

IMPLEMENTING A SCORE FUNCTION TO PRIORITIZE BUSINESS SURVEY EDIT FAILURES AT BLS

Mary Dorinda Allard, Gordon Mikkelson, Linda I. Unger, U.S. Bureau of Labor Statistics
 Gordon Mikkelson, Bureau of Labor Statistics, Suite 4985, 2 Massachusetts Avenue NE, Washington, DC 20212
Mikkelson_G@bls.gov

Key Words: Score function, data quality, prioritized data editing.

1. INTRODUCTION

The Covered Employment and Wages Program, commonly referred to as the ES-202 program, is a cooperative endeavor between the Bureau of Labor Statistics (BLS) of the U.S. Department of Labor and the State Employment Security Agencies in the 50 states, the District of Columbia, Puerto Rico, and the Virgin Islands. Each calendar quarter, the ES-202 program collects data and produces a comprehensive and detailed registry of establishment information for workers covered by state unemployment insurance (UI) laws or the Unemployment Compensation for Federal Employees (UCFE) program. This registry, comprised of almost 100 different data elements, serves as the sampling frame for BLS establishment surveys. Extensive employment and wage tabulations by industry at the national, state, and county level are also produced, published, and used for various public and private sector economic and analytical purposes. The data are also a large part of the wage and salary component of national income and gross domestic productions.

The ES-202 program has both micro and macro data. The micro data are submitted at the establishment level and are summed into macro cells made up of all units in a particular state, county, sector (public or private), and industry. These macro data are examined before being published by each state and BLS. The micro data are also extensively reviewed prior to being used for sampling, longitudinal research, and production of other state and BLS statistics.

Standardized state processing systems include over 150 microdata edits on more than 8.2 million records each calendar quarter. The micro level edit conditions are categorized into four groupings:

- (1) economic (employment and wage) data,
- (2) classification (industry and geographic) codes,
- (3) business identifiers (names, addresses, etc.), and
- (4) firm linkage information (mergers, acquisitions, etc.).

State staffs use the edit results to review, research, correct, and explain suspect data (approximately 7 to

10 percent of all records). Similar edits are used by BLS to validate the data.

Such a large number of edit failures to review and resolve is obviously a huge workload, especially since edited data are reviewed by both state staffs and BLS. Recent time-use studies have shown that approximately one-fourth of the states' staff time is spent performing this data review process.

Initially, state staffs sorted edit results by account number or county and worked through the list. Different staff members were frequently assigned parts of the list. Although this helped distribute the workload equitably, it did not help editors focus on the more significant or critical edit cases.

To assist edit reviewers in prioritizing edits, the BLS created the ABC List, which ranks edit codes by severity. A-level edits, which concentrate primarily on invalid classification codes and economic conditions, are the most severe conditions. B-level edits, which focus on assorted edit conditions such as inconsistencies in related data fields or ownership transactions between employers, are somewhat less severe, and C-level edits are even less so. For example, a record with a large employment fluctuation would receive an A-level edit, whereas a record with an invalid telephone number would receive a C-level edit. Thus, editors concentrate on A-level edits first, then B-level edits, and finally C-level edits.

The distribution of A, B, and C category edits are listed below:

Category ¹	% Edit Conditions	% Edit Flags ²
A	38	53
B	26	20
C	36	27

¹ ABC list as of April 2001 but subject to revision

² This approximation is based on fourth quarter 1999 BLS edit flags. It is possible for a record to have more than one edit flag, and not all edits are run against all records.

* The views expressed in this paper are those of the authors and do not represent official policy of the Bureau of Labor Statistics.

** The authors would like to thank Glen Read of the Utah Department of Workforce Services for developing and introducing scoring functions to the ES-202 program.

Although the ABC List helps editors focus on more serious edits, two major problems remain: (1) All records within each category on the ABC List are given equal weight and importance. Categorizing the importance of the edit does not adequately prioritize the review work based on significance of the data, impact of macro data, and critical data elements. For example, a B-level edit occurring in a record with large employment is probably more serious than a record with an A-level edit with little or no employment. (2) Since so many records are flagged for review within category A, there may not be adequate time and resources to review and resolve serious questions in categories B and C.

