Journal of Modern Applied Statistical Methods

Volume 19 | Issue 1 Article 10

6-8-2021

A Simple Random Sampling Modified Dual to Product Estimator for estimating Population Mean Using Order Statistics

Sanjay Kumar
Central University of Rajasthan, rahibhu@gmail.com

Priyanka Chhaparwal
Central University of Rajasthan, priyankachhaparwal4@gmail.com

Recommended Citation

Kumar, Sanjay and Chhaparwal, Priyanka (2021) "A Simple Random Sampling Modified Dual to Product Estimator for estimating Population Mean Using Order Statistics," *Journal of Modern Applied Statistical Methods*: Vol. 19 : Iss. 1 , Article 10.
DOI: 10.22237/jmasm/1608553620
Available at: https://digitalcommons.wayne.edu/jmasm/vol19/iss1/10

A Simple Random Sampling Modified Dual to Product Estimator for estimating Population Mean Using Order Statistics

Cover Page Footnote

The authors are grateful to the Editors and referees for their valuable suggestions which led to improvements in the article.

Journal of Modern Applied Statistical Methods
May 2020, Vol. 19, No. 1, eP2988.
doi: 10.22237/jmasm/1608553620

ISSN 1538 − 9472

A Simple Random Sampling Modified Dual to Product Estimator for Estimating Population Mean using Order Statistics

Sanjay Kumar
Central University of Rajasthan
Ajmer, India

Priyanka Chhaparwal
Central University of Rajasthan
Ajmer, India

Bandopadhyaya (1980) developed a dual to product estimator using robust modified maximum likelihood estimators (MMLE's). Their properties were obtained theoretically and supported through simulations studies with generated as well as one real data set. Robustness properties in the presence of outliers and confidence intervals were studied.

Keywords: Product estimator, dual to product estimator, simulation study, modified maximum likelihood, transformed auxiliary variable

Introduction

Estimating population parameters are common problems in almost all areas like management, engineering, and social science at the different stages of estimation procedure. Sometimes supplementary information on several variables is useful for estimating population parameters. In practice, when the correlation coefficient is negatively high between the study variable and auxiliary variables, a product type estimator is used to estimate population mean and the estimator is more efficient than the simple mean estimator under some realistic conditions. Further, the utilization of such supplementary information in sample surveys has been studied broadly by Yates (1960), Murthy (1967), Cochran (1977), Sukhatme et al. (1984), S. Singh (2003), Bouza (2008, 2015), Chanu and Singh (2014a, b), Gupta and Shabbir (2008, 2011), Diana et al. (2011), Choudhury and Singh (2012), H. P. Singh and Solanki (2012), Tato et al. (2016), Kumar (2015), Kumar and Chhaparwal (2016a), and Yadav and Kadilar (2013).

doi: 10.22237/jmasm/1608553620 | Accepted: October 4, 2018; Published: June 8, 2021.
Correspondence: Sanjay Kumar, rahibhu@gmail.com

Consider a finite population π: (π_1, π_2,…, π_N) of size N units. Let y_i and x_i are the values of the study (y) and the auxiliary (x) variable, respectively. Now, let

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i \quad \text{and} \quad \bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

be the population means, C_y and C_x be the coefficient of variations of the study (y) and the auxiliary (x) variables, respectively, and the correlation coefficient between the study and the auxiliary variables be ρ_{yx}. Murthy (1964) suggested the product estimator ($\bar{y}_p$) for the population mean $\bar{Y}$ given by

$$\bar{y}_p = \frac{\bar{y}}{\bar{x}}\bar{x}, \tag{1}$$

where

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i, \quad \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i,$$

and n is the number of units in the sample.

The expressions for bias and the mean square error (MSE) of the estimator $\bar{y}_p$ are as follows:

$$\mathrm{B}\left(\bar{y}_p\right) = \left(\frac{1-f}{n}\right)\bar{Y}C_{yx} \tag{2}$$

and

$$\mathrm{MSE}\left(\bar{y}_p\right) = \left(\frac{1-f}{n}\right)\bar{Y}^2\left(C_y^2 + C_x^2 + 2C_{yx}\right) \tag{3}$$

where

$$C_y^2 = \frac{S_y^2}{\bar{Y}^2}, \quad C_x^2 = \frac{S_x^2}{\bar{X}^2}, \quad C_{yx} = \frac{S_{yx}}{\bar{Y}\bar{X}}, \quad S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2,$$
$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2, \quad f = \frac{n}{N}, \quad \text{and} \quad S_{yx} = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)\left(y_i - \bar{Y}\right)$$

is the covariance between the study and auxiliary variables.

By taking a transformation,

$$x_i^* = \frac{N\bar{X} - nx_i}{N-n}, \quad \left(i = 1, 2, \ldots, N\right)$$

Bandopadhyaya (1980) studied a dual to product estimator given by

$$t_1 = \frac{\bar{y}}{\bar{x}^*}\bar{X}, \tag{4}$$

where

$$\bar{x}^* = \frac{N\bar{X} - n\bar{x}_i}{N-n},$$

and the correlations corr(y, x) and $\operatorname{corr}\left(y, x_i^*\right)$ are negative and positive, respectively.

The expressions for mean square error and bias of the estimator t_1 are

$$\mathrm{B}\left(t_1\right) = \left(\frac{1-f}{n}\right)\gamma\left(k+1\right)\bar{Y}C_x^2 \tag{5}$$

and

$$\mathrm{MSE}\left(t_1\right) = \left(\frac{1-f}{n}\right)\bar{Y}^2\left(C_Y^2 + \gamma^2 C_x^2 + 2\gamma\rho_{yx}C_yC_x\right), \tag{6}$$

where ρ_{yx} (< 0) is the correlation between y and x, $\gamma = n / (N - n)$, $k = C_{yx}\big/C_x^2 = \rho_{yx}\left(C_y/C_x\right)$.

The estimator t_1 is preferred to $\bar{y}_p$ when $k > -(1+\gamma)/2$, $(1-\gamma) > 0$, k being negative because $\rho_{yx} < 0$.

The studies mentioned above were limited to normal populations. The aim of this study is to consider the case where the population is not normal, i.e., real life situations. A new modified dual to product type estimator is proposed based on modified maximum likelihood (MML) methodology.

Long Tailed Symmetric Family

Let a linear regression model $y_i = \theta x_i + e_i$; $i = 1, 2, \ldots, n$. Consider a study variable y from the long tailed symmetric family

$$\mathrm{f}(y) = \mathrm{LTS}(p, \sigma) = \frac{\Gamma p}{\sigma\sqrt{K}\,\Gamma\left(\frac{1}{2}\right)\Gamma\left(p-\frac{1}{2}\right)}\left\{1+\frac{1}{K}\left(\frac{y-\mu}{\sigma}\right)^2\right\}^{-p}, \tag{7}$$

$-\infty < y < \infty$, where $K = 2p - 3$ and $p \geq 2$ is the shape parameter (p is known) with $\mathrm{E}(y) = \mu$ and $\mathrm{Var}(y) = \sigma^2$. Here the kurtosis of (7) can be obtained as

$$\frac{\mu_4}{\mu_2^2} = \frac{3K}{K-2}.$$

Note

$$t = \sqrt{\frac{v}{K}}\left(\frac{y-\mu}{\sigma}\right) \sim t_{v=2p-1}.$$

Assume p = 2.5, 3.5, 4.5, and 5.5, which correspond to a kurtosis of ∞, 6, 4.5, and 4.0. (7) reduces to a normal distribution when $p = \infty$. The likelihood function obtained from (7) is given by

$$\mathrm{LogL} \propto -n\log\sigma - p\sum_{i=1}^{n}\log\left\{1+\frac{1}{K}z_i^2\right\}; \quad z_i = \frac{y_i-\mu}{\sigma}. \tag{8}$$

The solution of the likelihood equation (assuming σ is known),

$$\frac{d\,\text{LogL}}{d\mu} = \frac{2p}{K\sigma}\sum_{i=1}^{n} \text{g}(z_i) = 0\,, \tag{9}$$

where

$$\text{g}(z_i) = \frac{z_i}{\left\{1 + \frac{1}{K}\left(z_i^2\right)\right\}}\,,$$

will produce the MLE of μ, which does not have explicit solutions.

For all the shape parameters $p < \infty$,Vaughan (1992a) and Oral (2010) showed that equation (8) has multiple unknown roots and the robust MMLE asymptotically equivalent to the MLE are obtained as

1. The likelihood equations are expressed in ordered variates:

$$y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)},$$

2. The function $\text{g}(z_i)$ are linearized by Taylor series expansion around

$$t_{(i)} = \text{E}\left(z_{(i)}\right), \quad z_{(i)} = \frac{y_{(i)} - \mu}{\sigma}, \quad 1 \leq i \leq n$$

   up to the first two terms.

