SSC Annual Meeting, June 2007
Proceedings of the Survey Methods Section

TREATING INFLUENTIAL VALUES IN A MONTHLY RETAIL TRADE SURVEY

Mary H. Mulry and Roxanne M. Feldpausch[1]

ABSTRACT

Influential observations occur infrequently in establishment surveys but are problematic when they do appear. An observation is considered influential if it is true and its weighted contribution has an excessive effect on the estimated total. The paper describes the results of a study that examines two methodologies that use such an observation but in a manner that allows a more appropriate impact on the estimated total. In particular, the study examines Winsorization and weighted M-estimation to adjust influential values when estimating monthly sales using data from the U.S. Monthly Retail Trade Survey.

Keywords: Outlier, Winsorization, M-estimation

ABSTRACT

Les observations influentes se produisent de façon peu fréquente dans les enquêtes sur les établissements, mais elles se révèlent problématiques lorsqu'elles se présentent. On considère qu'une observation est influente si elle est vraie et que sa contribution pondérée a un effet excessif sur le total estimé. L'article décrit les résultats d'une étude qui examine deux méthodologies qui utilisent une telle observation, mais d'une façon qui permet un effet plus approprié sur le total estimé. L'étude examine notamment la méthode d'estimation de Winsor est l'estimation de M pondérée pour ajuster les valeurs influentes lors de l'estimation des ventes mensuelles qui utilise des données de la U.S. Monthly Retail Trade Survey.

Keywords: M-estimation; valeur abérrante, Winsorization,

1. INTRODUCTION

This paper investigates two methods of identifying and treating influential observations whose weighted contribution has an excessive effect on the estimate of total monthly sales in the U.S. Monthly Retail Trade Survey (MRTS). The goal is to find methodology that improves upon current methodology and uses the observation but in a manner that assures its contribution does not have an excessive effect on the total. Influential observations tend to occur infrequently, but are problematic when they do occur.

Each month the MRTS has a sample of about 12,500 retail businesses with paid employees from which it collects sales from all and inventories from some. The stratification for sample selection is based on major industry and the amount of sales. The sample is selected every five years after the economic census and then updated as needed with a quarterly sample of births and removal of deaths.

When an influential observation appears in a month, the current corrective procedures depend on whether the analysts believe the observation is a one-time phenomenon or a recurring situation. If the influential appears to be a rare occurrence for the business, then the observation is replaced with an imputation. If the influential value represents a permanent change, then analysts adjust 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 outliers and create the imputation pool (Hunt, Johnson, and King 1999). The Hidiroglou-Berthelot algorithm designates observations that should be reviewed and

[1] Mary H. Mulry, Statistical Research Division, U.S. Census Bureau, Washington, DC 20233. mary.h.mulry@census.gov. Roxanne M. Feldpausch, Economic Programs Directorate, U.S. Census Bureau, Washington, DC 20233. roxanne.feldpausch@census.gov. 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.

sometimes suppressed from the imputation base. The intent is for the treatment of influential values that is developed to complement, not replace, the Hidiroglou-Berthelot algorithm. With the Hidiroglou-Berthelot algorithm detecting and compensating for reporting errors, the expectation is that the appearance of influential values will be fairly rare.

Our investigation applied the two selected methodologies to 38 months of data from the MRTS with the intent of proposing methodologies to run with new data every month on a trial basis. The evaluation criteria include the number of influential observations that are detected, including the number of true and false detections made. In addition, the evaluation will include an estimate of bias and an assessment of the impact on measures of change, in particular the month-to-month ratio of sales.

2. METHODS

A preliminary study (Mulry and Feldpausch 2007) 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.

Before describing the methods, we first describe the notation. For the i-th business in a survey sample of size n, Y_i is its revenue for the month of observation; w_i is its sample weight, and X_i is a variable highly correlated with Y_i such as previous month's revenue or its monthly revenue from a pre-entry questionnaire. Total monthly revenue T is estimated by

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

2.1 Clarke Winsorization

Winsorization 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 satisfying: $E(Y_i) = \mu_i$ and $var(Y_i) = \sigma_i^2$.

The Winsorized estimator of the total 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. 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 $\hat{D}_{(1)}, \hat{D}_{(2)}, \ldots.. \hat{D}_{(n)}$.

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

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 w_i is the survey weight, which may or may not be the inverse of the probability of selection. The variable h_i is a weight that may or may not be a function of x_i. Section 3.1 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. An example of a one-sided form, called the Type II Huber function, is

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

where φ is a positive tuning constant. This form is equivalent to a Winsorization of $r_i(\hat{B}^M)$.

