B. COLLECTION OF DATA EMPLOYING STATISTICAL METHODS

1. Universe and Sample Size

Universe and Sample--The Current Employment Statistics (CES) sample for the private sector is drawn from a private sector universe comprised of about 6.7 million in-scope U.S. private business establishments (worksites). Private households and all agriculture except forestry are excluded from the frame. 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 primary data source for the LDB is the Quarterly Contribution Reports filed by employers with their State's Unemployment Insurance Agency. The BLS has cooperative agreements with each State to access these data through the Quarterly Census of Employment and Wages (QCEW) program. The LDB contains microdata records with the name, location, NAICS industrial classification, 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 probability sample includes approximately 196,000 Unemployment Insurance (U.I.) accounts selected from the approximately 5,510,300 U.I. accounts in the private sector. The sampled U.I. accounts cover approximately 864,400 individual worksites. These U.I. accounts 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 basis, data are collected for a large percentage of the population employment. Data are collected for: approximately 84.1 percent (2.38 million employees) of total Federal civilian employment; approximately 77.8 percent (3.97 million employees) of total State employment; and approximately 59.6 percent (8.52 million employees) of total local government employment. The government sample units are selected from the 19,600 State, and 74,600 local government U.I. accounts on the LDB. Federal government employment is collected from the National Finance Center.

During 2011 approximately 6,000 additional local government U.I. accounts were identified from the LDB/QCEW and will be solicited to increase the employment coverage for all MSAs in an effort to improve the local government published estimates for MSAs.

Data from sample members are 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, industry (NAICS) major industry division, and employment size (BLS Handbook of Methods, Chapter 2, page 5, Bureau of Labor Statistics, 2004). Stratification groups population members together for the purpose of sample allocation and selection. With 13 industries and 8 size classes, there are 104 total strata per State (Erkens, Huff, and Gershunskaya, 2005). 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 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. Though the LDB is updated on a quarterly basis, the sample is redrawn annually due to budget constraints. The annual sample is redrawn to reflect the changes in the population or the sampling frame. These changes include, but are not limited to: removal of establishments that have gone out of business;

inclusion of new establishments (i.e., births); growth or decline in the employment of each establishment and consequently the U.I. account; and updated industrial classification for some establishments. The annual sample of U.I. accounts is drawn using a permanent random numbers technique in order to achieve the desired sample overlap with the U.I.s from previous samples (Butani, Robertson, and Mueller, 1998). The use of this technique keeps solicitation costs within budget and keeps the sample statistically valid in terms of updated probabilities of selection that reflect all recent changes (Crankshaw, Kujawa, and Stamas. 2002). 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.

Large U.I. accounts are sampled with virtual certainty. In addition, all units reporting through the Electronic Data Interchange (EDI) center 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 cost 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 2010)			Sample (March 2010)		
Industry	U.I. Accounts	Reporting Units	Employment	U.I. Accounts	Reporting Units	Employment
Natural Resources and Mining	30.2	35.4	666.0	2.0	5.1	317.6
Construction	603.5	626.1	5,191.2	20.4	36.9	1,019.3
Durable Manufacturing	183.5	200.7	6,943.3	9.3	18.8	2,932.1
Non-Durable Manufacturing	102.0	113.9	4,398.2	6.3	12.7	1,933.4
Trade, Transportation, and Utilities	1,254.4	1,676.8	23,991.1	36.8	276.6	10,984.8
Information	83.6	122.5	2,704.1	4.9	36.1	1,612.0
Financial Activities	491.5	719.6	7,375.7	12.3	127.8	3,035.6
Professional and Business Services	1,130.5	1,259.5	16,307.0	36.7	113.8	6,625.8
Education and Health Services	659.6	793.8	18,564.9	26.2	87.1	8,547.4
Leisure and Hospitality	536.1	670.6	12,529.2	31.3	119.6	4,849.5
Other Services	435.4	479.6	3,502.1	9.9	29.8	602.5
Total	5,510.4	6,698.6	102,172.7	196.0	864.4	42,460.1

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 (for detailed mathematical formulae of this estimator, see the BLS Handbook of Methods, Chapter 2, pages 5-7, Bureau of Labor Statistics, 2004). 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 the matched sample is used to reduce the variance on the over-the-month change. 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 non-sampling 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 about 40 percent of total universe employment, the sample error on the national, first closing, total nonfarm employment 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	0.1
Natural Resources and Mining	2.3
Construction	0.9
Manufacturing	0.4
Wholesale Trade	0.6
Retail Trade	0.3
Transportation and Warehousing	0.6
Utilities	1.2
Information	0.9
Financial Activities	0.5
Professional and Business Services	0.5
Educational and Health Services	0.3
Leisure and Hospitality	0.4
Other Services	0.8

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 estimates made from these subsamples. (For a detailed mathematical presentation of this method, see the BLS Handbook of Methods, Chapter 2, pages 8-9, Bureau of Labor Statistics, 2004.)

