

# Setting M-Estimation Parameters for Detection and Treatment of Influential Values

Mary H. Mulry<sup>1</sup>, Stephen Kaputa<sup>1</sup>, and Katherine J. Thompson<sup>1</sup>

Recent research on the use of M-estimation methodology for detecting and treating verified influential values in economic surveys found that initial parameter settings affect effectiveness. In this article, we explore the basic question of how to develop initial settings for the M-estimation parameters. The economic populations that we studied are highly skewed and are consequently highly stratified. While we investigated settings for several parameters, the most challenging problem was to develop an “automatic” data-driven method for setting the initial value of the tuning constant  $\varphi$ , the parameter with the greatest influence on performance of the algorithm. Of all the methods that we considered, we found that methods defined in terms of the accuracy of published estimates can be implemented on a large scale and yielded the best performance. We illustrate the methodology with an empirical analysis of 36 consecutive months of data from 19 industries in the Monthly Wholesale Trade Survey.

*Key words:* Outlier; economic surveys; Monthly Wholesale Trade Survey.

## 1. Introduction

The outlier detection and treatment in sample surveys is a recurrent challenge. Of course, the values provided by the sample units are intended to “represent” other – similar – units. However, in a business survey setting, the size of a business can change along with the economy. Along the same lines, it is not unprecedented for a business to experience an exceptionally good or poor month. The skewed nature of business populations requires stratification and differential sampling rates, with the largest units often selected with certainty to avoid bias and reduce the variance of the survey estimates. Occasionally, a sample unit in a stratum with a large weight has unexpectedly high or low value for the item collected, such as revenue. Business survey data undergo editing and review routinely prior to use in the formation of estimates such as totals or period-to-period change, and such values are frequently validated and left unadjusted until the final tabulations are reviewed. We define such an observation as influential if its value is correct, but its weighted contribution has an excessive effect on the estimated total or

<sup>1</sup> U.S. Census Bureau. 4600 Silver Hill Road, Washington, DC 20233, U.S.A. Emails: [mary.h.mulry@census.gov](mailto:mary.h.mulry@census.gov), [stephen.kaputa@census.gov](mailto:stephen.kaputa@census.gov), and [katherine.j.thompson@census.gov](mailto:katherine.j.thompson@census.gov).

**Acknowledgments:** This article 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. 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 Eric Slud, Carma Hogue, and William Abriatis are appreciated very much, as are the reviews by the Associate Editor and three referees.

period-to-period change. Failure to “treat” such verified influential values may lead to substantial over- or under-estimation of survey totals, and the resultant change estimates. “Treatment” reduces the variance in a classic bias versus variance trade-off, motivated by the desirability of low variance for economic indicators.

To illustrate the problem faced by survey staff when an influential value appears, we recount a story told to us as true although it may be better described as folklore. A T-shirt shop on an East Coast beach in the U.S. was in sample. The beach community advertised special events to attract college students on spring break, and they came *en masse*. Evidently, the T-shirt shop had the most popular T-shirt, and almost all the beach visitors bought one for themselves and some for their friends back home. The T-shirt shop had outsized revenue for the month that included spring break, and being a small business, its stratum had the largest weight in the sample. Imagine the survey staff’s surprise when they realized the T-shirt shop’s weighted value moved Gross Domestic Product. The shop’s good fortune appeared to be an isolated event and did not reflect economic conditions among T-shirt shops or other businesses. Even if the shop’s revenue was true, a T-shirt shop alone should not move Gross Domestic Product. The weight or the value needed an adjustment.

Currently in the Economic Directorate at the U.S. Census Bureau, when an influential value is detected, the mitigation strategy depends on whether the subject matter experts believe the observation is a one-time phenomenon or a permanent shift. If the influential value appears to be an atypical occurrence for the business, then the influential observation may be replaced with an imputed value or simply excluded from the group of observations eligible to be used as imputations. If the influential value persists, indicating a permanent change, then methodologists adjust its sampling weight. Ideally, the replacement (imputed) value or the adjusted weight should be set using a statistically rigorous, automated, data-driven method.

For several years, our goal has been to find such a method, under the strong constraint of minimizing additional trimming of previously validated values. The U.S. Census Bureau publishes imputation rates along with the estimates. Adjustments as part of influential value mitigation do fall under the umbrella of imputations; a high imputation rate is often considered indicative of poor data quality and is discouraging to potential respondents. We attempt to reduce the Mean Square Error (MSE) as much as possible, while changing only the influential value or values, thus preserving as much of the reported and validated data as possible. This is not equivalent to finding the optimal estimator in terms of achieving the minimum MSE over all possible estimates. In the latter case, one would attempt to guarantee a global solution for the problem at hand in the current data set, as recently studied by [Martinoz et al. \(2015\)](#) and [Clark et al. \(2017\)](#). These papers offer methodology for treating influential values in business surveys, but produce global solutions that adjust both influential and non-influential observations. In a similar vein, if preserving reported values is not an overarching concern, then one should consider robust estimation methods for sample surveys, like those described by [Duchesne \(1999\)](#), [Gwet and Rivest \(1992\)](#), [Hulliger \(1995, 1999\)](#), [Beaumont et al. \(2013\)](#), and [Mulry et al. \(2016\)](#).

Building on previous studies, we present research on the implementation of M-estimation ([Beaumont and Alavi 2004](#); [Beaumont 2004](#)) in an automated data-driven manner that can be implemented in an environment that processes a large number of surveys simultaneously. M-estimation is designed to produce an adjusted value that

minimizes the design-based estimator of the MSE. The procedure itself is extremely flexible, allowing for detection of high, low, or both high and low influential values, accommodating variations in the prediction models, and providing control over the outlier-detection region via numerous parameters (each with a specific function). However, this same flexibility complements implementation, especially when processing must be completed on a tight time schedule.

Previous studies compared M-estimation to Clark Winsorization (Clark 1995) on both empirical (historic) and simulated data. Mulry and Feldpausch (2007) applied these two methods – along with others – to empirical data in volatile industry from the Monthly Retail Trade Survey and found that only Clark Winsorization and M-estimation met the above-stated criteria of reducing bias and MSE in a high influential value situation. Mulry et al. (2012, 2014; Mulry et al. 2013) further studied the performance of these two methods using twelve months of data from two simulated populations constructed in a manner that realistically reflected two retail trade industries, one more volatile and one more stable. They examined the performance of both methods over repeated samples in terms of the bias, variance and MSE, conducting both conditional (only replicates with the influential value selected) and unconditional (all replicates) analyses. M-estimation performed better according to the evaluation criteria under selected parameter settings; Clark Winsorization performed comparably when an influential value was present, but when there was no influential value in the sample, the Winsorization changed the values of non-influential observations by small amounts to achieve a minimum MSE by reducing variance.

