

A Comparison of two Robust Estimation Methods for Business Surveys

Robert Graham Clark¹, Philip Kokic^{1,2} and Paul A. Smith³

¹ *National Institute for Applied Statistics Research, University of Wollongong, NSW 2522, Australia E-mail rclark@uow.edu.au*

² *CSIRO Digital Productivity Flagship, GPO Box 664, Canberra ACT 2601, Australia E-mail philip.kokic@csiro.au*

³ *Southampton Statistical Sciences Research Institute (S3RI), University of Southampton, SO17 1BJ, United Kingdom E-mail p.a.smith@soton.ac.uk*

Summary

Two alternative robust estimation methods often employed by National Statistical Institutes in business surveys are two-sided M-estimation and one-sided Winsorisation, which can be regarded as an approximate implementation of one-sided M-estimation. We review these methods and evaluate their performance in a simulation of a repeated rotating business survey based on data from the Retail Sales Inquiry conducted by the UK Office for National Statistics. One- and two-sided M -estimation are found to have very similar performance, with a slight edge for the former for positive variables. Both methods considerably improve both level and movement estimators. Approaches for setting tuning parameters are evaluated for both methods and this is a more important issue than the difference between the two approaches. M-estimation works best when tuning parameters are estimated using historical data but is serviceable even when only live data is available. Confidence interval coverage is much improved by the use of a bootstrap percentile confidence interval.

Key words: bootstrap; mean squared error; M-estimation; movement estimation; influential values; outliers; robustness; sample survey; Winsorisation; Winsorization

1 Introduction

Outliers in surveys arise from a number of sources, and can have a large effect on estimates of totals. This is especially true in business surveys where the populations from which samples are drawn are typically skewed and heavy-tailed. We will consider here representative outliers (Chambers 1986), that is extreme observations which have been confirmed by the respondent, and consequently under standard random sampling assumptions represent some similar observations in the non-sampled part of the population. Mulry et al. (2014) call these “verified influential values” and Martinoz et al. (2015) note that the aim is to infer about the total population including both outliers and “inliers”. Throughout, we refer to these influential values as outliers. Outliers are sometimes defined as values so far beyond other values in the sample that they must be given special treatment, otherwise survey estimates would be grossly inaccurate (Chambers et al. 2004). In addition, we confine attention to continuous survey variables and outliers in the response-space only. Outliers in the benchmark-space are sometimes treated by size stratification or may be the cause of response-space outliers (Hedlin et al. 2001), but otherwise their treatment is beyond the scope of the current paper.

In many National Statistical Institutes (NSIs), outliers are typically treated by post-stratification, in which they are moved into a “surprise stratum”, which is assumed to be completely enumerated (CEd). For example, a technique related to the surprise stratum approach has been considered by Statistics Norway, and is used in the majority of Statistics Canada and Statistics Netherlands business surveys as well as by Statistics Sweden in their business surveys (pers.comms). The choice of these outliers may be subjective, or this approach may be supplemented by some test which determines which observations should be treated by moving them to the surprise post-stratum; for example, Statistics Canada uses the detection method of Hidiroglou and Berthelot

(1986). In either case, determining the correct weights to use in the post-strata can be difficult, especially with a small number of outliers.

More transparent approaches have been proposed, where estimators are constructed to be robust to outliers, often involving tuning parameters optimised to achieve an appropriate trade-off between variance and bias. One such approach is M-estimation, where the influence of extreme values is curtailed by applying a bounded function to residuals in the estimating equations for a model coefficient. This paper focuses on two adaptations of M-estimation which are currently used by several NSIs. The first approach, one-sided M-estimation (or an approximate non-iterative version, one-sided Winsorisation), is robust only to large positive outliers, while the second (two-sided M-estimation) is robust to both positive and negative outliers. The first approach has the advantage that a theoretical result is available relating the optimal tuning parameters to the survey weights (Kokic and Bell 1994; Clark 1995). The second option deals with outliers in both directions, but lacks a simplifying theoretical rule and instead imposes a relationship between the tuning parameters and the survey weights (Beaumont and Alavi 2004). Both approaches minimise an estimator of the mean squared error (MSE) of the resulting estimator of total. Versions of MSE-optimal one-sided Winsorisation are used by the UK Office for National Statistics (Lewis 2007) and the Australian Bureau of Statistics (Preston and Mackin 2002). The two-sided alternative has been favourably evaluated for the US Bureau of the Census (Mulry et al. 2014).

Alternative robust estimation methods exist. The log-normal method developed by Karlberg (2000) has low variance when an assumed lognormal model is correct, but is sensitive to this assumption (see Beaumont and Rivest 2009 who cite Myers and Pepin 1990). Basak et al. (2014) propose and evaluate a bias-corrected version of Karlberg's (2000) estimator based on the log-normal distribution. Ståhl (2015) reviews estimators

for right-skewed data which make use of parametric models for the tail distribution. Beaumont et al. (2013) develop an alternative approach based on the conditional influence function. Martinoz et al. (2015) apply the idea of limiting the maximum conditional bias to Winsorisation, and evaluate its use for domain estimation. Gwet and Rivest (1992), Hulliger (1995) and Duchesne (1999) propose estimators related to the M-estimation approach, replacing non-robust regression coefficient estimators by robust alternatives (Beaumont and Rivest 2009). We do not consider these approaches, instead concentrating on MSE-optimal M-estimators, as these are the main methods other than the surprise outlier method used by NSIs.

Both one-sided and two-sided M-estimation are related to Winsorised estimation (Chambers and Kokic 1993). Winsorised estimators truncate the values of the variable of interest at upper and/or lower cutoffs. In M-estimation, the cutoffs depend on regression coefficients estimated using the truncated values, and so must be determined iteratively. In practice, this iteration has usually been avoided in one-sided Winsorisation by separately estimating regression coefficients robustly. It should be noted that both of the M-estimation approaches considered here are only applicable to continuous variables. For NSIs this requirement essentially confines the methods to business surveys.

The novelty and contribution of this paper is that it presents a succinct proof of the well-known result from Clark's unpublished thesis (Clark 1995), which relates the MSE-optimal tuning parameters in one-sided M-estimation to the survey weights. We compare the one- and two-sided methods in a realistic simulation setting involving sample rotation of a repeated survey based on real business survey data. A previous simulation (Mulry et al. 2014) compares the two approaches in a one-off sample which includes just one outlier in the population contributing only 1% to total estimates even when selected, whereas many business surveys are more outlier-prone than this. Here, we

simulate a greater number and contribution of outliers reflecting typical business survey data, and partially rotate the sample each month reflecting common practice. We consider the effect of different methods of selecting the tuning parameter in both methods and also examine the performance of bootstrap confidence intervals (CIs) compared to those based on variance formulae. We claim that this is a useful study for NSIs, as it is not obvious a-priori whether one- or two-sided M-estimation performs better when outliers are present, nor the influence of the tuning parameters on the estimators of level and movement.

In the next section we review M-estimation methods for surveys where tuning parameters are chosen to minimise the MSE. We present the key theoretical result for one-sided M-estimation mentioned above, and describe the two-sided approach. Section 3 discusses and contrasts the two options. Section 4 describes a simulation study based on data from the monthly Retail Sales Inquiry (RSI) conducted by the United Kingdom Office for National Statistics (ONS). One- and two-sided M-estimation and one-sided Winsorised estimators are compared, with different settings for both methods. Section 5 summarises the results of the simulation, and Section 6 is a discussion.

