ALTERNATIVE EDITING AND ESTIMATION PROCEDURES FOR THE ADVANCE MONTHLY RETAIL TRADE SURVEY

Julia L. Bienias, Timothy J. Braam, and William C. Davie, Jr., U.S. Bureau of the Census
William C. Davie, Jr., U. S. Bureau of the Census, SVSD Rm. 2653-3, Washington, DC 20233-6500

KEY WORDS: Outlier detection, Exploratory data analysis, Hidiroglou-Berthelot, Nonresponse adjustment

1. Overview

Editing is an important step in producing estimates from survey data. Accurately identifying the values most likely to be in error is an essential part of efficient editing. This is particularly important in establishment surveys, where the data are often highly skewed. Graphical methods have been used to improve the efficiency and accuracy of the editing process (e.g., Houston & Bruce, 1992; Hughes, McDermid, & Linacre, 1990). Previous work (e.g., Granquist, 1990; Lee, 1995; also Bienias et al., 1994) has demonstrated the applicability of various outlier detection methods to economic and skewed data.

For the Advance Monthly Retail Trade Survey (MARTS), accuracy and timeliness are crucial. From this survey are obtained the earliest estimates of month-to-month change in retail sales, published on the ninth working day following the reference month.

The current editing system includes both automated consistency-type edits and analyst review. Although many tools are used, there is still a fair amount of clerical review of listings required in the editing process. In this paper we report results from an extensive comparison of different methods for automated outlier detection for this survey. We focused on methods to be applied after data collection close-out, when there is no time in the tight MARTS time schedule for further follow-up of questionable cases.

We included current as well as new methods in our research, focusing on approaches that were multi-faceted and could be implemented easily in a production environment. Methods we discuss include: influence measures from a linear regression, graphical-based methods (e.g., based on box plots), multi-way scatter plots, and combining multiple criteria using the Mahalanobis distance statistic. We examined the impact of a given case on the overall estimate as well as its relationship to other cases in the sample.

We also present results from a modified estimation procedure that is designed to better account for the effect of nonresponse.

2. The Advance Monthly Retail Trade Survey

2.1 Scope, Sample Design, and Sample Size

The MARTS sample is a small probability subsample selected from units in the Monthly Retail Trade Survey (MRTS). The scope of both surveys is all employer firms engaged in retail trade, as defined by Major Groups 52-59 of the 1987 edition of the *Standard Industrial Classification Manual* (Ofc. of Mgmt. & Budget, 1987). The MRTS is a mail survey (with telephone follow-up) of approximately 13,300 firms (U.S. Bur. of the Census, 1996). The MARTS sample contains approximately 3,400 retail firms.

Businesses selected for the MARTS are asked to report their monthly retail sales right after the end of each month, much earlier than units only in the MRTS.

2.2 Estimation

The primary objective of the MARTS is the estimation of month-to-month change in sales. The estimation procedure is described here.

For each detailed kind-of-business (KB) level in the MARTS, a ratio is formed by dividing the sum of the weighted current month sales by the sum of the weighted previous month sales for all units that reported sales data for both the current and previous months. No attempt is made to impute sales data for the nonresponding businesses. The ratio is then multiplied by the previous month MRTS estimate of total sales at the appropriate KB level to arrive at a total sales estimate for the current month. This is a link-relative estimation procedure (a ratio multiplied by a benchmarked total); the ratio carries forward a more accurate benchmarked total (in this case, from the MRTS). At present there are 16 detailed KB levels, each of which covers several types of retail activity. Total sales estimates at broader KB levels (e.g., durables, nondurables, total) are obtained by aggregation.

Month-to-month change estimates at all KB levels are obtained by dividing the total sales estimate from the MARTS for the current month by the total sales estimate from the MRTS for the previous month. For a detailed KB, this is equivalent to the ratio based only on the MARTS data. The published trend is computed from the ratio (trend = (ratio - 1) * 100).

This paper reports general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau. We thank Howard Hogan, Pat Cantwell, Nash Monsour, Ruth Detlefsen, and Dave desJardins for their helpful comments.

