A Simulation Study of Treatments of Influential Values in the Monthly Retail Trade Survey

Mary H. Mulry and Broderick Oliver[1]
Statistical Research Division, U.S. Census Bureau, Washington, DC 20233
Office of Statistical Methods and Research for Economic Programs, U.S. Census Bureau, Washington, DC 20233

Abstract
Although influential values are rare in economic surveys, they are problematic when they occur. An observation is considered influential if its value is correct but its weighted contribution has an excessive effect on the estimated total or period-to-period change. Currently, the U.S. Census Bureau uses weight trimming to address influential values in several ongoing programs. Mulry and Feldpausch (2007a, 2007b) showed that two methodologies, M-Estimation and Clarke Winsorization, had potential for improvements in detection and treatment of influential values in the Monthly Retail Trade Survey. The program uses the Hidiroglou-Berthelot statistical edit to identify potential outliers on a flow basis prior to final identification of influential values. In this paper, we present a continuation of the earlier research by Mulry and Feldpausch by applying the recommended methods to simulated data to examine the statistical properties of the treated estimates obtained using each method over repeated sampling. This paper describes the simulation methodology, illustrates with the initial results from one realistic scenario, and describes the other scenarios that will be addressed in future work.

Key Words: M-estimation, Clarke Winsorization, outliers.

1. Introduction

Although influential values are rare in economic surveys, they are problematic when they occur. An observation is considered influential if its value is correct but its weighted contribution has an excessive effect on the estimated total or period-to-period change. Failure to "treat" such influential observations may lead to substantial over- or under-estimation of survey totals, which in turn may lead to overly large or small decreases in estimates of change.

This paper illustrates the use of a simulation methodology to investigate two statistical methods of identifying and treating influential observations: Clarke Winsorization and M-estimation. The research continues work presented in Mulry and Feldpausch (2007a, 2007b). The previous studies examined a variety of outlier detection and treatment methods on 38 months of empirical data from the U.S. Census Bureau's Monthly Retail Trade Survey (MRTS). The further examination of the statistical properties of the previously recommended methods employs a simulation study

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

designed to produce data similar to that collected by the MRTS. The goal is to find a method that improves or replaces current methodology and uses the observation but in a manner that assures its contribution does not have an excessive effect on the monthly totals or an adverse effect on the estimates of month-to-month change and year-to-year change.

Each month, the MRTS surveys a sample of about 12,000 retail businesses with paid employees to collect sales and inventories. The sample design is typical of business surveys, with stratification that is based on major industry further stratified by the estimated annual sales. The sample design includes differential allocation to strata. The sample is selected every five years after the Economic Census and then updated as needed with a quarterly sample of births (new businesses) and removal of deaths (failed businesses). A complete record is imputed for each nonresponding active business.

When an influential observation appears in a month's data, the current corrective procedures depend on whether the analysts believe the observation is a one-time phenomenon or a recurring situation. If the influential value appears to be a rare occurrence for the business, then the influential observation is replaced with an imputed value. If the influential value represents a permanent change, then methodologists adjust its sampling weight using principles of representativeness or move the unit to a different industry when the nature of the business appears to have changed. The MRTS processing already includes running the algorithm by Hidiroglou and Berthelot (1986) each month to identify within-imputation-cell outliers and create the imputation base (Hunt, Johnson, and King 1999). The Hidiroglou-Berthelot statistical edit designates observations that should be reviewed by an analyst and sometimes suppressed from the imputation base. Treatment of influential values is done as a final step of the estimate review process. Hence, the methods described here are developed to complement, not replace, the Hidiroglou-Berthelot algorithm. The expectation is that the appearance of influential values will be fairly rare.

This paper describes the simulation methodology, illustrates with the initial results from one scenario, and describes the scenarios that future work will address. Our analysis emphasizes the simulation's estimates of relative bias for estimates of total sales and measures of change, in particular month-to-month change, when the influential value is adjusted and when it is not. Additional evaluation criteria include the number of influential observations that are detected, including the number of true and false detections made.

2. Methods

A preliminary study (Mulry and Feldpausch 2007a) examined several methods with one month of MRTS data and identified two methods that appeared to have promise for the MRTS data, Clarke Winsorization and generalized M-estimation. Subsequent work with the two methods and MRTS data showed further promise (Mulry and Feldpausch 2007b).

Before describing the methods, we first introduce the notation. For the i^{th} business in a survey sample of size n for the month of observation t, Y_{ti} is its revenue, w_{ti} is its final weight (which may or may not be equivalent to the inverse probability of selection), and X_{ti} is a variable highly correlated with Y_{ti}, such as previous month's revenue or its monthly revenue from a pre-entry questionnaire. The total monthly revenue Y_t is estimated by $\hat{Y}_t$ defined by

$$\hat{Y}_t = \sum_{i=1}^{n} w_{ti} Y_{ti} .$$

For ease of notation, we suppress the index for the month of observation t in the remainder of this section.