2 Review of MSE-Optimal M-Estimators in Survey Estimation

2.1 M-Estimation applied to Survey Estimation

Consider a finite population U of N units in which we are interested in estimating the total t_y of a continuous variable y . This variable is observed over a sample $s \subseteq U$ of units selected from the finite population. The set of non-sampled units is denoted $r = U - s$. Auxiliary variables \mathbf{x}_i may also be available for $i \in U$ to support estimation. In this case

it is helpful to define a working model to motivate the estimators:

$$\left. \begin{aligned} y_i \text{ are independent continuous random variables with densities } f_i, \\ E(y_i) = \mu_i < \infty \\ \text{var}(y_i) = \sigma^2 v_i < \infty \end{aligned} \right\} \quad (1)$$

where v_i are known constants and $\mu_i = \mu_i(\mathbf{x}_i)$ is typically a linear function

$$\mu_i = \boldsymbol{\beta}^T \mathbf{x}_i \quad (2)$$

with unknown coefficients $\boldsymbol{\beta}$. Models such as (1) and (2) can be used in a model-assisted framework, where they motivate the choice of an approximately design-unbiased estimator of a total or other quantity, such as the generalized regression estimator (e.g. Särndal et al. 1992). Alternatively the model-based approach can be adopted. Best linear unbiased predictors of population totals can be derived, which are unbiased over repeated realisations of population values from the model, but are not necessarily design-unbiased (e.g. Chambers and Clark 2012).

As noted by Beaumont and Alavi (2004), under the model defined by (1) and (2), the total t_y may be estimated by

$$\hat{t}_y = \hat{\boldsymbol{\beta}}^T \mathbf{t}_x \quad (3)$$

(sometimes called the projective form) or

$$\hat{t}_y = t_{ys} + \hat{\boldsymbol{\beta}}^T \mathbf{t}_{xr} \quad (4)$$

(the predictive form) where $\mathbf{t}_x = \sum_{i \in U} \mathbf{x}_i$ is the population total of the auxiliary variables, $\mathbf{t}_{xr} = \sum_{i \in U-s} \mathbf{x}_i$ is the non-sample total, $t_{ys} = \sum_{i \in s} y_i$, and

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i \in s} c_i \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \sum_{i \in s} c_i \mathbf{x}_i y_i$$

where c_i are regression weights (some possibilities for c_i will be discussed shortly).

Equivalently, $\hat{\boldsymbol{\beta}}$ is the solution of the following estimating equation:

$$\sum_{i \in s} c_i \left(y_i - \hat{\boldsymbol{\beta}}^T \mathbf{x}_i \right) \mathbf{x}_i = \mathbf{0}. \quad (5)$$

The estimator \hat{t}_y is also equal to

$$\hat{t}_y = \sum_{i \in s} w_i y_i \quad (6)$$

where the weights w_i are equal to

$$w_i = \mathbf{t}_x^T \left(\sum_{j \in s} c_j \mathbf{x}_j \mathbf{x}_j^T \right)^{-1} c_i \mathbf{x}_i \quad (7)$$

when (3) is used, and

$$w_i = 1 + \mathbf{t}_{xr}^T \left(\sum_{j \in s} c_j \mathbf{x}_j \mathbf{x}_j^T \right)^{-1} c_i \mathbf{x}_i \quad (8)$$

when (4) is used. The model-assisted generalized regression (GREG) estimator is a special case of (3) and (7), with $c_i = v_i^{-1} \pi_i^{-1}$, provided that

$$v_i = \boldsymbol{\lambda}^T \mathbf{x}_i \quad (9)$$

for some $\boldsymbol{\lambda}$ (Särndal et al. 1992, section 6.5). Condition (9) is satisfied by many but not all of the models used in practice. For example, in the gamma population model (Chambers and Clark 2012, p49), it is satisfied when γ is 0 or 1, but not otherwise. The BLUP is the special case of (4) when $c_i = v_i^{-1}$, and also of (3) when (9) holds. Beaumont and Alavi (2004) propose a compromise with $c_i = v_i^{-1} \pi_i^{-\alpha} \left(\sum_{j \in s} \pi_j^{-1} \right) / \left(\sum_{j \in s} v_j^{-1} \pi_j^{-\alpha} \right)$, where $0 \leq \alpha \leq 1$.

Estimators of the form (4) and (3), or equivalently (6), are not outlier-robust and the influence of each particular value y_i is unbounded. Robust versions are given by replacing $\left(y_i - \hat{\boldsymbol{\beta}}^T \mathbf{x}_i \right)$ in (5) with bounded functions of these residuals, to give a robust estimator $\hat{\boldsymbol{\beta}}^*$:

$$\sum_{i \in s} c_i \psi_i \left(y_i - \hat{\boldsymbol{\beta}}^{*T} \mathbf{x}_i \right) \mathbf{x}_i = \mathbf{0}. \quad (10)$$

where $\psi_i(u)$ is a suitable non-decreasing but bounded function. Typically,

$\psi_i = Qa_i\psi(Q^{-1}a_i^{-1}u)$, where ψ is a bounded non-decreasing function approximating the identity function near the origin. The value of $Q > 0$ controls the sensitivity to outliers. If Q is large, then ψ_i will closely approximate the identity function, and $\hat{\beta}^* \approx \hat{\beta}$. As Q becomes closer to 0, $\hat{\beta}^*$ becomes increasingly robust to outliers. The values of a_i control differential sensitivity to outliers for different units.

In this paper, we consider two alternatives for the function ψ which have been proposed. Let $[u]_A^B$ denote the truncation of u to the interval $[A, B]$. The two functions considered are the two-sided Huber function, $\psi(u) = [u]_{-1}^1$, and the one-sided Huber function, $\psi(u) = \min(u, 1) = [u]_{-\infty}^1$. The former results in robustness to large positive and large negative outliers, while the latter only deals with large positive outliers. For both choices,

$$\hat{\beta}^* = \left(\sum_{i \in s} c_i \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \sum_{i \in s} c_i \mathbf{x}_i y_i^* \quad (11)$$

where

$$y_i^* = [y_i]_{\hat{\beta}^{*T} \mathbf{x}_i - Qa_i}^{\hat{\beta}^{*T} \mathbf{x}_i + Qa_i} \quad (12)$$

when ψ is given by the two-sided Huber function, and

$$y_i^* = \min \left(y_i, \hat{\beta}^{*T} \mathbf{x}_i - Qa_i \right) \quad (13)$$

when ψ is given by the one-sided Huber function. It follows that the robust estimator of t_y is equal to

$$\hat{t}_y^* = \sum_{i \in s} w_i y_i^* \quad (14)$$

in both cases.

For infinite populations, values of $c_i = v_i^{-1}$, $a_i = \sqrt{v_i}$ and $Q = k\sigma^2$ would be typical choices, where k is around 2 or 3. In sample surveys, it is less clear how these values should be chosen because:

- Extreme values y_i affect \hat{t}_y more if their weight w_i is large. This suggests that greater robustness to outliers is appropriate for these units, which is achieved by making a_i smaller when w_i is larger.
- The aim is to estimate the total t_y over all units, including outliers and non-outliers. It is assumed that extreme values due to measurement error are corrected prior to estimation, so that the outliers that remain are “representative outliers” (Chambers 1986). As a result, outlier robustness will generally reduce the variance but incur bias. Tuning parameters should therefore minimise the mean-squared error of \hat{t}_y^* . This is particularly important in large scale surveys where variances may be fairly small, so that biases due to outlier robustness could lead to the mean-squared error being made worse.

