STATISTICS IN TRANSITION-new series, July 2010
Vol. 11, No. 1, pp. 127—144

ROBUST ESTIMATION OF FINITE POPULATION TOTAL

Manoj Kumar Srivastava[1], Namita Srivastava[2]

ABSTRACT

The present paper deals with the robust prediction of finite population total under the superpopulation model G_R . The design is de-emphasized while developing these predictors under the superpopulation model and making comparison among all resistant estimators. The suggested proposals involve reweighed iterative algorithm for Robust Prediction. The discussion also involves the calculation of asymptotic bias and variance in terms of the influence function computed for these predictors. Two populations have been considered for simulation study to judge the performance of proposed predictors with conventional and model based existing alternatives.

Key words: Influence function, Prediction approach, M-estimator, Superpopulation models.

1. Introduction

Sample survey data often contains outliers. Initially according to literature, applied survey practitioners ignored the problem of outliers in survey sampling probably treating this as perennial problem. Initial advises have been made by Kish (1965, sec.11.4B) in survey sampling literature under the topic "Skew populations" in economic survey and survey of individuals. Hidiroglou and Srinath (1981) and Glasser (1962) form a separate stratum for outlying units and then combined the class means. Ernst (1980), Fuller (1991), and Searls (1966) showed that for skewed population, the mean squared error of the winsorised sample mean is smaller than that of the sample mean. Rivest (1993) studied the behavior of winsorization schemes under simple random sampling. Shoemaker and Rosenberger (1983) calculated the exact formula for the expected value and variance of trimmed mean and of median under the simple random sampling.

[1] Department of Statistics, Institute of Social Sciences, Dr. BRA University, Agra-282002, U.P. India. E-mail: *mks_iss@yahoo.co.in.*
[2] Department of Statistics, St. John's College, Agra-282002, U.P. India. E-mail: *drnamita.sjc@gmail.com.*

Fuller (1991) suggested estimators which are so constructed that the impact of the largest y-value is reduced only when a test for extreme values is significant. Smith (1987) highlighted and showed the importance of detecting and treating outliers under the superpopulation model.

Ichan (1984) suggested predictors of population mean in the presence of spurious observations under a superpopulation model and studied sensitivity and robustification aspects. Chambers (1986) distinguishes between representative and non representative outliers in a sample and suggested ϵ-contaminated model based outliers resistant predictors of population mean by using M-estimation technique. Gwet and Rivest (1992) discussed the problem of estimating the population mean using auxiliary information in the presence of outliers. Beat Hulliger (1995) suggested outlier robust alternatives for Horvitz and Thompson estimator.

In Sample survey the finite population is denoted by $U = \{U_1, U_2, \cdots, U_N\}$; the study variable y on U be denoted by $\mathbf{y} = \{y_1, y_2, \cdots, y_N\}$ and assumed to be unknown. Let an unordered sample s be drawn from F under a given design $p(s) > 0$ and the data $d = \{(k, y_k) : k \in s\}$ has been observed. For any given $s \in F$, we can write

$$T = \sum_s y_k + \sum_{\tilde{s}} y_k \tag{1}$$

where $\tilde{s}$ denote the set of labels not in s, *i.e.*, $\sum_{\tilde{s}} y_k$ is the total of non-sampled units. The problem of estimating T is recognized as one of predicting the sum of the unobserved random variable $\sum_{\tilde{s}} y_k$. This leads to the prediction of T. For this, the true value of the parameter $\mathbf{y}$ be a realization of $\mathbf{Y}$ under the superpopulation model given by

$$Y_k = \beta x_k + e_k; \ \ k = 1, 2, \cdots, N$$

such that $E(Y_k \mid x_k) = \beta x_k$, $V(Y_k \mid x_k) = \sigma^2 v(x)$

and $\operatorname{cov}(Y_k, Y_l \mid x_k, x_l) = 0 \ \ \forall \ \ k \neq l = 1, 2, \cdots, N.$ (2)

where $x_k > 0$ $(k = 1, 2, \cdots, N)$ are known, auxiliary values β and σ^2 are unknown and $v(\cdot)$ is known.

The sample s is used to estimate the parameter β in the model ξ in equation (2) which is then used to predict values of y_k for $k \notin s$, *i.e.*, for $N - v(s)$ unobserved coordinates of $\mathbf{y} = (y_1, y_2, \cdots, y_N)$. Therefore,

$$T(d) = \sum_{s} y_k + U(d)$$

is called the predictor of T, where U is considered as predictor of $\sum_{\tilde{s}} y_k$ the total of $N - v(s)$ unobserved coordinates of **y** (Cassel, et al. (1977); Chambers(1986)).

The predictor $T(d)$ is highly sensitive to the presence of outliers both in x's and residuals r's. In the present paper the objective is to obtain a robust optimal predictor of T.

2. Proposed outlier robust predictor of *T*

Brewer (1963) and Royall (1970) gave $UM\xi UMSEE$ (Uniformly minimum ξ -unbiased mean square error estimator) in the class of all homogeneous model unbiased predictors of T

$$T_{BR} = Y_s + \widehat{\beta} x_{\tilde{s}}$$

$$\xi MSE_p(T_{BR}) = E\left\{\left(\sum_{\tilde{s}} x_k\right)^2 V_\xi\left(\widehat{\beta}\right) + \sigma^2 \sum_{\tilde{s}} V_\xi(x_k)\right\} \tag{3}$$

where $x_{\tilde{s}} = \sum_{\tilde{s}} x_k$, $Y_s = \sum_s y_k$, $\widehat{\beta} = \dfrac{\sum_s x_k y_k / v(x_k)}{\sum_s x_k^2 / v(x_k)}$ and $V_\xi\left(\widehat{\beta}\right) = \dfrac{\sigma^2}{\sum_s x_k^2 / v(x_k)}$

Chambers (1986) among others noted the high changes in the distribution of $\widehat{\beta}$ when sample data deviated from the Gauss Markov assumptions resulting into high sensitivity in $\widehat{\beta}$. This sensitivity is carried over to the predictor T_{BR}. We have proposed an outlier robust predictor alternative to $T_{BR}(d)$ which is less sensitive to sample outliers.

