ST. XAVIER'S COLLEGE (AUTONOMOUS), KOLKATA

DEPARTMENT OF STATISTICS

Name : **ANIS PAKRASHI**

Roll No. : **406** **Semester : 6** **Year : 2021**

CIN No. : **18-300-4-07-0406**

Regn. No. : **A01-1112-0765-18**

* * * * * * * * * * *

Supervisor: **PROF. MADHURA DASGUPTA**

TITLE

EFFICIENT MEAN ESTIMATION IN LOGNORMAL LINEAR MODELS

I affirm that I have identified all my sources and that no part of my dissertation paper uses unacknowledged materials.

CONTENTS :

Serial Number	*Topic*	*Page Number*
1.	Introduction	2
2.	Objective	4
3.	Results related to Linear Models	5
4.	Estimators for Lognormal Mean	6
	i. Four Existing Estimators	9
	ii. Two Proposed Estimators	14
5.	Simulation Study	20
	i. Simulation Algorithm	21
	ii. ‘Sample Size’ wise comparison	23
	iii. ‘Estimator’ wise comparison	29
6.	Conclusion	36
7.	Future Scope	37
8.	References	37
9.	Acknowledgement	38

****

INTRODUCTION:

Statisticians seem to be the happiest when they have a “Normal” Distribution at their disposal! Let us deviate slightly and consider a case where a simple transformation can provide us with the celebrated ‘normality’ – the Lognormal Distribution. A random variable **X** is said to follow a Lognormal Distribution is **ln X** follows a Normal Distribution. More specifically, a random variable **X** is said to follow a Lognormal distribution if its probability density function is given by

$$f(x) = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} & , 0 < x < \infty \\ 0 & , \text{otherwise} \end{cases}$$

where, $\mu \in \mathbb{R}$ and $\sigma > 0$.

Lognormality has been prevalent and applied in varied fields like –

- Mining (Marcotte and Groleau, 1997)
- estimating insurance reserves (Doray, 1996)
- water quality control (Gilliom and Helsel, 1986)
- monitoring air pollution concentrations (Holland et al., 2000)
- estimating the sediment discharge (Cohn, 1995) , and so on and so forth.

These applications involve the concept of lognormal linear models, where regression models are fitted to logarithmic transformed response variables.

Let $\underset{\sim}{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$ be the lognormal response vector .

Let $\underset{\sim}{x}_i = \begin{pmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix}$ be the covariate vector for observation i.

Thus, a lognormal linear model can be represented (in Gauss Markov form) as below:

$$\underset{\sim}{Y} = \ln\left(\underset{\sim}{Z}\right) = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \quad \text{………… (1)}$$

where, $\underset{\sim}{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ is the design matrix, $\underset{\sim}{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$ is the parameter vector and $\underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{pmatrix}$ is the vector of errors of the model. We assume $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0,\sigma^2)$.

We are interested in predicting the response variable for a new set of covariate value $\underset{\sim}{x}_0$, that is, predicting $Z_0 = e^{\left(\underset{\sim}{x}_0^T \underset{\sim}{\beta} + \varepsilon_0\right)}$ where ε_0 is the corresponding normal error with mean zero and variance σ^2. If $\underset{\sim}{\beta}$ and σ^2 are known, the best predictor is the conditional expectation, given by,

$$\mu(x_0) = E(Z \mid x_0) = e^{\left(\underset{\sim}{x}_0^T \underset{\sim}{\beta} + \frac{\sigma^2}{2}\right)} \quad \text{…………(2)}$$

Here $\mu(x_0)$ is a function of the unknown parameters β and σ^2 , which have to be estimated. In this paper, the problem of efficient estimation of $\mu(x_0)$ is considered.

A common measure of the quality of an estimator in the statistical world is the Mean Squared Error (MSE) defined as

$$MSE[\hat{\mu}(x_0)] = E[\hat{\mu}(x_0)-\mu(x_0)]^2 = Var[\hat{\mu}(x_0)] + Bias^2[\hat{\mu}(x_0)] \quad \text{……….(3)}$$

where $Bias[\hat{\mu}(x_0)] = E[\hat{\mu}(x_0)] - \mu(x_0)$ is the bias of the estimator $\hat{\mu}\left(x_0\right)$. In terms of MSE, the UMVU estimator is the best estimator among all unbiased estimators of $\mu\left(x_0\right)$. However, it is possible to find better estimators in terms of MSE if one can tolerate a small bias.

OBJECTIVE:

Lognormal Linear Models have wide applications and often it finds use in predicting the response variable at the original scale for a new set of covariates. Here, we take up the problem of efficiently estimating the conditional mean of the response variable at the original scale, for lognormal linear models. We review and state several existing estimators, including :

- Maximum Likelihood Estimator
- Restricted Maximum Likelihood Estimator
- Uniformly Minimum Variance Unbiased Estimator
- Bias-corrected Restricted Maximum Likelihood Estimator

Now, we propose two estimators that aim at minimizing the asymptotic mean square error and asymptotic bias respectively. This gives rise to two more estimators :

- Minimum MSE Estimator
- Minimum Bias Estimator

Simulation studies are used to compare the estimators. This reveals the superiority of one estimator over others in terms of MSE or bias.

Hence, our primary objective is to make graphical comparisons between estimators on the basis of the performance of estimates with respect to true lognormal mean and the worth of different estimators with respect to the Mean Square Error and Bias in estimation.

RESULTS RELATED TO LINEAR MODELS

Firstly, we state the several well-known results about the distributions for the OLS (Ordinary Least Square) estimators for β and the Residual Sum of Squares (RSS).

Result 1:-

Gauss – Markov Theorem : In the model $\left(\underset{\sim}{Y}, X\underset{\sim}{\beta}, \sigma^2 I\right)$, the BLUE of an estimable LPF $\underset{\sim}{\lambda}^T\underset{\sim}{\beta}$, $\underset{\sim}{\lambda} \in R^p$ is $\underset{\sim}{\lambda}^T\hat{\underset{\sim}{\beta}}$, where $\hat{\underset{\sim}{\beta}}$ is the solution of the equations $X^T X\underset{\sim}{\beta} = X^T\underset{\sim}{Y}$, obtained by minimizing $\left(\underset{\sim}{Y} - X\underset{\sim}{\beta}\right)^T\left(\underset{\sim}{Y} - X\underset{\sim}{\beta}\right)$ with respect to $\underset{\sim}{\beta}$.

Result 2:-

The OLS estimator for β is

$$\hat{\underset{\sim}{\beta}} = (X^T X)^{-1} X^T \underset{\sim}{Y} \sim N(\underset{\sim}{\beta}, \sigma^2 (X^T X)^{-1})$$

As a result, $\underset{\sim}{x}_0^T\hat{\underset{\sim}{\beta}} \sim N(\underset{\sim}{x}_0^T\underset{\sim}{\beta}, \sigma^2 v_0)$ where $v_0 = \underset{\sim}{x}_0^T (X^T X)^{-1} \underset{\sim}{x}_0$.

Result 3:-

Let m = n - (p+1). The Residual Sum of Squares is

$$\text{RSS} = Y^T[I - X (X^T X)^{-1} X^T]Y \sim \sigma^2\chi^2_m.$$

Thus, the MGF of RSS is

$$E(e^{t.RSS}) = E(e^{t\sigma^2\chi^2_m}) = \frac{1}{(1\text{-}2t\sigma^2)^{\frac{m}{2}}}, \text{ for } t < \frac{1}{2\sigma^2}.$$

ESTIMATORS FOR LOGNORMAL MEAN:

[***The naïve back-transform estimator*** : -

The UMVUE of log(Z_0) is $x_0^T \hat{\beta}$. Thus, it may seem apparently reasonable to estimate $\mu\left(\underset{\sim}{x}_0\right)$ by $\hat{\mu}_{BT}\left(\underset{\sim}{x}_0\right) = e^{\underset{\sim}{x}_0^T \hat{\underset{\sim}{\beta}}}$. This estimator is called the back-transform (BT) estimator. However, by comparing it to eqn. (2), we can show that this estimator is not consistent. Even for large sample sizes, there exists an asymptotic multiplicative bias of $e^{-\sigma^2/2}$, which is less than unity. Thus, there is a subtle underestimation of the value of $\mu\left(\underset{\sim}{x}_0\right)$ with a large bias for a high value of σ^2. Thus, we do not include this estimator in our comparison.

To check the effectiveness of the back-transform estimator, we perform a small simulation as given below:

Step 1: Start the main function , taking parameters n, sigma_sq, and covariate value x_0.

1) Define a two-component vector x0 with elements 1 and x_0.
2) Define the true parameter vector as (1,1)' and the true mean.
3) Form the design matrix of order (n x 2) by taking first column as all 1 and the second column as a random sample from uniform(0,1) distribution.
4) Start a replication loop to simulate 100 times.
5) Obtain the response vector Y and the OLS estimates as beta_hat.
6) Obtain the back-transform estimate.
7) End the replication loop and export the estimate to the function call.

Step 2: We set the seed to 12345678.

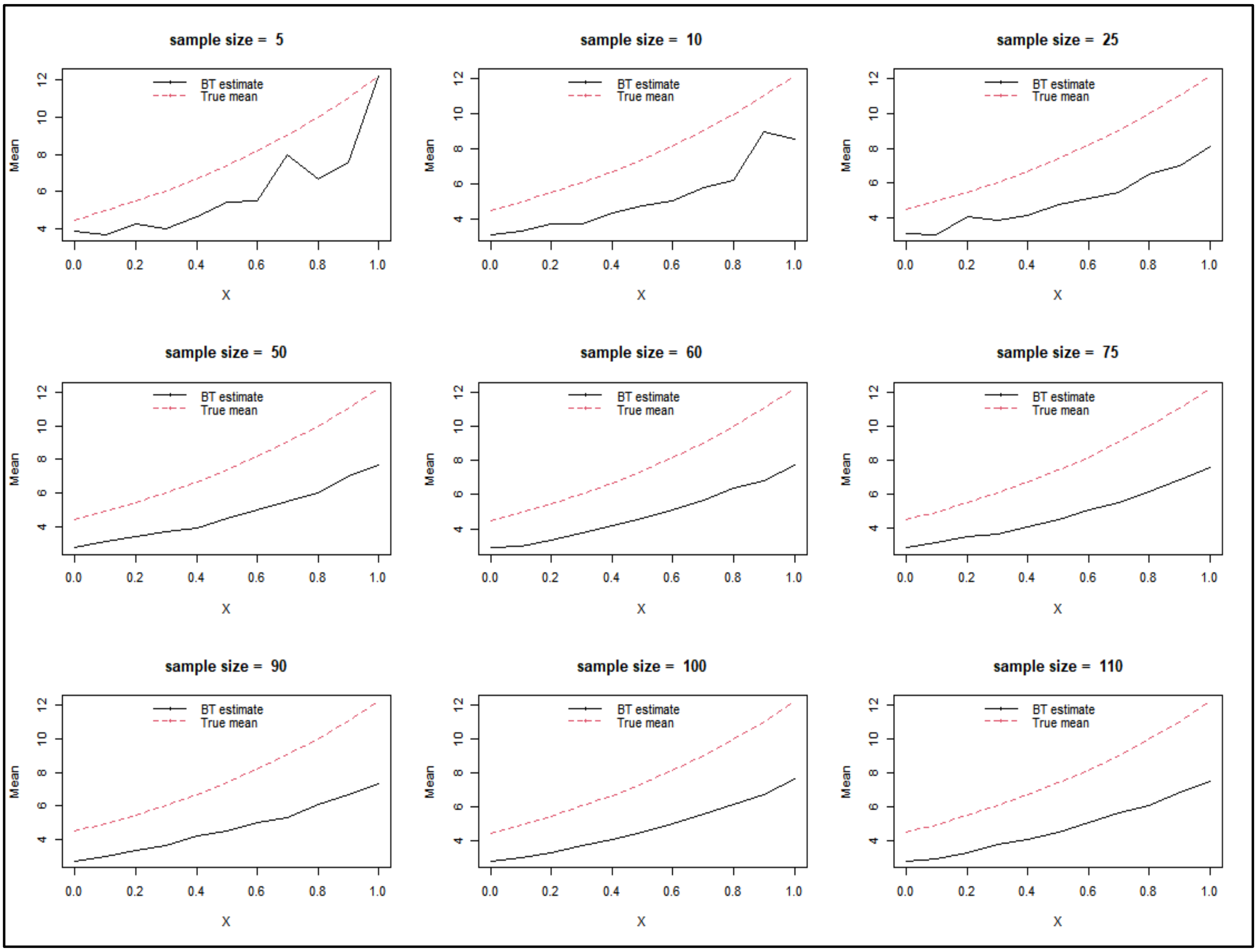
Step 3: The value of the error variance is initialized to 1.

