

# International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452  
NAAS Rating (2026): 4.49  
Maths 2026; 11(3): 40-43  
© 2026 Stats & Maths  
<https://www.mathsjournal.com>  
Received: 03-01-2026  
Accepted: 08-02-2026

**TA Raja**  
Division of Agricultural  
Statistics, Sher-e-Kashmir  
University of Agricultural  
Sciences and Technology,  
Shalimar, Jammu and Kashmir  
(UT), India

**S Maq**  
Division of Agricultural  
Economics and Statistics, Sher-e-  
Kashmir University of  
Agricultural Sciences and  
Technology, Wadura, Sopore,  
Jammu and Kashmir (UT), India

**Subzar Mir**  
Division of Agricultural  
Statistics, Sher-e-Kashmir  
University of Agricultural  
Sciences and Technology,  
Shalimar, Jammu and Kashmir  
(UT), India

**Nagina Nazir**  
Division of Agricultural  
Statistics, Sher-e-Kashmir  
University of Agricultural  
Sciences and Technology,  
Shalimar, Jammu and Kashmir  
(UT), India

**IA Shah**  
Division of Agricultural  
Statistics, Sher-e-Kashmir  
University of Agricultural  
Sciences and Technology,  
Shalimar, Jammu and Kashmir  
(UT), India

**Corresponding Author:**  
**TA Raja**  
Division of Agricultural  
Statistics, Sher-e-Kashmir  
University of Agricultural  
Sciences and Technology,  
Shalimar, Jammu and Kashmir  
(UT), India

## On regression technique and adaptive cluster sampling for estimation of population mean yield in case of clustered population

**TA Raja, S Maq, Subzar Mir, Nagina Nazir and IA Shah**

**DOI:** <https://www.doi.org/10.22271/math.2026.v11.i3a.2273>

### Abstract

Adaptive cluster sampling (ACS) provides an effective framework for including neighboring or related units while sampling rare and highly clustered populations. Introduced by Thompson, ACS is a reliable sampling technique under the assumption that the data are free from outliers. However, conventional ACS-based estimators can yield distorted results in the presence of outliers. Motivated by this limitation, the present study develops adaptive estimators within the ACS framework using ordinary least squares (OLS), Huber M, and Mallows GM estimation methods. The proposed regression-type estimators incorporate these functions to improve performance. For illustration, real-life numerical data of the Blue-winged teal has been utilized. The mean square error (MSE) properties of both the adapted and proposed estimators are derived and compared to assess their efficiency.

**Keywords:** Adaptive cluster sampling (ACS), ordinary least square (OLS) regression, Huber M, mallows GM, mean square error

### Introduction

Adaptive cluster sampling (ACS), conceptualized by Thompson, S. K. (1990) <sup>[4]</sup>, is an optimal sampling method for detecting rare and highly clustered population units. It is a data-driven approach that offers substantial potential for producing more reliable inferences when estimating parameters of populations characterized by both rarity and strong spatial clustering. ACS is commonly implemented as a two-phase sampling procedure for estimating population totals and means of rare, hidden, and clustered populations. This approach is particularly appropriate in situations where rare species tend to occur in localized clusters within a study area.

In the first phase, an initial sample is selected using a conventional sampling design, such as simple random sampling without replacement or stratified sampling, subject to a predefined condition that at least one unit containing the rare species is detected. In the second phase, all neighbouring units surrounding the identified units are incorporated into the sample.

Populations suited to this sampling framework include rare and endangered plant and animal species, fisheries, pollution hotspots, sporadic disease occurrences, mineral deposits, criminal activities, and populations affected by hidden or stigmatized diseases. Incorporating auxiliary information alongside the primary survey variable can substantially improve the precision of ACS-based estimates. For example, when the survey variable represents the count of a particular species in a given locality, auxiliary variables such as food sources, habitat characteristics, rainfall, and temperature in the same area can be used to estimate population parameters including the mean, total, and variance.