To replace $\widehat{\beta}$ by an outlier robust alternative, using general robust procedure developed by Maronna and Yohai (1981) (c.f. Hampel et al. (1987)) for hypothetical populations, and by taking its finite population analogous for superpopulation model (Gwet and Rivest. (1992)). The M-estimator of the parameter β in G_R is obtained by solving the equation implicitly

$$\sum_{k=1}^{n} \eta\left(x_k, \frac{y_k - x_k \beta}{\sigma v(x_k)}\right) x_k' = 0 \tag{4}$$

where $x'_k = \dfrac{x_k}{\sigma\sqrt{v(x_k)}}$ and $\eta : R^1 \times R^1 \to R^1$ which satisfies some regularity conditions (Hampel, et al. (1987), p 315). The solution of the above equation gives regression *M*-estimator. η-function is a function ψ of x_k, y_k and β which preserves the sign as of x_k, and depends on β only through $r = y - x'\beta$. These restrictions derive their requirements from equivariance considerations. If one writes η as

$$\eta(x,r) = w(x).\psi(r, v(z)) \tag{5}$$

where $\psi: R^1 \to R^N$, $w: R^P \to R^+$, $v: R^P \to R^+$ (w and v are known as weight functions). In particular, this paper uses Merrill and Schweppe's proposal (Hampel, et al. (1987), p 315) which gives $\eta(x,r) = w(x)\psi\left(r, \dfrac{1}{w(x)}\right)$ *i.e.,* $v(x) = 1/w(x)$ in equation (5). Large x is assigned a low weight and if this results into large residual r it assigns high $\dfrac{1}{w(x)}$ *i.e.* high $r.\dfrac{1}{w(x)}$, ψ is defined as diminishing or nearly zero outside a range and identity near the origin. Such a $\psi(\cdot)$ function contribute much less for large argument value *i.e.* $r.\dfrac{1}{w(x)}$. The above proposal suggests using redescending ψ - function.

3. Estimating equation for proposed predictor

Let the residuals r and x be standardized by dividing by $\sqrt{V(y/x)} = K.\sigma\sqrt{v(x)}$. Using this transformation the present set up of the model G_R is changed to the Markovian set up. The defining equation can now be written as

$$\sum_s \frac{x_k}{K\sigma\sqrt{v(x_k)}}.\frac{1}{w(x_k)}\psi\left(\frac{y_k - \beta x_k}{K\sigma\sqrt{v(x_k)}}.w(x_k)\right) = 0 \tag{6}$$

For a proper choice of ψ, the following conditions are imposed in the defining equation

1. $\psi(t)$ is real and skew-symmetric, i.e. $\psi(-t) = -\psi(t)$.
2. $\psi(t)$ is bounded.

3. $\frac{\psi(t)}{t} \to 1$ as $|t| \to 0$, $\psi(t) \to t$ as $|t| \to 0$, an identity transformation for small values of *t*.

4. $w(t)$ is positive and non-decreasing function of *t*.

we get, for $v(x_k) = x_k$, $\hat{\beta}_{LS} = \frac{\sum_s y_k/n}{\sum_s x_k/n} = \frac{\bar{y}_s}{\bar{x}_s}$ a ratio estimator of the population ratio R and if $v(x_k) \propto x_k^2$, $\hat{\beta}_{LS} = \frac{1}{n}\sum_s \frac{y_k}{x_k}$ and if $v(x_k) \propto$ constant, $\hat{\beta}_{LS} = \frac{\sum_s y_k x_k}{\sum_s x_k^2}$. They all are sensitive to the presence of sample outliers. The proper choice of $\psi(\cdot)$ and weight functions, reduces the effect of presence of sample outliers in *x* and in residual *r,* to make the $\hat{\beta}$ predictors robust. If $\psi(\cdot)$ is bounded and $w(\cdot)$ is constant, the *M*-estimator obtained may still be sensitive to large values of $x's$ (note the factor $x/v(x)$ in defining equation). Such an *M*-estimator, although being insensitive/robust against the presence of outliers among the residuals, is still non-robust against the presence of outliers in *x.* (refer Hampel et al., 1986, pp 314; also the example given there.)

In order to obtained a robustified estimator of β when sample outliers are present both in residuals and among $x's$, a bounded $\psi(\cdot)$ is considered and the weight is taken as $w(x) = \frac{x/\bar{x}}{\sqrt{v(x_k)}}$ to get the estimating equation of β as

$$\sum_s \psi\left(\frac{y_k - \beta x_k}{K\sigma\sqrt{v(x_k)}} \cdot \frac{x_k}{\bar{x}\sqrt{v(x_k)}}\right) = 0.$$

3.1.Solving estimating equation

To solve the estimating equation (6) in order to get the robust estimate of β against the presence of outlying observations in the sample data we adopted Tukey's (1970, 71) and Andrews et al., (1972, pp 239) proposal on the estimating equation. After some mathematical simplification the solution algorithm could be obtained which is termed as computational iteratively re-weighted least squares algorithm (a generalized M-estimator). The results are being stated as under:

