

REVSTAT
Statistical Journal
vol. 24 - n. 1 - January 2026



REVSTAT — Statistical Journal, vol.24, n. 1 (January 2026)

vol.1, 2003- . - Lisbon : Statistics Portugal, 2003- .

Continues: Revista de Estatística = ISSN 0873-4275.

ISSN 1645-6726 ; e-ISSN 2183-0371

Editorial Board (2025-2026)

Editor-in-Chief – *Manuel SCOTTO*

Co-Editor – *Cláudia NUNES*

Associate Editors

Abdelhakim AKNOUCHE

Andrés ALONSO

Barry ARNOLD

Wagner BARRETO-SOUZA

Francisco BLASQUES

Paula BRITO

Rui CASTRO

Valérie CHAVEZ-DEMOULIN

David CONESA

Charmaine DEAN

Fernanda FIGUEIREDO

Jorge Milhazes FREITAS

Stéphane GIRARD

Sónia GOUVEIA

Victor LEIVA

Artur LEMONTE

Shuangzhe LIU

Raquel MENEZES

Fernando MOURA

Cláudia NEVES

John NOLAN

Carlos OLIVEIRA

Paulo Eduardo OLIVEIRA

Pedro OLIVEIRA

Rosário OLIVEIRA

Gilbert SAPORTA

Alexandra M. SCHMIDT

Lisete SOUSA

Jacobo de UÑA-ÁLVAREZ

Christian WEIß

Executive Editor – *Olga BESSA MENDES*

Publisher – Statistics Portugal

Layout-Graphic Design – Carlos Perpétuo | Cover Design* – Helena Nogueira

Digital edition



Creative Commons Attribution 4.0 International (CC BY 4.0)






© Statistics Portugal, Lisbon. Portugal, 2026

* *image*: stain glass window by Abel Manta (1888-1982)

INDEX

Statistical Inference for the Inverse Lindley Distribution Based on Lower Record Values	
<i>Bahareh Etemad Golestani, Ehsan Ormoz and S. M. T. K. MirMostafae</i>	1
On the Construction of Third-Order Rotatable Designs in Smaller Runs	
<i>Ankita Verma, Seema Jaggi, Eldho Varghese, Cini Varghese, Arpan Bhowmik, Anindita Datta, Hemavathi M. and Ashutosh Dalal</i>	27
A New General Class of Ridge-type Estimator in Linear Regression Models	
<i>Kadri Ulaş Akay, Esra Ertan, Ali Erkoç and Ferhat Taş</i>	45
Comparison and Equality of Bajraktarević-type ψ-estimators	
<i>Mátyás Barczy and Zsolt Páles</i>	71
A Robust Variable Screening Approach with Application to Gene Expression Data	
<i>Mohammad Kazemi</i>	91
<i>E</i>-value Formulae for the Odds Ratio	
<i>Kalle Leppälä</i>	107

Statistical Inference for the Inverse Lindley Distribution Based on Lower Record Values

- Authors: BAHAREH ETEMAD GOLESTANI  
– Department of mathematics and Statistics, Mashhad Branch, Islamic Azad University,
Mashhad, Iran
etemad.bahar@gmail.com
- EHSAN ORMOZ  
– Department of mathematics and Statistics, Mashhad Branch, Islamic Azad University,
Mashhad, Iran
Ehsanormoz@mshdiau.ac.ir
- S. M. T. K. MIRMOSTAFEE 
– Department of Statistics, Faculty of Mathematical Sciences, University of Mazandaran,
Babolsar, Iran
m.mirmostafae@umz.ac.ir

Received: November 2022

Revised: January 2024

Accepted: January 2024

Abstract:

- In this paper, we discuss the problem of classical and Bayesian estimation of the parameter of the inverse Lindley distribution based on lower records, as well as the prediction of a future record value. We obtain the maximum likelihood estimator, the approximate confidence interval, as well as two bootstrap-type confidence intervals for the parameter based on the inverse Lindley distribution records. In the context of Bayesian estimation, we use the Tierney and Kadane's method and two Markov chain Monte Carlo approaches. The future record values are also explored using the maximum likelihood and Bayesian approaches. The highest conditional density, as well as Bayesian intervals, are also constructed for a future lower record. A simulation study and a real data example are also given for the sake of comparison and illustration.

