

Measuring Quality in the BLS Business Register

Richard L. Clayton

David M. Talan

U.S. Bureau of Labor Statistics

Joint UNECE/OECD/Eurostat Meeting of the

Group of Experts on Business Registers

3-4 September 2013, Geneva

The views expressed in this paper are solely those of the authors and do not necessarily reflect the official positions or policies of the US Bureau of Labor Statistics or the views of other staff members.

Motivation

One of the most important concerns involving large datasets is how the integrity of the data is maintained. The U.S. Bureau of Labor Statistics (BLS) maintains the QCEW Business Register which serves as the Bureau's establishment-based sampling frame. The QCEW Business Register is housed in a relational database of 9.2 million business establishments linked longitudinally from 1990 through 2012 and based on the microdata submitted quarterly by States from Unemployment Insurance (UI) tax files. Data elements on these files include information on monthly employment, quarterly wages, business name and addresses, industry classification, geocodes, and other administrative data. Every business establishment contains a unique identifier that allows it to be tracked at the micro-level across time. The Business Register has three critical functions: (1) serving as a sampling frame for establishment-based surveys, (2) producing longitudinal Business Employment Dynamics (BED) statistics, and (3) serving as an important resource for labor market research.

Because these data are used to generate high quality, high frequency, timely, and historically consistent data on businesses and employment, BLS uses measurement methods to quantify the quality of these data. This paper demonstrates how BLS measures the quality of its business register based on the dimensions of relevance, accuracy, timeliness, and accessibility to meet the needs of a wide range of different users.

Objectives

This paper describes data quality objectives in four sections. The first section (1) details the inputs used to create the QCEW Business Register and the review process used to remove errors from the input data. The second section (2) describes the construction of the QCEW Business Register. The third section (3) describes the metrics used to measure data quality and describes how quality controls are maintained through management directives and oversight. Finally, the fourth section (4) describes the uses and users of these data.

Investment in Data Quality

Since QCEW data are based in large part on administrative data, a significant investment in data quality is deployed to ensure that accurate employment, wages, industry, geography, and other information are reported each quarter for every business establishment in the United States. This investment includes: (1) the development of two separate BLS data collection surveys to supplement the administrative data, (2) the development of advanced computer processing systems to process, edit, and publish accurate employment and wage data at the detailed national, state and county level each quarter, and (3) the deployment of a federal/state management structure and staff at state, regional, and national offices to continuously improve the administrative data by setting quality goals and standards, and identifying, correcting, and preventing errors in the source data.

Section I: Inputs of the QCEW Business Register

The QCEW Business Register is a list of active employer business establishments in the United States, Puerto Rico, and the Virgin Islands. Its principal sources of information are the mandatory quarterly reports filed by all employers covered by the Unemployment Insurance system of the fifty States and the District of Columbia (and Puerto Rico and the Virgin Islands). Employers report to their State Employment Security Agencies (SESAs) in compliance with State Unemployment Insurance (UI) laws, and for Federal civilian workers, in compliance with the Unemployment Compensation for Federal Employee (UCFE) program. Each quarter, business and government employers report monthly employment and quarterly wages.

The QCEW Business Register contains data collected from four forms required to meet program needs. Two of these are state Unemployment Insurance based forms: the first is the Status Determination Form (SDF) and the second is the Quarterly Contribution Report (QCR). In addition, two BLS supplemental collections were developed to enhance the administrative data to meet the needs of the QCEW program. The first of the BLS forms is the Annual Refiling Survey (ARS) and the second is the Multiple Worksite Report (MWR). The purposes of these forms will be explained in more detail later in this section.

Maintenance Workload

To gain some perspective of the size and workload associated with maintaining the QCEW Business Register, there are currently 9.2 million worksites representing 133 million employees reported by the fifty states, the District of Columbia, Puerto Rico, and the Virgin Islands as of the fourth quarter of 2012.

Approximately 7.5 million worksites were reported as being single establishment employers, whose employment and wages data were collected from the state QCRs. The employment and wages data and other business identification information for the remaining 1.9 million worksites were collected from the MWRs.

There are approximately 7 million legal entities with only about 136,000 of those providing the MWR data. The term legal entity is used here since many large employers have accounts in more than one state and are thus counted more than once. The MWR employers only represent about 1.4% of total employers, but they constitute 17% of the total number of worksites and a stunning 41% of the Nations' employment. Thus, without the MWR, we could not accurately measure data about establishment births and deaths.

Status Determination Form (SDF)

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 (or screen for respondents who report via the web) is used to determine an employer's tax liability under the State's UI laws and

to collect administrative information. The form includes basic identification information, including business name, mailing and physical location address, type of organization, Federal Employer Identification Number (EIN), and more. The Status Determination Form also requests information to identify the establishment's industrial activity (NAICS) code, county code (or township in the New England area), and the ownership code (private sector, or Federal, State or local government). These codes are assigned by the SESA staff. The assignment of the industry 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 an Industry Verification Form 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.

Quarterly Contribution Report (QCR)

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. All of the QCR forms 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 source of statistical data. 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 reports a summary of these data for all of their establishments covered under the same State UI account on the QCR.

Multiple Worksite Report (MWR)

Business enterprises with more than one establishment in a State under a single UI account file a Multiple Worksite Report (MWR) so that data for each of its establishments is reported separately. The MWR is mandatory in 26 states and compliance in the voluntary states is also very high. The EIN provides linkages among establishments of the same business enterprise across States. The EIN for establishments on the QCEW Business Register is obtained from the

initial Status Determination Form and updated, if necessary, based on the quarterly UI tax form. This more comprehensive disaggregation of multi-establishment accounts is available in the QCEW Business Register which is almost entirely at the establishment level and thus provides more accurate industrial and geographic information for all establishments.

The MWR is collected each quarter to disaggregate the employment and wages of numerous establishments owned by an employer that are reporting under the same UI account number in a State. The summary information for this employer is reported on the State's QCR. This procedure allows the employment and wages for each worksite to be placed in their correct industrial and geographical category. Also collected are the physical location address of each worksite, a worksite description (normally a store or unit number or other information meaningful to the employer), and various other business identification information. Since UI coverage is virtually 100 percent in all industries, the employment and wages from this program represent a virtual census of the Nation's employment and wages each quarter.

Data collection procedures for multi-establishment employers differ from those for single units. For multi-establishment employers, the SESA unit 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 added to the MWR by employers. The SESA unit then adds these worksites to their database. Thus, new births and deaths and their associated employment are identified each quarter in a timely manner. This rapid identification of births and deaths improves the sample frame. When an 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 SESA 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.

