Selective and macro-editing of a large business based administrative data set

Prepared by David R. H. Hiles, Department of Labor, Bureau of Labor Statistics (BLS), USA

I. Introduction
Edits are applied to the United States Quarterly Census of Employment and Wages (QCEW) by the Bureau of Labor Statistics (BLS) to meet sample frame and economic publication requirements. The QCEW data stream is an enhanced version of the administrative data resulting from quarterly unemployment insurance (UI) reports. Data items include monthly employment and quarterly wages. Almost every employer in the United States, Puerto Rico, and the U.S. Virgin Islands files this report, as required by law. The resulting 9.5 million quarterly records are the basis of the BLS business register. These records are aggregated to provide employment and wage benchmarks for the Bureau of Labor Statistics and the Bureau of Economic Analysis and for the most current published employment and wage totals at the county or local level. These multiple uses are made possible by an edit suite that has both selective and macro-editing components. The vast size of the QCEW file and strict quarterly publication schedules force a rigorous series of edits that are both cross-sectional and longitudinal. The quarterly nature of the QCEW data provides opportunities for targeted edits, such as over the month, over the quarter and over the year edits.

The paper assesses the value and shortcomings of portions of the edit suite in current operations, as well as prospects for improvement.

Employers report in numerous modes to the QCEW program. Some report as a single account for a given state, the vast majority report each quarter at the worksite level, and a few establish a worksite distribution which BLS then uses to prorate a single account across worksites. In this paper, worksite, employer, establishment, and report are used interchangeably to refer to the most detailed worksite level data available. While QCEW is often referred to as a business register, it also includes reports from federal, state, and local government employers.

II. QCEW Basics
A. Uses of the QCEW
The QCEW is a foundation stone of the U.S. statistical system. It serves many other uses as seen in Appendix A. The QCEW business register is the sample frame for BLS business surveys. QCEW employment data serves as the benchmark for Current Employment Statistics estimates and for other BLS business surveys. BLS and the states publish establishment counts, employment and wages by industry five and a half months after the end of each quarter.

Data from the QCEW program serve as an input to other Federal and State programs. The Bureau of Economic Analysis (BEA) of the Department of Commerce uses QCEW data as the base for developing the wage and salary component of personal income. The Employment and Training Administration (ETA) of the Department of Labor, as well as state workforce agencies, use QCEW data to administer the employment security program. The QCEW data accurately reflect the extent of coverage of the State UI
laws and are used to measure UI revenues; national, State and local area employment; and total and UI taxable wage trends.

B. Data stream
The QCEW data stream is an enhanced version of the administrative data resulting from quarterly unemployment insurance reports. It is a business register that is updated quarterly. In addition to data on private sector employers, the QCEW also receives reports from employers in all levels of government, with federal data coming from the Unemployment Compensation for Federal Employees program (UCFE). BLS works with its state partners to encourage employers to report at the worksite level. BLS has a successful history of working with payroll service providers to make it possible to report worksite level data electronically.

Data elements are simple, reflecting the administrative data roots of the QCEW. Employers report monthly employment for employees who worked or received pay for the pay period that included the 12th of each month, and total quarterly wages for all employees who worked any time during the quarter. When employers join the unemployment insurance system, they provide information on their physical location and primary economic activity. They also provide contact information such as telephone numbers, mailing addresses and increasingly, email addresses.

Classification and location information is maintained largely via the QCEW Annual Refiling Survey (ARS). The ARS is conducted on a three-year cycle, with approximately one-third of all business establishments sampled each year. Establishments in certain NAICS industries that traditionally have low volatility are put on a six-year cycle. Establishments without an industrial classification are surveyed annually. The ARS also asks employers to identify new worksites in the State. Classification information can also be updated based on communication with the reporter during the editing process.

Worksite information is collected each quarter via the Multiple Worksite Report (MWR). Worksite reporting is mandatory in 28 states, and voluntary in the rest. Employers participating in the MWR provide monthly employment and quarterly wages at the worksite level. This built-in capacity to obtain worksite level reporting each quarter for virtually all businesses may be unique in the world in its frequency and comprehensiveness. The information collected via the MWR is combined with the ARS and the UI agency reports to create the QCEW.

