RESEARCH REPORT SERIES
(Statistics #2007-20)

Selective Editing Strategies for the U.S. Census Bureau Foreign Trade Data

Maria M. Garcia
Alison Gajcowski*
Andrew Jennings*

Foreign Trade Division*

Statistical Research Division
U.S. Census Bureau
Washington, DC 20233

Report Issued: December 10, 2007

Selective Editing Strategies for the U. S. Census Bureau Foreign Trade Data[1]

Maria M. Garcia, Alison Gajcowski, and Andrew Jennings

ABSTRACT

The U.S. Census Bureau Foreign Trade Division (FTD) publishes monthly import and export statistics for the shipment of merchandise goods between foreign countries and the United States customs territories and foreign trade zones. The FTD publishes these statistics for over 17,000 import and 10,000 export commodity classifications. These data are not survey based, but collected from forms upon arrival or departure of merchandise goods. Data are edited and checked at every step of collection, processing, and tabulation. However, due to the limitations of the merchandise trade statistics program, monthly publication cells may still be subject to errors. We present score functions to rank edit failing records according to their potential impact on publication totals. In this report we present four separate score functions and the results of testing these score functions with the 2004 exports data.

1. INTRODUCTION

The Foreign Trade Division (FTD) at the U.S. Census Bureau is the official source for international merchandise trade statistics for the United States. The FTD publishes monthly import and export statistics for the shipment of merchandise goods between the United States and its international trading partners providing a comprehensive enumeration. Transactions are filed via electronic or paper means, mostly through the U.S. Customs and Border Protection (Customs). The collection of these data is unusual at the Census Bureau because they are filed upon arrival or departure of merchandise goods and are not based on surveys or censuses that are sent to respondents soliciting responses. The data are classified using the Harmonized Commodity Classification System (HS) by assigning a 10-digit code to each commodity. As of 2004, the division received about 3.4 million import records and 1.8 million export records every month covering 17,000 imports and 10,000 export commodity classifications.

These data are edited and checked for errors and quality assurance at every step of collection and processing. The data are subject to extensive micro-editing using the division's automated edit and imputation system that uses a parameter file called the Edit Master (EM). The EM verifies that numeric data fall within the prescribed ranges and that the ratios of highly correlated items fall within prescribed commodity bounds. Records that do not pass the edits are automatically imputed or manually reviewed and corrected based on a value of shipments threshold. However, imputation may not be successful for a small portion of the edit failing records. Records for which imputation failed are marked as "rejects" and require manual resolution. Each month, fewer than 0.5% of the five million import/export records are rejects. The analysts use their commodity expertise to manually adjust rejected records. They may also call back filers in an attempt to correct erroneous data.

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

Manual review and follow-up of suspicious units consume a large amount of the data editing resources. In selective editing, this cost is reduced by concentrating the review effort on erroneous units with a large potential impact on the publication totals. In this report we present research investigating the use of selective editing methodologies for the foreign trade statistics programs. Section 2 provides background on foreign trade data editing procedures. Section 3 describes the different methodologies we used and a weighting scheme for our data. In Section 4 we provide a discussion along with results of an application to the 2004 export transactions reporting. We close with a short summary in Section 5.

2. EDITING THE U.S. CENSUS BUREAU TRADE STATISTICS DATA

The U.S. Census Bureau processes import and export transactions and publishes the official international merchandise trade statistics for the United States. Data items collected include commodity, country of origin or destination, port of arrival or dispatch, value, quantity, and shipping weight. Data processing at the Census Bureau begins with extensive micro-editing using a parameter file called the Edit Master. Fescina et al. (2004) use historical data (currently using up to five years of data) to automatically update edit parameters based on the distributions of the data at the individual commodity level. For the value of shipments (V) and quantity (Q) items, they first compute the unit price of an observation, $p=V/Q$; next symmetrize the data using the log transform and compute quartiles of unit prices, q_1 and q_3 for each commodity's log-transformed unit price. Their method then identifies a record as being suspicious if the unit price p falls outside the interval $(q_1 - k(q_3 - q_1),\ q_3 + k(q_3 - q_1))$, where k is a constant. Recent analysis indicate that more reasonable limits may be produced when data are not symmetrized, and research is currently being done to fine tune any methodology to update edit master parameters.

More than 99 percent of edit failing records are corrected using the automated system. Records for which imputation is not successful are distributed by commodity and sent to subject matter experts for manual review. The commodity experts review a large number of records under tight time-constraints before the publication of monthly statistics deadline. Due to time and resource constraints, the division has an ongoing effort to improve the current procedures while preserving (or improving) data quality. These efforts focus on automatically updating edit parameters, outlier detection, and the research on selective editing described in this report.