Alternative approaches to prioritizing records within the ABC List were considered. Glen Read of the Utah Department of Workforce Services first introduced the concept of using a score function to prioritize the edit failure listings for the ES-202 program (Read 2000).

This research was expanded and score functions incorporating both the ABC List as well as the impact of micro level data on macro level data were developed. The various functions were empirically tested and the results are presented in this paper.

2. METHODOLOGY

Because it was not clear what type of function would work best for ES-202 data, four different scoring functions were developed and tested. The test results were compared to each other in order to determine the function most suitable for ES-202.

Scores were calculated for all micro data records failing at least one edit. No ES-202 micro data are published by BLS; instead, micro data are summed into macro cells made up of all units in a particular state, industry, county, and sector.

The various score functions tested take into account elements considered important when ranking edited data. Because the two most important micro data elements are monthly employment and total quarterly wages, all of the score functions consider whether the record failed edits associated with these data elements. The ABC List is a valuable editing tool, so two of the scores are computed by examining which edits failed for a record, assigning weights to the records based on the ABC List, and then summing the weights to produce an overall score for the record. Since data are used both at the micro and macro level, three scoring functions also reflect the contribution that the record makes at the macro level.

Notation

The following notation will be used in the score function formulas:

$\max Emp(cq,pq)$ = the largest monthly employment in the current and prior quarters

wt_i = the weight associated with the i-th edit for the record

$wt_i = 0$, if the *ith* edit passed;
 1, if the *ith* edit failed and is on the A list;
 0.4, if the *ith* edit failed and is on the B list;
 0.2, if the *ith* edit failed and is on the C list.

$\max Wage(cq,pq)$ = the larger of the total quarterly wages in the current and prior quarter

$\max Empchange(cq,pq)$ = the maximum difference between consecutive months in the current and prior quarters

$Wagechange(cq,pq)$ = Absolute value of the difference between the total quarterly wages in the current and prior quarters

$microemp$ = the employment of the third month of the current quarter

$microwage$ = the total quarterly wages of the current quarter

$macroemp$ = the macro employment of the third month of the current quarter for the macro cell corresponding to the record

$macrowage$ = the macro level total quarterly wages for the macro cell corresponding to the record for the current quarter

Ie and Iw are indicator variables defined as follows:

If the monthly employment edits pass and the wage edits fail, Ie = 0 and Iw = 1;

If both the employment and wage edits fail, Ie = Iw = 0.5;

Else Ie = 1 and Iw = 0.

Score Functions

The first score function considered is based on a variation of the FLAG function described in LaTouche and Berthelot (1992):

$$\text{FLAG1} = \sqrt{\max \text{Emp}(cq, pq)} \times \sum wt_i \times Ie + \sqrt{\max \text{Wage}(cq, pq)} \times \sum wt_i \times Iw / 100.$$

This function incorporates two key micro data elements, the size of the record and the severity and number of the edits. For example, a record with 1000 employees and an A-level edit would receive a higher score than a record with 100 employees and an A-level error. Also, a record with 100 employees and two A-level edits would receive a higher score than a record of the same size with only one A-level edit.

FLAG1 does not take into account the effect that the record in question has on the macro cell. To incorporate macro data into our score function, two variations of a score function, FLAG2 and FLAG3, from an article by Farwell and Raine (2000) were developed. The Farwell and Raine method gives higher scores to micro records that represent a larger proportion of their macro level cell. We used summed employment and wages at the state, industry, county, and sector level as our macro cells.

$$\text{FLAG2} = \frac{\sqrt{\max \text{Emp}(cq, pq)}}{\text{macroemp}} \times Ie + \frac{\sqrt{\max \text{Wage}(cq, pq)}}{\text{macrowage}} \times Iw / 100.$$

The other score function based on the Farwell and Raine method, FLAG3, used the change in employment and quarterly wages instead of using monthly employment and quarterly wages:

$$\text{FLAG3} = \frac{\sqrt{\max \text{Empchange}(cq, pq)}}{\text{macroemp}} \times Ie + \frac{\sqrt{\text{Wagechange}(cq, pq)}}{\text{macrowage}} \times Iw / 100.$$

