## B. COLLECTION OF DATA EMPLOYING STATISTICAL METHODS

### 1. Universe and Sample Size

**Universe and Sample--**The Current Employment Statistics (CES) sample is drawn from a universe frame containing over 8 million U.S. business establishments. The universe frame, known as the Longitudinal Database (LDB), serves as the sampling frame and research database of the Bureau of Labor Statistics (BLS). The LDB contains microdata records with the name, location, employment, and wages of nearly all nonfarm establishments in the United States. Each quarter, the LDB is updated with the most recent universe data available.

The CES sample includes approximately 194,000 Unemployment Insurance accounts selected from the private sector, which cover approximately 750,000 individual worksites. These UI's are selected on a random probability basis as described in section 2a below.

In addition, CES collects data from Federal, State, and Local governments. Although the governments sample is not selected on a probability basus, data is collected for a large percentage of the population employment. Data is collected for: approximately 85% (2,300,000 employees) of total Federal civilian employment; approximately 89% (4,000,000 employees) of total State employment; and approximately 66% (9,300,000 employees) of total Local government employment. Sources of the data vary, however, the Local government sample units are selected from the 74,000 Local government UI Accounts on the LDB.

Data from sample members is collected each month on employment, hours, and earnings. The survey is a Federal-State cooperative survey, with the national sample being a composite of the State samples.

### 2a. Sample Design

The CES probability sample is a stratified, simple random sample, where the strata are defined by state, super sector, and employment size. Stratification groups population members together for the pupose of sample allocation and selection. With 13 industries and 8 size classes, there are 104 total strata per State. The sampling rates for each stratum are determined through a method known as 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 employment estimates at a state level.

The Unemployment Insurance (U.I.) Account is the basic sampling unit of the CES survey. U.I. account numbers are unique within a State and include all individual establishments within a firm. Each year a new sample of U.I. accounts is drawn from the LDB. Because the LDB is updated on a quarterly basis, the annual redraw of the sample provides the most recent information available on industry, size class, and location of each unit. In addition to the annual redraw, the CES sample is updated on a semi-annual basis, as more recent universe data becomes available. The semi-annual update provides the opportunity to sample birth units that were not previously on the sampling frame during the annual redraw.

All U.I. accounts with 1000 or more employees are sampled with certainty. In addition, all units reporting through the Electronic Data Interchange (EDI) are sampled with certainty. EDI units are collected via direct file transmission from large, multi-unit employers for whom conventional data collection methods would be less effective.

The size of the currently selected probability sample and sample coverage is shown in the following table.

CES Survey Universe and Sample Size Comparison on NAICS basis, Private Sector
(in thousands)

| | Universe (March 2003) | | | Sample (March 2003) | | |
|---|---|---|---|---|---|---|
| Industry | UI Accounts | Reporting Units | Employment | UI Accounts | Reporting Units | Employment |
| Natural Resources and Mining | 28.3 | 33.5 | 553.8 | 1.8 | 5.2 | 273.8 |
| Goods Production | 976.2 | 1,028.9 | 20,890.5 | 41.5 | 73.0 | 7,874.6 |
| Trade, Transportation, and Utilities | 1,315.4 | 1,695.5 | 24,690.7 | 39.6 | 235.1 | 10,465.6 |
| Information | 89.7 | 127.2 | 3,200.9 | 5.1 | 33.7 | 1,888.9 |
| Financial Activities | 492.2 | 688.3 | 7,751.3 | 12.6 | 108.2 | 3,114.0 |
| Professional and Business Services | 1,010.2 | 1,113.9 | 15,629.2 | 33.9 | 95.4 | 6,684.6 |
| Education and Health Services | 576.9 | 679.3 | 15,671.1 | 21.9 | 64.2 | 7,123.0 |
| Leisure and Hospitality | 487.0 | 603.5 | 11,721.9 | 28.0 | 102.0 | 4,502.6 |
| Other Services | 434.3 | 476.8 | 3,586.7 | 9.8 | 26.6 | 578.5 |
| Total | 5,410.3 | 6,446.9 | 103,695.9 | 194.4 | 743.7 | 42,505.5 |

## 2b. Estimation Procedure

The estimation technique used in estimating All Employees (AE) is a weighted link-relative estimator, which is a form of ratio estimation. From a sample composed of establishments reporting for both the previous and current months, the ratio of current month, weighted employment to that of the previous month, weighted employment is computed. The weights are defined to be the inverse of the probability of selection in the sample. The weight is calculated based on the number of U.I. accounts actually selected within each allocation cell. Estimates are calculated within each estimation cell and then summed across appropriate cells to form estimates for aggregate levels.

The weighted link and taper estimator used for non-AE datatypes accounts for the over-the-month change in the sampled units, but also includes a tapering feature used to keep the estimates close to the overall sample average over time. The taper is considered to be a level correction. Like the estimator for AE, only matched sample is used. The estimator tapers the estimate toward the sample average for the previous month of the current matched sample before applying the current month's change.