Step 4: Initialize the vector containing the values of sample sizes : 5, 10, 25, 50, 60, 75, 90, 100, 110 and store the length of the vector. Store the covariate values in the vector x and store its length.

Step 5: Compare the back-transform estimates with the true mean. Plot the estimates along with the true mean for different sample sizes as mentioned above.

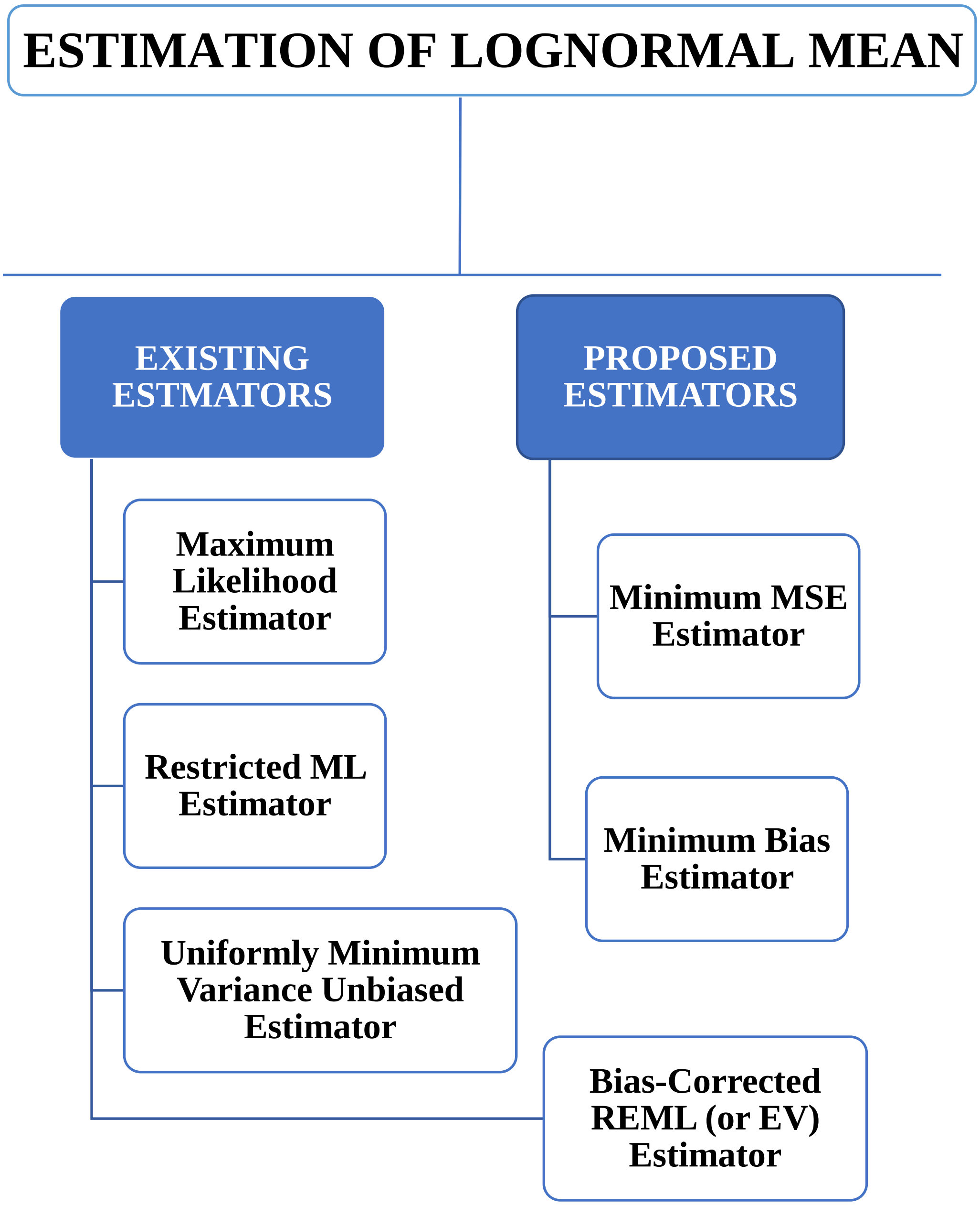
Consider the *Figure 1* below :

FIGURE 1



OBVERVATIONS:

From the figure, we observe that the back-transform estimator fluctuates highly for different values of x, for smaller samples, while still underestimating the true mean. With large samples, the random nature is subdued. However, even for large samples, the estimator largely underestimates the true mean value. Hence, it is not worthwhile to include such an estimator in our study.]



FOUR EXISTING ESTIMATORS

First, we put forward 4 existing estimators ('existing' as proposed in the paper by Haipeng Shen and Zhengyuan Zhu, referred to in **References**).

Estimator 1:- **THE MAXIMUM LIKELIHOOD ESTIMATOR (or ML Estimator)**

Note that the ML estimator for $\underset{\sim}{\beta}$ is same as the OLS estimator. Moreover, the ML estimator of σ^2 is $\hat{\sigma}^2_{ML} = \frac{RSS}{n}$. Since the MLE of a parametric function is the function of the MLE of the parameters, the ML estimator of $\mu\left(\underset{\sim}{x}_0\right)$ is given by

$$\hat{\mu}_{ML}\left(\underset{\sim}{x}_0\right) = e^{\left(\underset{\sim}{x}_0^T\hat{\underset{\sim}{\beta}} + \frac{\hat{\sigma}^2_{ML}}{2}\right)}$$

……….(4)

Now, using the moment generating functions of normal and chi-square variates, we obtain the MSE and Bias of the MLE. Here, we have

$$\left.\begin{aligned} E\left(e^{2\underset{\sim}{x}_0^T\hat{\underset{\sim}{\beta}}}\right) &= e^{2\underset{\sim}{x}_0^T\underset{\sim}{\beta} + 2\sigma^2 v_0} \\ E\left(e^{\underset{\sim}{x}_0^T\hat{\underset{\sim}{\beta}}}\right) &= e^{\underset{\sim}{x}_0^T\underset{\sim}{\beta} + \frac{\sigma^2 v_0}{2}} \end{aligned}\right\} \text{From MGF of Normal variate}$$

$$\left.\begin{aligned} E\left(e^{\frac{\hat{\sigma}^2_{ML}}{2}}\right) &= (1-\sigma^2/n)^{-m/2} \\ E\left(e^{\hat{\sigma}^2_{ML}}\right) &= (1-2\sigma^2/n)^{-m/2} \end{aligned}\right\} \text{From MGF of } \chi^2 \text{ variate}$$

[from Results 2 and 3]

Also, we have,

$$\text{MSE} = E\left[e^{\left(\underset{\sim}{x}_0^T \hat{\underset{\sim}{\beta}} + \frac{\hat{\sigma}_{ML}^2}{2}\right)} - e^{\left(\underset{\sim}{x}_0^T \underset{\sim}{\beta} + \frac{\sigma^2}{2}\right)} \right]^2$$

and

$$\text{Bias} = E\left[e^{\left(\underset{\sim}{x}_0^T \hat{\underset{\sim}{\beta}} + \frac{\hat{\sigma}_{ML}^2}{2}\right)} \right] - e^{\left(\underset{\sim}{x}_0^T \underset{\sim}{\beta} + \frac{\sigma^2}{2}\right)}$$

Therefore, using the above relations, the MSE and Bias are given by

$$\text{MSE}\left[\hat{\mu}_{ML}\left(\underset{\sim}{x}_0\right)\right] = \mu^2\left(\underset{\sim}{x}_0\right)\left[e^{(2v_0-1)\sigma^2} (1 - 2\sigma^2/n)^{-m/2} - 2e^{(v_0-1)\sigma^2/2} (1 - \sigma^2/n)^{-m/2} + 1 \right]$$

……(5)

and

$$\text{Bias}\left[\hat{\mu}_{ML}\left(\underset{\sim}{x}_0\right)\right] = \mu\left(\underset{\sim}{x}_0\right)\left[e^{(v_0-1)\sigma^2/2} (1 - \sigma^2/n)^{-m/2} - 1 \right] \qquad ……(6)$$

<u>Estimator 2:-</u> THE RESTRICTED MAXIMUM LIKELIHOOD ESTIMATOR (or REML Estimator)

This is basically an RSS- based degree of freedom adjusted estimator, similar to a maximum likelihood estimator. Here too, the REML estimator for $\underset{\sim}{\beta}$ is same as the OLS estimator. However, the estimator of the error variance involves the degrees of freedom of the RSS and not all the total number of observations. Thus, the REML estimator of σ^2 is $\hat{\sigma}_{REML}^2 = RSS/m$, which is, in fact, same as the OLS estimator. The REML estimator of $\mu\left(\underset{\sim}{x}_0\right)$ is given by

$$\hat{\mu}_{REML}\left(\underset{\sim}{x}_0\right) = e^{\left(\underset{\sim}{x}_0^T \underset{\sim}{\hat{\beta}} + \frac{\hat{\sigma}^2_{REML}}{2}\right)}$$

…….(7)

The MSE and Bias are given by

$$MSE\left[\hat{\mu}_{REML}\left(\underset{\sim}{x}_0\right)\right] = \mu^2\left(\underset{\sim}{x}_0\right)\left[e^{(2v_0-1)\sigma^2}\left(1-2\sigma^2/m\right)^{-m/2} - 2e^{(v_0-1)\sigma^2/2}\left(1-\sigma^2/m\right)^{-m/2} + 1\right]$$

……..(8)

and

$$Bias\left[\hat{\mu}_{REML}\left(\underset{\sim}{x}_0\right)\right] = \mu\left(\underset{\sim}{x}_0\right)\left[e^{(v_0-1)\sigma^2/2}\left(1-\sigma^2/m\right)^{-m/2} - 1\right] \qquad …….(9)$$

__Estimator 3:-__ THE UNIFORMLY MINIMUM VARIANCE UNBIASED ESTIMATOR (or UMVU Estimator)

Finney (1941) developed the UMVU estimator of parameters of the lognormal distribution and Heien (1968) extended Finney’s method to simple lognormal regression models. Bradu and Mundlak (1970) derived the UMVU estimator and its variance for generalized lognormal regression models. The estimator of $\mu\left(\underset{\sim}{x}_0\right)$ was derived using the Lehmann-Scheffe Theorem.

*[***Lehmann-Scheffe Theorem :**

The Lehmann-Scheffe Theorem states that if T is a complete sufficient statistic for {f(x,θ),θ∈Θ} and if h(T) is unbiased for g(θ), then h(T) is the UMVUE of g(θ)].