Most employers file paper MWRs either by inserting the monthly employment and wages in the appropriate boxes on the form next to the preprinted information (trade name, physical location address, and worksite description) for each of their worksites or they provide a computer printed facsimile of these MWR data.

MWR Collection via Contractor

To minimize costs and manual workload associated with printing and mailing MWR forms by the individual SESA units, a centralized printing, mailing and processing was initiated by QCEW in 2011. This process is conducted by a contractor, who receives a print file from QCEW and

prints and mails the MWR forms on behalf of the states. The forms are returned to the contractor's processing facility where they are opened, sorted, and scanned. The employment and wages from the forms are read and provided to states on a file that they can load directly into their systems, eliminating the need for manual data entry. The form images are provided to the state electronically, so that states can capture and additional information supplied by the respondent, such as address changes or location openings and closings.

The contractor collection of MWR has been successful and well received by the participating SESA units. As of 2013, 14 SESA units are participating in the process and more are expected to join as the program becomes more efficient and costs are further reduced.

MWR Web Collection

In an effort to make MWR reporting more efficient and cut down on cost, employers can report quarterly worksite data via the Internet rather than paper, using the Bureau of Labor Statistics' secure data collection website. The number of employers using the web-based system for filing the Multiple Worksite Report has increased tremendously since its development in 2007.

Employers can report all states' information for multiple worksites at one time, and BLS sends the information to the respective state agencies. Employers with multiple unemployment insurance (UI) accounts in one state, or multiple worksites in several states, can also use the web-based system.

Once an employer chooses the web-reporting option, they no longer receive paper forms to complete. Rather, BLS sends a reminder via e-mail stating that data collection for the next quarter has started. A few days before the official due date, employers who have not entered data for all of their worksites will receive an e-mail reminder that the due date is approaching.

The MWR web-based system has the capability to display the previous quarter's employment and wages, performs edits of employment and wages, prompts the respondent for new worksites, and advises the respondent of missing or incomplete data for a particular worksite, thereby enhancing data quality. In addition, the system can save partially completed reports. For example, if a respondent is interrupted by a telephone call or meeting while in the process of filing a report, even if the system closes down, the previously-entered data will be saved.

During fourth quarter 2012, there were 38,394 employers who provided their data via MWR Web. Data collected from respondents accounted for 13 million employees (25 percent of total employment) representing 381,536 individual worksite locations (23 percent of total worksites).

MWR Collection via Electronic Data Interchange (EDI)

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 continues to work closely with the companies providing payroll/tax services and those who develop payroll/tax software who have added the electronic transmission of the MWR data to their services or software.

On February 22, 1995, BLS opened an Electronic Data Interchange (EDI) Collection Center in Chicago, Illinois. The EDI center was initially created to collect data in a timely and cost-effective manner for the Current Employment Statistics (CES) survey. In 1996, the MWR was added to the EDI Center's list of data that could be collected from employers. EDI, as used in BLS, is the transmission of data electronically from a respondent to BLS. BLS provides the employer with a standard data transmission format that is used to process the data at the EDI center.

Once the standard format is created, the recurring data files can be provided to the EDI Center quickly, accurately and at less burden for the employer. Thus EDI can dramatically reduce ongoing respondent burden, particularly for the largest, multi-establishment firms, which are requested to participate in most surveys.

The EDI Collection Center opened two new doors of opportunity for firms. First, the EDI Collection Center offers large multi-State firms the option to transmit their paper data for all States to one facility rather than potentially fifty different sites. The EDI Collection Center edits and distributes these data to the respective States. The responses to this new reporting option were very enthusiastic. The second opportunity was the ability to transmit these data electronically, rather than using a magnetic medium-tapes, cartridges, or diskettes. Both of these opportunities demonstrate a reduction in the burden on the part of the employer. The employer would no longer have to receive the MWR forms each quarter, prepare the information that would be manually posted on the MWR forms, and then mail the appropriate information to the different States.

Firms obviously incur some start-up costs to develop the initial data file, but once that is developed the cost savings and reduction in respondent burden are substantial. This collection procedure also has enormous cost savings in the following areas:

- (1) Less postage required for mailing MWR forms to firms and its subsequent return
- (2) Less staff required to process the MWR forms (opening the mail, sorting the forms, etc.)
- (3) Less data entry of the MWRs
- (4) Reduction in the errors caused by posting errors on the employer end or data entry errors

During fourth quarter 2012, there were 14,009 employers who provided their data to the EDI center. Data collected from respondents accounted for 12 million employees (21 percent of total employment) representing 397,155 individual worksite locations (24 percent of total worksites).

Electronic MWR collection has steadily grown over time. While 61 percent of respondents report their MWR via paper, electronic MWR data reporting continues to steadily increase since its inception in 2007. As of fourth quarter 2012, 39 percent of MWRs are reported electronically. Of

those, 10 percent were collected through the EDI Center and 29 percent through the MWR website.

Report of Federal Employment and Wages (RFEW)

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 unit. Since 1993, all States have been using a standardized form, the RFEW, which 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. The coverage for this area is thorough: 99.5 percent of federal employment is collected each month and used in both the QCEW program and the Current Employment Statistics (CES) program. A vast majority of Federal agencies report their MWR data electronically via EDI. The RFEW forms are used for the remaining few agencies that are not currently providing data electronically.

Annual Refiling Survey (ARS)

The purpose of the BLS Annual Refiling Survey is to review and, if necessary, update the classification codes (industrial, geographical, and ownership) currently assigned to the establishments stored on the QCEW Business Register. The survey is initiated in October of each year with approximately one-third of the establishments in the QCEW Business Register being reviewed annually. The establishments are selected at random based on the Federal EIN range. 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 brief but specific description of the economic activities for that 6-digit NAICS code followed by some specific economic index items that comprise the industry. If the statement reflects the establishment's main economic activity for the previous 12 months, the respondent simply checks the "Yes" box. If the employer thinks that the description is not correct or is unsure, they are requested to check the "No" box and provide a description of the main economic activities of their business along with an approximate percentage of the sales or revenue for each activity listed. The SESA staff review this information and determine whether the NAICS code needs to be updated. The current NAICS 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, township, island, or parish in which the establishment is located.

