

# QUALITY OF THE BUREAU OF LABOR STATISTICS' BUSINESS ESTABLISHMENT LIST AS A SAMPLING FRAME

Michael A. Searson, Tracy E. Farmer, U. S. Bureau of Labor Statistics  
Michael A. Searson, BLS, 2 Massachusetts Ave. NE, Room 4840 Washington, DC 20212

**Key Words:** coverage, establishments, firms, addresses, editing

## I. Introduction

The Business Establishment List (BEL) is one of the outputs of the Bureau of Labor Statistics' (BLS) Covered Employment and Wages (ES-202) program. The ES-202 program is a Federal/State statistical program managed by BLS and the State Employment Security Agencies (SESA) in the 50 States, the District of Columbia, Puerto Rico, and the Virgin Islands. The BEL is used as the sampling frame for most of the BLS' programs. Nationwide, in 1996, Federal and State Unemployment Insurance (UI) programs covered 118 million full- and part-time workers who received \$3.42 trillion in pay.<sup>1</sup> ES-202 program data were collected for approximately 7.2 million employer establishments in 1996.<sup>2</sup> The BEL is the most complete and timely source of business establishment information used as a sampling frame.

The SESAs conduct four principal ES-202 program activities: data collection; assigning industrial and geographical codes; periodic review of these codes including review of the business identification information (names, addresses, etc.); and, micro/macro data editing. The SESAs submit both micro-level data (Quarterly Unemployment Insurance Name and Address File-QUI) and macro-level data (ES-202 Report) to BLS. The QUI is an input to the BEL.

Beginning with the 1997 third quarter data (due February 22, 1998), the SESAs will submit only one file, the Enhanced Quarterly Unemployment Insurance (EQUI), to the BLS each quarter. The EQUI is similar to the QUI; however, it contains additional data elements extracted from the SESA UI files. BLS will use the EQUI file to create the macro level data for the ES-202 Report. The code change information which has been submitted on a separate file (Annual Refiling Survey Control File) to BLS, will now be captured on the first quarter 1998 EQUI file (due August 22, 1998).

New data processing and editing procedures both in the SESAs and BLS will be implemented with the

production of the 1997 third quarter EQUI. These new systems will accommodate the approaching implementation of the North American Industry Classification System (NAICS). NAICS will replace the current SIC system. More details on the editing procedures and the EQUI file will be presented later in the paper.

## II. Data Sources of the Sampling Frame

As stated above, the QUI File and the ES-202 Report are products sent by the SESAs to BLS each quarter which contain data from the administrative files of the State UI programs. They are supplemented by data from the Multiple Worksite Report and the Report of Federal Employment and Wages both, of which are described in detail in Section 6.

The employment and wages data produced by the ES-202 program represent the universe of workers covered under State UI laws (this includes the private sector, State and local governments) as well as civilian workers covered by the program of Unemployment Compensation for Federal Employees (UCFE). Since coverage is so broad (approximately 98 percent of all non-farm wage and salary employment), the ES-202 program provides a virtual census of these employees and their wages. It is the most complete and timely source of monthly employment and quarterly wages information by detailed industry and county. Consequently, ES-202 data are used extensively in many economic and statistical applications. These include UI program administration, macro-economic research, survey employment benchmarking, and micro-economic analysis. The Bureau of Economic Analysis uses the macro level ES-202 data in the preparation of the Personal Income component of the Gross Domestic Product. BLS programs and surveys use the micro level ES-202 data for sampling purposes. These BLS programs and surveys are Current Employment Statistics; Occupational Employment Statistics; Producer Price Index; Occupational Safety and Health Surveys; Occupational Compensation Survey; Employer Benefits Survey; Employment Cost Index Survey; and, Productivity Surveys. BLS also will use the micro level ES-202 data as an input to the Longitudinal Database (LDB). In 1998, the LDB will produce a new data series to analyze the job creation and job destruction process.

<sup>1</sup> Preliminary 1996 Employment & Wages Data

<sup>2</sup> Preliminary 1996 Employment & Wages Data

BLS, a user of UI program data, does not control the procedures and processes SESAs use for collecting and maintaining the UI administrative and accounting data. For example, BLS has no control over the assignment of the UI account numbers used by the SESAs to identify employers. BLS does, however, contract with each State's LMI unit to provide selected UI data for statistical purposes. The contract, called the LMI Cooperative Agreement, specifies the quality standards and the types and timing of data submitted for the five Federal/State statistical programs.

