Outlier Detection for the Manufacturing, Mining, and Construction Sectors in the 2012 Economic Census

Nicole M Czaplicki[1] and Katherine J. Thompson
[1]U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Abstract
In 2002, the U.S. Census Bureau began using a modified Hidiroglou-Berthelot (HB) edit for outlier identification to find outlying tabulations in the Geographic Area Series (GAS) reports of the Economic Census. This outlier-detection procedure compares ratios of tabulations, either of the same item over two time periods (historic ratios) or of two different but related items from the current time period (current cell ratio). The methodology implemented in production was developed by a group of subject matter experts and methodologists from five of the eight trade areas covered by the Economic Census. Seeking to expand the use of this methodology for the 2012 Economic Census, we conducted a feasibility study for the manufacturing, mining, and construction sectors to see if they could also use this approach or a further modified version. The data collected by these sectors differ from the service sectors in several meaningful ways, such as the number of the collected items and the correlation between historic ratio pairs. This paper presents the results of our empirical investigation along with our conclusions.

Key Words: Hidiroglou-Berthelot edit, macro editing

1. Introduction

The U.S. Census Bureau conducts an Economic Census in years ending in 2 and 7, mailing out over four million census forms to business establishments that provide commercial services to the public and other businesses. Data are collected at the establishment level and are classified according to the North American Industry Classification System (NAICS). The Economic Census coverage extends to establishments in eighteen non-farm economic sectors, including wholesale trade, retail trade, finance, insurance, real estate, services, transportation, communication, utilities, manufacturing, mining and construction – with all but the last three sectors being referred to collectively as the "services sectors." Economic Census statistics provide a more comprehensive view of the economy than the ongoing monthly and quarterly economic surveys conducted by the U.S. Census Bureau, but are less timely. Businesses receive their census forms between October and December of the census year. Subject matter experts begin reviewing responses in February. The Census Bureau releases the Economic Census data on a flow basis. The Advance Report is released first, approximately one year after the forms are mailed out. This is followed by the Industry Series reports, which present national-level industry totals. Next, the Census Bureau releases the Geographic Area Series (GAS) reports, which contain industry totals for states and for other selected geographic breakdowns. The subject/summary series reports are the final data product.

[1] This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

The Economic Census micro-data are edited and imputed using the Plain Vanilla subsystem (Sigman, 1997), which validates and imputes consistent records using ratio and balance edits. Other sector and industry-specific edit rules are also applied to the micro-data to ensure consistency within the record (reporting unit) itself and within the industry where the unit is classified. Consequently, the summarized data used in the Advance Report are validated at the national industry level. Unfortunately, the industry level tabulation parameters may not necessarily be applicable at the more detailed geographic levels: for example, the industry average wage per employee ratio in a given industry could vary by geography. Moreover, the finer cells used for the Industry and GAS reports contain fewer establishments than those used for the national report, thus increasing the probability that an atypically large or small establishment could have a substantive impact on the tabulation. Given the size of the Economic Census in terms of establishments and the volume of tabulations published in the Industry and GAS reports (over 2.5 million summaries), it would be impossible to validate every establishments' data at the industry by geography level in a timely manner; likewise, it would be impossible to conduct clerical review of every single tabulation. Instead, methodologists at the Census Bureau developed automatic macro review procedures that use statistical methods or influence functions to identify a subset of "suspicious' tabulations or establishment data within tabulation, and the subject matter experts investigate this selected subset.

Sigman (2005) presents the modified Hidiroglou-Berthelot (HB) edit used to identify outlying tabulations in the Industry and GAS reports produced for the service sector data. The methodology was introduced in a production setting for the 2002 Economic Census and was implemented with software called the SODS (Service Sector Statistics Division Outlier Detection System). The SODS has proved highly effective in both the 2002 and 2007 Economic Censuses. Given these successes, we hoped to apply this methodology to the entire Economic Census in 2012. This report presents the results of a feasibility study conducted to determine whether the SODS macro-editing outlier detection methodology could be applied to the manufacturing, mining, and construction sectors.

Section 2 provides background information on the manufacturing, mining, and construction sectors of the Economic Census. Section 3 describes the methodology used in the Outlier Review Tool (ORT), the software program that replaces the SODS in 2012. Section 4 presents the evaluation study. Section 5 discusses the reasons behind the very different outlier detection results/recommendations for the mining sector tabulations. Section 6 offers some concluding remarks.

