Original Research Article

Robust Ratio Estimation with an Application to Covid-19 Data from Louisiana

ABSTRACT

Traditional ratio estimator loses its efficiency when there are outliers in the data or when the error term is not normally distributed. Specifically in health-related data, many biological processes can be modeled by Laplace distribution. We propose a novel robust ratio estimator that utilizes Lloyd's estimator for the cases where the error term is from the Laplace distribution. We derive the mean square error of the proposed estimator and compare it with some other existing estimators using extensive simulations. We use the proposed estimator to estimate Covid-19 cases and deaths in Louisiana and demonstrate its performance.

Keywords: Auxiliary variable, Covid-19, Generalized least square estimator, Modified maximum likelihood, Robust ratio estimator

1. INTRODUCTION

The ratio estimator improves the precision of the sample mean $\bar{y}$ when the study variable is linearly correlated with an auxiliary variable whose values are known for each unit of the population (Cochran, 1940). The ratio estimator is a good choice to estimate the population mean $\bar{Y}$ when the study variable has a Normal distribution (Cochran, 1977). Let the relationship between the study variable y_j, and the auxiliary variable z_j in the population is given as

$$y_j = \beta z_j + h(z_j)e_j \ , \tag{1.1}$$

where $h(z_j)$ is a function of z_j, e_j is an iid random variable with mean zero and variance σ_e^2, and $g(e_j)$ is the probability density function of e_j $(j = 1,2,3, \ldots, N)$. We refer to the assumed model in Eq. (1.1) as the *true model*. The accuracy of the estimates and statistical inferences depend on the accuracy of the assumptions made on the true model. The Gaussian assumption on $g(e_j)$ provides the well-known ratio estimator,

$$\bar{y}_1 = (\bar{y}/\bar{z})\bar{Z}, \tag{1.2}$$

with the mean square error (MSE)

$$MSE(\bar{y}_1) = var(\bar{y}) - 2Rcov(\bar{y}, \bar{z}) + R^2 var(\bar{z}), \tag{1.3}$$

where $\bar{z} = \sum_{i=1}^{n} z_j/n$, $\bar{Z} = \sum_{i=1}^{N} Z_j/N$, $var(\bar{z}) = \left[\frac{1-f}{n}\right] S_z^2$, $var(\bar{y}) = \left[\frac{1-f}{n}\right] S_y^2$, $cov(\bar{y}, \bar{z}) = \left[\frac{1-f}{n}\right] S_{zy}$, $f = n/N$ and $R = \bar{Y}/\bar{Z}$. The traditional ratio estimator in Eq. (1.2) is more efficient than the sample mean $\bar{y}$ if $\rho > V_z/2V_y$, where ρ is the population correlation coefficient between z *and* y, $V_z = S_Z/\bar{Z}$, $V_y = S_y/\bar{Y}$, S_Z and S_y are the population standard deviations of $z\, and\, y,$ respectively. There are real-life situations, however, where the assumptions on a true model might be violated. For example, model misspecification occurs when one assumes a specific theoretical model for the population under study, whereas a different model describes it better in reality. The estimates might not be reliable when there is misspecification in the model. Contamination occurs when a few values of the sample data are extreme. If these outliers have a potential to be influential points, ignoring them or proceeding with standard procedures can lead to a seriously biased inference. When there are such model violations, one needs to modify the estimation procedures and robustify the estimators. This study explores a novel estimation method in simple random sampling (SRS) to maintain the quality of the statistical inferences based on the true model when there is misspecification or contamination in the data.

In survey sampling, robust estimation of the mean in the presence of outliers under normality has been discussed by several authors. Farrel and Barrera (2006); Kadilar et al. (2007) and Subzar et al. (2019) utilized the M-estimation technique to create robust ratio-type estimators. Oral and Kadilar (2011 a and b) and Oral and Oral (2011) integrated Tiku's Modified Maximum Likelihood Estimator (MMLE) into various ratio-type estimators, and studied their properties under misspecification and contamination by following the model (1.1) in which $g(e_j)$ was assumed to be from a long tailed symmetric (LTS) family. LTS distribution is a symmetrical distribution with a shape parameter *p*, and its kurtosis changes from ∞, 9, 5, to 4.2 for *p*=2.5, 3, 4, and 5, respectively; when *p* tends to ∞, it approaches to a normal distribution. Oral and Kadilar (2011 a and b) and Oral and Oral (2011) showed that, when $g(e_j)$ follows a LTS distribution, or the data has outliers, their estimators provide more efficient estimates than the traditional ratio estimator; see also Tiku and Bhasin (1982) and Tiku and Vellaisamy's (1996). More recently, Azaz et al. (2019) highlighted several problems with MMLE's weight function, and suggested using the Generalized Least Squares Estimation (GLSE) instead of MMLE for the robustification process in SRS. They showed that when $g(e_j)$ follows a LTS distribution, integrating GLSE into the traditional ratio estimator yields more efficient ratio estimators with respect to the traditional, and MMLE integrated ratio estimator; see also Azaz et al. (2022).

Specifically, Oral and Oral (2011) proposed the following robust ratio estimator

$$\bar{y}_{OO} = \frac{\hat{\mu}_y}{\bar{z}}\bar{Z}$$

where $\hat{\mu}_y$, which is a weighted mean, is estimated using the MMLE from the LTS distribution. Azaz et al. (2019) showed that although the above estimator works very well for large samples, for very small sample sizes it does not assign small enough weights to the extreme values. Thus, they modified this estimator such that the weights would be smaller for extreme values and improve the robustness. Later, Saenaullah et al. (2021) replaced the $\hat{\mu}_y$ above with its Best Linear Unbiased Estimator (BLUE) of μ_y, and studied the properties of their proposed estimator.

Although the LTS family provides a flexible symmetric family of distributions for modeling, in many real-life applications the Laplace distribution might provide a better choice as it generally has a sharper peak depending on its scale parameter. As an example, Purdom and Holmes (2005) showed that the error distribution for gene expression data from microarray experiments can be fitted nicely with the Laplace distribution. Bottai and Zhang (2010) used Laplace distribution to model survival time in patients with small cell lung cancer. Biological processes and health-related data every so often reveal heavy-tailed distributions with a sharp peak. In such cases, Laplace distribution might be the natural choice. In fact, statistical models and applications based on Laplace distribution have been rapidly developed in the recent years: Yang (2014) provided a robust mean change point estimation in linear regression assuming that the errors follow the Laplace distribution. Song et al. (2014) proposed a robust estimation for mixture linear regression models under the assumption that the errors are from the Laplace distribution. More recently, Lu and Chang (2020) proposed a robust algorithm for multiphase regression models to deal with data drawn from heavy-tailed distributions. Thus, in this study, we extend the work of Azaz et al. (2019) to the case where the error term in Eq. (1.1) is from the Laplace distribution. We integrate the GLSE into the traditional ratio estimator, and study its properties and robustness under both misspecification and contamination models assuming that the true model's $g(e_j)$ is characterized by the Laplace distribution. We also use the publicly available Covid-19 data for Louisiana, and by estimating **i)** the average number of deaths using the average number of cases as the auxiliary information, **ii)** fourth wave's average number of deaths using the third wave's average number of deaths, **iii)** fourth wave's average number of cases using third wave's average number of cases, we demonstrate that the proposed estimator is superior to the traditional ratio estimator.