All of the ES-202 program data originates from employers. Since these data are collected and processed by both UI and LMI staff, it is important to note the distinction between the two units because the UI unit is not under the control of BLS. Each State's UI program collects and maintains administrative record data and business identification information on employers who are subject to State UI laws. The UI program provides these micro, or employer-level, data to the LMI unit of the respective SESAs. The LMI unit then supplements, edits, and processes these data for use in the ES-202 program. Included in these administrative data are the registration information that is obtained from every new employer. All new employers are required to file a "Status Determination Form" to obtain a UI account number and UI tax rate. Information from this form is also used to assign to each establishment a set of classification codes (industrial, geographical, ownership, and auxiliary).

All covered employers are also required to file a "Quarterly Contribution Report (QCR)." The employers must report the total wages paid to all workers during the calendar quarter; the wages that are subject to UI taxes; and the UI taxes that are due to the State. These forms are described in more detail later in this paper.

The LMI unit's activities are centered on four primary areas from the UI administrative files. First, the LMI staff uses information collected on the Status Determination Form to assign the classification codes previously mentioned. Second, the LMI staff supplements the data provided by UI to obtain a greater level of geographic and industry detail for selected employers' classification codes. Specifically, they collect employment, wages, and business identification information for the individual establishments of most multi-establishment employers, which are not separately reported on UI administrative records (see section on the Multiple Worksite Report (MWR)).

Third, the LMI staff directly collects data from Federal government agencies each quarter to add the employment, wages and business identification information of workers covered under the UCFE program, as they are not included in the State UI administrative database (see section on the Report of Federal Employment and Wages (RFEW)). Fourth, the LMI staff is responsible for conducting the Annual Refiling Survey (ARS). The survey's purpose is to contact one-third of all the establishments each year to verify and update, if necessary, their location (county code), industrial activity, auxiliary status, type of ownership, and both physical location and mailing address information. The updated classification information from the survey is incorporated in the ES-202 program with the data submitted for the first quarter of each year (see section on the ARS form). In addition to the ARS, the BLS regional staff conduct an annual quality assurance review in each SESA for SIC coding for two purposes: 1) to improve the quality and consistency of assigned SIC codes; and, 2) to provide information regarding quality measures of these codes at the national level. The accurate assignment of SIC codes is crucial since most BLS data collection programs derive samples based upon these characteristics.

### III. Data Element Definitions

All of the data elements listed below have the potential to be updated each quarter, since the main source of the data for the BEL is the administrative tax records that the employers file with the SESAs for tax purposes each quarter. Generally, the SIC, geographical and ownership codes are only updated with the data being submitted for the first quarter of each year. Thus, any break in the series caused by an update to one of these data elements will occur between calendar years.

BLS follows the *1987 SIC Manual's* definition of **establishment** (including auxiliary establishments) in its BEL. Thus, an establishment is defined as an economic unit, generally located at a single physical location, where business is conducted or services or industrial operations are performed. If different business activities are conducted at the same site and individual payrolls are maintained, then these units will be reported separately.

An **auxiliary establishment** is primarily engaged in performing management or support services for other establishments of the same enterprise. Some examples of activities commonly performed by

auxiliaries are management and other general administrative functions, such as accounting, data processing, and legal services; research, development, and testing; and warehousing. Each auxiliary establishment is assigned a four-digit SIC code on the basis of the primary activity of the operating establishments that they serve. In addition, they are sub-classified further through the assignment of a one-digit auxiliary code based on the primary activity performed by the auxiliary establishment.

The BEL also contains a **Multi-Establishment Employment Indicator (MEEI)** code which designates whether the employer is one of the following: a) single establishment employer; b) multi-establishment employer reporting their sub-units on the MWR or RFEW (i.e., master record or parent record); c) establishment of a multi-establishment employer; d) sub-unit of a multi-establishment employer that includes more than one establishment for reporting purposes; e) multi-establishment employer reporting as a single unit because it does not meet the minimum criteria for filing the MWR or RFEW; and, f) multi-establishment employer reporting as a single unit because the employer refuses to complete the MWR or RFEW.

Also collected on the BEL are two types of **business names**. The first is the **legal or corporate name** whereas the second is the **trade name**, which is essentially the name that the employer is using to conduct its business. The trade name is also frequently referred to as the "trading as," "known as," or "doing business as" name. In addition to these names, the BEL also collects a **Reporting Unit Description (RUD)** for each establishment of a multi-establishment employer. The purpose of this field is to further delineate these units by using terminology (store or unit number, plant name, etc.,) provided by the employer, and thus familiar to them, to separately identify each of the establishments in the event that only one of their establishments is selected for a survey. In the event that an employer is reporting data for more than one establishment as a combined sub-unit (see item d above), the RUD should provide information on the establishments covered in this sub-unit.