3. A unique solution (MMLE) is obtained after the solving the equation.

The values of $t_{(i)}$; $1 \leq i \leq n$ were suggested by Tiku and Kumra (1985) for p =2 (0.5) 10 and Vaughan (1992b) for $p = 1.5$, $n \leq 20$. For $n > 20$, the values of $t_{(i)}$ can be approximated from the equations

$$\frac{\Gamma p}{\sigma\sqrt{K}\Gamma\left(\frac{1}{2}\right)\Gamma\left(p - \frac{1}{2}\right)}\int_{-\infty}^{t_{(i)}}\left\{1 + \frac{1}{K}z^2\right\}^{-p} dz = \frac{i}{n+1}; \quad 1 \leq i \leq n\,, \tag{10}$$

$$\frac{d\,\text{LogL}}{d\mu} = \frac{2p}{K\sigma}\sum_{i=1}^{n} \text{g}(z_i) = 0, \text{ since } \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} y_{(i)}\,. \tag{11}$$

A Taylor series expansion of $g(z_{(i)})$ around $t_{(i)}$ up to the first two terms of expansion gives

$$g\left(z_{(i)}\right) \cong g\left(t_{(i)}\right) + \left\{z_{(i)} - t_{(i)}\right\} \left\{ \left. \frac{d\left\{g\left(z\right)\right\}}{dz} \right|_{z=t_{(i)}} \right\} = \alpha_i + \beta_i z_{(i)}; \quad 1 \le i \le n, \tag{12}$$

where

$$\alpha_i = \left(\frac{2}{K}\right) \frac{t_{(i)}^3}{\left\{1 + \frac{1}{K} t_{(i)}^2\right\}^2} \quad \text{and} \quad \beta_i = \frac{1 - \frac{1}{K} t_{(i)}^2}{\left\{1 + \frac{1}{K} t_{(i)}^2\right\}^2}. \tag{13}$$

Further, for symmetric distributions, it may be noted that $t_{(i)} = -t_{(n-i+1)}$ and hence

$$\alpha_i = -\alpha_{(n-i+1)}, \quad \sum_{i=1}^{n} \alpha_i = 0, \quad \beta_i = \beta_{(n-i+1)}. \tag{14}$$

Now, (11) along with (12) and (13) give the modified likelihood equation given by

$$\frac{d\,\mathrm{LogL}}{d\mu} \cong \frac{d\,\mathrm{LogL}^*}{d\mu} = \frac{2p}{K\sigma} \sum_{i=1}^{n} \left(\alpha_i + \beta_i z_{(i)}\right) = 0. \tag{15}$$

Hence, (15) provides the MMLE $\hat{\mu}$ given by

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \beta_i y_{(i)}}{m} \tag{16}$$

where

$$m = \sum_{i=1}^{n} \beta_i .$$

Tiku and Vellaisamy (1996) and Oral and Oral (2011) showed

$$\mathrm{E}\left(\hat{\mu}-\bar{Y}\right)=0 \tag{17}$$

and

$$\mathrm{E}\left(\hat{\mu}-\bar{Y}\right)^2=\mathrm{V}\left(\hat{\mu}\right)-\frac{2n}{N}\mathrm{Cov}\left(\hat{\mu},\bar{y}\right)+\frac{\sigma^2}{N}. \tag{18}$$

The exact variance of $\hat{\mu}$ is given by $\mathrm{V}\left(\hat{\mu}\right)=\left(\boldsymbol{\beta}'\Omega\boldsymbol{\beta}\right)\left(\sigma^2/m^2\right)$, where $\boldsymbol{\beta}' = (\beta_1, \beta_2, \beta_3, \ldots, \beta_n)$ and

$$\mathrm{Cov}\left(z_{(i)}=\frac{y_{(i)}-\mu}{\sigma}\right)=\Omega, \quad 1\le i\le n.$$

$\mathrm{Cov}\left(\hat{\mu},\bar{y}\right)=\left(\boldsymbol{\beta}'\Omega\boldsymbol{\omega}\right)\left(\sigma^2/m\right)$, where $\boldsymbol{\omega}' = (1/n, 1/n, \ldots, 1/n)_{1\times n}$. Tiku and Kumra (1985) and Vaughan (1992b) tabulated the elements of $\boldsymbol{\Omega}$.

Tiku and Suresh (1992) and Tiku and Vellaisamy (1996) studied the MMLE $\hat{\sigma}$ (assuming σ is unknown), i.e.,

$$\hat{\sigma}=\frac{F+\sqrt{F^2+4nC}}{2\sqrt{n\left(n-1\right)}}, \tag{19}$$

where

$$F=\frac{2p}{K}\sum_{i=1}^{n}\alpha_i y_{(i)}, \quad C=\frac{2p}{K}\sum_{i=1}^{n}\beta_i\left(y_{(i)}-\hat{\mu}\right)^2.$$

Puthenpura and Sinha (1986), Tiku and Suresh (1992), Oral (2006, 2010), Oral and Oral (2011), Oral and Kadilar (2011), and Kumar and Chhaparwal (2016b, c, 2017) have studied the methodology of MML, where maximum likelihood (ML) estimation is intractable. Vaughan and Tiku (2000) discussed that MMLEs and ML estimators (MLEs) have the same asymptotic properties under certain regularity conditions, and both are as efficient as MLEs for small n values.

The Proposed Dual to Product Estimator and its Bias and Mean Square Error (MSE)

In the field of sample surveys, MMLE (16) was used by Tiku and Bhasin (1982) and Tiku and Vellaisamy (1996) to improve efficiencies in estimators. Using such methodology, a new dual to product estimator is proposed:

$$T_1 = \frac{\hat{\mu}}{\bar{x}^*}\bar{X}\ , \tag{20}$$

where $\bar{X}$ is known. The expressions for bias and MSE of the proposed estimator T_1, up to the terms of order n^{-1}, are given as follows:

Let $\hat{\mu} = \bar{Y}\left(1+\epsilon_0\right), \bar{x}^* = \bar{X}\left(1+\epsilon_1\right)$, such that $E(\epsilon_0) = 0 = E(\epsilon_1)$, $|\epsilon_1| < 1$. Under SRSWOR method of sampling,

$$\mathrm{E}\left(\epsilon_0^2\right) = \frac{1}{\bar{Y}^2}\mathrm{E}\left(\hat{\mu}-\bar{Y}\right)^2 = \frac{1}{\bar{Y}^2}\left\{\mathrm{V}\left(\hat{\mu}\right) - \frac{2n}{N}\mathrm{Cov}\left(\hat{\mu},\bar{y}\right) + \frac{\sigma^2}{N}\right\},$$

$$\mathrm{E}\left(\epsilon_1^2\right) = \frac{1}{\bar{X}^2}\mathrm{V}\left(\bar{x}^*\right) = \frac{1}{\bar{X}^2}\left(\frac{n}{N-n}\right)^2\mathrm{V}\left(\bar{x}\right) = \frac{1}{\bar{X}^2}\left(\frac{n}{N-n}\right)^2\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i-\bar{X}\right)^2$$

$$= \frac{1}{\bar{X}^2}\frac{n}{\left(N-n\right)N\left(N-1\right)}\sum_{i=1}^{N}\left(x_i-\bar{X}\right)^2,$$

$$\mathrm{E}\left(\epsilon_0,\epsilon_1\right) = \frac{1}{\bar{Y}\bar{X}}\mathrm{Cov}\left(\hat{\mu},\bar{x}^*\right) = -\frac{1}{\bar{Y}\bar{X}}\gamma\,\mathrm{Cov}\left(\hat{\mu},\bar{x}\right),$$

$$\mathrm{B}\left(T_1\right) = \frac{\gamma}{\bar{X}}\left\{R\gamma\,\mathrm{V}\left(\bar{x}\right) + \mathrm{Cov}\left(\hat{\mu},\bar{x}\right)\right\} \tag{21}$$

and

$$\mathrm{MSE}\left(T_1\right) = \mathrm{E}\left(\hat{\mu}-\bar{Y}\right)^2 + R^2\gamma^2\,\mathrm{V}\left(\bar{x}\right) + 2R\gamma\,\mathrm{Cov}\left(\hat{\mu},\bar{x}\right), \tag{22}$$

where the term $\mathrm{Cov}\left(\hat{\mu},\bar{x}\right)$ is calculated by Oral and Oral (2011) as

$$\mathrm{Cov}\left(\hat{\mu},\bar{x}\right) = \frac{1}{\theta}\left\{\mathrm{Cov}\left(\hat{\mu},\bar{y}-\bar{e}\right)\right\} = \frac{1}{\theta}\left\{\mathrm{Cov}\left(\hat{\mu},\bar{y}\right) - \mathrm{Cov}\left(\theta\bar{x}_{[\cdot]}+\bar{e}_{[\cdot]},\bar{e}\right)\right\},$$

where

$$\bar{x}_{[\cdot]} = \sum_{i=1}^{n} \frac{\beta_i \bar{x}_{[i]}}{m}, \bar{e}_{[\cdot]} = \sum_{i=1}^{n} \frac{\beta_i \bar{e}_{[i]}}{m}, \bar{e}_{[i]} = y_{(i)} - \theta x_{[i]},$$

and $x_{[i]}$ is the concomitant of $y_{(i)}$. Here x in $y = \theta x + e$ is assumed to be non-stochastic (Oral & Oral, 2011) and hence $\mathrm{Cov}(x_i, e_j)$ is not affected by the ordering of the y values for $1 \le i \le n$ and $1 \le j \le n$; therefore

$$\mathrm{Cov}(\hat{\mu}, \bar{x}) = \frac{1}{\theta}\left\{\mathrm{Cov}(\hat{\mu}, \bar{y}) - \mathrm{Cov}\left(\bar{e}_{[\cdot]}, \bar{e}\right)\right\},$$

where $\mathrm{Cov}\left(\bar{e}_{[\cdot]}, \bar{e}\right) = \left(\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\omega}\right)\left(\sigma_e^2 / m\right)$. Note in the case of exceeding 5% of the sampling fraction n / N, the finite population correction $(N - n) / N$ can be presented as

$$\mathrm{Cov}(\hat{\mu}, \bar{x}) = \frac{N-n}{N\theta}\left\{\mathrm{Cov}(\hat{\mu}, \bar{y}) - \mathrm{Cov}\left(\bar{e}_{[\cdot]}, \bar{e}\right)\right\}.$$

Monte Carlo Simulation

R is used as the simulation platform. The model in the generated super-population models is given by

$$y_i = \theta x_i + e_i, \quad i = 1, 2, \ldots, N. \tag{23}$$

The error term e_i, $i = 1, 2, \ldots, N$, with $\mathrm{E}(e) = 0$ and $\mathrm{V}(e) = \sigma_e^2$, and the auxiliary variable x_i are generated independently from each other and then y_i is calculated using (23). The calculations for the mean square error of (20) are performed as follows:

Consider the size of the population $N = 500$ and select a sample of size n (= 5, 11, 15, 21, 31, 51) from the finite population by SRSWOR. Out of the possible 500 choose n SRSWOR samples of size n (= 5, 11, 15, 21, 31, 51), select S = 1,00,000 random samples and calculate the values of mean square error (MSE) of different estimators as follows:

$$\mathrm{MSE}(T_1)=\frac{1}{S}\sum_{j=1}^{S}\left(T_{1j}-\bar{Y}\right)^2,\mathrm{MSE}(t_1)=\frac{1}{S}\sum_{j=1}^{S}\left(t_{1j}-\bar{Y}\right)^2,\mathrm{MSE}\left(\bar{y}_p\right)=\frac{1}{S}\sum_{j=1}^{S}\left(\bar{y}_{pj}-\bar{Y}\right)^2$$

Now, in the model $y = \theta x + e$, the value of θ is chosen by following Rao and Beegle (1967), Oral and Oral (2011), and Oral and Kadilar (2011) in such a way that the correlation coefficient between the study (y) and the auxiliary (x) variables is ρ_{yx} = -0.55. The value of θ is calculated using $\sigma^2 = 1$ without loss of generality.

