Investigation of Selective Editing Procedures for the Annual Survey of Government Finances

Loretta A. McKenzie, Terri L. Craig, Carma R. Hogue
Office of Statistical Methods and Research for Economic Programs,
U.S. Census Bureau,[1] Washington, DC 20233-9100

Abstract

The U.S. Census Bureau re-engineered the edit procedures for the Annual Survey of Government Finances. Beginning with the 2004 survey, the Census Bureau investigated the use of selective editing. A score was assigned to each individual government unit based on the effect that a change in the government's reported data would have on the final estimates. Using predetermined critical score values, those government units with the largest potential impact on the estimates were pinpointed as candidates for manual review. Because the survey form covers a large number of variables, some of which are very volatile from one year to the next, determining which variables to use to develop a score for a governmental unit was problematic. We researched several score functions involving different combinations of these variables. For each score function we set edit acceptance rates and compared absolute pseudo-biases of resulting estimates to identify edit acceptance rates that would reduce edit burden and maintain quality. This paper reviews the selective editing technique, the edit research process, and the problems that we encountered in attempting to apply it to the Annual Survey of Government Finances. We used empirical research methods to conclude that other micro-editing techniques should be used prior to selective editing. Data sources were the 2002 Census of Governments Finances and the 2004 Annual Survey of Government Finances.

Keywords: selective editing, score function, edit referral rate, edit acceptance rate

1. Introduction

In an effort to improve the estimates for the Annual Survey of Government Finances (ASGF) while optimizing costs, a team of analysts and mathematical statisticians examined various editing methodologies. Analysts were spending much of their time making data corrections that had little or no impact on the final estimates. Their valuable time could be spent in other survey-related activities, like non-respondent follow-up activities.

The team examined the consistency edits to determine if they were overlapping or totally covered by other consistency or ratio edits. After unduplicating edits, the team examined various methods (Hidiroglou-Berthelot, resistant fences, asymmetric fences) of calculating the ratio edit bounds. The team also examined the possible use of selective editing.

In this paper, Section 2 gives the background of the survey. Section 3 is a review of the edit research process. Section 4 covers the selective editing methodology. Section 5 covers our case study. Sections 6 and 7 cover the results and conclusions, respectively. Section 8 covers future research.

2. Background

The ASGF publishes detailed financial data for revenue, expenditures, assets, and debt from state and local governments (cities, counties, townships, special districts, and independent school districts) each year. The survey mails out in October and requests data for the government's fiscal year ending between July 1 of the previous year and June 30 of the survey year.

In years ending in '2' and '7' , a census of all state and local governments is done. In all other years, a sample of cities, counties, townships, and special districts is surveyed. All independent school districts (about 14,000) and state governments are surveyed every year. The total sample size including all types of government (states, cities, counties, towns, special districts, and independent school districts) is about 25,000.

---

[1]This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

Cities, counties, and townships are called general purpose governments since they perform multiple functions (fire protection, highway administration, welfare, police protection, etc.). Special district governments are fiscally and administratively independent entities that generally have only one purpose. These governments include airport authorities, sewer districts, utilities, etc. Each special district is categorized by a function code that indicates its function. There are some special districts that have more than one function, like sewer and water supply or natural resources and water supply. These are referred to as multi-function special districts. All other special districts are single-function.

The items collected on the questionnaire are the same for census and sample years. Data on several hundred variables are provided by state and local governments. The variables include various taxes, charges, assessments, intergovernmental revenue, utility revenue, liquor store revenue, and insurance trust revenue. Capital outlay and current operation expenditures are collected for various functions (governmental administration, utilities, safety, transportation, etc.). Long-term and short-term debts, as well as data on various cash and security holdings, are also included in the data collection. For more information on the survey, go to http://www.census.gov/govs/www/financegen.html. For information on the sample design, see "Sampling Procedure for the Annual Survey of Government Finances" by McLaughlin and McKenzie (2006) which is available upon request.

For survey year 2005, all of the questionnaires used to collect the data were redesigned. With a major redesign of the questions, we also needed to redesign the edit system. The aforementioned team was formed to overhaul the existing edit system and to research the best methods for detecting edit failures.