From the MARTS and the MRTS come three different estimates of month-to-month trend. The earliest, based on data collected from the MARTS sample, we will refer to as the **Advance trend**. Approximately one month later the **Preliminary trend** is published, based on data collected from the MRTS sample. The Preliminary trend is considered a revision to the Advance trend, and large revisions are highly publicized in the press (e.g., *The Wall Street Journal*, *USA Today*). From the MRTS also comes a **Final trend**, which is published approximately one month after the Preliminary trend.

3. Alternative Editing Methods

3.1 Overview

The purpose of this research was to identify an automated (to the extent possible) outlier detection method for units reporting in the MARTS. By "outlier," we mean any unit whose month-to-month ratio is either extreme with respect to the ratio of other similar units or has a large influence on the resultant Advance estimate of trend. Our goal was to find a reliable method that could be used to identify any outlying cases still remaining after all analyst review and follow-up had taken place. Because of the time constraints, these cases would then be excluded from the estimation procedure.

Our approach was to first investigate both univariate and multivariate outlier detection methods in a few detailed kinds of business and one or two months. If any patterns emerged, we then attempted to generalize the ideas and apply them to other months and then to other KBs. For each method, we re-computed the Advance trend after excluding those cases identified as outliers, and compared the result to the Preliminary trend. We used the difference between these two trends as a basis for comparing the outlier detection methods we examined.

For all of the analyses described here, we used data **unadjusted for seasonality**.

3.2 Outlier Detection Methods Considered

We considered the following methods to identify outliers. Define "CM" to be the weighted sales for a given case for the current (latest reference) month. "PM" is the same quantity for a given case for the prior month.

- **Ratio for individual case** k: $R_k = \dfrac{CM_k}{PM_k}$.

- **Transformed ratio for individual case** k:

  This is based on a two-step method proposed by Hidiroglou and Berthelot (1986; see also Lee, 1995). Let R_k be the ratio for the individual case, as defined above. Let R_M be the median of the ratios. Then for each case k define

$$S_k = \begin{cases} 1 - \dfrac{R_M}{R_k}, & \text{if } 0 < R_k < R_M \\[2ex] \dfrac{R_k}{R_M} - 1, & \text{if } R_k \geq R_M \end{cases} .$$

Next, compute

$$T_k = S_k \, [\max(CM_k, PM_k)]^V , \; 0 \leq V \leq 1 .$$

The exponent V in the transformation provides control over the importance of the magnitude of the data. We used a value of 0.5 for V.

- **Effect** of case k within a particular KB group is defined to be the difference between the overall ratio and the ratio excluding case k:

$$E_k = \frac{\sum_{i=1}^{n} CM_i}{\sum_{i=1}^{n} PM_i} - \frac{\left(\sum_{i=1}^{n} CM_i\right) - CM_k}{\left(\sum_{i=1}^{n} PM_i\right) - PM_k} ,$$

where the sum is taken over all n cases having both CM and PM data.

- **Market share** for current month attributed to case k:

$$M_k = \frac{CM_k}{\sum_{i=1}^{n} CM_i} .$$

We included this as a measure of the importance of a case with respect to its segment of the economy.

- **Cook's distance** (an influence measure) from a linear regression of CM on PM:

$$D_k = \left\{ \hat{\beta} - \hat{\beta}_{(k)} \right\}' X'X \left\{ \hat{\beta} - \hat{\beta}_{(k)} \right\} ,$$

where β is the $p \times 1$ vector of regression parameters, **X** is the $n \times p$ design matrix, and "(k)" means the parameters were estimated excluding case k. (See Draper & Smith, 1981, p. 170.)

- **Mahalanobis distance** of case k, combining several univariate criteria:

$$H_k = \sqrt{(\mathbf{x}_k - \bar{\mathbf{x}}) \, \hat{\Sigma}^{-1} \, (\mathbf{x}_k - \bar{\mathbf{x}})^T} ,$$

where $\bar{\mathbf{x}}$ is a vector of means of all the cases on several univariate criteria, $\mathbf{x}_k$ is a vector containing case k's values on those criteria, and $\hat{\Sigma}$ is the usual variance-covariance matrix. See Bienias (1995) for an application of this method to outlier detection.