Our previous studies highlighted the importance of setting the appropriate parameters for the one-sided M-estimation algorithm. We decided not to pursue the two-sided M-estimation algorithm, because in our simulations, it had convergence problems when the one-sided version did not. Furthermore, low influential values occur rarely in the studied surveys because the industry totals are very large (USD billions). This is further discussed in Subsection 3.2. This article proposes methods for setting parameters for the one-sided M-estimation algorithm, using a variety of common statistical data analysis tools, thus synthesizing our experiences with one-sided M-estimation gained from previous simulation studies and empirical analyses. As mentioned above, we are applying these methods to estimates from a monthly survey, which has limited processing and production review time from data collection to estimate release. The proposed methods complement the existing subject matter review procedures. However, to be useful, parameter development should not be a part of the monthly review process, although the subject matter experts and methodologists need to be advised on viable troubleshooting approaches in the case of failure (e.g., over half of observations identified as influential). The methods have the advantage that they can be implemented on programs that publish statistics for many varied domains (such as industry) on a limited time schedule. We explore the effectiveness of the resultant parameters on empirical data from the Monthly Wholesale Trade Survey (MWTS) and identify the method that yields the best performance.

## 2. M-Estimation Method

The description of the M-estimation method (Beaumont and Alavi 2004) in our application follows Mulry et al. (2012, 2014). First, we introduce the notation. For the

$i$ th business in a survey sample of size  $n$  for the month of observation  $t$ ,  $Y_{ti}$  is the collected characteristic (e.g., 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 the previous month's revenue. The monthly total  $Y_t$  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 our empirical applications, the survey weight  $w_{ti}$  is the design weight, since the missing data treatment is imputation and no other weight adjustments are made; see [Mulry et al. \(2012, 2014\)](#).

The M-estimation regression technique proposed by [Beaumont and Alavi \(2004\)](#) uses the Schweppe version of the weighted generalized technique ([Hampel et al. 1986](#), 15–316). This approach results in a consistent estimator of the total for a finite population because it equals the finite population total when a census is conducted ([Särndal et al. 1992](#), 68).

A key assumption of the M-estimation approach is that  $y_i$  given  $x_i$  is distributed under the prediction model  $m$  with  $E_m[y_i|x_i] = x_i'\beta$  and  $V_m[y_i|x_i] = v_i\sigma^2$ . In our application,  $y_i$  is the current month's value;  $x_i$  is the previous month's value, and the regression model does not include an intercept. In retail trade, the regression of the current month's sales on the previous month's sales tends to go through the origin ([Huang 1984, 1986](#)).

Briefly, the method estimates  $\hat{B}^M$  using

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

where  $v_i = \lambda x_i$ ,  $w_i^*(\hat{B}^M) = w_i \psi\{r_i(\hat{B}^M)\}/r_i(\hat{B}^M)$ , and  $r_i(\hat{B}^M) = h_i e_i(\hat{B}^M)/Q\sqrt{v_i}$  with  $e_i(\hat{B}^M) = y_i - x_i\hat{B}^M$ .

The variable  $x_i$  may be a vector and the regression estimation model  $e_i(\hat{B}^M) = y_i - x_i\hat{B}^M$  may or may not include an intercept. Our applications use a no-intercept linear regression model, where the independent variable is the previous month's tabulated value for the same item. This ratio model is commonly used for item imputation in business surveys, as previous period values are often very good predictors of the current period value when data collection is fairly frequent (e.g., weekly, monthly, or quarterly) and the intercept term is usually not significant.

The role of the Huber function  $\psi$  ([Huber 1964](#)) 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  $\psi$ , Huber I function  $\psi_I$  and Huber II function  $\psi_{II}$ , and describe their one- and two-sided-forms. The one-sided form of each Huber function is:

$$\psi_I\{r_i(\hat{B}^M)\} = \begin{cases} r_i(\hat{B}^M), & r_i(\hat{B}^M) \leq \varphi \\ \varphi, & \text{otherwise} \end{cases}$$

$$\psi_{II}\{r_i(\hat{B}^M)\} = \begin{cases} r_i(\hat{B}^M), & r_i(\hat{B}^M) \leq \varphi \\ \frac{1}{w_i} r_i(\hat{B}^M) + \frac{(w_i - 1)}{w_i} \varphi, & \text{otherwise} \end{cases}$$

where  $\varphi$  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 both Huber

functions occurs when  $r_i(\hat{B}^M) > \varphi$ . In the two-sided Huber I and Huber II functions,  $r_i(\hat{B}^M)$  is replaced by its absolute value  $|r_i(\hat{B}^M)|$ .

With M-estimation, the user has a choice of adjusting the weight of the influential value or adjusting its value, according to [Beaumont and Alavi \(2004\)](#). The weight adjustments corresponding to the Type I Huber function  $\psi_I$  and Type II Huber function  $\psi_{II}$  above are

$$w_{II}^*(\hat{B}^M) = \left\{ \begin{array}{l} w_i, r_i(\hat{B}^M) \leq \varphi \\ \frac{\varphi}{r_i(\hat{B}^M)}, otherwise \end{array} \right\} \quad w_{III}^*(\hat{B}^M) = \left\{ \begin{array}{l} w_i, r_i(\hat{B}^M) \leq \varphi \\ 1 + (w_i - 1)\frac{\varphi}{r_i(\hat{B}^M)}, otherwise \end{array} \right\}.$$

An undesirable feature of using the 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 ensures that all adjusted units are at least fully represented in the estimate.

To solve for  $\hat{B}^M$ , [Beaumont and Alavi \(2004\)](#) present an Iteratively Reweighted Least-Squares algorithm, which we use in our application of M-estimation. For certain choices of the weights and variables, the solution is the standard least-squares regression estimator. The M-estimation algorithm takes into account both the size of an observation’s weight and its weighted value when designating influential values. Typically, the sampling rate for small businesses is lower than for larger businesses because there are more small businesses. Therefore, the smaller businesses typically have higher weights. If two sample units have an equal unusually high amount of weighted difference between their current and previous months’ values, the M-estimation method is less likely to designate the one with the lower weight as an influential value.

For an adjustment to the influential value using either Huber function, [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

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

Using the Newton-Raphson algorithm from numerical analysis, [Beaumont \(2004\)](#) finds an optimal value of the tuning constant  $\varphi$  by deriving, and then minimizing, a design-based estimator of the MSE. At each iteration, the algorithm estimates the bias by comparing the predicted total to the original total and estimates the variance using the residuals from the robust regression.

### 3. Setting Algorithm Parameters

The M-estimation algorithm discussed in Section 2 requires settings for  $Q$ ,  $h_i$ ,  $v_i$ , the function  $\psi$ , and an initial value of the tuning constant  $\varphi$ . In this section, we propose methods for setting the parameters for the M-estimation algorithm discussed in Section 2, providing illustrative examples for each proposal. [Table 1](#) summarizes the parameters for the M-estimation algorithm.

We suggest using the default settings for the parameters  $Q = 1$  and  $h_i = (w_i - 1)\sqrt{x_i}$  in the SAS software for implementing the method that we received from the developer

Table 1. M-estimation algorithm parameters.