3. SCORE FUNCTIONS FOR THE CENSUS BUREAU TRADE DATA

Records that are labeled suspicious using the method by Fescina et al. (2004) pass through the Edit Master and will either be automatically imputed or sent to the analysts for review. Our aim is not to re-engineer the current editing procedures but to add selective editing strategies for prioritizing manual review of suspicious records for which the Edit Master imputation procedure is not successful. In selective editing, a score function is used to rank edit failing records; records are then prioritized for review according to their score. The overall objective is to spend manual review resources on suspicious records that may have a significant impact on the estimates without affecting overall data quality. Greenberg and Petkunas (1986) report research for an economic survey in which as much as five percent of the erroneous units were responsible for 90 percent of the published estimates. They conclude that a thorough follow-up of all erroneous units had little effect on the final publication totals. Lindell (1997) reports on a study in which the highest ranked 20 percent of the erroneous records contribute to 90 percent of the total adjustment. Granquist and Kovar (1997) showed that selective editing can produce savings of 50 percent or more of the total editing cost while having a small impact on the final publication.

Latouche and Berthelot (1994) developed score functions for an annual retail trade survey and Lawrence and McDavitt (1994) presented a score function for a quarterly average weekly earnings survey. In both studies, data from previous survey cycles are required for developing score functions and the corresponding cut-off values. Thompson and Hostetter (2000) developed score functions for the U.S. Census Bureau Annual Survey of Manufactures using both data from previous collection cycle and administrative data when available. Jäder and Norberg (2005) developed a score function including measures of suspicion and potential impact for the Swedish foreign trade survey.

3.1 Flagging the most important variable

In developing score functions for prioritizing manual review of the Census Bureau's trade statistics data we considered scores previously tested at other institutes. For this data, it is not possible to assign scores at the commodity level due to the large number of commodities. Also, assigning scores at this level of aggregation is not feasible as there are not enough records for a meaningful ranking. Despite the large number of commodity classifications, we needed a score function to prioritize review of some observations at the 10-digit classification level since users may more closely monitor and scrutinize the statistics for particular types of commodities. Latouche and Berthelot (1992) suggest a simple score function that gives prominence to the most important variable. For our trade data, the analysts give higher importance to the variable representing the value of shipments (*V)* over the variables representing quantity (*Q*) and shipping weight (*SW*) of a shipment. For every observation *i*, let *r* index reported data, *e* edited data, and *cm* and *pm* be indexes representing current and previous month data respectively. Let Z_i be a counter for the number of items flagged as erroneous for record *i*. With *V* marked as the most important variable, the *Flag* score as described by Latouche and Berthelot (1992) is,

$$Flag_i = sqrt(\max(V^r_{i,cm}, V^e_{i,pm})) * Z_i \,.$$

For our trade data it is not possible to use this score function as described. For most commodities previous month data may not be available or comparable to current month data as companies may have *m* number of shipments the current month and $n \neq m$ (or no shipments) the previous month. Thus, we need to adapt *Flag* to using only current month data. *Flag* compares the current reported value of *V* for the unit, $V^r_{i,cm}$, with the unit's final data from the previous month, $V^e_{i,pm}$, as the best possible anticipated value of *V*. On his research on outlier detection for the Economic Census using the Hidiroglou-Berthelot method, Sigman (2005) noted that when using only current cycle ratios the median of ratios and reported data can be used to estimate an anticipated value for current month data. Let p_2 denote the median of current month unit price ratios $p_i = V^r_{i,cm} / Q^r_{i,cm}$. We use $p_2 * Q^r_{i,cm}$ as the best possible anticipated value of $V^r_{i,cm}$ instead of $V^e_{i,pm}$ in the maximization part of *Flag*,

$$Flag_i = sqrt(\max(V^r_{i,cm}, p_2 * Q^r_{i,cm})) * Z_i .$$

We also considered a composite *Flag* using the two most important variables, value (*V*) and quantity (*Q*). In this case we need to compute an estimate of *Q* for the current month. We first compute m_2, the median of quantity/shipping weight ratios $m_i = Q^r_{i,cm} / SW^r_{i,cm}$ for the current month. We then assume $m_2 * SW^r_{i,cm}$ is the best possible available estimate of $Q^r_{i,cm}$ and compute

$$CFlag_i = \{sqrt(\max(V^r_{i,cm}, p_2 * Q^r_{i,cm})) + sqrt(\max(Q^r_{i,cm}, m_2 * SW^r_{i,cm}))\} * Z_i .$$

Note that items in *CFlag* must be in the same unit of measurement before computing the score. We do this by using imputation factors which have been computed using commodity averages and are available in the foreign trade data Edit Master.

As we mentioned before, *Flag* and *CFlag* are to be applied to a set of key commodities at the 10-digit commodity classification level as chosen by the subject matter experts. Using this type of score at this level of aggregation is not feasible for the whole set of data.

3.2 Effect on publication totals

Our next score function is adapted from the *Diff* function described by Latouche and Berthelot (1992). *Diff* examines the effect of changes in the variables *V* and *Q* on the final totals by looking at the absolute difference between the current month reported values and the final values from the previous month. Since previous month data may not be available, the score function must be adapted to using only current month data. As in *Flag*, we assume $p_2 * Q^r_{i,cm}$ and $m_2 * SW^r_{i,cm}$ are the best possible available estimates of the anticipated current month value and quantity of shipments for unit *i* respectively. The *Diff* score as applied to current month ratios is,

$$Diff_i = \frac{abs(V^r_{i,cm} - p_2 * Q^r_{i,cm})}{Total(V_{cm})} + \frac{abs(Q^r_{i,cm} - m_2 * SW^r_{i,cm})}{Total(Q_{cm})} .$$

The estimated totals $Total(V_{cm})$ and $Total(Q_{cm})$ are calculated using final data for records accepted or automatically imputed by the automated system and reported data for the current month rejects for every observation within the commodity.