Here $\hat{\underset{\sim}{\beta}}$ and $\hat{\sigma}^2_{REML}$ are complete sufficient statistics for β and σ^2; hence any unbiased function of these statistics is the UMVU estimator of $\mu\left(\underset{\sim}{x}_0\right)$. To obtain the UMVUE of $\mu\left(\underset{\sim}{x}_0\right)$, one needs to find f(x) subject to

$$E[e^{\underset{\sim}{x}_0^T\hat{\underset{\sim}{\beta}}} f(\hat{\sigma}^2_{REML})] = e^{\left(\underset{\sim}{x}_0^T\underset{\sim}{\beta} + \sigma^2/2\right)}$$

which leads to

$$f(\hat{\sigma}^2_{REML}) = \sum_{i=0}^{\infty} \frac{\Gamma(m/2)}{i!\Gamma(m/2+i)}\left[\frac{m(1-v_0)}{4}\hat{\sigma}^2_{REML}\right]^i$$

$$= {}_0F_1\left(\frac{m}{2};\frac{m(1-v_0)}{4}\hat{\sigma}^2_{REML}\right).$$

where, ${}_0F_1(\alpha;z)$ is the Hypergeometric function.

The UMVUE and its variance are given by

$$\hat{\mu}_{UMVU}\left(\underset{\sim}{x}_0\right) = e^{\underset{\sim}{x}_0^T\hat{\underset{\sim}{\beta}}}\,{}_0F_1\left(\frac{m}{2};\frac{m(1-v_0)}{4}\hat{\sigma}^2_{REML}\right)$$

…….(10)

and

$$Var\left[\hat{\mu}_{UMVU}\left(\underset{\sim}{x}_0\right)\right] = \mu^2\left(\underset{\sim}{x}_0\right)\left[e^{v_0\sigma^2}\,{}_0F_1\left(\frac{m}{2};\frac{(1-v_0)}{4}\sigma^4\right)-1\right]$$

…….(11)

Note that the UMVUE estimator is an unbiased estimator, that is, the bias is zero. Thus, the MSE is same as the variance. It is also to be noted that the UMVU estimator has the smallest MSE among all unbiased estimators.

Estimator 4:- **THE BIAS-CORRECTED RESTRICTED MAXIMUM LIKELIHOOD ESTIMATOR (or EV Estimator)**

In general, the ML or REML estimators are not unbiased. El-shaarawi and Viveros (1997) proposed to correct this bias using the leading terms of the Taylor's expansion of $\text{Bias}\left[\hat{\mu}_{REML}\left(\underset{\sim}{x}_0\right)\right]$ with respect to σ^2/m, which gives us the "bias-corrected REML estimator". To give justice to the fact that the two mathematicians proposed the bias correction, we henceforth refer to this estimator as the EV estimator. The estimator is as given below:

$$\hat{\mu}_{EV}\left(\underset{\sim}{x}_0\right) = e^{\left[\underset{\sim}{x}_0^T\underset{\sim}{\beta} + \frac{(1-v_0)\hat{\sigma}^2_{REML}}{2} - \frac{\hat{\sigma}^4_{REML}}{4m} - \frac{\hat{\sigma}^6_{REML}}{6m}\right]} \quad \ldots\ldots(12)$$

Define

$$f_{EV}(x) = e^{\left[\frac{(1-v_0)x}{2} - \frac{x^2}{4m} - \frac{x^3}{6m}\right]}$$

The MSE and Bias of this estimator are given by:

$$\text{MSE}\left[\hat{\mu}_{EV}\left(\underset{\sim}{x}_0\right)\right] = \mu^2\left(\underset{\sim}{x}_0\right)\left[e^{(2v_0-1)\sigma^2}E\left[f^2_{EV}\left(\hat{\sigma}^2_{REML}\right)\right] - 2e^{(v_0-1)\sigma^2/2}E\left[f_{EV}\left(\hat{\sigma}^2_{REML}\right)\right] + 1\right]$$

…………….(13)

and

$$\text{Bias}\left[\hat{\mu}_{EV}\left(\underset{\sim}{x}_0\right)\right] = \mu\left(\underset{\sim}{x}_0\right)\left[e^{(v_0-1)\sigma^2/2}E\left[f_{EV}\left(\hat{\sigma}^2_{REML}\right)\right] - 1\right]$$

…………….(14)

The expectations in the above expressions need to be evaluated using numerical integration methods.

TWO PROPOSED ESTIMATORS:

Now, we put forward two new estimators, by minimizing the MSE and Bias respectively.

<u>Proposed Estimator 1:</u> **MINIMUM MSE ESTIMATOR (or MM Estimator)**

This estimator has been formed by minimizing the MSE and is given by

$$\hat{\mu}_{MM}\left(\underset{\sim}{x}_0\right) = e^{\left(\underset{\sim}{x}_0^T \underset{\sim}{\hat{\beta}} + \frac{m.RSS}{2(n-p+1+3nv_0)m+3.RSS}\right)} \quad \ldots\ldots\ldots\ldots(15)$$

where, m=n - (p+1).

Define

$$f_{MM}(RSS) = e^{\left[\frac{m.RSS}{2(n-p+1+3nv_0)m+3.RSS}\right]}$$

The MSE and Bias of this estimator are given by:

$$MSE\left[\hat{\mu}_{MM}\left(\underset{\sim}{x}_0\right)\right] = \mu^2\left(\underset{\sim}{x}_0\right)\left[e^{(2v_0-1)\sigma^2} E\left[f_{MM}^2\left(RSS\right)\right] - 2e^{(v_0-1)\sigma^2/2} E\left[f_{MM}\left(RSS\right)\right] + 1\right]$$

$$\ldots\ldots\ldots\ldots\ldots(16)$$

and

$$Bias\left[\hat{\mu}_{MM}\left(\underset{\sim}{x}_0\right)\right] = \mu\left(\underset{\sim}{x}_0\right)\left[e^{(v_0-1)\sigma^2/2} E\left[f_{MM}\left(RSS\right)\right] - 1\right]$$

$$\ldots\ldots\ldots\ldots\ldots(17)$$

Proposed Estimator 2: **MINIMUM BIAS ESTIMATOR (or MB Estimator)**

This estimator has been formed by minimizing the bias and is given by

$$\hat{\mu}_{MB}\left(\underset{\sim}{x}_0\right) = e^{\left(\underset{\sim}{x}_0^T\underset{\sim}{\hat{\beta}} + \frac{m.RSS}{2(n-p-1+nv_0)m+RSS}\right)} \quad \ldots\ldots\ldots\ldots\ldots(18)$$

where, m=n - (p+1).

Define

$$f_{MB}(RSS) = e^{\left[\frac{m.RSS}{2(n-p-1+nv_0)m+RSS}\right]}$$

The MSE and Bias of this estimator are given by:

$$MSE\left[\hat{\mu}_{MB}\left(\underset{\sim}{x}_0\right)\right] = \mu^2\left(\underset{\sim}{x}_0\right)\left[e^{(2v_0-1)\sigma^2}E\left[f_{MB}^2\left(RSS\right)\right] - 2e^{(v_0-1)\sigma^2/2}E\left[f_{MB}\left(RSS\right)\right]+1\right]$$

$$\ldots\ldots\ldots\ldots\ldots\ldots(19)$$

and

$$Bias\left[\hat{\mu}_{MB}\left(\underset{\sim}{x}_0\right)\right] = \mu\left(\underset{\sim}{x}_0\right)\left[e^{(v_0-1)\sigma^2/2}E\left[f_{MB}\left(RSS\right)\right]-1\right]$$

$$\ldots\ldots\ldots\ldots\ldots\ldots..(20)$$

[**Note that**, in case of the above estimators, the MSE and Bias have such forms which cannot be evaluated algebraically. We require numerical integration to obtain their values.]

DERIVATION of the Proposed Estimators:

Consider the class of estimators,

$$\left\{\hat{\mu}_c(x_0):\hat{\mu}_c(x_0)=e^{\underset{\sim}{x}_0^T\underset{\sim}{\hat{\beta}}+cRSS/2},c=1/_{n-a},a<n\right\} \quad \text{.........(21)}$$

The above can be described as the class which is derived from eqn. (2) by plugging-in estimators $\hat{\beta}$ and c.RSS = $\frac{RSS}{n-a}$ in place of β and σ^2, respectively. Our goal is to find estimators from this class that can asymptotically minimize the MSE or the bias and have better or comparable performances as the existing estimators.

I. Let c < $1/_{2\sigma^2}$. Using similar steps as we had used to derive the MSE and Bias of ML estimator, in eqn. (5) and (6), we can derive the following general results for the class of estimators in eqn. (21):

$$\text{MSE}\left[\hat{\mu}_c\left(\underset{\sim}{x}_0\right)\right]=\mu^2\left(\underset{\sim}{x}_0\right)\left[e^{(2v_0-1)\sigma^2}(1-2c\sigma^2)^{-m/2}-2e^{(v_0-1)\sigma^2/2}(1-c\sigma^2)^{-m/2}+1\right]$$

......(22)

and

$$\text{Bias}\left[\hat{\mu}_c\left(\underset{\sim}{x}_0\right)\right]=\mu\left(\underset{\sim}{x}_0\right)\left[e^{(v_0-1)\sigma^2/2}(1-c\sigma^2)^{-m/2}-1\right]$$

......(23)

II. Let $c=\frac{1}{n}+\frac{a}{n^2}+\frac{b}{n^3}+o\left(\frac{1}{n^3}\right)$. Consider the following Taylor series expansion,

$$\log(1-t)=-\sum_{i=1}^{\infty}\frac{t^i}{i}$$

Define $V_1=e^{(2v_0-1)\sigma^2}(1\text{-}2c\sigma^2)^{-m/2}$ and $V_2=e^{(v_0-1)\sigma^2/2}(1\text{-}c\sigma^2)^{-m/2}$.

Expanding V_1 and V_2 using the above Taylor Series :

$$V_1 = e^{(2v_0-1)\sigma^2}\left(1-2c\sigma^2\right)^{-m/2}$$

$$= e^{(2v_0-1)\sigma^2 - \frac{m}{2}\ln(1-2c\sigma^2)}$$

$$= e^{(2v_0-1)\sigma^2 + \frac{m}{2}\left(2c\sigma^2 + 2c^2\sigma^4 + \frac{8}{3}c^3\sigma^6 + o\left(\frac{1}{n^3}\right)\right)}$$

$$= e^{\frac{2n^2v_0 + n(a-p-1) + b - a(p+1)}{n^2}\sigma^2 + \left(\frac{1}{n} + \frac{2a-p-1}{n^2}\right)\sigma^4 + \frac{4}{3n^2}\sigma^6 + o\left(\frac{1}{n^2}\right)}$$

$$= 1 + \left[2v_0 + \frac{a-p-1}{n} + \frac{b-a(p+1)}{n^2}\right]\sigma^2 + \left(\frac{1}{n} + \frac{2a-p-1}{n^2}\right)\sigma^4 + \frac{4}{3n^2}\sigma^6$$

$$+ \frac{\left(2nv_0 + a - p - 1\right)^2}{2n^2}\sigma^4 + \frac{1}{2n^2}\sigma^8 + \frac{2nv_0 + a - p - 1}{n^2}\sigma^6 + o\left(\frac{1}{n^2}\right)$$

$$= 1 + \left(2nv_0 + a - p - 1 + \sigma^2\right)\frac{\sigma^2}{n} + \left[b - a(p+1)\right]\frac{\sigma^2}{n^2}$$

$$+ \left[2a - p - 1 + 2n^2v_0^2 + 2nv_0(a-p-1) + \tfrac{1}{2}(a-p-1)^2\right]\frac{\sigma^4}{n^2}$$

$$+ \left(\frac{4}{3} + 2nv_0 + a - p - 1\right)\frac{\sigma^6}{n^2} + \frac{\sigma^8}{2n^2} + o\left(\frac{1}{n^2}\right)$$

$$V_2 = e^{\frac{1}{2}(v_0-1)\sigma^2}\left(1-c\sigma^2\right)^{-m/2}$$

$$= e^{\frac{1}{2}(v_0-1)\sigma^2 - \frac{m}{2}\ln(1-c\sigma^2)}$$

$$= e^{\frac{1}{2}(v_0-1)\sigma^2 + \frac{m}{2}\left(c\sigma^2 + \frac{c^2\sigma^4}{2} + \frac{c^3\sigma^6}{3} + o\left(\frac{1}{n^3}\right)\right)}$$

$$= e^{\frac{n^2v_0 + n(a-p-1) + b - a(p+1)}{2n^2}\sigma^2 + \left(\frac{1}{4n} + \frac{2a-p-1}{4n^2}\right)\sigma^4 + \frac{1}{6n^2}\sigma^6 + o\left(\frac{1}{n^2}\right)}$$

$$= 1 + \left[\frac{v_0}{2} + \frac{a-p-1}{2n} + \frac{b-a(p+1)}{2n^2}\right]\sigma^2 + \left(\frac{1}{4n} + \frac{2a-p-1}{4n^2}\right)\sigma^4 + \frac{1}{6n^2}\sigma^6$$

$$+ \frac{\left(nv_0 + a - p - 1\right)^2}{8n^2}\sigma^4 + \frac{1}{32n^2}\sigma^8 + \frac{nv_0 + a - p - 1}{8n^2}\sigma^6 + o\left(\frac{1}{n^2}\right)$$

$$= 1 + \left(nv_0 + a - p - 1 + \frac{\sigma^2}{2} \right) \frac{\sigma^2}{2n} + \left[b - a\left(p+1\right) \right] \frac{\sigma^2}{2n^2}$$

$$+ \left[a - \frac{p+1}{2} + \frac{n^2 v_0^2}{4} + \frac{nv_0\left(a-p-1\right)}{2} + \frac{\left(a-p-1\right)^2}{4} \right] \frac{\sigma^4}{2n^2}$$

$$+ \left(\frac{1}{3} + \frac{nv_0}{4} + \frac{a-p-1}{4} \right) \frac{\sigma^6}{2n^2} + \frac{\sigma^8}{32n^2} + o\left(\frac{1}{n^2}\right)$$