3. Edit Research

The goal of survey editing has been for the analyst to isolate and correct as much of the non-sampling error as possible. In the past few years, survey methodologists in government agencies have been researching how to reduce survey costs, produce estimates more efficiently, and reduce respondent burden all at the same time. Researchers have been questioning the value of this extensive editing. If the effort to validate the data does not noticeably affect the final tabulated results, the cost is not justified. Current edit procedures for the ASGF consist of a thorough micro-editing of each survey unit. Micro-editing, which is very costly, finds errors by inspecting individual units.

Several edit methodologies can be used that could potentially increase the efficiency of the ASGF edit process. For the ASGF we researched three edit methodologies: the Hidiroglou and Berthelot (HB) edit parameter determination, the resistant fences (RF) methods of determining parameter bounds, and selective editing to determine units that should be reviewed. We analyzed the effectiveness of each method to determine which method would best suit the ASGF data. Our research is explained in more detail in Cornett, McLaughlin, and Hogue (2006). This paper looks at the HB and RF methods of determining edit bounds.

Selective editing is designed to identify those possible erroneous units that have the most impact on the survey estimates. It identifies a whole unit for review, but all of the variables on the questionable record must be reviewed. Selective editing does not isolate erroneous variables for questionable records. Since the ASGF has numerous variables for review, selective editing was researched to see if it could be used as an editing tool to prioritize previously identified edit failures for edit follow-up.

4. Selective Editing Methodology

The edited micro-data is passed to a macro-editing process, which detects errors at some aggregate level based on the responses of a number of units. Selective editing can help connect the two by focusing on editing an individual unit response, but prioritizing error resolution by determining how much the error potentially impacts the survey estimates. The goal is to improve quality by targeting editing resources on the responses that have the greatest effect on the survey estimates. Selective editing can then rank survey units by the importance that their possible error affects the survey outcome. It is a tool that helps move valuable resources away from cleaning unimportant errors to improving other parts of the survey process, such as additional macro-editing.

Selective editing scores each individual government unit on its possible effect on the estimates, and determines critical values used to isolate the units with the largest potential impact on the estimates as candidates for manual review. Every unit is allocated a score that measures the relative importance of resolving its error in comparison to the other survey responses. Therefore, this procedure targets only a small part of the micro-data for the analyst to re-contact and validate. Surveys that require clean micro-data for all units should not use this method. Selective editing is useful when the reason to collect the data is to produce estimates rather than micro-data.

Using the simulation approach detailed and described in Section 5 by Lawrence and McKenzie (2000), we investigated the effectiveness of selective editing using several score functions. This approach evaluates, in terms of the final survey estimates, several editing models. In particular, it focuses on each model's effectiveness in isolating units for review.

Expected amended values are estimates of an item response; they are not necessarily the correct or final values. Expected amended values are calculated for the raw (unedited, reported) data. Scarrott (2005) says these values can be based on historical data or responses from similar units. A local score for each survey item of interest is determined. The local score is defined to be the expected change for a particular variable that would come about if the raw data were amended (i.e., not the final data after editing). Fatal edits, such as blank required items, often are resolved prior to selective editing. A single measurement, called a global score, is calculated for each survey unit. A global score is a combination of a unit's local scores. Cell specific critical values, or cut-off scores, are computed using a simulation study and the global scores. Examples of cells are size-class categories, type of government, or industry. The distribution of the global scores is used to determine the cell specific critical values. The unit's global score is compared to its cell specific critical value. A unit with a global score greater than its cell specific critical value is selected for manual review, otherwise the data are accepted. Follow-up attention is then given to the survey returns with the highest global scores.

Research has shown selective editing to be an effective method of reducing survey editing without affecting the quality of the final survey estimates. Latouche and Berthelot (1992) found that for a business survey of 987 questionnaires, 40% can be computer edited, 20% re-contacted (based on the selective editing score), and 40% left uncorrected, and the simulated survey estimates will be within 0.1% of the final estimates for frequently reported variables. Latouche and Berthelot (1992) also show selective editing to be effective for continuous variables with skewed error distributions. In skewed error distributions, a small number of units have error that is so large in magnitude that it impacts the survey estimates. When Thompson and Hostetter (2000) performed a feasibility study for U.S. Census Bureau economic programs, they found a local and global score function that worked well for two sequential years of the Annual Survey of Manufacturers. Lawrence and McDavitt (1994) showed significance editing reduced respondent burden and re-contact, while also providing an ordering so that the most important edit queries could be addressed first. Selective editing can be implemented in batches when the data are received, since most implementations only require a unit's current period and prior period responses. This is important for time critical applications.

