Several Scenarios for Influential Observations and Methods for Their Treatment

Mary H. Mulry, Broderick Oliver, Stephen Kaputa[1]
U. S. Census Bureau, Washington, DC 20233

Abstract
In survey data, an observation is considered influential if it is reported correctly and its weighted contribution has an excessive effect on a key estimate, such as an estimate of total or change. Influential observations occur infrequently in economic surveys but have a detrimental effect on the key estimate when they do appear. In previous research with data from the U.S. Monthly Retail Trade Survey (MRTS), two methods, Clark Winsorization and weighted M-estimation, have shown potential to detect and adjust influential observations. This paper discusses results of the application of an improved simulation methodology that generates more realistic population data. The analyses consider several scenarios for the occurrence of influential observations in the MRTS and assess the performance of the two methods in detecting influential values under the different scenarios.

Key words: Outlier, Winsorization, M-estimation

1. Introduction

Although influential values occur infrequently in economic surveys, they are problematic when they appear. 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. To be clear, our focus is on influential values that remain after all the data have been verified and corrected so these unusual values are true and not the result of reporting or recording errors. 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 increases or exceedingly small decreases in estimates of change.

Each month, the U.S. Census Bureau's Monthly Retail Trade Survey (MRTS) surveys a sample of about 12,000 retail businesses with paid employees to collect data on sales and inventories. The MRTS is an economic indicator survey, whose monthly estimates are inputs to the Gross Domestic Product estimates. Moreover, significant changes in levels are important to monetary and budgetary decision makers, economists, business analysts, and economic researchers in assessing the health of the economy, and in making corporate investment decisions. The MRTS sample design is typical of business surveys, with stratification based on major industry, further stratified by the estimated annual sales. The sample design requires the sampling rates to be higher in the strata with the larger units than in the strata with the smaller units. 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 (businesses no longer in operation).

When an influential observation appears in a month's data, the current corrective procedures depend on whether the subject-matter experts believe the observation is a one-time phenomenon

[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.

or a permanent shift. If the influential value appears to be an atypical 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 (Black 2001). 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). 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 infrequent.

The objective of this research is to find an automated statistical procedure for detecting and treating influential values to replace the current subjective procedure performed by analysts. 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.

In previous research with data from the U.S. Monthly Retail Trade Survey (MRTS), two methods, Clark Winsorization and weighted M-estimation, have shown potential to detect and adjust influential observations (Mulry and Feldpausch 2007a 2007b, Mulry and Oliver 2009). Therefore, the focus of this paper is the use of simulation methodology to investigate two robust statistical methods of identifying and treating influential observations: Clark Winsorization (Clark 1995, Chambers et al. 2000) and M-estimation (Beaumont and Alavi 2004, Beaumont 2004). In a sample survey setting, robust methods offer an appealing approach for dealing with influential values that introduce bias and increase sampling variance of estimates because survey data are generally not from a simple random sample, and consequently it is difficult to validate any assumed originating distribution.

This paper focuses on the ability of Clark Winsorization and M-estimation to detect influential values in several scenarios. This work compliments other work on performance measures for the two methods (Mulry, Oliver, and Kaputa 2012). The scenarios include a single influential value that is either unusually high or unusually low, and multiple influential values, specifically, two high values and the combination of one high and one low value.

2. Methods

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 survey 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. In MRTS, the survey weight w_{ti} is the design weight since the missing data treatment is imputation and no other weight adjustments are made.

2.1 Clark 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 could be used to derive an adjustment for the survey weight if that is needed instead. Winsorization procedures may be one-sided or two-sided, but the method developed by Clark (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

$$\hat{Y}^* = \sum_{i=1}^{n} w_i Z_i \qquad \text{where } Z_i = \min\{Y_i, K_i + (Y_i - K_i)/w_i\}.$$

Detection of observation i as an influential value by Clark Winsorization occurs when $Z_i \neq Y_i$. To implement the method, Clark 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. Clark's approach builds on a method developed by Kokic and Bell (1994) that derived a K for each stratum rather than for each individual unit.

For an estimate of μ_i, Chambers et al (2000) suggest using the results of a robust regression. In our application, we used the least median of squares (LMS) robust regression method because it seemed to perform the best of all the methods considered (Mulry and Feldpausch 2007a).