Parameter	Parameter function	Values	Discussed
$Q$	Constant	= 1 (default)	Section 3
$h_i$	Unit weight	= $(w_i - 1)\sqrt{x_i}$ (default)	Section 3
$v_i$	Model error underlying regression estimator	= 1 or $x_i$	Subsection 3.1
$\psi$	Huber function	Huber I or Huber II	Subsection 3.2
$\varphi$	Tuning constant (determines starting point for detection region)	User provides initial value and program calculates optimal value	Subsection 3.3

Jean-Francois Beaumont through personal communication. The following sections explore the potential impact of different settings for the other  $v_i$ ,  $\psi$ , and  $\varphi$ . To illustrate the importance of the selections of  $v_i$ , and  $\varphi$  on the effectiveness of the algorithm, we use data from two simulated MRTS industries, which are discussed in detail in [Mulry et al. \(2014\)](#).

For our investigation of a data-driven method for setting the initial  $\varphi$ , we use 36 consecutive months of empirical (edited/imputed) values of sales and inventory from 19 industries in the MWTS for the presented analyses. The MWTS is a monthly survey that collects sales and inventories data and uses a stratified SRS-WOR design, with industry as the primary strata and unit size group strata defined within the industry strata. Updating the sample to include new businesses and remove failed businesses reduces coverage bias and keeps the sample from attrition. A new sample for each industry is selected approximately every five years. There is very little overlap in small businesses in samples selected for adjacent periods, but the overlap can be quite high for large businesses with substantive inventories. The MWTS publishes *industry level* tabulations. Influential values are considered at the industry level rather than at the industry-size-stratum level. Treatment of influential values is the final step of the estimate review process. Hence, the methods described here are developed for implementation subsequent to editing of the data. For more details on the MWTS estimation and review procedures, see [http://www.census.gov/wholesale/www/how\\_surveys\\_are\\_collected/monthly\\_methodology.html](http://www.census.gov/wholesale/www/how_surveys_are_collected/monthly_methodology.html).

### 3.1. Parameter $v_i$

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, denoted by  $e_i(\hat{B}^M) = y_i - x_i\hat{B}^M$  in Section 2. To determine the value of  $v_i$ , we suggest fitting the following weighted linear regression models at the adjustment cell level:

	Regression model	M-estimation parameter setting
(1)	$y_i = \beta_i x_i + \varepsilon_i, \varepsilon_i \sim (0, \sigma^2)$	$v_i = 1$
(2)	$y_i = \beta_i x_i + \varepsilon_i, \varepsilon_i \sim (0, x_i \sigma^2)$	$v_i = \sqrt{x_i}$
(3)	$y_i = \beta_i x_i + \varepsilon_i, \varepsilon_i \sim (0, x_i^2 \sigma^2)$	$v_i = x_i$

The first model is a typical ordinary least-squares regression, appropriate if all units in the adjustment have the same expected value and variance, and equal sampling weights (e.g., the adjustment cells are sampling strata in a stratified simple random sample). The other two weighted regression models have proved useful in numerous business survey settings (Thompson and Sigman 1996; Huang 1984, 1986; Binder et al. 2000). After fitting each model separately within the adjustment cell, find the model that best corrects heteroscedasticity (unequal error variances) in the majority of adjustment cells. We suggest using White’s Test for heteroscedasticity (White 1980) and an examination of the regression residuals when determining which model is best. In general, economic surveys have stratified sample designs. Although the selection within strata is often equal probability, the probabilities of selection usually vary across the strata. If no such model can be found, then the experience of other researchers has shown that the ratio model  $v_i = x_i$  (Model 3) appears to be fairly robust to model misspecification with economic data (Huang 1986, Beaumont 2004). For further discussion on the choice of  $v_i$  in the application of M-estimation in this article, see Mulry et al. (2014).

In our MWTS applications, we use the previous month’s (edited/imputed) value of  $y$  as the independent variable and fit a no-intercept regression model (the ratio estimation model). We found that Model (3) was the best fit for the MWTS data.

The selection of  $v_i$  can be the factor determining the effectiveness of the algorithm, as demonstrated in Figure 1, where the only difference in the two applications is the setting of  $v_i$ . In Figure 1, when  $v_i = x_i$ , the MSE was a concave function of  $\varphi$ . A minimum MSE was achieved and an adjustment produced for the influential value. However, when  $v_i = 1$  was applied to the same sample data, the MSE was a strictly decreasing function of  $\varphi$ . The minimum MSE occurred at the influential value, and no adjustment was made because the

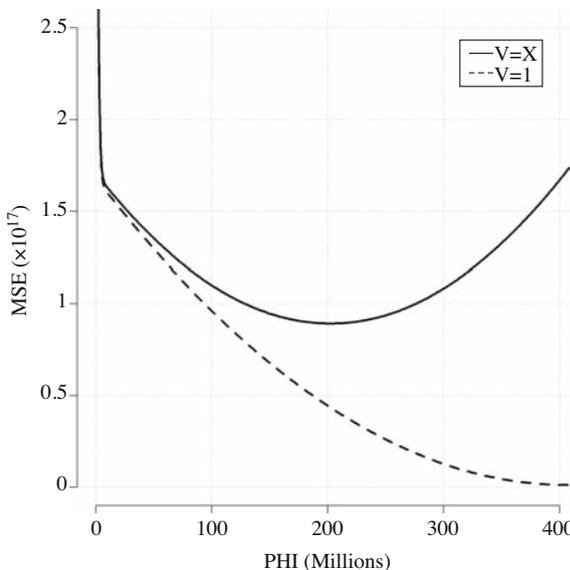


Fig. 1. M-estimation MSE vs.  $\varphi$  for a sample for one month with an influential value selected from a simulated MRTS industry using the one-sided Huber II function  $\psi$ .

algorithm failed to detect the influential value. We have encountered other analogous situations with in retail trade industries in the MRTS, so the configuration is not unique.

For some insight on the difference in outcome when  $v_i = x_i$  and  $v_i = 1$ , note that with the defaults  $Q = 1$  and  $h = (w_i - 1)\sqrt{x_i}$ , setting  $v_i = 1$  tended to give the residuals for large weighted values of  $x_i$  more influence in fitting the M-estimation regression line than when  $v_i = x_i$ . The variable  $Q$  may be used as a scaling factor to dampen the effect of unequal variances and is set to equal 1 for equal or approximately equal variances. The variable  $h$  is a weighting factor, where using the default setting excludes certainty units from the regression.

### 3.2. Function $\psi$

The first decision regarding the function  $\psi$  is whether to use the Huber I or Huber II function, as described in Section 2. In [Mulry et al. \(2014\)](#), results were comparable for both one-sided and two-sided Huber I and Huber II functions when  $v_i = x_i$  and  $v_i = 1$ . To avoid the risk of obtaining adjusted weights with values less than one, we hereafter consider only the Huber II function.