Result 1

Under the model ξ using Andrews (1972) and Andrew's et al. (1972) procedures, the solution of defining implicit equation is

$$\beta_n = \frac{\sum_s x'_k y'_k .w\left(r'_k .w\left(x_k\right)\right)}{\sum_s x'^2_k .w\left(r'_k .w\left(x_k\right)\right)} \tag{7}$$

where $x'_k = \dfrac{x_k}{K\sigma\sqrt{v(x_k)}}$ and $r' = \dfrac{y_k - \beta x_k}{K\sigma\sqrt{v(x_k)}}$

Starting with $\beta_n^{(0)} = \dfrac{med\left(x_k y_y \big/ S_n^{(0)2} v(x_k)\right)}{med\left(x_k^2 \big/ S_n^{(0)2} v(x_k)\right)}$

where $S_n^{(0)} = 1.483\ MAD(y_k) = 1.483\ med_s |\ y_k - med_s(y_l)|$ (Andrews et al., 1972, p. 239), and weights $w^{(0)}(\cdot) = w^{(0)}\left(r'^{(0)}_k \cdot w(x_k)\right) = \dfrac{\psi\left(r'^{(0)}_k \cdot w(x_k)\right)}{r'^{(0)}_k \cdot w(x_k)}$.

Iteration are continued by recalculating the weights, and modifying the value of β_n by computing

$$\beta_n^{(j+1)} = \frac{\sum_s y'_k x'_k \cdot w^{(j)}\left(r'^{(j)}_k\right)}{\sum_s x'^2_k \cdot w^{(j)}\left(r'^{(j)}_k\right)} \tag{8}$$

from the second iteration and onwards $S_n^{j+1} = sqrt\left\{\sum_s w\left(r'^{(j)}_k\right)\left(r'^{(j)}_k - \overline{r}_s\right)^2\right\}$, the square root of weighted residual sum of squares. The iterations are continued until $|\ \beta_n^{(k)} - \beta_n^{(k-1)}| \le \varepsilon, \beta_n^{(k)}$ is called k-step M-predictor of β. The rate of convergence depends on the proper robust choices of scale $S_n^{(0)}$ and the slope $\beta_n^{(0)}$. The behavior of one step and fully iterated M-predictor of β essentially the same when ψ is odd and underlying distribution F is symmetric.

Result 2

When the function $\psi(\cdot)$ be differentiable once in the neighborhood $|\ r_k - r_k^{(0)}| \le k$ the M-predictor calculated by using

$$T_n^{(j+1)} = T_n^{(j)} + \sum_s \psi\left(\frac{y_k - \beta_n^{(j)} x_k}{KS_n^{(j)}\sqrt{v(x_k)}}\right) \Bigg/ \frac{x_k}{KS_n^{(j)}\sqrt{v(x_k)}} \cdot \psi'\left(\frac{y_k - \beta_n^{(j)} x_k}{KS_n^{(j)}\sqrt{v(x_k)}}\right)$$

the iterations are continued until $T_n^{(\cdot)}$ converge, i.e. $|\ T_n^{(k)} - T_n^{(k-1)}| \le \varepsilon$.

In the following discussion the above result have been applied for different bounded and redecending ψ functions

1. Considering Huber ψ -function

$\psi(t) = \text{sgn}(t) \ \ \min(k_0, |t|) = t \ \min\left(1, \frac{k_0}{|t|}\right)$, weight becomes

$$w_k(\beta, S_n, k_0) = \min\left\{1, \frac{k_0}{\left|\frac{y_k - \beta x_k}{KS_n\sqrt{v(x_k)}}\right| \cdot w(x_k)}\right\}$$

And $\beta_n = \sum_s \frac{y_k x_k}{K^2 S_n^2 v(x_k)} \cdot w_k(\beta, S_n, k_0) \Big/ \sum_s \frac{x_k^2}{K^2 S_n^2 v(x_k)} \cdot w_k(\beta, S_n, k_0)$

where k_0 is some +*ve* constant, S_n is some predictor of σ, and robustification constant K in the scale parameter is assumed to be known.

Using iteratively reweighted least square algorithm and starting with appropriate robust predictors of σ and β as $S_n^{(0)}, \beta_n^{(0)}, S_n^{(0)} = 1.483 \text{ MAD}(y_k)$ and

$$\beta_n^{(0)} = \text{med}_k \left\{\frac{y_k x_k}{K_{(0)}^2 S_n^{(0)^2} v(x_k)}\right\} \Big/ \text{med}_k \left\{\frac{x_k^2}{K_{(0)}^2 S_n^{(0)^2} v(x_k)}\right\}$$

One computes

$$\beta_n^{(j+1)} = \frac{\sum_s \frac{y_k x_k}{K_{(0)}^2 S_n^{(j)^2} v(x_k)} \cdot w_k\left(\beta_n^{(j)}, S_n^{(j)}, k_0\right)}{\sum_s \frac{x_k^2}{K_{(0)}^2 S_n^{(j)^2} v(x_k)} \cdot w_k\left(\beta_n^{(j)}, S_n^{(j)}, k_0\right)} \tag{9}$$

until β_n converges. This predictor features the bounded or limited impact of outliers on the predictor calculated under the model $\xi(G_R)$.

2. Cauchy ψ -function which is a typical redescending ψ function defined as

$$\psi_c(t) = \frac{t}{1+t^2}; \tag{10}$$

The shape of this Cauchy function is a pure redescending function without any tuning constant. For this ψ - function the weight function is calculated as

$$w_k(\beta, \sigma) = \frac{1}{\left\{1 + \left\{\frac{y_k - \beta x_k}{K\sigma\sqrt{v(x_k)}} \cdot w(x_k)\right\}^2\right\}} \tag{11}$$

and the iteratively reweighing least square algorithm gives

$$\beta_n{}^{(j+1)} = \frac{\sum_s \frac{y_k x_k}{K_{(0)} S_n^{(j)^2} v(x_k)} \cdot w_k\left(\beta_n^{(j)}, S_n^{(j)}\right)}{\sum_s \frac{x_k^2}{K_{(0)} S_n^{(j)^2} v(x_k)} \cdot w_k\left(\beta_n^{(j)}, S_n^{(j)}\right)} \tag{12}$$

taking $S_n^{(0)}$ and $\beta_n^{(0)}$ same as has been considered earlier and iterative procedure is continued until $\beta_n^{(j)}$ converges.