C. Edit process
Employer reports (microdata) can be received initially by state UI offices, by state QCEW offices, or by BLS, depending on the reporting option selected to the employer. Reports received by the state UI office are processed with a priority placed on correct payment of unemployment insurance contributions and are then shared with the state QCEW office. After the state QCEW phase is complete, processing moves to BLS.

State phase
The state QCEW office subjects these reports to a comprehensive editing process approved by BLS. The basic monthly employment edit consists of a six step statistical test that includes the use of multiple t-tests for: month-to-month change, over-the-year change, and a 12 month variation in data; some tests are conducted on levels while others are conducted on rate of change. The wage edit includes the use of inter-quartile tests developed by Hoaglin, Iglewicz, and Tukey. Reports received directly by the state QCEW office are subjected to the same editing process applied to data provided by the UI office.

Reports covering a range of the largest multi-unit enterprises in the U.S are received directly by the BLS Electronic Data Interchange (EDI) center. BLS staff at the EDI center use a subset of the edits used by states. After this first set of edits, the data from the EDI center are sent to the state QCEW office for comprehensive editing and correction. Late and missing reports are imputed to create a complete file.

In addition to the usual correction of classification information and employment and wage errors, the QCEW edit and review process substantially improves the UI data by allowing reviewers to add comments regarding particular worksites. Reviewers can add up to three comment codes for a worksite per quarter. Reviewers can use these codes to note economic events such as weather-related shutdowns, more or less business, intra-plant transfers, improved reporting, bonuses paid, and so on. The QCEW file
also includes a narrative comment field which reviewers can use to store information that does not fit the comment code system. Comment codes are used by some of the editing and review systems. The FESTER state system, discussed below, allows records with comment codes to be bypassed. The QCEW over-the-year adjustment system, also discussed below, uses comment codes to identify candidate worksites for adjustment. Comment codes and narrative comments are also available to all BLS sample frame users.

Employer reports are aggregated into macrodata which are also edited by the state QCEW office. This macrodata edit uses comprehensive processes approved by BLS. Editing of the macrodata is typically done after the microdata editing and correction phase is complete. Macrodata edit failures are traced back to the responsible microdata records for correction, acceptance as false edit results, or documentation as real economic events. After the completion of the state macrodata review and correction phase, each state delivers the microdata to BLS for further processing. There is some variation across states regarding when macrodata edits are first reviewed.

**BLS phase**
Output edits are conducted by BLS staff using the QCEW microdata as delivered by the states. BLS review is focused on ensuring the quality of the sample frame and publication products based on the QCEW file. Items questioned by BLS reviewers are returned to the state QCEW office for comment, explanation, and correction as needed. Corrections may be made by state or BLS staff. All corrections are made at the microdata level, with reaggregation performed to integrate the correction into publication products.

**III. Selective Edits**

**FESTER**
Selective edits identify a limited number of records with influential errors for which manual review is worthwhile. The QCEW state editing system includes a selective edit known as FESTER (Focused Error Scoring Targeted Edit Report). This report lists records with either high volatility of employment and wage data or with specific, (targeted) errors. FESTER is typically run during the microdata portion of editing. This allows the identification and correction of many large errors before the beginning of output editing by the states. It can be run repeatedly to assess the progression of editing. FESTER itself is not an edit, but a means of sorting edit results to ensure that large issues can be prioritized. As is the case with most sections of state review in QCEW, the method is approved by BLS while the state is accorded substantial latitude in the setting of parameters and organization of work. States may tighten parameters but they cannot loosen them beyond the BLS-established requirements.

FESTER assigns an edit score and presents records in a sorted report by score-indicated severity. States can subdivide the report by using sorts that match the assignment of work within the state (such as by county, ownership (private, federal, state, local), metropolitan area, town, or account number range).

**FESTER scoring**
The base score of a record is calculated based on the employment size of the record, the size of the change in employment or wages, and the number and class of edits that have failed for the given record. The size of employment and wages is part of the score because large records may have little change but may have invalid classification codes. Monthly employment and quarterly wage changes are calculated between the selected quarter and the immediately prior quarter. The use of quarterly change can result in normal seasonal movements receiving high scores. Employment changes are assigned a higher weight in light of the higher volatility observed in wages.

This base score is assigned to each edit failure for the particular record. Records that fail numerous edits get a higher score. The base scores for the failed edit are adjusted for the priority class of the edit (A, B, or C) and then summed to yield a total score for the record.