2.2 *M-Estimation using the One-Sided Huber Function*

We now consider the estimator (14) using the one-sided Huber function, so that y_i^* are given by (13). A theorem by Clark (1995) on Winsorised estimators sheds light on how the tuning parameters a_i and Q should be chosen.

Searls (1966) compares the performance of a number of outlier-robust estimators of total, and suggested replacing y_i by $y_i^* = \min(y_i, K)$ where K is a cutoff, with values of y_i above K deemed to be outliers. This method is described as type 1 Winsorisation and is unattractive because it is possible for the Winsorised estimate of the population total of a survey variable to be less than the sum of the sample values. Gross et al. (1986) instead recommended the type 2 Winsorised estimator:

$$\hat{t}_y^* = t_{ys} + \sum_{i \in s} (w_i - 1) y_i^*. \quad (15)$$

where $y_i^* = \min(y_i, K)$. Each observation (including outliers) is allowed to represent itself

in the first term of the right hand side of (15), but is truncated in the second term when weighted up by $(w_i - 1)$ to represent non-sampled units.

Clark (1995) considered the one-sided type 2 Winsorised estimator (15) where $y_i^* = \min(y_i, K_i)$ and K_i may be different for every unit i in sample. This estimator has n tuning parameters K_i which is excessive. Fortunately this can be reduced to just one due to the following result which extends a theorem in Kokic and Bell (1994). The theorem uses a model-based framework, but it seems reasonable to apply it even for model-assisted approaches such as generalized regression estimators, because the model assumptions are relatively weak and do not depend on the distributions f_i in (1).

Theorem 1. Suppose that a population is generated by model (1) and (2), and that $\hat{t}_y = \sum_s w_i y_i$ is model-unbiased with $E_M [\hat{t}_y - t_y] = 0$. Let $\hat{t}_y = t_{ys} + \sum_{i \in s} (w_i - 1) y_i^*$ where $y_i^* = \min(y_i, K_i)$. Then $E_M [(\hat{t}_y^* - t_y)^2]$ is minimised with respect to $\{K_i : i \in s\}$ when:

$$K_i = \mu_i^* + Q (w_i - 1)^{-1} \quad (16)$$

where

$$\mu_i^* = E [\min(y_i, K_i)]$$

and Q is the solution to

$$Q = -B(Q) = \sum_{i \in s} w_i (\mu_i - \mu_i^*)$$

■

As the theorem has only been published in thesis form (Clark 1995), we include the proof in Appendix 1. Remarkably, Theorem 1 does not require any restriction on the densities f_i in (1), which may be different for every i , and they do not need to be modelled in order to apply the theorem.

The values of μ_i^* are not known, but reasonable values of μ_i^* can be constructed by assuming $\mu_i^* = \boldsymbol{\beta}^{*T} \mathbf{x}_i$, where $\hat{\boldsymbol{\beta}}^*$ is obtained by regressing $y_i^* = \min \left(y_i, \hat{\boldsymbol{\beta}}^{*T} \mathbf{x}_i + Q (w_i - 1)^{-1} \right)$ on \mathbf{x}_i . Since y_i^* itself depends on $\hat{\boldsymbol{\beta}}^*$, the result must be calculated iteratively. It is easy to see that $\hat{\boldsymbol{\beta}}^*$ is the M-estimator (10) where ψ is the one-sided Huber function and $a_i = (w_i - 1)^{-1}$. The estimation of optimal values of Q can be performed using the `surveyoutliers` package (Clark 2015) for the R statistical environment (R Core Team 2014).

Fast Approximate Solution

In the literature and applications of Winsorising (e.g. Kokic and Bell 1994; Preston and Mackin 2002), calculation of the cutoffs K_i is simplified by replacing μ_i^* in (16) by $\tilde{\mu}_i$, where $\tilde{\mu}_i$ can be defined to be a robust estimator not involving Q , or set to 0. The values of $Q (w_i - 1)^{-1}$ would normally dominate the values of μ_i^* , so using an approximation of μ_i^* probably has little impact. We then obtain the easier to solve system:

$$K_i = \tilde{\mu}_i + Q (w_i - 1)^{-1} \quad (17)$$

$$Q = \sum_{i \in s} (w_i - 1) \left\{ \mu_i - E \left[\min \left(y_i, \tilde{\mu}_i + Q (w_i - 1)^{-1} \right) \right] \right\} \quad (18)$$

To obtain optimal cutoffs in practice, (18) is solved using an estimator $\hat{B}(Q)$ based on sample data either from the current survey or one or more historical survey datasets:

$$\begin{aligned} \hat{B}(Q) &= \sum_{i \in s} (w_i - 1) \left\{ \min \left(y_i, \tilde{\mu}_i + Q (w_i - 1)^{-1} \right) \right\} \\ &= \sum_s \left\{ \min \left(\hat{D}_i, Q \right) - \hat{D}_i \right\} \end{aligned} \quad (19)$$

where $\hat{D}_i = (y_i - \tilde{\mu}_i) (w_i - 1)$. The aim is to solve $\hat{F}(Q) = 0$ where

$$\hat{F}(Q) = Q + \hat{B}(Q) = \sum_s \left\{ \min \left(\hat{D}_i, Q \right) - \hat{D}_i \right\} + Q. \quad (20)$$

It is clear that $\hat{F}(Q)$ is piecewise linear and decreasing with joints at the values of \hat{D}_i .

Hence the equation can easily be solved by evaluating it at $\hat{D}_{(1)}, \hat{D}_{(2)}, \dots$ (the ordered

\hat{D}_i s from largest to smallest) until the sign of $\hat{F}(Q)$ changes from positive to negative, then using linear interpolation to obtain the solution (Kokic and Bell 1994; Clark 1995). Typically, only a small number of evaluations will be needed, because the number of values of $\hat{F}(\hat{D}_{(k)})$ which are positive corresponds to the number of observations above the cutoff.

If the historical dataset differs much in size or design from the data at hand, then (19) should be modified to:

$$\hat{B}(Q) = \sum_{i \in s_{old}} u_i \left\{ \min(\hat{D}_i, Q) - \hat{D}_i \right\} \quad (21)$$

where u_i are weights scaling from s_{old} to s . Setting $u_i = (\pi_{i(old)}^{-1} - 1) / (\pi_i^{-1} - 1)$ would be a reasonable choice and is used in the simulation in the next section.

2.3 M-Estimation using the Two-Sided Huber Function

Beaumont and Alavi (2004) propose the robust estimator (14) using the two-sided Huber function. Several suggestions for c_i are given; we restrict ourselves here to $c_i = v_i^{-1} \pi_i^{-\alpha} \left(\sum_{j \in s} \pi_j^{-1} \right) / \left(\sum_{j \in s} v_j^{-1} \pi_j^{-\alpha} \right)$ which is used in their simulation study. Choosing a value of α between 0 and 1 gives a compromise between the BLUP's greater efficiency under the model and the GREG's robustness when the model's assumptions are not justified.