2d. Benchmark Revisions

The sum of sampling and nonsampling error can be considered the total survey error. Most sample surveys are only able to publish the sampling error as their only measure of error. CES has the ability to produce an approximation of the total error, on a lagged basis, because of the availability of the independently derived universe data from the QCEW program. While the benchmark error is used as a measure of the 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, the benchmark error has averaged 0.27 percent, with an absolute range from 0.1 percent to 0.7 percent.

2e. Specialized Procedures

The BLS 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 U.I. 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 BLS shows the net employment (employment gained minus employment lost) is small because the gains and losses offset each other (Mueller, 2006). The sample design accounts for the majority of the employment gain from new businesses by imputing for U.I. accounts that have gone out-of-business (Kropf, Strifas, and Traetow, 2002). 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 BLS 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 private-industry estimates.

2f. 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 70 percent. After enrollment, sample attrition averages about 1 percent 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 25 to 30 percent of the units are replaced each year. The sample selected from the 2009 QCEW sample frame, which was placed in production for making January 2011 estimates, rotated out of the sample all U.I. accounts that had been in the sample for four years or longer, except for those certainty or near certainty sample units. This will have a positive effect on response rates since the response rates generally trend down over time, as units remain in the sample over a long period of time.

The response rate (based on weighted employment) used in making final estimates for the private sector probability sample is about 50 percent. As indicated earlier, these sample respondents from the private sector are combined with government reports that cover 84.1percent of Federal government employment, 77.8 percent of State government employment, and 59.6 percent of local government employment to make estimates for total nonfarm employment.

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

CES 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 BLS has implemented procedures to limit the size of revisions in these preliminary estimates. Automated collection methods, such as CATI, Web, and TDE, have been identified as the best possible means of overcoming the revision problem in the first estimate. These methods have been found to consistently maintain collection rates for preliminary estimates between 65 and 75 percent.

The BLS and the cooperating Territories conduct an extensive and vigorous program of notification and non-response follow-up. These include:

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

In a typical month, the BLS conducts over 20,000 non-response prompt phone calls, and sends over 20,000 fax messages to non-respondents.

In addition, the BLS follows an aggressive refusal conversion protocol. Each month the BLS Data Collection Centers 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 the BLS, also provides a high level of response and stability. The BLS currently collects 133,000 reports from 93 large firms via EDI. For final estimates, virtually all of these firms provide data. EDI also experiences very few refusals.

In Fiscal Year 2011 BLS began to implement a restructuring of the CES program. This initiative included three components: (1) centralize the production of state and metropolitan area estimates, (2) reduce program costs significantly, and (3) utilize a portion of the savings to improve survey response rates. The first two components have been completed. In order to implement the third component, the program plans to hire additional CATI interviewers during Fiscal Year 2012. With more CATI interviewers on staff, several objectives will be targeted. First, fewer new respondents will be rotated to self-reporting methods. Keeping more survey respondents on interviewer-based collection will enhance the overall response rate, reducing the total survey error. Second, additional interviewers will allow the BLS to obtain more of the total response in time for the first release of estimates, thereby reducing the average size of the over-the-month revision.

Each year the BLS conducts analyses of the survey estimates and decomposes the total survey error into its various components including sampling error, non-response error, reporting error, and errors from other sources. These analyses are possible since the LDB provides employment data for all units in the population. It is possible to use this employment information to calculate CES estimates using all of the units selected in the sample (100 percent response) and compare it with the CES estimates using only those units that responded to CES. This provides a measure of the non-response error. Similar methods are used to measure other sources of survey error which can then be aggregated to the total survey error. See Gershunskaya, Eltinge, and Huff, 2002, for detailed mathematical formulae and numerical results on analysis of multiple error components in small domain estimates for the CES program. These analyses are useful in determining the source of errors for past estimates; however, they have not yet proven useful in predicting or limiting the errors in current estimates. The attached research paper presented at the 2009 American Statistical Association Annual Meeting showed that the non-response error was not a principal factor in the total survey error and that the non-response error is not unidirectional or particularly related to the response rates. The study covers the CES estimates

from 2004 through 2008. These error decomposition studies and the effect of non-response on the survey error will continue to be monitored for any changes. Detailed tables of response rates for the private sector probability sample are available for earlier years and are updated on an annual basis. Details are available on request.

4. Tests

The 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 (American Statistical Association, 1994). The BLS has adopted most of the report's recommendations. The BLS has also conducted a Response Analysis Survey of Payroll Processing Firms (Goldenberg, Moore, and Rosen, 1994). The purpose of the survey was to identify practices that can affect the data collected by the CES program and the U.I. QCEW 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. The BLS has also conducted a Response Analysis Survey (RAS) of CES and U.I. covered employment reporting (Werking, Clayton, and Rosen, 1995). The survey identified factors that affect both CES and U.I. reporting within the same firm. Based on these RAS studies, the 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. A new RAS was conducted in 2008, and a report detailing new findings is attached.