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 specification of the function ψ leads to three choices for adjusting the survey weights. The Type II weight adjustment, which is the default in Beaumont's program, for the ψ above is

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

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

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

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 . Beaumont's program requires setting an initial value for the tuning constant φ .

The adjustment corresponding to the Type II Huber function is

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

When the set of weights that includes the adjusted weight are calibrated to maintain their total, then the weighted sum of the original y-values equals the sum of the y-values weighted by the original weights when the influential value is replaced by the adjusted y-value.

3. RESULTS

We investigated M-estimation methods to detect influential values and calculate adjustments using 38 months of four industries in the MRTS data. One particular industry was more volatile than the others so we focused on it. We explored several different settings of the parameters in the M-estimation algorithm that finds an optimal value of the tuning constant φ, which is the cut-off value for the weighted regression residuals. The user sets an initial value for the tuning constant φ, and the algorithm finds the value that minimizes the mean square error (MSE).

Sometimes the algorithm failed to converge for two-sided methods, which Beaumont (2004) also noted in his simulations. The investigation with the MRTS data found that sometimes the two-sided methods did not converge, and that for some months, one-sided methods also failed to converge for some parameter settings. However, there has always been at least one set of parameter settings where the one-sided methods converge.

A new wrinkle is that we have found that sometimes the algorithm converges, but the results are not helpful. In these cases, the MSE is either a strictly decreasing or a strictly increasing function of the tuning constant φ . When the MSE is strictly decreasing, the bias dominates the MSE. In this case, the optimal value of the tuning constant φ is the residual of the outlier and the adjusted value equals the outlier. When the MSE is strictly increasing, the variance dominates the MSE. Therefore, the optimal value of the tuning constant φ is very small or zero, and most if not all the observations are designated outliers.

The Winsorization, which is a one-sided method, produced some reasonable initial results for a month where the M-estimation method converged but was not helpful. The Winsorization minimizes the MSE under model assumptions where the M-estimation method minimizes the design-based MSE. The weight of the observation effects whether the M-estimation method designates it as influential. The sample units with low weights will be designated as influential less often than the sample units with high weights. The weight does not affect whether the Winsorization calculates an adjustment.

3.1 Application of methods

For the Winsorization, we developed the software in SAS. For the M-estimation, we used SAS software developed by Jean-Francois Beaumont. The program finds the optimal φ but calls for an initial value that we set equal to 200 million. The 11 methods considered are the one- and two- sided Huber I and II functions that vary the values of the weighting parameter v and whether certainty observations are included in the data sets. The values of the weighting parameter for the residuals were varied so that $v = x, \sqrt{x}, 1$.

Notice that when we used 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, $r_i = (w_i - 1)(y_i - x_i\hat{B}^M)$. Now r_i now 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. With $Q = 1$ and $h_i = (w_i - 1)\sqrt{x_i}$, setting $v_i = 1$ tends to give the residuals for large weighted values of x_i more influence in fitting the regression line. Setting $v_i = \sqrt{x_i}$ also gives more influence to large weighted values of x_i but not as much as setting $v_i = 1$.

Results of the application of M-estimation with and without the units selected with certainty were similar. Therefore, this paper reports the results for M-estimation without the certainty units. The following M-estimation treatments with the default $Q = 1$ were applied: (1) two-sided Huber I ψ function with $v_i = x_i$, (2) two-sided Huber II ψ function with $v_i = x_i$, (3) two-sided Huber I ψ function with $v_i = 1$ (the default), (4) two-sided Huber II ψ function with $v_i = 1$ (the default), (5) two-sided Huber II ψ function with $v_i = \sqrt{x_i}$ for all sample observations *i*.

3.2 Illustration with 5 months

We found that the nature of the MSE as a function of the tuning constant φ for some parameter settings appears to influence whether the algorithm converges and whether the results are helpful when it does converge. Five months of the MRTS data for the selected industry illustrate the performance of the algorithm under different circumstances. These months include data configurations where the algorithm converges, converges but is not helpful, and does not converge for the different parameter settings. When the methods did produce adjusted values in the illustrative months the absolute value of the change in the total was in the range of 0.2 to 0.65 percent.

Table 1 contains results for the five methods for the five months, Months 17, 22, 27, 32, and 33. Table 1 shows the apparent influential value(s) in the month and the total before any adjustments. For each method, the table includes the adjusted value, the residual, whether the method detected the outlier, the value of the tuning constant φ, the total with the adjusted value, the number of influential values detected, and the root mean square error (RMSE). "NC" denotes the algorithm did not converge for a method. A discussion of the results for the five months follows.