4. Comparison of Editing Methods

We examined box plots and multi-way scatter plots of the individual statistics and box plots of the Mahalanobis distance statistic. Box-and-whisker plots allow quick visual analysis of the location, spread, and shape of a distribution (see Tukey, 1977; Hoaglin, Mosteller, and Tukey, 1983). Automating this, we flagged as potential outliers those cases that would fall beyond an outer fence (or extended whisker) on a box plot. For each criterion (cutoff value) considered, we computed the Advance trend after dropping cases that were past the cutoff. (An extension of this would be to exclude cases based on the box plot criterion, but never drop more than a specified % of n cases.) Then we compared the difference between the Preliminary and re-computed Advance trends. Ideally, we would have liked to have seen the difference become essentially zero or a non-zero constant value for a particular measure and outlier detection cutoff.

Below we report results from the measures described in Section 3.2 for the Apparel kind of business.

- **Ratio for individual case k:**
  We examined box plots of this measure. Although very intuitive, this measure has a number of serious drawbacks. Distributions of month-to-month change ratios are usually skewed (e.g., outliers on the left tail of the distribution may be harder to identify than those on the right tail of the distribution) and are often more variable for small businesses than for large businesses (Hidiroglou & Berthelot, 1986; Lee, 1995). This means that ratios for small businesses are more likely to be identified as outliers. However, the large businesses contribute more to the ratio and thus should get more attention. This is called the *size masking effect*. Figure 1 is a plot of the ratio of an individual case versus its market share for one month for Apparel.

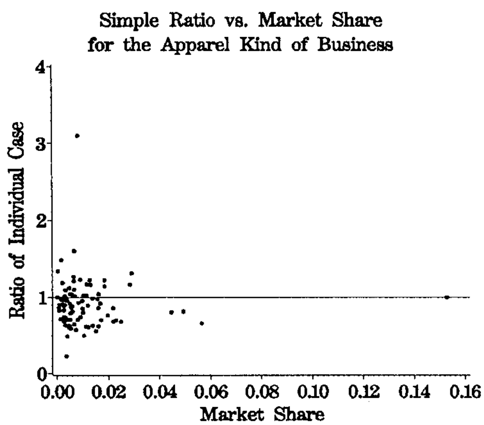

Figure 1

This plot illustrates the inability of this measure to account for the relative size of a case. Therefore, we rejected this as an effective outlier detection method.

- **Transformed ratio for individual case k:**
  The Hidiroglou-Berthelot (H-B) transformation attempts to correct for the problems of using the ratio of an individual case as a measure for outlier detection. As with the preceding method, we examined box plots of this statistic.

  We compared the untransformed and transformed ratios. For each measure, we computed the Advance trend after excluding cases based on whisker lengths of 7.5, 7, 6.5, ..., 0.5 times the interquartile range. In Figure 2, we plot the difference between the Preliminary and the re-computed Advance trends against the multiplier of the interquartile range for the untransformed ratio for a single month. Each point represents the number of cases used to compute the estimate out of a total of 85. Figure 3 shows the same for the H-B method.

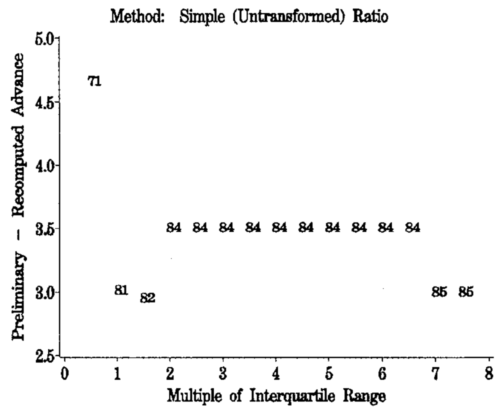

Figure 2

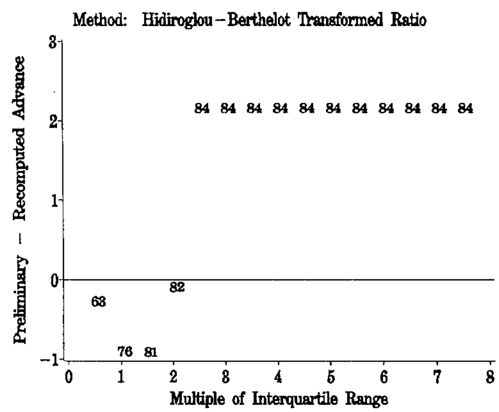

Figure 3

The measure that resulted in the smallest difference and excluded the fewest number of points was considered superior. For this particular month and kind of business, the H-B transformation was clearly superior.