2. GLSE OF THE MEAN AND VARIANCE FOR THE LAPLACE DISTRIBUTION

Following the Lloyd (1952)'s GLSE procedure, let $y_1, y_2, y_3, \dots, y_n$ be a SRS from the Laplace distribution

$$f(y): L(\mu_y, \sigma_y) = \frac{1}{2\sigma} Exp\left[-\frac{|y-\mu_y|}{\sigma_y}\right], \quad -\infty < y < \infty \tag{2.1}$$

where μ_y and σ_y are the location and scale parameters, respectively. Let $y_{(1)} \le y_{(2)} \le \dots \le y_{(n)}$ be the order statistics from the sample above, and $z_{[1]} \le z_{[2]} \le \dots \le z_{[n]}$ be their concomitants in the true model (1.1). Let $V_{(j)} = \frac{y_{(j)}-\mu_y}{\sigma_y}$ $(j = 1,2,3, \dots, n)$ be the standardized variate in (2.1), and the means, variances and covariances of the order statistics $V_{(j)}$ are denoted as $t_{(j)}$, ω_{jj} and ω_{ji}, respectively. Further, let $\boldsymbol{y}' = (y_{(1)}, y_{(2)}, \dots, y_{(n)})$, $\boldsymbol{t}' = (t_{(1)}, t_{(2)}, \dots, t_{(n)})$, $\mathbf{1}' = (1,1,\dots,1)$ and $\boldsymbol{\Omega} = \omega_{ji}$ for $j, i = 1,2,\dots,n$. The BLUE of μ_y and σ_y for $L(\mu_y, \sigma_y)$ are given by,

$$\hat{\mu}_y^* = \frac{\mathbf{1}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}}{\mathbf{1}'\boldsymbol{\Omega}^{-1}\mathbf{1}} \text{ or } \left(\hat{\mu}_y^* = \sum_{j=1}^{n} \phi_j y_{(j)}\right) \tag{2.2}$$

and

$$\hat{\sigma}_y^* = \frac{\boldsymbol{t}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}}{\boldsymbol{t}'\boldsymbol{\Omega}^{-1}\boldsymbol{t}}, \tag{2.3}$$

where $\phi_j = \sum_{i=1}^{n} v_{ji} / \sum_{i=1}^{n} \sum_{j=1}^{n} v_{ji}$; v_{ji} are the elements of the inverse matrix $\boldsymbol{\Omega}$, and their variances are

$$var(\hat{\mu}_y^*) = \frac{\sigma_y^2}{\mathbf{1}'\boldsymbol{\Omega}^{-1}\mathbf{1}} \text{ and } var(\hat{\sigma}_y^*) = \frac{\sigma_y^2}{\boldsymbol{t}'\boldsymbol{\Omega}^{-1}\boldsymbol{t}} \tag{2.4}$$

(Govindarajulu, 1966). The exact values of $t_{(j)}$ and $\boldsymbol{\Omega}$ were tabulated by Govindarajulu (1966) for $n \le 20$. For large sample sizes the elements of $t_{(j)}$ may be calculated with the formula

$$\int_{-\infty}^{t_{(j)}} f(v)dv = \frac{j}{n+1}, \quad f(v) = \frac{1}{2} exp[-|v|], -\infty < v < \infty \tag{2.5}$$

and the elements of $\boldsymbol{\Omega}$ are determined from the equation

$$\omega_{ij} \cong p_j(1-p_j)/\{(n+2)f(u_i)f(u_j)\} \tag{2.6}$$

where $p_i = i/(n+1)$, $p_j = j/(n+1)$, $F(u_i) = p_i$ and $f(\mathrm{v}) = F'(\mathrm{v})$; see, David and Nagaraja, (2004). Using $t_{(j)}$ and $\boldsymbol{\Omega}$, one may get the solutions for Eq. (2.2)-(2.4).

To evaluate the weight function of GLSE, we calculate the values of the coefficients ϕ_j for n = 5, 10, 12 and 15 and present them in Figure 1. As can be seen from the Figure, the function ϕ_j allocates higher weights to the central observations and lower weights to the extreme observations, so the extreme values get minimum weights, and the effects of the outliers are minimized.

-FIGURE 1-

3. PROPOSED ROBUST RATIO ESTIMATOR

Assuming that the true model (1.1) follows the Laplace distribution given in (2.1) with $E(y_j|z_j) = \mu_j = \beta z_j$ and $var(y_j|z_j) = \sigma_e^2$ $(j = 1,2,3, \dots, n)$, we propose the robust ratio estimator

$$\bar{y}_p = \frac{\hat{\mu}_y^*}{\bar{z}} \bar{Z}, \tag{3.1}$$

where $\hat{\mu}_y^*$ is the BLUE given in (2.2). The approximate MSE of (3.1) under the true model (1.1) can be derived as follows. Let

$\bar{y}_p - \bar{Y} = \frac{\hat{\mu}_y^*}{\bar{z}}\bar{Z} - \bar{Y}$, or $\bar{y}_p - \bar{Y} = \bar{Z}(\hat{R} - R)$,

where

$\hat{R} = \frac{\hat{\mu}_y^*}{\bar{z}}$ and $R = \frac{\bar{Y}}{\bar{Z}}$. If we denote $g(\bar{z}, \hat{\mu}_y^*) = \hat{R}$ and $g(\bar{Z}, \bar{Y}) = R$, by applying the Taylor series approximation to $\hat{R} - R$ around $(\bar{Z}, \bar{Y})$ we get

$$g(\bar{z}, \hat{\mu}_y^*) \cong g(\bar{Z}, \bar{Y}) + (\bar{z} - \bar{Z})\left|\frac{\partial g(\bar{z}, \hat{\mu}_y^*)}{\partial \bar{z}}\right|_{\substack{\bar{z}=\bar{Z} \\ \hat{\mu}_y^*=\bar{Y}}} + (\hat{\mu}_y^* - \bar{Y})\left|\frac{\partial g(\bar{z}, \hat{\mu}_y^*)}{\partial \hat{\mu}_y^*}\right|_{\substack{\bar{z}=\bar{Z} \\ \hat{\mu}_y^*=\bar{Y}}},$$

$$\left(\hat{R} - R\right) \cong \left(\hat{\mu}_y^* - \bar{Y}\right)\frac{1}{\bar{Z}} - (\bar{z} - \bar{Z})\frac{\bar{Y}}{\bar{Z}^2},$$

$$\bar{Z}\left(\hat{R} - R\right) \cong \left(\hat{\mu}_y^* - \bar{Y}\right) - R(\bar{z} - \bar{Z}),$$

$$\bar{Z}^2 E\left(\hat{R} - R\right)^2 \cong E\left(\hat{\mu}_y^* - \bar{Y}\right)^2 + E(\bar{z} - \bar{Z})^2 R^2 - 2RE\left(\left(\hat{\mu}_y^* - \bar{Y}\right)(\bar{z} - \bar{Z})\right) \tag{3.2}$$

Thus, from (3.2) the MSE of the proposed estimator can be written as

$$MSE\left(\bar{y}_p\right) \cong var\left(\hat{\mu}_y^*\right) + R^2 var(\bar{z}) - 2R cov(\hat{\mu}_y^*, \bar{z}) \tag{3.2}$$

The variance $var(\hat{\mu}_y^*)$ in Eq. (3.3) can be written as $var(\hat{\mu}_y^*) = \sigma_y^2 \phi' \mathbf{\Omega} \phi$, where ϕ is the vector consisting of the elements of the coefficients ϕ_j, $var(\bar{z})$ is the same as given above, and $cov(\hat{\mu}_y^*, \bar{z})$ can be given as

$$cov(\hat{\mu}_y^*, \bar{z}) = cov\left(\sum_{j=1}^n \phi_j y_{(j)}, \frac{\sum_{j=1}^n z_j}{n}\right).$$