Comparison of Efficiencies of the Proposed Estimator

The conditions under which the proposed estimator T_1 is more efficient than the corresponding estimators $\bar{y}_p$ and t_1 are given as follows:

$$\begin{aligned}&\mathrm{MSE}(T_1)\le\mathrm{MSE}(t_1)\le\mathrm{MSE}\left(\bar{y}_p\right)\text{ if}\\&\frac{1}{2R\gamma}\left\{\mathrm{E}\left(\hat{\mu}-\bar{Y}\right)^2-\mathrm{E}\left(\bar{y}-\bar{Y}\right)^2\right\}+\mathrm{Cov}\left(\hat{\mu},\bar{x}\right)\le\mathrm{Cov}\left(\bar{y},\bar{x}\right)\\&\qquad\le\frac{\left(1-\gamma^2\right)}{2\gamma}R\,\mathrm{V}\left(\bar{x}\right)+\frac{1}{\gamma}\mathrm{Cov}\left(\bar{y},\bar{x}\right)\end{aligned}\tag{24}$$

for $R > 0$,

$$\begin{aligned}&\mathrm{MSE}(T_1)\le\mathrm{MSE}(t_1)\le\mathrm{MSE}\left(\bar{y}_p\right)\text{ if}\\&\frac{\left(1-\gamma^2\right)}{2\gamma}R\,\mathrm{V}\left(\bar{x}\right)+\frac{1}{\gamma}\mathrm{Cov}\left(\bar{y},\bar{x}\right)\le\mathrm{Cov}\left(\bar{y},\bar{x}\right)\\&\qquad\le\frac{1}{2R\gamma}\left\{\mathrm{E}\left(\hat{\mu}-\bar{Y}\right)^2-\mathrm{E}\left(\bar{y}-\bar{Y}\right)^2\right\}+\mathrm{Cov}\left(\hat{\mu},\bar{x}\right)\end{aligned}\tag{25}$$

for R < 0, where

$$\mathrm{Cov}\left(\bar{y},\bar{x}\right)=\left(\frac{1}{n}-\frac{1}{N}\right)S_{yx}.$$

Two different super-population models as suggested by Oral and Kadilar (2011) are given below to observe the performance of the proposed modified estimator. Model 2 is taken for knowing the effeteness of outliers.

Model 1. $x \sim U(1, 2.5)$ and $y \sim LTS(p, 1)$
Model 2. $x \sim \exp(1)$ and $y \sim LTS(p, 1)$

For Models 1 and 2, the values of θ are given in Table 1. A scatter graph and a histogram for the underlying distribution of Model 2 for $p = 3.5$ are provided in Figure 1.

Table 1. Parameter values of θ used in Models 1 and 2 that give $\rho_{yx} = -0.55$

	p		
Population	**2.5**	**4.5**	**5.5**
Model 1	-1.521	-1.521	-1.521
Model 2	-0.659	-0.659	-0.659

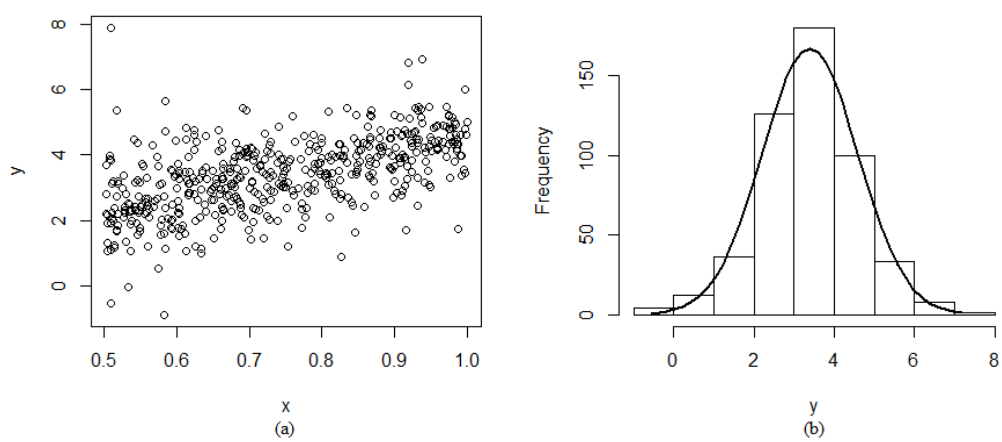

Figure 1. (a) Scatter graph of the study variable and auxiliary variable; (b) Underlying distribution of the study variable obtained from Model 2 for $p = 3.5$

Table 2. Mean square error and efficiencies of the estimators under super-populations 1 and 2

Model 1: x ~ U(1, 2.5) and y ~ LTS(p, 1)

		n					
p	**Estimator**	**5**	**11**	**15**	**21**	**31**	**51**
2.5	T_1	201.97	203.80	208.33	206.02	192.55	190.00
		(0.1266)	(0.0526)	(0.0360)	(0.0266)	(0.0188)	(0.0120)
	t_1	190.25	188.07	182.04	186.39	187.56	182.40
		(0.1344)	(0.0570)	(0.0412)	(0.0294)	(0.0193)	(0.0125)
	$\bar{y}_p$	100.00	100.00	100.00	100.00	100.00	100.00
		(0.2557)	(0.1072)	(0.0750)	(0.0548)	(0.0362)	(0.0228)
4.5	T_1	197.65	189.04	192.04	186.97	184.06	178.40
		(0.1320)	(0.0602)	(0.0377)	(0.0307)	(0.0207)	(0.0125)
	t_1	197.50	188.72	190.53	183.97	183.17	175.59
		(0.1321)	(0.0603)	(0.0380)	(0.0312)	(0.0208)	(0.0127)
	$\bar{y}_p$	100.00	100.00	100.00	100.00	100.00	100.00
		(0.2609)	(0.1138)	(0.0724)	(0.0574)	(0.0381)	(0.0223)
5.5	T_1	194.18	187.95	191.45	192.23	184.13	177.34
		(0.1322)	(0.0614)	(0.0399)	(0.0309)	(0.0208)	(0.0128)
	t_1	193.59	185.83	189.58	190.10	182.38	175.97
		(0.1326)	(0.0621)	(0.0403)	(0.0311)	(0.0210)	(0.0129)
	$\bar{y}_p$	100.00	100.00	100.00	100.00	100.00	100.00
		(0.2567)	(0.1154)	(0.0764)	(0.0594)	(0.0383)	(0.0227)

Model 2: x ~ exp(1) and y ~ LTS(p, 1)

		n					
p	**Estimator**	**5**	**11**	**15**	**21**	**31**	**51**
2.5	T_1	260.35	261.64	263.23	233.28	222.76	209.14
		(0.5523)	(0.2474)	(0.1727)	(0.1331)	(0.0883)	(0.0536)
	t_1	235.64	221.07	217.62	204.14	194.75	190.65
		(0.6102)	(0.2928)	(0.2089)	(0.1521)	(0.1010)	(0.0588)
	$\bar{y}_p$	100.00	100.00	100.00	100.00	100.00	100.00
		(1.4379)	(0.6473)	(0.4546)	(0.3105)	(0.1967)	(0.1121)
4.5	T_1	265.72	228.89	230.09	209.50	210.86	184.40
		(0.6520)	(0.2831)	(0.2087)	(0.1494)	(0.0976)	(0.0609)
	t_1	259.40	220.63	221.39	198.10	198.84	179.11
		(0.6679)	(0.2937)	(0.2169)	(0.1581)	(0.1035)	(0.0627)
	$\bar{y}_p$	100.00	100.00	100.00	100.00	100.00	100.00
		(1.7325)	(0.6480)	(0.4802)	(0.3130)	(0.2058)	(0.1123)
5.5	T_1	287.83	238.14	233.36	223.44	205.30	191.11
		(0.6928)	(0.2892)	(0.2218)	(0.1553)	(0.1019)	(0.0630)
	t_1	283.13	230.41	220.35	211.20	194.42	182.98
		(0.7043)	(0.2989)	(0.2349)	(0.1643)	(0.1076)	(0.0658)
	$\bar{y}_p$	100.00	100.00	100.00	100.00	100.00	100.00
		(1.9941)	(0.6887)	(0.5176)	(0.3430)	(0.2092)	(0.1204)

Note: Mean square errors are in parenthesis

Relative efficiencies (RE) are obtained as

$$RE = \frac{\mathrm{MSE}\left(\bar{y}_p\right)}{\mathrm{MSE}\left(\cdot\right)} * 100,$$

where MSE(·) and RE are given in Table 2 for Models 1 and 2.

From Table 2, note that the proposed estimator T_1 is more efficient than the corresponding estimators $\bar{y}_p$ and t_1. We also observe that when sample size increases, mean square error decreases. Further, we observe that due to the presence of outliers, mean square errors of the estimators increase for Model 2 as compared to Model 1. Next, the values of mean square errors of different estimators for different values of n and p are plotted and shown in Figures 2 and 3.