The relationship between ecological variables and environmental covariates may be either positive or negative. Accordingly, ratio-type estimators are used for positive correlations, while product-type estimators are employed for negative correlations. The ACS procedure initiates the sampling process through standard sampling methods like simple, stratified or systematic sampling, and then seamlessly expanding it into the realm of ACS.

Table 1 illustrates a cluster of V variables within a network structure. The unit indicated by a star represents the initial selection, while the units located to its south, north, east, and west are collectively referred to as the first-order neighbourhoods. In the context of the auxiliary variable V and the study variable U, the adaptation criterion is pre-defined as

**Table 1:** Supplementary variable (v) values of population-1

|   |   |   |     |    |   |   |    |       |     |
|---|---|---|-----|----|---|---|----|-------|-----|
| 0 | 0 | 3 | 5   | 0  | 0 | 0 | 0  | 0     | 0   |
| 0 | 0 | 0 | 24* | 14 | 0 | 0 | 10 | 103   | 0   |
| 0 | 0 | 0 | 0   | 3  | 3 | 2 | 0  | 13639 | 1   |
| 0 | 0 | 0 | 0   | 0  | 0 | 0 | 0  | 14    | 122 |
| 0 | 0 | 0 | 0   | 0  | 0 | 2 | 0  | 0     | 177 |

**Table 2:** Response variable (u) values of population-1

|   |   |    |     |    |    |    |    |       |     |
|---|---|----|-----|----|----|----|----|-------|-----|
| 0 | 0 | 14 | 16* | 63 | 0  | 0  | 0  | 0     | 0   |
| 0 | 0 | 0  | 0   | 97 | 10 | 0  | 23 | 383   | 0   |
| 0 | 0 | 0  | 0   | 0  | 10 | 10 | 0  | 34837 | 8   |
| 0 | 0 | 0  | 0   | 0  | 0  | 2  | 0  | 64    | 427 |
| 0 | 0 | 0  | 0   | 0  | 0  | 8  | 0  | 0     | 627 |

Table 2 follows the pattern of Table 1, and is also shaded in yellow and blue to indicate the network and cluster of U, respectively. The size of a network is denoted by h. Hence, the average value of a network is  $\bar{h} = (\sum(\cdot)/h)$ . Here  $(\cdot)$  represents the sum of V or U values in a network.

Measures of central tendency, owing to their broad applicability across scientific and artistic disciplines, have created a growing need for improved mean estimation, particularly in situations where adaptive cluster sampling (ACS) is employed. The pioneering work on adaptive cluster sampling (ACS) was carried out by Thompson (1990) [4], who proposed several unbiased estimators. Subsequently, numerous researchers have contributed to improving the efficiency of mean and variance estimation by incorporating ancillary information, using either a single auxiliary variable or bivariate auxiliary variables, under various sampling designs. For details, see Thompson (1991) [5], Chao (2004) [7], Chutiman and Kumphon (2008) [11], Chutiman (2013) [13], Qureshi *et al.* (2018, 2020) [24], Yasmeen and Thompson (2020) [25], Shahzad *et al.* (2022) [26], Raja *et al.* (2023) [20], Raja and Maqbool (2024) [22]. For mean estimation, Kadilar *et al.* (2007) [10] and Subzar *et al.* (2025) [27] have utilized robust regression. Following this work, many authors have developed mean estimators using robust regression; for details, see Zaman and Bulut (2020) [16] under SRS and stratified sampling schemes, and quantile regression by Shahzad *et al.* (2023) [21].

Accordingly, the present study aims to develop adapted and newly proposed estimators within the adaptive cluster sampling (ACS) framework by employing robust techniques. These estimators are designed to enhance resistance to outliers, non-normality, model misspecification, small sample sizes, heteroscedasticity, and complex population structures, thereby addressing the limitations of conventional estimation methods. Robust regression methods are particularly valuable for studies involving rare and clustered populations, as they yield more accurate and reliable estimates.