Now, according to eqn. (22), we know that

$$\mathrm{MSE}\left[\hat{\mu}_c\left(\underset{\sim}{x}_0\right)\right] = \mu^2\left(\underset{\sim}{x}_0\right)\left(V_1 - 2V_2 + 1\right)$$

Thus, incorporating the above expressions for V_1 and V_2, we obtain

$$\mathrm{MSE}\left[\hat{\mu}_c\left(\underset{\sim}{x}_0\right)\right]$$

$$= \mu^2\left(\underset{\sim}{x}_0\right) \frac{\sigma^2}{n} \left\{ 1 + \frac{\sigma^2}{2} + \frac{\sigma^2}{4n}\left[a^2 + \left(2 - 2p + 6nv_0 + 3\sigma^2\right)a + f\left(p, n, \sigma^2, v_0\right) \right] \right\} + o\left(\frac{1}{n^2}\right)$$

………(24)

where, $f\left(p, n, \sigma^2, v_0\right) = -1 + p^2 - 6nv_0\left(p+1\right) + 7n^2v_0^2 + \left(1 - 3p + 7nv_0\right)\sigma^2 + {7\sigma^2}/{4}$

Now, eqn. (23) gives

$$\mathrm{Bias}\left[\hat{\mu}_c\left(\underset{\sim}{x}_0\right)\right] = \mu\left(\underset{\sim}{x}_0\right)\left(V_2 - 1\right)$$

Incorporating the expression for V_2, we obtain

$$\mathrm{Bias}\left[\hat{\mu}_c\left(\underset{\sim}{x}_0\right)\right] = \mu\left(\underset{\sim}{x}_0\right) \frac{\sigma^2}{2n}\left(nv_0 + a - p - 1 + \frac{\sigma^2}{2} \right) + o\left(\frac{1}{n^2}\right) \quad ………(25)$$

Suppose we consider finding **c** that can minimize MSE up to the order of ${1}/{n^2}$. Eqn. (24) suggests that it is sufficient to find **a** to minimize $a^2 + \left(2 - 2p + 6nv_0 + 3\sigma^2\right)a$. According to the quadratic form, the minimizer depends on σ^2 and is $-\left(1-p+3nv_0+\frac{3\sigma^2}{2}\right)$. This means that the

c which minimizes the approximate MSE must be of the order of $\frac{1}{\left(n+1-p+3nv_0+\frac{3\sigma^2}{2}\right)}$.

In reality, the true value of σ^2 is usually unknown and an *adaptive* estimator is proposed by replacing σ^2 by its consistent estimator, $\hat{\sigma}^2_{REML}=\frac{RSS}{m}$. As a result, our proposed estimator:

$$\hat{\mu}_{MM}(\underset{\sim}{x}_0)=e^{\left(\underset{\sim}{x}_0^T\hat{\underset{\sim}{\beta}}+\frac{m.RSS}{2(n-p+1+3nv_0)m+3.RSS}\right)}$$

On the other hand, suppose we consider finding **c** that can reduce bias up to the order of $\frac{1}{n}$. Eqn. (25) suggests that it is sufficient to find **a** to satisfy $nv_0+a-p-1+\frac{\sigma^2}{2}=0$, which leads to $a=p+1-nv_0-\frac{\sigma^2}{2}$. This means that the **c** which minimizes the approximate bias must be of the order of $\frac{1}{\left(n-1-p+nv_0+\frac{\sigma^2}{2}\right)}$.

In reality, the true value of σ^2 is usually unknown and an *adaptive* estimator is proposed by replacing σ^2 by $\hat{\sigma}^2_{REML}=\frac{RSS}{m}$. As a result, our proposed estimator is

$$\hat{\mu}_{MB}(\underset{\sim}{x}_0)=e^{\left(\underset{\sim}{x}_0^T\hat{\underset{\sim}{\beta}}+\frac{m.RSS}{2(n-p-1+nv_0)m+RSS}\right)}$$

Likewise, we have the MSE and bias of the above estimators.

SIMULATION STUDY:

We compute the MSE and Bias of the stated and proposed estimators for different sample sizes (n).

- We assume the data modelled in eqn. (1), with only one covariate x, taking values uniformly distributed between 0 and 1.
- The regression coefficient vector $\underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ is taken to be $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.
- We consider the error variance σ^2 to take the value 1. (specific to our study)
- We assume sample size n to takes certain values over the range 1 to 110.
- We consider estimation of $\mu\left(\underset{\sim}{x}_0\right)$ for covariate values { 0, 0.1, 0.2, ……, 1 }. For our study, we consider this as the valid range of covariate values.

We consider the relative MSEs of the estimators which is nothing but the ratio between the MSEs of a given estimator and the MM estimator. Hence, by definition, the relative MSE of the MM estimator is always 1. The values for the other estimators represent their MSEs as percentages of the MSE of the MM estimator. Hence, by definition, the relative MSE of an estimator T is

$$\text{Relative MSE of T} = \frac{\text{MSE(T)}}{\text{MSE}\left(\text{T}_{\text{MM}}\right)}$$

, where $\text{MSE}\left(\text{T}_{\text{MM}}\right)$ is the MSE of the MM estimator

We also study the relative absolute bias of the estimators, defined as the absolute bias of an estimator of $\mu(x_0)$ divided by $\mu(x_0)$. Thus, it has the nice interpretation as being the bias in terms of the percentage of the parametric function to be estimated. By definition, the relative absolute bias of T is

$$\text{Relative absolute bias of T} = \frac{\left|\text{Bias}\left(\text{T}\right)\right|}{\mu\left(x_0\right)}$$

, where $\mu\left(x_0\right)$ is the lognormal mean

Now, we have performed two types of comparisons:-

- Comparing all the estimators for each of the six different sample sizes {10, 25, 50, 80, 100, 110}, in terms of the estimates, relative MSEs and the relative absolute biases.
- Studying the effect of different sample sizes (mentioned in the aforesaid point) on each of the six estimators, with respect to the estimates, MSEs, relative MSEs, biases and relative absolute biases.

SIMULATION ALGORITHM:

Step 1: Call the package "hypergeo", to use its functions.

Step 2: Start the main function , taking parameters n, sigma_sq, and covariate value x_0.

1) Define a two-component vector x0 with elements 1 and x_0.
2) Define the true parameter vector as (1,1)'.
3) Define true mean.
4) Form the design matrix of order (n x 2) by taking first column as all 1 and the second column as a random sample from uniform(0,1) distribution.
5) Obtain the vector v0 as the factor attached to sigma square in the variance of the linear estimator.
6) Start a replication loop to simulate 100 times.
7) Obtain the response vector Y.
8) Obtain the OLS estimates as beta_hat.
9) Obtain number of independent observations 'm'.
10) Compute the Residual Sum of Squares.
11) Compute the estimates of error variance for the Maximum Likelihood and the REML estimators.

12) Compute the estimate of mean, the MSE and the bias of the Maximum Likelihood (ML) estimator.

13) Compute the estimate of mean, the MSE and the bias of the Restricted Maximum Likelihood (REML) estimator.

14) Compute the estimate of mean and the variance of the Uniformly Minimum Variance Unbiased (UMVU) estimator, using the genhypergeo() function.

15) Compute the estimate of mean, the MSE and the bias of the Bias-corrected REML (EV) estimator, with the help of integrate() function for numerical integration.

16) Compute the estimate of mean, the MSE and the bias of the Minimum MSE (MM) estimator, with the help of integrate() function for numerical integration.

17) Compute the estimate of mean, the MSE and the bias of the Minimum Bias (MB) estimator, with the help of integrate() function for numerical integration.

18) End the replication loop.

19) Combine all these values as a data frame and export it to the function call statement.

Step 3: We set the seed to 12345678.

Step 4: The value of the error variance is initialized to 1.

Step 5: Initialize the vector containing the values of sample sizes : 10, 25, 50, 80, 100, 110 and store the length of the vector.

Step 6: Store the covariate values in the vector x and store its length.

(SAMPLE SIZE WISE COMPARISON)

Step 7: Compare estimates of mean with true mean, through graphs. Plot all the mean estimates along with the true mean for each sample size n.

Step 8: Compare Relative MSE w.r.t the MM estimator, through graphs. Plot all the relative MSEs for each sample size n.

Step 9: Compare Relative Absolute Bias, through graphs. Plot the relative absolute biases for each n.

(ESTIMATOR WISE COMPARISON)

Step 10: Compare estimates of mean with true mean, through graphs. Plot the mean estimates for different sample sizes along with the true mean for each estimator.

Step 11: Compare the MSEs and Relative MSE w.r.t the MM estimator, through graphs. Plot all the MSEs and relative MSEs for different sample sizes for each estimator.

Step 12: Compare the bias and relative absolute bias, through graphs. Plot the biases and the relative absolute biases for different sample sizes for each estimator.