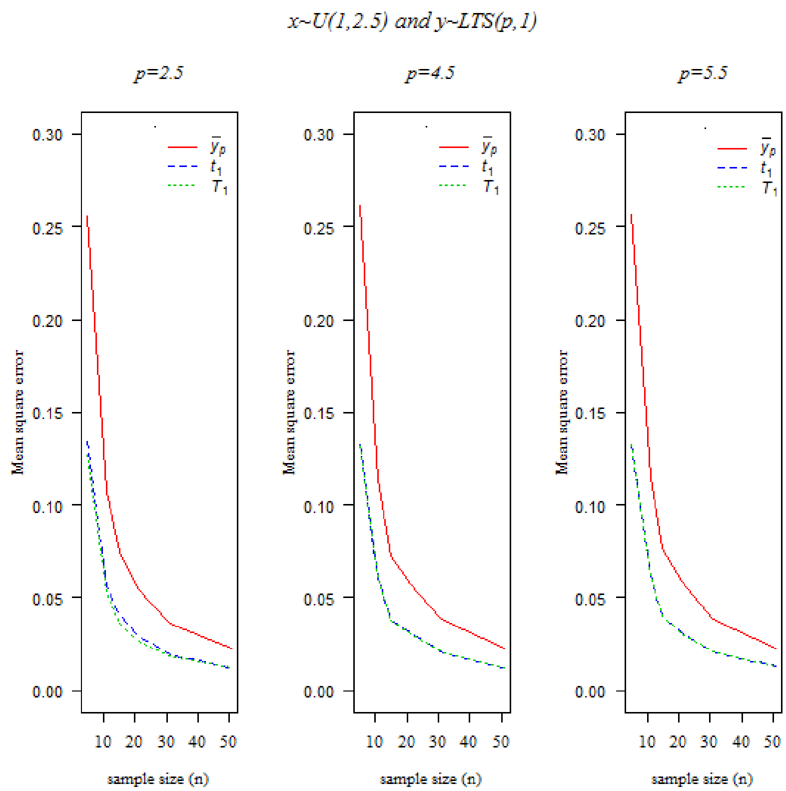

Figure 2. Mean square errors of different estimators for different values of n and p

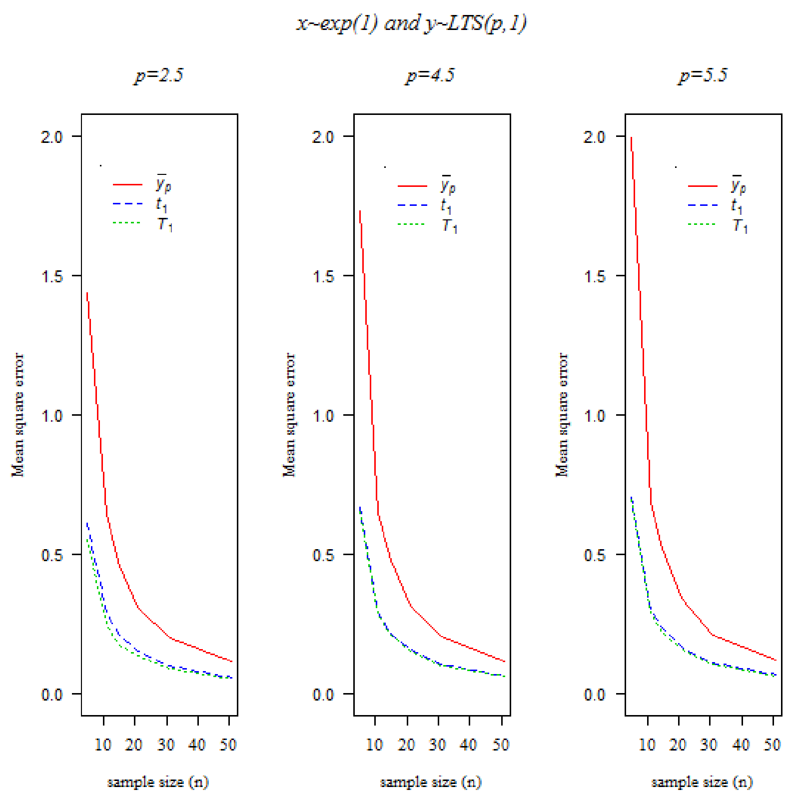

Figure 3. Mean square errors of different estimators for different values of *n* and *p*

The mean square error of the proposed estimator T_1 is more efficient than the corresponding estimators $\bar{y}_p$ and t_1. Also, when sample size increases, mean square error decreases. Further, when *p* increases, mean square error of the proposed estimator increases and becomes close to t_1. Absolute biases are calculated via

$$\mathrm{B}(T_1)=\frac{1}{S}\left|\sum_{j=1}^{S}\left(T_{1j}-\bar{Y}\right)\right|,\mathrm{B}(t_1)=\frac{1}{S}\left|\sum_{j=1}^{S}\left(t_{1j}-\bar{Y}\right)\right|,\text{ and }\mathrm{B}\left(\bar{y}_p\right)=\frac{1}{S}\left|\sum_{j=1}^{S}\left(\bar{y}_p-\bar{Y}\right)\right|.$$

The simulated bias of the proposed estimator T_1 is less than the corresponding estimators t_1 and $\bar{y}_p$. We also observe that when sample size increases, bias decreases. Further, observe that the biases of the estimators increase for Model 2 as compared to Model 1 due to the presence of outliers. Next, the values of absolute bias of different estimators for different values of *n* and *p* are plotted and are shown in Figures 4 and 5.

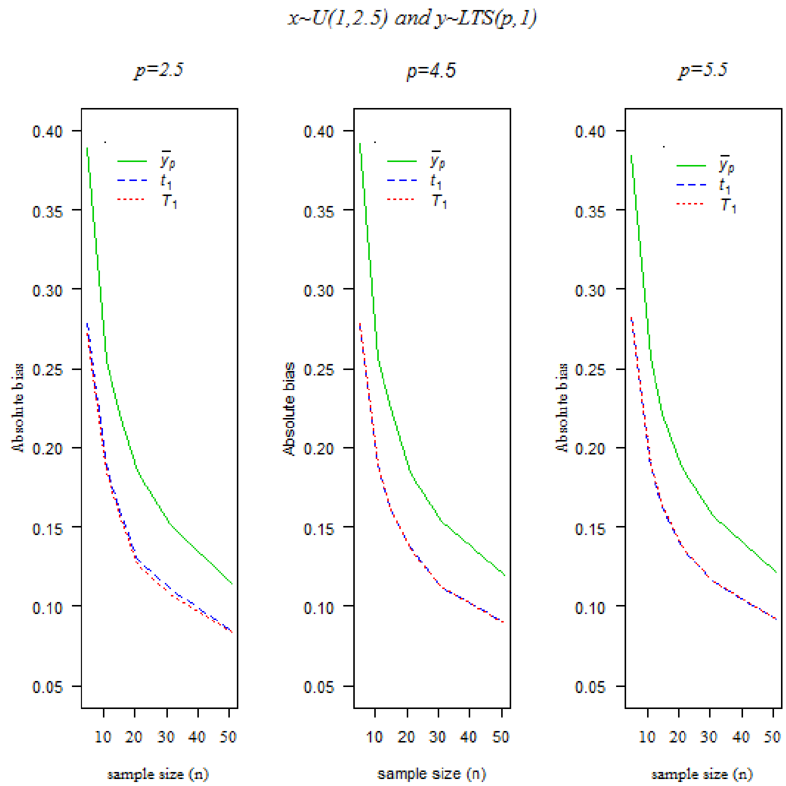

Figure 4. Absolute bias of different estimators for different values of n and p

---

Table 3. Simulated absolute bias of the estimators T_1, t_1, and $\bar{y}_p$ under super-populations 1 and 2

Model 1: x ~ U(1, 2.5) and y ~ LTS(p, 1)

		n					
p	**Estimator**	**5**	**11**	**15**	**21**	**31**	**51**
2.5	T_1	0.2719	0.1847	0.1580	0.1260	0.1082	0.0838
	t_1	0.2787	0.1888	0.1616	0.1303	0.1116	0.0851
	$\bar{y}_p$	0.3893	0.2552	0.2211	0.1855	0.1517	0.1142
4.5	T_1	0.2779	0.1887	0.1615	0.1363	0.1123	0.0897
	t_1	0.2786	0.1891	0.1609	0.1369	0.1126	0.0902
	$\bar{y}_p$	0.3918	0.2564	0.2245	0.1843	0.1541	0.1195
5.5	T_1	0.2820	0.1894	0.1636	0.1383	0.1158	0.0919
	t_1	0.2823	0.1890	0.1631	0.1377	0.1157	0.0920
	$\bar{y}_p$	0.3847	0.2570	0.2210	0.1876	0.1576	0.1212

Table 3 (continued).

Model 2: x ~ exp(1) and y ~ LTS(p, 1)

		n					
p	**Estimator**	**5**	**11**	**15**	**21**	**31**	**51**
2.5	T_1	0.5859	0.3956	0.3378	0.2861	0.2375	0.1893
	t_1	0.6103	0.4355	0.3723	0.3142	0.2551	0.2006
	$\bar{y}_p$	0.8972	0.5984	0.5281	0.4361	0.3517	0.2676
4.5	T_1	0.6105	0.4200	0.3468	0.3085	0.2453	0.1924
	t_1	0.6231	0.4252	0.3524	0.3192	0.2554	0.1961
	$\bar{y}_p$	0.9112	0.6117	0.4816	0.4462	0.3585	0.2337
5.5	T_1	0.6176	0.4348	0.3631	0.3205	0.2506	0.1955
	t_1	0.6234	0.4406	0.3669	0.3256	0.2569	0.1981
	$\bar{y}_p$	0.8870	0.6244	0.5290	0.4490	0.3542	0.2658

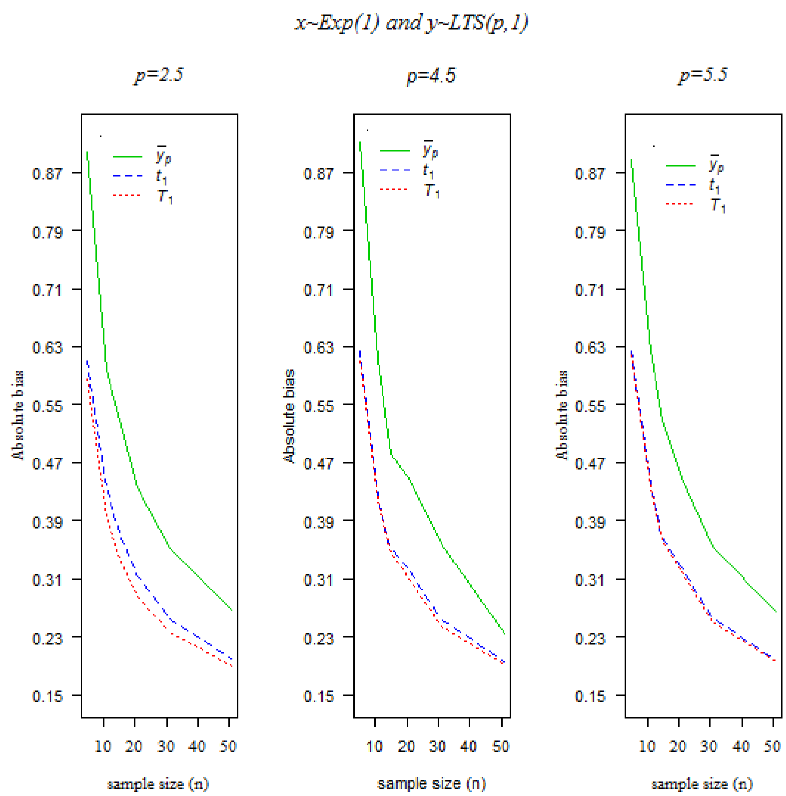

Figure 5. Absolute bias of different estimators for different values of n and p

The absolute bias of the proposed estimator T_1 is less than the corresponding estimators $\bar{y}_p$ and t_1. Also, when sample size increases, absolute bias decreases. When p increases, absolute bias of the proposed estimator increases and becomes close to the bias of t_1.