The paper is structured as follows. “*OLS Regression and Adapted Estimators*,” “*Huber M Regression and Adapted Estimators*,” “*Generalized M Regression and Adapted Estimators*,” and “*Mallows GM Regression and Adapted Estimators*” examine these regression techniques within the ACS framework. The section “*Proposed Robust Estimator*” is devoted to the development of a robust estimator tailored

follows: a unit will be included in the network if its value is greater than or equal to 1 (i.e.,  $V \geq 1$ ). Units within the yellow shaded area collectively constitute the network, while those indicated with blue numbers represent respective edge units of this network. Together, the network and its associated edge units form a cluster.

specifically to the ACS setting, drawing on the Huber and Mallows GM methodologies.

### Adapted estimators

#### Ordinary Least Square (OLS)

The OLS method is a widely recognized parametric technique with a strong reputation. Its popularity arises from its simple computational procedure, which is often used for parameter estimation. The OLS regression coefficient is commonly applied to estimate the population mean. For example, Kadilar and Cingi (2004, 2006) [8, 9] proposed ratio-type regression estimators for mean estimation by utilizing well-known properties of auxiliary information. Likewise, Yan and Tian (2010) [12] developed two modified ratio estimators based on the coefficients of skewness and kurtosis. Also, Raja and Maqbool (2021) [17], Rather *et al.* (2022) [18], and Subzar *et al.* (2025) [27] have proposed ratio-type regression estimators. Inspired by these approaches, we have suggested estimator within the framework of adaptive cluster sampling (ACS).

$$At1 = [hu^- + \lambda_{ols}(V^- - hv^-)]/h v^{-*}(V^-) \dots \dots \dots (1)$$

$$At2 = [hu^- + \lambda_{ols}(V^- - hv^-)]/h c v^{-*}(V^-_{cv}) \dots \dots \dots (2)$$

$$At3 = [hu^- + \lambda_{ols}(V^- - hv^-)]/[h \beta_2 v]^{-*}(V^-_{\beta_2 v}) \dots \dots (3)$$

#### Huber M regression and adapted estimators

Peter J. Huber introduced the concept of M-estimation for regression, offering a more direct approach than M-estimation for location. In M-estimation, which involves minimizing a specified function of the residuals—as in location estimation—the robustness of the estimator depends largely on the choice of the weight function.

The OLS estimator that is generated by minimizing the sum of squares function is identical to the maximum likelihood estimator under the conditions of linearity, homoscedasticity, and uncorrelated errors:

$$\min \sum_{i=1}^n (u_i - \sum v_{ij} \lambda_j)^2 = \min \sum_{i=1}^n (e_i)^2 \quad (4)$$

Expanding on principles of M-estimation for location, a robust regression M-estimation minimizes sum of a function of residuals that increases less rapidly

$$\min \sum_{i=1}^n \rho(u_i - \sum v_{ij} \lambda_j)^2 = \min \sum_{i=1}^n \rho(e_i)^2 \quad (5)$$

Since the solution lacks scale equivariance, residuals need to be standardized using a robust scale estimate  $\sigma_{ci}$ , which is determined simultaneously. Similar to M-estimation for location, median absolute deviation (MAD) is frequently employed for this purpose. Deriving the score function involves taking the derivative of equation (5) and solving:

$$\min \sum_{i=1}^n \psi(u_i - \sum v_{ij} \lambda_j / \hat{\sigma}) v_{ik} = \sum_{i=1}^n \psi(e_i / \hat{\sigma} e_i) v_i = 0 \quad (6)$$

With  $\psi = \hat{\rho}$ , there is now a system of  $K+1$  equations, for which  $\psi$  is replaced by appropriate weights that decrease as size of residual increases

$$\sum_{i=1}^n w_i (e_i / \hat{\sigma} e_i) v_i = 0 \quad (7)$$

Huber regression coefficient proves highly beneficial for mean estimation under simple random sampling (SRS), Zaman, T (2019) [15].