While these two functions incorporated macro data and the size of the record into the score, they did not include information about the severity or number of edits. To try to incorporate all these elements into one score, a variation of FLAG1 was developed. Two weights were added to the score function:

$$\text{FLAG4} = \text{FLAG1} \times \text{Wt1} \times \text{Wt2},$$

where

$$\text{Wt1} = \begin{cases} 1.50, & \text{if } 0.80 < \max(\text{microemp}/\text{macroemp}, \text{microwage}/\text{macrowage}) \leq 1.00; \\ 1.25, & \text{if } 0.40 < \max(\text{microemp}/\text{macroemp}, \text{microwage}/\text{macrowage}) \leq 0.80; \\ 0.75, & \text{if } 0.10 < \max(\text{microemp}/\text{macroemp}, \text{microwage}/\text{macrowage}) \leq 0.40; \\ 0.25, & \text{if } 0.10 \leq \max(\text{microemp}/\text{macroemp}, \text{microwage}/\text{macrowage}); \end{cases}$$

and

$$\text{Wt2} = \begin{cases} 1.25, & \text{if } \text{macroemp} \geq 1000; \\ 1.0, & \text{if } 100 \leq \text{macroemp} < 1000; \\ 0.75, & \text{if } \text{macroemp} < 100. \end{cases}$$

3. RESULTS

In order to compare the four different score functions and determine whether a scoring function would be a helpful tool for data reviewers, ES-202 data from Texas, California, North Carolina, Florida, Maine, Vermont, and the District of Columbia were scored. Results presented here are from Texas, and although they are not included in this paper, other states showed similar patterns.

The scored data were reviewed thoroughly to determine the best scoring function for the ES-202 program. Three characteristics were deemed to be important in determining the most effective score: (1) the scoring function should in general rank errors in records with large employment or wages higher than records with similar errors and smaller employment and wages. (2) The scoring function should rank more serious errors over less serious errors. Since the ES-202 program uses the ABC list to identify three levels of severity of errors, an A-level error should be ranked higher than a B- or a C-level error with similar employment and wages. (3) Records whose employment and wages make up a large percent of their macro cell should be ranked over records with employment and wages that make up a smaller percentage of their cells.

To compare the four different functions, the scores of each were divided into quantiles of 10, and the averages of selected variables were computed.

Table 1. Average Employment in Texas edit failures (Data from the third quarter of 2000)

Quantile	FLAG1	FLAG2	FLAG3	FLAG4
1	743.6	40.2	40.2	782.9
2	198.2	138.8	109.2	231.3
3	118.8	166.2	133.4	89.3
4	71.1	155.8	148.9	40.9
5	37.1	206.4	236.1	29.3
6	16.9	92.1	159.8	19.2
7	12.2	43.1	85.7	10.7
8	13.9	247.6	234.1	9.4
9	20.8	119.9	84.1	20.4
10	3.4	26.8	5.7	3.8

There is a clear trend in FLAG1 and FLAG4 to rank records with larger employment higher than records with smaller employment (see Table 1). The other two functions, FLAG2 and FLAG3, do not appear to have very obvious patterns; the top quantile has records with fairly small employment.

Table 2. Percentage of A-level flags in Texas edit failures (Data from the third quarter of 2000)

Quantile	FLAG1	FLAG2	FLAG3	FLAG4
1	99.7	54.5	71.3	97.2
2	98.6	64.1	70.6	96.0
3	97.2	68.8	73.0	97.9
4	96.9	72.3	76.7	98.1
5	97.9	71.3	75.5	97.0
6	97.6	71.8	72.3	97.0
7	97.2	74.1	91.9	95.8
8	84.1	100.0	100.0	89.3
9	0	100.0	100.0	1.2
10	0	92.5	38.2	0

FLAG1 has the largest number of A-level edits in the top quantile, followed by FLAG4 (see Table 2). It is not surprising that FLAG2 and FLAG3 have a smaller number of A-level edits in the top quantile, since the ABC list was not incorporated into their formulas. Upon closer examination, many of the records that were ranked in the top quantile in FLAG2 and FLAG3 were relatively small records with, according to ABC List priorities, relatively insignificant edits.