Because of the success of the H-B transformation for this particular month, we examined several more months (July 1996 - June 1997) for this KB. We chose a whisker length of 2 times the interquartile range. However, improvement in the Advance estimate was not consistent across these months.

- **Effect** of case k:

  Analysts frequently use this measure to detect outliers; therefore, we wanted to evaluate its effectiveness relative to the other outlier detection methods. We used two different approaches. One approach was to identify cases with the largest absolute effect as outliers and exclude them from the computation of the Advance trend one at a time. Trying this in several KBs, we found no systematic pattern between the number of cases so excluded and the difference between the Preliminary and Advance trends. An alternative approach was to exclude the case with the largest absolute effect, and then recompute the Advance trend and effects of the remaining cases and exclude the case with the largest absolute effect, and so on. This more closely mimics the editing process used during data collection (although in the actual process, an excluded case can be re-included later). We compared these two approaches and found minimal differences. Typically, the first few cases to be discarded were the same in both situations. Furthermore, we found no consistent improvement in the Advance estimate.

- **Market share** for case k:

  This was used as a measure of the importance of a case with respect to its segment of the economy. It was especially useful in multivariate graphical displays of data (see Figure 1).

- **Cook's distance** for case k:

  Based on earlier work that showed a fourth root transformation worked well with sales data (Bienias et al., 1994), we modeled the fourth root of the current month sales as a function of the fourth root of the previous month sales. Although we considered this to be a valid measure for detecting outliers, we did not find it to be effective by itself. However, we used it in the Mahalanobis distance statistic.

- **Mahalanobis distance** of case k, combining several univariate criteria:

  We combined the following univariate criteria to compute the Mahalanobis distance: the H-B transformation of the ratio, the effect, and the Cook's distance for each case. We did not have a lot of success with combining these three measures, and we thought the problem was that "effect" was somewhat unstable.

5. An Alternative Estimation Procedure

At the moment, cases for which no response has been recorded or those for which the response is unusable (for example, in a given month a case might only be able to report for a 4-week period that does not overlap sufficiently with the current month) are not included in the estimation. This effectively means their trend is imputed to be the trend of the overall KB, as the estimation is done at the KB level (see Section 2.2). This can be problematic for two reasons. First, it is not necessarily the case that all cases in the KB have similar patterns of retail activity from month to month. Certain KBs, such as "Balance of Other Durable Goods," contain such a mix of types of businesses (e.g., toy stores, bookstores, newsstands, optical goods stores, luggage stores) that it is not reasonable to assume the retail trend of one can be used to impute that of another. Second, high nonresponse makes any imputation method potentially dangerous.

A possible competitor (or complement) to the use of one or more criteria to identify cases for exclusion is to use an estimator that allows for multiple imputation classes within a KB. For example, one might use the following estimator:

$$\text{Sales} = \frac{CM_A}{PM_A} \times P_A + \frac{CM_{A'}}{PM_{A'}} \times P_{A'} \quad ,$$

where P is the MRTS Preliminary **total** sales estimate for the previous month, and the set A is defined by some auxiliary classification variable. This could be extended to more than two classes.

Using data from January 1994 to December 1996, we applied this proposed estimator to three MARTS KBs and compared the resulting estimates of trend to those obtained using the current MARTS estimator. The goal was to determine which of these estimators produced estimates most consistent with those obtained in the MRTS. As in the earlier work, we used the difference between the MRTS Preliminary trend and the re-computed Advance trend as a basis for comparing the two estimators. Two or three subclasses (one or two subclasses of interest, and a subclass to cover the remaining SICs making up the MARTS KB) were created within each of the MARTS KBs based on the respondents' retail business subclassification (defined by their 4-digit SIC code; see Office of Management and Budget, 1987).

