Journal of Modern Applied Statistical Methods
https://jmasm.com

Article

Mean Estimation Using Factor Type Estimator in Presence of Measurement Errors in Systematic Sampling

Amita Yadav [1], Sarla Pareek [1] and Narendra Singh Thakur [2]

[1] Department of Mathematics, Statistics and Computing, Banasthali Vidyapith, Banasthali 304022, Rajasthan, India; amitayadav73@yahoo.in or psarla13@gmail.com
[2] Department of Statistics, Government Adarsh Girls College, Sheopur 476337, Madhya Pradesh, India; nst_stats@yahoo.co.in
* Correspondence: amitayadav73@yahoo.in

How To Cite: Yadav, A.; Pareek, S.; Thakur, N.S. Mean Estimation Using Factor Type Estimator in Presence of Measurement Errors in Systematic Sampling. *Journal of Modern Applied Statistical Methods* **2025**, *24*(2), 3. https://doi.org/10.53941/jmasm.2025.100003

Abstract: This article proposes estimation of population mean using factor-type (F-T) estimator in the presence of measurement errors under systematic sampling scheme. The factor-type (F-T) estimator is biased and the expression of bias, MSE and optimum MSE of proposed estimator is obtained up to first order of approximation under the concept of large sample approximations and a comparative study of this estimator along with related pre-existing estimators is taken out. A simulation study has been performed to ratify the performance of proposed estimator in systematic sampling. The proposed factor-type (F-T) estimator is found better than other existing estimators as considered under in this study.

Keywords: measurement errors; systematic sampling; ratio estimator; factor-type (F-T) estimator; R software; FESO; NESO; RMSE; electronic throttle volve; least square algorithm

MSC: 62D05

1. Introduction

Systematic sampling is a structured approach that ensures that every element has an equal chance of being selected in the sample. So, random numbers are used initially in systematic sampling and then it follows a regular pattern of selection that defines the method. It ensures equal possibility for each unit in the population to be selected in the sample by taking first one randomly and rest units get selected automatically following a pre-defined pattern. The importance of this sampling strategy is due to its simplicity that emphasized it one of widely use sampling strategy.

In other words, let a random sample of size n is desired from a population of size N. in systematic sampling, the first element i is selected randomly from the first k elements (generally, i $(1 \leq i \leq k)$ is known as *random start* and k is defined as $N = nk$ and known as sampling interval), and thereafter every k^{th} element $[i + k, i + 2k, i + 3k, \ldots\ldots.., i + (n-1)k]$ is selected in the sample. According to Singh and Chaudhary (2009) [1], systematic sampling is simple and fool proof. Apart from its simplicity, this procedure, in many situations, provides estimates more efficient than simple random sampling and is widely used in various surveys.

Systematic sampling has fine features of selecting samples. Due to its simplicity and tendency to give better results, many authors have done their research in this scheme [2–8].

Non-response and measurement errors are major factors of non-sampling errors in survey sampling. However, it is possible that a sample may be defective by the measurement errors as well due to self-interest, failure of memory, fatigue, careless handling of data, etc. Measurement errors are those errors in which the recorded observations deviated from the true observations. For example, an income tax payee may hide his actual income to decrease self-tax liabilities.

A study by Humaidi et al. [9] of observer performance in state estimation for a rotary inverted pendulum system, focusing on the impact of measurement errors, reveals that Finite-Time Extended State Observers (FESO) generally outperform other observer types like Nonlinear Extended State Observers (NESO) and Linear Extended State Observers (LESO) in terms of estimation accuracy. This is particularly evident when considering parameter uncertainties and external disturbances, with FESO achieving a significantly lower Root Mean Squared Errors (RMSE).

Measurement errors in the context of Least Squares (LS) algorithms, particularly when applied to parameter estimation in canonical state-space models and incorporating bias compensation principles, are a critical concern influencing the accuracy and reliability of the estimated parameters [10,11]. Studied measurement errors in an electronic throttle valve can significantly affect engine performance and drivability. These errors can arise from various sources, including sensor inaccuracies, actuator limitations, and environmental factors. Addressing these errors are crucial for precise throttle control and optimal engine operation.

If the measurement errors occur in sampled data, then it can lead to an imprecise estimate. Several methods are found for controlling these types of errors like–ratio, product, regression, etc. Several authors did their contributions in this field to increase the precision of different estimators under diverse sampling schemes. Shalabh and Tsai [12] proposed a class of ratio and product method of estimation in the presence of correlated measurement errors. Singh et al. [13] have studied the effect of measurement errors on ratio, product, and mean estimator simultaneously. Singh and Vishwakarma [14] studied the simultaneous effect of measurement errors and non-response on mean estimation. These all works used simple random sampling but not in the context of systematic sampling [15–21].

Singh and Shukla [22] proposed a family of factor-type (F-T) estimator $\bar{y}_{FT}$ in simple random sampling using the auxiliary variable. The beauty of factor-type (F-T) estimator is that this estimator provides different estimators like, ratio, product, etc. at some specified values of its constant k_1. This estimator is bias controlled because it provides choices of constant k_1 for optimum mean squared errors as well. The estimator $\bar{y}_{FT}$ is biased and the expressions of bias, mean squared errors and optimum mean squared errors can obtain simply using the concept of large sample approximations up to first order.

Again, Shukla and Thakur [23] developed factor-type estimator as a device of imputation used for dealing missingness of the data. This estimator is found better than other existing methods of imputation under the simple random sampling design. Thakur and Shukla [24] applied the factor-type (F-T) estimator for chaining of two auxiliary variables and found that the factor-type (F-T) estimator is performed better in the same setup. A number of manuscript available in the literature in which importance and applicability of factor-type (F-T) estimator is proved with empirical study under different sampling strategies like, post-stratification, two-phase sampling, stratified random sampling, etc.

Vishwakarma and Singh [25] presented a generalized class of mean estimators under simple random sampling using auxiliary variable. The observations on both the study variable and the auxiliary variable are supposed to be recorded with measurement errors. The mean square errors of the proposed class of estimators is derived and studied under measurement errors.

To provide a brief overview, this paper organized in different sections. In Section 2, we have described the notations and terminologies used in this paper. Also, layout of systematic random sampling has been explained in this section. In Section 3, the focus is towards the estimators already existing in literature and then in Section 4, we have proposed an estimator to solve the problem and its properties in case of systematic sampling in presence of measurement errors. In Section 5, the comparative study of the proposed estimator with existing estimators is focused. In Section 6, we have performed simulation study using R studio and has done comparison of the efficiencies of different estimators with proposed estimator. In Section 7, a result has been made regarding the study as obtained and finally in Section 8, we conclude and discussed over all study.