2.1 Clarke Winsorization

Winsorization procedures replace extreme values with other, less extreme values, effectively moving the original extreme values toward the center of the distribution. Winsorization methods offer adjustments for the observed influential value but may be interpreted as inspiring how to adjust the survey weight if that is needed instead. Winsorization procedures may be one-sided or two-sided, but the method developed by Clarke (1995) and described by Chambers et al (2000) is one-sided. The approach assumes a general model where the Y_i are characterized as independent realizations of random variables with $E(Y_i) = \mu_i$ and $var(Y_i) = \sigma_i^2$.

The general form of the one-sided Winsorized estimator of the total is designed for large values and is written as

$$\hat{Y}^* = \sum_{i=1}^{n} w_i Z_i$$

where $Z_i = \min\{Y_i, K_i + (Y_i - K_i)/w_i\}$.

Clarke suggests approximating the K_i that minimizes the mean squared error under the general model by $K_i = \mu_i + L(w_i - 1)^{-1}$, which requires estimating μ_i and L.

For an estimate of μ_i, Chambers et al (2000) suggest using the results of a robust regression. Then the estimate of μ_i is bX_i where b is the regression coefficient and X_i is the previous month's observation. To estimate L, the Clarke Winsorization first uses the estimate of μ_i to estimate weighted residuals

$D_i = (Y_i - \mu_i)(w_i - 1)$ by $\hat{D}_i = (Y_i - bX_i)(w_i - 1)$

Next the method arranges the estimates of the residuals in decreasing order $\hat{D}_{(1)}, \hat{D}_{(2)}, \ldots\ldots \hat{D}_{(n)}$. Then the Clarke method finds the last value of k, called k^*, such that $(k+1)\hat{D}_{(k)} - \sum_{j=1}^{k} \hat{D}_{(j)}$ is positive. Finally, the method estimates L by $\hat{L} = (k^* + 1)^{-1} \sum_{j=1}^{k^*} \hat{D}_{(j)}$.

2.2 Weighted M-Estimation

Robust methods are useful for studying influential observations because they relax the assumption in parametric statistics of a strict parametric model to allowing a "neighborhood" of parametric models (Hampel et al 1986). Thus, robust methods can be applied to a variety of parametric models instead of being strictly limited to one. In contrast, nonparametric statistics makes much weaker model assumptions, such as continuity and symmetry. In a sample survey setting, robust methods or nonparametric methods are quite appealing because the survey data are generally not from a simple random sample, and consequently it is difficult to validate any assumed originating distribution.

M-estimators (Huber 1964) are robust estimators that come from a generalization of maximum likelihood estimation. The application of M-estimation examined in this investigation is regression estimation. The weighted M-estimation technique proposed by Beaumont and Alavi (2004) is able to modify the weights for influential observations or adjust values of the influential observations. The approach for adjusting the values uses a compromise between the generalized regression estimator and the best linear unbiased estimator of the population total (Beaumont and Alavi 2004, Beaumont 2004). Briefly, the method estimates $\hat{B}^M$, which is implicitly defined by

$$\sum_{i \in S} w_i^*(\hat{B}^M)(y_i - x_i\hat{B}^M)\frac{x_i}{v_i} = 0$$

where

$$v_i = \lambda x_i$$

$$w_i^* = w_i\psi\{r_i(\hat{B}^M)\}/r_i(\hat{B}^M)$$

$$r_i(\hat{B}^M) = h_i e_i(\hat{B}^M)/Q\sqrt{v_i}$$

$$e_i(\hat{B}^M) = y_i - x_i\hat{B}^M$$

Q is a constant that is specified. The variable h_i is a weight that may or may not be a function of x_i. Section 4.1 contains a discussion of the settings for these parameters used in this investigation.

The function ψ may have a two-sided or one-sided form. We focus on two choices for the function ψ, Type I and Type II Huber functions, and investigate both the one- and two-sided-forms. The one-sided Type I Huber function is

$$\psi\{r_i(\hat{B}^M)\} = \begin{Bmatrix} r_i(\hat{B}^M), r_i(\hat{B}^M) \le \varphi \\ \varphi, otherwise \end{Bmatrix}$$

where φ is a positive tuning constant. This form is equivalent to a Winzorization of $r_i(\hat{B}^M)$. In the two-sided Huber I function $r_i(\hat{B}^M)$ is replaced by its absolute value $\left|r_i(\hat{B}^M)\right|$.

The weight adjustment corresponding to the Type I Huber function ψ above is

$$w_i^*(\hat{B}^M) = \begin{Bmatrix} w_i, r_i(\hat{B}^M) \le \varphi \\ \dfrac{\varphi}{r_i(\hat{B}^M)}, otherwise \end{Bmatrix}$$

An undesirable feature of using Type I Huber function is that the unit's adjusted weight may be less than one if the influential value is very extreme, thereby not allowing the influential value to represent itself in the estimation. The Type II Huber function ψ ensures that all adjusted units are at least fully represented in the estimate. The one-sided Type II Huber function is

$$\psi\{r_i(\hat{B}^M)\} = \begin{cases} r_i(\hat{B}^M), r_i(\hat{B}^M) \le \varphi \\ \dfrac{1}{w_i} r_i(\hat{B}^M) + \dfrac{(w_i - 1)}{w_i}\varphi, otherwise \end{cases}$$

where φ is a positive tuning constant. In the two-sided Type II Huber function $r_i(\hat{B}^M)$ is replaced by its absolute value $\left|r_i(\hat{B}^M)\right|$. This form is equivalent to a Winzorization of $r_i(\hat{B}^M)$, cf. the Type I Huber function.

