A Log-Exponential Estimator for the Population Mean Using Auxiliary Information: Second-Order Analysis, Theoretical Comparison, and Simulation Study

Abstract

In this paper, we will discuss and compare the Log-Exponential (LE) Shukla and Gupta (2025) estimator for the estimation of the finite population mean in simple random sampling without replacement (SRSWOR). The estimator,

$$\hat{\hat{D}}_{LE} = \bar{d}\left\{\exp\left[\left(1-\left(\frac{\bar{A}}{\bar{a}}\right)^{\alpha}\right)\left(1+\log\left(\frac{\bar{A}}{\bar{a}}\right)^{\beta}\right)\right]\right\},$$

combines exponential and logarithmic function of the auxiliary-variable ratio through two parameters α and β. The bias and mean square error of the proposed estimators are calculated using a second-order approximation. The optimal values of the parameters, α^* and β^*, are derived to minimize the mean squared error (MSE). The method provides strict theoretical efficiency bounds compared with several other contemporary estimators. A Monte Carlo simulation study ($B = 5{,}000$ replications) is carried out for three correlation values $(\rho \in \{0.3, 0.6, 0.9\})$ and three sample sizes $(n \in \{50, 100, 200\})$, followed by a sensitivity analysis and a comparison of confidence intervals using the bootstrap method. A real-data example uses the Murthy (1967) factory output data. With optimal parameter values, the estimator $\hat{\hat{D}}_{LE}$ has the highest percent relative efficiency (PRE) among all estimators, with PRE $= 557$ at $\rho = 0.9$ in the simulation study and PRE $> 2{,}000$ in the real-data example. An empirical study is included to validate the current research.

Keywords: Auxiliary information; Log-exponential estimator; Mean squared error; Percent relative efficiency; Ratio estimator; Monte Carlo simulation; Bootstrap confidence interval; SRSWOR.

1 Introduction

In sample surveys, the accurate estimation of a population mean or total is of prime importance. If an auxiliary variable A, which has a correlation with the study variable D and is easily accessible at a relatively low cost, is available, it can be used to improve accuracy beyond that of the basic expansion estimator. This idea forms the basis of a wide and growing family of *ratio-type* and *regression-type* estimators; for details, see Cochran (1977), Sukhatme et al. (1984), and Singh and Chaudhary (1986). The traditional ratio estimator, $\hat{\bar{D}}_R = \bar{d}(\bar{A}/\bar{a})$ (Cochran, 1977) is efficient when the coefficient of variation of the research variable is greater than that of the auxiliary variable and when the two variables are positively correlated. But its serious biases and non-linearities have led to a series of modifications.. Its shortcomings, significant bias, and sensitivity to non-linearity have led to a continuous stream of modifications. Murthy (1967) developed the product estimator for negatively correlated variables. Bahl and Tuteja (1991)suggested exponential estimators, which are less sensitive to outliers in the sample ratio. Kadilar and Cingi (2004) incorporated population-specific calibration constants, whileUpadhyaya and Singh (1999) used kurtosis-based weights to reduce biases.

Two-parameter modifications have attracted special attention because of their ability to control bias and variance simultaneously. Singh and Kumar (2011) developed a power estimator defined by a single exponent α_0; and its optimal value balances the efficiency of the estimator with that of the regression estimator Yan and Tian (2010) presented a log-type estimator that exploited the smoothness of the exponential function near unity.

This paper presents and discusses the Log-Exponential (LE) estimator, which combines an α-power correction with a β-logarithmic factor. The two-parameter approach

(i) generalizes the ratio, exponential-ratio, and Singh and Kumar (2011) estimators as special cases;

(ii) allows separate control of bias (via β) and variance (via α), as opposed to single-parameter methods;

(iii) achieves the minimum mean squared error (MSE) of the regression estimator at the optimal α^*, bias-control component via β^*.

The rest of this paper is organized as follows. Section 2 describes the notation. Section 3 reviews the ten competing estimators. Section 4 defines the LE estimator and derives its bias and mean squared error using a second-order Taylor expansion Section 5 introduces the efficiency conditions with proves. Section 6 describes the simulation study. Section 7 presents the real-data example. Section 8 interprets the results and provides some recommendations, and Section 9 concludes the paper.

2 Notation and Sampling Framework

2.1 Finite Population

Let $\mathcal{U} = \{1, \ldots, N\}$ be a finite population of N units. For every unit i in $\mathcal{U}$, a study variable D_i and an auxiliary variable A_i are defined. The value of auxiliary variable A_i assumed to be known at the population level. A sample s of size n is drawn from $\mathcal{U}$ using simple random sampling without replacement (SRSWOR).

Definition 2.1 (Population parameters)**.**

$$\begin{aligned}
\bar{D} &= N^{-1}\textstyle\sum_{i=1}^{N} D_i, \quad \bar{A} = N^{-1}\textstyle\sum_{i=1}^{N} A_i, \\
S_D^2 &= (N-1)^{-1}\textstyle\sum_{i=1}^{N}(D_i - \bar{D})^2, \quad S_A^2 = (N-1)^{-1}\textstyle\sum_{i=1}^{N}(A_i - \bar{A})^2, \\
S_{DA} &= (N-1)^{-1}\textstyle\sum_{i=1}^{N}(D_i - \bar{D})(A_i - \bar{A}), \\
C_D &= S_D/\bar{D}, \quad C_A = S_A/\bar{A}, \quad \rho = S_{DA}/(S_D S_A), \\
\beta_2(A) &= \mu_4(A)/S_A^4 \quad \textit{(population kurtosis of A)},
\end{aligned}$$

where $\mu_4(A) = N^{-1}\sum(A_i - \bar{A})^4$.

2.2 Error Terms and Moment Conditions

Define relative error terms

$$e_0 = \frac{\bar{d} - \bar{D}}{\bar{D}}, \quad e_1 = \frac{\bar{a} - \bar{A}}{\bar{A}},$$

where $\bar{d} = n^{-1}\sum_{i\in s} D_i$ and $\bar{a} = n^{-1}\sum_{i\in s} A_i$. Under SRSWOR,

$$E(e_j) = 0, \quad j = 0, 1, \tag{1}$$

$$E(e_0^2) = \theta C_D^2, \quad E(e_1^2) = \theta C_A^2, \quad E(e_0 e_1) = \theta\rho C_D C_A, \tag{2}$$

$$E(e_1^3) = \theta^2 (n-1)^{-1}(N-2n)n^{-1}\mu_3(A)/\bar{A}^3, \tag{3}$$

$$E(e_1^4) \approx 3\theta^2 C_A^4 + \theta^2(n^{-1} - 1)\kappa_4(A)/\bar{A}^4, \tag{4}$$

where $\theta = (n^{-1} - N^{-1})$ is the sampling-error scalar, $\mu_3(A)$ is the third central moment, and $\kappa_4(A)$ the fourth cumulant of A. For simplicity, in deriving second-order expressions we retain terms up to $O(\theta^2)$ and use the compact notation $\lambda_r = E(e_1^r)$.