Keywords:

- *Bayesian estimation and prediction; general entropy loss function; inverse Lindley distribution; lower record values; maximum likelihood estimation.*

AMS Subject Classification:

- 62F10, 62C10.

1. INTRODUCTION

In recent years, the inverse Lindley distribution (*ILD*) has attracted the attention of several authors. It was first introduced by [Sharma et al. \(2015\)](#) and its stress-strength reliability was explored under classical and Bayesian models. In addition, its application to head and neck cancer data was demonstrated. Let Y have a Lindley distribution with parameter θ , and define $X = \frac{1}{Y}$, then X has an inverse Lindley distribution with parameter θ (notationally $X \sim ILD(\theta)$) and the probability density function (*PDF*) of X is obtained to be

$$(1.1) \quad f(x; \theta) = \frac{\theta^2}{1 + \theta} \left(\frac{x+1}{x^3} \right) e^{-\frac{\theta}{x}}, \quad x > 0, \quad \theta > 0.$$

The corresponding cumulative distribution function (*CDF*) is given by

$$(1.2) \quad F(x; \theta) = \left(1 + \frac{\theta}{(1 + \theta)x} \right) e^{-\frac{\theta}{x}}, \quad x > 0, \quad \theta > 0.$$

The *PDF* (1.1), is a mixture of the inverse exponential distribution with parameter θ and the inverse gamma distribution with shape parameter 2 and scale parameter θ , namely (1.1) can be written as

$$f(x; \theta) = pf_1(x) + (1 - p)f_2(x),$$

where

$$f_1(x) = \theta x^{-2} e^{-\frac{\theta}{x}}, \quad f_2(x) = \theta^2 x^{-3} e^{-\frac{\theta}{x}}, \quad \text{and} \quad p = \frac{\theta}{1 + \theta}.$$

[Basu et al. \(2017\)](#) and [Basu et al. \(2019\)](#) discussed the problem of estimation of the parameter of the *ILD* under Type-I censored data and hybrid censored data, respectively. Many generalizations of the *ILD* have been introduced. For example, [Asgharzadeh et al. \(2017\)](#) studied the generalized inverse Lindley distribution and presented an application to Danish fire insurance data.

Record values are of great significance in many real-life situations such as in industry, weather, and life-testing events. Record values and their basic properties have been discussed by [Chandler \(1952\)](#), [Resnick \(1973\)](#), [Nevzorov \(1988\)](#), and [Arnold et al. \(1998\)](#), among others. Recently, [Asgharzadeh et al. \(2018\)](#) and [Fallah et al. \(2018\)](#) worked on the inferential problems for the Lindley distribution, and [Singh et al. \(2020\)](#) focused on the inference for the generalized Lindley distribution based on record data. Let X_1, X_2, \dots be a sequence of independent and identically distributed (iid) random variables with *CDF* $F(x; \theta)$ and *PDF* $f(x; \theta)$. Then the observation X_j is a lower record value if it is smaller than all its preceding observations, namely $X_j < X_i, \forall i < j$. In other words, let $L(1) = 1$ and $L(m) = \min\{j | j > L(m-1), X_j < X_{L(m-1)}\}$ for $m > 1$. Then $X_{L(m)}$ is the m -th lower record value, and the sequence $\{L(m), m \geq 1\}$ represents the record times. The *PDF* of $X_{L(m)}$ for $m \geq 1$ is given by (see e.g. [Arnold et al., 1998](#))

$$f_{X_{L(m)}}(x) = \frac{1}{(m-1)!} [-\ln(F(x; \theta))]^{m-1} f(x; \theta), \quad x > 0, \quad m \geq 1.$$

