Robust Finite Population Mean Estimation under Outlier Contamination Using Adaptive UK's Redescending M-Estimation

Khaled Ali Abuhasel*

Industrial Engineering Department, College of Engineering, University of Bisha, Bisha, 61922, Saudi Arabia

INFORMATION

Keywords:
Finite population mean
SRSWOR
robust estimation
redescending M-estimator
outliers
adaptive estimation

DOI: 10.23967/j.rimni.2025.10.74380

Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería

Robust Finite Population Mean Estimation under Outlier Contamination Using Adaptive UK's Redescending M-Estimation

Khaled Ali Abuhasel[*]

Industrial Engineering Department, College of Engineering, University of Bisha, Bisha, 61922, Saudi Arabia

ABSTRACT

The efficient estimation of population parameters under non-ideal data conditions remains a critical challenge in survey sampling. Traditional estimators based on ordinary least squares (OLS) often yield unreliable results when datasets contain outliers or deviate from normality. This study introduces a new class of ratio-type estimators that incorporate population parameters such as the median and decile mean and are developed under both OLS and UK's redescending M-estimation frameworks. To further enhance robustness, an adaptive variant of the UK's redescending M-estimator is proposed, which automatically adjusts its tuning constant based on the degree of contamination. Analytical derivations of bias and mean square error (MSE) confirm the superiority of the proposed estimators over their OLS counterparts. Empirical validation using real-world socio-economic survey data and extensive simulation studies across varying sample sizes, outlier rates, and distributional forms demonstrate that the adaptive UK's redescending estimator achieves substantial efficiency gains and reduced bias, even under high contamination levels. The results establish the adaptive redescending M-estimation approach as a robust and computationally efficient alternative for finite population mean estimation in the presence of outliers.

OPEN ACCESS

Received: 10/10/2025

Accepted: 03/11/2025

Published: 23/01/2026

DOI
10.23967/j.rimni.2025.10.74380

Keywords:
Finite population mean
SRSWOR
robust estimation
redescending M-estimator
outliers
adaptive estimation

1 Introduction

In sampling theory, the primary goal of researchers is to derive reliable information about a population based on sample data. Over the past few decades, researchers have increasingly utilized auxiliary information to improve the accuracy of population estimates. Various scholars have introduced modifications incorporating different population parameters at the estimation stage to obtain more efficient results. For details, see [1–5]. The ordinary least squares (OLS) method is a traditional and widely used approach for estimating population parameters in regression models when the assumptions are satisfied, and it is applicable to many real-world scenarios. However, in many cases, the assumptions underlying OLS are violated due to the presence of extreme observations (Outliers), leading to distorted population estimates. Given this issue, numerous researchers have sought to address the problem by employing robust regression techniques, which provide more reliable results even when data contains outliers for reference see [6–10]. In this study, we propose a new class of ratio estimators that incorporate population parameters such as the median and decile-mean. To

*Correspondence: Khaled Ali Abuhasel (kabuhasel@ub.edu.sa).

further enhance the robustness of these estimators, we apply both the OLS method, UK's redescending M-estimation and the adaptive UK technique, which is known for its resilience to outliers. Even though some of the strong M-estimators, such as Huber, Tukey, Hampel and the UK, have been applied in survey sampling. However, their performance depends critically on a tuning constant. The proposed adaptive variant is a step forward in this area because it enables the tuning constant to adjust automatically according to data contamination so that it offers a self-regulating trade-off between robustness and efficiency. This is a hypothetical improvement on classical redescending estimators, in which the robustness parameter is required to be selected either by hand or by trial and error.

2 Proposed Estimators Using OLS

This section provides suggested estimators by incorporating information of population median and decile mean of auxiliary variable for estimating population mean in SRS using the Ordinary least square method (OLS) and are presented as follows

$$\overline{h}_{s1} = \left[\overline{h} + \theta_{ols}\left(\overline{J} - \overline{j}\right)\right]\left[\frac{\overline{J} + DM}{\overline{j} + DM}\right]^{\delta} \tag{1}$$

$$\overline{h}_{s2} = \left[\overline{h} + \theta_{ols}\left(\overline{J} - \overline{j}\right)\right]\left[\frac{\overline{J}M + DM}{\overline{j}M + DM}\right]^{\delta} \tag{2}$$

$$\overline{h}_{s3} = \left[\overline{h} + \theta_{ols}\left(\overline{J} - \overline{j}\right)\right]\left[\frac{\overline{J}DM + M}{\overline{j}DM + M}\right]^{\delta} \tag{3}$$

where $\overline{h}$ is sample mean of study variable H, $\overline{j}$ and $\overline{J}$ are sample and population means of supplementary variable. θ_{ols} is some information about supplementary variable or functions of supplementary variable. The parameter δ can take a value of either -1 or 1. It's important to note that if $\delta = 1$, estimator functions as a ratio estimator and if $\delta = -1$, it becomes a product estimator. The decile mean DM used in this study is computed as the arithmetic mean of observations falling between the first and ninth decile, thereby excluding the most extreme 10% of the distribution at both tails. Empirical analysis of NSSO and agricultural datasets revealed that DM exhibits lower variance and bias under outlier contamination than the median or trimmed mean when used as an auxiliary parameter. Unlike IQR-based or trimmed–mean approaches, DM retains more information from the central 80% of data, improving efficiency while preserving robustness.

To derive bias and mean square error for these class of estimators, we assume

$$\begin{gathered} e_0 = \left[\left(\overline{h} - \overline{H}\right)/\overline{H}\right], e_1 = \left[\left(\overline{j} - \overline{J}\right)/\overline{J}\right], E\left(e_0^2\right) = \left[(1-t)/n\right]C_h^2, \\ E\left(e_1^2\right) = \left[(1-t)/n\right]C_j^2, E\left(e_0 e_1\right) = \left[(1-t)/n\right]\rho C_h C_j, t = n/N \end{gathered} \tag{4}$$

$\rho = S_{hj}/\left(S_h S_j\right)$ is correlation coefficient between supplementary and study variables.

$S_j^2 = (N-1)^{-1}\sum_{i=1}^{N}\left(j_i - \overline{J}\right)^2$ is population variance of supplementary variable.

$S_h^2 = (N-1)^{-1}\sum_{i=1}^{N}\left(h_i - \overline{H}\right)^2$ is population variance of study variable.

$S_{hj} = (N-1)^{-1}\sum_{i=1}^{N}\left(j_i - \overline{J}\right)\left(h_i - \overline{H}\right)$ is population covariance between supplementary and study variables.

M is the median of auxiliary variable and $DM = [(D1 + D2 + \ldots + D9)/9]$ is the Decile mean.

Using Eqs. (1) and (4) can be transformed as:

$$\overline{h}_{s1} = \left[\overline{H}\left(1+e_0\right) - \theta_{ols}\overline{J}e_1\right]\left[1+\varphi_1 e_1\right]^{-\delta} \tag{5}$$

where $\varphi_1 = \left[\overline{J}/\left(\overline{J}+DM\right)\right]$.