ARS Collection via Contractor

In an effort to reduce costs and manual workload associated with printing and mailing ARS forms by the individual SESA units, a centralized printing, mailing and processing was initiated by QCEW. This process is conducted by a contractor, who receives a print file from QCEW and prints and mails the MWR forms on behalf of the states. The forms are returned to the contractor's processing facility where they are opened and sorted into groups based on the information provided by respondent. The forms are provided to the states, bundled by change type, to facilitate ease of review and data entry. The contractor collection of ARS has been successful and well received by the participating SESA units. As of 2013, 48 SESA units are participating in the process.

ARS Web Collection

Due to the continuing growth of the number of establishments on the QCEW Business Register and increasing postage rates, a more cost efficient method of collecting ARS data was developed. There were three major objectives for the design of the ARS Web system. First was to reduce the overall cost. This includes directing ARS respondents to the most cost effective response method for their situation: Touchtone Data Entry (described below) for respondents with no changes and web or paper for respondents with changes to report. Second, since the Office of Management and Budget estimates an average time of completion for the ARS form of five minutes, the ARS Web system average time of completion is comparable or reduced. The third objective was to improve the quality of the data received. This was achieved with live data edits that require respondents to answer each question on the form and providing respondents additional information about industry codes. Additionally, with approximately 1.4 million the ARS surveys being mailed annually, it was critical to create a simple and straightforward web system to facilitate respondent ease of reporting. Large scale confusion and respondent questions about web reporting would require a costly helpdesk, which would undermine the first goal of reducing survey costs.

During the first year of implementation (September 2012-August 2013), approximately 180,000 responses were received on the ARS Web system. About 84% of ARS web responses came back with some change to industry or geographic code, which met the goal of promoting the most cost effective response method. QCEW will continue to expand on the ARS Web program as a means of reducing costs and increasing data quality.

Touchtone Data Entry (TDE)

Touchtone Data Entry or "TDE" refers to the system that provides users the convenience of calling a toll-free number to access the system any time of the day or night, seven days a week. The call activates an interview session where the respondent is prompted by a pre-recorded series of questions. The respondent first identifies their Unemployment Insurance Number and State by pressing touchtone phone buttons. Each answer is then read back for respondent verification. Upon verification, the system asks the caller a number of questions to verify that the information

provided to the respondent on the pre-printed ARS form is all correct. Upon verifying all applicable questions, the respondent is directed to simply hang up and the verified information is stored and provided to the states.

Advantages of Touchtone Data Entry (TDE)

TDE is intended to reduce data collection costs and is a cost-effective alternative to mail. Rising labor costs and postal rates, and declining telephone and technology costs offer incentives to redesign collection processes.

It is quite true that the start-up costs for implementing a TDE system are relatively high. The cost of hardware, software, communications infrastructure, reliable backups, and labor must be borne, but the operational cost over the life of the system is marginal. The savings realized by process automation alone may justify the adoption of a touchtone system. These savings are in terms of cost as well as efficiency and accuracy.

For the federal government, the savings are particularly high. With the U.S. Postal Service having to quarantine mail addressed to government facilities, touchtone reduced both the data collection time and the quantity of mail handled by government agencies. Waiting for re-delivery of quarantined mail created a dilemma for both the sender and recipient. Also, because the government pays postage for survey forms mailed back by respondents, the cost savings is substantial.

In FY 2002 BLS returned to using a Touch-tone Data Entry system for the ARS. It had been tested in a few States in the late 1990s with mixed results. The TDE system was not suited to address the needs of the conversion from the Standard Classification System (SIC) to the North American Industry Classification System (NAICS) that began in October 1998. When reintroduced in October 2001 as a test in five States, the initial approach was modified so that only those employers meeting certain criteria were allowed to participate. The project's name was also changed to the "Touch-Tone Response System" (TRS). This selectivity feature raised the response rates and also reduced the number of phone calls requesting clarification updates. In 2003 the TRS system expanded to all states including the District of Columbia, Puerto Rico, and the Virgin Islands with cost savings of around \$200,000 per collection cycle. In 2013, of the nearly 1,291,550 respondents eligible to use the TRS system, around 350,000, or 27 percent of respondents used this system for the Annual Refile Survey.

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 are quite similar to that for single establishment employers. The main difference is that the former group will receive a separate 6-digit industry description for each NAICS 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 an 80 percent usable response rate for employment. Most States conduct at least one, and sometimes two, follow-ups to achieve the required response rates. Also, telephone calls are used to gain more information from the respondent.

Editing Procedures

Micro data collected on the QCR, MWR, and RFEW are edited by the SESA staff and corrected, as necessary. The micro data, including imputed values, are then aggregated to the appropriate macro-level cells. The SESA unit then edits and updates these macro records. There are 150 separate edits designed to detect a wide range of invalid and inconsistent values. These edits have been refined and enhanced over time reflecting the many years of data editing experience of state, regional and national office staff. 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 SESA unit pursues delinquent MWRs and RFEWs. 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 SESA 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. Imputation rates are very low and have been trending down in the QCEW program. Since 2001, the imputation rate for employment has steadily declined from 6 percent to under 3 percent in 2012.

After making corrections and adding comments to the summarized file, States submit the micro data to the BLS national office, using various delivery methods. Most states transmit their data electronically to BLS, while a few submit their files via other electronic media such as tape cartridges or CDs. Employers are required by law to file their UI taxes 30 days after the end of the quarter. These data are then extracted by QCEW state staff where they spend around 2.5 months editing and reviewing the files. The files are due approximately four months after the end of the reference quarter.

Standardized processing systems were developed to ensure standardized and consistent processing, editing, and imputation methodology in the states, improve the data quality of the QCEW Business Register, and control program maintenance costs. This approach permits a

timely introduction of program changes in a more cost effective manner. These original system requirements were developed by the State staff and approved by BLS. The system requirements for these versions are developed jointly by BLS and the State developers.

Section II: Construction of the QCEW Business Register

In order to assure accurate linkages between new and existing establishments, there are two components of QCEW Business Register matching process: automated matching and analyst matching. The computer automated matching links approximately 97 percent of all establishments each quarter through a process which links State Employment Security Agencies' identification numbers (SESA-ID). Less than one percent of records are linked by statistical weighted matching or analyst matching. Every time a link is made, an identifier is assigned to the link to distinguish which type of match was made such as a breakout or consolidation.