We may consider measuring only the effect of changes in the variable value of shipments *V* over the total value of shipments,

$$Diff_i = abs(V_{i,cm}^r - p_2 * Q_{i,cm}^r) / Total(V_{cm}) .$$

In this case *Diff* is similar to the measure of impact within the score function developed by Jäder and Norberg (2005) for the Swedish trade data.

3.3 Hidiroglou-Berthelot method

The Hidiroglou-Berthelot (HB) method uses historical ratios to identify suspicious records (Hidiroglou and Berthelot, 1986). The HB edit as applied to our data begins with the current month unit price ratio, $p_i = V_{i,cm} / Q_{i,cm}$, and p_2, the median of unit prices. The unit price ratios are then transformed to ensure outliers are identified at both end of the distributions,

$$S_i = \begin{cases} p_i / p_2 - 1 \; if \;\; p_i \geq p_2 \\ 1 - p_2 / p_i \;\; if \;\; 0 < p_i < p_2 \end{cases},$$

Hidiroglou and Berthelot suggest applying another transformation that ensures more importance will be placed on a small deviation within a large unit as opposed to a large deviation within a small unit. As with *Flag* and *Diff*, for application to this data the transformation must be adapted to using only current month ratios. The transformation as applied to our current month unit price ratios is,

$$E_i = S_i * \{\max(V_i, p_2 * Q_i)\}^u ,$$

where $0 \leq u \leq 1$. A value of $u = 0.5$ as in the maximization part of *Flag* seems to work well for our data.

We then calculate a measure of the distance of the first and third quartile of the transformed unit price ratios from the median. Let q_1, q_2, and q_3 be the first quartile, the median and the third quartile of the transformed unit price ratios respectively. Calculate

$$d_{q_1} = \max(q_2 - q_1, abs(a * q_2))$$
$$d_{q_3} = \max(q_3 - q_2, abs(a * q_2))$$

Then, assign to every observation a score that is a ratio with a factor measuring displacement of the transformed unit prices from the median, weighted by the appropriate distance from the median,

$$Ratio_i = \begin{cases} (q_2 - E_i) / d_{q_1} \;\; if \;\; E_i < q_2 \\ (E_i - q_2) / d_{q_3} \;\; if \;\; E_i > q_2 \end{cases}$$

According to Hidiroglou and Berthelot the $abs(a * q_2)$ term in the calculation of the distances ensures that d_{q_1} and d_{q_3} are not too small for observations clustered about the median. In our application we used a value of a = .05 as suggested by Hidiroglou and Berthelot.

We also considered a simple variation of *Ratio* using the log transformed unit price ratios as are computed during parameter development (Fescina et al., 2004). In this case we calculate the quartiles of the transformed unit price ratios q_1, q_2, and q_3 using the log transformed ratios instead of the E_i 's described previously before computing d_{q_1}, d_{q_3}, and *Ratio*.

3.4 Combine Hidiroglou-Berthelot edit and Effect on publication totals

We wanted to include another score function based on the score function developed by Jäder and Norberg (2005). They report on successfully implementing a score function for the Swedish trade data as a weighted geometric mean of measures for suspicion, suspicion of errors in *V* over errors in *Q*, and potential impact (See Jäder and Norberg (2005) for details), $Score = Suspicion * Suspicion(V\ over\ Q)^{weight(SuspV)} * Impact^{weight(Imp)}$. For our data, it was decided to not include a measure of suspicion of errors in value of shipments *V* over errors in quantity of shipment *Q* into the scores, but we can use the idea of using the geometric mean to calculate an alternative score using the product of the score based on the HB method (*Ratio*) and the effect on publication totals (*Diff*) described above. [Note: We have not tried assigning weights to separate contributions of *Ratio* and *Diff* yet.] For every observation we compute a new score as, $RatioDiff_i = Ratio_i * Diff_i$.

3.5 Weights

Our team worked together with senior analysts assigned to review and resolve edit failures to design a suitable weighting scheme for the data. The analysts reported considering the size of the company in terms of value of shipments when resolving suspicious records: they recognize when value of shipments is large enough to affect tabulations and thus warrants more attention during manual review. Also, analysts' workload is stacked by section (a section is an aggregate of commodities). Using the analysts recommendations, we assign to every observation an importance weight (W_v) based on the value of the variable value of shipments by commodity and a separate weight measuring the importance of key commodities (W_c) (key commodities as classified by senior analysts) by section. We incorporate these weights into the score functions. For example, in the case of *Flag*, the score will includes the importance weight for the variable *V* (W_v) and the importance weight for the commodity (W_c),

$$Flag_i = sqrt(\max(V^r_{i,cm}, p_2 * Q^r_{i,cm})) * Z_i * W_v * W_c$$

4. PRELIMINARY RESULTS and DISCUSSION

In this paper we presented score functions for prioritizing manual review of foreign trade data records identified as suspicious by the automated editing system. Selective editing is probably a misleading term: Our aim is different from the traditional selective editing goal of manually reviewing records with a significant impact on tabulations with all other records handled by an automated system. We did not start out to develop a new editing strategy to identify suspicious records, but to prioritize manual review of records that had already been labeled as suspicious. The editing process will be a two-tiered flow system in which the current month suspicious records are identified and fields marked to be changed are automatically imputed. Then, records for which automatic imputation is not successful would be marked as rejects and assigned a score to provide a ranking for guiding clerical review. The bulk of manual review resources would be spent on the most important records; however we have the mandate that all rejected records are to be reviewed.