We chose three current MARTS KBs that were believed to contain subclasses of businesses with disparate patterns of retail activity. We separated "toys" from "Balance of Other Durable Goods," creating two subclasses; "computer equipment" and "home electronics" from "Furniture" creating three subclasses; and "nurseries" and "mobile home dealers" from "Building Materials," again creating three subclasses. The results obtained using the new estimator for Balance of Other Durables and Furniture were encouraging, whereas the performance for Building Materials was less striking.

Figure 4 shows a plot of the revision based on the new estimator (Preliminary - Proposed) against that based on the current estimator (Preliminary - Current) for Balance of Other Durables. Reference lines at 0 and at absolute equality (45° and -45°) have been added for clarity. As can be seen, the two estimators perform similarly (most of the points are near the line of equality). However, when the current estimator performs poorly and the revision is large (especially toward the left part of the graph), the new estimator is an improvement, bringing the revision closer to 0. Plots similar to this one were made for Furniture (Figure 5) and Building Materials (Figure 6).

Balance of Other Durable Goods was of particular interest to us. In the months around Christmas, toy store sales are very strong, much more so than the remainder of the KB, potentially making the "implicit imputation" problematic. Considering only November, December, and January in the three years we examined, the proposed estimator outperformed the current estimator in every one of these nine months. On average, the new estimator of trend was 5.5 points closer to the MRTS preliminary trend than the trend obtained using the current estimator.

Based on this work, we plan to use the new estimator for Balance of Other Durables and Furniture but not for Building Materials.

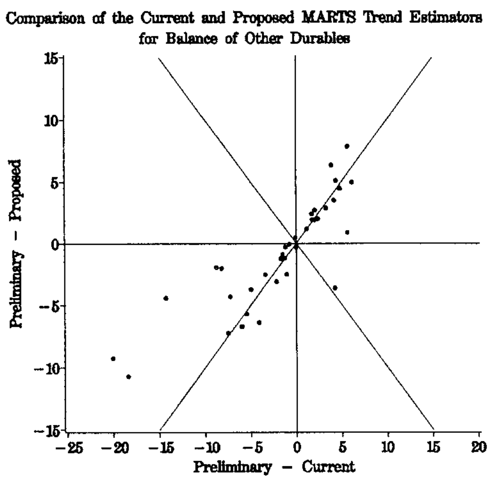

Figure 4

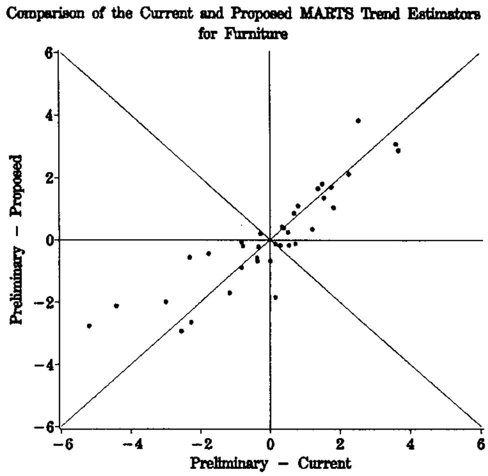

Figure 5

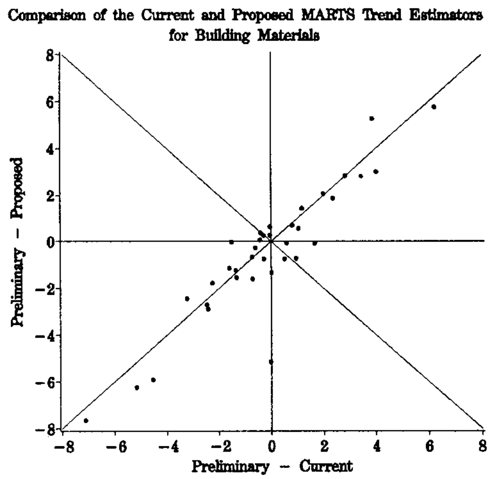

Figure 6

6. Summary and Extensions

Although the results of our research into an alternative estimation procedure were encouraging, we do not yet have a clear vision for an automated outlier detection procedure. We plan to extend the research described here to more kinds of business and months. One additional method we wish to consider is the change in market share for case k:

$$\Delta_k = \left| \frac{CM_k}{\sum_{i=1}^{n} CM_i} - \frac{PM_k}{\sum_{i=1}^{n} PM_i} \right| .$$

This measure is currently used in the MRTS for noncertainty cases, but we have not yet examined the properties of this measure with respect to the MARTS data.