Identifying Births and Deaths

The QCEW Business Register classifies establishment relationships into births and deaths, breakouts and consolidations, and one-to-one matches. Births are new establishments that are units that came into existence after the creation of the QCEW Business Register universe, while deaths are establishments that have gone out of business, or have had four consecutive quarters of zero employment. Breakouts are records that spin-off employment to a new reporting unit. The original unit can do one of two things: transfer employment and close, or transfer employment and stay open. Consolidations are records that close establishments and merge employment from many units to one or more units. In this instance, like the breakout scenario, the original unit can do one of two things: transfer employment and close, or transfer employment and stay open. Many times, a transfer of employment may be directly from one establishment to another and that is called a one-to-one match.

The QCEW Business Register captures approximately 200,000 new establishment births and deaths every quarter. By collecting timely data on births and deaths, a number of BLS survey programs are able to include these establishments in their sample, thus making them more representative of the universe. Since these new births and deaths are available immediately, this reduces survey costs and leads to higher initial response rates.

SESA ID Matching

Ninety seven percent of all BLS establishment records are linked through a process by State Employment Security Agencies' identification numbers (SESA-ID). The SESA-ID is the establishment's unique identifier that the State Employment Security Agencies transmit to BLS, and the QCEW Business Register initially matches all units with identical SESA IDs; these matches are continuous establishments from the previous quarter. SESA ID matches take precedence over any other type of matching. After identifying continuous SESA IDs, the linkage system identifies four additional linkage steps: predecessor linkages, successor linkages,

breakout identifications, and consolidation identifications. These linkages are made by matching previous quarter SESA IDs to the current quarter SESA ID. After the match is made the system flags both records and copies the QCEW Business Register number, date of first positive employment (DFPE), and date of last positive employment (DLPE) from the previous quarter to the current quarter record.

Probability Weighted Matching

Probability weighted matching is a process by which related units are linked based on the similarity of various characteristics between two records. Pairs of records having enough in common are scored to exceed a specific point value and are then identified as valid matches. This point value is called a cutoff weight, and do not vary upon State, industry, or other characteristics, and create a consistent measure across the nation.

Analyst Manual Matching

As a final quality control measure of QCEW Business Register data, each quarter an analyst reviews data that contain records that were not matched by the automated process. The reasons for this additional review are that data elements may be miscoded, whereby inaccurate information is placed in system identifiers that the automated processes targets and therefore the system cannot make a linkage. Another reason for the supplementary analyst review is that certain records should not be linked, even though the system identifies a weighted match, and should be added to the QCEW Business Register as an establishment birth or removed as an establishment death. The analyst manual review is essential to maintaining proper linkages and to preserve the high data quality produced by the QCEW Business Register.

Section III: The Quality Metrics of the QCEW Business Register

QCEW Business Register metrics are based on statistical quality dimensions of relevance, accuracy, timeliness, interpretability, coherence, and accessibility in order to insure high quality data that meets the varied requirements of different user groups. Since the QCEW Business Register is the foundation for the Bureau's employment statistics, it is essential that accurate linkages are made between establishments to preserve data integrity.

Management Review Mechanisms

The QCEW Business Register has a comprehensive management hierarchy to ensure that data quality permeates all levels of the QCEW structure. One of the tools used in the management structure is the BLS Co-operative agreement which is a contract that sets quality standards for each State Workforce Agency. The co-operative agreement sets goals for each State which is tied to QCEW fund allocation.

Since the inputs to the QCEW Business Register are derived from the Federal-State cooperative, it is vital that communication between and among States and the BLS saturates the system to

promote coordination and cooperation to increase data quality. Frequent meetings for State and Regional offices are management coordination tools which allow these branches of the QCEW program to communicate questions, concerns, and standards in data quality. Regional offices ensure data quality by holding states accountable to data quality standards, providing them with training opportunities, and coordinating program directives. States also offer another management layer to ensure data quality. Each state has their own organizational structure with its own set of data users and stakeholders.

In order to coordinate all agencies involved in the QCEW data production process, the QCEW policy council was created to provide a hierarchical structure for the program. The QCEW policy council is comprised of six representatives from BLS and ten State representatives. Their mission is to prioritize and coordinate QCEW improvement projects. In addition, the policy council seeks the input of all States in establishing policy.

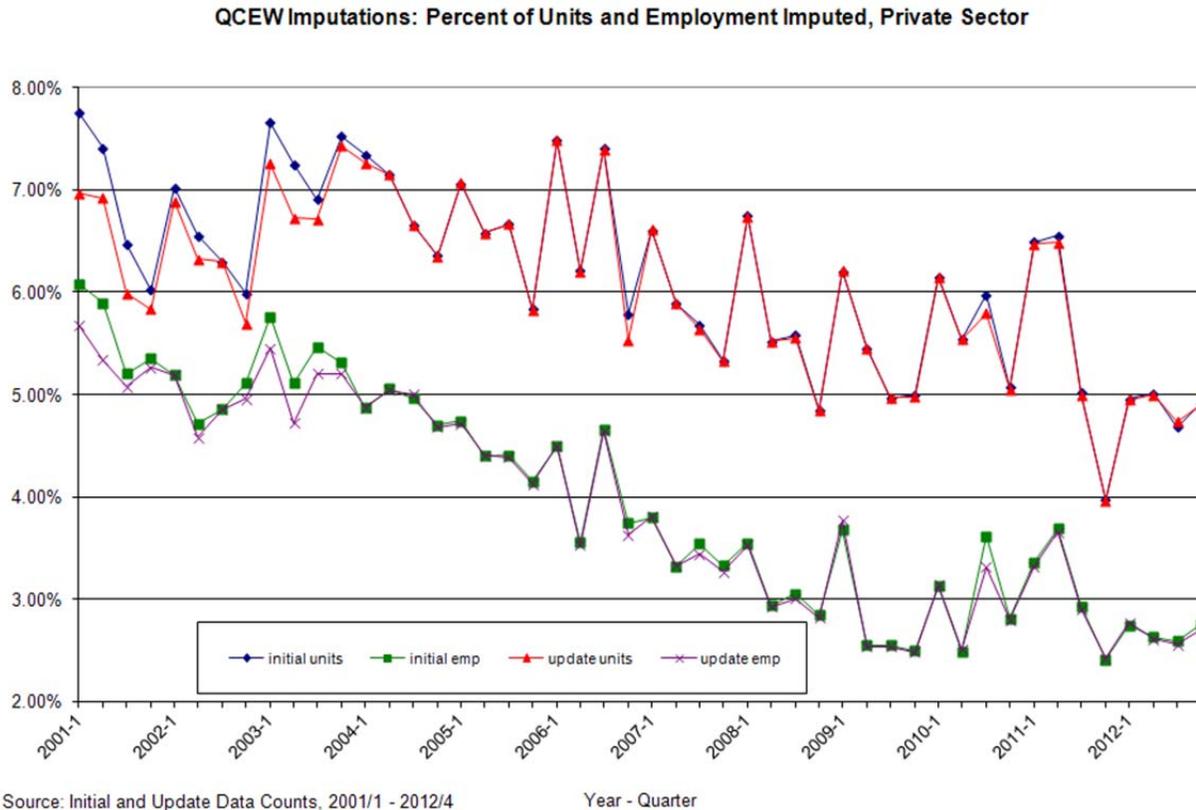
Review Mechanisms of QCEW Business Register Systems

There are two quality control documents which are produced to review the Business Register process and systems: the Flash Report and the Business Register Metrics Report.