2. Background on the Manufacturing, Mining, and Construction Sectors of the Economic Census

Although there is one Economic Census, the collected data items differ by sector. All sectors included in the Economic Census collect annual payroll, first quarter payroll, total employment, and sales/receipts from each establishment. Administrative data are often available for these four key items and are included in micro-editing validation and imputation processes. For the services sectors, up to an additional four "basic" (or core) variables may be collected from each establishment, depending on the sector. In contrast, the manufacturing, mining, and construction sectors collect 29, 23, and 21 basic data items from each establishment, respectively, and usable administrative data are only available for the payroll, employment, and receipts items. Thus, these sectors' micro and macro review procedures rely on historic comparisons to prior census data and to Annual Survey of Manufactures (ASM) data when available (manufacturing only).

In general, economic data are highly skewed, with a relatively small number of very large establishments that make up most of the activity in that sector and a much larger number of small establishments that make up the remainder. The services sectors industries tend to have heavy right tailed distributions, whereas the kurtosis of the manufacturing, mining, and construction sector populations are fairly low due to the scarcity of large plants, mines, and builders in specialized sectors. Like the services sectors industries, the manufacturing sector publishes both county and place level geographic series. To avoid disclosure, mining and construction sectors do not publish county or place level geographic industry series. Of these three sectors, the manufacturing sector is by far the largest. The mining and construction sectors each correspond to one unique 2-digit NAICS code. However, the manufacturing sector comprises three 2-digit NAICS codes[2]. We perform the analysis described in Section 4 separately by 2-digit NAICS code. This retained the homogeneity within sets of 6-digit NAICS tabulations with their associate 2-digit NAICS sector code.

Lastly, the manufacturing and mining sectors attempt a complete collection of establishments that exceed a predetermined size (Lineback et al., 2012). The construction sector uses a probability sample (see http://www.census.gov/econ/census07/www/methodology/).

The Advance Report for these three sectors is scheduled for release in December 2013, the Industry Series Reports are scheduled for November 2014, and the Geographic Area Series reports are targeted for August 2015[3]. This provides more than adequate time for development of new methods. It also provides a large time window in the project schedule for review. Historically the analysts have extensively reviewed the micro- and macro-data by industry for a majority of the collected items. A risk of this extensive review is overediting, where several values can be changed by a very small amount, with little perceivable effect on the tabulations while inducing an unmeasurable potential bias in the micro-data. For 2012, the program managers wanted a more objective procedure that would balance statistical processes with analyst input and would allow for more rigorous control of review time and workload. Given the production success with SODS, it made sense to evaluate the feasibility of expanding the software to include the additional three sectors.

3. Outlier Review Tool (ORT) Methodology

The macro-editing procedure described below utilizes two types of ratios of tabulations: historic cell ratios and current cell ratios. Historic cell ratios compare the tabulation of a variable from the current Economic Census to the same tabulation from the previous Economic Census. Current cell ratios compare two different but related tabulations of variables from the current Economic Census. Both items in the ratio must be strictly positive.

[2] Some internal documentation refers to the manufacturing component of the Economic Census as a trade area instead of a sector because it comprises collections from more than one sector.
[3] These dates are subject to change pending budgetary constraints.

The Outlier Review Tool (ORT) uses the modification the Hidirolgou-Berthelot edit (Hidiroglou and Berthelot, 1986) presented in Sigman (2005). Hereafter, we refer to this procedure as the ORT method and to the Hidiroglou-Berthelot edit as the HB edit. Both the ORT and the HB edit are performed as follows:

Step 1: Centering transformation

This transformation centers the distribution of ratios around zero.

$$SR = \begin{cases} R/R_m - 1 & R \geq R_m \\ 1 - R_m/R & 0 < R < R_m \end{cases}$$

Here R_m = median value for ratio R over a set of cell ratios within the same industry and level of geographical detail (e.g., state).

Step 2: Size-effect transformation

This transformation accounts for the size of the tabulations.

For historic cell ratios

$$ESR = (SR)[\max(T_C, T_P)]^u$$

T_C = current period cell total
T_P = prior period cell total

For current cell ratios

$$ESR = (SR)[\max(T_{Num}, R_m T_{Den}]^u$$

T_{Num} = numerator cell total of the current ratio
T_{Den} = denominator cell total of the current ratio

With the current cell ratios, the denominator value is rescaled by the median ratio value described in step one to account for potential differences in unit (e.g., wage per employee).

"The parameter *u* controls the importance associated with the magnitude of the data" (Hidiroglou and Berthelot, 1986) by giving small changes in "large" units greater importance than large changes in "small" units. For example, the choice of *u*= 0.3 (approximately a cube-root) will bring large observations closer to the center of the distribution, while leaving the smaller ratio values nearly untouched. Using *u* = 0 creates an outlier detection region from the symmetrized distribution created in the previous step without incorporating any rescaling for unit size.

Step 3: Quartile Transformations

The quartile transformation provides the statistics used to identify outliers and is applied to both the centered observations (SR) and the size-effect-centered values (ESR), also called the "effects."