5. Case Study

Our study used raw and edited data from the 2002 Census of Government Finances and the 2004 ASGF sample weights. The 2002 Finance Census raw and edited data files and the 2004 sample unit file were used to compute the local and global scores. We only used the units from the 2002 Census that were in the 2004 sample. The data file created from the 2002 edited file with the 2004 sample weights was considered to be the expected amended values.

The survey form covers a large number of variables, including governmental functions, and some are very volatile. Determining which variables to use to develop a global score for a governmental unit was problematic. Several global scores were made using different categories of variables. They were compared in terms of absolute pseudo-bias (APB) and acceptance rate. The acceptance rate is 100 times the number of cases accepted as reported divided by the number of sample cases. One score contained five fundamental aggregate variables, including all revenue and expenditure variables at a high aggregate level. Other scores considered different aggregate variables to specifically research volatility; others included more variables in the global score.

The survey units were ranked by the global score. We chose several arbitrary values of percentile *p*, where *p* denotes the editing workload. A smaller *p* means a larger editing workload, resolving more cases. For example at the 45^{th} percentile, 55% of the units are referred for review. Critical values were computed for each global score and value of *p*. This procedure was done for *p*=45, 55, 65, 75, and 85. Critical values were set in accordance with the amount of editing that will be performed. A unit was flagged for review if its global score was above the critical value or cut-off. The units with the highest global score were selected for editing. A composite survey returns file was created, for each value of *p*, containing the edited 2002 ASGF data for the top *(100- p)%* of governments and the raw 2002 ASGF data for all other governments (*p*%). We re-computed the estimates using each composite file.

Thompson and Hostetter (2000) discuss fatal versus nonfatal edit failures. Examples of fatal referrals include blank items and misclassifications which should always be resolved by imputation or analyst investigation. Thompson and Hostetter (2000), define z_j to be a 0/1 indicator for edit failures. An item is said to fail an edit if the reported value ≠ edited value for non blank reported cases, then z_j is set to one. Otherwise, z_j is set to zero. LaTouche and Berthelot (1992) also set $z_j = 1$ if it is labeled as suspicious, otherwise it is 0. The ASGF does not have an "in-between" file that identifies the edit failures. We do not currently flag cases that fail edits but are verified to be good data. Since the data for many of the edit failures are not changed, it was difficult to distinguish failures. Therefore, we set all $z_j = 1$.

We compared all the different score functions in terms of the APB and acceptance rate. The APB due to editing only select units was defined to be:

$$\frac{\hat{Y}_p - \hat{Y}_{100}}{\hat{Y}_{100}},$$

where: $\hat{Y}_p$ is the estimate of the data item calculated by replacing all unedited reported values where a global score function was larger than the critical value with their edited values and leaving the rest of the unedited reported values unchanged.

$\hat{Y}_{100}$ is the estimate of the data item calculated by replacing all of the unedited, reported values with edited values.

We considered an APB of less than 10% due to not editing to be acceptable.

5.1 Defining the Local and Global Score Functions

Reviewing past literature shows most global scores are comprised of four to five local scores. Latouche and Berthelot (1992) suggests considering four major elements when developing score functions. These elements are the size of the responding unit, the size and number of suspicious data items, and the relative importance of variables as determined by the subject matter specialists. We used the following categories, *c*, of aggregate variables for global scores.

Category A: non-trust revenue, general expenditures, total taxes, total charges, current operations expenditures

Category B: non-trust revenue, general expenditures, total taxes, total charges, highway expenditures, health and hospital expenditures, criminal justice expenditures, housing expenditures, welfare expenditures, utilities expenditures

Category C: total long term debt, total capital outlay, intergovernmental revenue, intergovernmental expenditures

Category D: non-trust revenue, general expenditures, total long term debt, total assets

Category A scores consist of aggregate revenue and expenditure variables. Category B scores consist of 10 of the more important detail aggregate variables. We felt this could help reduce the APB, since more detail about the governmental unit was being included in its global score. Category C scores consist of highly volatile variables. Category D contains variables that were to be collected during a third non-response follow-up. We expected to have these variables for total and partial respondents.