2. Notations and Terminologies

Let population U consist of N units numbered as $1, 2, 3, \ldots\ldots, N$. For drawing a sample of size n, first unit will be selected randomly between 1 and k and the other units will be selected by pre-defined interval k or we can say element of that column in the Table 1. Let (μ_{sy}, μ_{sx}) and $(\sigma_{sy}^2, \sigma_{sx}^2)$ be the population mean and variance of study and auxiliary variable respectively. The symbol ρ_s be the correlation coefficient between auxiliary and study variable. However, the auxiliary variable, X, is fully available in the complete population either due to past experience or due to any other reasons.

Table 1. Layout of systematic sampling.

Random Start	1	2		i		k
1	1	2		i		k
2	k + 1	k + 2		k + i		2k
	⋮		…..	⋮	…..	⋮
j	(j − 1) k + 1	(j − 1) k + 2		(j − 1) k + i		jk
	⋮			⋮		⋮
n	(n − 1) k + 1	(n − 1) k + 2		(n − 1) k + i		nk

Suppose a systematic random sampling strategy with $(y_{ij} = Y_{ij} + u_{ij},\ x_{ij} = X_{ij} + v_{ij})$ be the observed values of study and auxiliary variables for $i = 1, 2, \ldots, k; j = 1, 2, \ldots, n$. (Y_{ij}, X_{ij}) be their true values of study variable and auxiliary variable respectively and (u_{ij}, v_{ij}) be the deviations between observed values and true values. The mentioned errors are stochastic in nature i.e., with zero mean and variances $(\sigma_{su}^2, \sigma_{sv}^2)$ and these errors are uncorrelated in nature. That implies cov $(X,Y) \neq 0$ and cov (X,u) = cov (X,v) = cov (Y,u) = cov (Y,v) = cov $(u,v) = 0$.

Let sample mean of study and auxiliary variable will be $(\bar{y}_{si}, \bar{x}_{si})$ for the observations with measurement errors are defined as

$$\bar{y}_{si} = \frac{1}{n}\sum_{j=1}^{n} y_{ij},\ i = 1, 2, \ldots, k; j = 1, 2, \ldots, n.$$

$$\bar{x}_{si} = \frac{1}{n}\sum_{j=1}^{n} x_{ij},\ i = 1, 2, \ldots, k; j = 1, 2, \ldots, n.$$

$$\bar{u}_{si} = \frac{1}{n}\sum_{j=1}^{n} u_{ij},\ i = 1, 2, \ldots, k; j = 1, 2, \ldots, n.$$

$$\bar{v}_{si} = \frac{1}{n}\sum_{j=1}^{n} v_{ij},\ i = 1, 2, \ldots, k; j = 1, 2, \ldots, n.$$

Under the concept of large sample approximations as $n \longrightarrow N$

$$\bar{y}_{si} = \mu_{sy}(1 + \epsilon_0) \Rightarrow \epsilon_0 = \frac{\bar{y}_{si} - \mu_{sy}}{\mu_{sy}}$$

$$\bar{x}_{si} = \mu_{sx}(1 + \epsilon_1) \Rightarrow \epsilon_1 = \frac{\bar{x}_{si} - \mu_{sx}}{\mu_{sx}} \quad \text{such that } |\epsilon_i| \leq 1;\ i = 0,1.$$

$$E(\epsilon_0) = 0, \qquad E(\epsilon_1) = 0,$$

$$E(\epsilon_0^2) = \frac{\sigma_{sy}^2 + \sigma_{su}^2}{k\,\mu_{sy}^2}, \qquad E(\epsilon_1^2) = \frac{\sigma_{sx}^2 + \sigma_{sv}^2}{k\,\mu_{sx}^2}, \qquad E(\epsilon_0\epsilon_1) = \frac{\rho_s\,\sigma_{sx}\,\sigma_{sy}}{k\,\mu_{sx}\,\mu_{sy}}$$

3. Estimators in Literature

This section includes some existing estimators for systematic sampling with measurement errors in study and auxiliary variables. Singh and Vishwakarma (2021) [8] studied the effect of measurement errors on sample mean estimator, ratio estimator, product estimator and the difference estimator while considering that the study and auxiliary information both are suffering by measurement errors under systematic sampling design.

The Table 2 summarize the bias and mean squared errors along with optimum mean squared errors (if required) of sample mean, ratio, product and difference estimators in the presence of measurement errors up-to first order of approximation.

Table 2. Properties of existing estimators.

Estimator	Bias	Mean Squared Errors
$t_1 = \bar{y}_{si}$	$B(t_1) = 0$	$V(t_1) = \frac{\sigma_{sy}^2 + \sigma_{su}^2}{k}$
$t_2 = \bar{y}_{si}\frac{\mu_{sx}}{\bar{x}_{si}}$	$B(t_2) = \frac{R_s^2\,(\sigma_{sx}^2 + \sigma_{sv}^2)}{k\,\mu_{sy}} - \frac{R_s\,\rho_s\,\sigma_{sx}\,\sigma_{sy}}{k\,\mu_{sy}}$	$M(t_2) = \frac{\sigma_{sy}^2 + \sigma_{su}^2}{k} + \frac{R_s^2\,(\sigma_{sx}^2 + \sigma_{sv}^2)}{k} - \frac{2R_s\,\rho_s\,\sigma_{sx}\,\sigma_{sy}}{k}$
$t_3 = \bar{y}_{si}\frac{\bar{x}_{si}}{\mu_{sx}}$	$B(t_3) = \frac{2R_s\,\rho_s\,\sigma_{sx}\,\sigma_{sy}}{k}$	$M(t_3) = \frac{\sigma_{sy}^2 + \sigma_{su}^2}{k} + \frac{R_s^2\,(\sigma_{sx}^2 + \sigma_{sv}^2)}{k} + \frac{2R_s\,\rho_s\,\sigma_{sx}\,\sigma_{sy}}{k}$
$t_4 = \bar{y}_{si} + d(\mu_{sx} - \bar{x}_{si})$	$B(t_4) = 0$	$M(t_4)_{Opt} = \frac{\sigma_{sy}^2 + \sigma_{su}^2}{k} - \frac{\rho_s^2\,\sigma_{sx}^2\,\sigma_{sy}^2}{k(\sigma_{sx}^2 + \sigma_{sv}^2)}$

4. Proposed Estimator

Singh and Shukla [22] proposed a family of factor-type (F-T) estimator in simple random sampling using the auxiliary variable. The factor-type estimator of Singh and Shukla (1987) is

$$\bar{y}_{FT} = \bar{y}\,\frac{(A_1+A_3)\bar{X}+fA_2\bar{x}}{(A_1+fA_2)\bar{X}+A_3\bar{x}}$$

where $A_1 = (k_1-1)(k_1-2), A_2 = (k_1-1)(k_1-4), A_3 = (k_1-2)(k_1-3)(k_1-4),\ f = \frac{n}{N}$ and k_1 = constant such that $0 < k_1 < \infty$.