Quartile transformation with u = 0 (*QSR*)

$$QSR = \begin{cases} (SR_m - SR)/D_{SR,Q_1} & SR \leq SR_m \\ (SR - SR_m)/D_{SR,Q_3} & SR > SR_m \end{cases}$$

where SR_{Q_1} = first quartile of SR
SR_m = median SR
SR_{Q_3} = third quartile of SR
$D_{SR,Q_1} = max\{SR_m - SR_{Q_1}, |A * SR_m|\}$

$$D_{SR,Q_3} = max\{SR_{Q_3} - SR_m, |A * SR_m|\}$$

where A is a small value that ensures that the outlier region does not automatically include the largest or smallest case when $SR_m - SR_{Q_1} = 0$ or $SR_{Q_3} - SR_m = 0$

HB Edit Quartile transformation (QESR)

$$QESR = \begin{cases} (ESR_m - ESR)/D_{ESR,Q_1} & ESR \leq ESR_m \\ (ESR - ESR_m)/D_{ESR,Q_3} & ESR \geq ESR_m \end{cases}$$

where ESR_{Q_1} = first quartile of ESR associated with a selected value of u.

ESR_m = median ESR

ESR_{Q_3} = third quartile of ESR

$$D_{ESR,Q_1} = max\{ESR_m - ESR_{Q_1}, |A * ESR_m|\}$$
$$D_{ESR,Q_3} = max\{ESR_{Q_3} - ESR_m, |A * ESR_m|\}$$

where A is a small value that ensures that the outlier region does not automatically include the largest or smallest case when
$ESR_m - ESR_{Q_1} = 0$ or $ESR_{Q_3} - ESR_m = 0$

The outlier region is determined after applying the appropriate quartile transformation. The outlier-detection region essentially *resembles* a robust confidence interval, centered around the *median* effect and uses the distance from the first or third quartile to the median as robust proxy standard error. We can rewrite the acceptance region in the following way.

Modified HB edit (u=0) Acceptance Interval

$$(SR_m - cD_{SR,Q_1}, SR_m + cD_{SR,Q_3})$$

HB Edit Acceptance Interval

$$(ESR_m - cD_{ESR,Q_1}, ESR_m + cD_{ESR,Q_3})$$

Note that the effect incorporates the size transformation via the u parameter: the ORT method considers the intervals determined both by $u = 0$ (no size effect) and $0 < u < 1$, whereas the HB edit only considers the latter. However, the value of c is determined subjectively and can differ from ratio test to ratio test. The quartile transformation makes the critical value c into the tolerance for the ratio test.

With the HB edit, a cell ratio is flagged as an outlier when its QESR value is greater than c. In ORT, a cell ratio is flagged as an outlier if it fails the "very small cell test"[4] and the absolute value of QSR and QESR are **both** greater than the critical value, c. This is a very conservative outlier identification approach, as a tabulation ratio must fail two different tests to be flagged as an outlier.

---

[4] The very small test verifies that the number of establishments used to create the tabulation is less than a predetermined threshold provided by the subject matter experts. Cells that satisfy the conditions of the "very small cell test" are not considered in the ORT outlier-identification process.

4. Empirical Study

4.1 Data

This section presents our evaluation study of that applied the ORT outlier detection algorithm and the HB edit methodology separately to tabulations of "unedited" data from the manufacturing, mining, and construction sectors. For these analyses, we used historic data from the 2002 and 2007 Economic Censuses. Our tabulations used reported data values, after correcting all "rounding" errors (values reported in units instead of thousands) because those types of errors are usually corrected during micro-data review. The tabulations used for this analysis are not necessarily representative of those that would be examined in a production setting, as macro-editing procedures usually begin after the first stage of micro-editing is completed, and our input data are consequently much more subject to reporting errors than the production data.

Subject-matter experts provided us with a candidate set of current cell and historic cell ratio tests. Initially, these lists contained over twenty ratio tests per sector. However, some ratio tests involved real-valued items and others used derived items (items that are mathematical functions of other items), which eliminated them from consideration for ORT. After eliminating the ratio tests that contained these items, we requested that the subject matter experts select a "core list" of ten ratios per sector for the evaluation. A ratio model implies a linear no-intercept regression model, and the most effective ratio tests use highly correlated items. To help the subject matter experts select strong ratio tests, we provided correlation analyses for each ratio test obtained from tabulations of final edited data, along with linear regression analysis to determine whether an intercept should be included in the model.

We needed a "gold-standard" to evaluate the effectiveness of the candidate outlier-detection methods on the manufacturing, mining, and construction tabulations. Thompson and Sigman (1999) describe a classification procedure for flagging individual ratios as "good" or "bad." First, the numerator and denominator items for the ratio are independently classified as "good," "bad," or "questionable. These classifications are then used to classify the ratio as "good" if both the numerator and denominator are classified as "good" and "bad" otherwise.