On the Multiple Worksite Report, employers that have employees that move from place to place on a daily or weekly basis (e.g., utility repair persons, sales representatives, etc.) are instructed to report these persons at the worksite to which they are based or

receive supervision. If the sales representatives have no base within the State and operate out of their residences, the employer is instructed to report these employees as "sales representatives-statewide" in the RUD on their MWR.

The BEL does not currently collect information on the **type of legal entity** that the business or employer represents. The EQUI file that the States will provide in early 1998 will capture this information, if it is available from the State's employer master file. The types of legal entities that will be included on the EQUI are 1) Individual; 2) Partnership; 3) Corporation; and, 4) Other (limited partnership, S-corporation, household, etc.).

A number of **addresses** for each establishment are also stored on the BEL. One of these addresses can be described as the **tax mailing address** that is carried on the State's UI employer master file. This is the address that the employer has supplied for the purpose of communicating with the agency for filing UI reports and other correspondence. Two other addresses are obtained by the LMI staff as part of the ARS. The ARS questionnaire asks the respondent if the address to which the form was mailed is the correct **mailing address** for that employer and to update it, if necessary. The employer is also requested to review and update a **physical location address** that was preprinted on the form for each establishment.

#### IV. Measures of Size

With regard to measures of size, the BEL contains values that assist the mathematical statisticians in their selection of establishments for various surveys. Data for each establishment includes an **employment size class code** in addition to the posting of **employment** for all three months of each quarter. The first ten size class values designated by the Office of Management and Budget (OMB) are used in these assignments. Although **wage size class values** are not assigned to each establishment, the BEL does include an **average monthly wage value** for each quarter. **Total Wages** for the establishment is not stored but an approximation for this figure can be derived by multiplying the average employment for the quarter by the average monthly wage for the quarter. This product is then multiplied by three to derive total wages.

**Employment, wage, or establishment count** data are not included on the BEL for a firm at the State level or nationwide; however, it is easily derivable by

extracting the firm's establishments using the State UI account number for a State total or Federal Employer Identification Number (EIN) for a National total. These records would then be summed to determine the values for this firm at the State and National levels. To the extent that a firm has only one EIN, this method has proven quite successful for a nationwide count. It does have some limitations if the firm has more than one EIN at the national level or more than one UI account number at the State level and, all of the EINs or UI account numbers are not known. The BEL does not contain any **corporate links** other than the EIN. Initial work in this area several years ago using the Dun & Bradstreet Corporate Affiliate File looked promising but budget cutbacks forced elimination of these research activities. **Sales volume** data are not collected by any BLS survey nor is this information collected by the SESAs; therefore, the BEL does not store any data on sales or sales volume.

## V. Data Coverage

UI coverage is quite broad and comprehensive covering approximately 98 percent of all non-farm wage and salary workers. All employers who have one or more workers employed on one day in 20 weeks or more in a calendar year or who have a payroll of more than \$ 1,500 in a calendar quarter are required to have their workers covered by UI. Under the Federal Unemployment Tax Act (FUTA), **nonprofit organizations** who employ more than 4 workers during the same time period just cited or who meet the same payroll requirement must also cover their workers. Many States, however, have extended this coverage to the "one or more worker criteria" and thus draw no distinction between profit and nonprofit businesses.

**Agricultural coverage** also has different coverage requirements from nonagricultural employers. The agricultural employers are required to be covered if they employ ten or more workers on one day in 20 weeks in a calendar year or have more than a \$20,000 payroll in a quarter. Some of the more densely populated States have also chosen to extend coverage beyond the Federal minimum, resulting in an agricultural coverage rate of 47 percent.

Private households, social clubs, and college fraternities and sororities which employ **domestic help** and pay wages less than \$1,000 in a quarter are excluded from UI laws.

State and local government employers are also required to cover their workers with only a few minor exceptions. These exceptions are principally **students who work part-time at the colleges or universities that they attend** (approximately 495,000 for State and 40,200 for local); **student nurses and interns** (approximately 17,600 for State and 21,600 for local); **elected officials** (approximately 18,200 for State and 108,700 for local) and **temporary emergency employees** hired to handle disaster relief activities (varies from year to year). Many States also exclude from coverage the spouses of students who work at the university if the employment is part of a program to provide financial assistance to the student. The same exceptions apply to these type of students, nurses and interns owned by private sector interests (approximately 445,000 and 29,200, respectively). Other State and local employees exempted from UI coverage include members of the judiciary; State national and air national guardsmen; and policy and advisory positions.