The tuning parameters a_i are defined by Beaumont and Alavi (2004) to equal c_i^{-1} . Both c_i and w_i can be regarded as survey weights, so this rule for a_i is somewhat similar to the values $a_i = (w_i - 1)^{-1}$ from Theorem 1 in the previous section. Beaumont and Alavi (2004) note that the M-estimator of β can be expressed in terms of a weight modification, since (10) becomes

$$\mathbf{0} = \sum_{i \in s} c_i \psi_i \left(y_i - \hat{\beta}^{*T} \mathbf{x}_i \right) \mathbf{x}_i$$

$$\begin{aligned}
\mathbf{0} &= \sum_{i \in s} c_i \pi_i^{-\alpha} \left[Q^{-1} v_i^{-1} \pi_i^{-\alpha} \left(y_i - \hat{\boldsymbol{\beta}}^{*T} \mathbf{x}_i \right) \right]_{-1}^1 v_i \pi_i^\alpha Q \mathbf{x}_i \\
\mathbf{0} &= \sum_{i \in s} c_i \left[y_i - \hat{\boldsymbol{\beta}}^{*T} \mathbf{x}_i \right]_{-Q v_i \pi_i^\alpha}^{Q v_i \pi_i^\alpha} \mathbf{x}_i \\
\mathbf{0} &= \sum_{i \in s} c_i^* \left(y_i - \hat{\boldsymbol{\beta}}^{*T} \mathbf{x}_i \right) \mathbf{x}_i
\end{aligned} \tag{22}$$

where

$$c_i^* = c_i \left[y_i - \hat{\boldsymbol{\beta}}^{*T} \mathbf{x}_i \right]_{-Q v_i \pi_i^\alpha}^{Q v_i \pi_i^\alpha} / \left(y_i - \hat{\boldsymbol{\beta}}^{*T} \mathbf{x}_i \right). \tag{23}$$

It is clear that $c_i^* = c_i$ when $\left| y_i - \hat{\boldsymbol{\beta}}^{*T} \mathbf{x}_i \right| \leq Q v_i \pi_i^\alpha$, otherwise $c_i^* < c_i$. It is occasionally possible that c_i^* can be less than 1, which is counter-intuitive in a survey context since representative outliers should at least be allowed to represent themselves. To counter this, Beaumont and Alavi (2004) redefine c_i^* to be the maximum of 1 and (23). Since both c_i^* and (22) depend on $\hat{\boldsymbol{\beta}}^*$, iteration is required.

We slightly simplify the setup of Beaumont and Alavi (2004), in that they also allow the threshold of ψ_i to be an unknown parameter ϕ rather than 1, but this makes either ϕ or the tuning parameter Q redundant. We set $\phi = 1$ and leave Q free, as this is more in line with the mainstream robustness literature (e.g. Huber and Ronchetti 2009), whereas Beaumont and Alavi (2004) do the reverse. Also we assume $a_i = c_i^{-1}$ which is one of the preferred special cases of the more general form in Beaumont and Alavi (2004).

The tuning parameters α and then Q are chosen to minimise the estimated MSE of the M -regression estimator

$$\hat{t}_{y(M)} = \mathbf{t}_x^T \hat{\boldsymbol{\beta}}_M. \tag{24}$$

A number of MSE estimators are discussed by Beaumont and Alavi (2004). One of the simpler options, which is adopted in their simulation study, is

$$\widehat{MSE} [\hat{t}_{y(M)}] = \widehat{var} [\hat{t}_{y(M)}] + \{ \hat{t}_{y(M)} - \hat{t}_y \}^2 \tag{25}$$

where \hat{t}_y is a non-outlier-robust GREG estimator (see equation 5.3 of Beaumont and

Alavi 2004).

3 Remarks on the Two M-Estimation Approaches

3.1 Issues common to both Two Approaches

The value of Q should be updated only infrequently in a repeated survey, because then the bias introduced will remain approximately constant, and estimates of change will have smaller variance with approximately no bias (although measures of level will still be biased in order to achieve a lower variance).

There are two main issues in the use of M-estimation which stop it from being an ideal transparent outlier detection and treatment method in practice. The first is the level at which estimates are produced, and the second how to treat derived variables.

As an example of what effect the level has, in a survey of retail trade, we may determine a value \hat{Q} based on the whole survey. We may also produce several further values \hat{Q}_g for subsectors $g = 1, \dots, G$ of the retail sector. The results from using either the former value, or the set of latter values will be different. In fact, the first will give an optimum mean squared error (MSE) estimate of the total for the whole survey, but sub-optimal estimates for any lower level of aggregation, and the second will give optimal estimates of the totals for the subsectors, but suboptimal estimates of the survey total and any aggregates at a lower level than subsector.

The way this works is easily seen by considering the usual bias-variance trade-off. Processing outliers involves introducing a small amount of bias into the estimate in order to achieve a large gain in precision (drop in variance). At the whole survey level, the truncation in M-estimation affects relatively few outliers, and introduces a small bias for a good gain in precision, because these are the most extreme observations in the whole survey. If these same outlier weights are used in constructing the lower level subsector

totals, the small amount of bias is apportioned among them (usually unevenly), so that each gets a smaller introduction of bias, with a consequent smaller reduction in variance.

Conversely if M-estimation is applied at the subsector level, each subsector has some small bias introduced for a substantial reduction in variance. These biases are always negative for one-sided M-estimation. They may be positive or negative for the two-sided alternative, but they will almost always be negative in practice, because large positive residuals are much more common than large negative residuals in business surveys. As a result, when the subsector estimates are added together to give a whole survey estimate, they introduce too much bias for a minimum-MSE estimator at the whole survey level. This is implied in the one-sided case by the fact that for all g , $\hat{Q}_g \leq \hat{Q}$, which follows from (21).

In practice it is not possible to reconcile these measures, so a decision must be made as to which level to choose. An initial reaction is that the total level is the best, as it is the most important, and introduces the least bias overall. On the other hand, the estimated total may be more accurate than needed, whereas subsector estimates may be in greater need of improvement, suggesting calculating cutoffs at the subsector level. This is explored more fully in the simulation study. Martinoz et al. (2015) suggest separately estimating domain totals and the grand total, and then forcing them to agree using a method similar to calibration. Another option is to choose the values of \hat{Q}_g to be the simple or geometric mean of \hat{Q} optimised for overall estimates and \hat{Q}_g optimised for subsector estimates.

The second issue is how to construct a robust estimator for a derived variable, which is a linear function of other survey variables. For example, net capital expenditure is the difference of acquisitions and disposals of capital items. In a similar way to the level of M-estimation described above, estimating acquisitions and disposals separately leads to a

different estimate of net capital expenditure from that obtained when forming net capital expenditure first and then applying M-estimation. If the difference between these two estimates is not important, the former method is the ideal because it uses the distributions of the components separately. Typically, however, the estimates need to correspond in publications. Derived variables may also be non-linear combinations of other survey variables, but in this case perfect consistency between the various estimates is not essential, because the derivation process would not be obvious to readers of the survey publication.

The first method, separate M-estimators of the components of the derived variable, introduces a bias into the estimate for each component. If these biases are summed in the derived variable (that is if they are all or mostly in the same direction), the resultant estimator will have a large bias, and will not normally be useful. However, in our example, where disposals is negative in the derived variable, we may hope that the biases partly cancel out, and this procedure may be quite reasonable. Consequently, where some of the terms in the linear combination are negative, this approach should be considered.

The second method, M-estimation of the derived variable, gives an additional challenge, which is how to allocate the bias to the component variables so as to maintain the consistency of the derivation. There are several ways of doing this, which are discussed in Cruddas and Kokic (1996) and Preston and Watmuff (2005). However, to the authors' knowledge, these methods have not been applied in any NSI, presumably due to their complexity.

3.2 Differences between the Two Approaches

It is not clear which of the one-sided or two-sided approaches performs better in practice. Both make the truncation cutoff for the model residual approximately inversely proportional to the weight. The latter handles both positive and negative outliers. On

the other hand, negative outliers are rare in many business surveys, and approximately optimal values of a_i are available in the one-sided approach. In contrast, Q is optimised in the two-sided approach by using an MSE estimator (25) which is potentially biased and may also be volatile. Given these strengths and weaknesses of the two methods, their relative performance is an empirical question.