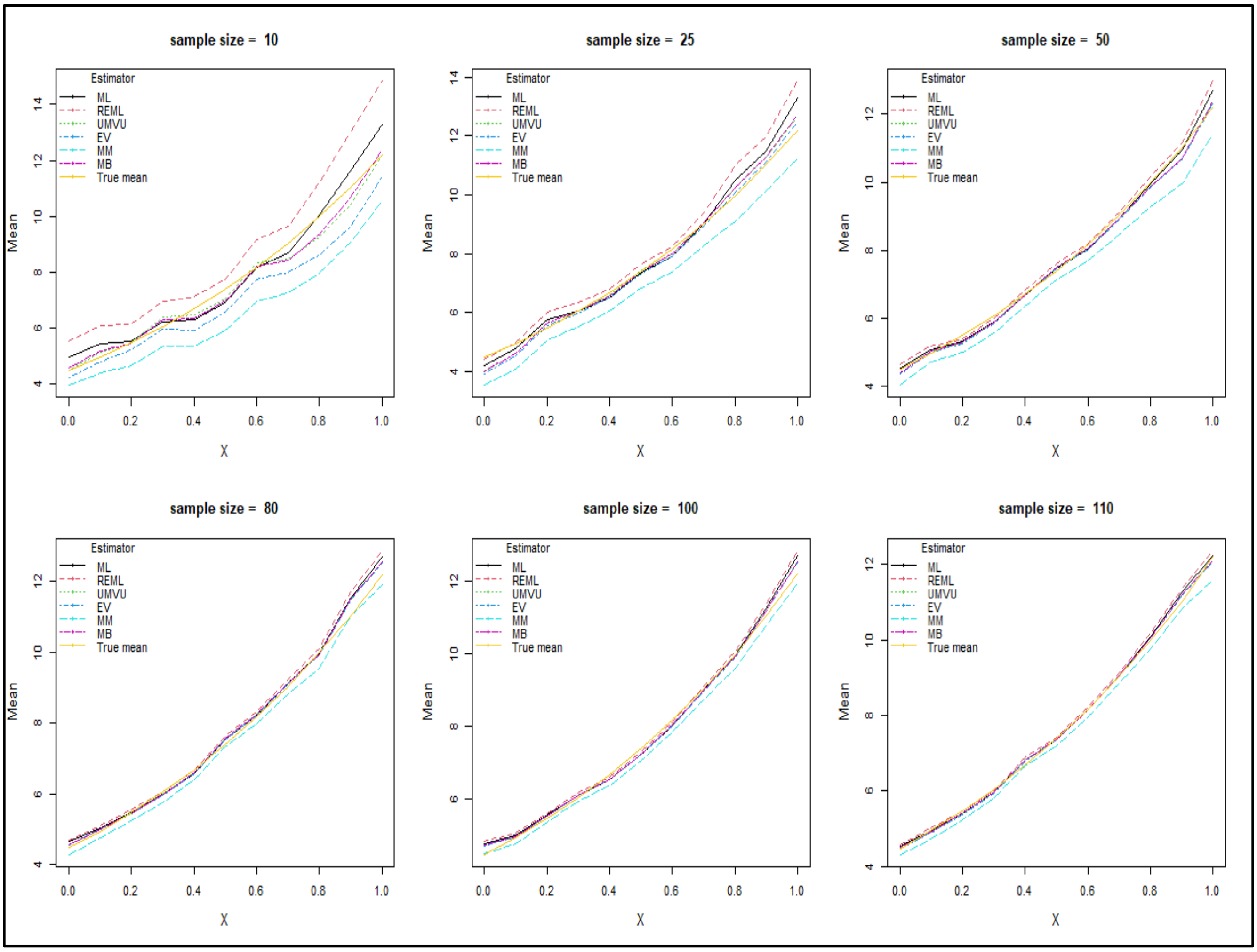
******

(SAMPLE SIZE WISE COMPARISON)

Here our objective is to make a graphical comparison among six different estimators over six distinct sample sizes. Six plots, one for each of the six sample sizes, will each depict the comparison among the estimators, with respect to the estimates, their relative MSEs and relative absolute biases.

1. <u>COMPARISON OF THE DIFFERENT ESTIMATES WITH THE TRUE MEAN</u> :

<u>FIGURE 2</u>



OBVERVATIONS:

❖ It is observed that as sample size increases, all estimators, on an average, traces the value of the true mean more and more closely.

❖ For lower sample sizes, the undulations in estimation of the lognormal mean are higher.

❖ It may also be perceived that the UMVU estimator as well as the ML estimator performs quite well in terms of estimation as they are seen to lie close to the true value for almost the entire range of covariates.

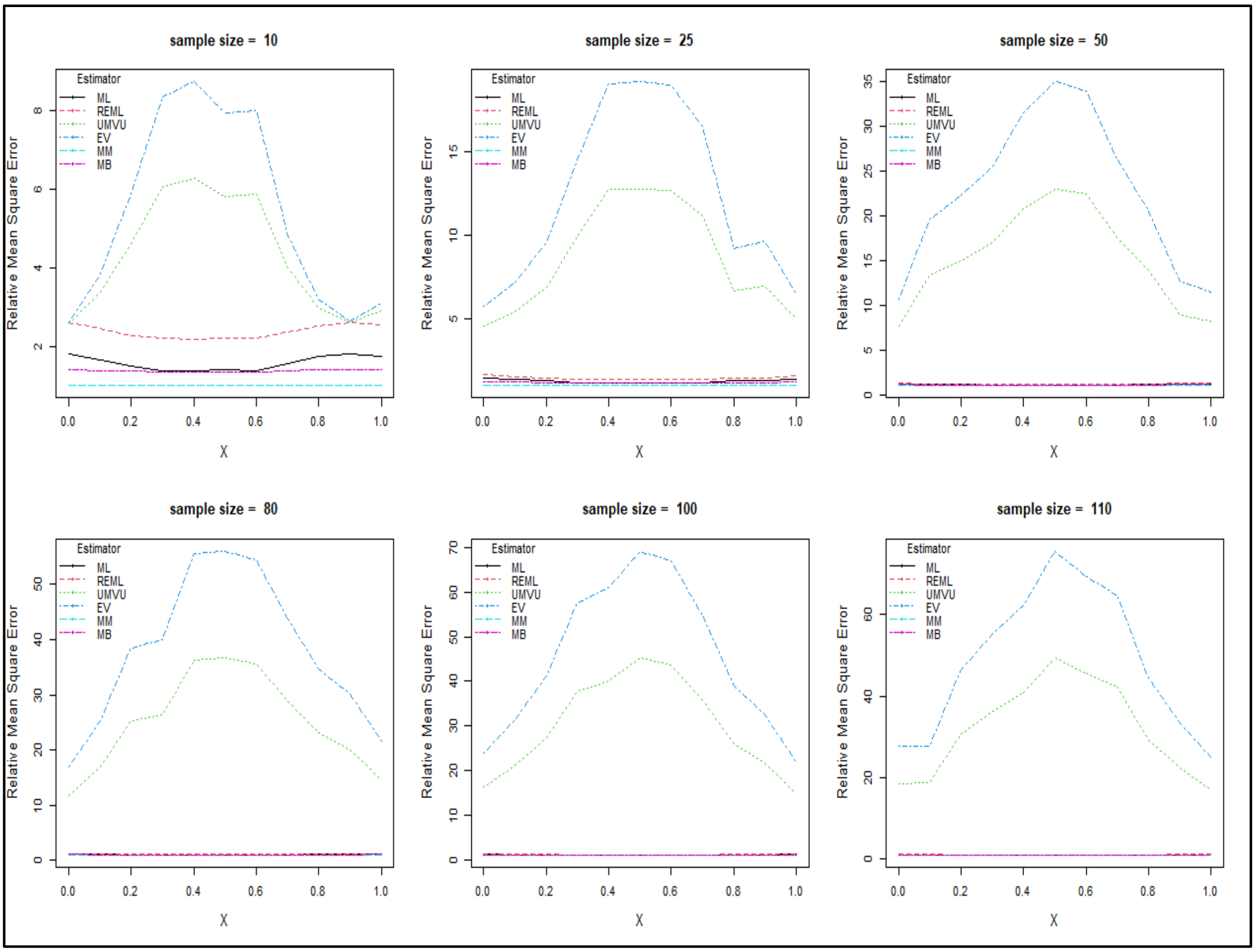
❖ The REML estimator, followed by the MM estimator, seem to be the least effective among all the above estimators, in estimating the true lognormal mean as they appear the most deviated, in any of the six plots, compared to others.

❖ In this regard, we may say that as the value of the covariate increases, the estimates seem to be more dispersed among themselves and about the true mean. This feature is prominent for small samples but fades gradually for large samples

***

2. COMPARISON OF THE RELATIVE MSE WITH RESPECT TO MM ESTIMATOR :

FIGURE 3



OBVERVATIONS:

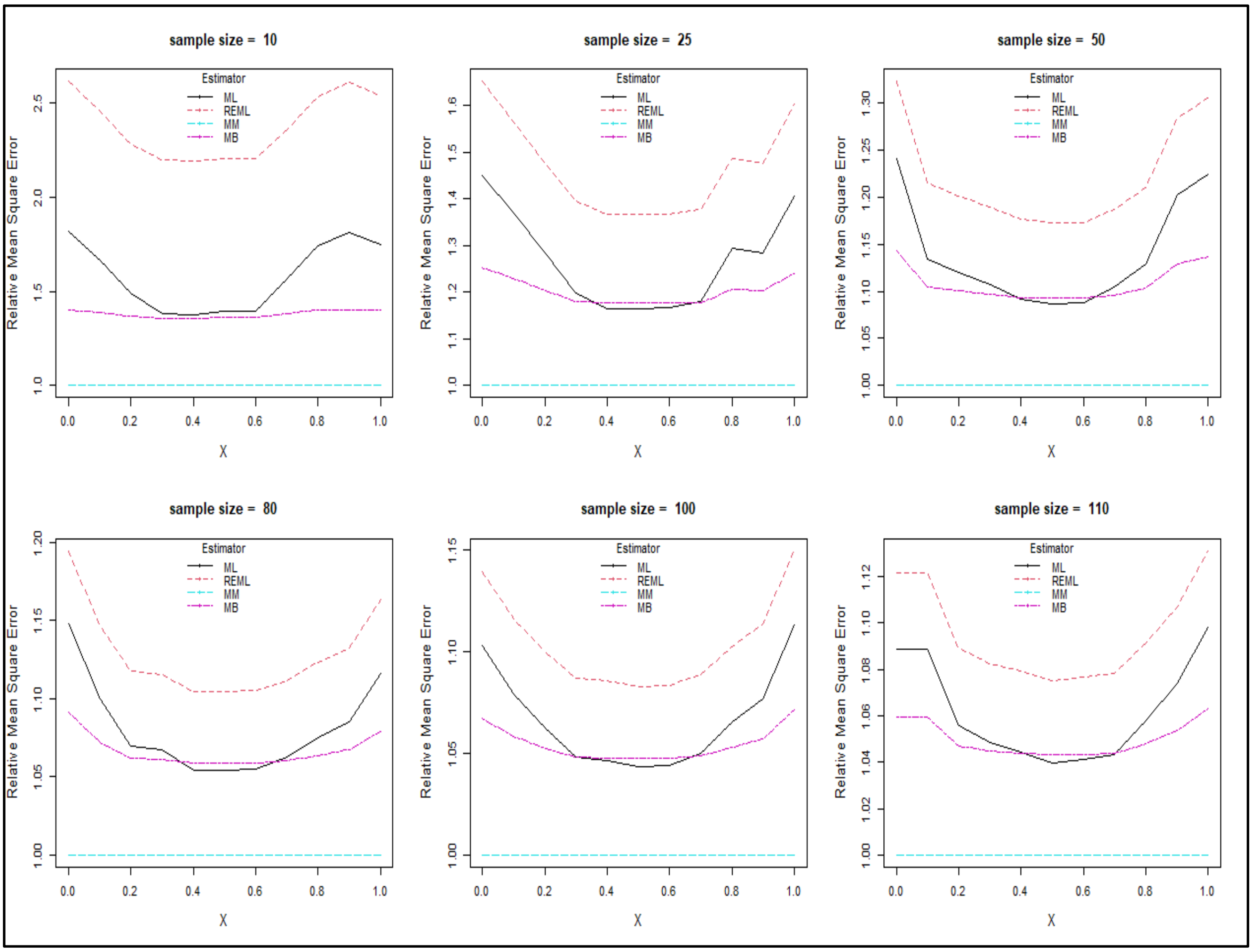
❖ It is clearly observed from all the plots that the EV estimator, followed by the UMVU estimator, have the highest values of the relative MSE among all the estimators and the magnitude amplifies with increasing sample size.

- ❖ It is also important to note that since UMVUE has the minimum variance among all estimators, we cannot obtain any other *unbiased* estimator with a lower relative MSE than the UMVUE.
- ❖ As desired, the MM estimator, which has relative MSE equal to 1, lies below all other estimators. However, this fact is not pretty clear from the plots other than that for sample size 10.
- ❖ For higher and higher sample sizes, the MSEs of estimators other than that of the EV and the UMVU estimator tend to superimpose near the zero mark.

Thus, for getting a clearer picture of the performance of the other estimators, we remove the data on the UMVU and the EV estimators from our study and re-plot the relative MSEs.