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 Clark 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 $\hat{D}_{(1)}, \hat{D}_{(2)}, ...\hat{D}_{(n)}$. Then the Clark 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.

Next, the method estimates L by $\hat{L} = (k^* + 1)^{-1} \sum_{j=1}^{k^*} \hat{D}_{(j)}$.

Then the estimate of K_i is formed by $\hat{K}_i = bX_i + \hat{L}(w_i - 1)^{-1}$, which is used to determine the values of Z_i for the estimate of the total $\hat{Y}^*$.

2.2 Weighted M-Estimation

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) uses the Schweppe version of the weighted generalized technique (Hampel et al. 1986, p. 315 – 316). The estimator of the total using this approach is consistent for a finite population since it equals the finite population total when a census is conducted (Sarndal et al. 1992, p. 168).

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 contains a discussion of the settings for these parameters used in this investigation. The variable x_i may be a vector, but in our application, it is the previous month's value. The regression model does not include an intercept because with retail trade, the regression of current month's sales on the previous month's sales tends to go through the origin.

The role of the function ψ is to reduce the influence of units with a large weighted residual $r_i(\hat{B}^M)$. We focus on two choices for the function ψ, Type I and Type II Huber functions, and investigate their 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 Winsorization of $r_i(\hat{B}^M)$. Detection of observation i as an influential value by M-estimation with the Huber I function occurs when $r_i(\hat{B}^M) > \varphi$. 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

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

where φ is a positive tuning constant. Detection of observation i as an influential value by M-estimation with the Huber II function occurs when $r_i(\hat{B}^M) > \varphi$. 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 Winsorization 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 Clark 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. For certain choices of the weights and variables, the solution is the standard least-squares regression estimator.

The weight adjustment for the Type II Huber function above is

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

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. Note that the MRTS does not perform any calibration weighting, so that M-estimation procedures are performed using the unadjusted design weights.

The adjusted value corresponding to the Type II Huber function is

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

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 Clark 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. Research methodology

To assess how well M-estimation and Clark Winsorization identify influential values in MRTS data, we conduct a simulation study using different scenarios. Consequently, the simulated population data presents "realistic" monthly sales estimates, modeled from two industries with different natures. One that we refer to as Industry 1 has monthly sales of approximately 46.1 billion and one of the most volatile patterns for influential values. The other that we refer to as Industry 2 has a more stable pattern and has monthly sales of approximately 2.5 billion. The sample sizes in our simulations are 1,161 for Industry 1 and 147 for Industry 2.

In practice, influential observations are infrequent so for most months, no influential value is present in any of the MRTS monthly samples. The common scenario for an influential value is an observation that is much higher than previous measurements, and it has a high weight that greatly amplifies its impact on the estimates. Failure to address this scenario properly can have far-reaching consequences in interpreting the state of the economy, so we focus first on this scenario, although we investigate other possible scenarios. These other scenarios include an influential value that is much lower than previous and two influential values, either both very high or one very high and one very low.

The models used to simulate populations for Industries 1 and 2 use the MRTS data for these industries. Recall that the MRTS is a stratified sample, with strata defined by unit size within industry where the measure of size is sales. An exploratory empirical analysis for both studied industries confirmed that the strata-level means differ by within-industry-strata. To obtain realistic level estimates, we apply the nonparametric resampling algorithm described in Thompson (2000) by industry-strata to empirical MRTS data to obtain Month 1 data, thus ensuring that the strata means are different and the industry totals equal the survey estimates. Then, we generate 19 additional months of the population data for each sampling stratum *h* in the industry using ARMA modeling to form a stationary series for that stratum, so that $\hat{y}_{hi,t} = \beta_h \hat{y}_{hi,t-1} + \varepsilon_{hi}, \varepsilon_{hi} \sim (0, w_{ti}\sigma^2_{hit}), t > 1$. Therefore, each of the two populations is a stationary series within strata, but not at the industry level. Since the time series is stationary, the strata-level means are approximately the same over time although in practice the strata-level means may vary over a similar period. Using a stationary series avoids the possibility of a trend confounding the effects of the influential values.