Vishwakarma and Singh (2020) [25] proposed a generalized class of estimators for the estimation of population mean, when both study, as well as auxiliary variables, are commingled with measurement errors as $\bar{y}_M = \bar{y}M(v)$; where, $v = \frac{\bar{X}}{\mu_x}$ is such that $M(v)$ is continuous and bounded in R. Also, its first and second-order derivative are existing. Ratio, Product, unbiased estimator and some other pre-existing estimators can be the members of this family as well.

Motivated by Vishwakarma and Singh (2020) [25] and Singh and Shukla (1987) [22], we decided to experiment on a new estimation strategy in systematic sampling while considering that the data collected in the sample is suffering with measurement errors due to many reasons. The proposed new factor type (F-T) estimator under the same strategy is:

$$t_{FT} = \bar{y}_{si}\exp\left(\frac{\mu_{sx}-\bar{x}_{si}}{\mu_{sx}+\bar{x}_{si}} + \frac{(A_1+A_3)\mu_{sx}+fA_2\bar{x}_{si}}{(A_1+fA_2)\mu_{sx}+A_3\bar{x}_{si}} - 1\right) \tag{1}$$

where $A_1 = (k_1-1)(k_1-2), A_2 = (k_1-1)(k_1-4)$,
$A_3 = (k_1-2)(k_1-3)(k_1-4),\ f = \frac{n}{N}$ and k_1 is a constant such that $0 < k_1 < \infty$.
Define, $\theta_1 = \frac{fA_2}{A_1+fA_2+A_3}$, $\theta_2 = \frac{A_3}{A_1+fA_2+A_3}$, $\theta = \theta_1 - \theta_2$ and $R_s = \frac{\mu_{sy}}{\mu_{sx}}$.
Then, Equation (1) can be expressed in terms of errors as-

$$t_{FT} = \mu_{sy}(1+\epsilon_0)\exp\left[\frac{\mu_{sx}-\mu_{sx}(1+\epsilon_1)}{\mu_{sx}+\mu_{sx}(1+\epsilon_1)} + \frac{(A_1+A_3)\mu_{sx}+fA_2\,\mu_{sx}(1+\epsilon_1)}{(A_1+fA_2)\mu_{sx}+A_3\,\mu_{sx}(1+\epsilon_1)} - 1\right]$$

$$= \mu_{sy}\left[1+\epsilon_0+\left(\theta-\frac{1}{2}\right)\epsilon_1+\left(\theta-\frac{1}{2}\right)\epsilon_0\,\epsilon_1+\left(\theta-\frac{1}{2}\right)^2\frac{\epsilon_1^2}{2}+\left(\frac{1}{4}-\theta\theta_2\right)\epsilon_1^2\right] \tag{2}$$

$$t_{FT}-\mu_{sy} = \mu_{sy}\left[\epsilon_0+\left(\theta-\frac{1}{2}\right)\epsilon_1+\left(\theta-\frac{1}{2}\right)\epsilon_0\,\epsilon_1+\left(\theta-\frac{1}{2}\right)^2\frac{\epsilon_1^2}{2}+\left(\frac{1}{4}-\theta\theta_2\right)\epsilon_1^2\right]$$

To obtain bias of estimator t_{FT}, take expectations on both sides of Equation (2),

$$B(t_{FT}) = E(t_{FT}-\mu_{sy})$$

$$B(t_{FT}) = \mu_{sy}\left[\left(\theta-\frac{1}{2}\right)\frac{\rho_s\,\sigma_{sx}\,\sigma_{sy}}{k\,\mu_{sx}\,\mu_{sy}}+\frac{1}{2}\left(\theta-\frac{1}{2}\right)^2\frac{(\sigma_{sx}^2+\sigma_{sv}^2)}{k\,\mu_{sx}^2}+\left(\frac{1}{4}-\theta\theta_2\right)\frac{(\sigma_{sx}^2+\sigma_{sv}^2)}{k\,\mu_{sx}^2}\right]$$

$$B(t_{FT}) = \left(\theta-\frac{1}{2}\right)\frac{\rho_s\,\sigma_{sx}\,\sigma_{sy}}{k\,\mu_{sx}}+\frac{1}{2}\left(\theta-\frac{1}{2}\right)^2\frac{R_s\,(\sigma_{sx}^2+\sigma_{sv}^2)}{k\,\mu_{sx}}+\left(\frac{1}{4}-\theta\theta_2\right)\frac{R_s\,(\sigma_{sx}^2+\sigma_{sv}^2)}{k\,\mu_{sx}} \tag{3}$$

$$= \left(\theta-\frac{1}{2}\right)\frac{\rho_s\,\sigma_{sx}\,\sigma_{sy}}{k\,\mu_{sx}}+\left[\frac{1}{2}\left(\theta-\frac{1}{2}\right)^2+\frac{1}{4}-\theta\theta_2\right]\frac{R_s\,(\sigma_{sx}^2+\sigma_{sv}^2)}{k\,\mu_{sx}}$$

Squaring on both sides of Equation (2) along with ignoring higher terms of $O(n^{-1})$ and then taking expectation, we get the mean squared errors of t_{FT} as

$$M(t_{FT}) = \mathrm{E}(t_{FT}-\mu_{sy})^2 = E\left[\mu_{sy}{}^2\left\{\epsilon_0^2+\left(\theta-\frac{1}{2}\right)^2\epsilon_1^2+2\left(\theta-\frac{1}{2}\right)\epsilon_0\,\epsilon_1\right\}\right]$$

$$= \mu_{sy}{}^2\left[\frac{(\sigma_{sy}^2+\sigma_{su}^2)}{k\,\mu_{sy}^2}+\left(\theta-\frac{1}{2}\right)^2\frac{(\sigma_{sx}^2+\sigma_{sv}^2)}{k\,\mu_{sx}^2}+2\left(\theta-\frac{1}{2}\right)\frac{\rho_s\,\sigma_{sx}\,\sigma_{sy}}{k\,\mu_{sx}\,\mu_{sy}}\right] \tag{4}$$

$$= \frac{(\sigma_{sy}^2+\sigma_{su}^2)}{k}+\left(\theta-\frac{1}{2}\right)^2\frac{R_s^2\,(\sigma_{sx}^2+\sigma_{sv}^2)}{k}+2\left(\theta-\frac{1}{2}\right)\frac{R_s\,\rho_s\,\sigma_{sx}\,\sigma_{sy}}{k}$$