An interesting feature of using the one-sided Type II Huber function in the M-estimation method is that the parameters can be set to mimic the assumptions of the Clarke Winsorization outlined in Section 2.1 (Beaumont 2004). However, the results will not be identical because the method used to estimate $\hat{B}^M$ is different.

Solving for $\hat{B}^M$ requires the Iteratively Reweighted Least-Squares algorithm in many circumstances when using the M-estimation method. For certain choices of the weights and variables, the solution is the standard least-squares regression estimator.

The weight adjustment for the above Type II Huber function above, which is the default in Beaumont's program, is

$$w_i^*(\hat{B}^M) = \begin{cases} w_i, r_i(\hat{B}^M) \le \varphi \\ 1 + (w_i - 1)\dfrac{\varphi}{r_i(\hat{B}^M)}, otherwise \end{cases}$$

For an adjustment to the influential value, Beaumont and Alavi (2004) use a weighted average of the robust prediction $x_i\hat{B}^M$ and the observed value y_i of the form

$$y_i^* = a_i y_i + (1 - a_i)x_i\hat{B}^M \text{ where } a_i = \frac{w_i^*(\hat{B}^M)}{w_i}.$$

When the set of weights that includes the adjusted weight $\{w_i^*(\hat{B}^M)\}$ are calibrated to maintain their total, then the sum of the original y-values weighted by the calibrated adjusted weights equals the sum of the y-values weighted by the original weights when the influential value is replaced by the adjusted y-value.

The adjusted value corresponding to the Type II Huber function is

$$y^* = \frac{1}{w_i} y_i + \frac{(w_i - 1)}{w_i}\{x_i\hat{B}^M + \frac{\sqrt{v_i}}{h_i}Q\varphi\}.$$

Beaumont (2004) finds an optimal value of the tuning constant φ by deriving and then minimizing a design-based estimator of the mean-square error that does not require a model to hold for all the data as in the Clarke Winsorization. It does not require a model to hold for the influential value, in particular. Beaumont uses numerical analysis to solve for the optimal value of the tuning constant φ.

3. Simulation Methodology

We used simulation methodology to investigate the statistical properties of the Clarke Winsorization and M-estimation methods that detect influential values and calculate adjustments. This approach allows us to estimate performance measures for the methods over repeated samples. We focused on one particular industry that was recommended by subject-matter (MRTS) experts because of its volatility. The simulation used one month of empirical sample data and sample estimates of autocorrelation to generate a 16-month series for a bivariate population of retail sales and inventories, (Y_1,V_1), …., $(Y_{16,} V_{16})$ where Y_t is the total sales and V_t is the total inventory for the t^{th} month for a particular industry.[1] The series is stationary except for Month 4, where we induced outliers in two percent of the population. To easily maintain stationarity in the simulated population, we did not include births and deaths in the simulation procedure.

3.1 Population

To obtain the simulated population, we began by generating the population values for Month 1 by using U.S. Census Bureau software that applies a nonparametric resampling algorithm described in Thompson (2000) to the (real) training data. To obtain the complete industry population, we applied the resampling algorithm separately in each MRTS stratum. Subsequent data for Months 2 –16 were generated as a stationary time series essentially as a forecast going forward from Month 1. The series was generated using an ARMA series with historical standard errors and autocorrelations to develop the AR(1) models.

The AR(1) model for the stationary time series for Months 2 to 16 is given by

$$y_t - m = \Phi * (y_{t-1} - m) + a_t \text{ , for } t = 2, \ldots 16.$$

where

$y_1 - m = 0$ and m is the series mean;

$a_t \sim N(0, \sigma^2)$ (white noise process)

Φ is estimated using the sample-based lag one autocorrelation for the selected industry.

The scenario we chose to investigate first is encountered most often in practice: a unit with a large sampling weight has an excessively high value in one month. To study the performance of

[1] Although inventory data were simulated, the analyses presented in Section 4.2 investigate only sales.

the considered methods under this scenario, after developing the population that was "influential value free," we induced extreme values in Month 4 by randomly adding two percent high outliers to the population data. The strategy was to add a large value to the simulated value of sales to the appropriate number of units in the stratum with the lowest sampling rate, which also have the largest weights in the sample. These high outliers only occur in Month 4, but are large enough to affect the total for that month.

Finally, we selected 5,000 stratified simple random samples without replacement of size *n* from the simulated population, where *n* is also the sample size for the industry in the MRTS. The sample followed the MRTS sample design as closely as possible (i.e., the same allocations to strata).