Robustness of the Proposed Estimator

Oral and Oral (2011) and Oral and Kadilar (2011) studied the problem of outliers in sample data and hence the shape parameter p in LTS(p, σ) might be mis-specified in experiments. Thus, it is important for estimators to be studied for plausibility to the assumed model. Consider the robustness property under different outlier models for $N = 500$ and $\sigma^2 = 1$ without loss of generality. Assume $x \sim \mathrm{U}(1, 2.5)$ as well as $x \sim \exp(1)$ and $y \sim \mathrm{LTS}(p = 3.5, \sigma^2 = 1)$. Super-population models are determined as follows:

Model 3. True model: LTS($p = 3.5$, $\sigma^2 = 1$)
Model 4. Dixon's outliers model: $N - N_o$ observations from LTS(3.5, 1) and N_o (we don't know which) form LTS(3.5, 2.0)
Model 5. Mis-specified model: LTS(4.0, 1)

Here, Model 3 is assumed as a super population model and Models 4 and 5 are taken as its plausible alternatives. N_o in Model 4 is calculated by $|0.5 + 0.1 * N| = 50$ for $N = 500$. The generated e_i's , ($i = 1, 2, \ldots, N$) are standardized in all the models to have the same variance as LTS(3.5, 1), i.e., it should be equal to 1. The simulated values of MSE and relative efficiency are given in Table 4.

Table 4. Mean square errors and efficiencies under super-populations 3 to 5 for LTS family

	n			***n***		
	5	**11**	**15**	**21**	**31**	**51**
Estimator	**Model 3**			**Model 4**		
T_1	195.90	189.38	199.44	186.39	211.52	221.34
	(0.1292)	(0.0593)	(0.0354)	(0.2771)	(0.0755)	(0.0464)
t_1	193.80	186.24	191.85	156.71	160.83	170.32
	(0.1306)	(0.0603)	(0.0368)	(0.3296)	(0.0993)	(0.0603)
$\bar{y}_p$	100.00	100.00	100.00	100.00	100.00	100.00
	(0.2531)	(0.1123)	(0.0706)	(0.5165)	(0.1597)	(0.1023)

Table 4 (continued).

Estimator	*n* = 5	11	15	*n* = 21	31	51
	Model 5			**Model 3**		
T_1	196.60	200.00	224.28	276.33	238.84	248.12
	(0.1265)	(0.0528)	(0.0383)	(0.6260)	(0.2698)	(0.1970)
t_1	194.30	199.25	166.80	266.70	217.63	224.53
	(0.1280)	(0.0530)	(0.0515)	(0.6486)	(0.2961)	(0.2177)
$\bar{y}_p$	100.00	100.00	100.00	100.00	100.00	100.00
	(0.2487)	(0.1056)	(0.0859)	(1.7298)	(0.6444)	(0.4888)
	Model 4			**Model 5**		
T_1	313.11	222.34	225.46	302.96	231.61	228.78
	(0.9839)	(0.3093)	(0.2239)	(0.6145)	(0.2664)	(0.2081)
t_1	278.14	202.74	206.21	294.57	217.94	210.48
	(1.1076)	(0.3392)	(0.2448)	(0.6320)	(0.2830)	(0.2262)
$\bar{y}_p$	100.00	100.00	100.00	100.00	100.00	100.00
	(3.0807)	(0.6877)	(0.5048)	(1.8617)	(0.6170)	(0.4761)

Note: Mean square error are in parenthesis

The proposed estimator T_1 is more efficient than the estimators $\bar{y}_p$ and t_1 and, as sample size increases, mean square error decreases. Due to the presence of outliers, mean square errors of the estimators increase for Model 2 as compared to Model 1.

Real Life Application

For studying the performance of the product estimator in (7), consider the real-life problem of the Auto MPG Data Set (Ramos et al., 1993). It pertains to the acceleration (m/s^2) of a car as a study variable (y) and weight (pounds) of the car as an auxiliary variable (x). The summary of the data on y is as follows:

$$N = 240, \text{Median} = 15.20, \text{Mean} = 15.34, \text{Kurtosis} = 3.5, \text{Skewness} = 0.20,$$
$$\rho_{yx} = -0.43$$

The data on y follows the long tailed symmetric distribution with $p = 8.5$, which can be obtained using $K = 2p - 3$. The scatter plot, histogram between the study variable and the auxiliary variable, and the Q-Q plot for the data on the study variable are given in Figure 6, which shows the nature (negative correlation, normality etc.) of the data.

For the simulation study using this data set, R was used and the MSE of the proposed estimator in (7) was calculated. The Monte Carlo study proceeded as follows: From the real-life population of size 240, S = 1,00,000 samples of size n (= 5, 10, 15, 20) are selected by SRSWOR, which gives 1,00,000 values of T_1.

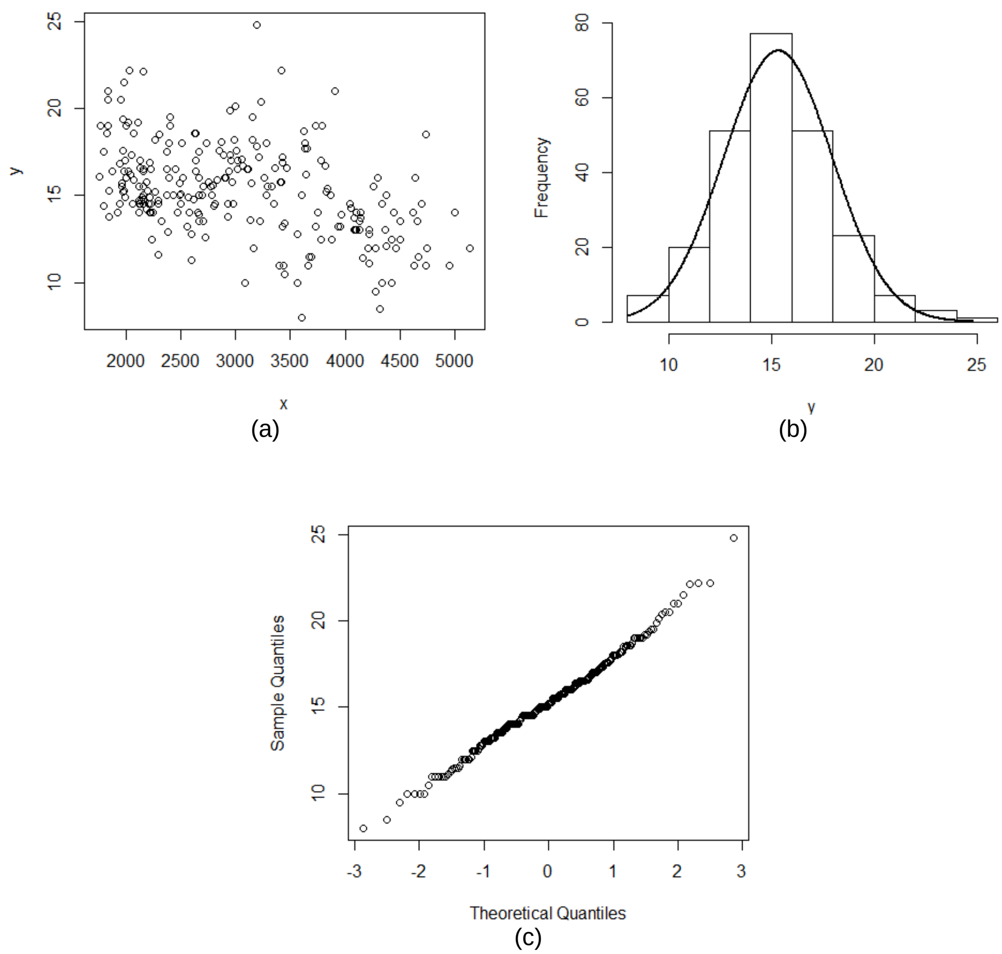

Figure 6. (a) Scatter graph of study and auxiliary variables; (b) Histogram for underlying distribution of study variable; (c) Q-Q plot for underlying distribution of study variable

The proposed estimator T_1 has minimum mean square error as well as minimum absolute bias compared to those of the relevant estimators for the true value of the shape parameter p = 8.5. However, sample data always have outliers. In practice, there might be mis-specification of the shape parameter p in LTS(p, σ). Therefore, an estimator must have efficiency robustness. So, consider the robustness property of the proposed estimators under mis-specification of the shape parameter which are given as follows:

Model 6. True model: LTS(p = 8.5, σ^2 = 7.0)
Model 7. Mis-specified model: LTS(7.0, 7.0)
Model 8. Mis-specified model: LTS(9.5, 7.0)
Model 9. Mis-specified model: LTS(10.0, 7.0)

As noted in Table 5, the proposed estimator T_1 is more efficient than the estimators $\bar{y}_p$ and t_1 and the mean square error decreases as sample size increases.