Cauchy ψ function is an identity function near the origin and it converges to zero when $|\ t| \to \infty$. This is used when impact of outliers on the residual need to be controlled.

3. Chamber type $\psi(\cdot)$ function which is a redescending ψ -function

$$\psi(t) = t \exp\left\{-\frac{a}{2}\{|t| - b\}^2\right\} \tag{13}$$

where b is the location and a is scale constant or shape constant.

For the above ψ -function the weight function is calculated as

$$w_k\left(r_k'(\beta), \sigma, a, b\right) = \exp\left\{-\frac{a}{2}\left(|r_k'| w(x_k) - b\right)^2\right\} \tag{14}$$

using these weights, the iterative least squares reweighted algorithm suggests

$$\beta_n^{(j+1)} = \frac{\sum_s x_k' y_k' . w_k\left(r_k'\left(\beta_n^{(j)}\right), S_n^{(j)}, a, b\right)}{\sum_s x_k'^2 w_k\left(r_k'\left(\beta_n^{(j)}\right), S_n^{(j)}, a, b\right)} \tag{15}$$

until $\beta_n^{(\cdot)}$ converges.

If $|r_k'| w(x_k) \le b,$ the weight in the estimating process is $w_k(\cdot) \propto x_k' = x_k \Big/ \left(K \sigma \sqrt{v(x_k)}\right).$ If $|r_k'| w(x_k)$ is low then the weight $w_k \propto x_k'$ and if $|r_k'| w(x_k)$ is large and further becomes larger, $w_k(\cdot)$ becomes lower and lower contributing almost zero or asymptotically insignificant to the process of estimation.

4. If ψ is Biweight function due to Beaton and Tukey (1974) (Hampel (1986), section 2.6, p 151) which is a smooth (in the sense of continuity) function defined as $\psi_{bi,(k_0)} = t\left(k_0^2 - t^2\right)^2 \mathrm{I}_{[-k_0, k_0]}(t)$ the weights are calculated as

$$w_k\left(r_k'(\beta),\sigma,b\right)=\left[k_0^2-\left(r_k'.w\left(x_k\right)\right)^2\right]^2 \mathrm{I}_{[-k_0,k_0]}(r_k'.w_k) \tag{16}$$

and iterative least squares reweighted algorithm be given by

$$\beta_n^{(j+1)}=\frac{\sum_s x_k' y_k' \cdot w_k\left(r_k'\left(\beta_n^{(j)}\right),S_n^{(j)},k_0\right)}{\sum_s x_k'^2 \cdot w_k\left(r_k'\left(\beta_n^{(j)}\right),S_n^{(j)},k_0\right)} \tag{17}$$

until β_n converges. Note that in estimation process outlying observations $\left|r_k' w\left(x_k\right)\right|>k_0$ are totally rejected by assigning them zero weights and well behaved observations are given full weight.

The constant k_0 in ψ is determined such that $\rho^*(\psi)\le k_0$ where

$$\rho^*=\inf\left\{c>0:\left(\psi\left(y,\xi\left(G_R\right)\right)=0\right) \text{ when } |\mathrm{y}|>\text{constant}\right\} \tag{18}$$

which is based on the rejection policy due to Daniel Bernoulli in 1769 (Stigler, 1980) of rejecting contaminated values of y entirely from the estimation process. This corresponds to $IF\left(x,\psi(\cdot),\xi\right)=0$. The condition $\psi(\cdot)=0$ in (18) comes from $IF(\cdot,\cdot,\cdot)=0$.

The choice of robust constant K should be such that it balances the choice of S_n so that contaminations in residuals are properly downweighed instead rejected entirely leading to wrong conclusions. In other words, K allows the choice of level of robustness in ψ-based predictors. For example, when ψ is redescending selecting large K amounts to $\psi(t)=t$, leading to LS-predictors. For bi-weight ψ-robustness procedure K = 2 is recommended for $\left|r_k'\right|\le w\left(x_k\right)$ so that $\frac{r_k'}{w\left(x_k\right)}=\pm 1$ falls roughly at $\pm 2.7\sigma$, other values of K = 3,4, or 5 are also recommended in the literature.

In case of Cauchy ψ-function the robust constant K is calculated by the relation

$$K_{\text{Cauchy}}^{\text{GM}}=K_{\text{Hub}}^{\text{GM}}\ \frac{K_{\text{Cauchy}}^{\text{M}}}{K_{\text{Hub}}^{\text{M}}} \tag{19}$$

Notionally, all ψ-functions fall in between Cauchy and Huber ψ-functions with their corresponding robustifying constant K values 2.34 and 4.54. The choice of K by different statisticians for their ψ-function could be made between this range by using some rationality or optimality criterion.