In each of the 5,000 samples, we applied M-Estimation and Clarke Winsorization to Months 2 through 16 using the prior month data as auxiliary data. In each sample, we computed total and change estimates using unadjusted (untreated) and adjusted data from each method. This allowed us to compute relative bias measures for each method.

3.2 Measures of Performance

We used the measures of performance outlined below to assess each method's performance over repeated samples. Due to the nature of the considered scenario (which only contained one month of data with influential values), we calculated the Type II error rate and the Hit rate for only Month 4, but the Type I error rate was calculated for all months.

Detection Measures

- Hit Rate = the percentage of induced influential values that were detected.
- Type I error rate = the percentage of observations that were not induced influential values that were designated as influential (false positive).
- Type II error rate = the percentage of induced influential values that were not detected (false negative). Note that the Type II error rate is equal to 0 in Months 1-3 and 5-16 since no influential values were induced in those months.

All measures are averaged over the 5,000 samples.

Relative bias estimates

To estimate the relative bias, we needed to define the estimate of monthly population total for sales when no treatment is applied and the estimate of monthly population total for sales when a treatment is applied. We also needed a measure of the "true" population total (truth).

For the definitions, we let

$\hat{Y}_{U,t}$ = Untreated estimate of total Y from month t

$\hat{Y}_{T,t}$ = Treated estimate of total Y from month t

Since we started with a stationary series and then induced large (influential) values in two percent of the population in Month 4, the expected value of the total sales is equal for all months except for Month 4. This large percentage of influential values in the population is not necessarily realistic, but it ensures that a high proportion of replications will contain an influential value, thus facilitating the simulation. The induced influential values in Month 4 yield an atypically large population total for the series by design. They also illustrate the problem in our programs' "reality," i.e., exaggerated estimates of change for the months preceding and following the month containing the influential value. With our simulation data, the high values in Month 4 are an appreciable phenomenon in the population – for that month and that month only – rather than an isolated occurrence. To strike a balance between the ease of the simulation and finding a realistic standard of comparison for the performance measures, we define the "true" monthly total sales to be the mean of the population total sales over the 16 months. This definition allows the influential values in Month 4 to increase the overall level of the true monthly total sales slightly as it would in any time series with a single divergent estimate. Therefore, we define the "true" monthly total sales to be

$$\hat{\hat{Y}} = \frac{\sum_t Y_t}{16} = \text{True monthly total sales.}$$

Next we define the relative bias estimates for the untreated and treated data from a particular sample in month t to be

$$\frac{(\hat{Y}_{U,t} - \hat{\hat{Y}})}{\hat{\hat{Y}}} = \text{Relative bias in month } t \text{ of } \textbf{untreated} \text{ estimate}$$

$$\frac{(\hat{Y}_{T,i} - \hat{\hat{Y}})}{\hat{\hat{Y}}} = \text{Relative bias in month } t \text{ of } \textbf{treated} \text{ estimate.}$$

The relative bias for each method is obtained by averaging these sample statistics over the 5,000 samples.

Month-to-Month Change Estimates

We define the true month-to-month change to be

$$\hat{\hat{Y}}_C = \frac{1}{15}\sum_{t=2}^{16}\frac{Y_t}{Y_{t-1}} = \text{True month-to-month change.}$$

The month-to-month change may be estimated for Months 2 to 16 using

$$\frac{\left(\dfrac{\hat{Y}_{U,t}}{\hat{Y}_{U,t-1}}\right) - \hat{\hat{Y}}_C}{\hat{\hat{Y}}_C} = \text{Relative bias in month } t \text{ of } \textbf{untreated} \text{ estimate}$$

$$\frac{\left(\dfrac{\hat{Y}_{T,t}}{\hat{Y}_{T,t-1}}\right)-\hat{\hat{Y}}_C}{\hat{\hat{Y}}_C}$$ = Relative bias in month t of **treated** estimate.

The relative bias in the estimate of month-to-month change for each method is obtained by averaging these sample statistics over the 5,000 samples.

<u>Year-to-Year Change Estimates</u>

We define the true monthly year-to-year change to be

$$\hat{\hat{Y}}_{CY} = \frac{1}{4}\sum_{t=13}^{16}\frac{Y_t}{Y_{t-12}}$$ = True monthly year-to-year change.

The monthly year-to-year change may be estimated for Months 13 to 16 using

$$\frac{\left(\dfrac{\hat{Y}_{U,t}}{\hat{Y}_{U,t-12}}\right)-\hat{\hat{Y}}_{CY}}{\hat{\hat{Y}}_{CY}}$$ = Relative bias in month t of untreated estimate

$$\frac{\left(\dfrac{\hat{Y}_{T,t}}{\hat{Y}_{T,t-12}}\right)-\hat{\hat{Y}}_{CY}}{\hat{\hat{Y}}_{CY}}$$ **=** Relative bias in month t of treated estimate**.**

The relative bias in the estimate of year-to-year change for each method is obtained by averaging these sample statistics over the 5,000 samples.