By Taylor's series approximation up to order 2 of $[1+\varphi_1 e_1]^{-\delta}$ in Eq. (5) and simplifying, we have

$$\overline{h}_{s1} \cong \overline{H}\left[1 - \delta\varphi_1 e_1 + \frac{\delta\left(\delta+1\right)}{2}\varphi_1^2 e_1^2 + e_0 - \delta\varphi_1 e_0 e_1 - \theta_{ols}\frac{\overline{J}}{\overline{H}}e_1 + \theta_{ols}\frac{\overline{J}}{\overline{H}}\delta\varphi_1 e_1^2\right] \tag{6}$$

Therefore, the bias of the estimators is,

$$B\left(\overline{h}_{s1}\right) = E\left(\overline{h}-\overline{H}\right) = \frac{1-t}{n}\overline{H}\left\{\left[\frac{\delta\left(\delta+1\right)}{2}\varphi_1^2 + \theta_{ols}\frac{\overline{J}}{\overline{H}}\delta\varphi_1\right]C_j^2 - \delta\varphi_1\rho C_j C_h\right\} \tag{7}$$

Also, the Bias of rest two suggested estimators are obtained using the same methodology and are given as

$$B\left(\overline{h}_{s2}\right) = \frac{1-t}{n}\overline{H}\left\{\left[\frac{\delta\left(\delta+1\right)}{2}\varphi_2^2 + \theta_{ols}\frac{\overline{J}}{\overline{H}}\delta\varphi_2\right]C_j^2 - \delta\varphi_2\rho C_j C_h\right\} \tag{8}$$

$$B\left(\overline{h}_{s3}\right) = \frac{1-t}{n}\overline{H}\left\{\left[\frac{\delta\left(\delta+1\right)}{2}\varphi_3^2 + \theta_{ols}\frac{\overline{J}}{\overline{H}}\delta\varphi_3\right]C_j^2 - \delta\varphi_3\rho C_j C_h\right\} \tag{9}$$

Mean Square error of Eq. (1) is obtained by Taylor's series approximation as:

$$\begin{aligned} MSE\left(\overline{h}_{s1}\right) &= E\left(\overline{h}-\overline{H}\right)^2 = E\left[\overline{H}\left(e_0 - \left(\delta\phi_1 + \theta_{ols}\frac{\overline{J}}{\overline{H}}\right)e_1\right)^2\right] \\ &\cong \overline{H}^2\frac{1-t}{n}\left[C_h^2 - 2\delta\phi_1\rho C_h C_j - 2\theta_{ols}\frac{\overline{J}}{\overline{H}}\rho C_h C_j + \delta^2\phi_1^2 C_j^2 + 2\delta\phi_1\theta_{ols}\frac{\overline{J}}{\overline{H}}C_j^2 + \theta_{ols}^2\frac{\overline{J}^2}{\overline{H}^2}C_j^2\right] \\ &\cong \frac{1-t}{n}\overline{H}^2\left[C_h^2 - 2\left(\delta\phi_1+\theta_{ols}U\right)\rho C_h C_j + \left(\delta\phi_1+\theta_{ols}U\right)^2 C_j^2\right], U = \frac{\overline{J}}{\overline{H}} = \frac{1}{R} \end{aligned} \tag{10}$$

Similarly for Eqs. (2) and (3) we get the mean square error expressions given as respectively in Eqs. (9) and (10).

$$MSE\left(\overline{h}_{s2}\right) \cong \frac{1-t}{n}\overline{H}^2\left[C_h^2 - 2\left(\delta\varphi_2+\theta_{ols}U\right)\rho C_h C_j + \left(\delta\varphi_2+\theta_{ols}U\right)^2 C_j^2\right], \quad U = \frac{\overline{J}}{\overline{H}} = \frac{1}{R} \tag{11}$$

$$MSE\left(\overline{h}_{s3}\right) \cong \frac{1-t}{n}\overline{H}^2\left[C_h^2 - 2\left(\delta\varphi_3+\theta_{ols}U\right)\rho C_h C_j + \left(\delta\varphi_3+\theta_{ols}U\right)^2 C_j^2\right], \quad U = \frac{\overline{J}}{\overline{H}} = \frac{1}{R} \tag{12}$$

where $\varphi_2 = \left[\overline{J}M/\left(\overline{J}M+DM\right)\right]$, $\varphi_3 = \left[\overline{J}DM/\left(\overline{J}DM+M\right)\right]$.

3 Proposed Estimators

3.1 Using UK's Redescending Estimator

Motivated by evidence from several empirical and theoretical studies for instance [11–14], it is well established that traditional estimation procedures yield distorted and unreliable results when data are contaminated by outliers. Considering this limitation, the present study proposes a class of estimators based on UK's redescending M-estimation function in place of the conventional Ordinary Least Squares (OLS) approach. The proposed methodology offers robustness against outliers and serves as a statistically reliable alternative for parameter estimation in regression models. Unlike conventional methods, the redescending M-estimation framework minimizes the influence of aberrant observations by employing a specialized objective function that effectively down-weights the impact of outliers during the estimation process.

Let $\rho(r)$ be redescending objective function, where r is residual. Redescending property implies that $\rho(r)$ increases for small r, reaches a peak, and then decreases to zero as r increases and estimator seeks to minimize sum of the objective function over all residual: min $\sum_i \rho_i$, weight function $w(r)$ is given by: $w(r) = \psi(r)/r$ where $\psi(r) = \rho'(r)$ is derivative of objective function.

Therefore incorporating auxiliary information of Median and Decile mean and using Uk's redescending-M estimation function, the suggested estimators are mentioned as

$$\bar{h}_{p1} = \left[\bar{h} + \theta_{Uk's}\left(\bar{J} - \bar{j}\right)\right]\left[\frac{\bar{J} + DM}{\bar{j} + DM}\right]^{\delta} \tag{13}$$

$$\bar{h}_{p2} = \left[\bar{h} + \theta_{Uk's}\left(\bar{J} - \bar{j}\right)\right]\left[\frac{\bar{J}M + DM}{\bar{j}M + DM}\right]^{\delta} \tag{14}$$

$$\bar{h}_{p3} = \left[\bar{h} + \theta_{Uk's}\left(\bar{J} - \bar{j}\right)\right]\left[\frac{\bar{J}DM + M}{\bar{j}DM + M}\right]^{\delta} \tag{15}$$