Our evaluation procedure was very similar. Using the industry tabulations for each item obtained by summing the "unedited" census data, we had our subject matter experts classify each *tabulation* as "outlier" or "non-outlier." To do this, we provided spreadsheets containing item tabulations created from the unedited and the final edited data for each variable in each industry as well as the difference and percentage difference between the two values and gave the following instructions:

> *For our evaluation, we are defining an "outlier" as an industry tabulation of an item whose final tabulation has a very large percentage difference or absolute difference from the same tabulation computed from unedited data.*

Analysts used their own judgment to decide what constituted a large difference between edited and unedited values when flagging outliers and were allowed to change their criteria across variables or sectors if deemed necessary.

4.2 Outlier Detection Evaluation

To parallel the research approach presented in Sigman (2005), we evaluated the HB edit and the modified version HB method (the ORT method). The ORT method will never flag more outliers than the HB edit alone, so testing both methods provides an alternative approach in case the HB edit flags many incorrect outliers or the ORT method flags too few. We examined the performance of the individual ratio tests as well as the collective set of ratio tests for each sector.

Recall that the HB edit has two main parameters. The u parameter controls the effect of unit (or tabulation) size and the c parameter controls the width of the acceptable range. For each ratio test, we varied these two parameters to identify the combination that maximized the number of actual outliers flagged and minimized the number of flagged non-outliers. Following the analytic approach provided in Hunt et al. (1999), we systematically varied the parameter u between 0 and 1.0 (incrementing by 0.1) and the parameter c between 3 and 7 (incrementing by 1). Individual ratios tests were evaluated based on the Type I error rate, Type II error rate and the Hit Rate, defined in Thompson and Sigman (1999) as

> **Type I Error Rate**. The proportion of true non-outliers flagged as outliers by a given procedure.
> **Type II Error Rate.** The proportion of true outliers not flagged by a given procedure.
> **Hit Rate.** The proportion of flagged tabulations that are true outliers.

For each individual ratio test, we tried to select parameters that balanced the simultaneous goals of obtaining low Type I and Type II error rates and high Hit Rates. Of course, this balancing is highly dependent on the subjective tabulation classification procedure. Future applications could consider the heuristic approach presented in Belcher (2003).

In general, the ORT method with u=0.3 and c=5 generally yielded reasonable results in the studied sectors. However, we found that some of the individual ratio test results were improved with different parameter choices. In fact, there was no single combination of u and c that was the best choice for all the construction sector ratios. For construction sector ratio tests, the best parameters for each test varied between 0.3 and 0.7 for u and 4 and 7 for c. Fortunately, ORT allows different parameters for each ratio test.

Because tabulations are often compared in more than one ratio test, the individual ratio test Type II error is a poor measure of the overall proportion of unidentified bad tabulations left remaining in the final report. To evaluate the results from the selected outlier detection methods for the *complete* set of ratio tests, we calculated the following statistics for each sector:

> **All-item Type II error rate.** The proportion of true outlier estimates that are not flagged as outliers by any ratio test.
> **All-item hit rate.** The proportion of flagged estimates that are true outliers.

Table 1 presents the evaluation statistics that were obtained using u = 0.3 and c = 5 for *all* current cell ratio tests. We conducted the evaluation within 2-digit NAICS code, mimicking the production set-up. Within in each sector/NAICS code, the 3rd through 5th columns provide a two-way classification table, with the rows providing counts of ratio test results from ORT (not outlier/outlier) and the columns providing the "gold standard" classification counts described in Section 4.1.

Table 1: All-item Current Ratio Test Results with the ORT method
u=0.3 and *c*=5 Used for All Ratio Tests

Sector (NIACS)	*Total Tabs*	*Outlier Outcome*	*True Outlier*	*True Not Outlier*	*All-Item Type II error*	*All-Item Hit Rate*
Mining (21)	29	Not Outlier	2	9	0.167	0.556
		Outlier	10	8		
Construction (23)	31	Not Outlier	5	10	0.313	0.688
		Outlier	11	5		
Manufacturing (31)	109	Not Outlier	19	35	0.306	0.782
		Outlier	43	12		
Manufacturing (32)	126	Not Outlier	5	60	0.091	0.656
		Outlier	40	21		
Manufacturing (33)	236	Not Outlier	24	106	0.217	0.783
		Outlier	83	23		

For the manufacturing sector, *u*=0.3 and *c*=5 worked well for all considered ratio tests, and the ORT outlier-detection method performed well overall. The results for the construction sector were similar, with better results obtained using ORT than the HB edit method alone after varying the *u* and *c* parameters by test. In contrast, with several of the mining sector ratio tests, we found that the HB edit alone identified more correct outliers while retaining essentially the Type I error rate as the ORT. Like the construction sector, no single set of parameters proved to be the best choice for all mining sector ratio tests. For mining, the best parameters for each test varied between 0.3 and 0.6 for *u* and 4 and 6 for *c*.