**Alternate ranking options**
States have some ability to alter the sort of FESTER records by adding or subtracting points programmatically. They can add points to records with invalid classification codes or to records with up to 10 particular edits that the state staff wish to emphasize. Similarly, state staff can select two edits that
can be de-emphasized. These options allow state staff to adjust the rankings to help see the issues that are most critical in the particular state or processing period. Most states use the default settings.

**Bypass options**
States can bypass records that have a comment code. Comment codes are entered manually. They are an indicator that these records have been reviewed and accepted earlier in the process. This recognizes the fact that many large movements are legitimate economic or seasonal events even though they fail an edit.

### IV. Macro-edits

#### A. BLS QCEW macro-edits are driven by publication requirements

One QCEW product is a quarterly BLS news release *(County Employment and Wages)* which ranks large counties in the US by employment and wage change. There are more than 3,000 counties in the US, which if all included would make for an unwieldy publication. This release is limited to counties with more than 75,000 employees in the most recent year. For the 2014 reference year, this release included 339 counties, which together accounted for 72 percent of US employment. This news release is a relatively new product (first issued on October 31, 2003) of the venerable QCEW program. The release has stimulated the continuing development of new approaches to QCEW macro-edits.

QCEW macrodata are the result of the summation of the QCEW universe file of employer worksite reports. Macrodata is not produced by the usual means of pushing sample responses through an estimator. This use of addition as a replacement for an estimator is simple. However, this method is also subject to distortion based on changes in the basis of reporting as well as to incorrect employment and wage levels. Reporting basis change in the QCEW data can originate from the unemployment data or from the worksite data developed by QCEW operations.

The publication of a ranked list of counties based on change in employment and wages required BLS to improve the means of handling these aspects of the QCEW process, discussed below. The news release ranks counties by over-the-year change in employment and wages. This is simpler than using a seasonal adjustment process. Simple over-the-year change accounts for most of the largest seasonal movements that occur in the U.S. economy, such as the swings in residential construction and retail trade. However, using the traditional macrodata aggregates would result in numerous cases of counties being ranked incorrectly due to changes in the basis of reporting.

#### B. Over-the-year change adjusted for reporting differences

BLS adjusts year-ago data to derive more representative measures of change. In addition to producing better measures, this adjustment improves the efficiency of editing by reducing some of the noise that may otherwise make errors more difficult to identify. Ranking serves as a useful component of editing, because unusual rankings are often caused by data errors that have made it past the rest of processing. The review of these over-the-year rankings uncovers problems that may not be seen in the over-the-quarter edits used by state QCEW staff.

**Classification changes**
Classification changes have long been identified within state QCEW operations as being either economic or non-economic. Worksites with economic changes are not adjusted so that the economic change appears in the rankings. Worksites with non-economic changes are understood to have correct classification in the current period. The new classification is applied to the year-ago reports which are then reaggregated for ranking purposes.

**Employers start reporting by worksite**
One of the major features of the QCEW program is the successful development of quarterly reports at the worksite level. This is a continuing process. Each year many employers begin reporting at the worksite level. This means that year-ago worksite data is not available for these employers. This conversion usually results in jobs that were counted in one location and industry being distributed across the state, causing a large drop in one area and smaller increases elsewhere. BLS creates synthetic year-ago
worksite data by applying the new worksite distribution to the year-ago employer total, which is then deactivated. The synthetic year-ago records are included in the reaggregation used for the news release.

Worksite distribution improves
Some worksite reports are based on an initial report by worksite, which can be carried forward for a substantial time. These worksite reports are often updated to reflect a more current distribution provided by the employer or to convert to actual quarterly worksite reporting. These distributions can be substantially different than the initial report. Most of these changes in the basis of reporting are identified by state QCEW staff who add an appropriate comment code to the affected records. These reports can also be identified by BLS reviewers. The identified changes are accepted as correct and are applied to the year-ago records for reaggregation. Care is taken to limit this adjustment to records that change for non-economic reasons.

V. Value added by the QCEW edit process
The edit process adds enormous value to the QCEW file as received from its administrative source, state unemployment insurance agencies. To measure the effect of the edit and review cycle, BLS developed a study that compared the raw file, as initially delivered by UI, to the QCEW data as later delivered to BLS. Thirty-nine states participated in this study by submitting data at both points of processing for the first quarter of 2008. Tables 1 and 2 below show the impact of QCEW editing on classification information, addresses, and economic data. Sixteen percent of the records in the study population of 4.6 million reports were changed in one way or another.