(RE-PLOTTING THE ABOVE FIGURE AFTER NECESSARY REMOVALS)

<u>FIGURE 4</u>



OBVERVATIONS:

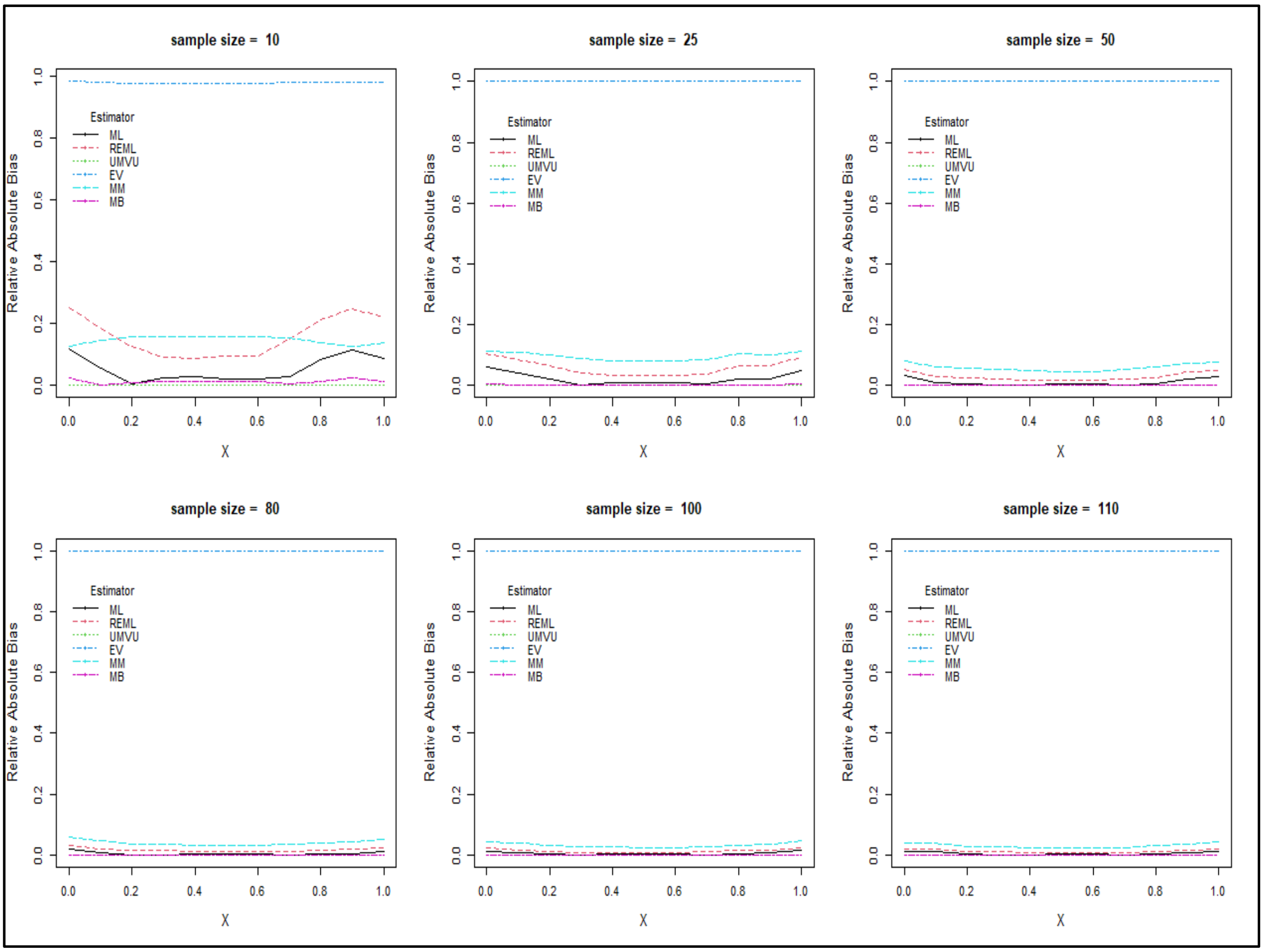
- ❖ The estimators, in general, have lower MSE values for moderate values of the covariates.
- ❖ Now, from the above figure, it is evident the MM estimator is indeed one which minimizes the MSE among all the estimators.
- ❖ The relative MSEs of all other estimators clearly lie way above 1, with the closest to 1 being the minimum bias estimator and the farthest being the REML estimator.

Thus, our proposed MM estimator is the desired one with respect to minimizing the MSE.

***

3. <u>COMPARISON OF THE RELATIVE ABSOLUTE BIAS</u> :

<u>FIGURE 5</u>



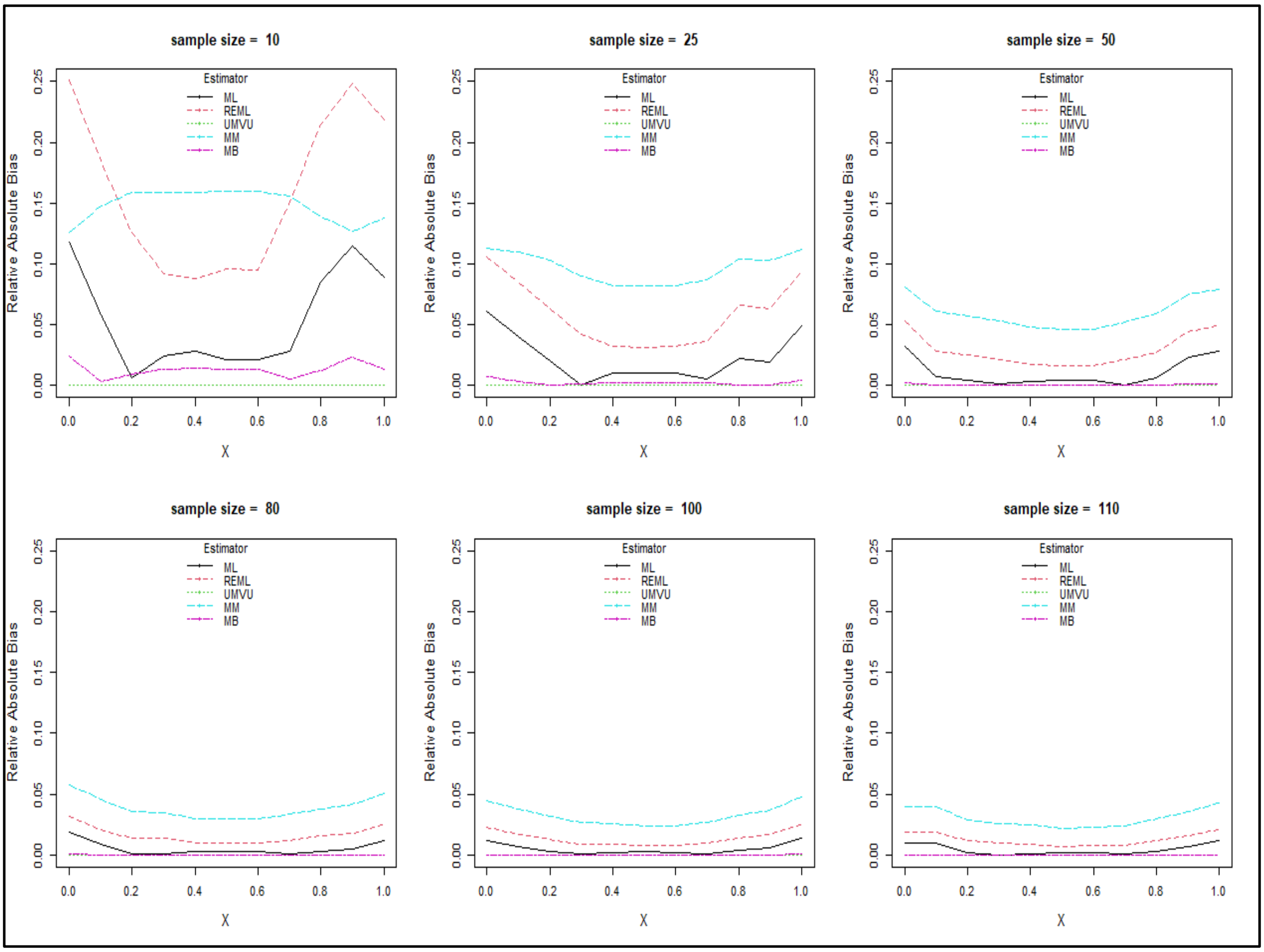
OBVERVATIONS:

- The above plots are indicative of the fact that the EV estimator has relative bias with a magnitude close to 1, for any value of the covariate in the entire range of covariates. With increase in sample size, the nature of the relative absolute bias of the EV estimator is almost unchanged.
- Also note that the bias for an UMVUE is necessarily 0.
- The values of the relative bias for the other estimators are closer to null. For increasing sample size, the values tend to override near 0.

To analyse the nature of variation of the relative bias with covariate values and sample size for the other estimators, we remove the data on EV estimator and re-plot the other relative absolute biases.

(RE-PLOTTING THE ABOVE FIGURE AFTER NECESSARY REMOVALS)

<u>FIGURE 6</u>



OBVERVATIONS:

- For a sample size of 10, we observe a lot of fluctuations, with the largest fluctuations being observed in the maximum likelihood estimators, both restricted and non-restricted. Interestingly, as sample size increases, the fluctuations smooth out very quickly and the rapidity in smoothness is best observed in the ML and REML estimators.
- For sample sizes 25 and above, we observe that the MM estimator has the highest absolute bias, although the magnitude is small.
- The most important thing that needs attention is that the proposed MB estimator, which aimed at minimizing the bias in estimation, indeed minimizes the bias and lies below the other estimators, for almost all values of covariate and all sample sizes, and the values almost trace the zero line.

Thus, our MB estimator is the desired one with respect to minimizing the Bias.

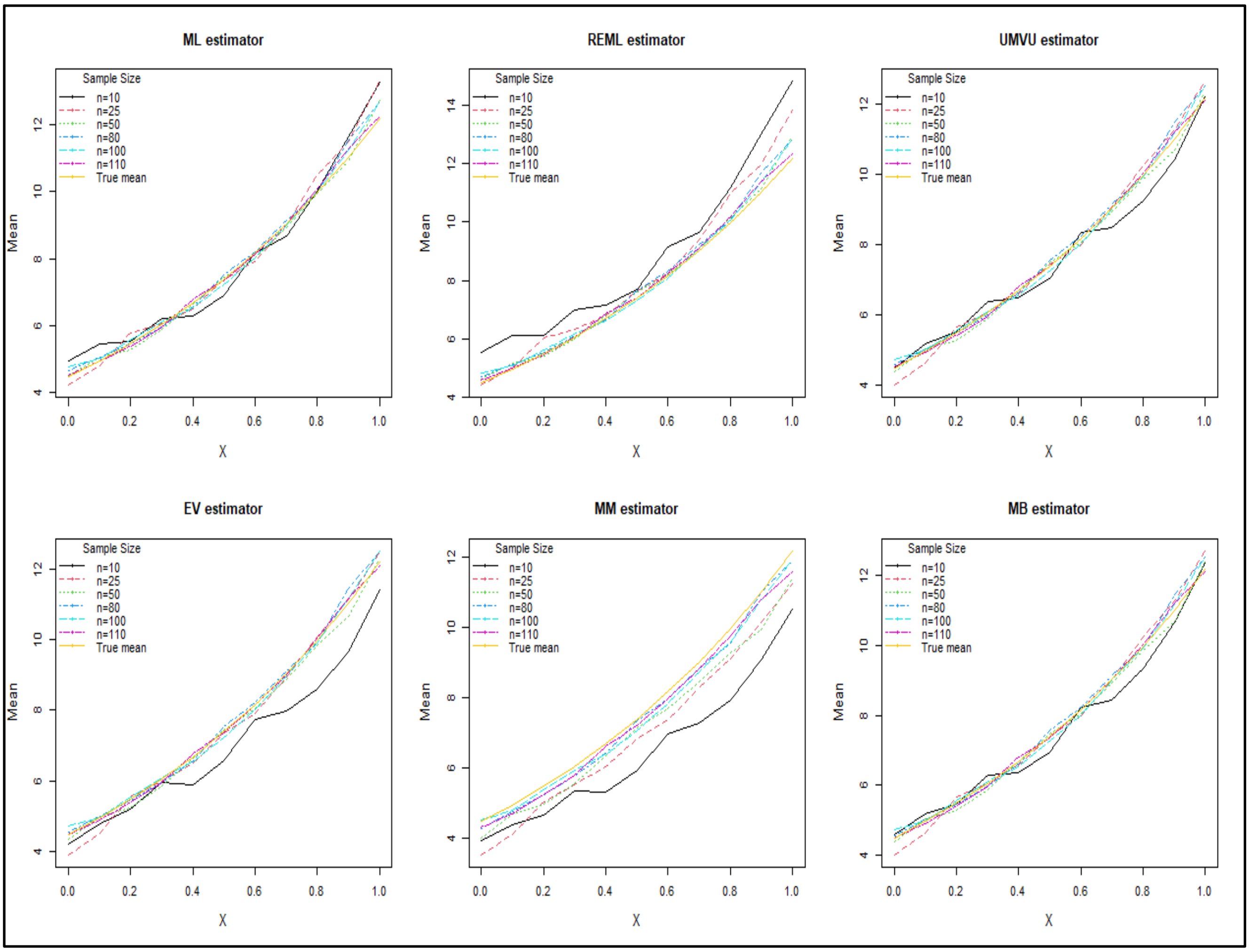
*****

(ESTIMATOR WISE COMPARISON)