To determine the overall effect of using different parameter settings by ratio test, we repeated our original evaluation using the differing (best choice) values of *u* and *c* for each ratio test with those obtained from using *u*=0.3 and *c*=5 for all current cell ratio tests. Table 2 summarizes the results, applying the ORT method to the all manufacturing and construction ratio tests and the HB edit to the mining ratio tests.

Table 2: All-item Current Ratio Test Results with Recommended *u* and *c*
(Differing by Test)

Sector (NAICS)	*Total Tabs*	*Outlier Outcome*	*True Outlier*	*True Not Outlier*	*All-Item Type II error*	*All-Item Hit Rate*
Mining[5] (21)	29	Not Outlier	1	10	0.083	0.611
		Outlier	11	7		
Construction (23)	31	Not Outlier	5	10	0.313	0.688
		Outlier	11	5		
Manufacturing (31)	109	Not Outlier	19	35	0.306	0.782
		Outlier	43	12		
Manufacturing (32)	126	Not Outlier	5	60	0.091	0.656
		Outlier	40	21		
Manufacturing (33)	236	Not Outlier	24	106	0.217	0.783
		Outlier	83	23		

[5] Using HB edit alone.

Interestingly, although using different values of *u* and *c* for each construction sector ratio yielded better *individual* test results (in terms of Type I error rates and hit rates), the *aggregate* results are the same as those presented in Table 1. This similarity is a consequence of the *set* of ratio tests, which include some tests with weakly correlated items. Initially, we recommended dropping these tests, but ended up retaining them at the subject matter experts' request. Modifying the parameters for these tests improved their efficiency somewhat, but the remaining outlier-detection regions are quite wide. Because there are several effective ratio tests, the all-item Type II error rate is quite low. However, we believe that the same set of "true" outliers could be identified with a subset of the ratio tests used (confining the tests to those with highly correlated items). In other words, the stronger ratios are identifying the true outliers in our analysis, whereas the less correlated ratios contribute very little.

Implementing macro-level outlier detection for the mining sector was problematic. Using the HB method alone (instead of ORT) and varying ratio test parameters improved the individual ratio test results and the aggregate results over the ORT application. However, the choice of method was not as clear-cut. Ultimately, the mining sector subject matter experts preferred the most conservative approach obtained using the HB edit.

After completing the current cell ratio evaluation, we repeated the evaluation procedure on the manufacturing sector historic ratio tests, again using the parameters u= 0.3and c=5 for all ratio tests. Table 3 summarizes these results.

Table 3: Historic Ratio Test Results (Manufacturing Sector Only)

Sector (NAICS)	*Total Tabs*	*Outlier Outcome*	*True Outlier*	*True Not Outlier*	*All-Item Type II error*	*All-Item Hit Rate*
Manufacturing (31)	109	Not Outlier	16	46	0.296	0.809
		Outlier	38	9		
Manufacturing (32)	126	Not Outlier	6	65	0.176	0.509
		Outlier	28	27		
Manufacturing (33)	236	Not Outlier	26	116	0.306	0.628
		Outlier	59	35		

Finally, we combined historic and current ratio tests for the manufacturing sector to obtain the results presented in Table 4.

Table 4: Combined Current Cell Ratio and Historic Test Results (Manufacturing Sector Only)

Sector (NAICS)	*Total Tabs*	*Outlier Outcome*	*True Outlier*	*True Not Outlier*	*All-Item Type II error*	*All-Item Hit Rate*
Manufacturing (31)	109	Not Outlier	16	33	0.258	0.767
		Outlier	46	14		
Manufacturing (32)	126	Not Outlier	5	59	0.111	0.645
		Outlier	40	22		
Manufacturing (33)	236	Not Outlier	20	103	0.187	0.770
		Outlier	87	26		

Before presenting our results and making recommendations to the subject matter experts, we investigated the sources of the Type II errors. Recall that the outlier classification procedure was applied to item-level tabulations. An outlying item tabulation may not necessarily be in an outlying ratio test. Indeed, if the two ratio items are highly correlated, the ratio test value may be consistent with the remainder of the distribution.

Our Type II error investigation used exploratory data analysis, graphing scatterplots of the ratio item pairs. The discussion below uses two current cell ratio tests from the mining sector. In both examples, the red circles, O, are the “true” outliers and the blue pluses, +, are “true” non-outliers.