Since $\sum_{j=1}^n z_j = \sum_{j=1}^n z_{[j]}$, we can write

$$cov(\hat{\mu}_y^*, \bar{z}) = cov\left(\sum_{j=1}^n \phi_j y_{(j)}, \frac{\sum_{j=1}^n z_{[j]}}{n}\right),$$

or alternatively,

$$cov(\hat{\mu}_y^*, \bar{z}) = cov\left(\sum_{j=1}^n \phi_j [\mu_y + \sigma_y V_{(j)}], \sum_{j=1}^n [\mu_z + \sigma_z U_{[j]}]/n\right) \tag{3.4}$$

where $V_{(j)} = \frac{y_{(j)} - \mu_y}{\sigma_y}$, $U_{(j)} = \frac{z_{[j]} - \mu_z}{\sigma_z}$, and the covariance between $V_{(j)}$ and $U_{(j)}$ can be given as $cov(V_{(j)}, U_{(j)}) = \rho\sigma_z\sigma_y\mathbf{\Omega}$; see Wang (2008). The expression in Eq. (3.4) may also be written by using a matrix notation,

$$cov(\hat{\mu}_y^*, \bar{z}) = \rho\sigma_y\sigma_z\phi'\mathbf{\Omega}\delta, \tag{3.5}$$

where $\boldsymbol{\delta}$ is the $n \times 1$ vector with elements $1/n$ and $\phi' = (\phi_1, \phi_2, \ldots, \phi_n)$. Finally, the expression of $MSE\left(\bar{y}_p\right)$ is obtained as,

$$MSE\left(\bar{y}_p\right) \cong \sigma_y^2 \phi' \mathbf{\Omega}\, \phi + R^2 var(\bar{z}) - 2R\,\rho\sigma_y\sigma_z\phi'\mathbf{\Omega}\delta. \tag{3.6}$$

4. EFFICIENCY COMPARISONS

Suppose that the underlying super-population, i.e. the true model, is from (2.1). The GLS estimator $\hat{\mu}_y^*$ in (2.2) is calculated from the order statistics $y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}$ for the random sample of size n. In order to show that the proposed estimator is more efficient than the traditional sample mean, we first establish the conditions where GLSE $\hat{\mu}_y^*$ is more efficient than $\bar{y}$, which we denote by $\bar{y}_0$ and solve the inequality below

$$E(\bar{y}_o - \bar{Y})^2 > E(\hat{\mu}_y^* - \bar{Y})^2. \quad \textbf{(4.1)}$$

Here, $E(\bar{y}_o - \bar{Y})^2 = \frac{\sigma_y^2}{n}(1-f)$ and

$$\begin{aligned} E(\hat{\mu}_y^* - \bar{Y})^2 &= E\left[\left(\hat{\mu}_y^* - \mu_y\right) - (\bar{Y} - \mu_y)\right]^2 \\ &= E\left[\left(\hat{\mu}_y^* - \mu_y\right)^2 + (\bar{Y} - \mu_y)^2 - 2\left(\hat{\mu}_y^* - \mu_y\right)(\bar{Y} - \mu_y)\right] \\ &= var(\hat{\mu}_y^*) - 2cov(\hat{\mu}_y^*, \bar{Y}) + var(\bar{Y}) \end{aligned} \quad \textbf{(4.2)}$$

where $cov(\hat{\mu}_y^*, \bar{Y})$ can be obtained by using the identity $\sum_{j=1}^{N} Y_j = \sum_{j=1}^{n} y_j + \sum_{j=1}^{N-n} Y_j$ as

$$cov(\hat{\mu}_y^*, \bar{Y}) = \frac{n}{N} cov(\hat{\mu}_y^*, \bar{y}_o). \quad \textbf{(4.3)}$$

Since $Y_1, Y_2, \ldots, Y_N$ are independently and identically distributed random variables with

$$var(\bar{Y}) = \frac{\sigma_y^2}{N}, \quad \textbf{(4.4)}$$

integrating (4.3) and (4.4) into (4.2) provides

$$E(\hat{\mu}_y^* - \bar{Y})^2 = var\left(\hat{\mu}_y^*\right) - 2\frac{n}{N} cov\left(\hat{\mu}_y^*, \bar{y}_o\right) + \frac{\sigma_y^2}{N}. \quad \textbf{(4.5)}$$

Thus, the inequality in (4.1) becomes

$$\frac{\sigma_y^2}{n} > var\left(\hat{\mu}_y^*\right) + 2\frac{n}{N}\left(\frac{\sigma_y^2}{n} - cov\left(\hat{\mu}_y^*, \bar{y}_o\right)\right). \quad \textbf{(4.6)}$$

It can be showed that inequality (4.6) always satisfies, and hence $\hat{\mu}_y^*$ is always more efficient than $\bar{y}_o$. For demonstration, the results from a simulation study are given below for sample sizes n = 10, 20, 30 and 500.

n	$var(\bar{y}_o)$	$var(\hat{\mu}_y^*)$
10	0.2040	0.1460
20	0.0988	0.0662
30	0.0672	0.0425
500	0.0042	0.0021

Note that, the exact expressions of $var(\hat{\mu}_y^*)$ and $cov\left(\hat{\mu}_y^*, \bar{y}_o\right)$ are $var(\hat{\mu}_y^*) = \sigma_y^2 \phi' \mathbf{\Omega} \phi$ and $cov\left(\hat{\mu}_y^*, \bar{y}_o\right) = \phi' \mathbf{\Omega} \omega \sigma_y^2$.

For large samples, the proposed estimator in (3.1) has smaller MSE with respect to the MSE of the traditional ratio estimator in (1.2), when y is from Laplace distribution. To show that $\bar{y}_p$ is always more efficient than $\bar{y}_1$ for large samples, we write

$MSE(\bar{y}_1) = E(\bar{y}_1^2) - \bar{Y}^2$ and $MSE(\bar{y}_p) = E(\bar{y}_p^2) - \bar{Y}^2$. **(4.7)**

Considering two correlated random variables *y* and *z*, we can write $E(r(z)h(y)) = E(r(z)E(h(y)|z))$ and we obtain the expected values in Eq. (4.7) as

$$E(\bar{y}_1^2) = E\left(\left(\frac{\bar{Z}}{\bar{z}}\right)^2 \bar{y}_o^2\right) = \bar{Z}\, E\left[\left(\frac{1}{\bar{z}}\right)^2 E(\bar{y}_o^2|z)\right], \quad \textbf{(4.8)}$$

and

$$E(\bar{y}_p^2) = E\left(\left(\frac{\bar{Z}}{\bar{z}}\right)^2 \left(\hat{\mu}_y^*\right)^2\right) = \bar{Z}\, E\left[\left(\frac{1}{\bar{z}}\right)^2 E\left(\left(\hat{\mu}_y^*\right)^2 \Big| z\right)\right]. \quad \textbf{(4.9)}$$

Since $E(\bar{y}_o^2) > E\left(\hat{\mu}_y^*\right)^2$ from (4.1), we have $E(\bar{y}_o^2|z) > E\left(\left(\hat{\mu}_y^*\right)^2\Big|z\right)$; this follows from the fact that, on a probability space (Ψ, ξ, P), $P(E(y) \le E(z)) = 1$ implies $P\left(E(y\,|\,\mathrm{H}) \le E(z\,|\,\mathrm{H})\right) = 1$ for any $\mathrm{H} \subset \xi$ (Loeve, 1977). Consequently,

$$\left(\frac{1}{\bar{z}^2}\right) E(\bar{y}^2|z) > \left(\frac{1}{\bar{z}^2}\right) E\left(\left(\hat{\mu}_y^*\right)^2 \Big| z\right)$$

and

$$\bar{Z}^2 E((1/\bar{z}^2) E(\bar{y}^2|z) > \bar{Z}^2 E\left((1/\bar{z}^2) E\left(\hat{\mu}_y^*\right)^2 \Big| z\right);$$

therefore, it is clear from (4.7)-(4.9) that $MSE(\bar{y}_1) > MSE(\bar{y}_p)$ for large sample sizes. The proposed estimator $\bar{y}_p$ performs better than $\bar{y}_1$ for small samples, if