The joint *PDF* of $X_{L(m)}$ and $X_{L(n)}$, for $1 \leq m < n$ and $x < y$, is

$$f_{X_{L(m)}, X_{L(n)}}(x, y; \theta) = \frac{1}{(m-1)!(n-m-1)!} [-\ln(F(x; \theta))]^{m-1} \\ \times [\ln(F(x; \theta)) - \ln(F(y; \theta))]^{n-m-1} \frac{f(x; \theta)}{F(x; \theta)} f(y; \theta).$$

In addition, suppose that $\mathbf{x} = (x_1, \dots, x_m)$ is the observed vector of $(X_{L(1)}, \dots, X_{L(m)})$, then the likelihood function of θ given the m lower records can be expressed as

$$(1.3) \quad L(\theta|\mathbf{x}) = f(x_m; \theta) \prod_{i=1}^{m-1} \frac{f(x_i; \theta)}{F(x_i; \theta)}, \quad x > 0, \quad m \geq 1.$$

So, for the *ILD*, the *PDF* of m -th lower record is given by

$$(1.4) \quad f_{X_{L(m)}}(x) = \frac{1}{\Gamma(m)} \left[-\ln \left(1 + \frac{\theta}{(1+\theta)x} \right) + \frac{\theta}{x} \right]^{m-1} \frac{\theta^2}{1+\theta} \left(\frac{1+x}{x^3} \right) e^{-\frac{\theta}{x}}, \quad x > 0, \quad \theta > 0.$$

The main aim of this article is to present both frequentist and Bayesian methodology to estimate the parameter of the *ILD* based on lower records and to predict a future record based on past observed record values. The rest of the paper is organized as follows: In Section 2, we use the maximum likelihood (*ML*) method as a frequentist methodology to obtain a point estimator of the parameter. Besides, the asymptotic confidence interval (*CI*) as well as two different bootstrap-type *CI*s are obtained. We also consider the problem of Bayesian estimation of the unknown parameter in this section. In Section 3, the problem of predicting a future record value is discussed based on using both classical and Bayesian procedures. In Section 4, a Monte Carlo simulation study is conducted to evaluate the performances of the proposed estimators in the sense of estimated bias and their associated estimated risks. In Section 5, the applicability of the paper results, is shown using an application to real data. Finally, the paper ends with some conclusions in Section 6.

2. PARAMETER ESTIMATION

In this section, we use both classical and Bayesian methods of estimation to evaluate the parameter of the inverse Lindley distribution based on lower records.

2.1. Maximum likelihood estimation

In this subsection, we discuss the process of obtaining the *ML* estimator of parameter θ based on lower record values for *ILD*(θ). Suppose that $X_{L(1)}, \dots, X_{L(m)}$ are the first m record statistics arising from a sequence of iid random variables from *ILD*(θ) with *PDF* (1.1) and $\mathbf{x} = (x_1, x_2, \dots, x_m)$ is the observed vector of $(X_{L(1)}, \dots, X_{L(m)})$. The likelihood function of the parameter given \mathbf{x} is as follows:

$$L(\theta|\mathbf{x}) = \frac{\theta^{2m} e^{-\frac{\theta}{x_m}}}{x_m(1+\theta)} \prod_{i=1}^m \frac{1+x_i}{x_i^2} \prod_{i=1}^{m-1} \frac{1}{\theta(1+x_i) + x_i}.$$

Hence, the log-likelihood function is

$$(2.1) \quad l(\theta) = \ln L(\theta|\mathbf{x}) = 2m \ln \theta - \ln(1 + \theta) - \frac{\theta}{x_m} - \sum_{i=1}^{m-1} \ln(\theta(1 + x_i) + x_i) + A(\mathbf{x}),$$

where $A(\mathbf{x}) = \sum_{i=1}^m \ln(1 + x_i) - \ln x_m - 2 \sum_{i=1}^m \ln x_i$.