Table 5. Mean square error and efficiencies of the estimators T_1, t_1, and $\bar{y}_p$

	Estimators					
			T_1			
n	$\bar{y}_p$	t_1	p = 7.0	p = 8.5	p = 9.5	p = 10
5	100.00	633.37	639.14	638.25	637.79	637.58
	(7.8620)	(1.2413)	(1.2301)	(1.2318)	(1.2327)	(1.2331)
10	100.00	619.81	632.07	630.44	629.52	629.11
	(3.8961)	(0.6286)	(0.6164)	(0.6180)	(0.6189)	(0.6193)
15	100.00	563.43	578.26	576.22	575.20	574.62
	(2.2847)	(0.4055)	(0.3951)	(0.3965)	(0.3972)	(0.3976)
20	100.00	602.43	627.51	624.11	622.42	621.70
	(1.6127)	(0.2677)	(0.2570)	(0.2584)	(0.2591)	(0.2594)

Note: Mean square error are in parenthesis

Table 6. Simulated absolute bias of the estimators T_1, t_1, and $\bar{y}_p$

	Estimators					
			T_1			
n	$\bar{y}_p$	t_1	p = 7.0	p = 8.5	p = 9.5	p = 10
5	2.2273	0.9178	0.9117	0.9128	0.9133	0.9135
10	1.4841	0.6574	0.6466	0.6484	0.6493	0.6497
15	1.1889	0.5145	0.5035	0.5050	0.5058	0.5062
20	1.0129	0.4210	0.4148	0.4155	0.4159	0.4161

From Table 6, note the simulated absolute bias of the proposed estimator T_1 is less than the corresponding estimators t_1 and $\bar{y}_p$. When sample size increases, bias decreases.

From the Figures 7 and 8, note the absolute bias of the proposed estimator T_1 is less than the corresponding estimators $\bar{y}_p$ and t_1. Also, when sample size increases, absolute bias decreases. When p increases, absolute bias of the proposed estimator increases and becomes close to the bias of t_1.

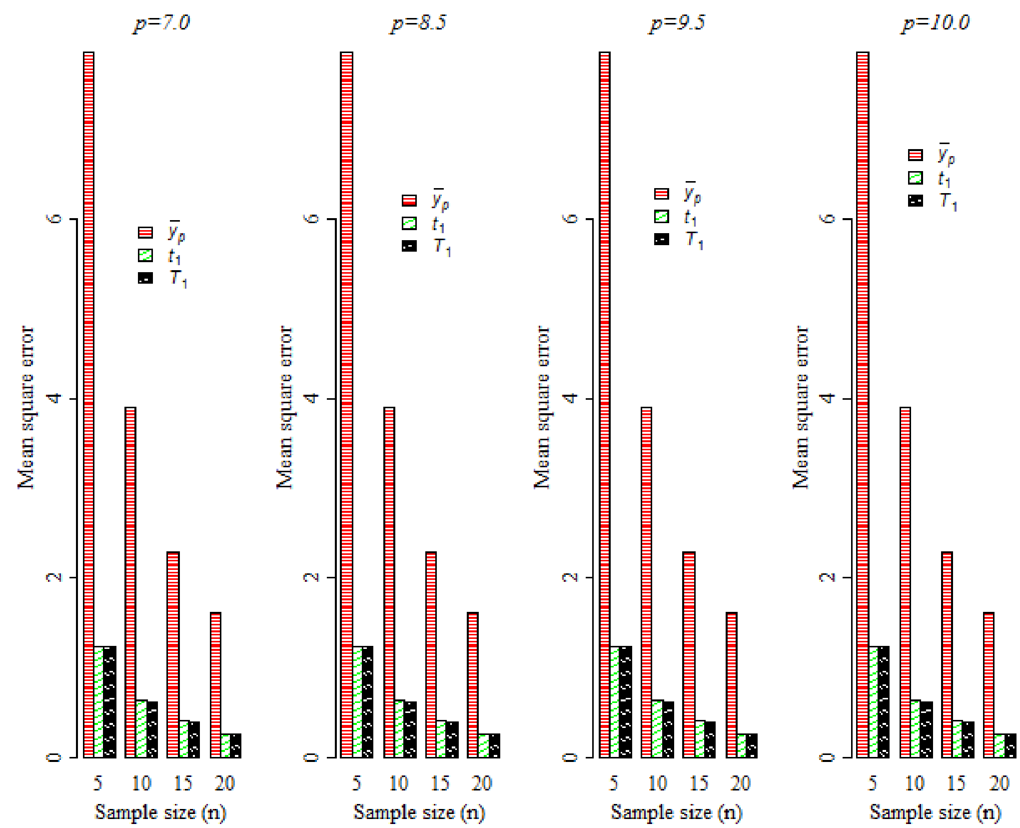

Figure 7. Mean square errors of different estimators for different values of n and p

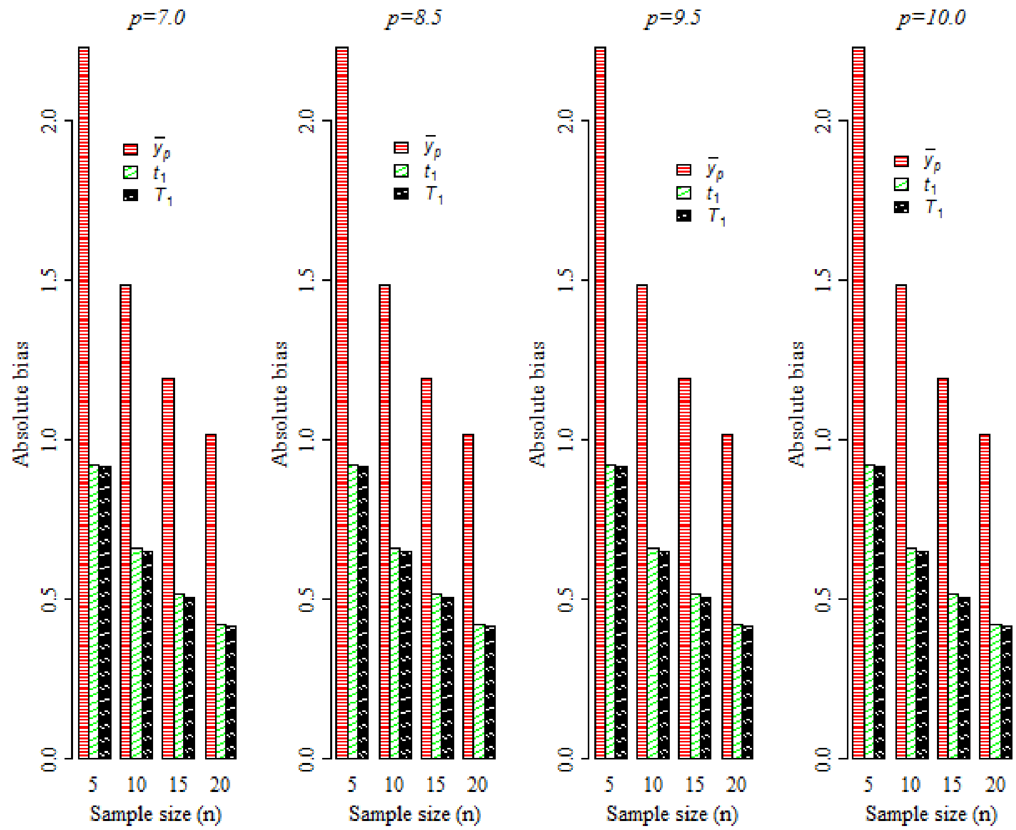

Figure 8. Absolute bias of different estimators for different values of *n* and *p*

Confidence Interval

The $100(1-\alpha)$ percent confidence intervals for the estimators T_1, t_1, and $\bar{y}_p$ are given by

$$T_1 \pm t_\vartheta(\alpha)\sqrt{\text{MSE}(T_1)}, t_1 \pm t_\vartheta(\alpha)\sqrt{\text{MSE}(t_1)}, \text{ and } \bar{y}_p \pm t_\vartheta(\alpha)\sqrt{\text{MSE}(\bar{y}_p)},$$

where $t_\vartheta(\alpha)$ is the $100(1-\alpha)\%$ point of the Student t distribution with $\vartheta = n-1$ degrees of freedom. The confidence interval $T_1 \pm t_\vartheta(\alpha)\sqrt{\text{MSE}(T_1)}$ is considerably shorter than the classical intervals $t_1 \pm t_\vartheta(\alpha)\sqrt{\text{MSE}(t_1)}$ and

$\bar{y}_p \pm t_g(\alpha)\sqrt{\text{MSE}(\bar{y}_p)}$. For $p = \infty$, the confidence interval $T_1 \pm t_g(\alpha)\sqrt{\text{MSE}(T_1)}$ reduces to the confidence interval $t_1 \pm t_g(\alpha)\sqrt{\text{MSE}(t_1)}$. Here, we consider $\alpha = 5\%$ level of significance.

The coverage of the estimates of the different estimators are now compared, and the standard deviation, lower and upper quartile, and the median are obtained from the 1,000,000 simulations. Violin plots are shown for the different estimators (the red line indicates the value of $\bar{Y}$); the dashed green line indicates the lower limit and the dotted blue line indicates the upper limit for the usual estimator ($\bar{y}_p$) at the 95% confidence interval for getting a visual conformation of the numbers just presented.