Differentiating Equation (4) w.r.t. θ and equating that equation to zero for getting minimum MSE

$$2\left(\theta-\frac{1}{2}\right)\frac{R_S^2\,(\sigma_{sx}^2+\sigma_{sv}^2)}{k}+2\frac{R_S\,\rho_S\,\sigma_{sx}\,\sigma_{sy}}{k}=0$$

$$\Rightarrow\left(\theta-\frac{1}{2}\right)=-\frac{\rho_S\,\sigma_{sx}\,\sigma_{sy}}{R_S(\sigma_{sx}^2+\sigma_{sv}^2)}\qquad\Rightarrow\theta=\frac{1}{2}-\frac{\rho_S\,\sigma_{sx}\,\sigma_{sy}}{R_S(\sigma_{sx}^2+\sigma_{sv}^2)}=\theta_{opt}\ \text{(let)} \tag{5}$$

By replacing the value of θ_{opt} in Equation (4), we have minimum MSE of t_{FT}

$$minM(t_{FT})=\frac{\sigma_{sy}^2+\sigma_{su}^2}{k}+\left(\theta_{opt}-\frac{1}{2}\right)^2\frac{R_S^2\,(\sigma_{sx}^2+\sigma_{sv}^2)}{k}+2\left(\theta_{opt}-\frac{1}{2}\right)\frac{R_S\,\rho_S\,\sigma_{sx}\,\sigma_{sy}}{k}$$

$$minM(t_{FT})=\frac{\sigma_{sy}^2+\sigma_{su}^2}{k}-\frac{\rho_S^2\,\sigma_{sx}^2\sigma_{sy}^2}{k(\sigma_{sx}^2+\sigma_{sv}^2)} \tag{6}$$

Remark 1: ***Bias control estimator t_{FT}*:**

The condition of optimality provides from Equation (5)

$$A_1\theta_{opt}+(\theta_{opt}+1)fA_2+(\theta_{opt}-1)A_3=0$$

$$(\theta_{opt}+1)k_1^3+\left((\theta_{opt}-1)f-8\theta_{opt}-9\right)k_1^2+\left(23\theta_{opt}+26-5(\theta_{opt}-1)f\right)k_1+4(\theta_{opt}-1)f-22\theta_{opt}-24=0\ \ldots(*)$$

This equation (*) is an equation of degree 3 in terms of k.

Obviously, at most three values of k_1 (k_{11}, k_{12}, k_{13}) are possible for which mean squared error is optimum.

The choice criteria for best estimation is

(i) Compute

$$\left|B(t_{FT})_{k_{1j}}\right|\ \text{for}\quad j=1,2,3$$

(ii) From computed values, choose k_j as

$$\left|B(t_{FT})_{k_{1j}}\right|=min\left[\left|B(t_{FT})_{k_{1j}}\right|\right];\ j=1,2,3$$

So, it is clear that the estimator t_{FT} is bias control at the optimum level of MSE.

5. Comparative Study with Other Existing Estimators

Here, we have compared the proposed estimator t_{FT} with existing estimators in literature as discussed in Section 3 of this paper.

When the relative efficiency of two estimators T_1 and T_2 are compared for the same population parameter under similar conditions, then in that case their variances are compared by us. If the variance of T_2 is greater than T_1, then the estimator T_1 will be better than T_2.

Using this concept of theoretical comparison between two estimators, we have derived the conditions of better performance of t_{FT} as follows:

$$V(t_1)-minM(t_{FT})=\frac{\sigma_{sy}^2+\sigma_{su}^2}{k}-\left(\frac{\sigma_{sy}^2+\sigma_{su}^2}{k}-\frac{\rho_S^2\,\sigma_{sx}^2\sigma_{sy}^2}{k(\sigma_{sx}^2+\sigma_{sv}^2)}\right)$$

$$=\frac{\rho_S^2\,\sigma_{sx}^2\sigma_{sy}^2}{k(\sigma_{sx}^2+\sigma_{sv}^2)}>0 \tag{7}$$

$$=\rho_S^2\,\sigma_{sx}^2\sigma_{sy}^2>0\text{(always)}$$

So, the estimator t_{FT} performs better than t_1.

$$M(t_2)-minM(t_{FT})=$$

$$=\frac{\sigma_{sy}^2+\sigma_{su}^2}{k}+\frac{R_S^2\,(\sigma_{sx}^2+\sigma_{sv}^2)}{k}-\frac{2R_S\,\rho_S\,\sigma_{sx}\,\sigma_{sy}}{k}-\left(\frac{\sigma_{sy}^2+\sigma_{su}^2}{k}-\frac{\rho_S^2\,\sigma_{sx}^2\sigma_{sy}^2}{k(\sigma_{sx}^2+\sigma_{sv}^2)}\right)$$

$$=\frac{R_S^2\,(\sigma_{sx}^2+\sigma_{sv}^2)}{k}-\frac{2R_S\,\rho_S\,\sigma_{sx}\,\sigma_{sy}}{k}+\frac{\rho_S^2\,\sigma_{sx}^2\sigma_{sy}^2}{k(\sigma_{sx}^2+\sigma_{sv}^2)} \tag{8}$$

$$=\left(R_S^2\,(\sigma_{sx}^2+\sigma_{sv}^2)-\rho_S\,\sigma_{sx}\,\sigma_{sy}\right)^2>0\qquad\text{(always)}$$

So, the estimator t_{FT} performs better than t_2.

$$M(t_3) - minM(t_{FT}) =$$

$$= \frac{\sigma_{sy}^2 + \sigma_{su}^2}{k} + \frac{R_S^2\,(\sigma_{sx}^2 + \sigma_{sv}^2)}{k} + \frac{2R_S\,\rho_S\,\sigma_{sx}\,\sigma_{sy}}{k} - \left(\frac{\sigma_{sy}^2 + \sigma_{su}^2}{k} - \frac{\rho_S^2\,\sigma_{sx}^2\sigma_{sy}^2}{k(\sigma_{sx}^2 + \sigma_{sv}^2)}\right)$$

$$= \frac{R_S^2\,(\sigma_{sx}^2 + \sigma_{sv}^2)}{k} + \frac{2R_S\,\rho_S\,\sigma_{sx}\,\sigma_{sy}}{k} + \frac{\rho_S^2\,\sigma_{sx}^2\sigma_{sy}^2}{k(\sigma_{sx}^2 + \sigma_{sv}^2)} \tag{9}$$

$$= \left(R_S^2\,(\sigma_{sx}^2 + \sigma_{sv}^2) + \rho_s\,\sigma_{sx}\,\sigma_{sy}\right)^2 > 0 \quad \text{(always)}$$

So, the estimator t_{FT} performs better than t_3.

$$M(t_4) - minM(t_{FT}) =$$

$$= \frac{\sigma_{sy}^2 + \sigma_{su}^2}{k} - \frac{\rho_S^2\,\sigma_{sx}^2\,\sigma_{sy}^2}{k(\sigma_{sx}^2 + \sigma_{sv}^2)} - \left(\frac{\sigma_{sy}^2 + \sigma_{su}^2}{k} - \frac{\rho_S^2\,\sigma_{sx}^2\sigma_{sy}^2}{k(\sigma_{sx}^2 + \sigma_{sv}^2)}\right) \tag{10}$$

$$= 0$$

So, the estimator t_{FT} performs better than t_4.