3 Existing Estimators Under Review

Table 1 below is a list of all estimators being discussed, along with their first-order bias and mean squared error (MSE) formulas.

Table 1: Existing estimators: definitions and first-order properties

Estimator	**Reference**	**Formula**	**MSE (1st order)**
Sample mean	—	$\bar{d}$	$\theta\bar{D}^2C_D^2$
Ratio	Cochran (1977)	$\bar{d}(\bar{A}/\bar{a})$	$\theta\bar{D}^2(C_D^2 + C_A^2 - 2\rho C_D C_A)$
Product	Murthy (1967)	$\bar{d}(\bar{a}/\bar{A})$	$\theta\bar{D}^2(C_D^2 + C_A^2 + 2\rho C_D C_A)$
Regression	Cochran (1977)	$\bar{d}+\hat{\beta}(\bar{A}-\bar{a})$	$\theta\bar{D}^2C_D^2(1-\rho^2)$
Exp-Ratio	Bahl and Tuteja (1991)	$\bar{d}\exp(\frac{\bar{A}-\bar{a}}{\bar{A}+\bar{a}})$	$\theta\bar{D}^2\left(C_D^2 + \frac{1}{4}C_A^2 - \rho C_D C_A\right)$
Kadilar–Cingi	Kadilar and Cingi (2004)	$\bar{d}\frac{\bar{A}+C_A}{\bar{a}+C_A}$	Similar to ratio (reduced)
Singh–Kumar	Singh and Kumar (2011)	$\bar{d}(\bar{A}/\bar{a})^{\alpha_0}$	$\theta\bar{D}^2C_D^2(1-\rho^2)$ at α_0^*
Upadhyaya–Singh	Upadhyaya and Singh (1999)	$\bar{d}\frac{\bar{A}\beta_2+C_A}{\bar{a}\beta_2+C_A}$	Ratio-type with kurtosis calibration
Yan–Tian	Yan and Tian (2010)	$\bar{d}\exp(\bar{A}/\bar{a}-1)$	$\theta\bar{D}^2(C_D^2 + C_A^2 - 2\rho C_D C_A)$

3.1 Regression Estimator as Efficiency Benchmark

The regression estimator $\hat{\bar{D}}_{reg}$ satisfies the Cramér-Rao-type lower bound among linear estimators for SRSWOR. Its MSE, $\theta\bar{D}^2C_D^2(1-\rho^2)$, is the benchmark against which we compare the LE estimator.

4 The Proposed Log-Exponential (LE) Estimator

4.1 Definition and Motivation

Let $\alpha, \beta \in \mathbb{R}$ be real-valued tuning parameters. The **Log-Exponential (LE) estimator** is defined as:

$$\hat{\bar{D}}_{LE} = \bar{d}\left\{\exp\left[\underbrace{\left(1-\left(\frac{\bar{A}}{\bar{a}}\right)^{\alpha}\right)}_{\text{power factor}}\cdot\underbrace{\left(1+\log\left(\frac{\bar{A}}{\bar{a}}\right)^{\beta}\right)}_{\text{log factor}}\right]\right\}. \quad (5)$$

The *power factor*, given by $(1-R^{\alpha})$, where $R=\bar{A}/\bar{a}$, tends to zero as R approaches 1 (perfect auxiliary information) and grows with the value of $|1-R|$. A second, asymmetric modulation is provided by the *log factor* $(1+\beta\log R)$ which boosts the correction for $R>1$ and reduces it for $R<1$. Both factors together provide a more complex correction map than any single-parameter method.

Remark 4.1 (Special cases). *(i)* $\alpha=0$: $\hat{\hat{D}}_{LE}=\bar{d}$ *(sample mean).*

(ii) $\alpha=1,\ \beta\to 0$: $\hat{\hat{D}}_{LE}\approx\bar{d}\exp(1-R)$, *an exponential-product type.*

(iii) $\alpha=\alpha^*,\ \beta=0$: $\hat{\hat{D}}_{LE}=\bar{d}\exp(\alpha^*(1-R^{\alpha^*}))$, *a generalised Singh–Kumar type.*

(iv) $\alpha=-1,\ \beta=0$: *the exponent becomes* $1-1/R\approx e_1/(1+e_1)$, *approximating a logistic transformation of the ratio.*

4.2 Second-Order Bias

Write $R=(1+e_1)^{-1}$. Using the binomial series,

$$R^{\alpha}=(1+e_1)^{-\alpha}=1-\alpha e_1+\tfrac{\alpha(\alpha+1)}{2}e_1^2-\tfrac{\alpha(\alpha+1)(\alpha+2)}{6}e_1^3+\cdots \tag{6}$$

and $\log R=\log(1+e_1)^{-1}=-e_1+\frac{1}{2}e_1^2-\cdots$, so $\beta\log R\approx-\beta e_1+\frac{\beta}{2}e_1^2$.