Here our objective is to make a graphical comparison among six different sample sizes individually and overally for six distinct sample sizes. Six plots, one for each of the six estimators, will each depict the comparison among the sample sizes, with respect to the estimates, their MSEs and relative MSEs, biases and relative absolute biases.

1. COMPARISON OF THE DIFFERENT ESTIMATES WITH THE TRUE MEAN :

FIGURE 7



OBVERVATIONS:

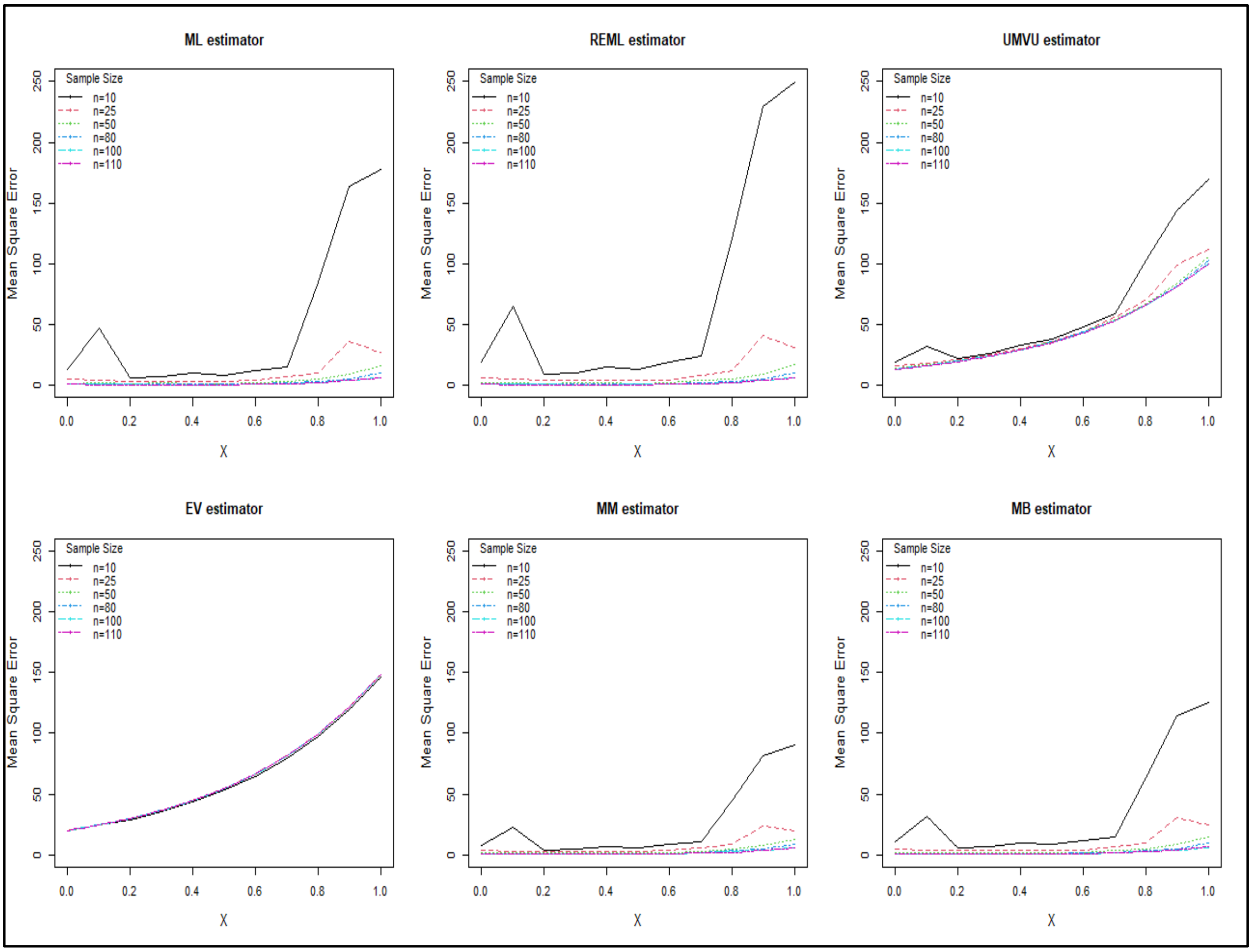
- It is observed that for all the estimators, the estimates tend to fluctuate along the entire range of the covariates, when sample size of 10 is employed. As per large sample considerations, with increase in sample size, the estimates tend to closely resemble the true lognormal mean, for more or less all estimators.
- Closer investigation will lead us to conclude that the UMVU estimates, the ML estimates and the MB estimates closely trace the true mean for almost the entire range of sample sizes and covariates.
- The REML estimator can be seen to overestimate the true mean for all sample sizes.

❖ The MM estimator underestimates the true mean value, for almost the entire range of sample sizes.

***

2. COMPARISON OF THE MEAN SQUARE ERRORS:

FIGURE 8



OBVERVATIONS:

❖ In general, for very large magnitude of the covariate, the MSE has very high values, mainly in small samples.

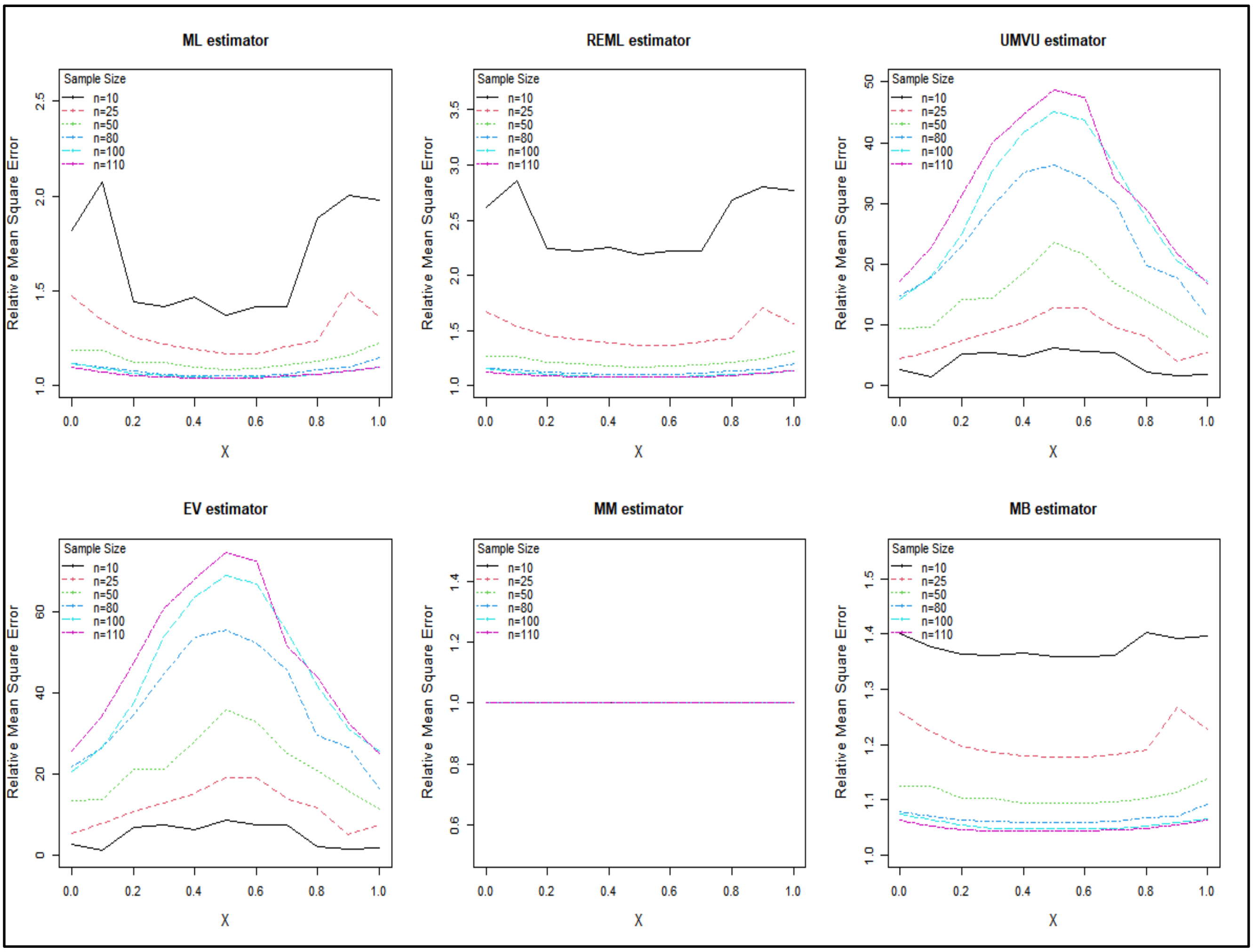
❖ For lower values of sample size, the MSE of the REML estimator has the highest value when covariate takes larger values.

- For lower covariates, all the estimators have nearly the same MSE.
- Another observation is that the MSE values tend to increase functionally with increase in covariate values in case of the UMVU and the EV estimators.
- Note that, for the EV estimator, with increase in x values, the MSE values maintain a decent functional relation, whatever the sample size may be.
- The least MSE, overall, is observed for the minimum MSE estimator, as the plot attains the least height, for any sample size.