$MSE(\bar{y}_p) < MSE(\bar{y}_1)$, or

$$cov(\bar{y}_o, \bar{z}) < C_1, \text{ where } C_1 = \frac{1}{2R}\left[var(\bar{y}_o) - var(\hat{\mu}_y^*)\right] + cov(\hat{\mu}_y^*, \bar{z}). \quad \textbf{(4.10)}$$

5. ROBUSTNESS OF THE PROPOSED ESTIMATOR

To evaluate the robustness properties of the proposed estimator, we conduct an extensive simulation study as follows. We consider a population from the model (1.1), and generate e_j and z_j independently, where the random error e_j is from $L(0, \sigma_e^2)$ and z_j is from $U(0,1)$ $(1 \le j \le N)$. Let Π_N represent the super-population of size *N* consisting of $(z_1, y_1), (z_2, y_2), ..., (z_N, y_N)$. To assure that the correlation coefficient ρ is sufficiently high, the values of the parameter β in the model (1.1) are chosen such that the correlation coefficient is 0.60. The value of β which satisfies this condition is determined by $\beta^2 = \mathrm{var}(e)\rho^2 / (1-\rho^2)\,\mathrm{var}(z)$; see Oral and Kadilar (2011 a) and Oral and Oral (2011). To calculate the MSE of the estimators in (1.2) and (3.1), one has to calculate the $\bar{y}_o$, $\bar{y}_1$ and $\bar{y}_p$ from all ${}_N C_n$ possible samples of size n from Π_N. We consider the values $N = 500$ and $n = 10$, 30, 70 and 100. Since ${}_{500}C_n$ is extremely large, we choose *M*=50,000 possible simple random samples of size n which then give 50,000 values for each estimator, i.e., $\bar{y}_0$, $\bar{y}_1$ and $\bar{y}_p$. In calculating $\bar{y}_p$ using *n*=10, 30, 70 and 100, we also need to calculate the coefficients ϕ_j and integrate them into Eq. (3.1). To compare the efficiencies, we calculate the values of the MSEs of each estimator from the expressions $MSE(\bar{y}_0) = \sum_{i=1}^{M} (\bar{y}_0 - \bar{Y})^2 / M$,

$MSE(\bar{y}_1)=\sum_{i=1}^{M}(\bar{y}_1-\bar{Y})^2/M$ and $MSE(\bar{y}_p)=\sum_{i=1}^{M}(\bar{y}_p-\bar{Y})^2/M$, under the true model and under eleven different contamination or misspecification models. The description of each model is given below

True model: All N observations are from $L(0,1)$ and no outlier is present.

Dixon's outlier model-I: $N-N_0$ observations are from $L(0,1)$ and N_0 (we do not know which) are from $L(0,d)$, where N_0 is calculated from the formula $\left[\left|\frac{N}{10}+\frac{1}{2}\right|\right]$.

Dixon's outlier model-II: $N-N_0$ observations are from $L(0,1)$ and N_0 (we do not know which) are from $N(0,d)$, where N_0 is calculated from the formula $\left[\left|\frac{N}{10}+\frac{1}{2}\right|\right]$.

Tatum's, (1997) localized scale disturbances model-I: A proportion $1-\varphi$ of observations are from a population $L(0,1)$ and a proportion φ of the observations are from $L(0,d)$.

Tatum's, (1997) localized scale disturbances model-II: A proportion $1-\varphi$ of observations are from a population $L(0,1)$ and a proportion φ of the observations are from $N(0,d)$.

Amiri and Allahyari's (2012) single step shift in the scale model-I: the first $N(1-\varphi)$ observations are from $L(0,1)$ and last $N(\varphi)$ observations are from $L(0,d)$.

Amiri and Allahyari's (2012) single step shift in the scale model-II: the first $N(1-\varphi)$ observations are from $L(0,1)$ and last $N(\varphi)$ observations are from $N(0,d)$.

Misspecification model: $L(0,d=1.07)$,

where $\varphi(=0.10,0.15)$ denotes the percentage, and *d(=3,4,5)* refers to the extremity of the contamination. Realize that the first model, i.e., the true model, is given for the sake of comparisons, and all other models are its plausible alternatives. In order to make direct comparisons between these models, the generated e_j's ($j=1,2,3,...,N$) were standardized to have the same variance as that of the true model. The simulated values of the MSEs and their corresponding relative efficiencies $E_{0,h}$ are given in Tables 1 and 2, respectively, where $E_{0,h}=MSE(\bar{y}_0)/MSE(\bar{y}_h)$ for *h*=1, *p*, where $E_{0,1}$ denotes the relative efficiency of the traditional ratio estimator and $E_{0,p}$ denotes the relative efficiency of proposed robust ratio estimator.

-TABLE 1-

-TABLE 2-

It can be seen from Table 1 that, the MSE of the proposed robust ratio estimator slightly increases as the percentage of contamination increases (from a low value of $\varphi=0.10$ to a comparatively high value of $\varphi=0.15$), which is expected, but the increase in the MSE of the traditional ratio estimator is larger compared to the proposed estimator. Besides, for a given sample size, the MSE values of the proposed estimator almost stay the same compared to the true model demonstrating its robustness to its plausible deviations from the assumed Laplace distribution. Tables 1 and 2 together show that the MSEs of the proposed robust ratio estimator are smaller compared to MSEs of the traditional ratio estimator under all types of model violations. We conclude that the proposed robust ratio estimator is both robust and more efficient than the traditional ratio estimator when the error term is from Laplace distribution.

6. COVID-19 IN LOUISIANA: ESTIMATING THE MORTALITY AND CASE NUMBERS

The first case of Covid-19 was identified in Wuhan, China in December 2019. Since then, it has spread worldwide causing an ongoing pandemic. It has affected all nations profoundly, causing issues from crippling economies to mental health problems (Baldwin and Werner di Mauro, 2020; Roy et al., 2021; O'Connor et al., 2021; Liu et al, 2022).

Consequently, accurately estimating the Covid-19 infections and deaths has been crucial for epidemiologists, public health workers, and federal governments to be able to make public policies and combat the disease. There are not many studies that use survey sampling procedures to estimate Covid-19 cases. Recently, Chandra et al. (2021) have used adaptive cluster sampling to estimate Covid-19 cases in the Uttarakhand and Kerala states; see also Chandra et al. (2019). Epidemiologists generally use infectious disease models to estimate cases and deaths. However, with novel infectious diseases, such as Covid-19, the initial estimates regarding the cases and deaths can be erroneous (Biggs and Littlejohn, 2021). Besides, due to the evolving nature of Covid-19, it has been challenging to predict cases and deaths for different waves caused by different mutations. For example, while the Omicron variant was found to be more transmissible than the Delta variant, it was also observed to be less severe (Mahase, 2022). Some of these infectious disease models, such as the agent-based model, have been criticized for their unrealistic assumptions. Therefore, health researchers might want to consider utilizing more simplistic approaches in estimating the cases or deaths, especially when there are too many unknowns at the beginning of pandemics.

Louisiana (LA) state reported its first case of Covid-19 in March 2020 and immediately became one of the hot spots of the pandemic in the U.S. (KC et al., 2020). Two years after the first reported case, in March 2022, there were 1,229,511 total confirmed Covid-19 cases, and 16,862 total confirmed deaths from Covid-19 in the LA state. As of July 2023, these numbers increased to a total of 1,611,869 confirmed cases and a total of 19,049 deaths[‡]. Considering that the number of deaths due to Covid-19 is related to the number of confirmed cases, we analyze the publicly available data reported by New York Times[‡‡]. More specifically, we consider the newly confirmed Covid-19 daily cases and deaths in LA. As doctors and epidemiologists are interested more in average numbers than the raw numbers, we consider the average number of confirmed cases and average number of deaths**.** Our analyses include three different situations:

Case-I

We estimate the average number of Covid-19 deaths in LA between 7/1/2021 and 8/14/2021, using the average number of confirmed cases for the same time period as the auxiliary information.