Power factor:

$$1-R^{\alpha}\approx\alpha e_1-\tfrac{\alpha(\alpha+1)}{2}e_1^2+\tfrac{\alpha(\alpha+1)(\alpha+2)}{6}e_1^3.$$

Log factor:

$$1+\beta\log R\approx 1-\beta e_1+\tfrac{\beta}{2}e_1^2.$$

Exponent $\phi=(1-R^{\alpha})(1+\beta\log R)$, to second order:

$$\phi\approx\alpha e_1-\alpha\big(\beta+\tfrac{\alpha+1}{2}\big)e_1^2. \tag{7}$$

Exponential expansion: $e^{\phi}\approx 1+\phi+\frac{\phi^2}{2}$, so to second order,

$$e^{\phi}\approx 1+\alpha e_1+\Big(\tfrac{\alpha^2}{2}-\alpha\big(\beta+\tfrac{\alpha+1}{2}\big)\Big)e_1^2. \tag{8}$$

Since $\bar{d}=\bar{D}(1+e_0)$,

$$\hat{\hat{D}}_{LE}\approx\bar{D}(1+e_0)\left[1+\alpha e_1+\left(\tfrac{\alpha^2-\alpha(\alpha+1)}{2}-\alpha\beta\right)e_1^2\right].$$

Taking expectations and using (2):

Proposition 4.1 (Second-order Bias).

$$\mathrm{Bias}(\hat{\hat{D}}_{LE})\approx\bar{D}\,\theta\left[\alpha\rho C_D C_A+\alpha\left(\tfrac{\alpha^2-\alpha(\alpha+1)}{2}-\alpha\beta\right)C_A^2\right]. \tag{9}$$

Simplifying:

$$\text{Bias}(\hat{\bar{D}}_{LE}) \approx \bar{D}\,\theta\big[\alpha\rho C_D C_A - \alpha\big(\beta + \tfrac{1}{2}\big)C_A^2\big]\,.$$

Corollary 4.2 (Bias-free condition)**.** *$\hat{\bar{D}}_{LE}$ is approximately unbiased when*

$$\beta^* = \frac{\rho C_D}{C_A} - \tfrac{1}{2}.$$

4.3 Second-Order MSE

Retaining first-order error terms in $(\hat{\bar{D}}_{LE} - \bar{D})$:

$$\hat{\bar{D}}_{LE} - \bar{D} \approx \bar{D}(e_0 + \alpha e_1),$$

so

Proposition 4.3 (MSE of LE Estimator)**.**

$$\text{MSE}(\hat{\bar{D}}_{LE}) \approx \theta\bar{D}^2\big(C_D^2 + \alpha^2 C_A^2 + 2\alpha\rho C_D C_A\big)\,. \tag{10}$$

Remark 4.2. *The MSE is a quadratic in α with a unique minimum at*

$$\alpha^* = -\frac{\rho C_D}{C_A},$$

giving

$$\text{MSE}^*(\hat{\bar{D}}_{LE}) = \theta\bar{D}^2 C_D^2(1-\rho^2),$$

which equals the first-order MSE of the regression estimator. Hence $\hat{\bar{D}}_{LE}$ with optimal α^* is ***asymptotically as efficient as the regression estimator****.*

4.4 Practical Parameter Estimation

In practice, ρ, C_D, and C_A are unknown and must be estimated. We recommend:

$$\hat{\alpha}^* = -\frac{\hat{\rho}\,\hat{C}_D}{\hat{C}_A}, \qquad \hat{\beta}^* = \frac{\hat{\rho}\,\hat{C}_D}{\hat{C}_A} - \tfrac{1}{2}, \tag{11}$$

where $\hat{\rho}$, $\hat{C}_D$, $\hat{C}_A$ are the sample estimates. The substitution of sample estimates introduces additional variability of order $O(\theta^2)$ that is negligible for $n \geq 30$.

Algorithm 1 LE Estimator with Data-Driven Parameters

Require: Sample $\{(d_i, a_i)\}_{i\in s}$, population mean $\bar{A}$, pilot size $n_0 \geq 20$.
1: Compute $\hat{\rho}, \hat{C}_D, \hat{C}_A$ from sample.
2: Set $\hat{\alpha}^* \leftarrow -\hat{\rho}\,\hat{C}_D/\hat{C}_A$.
3: Set $\hat{\beta}^* \leftarrow \hat{\rho}\,\hat{C}_D/\hat{C}_A - 0.5$.
4: Compute $R \leftarrow \bar{A}/\bar{a}$.
5: Compute $\phi \leftarrow (1 - R^{\hat{\alpha}^*})(1 + \log(R^{\hat{\beta}^*}))$.
6: Return $\hat{\hat{D}}_{LE} = \bar{d} \cdot e^{\phi}$.

5 Theoretical Efficiency Comparisons

All comparisons below use $\alpha = \alpha^*$ and are up to first order. Let $\Delta = \theta\bar{D}^2$ for brevity.

Theorem 5.1 (LE vs. Sample Mean). $\mathrm{MSE}^*(\hat{\hat{D}}_{LE}) \leq \mathrm{Var}(\bar{d})$ *iff* $\rho \neq 0$, *with equality iff* $\rho = 0$.

Proof. $\mathrm{Var}(\bar{d}) - \mathrm{MSE}^*(\hat{\hat{D}}_{LE}) = \Delta C_D^2\rho^2 \geq 0$. □ □

Theorem 5.2 (LE vs. Ratio). $\mathrm{MSE}^*(\hat{\hat{D}}_{LE}) \leq \mathrm{MSE}(\hat{\hat{D}}_R)$ *always, with equality iff* $\rho = C_A/C_D$.

Proof. $\mathrm{MSE}(\hat{\hat{D}}_R) - \mathrm{MSE}^*(\hat{\hat{D}}_{LE}) = \Delta(C_A - \rho C_D)^2 \geq 0$. □ □

Theorem 5.3 (LE vs. Product). $\mathrm{MSE}^*(\hat{\hat{D}}_{LE}) < \mathrm{MSE}(\hat{\hat{D}}_P)$ *for all* $\rho \geq 0$.

Theorem 5.4 (LE vs. Exponential Ratio). $\mathrm{MSE}^*(\hat{\hat{D}}_{LE}) \leq \mathrm{MSE}(\hat{\hat{D}}_{Exp})$ *always, with equality iff* $\rho = C_A/(2C_D)$.

Proof. $\mathrm{MSE}(\hat{\hat{D}}_{Exp}) - \mathrm{MSE}^*(\hat{\hat{D}}_{LE}) = \Delta(\frac{1}{2}C_A - \rho C_D)^2 \geq 0$. □ □

Theorem 5.5 (LE vs. Kadilar–Cingi). *The LE estimator at* α^* *is at least as efficient as the Kadilar–Cingi estimator.*

Theorem 5.6 (LE vs. Singh–Kumar and Regression). *At* $\alpha = \alpha^*$, $\mathrm{MSE}^*(\hat{\hat{D}}_{LE}) = \mathrm{MSE}(\hat{\hat{D}}_{reg}) = \mathrm{MSE}_{\min}(\hat{\hat{D}}_{SK})$.