4. Results

In this section, we examine the simulation results regarding the performance of the two treatments and quality of the estimates they produce. The Clark Winsorization algorithm does not require parameter settings, but the M-estimation algorithm does. First we investigate the settings of the parameters for the M-estimation algorithm to determine which options produce the best estimates. Then we use those settings for the M-estimation in the comparison with Clark Winsorization. For the Winsorization, we developed the software in SAS. For the M-estimation, we used SAS software developed by Jean-Francois Beaumont (2007).

4.1 M-estimation algorithm settings

The M-estimation algorithm discussed in Section 2.2 requires settings for Q, h_i, v_i, the function ψ, and an initial value of the tuning constant φ. We use the default settings for the parameters Q and h_i, but explore different settings for the other options. We also consider whether to include the observations selected with certainty in fitting the regression line. Table 12 summarizes the parameters for the M-estimation algorithm.

Table 1. M-estimation algorithm parameters

Parameter	Parameter Function	Values
Q	Constant	=1 (default)
h_i	Unit weight	$= (w_i - 1)\sqrt{x_i}$ (default)
v_i	Model error underlying regression estimator	= 1 or x_i
ψ	Huber function	Huber I or Huber II
φ	Tuning constant (determines starting point for detection region)	User provides initial value and program calculates optimal value

The M-estimation program default settings for Q and h_i are $Q = 1$ and $h_i = (w_i - 1)\sqrt{x_i}$. Our investigation considers two values of the weighting parameter for the residuals v_i, namely $v_i = x_i$

and $v_i =$ 1. Ideally, the choice of the setting for v_i should be a data-driven decision because v_i essentially specifies the variance of the model errors underlying the regression estimator for M-estimation. As expected, neither $v_i= x_i$ nor v_i=1 provide a good model of the MRTS data Notice that when we used the default settings for Q and h_i along with setting $v_i = x_i$ for all units in sample, $r_i = (w_i - 1)(y_i - x_i\hat{B}^M)$. Now r_i has the same form as $\hat{D}_i$ in the Clark Winsorization. However, the b in the Winsorization estimation method and $\hat{B}^M$ in the M-estimation method usually are not going to be equal because they use different estimation methods. With $Q = 1$ and $h = (w_i - 1)\sqrt{x_i}$, setting $v_i = 1$ tends to give the residuals for large weighted values of x_i more influence in fitting the M-estimation regression line than when $v_i = x_i$.

Since our investigation found that there is some Type II error when $v_i = 1$ and none when $v_i = x_i$, and the two settings produce about the same results regarding Type I error, we decided to pursue only $v_i = x_i$. In the next section we explore the detection properties when the initial φ is set to a low value and when it is set to a high value.

4.2 Detection regions

In the previous section, we investigated the ability of each method to detect an influential observation with one particular value. Now, we examine the range of influential values that the methods designate as influential, called the detection region. We investigate the detection regions when only one influential value is present in the sample and when a sample has two influential values. In addition, we investigate the effect of the weight of the sample unit and the choice of the initial φ on the detection region.

With M-estimation, the choice of the initial φ also may affect the size of the detection region. This is important because staff usually investigates observations flagged as influential to be certain that the published numbers represent a change in the economic measure and are not the consequence of an “unrepresentative” unit. The ability to set the initial φ allows some control over the size of the values that analysts will check and thereby some control over the amount of staff time that has to be devoted to the checking. This control is not available with Clark Winsorization.

4.2.1 Results for one influential value

The size of an observation’s weight as well as its weighted value both affect to whether it will be designated as influential by M-estimation. Typically, the sampling rate for small businesses is lower than for larger businesses because there are more small businesses than larger businesses. Therefore, the smaller businesses typically have higher weights. If two observations have the same unusually high amount of weighted month-to-month change, the M-estimation method is less likely to designate the one with the lower weight as an influential value. Figure 1 and 2 use unweighted data to illustrate how the critical region varies by the size of the business. Table 2 shows the weights for the points on the grid based on their value in the previous month. The increment between adjacent vertical lines on a grid is 100,000.