<u>Month 32</u>

For Month 32, all the methods converge and identify the influential value, but two of the methods adjust the observation only slightly. The apparent weighted outlier is 410 million with a weight of 55 and a weighted residual 380 million. Figures 1 and 2 illustrate M-estimation using the Huber I ψ with $v_i = x_i$ which recommends an adjusted value of 110 million based on an optimal value of φ of 77 million and a RMSE equal to 504 million. Figure 1 shows the weighted current month versus the weighted previous month with the robust regression line and the influential value and adjusted value from both M-estimation and Winsorization. Figure 2, which plots the MSE versus the tuning constant φ, illustrates the minimum. Using Huber II instead of Huber I produces the same results. The methods with $v_i = 1$ converge and identify the outlier but are not helpful since they produce little change in the outlier. Using Huber I and setting $v_i = \sqrt{x_i}$ produce an adjusted value of 290 million and a total of 46,020 million which is a reduction of 0.26 percent, and RMSE of 415 million.

Applying the Winsorization in Month 32 produces an adjusted value of about 220 million and a total of 45,950 million, a reduction of 0.4 percent.

Table 1. Two-sided M-estimation for influential values (in millions) for five months

	observed	Huber I, v=x	Huber II, v=x	Huber I, v=1	Huber II ,v=1	Huber I, v=sqrt(x)
Month 17						
wgt adj value	780	780	780	760	760	780
wgt residual		510	510	510	510	510
found?		yes	yes	yes	yes	yes
Phi		510	510	500	480	510
wgt total	46,870	46,870	46,870	46,840	46,840	46,860
outliers ID		1	1	1	1	1
RMSE		328	327	376	375	333
Month 22						
wgt outlier	310					
wgt adj value		180	180	96	**NC**	120
wgt residual		200	210	210	**-**	210
found?		yes	yes	yes	-	yes
Phi		100	76	41	-	62
wgt total	40,220	40,090	40,090	40,010	-	40,040
outliers ID		1	1	1	-	1
RMSE		345	345	418	-	375
Month 27						
wgt outlier	150					
wgt adj value		-	-	230	230	-
wgt residual		-270	-270	-230	-230	-260
found?		no	no	yes	yes	no
Phi		270	270	160	150	260
wgt total	46,220	-	-	46,300	46,300	-
outliers ID		0	0	1	1	0
RMSE		-	-	484	466	-
Month 32						
wgt outlier	410					
wgt adj value		110	110	400	400	290
wgt residual		380	380	380	380	380
found?		yes	yes	yes	yes	yes
Phi		77	72	360	360	260
wgt total	46,140	45,840	45,840	46,130	46,130	46,020
outliers ID		1	1	1	1	1
RMSE		504	504	416	416	415
Month 33						
wgt outlier	370					
wgt adj value		36	-	36	54	-
wgt residual		317	-	317	317	-
found?		yes	**NC**	yes	yes	**NC**
wgt outlier	42					
wgt adj value		397	-	391	380	
wgt residual		-348	-	-342	-340	
found?		yes	**NC**	yes	yes	**NC**
Phi		0	-	0	2	-
wgt total	44,510	44,520	-	43,920	43,900	-
outliers ID		691	**-**	691	561	**-**
RMSE		5.43E+84	-	3.46E+91	857	-