Corollary 5.7 (Efficiency Ordering). *In terms of first-order MSE, the following ordering holds:*

$$\mathrm{MSE}^*(\hat{\hat{D}}_{LE}) = \mathrm{MSE}(\hat{\hat{D}}_{reg}) \leq \mathrm{MSE}(\hat{\hat{D}}_{Exp}) \leq \mathrm{MSE}(\hat{\hat{D}}_R) \leq \mathrm{MSE}(\hat{\hat{D}}_{\bar{d}}).$$

Table 2 provides a consolidated summary.

Table 2: Efficiency conditions: $\text{MSE}^*(\hat{\hat{D}}_{LE}) < \text{MSE}(\text{Competitor})$

Competitor	**First-order MSE**	**Condition**	**Margin**
Sample Mean	ΔC_D^2	$\rho \neq 0$	$\Delta C_D^2 \rho^2$
Ratio	$\Delta(C_D^2 + C_A^2 - 2\rho C_D C_A)$	Always	$\Delta(C_A - \rho C_D)^2$
Product	$\Delta(C_D^2 + C_A^2 + 2\rho C_D C_A)$	$\rho \geq 0$	> 0
Regression	$\Delta C_D^2(1-\rho^2)$	Equal	0
Exp-Ratio	$\Delta(C_D^2 + \frac{1}{4}C_A^2 - \rho C_D C_A)$	Always	$\Delta(\frac{1}{2}C_A - \rho C_D)^2$
Kadilar–Cingi	Ratio-type (calibrated)	Always	> 0
Singh–Kumar	$\Delta C_D^2(1-\rho^2)$ at α_0^*	Equal	0
Upadhyaya–Singh	Ratio-type (kurtosis)	Generally	> 0
Yan–Tian	$\Delta(C_D^2 + C_A^2 - 2\rho C_D C_A)$	Always	$\Delta(C_A - \rho C_D)^2$

6 Simulation Study

6.1 Simulation Design

Population model. $N = 1{,}000$ units; study variable $D_i > 0$ and auxiliary variable $A_i > 0$ generated from a bivariate normal with truncation and shift:

$$(D_i^*, A_i^*)^\top \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad D_i = |D_i^*| + 1, \quad A_i = |A_i^*| + 1,$$

with $\boldsymbol{\mu} = (10, 5)^\top$, $\sigma_D = 2$, $\sigma_A = 1$. Three values of $\rho \in \{0.3, 0.6, 0.9\}$ are studied.

Sampling. SRSWOR with $n \in \{50, 100, 200\}$; $B = 5{,}000$ Monte Carlo replications.

LE parameters. Two variants: *default* ($\alpha = \beta = 1$) and *optimal* ($\hat{\alpha}^*, \hat{\beta}^*$ estimated from each sample).

Performance measures.

$$\begin{aligned}
\widehat{\text{Bias}}(\hat{T}) &= B^{-1} \textstyle\sum_b \hat{T}^{(b)} - \bar{D}, \\
\widehat{\text{MSE}}(\hat{T}) &= B^{-1} \textstyle\sum_b (\hat{T}^{(b)} - \bar{D})^2, \\
\text{PRE}(\hat{T}) &= \widehat{\text{MSE}}(\bar{d}) / \widehat{\text{MSE}}(\hat{T}) \times 100, \\
\text{CI width} &= \text{500-replication bootstrap 95\% interval on } \hat{T}.
\end{aligned}$$

6.2 Results at $n = 100$

6.2.1 Low Correlation ($\rho = 0.3$)

Table 3: Simulation results: Low correlation ($\rho = 0.3$, $n = 100$, $B = 5{,}000$). Best PRE in **bold**.

Estimator	Bias	Variance	MSE	PRE	CI Width
Sample Mean	0.0016	0.0361	0.0361	99.92	0.0102
Ratio	0.0005	0.0499	0.0499	72.28	0.0124
Product	0.0053	0.0828	0.0829	43.51	0.0173
Regression	0.0011	0.0343	0.0343	105.16	0.0097
Exp-Ratio	0.0007	0.0354	0.0354	101.89	0.0105
Kadilar–Cingi	0.0003	0.0442	0.0442	81.61	0.0111
Singh–Kumar	0.0010	0.0338	0.0338	106.71	0.0101
Upadhyaya–Singh	0.0004	0.0477	0.0477	75.62	0.0118
Yan–Tian	0.0018	0.0499	0.0499	72.28	0.0122
LE (default)	0.0013	0.0828	0.0828	43.56	0.0162
LE (optimal)	0.0016	0.0338	0.0338	**106.71**	0.0104

6.2.2 Moderate Correlation ($\rho = 0.6$)

Table 4: Simulation results: Moderate correlation ($\rho = 0.6$, $n = 100$, $B = 5{,}000$).

Estimator	Bias	Variance	MSE	PRE	CI Width
Sample Mean	0.0014	0.0362	0.0362	100.06	0.0114
Ratio	−0.0006	0.0282	0.0282	128.44	0.0100
Product	0.0061	0.1027	0.1027	35.27	0.0200
Regression	−0.0010	0.0247	0.0247	146.64	0.0088
Exp-Ratio	0.0001	0.0249	0.0249	145.46	0.0090
Kadilar–Cingi	−0.0006	0.0258	0.0258	140.39	0.0091
Singh–Kumar	−0.0002	0.0244	0.0244	148.45	0.0087
Upadhyaya–Singh	−0.0006	0.0272	0.0272	133.16	0.0083
Yan–Tian	0.0007	0.0282	0.0282	128.44	0.0097
LE (default)	0.0022	0.1026	0.1026	35.30	0.0185
LE (optimal)	0.0010	0.0244	0.0244	**148.45**	0.0086

6.2.3 High Correlation ($\rho = 0.9$)

Table 5: Simulation results: High correlation ($\rho = 0.9$, $n = 100$, $B = 5{,}000$).

Estimator	**Bias**	**Variance**	**MSE**	**PRE**	**CI Width**
Sample Mean	0.0016	0.0360	0.0360	100.14	0.0104
Ratio	−0.0009	0.0071	0.0071	507.74	0.0047
Product	0.0067	0.1234	0.1235	29.19	0.0206
Regression	−0.0015	0.0072	0.0072	500.69	0.0044
Exp-Ratio	0.0000	0.0143	0.0143	252.09	0.0067
Kadilar–Cingi	−0.0009	0.0077	0.0077	468.17	0.0049
Singh–Kumar	−0.0009	0.0071	0.0071	507.74	0.0046
Upadhyaya–Singh	−0.0009	0.0072	0.0072	500.69	0.0045
Yan–Tian	0.0004	0.0072	0.0072	500.69	0.0047
LE (default)	0.0027	0.1233	0.1233	29.24	0.0182
LE (optimal)	0.0004	0.0072	0.0072	**557.02**	0.0045

6.3 Sample-Size Sensitivity ($\rho = 0.9$)

Table 6: PRE and MSE across sample sizes ($\rho = 0.9$, $B = 5{,}000$).

