Quarterly Census of Employment and Wages (QCEW) Business Register Metrics August 2005

Sheryl Konigsberg, Merissa Piazza, David Talan, and Richard Clayton
Bureau of Labor Statistics

Abstract

One of the most important concerns involving large datasets is how the integrity of the data is maintained. In 1999, the Bureau of Labor Statistics (BLS) launched the QCEW Business Register; the Bureau's establishment based business register. The QCEW Business Register is a relational database of 8.4 million business establishments linked longitudinally based on the micro-data submitted quarterly by States from Unemployment Insurance (UI) tax files. Data elements on these files include information on monthly employment, quarterly wages, and other administrative data. Every business establishment on the database contains a unique identifier that allows for tracking of individual establishments at the micro-level across quarters for the untied States. The QCEW Business Register has three critical functions: producing longitudinal Business Employment Dynamics (BED) statistics, serving as a sampling frame for establishment-based surveys for BLS, and serving as an important source for labor market research.

Because these data are used to generate high quality, high frequency, timely and historically consistent data on business and employment, BLS uses measurement methods to quantify the standards of these data. This paper demonstrates how BLS uses business register metrics based on the statistical quality dimensions of relevance, accuracy, timeliness, interpretability, coherence and accessibility to insure high quality data that meets the varied requirements of different user groups.

Keywords: QCEW Business Register, Longitudinal Database (LDB), QCEW Management Review Mechanisms

Introduction

The QCEW Business Register is a rich source of employment and establishment data for the BLS and high quality, high frequency, timely and historically consistent data on businesses, employment and wages. Every business establishment on the database contains a unique identifier (QCEW Business Register Number) that allows for tracking of individual establishments and employment at the micro-level across quarters for the United States. BLS uses measurement methods to quantify the quality of these data. The QCEW Business Register has three critical functions: to produce longitudinal Business Employment Dynamics (BED) statistics, to serve as a sampling frame for establishment-based surveys for BLS, and to provide a new resource for labor market research.

What is the QCEW Business Register?

The predecessor to the QCEW Business Register was the Universe Database (UDB)[1]. The UDB had several limitations; in 1989 only 20 percent of the data on the frame was categorized as multi-location entities. Through the Business Establishment List (BEL), improvement project standards were set on reporting for multi-establishment employers, whereby each individual establishment reported employment for their sites. This allowed the UDB to emerge as the QCEW Business Register and accurately report employment at the establishment level, involving multiple worksites.

The QCEW Business Register is a relational database of 450 million records dating back to 1990 which are linked longitudinally to create a time-series of employment and wages. The QCEW Business Register is based on the micro-data submitted quarterly by States from Unemployment Insurance (UI) tax files to BLS. Each firm is given a UI number which contains the data elements of monthly employment and wages, business name and addresses, industry classification, geocodes, and other administrative data. With these files, the QCEW Business Register provides reliable establishment level data on a quarterly basis to its users.

Definitions and Concepts

The establishment birth and death concept is not new to economics, but in order for the QCEW Business Register to truly capture the dynamic relationships of establishments more information and codification is needed. Two important definitions deal with the relationships between establishments: a predecessor and successor. A predecessor is an establishment that previously reported as one UI account number or reporting unit number and is now being reported under a different UI account number or reporting unit number. A successor is an establishment that is now reported (or will be reporting) under a UI account number or reporting unit number, that was being reported under a different UI account configuration. The purpose of predecessor and successor information is to maintain establishments as continuous, especially when they change ownership or UI number, in order to preserve the QCEW Business Register time-series.

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.

QCEW Business Register Metrics Paper Objective

One of the most important concerns involving large databases is how the integrity of the data is maintained. This paper outlines the QCEW Business Register metrics in three sections. The first section (1) details the inputs and outputs of the QCEW Business Register, the data used in creating the QCEW Business Register, the review process to remove impurities from the input data, the output generated from QCEW Business Register data and the metrics used in the QCEW Business Register linkage process. The second section (2) covers the business register metrics and examines the methodology of the QCEW Business Register linkage system and the control mechanisms to ensure accurate linkages and data quality. Finally, the third section (3) sets the standards and performance management of the QCEW Business Register and describes how quality controls are maintained through management directives and cooperation.