6. Simulation Study

Using R studio [visit R-Core Team [26], a comparative study of the proposed estimator and other estimators listed in Section 3 has been performed in this section. The characteristics of population regarding study and auxiliary variables are:

$$\mu_{sy} = 15,\ \mu_{sx} = 13,\ \sigma_{sy}^2 = 30,\ \sigma_{su}^2 = 2,\ \sigma_{sx}^2 = 20,\ \sigma_{sv}^2 = 2$$

The following steps are performed under simulation:

(1) Generate $N = 1000$ sized population data related to variables X, Y on R studio using multivariate normal distribution with mean and covariance matrix and measurements errors u and v.
(2) Draw $n = 25$ sized sample using systematic sampling.
(3) Compute mean and variance of random sample drawn.
(4) Calculate mean squared errors for all estimators under consideration.
(5) Repeat steps 1 to 4 for 40,000 times.
(6) Now, calculate efficiency, i.e., $e(t) = \frac{V(t_1)}{M(t)} \times 100$

Under above procedure of simulation for ρ_s = 0.9 the mean squared errors and efficiency of the estimators under consideration are shown in Table 3.

Table 3. MSEs and efficiency of proposed and existing estimators when ρ_s = 0.9.

Estimator	MSE	Efficiency
t_1	32.5738	100
t_2	32.59393	99.93823
t_3	21.59877	150.8132
t_4	10.01299	325.3154153
t_{FT} (k_{11} = 4.37589)	13.21567	246.47854
t_{FT} (k_{12} = 0.63069)	4.633612	702.9894
t_{FT} (k_{13} = 9.97968)	9.9981237	325.7991297

From the above Table 3, we have obtained that t_{FT} (k_{12}) is the best estimators among all the estimators considered in this study. And estimator t_{FT} (k_{13}) gives almost same results as estimator t_4 that is obtained in comparison Section 5 but between these two estimators t_{FT} (k_{13}) have less MSE than estimator t_4.

The Figure 1 represents the diagrammatic representation of comparison of the performance of all estimators from the figure again it is clear that t_{FT} (k_{12}) is the best among all the estimators considered in this study. While t_{FT} (k_{13}) gives approximate same result as estimator t_4.

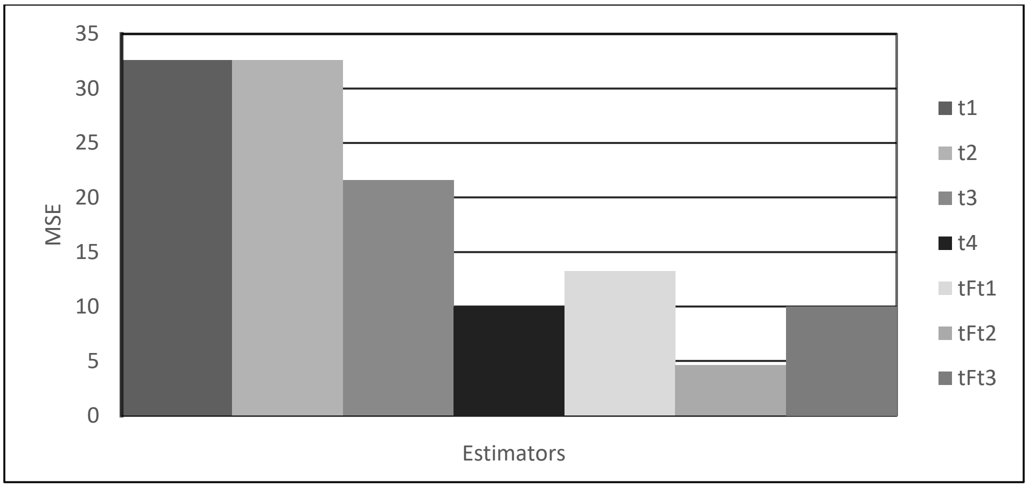

Figure 1. Histogram for Table 3.

Again, we considered correlation between study and auxiliary variable as ρ_s = 0.5 and study the performance of estimators. The Table 4 shows the MSE and the efficiency of estimators.

Table 4. MSEs and efficiency of proposed and existing estimators when ρ_s = 0.5.

Estimator	MSE	Efficiency
t_1	33.5125	100
t_2	33.5832	99.78947807
t_3	20.9877	159.6768584
t_4	11.1095	301.6562402
t_{FT} $(k_{11} = 4.13125)$	12.0215	278.771368
t_{FT} $(k_{12} = 0.56658)$	4.5312	739.5943679
t_{FT} $(k_{13} = 9.75988)$	10.9425	306.2599954

From the above Table 4, we have obtained that t_{FT} (k_{12}) is the best estimators among all the estimators considered in this study. And estimator t_{FT} (k_{13}) gives close results as estimator t_4 that is obtained in comparison Section 5 but between these two estimators t_{FT} (k_{13}) have less MSE than estimator t_4.

The Figure 2 represents the diagrammatic representation of comparison of the performance of all estimators from the figure again it is clear that t_{FT} (k_{12}) is the best among all the estimators considered in this study. While t_{FT} (k_{13}) gives approximate same result as estimator t_4.

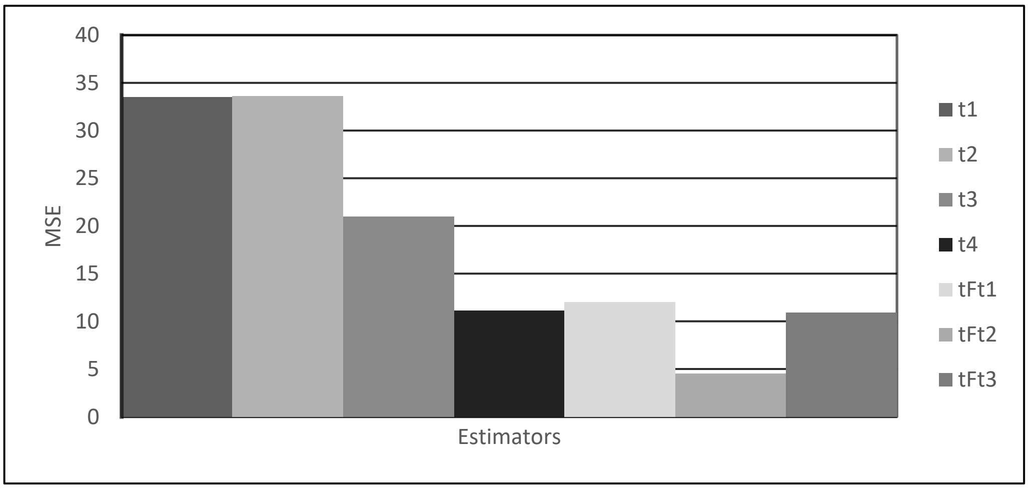

Figure 2. Histogram for Table 4.