We also plan to examine the Mahalanobis distance statistic further using Cook's distance and the H-B transformed ratio. One of the drawbacks to the Mahalanobis distance statistic is that it is not robust, as it relies on the estimated means, variances, and covariances of the individual variables. Methods exist for robust estimation of these quantities; see Campbell (1980; also described in Bienias, 1995) and Barnett and Lewis (1994).

Finally, we plan to monitor both the new estimation procedure and proposed outlier detection methods in a production setting. All of this work is part of a larger effort that includes modeling the relationship between the Advance trend and the Final MRTS trend (see Hogan & Cantwell, 1997).

References

Barnett, V., & Lewis, T. (1994). *Outliers in Statistical Data*, 3rd. ed. NY: Wiley.

Bienias, J. L. (1995, Jan.). *Methods for Outlier Detection for the Quarterly Financial Report*. Washington, DC: U. S. Bureau of the Census. (Technical report #ESMD-9407, available from Economic Statistical Methods and Programming Division.) See also pp. 368-75 of *Proceedings of the 1995 Bureau of the Census Annual Research Conference* for a briefer version.

Bienias et al. (1994, Dec.). *Graphical Review of Economic Data* (Final Report of the Graphical Data Analysis Working Group). Washington, DC: U. S. Bureau of the Census. (Technical report #ESMD-9406, available from Economic Statistical Methods and Programming Division.). See also pp. 1140-43 of Vol. II of the *American Statistical Association 1994 Proceedings of the Annual Meeting of the American Statistical Association, Survey Research Methods Section* for a briefer version.

Campbell, N. A. (1980). Robust procedures in multivariate analysis I: Robust covariance estimation. *Applied Statistics*, **29**, 231-37.

Draper, N. R., & Smith, H. (1981). *Applied Regression Analysis* (2nd Ed.). NY: Wiley.

Granquist, L. (1990). A review of some macro-editing methods for rationalizing the editing process. *Proceedings of Statistics Canada Symposium 90, Measurement and Improvement of Data Quality*, pp. 225-34. Ottawa: Statistics Canada.

Hidiroglou, M.A., & Berthelot, J.-M. (1986). Statistical editing and imputation for periodic business surveys. *Survey Methodology*, **12**, 73-83.

Hoaglin, D.C., Mosteller, F., & Tukey, J. W. (Eds.) (1983). *Understanding Robust and Exploratory Data Analysis*. NY: Wiley.

Hogan, H., & Cantwell, P. J. (in press). Predicting final retail sales estimates from advance reports. *American Statistical Association 1997 Proceedings of the Annual Meeting of the American Statistical Association, Survey Research Methods Section*. Alexandria, VA: American Statistical Association

Houston, G., & Bruce, A. G. (1992, February). Graphical editing for business and economic surveys. Technical report, New Zealand Department of Statistics, Mathematical Statistical Branch.

Hughes, P.J., McDermid, I., & Linacre, S. J. (1990). The use of graphical methods in editing (with discussion). *Proceedings of the 1990 Bureau of the Census Annual Research Conference*, pp. 538-54. Washington, DC: U.S. Department of Commerce.

Lee, H. (1995). Outliers in survey sampling. In B. Cox et al. (Eds.), *Survey Methods for Business, Farms, and Institutions*. NY: Wiley, pp. 503-526.

Office of Management and Budget. (1987). *Standard Industrial Classification Manual*. Available from National Technical Information Service, Springfield, VA (Order no. PB 87-100012).

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

U.S. Bureau of the Census. (1996). *Combined Annual and Revised Monthly Retail Trade, January 1986-December 1995*. Washington, DC: U.S. Government Printing Office (Current Business Reports, Item BR/95-RV).