**Military personnel** are covered under a separate program, Unemployment Compensation for Ex-Service members (UCX) and their counts are not included on the BEL. Civilian defense workers, however, and all other Federal employees are covered under the Unemployment Compensation for Federal Employees (UCFE). These employees are included on the BEL. Exclusions to coverage for Federal workers are minor.

The major **exclusions to coverage** are workers in selected industries, such as **nonprofit organizations** that employ less than 4 workers (approximately 1,358,000 workers). **Railroad workers** (approximately 252,200) are also excluded from UI coverage. These workers are covered under the Railroad Unemployment Insurance Act and their establishment's data are not included on the BEL. Also excluded are workers in **school systems** that are **owned and operated by religious institutions** (approximately 193,200). Selected classes of workers are also excluded from coverage with the major group being **straight commission life insurance agents** (approximately 135,400). These agents are viewed as being **self-employed**, which also represents another class of workers excluded from Federal coverage requirements. Proprietors of the businesses that they own are also excluded from coverage. The total number of self-employed that are not included on the BEL is estimated to be approximately 8.9 million workers. **Unpaid family workers** (approximately

200,000) who work at a business owned by the family are also excluded from coverage.

Employees of firms who provide contract services to other companies and who work on-site at the companies purchasing the services are considered employees of the contractor. The contractor would be requested to file an MWR showing the work locations of their employees and then providing the employment and wages for these locations on the MWR each quarter. Their SIC code would be determined by the type of services that they are providing (e.g. programming or systems analysts services) rather than the SIC of the purchasing company. An exception to this policy is work performed by temporary help agencies. These workers are assigned SIC code 7363—Help Supply Services and are reported geographically from where they are based or supervised.

## **VI. Data Collection**

### **A. Status Determination Form**

All new employers which become subject to UI coverage are required to file a Status Determination Form with the UI unit of the SESA. This form, which varies in content from State to State, is used to determine an employer's tax liability under the State's UI laws and to collect administrative information such as the employer's EIN. The Status Determination Form also requests information on which the classifications for industrial activity, county (township in the New England area), auxiliary (i.e., warehouse, central administrative office, research and development office, etc.), and ownership (private sector, or Federal, State or local government) codes are based. These codes are assigned by the LMI staff. The assignment of the SIC code is based on the establishment's primary economic activity, which is determined by its principal product or group of products produced or distributed, or services rendered. If there is insufficient information on the Status Determination Form, the employer will either be contacted by telephone or mailed the Industry Classification Statement to obtain the necessary information. In a sense, it is the State's equivalent of Social Security's Form SS-4, Application for a Federal EIN.

After employer liability is determined, the SESA UI unit assigns a UI account number. Most new employers are aware of their UI liability and request the SESA to supply a Status Determination Form when they begin their business operations. Some liable employers, usually small ones, fail to file a Status

Determination Form. These employers may be discovered through information on new firms applying for EINs supplied by the Internal Revenue Service (IRS) to the SESA each quarter. Other means of discovering liable employers are through the UI claims process, the UI field auditor investigations, and an initial filing of the QCR without an UI account number.

### **B. Quarterly Contribution Report**

All liable employers are required to file a QCR with the SESAs for their UI accounts. These reports, like the Status Determination Forms, are administered by the UI program and also differ slightly in design for each State. All of the QCR forms, however, request employment values for each month of the quarter and total wages, taxable wages, and UI taxes due for the quarter. This information and the taxes that are due are necessary for the operation of the UI tax system but they are also important for statistical purposes for the ES-202 program. Employers are asked to report, among other items, the total number of covered workers (full and part-time) who earned wages (subject to UI taxes) during the pay period(s) which includes the 12th of each month in the quarter and the total payroll for the quarter. This report is mandatory for employers with a single location as well as employers with multiple locations in the State. The latter group of employers report a summary of these data for all of their establishments covered under the same State UI account on the QCR. Therefore, establishment-level data for these employers do not exist in most State UI administrative files.

### **C. Multiple Worksite Report**

Major changes in the collection of data from multi-establishment employers were introduced with the implementation of the BLS Business Establishment List (BEL) project in 1990 and 1991. Prior to that time, the detail of data for employers operating in more than one location and/or industry in a State was limited. Only those multi-establishment employers who had 50 or more employees in the sum of secondary county locations and/or four-digit industries were asked to provide a breakdown of employment and wages data by county and industry. This information was collected on forms referred to as "Statistical Supplements" to the QCR. A different version of this form existed in each State.