Section I: Inputs and Outputs of the QCEW Business Register

Inputs to the QCEW Business Register

The QCEW is a Federal-State cooperative program that produces a comprehensive tabulation of employment and wage information for workers covered by State Unemployment Insurance (UI) laws and Federal workers covered by the Unemployment Compensation for Federal Employees (UCFE) program.[2] The QCEW is a statistical universe that contains quarterly data of employment and wages. Within these data, each UI record is longitudinally linked creating the QCEW Business Register, which contains a time series of establishments.

The main deliverable from State Workforce Agencies units is the EQUI file. The EQUI file contains seven quarters of employment and wage data. In addition it contains all administrative data such as addresses, ownership, and industry coding. The EQUI file is the raw micro-data used to create the longitudinal links for the QCEW Business Register.

Section II: Uses and Users of the QCEW Business Register

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 programs that use the QCEW Business Register for sampling purposes are the Producer Price Index (PPI), Job Openings and Labor Turnover Survey (JOLTS), and Occupational Employment Statistics (OES). The Local Area Unemployment Statistics (LAUS) program also uses the QCEW Business Register as its employment when CES estimates are not available.

Labor Market Research

The QCEW Business Register offers researchers a rich dataset of labor market data. Not only does the QCEW Business Register contain quarterly employment and wage data, but it also provides administrative data such as State, county, metropolitan statistical area (MSA) codes, physical location addresses, and North American Industry Classification System (NAICS) industry information.

One of the first uses of QCEW Business Register data by researchers external to BLS was undertaken by Card and Krueger (2000)[3] 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 Kruger to re-evaluate their pervious study and eliminate common survey errors because the QCEW Business Register data is derived from UI tax records. [4]

Significant labor market research has been done using the QCEW Business Register to show the importance of size class, firm survival, and industry specific analysis. The Monthly Labor Review (MLR) article on the importance of size class methodology by Okolie (2004)[5] was conducted using data from the QCEW Business Register. The QCEW Business Register aided Knaup (2005)[6] 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)[7] published in the MLR.

QCEW Business Register Output

The longitudinally linked establishment records on the QCEW Business Register provide a rich resource of establishment and employment data used in publishing the Business Employment Dynamics (BED). 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. The BED statistics are calculated with 6.5 million establishments[8]. In addition, these statistics are classified by industry under the North American Industry Classification System (NAICS) and published at the sectors level (two-digit NAICS) in order to give a more detailed perspective on the economy. The output produced by BED is used by many in labor market research and economic calculations, such as State and local governments, the U.S. Federal Reserve Board, the Bureau of Economic Analysis, National Bureau of Economic Research[9] and the academic community.

Section III: QCEW Business Register Linkage Process

In order to assure accurate linkages, there are two components of QCEW Business Register matching: automated matching and analyst matching. The computer automated matching links approximately 96 percent of all records 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 a consolidation.

SESA ID Matching

Ninety six 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 (See Chart 1). 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 (see Table 1).

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 deleted 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 IV: The 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 and coherence to this time-series.

Management Review Mechanisms

The QCEW Business Register has a comprehensive hierarchy to ensure that data quality permeates all levels of the QCEW structure. One of such tools used in the management structure is the BLS Co-operative agreement which is a contract that sets 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 saturate the system to promote coordination and cooperation to increase data quality. Annual meetings for State and Regional offices are management coordination tools which allow these branches of the QCEW program to communicate questions and concerns, as well as excellence in data quality and standards.

In order to coordinate all agencies involved in the QCEW 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 management decisions.

Review Mechanisms of QCEW Business Register Systems

There are three quality control documents which are produced to review the Business Register process and systems: the Flash Report, the Business Register Metrics Report, and the Employment and Training Administration (ETA) 581 Report.