Figure 1 presents a scatterplot of ending inventories to beginning inventories. This figure shows a strong linear relationship between the two variables. Moreover, the two values that are not on the regression line are clearly outliers. In this example, the subject matter experts identified these values as outliers, as did the ORT.

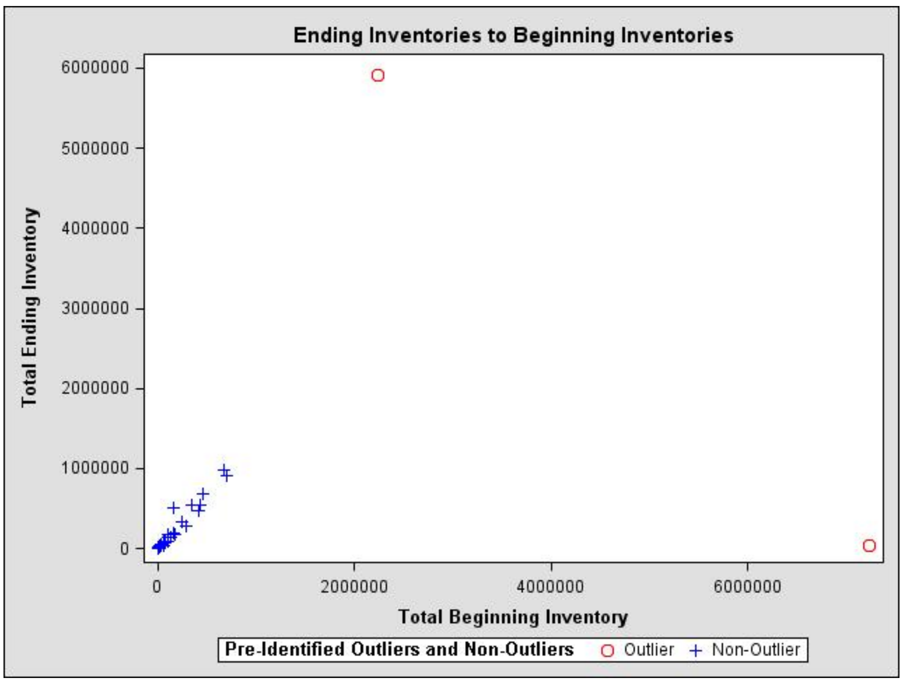

Figure 1: Mining total ending inventory to total beginning inventory

Figure 2 below presents a scatterplot of cost of materials to total receipts. Again, there is a strong linear relationship between the two variables. However, some of the “true” outliers are very close to the regression line and are relatively small in magnitude. The HB edit and the ORT will flag cell ratios that deviate from the median by a relatively large amount for small cells and a relatively small amount for large cells. In this distribution, some of the “true” outliers are near zero and are very close to the regression line. In fact, one of these “true” outliers equaled the median of this ratio test. These observations are unlikely to be identified as outliers by the ORT procedure (the median will never be identified), no matter what ratio test parameters are selected, nor should they be.

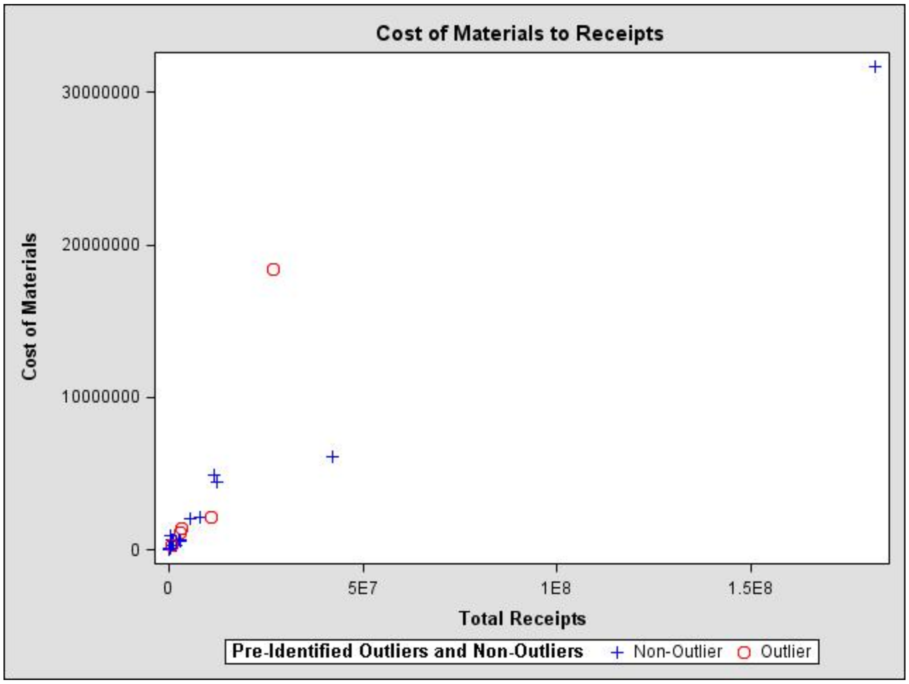

Figure 2: Mining cost of materials to total receipts

The graphical analysis was very useful in two ways. First, it helped us identify specific ratio tests whose efficiency could be improved by modifying the *u* and *c* parameters. More important, it demonstrated to us and to the subject matter experts that some of our "outliers' were actually "inliers" i.e., values that are incorrect but consistent with the remainder of the distribution (Cox et al, 1995, p. 393).