Flash Report

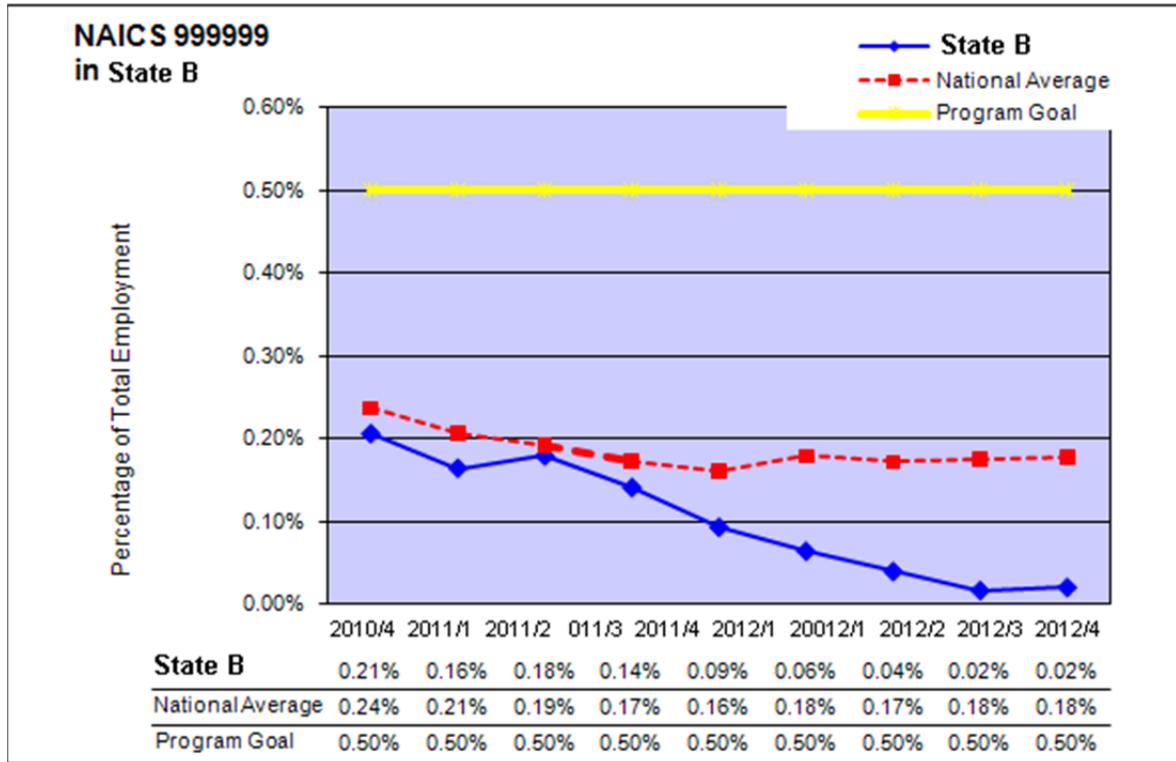
The Flash Report monitors the data quality received from the States. This includes the monitoring of the number of reported units, the number of imputations, the number of prorations, the number of missing units, and the number of records with invalid county and NAICS codes. The Flash Report is a comprehensive management tool that ensures that data quality goals are being met, flags potential data quality problems, and monitors the progress of each State. The Flash Report combines information from many sources on the number and percentage of imputed records, prorated records, and missing records. If there are large fluctuations, increases or decreases in these numbers that might affect data quality, they are flagged and investigated. As seen in Chart 1, there has been a continuous improvement over the last decade in the amount of reported data from employers each quarter. The percent of reported employment increased from 94 percent in 2001 to 97 percent in 2012.

Chart 1. Continuous Improvement: Imputed, Prorated, and Missing Units as a Percentage of Reported Units: US National Average



Since the QCEW Business Register not only contains employment and wage data, but administrative data such as county and NAICS codes, it is essential that the quality of these fields is monitored because it is used as a sampling frame. These items are included in the Flash Report and allow QCEW management to watch trends in data quality. On occasion data is provided to the national BLS office from states with fields not populated, or with unclassified data. It is imperative that management monitor the quantity of these records in order to control data quality. An example of the use of management tools is shown in Figure 1 on the next page. In Figure 1, State B actively monitors their records with unclassified industry codes (NAICS 999999) and has a percentage of these records well below the national average. With proper monitoring of data quality fields, data quality issues are brought to QCEW management and resolved in a timely fashion. Through management administration, any upward trends will be noticed and actions will be taken to improve these particular fields. The graphics displayed above are not only included in the Flash Report, but are available historically to management at the national, regional, and State level. This allows management involved in the data production process to monitor data quality.

Figure 1. Percentage of Unclassified NAICS in State B in Comparison to the National Average of Unclassified NAICS