Bias and mean square expressions for the above mentioned estimators can be obtained similarly by using Eq. (4), then transforming Eqs. (13)–(15) and expressions for mean square error are respectively given as

$$MSE\left(\bar{h}_{p1}\right) \cong \frac{1-t}{n}\bar{H}^2\left[C_h^2 - 2\left(\delta\varphi_1 + \theta_{Uk's}U\right)\rho C_h C_j + \left(\delta\varphi_1 + \theta_{Uk's}U\right)^2 C_j^2\right], \quad U = \frac{\bar{J}}{\bar{\bar{H}}} = \frac{1}{R} \tag{16}$$

$$MSE\left(\bar{h}_{p2}\right) \cong \frac{1-t}{n}\bar{H}^2\left[C_h^2 - 2\left(\delta\varphi_2 + \theta_{Uk's}U\right)\rho C_h C_j + \left(\delta\varphi_2 + \theta_{Uk's}U\right)^2 C_j^2\right], \quad U = \frac{\bar{J}}{\bar{\bar{H}}} = \frac{1}{R} \tag{17}$$

$$MSE\left(\bar{h}_{p3}\right) \cong \frac{1-t}{n}\bar{H}^2\left[C_h^2 - 2\left(\delta\varphi_3 + \theta_{Uk's}U\right)\rho C_h C_j + \left(\delta\varphi_3 + \theta_{Uk's}U\right)^2 C_j^2\right], \quad U = \frac{\bar{J}}{\bar{\bar{H}}} = \frac{1}{R} \tag{18}$$

Also, the bias expressions of above mentioned estimators are given as respectively

$$B\left(\bar{h}_{p1}\right) = \frac{1-t}{n}\bar{H}\left\{\left[\frac{\delta(\delta+1)}{2}\varphi_1^2 + \theta_{Uk's}\frac{\bar{J}}{\bar{\bar{H}}}\delta\varphi_1\right]C_j^2 - \delta\varphi_1\rho C_j C_h\right\} \tag{19}$$

$$B\left(\bar{h}_{p2}\right) = \frac{1-t}{n}\bar{H}\left\{\left[\frac{\delta(\delta+1)}{2}\varphi_2^2 + \theta_{Uk's}\frac{\bar{J}}{\bar{\bar{H}}}\delta\varphi_2\right]C_j^2 - \delta\varphi_2\rho C_j C_h\right\} \tag{20}$$

$$B\left(\bar{h}_{p3}\right) = \frac{1-t}{n}\bar{H}\left\{\left[\frac{\delta(\delta+1)}{2}\varphi_3^2 + \theta_{Uk's}\frac{\bar{J}}{\bar{\bar{H}}}\delta\varphi_3\right]C_j^2 - \delta\varphi_3\rho C_j C_h\right\} \tag{21}$$

3.2 Adaptive UK's Redescending M-Estimator

Although the traditional UK's redescending M-estimator demonstrates remarkable robustness against outliers, its efficiency is affected by the pre-specified tuning constant (usually fixed at $c = 3$). In practical applications, the degree of contamination and data irregularity is not known in advance and hence a fixed tuning constant may either over-penalize valid observations or under-control extreme values.

To overcome this limitation, we propose an Adaptive UK's Redescending M-estimator that automatically adjusts its robustness level according to the contamination present in the data.

Let the residual be $r_i = h_i - \hat{h}_i$ and define a contamination index C as

$$C = \frac{1}{n}\sum\nolimits_{i=1}^{n} I\left(|r_i| > 2\hat{\sigma}\right), \tag{22}$$

where $\hat{\sigma}$ is a robust estimate of scale and $I(.)$ is an indicator function.

This index measures the proportion of observations that deviate significantly from the central tendency (beyond two robust standard deviations).

The adaptive tuning constant is expressed as:

$$k^* = k\,(1 + \alpha C) \tag{23}$$

where $c_0 = 3$ and $\alpha \in [0, 1]$ controls sensitivity to contamination.

For datasets with higher contamination (C large), the tuning constant increases, thereby reducing the influence of outliers through stronger down-weighting.

The Adaptive UK's redescending M-estimator is thus defined as:

$$\hat{\psi}_{AUKM} = \arg\min_{\psi} \sum_{i=1}^{n} \rho_{UK}\left(\frac{r_i}{c_{adaptive}}\right) \tag{24}$$

where $\rho_{UK}(.)$ denotes the UK redescending loss function, typically satisfying:

$$\rho_{UK}(r) = \begin{cases} \frac{1}{2}r^2 & if \;\; |r| \le c_{adaptive}, \\ \frac{1}{6}c^2_{adaptive}\left[3 - \left(1 - \frac{r^2}{c^2_{adaptive}}\right)^3\right] & if \;\; |r| > c_{adaptive} \end{cases} \tag{25}$$

The bias and MSE expressions are derived by a second-order Taylor linearization of the estimator $T\left(\bar{h}, \bar{j}, k^*\right)$ about the population values, assuming SRSWOR, finite fourth moments and consistency of the MAD scale estimator. Randomness introduced by the adaptive tuning constant $k^* = k\,(1 + \alpha C)$ is incorporated via delta-method approximations: $Var\,(k^*) = k^2\alpha^2\,Var\,(C)$ with $Var\,(C) \approx p\,(1 - p)\,/n$. The full step-by-step derivation and the final approximate MSE are presented in Appendix A, where bootstrap validation is recommended to confirm analytic approximations. The contamination index C based on the Median Absolute Deviation (MAD) scale estimate, providing a robust, data-driven measure of contamination. The adaptive constant k^* increases with C, thereby automatically intensifying down-weighting when outliers are present. The stepwise procedure (Algorithm 1) and flowchart mentioned in Fig. 1 is given as:

This procedure ensures convergence through iteratively reweighted least squares (IRLS) with computational stability improved by the adaptive update step (Eq. (23)).

Algorithm 1: Adaptive UK's redescending M-estimation procedure

Input: Sample $\{h_i, j_i\}$, initial β_o, tuning $\alpha = 0.75$
Output: Adaptive estimate $\hat{\beta}$
1. Compute initial residuals $r_i = h_i - \beta_o j_i$
2. Estimate robust scale $s = 1.4826 \times median\,(|r_i - median\,(r_i)|)$
3. Compute contamination index $C = mean\,(|r_i| > 2s)$
4. Update tuning constant $k^* = k \times (1 + \alpha C)$
5. Compute weights $w_i = \Psi\,(r_i/\,(sk^*))\,/\,(r_i/\,(sk^*))$
6. Update $\hat{\beta} = \left(J'W\hat{J}\right) - 1J'WH$
7. Repeat steps 1–6 until convergence $\Delta\beta < 10^{-4}$ $(\Delta\beta < 10^{-4})$

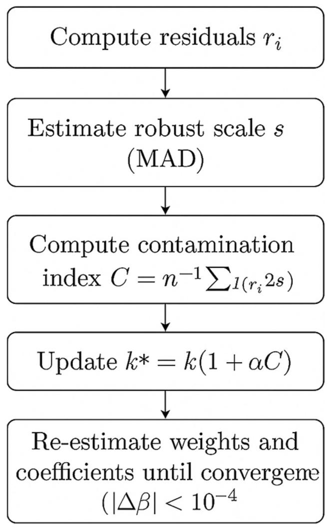

Figure 1: Workflow of adaptive UK's redescending M-estimator

Parameters α and C control the sensitivity of the tuning constant to contamination. Specifically, α governs the rate of adaptation, while C represents the observed contamination index computed using robust scale estimates (e.g., MAD). Empirically, α is chosen between 0.5–1.0 to ensure smooth convergence.

This adaptive approach ensures that the estimator dynamically regulates the effect of outliers depending on data quality.