Table 7. Simulated confidence intervals, coverage (%) of the estimates, simulated estimates, and quartiles of the estimators T_1, t_1, and $\bar{y}_p$ for the generated and real data

Exp(1): $p = 2.5$, $\bar{Y} = -0.990$

		Confidence interval			Coverage	Sim.	Std.	Lower		Upper
n	**Est.**	**L limit**	**U limit**	**U – L**	**(%)**	**est.**	**dev.**	**quartile**	**Median**	**quartile**
5	T_1	-2.648	0.702	3.350	99.723	-0.970	0.769	-1.455	-0.949	-0.464
	t_1	-2.748	0.755	3.503	99.491	-1.000	0.811	-1.502	-0.971	-0.473
	$\bar{y}_p$	-3.737	1.351	5.087	94.860	-1.190	1.328	-1.687	-0.847	-0.322
10	T_1	-2.107	0.222	2.328	99.858	-0.940	0.526	-1.282	-0.929	-0.587
	t_1	-2.243	0.262	2.505	99.602	-0.990	0.573	-1.357	-0.980	-0.609
	$\bar{y}_p$	-2.876	0.690	3.566	95.741	-1.090	0.876	-1.504	-0.915	-0.486
15	T_1	-1.877	0.013	1.890	99.898	-0.930	0.423	-1.209	-0.923	-0.645
	t_1	-2.012	0.031	2.043	99.622	-0.990	0.466	-1.292	-0.982	-0.681
	$\bar{y}_p$	-2.500	0.383	2.884	96.165	-1.060	0.690	-1.411	-0.939	-0.574

Real data: $p = 8.5$, $\bar{Y} = 15.336$

		Confidence interval			Coverage	Sim.	Std.	Lower		Upper
n	**Est.**	**L limit**	**U limit**	**U – L**	**(%)**	**est.**	**dev.**	**quartile**	**Median**	**quartile**
5	T_1	13.398	17.256	3.859	99.108	15.330	1.145	14.550	15.300	16.080
	t_1	13.390	17.273	3.883	99.096	15.330	1.151	14.550	15.310	16.090
	$\bar{y}_p$	12.205	18.309	6.105	91.330	15.260	1.794	13.990	15.190	16.440
10	T_1	13.995	16.654	2.659	99.220	15.320	0.787	14.790	15.310	15.840
	t_1	13.989	16.679	2.690	99.182	15.330	0.796	14.790	15.320	15.860
	$\bar{y}_p$	13.179	17.420	4.241	91.194	15.300	1.250	14.440	15.270	16.120
15	T_1	14.257	16.378	2.121	99.292	15.320	0.627	14.890	15.310	15.740
	t_1	14.255	16.407	2.152	99.232	15.330	0.636	14.900	15.320	15.750
	$\bar{y}_p$	13.600	17.020	3.420	90.970	15.310	1.010	14.610	15.280	15.980

In Table 7, the confidence intervals are presented for the estimators T_1, t_1, and $\bar{y}_p$ along with corresponding coverage (%) of the estimates in the intervals, the simulated estimates, standard deviations, lower quartiles, medians, and the upper quartiles for both the generated data (p = 2.5) and the real data set (p = 8.5) for different sample sizes (n = 5, 10, 15).

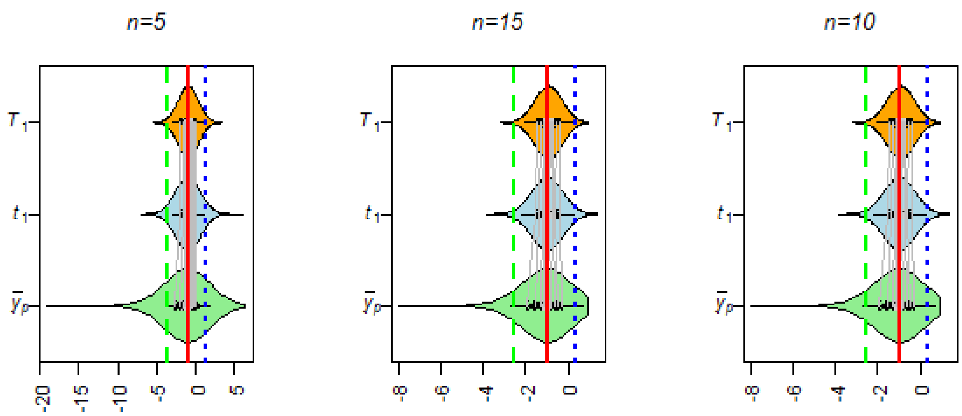

Figure 9. Coverage (%) of different estimators for different values of n

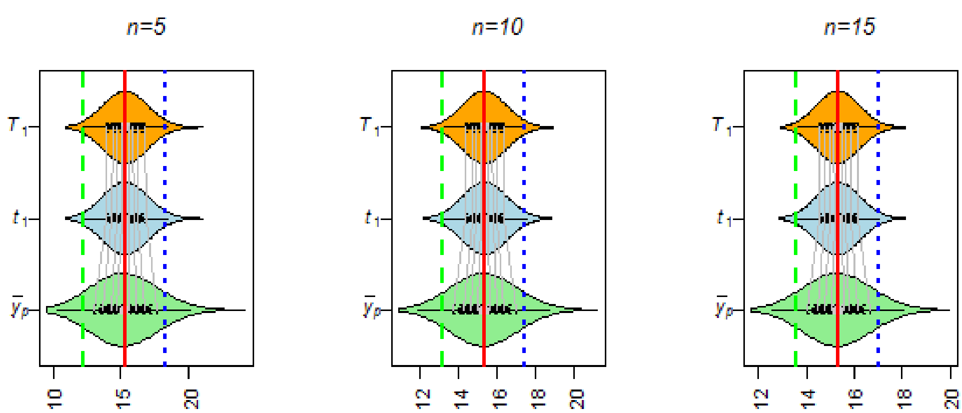

Figure 10. Coverage (%) of different estimators for different values of *n*

From Table 7, we observe that the confidence interval of the proposed estimator is shorter than that of the relevant estimators. Also, the standard deviation of the proposed estimator is less than that of the other estimators. The coverage of the estimate of the proposed estimator is more than the others. When the sample size is increased via more information, the confidence interval becomes shorter, the standard deviation decreases, the coverage of the estimate increases, and the lower as well as the upper quartiles tend to the median value.

In Figures 9 and 10, violin plots are presented for the coverage (%) of the estimates in the confidence interval of the traditional product estimator and we observe that the coverage of the estimate of the proposed estimator is more than that of the others. Note when increasing the sample size, the coverage of the estimate increases.

Table 8. Simulated confidence intervals, coverage (%), simulated estimates, and quartiles for the generated and real data

Exp(1): n = 10

			Confidence interval								
$\bar{Y}$	p	Est.	L limit	U limit	U – L	Cov. (%)	Sim. est.	Std. dev.	Lower quartile	Median	Upper quartile
-0.990	2.5	T_1	-2.648	0.702	3.350	99.723	-0.970	0.769	-1.455	-0.949	-0.464
		t_1	-2.748	0.755	3.503	99.491	-1.000	0.811	-1.502	-0.971	-0.473
		$\bar{y}_p$	-3.737	1.351	5.087	94.860	-1.190	1.328	-1.687	-0.847	-0.322
-0.990	4.5	T_1	-2.107	0.222	2.328	99.858	-0.940	0.526	-1.282	-0.929	-0.587
		t_1	-2.243	0.262	2.505	99.602	-0.990	0.573	-1.357	-0.980	-0.609
		$\bar{y}_p$	-2.876	0.690	3.566	95.741	-1.090	0.876	-1.504	-0.915	-0.486
-1.000	5.5	T_1	-1.877	0.013	1.890	99.898	-0.930	0.423	-1.209	-0.923	-0.645
		t_1	-2.012	0.031	2.043	99.622	-0.990	0.466	-1.292	-0.982	-0.681
		$\bar{y}_p$	-2.500	0.383	2.884	96.165	-1.060	0.690	-1.411	-0.939	-0.574

Real data: n = 10, $\bar{Y}$ = 15.336

		Confidence interval								
p	Est.	L limit	U limit	U – L	Cov. (%)	Sim. est.	Std. dev.	Lower quartile	Median	Upper quartile
7.0	T_1	13.398	17.256	3.859	99.108	15.330	1.145	14.550	15.300	16.080
	t_1	13.390	17.273	3.883	99.096	15.330	1.151	14.550	15.310	16.090
	$\bar{y}_p$	12.205	18.309	6.105	91.330	15.260	1.794	13.990	15.190	16.440
8.5	T_1	13.995	16.654	2.659	99.220	15.320	0.787	14.790	15.310	15.840
	t_1	13.989	16.679	2.690	99.182	15.330	0.796	14.790	15.320	15.860
	$\bar{y}_p$	13.179	17.420	4.241	91.194	15.300	1.250	14.440	15.270	16.120
9.5	T_1	14.257	16.378	2.121	99.292	15.320	0.627	14.890	15.310	15.740
	t_1	14.255	16.407	2.152	99.232	15.330	0.636	14.900	15.320	15.750
	$\bar{y}_p$	13.600	17.020	3.420	90.970	15.310	1.010	14.610	15.280	15.980

In Table 8, confidence intervals are presented for the estimators T_1, t_1, and $\bar{y}_p$ along wtih corresponding coverage (%) of the estimates in the intervals, the simulated estimates, standard deviations, lower quartiles, medians, and the upper quartiles for the fixed sample size ($n = 10$) and for different shape parameters $p = 2.5$, 4.5, 5.5 and $p = 7.0$, 8.5, 9.5 for the generated data and real data, respectively. The confidence interval of the proposed estimator is shorter than the other relevant estimators. Also, the standard deviation of the proposed estimator is less than that of the other estimators. The coverage of the estimate of the proposed estimator is more than that of the others. When the shape parameter is increase, i.e., tends to normality, the confidence interval of the proposed estimator T_1 becomes closer to the estimator t_1, the standard deviation increases, the coverage of the estimate of the proposed estimator T_1 decreases and becomes closer to that of the estimator t_1, and the lower as well as the upper quartiles tend far from the median value.

In Figures 11 and 12, violin plots are presented for the coverage (%) of the estimates in the confidence interval of the traditional product estimator, and the coverage of the estimate of the proposed estimator is more than the others. When the shape parameters increase, the coverage of the estimate is decreasing and the coverage of the estimate of the proposed estimator T_1 becomes closer to that of the estimator t_1.