4 Simulation of MSE-Optimal M-Estimation in the UK RSI

The purpose of this section is to assess one and two-sided M-estimation approaches in a realistic simulation. We use data collected in the UK monthly Retail Sales Inquiry (RSI). Our aim is to mimic the sample rotation procedure used in ONS business surveys so as to assess the amount of information required to estimate the tuning parameters accurately. An examination is made of the level at which the cut-off parameters should be set, and their effects on the accuracy and precision of estimators both of total and of month on month change are determined.

4.1 Description of the RSI Data

The RSI is a monthly survey covering all retail businesses in the UK with at least 10 employees. The examples here are based on data from 1995-96, at which time the sample design could be summarised as follows (many of these features continue in the present design). The outcome variable is average weekly sales, calculated over a four or five week period depending on the month. The auxiliary variable used in estimation was register turnover, taken from the ONS inter-departmental business register (IDBR). It is an annual total derived from value added tax returns, closely correlated with sales. The survey was stratified into 27 industry strata and 6 size strata according to the value of register turnover. The top one or two size strata are CED and within stratum ratio estimation was used in the remaining strata. The industry strata are defined in terms of

the main activity of the business, see Central Statistical Office (1992). Simulations are based on the five months of RSI data from October 1995 until February 1996. Table 1 shows the average sample and population sizes for the main industry groups covered by the RSI over this time period. For the purposes of this simulation exercise data from industry groups 526 and 527 are combined.

4.2 Modelling The Population

Since complete real survey population data are not available (only samples from previous surveys) the population is constructed from a model. The actual register turnover auxiliary variable is available for all the population units and so this was used. The sales data (y) are modelled according to the following model:

$$\log(y_i) = \alpha_k + \beta_k \log(x_i) + \epsilon_i$$

where unit i belongs to industry group k , x_i is register turnover and ϵ_i are independent observations from a normal random variable with mean 0 and variance σ^2 . The most extreme values are pre-treated in the dataset available for modelling, so that some further adjustment is needed to give a realistic simulation. To give a sufficient number of extreme values, 0.2% of the units are selected at random and their y -values are multiplied by 10. With this approach, approximately 10 extreme outliers occur in the sample each month, which is close to the number observed in practice. The parameters in the model are estimated from the survey data using robust estimation procedures. They are shown in Table 2, along with industry group population totals of the simulated sales data y .

4.3 Method of Sampling

The sampling method used in ONS business surveys and applied to the current simulation exercise is essentially fixed size simple random sampling, but the samples are

selected using a permanent random number (PRN) rotation system (Kokic 1997). Each unit is assigned a random number (that is a PRN) selected from a uniform distribution between zero and one. Within each stratum units are arranged (from left to right) on the zero-one interval according to the value of their random numbers and the leftmost units are included in the sample, where n_h is the sample size for stratum h . A sampling interval is thus formed by constructing a closed interval which covers all units selected in the sample within each stratum. The sample is rotated by moving both ends of the sampling interval to the right in such a way as to include a given number of new units in the sample each month and to remove exactly the same number of units from the sample. The fixed proportion of new units is referred to as the rotation rate of the sample. For the RSI the 1995/96 rotation rate was set in such a way that newly selected units remain in the sample for approximately 15 months before being rotated out. Of course, it is unnecessary to rotate the sample in the CEd strata.

4.4 The Simulation Procedure

A single population is generated from the model described in Subsection 4.2 using the population sizes indicated in Table 1, and the resulting population values of y and x are held fixed over the whole simulation process. The procedure for simulation is as follows.

- a) Each unit is given a PRN and then units are sorted within strata according to the PRN.
- b) The sample for each month is selected using a rotation period of 15 months within sampled strata (see Subsection 4.3). Samples for 29 consecutive months are selected in this way.
- c) For one-sided M-estimation, Q is calculated to satisfy (20) from Subsection 2.2. This is done using either pooled data from months 1-12, or data from the current month

only. A single Q parameter is estimated for estimating the total as well as separate values of Q for estimating each industry group total. The fast approximate solution described in Subsection 2.2 is also calculated, using $\tilde{\mu}_i = \tilde{\beta}_h x_i$ in (17) and (18) where $\tilde{\beta}_h$ is the ratio of sample means of y_i and x_i , the ratio of sample medians, or the ratio of Huber means (using the `huber` function in R), for stratum h .

- d) For two-sided M-estimation, Q is calculated numerically using the `optimize` function in R, to minimize the estimate of the mean squared error in (25). As for one-sided M-estimation, a single Q parameter is estimated for estimating the total as well as separate values of Q for estimating each industry group total. Q is calculated using either pooled month 1-12 data, or data from the current month only.
- e) These parameters are then used to produce outlier-robust and non-robust ratio estimates of total from the samples from months 28 and 29 only. The sample from these months are used because they are entirely non-overlapping with those samples used to optimize Q except in CEd strata, which will correspond with practice. At the same time, estimates of change between months 28 and 29 are produced by taking the difference between the corresponding estimates of total. Variance estimates for each of the estimates of total are also produced by using truncated values in the standard ratio variance estimator (expression 6.9 in Cochran, 1977, p 155). Variances of estimates of change are derived from the formulae presented in Cochran (1977, pp 351-353).
- f) Confidence intervals are calculated using two methods. Firstly, standard errors are calculated as described in (d), and asymptotic 95% confidence intervals are given by the estimate plus or minus 1.96 times the standard error. Secondly, non-parametric studentized bootstrap percentile confidence intervals are calculated (Davison and Hinkley 1997, formula 2.12, p29), with 1000 bootstrap replicates, again using the

standard errors from (d). Tuning parameters and cutoffs are treated as fixed so that computation time was feasible, rather than also bootstrapping tuning parameter estimation as recommended by Beaumont and Rivest (2009, section 4.4).

g) The process (a)-(f) above is repeated for each of 1000 simulations.

Using the simulated estimates and their corresponding variance estimates, the bias, root mean squared error (RMSE) and coverage properties of nominal 95 per cent confidence intervals (CIs) for the estimates of total and of change are calculated. These results are presented in Section 6 below. In this data, where there are large positive outliers, but no small outliers, there is, in principle, little difference between the one-sided and two-sided M-estimators. The major difference is likely to be that a theoretical result is available to guide the choice of Q in the one-sided case (as discussed in 2.2), while Q minimises an approximate estimator of the MSE in the two-sided case (see 2.3).

All computations are carried out in the R statistical environment version 3.0.3 (R Core Team 2014). The `survey` package is used to calculate survey-weighted estimates (Lumley 2014). Parallelisation is used to speed up the simulation using the R packages `doParallel` (Revolution Analytics and Weston 2014) and `doRNG` Gaujoux (2014) with 7 separate processes on a desktop computer using an i7 chip running an Ubuntu UNIX operating system. All code is provided as supplementary material to the published paper, but access restrictions mean that the source dataset cannot be shared.

4.5 Effects not Covered by the Simulations

A number of effects are deliberately omitted from the simulations so that the impact of those that were included can be measured without being confounded. Firstly, changes in the register and births and deaths of businesses are excluded so that the population is unchanging. Only one set of auxiliary data is used for the whole simulation period.

