Part B. Collections of Information Employing Statistical Methods

1. Description of Universe

The 2007 Business Expenses Supplement (BES) covers businesses classified in Wholesale Trade (excluding manufacturers' sales branches and offices and agents and brokers), Retail Trade, and Accommodation and Food Services, as defined by the 2002 North American Industry Classification System (NAICS).

The following NAICS sectors, previously covered by Business Expenses Survey, will be covered as part of the 2007 Service Annual Survey: Transportation and Warehousing; Information; Finance and Insurance; Real Estate, Rental and Leasing; Professional and Technical Services; Administrative, Support, and Waste Services; Business and Trade Schools; Health and Social Assistance Services; Arts, Entertainment, and Recreation; and Other Services.

The 2007 BES is a one-time extension of the Annual Wholesale Trade Survey (AWTS) and the Annual Retail Trade Survey (ARTS), together consisting of 28,363 firms. The samples represent a universe of approximately 2.1 million establishments, based on the Census Bureau's Business Register (formerly the Standard Statistical Establishment List).

2. Sampling Methodology and Estimating Procedures

   a. Sampling Methodology

      The sample for the Business Expenses Supplement is comprised of businesses selected for the ARTS and a subset of businesses selected for the AWTS. The methods described below are used to design and select the sample for each of these surveys.

      The initial sample for each survey is selected using a stratified, one-stage design where strata are defined by industry and size. Within each of the size strata, units are selected using simple random sampling without replacement. Stratification is used because estimates are desired for specific industries and stratification increases the efficiency of the sample (i.e., fewer units are needed to meet desired sampling variance constraints).

      Sample sizes are computed to meet coefficient of variation (CV) constraints on estimated annual sales/receipts (and inventories for wholesale and retail) totals at various industry levels. Survey analysts specify desired publication levels based on the 2002 North American Industry Classification System (NAICS) and give sampling variability constraints for these levels (in the form of CVs for estimated totals and standard errors for period-to-period change estimates). Publication levels may be defined by a specific NAICS industry or a group of NAICS industries.
      Primary strata are defined by industry based on 2002 NAICS detail. The primary strata are substratified into 4, 7, 10, or 13 strata based on estimated annual sales/receipts. The

largest sales/receipts size stratum within each primary stratum consists of companies, all of which are selected with certainty (probability equal to one). All firms not selected with certainty are subjected to sampling on an employer identification number (EIN) basis. If a firm had more than one EIN at the time of initial sampling, we treat each of its EINs as a separate sampling unit.

To redistribute the reporting burden for small and medium-sized businesses, we exclude from the noncertainty sampling operations particular noncertainty sampling units that were selected in the previous sample. We call this procedure, "controlled nonselection." The first part of the controlled nonselection procedure specifies the creation of a data set of noncertainty sampling units in the previous sample. The second part of the procedure identifies a subset of these units that meets the following two conditions. First, the estimated amount of bias in our estimates of sales (and inventories, for the wholesale samples) and their corresponding variance estimates due to the controlled nonselection of these sampling units in each primary stratum is negligible. Second, the number of sampling units in the substratum that are not in the previous sample is greater than or equal to the required substratum sample size.

The sampling frame has two types of sampling units represented -- Employer Identification Numbers (EINs) and large, multiple-establishment firms (also called companies). Both sampling units represent clusters of one or more establishments owned or controlled by the same firm. The information used to create these sampling units was extracted from data collected as part of the 2002 Economic Census and from establishment records contained on the Census Bureau's Business Register updated with results of the annual Company Organization Survey and the latest available administrative data obtained from other Federal agencies.

To create the sampling frame, establishments are assigned to primary strata based on their NAICS industry. The primary strata correspond to the industry levels where analysts have provided sampling variability constraints. Measures of size (estimated annual sales/receipts) for establishments are aggregated according to their company affiliation and separately, according to their EIN affiliation. This aggregation gives the total measure of size for the company sampling units and the EIN sampling units. The company sampling unit is then assigned to the sampling recode that contributes the most to the unit's measure of size (i.e., annual sales/receipts). A similar procedure is performed for EIN sampling units.

The first step in the sample selection identifies firms selected with certainty. If a firm's estimated annual sales (or end-of-year inventories for retail and wholesale) are greater than the corresponding certainty cutoff, that firm is selected into its respective sample with certainty.

All firms not selected with certainty are subjected to sampling on an EIN basis. If a firm has more than one EIN, we treat each of its EINs as a separate sampling unit. The EINs are stratified according to their major industry and their estimated annual sales (on a

2002 basis). Within each noncertainty stratum, a simple random sample of EINs is selected without replacement.