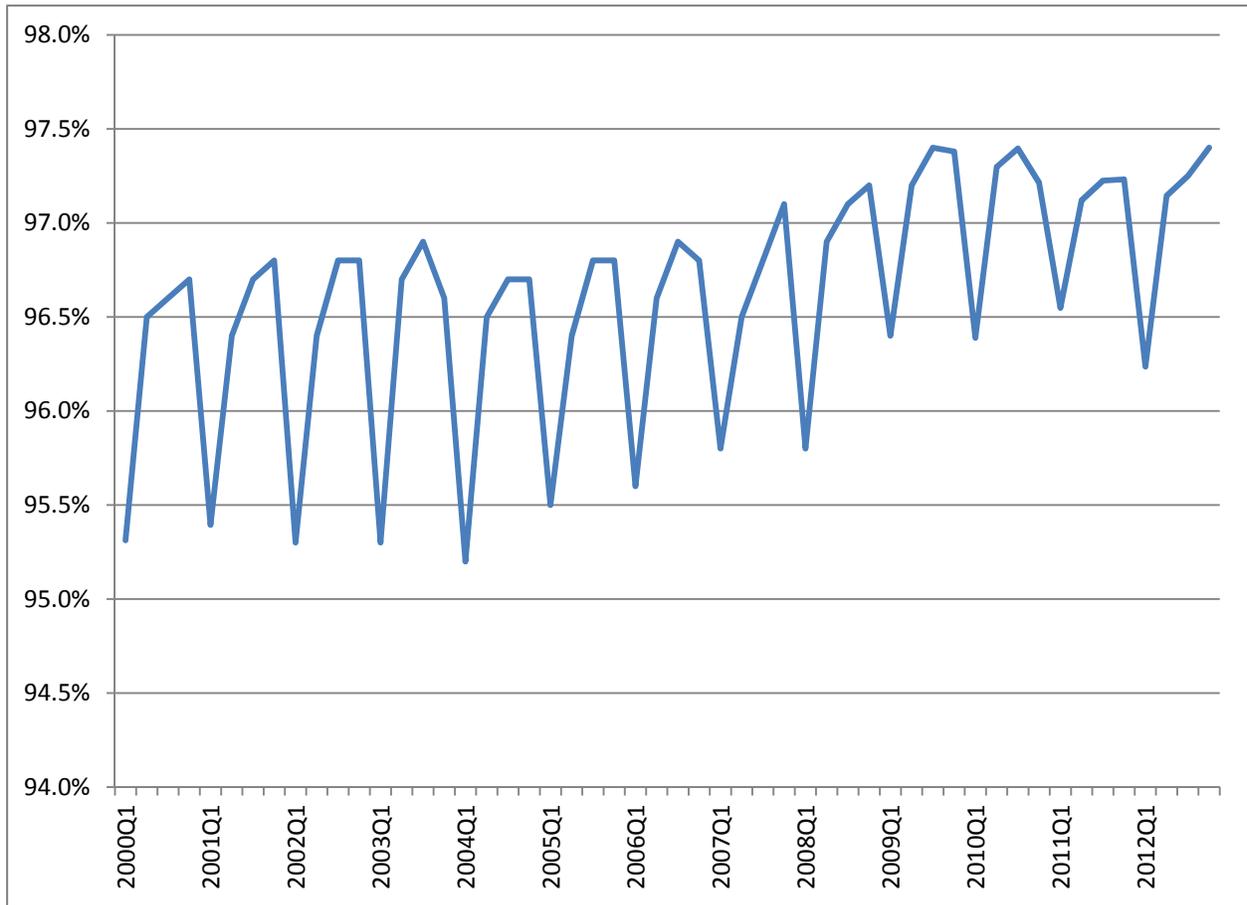


Business Register Metrics Report

The second review mechanism of the QCEW Business Register is the Business Register Metrics Report, which monitors the system software, data quality, and quality of analyst manual matches. The Business Register Metrics Report is produced after the longitudinal linkage process is complete and evaluates the amount and type of record matches, the number of new QCEW Business Register numbers issued each quarter, the number of births and deaths of establishments each quarter, the number of unmatched records, and the number of weighted matches made by the system software.

Monitoring the fields mentioned above allows the BLS to ensure high data quality of the QCEW Business Register. As one can see in Chart 2, the percentage of SESA ID matches that take place from quarter to quarter average between 96 percent and 97 percent. This high percentage indicates the reliability of the QCEW Business Register linkage system and illustrates the small amount of variability in the percentage of records that were linked by SESA.

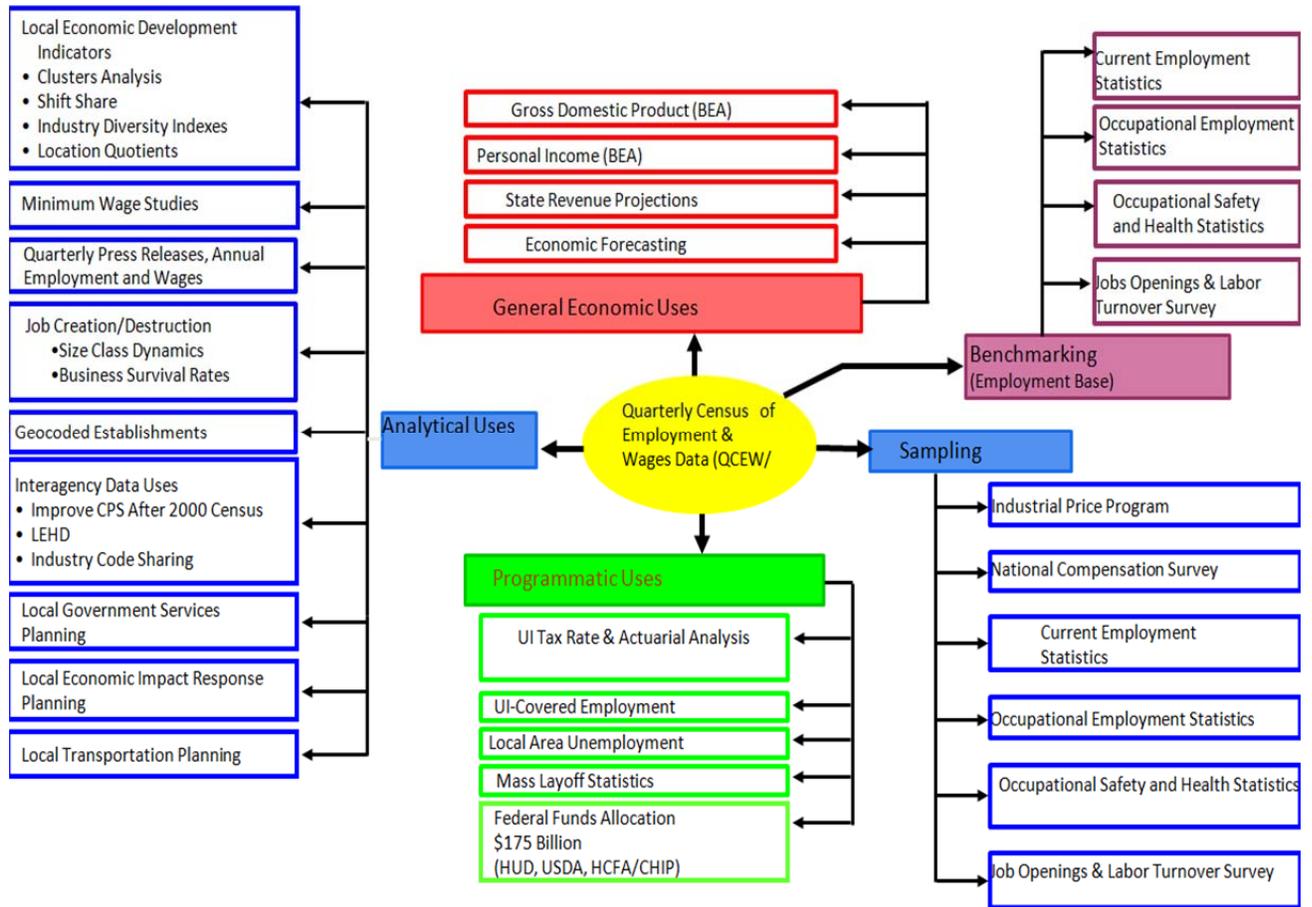
Chart 2. Percentage of Records Matched by SESA ID: 2000/Q1 to 2012/Q4