A local score was computed for each aggregate variable included in any global score. For every unit and each of the 16 unique variables above, we calculated the local score two ways. LS will refer to local score and GS will refer to global score. The first local score is the scaled absolute difference between the edited 2002 data and unedited 2002 data (LS Equation 1). For the second local score, we calculated the scaled square root of the maximum of edited 2002 data and unedited 2002 data (LS Equation 2). The former will be known as DIFF, and the latter known as SQRT.

LS Equation 1 (DIFF): $LS_{1ij} = \dfrac{\left|x_{ij} - x_{ij}^{*}\right|}{z_j \left|\sum_j w_j x_{ij}^{*}\right|}$ and

LS Equation 2 (SQRT): $LS_{2ij} = \dfrac{\left[MAX(x_{ij}, x_{ij}^{*})\right]^{1/2}}{z_j \left|\sum_j w_j x_{ij}^{*}\right|}$,

where: i = 1,... ,16 (the variable of interest),
j = 1,... , n (the unit or ID of interest),

x_{ij} is the 2002 amended data for variable i, unit j,

x_{ij}^{*} is the 2002 reported data for variable i, unit j,

z_j = 1 for edit failures and 0 otherwise, and

w_j = weight for unit j.

Dividing by the absolute value, to scale the local score, compensates for variables that have a different order of magnitude. SQRT was found promising by Thompson and Hostetter (2000). DIFF emphasizes the absolute discrepancy between the reported and amended data and is suggested in Lawrence and McKenzie (2000).

We chose two methods to calculate the global score from each type of local score. We used the maximum local score for each unit (GS Equation 1) and the sum of the local scores for each unit (GS Equation 2). The former will be called MAX and the latter SUM. MAX was used by Lawrence and McDavitt (1994) and gives attention to a substantial error in any one variable.

GS Equation 1 (MAX): $GS_{1kc \cdot j} = MAX\left(LS_{kij}\right)$, $i \in c$, and

GS Equation 2 (SUM): $GS_{2kc \cdot j} = \sum_i LS_{kij}$, $i \in c$,

where: k = 1,2 (the local score),
c = A, B, C, or D (the category),
i = 1,... ,16 (the variable of interest), and
j = 1,... , n (the unit or ID of interest).

We had four categories, two types of local scores, and two types of global scores, so 16 global scores were computed per unit. We prepared 16 sets of estimates, using the composite files, to compare in terms of APB for each value of p.

6. Results

General purpose governments were researched separately from special districts and supplements. General purpose governments, which perform multiple governmental functions, can have over a hundred variables because of the myriad of functions they perform. On the other hand, a special district government usually has one function such as an airport authority, so there are not as many reported variables. A supplement supplies additional data for a general purpose government. Examples would be hospitals that would supply expenditures, revenues, etc. Therefore, a supplement is much like a special district in the amount of data that it reports. For general purpose governments, cut-off scores (critical values) were set at the state by type of government level. However, due to time and resource constraints, we analyzed the composite estimates at the national by type of government level. For special district governments, cut-off scores were set and the composite estimates were analyzed at the function level. This analysis was as detailed as we could get for special districts, because of the scantiness of the data for some functions in the sample.

6.1 General Purpose Governments

We considered using selective editing as an editing tool by itself for general purpose governments for fiscal year 2005. There are 16 variables of interest that we analyzed at the estimate level for each global score. Forty-eight states have counties, 49 have cities, and 20 states have towns. We looked at 5 values of acceptance rates (for p = 45,55,65,75,85). We computed an APB for each type of government, state, variable, global score, and acceptance rate. So, for counties there were 768 (48 x 16) APB calculations per global score and acceptance rate. There were 784 APB calculations for cities, and 320 for towns.

We found for all levels and for all types of general purpose governments, category B (having the largest number of variables), produced the greatest number of score calculations with an APB less than 10%. Global scores computed using category B almost always ranked in the top two of 16 global scores. Using a score with more variables (category B) reduced the APB. The global score calculated using the sum of the local scores tended to work better than the maximum of the local scores. Using the scaled difference local score worked better than the scaled square root of the maximum. The SUM and DIFF scores gave more APB calculations less than 10% for almost every value of p and type of government. Table 1 shows the top performing score by type of government and acceptance rate. All scores in the table use category B since it always performed the best.