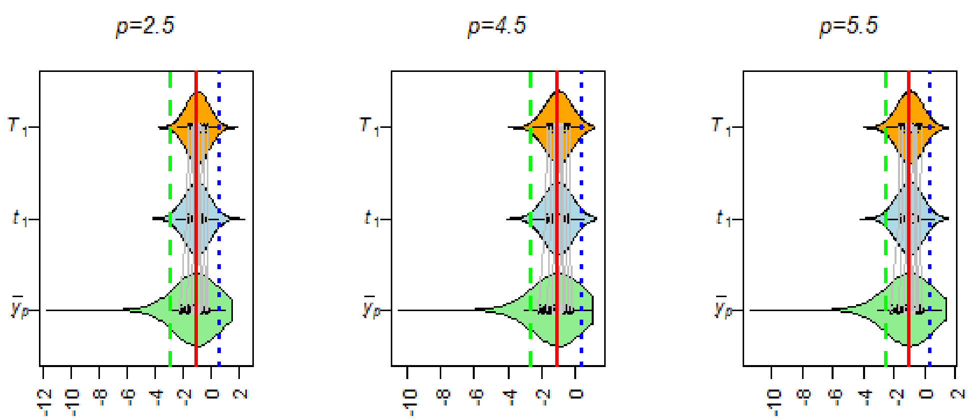

Figure 11. Coverage (%) of different estimators for different values of p

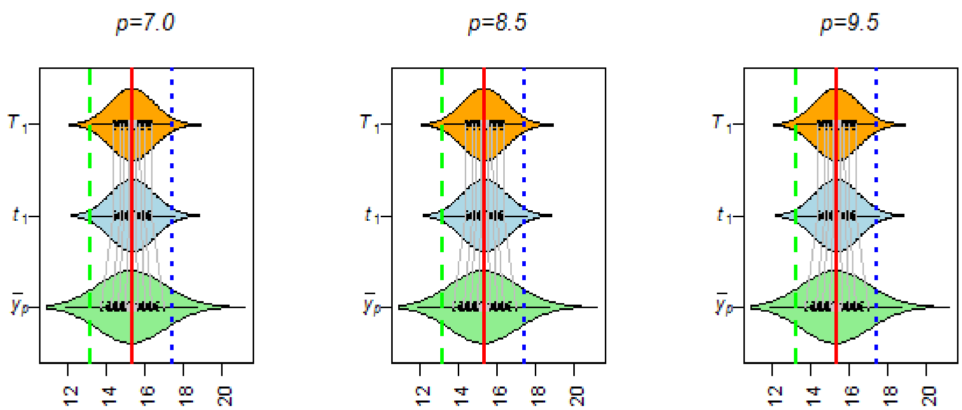

Figure 12. Coverage (%) of different estimators for different values of *p*

Determination of Shape Parameter

Sometimes the shape parameter p is not known, and hence to determine whether a particular density is suitable for the underlying distribution of the study variable y, make a Q-Q plot by plotting the population quantiles for the density against the ordered values of y, where the population quantiles $t_{(i)}$ are calculated from

$$\int_{-\infty}^{t_{(i)}} t(u)\,du = \frac{i}{n+1}, 1 \leq i \leq n.$$

The Q-Q plot that closely approximates a straight line would be assumed to be the most appropriate. Using such a procedure, a plausible value may be obtained for the shape parameter.

Conclusion

The modified dual to product estimator (T_1) can improve the efficiency of the Bandopadhyaya dual to product estimator t_1 when the underlying population is not normal. The proposed estimator T_1 is also more efficient than the estimator $\bar{y}_p$ and the dual to product estimator T_1 is robust to outliers. The confidence interval of the proposed estimator is shorter than competitors. Also, the standard deviation of the proposed estimator is at a minimum compared with the other estimators, and the coverage is greater.

References

Bandopadhyaya, S. (1980). Improved ratio and product estimators. *Sankhyā, Series C, 42*(1-2), 45-49.

Bouza, C. N. (2008). Ranked set sampling for the product estimator. *Investigación Operacional, 29*(3), 201-206.

Bouza, C. N. (2015). A family of ratio estimators of the mean containing primals and duals for simple random sampling with replacement and ranked set sampling designs. *Journal of Basic and Applied Research International, 8*(4), 245-253.

Chanu, W. W., & Singh, B. K. (2014a). An efficient class of double sampling dual to ratio estimators of population mean in sample surveys. *International Journal of Statistics & Economics, 14*(2), 25-40.

Chanu, W. W., & Singh, B. K. (2014b). Improved class of ratio-cum-product estimators of finite population mean in two phase sampling. *Global Journal of Science Frontier Research, 14*(2-1), 69-81.

Choudhury, S., & Singh, B. K. (2012). A class of chain ratio-cum-dual to ratio type estimator with two auxiliary characters under double sampling in sample surveys. *Statistics in Transition New Series, 13*(3), 519-536.

Cochran, W. G. (1977). *Sampling techniques* (3rd edition). New York: John Wiley & Sons.

Diana, G., Giordan, M., & Perri, P. F. (2011). An improved class of estimators for the population mean. *Statistical Methods & Applications, 20*(2), 123-140. doi: 10.1007/s10260-010-0156-6

Gupta, S., & Shabbir, J. (2008). On the improvement in estimating the population mean in simple random sampling. *Journal of Applied Statistics, 35*(5), 559-566. doi: 10.1080/02664760701835839

Gupta, S., & Shabbir, J. (2011). On estimating finite population mean in simple and stratified sampling. *Communications in Statistics – Theory and Methods, 40*(2), 199-212. doi: 10.1080/03610920903411259

Kumar, S. (2015). A robust regression type estimator for estimating population mean under non normality in the presence of non-response. *Global Journal of Science Frontier Research, 15*(7-1), 43-55.

Kumar, S., & Chhaparwal, P. (2016a). A generalized multivariate ratio and regression type estimator for population mean using a linear combination of two auxiliary variables. *Sri Lankan Journal of Applied Statistics, 17*(1), 19-37. doi: 10.4038/sljastats.v17i1.7843

Kumar, S., & Chhaparwal, P. (2016b). A robust dual to ratio estimator for population mean through modified maximum likelihood in simple random sampling. *Journal of Applied Probability and Statistics, 11*(2), 67-82.

Kumar, S., & Chhaparwal, P. (2016c). A robust unbiased dual to product estimator for population mean through modified maximum likelihood in simple random sampling. *Cogent Mathematics, 3*(1), 1168070. doi: 10.1080/23311835.2016.1168070

Kumar, S., & Chhaparwal, P. (2017). Robust exponential ratio and product type estimators for population mean using order statistics in simple random sampling. *International Journal of Ecological Economics and Statistics, 38*(3), 51-70.

Murthy, M. N. (1964). Product method of estimation. *Sankhyā, Series A, 26*(1), 69-74

Murthy, M. N. (1967). *Sampling theory and methods*. Calcutta: Statistical Publishing Society.

Oral, E. (2006). Binary regression with stochastic covariates. *Communications in Statistics – Theory and Methods, 35*(8), 1429-1447. doi: 10.1080/03610920600637123

Oral, E. (2010). Improving efficiency of ratio-type estimators through order statistics. In *JSM Proceedings, Section on Survey Research Methods* (pp. 4231-4239). Alexandria, VA: American Statistical Association.

Oral, E., & Kadilar, C. (2011). Robust ratio-type estimators in simple random sampling. *Journal of the Korean Statistical Society, 40*(4), 457-467. doi: 10.1016/j.jkss.2011.04.001

Oral, E., & Oral, E. (2011). A robust alternative to the ratio estimator under non-normality. *Statistics and Probability Letters, 81*(8), 930-936. doi: 10.1016/j.spl.2011.03.040

Puthenpura, S., & Sinha, N. K. (1986). Modified maximum likelihood method for the robust estimation of system parameters from very noisy data. *Automatica, 22*(2), 231-235. doi: 10.1016/0005-1098(86)90085-3

Ramos, E., Donoho, D., & UCI Machine Learning Repository. (1993). *Auto MPG data set* [Data set]. Retrieved from https://archive.ics.uci.edu/ml/datasets/Auto+MPG

Rao, J. N. K., & Beegle, L. D. (1967). A Monte Carlo study of some ratio estimators. *Sankhyā, Series B, 29*(1/2), 47-56.

Singh, H. P., & Solanki, R. S. (2012). An alternative procedure for estimating the population mean in simple random sampling. *Pakistan Journal of Statistics and Operations Research, 8*(2), 213-232. doi: 10.18187/pjsor.v8i2.252

Singh, S. (2003). *Advanced sampling theory with applications* (Vol. 1). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Sukhatme, P. V., Sukhatme, B. V., & Asok, C. (1984). *Sampling theory of surveys with applications* (3rd edition). New Delhi: Indian Society Agricultural Statistics.

Tato, Y., Singh, B. K., & Chanu, W. W. (2016). A class of exponential dual to ratio cum dual to product estimator for finite population mean in presence of non-response. *International Journal of Statistics & Economics, 17*(2), 20-31.

Tiku, M. L., & Bhasin, P. (1982). Usefulness of robust estimators in sample survey. *Communications in Statistics – Theory and Methods, 11*(22), 2597-2610. doi: 10.1080/03610918208828409

Tiku, M. L., & Kumra, S. (1985). Expected values and variances and covariances of order statistics for a family of symmetric distributions (Student's t). In W. J. Kennedy, R. E. Odeh, J. M. Davenport, & Institute of Mathematical Statistics (Eds.), *Selected tables in mathematical statistics* (Vol. 8) (pp. 141-270). Providence, RI: American Mathematical Society.

Tiku, M. L., & Suresh, R. P. (1992). A new method of estimation for location and scale parameters. *Journal of Statistical Planning and Inference, 30*(2), 281-292. doi: 10.1016/0378-3758(92)90088-A

Tiku, M. L., & Vellaisamy, P. (1996). Improving efficiency of survey sample procedures through order statistics. *Journal of Indian Society Agricultural Statistics, 49,* 363-385.

Vaughan, D. C. (1992a). On the Tiku-Suresh method of estimation. *Communications in Statistics – Theory and Methods, 21*(2), 451-469. doi: 10.1080/03610929208830788

Vaughan, D. C. (1992b). Expected values, variances and covariances of order statistics for Student's t-distribution with two degrees of freedom.

Communications in Statistics – Simulation and Computation, 21(2), 391-404. doi: 10.1080/03610919208813025

Vaughan, D. C., & Tiku, M. L. (2000). Estimation and hypothesis testing for non-normal bivariate distribution with applications. *Journal of Mathematical and Computer Modelling*, *32*(1-2), 53-67. doi: 10.1016/S0895-7177(00)00119-9

Yadav, S. K., & Kadilar, C. (2013). Improved class of ratio and product estimators. *Applied Mathematics and Computation, 219*(22), 10726-10731. doi: 10.1016/j.amc.2013.04.048

Yates, F. (1960). *Sampling methods in censuses and surveys* (3rd edition). London: Charles Griffin & Co.