Case-II

Considering LA's third wave's (between 11/5/2021 and 3/5/2021) average number of deaths as the auxiliary information, we estimate LA's fourth wave's (between 6/1/2021 and 9/29/2021) average number of deaths.

Case-III

Considering LA's third wave's average number of confirmed Covid-19 cases as the auxiliary information, we estimate LA's fourth wave's average number of confirmed Covid-19 cases.

-FIGURE 2-

-FIGURE 3-

See Figure 2 for the first five waves of Covid-19 in LA; scatter plots of the variables described for each situation above is given in Figure 3. Reviewing several normality tests, we conclude that none of the residuals are normally distributed (Table 3); the normal probability plots also confirms non-normality as well as the existence of several outliers (Figure 4). On the other hand, Laplace distribution fits very well to each case, see Figure 5. Goodness of fit tests from Laplace distribution also provide evidence that we can model the residuals using the Laplace distribution, see Table 4.

-TABLE 3-

-FIGURE 4-

-FIGURE 5-

-TABLE 4-

To evaluate the performance of the proposed robust ratio estimator with its competitors, we report the mean estimates and the MSEs given in (1.3) and (3.6) along with the relative efficiencies $E_{0,h} = \mathrm{var}(\bar{y}_0) / MSE(\bar{y}_h)$ where $h = 1, p$. The results are presented in Table 5. The R code that was used for calculations can be obtained from the contact author upon request.

-TABLE 5-

From Table 5, it can be seen that the MSE of the traditional ratio estimator is inflated. The MSE of the proposed robust ratio estimator is stable in the presence of outliers. Besides, the proposed robust ratio estimator is more efficient compared to the traditional ratio estimator. This result is expected since the condition (3.10) is satisfied for all cases studied (see, Table 4). We conclude that the proposed estimator based on GLSE is a better estimator than its counterparts.

7. DISCUSSION AND CONCLUSION

The traditional ratio estimator becomes unstable and inefficient when there are outliers in the data, or if the underlying distribution is not Normal. To increase the efficiency of the traditional ratio estimator when the error distribution is from Laplace, we utilized the GLS estimation, which also provides robustness by assigning small weights to the outliers and large weights to the central observations. We first studied the efficiency and robustness properties of the proposed ratio estimator via an extensive simulation study. We then applied the novel robust estimator to Covid-19 data from Louisiana and showed that it is more efficient than the sample mean and the traditional ratio estimator. Using the Covid-19 data we also showed that the results from the traditional and the proposed ratio estimators are not close to each other; one cannot trust the over-estimated results from the traditional ratio estimator since it is highly affected by the outliers. To our knowledge, this is the first time in survey sampling literature that a ratio-type estimator was used to analyze health-related data. The novel estimator we proposed in this study can specifically help researchers to obtain robust estimates in analyzing medical and biological data when there is auxiliary information available about the population and the underlying distribution is Laplace.

‡ Source: Louisiana Department of Health
‡‡ Covid-19 cases and deaths for all Parishes of the Louisiana State was obtained from the New York Times, based on reports from state and local health agencies. See References Section for full citation.

References

1. Ahmad A, Sanaullah A, Hanif M. Modified maximum likelihood integrated robust ratio estimator in simple random sampling. *Punjab University Journal of Mathematics* 2019; 51:3, 133-145.
2. Ahmad A, Sanaullah A, Oral E, Hanif M. Robust ratio estimators of population mean for skewed and contaminated population. *Journal of Statistical Computation and Simulation* 2022; 93:5, 800-817.
3. Amiri A, Allahyari S. Change point estimation methods for control chart post signal diagnostics: a literature review. *Quality and Reliability Engineering International* 2012; 28:7, 673-685.
4. Baldwin R, Weder di Mauro B. *Economics in the Time of Covid-19*, London: CEPR Press, 2020
5. Biggs AT, Littlejohn LF. Revisiting the initial COVID-19 pandemic projections. *The Lancet. Microbe* 2021; 2:3, e91–e92. https://doi.org/10.1016/S2666-5247(21)00029-X
6. Brownlee KA. *Statistical theory and Methodology in science and Engineering.* 2nd Ed. New York: Wiley and Sons, 1965.
7. Cabilio P, Masaro J. A simple test of symmetry about an unknown median. *The Canadian Journal of Statistics* 1996; 24, 349-361.
8. Chambers RL. Outlier robust finite population estimation. *Journal of American Statistical Association* 1986; 81, 1063–1069.
9. Chandra G, Tiwari N, Nautiyal R. Two stage adaptive cluster sampling based on ordered statistics. *Metodoloski Zvezki* 2019; *16(1),* 43-60.
10. Chandra G, Tiwari N, Nautiyal R. Adaptive cluster sampling-based design for estimating Corona COVID-19 cases with random samples, *Current Science* 2021; 120(7), 1202-1210.
11. Cochran WG. *Sampling techniques.* New York: John Wiley and Sons, 1977.
12. Cochran WG. The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *Journal of Agriculture and Sciences* 1940; 30, 262-275.
13. O'Connor RC, Wetherall K, Cleare S, McClelland H, Melson AJ, Niedzwiedz CL, O'Carroll RE, O'Connor DB, Platt S, Scowcroft E, Watson B. Mental health and well-being during the COVID-19 pandemic: Longitudinal analyses of adults in the UK COVID-19 Mental Health & Wellbeing study. *The British Journal of Psychiatry* 2021; 218:6, 326–333.
14. David HA, Nagaraja HN. Order statistics. *Encyclopedia of Statistical Sciences*, 9. John Wiley and Sons, 2004.
15. Farrella JP, Barrera MS. A comparison of several robust estimators for a finite population mean. *Journal of Statistical Studies* 2006; 26, 29-43.
16. Govindarajulu Z. Best linear estimates under symmetric censoring of the parameters of a double exponential population. *Journal of the American Statistical Association* 1966; 61:313, 248-258.
17. Kadilar C, Candan M, Cingi H. Ratio estimators using robust regression. *Hacettepe Journal of Mathematics and Statistics* 2007; 36:2, 181-188.
18. Liu Q, Liu Z, Lin S, Zhao P. Perceived accessibility and mental health consequences of COVID-19 containment policies. *J Transp Health* 2022; 25:101354. doi: 10.1016/j.jth.2022.101354
19. Lloyd EH. Least-squares estimation of location and scale parameters using order statistics. *Biometrika* 1952; 39:1/2, 88-95.
20. Loeve M. *Probability Theory I*. New York : Springer, 1977.
21. Lu KP, Chang ST. Robust algorithms for multiphase regression models. *Applied Mathematical Modelling* 2020; *77*, 1643-1661.
22. Mahase E. Covid-19: What do we know about Omicron sublineages? *BMJ* 2022; 376:o358. doi: https://doi.org/10.1136/bmj.o358
23. Miao W, Gel YR, Gastwirth JL. A new test of symmetry about an unknown median. *Random Walk, Sequential Analysis and Related Topics* 2006; 199-214.
24. Mira A. Distribution-free test for symmetry based on Bonferroni's measure. *Journal of Applied Statistics* 1999; 26, 959-972.
25. Oral E, Kadilar C. Robust ratio-type estimators in simple random sampling. *Journal of the Korean Statistical Society* 2011a; 40:4, 457-467.
26. Oral E, Kadilar C. Improved Ratio Estimators via Modified Maximum Likelihood. *Pak. J. Statist.* 2011b; 27:3, 269-282.
27. Oral E, Oral E. A robust alternative to the ratio estimator under non-normality. *Statistics and Probability Letters* 2011; 81:8, 930-936.
28. Purdom E, Holmes SP. Error distribution for gene expression data. *Statistical Applications in Genetics and Molecular Biology* 2005; 4:1,1-33.
29. Rousseeuw PJ, Leroy AM. *Robust Regression and Outlier Detection*. New York: John Wiley and Sons, 1987.
30. Roy A, Singh AK, Mishra S, Chinnadurai A, Mitra A, Bakshi O. Mental health implications of COVID-19 pandemic and its response in India. *Int J Soc Psychiatry* 2021; 67:5:587–600. doi: 10.1177/0020764020950769.
31. Saenaullah A, Ahmed A, Hanif M. A new robust ratio estimator with reference to non-normal distribution. Commun Stat- Theory and Methods 2021; 50:5 1099-1116.