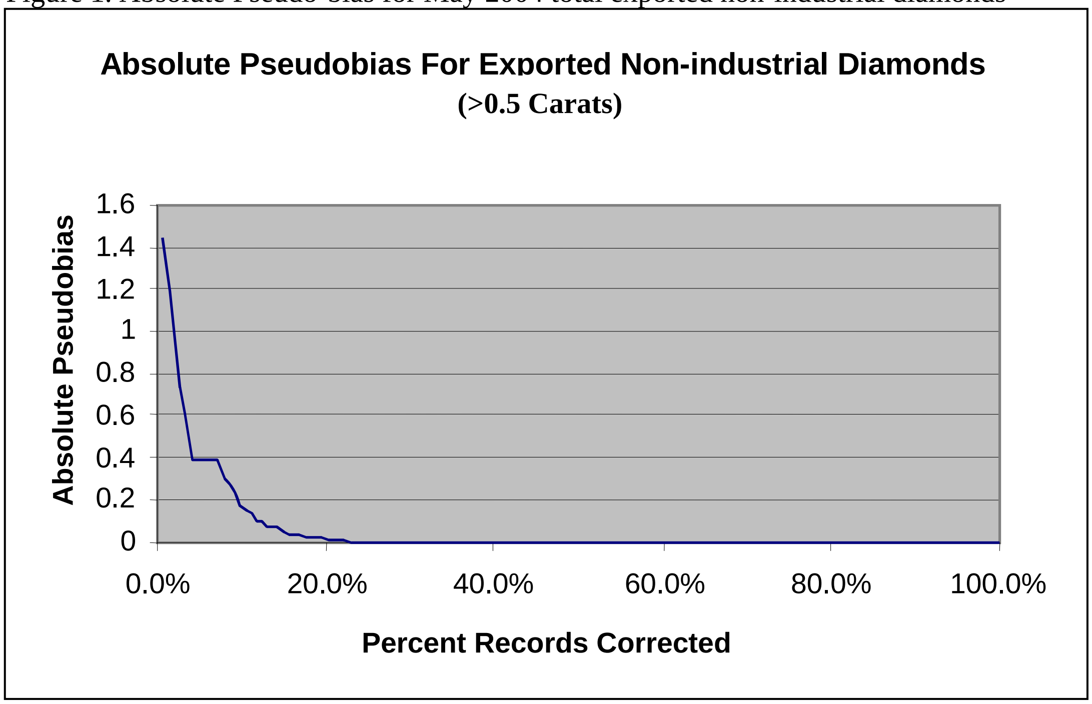
At the beginning of this project we were to focus on prioritizing manual review of records within a small group of commodities requiring more thorough scrutiny. The *Flag* score is applied only at the 10-digit classification level for this small number of products for which end-users need more detailed statistics. Since we are scoring only rejected records, for most commodities there are not enough observations to apply a score function at this level of aggregation. We must group rejected records at higher aggregation levels. Rejects are classified by commodities and sent to the analysts by sections (commodity groupings). On average each analyst must correct over 600 records per month. Since analysts' workload is stacked by section, we calculate the scores to rank records by section. We do not cross-classify the data with other fields (e.g. country of export, mode of transport, port of dispatch) as this will further reduce the number of records for most commodity groupings. The scores *Diff*, *Ratio*, and *RatioDiff* can be computed at different domains or levels of aggregation.

We have available archived raw and final data for the 2004 monthly exports transactions reports for products shipments from the United States to foreign countries. The data file has flags for fields marked to be imputed and for records for which automatic imputation was not successful (rejects). The file also contains data for several items including value of shipments, quantity of shipments, shipping weight, country of destination, mode of transport, and port of dispatch among others. The medians and quartiles of items and estimated totals in score functions are computed using current month raw data for rejected records and final current month data for all other records. This is possible in our application as items marked to be changed in suspicious records are automatically imputed before labeling records as rejects. (Note: Historical data is used for computing quartiles of data in the measures of suspicion; see Fescina et al., 2004)

Since we are mandated to review all rejected records, comparisons between the total value of shipments and quantity of shipments obtained after reviewing records using the different scores are not possible (all records are manually reviewed). A question remains: how do we evaluate the effectiveness of the score functions? In the case of *Flag*, which is used when there are enough records within a particular commodity (say at least 50) for which more detailed statistics are needed we looked at the absolute pseudo-bias. Latouche and Berthelot (1992) define absolute pseudo-bias as $abs(T_E - T_F)/T_F$, where T_F is the final publication total and T_E is the estimated total obtained by replacing raw values in records with a score larger than a certain cut-off value with the final data while keeping raw values for records with a score lower than the cut-off value.

