Setting the M-estimation Tuning Constant for Detection and Treatment of Influential Values

Mary H. Mulry, Stephen Kaputa, Katherine J. Thompson[1]

Abstract

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 paper, 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 φ, 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

1. Introduction

For the most part, business surveys publish totals and period-to-period change estimates. Some of the estimates are economic indicators while others may contribute to estimates of Gross Domestic Product. When the survey estimates are released monthly or quarterly, much of the data review is performed on a flow basis on the micro-data. In general, business populations are highly skewed, and the bulk of the analyst review therefore focuses on the larger units that contribute the most to the overall levels (Thompson and Oliver 2012). It is possible for a smaller business with a large sampling weight to report an unexpectedly large or small value for a collected item. 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 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.

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

[1] Mary H. Mulry (mary.h.mulry@census.gov), Stephen Kaputa (stephen.kaputa@census.gov), Katherine J. Thompson (katherine.j.thompson@census.gov), U.S. Census Bureau, Washington, DC 20233. 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.

from the imputation base. 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 determined using an objective criterion that optimizes a precision measure.

M-estimation produces an adjusted value that minimizes the design-based estimator of the mean squared error (MSE). The procedure itself is extremely flexible, allowing for detection of large, small, or both large and small influential values, accommodating variations in the prediction models, and providing control over the outlier-detection region via numerous parameters (each with a specific function). Beaumont and Alavi (2004) and Beaumont (2004) developed M-estimation procedures for complex survey data and demonstrated the effectiveness of these methods using simulated data.

The flexibility of M-estimation and the option for two influential value "treatments" provides survey analysts with repeatable solutions. Mulry, Oliver, and Kaputa (2012, 2013, 2014) tested several variants of M-estimation on simulated data modeled from two industries in the Monthly Retail Trade Survey (MRTS) conducted by the U.S. Census. The results over repeated samples were excellent in terms of reduced MSE after extensive fine-tuning of all input parameters.

The previous studies highlighted the importance of setting the appropriate parameters for the M-estimation algorithm. This paper proposes methods for setting parameters for the M-estimation algorithm, using a variety of common statistical data analysis tools. The methods have the advantage that they can be implemented on a large scale. 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 M-estimation method (Beaumont and Alavi 2004) in our application follows Mulry, Oliver, and Kaputa (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 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, Oliver, and Kaputa (2012, 2014).

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 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).

A key assumption of the M-estimation approach is that y_i given x_i is distributed under the prediction model m with $E_m\left[y_i \mid x_i\right] = x_i'\beta$ and $V_m\left[y_i \mid x_i\right] = v_i\sigma^2$. 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. With retail trade, the regression of current month's sales on the previous month's sales tends to go through the origin (Huang 1986, 1984).

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

$$v_i = \lambda x_i$$

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

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

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

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 prior 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 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 describe their one- and two-sided-forms. The one-sided Type I Huber function is

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

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

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

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 ψ ensures that all adjusted units are fully represented in the estimate. The one-sided Type II Huber function is

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

where φ is a positive tuning constant. 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.

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 M-estimation algorithm takes into account both the size of an observation's weight as well as 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 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.

With M-estimation, the user has a choice of adjusting the weight of the influential value or adjusting its value. The weight adjustment for the Type II Huber function above has the appealing feature of always being greater than one and is given by

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

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

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

Using numerical analysis, 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 (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. The minimization does not require a generating data model that holds for all observations (including the influential value).

3. Setting algorithm parameters

The M-estimation algorithm discussed in Section 2 requires settings for Q, h_i, v_i, the function ψ, and an initial value of the tuning constant φ. 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 developed by Beaumont (2007) for implementing the method. The following sections explore the potential impact of different settings for the other v_i, ψ, and φ. The importance of the selections of v_i and φ on the effectiveness of the algorithm is discussed in detail in Mulry, Oliver, and Kaputa (2014). In this paper, we focus on setting the most important parameter, the initial φ.

Table 1. M-estimation algorithm parameters

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

For our investigation of a data-driven method for setting the initial φ, 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 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. Micro-review procedures are primarily ratio edits, followed by the Hidiroglou Berthelot (1986) edit used to identify within-imputation-cell outliers and determine their treatment, as well as to create the imputation base (Hunt, Johnson, and King 1999) when sufficient returns in an industry have been processed. The MWTS publishes *industry level* tabulations. Influential values are considered at the industry level rather than 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 to complement, not replace, the HB edit. For more details on the MWTS estimation and review procedures, see http://www.census.gov/wholesale/www/how_surveys_are_collected/monthly_methodology.html

3.1 Tuning constant φ

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 φ by minimizing the design-based estimator of the MSE. An observation's weighted residual has to exceed the initial φ for the algorithm to consider it as a possible influential value. Viewing the MSE as a function of φ implies that a value of φ corresponds to a "treated" value for each observation with a weighted residual greater than φ. When a minimum exists, the initial value of φ has to be "close enough" to the minimum for the algorithm to find it. If the initial φ 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 φ 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 φ.

