment*.

Ken Shipp (202) 691-6519

Estimation:

The estimating methods for preliminary estimates are essentially identical to those used for calculating final estimates. The CES program does no explicit imputation to correct for non-response in preliminary or final estimates. During estimate review, however, analysts can adjust

for major, clearly defined out-of-sample events, such as strikes, and they also adjust estimates for outliers whose employment change is considered truly atypical.

Although preliminary estimates are calculated at all levels of industry detail, only more aggregate preliminary estimates are published. For employment, initial estimates are published for all 3-digit NAICS industries and above, as well as for selected 4- and 5-digit industries. Employment estimates for the remaining 4-, 5-, and 6-digit industries are published the following month. A similar pattern holds for estimates of women worker employment, production worker employment, hours, and earnings.

The sample-based employment estimates are adjusted each month by a statistical model designed to reduce a primary source of non-sampling error, the inability of the sample to capture on a timely basis, employment growth generated by new business formations. There is an unavoidable lag between an establishment opening for business and it appearing on the sample frame and being available for sampling. Because new firm births generate a portion of employment growth each month, non-sampling methods must be used to estimate this growth.

Earlier research indicated that while both the business birth and death portions of total employment are generally significant, the net contribution is relatively small and stable. To account for this net birth/death portion of total employment, the CES program uses an estimation procedure with two components: the first component uses business deaths to impute employment for business births. This is incorporated into the sample-based link relative estimate procedure by simply not reflecting sample units going out of business, but imputing to them the same trend as the other firms in the sample. The second component is an ARIMA time series model designed to estimate the residual net birth/death employment not accounted for by the imputation. The historical time series used to create and test the ARIMA model was derived from the UI universe micro level database, and reflects the actual residual net of births and deaths over the past five years.

The ARIMA model component is updated and reviewed on a quarterly basis. The net birth/death model component figures are unique to each month and exhibit a seasonal pattern that can result in negative adjustments in some months. These models do not attempt to correct for any other potential error sources in the CES estimates such as sampling error or design limitations. The most significant potential drawback to this or any model-based approach is that time series modeling assumes a predictable continuation of historical patterns and relationships and therefore is likely to have some difficulty producing reliable estimates at economic turning points or during periods when there are sudden changes in trend. CES continues to research alternative model-based techniques for the net birth/death component; it is likely to remain as the most problematic part of the estimation process.

Angie Clinton (202) 691-6551

Dissemination:

Subject to embargo, no information is released prior to the official press release. The data are released simultaneously to all interested parties by *The Employment Situation* news release, which normally is issued at 8:30 a.m. EST Friday, three weeks after the survey reference week. It is posted on the CES homepage together with the *Commissioner's Statement on the Economic Situation Release*. This material is given to approved media 30 minutes before release under lock-up conditions.

All published CES estimates are posted monthly on the internet on the morning of the release. They are available through an online database called Labstat and in a variety of tables. Estimates also are published in *Employment and Earnings*. The summary data are in the issue available about five weeks after the survey reference week; final (pre-benchmarked) figures are issued approximately nine weeks after the survey reference week.

Release schedules giving the precise release dates for the following year are published in the December issue of "Employment Situation". Release schedules giving the precise release dates for the following six months are published in each issue of "Employment and Earnings".

National data also are disseminated in the publications or on-line databases of other Federal agencies, such as the Department of Commerce, the Board of Governors of the Federal Reserve System, and the Council of Economic Advisers. Data also are regularly republished in summary form or for specific industries in many trade association journals, the labor press, and in general reference works.

Angie Clinton (202) 691-6551

Evaluation:

Data may be revised as a result of improved sample receipts. Revised estimates are released over the two months following their initial release, coinciding with the initial release for the next two months. (For example, preliminary January data are released in February, and revised data for January are released in March, together with the preliminary data for February. In April, the preliminary estimates for March are published together with revised February estimates and final [pre-benchmarked] estimates for January.) These revised estimates are subject to further revision each year when the sample-based estimates are realigned to reflect more accurate available universe counts.

All estimates back to the most recent benchmark month are subject to revision each year when new benchmarks become available. National benchmarks are published 15 months after the benchmark month (March). For example, the revised estimates based on the March 2002 benchmarks were released in June 2003. The inter-benchmark revision period extended from April 2001 through February 2002. Estimates based on the new benchmark level also were released at that time for the post-benchmark period—April 2002 through May 2003. Subsequent estimates also are based on the 2002 benchmark levels until release of the 2003 benchmark.

For the few industries exempt from mandatory UI coverage, other sources are used for benchmark information. For example, data on employees covered under Social Security laws, published by the Bureau of the Census in *County Business Patterns*, augments the UI data for religious organizations, private schools, and interns and trainees in hospitals. The Surface Transportation Board provides data for interstate railroads.

To determine the appropriate revisions, the new benchmarks for March are compared to the estimates previously made for that month. The differences represent: 1) Estimating errors that accumulated since the previous benchmark revision and 2) corrections to establishments' industry classification. These differences are assumed to have accumulated at a regular rate. The all-employee estimates are wedged, or tapered, in order to smooth out the differences between the new and old benchmarks. Estimates for the seven months subsequent to the benchmark month (April – September) are revised by applying the previously computed sample movement and new birth/death factors to the new benchmark level. Estimates for women workers and production workers are recomputed using the revised all-employee estimates and the previously computed sample ratios of these workers to all employees. Estimates for the subsequent months (October – January) after the benchmark include sample updates and new birth/death factors, as well as the effects of linking from a new benchmark level.