Table 3. Average Percentage of Macro Cell in Texas edit failures (Data from the third quarter of 2000)

Quantile	FLAG1	FLAG2	FLAG3	FLAG4
1	20.6	64.2	57.0	35.5
2	17.2	26.9	26.1	18.6
3	13.1	9.4	9.7	14.4
4	12.6	2.9	4.8	13.4
5	10.3	1.2	1.7	8.3
6	11.3	0.4	0.6	4.8
7	7.0	9.8	19.3	2.0
8	7.5	12.7	3.8	4.9
9	16.5	1.0	0.4	21.2
10	12.8	0.3	5.6	5.7

FLAG1, which does not include macrodata in its formula, ranks records that do not make up much of the macro cell in the top quantile (see Table 3).

When a small state with approximately 100 micro edit failures was tested, the records with the highest values for FLAG2 and FLAG3 were records from macro cells that had less than five micro records. The record with the highest score for FLAG2 and FLAG3 was one that had only one record in the macro cell, and the average monthly employment had gone from 4 in the prior quarter to 62 in the current quarter. That record had the 29th highest score using FLAG1 and the 16th highest flag when using FLAG4. The records that had higher scores than this record using FLAG1 and FLAG4 tended to be records with much larger employment, and come from macro cells that had more micro records. The record with the highest score using either FLAG1 or FLAG4 had the 43rd highest score using FLAG2 or FLAG3. This record failed the monthly employment edit when the employment dropped to 589 in the first month of the current quarter after being over 800 in the prior quarter.

4. CONCLUSIONS

The research and testing demonstrated that a score function could indeed be a helpful tool for edit review. In order to be an effective tool, however, the needs of the program have to be carefully taken into account. The score functions FLAG2 and FLAG3, for instance, ranked records with small employment and wages highest. A score function such as this implemented as a data review tool would result in reviewers spending a great deal of time on small records.

Based on the quantile comparisons and review of scored micro data, it was determined that FLAG4 was the most promising of the four formulas because it ranked records highest when they had A-level edits, had large employment or wages, and made up a large percentage of their macrodata cells. FLAG1 ranked larger records and records with A-level edits higher, but it did not do as well ranking records that made up a large percent of their macro cell. FLAG2 and FLAG3 did not rank records high when they had either large employment or wages or A-level edits.

5. FUTURE PLANS

The macro data cells used in this paper were summed employment and wages in the same state, industry, county, and sector. This is the most useful macro data definition for the ES-202 program because its data review and publication are done at that level, but we would like to experiment with other definitions of macro cells.

It would be useful to test some variations to the weights for FLAG4 for records that have employment or wages above some tolerance or below another tolerance. This will reduce the weight of records that are a high percentage of a small macro cell and increase the weight of large records that are a small percentage of a very large cell.

Future testing will use raw data from the states from different times during the data collection process. This will ensure that the score function chosen for implementation in the ES-202 Program will be an effective review tool throughout the year.

The scoring research will be further expanded to incorporate additional aspects of the ES-202 program. A scoring function will be developed for macro data to expedite the review and validation of aggregate data for publication. Also, future scoring functions will incorporate the use of historical data. Since states continue to receive new or updated information for prior calendar quarters of data, significant edit questions may exist on the older data that are more critical than more recent information.

REFERENCES

Farwell, K. and Raine, M., (2000), "Some Current Approaches to Editing in the ABS", Proceedings of the Second International Conference on Establishment Surveys, ASA, pp. 529-538.

Latouche, M and Berthelot, J.-M., (1992), "Use of a Score Function to Prioritize and Limit Recontacts in Editing Business Surveys", Journal of Official Statistics, Statistics Sweden, vol. 8, no. 3, pp. 389-400.

U.S. Department of Labor, Bureau of Labor Statistics, Bulletin 2490 (1997), *BLS Handbook of Methods*, pp.42-47.

U.S. Department of Labor, Bureau of Labor Statistics, ES-202 Operating Manual (July 2000), *Edit Conditions and Formulas*, Appendix F.

Read, Glen (October 2000), Utah Department of Workforce Services, in cooperation with U.S. Department of Labor, Bureau of Labor Statistics, EXPO-202 and EARS Version 6.0 System Documentation, *Edit Scoring Function*, Appendix E.