Despite the large number of commodity classifications, we needed a score function to prioritize review of some observations at the 10-digit commodity classification level since users may more closely monitor and scrutinize the statistics for particular types of products that are shipped regardless of their value. For example, import and export of diamonds merit a thorough review at this level of aggregation. The Kimberly Process Act was designed to stop the trade in conflict diamonds, diamonds that may have contributed to violent conflict and human rights abuses in their countries of origin. Filers are required to report the Kimberly Process Certificate number for the exports/imports and re-exports/ re-imports of rough diamonds. Figure 1 displays the absolute pseudo-bias for the variable quantity of shipment (*Q*) of exported non-industrial diamonds in May 2004 using the score function *Flag*. The graph illustrates how the pseudo-bias rapidly decreases as the percentage of records marked for clerical review increases, and review of more than 20 percent of the rejected records with the highest scores does not affect the final estimate. Technically we could stop reviewing records at the 20 percent level of review when the effect of changes on the absolute pseudobias approaches zero and the estimated total approaches the final publication total.

Figure 1. Absolute Pseudo-bias for May 2004 total exported non-industrial diamonds



We note that in production all records are reviewed; however comparing the estimated and final totals gives us an idea of how well the score function is tracking the most influential observations. Since we are looking at the most detailed level (commodity level) the distribution of scores using *Flag* tend to have a similar shape from month to month. In the above example it is possible to expect that if manual review of rejected records proceeds up to the top 20 percent of the ranked records, then the same 20 percent cut-off value could be used at the next cycle. For the other score functions, where records are ranked at higher levels of aggregations, we expect distributions of scores to change in consecutive months due to the magnitude and complexity of the data, thus a similar analysis cannot be used for determining cut-off values.

Figures 2 and 3 display distributions of the *RatioDiff* scores for the Foods sector for May and June 2004 (Note: there are some hidden observations at both ends of the distributions). Using Excel graphical tools we zoom in an area of the graph in which there is an inflection point and the slope of the graph had begun to approach zero. We use the graphical tool to fit a trend line to the distribution of scores, and then fit a 95% Confidence Interval around the trend line by finding a confidence interval for the trend values corresponding to each point in the distribution of scores. To find confidence intervals for the trend points in the fitted trend line, we calculate $trend\ point \pm 1.96 * std\ dev / \sqrt{n}$, where n is the number of observations in the area of the graph we are zooming in on. Connecting the upper and lower 95% confidence intervals for the trend values gives us the range of possible scores in which we are 95% sure that the scores will fall within. We can calculate an ad-hoc cut-off value depending upon where the distribution of scores stopped leaving this 95% Confidence Interval. As the trend line would change depending on the area of the graph the graphing tool is zooming in on, estimating a cut-off value this way is rather arbitrary. However, since we have a mandate to review all rejected records determining cut-off values for the proportion of records to follow-up is no longer an issue.

Figure 2: Trend Lines for the distribution of scores for the May 2004 Food Sector using *RatioDiff*