Section IV: Uses and Users of the QCEW Business Register

Since coverage is so broad (approximately 98 percent of all non-farm wage and salary employment), the QCEW Business Register 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, QCEW data are used extensively in many economic and statistical applications. These include Unemployment Insurance program administration (determining the UI tax rate for establishments for states which base taxes on NAICS codes, assisting UI with planning, and actuarial analysis), macro-economic research, survey employment benchmarking, and micro-economic analysis. BLS programs and surveys use the micro data for sampling purposes. (See Figure 2. below)

Figure 2. The Uses of Quarterly Census of Employment and Wages Data (QCEW)



Sampling Frame

The QCEW Business Register is used as a sampling frame for key economic surveys published by the BLS. One of the many programs that use the QCEW Business Register as a sampling frame is the Current Employment Survey (CES), which is a key survey utilized by the BLS for the publication of the monthly “Employment Situation Report.” Other BLS programs that use the QCEW Business Register for sampling purposes are the Job Openings and Labor Turnover Survey (JOLTS), Occupational Employment Statistics (OES), Producer Price Index (PPI), Occupational Safety and Health Surveys, Occupational Compensation Survey, Employer Benefits Survey, Employment Cost Index Survey, and Productivity Surveys. The Local Area Unemployment Statistics (LAUS) program also uses the QCEW Business Register as its source of employment when CES estimates are not available.

Data Sharing

Data from the QCEW Business Register are shared routinely with other U.S. statistical agencies.

Data Sharing with the Bureau of Census

Since 1990, BLS has shared NAICS codes with the Census Bureau to improve industry coding consistency and reduce respondent burden and costs. This process is as follows: Every quarter, the Census Bureau electronically transfers a file to BLS containing an Employer Identification Number (EIN) which is matched against the QCEW Business Register. Records with matching EINs are supplied with the BLS NAICS code along with supplemental information such as whether the unit is a single or part of a multi-establishment company, ownership code, county code, and detailed address information.

Over a million detailed 6-digit NAICS codes are routinely provided to the Census Bureau each quarter. For the recent reference quarter 2012Q4, the Census Bureau provided BLS with 1.37 million EINs and 1 million were matched to establishments with identical EINs on the QCEW Business Register (a 77 percent match rate). In FY2012, the Census Bureau provided BLS with 4.4 million EINs and BLS matched 3.3 million (75 percent) records; this work includes 13.2 million EINs provided by the Census Bureau over the past 3 years, of which BLS matched 10 million records, a 75 percent match rate. The Census Bureau uses these industry data in their Business Register that serves as a source of sampling frames for frequent business surveys (such as the Annual Survey of Manufacturers) and as a basis for statistical tabulations. The most important benefits of this data sharing project are relieving American businesses of unnecessary response burden, improving industry coding for the Census Bureau, improving usability and promoting consistency between federal statistical products, and reducing redundancy between agency statistical programs, to the exceptional benefit of the American taxpayer.

QCEW data are also used to calibrate the Current Population Survey after each decennial census. In addition, QCEW micro data serve as the primary input for the Longitudinal Employer-Household Dynamics (LEHD) program.

Beginning a three-year period data sharing project in 2012, BLS and Census have also exchanged information on multi-unit establishments. BLS recently shared multi-unit data with the Census Bureau from the QCEW and detailed establishment based data on Professional Employer Organizations (PEOs). Census also provided BLS with multi-unit data. BLS plans to match identical units and adjudicate industry coding differences. The resulting use of consistent industry coding will help improve the quality of the QCEW Business Register and its outputs.

Data Sharing with the Bureau of Economic Analysis

The Bureau of Economic Analysis routinely receives macro level QCEW data each quarter in the preparation of the Personal Income component of the National Income and Product Accounts at the national, state, and county level. In addition, QCEW data are used as input for the Gross Domestic Product. For example, in 2012, covered workers received \$6.497 trillion in pay, representing 93.8 percent of the wage and salary component of personal income and 39.9 percent of the gross domestic income.

National Oceanic and Atmospheric Administration

The National Oceanic and Atmospheric Administration (NOAA) is using QCEW Business Register data to develop a set of statistics on employment, wages, and gross domestic product that provides insights into the contributions of coastal and ocean resources to the national economy.

QCEW micro-data was used in research led by the National Ocean Economics Program (NOEP) that reclassified economic data in a way that focuses on six ocean-linked sectors: ocean-linked construction, living resources, minerals, ship and boat building, ocean-linked tourism and recreation, and marine transportation. Under this new data sharing agreement, NOAA and BLS will collaborate to build on the successful proof-of-concept work of NOEP, updating the historical time series annually as new data become available.

Future Data Sharing Plans

BLS and the Economic Research Service (ERS) with the Department of Agriculture have a data sharing agreement that will use the QCEW Business Register as its sample frame. The Rural Establishment Innovation Survey (REIS) will collect data over a 6-month period for 30,000 respondents. The sample will be largely from rural (nonmetropolitan) areas, but an urban (metropolitan) sample will also be surveyed for comparison purposes. This information will contribute to a better understanding of how increasing international competition and the increasing knowledge intensity of economic activity in the U.S. are affecting the economic vitality of rural areas—and the conditions associated with effective adjustment to these pressures.

Business Employment Dynamics

The individually linked microrecords can be linked across time to publish Business Employment Dynamics (BED) data. Each quarter, the BED uses QCEW Business Register data to produce its quarterly publication that measures the net change in employment at the establishment level and at the firm level.