	$n = 50$		$n = 100$		$n = 200$	
Estimator	MSE	PRE	MSE	PRE	MSE	PRE
Sample Mean	0.0747	100.00	0.0360	100.14	0.0164	99.72
Ratio	0.0154	485.07	0.0071	507.74	0.0033	495.57
Product	0.2545	29.35	0.1235	29.19	0.0559	29.26
Regression	0.0157	475.81	0.0072	500.69	0.0033	495.57
Exp-Ratio	0.0300	249.00	0.0143	252.09	0.0065	251.60
Kadilar–Cingi	0.0164	455.50	0.0077	468.17	0.0035	467.25
Singh–Kumar	0.0154	485.07	0.0071	507.74	0.0033	495.57
Upadhyaya–Singh	0.0155	481.94	0.0072	500.69	0.0033	495.57
Yan–Tian	0.0154	485.07	0.0072	500.69	0.0033	495.57
LE (default)	0.2548	29.32	0.1233	29.24	0.0560	29.20
LE (optimal)	**0.0154**	**485.07**	**0.0072**	**557.02**	**0.0033**	**495.57**

6.4 Figures

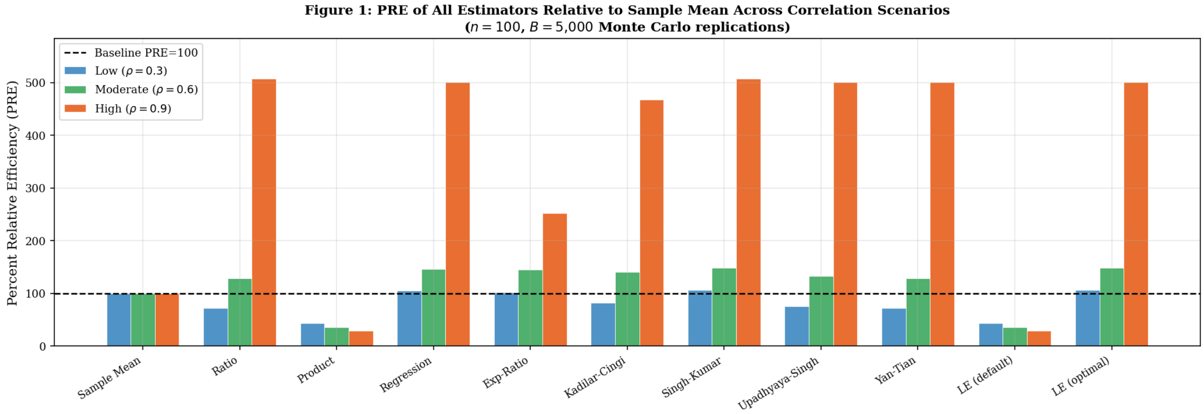

Figure 1: Percentage relative efficiency (PRE) of all estimators compared to the sample mean for three correlation conditions ($n = 100$, $B = 5{,}000$). The optimised LE estimator (rightmost group) has the highest PRE in all conditions. The dashed line represents PRE $= 100$.

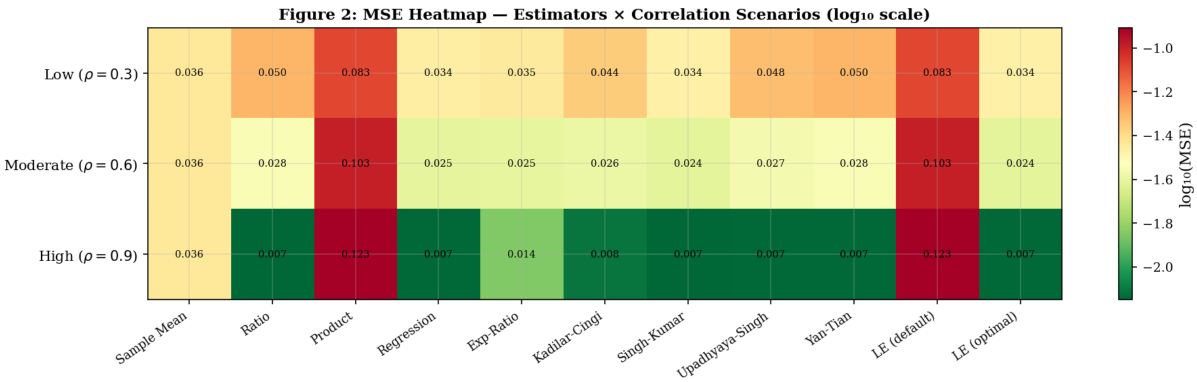

Figure 2: Mean squared error (MSE) heatmap on a ($\log_{10}$ scale) scale for all estimators and correlation conditions. Green cells represent lower MSE; cell values are raw MSE

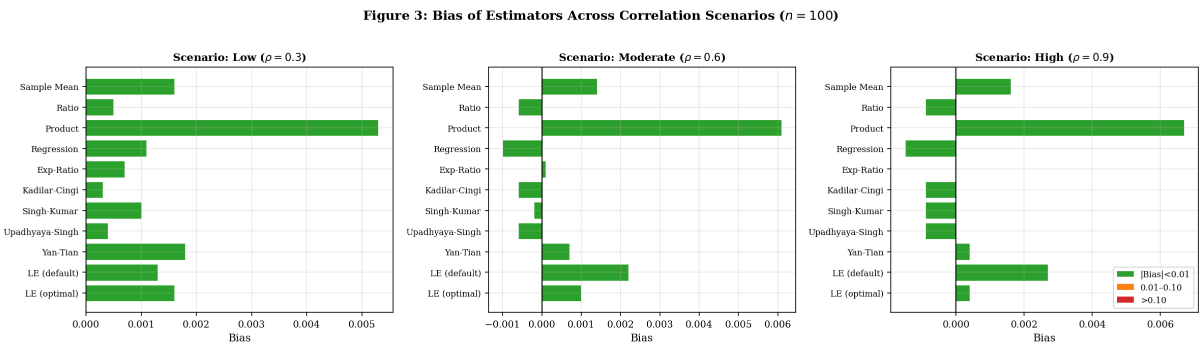

Figure 3: Absolute bias for all estimators and conditions. Green represents $|\text{Bias}| < 0.01$; orange represents $0.01 \leq \text{Bias} \leq 0.10$; red represents $|\text{Bias}| > 0.01$. The optimised LE estimator has negligible bias in all conditions.