Under the BEL project, multi-establishment employers which have 10 or more employees in the sum of their secondary physical locations and/or industrial activities covered under one UI account, are

requested to provide establishment level data. This more comprehensive disaggregation of multi-establishment accounts has created a database (the BEL) which is almost entirely at the establishment level and thus provides more accurate industrial and geographic information for all establishments. The BEL project instituted the use of a BLS-standardized collection form to gather the additional data from multi-establishment employers. The MWR replaced the State-specific Statistical Supplement forms that the LMI units previously used to collect county-level data from multi-establishment employers.

Data collection procedures for multi-establishment employers differ from those for single units. For multi-establishment employers, the LMI unit of the SESA is responsible for the mail-out, processing, and review of the MWR forms each quarter. As part of this process, multi-establishment employers are asked to verify the business identifying information (trade name, worksite description, and physical location address) for each establishment (worksite) that is pre-printed on the MWR. In addition, the employer is requested to provide the employment for each month (using the proper reference period) and total wages for each worksite for the given quarter. New worksites are manually added to the MWR by employers. The State LMI unit then adds these worksites to their database. When that employer receives the next quarter's MWR, these new worksites will be pre-printed on the MWR along with their other worksites. The situation for deaths or business transfers is handled in the same manner. The employer provides the information on the affected worksite and the LMI staff deletes the worksite from that employer's file. Thus, the MWR captures business births and deaths for these multi-establishment employers on an on-going quarterly basis.

In 1992, BLS began emphasizing the collection of the MWR in a standardized electronic/magnetic media format. By 1995, BLS provided employers located in multiple States with the option of sending their MWR for all States to the Electronic Data Interchange Center (EDIC) located in Chicago, IL. The EDIC receives and edits these data and then forwards them to the appropriate States. This process greatly reduces the employer costs (manually posting the MWR data to the forms) and the State's costs (printing, handling, postage, data entry, etc.). BLS has also been working closely with the companies providing payroll/tax services and those who develop payroll/tax software to have the electronic transmission of the MWR data added to their services or software. Two payroll

software developers have already released these programs to their clients whereas three others are scheduled to provide this option in 1997. Similarly, three payroll/tax provider firms have informed BLS of their intention to add this service in the near future.

#### **D. Report of Federal Employment and Wages**

Federal agencies, whose civilian employees are covered under the separate but comparable UCFE program, do not file QCR forms with State UI programs but instead report employment and wages data directly to the SESA's LMI unit. Since 1993, all States have been using a standardized form, the RFEW, that was developed by BLS to collect these data each quarter. The RFEW was modeled after the MWR to facilitate its use in the State processing systems. Likewise, the Federal agencies providing the RFEW data have been requested to use the electronic transmittal procedure.

#### **E. Annual Refiling Survey**

The purpose of the Annual Refiling Survey is to review and update, if necessary, the classification codes (industrial, geographical, ownership and auxiliary) currently assigned to the establishments stored on the BEL. The survey, conducted by the State LMI units, is initiated in October of each year with approximately one-third of the establishments being reviewed annually. The establishments are selected on the basis of the 7th and 8th digit of their Federal EIN, which are essentially random digits within the nine digit number. This selection process ensures that the industrial distribution of the survey respondents is proportional to the establishments in the economy. In other words, no industrial sector is specifically targeted in any one year.

For an employer currently coded as a single establishment, the ARS questionnaire requests that the respondent review an industrial classification statement. This statement is a general description of the economic activities for that 4 digit SIC code followed by some specific economic index items that comprise the industry. If the statement reflects the establishment's previous 12 month economic activity, then the respondent simply checks the "Yes" box and underlines the relevant wording that applies to their economic activity. If the employer thinks that the description is not correct or is unsure, then they are requested to check the "No" box and provide a description of the economic activities of their business along with an approximate percentage of the sales for each activity listed. The LMI staff then review this information and determine whether the SIC code needs to be updated. The current SIC code may have been

assigned from the Status Determination Form or it may have been updated from a previous ARS questionnaire. In addition to the industrial classification review, the respondent is also requested to review and update the following, or provide the information if it is not preprinted: 1) physical location address; 2) mailing address; and, 3) county in which the establishment is located.

The respondent is also requested to answer a question concerning whether they provide support services for other units of the same enterprise or to the general business community or the public. This information is used to update the auxiliary code, if necessary. The ARS questionnaire also helps identify new multi-establishment employers. Employers are asked if the establishment whose listed physical location address is the only establishment in that State under that UI account number. If no, then the employer is requested to complete the back of the ARS form which provides space to list the physical location address, economic activity and employment for each of the previously unidentified establishments. This information is then reviewed by State staff to determine if the employer should file a MWR form each quarter.