In the quarter studied, 751,000 records had a change in at least one field. Most of these changed either employment or wages, thus improving accuracy. The record count under-measures the value of editing. Many thousands of records were reviewed and found to be accurate, thus requiring no change. Records with substantial changes often resulted in respondent re-contact, which yielded either a confirmation or a correction. If QCEW editing, review, re-contact, and correction was not conducted, employment counts for the U.S. would have been in error for this period by 14.2 percent for the participating states. It is reasonable to expect that the error would be comparable for the remaining states. It is clear that a few records had large errors—easily identified and corrected for national use. However, the QCEW is also used for accurate local counts. Intensive edit and review assures local data accuracy.

Table 1. Value-added study, 2008 first quarter
Selected counts of records with changes

<table>
<thead>
<tr>
<th>Count</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>751,481</td>
<td>records with changes, mostly to employment and wages</td>
</tr>
<tr>
<td>Classification changes</td>
<td></td>
</tr>
<tr>
<td>60,582</td>
<td>county changes</td>
</tr>
<tr>
<td>41,903</td>
<td>from unclassified to known county</td>
</tr>
<tr>
<td>73,343</td>
<td>industry changes</td>
</tr>
<tr>
<td>40,452</td>
<td>from unclassified to known industry</td>
</tr>
<tr>
<td>Address changes</td>
<td></td>
</tr>
<tr>
<td>65,073</td>
<td>physical location changes</td>
</tr>
<tr>
<td>13,941</td>
<td>mailing address changes</td>
</tr>
<tr>
<td>Reviewer notes in file</td>
<td></td>
</tr>
<tr>
<td>109,022</td>
<td>comment codes</td>
</tr>
<tr>
<td>3,640</td>
<td>narrative comments</td>
</tr>
</tbody>
</table>
Table 2. Value-added study, 2008 first quarter
Revision in employment and wages, 39 states

<table>
<thead>
<tr>
<th>Employment in millions</th>
<th>Raw UI</th>
<th>QCEW</th>
<th>Change</th>
<th>%Change</th>
<th>Abs. Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>102.8</td>
<td>88.3</td>
<td>-14.6</td>
<td>-14.2 %</td>
<td>22.1 %</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wages in trillions of dollars</th>
<th>Raw UI</th>
<th>QCEW</th>
<th>Change</th>
<th>%Change</th>
<th>Abs. Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$979.7</td>
<td>$985.9</td>
<td>+ 6.3</td>
<td>+0.6%</td>
<td>3.0%</td>
<td></td>
</tr>
</tbody>
</table>

VI. Shortcomings

A. Systematic error spread across many small establishments
Although every one of the 9.5 million records in the QCEW file are subjected to mechanical edits, only a small fraction of edit failures are reviewed by humans. Resource constraints necessarily direct the review towards the largest errors and the largest employers. Small establishments make up a major share of the QCEW file. In the first quarter of 2014, there were 5.6 million establishments with fewer than 5 employees, together reporting 8 million employees.

How does QCEW guard against systematic error in these accounts? The QCEW administrative data flow starts with 53 separate UI offices (each state, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands). This provides an important cross check. Systems and processes are not identical across all UI offices, so a systemic error in one state is unlikely to be repeated in another. When states are combined at BLS and reviewed against each other, this cross check comes into play.

Systematic errors in small accounts can also be caught in macro editing. In 2014, a processing error resulted in a very large number of small accounts being overcounted. At the worksite level, this was commonly seen as a change from one to two employees. These records would not be looked at by any reasonable micro-edit. However, summed together they resulted in growth that was larger than expected. The larger than expected growth was spread out across many establishments. It was impossible to tie the unusual movement to the usual case of a few large movements. The movement in the state did not have an economic explanation, and the state did not behave similarly to its peers. The BLS reviewer kept looking. He eventually found that the double counted records did not match up to a secondary data resource. Once the discrepancy was discovered, the group of records was isolated, the systematic error was uncovered, and the overcount was corrected. The secondary administrative data resource was a count of wage records by employer. Employers have to file quarterly wage records for each employee with the state UI program. The wage record datastream is independent of the UI employer report provided to QCEW. Most state QCEW offices are able to aggregate wage records by UI employer to use as a cross check for the standard QCEW report.