4. Results

4.1 Application of Methods

To perform the Clarke Winsorization, we used the in-house SAS software developed by Roxanne Feldpausch for previous research. For the M-estimation, we used SAS software developed by Jean-Francois Beaumont of Statistics Canada (Beaumont 2007). The program has default settings for all but one of the parameters but also allows for specifying different values of the parameters. The program does not have a default for the required initial value for φ, the parameter for which it finds an optimal value, so we set the initial value equal to 200 million, as recommended by our prior research.

Notice that when we use the program default settings $Q = 1$ and $h_i = (w_i - 1)\sqrt{x_i}$ along with setting $v_i = x_i$ for all units in sample, then $r_i = (w_i - 1)(y_i - x_i\hat{B}^M)$. Under these conditions, r_i has the same form as $\hat{D}_i$ in the Clarke Winsorization. However, the b in the Winsorization and

$\hat{B}^M$ in the M-estimation method usually are not going to be equal because they use different estimation methods.

The program default setting for v_i is $v_i = 1$. However, we also consider $v_i = x_i$ and $v_i = \sqrt{x_i}$. When setting $Q = 1$ and $h_i = (w_i - 1)\sqrt{x_i}$, using $v_i = 1$ tends to give the residuals for large weighted values of x_i more influence in fitting the regression line than when $v_i = x_i$. Setting $v_i = \sqrt{x_i}$ also gives more influence to large weighted values of x_i but not as much as setting $v_i = 1$.

We apply the following M-estimation treatments with the default $Q = 1$: (1) two-sided Huber I ψ function with $v_i = x_i$, (2) two-sided Type II Huber function ψ with $v_i = x_i$, (3) two-sided Type I Huber function ψ with $v_i = 1$ (the default), (4) two-sided Type II Huber function ψ with $v_i = 1$ (the default), (5) two-sided Type II Huber function ψ with $v_i = \sqrt{x_i}$, for all sample units i.

4.2 Simulation Estimates

Table 1 shows performance measures for estimates of total sales from Clarke Winsorization and two-sided M-estimation methods used to detect influential values. The true total sales for each month, as defined in Section 3.2, are 48,395 million. The Type II error rates and the Hit rates are not shown because all methods detected the induced influential values (Hit rate = 100 percent) and none of the methods failed to detect induced outliers (Type II error rate = 0 percent). Although previous research found that the algorithm used by the M-estimation method does not converge in every situation (Mulry and Feldpausch 2007b), the algorithm converged for all five versions of M-estimation in every replicate in every month.

The Clarke Winsorization and all five two-sided versions of the M-estimation method detect the influential values induced in Month 4. All six methods have a relative bias in the estimate for Month 4 that is closer to zero than the relative bias in the unadjusted estimates, which was 0.57 percent. The relative bias closest to zero occurs for the Clarke Winsorization at 0.34 percent. Since the true estimate is 48, 395 million, the difference in the relative bias translates into a difference of 111 million in the estimate of total sales between the unadjusted and Clarke Winsorization estimates. The three variations of the Type I Huber M-estimation method has the same relative bias at 0.38 percent and therefore, are 20 million higher than the Clarke Winsorization estimate. The relative bias observed for the two Type II Huber M-estimation methods is 0.39 percent, implying they are 21 million higher than the Clarke estimate.

Since both M-estimation and Clarke Winsorization use observations from the previous month in fitting their robust regression lines, the adjustment of influential values in Month 4 may affect the results of both methods in Month 5. The relative bias for all five versions of M-estimation in Month 5 is 0.07 percent. Although the relative bias of the unadjusted estimate, which is –0.05 percent, has a different sign, the absolute difference in the relative bias is only 0.12 percent. For the Clarke Winsorization in Month 5, the relative bias is -0.06 percent.

The adjustment of influential values in Month 4 affects the results for Month 5 on rare occasions for the M-estimation methods. All five versions of the M-estimation method have a Type I error rate of 0.06 percent in Month 5, which translates into detecting non-induced influential values in three of the 5,000 replicates. The Type I error is zero for all the months other than Month 4 for all versions of the M-estimation method.

For Clarke Winsorization, the effect of the adjustment in Month 4 on the results for Month 5 is not clear because the Type I error for all months is non-zero, implying that non-induced outliers are being detected. For Month 5, the Type I error is 0.27 percent and is within the range of 0.25 percent to 0.32 percent observed for months other than Month 4 and 5. The lowest Type I error rate, is 0.14 percent in Month 4. The Type I error rates for the other months range from 0.25 percent in Month 10 to 0.32 percent in Month 2. From the 5,000 replicates, the number of replicates where the Clarke Winsorization detected non-induced influential values ranged from 3 in Month 4 to 16 in Month 2. These adjustments in all the months except Month 4 cause the relative bias of the Clarke Winsorization to be 0.01 percent to 0.02 percent further from zero than the relative bias of the unadjusted estimate.