The *ML* estimate of θ can be obtained by maximizing (2.1) with respect to θ . Upon differentiating (2.1) with respect to θ and equating it with zero, we have

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{2m}{\theta} - \frac{1}{1 + \theta} - \frac{1}{x_m} - \sum_{i=1}^{m-1} \frac{1 + x_i}{\theta(1 + x_i) + x_i} = 0.$$

It can be shown that the solution of (2.1) can be obtained as a fixed point solution of $h(\theta) = \theta$ where

$$h(\theta) = 2m \left(\frac{1}{1 + \theta} + \frac{1}{x_m} + \sum_{i=1}^{m-1} \frac{1 + x_i}{\theta(1 + x_i) + x_i} \right)^{-1}.$$

Next, we show the uniqueness and existence of the *ML* estimate of θ . To this end, let $v_1(\theta) = h(\theta)$ and $v_2(\theta) = \theta$. It can be easily verified that $v_1(\theta)$ is an increasing function with

$$v_1(0) = 2m \left(\sum_{i=1}^m \frac{1}{x_i} + m \right)^{-1}, \quad v_1(\infty) = 2m x_m.$$

So $v_1(\theta)$ starts from a positive real value at 0 and increases to $2m x_m$, which is a finite value. For large θ , $v_1(\theta)$ is a finite value whereas $v_2(\theta) \rightarrow \infty$ as θ goes to ∞ . This implies that there exists one real positive root, say $\hat{\theta}$, such that $h(\hat{\theta}) = \hat{\theta}$.

2.2. Asymptotic confidence interval

It seems that the *ML* estimate of θ does not possess an explicit form, and therefore it is not easy to obtain the variance of $\hat{\theta}$, where $\hat{\theta}$ denotes the *MLE* of θ . Consequently, we cannot get the exact distribution of the *MLE* and the exact bounds for the parameter. The intent is to use the large-sample approximation. The asymptotic distribution of $\hat{\theta}$ is (Lawless, 2003)

$$(\hat{\theta} - \theta) \xrightarrow{D} N(0, I_{X_{L(1)}, \dots, X_{L(m)}}^{-1}(\theta)),$$

where $I_{X_{L(1)}, \dots, X_{L(m)}}^{-1}(\theta)$ is the inverse of the Fisher information of the first m lower records about the unknown parameter θ and \xrightarrow{D} stands for convergence in distribution. Since θ is unknown, we estimate the asymptotic variance of $\hat{\theta}$ based on the inverse of the observed Fisher information of the first m lower records, in other words, we have

$$\widehat{\text{Var}}(\hat{\theta}) = \left(\tilde{I}_{X_{L(1)}, \dots, X_{L(m)}}(\hat{\theta}) \right)^{-1},$$

where

$$\tilde{I}_{X_{L(1)}, \dots, X_{L(m)}}(\hat{\theta}) = \frac{2m}{\hat{\theta}^2} - \frac{1}{(1 + \hat{\theta})^2} - \sum_{i=1}^{m-1} \left(\frac{1 + X_{L(i)}}{\hat{\theta}(1 + X_{L(i)}) + X_{L(i)}} \right)^2.$$

Using the above element, one can derive the approximate $100(1 - \alpha)\%$ *CI* of the parameter θ as follows:

$$\hat{\theta} \pm z_{\frac{\alpha}{2}} \sqrt{\widehat{\text{Var}}(\hat{\theta})},$$

where $z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ quantile of the standard normal distribution.

2.3. Bootstrap confidence interval

In this subsection, two different bootstrap confidence intervals are proposed. The first one is the bootstrap percentile (*Boot - P*) *CI* and the second one is the basic bootstrap (*Boot - B*) *CI* (Davison and Hinkley, 1997; Efron, 1982). The following algorithm is used to generate parametric bootstrap samples.