Figure 1. Month 32 weighted y vs weighted x

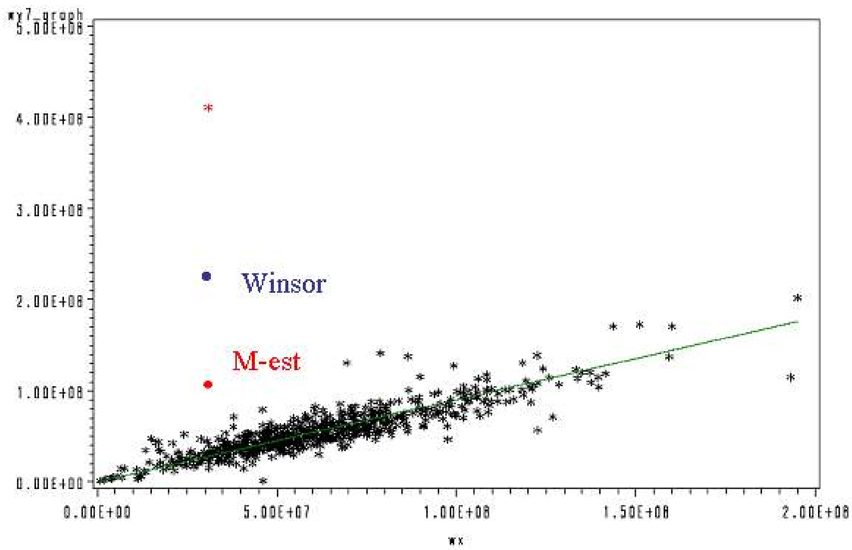
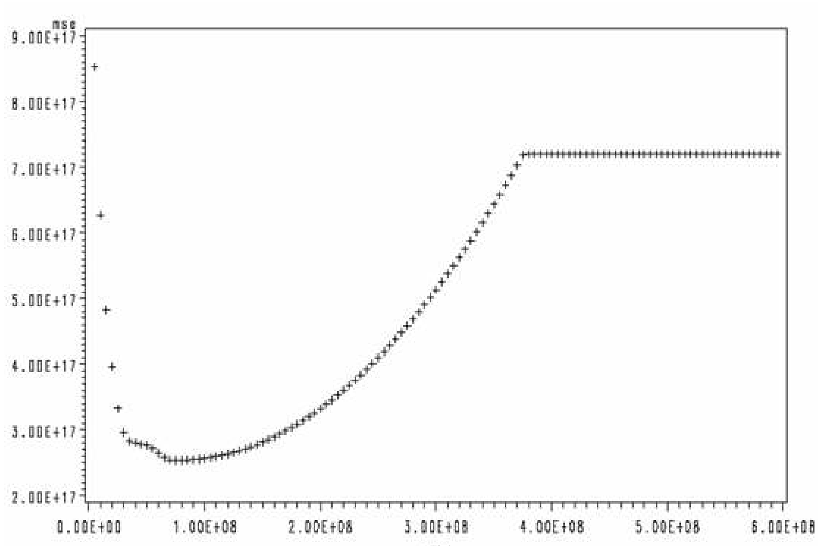


Figure 2. Month 32 M-est MSE vs φ



Month 22

All but one of the methods converged and produced adjusted values for the influential value in Month 22. Table 1 shows that the apparent weighted outlier is 310 million with a weight of 5 and a weighted residual of approximately 200 million. With the Huber I ψ and $v_i = x_i$, the adjusted value equals 180 million resulting in a total of 40,090 million, a reduction of 0.3 percent. The minimum MSE occurs when the value of φ is 100 million leading to a RMSE of about 345 million. The combination of the Huber I ψ and $v_i = 1$ produces an adjusted value of 96 million, a total of 40,010 million which is a reduction of 0.5 percent and a RMSE of 418 million. Using $v_i = \sqrt{x_i}$ instead produces an adjusted value of 120 million and a total of 40,040 million with a RMSE of 375 million. When the Huber II ψ is combined with $v_i = x_i$, the M-estimation algorithm does not converge.

Applying Winsorization in Month 22 produces an adjusted value of about 200 million which implies a total of 40,100 million.

Month 17

Month 17 is an example of the M-estimation methods converging but not being helpful while Winsorization produces an adjusted value. Table 1 shows that the apparent weighted outlier is 780 million with a weight of 10 and the residual is about 510 million. Table 1 also shows that the optimal tuning constant φ, with the Huber I ψ and $v_i = x_i$, is also close to 510 million, which indicates that the optimal adjusted value produced by the algorithm is close to the outlier. The MSE is a decreasing function of the tuning constant φ and is dominated by the bias. Table 1 shows that the other four methods also produce an adjusted value that is within 15 million of the original value causing little change in the total. There is not much difference in the RMSE values which range from 327 million to 376 million.

Applying Winsorization to Month 17 produces an adjusted value of about 520 million implying an adjusted total of 46,610 million, a reduction of 0.55 percent.

Month 27

Month 27 has only one influential value and it is too low, not too high. Only the methods where $v_i = 1$ identify the outlier. The apparent weighted outlier is 150 million with a weight of 6 and a weighted previous month 300 million. Using the Huber I ψ and $v_i = x_i$, the weighted residual is approximately -270 million and the optimal φ is 270 million, but the algorithm does not detect the potential outlier. If it did, the optimal adjusted value produced by the algorithm would have been equal to the outlier. The MSE is a strictly decreasing function of φ. Both methods where $v_i = x_i$ do converge and produce an adjusted value of 230 million, a total of 46,300 million which is an increase of 0.2 percent, and RMSE estimates of 484 million and 466 million. Rounding causes the adjusted values to appear equal while the RMSE values are not equal.

The Winsorization is one-sided so it is not appropriate for producing an adjusted value in this case.