After the choice of the function  $\psi$  is made, the second decision is whether to use the one-sided version or the two-sided version. The one-sided version, by design, detects only unusually high influential values, whereas the two-sided version is able to detect both unusually high and unusually low influential values. On the surface, the two-sided function is very appealing, as the potential for high or low outliers can exist. However, one issue to consider in choosing the one-sided or two-sided function  $\psi$  is that the M-estimation algorithm using a two-sided function  $\psi$  may experience some problems with convergence for some scenarios where the second value is 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.

To study how such convergence issues might affect our business survey data, we investigated the detection region for the scenarios with two influential values with simulated sample data. The sample initially contained one high fixed influential value. We systematically added a second value as a candidate for detection and thereby determined an influential value detection region, given the first fixed value. Interestingly, we found that the detection regions for a high and low second value are not symmetric around the robust regression line. When the second value was high, it had a positive residual and moved from not influential to influential as the value of its residual increased. When the second value was low, its residual was negative. As its value declines and is larger in absolute value, the second value moves from not being identified as influential into a range where the algorithm does not converge, and then to values identified as influential.

To summarize, when a sample contains both unusually high and unusually low influential values and the M-estimation algorithm does not converge, the researcher may want to consider the option of no adjustment, particularly in situations where bias is more important than variance. The rationale for choosing no adjustment is that the unusual values counterbalance each other in a manner that introduces minimal bias. With this perspective, the failure of the algorithm to identify the influential values is not necessarily

a handicap. That said, as one referee pointed out, this is a form of masking and can lead to intolerable levels of variance inflation, even if the bias level is tolerable. However, this situation is an excellent topic for further research with different weighting schemes.

With the monthly surveys that we are studying, a low influential value can be indicative of a failing business. Often, the subject matter experts are monitoring the largest cases in an industry (which all have sampling weights near 1) and have an effective down-weighting treatment strategy in place. A similar post-processing adjustment is applied to small businesses retrospectively. However, even with large sampling weights, failing small businesses rarely have an influential effect on the overall totals, as the lowest possible value is zero and the industry totals are very large. Inducing low influential values in the studied populations was often impossible; we would have needed to induce a large number of low influential values to affect the industry totals or a few large (unrealistic) negative values. The more pressing situation occurs with unusually high values that are not typical of the survey unit or its stratum. A large increase in the estimated total of an economic indicator (beyond sampling error) can have policy implications in both the public and private sectors, which in turn can have ramifications for the survey agency and policymakers if the total returns to its original level in the subsequent collection. Consequently, our applications are confined to the one-sided Huber II function.

### 3.3. Tuning Constant $\varphi$

#### 3.3.1. Effect of the Initial Value of the Tuning Constant on Detection Regions

Recall that the M-estimation algorithm finds an optimal value of  $\varphi$  by minimizing the design-based estimator of the MSE. An observation's weighted residual has to exceed the initial  $\varphi$  for the algorithm to consider it as a possible influential value. Viewing the MSE as a function of  $\varphi$  implies that a value of  $\varphi$  corresponds to a "treated" value for each observation with a weighted residual greater than  $\varphi$ . When a minimum exists, the initial value of  $\varphi$  has to be "close enough" to the minimum for the algorithm to find it. If the initial  $\varphi$  value is too low and the sample does not contain any influential values, the algorithm can fail to converge or can converge to zero (or a very small number). If the initial  $\varphi$  is larger than all the weighted residuals, the algorithm does not find the minimum MSE because the MSE is a constant function in the neighborhood of  $\varphi$ . In practice, we have found that the effectiveness of the Newton Raphson algorithm is highly sensitive to the choice of the initial tuning constant  $\varphi$  when there are zero or one influential values in the adjustment cell, since the initial  $\varphi$  determines the lower boundary of the detection region (Mulry et al. 2012).

Figure 2 uses sample data for one month selected from a simulated MRTS industry to illustrate the MSE as a function of  $\varphi$  in a sample without an induced influential value. Figure 2 also shows the effect of an influential value that is induced in the same sample by adding four different amounts to a unit selected at random from those with a sampling weight of 60 (i.e., a "small" business). When there is no influential value, the MSE function has a slope of zero since it is a constant function of  $\varphi$  equal to  $5.97 \times 10^{13}$  (although the MSE may appear equal to zero on the scale used in Figure 2). As the amount added to the unweighted observation to induce an influential value increases from two

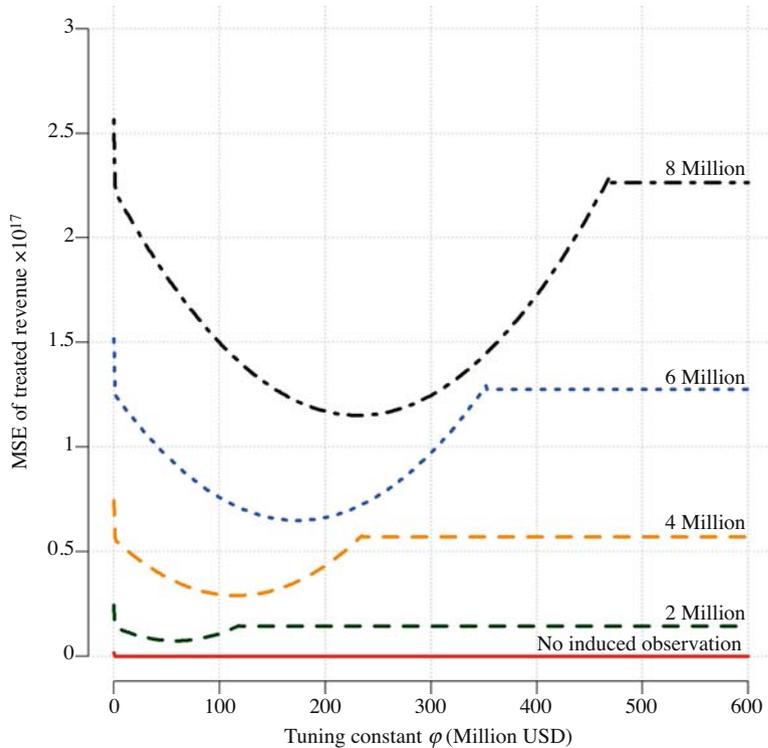


Fig. 2. MSE as a function of the tuning constant  $\phi$  when no influential value is present (MSE has a constant value of  $5.97 \times 10^{13}$ ) and when an influential value is induced by adding four different amounts to an unweighted observed value with weight 60. Data is from a sample selected for one month from a simulated MRTS industry.

million to eight million, the curve shifts, as does the value of  $\phi$ , where the minimum MSE occurs. Notice that with an initial  $\phi$  set at about 470 million, the algorithm would not find the minimum MSE when any one of the four induced influential values were present. Also, notice that for the algorithm to find the minimum MSE for all four induced influential values, the initial  $\phi$  must be approximately 100 million or less.