The values where the algorithm designates smaller businesses as influential tend to be lower than for larger businesses due to stratum weighting differences. However, in a particular month, the weights for some of the businesses with the small values may not be large. Although the design at the beginning of the five years between censuses tends to have large weights for the small businesses and small weights for the large businesses, adjustments as the sample matures may cause more variability in the weights for the smaller businesses. One reason the weights vary for small businesses is that seasonal businesses with high income some months and almost no income in others retain the same weight throughout the year. Another reason is that changes in the nature of a particular business that causes a continuing growth or decline in sales.

Table 2. Weights for variable points in detection region plots in Figures 1 and 2, based on the previous month's unweighted value.

Grid Division	Previous Month's Range (Month 3)		Column Weight
	Maximum	Minimum	
1	100	100	436.670
2	100,100	100,100	372.420
3	200,100	200,100	199.370
4	300,100	300,100	140.550
5	400,100	400,100	85.571
6	500,100	500,100	59.857
7	600,100	600,100	33.375
8	1,300,100	700,100	27.000
9	5,200,100	1,400,100	7.600
10	6,200,100	5,300,000	1.000

Note: If the maximum and minimum are the same, the increment in the grid division is only one column

On the other hand, detection by Clark Winsorization depends only on the weighted value of the observation. Figure 1 shows the detection region for Clark Winsorization with unweighted data. The levels at which Clark Winsorization starts designating values as influential are much closer to the regression line than where M-estimation begins its designations of influential. This reflects the trimming that the Clark Winsorization does to reduce the MSE through lowering the variance but introducing some bias.

Figure 3 uses weighted data to illustrate how the weight affects the M-estimation detection region by showing the detection region for three different weights of an observation for a particular sample from Industry 2 when the initial φ=150 million. When there is only one high observation, the algorithm designates it as influential if its weighted value is above the upper line corresponding to its weight. A single low observation receives the designation of influential if its weighted value is below the lower line corresponding to its weight. The underlying assumption for the lines indicating the detection regions in Figure 3 is that the unweighted induced potential influential value varies and its weight stays constant at 7.6, 27, or 59.9.

Figure 4 illustrates how the size of the M-estimation detection region varies with different values of the initial φ. In this figure, an observation with a weight of 27 will be designated as influential if its weighted value is above the boundary line corresponding to the setting of the initial φ. The boundary line for the detection region for high influential values moves up as the setting for the initial φ increases.

4.2.2 Results for two influential values

We also investigated the detection regions for scenarios with two influential values. Our approach uses one of the samples from Industry 2 that has the single fixed influential value induced and lets the second candidate for detection vary to determine when it is designated as an influential value. Then we are able to identify the detection region for the second value. This analysis parallels the study of masking in outlier detection where the presence of one outlier causes the failure of detecting a second outlier. Although the designs of the Clark Winsorization and M-estimation methods permit simultaneous detection and treatment of multiple influential values, investigating the detection regions for the second influential value provides more information about the performance of the two methods.

Figure 5 shows plots that illustrate the detection regions for the second influential value for Clark Winsorization, and for M-estimation. In these plots, the unweighted value of the induced potential influential value varies and its weight varies as shown in Table 2. Interestingly, the boundaries of the detection regions for the two methods are almost the same. In contrast, when only one influential value is present as in Figure 2, the M-estimation boundary and Clark Winsorization boundary are very different.

The M-estimation algorithm experienced some problems with convergence for some scenarios where the second value was too low. The combination of a high influential value and a low influential value causes the algorithm to be less likely to converge. Beaumont (2004) also noted some problems with convergence in his simulations in this situation. In Figure 6, the black dots mark the constant sample and the gray area of the grid marks the non-convergent region. The area in red contains values that are detected as influential while the values in the green area are not influential. The unweighted observations varied while their weight was constant at 7.6.

To explore the circumstances when the algorithm fails to converge, we focus on two samples. One is a sample where the M-estimation algorithm converged, called Sample A, and the other where the algorithm did not converge, called Sample B. Both samples contain the constant high influential value and Sample A contains another high influential value whereas sample B contains a low influential value. Figure 7 displays plots of these samples.