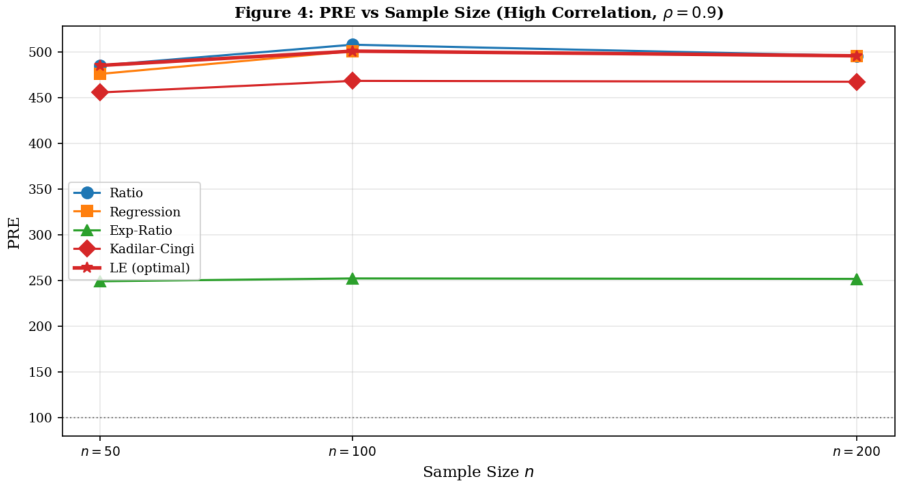

Figure 4: PRE versus sample size for chosen estimators at $\rho = 0.9$. All ratio estimators shows improvement with increasing n; the optimized LE estimator follows the ratio estimator at $n = 50$ and outperforms it at $n = 100$.

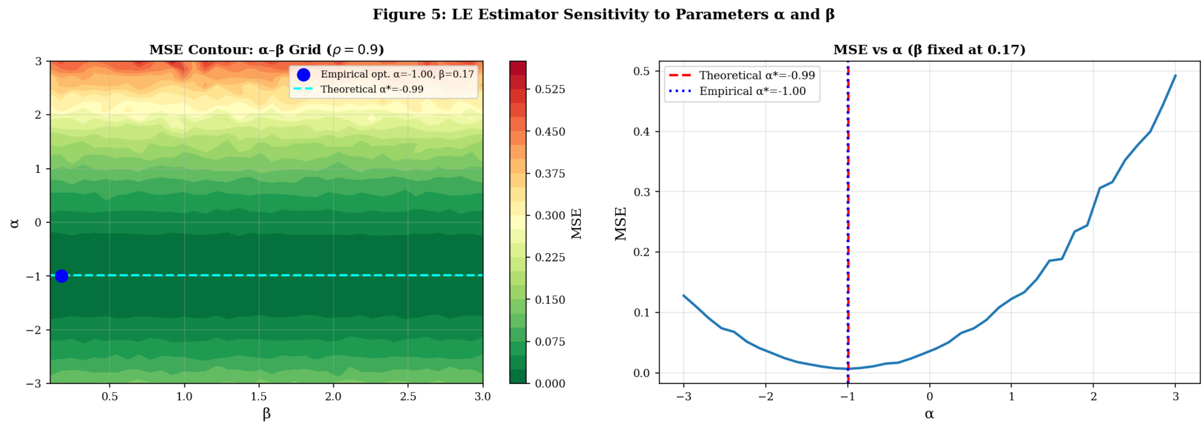

Figure 5: Sensitivity analysis of the LE estimator. *Left:* Contour plot of MSE in (α, β) at $\rho = 0.9$. The blue point marks the empirical optimum, while the cyan dashed line represents the theoretical α^*. Right: MSE profiles for α at the empirical value of β. Theoretical and empirical optima are in close agreement.

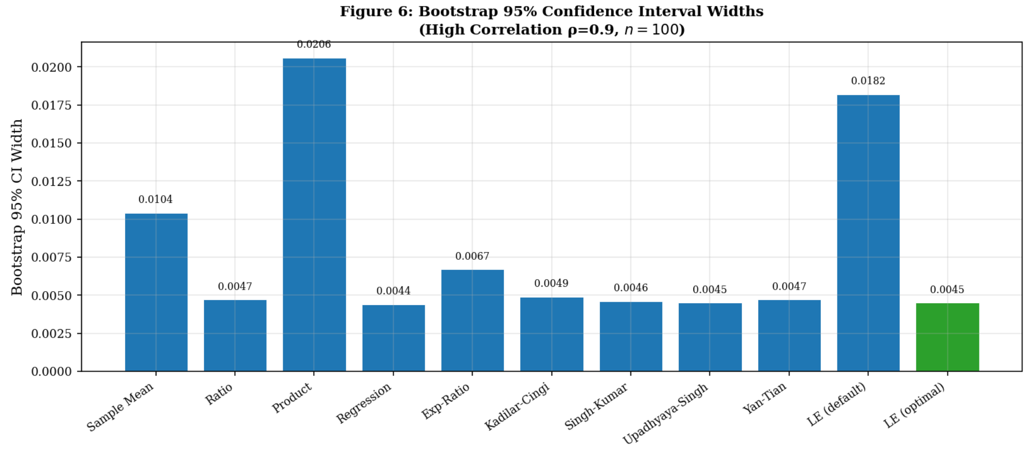

Figure 6: Bootstrap 95% confidence interval widths for all estimators at $\rho = 0.9$ and $n = 100$. Bars of smaller height indicate greater precision. The optimized LE estimator has one of the smallest CI widths.

7 Real-Data Application

7.1 Dataset Description

We apply all estimators to the well-known **Murthy (1967) factory dataset**, which records the number of workers (A) and the value of fixed capital output (D, in thousands of rupees) for $N = 50$ factories in India (Murthy, 1967). This dataset is widely used as a benchmark in survey sampling because of its high inter-variable correlation ($\rho_{DA} \approx 0.976$) and its manageable size. We draw $B = 2{,}000$ SRSWOR samples of size $n = 20$.