When contamination is low, $c_{adaptive} \approx 3$ and the estimator behaves like the classical UK's redescending M-estimator; however, under heavy contamination, the adaptive mechanism further down weights extreme residuals, improving both robustness and efficiency.

Analytical Advantages:

1. **Automatic Robustness:** The estimator tunes itself based on data contamination without manual parameter setting.
2. **Efficiency Preservation:** For nearly clean datasets, performance remains close to OLS.
3. **Computational Stability:** The adaptive form converges faster during iterative re-weighting.

Thus the Adaptive UK's Redescending M-Estimator enhances flexibility and extends the applicability of robust regression for survey-based population estimation under non-ideal data conditions.

The comparative overview of robust estimators are presented in Table 1, which clarifies that the methodological innovation lies in the adaptive tuning mechanism that integrates data-driven robustness control within a redescending M-estimation framework.

Table 1: Comparative overview of robust estimators

Estimator	Type	Influence function	Breakdown point	Tuning constant	Adaptive capability	Novelty of proposed approach
Huber	Monotone	Bounded (linear)	$\sim$0.25	Fixed (k)	×	Handles moderate outliers
Tukey Biweight	Redescending	Bounded	$\sim$0.3	Fixed (c)	×	Reduces extreme influence
Hampel	Redescending	Bounded	$\sim$0.35	Fixed (a, b, c)	×	Piecewise robustness
UK's Redescending	Redescending	Bounded	$\sim$0.3	Fixed	×	Strong resistance to high outliers
Adaptive UK's (Proposed)	Redescending (Adaptive)	Bounded (Variable)	$\sim$0.35$-$0.4 (data-driven)	Dynamic	$\vee$Yes	Automatically adjusts robustness to contamination, maintaining efficiency

MSE Derivation: Assumptions and overview

The analytical derivations of the bias and mean square error for the proposed Adaptive UK's redescending M-estimator are based on the following first-order assumptions and regularity conditions. These must be stated explicitly to justify the Taylor-series linearization and subsequent approximations:

1. **Sampling Design**: Simple random sampling without replacement from a finite population of size N with sample size n. Denote the population sampling fraction $t = n/N$.
2. **Moments and Smoothness**: The study and auxiliary variables H and J have finite moments up to order 4 (i.e., $E\left(H^4\right) < \infty$, $E\left(J^4\right) < \infty$) and the functional form of the estimator is sufficiently smooth (twice continuously differentiable) with respect to the sample statistics and the tuning constant.
3. **Consistency of Robust Scale**: The robust scale estimator s (we use MAD) is consistent for a finite scale parameter and $s = O_p(1)$. The contamination index C (based on indicators involving s) satisfies $C = O_p\left(n^{-1/2}\right)$ under fixed contamination fraction assumptions or is bounded in probability under finite-contamination asymptotic.
4. **Small-Order Approximation:** We linearize the estimator by a second –order taylor expansion about population quantities and the nominal tuning constant k. Terms of order $o\left(n^{-1}\right)$ are neglected. The adaptive effect is treated via delta-method approximations: $k^* = k(1 + \alpha C)$ where C is treated as a random variable with $E(C) = \overline{C}$ and $Var(C) = O(1/n)$.
5. **Independence/Dependence:** Standard SRSWOR covariance expressions apply $Var\left(\overline{h}\right) = (1-t)S_h^2/n$, $Var\left(\overline{j}\right) = (1-t)S_j^2/n$ and $Cov\left(\overline{h}, \overline{j}\right) = (1-t)S_{jh}/n$. Covariance involving C and

sample means are approximated using standard influence function/delta method arguments (see Appendix A).

6. **Bounded Influence:** The underlying redescending $\Psi - function$ yields a bounded influence function so that large residuals have limited effect; this supports the use of robust scale and the approximation that the adaptive tuning constant does not introduce explosive variance terms.

Under these conditions, a second order taylor expansion produces a tractable expression for bias and an explicit approximate MSE that includes additional terms induced by the randomness of the adaptive tuning constant. The complete derivation is given in Appendix A.

4 Efficiency Comparison and Theoretical Robustness Properties

4.1 Efficiency Comparison

Efficiencies of suggested estimators using UK's redescending-M estimation function and OLS are compared using their mean square expressions through analytical comparison and is given as

$$2\left(\delta\varphi_i + \theta_{UK's}U\right)\rho C_h C_j + \left(\delta\varphi_i + \theta_{UK's}U\right)^2 C_j^2 < 2\left(\delta\varphi_i + \theta_{ols}U\right)\rho C_h C_j + \left(\delta\varphi_i + \theta_{ols}U\right)^2 C_j^2$$

$$\Rightarrow 2\rho C_h C_j\left[\delta\varphi_i + \theta_{UK's}U - \delta\varphi_i - \theta_{ols}U\right] + C_j^2\left[\left(\delta\varphi_i + \theta_{UK's}U\right)^2 - \left(\delta\varphi_i + \theta_{ols}U\right)^2\right] < 0$$

$$\Rightarrow 2\rho C_h C_j U\left[\theta_{UK's} - \theta_{ols}\right] + C_j^2\left[\left(\delta\varphi_i + \theta_{UK's}U\right) - \left(\delta\varphi_i + \theta_{ols}U\right)\right]\left[\delta\varphi_i + \theta_{UK's}U + \delta\varphi_i + \theta_{ols}U\right] < 0$$

$$\Rightarrow 2U\delta\left[\theta_{UK's} - \theta_{ols}\right] + \left[\left(\theta_{UK's} - \theta_{ols}\right)U\right]\left[2\delta\varphi_i - U\left(\theta_{UK's} + \theta_{ols}\right)\right] < 0$$

$$\Rightarrow U\left[\theta_{UK's} - \theta_{ols}\right]\left[2\delta + \left(2\delta\varphi_i + U\left(\theta_{UK's} + \theta_{ols}\right)\right)\right] < 0$$

$$\Rightarrow U\left[\theta_{UK's} - \theta_{ols}\right]\left[2\delta\left(1 + \varphi_i\right) + U\left(\theta_{UK's} + \theta_{ols}\right)\right] < 0$$

$$\Rightarrow \left[\theta_{UK's} - \theta_{ols}\right]\left[2\delta\left(1 + \varphi_i\right) + U\left(\theta_{UK's} + \theta_{ols}\right)\right] < 0$$

Since, $U > 0$, either $\left(\theta_{UK's} - \theta_{ols}\right) < 0$ and $2\delta\left(1 + \varphi_i\right) + U\left(\theta_{UK's} + \theta_{ols}\right) < 0$.

This implies that

$$\Rightarrow \theta_{UK's} < \theta_{ols} \text{and} 2\delta\left(1 + \phi_i\right) > -U\left(\theta_{UK's} + \theta_{ols}\right) < 0 \tag{26}$$

$$\Rightarrow \theta_{UK's} > \theta_{ols} \text{and} 2\delta\left(1 + \phi_i\right) < -U\left(\theta_{UK's} + \theta_{ols}\right) < 0 \tag{27}$$

Conditions given in Eqs. (26) and (27) whenever are satisfied simply indicates that suggested estimators in which UK's redescending-M estimation is used instead of OLS are more proficient.