Figure 3 uses unweighted data from a sample for one month selected from a simulated MRTS industry to illustrate the detection regions for the application of the algorithm in the case where there is no predetermined (identified) influential value in the sample. In these figures, we selected a low value and a high value of the initial  $\phi$  via graphical analysis. The chosen low value was expected to force the algorithm to run on the studied data, whereas the high value was selected to be a value in the area where the MSE levels were a constant function of  $\phi$ .

The size of an observation's weight, as well as its weighted value, both affect whether it will be designated as influential by M-estimation. The unweighted values of sales from smaller businesses tend to be lower than from larger businesses due to stratum weighting differences, even for those cases identified as influential via M-estimation. However, this is not universally true in an ongoing sample. For example, when the new sample is introduced, at the time of frame determination the smaller businesses will likely have large

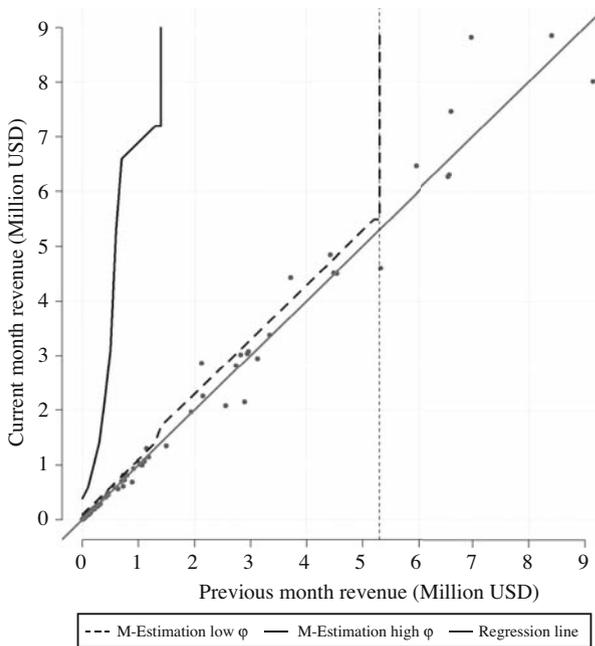


Fig. 3. Lower boundaries of detection regions for high and low initial  $\varphi$ . When using the lower initial  $\varphi$ , the algorithm will identify the addition of a single observation above the dotted line as influential. The algorithm also identifies the addition of a single observation above the solid line as influential when using the higher initial  $\varphi$ . Data is from a sample selected for one month from a simulated MRTS industry.

weights and low probability of selection, whereas the large businesses will be included with certainty or with high probability (small weights). However, as the sample matures, some sampled businesses may “outgrow” their original size strata - in some cases reporting the values of total receipts that would categorized them as large (certainty) businesses at the time of sample design. These “stratum jumpers” therefore influence estimated totals and their weights may be adjusted (reduced) accordingly, thus causing more variability in the weights for the smaller businesses.

Figure 3 overlays the boundaries of the detection regions obtained with M-estimation with a low initial  $\varphi$  and M-estimation with a high initial  $\varphi$ . The unweighted sample observations used to form the detection regions are shown as gray dots with the x-axis representing the previous month’s value and the y-axis representing the current month’s value. The least-squares regression line for the model used in the M-estimation application has been added. For the given sample, the addition of a single observation above the black line, which may be dashed or solid, will cause it to be flagged as influential and adjusted. The dotted vertical bar marks the largest observation with a weight greater than one in the sample and population; that is, all observations to the right are guaranteed to have a weight of one.

Figure 3 shows that the detection region obtained using the M-estimation-high  $\varphi$  is much more restrictive and does reduce bias. The close proximity to the regression line M-estimation-low  $\varphi$  reflects the trimming that both methods do to minimize the MSE by lowering the variance at the cost of introducing a small bias. Notice that the algorithm

rarely identifies large unweighted observations (those with  $w_i$  approaching 1) as influential. None of these large observations has a weighted residual that exceeds the initial high  $\varphi$ .

For small sampled businesses, large changes between the current and previous values for the same unit are not atypical. For this reason, it is crucial to set the initial  $\varphi$  to be the weighted distance between an observed and a *predicted* value in the current month expected to lead to a statistically significant change in the estimated total. The comparison to the prediction is especially important, as this reduces effects of industry-wide trends and seasonal effects. Therefore, the algorithm requires that the initial  $\varphi$  not be too high or too low, but, as Goldilocks says, just right.

The goal when setting the initial tuning constant  $\varphi$  is to be high enough to avoid detecting natural variation as influential, but low enough to detect truly influential values. Setting the initial  $\varphi$  too high may result in the algorithm failing to detect influential values lower than the initial  $\varphi$ . When none of the values in the sample is larger than the initial  $\varphi$ , the algorithm runs for one iteration and then stops. In this circumstance, the MSE is a constant function in a neighborhood of the initial  $\varphi$ , and the algorithm continues to run only when it detects a change in the MSE in the proximity of the initial  $\varphi$ .

On the other hand, setting the initial  $\varphi$  too low causes the algorithm to give the influential designation to observations not considered influential. This occurs because the algorithm achieves a minimum MSE when there is no influential value by trimming about 0.05 percent of the observations for a very small reduction in the MSE. In an ongoing survey, an initial  $\varphi$  that is too low may also cause convergence problems in a month following an adjustment if the unit returns to its more stable level from two months earlier. In this case, the adjusted value will appear to be unusually low. In some cases, both one-sided and two-sided functions  $\psi$  have convergence problems (Mulry et al. 2014).

Our objective is to minimize the design-based MSE, while limiting adjustments to previously validated items; therefore the initial  $\varphi$  is important for the success of the M-estimation implementation. This is largely achieved by selecting an initial  $\varphi$  that essentially defines the detection region – quickly – under an assumption of a rare prevalence of influential values. However, if M-estimation is the *only* (or primary) form of data review and trimming is not constrained, then the dependence on the initial  $\varphi$  should be minimal, as the adjustment cell would be expected to contain several outlying or inconsistent values with respect to the M-estimation regression model, and the algorithm should converge. That said, the M-estimation application could be enhanced to switch to a global search for an initial  $\varphi$  when there is no gradient in the estimated MSE to guard against the incidence of a single outlier/inconsistent value, as suggested by two referees. Continuing with our approach, the following section describes how we incorporate survey design requirements into the parameter settings.

### 3.3.2. Automated Data-Driven Methods of Obtaining Initial $\varphi$

The first attempts of our quest to find a general method for setting the initial  $\varphi$  relied on the M-estimation weighted residual distributions. Exploratory data analyses provided solid evidence against normality, so we attempted to find alternative distributions that provided a better fit for the residuals. As mentioned in Subsection 3.1, all the weighted regression models of current month's value on the previous month's value did not have an intercept

and used the regression weight equal to the sample design weight divided by the previous month's observation. The data for each model included the certainty units.