Month 33

Month 33 has both a very high apparent weighted outlier of 370 million with a weight of 20 and a very low apparent weighted outlier of 42 million with a weight of 55. With the Huber I ψ and $v_i = x_i$, the corresponding weighted residuals are approximately 320 million and -320 million. The MSE is dominated by the variance and is a strictly increasing function of φ so the optimal value of φ is zero. This implies 691 outliers, far too many. The two methods where $v_i = 1$ produce similar results. The algorithm does not converge for the Huber II ψ with $v_i = x_i$ or the Huber I ψ with $v_i = \sqrt{x_i}$. However, if a method adjusted both apparent outliers, the adjustments would largely offset each other, in which case the estimated total would be the same. Therefore, not adjusting these observations is likely an acceptable outcome.

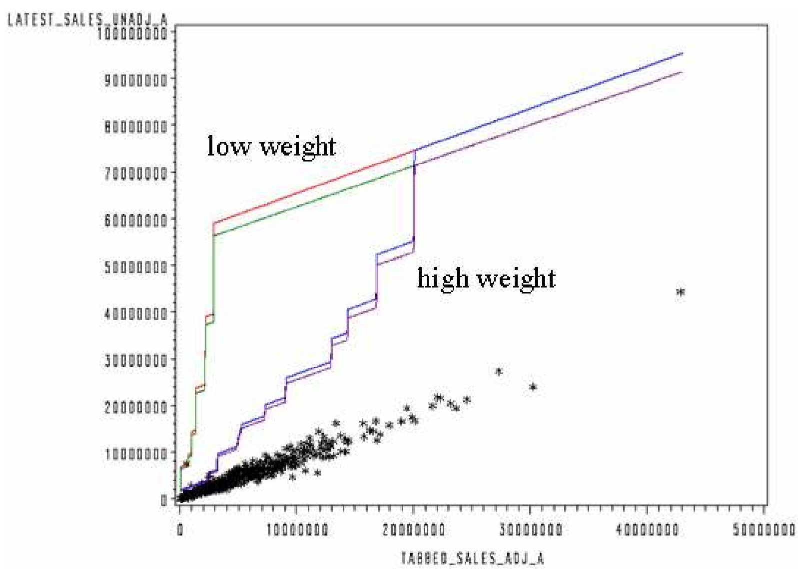
Winsorization is a one-sided method so it is appropriate for producing an adjusted value for only the large outliers. For that outlier, the adjusted value is about 210 million with a total of 44,350 million, a reduction of 0.36 percent.

3.3 Implications of weight of observation

To examine the effect of the weight on a method, we calculated the minimum value of the current month that the treatment identifies as influential given the value for the previous month. We use the term critical region to mean all the values greater than the minimum for being influential.

The minimum value for the weighted value to be designated influential is smaller for large weights than for small weights. Figure 3 shows the critical regions for unweighted data from Month 32. The lines denoting the critical regions for observations with high weights are in red for M-estimation and in green for Winsorization. The lines marking the critical regions for observations with low weights are in blue for M-estimation and purple for Winsorization.

Figure 3. Month 32 Unweighted critical region
for M-estimation (red & blue) and Winsorization (green & purple)



4. SUMMARY

The M-estimation method has the advantage over the Clarke method of being able to perform one-sided and two-sided adjustments although one could adapt the Clarke method to obtain a method that would adjust influential values that are too low. The M-estimation method also can be implemented with several settings. Setting the value produced by setting $v_i = x_i$ probably is preferable unless the percentage change in the total is deemed unacceptable. Setting $v_i = x_i$

does not give more influence to the larger weighted values of x_i. However, when there is an outlier that is too low, it will probably have a large weighted value of x_i and setting $v_i = 1$ may be necessary to obtain a reasonable adjustment. On the other hand, no adjustment may be preferable when no adjustment is recommended with the setting $v_i = x_i$. Running either method on an ongoing basis often will reduce the number of influential values. Adjusting an influential value in one month may reduce the potential for an influential value in the next month. For example, when we adjusted the influential value in Month 32, and then ran the M-estimation method for Month 33, the influential value that was too low in Month 33 no longer appeared influential.

Our next step is to develop a system to run in parallel with the ongoing data collection on a trial basis to see if the methods are helpful in practice. We plan to investigate two approaches for improving the convergence of the M-estimation algorithm. One strategy is to standardize the observations, and the other is using a fixed φ rather than estimating it every month. A related line of research is investigating whether the methods are helpful with resolving outliers when implementing a new sample.

REFERENCES

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.

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