4. Asymptotic properties

4.1. Influence function for the defined GM- estimator

The effect of small departure of the distribution of $\frac{\left(y-\beta(F)x\right)}{\left(K\sigma\sqrt{v(x)}\right)}$ from Gauss-Markov model due to the presence of outliers is measured by the influence function of β at *F(x, y)* is defined as

$$IF\left((x,y);\beta,F(x,y)\right)=\frac{\partial}{\partial t}\left[\beta\left(F_t(x,y)\right)\right]_{t=0} \tag{20}$$

where $F_t(x,y)=(1-t)F(x,y)+t\Delta_{(x,y)}$, differentiating the functional defining equation with respect to *t,* then, making use of the above definition we get

$$\frac{\partial}{\partial t}\left[\beta\left(F_t(x,y)\right)\right]_{t=0}=\frac{x'\cdot\frac{1}{w(x)}\cdot\psi\left(r'.w(x)\right)|_{\beta(F(x,y))}}{\int x'^2\cdot\frac{1}{w(x)}\cdot\psi'\left(r'.w(x)\right)|_{\beta(F(x,y))}.dF(x,y)}$$

4.2. Asymptotic bias and variance

Theorem 4.1. Under the model ξ, the ξ-bias in the predictor T_n is given by

$$\xi(T_n-T)=\frac{\sum_{\bar{s}}x_k}{\lambda_0}.\frac{1}{n}\sum_s\lambda_k-\sum_{\tilde{s}}\mu_k \tag{21}$$

where $\lambda_0=\xi\left[x'^2\frac{1}{w(x)}\psi'\left(r'.w(x)\right)\right]_{\beta(F(x,y))}$

and $\lambda_k=\xi\left[\frac{x_k'}{w(x_k)}\psi\left(r_k'.w(x_k)\right)\right]_{\beta(F(x,y))}$

Proof. If some distribution *G* is considered near *F,* then the first-order Von Mises expansion of β at *F* (which is derived from a Taylor Series) evaluated in *G* is given by

$T(G)=T(F)+\int IF(x;T,F)d(G-F)(x)$ + Remainder (refer Hampel et al. page 85)

$$\beta(G(x,y))=\beta(F(x,y))+\int IF((x,y);\psi,F(x,y))\times d\left(G(x,y)\right.$$
$$\left.-F(x,y)\right)+\text{Remainder}$$

If we assume that (x_k, y_k) are iid from $F(x, y)$ then the empirical distribution function $F_n(x,y) \xrightarrow{\pounds} F(x,y)$, by Glivenko-Cantelli theorem for large n. Also, note that $\int IF((x,y);\psi;F(x,y))dF(x,y)=0$. Under these conditions, we could now replace $G(x, y)$ by the sample empirical distribution function $F_n(x,y)$, we get

$$\beta_n\left(F_n(x,y)\right) \approx \beta\left(F(x,y)\right)+\int IF\left((x,y);\psi,F(x,y)\right)\times dF_n + \text{remainder}$$

If we further assume that $\beta(\cdot)$ is a functional, $\beta_n(F_n)$ then may be approximated by $\beta(F_n)$. Therefore,

$$\beta\left(F_n(x,y)\right)-\beta\left(F(x,y)\right) \approx \frac{1}{n}\sum_{k=1}^{n} IF\left((x_k,y_k);\psi,F(x_k,y_k)\right)+ \text{ remainder}$$

In most cases remainder becomes negligible for $n \uparrow \infty$. Rewriting this yields

$$\sqrt{n}\left(\beta\left(F_n(x,y)\right)-\beta\left(F(x,y)\right)\right) \approx \frac{1}{\sqrt{n}}\sum_{i=1}^{n} IF\left(\left(x_k,y_k\right);\psi,F\right)+R_n$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} IF\left(\left(x_k,y_k\right);\psi,F\right) \rightarrow N\left(0,V_\xi\left(\beta,F\left(x,y\right)\right)\right) \text{ as } n\rightarrow\infty$$

where $V_\xi\left(\beta,F\left(x,y\right)\right)=\int_{(x,y)} IF^2\left(\left(x,y\right);\psi,F\left(x,y\right)\right)dF\left(x,y\right)$

Therefore, the asymptotic bias in β_n shall be given by

$$(\beta_n-\beta) \approx \frac{1}{n}\sum_s \frac{\frac{x'_k}{w\left(x_k\right)}\psi\left(r'_k\cdot w(x_k)\right)\mid_{\beta(F(x,y))}}{\int x'^2_k\cdot\frac{1}{w(x)}\psi'\left(r'\cdot w(x)\right)\mid_{\beta(F(x,y))}\cdot dF(x,y)}$$

and ξ- bias is

$$B\left(\beta_n\right)=\xi\left(\beta_n-\beta\right)=\frac{1}{n}\sum_s \frac{\xi\left[\frac{x'_k}{w\left(x_k\right)}.\psi\left(r'_k.w(x_k)\right)\right]_{\beta(F(x,y))}}{\xi\left[x'^2.\frac{1}{w(x)}\psi'\left(r'.w(x)\right)\right]_{\beta(F(x,y))}}$$

$$=\frac{1}{\lambda_0}\cdot\frac{1}{n}\sum_s \lambda_k$$

The mean of the prediction error $\left(T_n-T\right)$ under the model $\xi\left(G_R\right)$, *i.e.* $\xi\left(T_n-T\right)$, also termed as ξ-bias in T_n, is given by

$$\xi(T_n - T) = \xi^{\xi}\xi^{\beta/\xi}\left\{-\sum_{\tilde{s}}(y_k - \beta_n x_k)\right\} = \frac{\sum_{\tilde{s}} x_k}{\lambda_0}.\frac{1}{n}\sum_{s}\lambda_k - \sum_{\tilde{s}}\mu_k$$

Theorem 4.2. *Under the model* ξ in the ξ*-mean square error in the predictor* T_n *is given by*

$$V_\xi(T_n - y_u) = \left(\sum_{\tilde{s}} x_k\right)^2 . \frac{\xi\left[\frac{x'}{w(x)}.\psi(r'.w(x))\right]^2_{\beta(F(x,y))}}{\xi^2\left[x'^2.\frac{1}{w(x)}\psi'(r'.w(x))\right]_{\beta(F(x,y))}}$$
$$+ (K\sigma)^2 \sum_{\tilde{s}} \xi^2 v(x_k) - 2\beta\sum_{\tilde{s}}\operatorname{cov}(x_k, y_k)$$