For month-to-month change, Table 2 shows the relative bias in the estimates and that the true month-to-month change, as defined in Section 3.2, is 99.995 percent. The adjustment of the influential values induced in Month 4 appear to affect the relative bias in the estimates of month-to-month change for Months 3 to 4 and Months 4 to 5 for the Clarke Winsorization and all five versions of M-estimation. The relative biases in the unadjusted estimates for Month 4 to Month 5 and for Month 5 to Month 6 are 0.760 percent and –0.610 percent, respectively. The relative bias for the Clarke Winsorization method's estimate of change from Month 3 to Month 4 at 0.547 percent has the opposite sign and is closer to zero than all the M-estimation methods where the relative biases range from –0.577 percent to –0.579 percent.

However, for the change from Month 4 to Month 5, the relative bias for Clarke Winsorization is considerably further from zero than the relative bias for all the M-estimation methods. For the estimate of change from Month 5 to Month 6, the adjustment in Month 4 may still have an effect on the relative bias in estimates from all five versions of the M-estimation method. The Type I error rates in Table 1 show that the Clarke Winsorization and the M-estimation methods detect influential values that were not induced in some replicates in Month 5. However, in five other instances, the identification by the Clarke Winsorization of influential values that were not induced appears to cause a small effect on estimate of month-to-month change. Twice the relative bias in the estimates of month-to-month change was slightly further from zero, and three times the relative bias was slightly closer to zero.

For year-to-year change, Table 3 shows the relative bias in the estimates and that the true year-to-year change, as defined in Section 3.2, is 99.774 percent. The adjustment of the influential values induced in Month 4 appears to affect the relative bias in the estimates of year-to-year change for Months 4 to 16. The relative bias in the unadjusted estimate of change from Month 4 to Month 16 is –0.586 percent. The relative bias in the estimated change from the Clarke method is –0.375 percent, which is closer to zero than the relative bias for all the M-estimation methods. The relative bias in the estimates of change from Month 4 to Month 16 are –0.405 percent for the three M-estimation methods that use the Type I Huber function and –0.407 for the two that use the Type II Huber function.

Table 1. Performance measures for estimates of total sales from Clarke Winsorization and two-sided M-estimation methods to detect influential values from simulation with 5,000 replicates when influential values are induced in Month 4.
True total sales = 48,395 (in millions).

months	2	3	**4**	**5**	6	7	8	9	10	11	12	13	14	15	16
Unadj est	48,322	48,269	**48,634**	**48,335**	48,268	48,267	48,276	48,266	48,273	48,288	48,311	48,295	48,296	48,300	48,239
rel bias unadj	-0.08%	-0.19%	**0.57%**	**-0.05%**	-0.19%	-0.19%	-0.17%	-0.19%	-0.18%	-0.15%	-0.10%	-0.13%	-0.13%	-0.12%	-0.25%
Clarke															
Adj est	**48,316**	**48,263**	**48,525**	**48,329**	**48,261**	**48,261**	**48,270**	**48,260**	**48,266**	**48,282**	**48,305**	**48,289**	**48,290**	**48,294**	**48,234**
rel bias	**-0.09%**	**-0.20%**	**0.34%**	**-0.06%**	**-0.20%**	**-0.20%**	**-0.18%**	**-0.21%**	**-0.19%**	**-0.16%**	**-0.11%**	**-0.15%**	**-0.14%**	**-0.13%**	**-0.26%**
Type I error	**0.32%**	**0.30%**	**0.14%**	**0.27%**	**0.25%**	**0.26%**	**0.28%**	**0.28%**	**0.25%**	**0.28%**	**0.28%**	**0.30%**	**0.27%**	**0.28%**	**0.28%**
Huber I $v = x$															
Adj est	48,322	48,269	**48,545**	**48,393**	48,268	48,267	48,276	48,266	48,273	48,288	48,311	48,295	48,296	48,300	48,239
rel bias	-0.08%	-0.19%	**0.38%**	**0.07%**	-0.19%	-0.19%	-0.17%	-0.19%	-0.18%	-0.15%	-0.10%	-0.13%	-0.13%	-0.12%	-0.25%
Type I error	0	0	**0**	**0.06%**	0	0	0	0	0	0	0	0	0	0	0
Huber II $v = x$															
Adj est	48,322	48,269	**48,546**	**48,393**	48,268	48,267	48,276	48,266	48,273	48,288	48,311	48,295	48,296	48,300	48,239
rel bias	-0.08%	-0.19%	**0.39%**	**0.07%**	-0.19%	-0.19%	-0.17%	-0.19%	-0.18%	-0.15%	-0.10%	-0.13%	-0.13%	-0.12%	-0.25%
Type I error	0	0	**0**	**0.06%**	0	0	0	0	0	0	0	0	0	0	0
Huber I $v = 1$															
Adj est	48,322	48,269	**48,545**	**48,392**	48,268	48,267	48,276	48,266	48,273	48,288	48,311	48,295	48,296	48,300	48,239
rel bias	-0.08%	-0.19%	**0.38%**	**0.07%**	-0.19%	-0.19%	-0.17%	-0.19%	-0.18%	-0.15%	-0.10%	-0.13%	-0.13%	-0.12%	-0.25%
Type I error	0	0	**0**	**0.06%**	0	0	0	0	0	0	0	0	0	0	0
Huber II $v = 1$															
Adj est	48,322	48,269	**48,546**	**48,393**	48,268	48,267	48,276	48,266	48,273	48,288	48,311	48,295	48,296	48,300	48,239
rel bias	-0.08%	-0.19%	**0.39%**	**0.07%**	-0.19%	-0.19%	-0.17%	-0.19%	-0.18%	-0.15%	-0.10%	-0.13%	-0.13%	-0.12%	-0.25%
Type I error	0	0	**0**	**0.06%**	0	0	0	0	0	0	0	0	0	0	0
Huber I $v = sqrt(x)$															
Adj est	48,322	48,269	**48,545**	**48,393**	48,268	48,267	48,276	48,266	48,273	48,288	48,311	48,295	48,296	48,300	48,239
rel bias	-0.08%	-0.19%	**0.38%**	**0.07%**	-0.19%	-0.19%	-0.17%	-0.19%	-0.18%	-0.15%	-0.10%	-0.13%	-0.13%	-0.12%	-0.25%
Type I error	0	0	**0**	**0.06%**	0	0	0	0	0	0	0	0	0	0	0