32. Song W, Yao W, Xing Y. Robust mixture regression model fitting by Laplace distribution. *Computational Statistics and Data Analysis* 2014; 71, 128-137.
33. Subzar M, Bouza CN, Maqbool S, Raja TA, Para BA. Robust ratio type estimators in simple random sampling using Huber M estimation. *Investigación Operacional* 2019; 40:2, 201-209.
34. Tatum LG. Robust estimation of the process standard deviation for control charts. *Technometrics* 1997; 39:2, 127-141.
35. The New York Times. Coronavirus (Covid-19) Data in the United States. Retrieved November-December 2021, from https://github.com/nytimes/covid-19-data.
36. Tiku ML, Bhasln P. Usefulness of robust estimators in sample survey. *Communications in Statistics Theory and Methods* 1982; 11, 2597-2610.
37. Tiku ML, Vellaisamy P. Improving efficiency of survey sample procedures through order statistics. *Journal of the Indian Society of Agricultural Statistics* 1996; 49, 363-385.
38. Tiku ML, Suresh RP. A new method of estimation for location and scale parameters. *Journal of Statistical Planning and Inference* 1992; 30:2, 281-292.
39. Wang K. *On concomitants of order statistics*. PhD Thesis, The Ohio State University, USA, 2008.
40. Yang F. Robust mean change-point detecting through Laplace linear regression using EM algorithm. *Journal of Applied Mathematics* 2014; 1-9. https://doi.org/10.1155/2014/856350

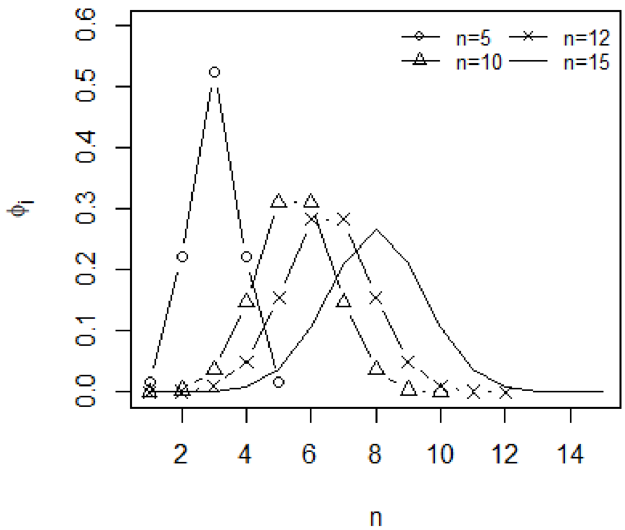

Figure 1. The weight function ϕ_j in GLSE to estimate the population mean

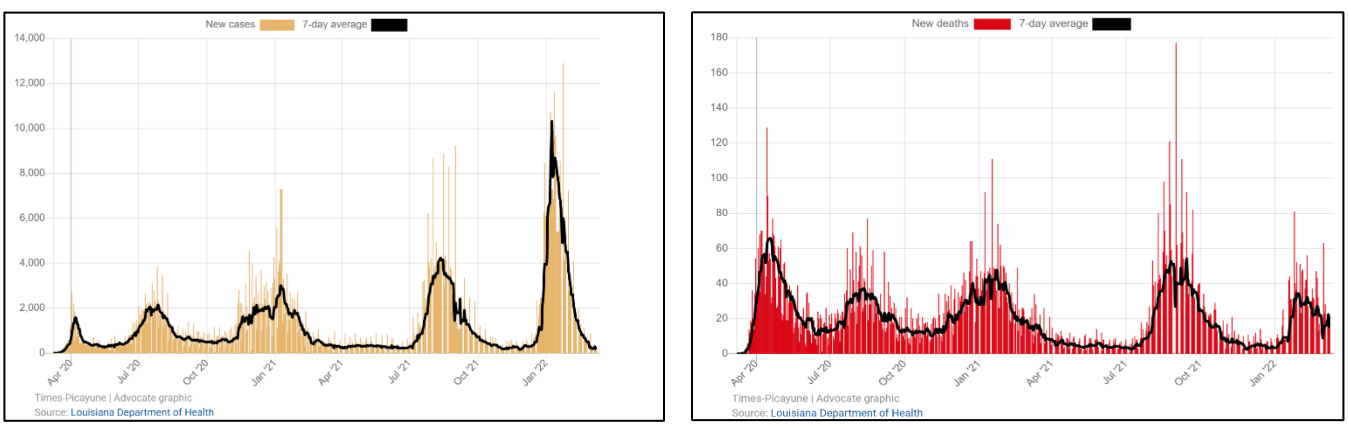

Figure 2. Covid-19 cases and deaths in LA; the first five waves.