Algorithm 1

- Step 1: Compute the *ML* estimate of θ , denoted by $\hat{\theta}$, based on the observed lower records.
 Step 2: Generate the bootstrap lower record sample $X_{L(1)}^*, \dots, X_{L(m)}^*$, from *ILLD*($\hat{\theta}$).
 Step 3: Compute the *ML* estimate of θ based on the generated bootstrap sample in Step 2, denoted by $\hat{\theta}_1^*$.
 Step 4: Repeat Steps 2 and 3, B times, and store $\hat{\theta}_i^*$ for $i = 1, \dots, B$, say $\{\hat{\theta}_1^*, \dots, \hat{\theta}_B^*\}$.
-

i) *Boot - P* method

Arrange $\hat{\theta}_i^*$'s in an ascending order and let θ_i^* be the i -th ordered member of $\{\hat{\theta}_1^*, \dots, \hat{\theta}_B^*\}$, then the $100(1 - \gamma)\%$ bootstrap percentile *CI* for θ is given by

$$\left(\theta_{(B+1)\frac{\gamma}{2}}^*, \theta_{(B+1)(1-\frac{\gamma}{2})}^* \right).$$

ii) *Boot - B* method

The $100(1 - \gamma)\%$ basic bootstrap *CI* for θ is given by

$$\left(2\hat{\theta} - \theta_{(B+1)(1-\frac{\gamma}{2})}^*, 2\hat{\theta} - \theta_{(B+1)(\frac{\gamma}{2})}^* \right).$$

2.4. Bayesian estimation

In this subsection, we work on Bayesian estimation of the unknown parameter θ in the *ILLD*, based on lower record values. It should be noted that all the relations given in this subsection hold for the general case of one-dimensional parameter θ . In the context of Bayes estimation, the parameter is assumed to be a random variable with a prior distribution, $\pi(\theta)$. Let \mathbf{X} denote the informative sample and $L(\theta, \delta(\mathbf{X}))$ denote the loss function, where $\delta(\mathbf{X})$ is an estimator of θ . The Bayes estimator of θ is derived through minimizing the posterior risk $E[L(\theta, \delta(\mathbf{X}))|\mathbf{X}]$ with respect to δ . In the literature, the squared error (*SE*) loss function is one of the common loss functions that has been frequently used for estimation

problems, which is defined as $L(\theta, \delta(\mathbf{X})) = (\delta(\mathbf{X}) - \theta)^2$. The Bayes estimator of θ is given by $\delta_{SE}(\mathbf{X}) = E(\theta|\mathbf{X})$ under the SE loss function, provided that the mentioned expectation exists and is finite. The SE loss function, as a symmetric function, allocates equivalent losses to the overestimation and underestimation. However, in some practical situations, overestimation and underestimation are not of the same importance, and the use of symmetric loss functions seems inappropriate. [Varian \(1975\)](#) proposed an asymmetric loss function, called the linear-exponential (LE or *linex*) loss function, which is defined as

$$L(\theta, \delta(\mathbf{X})) = b^* \left[e^{c(\delta(\mathbf{X}) - \theta)} - c(\delta(\mathbf{X}) - \theta) - 1 \right], \quad c \neq 0, \quad b^* > 0,$$

where b^* and c are the parameters of the function. Without loss of generality, we can assume $b^* = 1$ whereas c has to be determined carefully. Positive values of c are considered when the overestimation is more serious than underestimation, while the negative values are considered when the underestimation is more serious than overestimation (see e.g. [Zellner, 1986](#)). The Bayes estimator of θ under the LE loss function is given by

$$\delta_{LE}(\mathbf{X}) = \frac{-1}{c} \ln E(e^{-c\theta}|\mathbf{X}), \quad c \neq 0.$$

provided that the above expectation exists and is finite.