## 2c. Reliability

Like other sample surveys, CES 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. Because the CES sample covers over one-third of total universe employment, the sample error on the total nonfarm estimate is small. The relative standard error for the major industry divisions at the national level under the probability sample are given in the table below.

| Major Industry Division | Average Relative Standard Error for All Employment (in percent) |
|---|---|
| Total Private | 0.2 |
| Natural Resources and Mining | 2.1 |
| Construction | 0.8 |
| Manufacturing | 0.4 |
| Wholesale Trade | 0.7 |
| Retail Trade | 0.4 |
| Transportation and Warehousing | 0.7 |
| Utilities | 0.9 |
| Information | 1.0 |
| Financial Activities | 0.5 |
| Professional and Business Services | 0.8 |
| Educational and Health Services | 0.3 |
| Leisure and Hospitality | 0.4 |
| Other Services | 1.3 |

The estimation of sample variances for the CES survey is accomplished through the method of Balanced Half Samples (BHS). This replication technique uses half samples of the original sample and calculates estimates using those subsamples. The sample variance is calculated by measuring the variability of the subsample estimates.

**2d. Benchmark Revisions**

The sum of sampling and nonsampling error can be considered total survey error. Most sample surveys are only able to publish sampling error as their only measure of error. CES has the ability to produce an 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 processes, and thus reflects the errors present in each program. Historically, the benchmark revision has been very small for total nonfarm employment. Over the past decade, percentage benchmark error has averaged 0.2 percent, with an absolute range from less than 0.05 percent to 0.5 percent.

**2e. Specialized Procedures**

The Bureau has conducted extensive research into various ways to more directly capture the impact of new business births. This research included obtaining early records of new UI accounts and a pilot program to solicit from this frame. Operationally, a sample-based approach did not yield satisfactory results. This was mainly due to the lack of a comprehensive sampling frame on a timely basis. While both employment gains and losses from new and failed businesses are large in terms of over the year change, research

conducted by the Bureau shows that the net employment (employment gained minus employment loss) is small because the gains and losses offset each other. The sample design calls for accounting for the majority of the employment gain from new businesses by imputing for UI accounts that have gone out-of-business. On a semi-annual basis, the universe is reviewed to identify new births. A portion of the births are selected on a probability basis. Thus, only births (and deaths) since the semi-annual update (about a 15 month lag) must be imputed for. The Bureau has researched models to account for the residual birth employment not accounted for by the death imputation model. Models are currently in use for all industry estimates.

**2e. Data Collection Cycles**

The CES survey was mandated by Congress to be a monthly survey.

**3. Methods to Maximize Response Rates**

New firms are enrolled into the survey by interviewers working in BLS Data Collection Centers. The response rate for new enrollments is about 80%. After enrollment, sample attrition averages about 1% per month. CES rotates new units into the sample each year both to replace deaths and to re-align the sample by industry and size. Typically about 15% of the units are replaced each year.

The overall response rate for final estimates is about 65%.

The link-relative estimating technique implicitly adjusts for nonrespondents using the respondents' relationship of current to previous month's employment within each estimation cell.

Current Employment Statistics survey estimates are generated three times for each month to reflect additional sample received. (Estimates are revised two more times to reflect updated universe counts). Policy makers in both the private and public sectors rely extensively on the first estimate for the month. The Bureau has implemented procedures to limit the size of revisions in these preliminary estimates. Automated collection methods, CATI, TDE, and Electronic Data Interchange (EDI) have been identified as the best possible means of overcoming the revision problem in the first estimate. These methods have been found to consistently improve response rates for preliminary estimates by 30 percentage points over the mail rate.

BLS and the cooperating States conduct an extensive and vigorous program of notification and nonresponse follow-up. These include:

- Targeted advance notice faxes and postcards, sent to all sample units
- Time specific nonresponse prompting Fax messages, telephone calls, and postcards.

In a typical month, BLS and the States conduct over 15,000 nonresponse prompt phone calls, and send over 50,000 fax messages to nonrespondents.

In addition, BLS and the States follow an aggressive refusal conversion protocol. Each month the BLS Data Collection Center and the States target prior refusals for re-contact. About one-half of these refusals agree to resume reporting.

Growth of EDI, the direct transfer of data from the firm to BLS, also provides a high level of response and stability. BLS currently collects over 80,000 reports from nearly

100 large firms via EDI. For final estimates, virtually all of these firms provide data. EDI also experiences very few refusals.

## 4. Tests

BLS has undertaken several new initiatives in the area of research on control and measurement of non-sampling error. The 1991 benchmark revealed a substantial non-sampling error problem caused by payroll processing firms. The American Statistical Association formed a committee to review BLS procedures and issued a report in January 1994. BLS has adopted most of the report's recommendations. BLS has also conducted a Response Analysis Survey of Payroll Processing Firms. The purpose of the survey was to identify practices that can affect the data collected by the CES program and the U.I. Quarterly Census of Employment and Wages program (the benchmark source data) and educate payroll processors on proper reporting procedures. Payroll processing firms that report changes in procedures are asked to perform a dual run under old and new procedures for one or more months, if possible, in order to assess the impact of the change. BLS has also conducted a Response Analysis Survey (RAS) of CES and Unemployment Insurance covered employment reporting. The survey identified factors that affect both CES and U.I. reporting within the same firm. Based on these RAS studies, BLS has undertaken an extensive education program with CES respondents. This includes highlighting correct reporting of problem items on the CES report form and the inclusion of special notices on correct reporting on the monthly advance notice Fax message.

BLS has developed forms to allow respondents to report data by facsimile transmission ("fax"). These forms lessen reporting burden on large multi-unit reporters by allowing them to report information for their establishments on one form.

## 5. Statistical Responsibility

Ms. Shail Butani, Chief, Statistical Methods Division of the Office of Employment and Unemployment Statistics, is responsible for the statistical aspects of the CES survey. Ms. Butani can be reached on 202--691-6347.