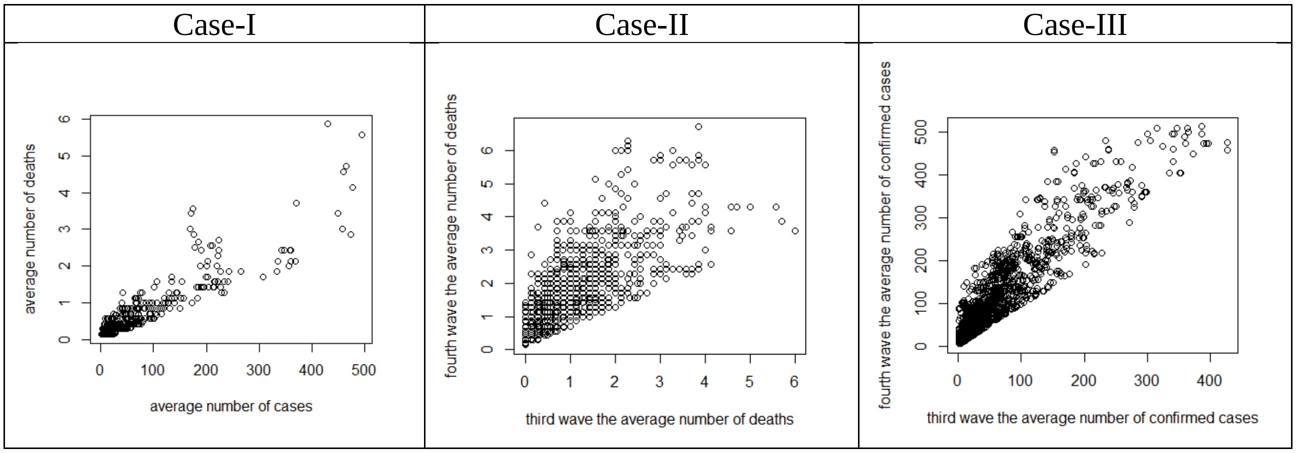

Figure 3. Scatter plot of the variables for Cases I-III

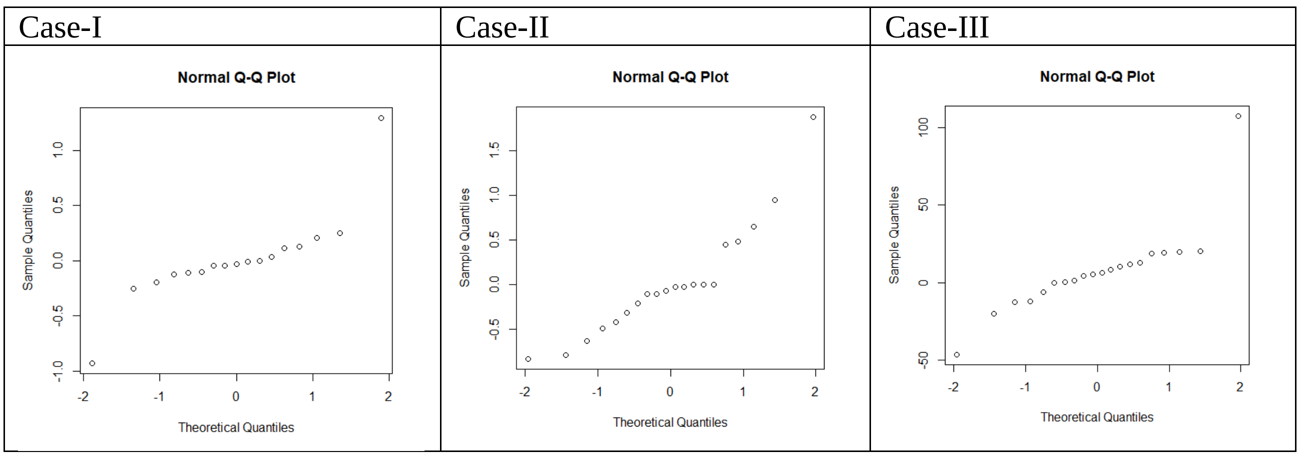

Figure 4. The Normal Probability plot of the residuals

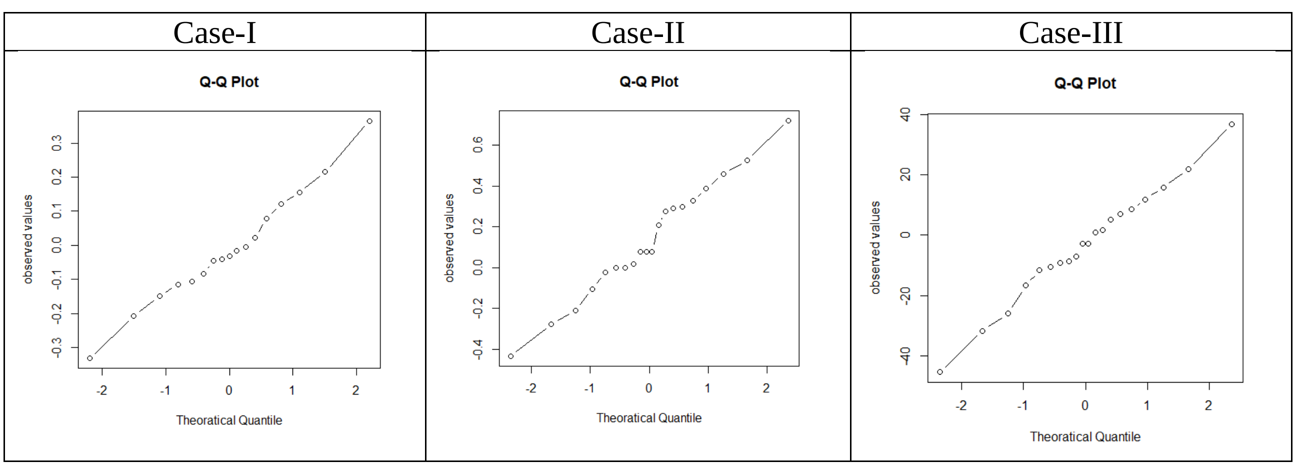

Figure 5. The Q-Q plot of the residuals where the theoretical quantiles are calculated from Laplace distribution

Table 1. The values of (1) $MSE(\bar{y}_o)$, (2) $MSE(\bar{y}_1)$ and (3) $MSE(\bar{y}_p)$

	n=10			*n*=30			*n*=70			*n*=100		
d	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
(i) True Model												
-	0.2081	0.2427	0.1836	0.0693	0.0665	0.0532	0.0298	0.0271	0.0237	0.0207	0.0191	0.0142
(ii) Dixon's outlier model-I												
3	0.3670	0.3622	0.1876	0.1223	0.1175	0.0610	0.0523	0.0449	0.0255	0.0367	0.0298	0.0179
4	0.5153	0.5015	0.1943	0.1677	0.1749	0.0678	0.0722	0.0618	0.0257	0.0504	0.0383	0.0170
5	0.6776	0.6599	0.1925	0.2297	0.2133	0.0601	0.0983	0.0908	0.0282	0.0672	0.0510	0.0172
(iii) Dixon's outlier model-II												
3	0.7976	0.8599	0.2301	0.2667	0.2564	0.0652	0.1132	0.1022	0.0276	0.0800	0.0661	0.0195
4	1.2591	1.2251	0.2133	0.4193	0.3983	0.0633	0.1807	0.1604	0.0276	0.1265	0.1055	0.0196
5	1.8796	1.9995	0.2394	0.6159	0.5538	0.0606	0.2653	0.2296	0.0273	0.1882	0.1477	0.0183
(iv) Tatum's, (1997) localized scale disturbances model-I using φ =0.10												
3	0.3672	0.3813	0.1961	0.1220	0.1087	0.0570	0.0525	0.0435	0.0244	0.0365	0.0289	0.0175
4	0.5216	0.6336	0.2380	0.1705	0.1677	0.0632	0.0718	0.0560	0.0234	0.0508	0.0419	0.0184
5	0.6796	0.7160	0.2065	0.2282	0.2171	0.0622	0.0982	0.0859	0.0265	0.0689	0.0514	0.0175
(v) Tatum's, (1997) localized scale disturbances model-II using φ =0.10												
3	0.7777	0.8149	0.2188	0.2641	0.2552	0.0652	0.1131	0.0948	0.0262	0.0789	0.0623	0.0187
4	1.2660	1.3443	0.2366	0.4213	0.4077	0.0655	0.1815	0.1709	0.0296	0.1286	0.1031	0.0191
5	1.8838	2.3185	0.2735	0.6170	0.6388	0.0705	0.2633	0.2266	0.0268	0.1872	0.1458	0.0183
(vi) Tatum's, (1997) localized scale disturbances model-I using φ =0.15												
3	0.4508	0.4622	0.2118	0.1497	0.1456	0.0677	0.0638	0.0557	0.0272	0.0451	0.0364	0.0190
4	0.6548	0.7470	0.2513	0.2174	0.2084	0.0675	0.0939	0.0756	0.0262	0.0647	0.0562	0.0214
5	0.9150	1.0337	0.2586	0.3075	0.2852	0.0669	0.1322	0.1132	0.0280	0.0932	0.0799	0.0209
(vii) Tatum's, (1997) localized scale disturbances model-II using φ =0.15												
3	1.0987	1.2493	0.2919	0.3682	0.3566	0.0731	0.1555	0.1347	0.0299	0.1087	0.1006	0.0244
4	1.7977	1.9547	0.2999	0.5938	0.5759	0.0727	0.2529	0.2201	0.0301	0.1805	0.1464	0.0210
5	2.7223	2.8950	0.3093	0.9140	0.8404	0.0705	0.3889	0.3373	0.0301	0.2716	0.2113	0.0203
(viii) Amiri and Allahyari's (2012) single step shift in the scale model-I using φ =0.10												
3	0.3656	0.3702	0.1908	0.1222	0.1192	0.0624	0.0525	0.0437	0.0245	0.0368	0.0276	0.0170
4	0.5138	0.5768	0.2169	0.1701	0.1626	0.0629	0.0716	0.0594	0.0250	0.0515	0.0418	0.0185
5	0.6963	0.7531	0.2173	0.2253	0.2218	0.0643	0.0979	0.0852	0.0266	0.0674	0.0510	0.0170
(ix) Amiri and Allahyari's (2012) single step shift in the scale model-II using φ =0.10												
3	0.8022	0.8517	0.2258	0.2633	0.2725	0.0699	0.1133	0.0981	0.0274	0.0780	0.0633	0.0188
4	1.2643	1.3365	0.2319	0.4212	0.4330	0.0706	0.1828	0.1676	0.0286	0.1256	0.1126	0.0213
5	1.8650	2.0652	0.2460	0.6303	0.6040	0.0654	0.2657	0.2510	0.0254	0.1873	0.1770	0.0228
(x) Amiri and Allahyari's (2012) single step shift in the scale model-I using φ =0.15												
3	0.4555	0.4725	0.2168	0.1492	0.1409	0.0649	0.1561	0.1243	0.0277	0.0443	0.0342	0.0178
4	0.6450	0.7436	0.2494	0.2195	0.2025	0.0650	0.2587	0.2252	0.0305	0.0662	0.0506	0.0190
5	0.9325	1.0102	0.2527	0.3073	0.2780	0.0645	0.3864	0.3809	0.0347	0.0921	0.0743	0.0197
(xi) Amiri and Allahyari's (2012) single step shift in the scale model-II using φ =0.15												
3	1.0821	1.3118	0.3134	0.3634	0.3519	0.0722	0.1557	0.1274	0.0283	0.1082	0.0978	0.0238
4	1.7935	1.9424	0.3016	0.5954	0.5821	0.0741	0.2532	0.2311	0.0310	0.1783	0.1472	0.0214
5	2.7065	2.8768	0.3097	0.9056	0.8810	0.0739	0.3904	0.3193	0.0294	0.2724	0.2248	0.0222
(xii) Misspecification Model												
1.07	0.2351	0.2412	0.1923	0.0788	0.0762	0.0570	0.0339	0.0299	0.0258	0.0234	0.0213	0.0197