Figure 1 uses sample data for one month selected from a simulated MRTS industry to illustrate the MSE as a function of φ in a sample without an induced influential value. Figure 1 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 φ equal to 5.97×10^{13} (although the MSE may appear equal to zero on the scale used in Figure 1). As the amount added to the unweighted observation to induce an influential value increases from 2 million to 8 million, the curve shifts as does the value of φ where the minimum MSE occurs. Notice that with an initial φ 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 φ must be approximately 100 million or less.

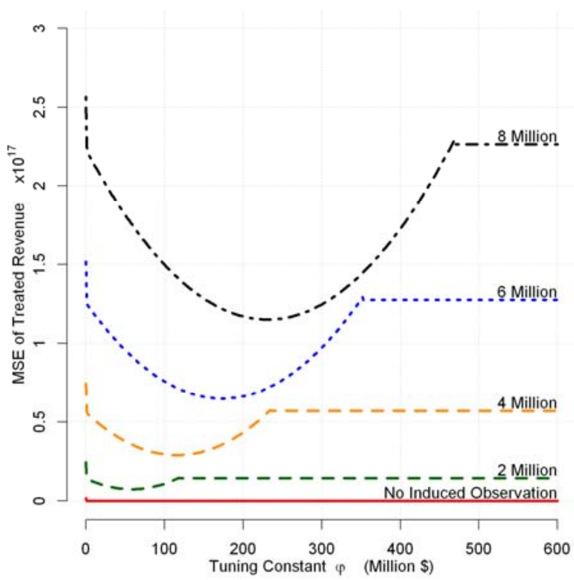

Figure 1. MSE as a function of the tuning constant φ when no influential value is present (MSE has a constant value of 5.97×10^{13}) and when an influential value is induced by adding four different amounts to an unweighted observed value with weight 60. Data is from a sample selected for one month from a simulated MRTS industry.

Figure 2 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 φ 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 φ.

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 for 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, the smaller businesses at the time of frame determination will likely have large 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, selected small businesses may become larger – in some cases reporting the same level of total receipts as the certainty large businesses. 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 2 overlays the boundaries of the detection regions obtained with M-estimation with a low initial φ and M-estimation with a high initial φ. 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 dash 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 2 shows that the detection region obtained using the M-estimation-high φ is much more restrictive and does reduce bias. The close proximity to the regression line M-estimation-low φ reflects the trimming that both methods do to minimize the MSE through 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 φ.

For small sampled business, large changes between the current and prior values for the same unit are not atypical. For this reason, it is crucial to set the initial φ to be the weighted distance between an observed and *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 φ not be too high or too low, but, as Goldilocks says, just right.

When only one influential value is present in simulated samples, the effectiveness of the M-estimation algorithm is sensitive to the choice of the initial tuning constant φ since the initial φ determines the lower boundary of the detection region. However, in simulations of samples with two high influential values, we found that when we fixed one influential value and let the second one vary, the detection region for the second one did not appear sensitive to the initial φ (Mulry, Oliver, and Kaputa 2012).

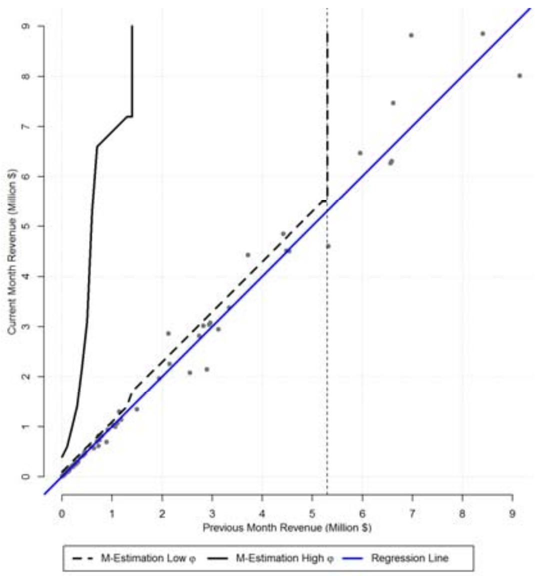

Figure 2. Lower Boundaries of Detection Regions for high and low initial φ. Data is from a sample selected for one month from a simulated MRTS industry.

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

On the other hand, setting the initial φ 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 φ 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 ψ have convergence problems (Mulry, Oliver, and Kaputa 2014).