Secondly, seasonality and real changes in the economy were excluded: the data generated from the super-population model are fixed for all months. Finally, non-response is only accounted for in its impact on the resulting sample sizes (the response rate in RSI, which is a compulsory survey, is about 90 per cent): its effects have not been specifically modelled in the simulation process. That is, data are missing at random. Such restrictions are expected to have little impact on the results reported below for estimates of total, but may tend to show that estimates of change perform better than they will in reality.

5 Simulation Results

The net effect of the trade-off between the bias and variance of the outlier-robust ratio estimators of total is presented in Table 3. The relative bias is defined as the bias divided by the true population total (as a percentage), and the relative root mean squared error (RRMSE) is the square root of the mean squared error divided by the true population total (as a percentage). When outlier treatment is applied at the total level:

- The absolute relative bias is about half of the RRMSE, for the one-sided and two-sided approaches.
- The RRMSE is reduced from 0.87 (no treatment) to 0.80 (one-sided M-estimation) and 0.82 (two-sided M-estimation).
- Greater relative reductions in RRMSE occur at industry group level than at the total level, for most groups.

When outlier treatment is applied at the industry group level:

- The reductions in the RRMSEs of industry group estimates is about double that achieved by tuning at the total level.

- Biases are much greater than when tuning is at the total level.
- The RRMSEs of total estimates are higher than if no outlier treatment is applied.
- One-sided M-estimation does better than two-sided for 4 groups and is very similar in the other 2.

As discussed in 2.2, a non-iterative approximate form of one-sided M-estimation is usually implemented in practice (one-sided Winsorisation). Table 4 shows the effect of substituting alternative estimators of μ_i^* into the optimal cutoffs in (17). The table shows that the choice of an estimator of the mean in the cutoff calculation has very little effect on RRMSEs. The ratio of the means appears to be the best option, in spite of its non-robustness to outliers. In practice, the ratio of Huber means may be preferable for safety in case of extreme outliers.

Table 5 shows the biases and mean squared errors achieved when the tuning parameters are chosen using live data from the current month, compared to pooled data from months 1-12. Both the one-sided and two-sided M-estimation approaches give reduced MSEs compared to no treatment when live data is used. However, the benefit is much reduced, and the reduction in RRMSE is only about half that achieved when historical pooled data is used.

Table 6 shows relative biases and RRMSEs for estimators of monthly movement. The outlier-robust estimators of change have much smaller biases than the corresponding estimators of level. This can be explained by the biases in each level estimate roughly cancelling out under subtraction. Presumably because of the resulting low biases, the gains in RRMSE from outlier treatment are much more substantial than for level estimates. For 5 out of 6 industry groups, one-sided M-estimation had lower RRMSE than two-sided, by 20% for 525 but with only small differences for the other groups.

Furthermore, both robust estimators have considerably smaller RRMSE than the non-robust estimator, up to as much as 36%, and there appears to be a considerable further reduction in RRMSE when cut-offs are chosen at group level.

Table 7 shows the coverage properties of confidence intervals (CIs) around the various estimators of level. The nominal non-coverage rate is 5% in all cases. The non-coverage rates for the asymptotic CIs (i.e. $\pm 1.96SE$) for industry group totals are 6.6-10.2% for no treatment, 11.1-15.8% for one-sided and 10.6-19.9% for two-sided M-estimation. The biases of the outlier robust methods apparently damage the confidence interval coverages. The use of bootstrap percentile intervals considerably improves the coverage under all 3 approaches. Non-coverage is now 4.1-7.0%, 6.6-11.1% and 7.1-15.4% for no treatment, one-sided M-estimation and two-sided M-estimation, respectively.

Table 8 shows the coverage of asymptotic confidence intervals for estimators of monthly movements. The percentile intervals are not calculated, to reduce the computation time for the simulation. The non-coverage rates of the robust estimators are much closer to the nominal 5% for movements than for levels. This is presumably because biases are much less for movements, due to the biases of the two monthly level estimators approximately cancelling out.

6 Discussion

Results presented in this paper suggest that there is considerable benefit in using an approximately MSE-optimal M-estimation when producing estimates of both level and movement. For positive variables, one-sided M-estimation performs slightly better for both level and movement estimates and for confidence interval coverage than the two-sided approach of Beaumont and Alavi (2004). This is probably due to the different methods for optimising the MSE. However, the differences between the two approaches

are generally quite small.

Nominal 95% confidence intervals have non-coverage up to 15.8% and 19.9% for asymptotic intervals about the one- and two-sided robust estimators of group total. This improves to worst case non-coverages of 11.1% and 15.4% respectively when a studentized percentile bootstrap interval is used. These non-parametric, asymmetric intervals are therefore recommended whenever outlier-robust methods are adopted, otherwise published confidence intervals will be excessively optimistic.

There is a very noticeable conflict between the optimal tuning parameters at total versus industry group levels in a simulated retail survey. Tuning at the total level still gives reductions in MSE at the industry group level. Greater gains in industry group estimates can be achieved by tuning at the finer level, at the cost of making the MSE of the total estimator higher than when no outlier treatment is applied. The appropriate choice depends on the relative priority of estimates at these two levels in any particular survey. This tradeoff applies equally to the one- and two-sided approaches.

An approximate non-iterative version of one-sided M-estimation is typically used in practice. The use of this approximation appears to have little impact, regardless of how the mean is estimated for the purpose of calculating Winsorising cutoffs.

For reasons of consistency with estimates of total, it may be preferable to estimate tuning parameters based on level estimates. However, it is possible to apply the benchmarking techniques described in Kokic and Jones (1997) to produce consistent estimates of totals and movements, probably with little loss of precision. The cost of such an improvement lies in the need to revise estimates. Even when tuning parameters are based on level estimates, movement MSEs are greatly improved by outlier treatment, more so than for level estimates.

Mulry et al. (2014) also conclude that two-sided M -estimation and one-sided

Winsorisation have very similar empirical properties. However, they find that the optimal tuning parameter in two-sided M-estimation is sensitive to the starting value for iteration and that Winsorisation trims too many non-outliers in their simulation. We find that the trimming of Winsorisation does not seem to be excessive in surveys with a more typical number of outliers.

In a situation where output from a NSI is under public scrutiny, it helps the public perception of the independence of the institute to have a written procedure which is based on a transparent criterion (in this case minimisation of the mean squared error), and not a procedure depending on the view of whoever is processing the data.

The main challenge in M-estimation is deciding between the options for the issues discussed in Section 3. It also relies on the assumption that the auxiliary data are correct. When negative values occur and may be outliers, two-sided M-estimation as developed by Beaumont and Alavi (2004) is a good option. For most economic survey variables, optimal one-sided M-estimation or Winsorisation seems to be slightly preferable and offers an automated, transparent and relatively simple-to-implement method of dealing with outliers.

Acknowledgements

The authors thank the Office for National Statistics for allowing RSI data to be used in this research. Ray Chambers and Alan Welsh provided helpful advice in the development of Theorem 1. Three anonymous referees made comments which improved the paper.