Figure 8 shows plots of the MSE as a function of the tuning constant φ for Samples A and B, respectively. The Sample A function clearly has a minimum around 250 million and detects both influential values. The Sample B function is a strictly increasing function so the minimum occurs at zero when the tuning constant φ is also zero. A tuning constant φ equal to zero means that every observation that is not on the M-estimation regression line is designated as influential, which is not helpful. The reason for a strictly increasing MSE as a function of φ is that the variance dominates the MSE. The influential values tend to offset each other and adjustments of the influential values tend to be symmetrical so the bias squared is essentially a constant function equal to zero. However, the variance is a strictly increasing function of φ causing the MSE also to be a strictly increasing function.

To wrap up, when a sample contains both unusually high and unusually low influential values and the M-estimation algorithm does not converge, no adjustment is probably the best choice. The reason is that the unusual values counterbalance each other in a manner that introduces minimal bias. Therefore, the failure of the algorithm to identify the influential values is not necessarily a handicap.

5. Summary

Our investigation examines the performance of weighted M-estimation and Clark Winsorization in identifying influential values in samples from populations generated to be similar to those of industries in the MRTS. Our simulation methodology permits investigating the performance of the methods on estimates of total sales, month-to-month change, and year-to-year change.

We find both methods to be effective, but each has advantages and disadvantages that may affect a decision about which to employ.

A big advantage of Clark Winsorization is the ease of implementation of its straightforward formulas. By design, the method identifies and treats only influential values that are unusually high so it does not identify or treat values that are influential values because they are unusually low. However, the major concern in economic surveys regarding influential values usually is the occurrence of high ones. When an influential value is present, Clark Winsorization always identifies it and offers an adjustment.

On the other hand, the Clark Winsorization appears to trim observations that are not unusually high but the trimming causes the estimates to have less variance, thereby reducing the MSE. The trimming is also disadvantageous because the staff usually researches whether observations flagged as influential are accurate. Unnecessary investigations are not an efficient use of staff time. In some situations, the ease of implementation of Clark Winsorization and the protection that it offers against unusual influential values could outweigh the small amount of bias introduced by trimming.

The weighted M-estimation methodology identifies both high and low influential values. High influential values usually are the major concern but low influential values do occur and can introduce bias. The M-estimation algorithm has flexibility in setting parameters to make assumptions appropriate for the underlying data. In addition, weighted M-estimation with a high value of the initial tuning constant φ performed the best overall of the three options considered for total sales and for month-to-month change in sales.

An attractive feature of M-estimation is that the algorithm allows an analyst to set the value of the initial tuning constant φ and thereby determine the minimum size of the weighted regression residuals that will be considered as potential influential values. When none of the weighted residuals is larger than the initial φ, the algorithm runs once and stops without identifying any influential values. However, when one or more weighted residuals are larger than the initial φ, the algorithm iterates until it finds the value of φ that minimizes the MSE and then identifies observations with weighted residuals larger than the final φ as influential. Therefore, setting the initial φ enables the analyst to control review criteria facilitates the efficient use of staff time in examining proposed adjustments.

The flexibility of weighted M-estimation has the disadvantage of introducing some complexity in implementation. There are situations when the algorithm has convergence issues, but careful setting of the parameters for the algorithm appears to reduce this problem and sometimes avoids it all together. These convergence issues tend to be more difficult to avoid when the algorithm uses a two-sided function ψ implementation than with a one-sided function. In the one-sided implementation, choosing an initial tuning constant φ higher than the non-influential weighted residuals tend to be appears to avoid convergence problems in the month with the influential value and in the succeeding month when the unit returns to its routine level. Avoiding convergence problems in the two-side implementation is not as clear. If the lack of convergence is caused by the occurrence of both an unusually high and an unusually low influential value in the same month, then an estimate with no adjustments is justified because the two influential values offset to result in the bias being approximately zero.

Acknowledgements

The authors thank Katherine Jenny Thompson for her useful suggestions. The authors are very grateful to Jean-Francois Beaumont for providing the SAS code for the M-estimation algorithm and for all his advice and consultations. The reviews by Lynn Weidman, Eric Slud, Scott Scheleur, and Bill Davie are appreciated very much.

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.

Black, J. (2001) "Changes in Sampling Units in Surveys of Businesses." *Proceedings of the Federal Committee on Statistical Methods Research Conference*. Office of Management and Budget. Washington, DC. http://www.fcsm.gov/01papers/Black.pdf

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.