Proof. The asymptotic variance of the predictor β at distribution function $F(x, y)$ is given by,

$$V_\xi(\beta, F(x,y)) = \int_{(x,y)} IF((x,y);\psi, F(x,y))^2\, dF(x,y)$$

$$V_\xi(T_n - y_u) = V_\xi\left\{(\beta_n - \beta)\sum_{\tilde{s}} x_k - \sum_{\tilde{s}}(y_k - \beta x_k)\right\}$$

Here $$V_\xi(\beta_n - \beta) = \frac{\xi\left[\frac{x'}{w(x)}.\psi(r'.w(x))\right]^2_{\beta(F(x,y))}}{\xi^2\left[x'^2.\frac{1}{w(x)}\psi'(r'.w(x))\right]_{\beta(F(x,y))}}$$

$$V_\xi(y_k - \beta x_k) = (K\sigma)^2\, \xi^X v(x_k)$$

and $$\operatorname{cov}\left\{\beta_n \sum_{\tilde{s}} x_k, \sum_{\tilde{s}} y_k\right\} = \beta\sum_{\tilde{s}}\operatorname{cov}(x_k, y_k)$$

$$\therefore V_\xi(T_n - y_u) = \left(\sum_{\tilde{s}} x_k\right)^2 . \frac{\xi\left[\frac{x'}{w(x)}.\psi(r'.w(x))\right]^2_{\beta(F(x,y))}}{\xi^2\left[x'^2.\frac{1}{w(x)}\psi'(r'.w(x))\right]_{\beta(F(x,y))}}$$

$$+ (K\sigma)^2 \sum_{\tilde{s}} \xi^2 v(x_k) - 2\beta\sum_{\tilde{s}}\operatorname{cov}(x_k, y_k)$$

5. Empirical study

Two populations have been considered for simulating the performance of robust predictors for different ψ functions discussed above. The first population shows the area under wheat in 1936 for 34 village in Lucknow subdivision (Sukhatme and Sukhatme (1970, p 185)), which is also used by Som (1973, p 73) and Gwet and Rivest (1992, p 1175). This population has a single outlier when superpopulation parameter is estimated by the Royall's prediction approach. The second population has been taken from Appendix E of Kish (1965) with $N = 235$. This population has 13 outliers. By simple examination of the data, the relationship between the two variable confirms the validity of the assumed kernel superpopulation model G_R as has been defined in early section, with variance function $v(x) = x$.

All the simulation results are based on over 400 repeated samples of sizes 4, 8, 12 and 20, 30, 40 from population 1 and 2 respectively. For the calculation of the superpopulation parameter β for ψ_{BT}, ψ_{CAU}, ψ_{HUB} in the estimating equation, the respective standard residuals were being calculated. On the basis of ψ-plots for these standardized residuals, the tuning constants have accordingly been chosen.

In Chambers predictor

	ψ_{BT}	ψ_{Hub}	ψ_{Cau}	ψ_{BT}	ψ_C
pop. 1	$c = 0.3$	$k = 0.2$	–	$a = 2.5, b = -0.09$	$a = 5, b = -0.09$
pop. 2	$c = 0.5$	$k = 0.3$	–	$a = 100, b = 0.0251$	$a = 100, b = 0.006$

In this performance evaluation study the following conventional and model based estimators and predictors have been included:

1. Simple expansion estimator $T_n = N\bar{y}_s$;
2. least square estimator, which is same as ratio estimator under the model $T_n(RAE) = \dfrac{y_s}{x_s} x_u$;
3. 1-winsorized estimator $T_n(WINS) = \dfrac{y_s^1}{x_s^1} x_u$
4. Robust Predictors $T_{RB}(BT) = T_1 + b_n \sum_2 x_k$,
   $T_{RB}(HUM) = T_1 + b_n(HUM)\sum_2 x_k$, $T_{RB}(HUMG) = T_1 + b_n(HUMG)\sum_2 x_k$,
   $T_{RB}(CAM) = T_1 + b_n(CAM)\sum_2 x_k$, $T_{RB}(CAGM) = T_1 + b_n(CAGM)\sum_2 x_k$;
5. Chambers predictor, $T_{CHAM}(BT) = T_1 + b_n(BT)\sum_2 x_k + \dfrac{\sigma}{n}\sum_2 x_k\left[\sum_1 \psi_c(t)\right]$,
   $T_{CHAM}(CAM) = T_1 + b_n(CAM)\sum_2 x_k + \dfrac{\sigma}{n}\sum_2 x_k\left[\sum_1 \psi_c(t)\right]$

It has been observed that by adopting the biweight reiterative algorithm, the above procedure resulted into the convergence of b_n, only after second to third iteration. This observation falls in agreement to the concept of one step M-estimator procedure. The various parametric values for population 1 and population 2 are summarized in table 5.

Table 4.1. Summary of Population Characteristics

	N	Y_u	X_u	β_u	Outliers
Population 1.	34	7426	26022	0.285374	1
Population 2.	235	4713	7463	0.6315	13

The performance of the predictor has been judged on the basis of relative bias (RB) = {average bias over large number of samples/ true value(T)} × 100, and relative coefficient of variation (RCV)= {square root of mean square error of the predictor over large number of samples/T} × 100. The simulation results for the suggested resistant alternatives with the existing conventional and those model based are summarized in the table 4.2 and table 4.3.