Flash Report

The Flash Report monitors the data quality received from the States submittal of QCEW micro-data. 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 2, these three topics are monitored to make sure they do not significantly increase from one quarter to the next nor do they continue on an upward trend.

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 and Figure 2. In Figure 1, State A is state which monitors their NAICS 99999 (unclassified) and has a percentage of these records well below the national average. While, State B in Figure 2 on the other hand, was submitting micro-data records with NAICS 99999 above the national average. In the first quarter of 2004 a problem was identified that the number of unclassified NAICS that State B was submitting was unacceptable. The problem was identified and corrected, and by the third quarter of 2004, State B reporting of NAICS 99999 returned to below the national average. With proper monitoring of data quality fields this issue was brought to QCEW management and resolved. 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 QCEW Business Register process to monitor the measures of data quality level of outputs.

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 linkage process is complete and evaluates the amount and type (breakout or consolidation) of analyst manual 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 Figure X3, 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 last 29 quarters of the percentage of records that were linked by SESA.

Employment and Training Administration (ETA) 581 Report

It is important that not only does the QCEW Business Register contain accurate employment and wages, but compiles accurately and in a timely fashion the number of establishment births quarterly. The ETA 581 Report monitors this issue. The ETA 581 is a comprehensive tool which analyzes the number of new Unemployment-Insurance liable employers each quarter. The main goal of this report is to determine if new births are being captured in a timely manner by the BLS QCEW program. As seen in Chart 3, over 90 percent of establishment births are captured within two quarters. Through this study, the ETA validates that the QCEW Business Register is maintaining its goal of capturing establishments as they enter the market in a timely manner.

Conclusion

The QCEW Business Register data is produced quarterly on a timely basis and published on a national level. States submit their data which is then reviewed and edited approximately 6½ months after the end of the reference quarter. The BLS then longitudinally links the QCEW microdata across quarters and publishes the Business Employment Dynamics press release 7½ months after the end of the quarter. BED publications consist of national aggregate numbers on a quarterly basis starting in the second quarter of 1992. In addition to the national publication, the BED program is working on methodologies to publish national size class data and State level BED data.

QCEW Business Register data provides a universe of establishments for the processing of Business Employment Dynamics Data, for researchers, and for general public usage. Since the QCEW Business Register is used for varying programs and surveys, it is essential to maintain high quality in the data produced for these sectors. By monitoring the QCEW Business Register linkage system and the data inputs the integrity of QCEW Business Register data is maintained.

Endnotes

[1] There were many inconsistencies with UDB data because of differences in the collection of data amongst States, such as disparities in timing and accuracy of reported employment, addresses, and industry coding. The BEL Improvement Project also aimed to improve the computer aided matching system and weighted matching system, which are now used in QCEW Business Register systems.

[2] Exclusions from UI coverage include self-employed workers, most agricultural workers on small farms, all members of the Armed Forces, elected officials in most states, most employees of railroads, some domestic workers, most student workers at schools, and employees of certain small nonprofit organizations.

[3] 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

[5] Okolie, Cordelia. (2004). "Why Size Class Methodology Matters in Analyses of Net and Gross Job Flows." *Monthly Labor Review.* 127 (7) 3-12.

[6] Knaup, Amy, E. (2005). "Survival and Longevity in the Business Employment Dynamics Data." *Monthly Labor Review*. 128 (5) 50-56.

[7] Colgan, Charles, S. (2004). "Employment and Wages for the U.S. Ocean and Coastal Economy." *Monthly Labor Review*. 127 (11) 24-30.

[8] BED data removes private households (NAICS 814110), establishments with zero employment for three quarters and establishments in Puerto Rico and the Virgin Islands.

[9] Clayton, Richard, and James R. Spletzer. "Business Employment Dynamics." NBER-CRIW Conference on Producer Dynamics. Bethesda, MD. April 8, 2005.