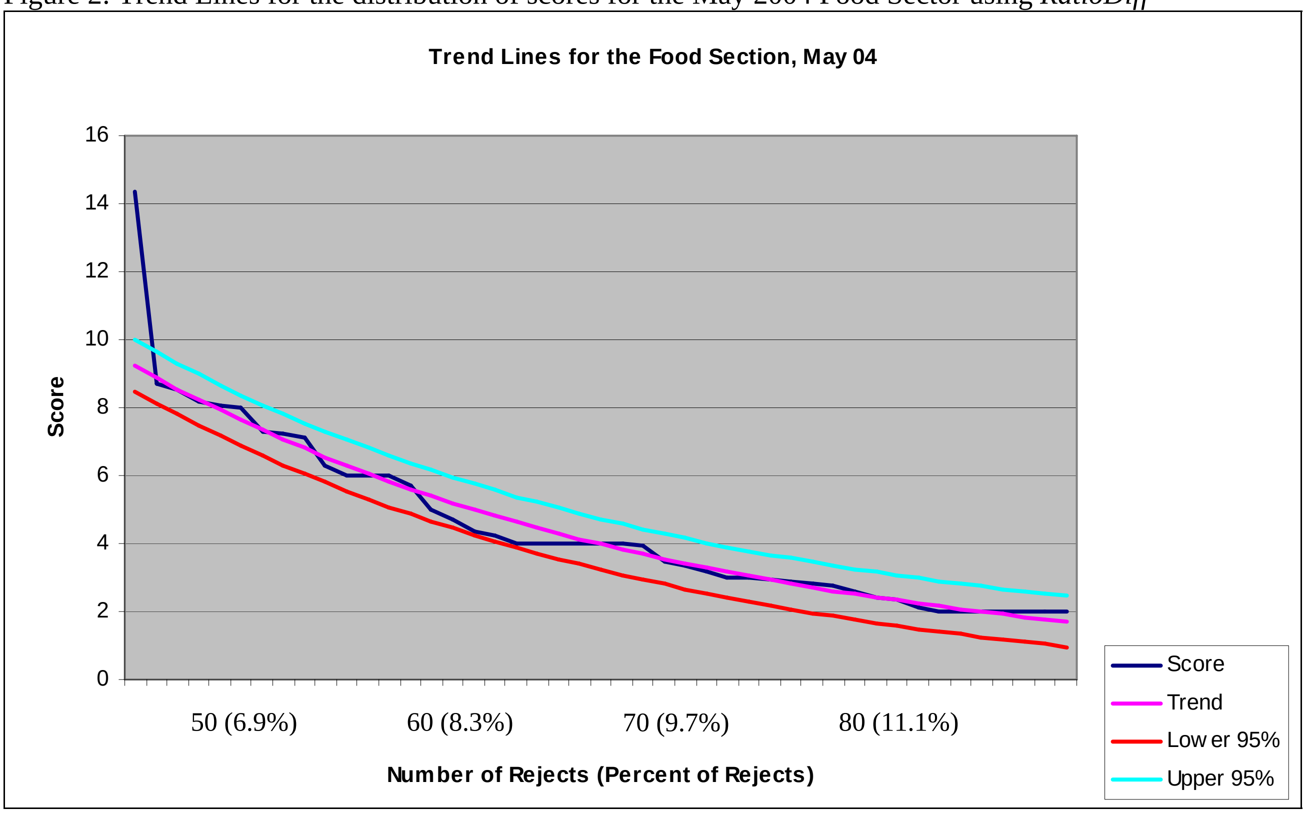
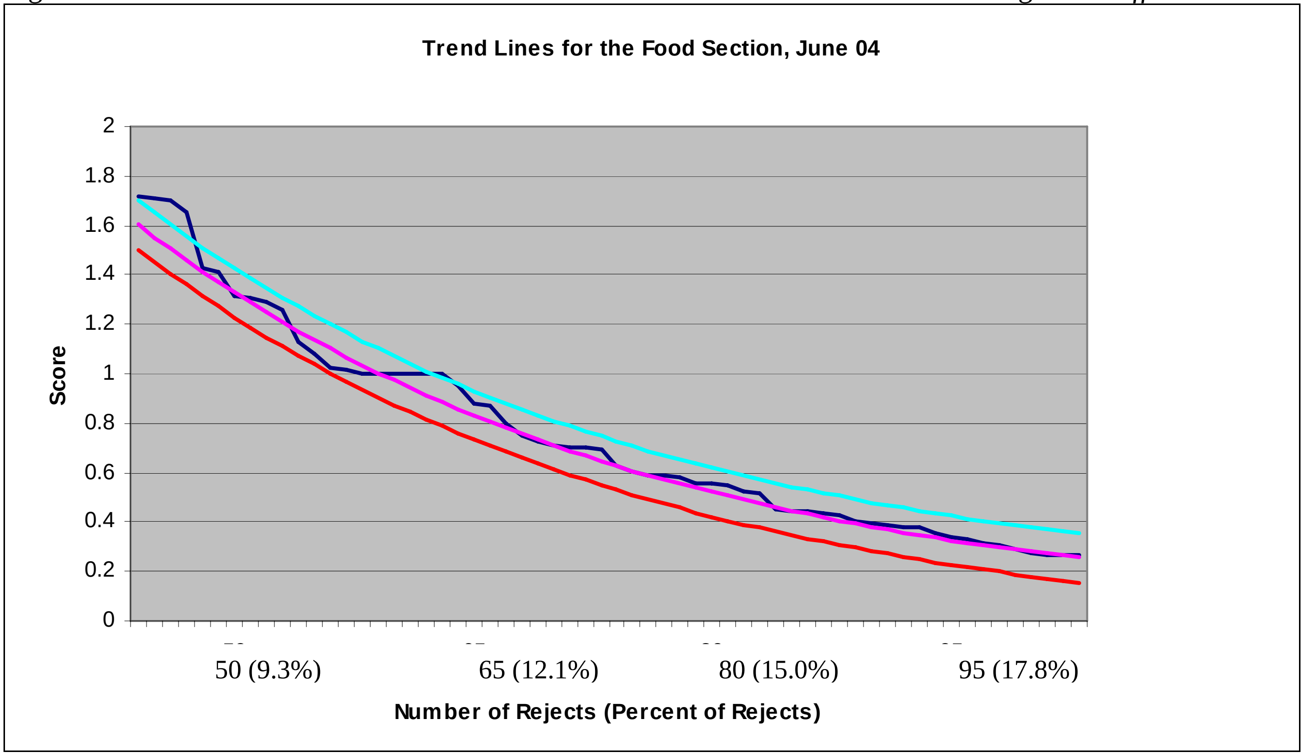
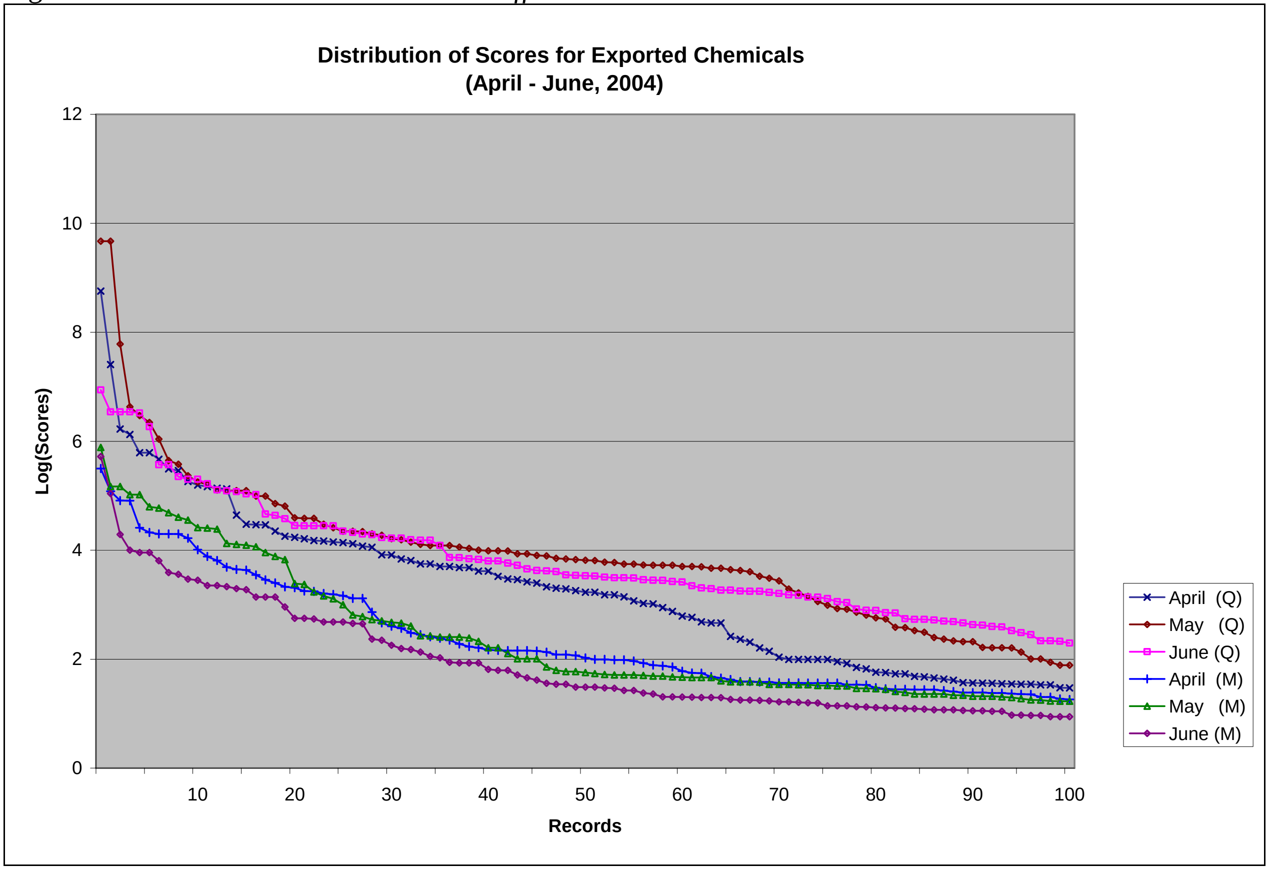