Here, we considered correlation between study and auxiliary variable as ρ_s = 0.1 and study the performance of estimators. The Table 5 shows the MSE and the efficiency of estimators.

Table 5. MSEs and efficiency of proposed and existing estimators when ρ_s = 0.1.

Estimator	MSE	Efficiency
t_1	33.9584	100
t_2	34.1452	99.45292457
t_3	18.8157	180.479068
t_4	11.9891	283.2439466
t_{FT} (k_{11} = 4.02154)	13.3042	255.2457119
t_{FT} (k_{12} = 0.45368)	3.8232	888.2192927
t_{FT} (k_{13} = 9.71258)	11.8458	286.6703811

From the above Table 5, we have obtained that $t_{FT}(k_{12})$ is the best estimators with efficiency (888.2192927) among all the estimators considered in this study. And estimator $t_{FT}(k_{13})$ gives close results as estimator t_4 but between these two estimators $t_{FT}(k_{13})$ have high efficency than estimator t_4.

The Figure 3 represents the diagrammatic representation of comparison of the performance of all estimators from the figure again it is clear that $t_{FT}(k_{12})$ is the best among all the estimators considered in this study. While t_{FT} (k_{13}) gives approximate same result as estimator t_4.

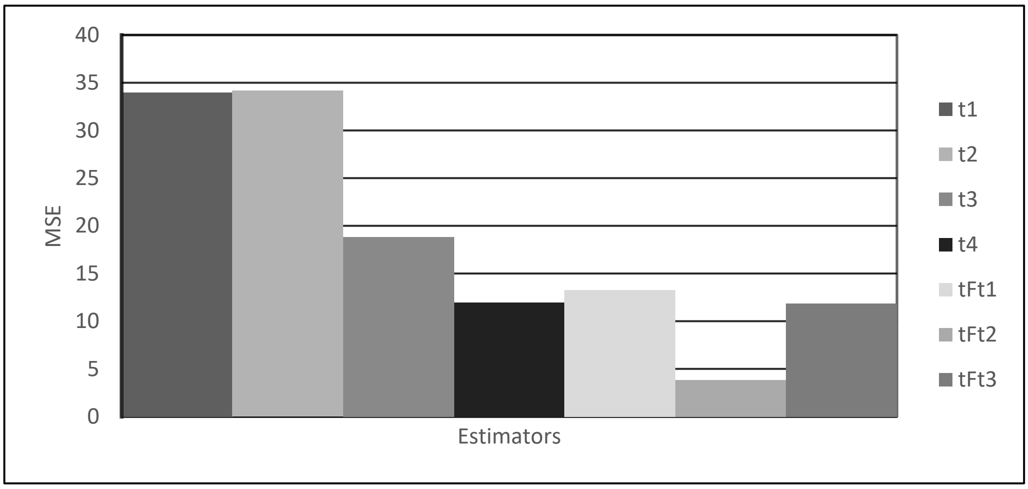

Figure 3. Histogram for Table 5.

Here, we considered correlation between study and auxiliary variable as ρ_s = −0.1 and study the performance of estimators. The Table 6 shows the MSE and the efficiency of estimators.

Table 6. MSEs and efficiency of proposed and existing estimators when ρ_s = −0.1.

Estimator	MSE	Efficiency
t_1	33.5514	100
t_2	34.1524	98.2402408
t_3	18.5714	180.6616626
t_4	11.0011	304.982229
t_{FT} (k_{11} = 4.13459)	12.0402	278.661484
t_{FT} (k_{12} = 0.51429)	3.9012	860.0276838
t_{FT} (k_{13} = 9.84567)	10.9808	305.546044

From the above Table 6, we have obtained that $t_{FT}(k_{12})$ is the best estimators with minimum MSE (i.e., 3.9012) among all the estimators considered here. And estimator $t_{FT}(k_{13})$ gives nearly close results as estimator t_4 but between these two estimators $t_{FT}(k_{13})$ have less MSE than estimator t_4.

The Figure 4 represents the diagrammatic representation of comparison of the performance of all estimators from the figure again it is clear that $t_{FT}(k_{12})$ is the best among all the estimators considered in this study. While t_{FT} (k_{13}) gives approximate same result as estimator t_4.

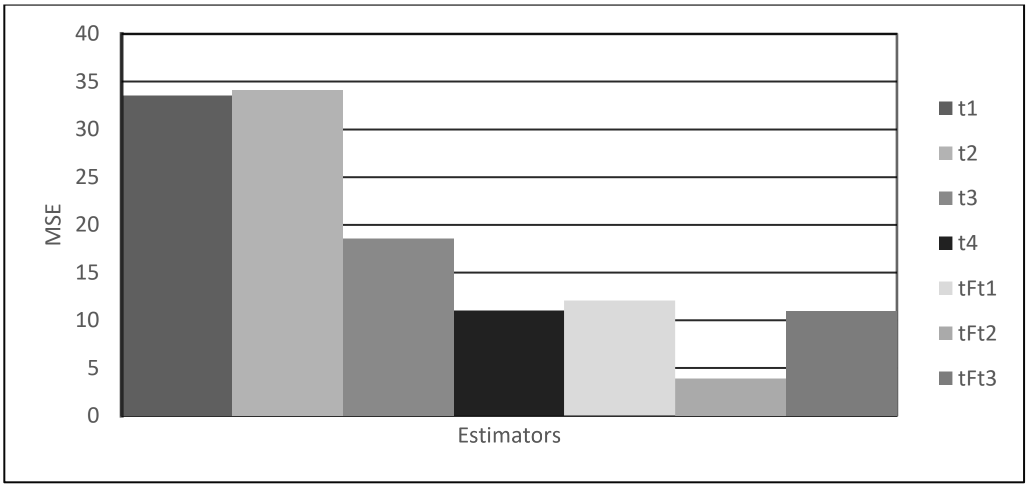

Figure 4. Histogram for Table 6.