5. Further Investigation of Mining Sector Application

For the manufacturing and construction sectors, the ORT method always performed as well as the HB edit alone in terms of Type I and Type II error. In fact, the ORT method often outperformed the HB edit in these two sectors. This was not the case for the mining sector, where the HB edit alone identified more "true" outliers without flagging additional "false" outliers. To understand why this occurred, we examined each step of the computed HB statistics by step (Centering, Size Effect, and Quartile), focusing on the "true" outliers that were not identified with the ORT but were identified with the HB method.

Figure 3 plots the QESR (HB edit effects) and QSR (modified HB edit effects) for the mining ratio comparison of cost of materials to total receipts. In this plot, the blue pluses (+ = QSR) and red circles (O = QESR) are *paired* industry values. The x-axis is the numerator ratio value on the original scale; the y-axis is the range of the QESR ∪ QSR values, with a horizontal asymptote at $y = 0$. The blue shaded region is the ORT acceptance region for this ratio test.

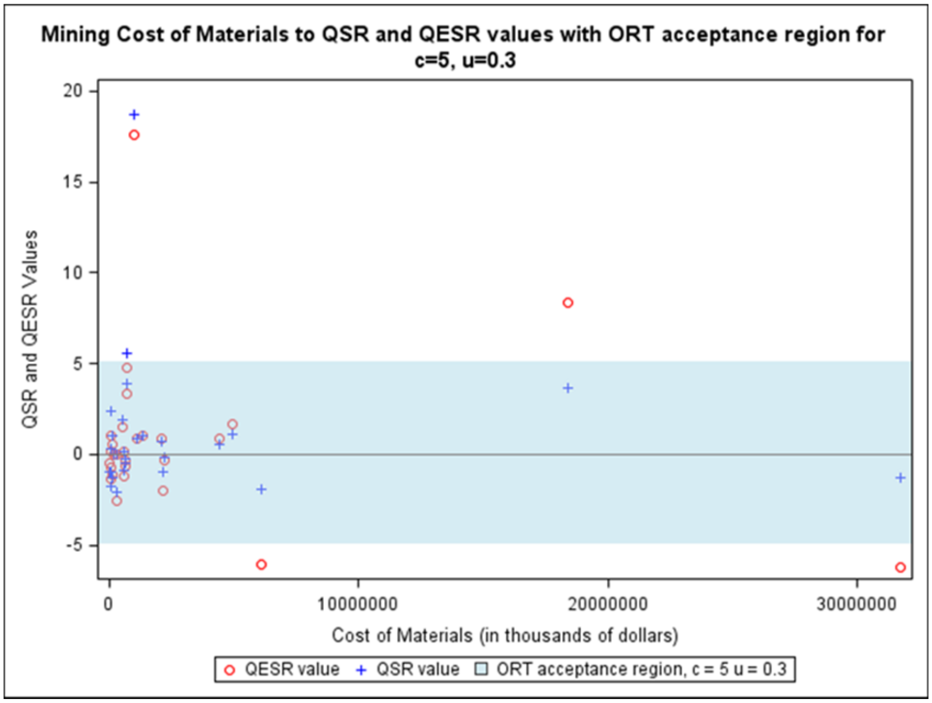

Figure 3: Analysis of mining ratio transformation for ORT and HB edit

Both the HB edit and the ORT method begin by centering the ratio at zero to obtain the SR values. In our HB edit application, the transformation step multiplies the SR by the size effect factor equal to the cube-root of max(numerator, $R_m \times$ denominator) to obtain the ESR. The mining sector contains 29 industries. With these data, the majority of the ratios are very close to zero after the centering transformation. Multiplying the centered values (the SR) by the size effect factor leaves these values relatively unchanged i.e., the ESR are very close to their corresponding SR value. However, as the size of the industry increases, the centered SR values are shifted by a large amount when multiplied by the size effect factor. Figure 3 illustrates this phenomenon. Here, the value of the cost of materials to total sales for the industries whose cost of materials value is greater than $10 billion are actually quite close to the ratio median by a modest amount, so the QSR values do not appear to be outlying. However, the tabulations themselves are relatively large, and the QESR values are likewise inflated. Since the ORT method only flags tabulations whose QSR *and* QESR values are outside their tolerances, these large values are not considered outliers by ORT, although they are flagged using the HB edit alone.