CES program's survey, like other sample surveys, is subject to two types of error, sampling and nonsampling error. The magnitude of sampling error, or variance, is directly related to the size of the sample and the percentage of universe coverage achieved by the sample. The establishment survey sample covers over one-third of total universe employment; this yields a very small variance on the total nonfarm estimates.

Unlike most sample surveys that publish sampling error as their only measure of error, CES can derive an annual approximation of total error, on a lagged basis, because of the availability of the independently derived universe data. While the benchmark error is used as a measure of total error for the CES survey estimate, it actually represents the difference between two independent estimates derived from separate survey processes (i.e., the CES sample process and the UI administrative process) and thus reflects the errors present in each program. Historically, benchmark revisions have been very small for total nonfarm employment.

Pat Getz (202) 691-6521

Summary of BLS techniques within STES timeliness framework categories

Broad statistical function	Statistical functions most relevant to timeliness	Methods to asset in improving timeliness and minimize loss in accuracy
Frame selection	Choice of frame source Requirements to use administrative data	Universe frame source (LDB) using UI (Unemployment Insurance tax accounts)-based data
		Twice yearly sample frame updates with new quarters of UI-based universe data
		Extensive detailed frame source so it doubles as a research database
Questionnaire design	Questionnaire and form design	One reporting form (BLS form 790 series) is used each month throughout the year so respondent can compare current and past month's data
		One page form - respondents enter data on the front side, definitions/directions available on the backside to assist respondent, if necessary
	Choice of data items	Large area for respondents to record data so CES can use OCR for mailed and faxed forms
Sample design and selection	Sophisticated sample designs	Respondents report employment and payroll figures; CES performs calculations
		Separate (yet similar) forms for industries with differing data types
	Sample designs for sub samples	Stratified, simple, random sample of worksites clustered by UI account number
	Sample selection methods to minimize sample size	New sample units enrolled each year to add new births, realign sample distribution, and rotate a portion of the sample
Reference period & due dates	Choice of reference periods for STES	Sample strata, or sub-populations, are defined by state, industry, and employment size
	Due date for businesses providing data	Sampling rates for each stratum determined through optimum allocation which distributes a fixed number of sample units across a set of strata to minimize overall variance on estimate
Data collection and validation	Efficient types of data collection methods using technology	Pay period that includes the 12th of the month encompasses all possible pay periods (monthly bimonthly, weekly, etc.)
		Second Friday after reference week for use in preliminary estimates; fifth Friday after reference week for use in revised estimates; eighth Friday after reference week for use in final estimates
		6 different reporting methods, all using the basic BLS form 790 series format
		Internet-based reporting
	Respondent relationships	Touchtone Data Entry (TDE) via telephone
		EDI collection center to minimize burden for large reporters
	Follow up of non respondents	Computer Assisted Telephone Interviews (CATI) to familiarize new respondents with reporting method
		FAX-based reporting - CES uses OCD to enter data
	Mail-based reporting - CES uses OCD to enter data	
	Send advance notices (postcards, emails, Faxes) prior to first deadline; track who does not w advance notice	
	Help Desk toll-free number available during business hours to assist respondents	
	Non Response Prompts sent to respondents dependent on reporting method (i.e., emails to W reporters, Faxes to FAX and TDE reporters, etc.)	
	Telephone calls to largest reporters (EDI center)	

		<p>Edit error tests to eliminate duplicate records, data and reporter identity problems, and internal inconsistencies</p> <p>Screening tests check the reported data against a variety of plausible patterns in reported data over time; rejected items flagged for validation or correction</p>
	Efficient data validation methods	
Organizational arrangements for input data processing	Organizational arrangements for input data processing which facilitate efficient resource use and improvements in timeliness	<p>Program divided into separate branches (data collection, National estimates, benchmarking) to streamline process</p> <p>Most data collection is completed by BLS-contracted data collection centers and by State partners</p> <p>SAS-based programs used for all National data production activities; supported by data production team</p>
		<p>Analyst screening at basic estimating cells; atypify or reject unrepresentative reports at the sub sample level before estimation</p>
		<p>During estimation, analysts adjust for large out-of-sample movements (such as strikes) and for atypical reporters</p> <p>Preliminary estimates are published only for 3-digit NAICS and for selected 4- and 5-digit NAICS industries; CES waits until revised estimates are published before releasing estimates for the remaining 4-, 5-, and 6-digit industries</p>
Estimation	<p>Early estimates from sub samples</p> <p>Estimation based on lower response rates</p>	
Dissemination	Streamlining dissemination processes	<p>Official release available on the CES homepage (www.bls.gov/ces) shortly after the 8:30 a.m. release time</p> <p>Release material given to approved media 30 minutes before release time under lock-up conditions</p>
	Target dates for dissemination	<p>Release schedules giving the precise release dates for the following year published in December</p> <p>Release schedules giving the precise release dates for the following six months are published monthly</p>
Evaluation	Revision analysis	<p>Revised estimates are published 1 month after preliminary estimates.</p> <p>Final (pre-benchmark) estimates are published 1 month after revised estimates (2 months after preliminary estimates)</p>
	Quality assessment	<p>Derive annual approximation of total error, on lagged basis, using UI data - "Benchmark error"</p>
	Benchmarking	<p>National benchmark revisions are published 15 months after the benchmark month and affect estimates subsequent to the prior benchmark month</p> <p>UI accounts are not available for certain industries; Bureau of the Census data substitutes for UI counts for those industries</p>