Figure 3: Trend Lines for the distribution of scores for the June 2004 Food Sector using *RatioDiff*



We previously mentioned the reporting patterns of filers are not consistent on a monthly basis: every month filers submit a different number of transactions (or no transactions). Thus, the monthly distributions of scores are expected to be different from one month to the next. Each month the trade data review process will use this methodology for scoring rejected records and assigning priority on a ranked list of rejects for guiding manual review. The score function *RatioDiff* uses measures of suspicion based on displacement of an observation from either the closest quartile or from the median to assign a score to every rejected record. Figure 4 displays a three-month historical distribution (there are hidden observations at both tails of the distributions) of the log(scores) for *RatioDiff* using both the median (M) and quartiles (Q) models for all chemicals exported from the United States to foreign countries during the period April – June, 2004. For all time periods the *RatioDiff* scores based on displacement from the median (labeled month-M) settled down into a banding pattern (around record number 45.) Meanwhile the *RatioDiff* scores based on quartiles (labeled month-Q) has larger fluctuations and the distance between the scores is larger from one month to the next. We observed similar behavior when plotting for other time periods. The *RatioDiff* distributions where suspicion is calculated using displacement from the median are more similar across consecutive time periods and may be a better choice for our data.

Figure 4: Historical Distribution of *RatioDiff*



5. CUSTOMER FEEDBACK

As part of our project we prepared a detailed briefing for our customers. Our team provided documentation including detailed output listings of the selective editing methodology with the ranked records for some selected sections. After careful review of the prioritized listings, the FTD methodology experts questioned the high rankings given to records that by experience they consider insignificant to final cell estimates. We examined some of these records along with all other records within the same data cell. Table 1 displays an example of this situation for commodity code representing "Blocks, Tiles, and Similar Refractory Ceramic Goods of Clay NESOI exported in March 2004". The selective editing methodology identified a record (shaded in gray) as having a large impact on the aggregated unit price (*V/Q*). The current editing system considers this record insignificant because it bases the impact the record will have on aggregated totals by value (*V*) alone: Analysts in general manually correct the record if the value field fails a range edit (i.e. it's below a fixed dollar value depending on the commodity.) Thus this record would be low priority during analysts' review. However, the difference between fixing the record and not fixing the record has a tremendous impact on the Quantity (*Q*) field. According to our methodology the impact this record has on the final total Quantity is,

$$abs(Final(Q) - Reported(Q)) / Total(Q) * 100 = 8595$$

In this case, although the record has a low value, it has a large impact (8595 percent) at the ten digit commodity level and will be identified by the selective editing methodology as having a very high priority on the ranked list of records as desired.