Periodically, we update the samples to represent EINs appearing on the Business Register. These new EINs, called births, are EINs recently assigned by the IRS, on the latest available IRS mailing list for FICA taxpayers, and assigned and industry classification (if possible) by the Social Security Administration (SSA).

EIN births are sampled on a quarterly basis using a two-phase selection procedure. To be eligible for selection, a birth must either have no industry classification or be classified in an industry within the scope of the annual surveys, and it must meet certain criteria regarding its number of paid employees or quarterly payroll.

In the first phase, births are stratified by industry and a measure of size based on expected employment or quarterly payroll. A relatively large sample is selected and canvassed to obtain a more reliable measure of size, consisting of sales/receipts in two recent months, company affiliation information, and a new or more detailed industry classification code. Births that have not returned their questionnaire after 30 days are contacted by telephone.

Using this more reliable information, the selected births from the first phase are subjected to probability proportional-to-size sampling with overall probabilities equivalent to those used in drawing the initial sample for its respective survey.

Because of the time it takes for a new employer firm to acquire an EIN from the IRS, and because of the time needed to accomplish the two-phase birth-selection procedure, births are added to the sample approximately nine months after they begin operation.

b. Estimation Procedure

Estimates are developed from the summation of weighted data (reported or imputed) for all selected sampling units eligible for tabulation. The assigned weight for each sampling unit is the reciprocal of its probability of selection into its respective sample. Estimates are adjusted using results of the latest available Economic Census, including estimates for retail nonemployers. Variance estimates are calculated using the method of random groups.

c. Measurable Accuracy

The accuracy of the survey results is determined by the joint effect of sampling variability and nonsampling errors.

- *Sampling variability* results because observations are being made on a sample of firms and not on the entire universe. Because each firm on the sampling frame has a known probability of being selected into the sample, it is possible to estimate the sampling variability of the survey estimates.

- *Nonsampling errors* can result from many sources: inability to obtain information from all cases in the sample; response errors; definitional difficulties; misinterpretation of questions; mistakes in recording or coding data obtained; and other errors of collection, response, coverage and estimation of missing data. Although no direct measurement of the biases due to nonsampling errors will be obtained, precautionary measures will be taken in all phases of the collection, processing, and tabulation of the data in an effort to minimize their influences. To further minimize sampling errors, all questionnaires will be reviewed for completeness and consistency. Extreme or unusual data will be verified and corrected if necessary.

3. Methods to Maximize Response and Accounting for Nonresponse

   a. Data Collection Procedures

   This information collection will maximize response through the following means: (i) mailing materials which state response to the survey is mandatory and that data reported to the Census Bureau are confidential, as provided by Title 13, USC; (ii) effective questionnaires, instructions, and an Internet reporting system that try to make reporting as simple as possible; (iii) customized mailing and reporting arrangements for selected large firms; (iv) toll-free and Internet assistance for any business that has questions about completing its report; and (v) systematic mail follow-ups for nonresponse, supplemented by two telephone follow-ups.

   We estimate the overall unit response rate to be at least 80 percent. We expect 90 percent of the response to be received by mail, fax, or Census Taker (Internet reporting). We estimate 10 percent to be received by telephone contact. This level of response will yield accuracy and reliability that are adequate for intended uses of survey results.

   b. Estimating for Missing Employer Data and for Nonemployer Data

   Missing data on expenses will be imputed to provide reliable estimates for the entire universe of employer firms. Data for nonemployers in all except the wholesale trade sector will be imputed based on administrative sales data and the distribution of expenses for employers in each industry. Estimates for wholesale trade nonemployers are not included in this program.

4. Testing of Procedures

Data collection procedures to be used in this program will be integrated with those used annually for the Bureau's annual trade surveys. Hence, no additional testing is required.

5.   Contacts for Statistical Aspects and Data Collection

The 2007 Business Expenses Supplement (BES) will be conducted in the Services Sector Statistics Division (SSSD), Bureau of the Census, Washington, DC 20233, under the direction of John R. Trimble, Chief, Annual Retail Branch (telephone 301-763-7223). Data collection, processing, and analysis, are managed by Sheldon G. Ziman, BES Section Chief (telephone 301-763-6559).

Carol S. King, Chief of the Statistical Methods Branch, SSSD, (telephone 301-763-2675); and William C. Davie, Jr., Chief of the Program Research and Development Branch, SSSD, (telephone 301-763-7182), are responsible for statistical methodology, including sample design and methodologies for nonresponse and estimation.

ATTACHMENTS
2007 Business Expenses Supplement to the Annual Wholesale and Retail Trade Surveys

A.   Contacts with Affected Agencies and Organizations

B.   Correspondence from the Bureau of Economic Analysis

C.   Prototype Report Form

D.   Prototype Cover Letter