The decision-making process for determining an initial value of φ is not difficult when the results are known – or when the number of studied subpopulations is fairly small. However, the MWTS application comprises 19 industries, each requiring its own set of parameters. To implement the algorithm in practice, we needed to find an automated data-driven method of setting these initial values.

Our first concern was seasonality. The MWTS series is seasonal, so we were uncertain whether this seasonality would extend to the M-estimation weighted residuals--which would require additional consideration in the parameter settings. However, only one of the 19 studied industries exhibited any consistent seasonal patterns in their residuals, as detected by the QS statistic test for seasonality at lags 12 and 24 after appropriate differencing at α=0.05 (Monsell and Blakely 2013). Of course, because one would expect to reject at least one test, we concluded that there is no convincing evidence of seasonality in the weighted MWTS residuals (Mulry, Kaputa, Thompson 2016).

3.2 Automated Data-Driven Methods of Obtaining Initial φ

The first attempts of our quest to find a general method for setting the initial φ 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 Section 3.1, all the weighted regression models of current month's value on previous month's value did not have an intercept and used the regression weight equal to the sample design weight divided by the prior 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 (see Arvanitis 2015). 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 (Mulry, Kaputa, Thompson 2016). 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 prior) 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 90% 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 φ:

1) (CV_EST) Set the initial φ as the product of the coefficient of variation of the estimated total (*cv*), the width of the 90% 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: Initial $\varphi = cv*1.7*\hat{T}_{t-1}$.

2) (CV_PRED) Use the same formula as in 1), but replace $\hat{T}_{t-1}$ by 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_{ti} on the previous month x_{ti}:
Initial $\varphi = cv*1.7*\sum \hat{\beta}_t x_{ti}$

3) (ST_EST) Set the initial φ as the width of the 90% 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 $\hat{T}_t$: Initial $\varphi = 1.65*StdErr(\hat{\hat{T}}_{t-1})$,
where $StdErr(\hat{T}_{t-1})$ is computed with the Taylor linearization estimator implemented in PROC SURVEYMEANS on the prior 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_{ti} on the previous month x_{ti}:
Initial $\varphi = 1.65*StdErr(\sum \hat{\beta}_t x_{ti})$,
where standard errors are computed analogously to ST_EST, but sample-weighted predicted current period values from the M-estimation regression replace the prior-period MWTS values in the computations.

4. Results

Our first concern was algorithm convergence issues with the four considered applications, specifically converging to zero (and flagging all observations) or failing to converge. Such problems appear in three of the 2,736 applications (4 options for setting the initial φ 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, one application of the CV_EST option in Industry 13 and 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 total number of values flagged in all 36 consecutive months when the algorithm converged by option for setting the initial φ 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 φ 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.
  - 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 1 month in Industry 16 when CV_EST does not.

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 φ for 19 MWTS industries

	CV_EST		*CV_PRED*		*SE_EST*		*SE_PRED*	
Industry	*Max*	*Sum*	*Max*	*Sum*	*Max*	*Sum*	*Max*	*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

Table 3 presents the mean and standard deviation of the settings of the initial φ 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 our indicators publish some form of average variance (or *cv*) to smooth away some of the noise.

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

	CV_EST		*CV_PRED*		*SE_EST*		*SE_PRED*	
Industry	*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

For further insight, Figure 3 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 φ from the four options by month for Industry 2. We see that the settings of the initial φ from the CV_EST and SE_EST are lower than the settings from CV_PRED and CV_PRED, but are still low enough that they are smaller than the maximum weighted observed residual, triggering the M-estimation algorithm. However, the values of the initial φ 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.