Here, we considered correlation between study and auxiliary variable as $\rho_s = -0.5$ and study the performance of estimators. The Table 7 shows the MSE and the efficiency of estimators.

Table 7. MSEs and efficiency of proposed and existing estimators when $\rho_s = -0.5$.

Estimator	MSE	Efficiency
t_1	33.4151	100
t_2	33.8342	98.76131252
t_3	20.7745	160.8467111
t_4	11.0125	303.4288309
t_{FT} $(k_{11} = 4.25358)$	12.1011	276.1327483
t_{FT} $(k_{12} = 0.57859)$	4.1142	812.1894901
t_{FT} $(k_{13} = 9.75416)$	10.9585	304.9240316

From the above Table 7, we have obtained that t_{FT} (k_{12}) is the best estimators among all the estimators considered in this study. And estimator t_{FT} (k_{13}) gives almost same results as estimator t_4 that is obtained in comparison Section 5 but between these two estimators t_{FT} (k_{13}) have less MSE than estimator t_4.

The Figure 5 represents the diagrammatic representation of comparison of the performance of all estimators from the figure again it is clear that t_{FT} (k_{12}) is the best among all the estimators considered in this study. While t_{FT} (k_{13}) gives approximate same result as estimator t_4.

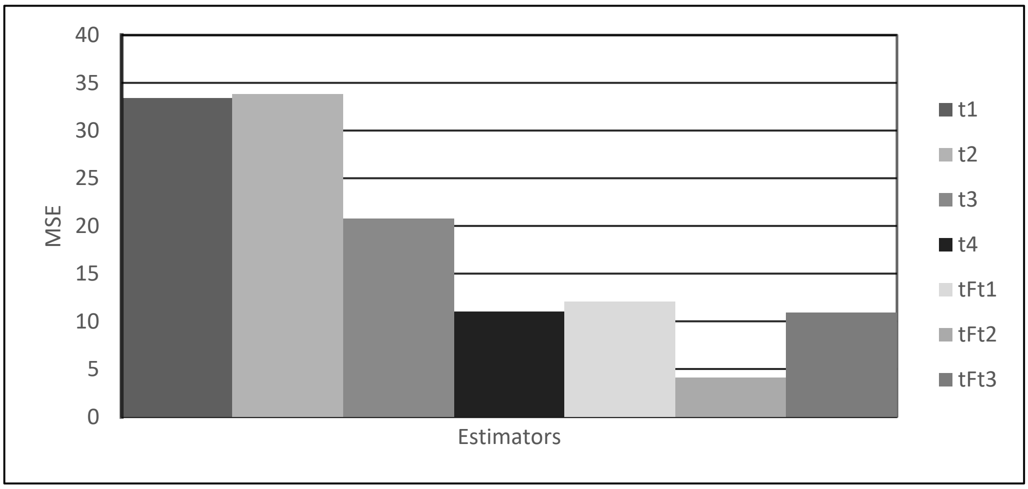

Figure 5. Histogram for Table 7.

Here, we considered correlation between study and auxiliary variable as $\rho_s = -0.9$ and study the performance of estimators. The Table 8 shows the MSE and the efficiency of estimators.

Table 8. MSEs and efficiency of proposed and existing estimators when $\rho_s = -0.9$.

Estimator	MSE	Efficiency
t_1	33.1597	100
t_2	33.1993	99.88072038
t_3	22.5877	146.8042342
t_4	10.1059	328.122186
t_{FT} $(k_{11} = 4.31489)$	12.0157	275.9697729
t_{FT} $(k_{12} = 0.61038)$	4.3612	760.3343117
t_{FT} $(k_{13} = 9.85988)$	9.9847	332.1051208

From the above Table 8, we have obtained that $t_{FT}\,(k_{12})$ is the best estimators among all the estimators considered in this study. And estimator $t_{FT}\,(k_{13})$ gives almost same results as estimator t_4 that is obtained in comparison Section 5 but between these two estimators $t_{FT}\,(k_{13})$ have less MSE than estimator t_4.

The Figure 6 represents the diagrammatic representation of comparison of the performance of all estimators from the figure again it is clear that $t_{FT}\,(k_{12})$ is the best among all the estimators considered in this study. While t_{FT} (k_{13}) gives approximate same result as estimator t_4.

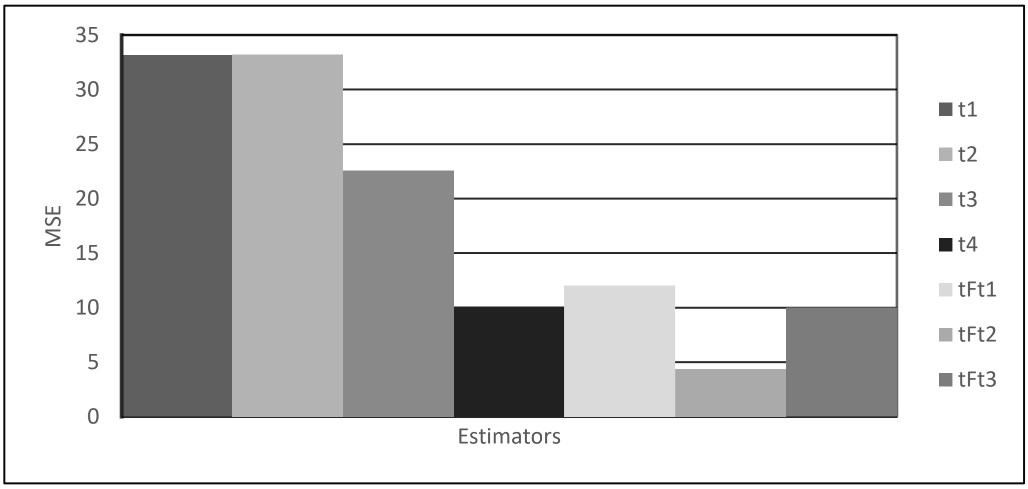

Figure 6. Histogram for Table 8.

7. Result

In this paper, we have proposed a factor type estimator in systematic sampling with measurement errors. The work pertains to minimize the mean squared errors to appropriately infer about the population parameter. The proposed factor-type estimator with measurement errors is theoretically compared with other estimators of population mean under systematic sampling. The conditions derived are verified through data simulated from multivariate normal distribution. The results of the simulation study are shown in above Tables 3–8. These tables show that MSEs and efficiencies of existing and proposed estimators. The proposed estimator t_{FT} is best estimator having highest efficiency and superiority of factor-type (F-T) estimator, having minimum MSE.

8. Discussion and Conclusions

In this manuscript, four well-known forms of estimators, namely ratio, product, difference, and mean estimator under systematic sampling have been proposed when study variable and auxiliary variable are observed with measurement errors.