Table 4.2. Simulation results

	$\beta_n(LS)$					$\beta_n(BT)$				
n	$E(\beta_n)$	Bias	RB	MSE	CV	$E(\beta_n)$	bias	RB	MSE	CV
4	.29789	.012	4.39	.006	27.45	.28124	–.004	–1.45	.003	19.51
8	.28657	.001	.43	.001	13.29	.27431	–.01	–3.88	.0009	10.43
12	.27955	–.006	–2.04	.0007	9.68	.27007	–.02	–5.3	.0006	8.89
20	.62684	–.005	–.74	.004	10.59	.63611	.004	.73	.008	14.43
30	.62680	–.005	–.75	.003	8.08	.64149	.01	1.58	.005	10.83
40	.63038	–.001	–.179	.002	6.82	.64147	.01	1.58	.005	11.39
	$\beta_n(HUB)$					$\beta_n(CAU)$				
n	$E(\beta_n)$	Bias	RB	MSE	CV	$E(\beta_n)$	bias	RB	MSE	CV
4	.02805	–.005	–1.70	.003	19.66	.297852	.01	4.37	.007	28.81
8	.27638	–.009	–3.15	.0008	9.66	.286705	.001	.47	.001	12.98
12	.27217	–.01	–4.62	.0006	8.74	.280722	–.005	–1.6	.0007	9.56
20	.64066	.009	1.45	.005	11.37	.647632	.02	2.55	.007	13.53
30	.64047	.009	1.41	.003	8.16	.646652	.02	2.40	.004	10.48
40	.64698	.02	2.45	.003	8.40	.653527	.02	3.49	.003	9.25

The table reveals the conventional way of outlier treatments (RAE, WINS) results into higher negative bias and increased coefficient of variation, comparing T_n (*WINS*) and T_n (*RAE*) for both the populations and for all sample sizes. This suggests the need to use the model based robust predictors.

Table 4.3. Simulation Results

N	T_n *(RAE)*		T_n *(WINS)*		T_n *(RAE)*		$T_n^{(5)}$ *(HUM]*		$T_n^{(5)}$ *(HUGM)*	
	RB	CV	RB	CV	RB	CV	RB	CV	RB	CV
4	1.38	22.39	–6.97	26.42	–2.75	15.55	–3.48	14.65	–3.03	15.59
8	–1.73	14.03	–3.46	16.82	–4.86	11.55	–4.53	11.35	–4.38	11.80
12	1.75	10.67	0.27	11.07	–2.42	7.37	1.83	7.73	–1.41	7.93
20	–1.58	11.50	–2.63	12.74	–1.67	13.50	0.04	11.02	0.075	10.59
30	–.49	8.20	–2.21	9.40	–4.19	11.76	1.03	7.23	0.76	7.55
40	–.85	7.58	–2.05	8.18	–2.47	9.11	1.31	7.05	0.83	6.90

	$T_n^{(5)}$ *(CAU)*		$T_n^{(5)}$ *(CAUGM)*		$T_n^{(5)}$ *(BT)*		$T_n^{(5)}$ *(CAU)*	
n	RB	CV	RB	CV	RB	CV	RB	CV
4	1.36	22.36	1.37	22.38	–2.23	16.59	0.77	21.26
8	–1.77	13.98	–1.74	14.01	–4.71	11.47	–2.68	12.84
12	1.69	10.60	1.73	10.65	–2.38	7.36	0.15	8.89
20	–1.23	11.33	–1.36	11.25	–1.56	13.67	–0.22	10.91
30	–0.05	7.88	–0.24	7.87	–4.09	11.89	0.60	7.45
40	–0.26	7.38	–.41	7.33	–2.46	9.20	0.38	7.03

The M-predictors (HUM and CAU) generally have smaller coefficient of variation as compared to GM-predictors (HUGM and CAGM) for population 1. Therefore, the use of down-weight functions for x results in loss in efficiency. On the contrary the *CV* of GM-predictors are small for population 2.

This has intuitive justification in case of population 2 since this contains 13 outliers in x's and GM predictors downweight large x's by suitable choice of downweight functions $w(x)$.

Practically, $T_n^{(5)}$ *(BT)* and $T_n^{(5)}$ *(HUM)* perform equally well whereas $T_n^{(5)}$ (*CAU*) is better in terms of bias at the cost of high coefficient of variation, in case of population 1. This confirms the non-demanding use of typical redescending type- $\psi(\cdot)$ function, when there is only one outlier in residuals, results in loss in efficiency. This is the reason why $T_n^{(5)}$ *(CAUGM)* and $T_n^{(5)}$ *(HUGM)* perform better as compared to $T_n^{(5)}$ (BT) in terms of RB and CV in case of population 2, where the use of such predictors are justified. $T_n^{(6)}(\cdot)$ the Chambers predictors do not gain substantially over $T_n^{(5)}(\cdot)$. It has been shown that Chambers predictors are one-step $T_n^{(5)}(\cdot)$ predictors. The robust estimator of $\beta(F)$ in the model G_R uses *M*-estimation procedure by adopting iteratively reweighted least square algorithm (Beaton and Tukey, 1974). Therefore, $T_n^{(5)}(\cdot)$ predictors effectively converge to $T_n^{(6)}(\cdot)$ predictors. That is why the performance of these predictors are usually so similar. In these simulation studies biases are not the important part of the mean squared errors. We note small bias even if the sampled outliers are given weight close to 1.

5.1.Conditional bias analysis

Large biases of $T_n^{(5)}$ estimators as compared to ratio estimator, RAE is observed in the simulation results, this is somewhat deceptive. The relative conditional biases of the resistant predictors have been calculated in the following table. Samples of size n = 30 are drawn from the population 2, conditioned on number of outliers sampled is 0, 1, 2, 3, 4, 5 respectively. Comparison figures RB and CV have been calculated for each of conditional situation.