Appendix 1 Proof of Theorem 1

Let $r = U - s$ refer to non-sampled units and $\sigma_i^{*2} = \text{var} [y_i^*]$. Note that by assumption,

$\sum_s (w_i - 1) \mu_i = \sum_r \mu_i$. The prediction MSE of \hat{t}_y^* is:

$$\begin{aligned}
 MSE &= E \left[(\hat{t}_y^* - t_y)^2 \right] \\
 &= \text{var} (\hat{t}_y^* - t_y) + \left\{ E (\hat{t}_y^* - t_y) \right\}^2 \\
 &= \text{var} \left(\sum_{i \in s} (w_i - 1) y_i^* - \sum_{i \in r} y_i \right) + \left\{ E \left(\sum_{i \in s} (w_i - 1) y_i^* - \sum_{i \in r} y_i \right) \right\}^2 \\
 &= \sum_{i \in s} (w_i - 1)^2 \sigma_i^{*2} + \left\{ \sum_{i \in s} (w_i - 1) (\mu_i^* - \mu_i) \right\}^2 + \text{var} \left(\sum_{i \in r} y_i \right) \\
 &= \sum_{i \in s} (w_i - 1)^2 \sigma_i^{*2} + B^2 + \text{var} \left(\sum_{i \in r} y_i \right)
 \end{aligned}$$

where $B = E (\hat{t}_y^* - t_y) = \sum_{i \in s} (w_i - 1) (\mu_i^* - \mu_i)$. Let $p_i = P [y_i > K_i]$. Then

$$\begin{aligned}
 \frac{\partial \mu_i^*}{\partial K_i} &= \frac{\partial}{\partial K_i} \left\{ \int_{-\infty}^{K_i} y f_i(y) dy + \int_{K_i}^{\infty} K_i f_i(y) dy \right\} \\
 &= K_i f_i(K_i) - K_i f_i(K_i) + \int_{K_i}^{\infty} f_i(y) dy = p_i \\
 \frac{\partial \sigma_i^{*2}}{\partial K_i} &= \frac{\partial}{\partial K_i} \left\{ \int_{-\infty}^{K_i} y^2 f_i(y) dy + \int_{K_i}^{\infty} K_i^2 f_i(y) dy - \mu_i^{*2} \right\} \\
 &= K_i^2 f_i(K_i) - K_i^2 f_i(K_i) + 2K_i \int_{K_i}^{\infty} f_i(y) dy - 2\mu_i^* p_i \\
 &= 2K_i p_i - 2\mu_i^* p_i = 2p_i (K_i - \mu_i^*)
 \end{aligned}$$

Hence

$$\begin{aligned}
 \frac{\partial MSE}{\partial K_i} &= \frac{\partial}{\partial K_i} \left(\sum_{i \in s} (w_i - 1)^2 \sigma_i^{*2} + B^2 \right) \\
 &= (w_i - 1)^2 \frac{\partial \sigma_i^{*2}}{\partial K_i} + 2B (w_i - 1) \frac{\partial \mu_i^*}{\partial K_i} \\
 &= 2 (w_i - 1)^2 p_i (K_i - \mu_i^*) + 2B (w_i - 1) p_i
 \end{aligned}$$

Setting to zero and dividing through by $2p_i (w_i - 1)$ gives $K_i = \mu_i^* - B (w_i - 1)^{-1}$.

References

- Basak, P., Chandra, H., and Sud, U. (2014), “Estimation of finite population total for skewed data,” *J. Indian Soc. Agricultural Statist.*, 68, 333–341.
- Beaumont, J.-F. and Alavi, A. (2004), “Robust generalized regression estimation,” *Surv. Method.*, 30, 195–208.
- Beaumont, J.-F., Haziza, D., and Ruiz-Gazen, A. (2013), “A unified approach to robust estimation in finite population sampling,” *Biometrika*, 100, 555–569.
- Beaumont, J.-F. and Rivest, L.-P. (2009), “Dealing with outliers in survey data,” in *Handbook of Statistics*, Elsevier, vol. 29, pp. 247–279.
- Central Statistical Office (1992), *Standard Industrial Classification of Economic Activity*, Newport: Central Statistical Office.
- Chambers, R. and Clark, R. (2012), *An Introduction to Model-Based Survey Sampling with Applications*, Oxford: Oxford University Press.
- Chambers, R., Hentges, A., and Zhao, X. (2004), “Robust automatic methods for outlier and error detection,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 167, 323–339.
- Chambers, R. and Kokic, P. (1993), “Outlier robust sample survey inference,” in *Proceedings of the 49th Session of the International Statistical Inference, Firenze*.
- Chambers, R. L. (1986), “Outlier robust finite population estimation,” *J. Amer. Statist. Assoc.*, 81, 1063–1069.
- Clark, R. G. (1995), “Winsorisation methods in sample surveys,” Master’s thesis, Australian National University, <http://hdl.handle.net/10440/1031>.

- (2015), “surveyoutliers: a package to help handle outliers in sample surveys,”
<http://CRAN.R-project.org/package=surveyoutliers>, R package version 0.0.
- Cruddas, M. and Kokic, P. (1996), “The treatment of outliers in ONS business surveys,”
in *Proceedings of the GSS(M) Methodology Conference, Office of National Statistics,
Newport UK*.
- Davison, A. and Hinkley, D. (1997), *Bootstrap Methods and their Application*, Cambridge:
Cambridge University Press.
- Duchesne, P. (1999), “Robust calibration estimators,” *Surv. Method.*, 25, 43–56.
- Gaujoux, R. (2014), *doRNG: Generic Reproducible Parallel Backend for foreach Loops*, R
package version 1.6.
- Gross, W. F., Bode, G., Taylor, J., and Lloyd-Smith, C. (1986), “Some finite population
estimators which reduce the contribution of outliers,” in *Proceedings of the Pacific
Statistical Congress. Elsevier Science Publishers BV, Amsterdam, The Netherlands*, pp.
386–390.
- Gwet, J.-P. and Rivest, L.-P. (1992), “Outlier resistant alternatives to the ratio
estimator,” *J. Amer. Statist. Assoc.*, 87, 1174–1182.
- Hedlin, D., Falvey, H., Chambers, R., and Kokic, P. (2001), “Does the model matter for
GREG estimation? A business survey example,” *J. Off. Stat.*, 17, 527–544.
- Hidiroglou, M. and Berthelot, J. (1986), “Statistical Editing and Imputation for
Periodical Business Surveys,” *Surv. Method.*, 12, 73–83.
- Huber, P. J. and Ronchetti, E. (2009), *Robust Statistics*, Hoboken: Wiley.

- Hulliger, B. (1995), “Outlier robust Horvitz-Thompson estimators,” *Surv. Method.*, 21, 79–87.
- Karlberg, F. (2000), “Survey estimation for highly skewed populations in the presence of zeroes,” *J. Off. Stat.*, 16, 229–242.
- Kokic, P. (1997), “Repeated sampling through panel rotation,” *Social and Community Planning Research, UK: Survey Methods Centre Newsletter*, 17, 6–8.
- Kokic, P. and Bell, P. (1994), “Optimal winsorizing cutoffs for a stratified finite population estimator,” *J. Off. Stat.*, 10, 419–435.
- Kokic, P. and Jones, T. (1997), “Comparing estimation methods for a monthly business inquiry,” in *Proceedings of the Statistics Canada conference: New directions in Surveys and Censuses, Ottawa, November 5-7, 1997*.
- Lewis, D. (2007), “Winsorisation for estimates of change and outstanding issues with the implementation of Winsorisation for level estimates,” ONS Report for 13th Meeting of the National Statistics Methodology Advisory Committee, available from <http://www.ons.gov.uk/ons/guide-method/method-quality/advisory-committee/2005-2007/thirteenth-meeting/index.html>.
- Lumley, T. (2014), “survey: analysis of complex survey samples,” R package version 3.30.
- Martinoz, C. F., Haziza, D., and Beaumont, J.-F. (2015), “A method of determining the winsorization threshold, with an application to domain estimation,” *Surv. Method.*, 41, 57–77.
- Mulry, M. H., Oliver, B. E., and Kaputa, S. J. (2014), “Detecting and Treating Verified Influential Values in a Monthly Retail Trade Survey,” *J. Off. Stat.*, 30, 721–747.