Clark, 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. http://www.amstat.org/sections/SRMS/Proceedings/papers/1999_093.pdf

Mulry, M. H., Oliver, B., and Kaputa, S. (2012) "Detecting and Treating Verified Influential Values in a Monthly Retail Trade Survey." Unpublished manuscript. Center for Statistical Research and Methodology. U.S. Census Bureau. Washington, DC.

Mulry, M. H. and Oliver, B. (2009) "A Simulation Study of Treatments of Influential Values in the Monthly Retail Trade Survey." *JSM Proceedings*, Survey Research Methods Section. American Statistical Association. Alexandria, VA. 2979-2993. http://www.amstat.org/sections/SRMS/Proceedings/y2009/Files/304328.pdf

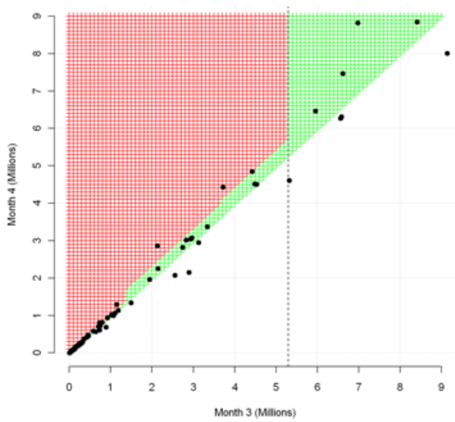
Mulry, M. H. and Feldpausch, R. (2007a) "Investigation of Treatment of Influential Values." *Proceedings of the Third International Conference on Establishment Surveys*. American Statistical Association. Alexandria, VA. 1173 – 1179.

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. http://www.ssc.ca/survey/documents/SSC2007_M_Mulry.pdf

Sarndal, C.-E., Swensson, B., and Wretman, J. (1992) Model *Assisted Survey Sampling.* Springer-Verlag. New York, NY.

Figure 1. Illustration of detection regions for a sample from Industry 2 using Clark Winsorization and one-sided Huber II M-estimation with a low initial φ when the unweighted induced potential influential value varies with weights as shown in Table 12. The algorithm designates values in red as influential.

Clark Winsorization



M-estimation, low initial φ

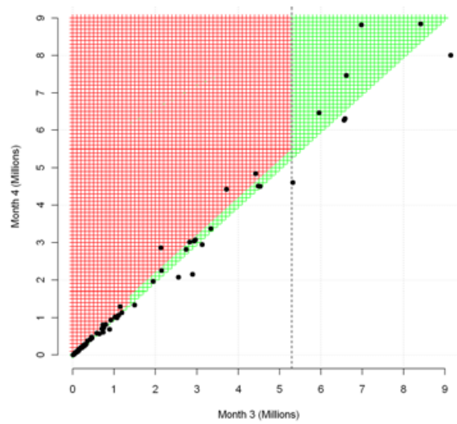
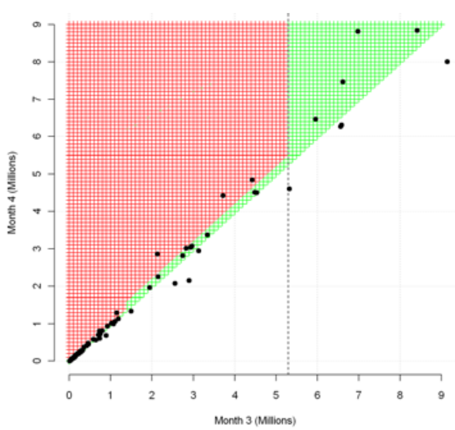


Figure 2. Illustration of detection region for a sample from Industry 2 using one-sided Huber II M-estimation with a low initial φ and a high initial φ when the unweighted induced potential influential value varies with weights as shown in Table 12. The algorithm designates values in red as influential.