Unfortunately, no single distribution worked well for all studied populations and, in many cases, the best fit appeared to be a complex blending distribution with no finite moments. We also explored fitting regression models within industry/strata. For this, the residuals from all the industry/strata level models were combined to form an industry level set that was comparable to the residuals from the industry level model. The industry level and industry/strata level weighted robust regression models were fit for all 19 MWTS industries over 36 months. The Kruskal Wallis test (non-parametric version of ANOVA) was used to compare the differences between the two groups. While differences did exist, no clear pattern of differences between the two groups appeared. Therefore, we decided to use the residuals from the weighted robust regression models fit at the industry level.

In the end, we settled on approaches that incorporate the survey design requirements into the parameter settings. The MWTS design has a national level coefficient of variation (*cv*) requirement of 0.01; selected industries have less restrictive *cv* requirements, ranging from 0.04 to 0.055. Assuming a constant variance (a usual assumption in ongoing surveys), an increase in the *cv* above the expected national level would likely be an attribute of a change in the total estimate (from the previous) period. An economic change could also contribute, as well as the effect of an influential value (or values), which should be investigated and possibly adjusted.

We use the half-width of the 90 percent confidence interval on the previous month's total. Values outside of the half-width should lead to a statistically significant change in the current month total. We considered two high-level methods of obtaining the confidence interval half-width: (1) use the *cv* publication requirement for national level totals to derive an estimated standard error; and (2) estimate the standard error directly from the predicted current month values. Thus, the standard errors obtained from (2) will be larger than the (1) counterparts.

This led to the following options considered for calculating the initial  $\varphi$ :

- 1) (CV\_EST) Set the initial  $\varphi$  as the product of the coefficient of variation of the estimated total (*cv*), the width of the 90 percent confidence interval using the t-distribution, and the previous month's estimated total  $\hat{T}_{t-1}$ , serving as an estimate of the current month's total:

$$\text{Initial } \varphi = cv * 1.7 * \hat{T}_{t-1}.$$

- 2) (CV\_PRED) Use the same formula as in 1), but replace  $\hat{T}_{t-1}$  with an estimate of the current month's total that uses the estimated coefficient  $\hat{\beta}_t$  from the weighted robust regression of the current month  $y_{it}$  on the previous month  $x_{it}$ :

$$\text{Initial } \varphi = cv * 1.7 * \sum \hat{\beta}_t x_{it}$$

- 3) (ST\_EST) Set the initial  $\varphi$  as the width of the 90 percent confidence interval of the previous month's estimated total  $\hat{T}_{t-1}$  using the normal distribution and assuming  $\hat{T}_{t-1}$  is serving as an estimate of  $\hat{T}_t$ :

$$\text{Initial } \varphi = 1.65 * \text{StdErr}(\hat{T}_{t-1})$$

where  $StdErr(\hat{T}_{t-1})$  is computed with the Taylor linearization estimator implemented in PROC SURVEYMEANS on the previous period sample-weighted MWTS microdata, with the MWTS design strata and incorporating the finite-population factor correction (SAS/STAT(R) 9.22 User's Guide 2016).

- 4) (ST\_PRED) Use the same formula as in 3), but replace  $\hat{T}_{t-1}$  by an estimate of the current month's total that uses the coefficient  $\hat{\beta}_t$  of the weighted robust regression estimate of the current month  $y_{it}$  on the previous month  $x_{it}$ :

$$\text{Initial } \varphi = 1.65 * StdErr(\sum \hat{\beta}_t x_{it})$$

where standard errors are computed analogously to ST\_EST, but sample-weighted predicted current period values from the M-estimation regression replace the previous-period MWTS values in the computations.

The next step was to apply the M-estimation algorithm using the four options for the initial  $\varphi$  to real historical data from the MWTS and compare the results for the options. The simulation studies described in Mulry et al. (2014) showed that the M-estimation algorithm met the criteria of detecting and treating only influential values that remained in the data after editing and verification in a manner that reduced bias and MSE for estimates of total and month-to-month change. Therefore, we proceeded with the knowledge that if the algorithm detected and treated influential values in historical MWTS data, the adjustment would improve the estimates of total and month-to-month change.

Our first concern was algorithm convergence issues with the four considered applications, specifically converging to zero (and flagging half of all observations) or failing to converge. Such problems appear in three of the 2,736 applications (four options for setting the initial  $\varphi$  to 36 consecutive months of data for 19 MWTS industries):

- For one application of the CV\_PRED option in Industry 11, the algorithm converged to zero.
- The algorithm failed to converge twice, in one application of the CV\_EST option in Industry 13 and in one application of the CV\_PRED option in Industry 2.

We do not see a pattern, so we conclude that these three occasions do not indicate a problem with the settings we have selected for the algorithm.

Table 2 shows the maximum number of influential values flagged in one month and the total number of values flagged in all 36 consecutive months when the algorithm converged by option for setting the initial  $\varphi$  for 19 MWTS industries. The results of Table 2 are summarized as follows:

- In nine MWTS industries, the results using the four options for settings of the initial  $\varphi$  agree by not flagging any influential values in any of the 36 consecutive months.
- SE\_PRED detects influential values on two occasions: once in Industry 2 when four influential values are flagged and once in Industry 15 when six influential values are flagged.
- SE\_EST does not detect any influential values in *any* month for the 19 industries.
- The options CV\_EST and CV\_PRED find more influential values than their respective SE counterparts do.

Table 2. Maximum number of influential values flagged in one month and total number of values flagged in all 36 consecutive months when the algorithm converged by option for setting the initial  $\varphi$  for 19 MWTS industries.

Industry	CV_EST		CV_PRED		SE_EST		SE_PRED	
	Maximum	Sum	Maximum	Sum	Maximum	Sum	Maximum	Sum
1	0	0	0	0	0	0	0	0
2	0	0	4	5	0	0	4	4
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	1	4	1	4	0	0	0	0
7	0	0	0	0	0	0	0	0
8	7	7	7	7	0	0	0	0
9	0	0	6	6	0	0	0	0
10	1	1	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0
12	1	2	1	2	0	0	0	0
13	1	2	1	3	0	0	0	0
14	0	0	0	0	0	0	0	0
15	2	3	2	3	0	0	6	6
16	0	0	4	4	0	0	0	0
17	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0
19	1	2	1	2	0	0	0	0

- In six industries, these two options share the same maximum detected in a month and had the same total number of detections.
- CV\_PRED finds influential values in two months in Industry 2 and one month in Industry 16 when CV\_EST does not.

Table 3 presents the mean and standard deviation of the settings of the initial  $\varphi$  from the four options when applied to the 36 consecutive months of data from 19 MWTS industries. Note that CV\_EST and the CV\_PRED have lower means and lower standard deviations than SE\_EST and SE\_PRED across the 36 months for the 19 MWTS industries, which is one reason that the CV\_EST and CV\_PRED flag influential values more often. By design, the SE methods yield higher estimated standard errors, since this approach computes industry-specific standard errors, more closely approximating the 0.04–0.05 industry reliability restrictions for MWTS alluded to at the beginning of this section. Another factor could be the standard error estimation procedure. With the monthly surveys, the standard errors can be quite variable due to the small sample size and the changing sample composition. Consequently, most of the indicators produced by the U.S. Census Bureau publish some form of average variance (or *cv*) to smooth away some of the noise.