Table 1: Top Score by Type of Government for General Purpose Governments. All global scores use category B.

Type	p (% of cases accepted as reported)	Top Global Score	Top Local Score	Percent of APB calculations < 10%	Percent Change of APB calculations < 10% from next value of p
Counties	45	SUM	DIFF	81.7%	5.97%
	55	SUM	DIFF	77.1 %	7.83%
	65	SUM	DIFF	71.5%	13.13%
	75	MAX	DIFF	63.2%	17.25%
	85	SUM	DIFF	53.9%	
Cities	45	SUM	DIFF	94.6%	5.00%
	55	SUM	DIFF	90.1%	5.26%
	65	SUM	DIFF	85.6%	9.18%
	75	SUM	DIFF	78.4%	19.15%
	85	SUM	DIFF	65.8%	
Towns	45	SUM	DIFF	95.9%	3.68%
	55	MAX	DIFF	92.5%	2.77%
	65	MAX	DIFF	90.0%	7.91%
	75	SUM	DIFF	83.4%	13.62%
	85	SUM	DIFF	73.4%	

As additional analyst burden is added, the percentage of APB calculations for the top score less than 10% increases. As discussed in Lawrence and McDavitt (1994), the key is to find that point at which continuing to query referrals will not substantially increase the accuracy of the final estimates. We proposed a 45% acceptance rate for counties and cities and a 55% acceptance rate for towns. When we looked at general purpose governments using 45% raw data and 55% edited data, for counties, the applicable variables had an APB less than 10% in 8 out of 10 cases. Cities (p=45) and towns (p=55) had better than 9 out of 10 cases with an APB less than 10%.

Since there were a lot of variables for general purpose governments, we found that it was difficult for an analyst to isolate what items to validate from a failed unit. Selective editing targets units for further review that exceed the critical value, but does not isolate the variables that should be checked. For this case selective editing would not really reduce the editing workload unless some other method was used to isolate the problem variables.

We next researched selective editing as a way to prioritize the ratio edit failures. Currently analysts are given all the ratio edit failures to review. When isolating the units that would have the most effect on the final estimates, we ranked the ratio edit failures on the amount of change that would result from resolving units with an edit referral. Only units that failed a ratio edit would be candidates for selective editing. The variable totals became those from all the units with a ratio edit failure. We re-evaluated all scores. Since all failed ratio edits are given back to the analyst for review, the value of p was not important. We set p=45 to determine the critical values. We didn't have enough cases to determine critical values at the state by type of government level so we used a grouping of states by type of government as determined by the subject matter specialists. See Attachment 1 for the state groupings used. Scores were again evaluated at the national by type of government level. For counties there were 12 state groupings, 13 for cities, and 6 for towns. There were, respectively by type of government, 192, 208, and 96 possible APB calculations per global score. Table 2 summarizes the results.

Table 2: Selective Editing Results for General Purpose Government Ratio Edit Failures, p=45

Counties				Cities				Towns			
Category	GS	LS	Percent APB < 10%	Category	GS	LS	Percent APB < 10%	Category	GS	LS	Percent APB < 10%
B	MAX	DIFF	96.88 %	A	MAX	DIFF	98.08 %	B	MAX	DIFF	94.79 %
D	SUM	DIFF	96.88 %	A	SUM	DIFF	98.08 %	B	SUM	DIFF	94.79%
A	SUM	DIFF	96.35 %	B	SUM	DIFF	97.60 %	C	SUM	SQRT	92.70%
D	MAX	DIFF	96.35 %	B	MAX	DIFF	97.12 %	D	SUM	DIFF	92.70%
A	MAX	DIFF	95.31 %	D	MAX	DIFF	96.63 %	B	MAX	SQRT	91.67%
D	MAX	SQRT	95.31 %	D	SUM	DIFF	96.63 %	B	SUM	SQRT	91.67%
B	SUM	DIFF	94.27 %	C	SUM	DIFF	96.15%	D	MAX	DIFF	91.67%
D	SUM	SQRT	94.27 %	C	MAX	DIFF	94.23%	A	MAX	DIFF	90.63%
C	MAX	DIFF	93.23 %	D	MAX	SQRT	93.75%	A	SUM	DIFF	90.63%
C	SUM	DIFF	93.23 %	D	SUM	SQRT	93.75%	C	MAX	SQRT	90.63%
A	MAX	SQRT	91.15 %	A	SUM	SQRT	92.30%	A	MAX	SQRT	89.58%
A	SUM	SQRT	90.62%	B	SUM	SQRT	92.30%	C	MAX	DIFF	89.58%
B	SUM	SQRT	90.62 %	B	MAX	SQRT	91.83%	C	SUM	DIFF	89.58%
B	MAX	SQRT	90.10 %	C	MAX	SQRT	90.38%	D	MAX	SQRT	89.58%
C	SUM	SQRT	86.46 %	C	SUM	SQRT	90.38%	D	SUM	SQRT	89.58%
C	MAX	SQRT	85.42 %	A	MAX	SQRT	89.90%	A	SUM	SQRT	88.54%