M-estimation, low initial φ



M-estimation, high initial φ

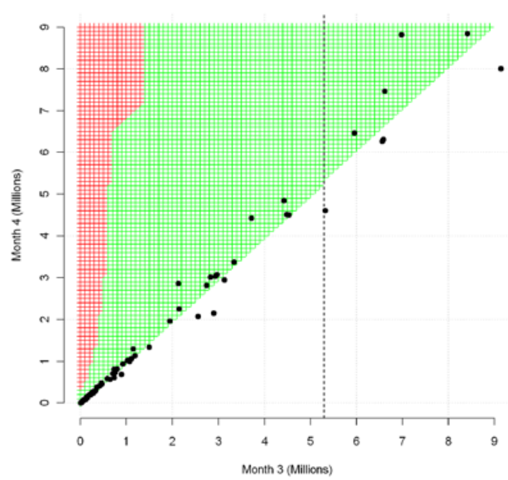


Figure 3. M-estimation. Illustration of detection regions for values with different weights for a sample from Industry 2. A high observation with a particular weight will be designated as influential if its weighted value is above the high line for that weight. A low observation with a particular weight will be designated as influential if its weighted value is below the high line for that weight.

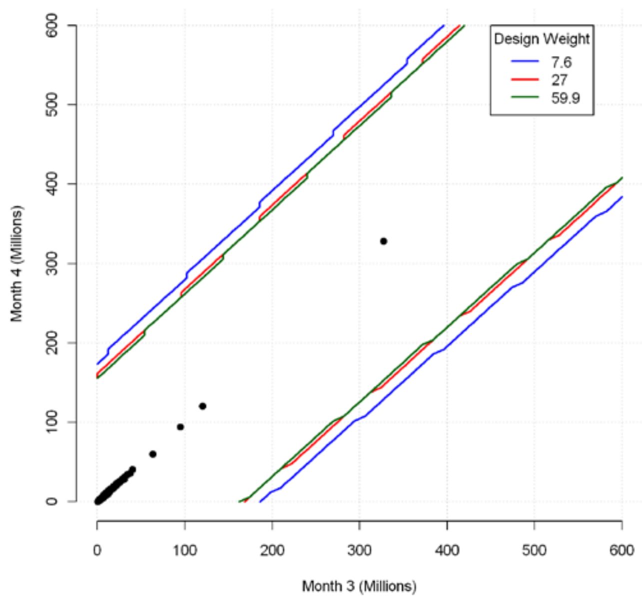


Figure 4. M-estimation. Illustration of detection regions corresponding to different values of the initial φ for a sample from Industry 2. An observation with a weight of 27 is designated as influential if its weighted value is above the line corresponding to the initial φ used in the application of the algorithm.

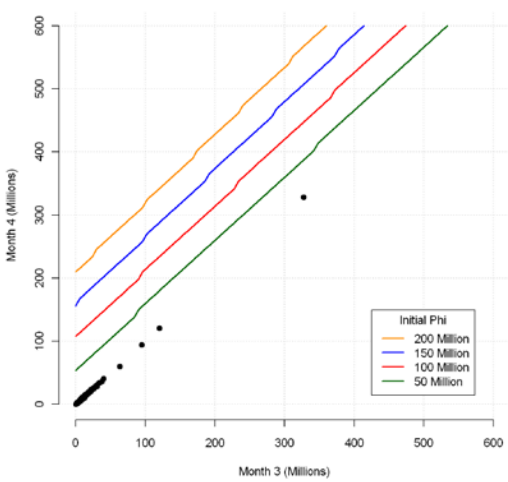


Figure 5. Illustration of Clark Winsorization and one-sided Huber II M-estimation with a high initial *φ* in a scenario with two high influential values in a sample from Industry 2. One influential value is fixed and the other varies. The unweighted induced potential influential value varies but its weight is a constant 27. The algorithm designated values in red as influential.