For further insight, Figure 4 shows the maximum weighted observed residual from the robust regression of the current month’s weighted observations on the previous month’s weighted observations and the settings of the initial  $\varphi$  from the four options by month for Industry 2. We see that the settings of the initial  $\varphi$  from the CV\_EST and SE\_EST are lower than the settings from CV\_PRED and SE\_PRED, but are still low enough that they are smaller than the maximum weighted observed residual, triggering the M-estimation

Table 3. Mean and standard deviation of the settings of the initial  $\varphi$  from the four options when applied to the 36 consecutive months of data from 19 MWTS industries (100 millions).

Industry	CV_EST		CV_PRED		SE_EST		SE_PRED	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
1	4.45	0.70	4.37	0.68	12.85	3.15	12.07	3.06
2	1.20	0.10	1.09	0.18	3.37	0.23	2.79	0.90
3	1.52	0.28	1.50	0.27	8.09	1.51	7.93	1.55
4	5.62	0.52	5.42	0.65	15.50	1.57	14.41	2.04
5	3.09	0.35	3.03	0.37	12.40	1.51	11.99	1.23
6	1.98	0.27	1.95	0.26	7.11	1.27	6.94	1.37
7	4.77	0.47	4.56	0.68	12.96	2.17	11.92	2.75
8	1.41	0.16	1.35	0.18	5.70	0.70	5.29	0.82
9	5.82	0.83	5.23	0.86	25.17	3.66	21.95	3.44
10	3.50	0.39	3.37	0.52	21.49	3.15	20.48	3.83
11	1.29	0.07	1.25	0.14	8.51	0.55	8.02	1.48
12	5.89	0.37	5.82	0.43	20.24	3.70	19.10	4.86
13	2.16	0.23	2.02	0.40	10.17	1.46	9.17	2.32
14	8.69	0.81	8.53	0.98	40.43	3.32	39.47	4.96
15	3.13	0.56	2.88	0.61	13.69	3.46	11.20	4.17
16	1.47	0.14	1.42	0.18	5.74	0.71	5.42	0.97
17	8.42	1.39	8.27	1.37	26.25	4.09	25.17	4.44
18	1.77	0.23	1.75	0.23	7.66	1.10	7.57	1.12
19	3.36	0.45	3.25	0.52	13.83	2.15	13.11	2.24

algorithm. However, the values of the initial  $\varphi$  are much lower in these months when the algorithm runs than in other months. We concluded that the options using the PRED values have high variability, which is an undesirable characteristic.

In these applications, both CV methods work approximately the same and appear to be better for MWTS than the SE methods. However, the CV\_EST option is more stable than the CV\_PRED option. Moreover, the CV\_EST is easy to implement and avoids a more complicated implementation of using the results of a robust regression of the current month on the previous month. Finally, the CV\_EST option is the easiest of the four approaches to explain and to modify. Thus, we recommend using the CV\_EST option for the MWTS.

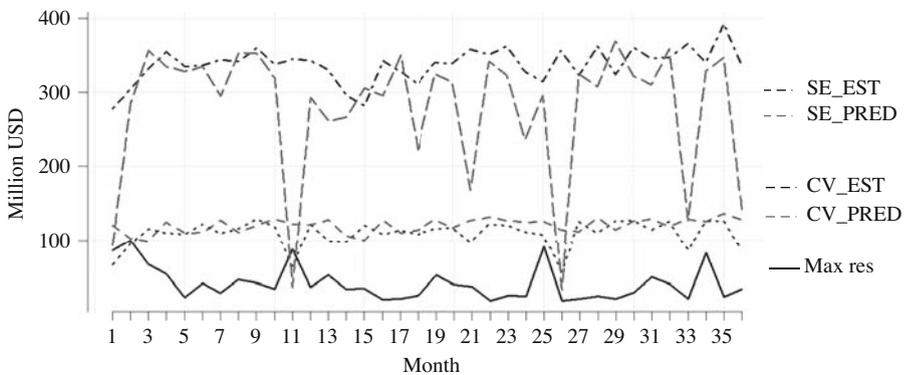


Fig. 4. Plot of maximum residual (solid black line), and values of initial  $\varphi$  from four methods for 36 consecutive months of MWTS Industry 2.

#### 4. Conclusion

Using M-estimation to identify and treat influential values in a survey setting is appealing from both methodological and statistical perspectives. The flexibility of weighted M-estimation makes it useful for a wide variety of data models, and our empirical results appear to support the algorithm's robustness to model misspecification. On the other hand, this same flexibility has the disadvantage of introducing some complexity in implementation. First, 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  $\psi$  implementation than with a one-sided function. If the occurrence of both an unusually high and an unusually low influential value in the same month causes lack of convergence, then an estimate with no adjustments is justified because the two influential values offset to result in the bias being approximately zero.

In this article, we explore the basic question of how to develop initial settings for the M-estimation parameters, focusing primarily on economic data applications. The populations that we studied are highly skewed and are consequently highly stratified. Because of this, the assumed data model that we use in our M-estimation application – a weighted robust regression model that uses survey weights and the predictor variable as regression weights – is misspecified when applied to population data. Even so, we found several advantages of using this data model over the simpler ordinary least-squares (equal variances) model.

Developing an “automatic” data-driven method for setting the initial value of the tuning constant  $\varphi$  posed a more challenging problem. The residuals of the weighted robust regression model exhibited only minimal seasonality when applied monthly. Since this parameter has the most impact on the performance of the detection of influential values, it is important to provide simple-to-use and data-based methods that are robust. Of all the methods that we considered, we found that methods defined in terms of the accuracy of published estimates yielded the best performance.

While we had success defining the initial  $\varphi$  in terms of estimated standard errors and coefficients of variation, another option to consider is whether the observation will, by itself, change the published estimate beyond what would be attributed to sampling error. Other researchers may want to examine the method for determining a data-dependent tuning constant developed by Wang et al. (2007). Alternatively, other prediction models could be considered, especially if correlated register data from the same statistical period were available.

The next step in our research is to apply the method in a side-by-side test. We will provide guidelines to the subject matter experts who have the responsibility of reviewing an adjustment proposed by the M-estimation algorithm and deciding on whether to incorporate it in the estimation each month. The dialog with subject matter experts during the test and the application of the algorithm in more industries may lead to refinements, but the basic approach appears very effective.