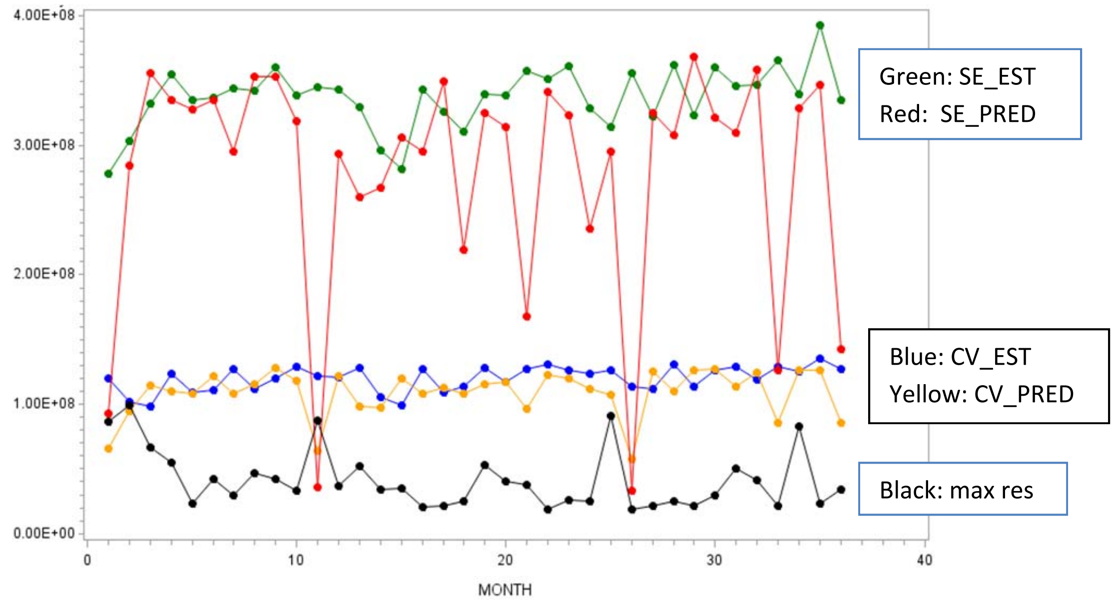

Figure 3. Plot of maximum residual (black) and values of initial φ from four methods for 36 consecutive months of MWTS Industry 2.

5. Summary

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 ψ 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 paper, 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 φ 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 φ 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). For more details, see Mulry, Kaputa, Thompson (2016).

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.

Acknowledgements

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.

References

Arvanitis, M. (2015). "Investigation of Alternative Distributions for M-Estimation Residuals." Unpublished draft manuscript. ESMPD. U.S. Census Bureau. Washington, DC.

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*. 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. Last accessed September 29, 2016 at https://fcsm.sites.usa.gov/files/2014/05/2001FCSM_Black.pdf

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

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. Last accessed September 29, 2016 at https://www.census.gov/srd/papers/pdf/rr86-09.pdf.

Huang, E. T. (1984). An Imputation Study for the Monthly Retail Trade Survey. *JSM Proceedings*, Survey Research Methods Section. American Statistical Association. Alexandria, VA. 610-615. Last accessed September 29, 2016 at http://ww2.amstat.org/sections/srms/Proceedings/papers/1984_117.pdf.

Huber, P. J. (1964). “Robust Estimation of a location parameter.” *Annals of Mathematical Statistics, 35*. Institute of Mathematical Statistics. Shaker Heights, OH. 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. Last accessed September 29, 2016 at http://ww2.amstat.org/sections/srms/Proceedings/papers/1999_093.pdf.

Monsell, B. and Blakely, C. (2013). “X-13ARIMA-SEATS and iMetrica.” *Proceedings of the 59th World Statistics Congress of the International Statistical Institute*, Hong Kong. The Hague, The Netherlands. Last accessed March 1, 2016 at http://www.census.gov/ts/papers/isipaperbcmcb.pdf.

Mulry, M.H., Kaputa, S., and Thompson, K. J. (2016). “Auxiliary Research on Influential Values in Economic Surveys.” draft research report dated March 14, 2016. CSRM. U.S. Census Bureau. Washington, DC.

Mulry, M. H., Oliver, B., and Kaputa, S. (2014). “Detecting and Treating Verified Influential Values in a Monthly Retail Trade Survey.” *Journal of Official Statistics, 30,* 4. Statistics Sweden. Stockholm, Sweden. 721–747.

Mulry, M. H., Oliver, B., Kaputa, S., and Thompson, K. J. (2013). “Setting M-estimation Parameters for Detection and Treatment of Influential Values.” *2013 JSM Proceedings,* Survey Research Methods Section. American Statistical Association. Alexandria, VA. 1424-1438. Last accessed September 29, 2016 at http://ww2.amstat.org/sections/srms/Proceedings/y2013/Files/308309_80627.pdf

Mulry, M. H., Oliver, B., and Kaputa, S. (2012). “Several Scenarios for Influential Values in Business Surveys and Methods for Their Treatment.” *2012 JSM Proceedings,* Survey Research Methods Section. American Statistical Association. Alexandria, VA. 4015-4029. Last accessed September 29, 2016 at http://ww2.amstat.org/sections/srms/Proceedings/y2012/Files/304652_73493.pdf.

Sarndal, C.E., Swensson, B., and Wretman, J. (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 Oliver, B. (2012). “Response Rates in Business Surveys: Going Beyond the Usual Performance Measure.” *Journal of Official Statistics, 28, 2*. Statistics Sweden. Stockholm, Sweden. 221–237.

Wang, Y.-G., Lin, X., Zhu, M. and Bai, Z. (2007). “Robust Estimation Using the Huber Function with a Data-Dependent Tuning Constant.” *Journal of Computational and Graphical Statistics, 16,2*. 468 – 481.