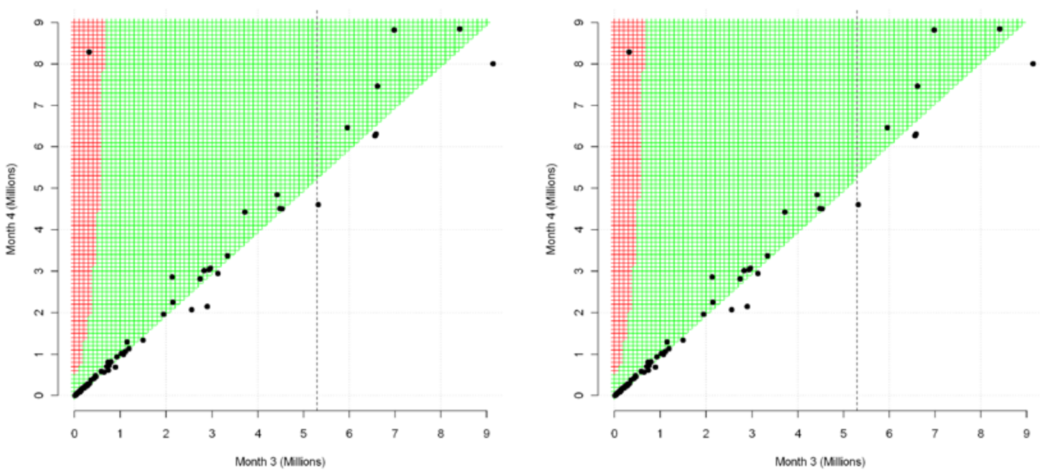


Figure 6. Illustration of when the two-sided Huber II M-estimation algorithm does not converge in the scenario of two influential values in a sample from Industry 2. One high influential value is fixed and the other varies but always has a weight of 7.6. The algorithm designated values in red as influential. The area in gray denotes the values of the second influential value that result in the M-estimation algorithm not converging.

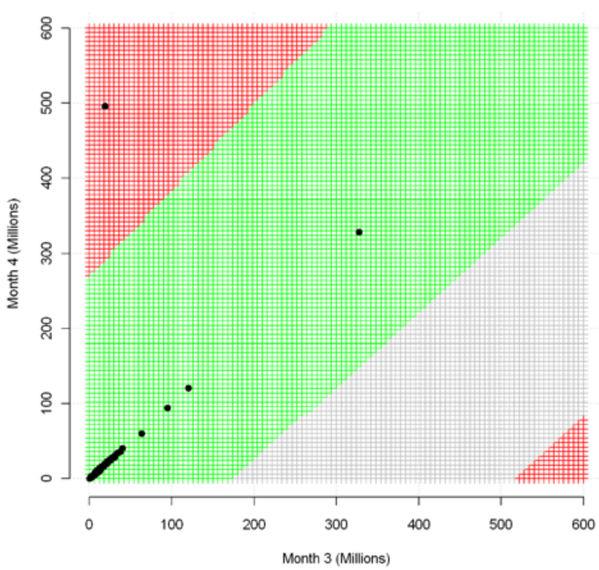


Figure 7. Illustration of samples from Industry 2 with two influential values.

Sample A, two high **Sample B, 1 high & 1 low**

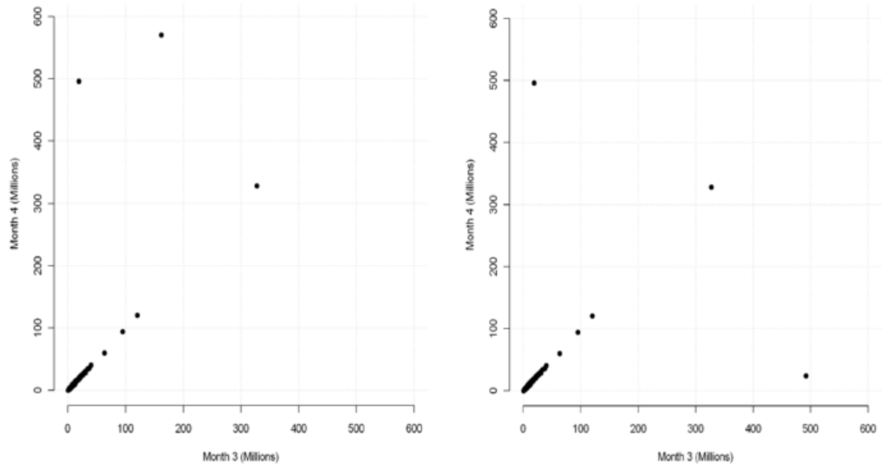


Figure 8. M-estimation. Graphs of bias squared, variance, and MSE vs. *φ* for Sample A and Sample B in Figure 22. For Sample A, the MSE reaches a minimum at about 250 million and detects both influential values. For Sample B, the MSE equals the variance since the bias squared is constantly zero so their plots coincide, and the MSE is strictly increasing function so minimum MSE occurs at *φ =0.*

Sample A, two high **Sample B, 1 high & 1 low**