From our study, we conclude that MSE has always been larger when study and auxiliary variables are commingled with measurement errors. Measurement errors highly affects the MSE as well as efficiency of the estimator when its value is high, but the properties of estimators do not change in the presence of measurement errors. Also, efficiency does not precisely show the effect of measurement errors as it is the ratio of two estimators, i.e., the ratio of MSE (t_1) and MSE (t). By simulation study of the tables, we can infer that MSE is minimum for proposed estimator than all other estimators and is more efficient than ratio, product and difference estimators. Measurement errors are inherent in data observation during systematic sampling, and inference based on these data may be misleading. The inference based on data when the amount of measurement errors is high may ruin the purpose of the study. In the presence of measurement errors, the proposed estimator is most efficient among all estimators under study. Thus, it can be used in the estimation of parameters of mean when the data under study are observed with measurement error.

Author Contributions

A.Y.: Conceptualization, Methodology, Software, Writing-Original draft preparation; S.P.: Conceptualization, Methodology, Visualization, Investigation; N.S.T.: Supervision, Validation, Writing—Reviewing and Editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Acknowledgments

Authors are deeply indebted to the editor-in-chief and learned referees of this journal for their valuable suggestions to improve the quality, contents and presentation of the research manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Singh, D.; Chaudhary, F.S. *Theory and Analysis of Sample Survey Designs*; New Age International (P) Ltd.: New Delhi, India, 2009.
2. Singh, H.P.; Tailor, R.; Jatwa, N.K. Modified ratio and product estimators for population mean in systematic sampling. *J. Mod. Appl. Stat. Methods* **2011**, *10*, 424–435.
3. Singh, R.; Malik, S.; Singh, V.K. An improved estimator in systematic sampling. *arXiv* **2013**, arXiv:1307.5145.
4. Singh, R.; Malik, S.; Chaudhary, M.K.; et al. A general Family of ratio-type estimators in systematic sampling. *J. Reliab. Stat. Stud.* **2012**, *5*, 73–82.
5. Khan, M.; Singh, R. Estimation of population mean in chain ratio type estimator under systematic sampling. *J. Probab. Stat.* **2015**, *2015*, 248374. https://doi.org/10.1155/2015/248374.
6. Pal, S.K.; Singh, H.P. Estimation of finite population mean using auxiliary information in systematic sampling. *Int. J. Syst. Assur. Eng. Manag.* **2017**, *8*, 1392–1398.
7. Singh, R.; Rai, S. Log-type estimators for estimating population mean in systematic sampling in the presence of non-response. *J. Reliab. Stat. Stud.* **2019**, *12*, 105–115.
8. Singh, N.; Vishwakarma, G.K. Computing the effect of measurement errors on the use of auxiliary information under systematic sampling. *Commun. Stat. Theory Methods* **2021**, *51*, 6919–6934. https://doi.org/10.1080/03610926.2020.1868513.
9. Humaidi, A.J.; Mohammad, A.A.; Hameed, A.H.; et al. State estimation of rotary inverted pendulum: A comparative study of observers performance. In Proceedings of the 2020 IEEE Congreso Bienal de Argentina (ARGENCON), Resistencia, Argentina, 1–4 December 2020. https://doi.org/10.1109/ARGENCON49523.2020.9505507.
10. Liu, L.; Long, Z.; Azar, A.T.; et al. Least square algorithm based on bias compensated principle for parameter estimation of canonical state space model. *Meas. Control* **2022**, *55*, 330–339. https://doi.org/10.1177/00202940211064179.
11. Mahmood, N.S.; Humaidi, A.J.; Al-Azzawi, R.S.; et al. Extended state observer design for uncertainty estimation in electronic throttle valve system. *Int. Rev. Appl. Sci. Eng.* **2024**, *15*, 107–115. https://doi.org/10.1556/1848.2023.00662.
12. Shalabh, S.; Tsai, J.R. Ratio and product methods of estimation of population mean in the presence of correlated measurement error. *Commun. Stat. Simul. Comput.* **2017**, *46*, 5566–5593.
13. Singh, N.; Vishwakarma, G.K.; Kim, J.M. Computing the effect of measurement error on efficient variant of ratio and product estimator using auxiliary variable. *Commun. Stat. Simul. Comput.* **2022**, *51*, 604–625. https://doi.org/10.1080/03610918.2019.1656742.
14. Singh, N.; Vishwakarma, G.K. A generalised class of estimator of population mean with the combined effect of measurement errors and non-response in sample survey. *Rev. Investig. Oper.* **2019**, *40*, 275–285.
15. Cochran, W.G. Errors of measurement in statistics. *Technometrics* **1968**, *10*, 37–666. https://doi.org/10.2307/1267450.
16. Shalabh, S. Ratio method of estimation in the presence of measurement error. *J. Indian Soc. Agric. Stat.* **1997**, *50*, 150–155.
17. Maneesha, S.; Singh, R.K. Role of regression estimator involving measurement errors. *Braz. J. Probab. Stat.* **2002**, *16*, 39–46.

18. Shukla, D.; Pathak, S.; Thakur, N.S. An estimator for mean estimation in presence of measurement error. *Res. Rev. A J. Stat.* **2012**, *2*, 1–8.
19. Shukla, D.; Pathak, S.; Thakur, N.S. Class(es) of factor-type estimator in presence of measurement error. *J. Mod. Appl. Stat. Methods* **2012**, *11*, 336–347.
20. Yadav, A.; Thakur, N.S.; Pareek, S. Determining the effect of measurement errors and estimation of mean by factor-type estimator in simple random sampling. *J. Mod. Appl. Stat. Methods* **2024**, *23*, 6. https://doi.org/10.56801/Jmasm.V23.i1.6.
21. Yadav, A.; Thakur, N.S.; Pareek, S. Ratio and product type estimators in post-stratified sampling with measurement errors. *Model Assist. Stat. Appl.* **2025**, *19*, 350–358. https://doi.org/10.1177/15741699241306823.
22. Singh, V.K.; Shukla, D. One parameter family of factor-type ratio estimator. *Metron* **1987**, *45*, 273–283.
23. Shukla, D.; Thakur, N.S. Estimation of mean with imputation of missing data using factor-type estimator. *Stat. Transit.* **2008**, *9*, 33–48.
24. Thakur, N.S.; Shukla, D. Missing data estimation based on the chaining technique in survey sampling. *Stat. Transit.* **2022**, *23*, 91–111. https://doi.org/10.2478/stattrans-2022-0044.
25. Vishwakarma, G.K.; Singh, N. A generalized class of estimators for mean in presence of measurement errors. *Rev. Investig. Oper.* **2020**, *41*, 961–969.
26. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021. Available online: https://www.R-project.org/ (accessed on 5 June 2022).