Huber regression-based ratio-type mean estimators under ACS are formulated as:

$$A_{t4} = \frac{[\bar{h}_u + \lambda_{Huber}(\bar{V} - \bar{h}_v)]}{\bar{h}_v} \bar{V} \quad (8)$$

$$A_{t5} = \frac{[\bar{h}_u + \lambda_{Huber}(\bar{V} - \bar{h}_v)]}{\bar{h}_{cv}} \bar{V}_{cv} \quad (9)$$

$$A_{t6} = \frac{[\bar{h}_u + \lambda_{Huber}(\bar{V} - \bar{h}_v)]}{\bar{h}_{\beta 2v}} \bar{V}_{\beta 2v} \quad (10)$$

### Generalized M regression and adapted estimators

M-estimators actually fail to account for leverage (Hampel *et al.*, 1986) [3]. GM estimators were proposed with the aim to create weights that will consider both leverage and vertical outliers. Standard M-estimators are used to deal with outliers, and their cap values are down-weighted to typically deal with leverage points. The GM class of estimators is defined by the function mentioned in equation (11):

$$\sum_{i=1}^n w_i(v_i) \psi \left\{ \frac{e_i}{\sqrt{\text{var}(v_i) \sigma_e}} \right\} v_i = 0 \quad (11)$$

$\psi$  represents score function and weights  $w_i$  and  $v_i$  are updated iteratively, depends on model matrix  $V$

### Mallows GM regression and adapted estimators

The GM estimator which includes only  $w_i$  weights (*i.e.*,  $\text{var}(v_i) = 1$  in equation 11) was proposed by Mallows; for reference, see Krasker and Welsch (1982) [2], where  $w_i$  are calculated from cap values and range from 0 to 1 with the assumption that observations receive less weight having high leverage than those with small leverage (*i.e.*, if  $s_i > s_j$ , then  $z_i < z_j$ ). To be resistant against high leverage outliers, Mallows proposed the Mallows GM estimate and defined it by:

$$\sum_{i=1}^n w_i \Re \left\{ \frac{e_i(\beta)}{\hat{\sigma}} \right\} v_i = 0, \text{ where } \Re(e) = \rho'(e) \text{ and } w_i = \sqrt{1 - s_i}$$

with  $s_i$  being leverage of the  $i_{th}$  observation. The adapted Mallows GM regression-based ratio-type mean estimators under ACS are formulated as:

$$A_{t7} = \frac{[\bar{h}_u + \lambda_{Mallows}(\bar{V} - \bar{h}_v)]}{\bar{h}_v} \bar{V} \quad (12)$$

$$A_{t8} = \frac{[\bar{h}_u + \lambda_{Mallows}(\bar{V} - \bar{h}_v)]}{\bar{h}_{cv}} \bar{V}_{cv} \quad (13)$$

$$A_{t9} = \frac{[\bar{h}_u + \lambda_{Huber}(\bar{V} - \bar{h}_v)]}{\bar{h}_{\beta 2v}} \bar{V}_{\beta 2v} \quad (14)$$

### Proposed robust estimator

A long-standing challenge in survey sampling is the efficient and effective estimation of population parameters for rare and clustered populations. Conventional sampling designs often fail in such settings, as many sampled units provide little information. Thompson's pioneering work on Adaptive Cluster Sampling (ACS) introduced a data-driven approach that improves sampling efficiency and yields more reliable inference for populations characterized by rarity and high spatial clustering. Here we introduce a regression estimator that consistently reveals indubitable efficacy. This estimator is a combination of Huber and Mallows and focuses on the regression component, excluding the ratio-biased factor present in the suggested versions. The proposed estimator is given as:

$$P_{\text{test}} = [\bar{h}_u + \lambda_{Huber}(\bar{V} - \bar{h}_v) + \lambda_{Mallows}(\bar{V} - \bar{h}_v)] \quad (15)$$