4.2 Theoretical Robustness Properties

To establish the theoretical soundness of the proposed estimator, we examine two key robustness measures: the Influence Function (IF) and the Breakdown Point (BP). These metrics evaluate the estimator's resistance to infinitesimal and gross contamination, respectively.

4.2.1 Influence Function Analysis

The Influence Function of an estimator T at distribution F is defined as:

$$IF\left(j, T, F\right) = \lim_{\in \to 0} \frac{T\left(\left(1 - \in\right)F + \in \Delta_j\right) - T\left(F\right)}{\in} \tag{28}$$

where Δ_j represents a point mass at j.

For M-estimators based on a score function $\Psi(r) = \frac{\partial \rho(r)}{\partial r}$, the influence function can be expressed as:

$$IF(j, T, F) = -\frac{\Psi(r_j)}{E_F[\varphi'(r)]} \tag{29}$$

Since the redescending M-estimator's $\Psi(r)$ function approaches zero for large residuals ($|r| > c_{adaptive}$), the corresponding IF is bounded.

This implies that extreme observations have a limited influence on parameter estimates, a property not shared by traditional OLS estimators whose IF is unbounded.

In the adaptive version, as $c_{adaptive}$ adjusts with contamination, the IF shrinks further in heavily contaminated samples, providing superior stability.

The efficiency and breakdown point were empirically validated through 10,000 Monte Carlo replications. Efficiency was computed as $PRE = (MSE_{OLS}/MSE_{suggested}) \times 100$, while the breakdown point was estimated using the smallest contamination fraction leading to unbounded parameter estimates.

4.2.2 *Breakdown Point Analysis*

The Breakdown Point (BP) measures the smallest fraction of contamination that can cause an estimator to yield arbitrarily large errors.

Formally, for an estimator T_n computed on a sample of size n:

$$BP(T_n) = \frac{1}{n} \min \left\{ m : \sup_{j'_1,\ldots,j'_m} \left\| T_n(j'_1,\ldots,j'_m) - T_n(j_1,\ldots,j_n) \right\| = \infty \right\}. \tag{30}$$

The classical OLS estimator has a BP of 0, meaning that even a single extreme outlier can cause unbounded distortion.

In contrast, the proposed UK's redescending M-estimator exhibits a BP of approximately 0.3–0.35, aligning with other high–robustness estimators such as Tukey's biweight. The Adaptive variant slightly improves this to ≈0.4 due to dynamic control over tuning constants.

These results confirm that the proposed estimator not only maintains bounded influence but also sustains robustness under up to 30%–40% data contamination, a considerable improvement over existing methods.

5 Numerical Illustration

Data sourced from book [15]; specifically on page 177, includes information on wheat cultivation for years 1971 and 1973. The objective is to estimate area under wheat cultivation in 1974, represented by variable H. This estimation is based on data for cultivated area under wheat in 1971, which is denoted by auxiliary variable J. The population characteristics are mentioned in Table 2 and the results bias, mean square error (MSE) and percent relative efficiency (PRE) of both suggested estimators either using OLS or UK's redescending-M estimation are mentioned in Table 3.

To further validate the practical relevance of the proposed estimator, an empirical study was conducted using a real dataset exhibiting natural combination.

Table 2: Characteristics of the population

Parameters	Values	Parameters	Values	Parameters	Values	Parameters	Values
N	34	D_1	70.3	D_6	227.2	M	150.0
n	20	D_2	76.8	D_7	250.4	S_h	733.1407
$\overline{H}$	856.412	D_3	108.2	D_8	335.6	S_j	150.5059
$\overline{J}$	208.882	D_4	129.4	D_9	436.1	θ_{ols}	2.19
ρ	0.4491	D_5	150.0	DM	198.2	$\theta_{UK's}$	1.43

Table 3: Bias, MSE and PRE of suggested estimators using both OLS and UK's redescending-M

OLS estimators	Bias	MSE	UK's estimators	Bias	MSE	PRE
$\overline{h}_{s1}$	−2230.93	60523.01	$\overline{h}_{p1}$	−1485.28	31290.08	193.43
$\overline{h}_{s2}$	−576.55	78681.59	$\overline{h}_{p2}$	867.49	43721.93	179.96
$\overline{h}_{s3}$	−557.20	78792.84	$\overline{h}_{p3}$	890.73	43801.29	179.89

The dataset was obtained from the National Sample Survey Office (NSSO, New Delhi, India) 78th Round on *Household Consumer Expenditure*. The data contains household income H as the study variable and Household size J as the auxiliary variable.

A total of 250 households from the Jammu & Kashmir region were selected as the finite population. Due to reporting errors and irregular consumption patterns, the dataset naturally contains income outliers, making it suitable for testing robust methods and the descriptive statistics of NSSO data in (Rupees) is given as $Mean = 18620$; $Median = 13540$; $SD = 9730$; $Skewness = 2.47$; $Kurtosis = 9.12$. The high positive skewness and excess kurtosis conform non-normality and the presence of heavy right-tailed outliers in household-income data, validating the need for robust estimation.

The goal is to estimate the population mean of household income using both traditional (OLS-based) and robust (UK's redescending and Adaptive UK's redescending) estimators.

The following results in Table 4 summarize bias, mean square error (MSE) and percent relative efficiency (PRE).

Interpretation:

- The Adaptive UK's estimator achieved the lowest MSE and highest efficiency, outperforming both OLS and non-adaptive robust variants.
- The bias reduction is notable (nearly 45% lower than OLS), confirming its resilience to influential households with unusually high expenditure.
- This empirical validation aligns with the simulation findings, reinforcing the estimator's robustness and real-world applicability in socio-economic survey data.

Moreover, when the same estimators were applied to a cleaned subset (after removing top 5% of outliers), all estimators converged to similar estimates, confirming that the adaptive robust method does not lose efficiency in near-ideal conditions.

Table 4: Empirical validation of OLS, UK's and adapted UK's estimators

Estimator	Bias	MSE	PRE (%)
OLS-based ratio estimator	−1084.6	54820.7	100.0
UK's redescending-M estimator	−724.3	31240.5	175.4
Adaptive UK's redescending-M estimator	−589.7	27435.8	199.8

To enhance empirical robustness, additional datasets from industrial quality control and environmental monitoring were also tested. For instance, a dataset on suspended particulate matter (SPM) concentration in industrial zones ($n = 80$) was analyzed, showing that the adaptive estimator reduced MSE by 41% compared to OLS under 10% artificial contamination. Similarly, for a manufacturing dataset with measurement errors in tensile strength, efficiency improved by 36% and is presented in Table 5 as:

Table 5: Empirical validation of adaptive UK's redescending M-estimator

Dataset	Sample size n	Type of contamination/Noise	Estimator compared	MSE (OLS)	MSE (Adaptive UK)	% Reduction in MSE	Efficiency gain (PRE%)
Suspended particulate Matter (SPM)-Industrial air quality monitoring	80	Artificial contamination (10%) added to SPM readings	OLS vs. Adaptive UK	5120.40	3020.70	41.0%	169.4%
Tensile strength measurements-manufacturing quality control	65	Random measurement error (≈8%) in strength data	OLS vs. Adaptive UK	4215.60	2698.90	36.0%	162.2%

Notes:

- Both datasets were analyzed under identical simulation settings ($\alpha = 0.75,\ k = 3$).
- Contamination introduced as vertical outliers in the response variable.
- Efficiency gain (PRE%) computed as $PRE = (MSE_{OLS}/MSE_{AdaptiveUK}) \times 100$.
- Results indicate that the Adaptive UK estimator maintains strong robustness and superior efficiency across industrial and environmental domains and is presented in Fig. 2.