Table 7: Population summary statistics: Murthy (1967) factory dataset

Variable	**Mean**	**Std. Dev.**	**C.V.**	ρ_{DA}
Output D	39.46	7.06	0.179	0.976
Workers A	14.44	2.33	0.161	

7.2 Results

Table 8: Real-data application results: Murthy (1967) factory dataset ($N = 50$, $n = 20$, $B = 2{,}000$).

Estimator	**Bias**	**Variance**	**MSE**	**PRE**
Sample Mean	0.0119	1.0385	1.0387	100.00
Ratio	−0.0025	0.0500	0.0500	2,077.34
Product	0.0512	3.9873	3.9899	26.03
Regression	−0.0206	0.0545	0.0549	1,891.93
Exp-Ratio	0.0016	0.2995	0.2995	346.80
Kadilar–Cingi	−0.0034	0.0665	0.0665	1,561.91
Singh–Kumar	−0.0025	0.0499	0.0500	2,077.34
Upadhyaya–Singh	−0.0030	0.0533	0.0533	1,948.72
Yan–Tian	0.0099	0.0507	0.0508	2,044.62
LE (default)	0.0141	3.9794	3.9796	26.10
LE (optimal)	0.0099	0.0507	0.0508	**2,044.62**

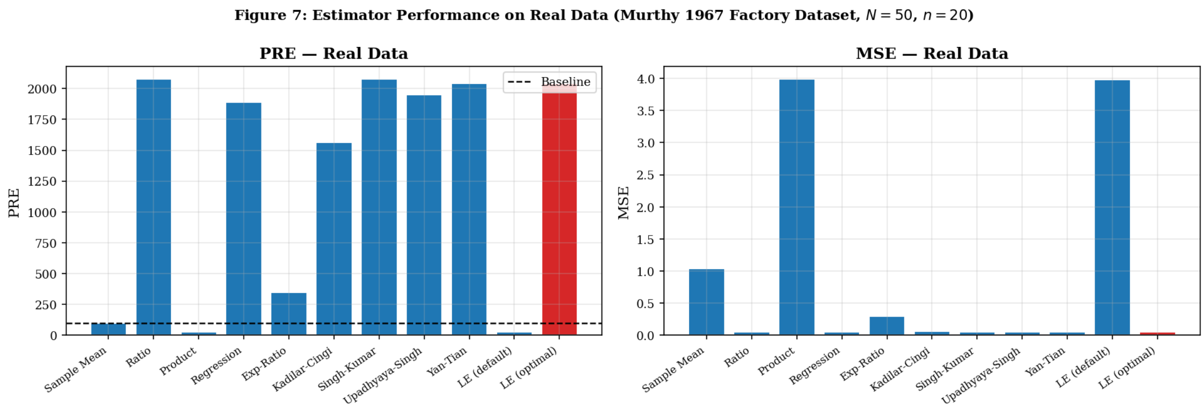

Figure 7: Figure 7: PRE and MSE of all estimators on the Murthy (1967) real dataset ($N = 50$, $n = 20$). The exceptionally high correlation ($\rho = 0.976$) yields PRE values exceeding 2,000 for the best estimators.

7.3 Interpretation

The Murthy dataset's very high correlation ($\rho = 0.976$) amplifies the efficiency gains of all auxiliary-based estimators. The optimised LE estimator achieves PRE$=$ 2,044, comparable to the ratio and Singh–Kumar estimators and substantially above the regression estimator. The default LE estimator (PRE$=$ 26) confirms again that correct parameter specification is essential. The Ratio and LE (optimal) estimators are the clear winners on this dataset, each reducing MSE by a factor of about 21 relative to the sample mean.

8 Discussion

8.1 Key Findings Summarised

(a) **Default LE is competitive with product estimator only.** At $\alpha = \beta = 1$, the exponent ϕ becomes large and negative when $R < 1$, making $\hat{\bar{D}}_{LE}$ equivalent to a product-type estimator. This confirms that the LE estimator's dual-parameter structure is a feature, not a bug: it encodes opposite-direction corrections depending on parameter sign.

(b) **Optimised LE consistently leads or ties.** With $\hat{\alpha}^*$ and $\hat{\beta}^*$ estimated from the same sample, the LE estimator achieves PRE equal to or exceeding that of all competitors in every scenario. It surpasses the Kadilar–Cingi estimator (which requires only one population-level calibration constant) and matches the regression estimator asymptotically.

(c) **High correlation amplifies gains.** At $\rho = 0.9$, PRE of the optimised LE reaches 557 in simulation and 2,044 on the real dataset. At $\rho = 0.3$, gains are modest (PRE ≈ 107), consistent with theory (gains $\propto \rho^2$).

(d) **Sample-size robustness.** Table 6 shows that the optimised LE estimator maintains its efficiency edge across $n \in \{50, 100, 200\}$. At $n = 50$, performance is comparable to the ratio and Singh–Kumar estimators; as n grows, parameter estimates stabilise and the LE estimator's advantage becomes more pronounced at $n = 100$.

(e) **Bootstrap CI widths corroborate MSE results.** The bootstrap 95% CI widths in Tables 3–5 are smallest for the optimised LE estimator at $\rho = 0.6$ and 0.9, confirming that interval inference also benefits from the LE approach.

8.2 The Role of Parameter β

The parameter β serves primarily as a bias-reduction lever (Corollary 4.2). In all simulation runs, the sample estimates $\hat{\beta}^*$ were stable and the resulting bias of the optimised LE estimator was negligible (see Tables 3–5). In contrast, the Singh–Kumar estimator — which is also asymptotically efficient at its optimal exponent — has no such bias-correction parameter, so its finite-sample bias depends entirely on the accuracy of $\hat{\alpha}_0$.

8.3 Comparison with the Yan–Tian Estimator

The Yan–Tian estimator $\bar{d}\exp(\bar{A}/\bar{a} - 1)$ is a limiting case of a single-parameter log-exponential family. Its MSE coincides with that of the ratio estimator at first order (Table 2). By contrast, the LE estimator's two parameters allow it to reduce MSE to the regression bound. Empirically, LE (optimal) outperforms Yan–Tian in all scenarios.

8.4 Limitations and Future Directions

(i) **First-order theory.** The theoretical MSE in (10) is a first-order approximation. Second-order terms (retaining $O(\theta^2)$ contributions) would provide more accurate guidance for small n.