The ARS procedures for known multi-establishment employers is quite similar to that for single establishment employers. The main difference is that the former group will receive a separate 4 digit industry description for each SIC code currently assigned to that employer in that State under that UI account number. The questions on the survey form are identical to those for a single establishment employer described previously. The States are required to achieve a 75 percent usable response rate to the ARS. Most States conduct at least one, and sometimes two, follow-ups to achieve the required response rates.

## **VII.Data Editing Procedures**

### **A. Current Editing Procedures**

Micro data collected on the QCR, MWR, and RFEW are edited by the State LMI staff and corrected, as necessary. The micro data, including imputed values, are then aggregated to the appropriate ES-202 macro-level cells, including the size class break-outs for first quarter. The State LMI unit then edits and updates these macro records, where necessary, using one of the standard data editing systems. Both the micro and macro edits include checks for invalid and inconsistent data as well as checks for large and unusual employment and wages fluctuations between and within quarters.

Every quarter, a relatively small number of employers fail to submit either a QCR, MWR, or RFEW. Others may submit incomplete reports, typically QCRs with missing employment data. Delinquent and missing data notices are sent to these employers, as appropriate. Usually the SESA unit which initially mailed the form is responsible for this follow-up. Therefore, the UI unit generally contacts employers who do not complete the QCR, while the LMI unit pursues delinquent MWRs and RFEWs. The follow-up procedures for delinquent and incomplete reports vary slightly in each State, but every attempt is made to minimize the amount of missing data on the ES-202 file. For those employers who fail to respond to follow-up requests, the data are imputed, generally by employing methods that use historical data for the establishment. This imputation procedure is automatic for one quarter. If the report remains delinquent for a second quarter, the record will be flagged for further review. A State LMI analyst will determine the status of the establishment (i.e., active or inactive). If it is determined that the establishment is still active, the data will be imputed. A new imputation methodology that emphasizes a current quarter industry's trend has been tested with data from several States. This new methodology will be included in a future revision of the States' standardized systems.

After making corrections and adding comments to the summarized file, States submit the macro data on a computer tape or cartridge in the standardized ES-202 Report format to the BLS national office, where it is due approximately five months after the end of the reference quarter. A long-term goal is to ultimately have a four month lag in the receipt date. This process began in FY 1996 but is currently on hold until after the implementation of the EQUI. The earlier submittal of these data would greatly assist the Bureau of Economic Analysis in their estimate of the Gross Domestic Product, and the BLS' Current Employment Statistics program.

Standardized ES-202 systems were developed to improve standardized processing, editing, and imputation methodology in the SESAs; improve the data quality of the BEL; and, control program maintenance costs. Furthermore, this approach permits a more timely introduction of program changes in a more cost effective manner. These standardized ES-202 State processing systems were developed by Utah and Maine staff. These original system requirements were developed by the State staff and approved by BLS. Improvements to the systems are implemented with

new versions. The system requirements for these versions are developed jointly by BLS and the State developers.

## **B. Future Editing Procedures**

BLS data processing systems and the SESA standardized data processing systems have been extensively revised to efficiently edit the micro and macro data and to produce only one file (the EQUI) for submittal to BLS. This file will replace the QUI, the ES-202 Report and the ARS Control File. The EQUI file contains more information on addresses, dates (setup date, initial date of liability, end of liability date, and reactivation date), old SIC codes, new SIC codes, NAICS codes, contributions, taxable wages, and types of coverage.

Two major goals for developing the EQUI file are to achieve consistency between micro and macro files and to ensure that the data are processed using the same editing criteria. All macro data will be system derived from micro data, eliminating independent macro corrections. Micro level corrections which are processed on the State files will be mechanically captured and electronically forwarded to BLS. A correction policy will be in place to ensure that updates to the State files (maximum of three **preliminary** quarters of data and the **current** quarter data) are forwarded to BLS.

Both BLS and State system developers (Utah and Maine staff) use the same edit specifications. As a part of the EQUI deliverable, the States will include information on their State-specific edit parameters and tolerances. These will be used by BLS to ensure that the same standards are used for BLS processing.