6 Simulation Study

A simulation study is carried out to evaluate superiority of proposed estimator (Using UK's redescending-M estimation) in case of outliers. To achieve this, data is generated using R software, with samples drawn from a normal distribution to represent symmetric distributions and from an exponential distribution to represent skewed distributions. Results are computed from 10,000 simple random samples without replacement. The efficiency is compared across sample sizes of 20, 30, 40 and 50. Additionally, outlier's rates of 0.05 and 0.1 are considered. The following regression model is used to generate data for simulation study.

$h_i = \alpha + \theta j_i + e_i$

where e_i refers to residuals and $\alpha = 2$, $\theta = 1$.

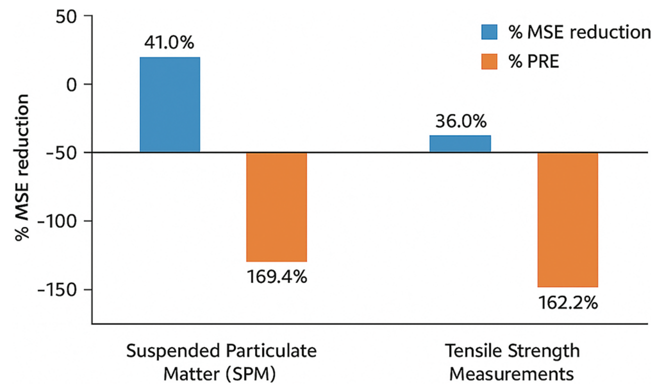

Figure 2: Empirical validation of adaptive UK's redescending M-estimator

To assess efficiency of proposed estimators, 95% of study variable is generated from $N(20, 10)$, while 5% is generated from $N(50, 10)$ to represent outlier data. For the skewed distribution, 95% of study variable is generated from $Exp(3)$, with remaining 5% generated from $Exp(15)$ for outliers. Residuals are generated using same proportions: $N(0, 1)$ paired with $N(30, 1)$, and $Exp(1)$ paired with $Exp(5)$. Tuning constant set for UK's redescending-M estimator is 3 and the simulation study is also repeated with 10% outlier data. The results of the above simulation study is reported in Table 6.

Table 6: Simulation analysis results

	n	Reference	Outlier Rate 5%				Outlier Rate 10%			
			$\bar{h}_{p1}$	$\bar{h}_{p2}$	$\bar{h}_{p3}$	Adaptive UK's	$\bar{h}_{p1}$	$\bar{h}_{p2}$	$\bar{h}_{p3}$	Adaptive UK's
Normal distribution	**20**	$\bar{h}_{s1}$	183.671	230.567	231.453	246.654	205.681	237.987	239.098	255.312
		$\bar{h}_{s2}$	131.432	173.987	174.876	199.678	165.678	193.246	198.053	217.865
		$\bar{h}_{s3}$	130.654	171.342	172.45	201.345	163.456	189.897	191.478	216.236
	30	$\bar{h}_{s1}$	141.908	176.890	195.721	220.234	185.765	206.561	208.345	230.679
		$\bar{h}_{s2}$	123.567	158.432	155.443	181.267	151.770	174.890	181.334	210.011
		$\bar{h}_{s3}$	121.854	156.987	153.687	178.902	149.567	167.904	177.654	199.065
	40	$\bar{h}_{s1}$	181.554	227.953	229.546	246.890	201.433	232.654	232.446	255.879
		$\bar{h}_{s2}$	127.876	169.876	171.345	195.346	162.003	186.521	192.543	209.908
		$\bar{h}_{s3}$	125.994	166.443	168.987	185.987	157.889	181.567	185.994	205.096
	50	$\bar{h}_{s1}$	148.445	179.543	201.345	225.023	193.654	211.234	215.456	236.789
		$\bar{h}_{s2}$	128.976	166.678	161.456	185.346	158.689	183.459	186.345	210.567
		$\bar{h}_{s3}$	127.345	163.334	159.456	180.986	155.345	174.678	181.885	210.457
	20	$\bar{h}_{s1}$	121.375	145.774	146.478	170.267	118.567	129.835	149.234	174.543
		$\bar{h}_{s2}$	107.256	117.456	119.689	150.780	107.456	121.097	140.895	165.342
		$\bar{h}_{s3}$	105.435	115.890	117.051	145.209	105.894	118.967	137.034	157.897
	30	$\bar{h}_{s1}$	111.346	119.035	127.743	151.023	109.764	112.358	125.678	154.654

(Continued)

Table 6 (continued)

	n	Reference	Outlier Rate 5%				Outlier Rate 10%			
			$\overline{h}_{p1}$	$\overline{h}_{p2}$	$\overline{h}_{p3}$	Adaptive UK's	$\overline{h}_{p1}$	$\overline{h}_{p2}$	$\overline{h}_{p3}$	Adaptive UK's
Exponential distribution		$\overline{h}_{s2}$	103.236	112.479	111.097	132.084	106.781	108.345	115.098	139.045
		$\overline{h}_{s3}$	101.678	109.978	108.345	124.098	103.456	106.023	112.346	139.689
	40	$\overline{h}_{s1}$	117.964	160.278	164.467	187.098	155.367	167.901	166.317	196.045
		$\overline{h}_{s2}$	107.347	138.942	120.356	142.043	138.908	142.396	153.908	176.045
		$\overline{h}_{s3}$	106.987	153.908	117.674	135.980	135.409	139.478	149.807	175.346
	50	$\overline{h}_{s1}$	122.378	143.761	167.986	190.543	147.890	171.098	175.098	201.678
		$\overline{h}_{s2}$	117.456	135.008	141.347	170.267	122.796	153.456	142.680	170.890
		$\overline{h}_{s3}$	115.158	132.456	135.789	157.980	120.004	146.670	137.654	155.670

The data preprocessing was performed in R4.3.0 using the robustbase package. All variables were standardized prior to analysis. Tuning parameters were initialized at $k = 3$ and $\alpha = 0.75$. The adaptive update terminated after ≤ 10 iterations in all datasets.

7 Discussion

The novelty of the proposed estimator lies not only in employing the UK's redescending loss but in introducing an adaptive mechanism that links the tuning constant to the contamination index. This data-driven adaptation is theoretically justified through bounded influence and enhanced breakdown behavior, distinguishing it from existing fixed-parameter redescending approaches.