For counties, we chose the category B score using MAX and DIFF. Although the SUM, DIFF for category D score was tied with the same number of APB calculations less than 10%, we selected category B, MAX and DIFF because the category B score had a slightly less problematic APB. Both scores had a high APB for utilities for state group 10 (100%). Variable IG (intergovernmental) expenditures did not fare well with our top scores for counties, but it is not used in any of our current consistency ratio edits. When looking at all scores together, the category B, A, and D scores performed better than category C, as expected. DIFF worked better than SQRT. MAX and SUM performed equally well.

For cities, two category A scores gave the highest number of APB calculations less than 10%. DIFF was better than SQRT, but MAX and SUM worked equally well. Two of the category B scores were close to the top score.

Notably, these scores had high APBs, for group 5 states, for total long term debt (13.86%) and total capital outlay (15.33%).

For towns we also chose the category B score using MAX and DIFF. Categories C and D scores performed similarly. The score using more variables appeared to help towns more than counties and cities. All four category B global scores were in the top six global scores for towns.

When prioritizing the ratio edit failures for general purpose governments, MAX was the better method to use for determining the global score, but when using all the data SUM was better. DIFF was always better than MAX for determining the local score. For general purpose governments category B produced the most APB calculations less than 10%, except for cities when only using the ratio edit failures, where category A was the best.

6.2 Special Districts

Since special district and supplements report on far fewer variables, it is easier to isolate variables to query for a failed unit. For each function code, we researched how each applicable variable performed amongst the 16 global scores. Setting p=85 (i.e., 85% raw and 15% edited), we found we could choose local and global scores in such a way that at the function level, most variables had an APB less than 10%. The most promising score used category A, the DIFF method for determining the local score, and the MAX method for calculating the global score. Unlike general purpose governments, it was rarely better to use category B with the larger number of variables.

We still had problematic variables and functions. The scores and functions where the category A, MAX, and DIFF did not perform the best are detailed in Table 3. For example, solid waste management had the most APB calculations less than 10% using category B, MAX, and DIFF. We sometimes found the APB for known volatile variables such as capital outlay was consistently high. Using the top score for solid waste management(B, MAX, and DIFF), there was still a high APB of 18.23% for the variable total capital outlay. For Other Single-Function Districts, two scores tied having the most APB calculations less than 10% using category B and MAX. Definitions for all function codes can be found at the link: http://www.census.gov/govs/www/class_ch4.html#s4.5.

Table 3: Functions where Selective Editing performed differently for Special Districts

Function	Category	Top Global Score	Top Local Score
Highways	D	MAX	DIFF
Sewerage	B	MAX	DIFF
Solid Waste Mgmt.	B	MAX	DIFF
Other Single-Function Special Districts	B B	MAX MAX	SQRT DIFF

For certain functions, such as water supply utility, we could refer less than 15% of the data and have all variables with an acceptable APB less than 10% using category A, DIFF, and MAX. For special districts and supplements, our composite estimates having 85% raw data and 15% edited data produced acceptable APBs.

7. Conclusions

Selective editing seems more promising for surveys with a smaller number of variables. For surveys with a large number of variables, selective editing may require referring more cases to produce acceptable APBs.