#### 5. 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*. Newport, U.K.: Office for National Statistics.

- Beaumont, J.-F. and A. Alavi. 2004. "Robust Generalized Regression Estimation." *Survey Methodology* 30(2): 195–208. Available at: <http://www.statcan.gc.ca/pub/12-001-x/2004002/article/7752-eng.pdf> (accessed August 2017).
- Beaumont, J.-F., D. Haziza, and A. Ruiz-Gazen. 2013. "A Unified Approach to Robust Estimation in Finite Population Sampling." *Biometrika* 100: 555–569. Doi: <https://doi.org/10.1093/biomet/ast010>.
- Binder, D.A., C. Babyak, M. Brodeur, M. Hidioglou, and W. Jocelyn. 2000. "Variance Estimation for Two-Phase Stratified Sampling." *The Canadian Journal of Statistics* 28(4): 751–764. Doi: <https://doi.org/10.2307/3315914>.
- Clark, R.G. 1995. "Winsorization Methods in Sample Surveys." Masters Thesis. Department of Statistics. Canberra: Australia National University. Available at: <http://hdl.handle.net/10440/1031> (accessed August 2017).
- Clark, R.G., P. Kokic, and P.A. Smith. 2017. "A Comparison of two Robust Estimation Methods for Business Surveys." *International Statistical Review* 85: 270–289. Doi: <https://doi.org/10.1111/insr.12177>.
- Duchesne, P. 1999. "Robust Calibration Estimators." *Survey Methodology* 25(1): 43–56. Available at: [http://hbanaszak.mjr.uw.edu.pl/TempTxt/Duchesne\\_1999\\_RobustCalibrationEstimators.PDF](http://hbanaszak.mjr.uw.edu.pl/TempTxt/Duchesne_1999_RobustCalibrationEstimators.PDF) (accessed August 2017).
- Gwet, J.-P. and L.-P. Rivest. 1992. "Outlier Resistant Alternatives to Ratio Estimator." *Journal of the American Statistical Association* 87: 1174–1182. Available at: <https://www.mat.ulaval.ca/fileadmin/mat/documents/lrivest/Publications/26-GwetRivest1992.pdf> (accessed August 2017).
- Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw, and S.A. Werner. 1986. *Robust Statistics. An Approach Based on Influence Functions*. New York, NY: John Wiley & Sons.
- Huang, E.T. 1984. "An Imputation Study for the Monthly Retail Trade Survey." In *JSM Proceedings*, Survey Research Methods Section, American Statistical Association, Philadelphia, Pennsylvania, August 13–16, 1984. Alexandria, VA: American Statistical Association. 610–615. Available at: [http://ww2.amstat.org/sections/srms/Proceedings/papers/1984\\_117.pdf](http://ww2.amstat.org/sections/srms/Proceedings/papers/1984_117.pdf) (accessed August 2017).
- Huang, E.T. 1986. "Report on the Imputation Research for the Monthly Retail Trade Survey." *Statistical Research Report Series No. CENSUS/SRD/RR-86-09*. U.S. Census Bureau. Washington, DC. Available at: <https://www.census.gov/srd/papers/pdf/rr86-09.pdf> (accessed August 2017).
- Huber, P.J. 1964. "Robust Estimation of a Location Parameter." *Annals of Mathematical Statistics* 35: 73–101. Doi: <https://doi.org/10.1214/aoms/1177703732>.
- Hulliger, B. 1995. "Outlier Robust Horvitz-Thompson Estimators." *Survey Methodology* 21(1): 79–87. Available at: <http://www.statcan.gc.ca/pub/12-001-x/1995001/article/14407-eng.pdf> (accessed August 2017).
- Hulliger, B. 1999. "Simple and Robust Estimators for Sampling." In *JSM Proceedings*, Survey Research Methods Section, American Statistical Association, Baltimore, MD, August 8–12. Alexandria, VA: American Statistical Association. 54–63. Available at: [http://ww2.amstat.org/sections/srms/Proceedings/papers/1999\\_009.pdf](http://ww2.amstat.org/sections/srms/Proceedings/papers/1999_009.pdf) (accessed August 2017).

- Martinoz, C.F., D. Haziza, and J.-F. Beaumont. 2015. "A Method of Determining the Winsorization Threshold, with an Application to Domain Estimation." *Survey Methodology* 41(1): 57–77.
- Mulry, M.H. and R. Feldpausch. 2007. "Investigation of Treatment of Influential Values." *Proceedings of the Third International Conference on Establishment Surveys*, Montreal, Quebec, Canada, June 18–21, 2007. Alexandria, VA: American Statistical Association. 1173–1179. Available at: <http://ww2.amstat.org/meetings/ices/2007/proceedings/ICES2007-000229.PDF> (accessed April 2018).
- Mulry, M.H., S. Kaputa, and K.J. Thompson. 2016. "A Cautionary Note on Clark Winsorization." *Survey Methodology* 42(2): 297–305. Available at: <http://www.statcan.gc.ca/pub/12-001-x/2016002/article/14676-eng.pdf> (accessed August 2017).
- Mulry, M.H., B. Oliver, and S. Kaputa. 2014. "Detecting and Treating Verified Influential Values in a Monthly Retail Trade Survey." *Journal of Official Statistics* 30(4): 721–747. Doi: <http://dx.doi.org/10.2478/JOS-2014-0045>.
- Mulry, M.H., B. Oliver, S. Kaputa, and K.J. Thompson. 2013. "Setting M-Estimation Parameters for Detection and Treatment of Influential Values." In *JSM Proceedings*, Survey Research Methods Section, American Statistical Association, Montréal, Québec, Canada, August 3–8, 2013. Alexandria, VA: American Statistical Association. 1424–1438. Available at: [http://ww2.amstat.org/sections/srms/Proceedings/y2013/Files/308309\\_80627.pdf](http://ww2.amstat.org/sections/srms/Proceedings/y2013/Files/308309_80627.pdf) (accessed August 2017).
- Mulry, M.H., B. Oliver, and S. Kaputa. 2012. "Several Scenarios for Influential Values in Business Surveys and Methods for Their Treatment." In *JSM Proceedings*, Survey Research Methods Section, American Statistical Association, San Diego, CA, July 28–August 2, 2012. Alexandria, VA: American Statistical Association. 4015–4029. Available at: [http://ww2.amstat.org/sections/srms/Proceedings/y2012/Files/304652\\_73493.pdf](http://ww2.amstat.org/sections/srms/Proceedings/y2012/Files/304652_73493.pdf) (accessed August 2017).
- Särndal, C.-E., B. Swensson, and J. Wretman. 1992. *Model Assisted Survey Sampling*. Springer-Verlag. New York, NY.
- SAS/STAT(R) 9.22 User's Guide. Web. 06 Apr. 2016.
- Thompson, K.J. and R.S. Sigman. 1996. *Evaluation of Statistical Methods for Developing Ratio Edit Module Parameters*. Technical report #ESM-9610. Washington, DC: U.S. Bureau of the Census. Available upon request.
- Wang, Y.-G., X. Lin, M. Zhu, and Z. Bai. 2007. "Robust Estimation Using the Huber Function with a Data-Dependent Tuning Constant." *Journal of Computational and Graphical Statistics* 16(2): 468–481. Doi: <http://dx.doi.org/10.1198/106186007X180156>.
- White, H. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48(4): 817–838. Doi: <https://doi.org/10.2307/1912934>.

Received November 2016

Revised May 2017

Accepted October 2017