B. Focus on quarterly change in edits
The QCEW edits used for employment and wages are dominated by a focus on change over the quarter. Establishments can be in business for many years, or they can be very short-lived. Quarterly edits are an important means for reviewing establishments with a short track record. This approach works well for industries with little seasonal movement. However, many false edit failures are generated for industries with large seasonal movements, such as those seen in construction and retail.

VII. Prospects for improvement

A. Wage record integration
The UI system in each state produces two reports: a summary report for each employer as described above, and an employee level report. This second report is referred to as a wage record. A wage record provides a worker’s unique identifier (SSN), total wages paid in the quarter, and the employer’s UI identification number.
Wage record use in state QCEW operations is still in development. Some state QCEW offices have good wage record access and use these records as a cross-check against UI employer summary reports of employment and wages. Many states use clusters of wage records to identify employers that change unemployment accounts or that merge with other firms. BLS is working with state staff to implement these improvements in all states. These uses also improve the quality of birth/death measures produced by QCEW. Individual wage records are not shared with BLS.

B. **Improved edits in QUEST**

QUEST is the BLS state system redesign project that is scheduled for implementation in September of 2018. This system is being written at BLS. It will assume the production role currently fulfilled by systems developed by state QCEW staff in Maine and Utah. This system will improve many aspects of state operations. It will also make major improvements in the edits used by state QCEW staff. Many changes will be made as part of this project. A select few are mentioned below.

1. **Improved FESTER-inspired scoring**
   The old FESTER method, discussed above, was dominated by employment size. The new version has separate scores for employment, wages, and other data elements. The highest of the separate scores for the record becomes the overall score. This allows other problems to score higher instead of just those with very high employment.

2. **New wage record tools**
   BLS will provide tools to ensure that all states will be able to sum individual employee wage records outside of the QCEW program and use the summarized data within the new QCEW system. Wage records are not carried within the state QCEW production system to avoid restrictions that accompany SSN-coded records. These tools add new fields that will help the editor review and evaluate the data. These include but are not limited to different wage record counts, wage record wages, counts of records with high wages, and the three largest individual wage records in a particular employer.

3. **Improved predecessor/successor edits**
   “Predecessor/successor” refers to reports that transfer ownership from a predecessor employer to a successor employer. Some of these transfers are simple one-to-one transactions, but others can be one predecessor to many successors, many to one, or many to many. These transactions commonly cause undercount or overcount conditions that may be expensive to find and correct.

   The improved predecessor/successor edits will better identify problems such as lost data, unexplained inclusion of additional employment and wage data, or strange shifts in classification. The improvements will be seen in cases that involve new worksites or that have more than one predecessor or successor.

4. **Improved employment edits**
   The improved employment edit will replace the *t*-test with a *seasonality test*. The *t*-test allows for large fluctuations in the current quarter based on a past fluctuation of similar magnitude, without regard for timing or direction of change. This results in records passing the edit with dubious monthly employment. The new edit will accept records with established seasonal trends or with small shifts in the timing of seasonal changes, while rejecting records with large changes that are not seasonal.

**References:**


*Exportable QCEW System Manual*

- Appendix E Edit Methods
- Appendix H Batch Job Parameters
Appendix I Batch Jobs

County Employment and Wages, Technical Note

Presentation on QCEW
Richard L. Clayton, NASWA LMI Conference
October 16, 2008, Indianapolis, Indiana

Data Sharing: Progress and Challenges at BLS
Richard L. Clayton, FESAC
December 12, 2014, Washington DC
Appendix A. Uses of the QCEW

Quarterly Census of Employment and Wages (QCEW)

General Economic Uses
- Gross Domestic Product (GDP)
- Personal Income (BEA)
- State Revenue Projections
- Economic Forecasting

Analytical Uses
- Minimum Wage Studies
- Job Creation/ Destruction
- Employment and Wages
- Location Quotients

Programmatic Uses
- U.S. Tax Rate & Actuarial Analysis
- UI-Covered Employment
- Local Area Unemployment
- Federal Funds Allocation

Local Economic Development
- Local Economic Impact Response Planning
- Local Government Services Planning
- Balanced Industry Code Sharing
- Local Economic Impact Response Planning

Current Employment Statistics
Occupational Employment Statistics
Occupational Safety and Health Statistics
Jobs Openings & Labor Turnover Survey

Benchmarking (Employment Base)

Sampling