***

3. COMPARISON OF THE RELATIVE MSE WITH RESPECT TO MM ESTIMATOR:

FIGURE 9



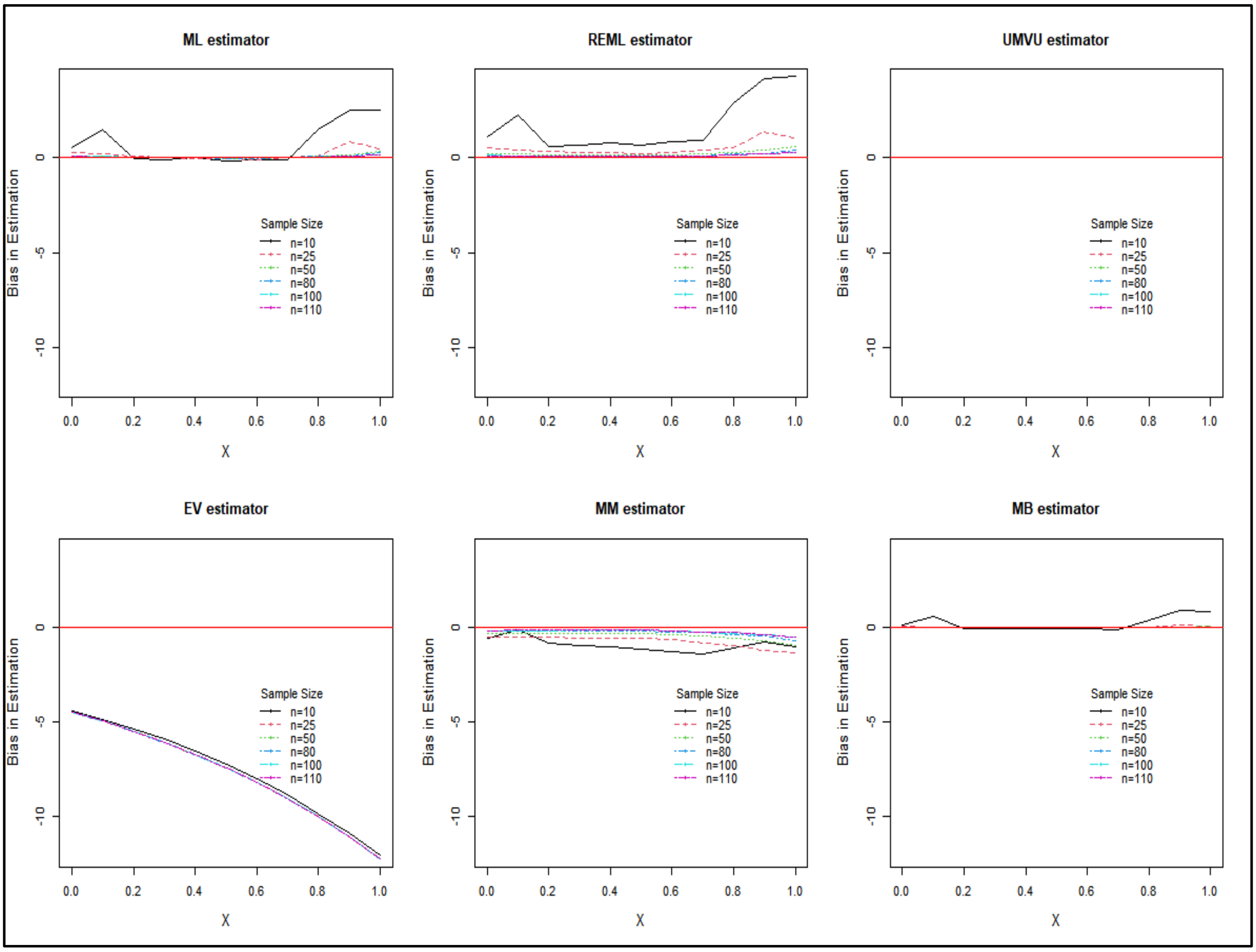
OBSERVATIONS:

- The plot for the MM estimator is always a straight line fixed at 1.
- If we observe the graphs for the ML, REML and the MB estimators, we notice a convexity towards origin, for all range of sample sizes, signifying low relative MSE for moderate x-values.
- For the UMVUE and the EV estimator, the nature of variation of the relative MSE is subtly concave towards the origin, indicating high relative MSE for moderate covariate values.
- With increase in sample size, the relative MSE falls in case of the ML, REML and MB estimators while it increases for the UMVU and EV estimators.

***

4. COMPARISON OF THE BIAS IN ESTIMATION:

FIGURE 10



OBVERVATIONS:

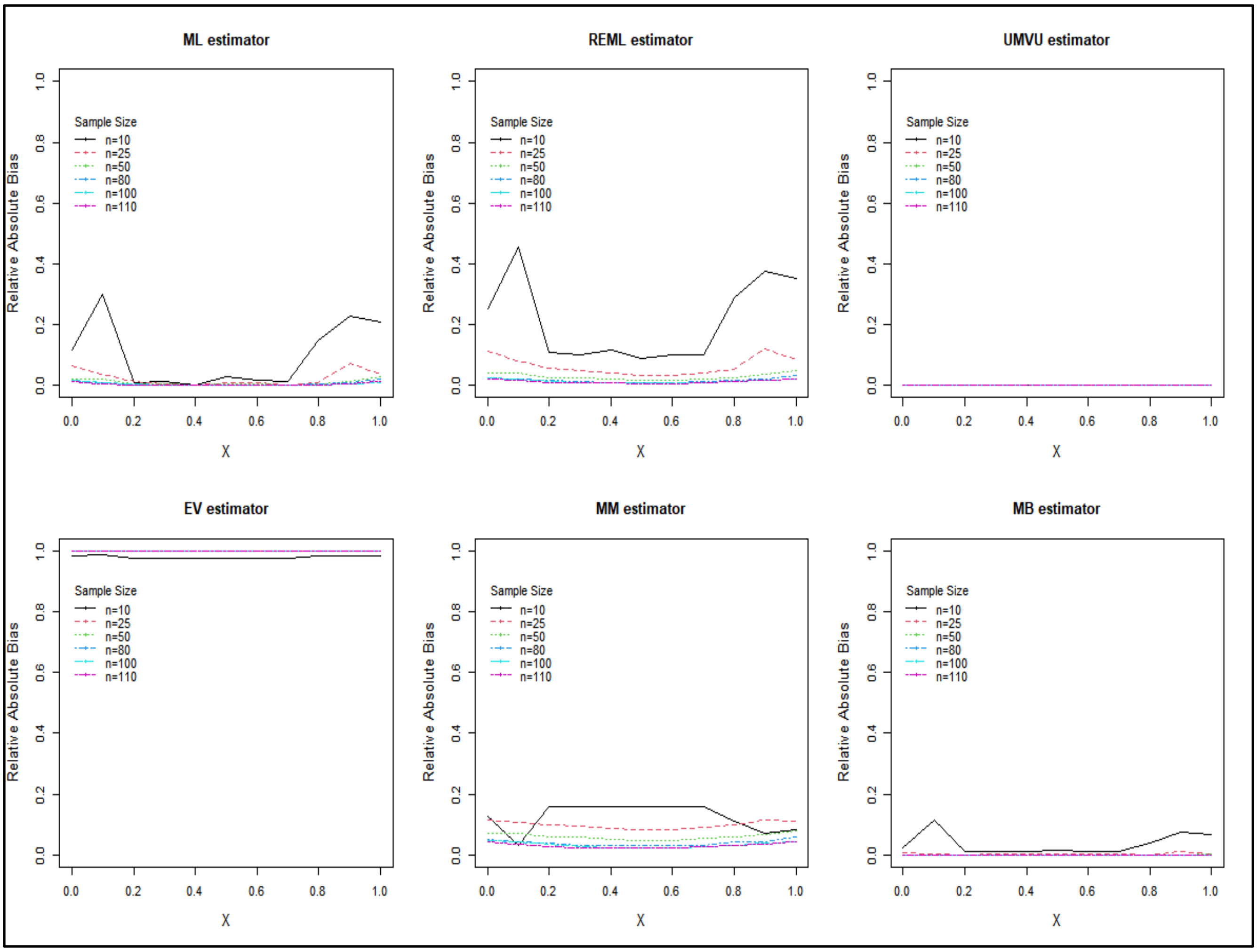
- The desired property of unbiasedness is indicated by the red line.
- We observe that the nature of bias for the EV estimation is strikingly distinct from the rest as it displays a large negative bias, for all sample sizes and this bias increases in magnitude with increase in the covariate values.
- The UMVUE has 0 bias, as the name suggests that the property of unbiasedness (or 0 bias) is innate to the estimator.
- The MM estimator also shows up a certain amount of negative bias, though the magnitude is less as compared to the EV estimator.
- The REML estimator has positive bias throughout the range of covariates and for all the sample sizes.
- For ML and MB estimators, the bias is neither necessarily positive nor necessarily negative throughout.
- Comparing all the plots, we observe that the MB estimator has the least bias for any covariate value, among all the estimators. This is evident from the fact that in the plot of bias for the MB estimator, the bias lies closer to the zero mark, compared to others (except for the UMVU Estimator).

***

And now, the final plot related to my arena of study is given by…….

5. COMPARISON OF THE RELATIVE ABSOLUTE BIAS :

FIGURE 11



OBVERVATIONS:

- The UMVU estimator with zero bias displays a graph with a straight line at 0.
- The EV estimator has remarkably high absolute bias, almost close to 1.
- Among other estimators, we observe that all of them have rather low relative absolute bias.
- The REML and the ML estimators show remarkable decrease and smoothing in relative bias with increase in sample size (For small samples, the bias fluctuates with change in x-values) .
- However, the MB estimator displays the least relative bias among all the estimators, as we observe that the values are quite close to null, for almost all sample sizes and all values of covariates.

*****

CONCLUSION:

We can make the following broad conclusions from our study:

- ♣ Going by the precision of estimates in estimating the true mean,

  the Maximum Likelihood Estimator and the Uniformly Minimum Variance Unbiased Estimator are the most effective ones, while the Restricted Maximum Likelihood Estimator is the least worthwhile.

- ♣ Judging the Mean Square Error (or the relative Mean Square Error with respect to the Minimum MSE Estimator),

  the Minimum MSE Estimator is the most effective one while the Bias-corrected REML Estimator performs the worst.

- ♣ Considering the Bias in estimation (or the Relative Absolute Bias),

  the Minimum Bias Estimator seems the best while the Bias-corrected REML Estimator appears the least effective.

Thus, the proposed estimators are proven to perform well and have indeed done justice to the way they have been named. The Minimum MSE indeed minimizes the MSE and the Minimum Bias truly minimizes the bias and are hence efficient estimators of lognormal mean. Thus, they can be treated as important estimators in further studies related to Lognormal Linear Models and Lognormal Regression.

* *

FUTURE SCOPE:

The simulation studies in my research have been performed under controlled conditions: the value of error variance was fixed at 1, selected samples sizes less than 110 were considered and covariates were allowed to vary between 0 and 1. These may be relaxed in future and new studies may be performed with an altogether different set of assumptions. Indeed, Statistics never fails to impress!!

****

REFERENCES :

PUBLICATION:

Efficient Mean Estimation in Lognormal Linear Models

by Haipeng Shen and Zhengyuan Zhu

BOOKS:

1. *An Introduction to Probability and Statistics*

   by V. K. Rohatgi and A. K. Md. Ehsanes Saleh

2. *Introduction to the Theory of Statistics*

   by A. M. Mood, F. A. Graybill and Duane C. Boes

3. *Linear Statistical Inference and its Applications*

   by C. R. Rao

4. *Statistical Inference*

   by George Casella and Roger L. Berger

5. *Log-Linear Models and Logistic Regression*

   by Ronald Christensen.

****

ACKNOWLEDGEMENT:

I would hereby like to express my gratitude towards

• **Rev. Dr. Father Dominic Savio SJ**, Principal, St. Xavier's College (Autonomous), Kolkata,

• **Prof. Bertram Da'Silva**, Vice-Principal, Arts and Science Departments,

• **Prof. Dr. Tapati Dutta**, Dean of Science ,

• **Prof. Dr. Arghya Bannerjee**, Dean of Arts

,for giving me an opportunity to carry out my dissertation.

I particularly thank **Prof. Madhura Dasgupta**, my mentor and supervisor, for her tireless efforts to make this dissertation a reality.

I thank all other professors of the department of Statistics (**Prof. Dr. Durba Bhattacharya**, HOD; **Prof. Dr. Surabhi Dasgupta**; **Prof. Dr. Surupa Chakraborty**; **Prof. Dr. Ayan Chandra**, **Prof. Debjit Sengupta** and **Prof. Pallabi Ghosh**) to support me all the way and imbibe in me the important aspects of Statistics which helped me in my research.

Last but not the least, I thank my family and friends who have been my constant moral support in completing my study.

- with due respect,

Anis Pakrashi.

****

THANK YOU !!