- Myers, R. and Pepin, P. (1990), “The robustness of lognormal-based estimators of abundance,” *Biometrics*, 1185–1192.
- Preston, J. and Mackin, C. (2002), “Winsorization for Generalised Regression Estimation,” Paper for the Methodological Advisory Committee, available from <http://www.abs.gov.au/ausstats/abs@.nsf/mf/1352.0.55.051>.
- Preston, J. and Watmuff, R. (2005), “Winsorization for linear related items,” in *Proceedings of the 55th Session of the ISI Conference, Sydney 5-12 April 2005*.
- R Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Revolution Analytics and Weston, S. (2014), *doParallel: Foreach parallel adaptor for the parallel package*, R package version 1.0.8.
- Särndal, C., Swensson, B., and Wretman, J. (1992), *Model Assisted Survey Sampling*, New York: Springer-Verlag.
- Searls, D. T. (1966), “An estimator for a population mean which reduces the effect of large true observations,” *J. Amer. Statist. Assoc.*, 61, 1200–1204.
- Ståhl, O. (2015), “Point estimation using tail modelling for right skew populations,” *Journal of Statistical Computation and Simulation*, 1–16.

Table 1: Sample and population sizes

Industry Group	Description	Sample	Population
52	Whole retail sector	4881	226550
521	Non-specialist stores	709	63588
522	Food, beverages & tobacco	787	41571
523	Pharmaceutical & cosmetics	190	7788
524	Other new-goods specialist stores	2683	99112
525	Second-hand stores	179	6202
526-7	Retail sales not in stores and repairs of personal goods	333	8289

Table 2: Model Parameters

Industry Group	α	β	σ	t_{yh} (/pounds 1,000,000)
521	3.07	0.98	0.43	1656
522	3.27	0.93	0.47	320
523	3.26	0.85	0.91	97
524	3.03	0.98	0.74	1588
525	3.25	0.94	0.75	41
526-7	2.94	1.00	0.74	167

Table 3: Relative bias (%) and relative root mean squared error (%) of estimators of level with tuning parameters chosen at either group or total level using historical training data

Industry Group	Ratio Estimator	M-Estimators			
		One-Sided		Two-Sided	
		total level	group level	total level	group level
RELATIVE BIAS (%)					
total	-0.04	-0.39	-0.71	-0.42	-0.74
521	0.02	-0.01	-0.15	-0.01	-0.13
522	0.03	-0.16	-1.04	-0.18	-0.91
523	-0.08	-0.47	-3.14	-0.50	-3.61
524	-0.10	-0.86	-0.92	-0.93	-1.11
525	-0.66	-0.79	-4.11	-0.80	-2.65
526-527	0.02	0.01	-1.29	0.00	-0.79
RELATIVE ROOT MEAN SQUARED ERROR (%)					
total	0.87	0.80	0.96	0.82	1.00
521	0.44	0.41	0.36	0.41	0.36
522	2.68	2.44	2.27	2.43	2.25
523	7.47	6.83	6.11	6.82	6.89
524	1.89	1.66	1.69	1.71	1.79
525	8.98	8.77	8.20	8.76	8.51
526-527	3.31	3.29	3.14	3.27	3.18

Table 4: Relative root mean squared error (%) of estimators of level based on one-sided M-estimation and approximate M-estimation with different estimators of level, with tuning parameters chosen at group level using historical training data

Industry Group	Ratio Estimator	One-sided Winsorisation with truncation based on			One-Sided M-estimation
		ratio of means	ratio of medians	ratio of Huber means	
total	0.87	0.94	0.97	0.96	0.96
521	0.44	0.36	0.36	0.36	0.36
522	2.68	2.25	2.27	2.27	2.27
523	7.47	6.08	6.12	6.12	6.11
524	1.89	1.68	1.70	1.69	1.69
525	8.98	8.18	8.25	8.25	8.20
526-527	3.31	3.11	3.19	3.16	3.14

Table 5: Relative bias (%) and relative root mean squared error (%) of estimators of level with tuning parameters chosen at group level using either live or historical training data

Industry Group	Ratio Estimator	M-Estimators			
		one-sided		two-sided	
		historical	live	historical	live
RELATIVE BIAS (%)					
total	-0.04	0.01	-0.52	0.01	-0.58
521	0.02	0.00	-0.10	0.00	-0.08
522	0.03	0.05	-0.71	0.05	-0.49
523	-0.08	0.37	-2.25	0.47	-3.57
524	-0.10	0.03	-0.71	0.03	-0.88
525	-0.66	0.67	-3.26	0.72	-3.05
526-527	0.02	0.10	-1.00	0.10	-0.39
RELATIVE ROOT MEAN SQUARED ERROR (%)					
total	0.87	0.96	0.91	1.00	0.93
521	0.44	0.36	0.39	0.36	0.39
522	2.68	2.27	2.42	2.25	2.40
523	7.47	6.11	6.75	6.89	6.58
524	1.89	1.69	1.77	1.79	1.79
525	8.98	8.20	8.71	8.51	8.21
526-527	3.31	3.14	3.27	3.18	3.20

Table 6: Relative bias (%) and relative root mean squared error (%) of estimators of movement with tuning parameters chosen at either group or total level using historical training data

Industry Group	Ratio Estimator	M-Estimators			
		One-Sided		Two-Sided	
		total level	group level	total level	group level
RELATIVE BIAS (%)					
total	0.01	0.01	0.01	0.01	0.01
521	0.00	0.00	0.00	0.00	0.00
522	0.01	0.01	-0.01	0.01	0.00
523	0.12	0.10	0.05	0.10	0.03
524	0.03	0.02	0.02	0.02	0.02
525	0.23	0.24	0.16	0.24	0.22
526-527	-0.05	-0.05	-0.06	-0.06	-0.06
RELATIVE ROOT MEAN SQUARED ERROR (%)					
total	0.34	0.27	0.25	0.27	0.25
521	0.16	0.15	0.12	0.15	0.12
522	0.99	0.89	0.70	0.90	0.72
523	3.14	2.72	2.00	2.71	2.03
524	0.75	0.56	0.55	0.56	0.54
525	4.00	3.89	2.88	3.87	3.46
526-527	1.58	1.57	1.31	1.57	1.42

Table 7: Non-coverage of asymptotic and bootstrap percentile confidence intervals of M-estimators of level with tuning parameters chosen at group level using historical training data and nominal non-coverage rate of 5%.

Industry Group	Ratio Estimator		One-Sided		Two-Sided	
	asypm	pctile	asypm	pctile	asypm	pctile
total	7.0	6.4	24.2	18.8	28.8	22.7
521	7.0	7.5	11.1	9.1	10.2	8.8
522	7.9	7.3	11.5	10.0	11.1	10.9
523	10.2	7.0	15.8	10.8	19.9	15.4
524	10.0	7.8	14.7	11.1	18.8	14.3
525	9.5	6.1	15.4	9.6	13.1	9.0
526-527	6.6	4.1	11.4	6.6	10.6	7.1

Table 8: Non-coverage of asymptotic confidence intervals of M-estimators of movements with tuning parameters chosen at group level using historical training data and nominal non-coverage rate of 5%.

Industry Group	Ratio Estimator	One-Sided	Two-Sided
total	8.1	6.8	6.9
521	6.2	5.7	5.7
522	7.2	4.8	5.4
523	9.0	7.3	6.6
524	9.1	7.0	7.0
525	10.2	9.8	9.7
526-527	12.0	10.6	12.7