We found that for general purpose governments, which have much longer questionnaires, using a longer score, such as category B, helps the APB. Even though selective editing was not a practical tool by itself, the subject matter analysts could use selective editing to review the historical ratio edits ranked by their global score. Selective editing indicated which ratio edit failures were the most important to validate, meaning they would most likely have the greatest effect on the estimates. Analysts felt it would be a promising tool to help them prioritize their work, especially during a census year when there is a heavier workload.

We found that for most special district government functions, referring less than 20% of the data yielded an APB less than 10% for most applicable variables. Selective editing by itself could be a valuable tool for reducing costs and analyst burden for special district governments.

8. Further Research

We recommend analyzing global scores at the state group and function level. State groups may behave differently when analyzed separately. In the future, we plan to compare the units isolated using selective editing with the ratio edit failures for special districts.

Acknowledgements

The authors would sincerely like to thank Rita Petroni, Joanna Fane McLaughlin, Yang Cheng, and Mary Christine Davies, for their assistance in reviewing this abstract and paper. The authors would also like to thank Steve Owens, Aaron Gray, Stan Batton, and Elizabeth Cornett for their participation on this project.

References

Cornett, E., McLaughlin, J. F., and Hogue, C. (2006). "A Comparison of Two Ratio Edit Methods for the Annual Survey of Government Finances." *Proceedings of the Joint Statistical Meetings, 6-10 August, Seattle.*

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*, Vol. 8, pp.389-400.

Lawrence, D., and McDavitt, C. (1994). "Significance Editing in the Australian Survey of Average Weekly Earnings." *Journal of Official Statistics*, Vol. 10, pp. 437- 447.

Lawrence, D., and McKenzie, R. (2000). "The General Application of Significance Editing." *Journal of Official Statistics*, 16, pp. 243-255.

McLaughlin, J.F., and McKenzie, L. (2006). "The Sampling Procedure for the Annual Survey of Government Finances." *Internal Document.*

Scarrott, C. (2005). "Feasibility Study: A Review of Selective Editing." *Technical Report, University of Canterbury, New Zealand.*

Thompson, K.J., 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 on Establishment Surveys, 17-21 June, Buffalo, NY.*

Attachment 1

State and Type Groups for Editing

For counties (type 1) the following groups were used:

Group 1:	Alabama, Arkansas, and Mississippi
Group 2:	Alaska, Maryland, North Carolina, Tennessee, and Virginia
Group 3:	Arizona, Nevada, Oklahoma, Texas and New Mexico
Group 4:	Colorado, Idaho, Montana, Utah, and Wyoming
Group 5:	Delaware, New Jersey, New York, and Pennsylvania
Group 6:	Florida, Georgia, Louisiana, and South Carolina
Group 7:	Illinois, Indiana, Iowa, Michigan, Minnesota, and Wisconsin
Group 8:	Kentucky, Ohio, and West Virginia
Group 9:	Maine, Massachusetts, New Hampshire, and Vermont
Group 10:	Hawaii, Oregon, and Washington
Group 11:	North Dakota, and South Dakota
Group 12:	Kansas, Missouri, and Nebraska

For municipalities (type 2) the following groups were used:

Group 1:	Alabama, Arkansas, and Mississippi
Group 2:	Alaska, Tennessee, and Virginia
Group 3:	Arizona, Nevada, Oklahoma, Texas and New Mexico
Group 4:	Colorado, Idaho, Montana, Utah, and Wyoming
Group 5:	Delaware, New Jersey, New York, and Pennsylvania
Group 6:	Florida, Georgia, Louisiana, and South Carolina
Group 7:	Illinois, Indiana, Michigan, Minnesota, and Wisconsin
Group 8:	Kentucky, Ohio, and West Virginia
Group 9:	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont
Group 10:	Oregon, and Washington
Group 11:	North Dakota, and South Dakota
Group 12:	Iowa, Kansas, Missouri, and Nebraska
Group 13:	Maryland, and North Carolina

For townships (type 3) the following groups were used:

Group 1:	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont
Group 2:	Illinois, Indiana
Group 3:	Pennsylvania, New Jersey, and New York
Group 4:	Kansas, Missouri, and Nebraska
Group 5:	North Dakota, and South Dakota
Group 6:	Michigan, Minnesota, Ohio, and Wisconsin