## **VIII. Linkages between Establishments and Firms**

Upon receipt each quarter, the records on each State's QUI file are edited by BLS for validity, accuracy and reasonableness. Following a correction process, the records are then matched to the prior quarter's data on the BEL using the Continuous Unit Match System, a multi-stage record linkage process. The first stage of the process matches the records by the "SESA ID," which consists of the appropriate Federal Information Processing Standard (FIPS) State code, the establishment's UI number and the establishment's Reporting Unit Number (RUN). The UI number is the number assigned to that employer by the UI staff to conduct its business with the SESA in that State. It is the State equivalent of the Federal EIN. The RUN is the number that is assigned by the State LMI staff to

make each establishment unique within the same UI number. The RUNs in the current BEL are three-digit alpha/numeric codes but will be converted to a five-digit numeric code with the submittal of the new EQUI file beginning in February, 1998.

The SESA ID component matches approximately 95.2 percent of the QUI records to a record on the BEL in the fourth quarter to first quarter match. Matching the other three quarters results in approximately 96.5 percent of the records matching based on the SESA ID. The resulting non-matches (potential births and deaths) are then processed through the second component which screens the predecessor and successor fields provided by the States on the QUI tapes. These relationships are examined and records that are matched using this process will have their Universe Data Base (UDB) number transferred from the old SESA ID to the new SESA ID. A **UDB number** is assigned to each record on the BEL and is the only means of tracking an establishment that is sold from one business to another on the BEL. Most States require that new UI numbers be assigned when business transfers occur. Between 7 and 11 percent of the potential QUI birth records match on predecessor codes, with more matching in the first quarter. This accounts for 17,000 to 40,000 matches each quarter. Some of these matches are legal changes because they occur at the beginning of the calendar year.

After the predecessor/successor relationships provided by the States have been linked, the remaining non-matched records enter the third stage of the record linkage process--the weighted matching portion of the system. This component is a statistical matching technique that assigns point values to matches on various characteristics of the non-matched units. For example, if an apparent establishment birth has certain characteristics that are very similar to an apparent establishment death, can it be assumed that it is the same unit? The matched characteristics include the physical location address, four digit SIC code, county (FIPS) code, telephone number, trade name, etc. A point value is assigned for each matching characteristic and if the total point value for a match meets or exceeds the minimum requirement pre-selected for a "true match," then the system assumes that there is a predecessor/successor relationship. The UDB number is then transferred from the death unit to the apparent birth unit, which now becomes a transfer of ownership record, rather than a business birth.



The final stage of the record linkage process attempts to link units that have undergone changes within the most recent quarter. The remaining records are matched using predecessor codes and the weighted matching routine. One requirement for units to link at this stage is that predecessors have zero employment in the third month and successors have zero employment in the first month. Between 11,000--16,000 matches are recorded from the weighted and within-quarter match processes each quarter with more discovered in the first to fourth quarter match. These matches account for 4.5 to 5.5 percent of potential births on the QUI file. The records for the current quarter that are not matched as a result of this matching effort are then determined to be "true births" and are assigned a new UDB number. The records for the prior quarter that did not match a record for the current quarter are considered deaths.

The Continuous Unit Match System is currently in the process of being revised to address the additional needs of the Bureau's new Establishment Longitudinal Database Program. The new system uses purchased software that standardizes the trade names and addresses of the establishments to be matched. Another purchased software package (AUTOMATCH) will be used to perform the actual matching. This software allows more flexibility in selecting sorting criteria used in the weighted match process. A research paper on this subject was recently presented by BLS staff at the 1997 Record Linkage Workshop. More details are presented in that paper titled "Improvements in Record Linkage Processes for the Bureau of Labor Statistics' Business Establishment List." The only linkage that exists between establishments of the same firm in the BEL are the UI account number within the State and the Federal EIN when examining the same firm across the States. To the degree that firms have more than one Federal EIN, the BEL has no corporate affiliate linking mechanism.

#### **IX. Timing of Entry of New Establishments and Exit of Dead Establishments**

Prior to 1989 BLS only required the States to provide the QUI data for the first quarter of the year. In May 1987, the Economic Policy Council recommended to OMB that BLS be given additional funding to collect these data each quarter and to move the data collection from a county/industry level to the establishment level. Congress provided the funds and from that point forward, BLS has been collecting these data each quarter. The advantages to this data collection are notable. Analyzing **quarterly** data

reveals more of the dynamics of the U.S. economy as opposed to **annual** data analysis. For instance, quarterly data analysis measures between 200,000 and 300,000 new establishments. Annual data analysis would mask some of these changes between quarters. Thus, collecting data on births and deaths as well as changes in the other continuing business establishments on a quarterly basis is superior to an annual update. For survey sampling purposes, new establishments that opened during the second quarter of the year at least have a chance of being selected as opposed to no chance if the sample frame is only updated once a year using data for the first quarter. The converse of this situation is also true in that units that died would still appear to be active if the sample frame is only updated for the first quarter of the year. This situation could result in closed establishments being selected for the survey because the frame did not receive quarterly updates.