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.

6. References

American Statistical Association (1994) "A Research Agenda to Guide and Improve the Current Employment Statistics Survey." American Statistical Association Panel for the Bureau of Labor Statistics' Current Employment Statistics Survey, January, 1994. Alexandria, VA: American Statistical Association.

Bureau of Labor Statistics. BLS Handbook of Methods Chapter 2: Employment, Hours, and Earnings from the Establishment Survey. Washington DC: Bureau of Labor Statistics, 2004, p.5. http://www.bls.gov/opub/hom/pdf/homch2.pdf

Bureau of Labor Statistics. BLS Handbook of Methods Chapter 2: Employment, Hours, and Earnings from the Establishment Survey. Washington DC: Bureau of Labor Statistics, 2004, pp. 5-7. http://www.bls.gov/opub/hom/pdf/homch2.pdf

Bureau of Labor Statistics. BLS Handbook of Methods Chapter 2: Employment, Hours, and Earnings from the Establishment Survey. Washington DC: Bureau of Labor Statistics, 2004, pp. 8-9. http://www.bls.gov/opub/hom/pdf/homch2.pdf

Butani, Shail, Kenneth W. Robertson, and Kirk Mueller (1998) "Assigning Permanent Random Numbers to the Bureau of Labor Statistics Longitudinal (Universe) Data Base." Proceedings of the Survey Research Methods Section, American Statistical Association, 9-13 August, 1998. Dallas: American Statistical Association, 1998.
http://www.bls.gov/ore/pdf/st980080.pdf

Crankshaw, Mark, Laurie Kujawa, and George Stamas (2002) "Recent Experiences in Survey Coordination and Sample Rotation within Monthly Business Establishment Surveys." Proceedings of the Survey Research Methods Section, American Statistical Association, 11-15 August, 2002. New York: American Statistical Association, 2002.
http://www.bls.gov/ore/pdf/st020290.pdf

Erkens, Gregory, Larry L. Huff, and Julie B. Gershunskaya (2005) "Alternative Sample Allocations for the U.S. Current Employment Statistics Survey." Proceedings of the Survey Research Methods Section, American Statistical Association, 7-11 August, 2005. Minneapolis: American Statistical Association, 2005, pp. 1-4.
http://www.bls.gov/ore/pdf/st050240.pdf

Gershunskaya, Julie, John L. Eltinge, and Larry L. Huff (2002) “Use of Auxiliary Information to Evaluate a Synthetic Estimator in the U.S. Current Employment Statistics Program.” Proceedings of the Section on Survey Research Methods, American Statistical Association, 11-15 August, 2002. New York, NY: American Statistical Association, 2002.
http://www.amstat.org/sections/SRMS/Proceedings/y2002/Files/JSM2002-000844.pdf

Gershunskaya, Julie, and Larry Huff (2009) “Components of Error Analysis in the Current Employment Statistics Survey.” Proceedings of the Survey Research Methods Section, American Statistical Association, 1-6 August 2009. Washington, DC: American Statistical Association, 2009.
http://www.bls.gov/osmr/abstract/st/st090050.htm

Goldenberg, Karen L., Susan E. Moore, and Richard J. Rosen (1994) "Commercial Payroll Software and the Quality of Employment Data." Proceedings of the Survey Research Methods Section, American Statistical Association, 13-18 August, 1994. Toronto: American Statistical Association, 1994.
http://www.amstat.org/sections/SRMS/Proceedings/papers/1994_178.pdf

Kristin Fairman, Margaret Applebaum, Chris Manning, and Polly Phipps (2009) “Response Analysis Survey: Examining Reasons for Employment Differences Between the QCEW and the CES Survey” Proceedings of the Survey Research Methods Section, American Statistical Association, 1-6 August 2009. Washington, DC: American Statistical Association, 2009.
http://www.bls.gov/osmr/abstract/st/st090240.htm

Kropf, Jurgen, Sharon Strifas, and Monica Traetow (2002) "Accounting for Business Births and Deaths in CES: Bias vs. Net Birth/Death Modeling." Washington DC: Bureau of Labor Statistics, 2002.
http://www.bls.gov/ore/pdf/st020090.pdf

Mueller, Kirk (2006) "Impact of business births and deaths in the payroll survey." Monthly Labor Review, Vol. 129, No. 5, May 2006, pp. 28-34.
http://www.bls.gov/opub/mlr/2006/05/art4full.pdf

Werking, George S., Richard L. Clayton, and Richard J. Rosen (1995) "Studying the Causes of Employment Count Differences Reported in Two BLS Programs." Proceedings of the Survey Research Methods Section, American Statistical Association, 13-17 August, 1995. Orlando: American Statistical Association, 1995.
http://www.amstat.org/sections/SRMS/Proceedings/papers/1995_137.pdf