Table 1: Blocks, Tiles, and Similar Refractory Ceramic Goods of Clay NESOI (March 2004)

	Total Value (*V*)	Total Quantity (*Q*)	Unit Price	V/Q	Bounds
				Lower	Upper
Reported (10 records)	$102,190	7,217	$14.15	90	3000
Reported Suspicious Record	$3,024	7,144	$0.42	90	3000
Final Suspicious Record	$3,024	10	$302.40	90	3000
Final (10 records)	$102,190	83	$1,231.20	90	3000

Our next example illustrates the opposite situation: fixing a record that has a slight impact on the final estimates. For the commodity code representing "Glass mirrors unframed, not vehicle rearview mirrors" there are three failing records identified as rejects by our current methodology (i.e. imputation was not successful). Analysts' review will fix all three records so that the aggregated unit price falls between the prescribed bounds (see second row). However the selective editing methodology identified only two records as having a significant impact on the final totals. Fixing only these two records (see third row) brings the aggregated unit price within the optimal bounds for this commodity as desired. Using the current editing methodology, analysts will fix all three records even though the third only had a slight impact on the aggregated publication cell.

Table 2: Glass Mirrors Unframed, Not Vehicle Rearview Mirror (March 2004)
Examining 3 Rejects Out of 128 Records

	Total Value	Total Quantity	Unit Price	V/Q	Bounds
				Lower	Upper
Three Rejects, 87 records imputed	$3,142,622	129,973,502	$0.02	0.25	50
Final with all 3 Rejects corrected	$3,142,622	1,230,629	$2.55	0.25	50
Selective Editing, two highest ranked records corrected	$3,142,622	1,804,699	$1.74	0.25	50

5. SUMMARY

In this report we presented research on developing selective editing strategies for the Census Bureau foreign trade statistics programs. In traditional selective editing methods, previous cycle data are used in the score functions; this is not possible for the trade data. We adapted available score functions to using only current cycle data by computing an estimate of the anticipated value of the variables. We presented four separate score functions that can be implemented at different levels of aggregations. Our computer program assigns a score to every observation and provides a ranked listing for guiding clerical review of rejected records. The ranking is based on a score function that includes measures for how suspicious the record is and the potential effect it has on the final estimates.

Feedback from our customers includes an interest in using the methodology earlier in the editing process. Current research includes investigating the feasibility of using the score function to identify which records should be imputed automatically and which records will go directly to the analysts for manual review. This means the most erroneous records are identified without the use of parameters. Records with low scores would go through the edit master and would be automatically imputed by the editing system. Records with a high potential impact on the final estimates corresponding to high score values would be sent directly to the analyst for manual correction. This process would minimize the number of rejects identified by the edit master. It would also allow a more efficient target of records for review in which analysts will only review records that will have a significant effect on publication cells. A ranked review process may also improve data quality: it provides a ranked order of rejected records which ensures the most resources are spent on the most significant observations.

REFERENCES

Fescina, R., Jennings A., Wroblewski, M. (2004). "Automated Production of Foreign Trade Data Parameters Using Resistant Fences". Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods, American Statistical Association.

Granquist, L. and Kovar, J. (1997). "Editing of Survey Data: How much is enough?" *Survey Measurement and Process Quality*, L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, and D. Trewin (Editors). Wiley, NY.

Greenberg, B. and Petkunas, T. (1986). "An Evaluation of Edit and Imputation Procedures Used in the 1982 Economic Censuses in Business Division," SRD Report RR-86/04, U.S. Census Bureau. Washington, D.C.

Hidiroglou, M., and Berthelot, J. (1986). "Statistical Editing and Imputation for Periodic Business Surveys". Survey Methodology, V. 12, No. 1, 1986.

Jäder, A. and Norberg, A., (2005). "A Selective Editing Method Considering Both Suspicion and Potential Impact Developed and Applied to the Swedish Foreign Trade Statistics," UNECE Work Session on Statistical Data Editing, Ottawa, Canada.

Latouche, M. and Berthelot, J., (1992). "Use of a Score Function to Prioritize and Limit Recontacts in Editing Business Surveys." JOS, V.8 No. 3.

Lawrence, D. and McDavitt, C.,(1994). "Significance Editing in the Australian Survey of Average Weekly Earnings. JOS, V.10 No.4.

Lindell, K. (1997). "Impact of Editing on the Salary Statistics for Employees in County Council." UN Economic Commission for Europe, Statistical Data Editing, vol. 2, Geneva, pp 2-7.

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

Thompson, K. and Hostetter, S., (2000). “Investigation of Selective Editing Procedures for the U.S. Bureau of the Census Economic Programs.” Proceedings of the Second International Conference of Establishment Surveys.

ACKNOWLEDGEMENTS

The authors would like to thank Debra Coaxum for her support to this project. We also wish to thank David Dickerson, Ryan Fescina, Yves Thibaudeau, and William Winkler for their helpful comments on earlier versions of this report.