Accordingly, using existing results and simple algebraic manipulations to avoid lengthy calculations, we obtain the MSE expressions for the proposed class of estimator up to order  $n^{-1}$ , as

$$MSE(P_{\text{est}}) = [V(\bar{h}_u) - 2(\lambda_{Huber} + \lambda_{Mallows})Cov(\bar{h}_v, \bar{h}_u) + \lambda_{Huber}^2 V(\bar{h}_u) + \lambda_{Mallows}^2 V(\bar{h}_v)] \dots \quad (16)$$

### Results and Discussion with numerical illustration

To support the employment of both adapted and proposed estimators, we employ a combination of empirical and theoretical analyses. In pursuit of this objective, we choose a real population for analysis: Blue-winged teal data. The effectiveness of the proposed estimator was evaluated using actual data from a population of blue-winged teal, as made available in Smith *et al.* (1995) [6] and Chutiman *et al.* (2014) [14]. The original dataset covered a total study area of 5,000 km<sup>2</sup>, which was partitioned into equal quadrants of 50-100 km<sup>2</sup> each in central Florida. Data presented in Table 1 were employed as a supplementary variable ( $v$ ) to generate values for the response variable ( $u$ ) using the model  $u = 4v + e$ , where  $e$  is normally distributed. Generated ( $u$ ) values are presented in Table 2.

Results obtained during the evaluation of real-world data are presented in Table 3 after applying robust methods instead of traditional techniques to estimate population parameters can substantially reduce variance. Conventional approaches typically assume that residuals follow a normal distribution, an assumption that is frequently violated in the presence of skewed or heavy-tailed data. In contrast, robust methods do not depend on the normality assumption, making them more reliable under such conditions.

**Table 3:** Parameters of population

|                            |                              |                          |
|----------------------------|------------------------------|--------------------------|
| $\bar{U} = 137.7203$       | $\bar{V} = 53.1315$          | $\lambda_{ols} = 2.5852$ |
| $\lambda_{Huber} = 2.5853$ | $\lambda_{Mallows} = 2.2506$ | $\beta_{2hv} = -3.2473$  |

Table 4 provides mean squared error (MSE) of the estimators for comparison.

**Table 4:** Mean square error of the adapted and proposed estimators

| Estimators        | Mean Square Error |
|-------------------|-------------------|
| At <sub>1</sub>   | 52,592            |
| At <sub>2</sub>   | 42,268            |
| At <sub>3</sub>   | 59,662            |
| At <sub>4</sub>   | 52,598            |
| At <sub>5</sub>   | 42,274            |
| At <sub>6</sub>   | 59,669            |
| At <sub>7</sub>   | 51,497            |
| At <sub>8</sub>   | 40,989            |
| At <sub>9</sub>   | 57,989            |
| P <sub>test</sub> | 21,567            |

## Conclusion

From the above results, we conclude that the Adaptive Cluster Sampling (ACS) technique is a highly effective method for estimating the mean of sparsely distributed and hidden clustered populations, yielding notably meaningful results. The findings from the numerical study clearly indicate that the use of ACS leads to the lowest mean square error (MSE) for the proposed estimator, outperforming all competing estimators considered. Consequently, the proposed estimator demonstrates performance relative to the other estimators examined in this study. Future research accomplishments may benefit from adopting the proposed estimator.