A number of BLS survey programs are now drawing their sample from the BEL, conducting their sample refinement procedures, then reviewing the next quarter's data to determine if any of the establishments selected have gone out of business. If that is true, then the supplemental sample establishments are immediately used. This procedure reduces survey costs and leads to higher initial response rates.

It would be difficult to measure the percentage of new employers that enter the BEL in the same quarter that the business actually started operations. It is important to note that the term "employers" was used rather than "establishments." The primary reason is that almost 100 percent of new establishments where employers are submitting a MWR will be recorded in the same quarter as they began operations. Delinquent MWR respondents would cause this to be less than 100 percent. As stated earlier, BLS is dependent upon the SESAs to capture the new employers. As part of their FUTA oversight responsibilities, the Employment and Training Administration has required the States to meet certain performance measures. The most appropriate measure for this activity requires that a minimum of 80 percent of status determinations of employer liability (new and successor employers) should be made within 180 days of the initial liability date. For the last measurement period available, 48 States met or exceeded this requirement with the national average equal to 86.2 percent.

#### **X. Limitations**

Some limitations of the data and areas that need improvement have been noted since BEL's inception. One of these areas is the need for better identification of predecessor/successor information in the States. State laws and practices regarding legal successors vary. Consequently, some establishments that would be designated as predecessor/successor relationships for statistical purposes are not identified as such for either legal reasons or the States regard this activity with less of a priority. The new matching system should help reduce the number of false births and deaths.

An area that needs additional attention is the inconsistency in SIC coding among the States for businesses that conduct the same economic activity in different States. The current decentralization of SIC coding activities can lead to these situations. A review of these situations is scheduled to begin in 1998.

Another opportunity for improvement is the lack of a formal survey feedback system. Users of the BEL as a sampling frame do not have a systematic procedure to transmit potential updates to the establishment data that were discovered during their survey or sample refinement activities. Prior discussions on this issue focused on the confidentiality aspect of providing this information to the States. Each survey would need to specify the potential sharing of these data (i.e., updates) with the States in future OMB clearances. The States would need these data to update their files for subsequent transmittals each quarter. Regardless of these issues, a formal mechanism should be established to utilize the additional information that is gleaned from the surveys.

## **XI. Summary**

The BEL is the most complete and timely source of business establishment data. The primary data source of the BEL is the State UI administrative records. These quarterly data are supplemented with multi-worksites information collected on the MWR and the Federal data collected on the RFEW. The industrial and geographic information is reviewed and updated, as necessary, every third year. Quality assurance programs are conducted to measure the accuracy of the State industrial coding. In addition, standardized State operating systems have been exported to the States to improve standardized processing, editing, and imputation methodology in the SESAs; improve the data quality of the BEL; and, control program maintenance costs. Beginning with the third quarter 1997 submittal of the first EQUI file, the BEL will

contain more data elements; be processed more efficiently; and, be edited in a more efficient manner.

**Note: Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.**

## **Acknowledgments**

We would like to thank Timothy R. Pivetz for his assistance in preparing this paper.

## **References**

EXECUTIVE OFFICE OF THE PRESIDENT AND OFFICE OF MANAGEMENT AND BUDGET (1987), Standard Industrial Classification Manual 1987, Springfield, Virginia,: National Technical Information Service.

FARMER, TRACY E. AND MICHAEL A. SEARSON (1995), "Use of Administrative Records in the Bureau of Labor Statistics' Covered Employment and Wages (ES-202) Program," 1995 Bureau of the Census Annual Research Conference, Washington, D.C., March 1995.

SEARSON, MICHAEL A. AND JOHN PINKOS (1990), "The Bureau of Labor Statistics' Business Establishment List Improvement Project," 1990 Bureau of the Census Annual Research Conference, Washington, D.C., March 1990.

U.S. DEPARTMENT OF LABOR, BUREAU OF LABOR STATISTICS (1992), "ES-202 Operating Manual," Employment Security Manual, Washington, D.C.: U.S. Department of Labor. (Includes revisions from BLS ES-202 Program Technical Memorandums from 1993 to 1996).

U.S. DEPARTMENT OF LABOR ETA (1997), Comparison of State Unemployment Insurance Laws, Washington, D.C.: U.S. Department of Labor.

U.S. DEPARTMENT OF LABOR ETA (1995), UCFE Instructions for Federal Agencies, Washington, D.C.: U.S. Government Printing Office.