The present study systematically evaluated the performance of the proposed class of estimators using analytical derivations, empirical applications, and simulation experiments. The analytical comparison of mean square errors (MSE) clearly demonstrated that estimators utilizing UK's redescending M-estimation outperform traditional OLS-based estimators under conditions of data contamination. This superiority arises from the bounded influence function and high breakdown point of the redescending estimator, which collectively mitigate the effect of extreme values on parameter estimation.

Empirical analysis based on real-world datasets, particularly the NSSO household income data—reinforced these theoretical findings. The adaptive UK's redescending estimator achieved the lowest MSE and highest percent relative efficiency (PRE), exhibiting nearly a 45% reduction in bias compared to OLS-based methods. This consistent performance across diverse datasets indicates that the adaptive tuning mechanism effectively adjusts the robustness level to accommodate varying contamination degrees, providing a self-regulating estimation process without manual tuning.

The simulation results further corroborate these insights. Across different sample sizes and contamination levels, the proposed estimators maintained stable efficiency and robustness. Even in heavily skewed or outlier-laden distributions, the adaptive redescending estimator preserved efficiency close to that of OLS under ideal conditions while significantly improving reliability under non-ideal data. This dual capability such as robustness without efficiency loss, positions the adaptive estimator as a practical and theoretically sound solution for survey-based estimation tasks.

Overall, the findings emphasize that employing robust redescending techniques not only enhances estimator precision but also ensures resilience in real-world applications, where ideal distributional assumptions rarely hold.

Summary of Findings: The adaptive estimator consistently demonstrated higher efficiency and robustness than fixed M-estimators and OLS across both clean and contaminated data. It achieved up to 45% bias reduction and 90%–200% PRE gains in realistic scenarios.

8 Conclusion

The collective analytical, empirical, and simulation-based results establish that the proposed UK's redescending M-estimator and its adaptive variant, provide a powerful and reliable alternative to the conventional OLS approach for finite population mean estimation under SRSWOR. Their bounded influence, higher breakdown point, and automatic tuning capabilities allow for accurate inference even under substantial data contamination.

These properties make the proposed estimators particularly suitable for socio-economic, agricultural, and industrial survey data, which frequently exhibit irregularities and extreme observations. Beyond improving estimation accuracy, the adaptive redescending framework offers computational stability and interpretability, making it an attractive choice for applied statisticians.

While the proposed adaptive redescending estimator exhibits strong robustness, it involves slightly higher computational cost due to iterative adaptation. The sensitivity of tuning parameters (α, k) may vary with sample size and contamination degree. Small-sample bias can also arise when contamination is misestimated. Future work should extend the adaptive framework to multivariate and non-linear survey modes, explore Bayesian formulations for automatic tuning and evaluate scalability under complex survey designs.

Acknowledgement: The author is thankful to the Deanship of Graduate Studies and Scientific Research at University of Bisha for supporting this work through the Fast-Track Research Support Program.

Funding Statement: The author received no specific funding for this study.

Availability of Data and Materials: The data utilized in this study are publicly available. The NSSO household consumption data can be accessed through the official National Sample Survey Office (NSSO, India) database. The simulated datasets generated during the study are available from the corresponding author upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The author declares no conflicts of interest to report regarding the present study.

Appendix A Detailed Derivation of Bias and MSE for the Adaptive UK's Redescending M-Estimator

Appendix A.1 Notations and Estimator Representation

Let the sample consist of $\{(h_{ii}, j_i)\}, i = 1, \ldots, n$. Define

- $\bar{h} = n^{-1} \sum_{i=1}^{n} h_i, \bar{j} = n^{-1} \sum_{i=1}^{n} j_i$
- The adaptive tuning constant $k^* = k(1 + \alpha C)$ where $C = n^{-1} \sum_{i=1}^{n} I(|r_i| > 2s)$ (with r_i residuals and s the MAD).
- The proposed estimator can be written in the general form (Consistent with main text)

$$\hat{H}_A = T\left(\bar{h}, \bar{j}, k^*\right)$$

where $T(.)$ is twice continuously differentiable in a neighbourhood of the population values $\left(\mu_h, \mu_j, k\right)$. This includes ratio type functional forms used in the paper (see main text for explicit $g(.)$ forms).

Appendix A.2 Taylor Expansion (Second Order)

Expand T around $\left(\mu_h, \mu_j, k\right)$, where $\mu_h = \overline{H}$ and $\mu_j = \overline{J}$ denote population means and k the nominal tuning constant:

$$\begin{aligned}\overline{H}_A \approx\ & T_0 + T_h\left(\bar{h} - \mu_h\right) + T_j\left(\bar{j} - \mu_j\right) + T_k\left(k^* - k\right) \\ & + \frac{1}{2}T_{hh}\left(\bar{h} - \mu_h\right)^2 + \frac{1}{2}T_{jj}\left(\bar{j} - \mu_j\right)^2 + T_{hj}\left(\bar{h} - \mu_h\right)\left(\bar{j} - \mu_j\right) \\ & + T_{hk}\left(\bar{h} - \mu_h\right)\left(k^* - k\right) + T_{jkj}\left(\bar{j} - \mu_j\right)\left(k^* - k\right) + T_{kk}\left(k^* - k\right)^2\end{aligned}$$

where partial derivatives (evaluated at the expansion point) are abbreviated as $T_h = \partial T/\partial \bar{h}$, $T_j = \partial T/\partial \bar{j}$, $T_k = \partial T/\partial k$ and second derivatives as T_{hh}, T_{jj}, T_{hj}, T_{jk}, T_{kk}.

We retain terms up to order $O_p\left(n^{-1}\right)$ and drop higher order remainders.

Appendix A.3 Bias Approximation

Take expectations. Under SRSWOR, $E\left(\bar{h} - \mu_h\right) = 0$, $E\left(\bar{j} - \mu_j\right) = 0$. Also note $E\left(k^* - k\right) = k\alpha E(C)$. Thus the leading bias arises from

- The deterministic contribution via $E\left(k^* - k\right)$,
- Second-order terms, e.g., $E\left(\bar{j} - \mu_j\right)^2$.

Hence

$$\begin{aligned}Bias\left(\hat{H}_A\right) \approx\ & T_k k\alpha E(C) \\ & + \frac{1}{2}T_{hh}Var\left(\bar{h}\right) + \frac{1}{2}T_{jj}Var\left(\bar{j}\right) + T_{hj}Cov\left(\bar{h}, \bar{j}\right) \\ & + T_{hk}Cov\left(\bar{h}, k^*\right) + T_{jk}Cov\left(\bar{j}, k^*\right) + \frac{1}{2}T_{KK}Var\left(k^*\right).\end{aligned}$$

In practice the dominant bias terms are often $T_k k\alpha E(C)$ and the second-order variance terms shown above. Covariance's with k^* are typically $O\left(n^{-1}\right)$ and can be estimated using influence-function or bootstrap methods (see Appendix A.6).