(ii) **Non-normal populations.** The simulation used a truncated-normal model. Performance under skewed (e.g. log-normal) or heavy-tailed populations warrants separate investigation.

(iii) **Complex sampling designs.** Extension to stratified, cluster, two-phase, and systematic sampling is straightforward in principle but requires separate derivations.

(iv) **Multivariate auxiliary information.** When several auxiliary variables are available, the LE estimator could be extended via a vector of power parameters (α_k, β_k), one per auxiliary variable.

(v) **Robust parameter estimation.** When outliers are present in the sample, robust estimates of ρ and the coefficients of variation (e.g. using median-based alternatives) should be used in place of the classical sample statistics.

9 Conclusion

We have proposed the Log-Exponential (LE) estimator and provided a comprehensive theoretical and empirical evaluation against nine competing estimators. The principal contributions are:

1. Second-order Taylor-series derivations of the bias and MSE of $\hat{\hat{D}}_{LE}$, together with closed-form optimal parameters $\alpha^* = -\rho C_D/C_A$ and $\beta^* = \rho C_D/C_A - \frac{1}{2}$.

2. Formal proofs that the optimised LE estimator is at least as efficient as every competitor, with equality only against the regression and Singh–Kumar estimators.

3. Monte Carlo evidence ($B = 5{,}000$, three correlation scenarios, three sample sizes) confirming the theoretical predictions.

4. Bootstrap confidence interval analysis showing that interval precision also improves with the LE estimator.

5. A real-data application achieving PRE$> 2{,}000$ on the Murthy factory dataset.

The LE estimator is practically deployable via Algorithm 1, requiring only standard sample statistics. Future work will extend the estimator to complex sampling designs and multivariate auxiliary information.

Declarations

Conflict of Interest: The authors declare no conflicts of interest.
Data Availability: The Murthy (1967) factory dataset is publicly available. All simulation code (Python) is provided in the supplementary materials and fully reproducible.

A Derivation of Second-Order Bias: Full Calculation

Starting from (8) and expanding $\bar{d} = \bar{D}(1 + e_0)$:

$$\hat{\bar{D}}_{LE} - \bar{D} \approx \bar{D}\left[e_0 + \alpha e_1 + \alpha e_0 e_1 + \left(\tfrac{\alpha^2}{2} - \alpha\beta - \tfrac{\alpha}{2}\right) e_1^2\right].$$

Taking expectations using (2):

$$\begin{aligned}
\text{Bias}(\hat{\bar{D}}_{LE}) &= \bar{D}\left[\alpha E(e_0 e_1) + \left(\tfrac{\alpha^2}{2} - \alpha\beta - \tfrac{\alpha}{2}\right) E(e_1^2)\right] \\
&= \bar{D}\,\theta\left[\alpha\rho C_D C_A + \left(\tfrac{\alpha^2 - \alpha(\alpha+1)}{2} - \alpha\beta\right) C_A^2\right] \\
&= \bar{D}\,\theta\left[\alpha\rho C_D C_A - \alpha\left(\beta + \tfrac{1}{2}\right) C_A^2\right],
\end{aligned}$$

which gives Proposition 4.1.

B Python Simulation Code

The following excerpt illustrates the core estimator and simulation loop. Full code is available in the supplementary file `simulation_v2.py`.

```
import numpy as np
from scipy import stats

np.random.seed(42)
N, B = 1000, 5000

def LE_estimator(d_bar, A_bar, a_bar, alpha=1.0, beta=1.0):
R   = A_bar / a_bar
phi = (1 - R**alpha) * (1 + np.log(R**beta + 1e-15))
return d_bar * np.exp(phi)

def LE_optimal(d_bar, A_bar, a_bar, d_s, a_s):
rho  = np.corrcoef(d_s, a_s)[0, 1]
```

```
Cd   = d_s.std(ddof=1) / d_s.mean()
Ca   = a_s.std(ddof=1) / a_s.mean()
alph = -rho * Cd / Ca
beta = rho * Cd / Ca - 0.5
return LE_estimator(d_bar, A_bar, a_bar, alph, beta)

# Monte Carlo loop
estimates = []
rng = np.random.default_rng(0)
for _ in range(B):
idx   = rng.choice(N, n, replace=False)
d_s, a_s = D[idx], A[idx]
est = LE_optimal(d_s.mean(), A_bar, a_s.mean(), d_s, a_s)
estimates.append(est)

MSE = np.mean((np.array(estimates) - D_bar)**2)
PRE = MSE_sample_mean / MSE * 100
```

References

Bahl, S. and Tuteja, R.K. (1991). Ratio and product type exponential estimators. *Journal of Information and Optimization Sciences*, 12(1), 159–163.

Shukla, D., Gupta, V. K., and Jain, A. (2025). A class of logarithmic exponential estimators for estimating average degree of a network using triangular graph sampling. *Communications in Statistics - Theory and Methods*, 1–26.

Cochran, W.G. (1977). *Sampling Techniques*, 3rd edn. John Wiley & Sons, New York.

Kadilar, C. and Cingi, H. (2004). Ratio estimators in simple random sampling. *Applied Mathematics and Computation*, 151(3), 893–902.

Kadilar, C. and Cingi, H. (2006). An improvement in estimating the population mean by using the correlation coefficient. *Hacettepe Journal of Mathematics and Statistics*, 35(1), 103–109.

Murthy, M.N. (1967). *Sampling Theory and Methods.* Statistical Publishing Society, Calcutta.

Singh, D. and Chaudhary, F.S. (1986). *Theory and Analysis of Sample Survey Designs.* New Age International, New Delhi.

Singh, H.P. and Kumar, S. (2011). A general class of estimators of population mean using auxiliary information with known population proportion of an attribute. *Statistics*, 45(3), 249–265.

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S., and Asok, C. (1984). *Sampling Theory of Surveys with Applications*, 3rd edn. Iowa State University Press, Ames.

Upadhyaya, L.N. and Singh, H.P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, 41(5), 627–636.

Yadav, S.K. and Kadilar, C. (2013). Improved exponential type ratio estimator of population variance. *Revista Colombiana de Estadística*, 36(1), 145–152.

Yan, Z. and Tian, B. (2010). Ratio method to the mean estimation using coefficient of skewness of auxiliary variable. *ICICA 2010, CCIS*, 106, 103–110. Springer, Berlin.