The mining sector data contain four very large industries (petroleum and natural gas industries). The ORT method prevents these large industry ratios from "dominating" the analysis, but fails to flag the large industries with moderately large ratios that are flagged by the HB edit alone.

Once we understood why the HB edit outperformed the ORT method in the mining sector, we briefly investigated the possibility of automatically selecting an outlier-detection method based on distributional properties such as skewness or kurtosis. However, our results were conditional on the analysts' identification of macro-level outliers, and their definitions of a "large change" differed by industry and by item as guided by their expertise. Our gold standard was subjectively determined and consequently, our results could differ using a different "gold standard" from the same data set.

In this case, the analysis *process* was more valuable than any global rule. Sigman (2005) applied seven different outlier-detection methods to 1997 Economic Census data from the services sectors (including three variations of the HB edit) and obtained feedback on the outlier-detection results from the subject matter experts. Preferences varied by sector. We did the converse for our analysis, and preferences likewise varied. The first approach would have several benefits in terms of identifying effective ratio edits and of setting tolerance thresholds (values of c). The second approach helped us to find a combination of ratio tests and parameters that minimized the all-Type II error, at the possible cost of individual ratio tests Type I error rates.

6. Conclusion

This paper presents the results of an evaluation study of macro-editing procedure used in production for services sectors portion of the Economic Census for the remaining three sectors. Our results illustrate the danger of applying a technique to a data set without careful analysis: the production method worked well for two of the three sectors but could be improved upon for the other. Moreover, in the two sectors that could use the production method, the input parameters were different.

The ORT has been modified for use by all sectors in the 2102 Economic Census. One of the original goals of the project sponsors was to reduce the amount of review while retaining similar (or better) quality. The ORT software identifies outlying tabulations, then provides micro-data review tools within the identified outliers (Sigman, 2005). This is a substantive departure from the current macro review procedures in the manufacturing, mining, and construction sectors, where the "top ten" individual values for each industry ratio are provided for review. Despite a favorable reaction by the managers to the ORT as applied to their data, many of the subject matter experts were still reluctant to "skip" the review of unidentified industries. Consequently, both ORT and the current production method will be used for the 2012 Economic Census.

The analysts' discomfort with a major change in review process left us reluctant to tackle another, possibly larger issue: the number of ratio tests being reviewed simultaneously in ORT by these sectors. We were provided with ten core current-cell ratios for the manufacturing and mining sectors and nine for the construction sector, along with ten historic-cell ratios for the manufacturing sector. There were also a number of non-core ratios that were not included in this investigation. Several of these ratio pairs have very poor correlation, making it difficult for any outlier-detection algorithm to be effective. We believe that this hypothesis is substantiated by the all-item Type II error results presented in Section 4.2. Perhaps after the analysts familiarize themselves with the ORT, they will be open to discussion on future modifications such as a reduced set of ratios.

Acknowledgements

The authors would like to thank Laura Bechtel, Xijian Liu, Demetra Lytras and Theresa Riddle for their careful review of earlier versions of this manuscript.

References

Belcher, R. (2003). Application of the Hidiroglou-Berthelot Edit to Periodic Business Surveys. *Proceedings of the Statistical Society of Canada Annual Meeting*.

Cox, B.G., Binder, D.A., Chinnappa, B.N., Christianson, A., Colledge, M.J., and Kott, P.S., eds (1995). Business Survey Methods. New York: Wiley.

Hunt, J., Johnson, J., and King, C. (1999). Detecting Outliers In The Monthly Retail Trade Survey Using the Hidiroglou Berthelot Method. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Hidiroglou, M.A. and Berthelot, J.M. (1986). Statistical Editing and Imputation for Periodic Business Surveys. *Survey Methodology*, **12**, pp. 73-83.

Lineback, J.F., Oliver, B., and Willimack, D.K. (2012) Developing Response Metrics for the Economic Census. *Proceedings of the 2012 Federal Committee on Statistical Methodology Research Conference*, Federal Committee on Statistical Methodology.

Thompson, K.J. and Sigman, R.S. (1999). Statistical Methods for Developing Ratio Edit Tolerances for Economic Data. *Journal of Official Statistics*, **15**, pp. 517 – 535.

Sigman, R. (1997). Development of a “Plain Vanilla” System for Editing Economic Census Data," UNECE Work Session on Statistical Data Editing, Working Paper no. 24, Prague, October 14-17, 1997.

Sigman, R.S. (2005). Statistical Methods Used to Detect Cell-Level and Respondent-Level Outliers in the 2002 Economic Census of the Services Sector. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.