Note: Cells in bold indicate a difference from the entry for the unadjusted estimate.

Table 2. Relative bias in estimates of month–to-month change rate for total retail sales.
True month-to-month change = 99.995 percent.

Months	unadjusted	Clarke	Huber I $v = x$	Huber II $v = x$	Huber I $v = 1$	Huber II $v = 1$	Huber I $v = \sqrt{x}$
2 to 3	-0.104%	-0.104%	-0.104%	-0.104%	-0.104%	-0.104%	-0.104%
3 to 4	**0.760%**	**0.547%**	**0.577%**	**0.579%**	**0.577%**	**0.579%**	**0.577%**
4 to 5	**-0.610%**	**-0.400%**	**-0.309%**	**-0.310%**	**-0.310%**	**-0.311%**	**-0.308%**
5 to 6	**-0.133%**	**-0.135%**	**-0.254%**	**-0.255%**	**-0.253%**	**-0.253%**	**-0.254%**
6 to 7	0.003%	**0.004%**	0.003%	0.003%	0.003%	0.003%	0.003%
7 to 8	0.024%	0.024%	0.024%	0.024%	0.024%	0.024%	0.024%
8 to 9	-0.016%	-0.016%	-0.016%	-0.016%	-0.016%	-0.016%	-0.016%
9 to 10	0.018%	**0.017%**	0.018%	0.018%	0.018%	0.018%	0.018%
10 to 11	0.037%	**0.038%**	0.037%	0.037%	0.037%	0.037%	0.037%
11 to 12	0.053%	**0.052%**	0.053%	0.053%	0.053%	0.053%	0.053%
12 to 13	-0.029%	**-0.028%**	-0.029%	-0.029%	-0.029%	-0.029%	-0.029%
13 to 14	0.007%	0.007%	0.007%	0.007%	0.007%	0.007%	0.007%
14 to 15	0.014%	0.014%	0.014%	0.014%	0.014%	0.014%	0.014%
15 to 16	-0.121%	-0.121%	-0.121%	-0.121%	-0.121%	-0.121%	-0.121%

Note: Cells in bold indicate a difference from the entry for the unadjusted estimate.

Table 3. Relative bias in estimates of year–to-year change rate for total retail sales.
True year-to-year change = 99.774 percent.

months	unadjusted	Clarke	Huber I $v = x$	Huber II $v = x$	Huber I $v = 1$	Huber II $v = 1$	Huber I $v = \sqrt{x}$
2 to 14	0.173%	0.172%	0.173%	0.173%	0.173%	0.173%	0.173%
3 to 15	0.291%	0.291%	0.291%	0.291%	0.291%	0.291%	0.291%
4 to 16	**-0.586%**	**-0.375%**	**-0.405%**	**-0.407%**	**-0.405%**	**-0.407%**	**-0.405%**

Note: Cells in bold indicate a difference from the entry for the unadjusted estimate.

4.3. Summary of initial results

Our initial simulation analyses examined the Clarke Winsorization method and five versions of the M-estimation method for the detection and adjustment of influential values under a scenario that is encountered the most often. Under this scenario, all methods detected the induced influential values.

However, the Clarke Winsorization appears to be more sensitive in that it identified non-induced influential values in several replicates. These values could be considered outliers in their respective strata but not for the whole population. The effect of the additional designations appears to be minor. All the methods identified non-induced influential values in the month after the month with the induced influential values.