## References

- Huber PJ. Robust estimation of a location parameter. *Ann Math Stat.* 1964;35:73-101.
- Krasker WS, Welsch RE. Efficient bounded-influence regression estimation. *J Am Stat Assoc.* 1982;77:595-604.
- Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA. *Robust statistics: the approach based on influence functions.* New York: Wiley; 1986.
- Thompson SK. Adaptive cluster sampling. *J Am Stat Assoc.* 1990;85(412):1050-1059.
- Thompson SK. Stratified adaptive cluster sampling. *Biometrika.* 1991;78(2):389-397.
- Smith DR, Conroy MJ, Brakhage DH. Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics.* 1995;51:777-788.
- Chao CT. Ratio estimation on adaptive cluster sampling. *Chinese Stat Assoc.* 2004;42(3):307-327.
- Kadilar C, Cingi H. Ratio estimators in simple random sampling. *Appl Math Comput.* 2004;151(3):893-902.
- Kadilar C, Cingi H. An improvement in estimating the population mean by using the correlation coefficient. *Haceteppe J Math Stat.* 2006;35(1):103-109.
- Kadilar C, Candan M, Cingi H. Ratio estimators using robust regression. *Haceteppe J Math Stat.* 2007;36(2):181-188.
- Chutiman N, Kumphon B. Ratio estimator using two auxiliary variables for adaptive cluster sampling. *Thailand Stat.* 2008;6(2):241-256.
- Yan Z, Tian B. Ratio method to the mean estimation using coefficient of skewness of auxiliary variable. In: *Information Computing and Applications: International Conference ICICA 2010, Tangshan, China, October 15-18, 2010.* Berlin: Springer; 2010. p. 103-110.
- Chutiman N. Adaptive cluster sampling using auxiliary variable. *J Math Stat.* 2013;9(3):249-255.
- Chutiman N, Chiangpradit M. Ratio estimator in adaptive cluster sampling without replacement of networks. *J Probab Stat.* 2014;2014:726398. doi:10.1155/2014/726398.
- Zaman T. Improvement of modified ratio estimators using robust regression methods. *Appl Math Comput.* 2019;348:627-631.
- Zaman T, Bulut H. Modified regression estimators using robust regression methods and covariance matrices in stratified random sampling. *Commun Stat Theory Methods.* 2020;49(14):3407-3420.
- Raja TA, Maqbool S. On modified ratio estimator using a new linear combination. *Int J Agriculture Stat Sci.* 2021;17(1):209-211.
- Rather KUI, Koçyiğit EG, Onyango R, Kadilar C. Improved regression in ratio type estimators based on robust M-estimation. *PLoS ONE.* 2022;17(12):e0278868. doi:10.1371/journal.pone.0278868.
- Alomair AM, Shahzad U. Optimizing mean estimators with calibrated minimum covariance determinant in median ranked set sampling. *Symmetry.* 2023;15(8):1581.
- Raja TA, Maqbool S, Mir S. On ratio estimation using auxiliary information in survey sampling. *Int J Agriculture Sci.* 2023;15(2):12199-12203.
- 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.
- Raja TA, Maqbool S. Ratio type estimators for estimating population mean by using some auxiliary information. *Int J Agriculture Sci.* 2024;16(1):12799-12801.
- Mir S, Al Quarashi T, Chandawat D, Tamboli S, Raja TA, Attri AK, *et al.* Generalized robust regression techniques and adaptive cluster sampling for efficient estimation of population mean in case of rare and clustered populations. *Sci Rep.* 2025;15:2069. doi:10.1038/s41598-025-85328-0.
- Qureshi A, Singh DK, Dwivedi S. Nano-fertilizers: a novel way for enhancing nutrient use efficiency and crop productivity. *International Journal of Current Microbiology and Applied Sciences.* 2018 Feb 10;7(2):3325-3335.
- Yasmeen U, Thompson M. Variance estimation in adaptive cluster sampling. *Communications in Statistics-Theory and Methods.* 2020 May 18;49(10):2485-2497.
- Shahzad M, Qu Y, Rehman SU, Zafar AU. Adoption of green innovation technology to accelerate sustainable development among manufacturing industry. *Journal of Innovation & Knowledge.* 2022 Oct 1;7(4):100231.
- Subzar M, Raja TA, Kumar M, Singh S, Maqbool S. Robust ratio type estimator as an alternative to regression estimator for population mean estimation in srswor. *Int. J. Agricult. Stat. Sci. Vol.* 2025;21(1):271-276.