The BED statistics are calculated for the private sector covering 7.0 million establishments. These statistics are classified by industry and published at the sector and sub-sector level for the nation back to 1992/Q3. Since the initial publication of these data in 2003, new dimensions of data have been added over time. Firm size and state level data are available at the total private sector. And annual over-the-year estimates have been published to allow users to compare these data internationally. “Size of employment change” data show how many established gained (or lost) 1 job, 2 jobs, 3 jobs, etc. Data are also available at the 3-digit NAICS level for the nation.

In August 2010, BLS published BED data by establishment age. The long-standing debate about the role and impact of small versus large businesses has expanded in recent years to consider the

contributions of younger versus older businesses. These new data from the Business Employment Dynamics (BED) program measure the age and establishment survival rates, which can help shed light on to the behavior of and contributions by of new and young businesses compared with their older and more well-established counterparts.

In February 2012, BLS published BED state data by 2-digit NAICS industry code.

Users have found great interest in the firm size data and continue to ask for more detailed information in this area. The output produced by BED is used by many labor market economists.

Labor Market Research

The QCEW Business Register offers researchers a rich dataset of labor market data.

One of the first uses of QCEW Business Register data by researchers external to BLS was undertaken by Card and Krueger (2000) in a study of the minimum wage and employment in fast-food restaurants in New Jersey and Pennsylvania. Using the QCEW Business Register data, the authors were able to select data by state, county, quarter, industry, trade name, and legal name. This rich dataset allowed Card and Krueger to reevaluate their previous study and eliminate common survey errors because the QCEW Business Register data is derived from UI tax records.

Significant labor market research has been done using the QCEW Business Register to show the importance of firm size class, firm survival, and industry analysis. The Monthly Labor Review (MLR) article on the importance of size class methodology by Okolie (2004) was conducted using data from the QCEW Business Register. The QCEW Business Register aided Knaup (2005) in researching establishments by industry classification for her MLR article on firm survival.

Not only can one research national aggregate employment but it is also possible to research particular sectors of the economy as shown in the paper on the employment and wages in the U.S. costal economy by Colgan (2004) published in the MLR. In addition, important research was conducted by Salamon and Sokolowski (2005) on the non-profit sector in the U.S. economy using the QCEW Business Register.

Timeliness of QCEW Business Register Outputs

In 2010, the QCEW quarterly press release on state and county employment and wages was accelerated by one week. For example, the County Employment and Wages press release for first quarter 2013 will be released Thursday, September 26, 2013 making these data available six months after the end of the reference quarter. (See <http://www.bls.gov/news.release/pdf/cewqtr.pdf> for the most recent release) Also, the QCEW program released a State and County Mapping application (<http://beta.bls.gov/maps>). This new feature of the BLS website provides users with super-sector industry employment and wages at

the national, state, and county levels. Data are presented in map, tabular, and downloadable formats.

In 2011, the Business Employment Dynamics press release was accelerated by two weeks, making these data available seven months after the end of the reference quarter. (See <http://www.bls.gov/news.release/pdf/cewbd.pdf> for the most recent release) These data lag the county press release due to additional time needed to link and review the micro data.

There may be further acceleration of these outputs in the next few years as procedures are refined, systems streamlined, and continuous improvement in all data quality assurance methods are improved.

Conclusion

The QCEW Business Register is the most complete and timely source of business establishment data. These quarterly data are supplemented with multiple worksite information and industry 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 and QCEW Business Register metrics are monitored quarterly to maintain a high level of data quality. In addition, standardized State operating systems have been exported to the States to improve standardized processing, editing, and imputation methodology; improve the data quality of the Business Register; and control program maintenance costs.

The QCEW Business Register serves as the source for Business Employment (BED) statistics published on a quarterly basis. The Business Employment Dynamics data are timely and high frequency job flow statistics available 7 months after the end of the reference quarter. BED data consist of national aggregate statistics from 1992Q2 through the most recent quarter available, 2012Q4. BLS first began publishing quarterly national total private data in September 2003. Since then, new national data series have been published by NAICS super-sector, firm size class, size of employment change, birth and death data, and annual data. In August 2007 BED data by State were released. Currently published State data includes quarterly and annual data on total gross job gains for openings, expansions, and births, and total gross job losses for closings, contractions, and deaths. Establishment age data was made available in August 2010 and state industry data was published in February 2012.

The BLS continues to leverage its investment in its QCEW Business Register by expanding its array of BED products, offering researchers free access to this rich set of microdata, and sharing these data with other statistical agencies.

References

- Shail J. Butani, Richard L. Clayton, Vinod Kapani, James R. Spletzer, David M. Talan, and George S. Werking, Jr., "Business Employment Dynamics: tabulation by employer size" *Monthly Labor Review*, February 2006, pp. 3–22.
- Card, David and Alan B. Krueger. (2000). "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply" *The American Economic Review*. 90 (5) 1397-1420. 4 Ibid. p. 1397
- Clayton, Richard L., R. Jason Faberman, Akbar Sadeghi, James R. Spletzer, and David M. Talan, "Business Employment Dynamics: New Data on Gross Job Gains and Losses." April 2004 *Monthly Labor Review*
- Colgan, Charles, S. (2004). "Employment and Wages for the U.S. Ocean and Coastal Economy." *Monthly Labor Review*. 127 (11) 24-30.
- R. Jason Faberman, Akbar Sadeghi, David M. Talan, and Richard L. Clayton, "Business Employment Dynamics," *Monthly Labor Review*, April 2004, pp. 29–42.
- 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, DC, March 1995.
- Knaup, Amy, E. (2005). "Survival and Longevity in the Business Employment Dynamics Data." *Monthly Labor Review*. 128 (5) 50-56.
- Robertson, Kenneth, Larry Huff, Gordon Mikkelson, Timothy Pivetz, and Alice Winkler. 1997. "Improvements in Record Linkage Processes for the Bureau of Labor Statistics' Business Establishment List." *Proceedings for the 1997 Record Linkage Workshop and Exposition*, pp. 212-221.
- Lester M. Salamon and S. Wojciech Sokolowski (2005). "Nonprofit organizations: new insights from QCEW data," *Monthly Labor Review*. 132 (9) 19-26.
- 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, DC, March 1990.
- Searson, M., "Strategies to Implement Electronic Collection of Multiple Worksite Report Data," *Federal Committee on Statistical Policy*, Washington, DC (November 2001)