Chart 1: Percentage of Records Matched by SESA ID

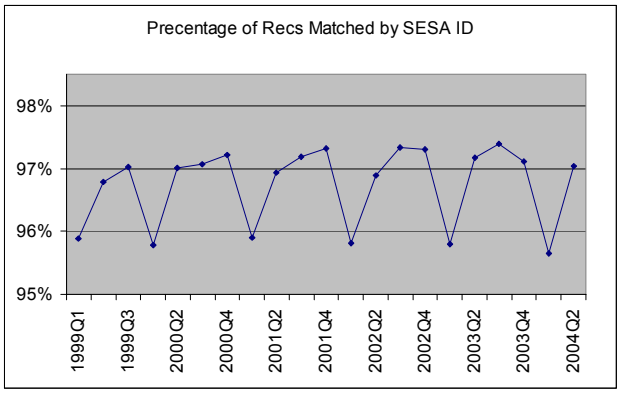

Table 1: Weighted Match Scoring Function

Match	Block	Blocking Variables
SET 1	Block 1	Business Name, Address , 6 Digit NAICS, County
	Block 2	Business Name , Address , County
	Block 3	Business Name , Address , 3 Digit NAICS, County
	Block 4	Business Name , Address , 6 Digit NAICS, Zipcode
SET 2		
	Block 1	Business Name , Address , 6 Digit NAICS, Phone Num
	Block 2	Business Name , Address , Phone Number
	Block 3	Business Name , 3 Digit NAICS, County, Phone Num
	Block 4	Business Name , Phone Number
	Block 5	Business Name , County, Phone Number
	Block 6	Address , 3 Digit NAICS, County, Phone Number
SET 3		
	Block 1	Business Name , 3 Digit NAICS, County
	Block 2	Business Name , 3 Digit NAICS, Zipcode
	Block 3	Business Name , Address , 6 Digit NAICS, County
	Block 4	Business Name , Address , County
	Block 5	EIN, Address , County
	Block 6	EIN, Zipcode
	Block 7	Business Name , Zipcode, County

Chart 2: Imputed, Prorated, and Missing Units as a Percentage of Reported Units (US Average)

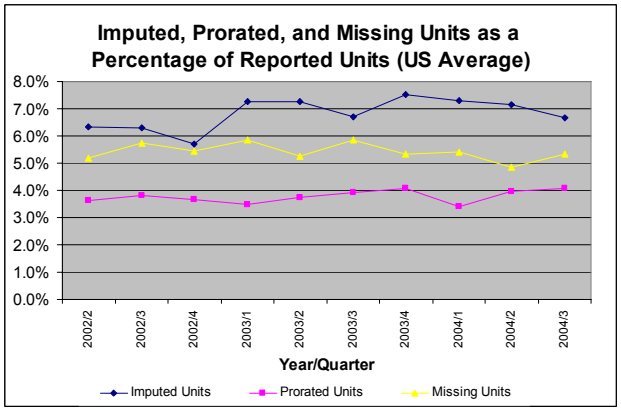

Figure 1: Percentage of Unclassified NAICS in State A in Comparison to the National Average of Unclassified NAICS

Figure 2: Percentage of Unclassified NAICS in State B in Comparison to the National Average of Unclassified

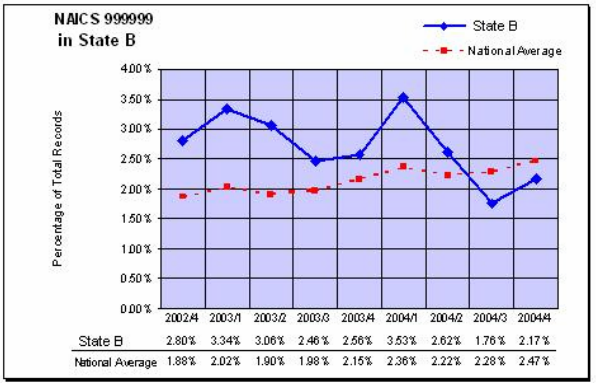

Chart 3: US Status Determination Promptness for Processing New Liable Employers: US Total