Tabl 5.1. Absolute Relative Conditional Bias for of size 30 drawn from population 2. The unconditional probabilities of getting 0, 1, 2, 3, 4, and 5 outliers are given by 0.16, 0.32, 0.29, 0.15, 0.05, and 0.01.

	Outliers	*(RAE)*	*(WINS)*	*(BT)*	*(HUM)*	*(HUGM)*
0	RB	8.37	6.68	3.92	12.33	12.51
	CV	(10.38)	(9.96)	(12.14)	(14.17)	(14.11)
1	RB	3.31	3.36	3.72	8.68	8.04
	CV	(7.53)	(8.55)	(11.55)	(11.54)	(10.99)
2	RB	–2.00	–3.08	0.99	3.55	2.59
	CV	(6.79)	(7.76)	(11.09)	(8.52)	(7.79)
3	RB	–5.77	–7.61	–0.65	–0.45	–1.53
	CV	(9.12)	(1.99)	(14.62)	(8.71)	(7.99)
4	RB	–11.37	–12.81	–3.34	–7.63	7.89
	CV	(13.39)	(15.16)	(13.75)	(11.83)	(11.44)
5	RB	–14.25	–15.68	–4.22	–11.95	10.82
	CV	(15.85)	(17.43)	(14.19)	(15.37)	(14.64)

	Outliers	CAU	CAGM	(CHM)BT	(CHM)CAU
0	RB	12.51	12.50	3.98	10.29
	CV	(14.21)	(14.08)	(11.42)	(12.20)
1	RB	6.85	6.59	3.48	6.31
	CV	(10.57)	(10.33)	(10.91)	(9.46)
2	RB	0.57	0.34	0.66	1.03
	CV	(8.15)	(7.96)	(10.44)	(7.01)
3	RB	–4.00	–4.25	–1.05	–2.98
	CV	(9.90)	(9.84)	(13.79)	(8.38)
4	RB	–10.82	–10.96	–4.01	–9.29
	CV	(11.01)	(14.02)	(13.21)	(12.19)
5	RB	–11.82	–14.64	–4.73	–12.83
	CV	(17.37)	(17.36)	(13.33)	(15.16)

The probabilities of getting 0, 1, 2, 3, 4, 5, outliers in the sample are 0.16, 0.32, 0.29, 0.15, 0.05 and 0.01 respectively. Therefore, expected number of outliers in a sample of size n = 30 would be approximately 2. Occurrences of more than 5 outliers in the sample is too low. The table 5.1 shows large biases of ratio and WINS estimators as compared to their resistant competitors, when too many outliers are sampled. In case when only two outliers are sampled, the conditional biases of $T_n^{(5)}$ predictors and the difference of biases grow rapidly as number of outliers sampled increases. Therefore, from the conditional point of view, ratio and WINS are most biased predictors.

REFERENCES

CASSEL, C. M. and SARNDAL C. E. (1977). Foundation of Inference in Survey Sampling, New York : John Wiley.

CHAMBERS, R. L. (1986). Outlier Robust Finite Population Estimation. *Journal of the American Statistical Association,* **81,** 1063–1069.

ERNST, L.R(1980). Comparison of estimator of the Mean which Adjust for large Observations, *'Sankhya, Ser.C.***42***,* 1–16.

FULLAR, W.A.(1991). Simple estimators of the mean of Skewed Populations, *Statistica Sinica,* **1**, 137–158.

GLASSER, G.J.(1962). On the complete coverage of large units in a statistical study. *International Statistical Review,* **30,** 28–32.

GWET, J. P. and RIVEST, L.P. (1992). Outlier resistant alternatives to the ratio estimator. *Journal of the American Statistical Association,* **87**, 1174–1182.

HAMPEL, F.R., RONCHETTI, E.M., ROUSSEEUW, P.J. and STAHEL, W.E. (1986): *Robust Statistics: The Approach Baaed on Influence Functions,* New york: John Wiley.

HIDIROGIOU, M. H., and SRINATH, K. P. (1981): Some Estimators of the Population Total from simple Random Samples Containing Large Units, *Journal of the American Statistical Association,* **76**, 690–695.

HULLIGER, B. (1995): Outlier Robust Horvitz-Thompson Estimators. Survey Methodology 21(1), 79–87.

ICHAN, R.(1984): Sampling strategies, robustness and efficiency: the state of art. *International Statistical Review,* **52**, 209–218.

KISH, L. (1965). *Survey Sampling,* New York : John Wiley.

RAO, J.N.K (1985). Conditional Inferences in Survey Sampling, *Survey Methodology,* **11**, 15–31.

RIVEST, L.P. (1993). Winsorization of Survey Data, *Proceedings of the 49th Session, International Statistical Institute*.

SCOTT, A. J., and BREWER, K. R. W., and HO, E. W. H. (1978). Finite Population Sampling and Robust Estimation, *Journal of the American Statistical Association,* **73** , 359–361.

SEARLS, D.T. (1966). An estimator for a population mean which reduces the effect of large observations, *Journal of the. American Statistical Association,* **61,** 1200–1204.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics,* New York : John Wiley.

SHOEMAKER, L.H. and ROSENBERGER, J.L. (1983). Moments and efficiency of the median and trimmed mean for finite population, *Communication in Statistics, Simulations and computation,* **12(4)**, 411–422.

SMITH, T.M.F. (1987). Influential observation in survey sampling, *Journal of Applied Statistics,* **14**,143–152.

SUKHATME, P. V., and SUKHATME, B. V. and ASOK, C. (1984). *Sampling Theory of Surveys With Applications (2nd ed.),* Rome: Food and Agriculture: Organization.

SARNDAL, C. E. and SWENSSON, J. W.(1991). *Model Assisted Survey Sampling,* Springer-Verlag, New York.