For the estimates of month-to-month change, we saw differences between the methods in relative biases for the estimates of change for Months 3 to 4, Months 4 to 5, and Months 5 to 6. Note that Month 5 to 6 effects demonstrate residual correction effects: for example, if the influential Month 5 value(s) are not sufficiently reduced, the monthly total will be too large, as will the monthly change estimate. For all five versions of the M-estimation, the relative bias in the change from Months 4 to 5 was closer to zero than observed for the Clarke Winsorization. The relative bias in the same change for the unadjusted estimates was further from zero than all the methods. However, the relative biases for all the M-estimation methods for the estimated changes from Months 3 to 4 and Months 5 to 6 were further from zero than the Clarke Winsorization. For all methods, the relative bias in the estimated change from Months 3 to 4 was closer to zero than the unadjusted estimates. However, the relative bias in the estimated change from Months 5 to 6 was further from for all methods than for the unadjusted estimates, demonstrating residual correction effects for all methods. The adjustment of the non-induced influential values by the Clarke Winsorization may be the reason its relative bias was only slightly further from zero than observed for the unadjusted estimates.

For the estimates of year-to-year change, we have examined only the change from Month 4 to Month 16. The relative bias from the Clarke Winsorization is closer to zero than the relative bias in all the M-estimation methods, but the difference is small and ranges from 0.030 percent to 0.032 percent. The relative bias in estimates from all the methods is closer to zero than the relative bias in the unadjusted estimates.

5. Future Work

Both the Clarke Winsorization and the five versions of M-estimation studied each appear to have advantages and disadvantages. Clarke Winsorization is very easy to use. However, it has a tendency to detect false influential values in this scenario although the adjustments for the false detections are small. Determining the parameters for the M-estimation is not easy. M-estimation parameters that work well for one scenario (e.g., unusually large values) may not be effective in others (e.g., unusually small values), which could impact effective usage in a production environment. Moreover, in previous research, the algorithms have not converged or have not offered a helpful adjustment in some circumstances. However, in this scenario, which is the most likely scenario to encounter, there were no problems with convergence of the algorithm. M-estimation has the advantage of the capacity to be implemented as a one-sided or two-sided method. Having the capability to adjust influential values deemed too small is also an important goal of the research. The design of the Clarke Winsorization only provides the ability to adjust an influential value that is too large. We intend to investigate whether the Clarke Winsorization can be adapted to treat influential values that are too small.

Having only investigated one scenario over repeated samples, we do not have enough information to recommend one method over another. We are presently continuing this research using simulations with other realistic influential value scenarios and are also examining year-to-year change further out in the series. The other influential value scenarios under investigation include

- Inducing influential values in Month 4 that are too low rather than too high.
- Inducing influential values in Month 4 in the stratum with the lowest weight and continuing through Month 10 as evidence of a permanent change rather than a one-time event.
- Inducing some influential values that are too high and some that are too low in Month 4.

To date, our analysis has focused on detection and treatment of influential values on one series. However, the MRTS publishes data on two items, and any recommended method for this survey should take this into consideration. Consequently, other scenarios may include inducing observations where the sales report has an influential value, but the inventory report does not and vice versa. In such cases, there will have to be a decision about whether to adjust the unit’s final weight for both or the individual item values.

Acknowledgements

The authors thank Jenny Thompson for many helpful conversations and suggestions. The authors also thank Lynn Weidman and Bill Davie for reviewing earlier drafts of the paper.

References

Beaumont, J.-F. (2007) personal communication.

Beaumont, J.-F. (2004) "Robust Estimation of a Finite Population Total in the Presence of Influential Units." Report for the Office for National Statistics, dated July23, 2004. Office for National Statistics, Newport, U.K.

Beaumont, J.-F. And Alavi, A. (2004) "Robust Generalized Regression Estimation." *Survey Methodology*, 30, 2, 195-208.

Chambers, R. L., Kokic, P., Smith, P. And Cruddas, M. (2000) "Winsorization for Identifying and Treating Outliers in Business Surveys." *Proceedings of the Second International Conference on Establishment Surveys.* Statistics Canada. Ottawa, Canada. 717-726.

Clarke, M. (1995) "Winsorization Methods in Sample Surveys." Masters Thesis. Department of Statistics. Australia National University.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Werner, S. A. (1986) *Robust Statistics. An Approach Based on Influence Functions.* John Wiley & Sons. New York, NY.

Huber, P. J. (1964) "Robust Estimation of a location parameter." *Annals of Mathematical Statistics*. Institute of Mathematical Statistics. 35. 73-101.

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

Hunt, J. W., Johnson, J. S., and King, C. S. (1999) "Detecting Outliers in the Monthly Retail Trade Survey Using the Hidiroglou-Berthelot Method." *Proceedings of the Section on Survey Research Methods.* American Statistical Association. Alexandria, VA. 539-543.

Mulry, M. H. and Feldpausch, R. (2007a) "Investigation of Treatment of Influential Values." *Proceedings of the Third International Conference on Establishment Surveys*. American Statistical Association. Alexandria, VA.

Mulry, M. H. and Feldpausch, R. (2007b) "Treating Influential Values in a Monthly Retail Trade Survey." *Proceedings of the Survey Methods Section, SSC Annual Meeting.* Statistical Society of Canada. Ottawa, Ontario, Canada.

Thompson, J. R. (2000) *Simulation: A Modeler's Approach*. John Wiley and Sons. New York. 87-110.