Another asymmetric loss function, proposed by [Calabria and Pulcini \(1994\)](#), is the general entropy (GE) loss function, which is defined as

$$L(\theta, \delta(\mathbf{X})) = w \left[\left(\frac{\delta(\mathbf{X})}{\theta} \right)^p - p \ln \left(\frac{\delta(\mathbf{X})}{\theta} \right) - 1 \right], \quad p \neq 0, \quad w > 0.$$

Without loss of generality, we assume $w = 1$. The Bayes estimator of θ under the GE loss function is given by

$$\delta_{GE}(\mathbf{X}) = [E(\theta^{-p}|\mathbf{X})]^{-\frac{1}{p}}, \quad p \neq 0,$$

provided that the above expectation exists and is finite. Now, assume that θ has a gamma prior distribution with the following PDF :

$$(2.2) \quad \pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}; \quad a > 0, b > 0, \theta > 0.$$

From the likelihood function (2.1) and the prior distribution (2.2), the posterior density function can be obtained to be

$$(2.3) \quad \begin{aligned} \pi(\theta|\mathbf{x}) &= \frac{L(\theta|\mathbf{x})\pi(\theta)}{\int_0^\infty L(\theta|\mathbf{x})\pi(\theta)d\theta} \\ &= m(\mathbf{x})\theta^{2m+a-1}e^{-\theta(\frac{1}{x_m}+b)} \left\{ (1+\theta) \prod_{i=1}^{m-1} (\theta(1+x_i) + x_i) \right\}^{-1}, \end{aligned}$$

where

$$m(\mathbf{x}) = \frac{1}{\int_0^\infty \theta^{2m+a-1}e^{-\theta(\frac{1}{x_m}+b)} \left\{ (1+\theta) \prod_{i=1}^{m-1} (\theta(1+x_i) + x_i) \right\}^{-1} d\theta}.$$

2.5. Tierney and Kadane's approximation

This subsection presents the approximate Bayes estimates of θ under the SE , LE , and GE loss functions using the Tierney and Kadane's (TK) approximation method. Tierney and Kadane (1986) used Laplace's formula to approximate posterior moments. To apply the TK approximation method, suppose that $F(\theta) = \frac{1}{m} \ln \pi(\theta) + \frac{1}{m} l(\theta)$ and $F^*(\theta) = F(\theta) + \frac{1}{m} \ln g(\theta)$ where $l(\theta)$ is the log-likelihood function of θ , $\pi(\theta)$ is the prior density, and $g(\theta)$ should be a smooth positive function on the parameter space. We know that posterior moment of $g(\theta)$ is

$$(2.4) \quad E(g(\theta)|\mathbf{x}) = \int_0^\infty g(\theta) \cdot \pi(\theta|\mathbf{x}) d\theta.$$

The expression (2.4) can be rewritten as

$$(2.5) \quad E(g(\theta)|\mathbf{x}) = \frac{\int_0^\infty e^{mF^*(\theta)} d\theta}{\int_0^\infty e^{mF(\theta)} d\theta}.$$

Using the TK method, the approximate form of (2.5) becomes

$$\widehat{E}(g(\theta)|\mathbf{x}) = \left(\frac{\sigma^*}{\sigma}\right) \exp\left(m(F^*(\tilde{\theta}^*) - F(\tilde{\theta}))\right),$$

where $\tilde{\theta}$ and $\tilde{\theta}^*$ are the modes of $F(\theta)$ and $F^*(\theta)$, respectively and

$$\sigma^2 = -\frac{1}{F''(\theta)|_{\theta=\tilde{\theta}}} \quad \text{and} \quad \sigma^{*2} = -\frac{1}{F^{*''}(\theta)|_{\theta=\tilde{\theta}^*}},$$

where $F''(\cdot)$ and $F^{*''}(\cdot)$ denote the second order derivatives of $F(\theta)$ and $F^*(\theta)$, respectively.

Now, let

$$G(\theta, k_1, k_2) = \frac{1}{m} \left[(2m + a - 1 + k_1) \ln(\theta) - \ln(1 + \theta) - \theta \left(b + k_2 + \frac{1}{x_m} \right) - \sum_{i=1}^{m-1} \ln(\theta(1 + x_i) + x_i) + B(\mathbf{x}) \right],$$

where $B(\mathbf{x}) = \sum_{i=1}^m \ln(1 + x_i) - \ln x_m - 2 \sum_{i=1}^m \ln x_i + a \ln b - \ln \Gamma(a)$ and k_1 and k_2 are real numbers. Then, $F(\theta) = G(\theta, 0, 0)$ and $F^*(\theta) = G(\theta, k_1^*, k_2^*)$, where

$$(2.6) \quad k_1^* = \begin{cases} 1 & \text{under } SE, \\ 0 & \text{under } LE, \\ -p & \text{under } GE, \end{cases} \quad \text{and} \quad k_2^* = \begin{cases} 0 & \text{under } SE \text{ and } GE, \\ c & \text{under } LE. \end{cases}$$