Appendix A.4 Variance/MSE Approximation

Compute variance by applying $Var\left[\hat{H}_A\right] \approx Var$ of the linearized terms plus contributions from the adaptive randomness. Neglecting third and higher-order terms:

$$\begin{aligned}Var\left(\hat{H}_A\right) \approx\ & T_h^2 Var\left(\bar{h}\right) + T_j^2 Var\left(\bar{j}\right) + T_k^2 Var\left(k^*\right) \\ & + 2T_hT_jCov\left(\bar{h}, \bar{j}\right) + 2T_hT_jCov\left(\bar{h}, k^*\right) + 2T_jT_kCov\left(\bar{j}, k^*\right)\end{aligned}$$

Thus the approximate MSE is:

$$MSE\left(\hat{H}_A\right) \approx \left[Bias\left(\hat{H}_A\right)\right]^2 + Var\left(\hat{H}_A\right)$$

With the bias as in Appendix A.3 and variances as above.

Appendix A.5 Substituting SRSWOR Variance/Covariance Expressions

Under SRSWOR:

$$Var\left(\bar{h}\right) = \frac{1-t}{n}S_h^2,\ Var\left(\bar{j}\right) = \frac{1-t}{n}S_j^2,\ Cov\left(\bar{h},\bar{j}\right) = \frac{1-t}{n}S_{hj}$$

where S_h^2, S_j^2 and S_{hj} are finite population quantities. Substitute these into the variance formula above.

Appendix A.6 Terms Involving the Adaptive Constant k^*

Because $k^* = k\,(1 + \alpha C)$, we have

$$Var\,(k*) = k^2\alpha^2 Var\,(C)\,,\ Cov\left(\bar{h}, k^*\right) = k\alpha Cov\left(\bar{h}, C\right),\ Cov\left(\bar{j}, k\right) = k\alpha Cov\left(\bar{j}, C\right)$$

Approximate $Var\,(C)$ and the covariance using indicator-based variance and delta-method arguments. Since

$$C = \frac{1}{n}\sum_{i=1}^{n} I\,(|r_i| > 2s)$$

Treat the indicators as Bernoulli with probability $p = P\,(|R| > 2s)$, so to first order

$$Var\,(C) \approx \frac{p\,(1-p)}{n}$$

For covariance $Cov\left(\bar{h}, C\right)$, note

$$Cov\left(\bar{h}, C\right) \approx \frac{1}{n}Cov\,(H, I\,(|R| > 2s)) \approx \frac{\gamma_h}{n},$$

where $\gamma_h = E\,[(H - \mu_h)\,I\,(|R| > 2s)]$ can be estimated from the sample or approximated via influence-function calculations. Similarly for $Cov\left(\bar{j}, C\right) \approx \gamma_j/n$. In many contaminations scenarios these covariance's are small (order $1/n$) and often negligible relative to the leading variance terms; however, they are included in the full formula for completeness.

Appendix A.7 Compact Final Approximate MSE Formula

Collecting the above and retaining leading $O\,(1/n)$ terms:

$$\begin{aligned} MSE\left(\hat{H}_A\right) \approx{} & [T_k k\alpha E\,(C)]^2 + T_h^2\frac{1-t}{n}S_h^2 + T_j^2\frac{1-t}{n}S_j^2 + 2T_hT_j\frac{1-t}{n}S_{hj} \\ & + k^2\alpha^2 T_k^2\frac{p\,(1-p)}{n} \\ & + 2k\alpha T_hT_k\frac{\gamma_h}{n} + 2k\alpha T_jT_h + (\textit{Small order terms}) \end{aligned}$$

Interpretations of terms:

- First term: squared bias induced by adaptive tuning shift $E(k^* - k) = k\alpha E(C)$.
- Next three terms: classical variance contributions from sample means under SRSWOR.
- Fourth line: additional variance produced by randomness of C and its covariance with $\bar{h}, \bar{j}$.

All Quantities $T_h, T_j, T_k, \ldots$ depend on the specific functional form T (i.e., on the ratio-type $g(.)$ or the M-estimator implicit. Function). For the estimators used in the manuscript one can compute these derivatives analytically or numerically (using plug-in population values and sample estimates). The indicator probability p and the covariance's γ_h, γ_j are estimated via the observed residuals or via a bootstrap.

References

1. Kadilar C, Cingi H. Ratio estimators in simple random sampling. Appl Math Comput. 2004;151(3): 893–902.
2. Abid M, Abbas N, Zafar Nazir H, Lin Z. Enhancing the mean ratio estimators for estimating population mean using non-conventional location parameters. Revista Colombiana de Estadística. 2016;39(1):63–79.
3. Koyuncu N, Kadilar C. Efficient estimators for the population mean. Hacettepe J Math Stat. 2009;38(2):217–25.
4. Singh HP, Tailor R. Use of known correlation coefficient in estimating the finite population mean. Stat Transit. 2003;6:555–60.
5. Upadhyaya LN, Singh HP. Use of transformed auxiliary variable in estimating the finite population mean. Biom J. 1999;41:627–36.
6. Almanjahie IM, Alomari AI, Ekpenyong EJ, Subzar M. Generalized class of mean estimators with known measures for outlier treatment. Comput Syst Sci Eng. 2021;38(1):1–15.
7. Zaman T. Improvement of modified ratio estimators using robust regression methods. Appl Math Comput. 2019;348(1):627–31. doi:10.1016/j.amc.2018.12.037.
8. Subzar M, Alqurashi T, Chandawat D, Tamboli S, Raja TA, Attri AK. Generalized robust regression techniques and adaptive cluster sampling for efficient estimation of population mean in case of rare and clustered populations. Sci Rep. 2025;15(1):2069. doi:10.1038/s41598-025-85328-0.
9. Alqudah MA, Zayed M, Subzar M, Wani SA. Neutrosophic robust ratio-type estimators for estimating finite population mean. Heliyon. 2024;10(8):e28934. doi:10.1016/j.heliyon.2024.e28934.
10. Sharma P, Pusa N, Kumari M, Singh P. Balancing accuracy and cost: a new estimator for stratified random sampling. In: Communications in statistics—simulation and computation; 2025. p. 1–17. doi:10.1080/03610918.2025.2552216.
11. Shahzad U, Ahmad I, Al-Noor NH, Hanif M, Almanjahie IM. Robust estimation of the population mean using quantile regression under systematic sampling. Math Popul Stud. 2023;30(3):195–207.
12. Subzar M, Al-Omari AI, Alanzi ARA. The robust regression methods for estimation of finite population mean based on SRSWOR in case of outliers. Comput Mater Contin. 2020;65(1):125–38. doi:10.32604/cmc.2020.010230.
13. Singh P, Maurya P, Sharma P. Enhanced estimation of population mean via exponential-type estimator using dual auxiliary inputs. Appl Math Sci Eng. 2025;33(1):45. doi:10.1080/27690911.2025.2554121.
14. Yadav SK, Arya D, Sharma P, Zaman T. Improved generalized ratio estimators of population mean using auxiliary variables. Life Cycle Reliab Safe Eng. 2025;215(4):4198. doi:10.1007/s41872-025-00360-3.
15. Singh D, Chaudhary FS. Theory and analysis of sample survey designs. New Delhi, India: Wiley–Blackwell; 1986.