Table 2. Relative efficiencies of the estimators with respect to $\bar{y}_0$

	n=10		*n*=30		*n*=70		*n*=100	
d	$E_{0,1}$	$E_{0,p}$	$E_{0,1}$	$E_{0,p}$	$E_{0,1}$	$E_{0,p}$	$E_{0,1}$	$E_{0,p}$
(i) True Model								
-	0.857	1.133	1.042	1.303	1.100	1.257	1.084	1.458
(ii) Dixon's outlier model-I								
3	1.013	1.956	1.041	2.005	1.165	2.051	1.232	2.050
4	1.028	2.652	0.959	2.473	1.168	2.809	1.316	2.965
5	1.027	3.520	1.077	3.822	1.083	3.486	1.318	3.907
(iii) Dixon's outlier model-II								
3	0.928	3.466	1.040	4.090	1.108	4.101	1.210	4.103
4	1.028	5.903	1.053	6.624	1.127	6.547	1.199	6.454
5	0.940	7.851	1.112	10.163	1.155	9.718	1.274	10.284
(iv) Tatum's, (1997) localized scale disturbances model-I using φ =0.10								
3	0.963	1.873	1.122	1.122	1.207	2.152	1.263	2.086
4	0.823	2.192	1.017	1.017	1.282	3.068	1.212	2.761
5	0.949	3.291	1.051	1.051	1.143	3.706	1.340	3.937
(v) Tatum's, (1997) localized scale disturbances model-II using φ =0.10								
3	0.954	3.554	1.035	1.035	1.193	4.317	1.266	4.219
4	0.942	5.351	1.033	1.033	1.062	6.132	1.247	6.733
5	0.813	6.888	0.966	0.966	1.162	9.825	1.284	10.230
(vi) Tatum's, (1997) localized scale disturbances model-I using φ =0.15								
3	0.975	2.128	1.028	1.028	1.145	2.346	1.239	2.374
4	0.877	2.606	1.043	1.043	1.242	3.584	1.151	3.023
5	0.885	3.538	1.078	1.078	1.168	4.721	1.166	4.459
(vii) Tatum's, (1997) localized scale disturbances model-II using φ =0.15								
3	0.825	3.453	1.033	1.033	1.222	5.502	1.106	4.546
4	0.923	5.947	1.023	1.023	1.096	8.168	1.211	8.332
5	0.941	8.739	1.028	1.028	1.223	13.279	1.212	12.270
(viii) Amiri and Allahyari's (2012) single step shift in the scale model-I using φ =0.10								
3	0.988	1.916	1.025	1.025	1.201	2.143	1.333	2.165
4	0.891	2.369	1.046	1.046	1.205	2.864	1.232	2.784
5	0.925	3.204	1.016	1.016	1.149	3.680	1.322	3.965
(ix) Amiri and Allahyari's (2012) single step shift in the scale model-II using φ =0.10								
3	0.942	3.553	0.966	0.966	1.155	4.135	1.232	4.149
4	0.946	5.452	0.973	0.973	1.091	6.392	1.115	5.897
5	0.903	7.581	1.044	1.044	1.059	10.461	1.058	8.215
(x) Amiri and Allahyari's (2012) single step shift in the scale model-I using φ =0.15								
3	0.964	2.101	1.059	1.059	1.256	5.635	1.295	2.489
4	0.867	2.586	1.084	1.084	1.149	8.482	1.308	3.484
5	0.923	3.690	1.105	1.105	1.014	11.135	1.240	4.675
(xi) Amiri and Allahyari's (2012) single step shift in the scale model-II using φ =0.15								
3	0.825	3.453	1.033	1.033	1.222	5.502	1.106	4.546
4	0.923	5.947	1.023	1.023	1.096	8.168	1.211	8.332
5	0.941	8.739	1.028	1.028	1.223	13.279	1.212	12.270
(xii) Misspecification Model								
1.07	0.975	1.222	1.034	1.381	1.135	1.316	1.102	1.190

Table 3. Tests of normality for residuals

	p-value		
Test	**Case-I**	**Case-II**	**Case-III**
Shapiro Wilk	0.0008	0.0240	0.0003
Anderson Darling	0.0002	0.0340	0.0003
K. Smirnov	0.0274	0.0009	0.0003

Table 4. Goodness of fit tests for Laplace distribution to assess residuals

Test	**Case-I**		**Case-II**		**Case-III**	
	Statistic	**C. R**	**Statistic**	**C. R**	**Statistic**	**C. R**
Anderson Darling	0.5561	0.9605	0.4423	0.930	0.4131	0.930
Cramer-von Mises	0.0697	0.1370	0.0711	0.131	0.0418	0.131
Watson	0.0629	0.0805	0.0705	0.080	0.0412	0.081

Table 5. Computational results of for estimating average Covid-19 cases and deaths

	Case-I	**Case-II**	**Case-III**
N	946	2662	3116
n	17	20	20
ρ_{zy}	0.91	0.81	0.93
R	0.0104	1.7894	1.8695
$\text{var}(\bar{z})$	353.5508	0.0312	155.2261
$\text{var}(\bar{y})$	0.0273	0.0515	361.4136
$\text{cov}(\bar{y},\bar{z})$	2.8488	0.0323	220.6684
R_p	0.0066	1.1065	1.5735
$\text{var}(\hat{\mu}_y^*)$	0.0249	0.0495	347.2143
$\text{cov}(\hat{\mu}_y^*,\bar{z})$	2.5956	0.0310	211.9987
$\bar{y}_1$	0.4296	0.3652	43.8501
$\bar{y}_p$	0.3601	0.1824	15.1357
$MSE(\bar{y}_1)$	0.0062	0.0358	78.8666
$MSE(\bar{y}_p)$	0.0059	0.0190	64.3833
$E_{0,1}$	4.3786	1.4372	4.5825
$E_{0,p}$	4.6264	2.7072	5.6135