Let $G^*(\theta, k_1, k_2) = \frac{\partial G(\theta, k_1, k_2)}{\partial \theta}$. Then, we have

$$G^*(\theta, k_1, k_2) = \frac{1}{m} \left[\frac{2m + a - 1 + k_1}{\theta} - \frac{1}{1 + \theta} - \left(b + k_2 + \frac{1}{x_m} \right) - \sum_{i=1}^{m-1} \frac{1 + x_i}{\theta(1 + x_i) + x_i} \right].$$

Note that $\frac{\partial^2 G(\theta, k_1, k_2)}{\partial \theta^2}$ is free of k_2 , so we let $G^{**}(\theta, k_1) = \frac{\partial^2 G(\theta, k_1, k_2)}{\partial \theta^2}$ and we have

$$G^{**}(\theta, k_1) = \frac{1}{m} \left[-\frac{2m + a - 1 + k_1}{\theta^2} + \frac{1}{(1 + \theta)^2} + \sum_{i=1}^{m-1} \left(\frac{1 + x_i}{\theta(1 + x_i) + x_i} \right)^2 \right].$$

Let $F'(\cdot)$ and $F^{*\prime}(\cdot)$ denote the first order derivatives of $F(\theta)$ and $F^*(\theta)$, respectively. Then, $F'(\theta) = G^*(\theta, 0, 0)$ and $F^{*\prime}(\theta) = G^*(\theta, k_1^*, k_2^*)$, where k_1^* and k_2^* are given in (2.6).

Moreover, $F''(\theta) = G^{**}(\theta, 0)$ and $F^{**\prime}(\theta) = G^{**}(\theta, k_1^*)$. Consequently, we get

$$\widehat{E}(g(\theta)|\mathbf{x}) = \sqrt{\frac{F''(\theta)|_{\theta=\tilde{\theta}}}{F^{**}(\theta)|_{\theta=\tilde{\theta}^*}}} \exp\left(m(F^*(\tilde{\theta}^*) - F(\tilde{\theta}))\right),$$

where $\tilde{\theta}$ and $\tilde{\theta}^*$ can be derived from $F'(\theta) = 0$ and $F^{*\prime}(\theta) = 0$, respectively, and

$$g(\theta) = \begin{cases} \theta & \text{under } SE, \\ \exp(-c\theta) & \text{under } LE, \\ \theta^{-p} & \text{under } GE. \end{cases}$$

Therefore, the approximate Bayes estimates of θ under the SE , LE and GE loss functions are given by

$$\begin{aligned} \tilde{\theta}_{SE} &= \widehat{E}(g_{SE}(\theta)|\mathbf{x}) \\ \tilde{\theta}_{LE} &= -\frac{1}{c} \ln [\widehat{E}(g_{LE}(\theta)|\mathbf{x})], \quad c \neq 0 \\ \tilde{\theta}_{GE} &= [\widehat{E}(g_{GE}(\theta)|\mathbf{x})]^{-\frac{1}{p}}, \quad p \neq 0, \end{aligned}$$

respectively, where $g_{SE}(\theta) = \theta$, $g_{LE}(\theta) = e^{-c\theta}$, and $g_{GE}(\theta) = \theta^{-p}$.

2.6. MCMC methods

In this subsection, we consider two Markov chain Monte Carlo (*MCMC*) methods to generate samples from the posterior distribution and then compute the approximate Bayes estimates of the parameter θ under the SE , LE , and GE loss functions. Two important subclasses of *MCMC* methods, which are considered here, are importance sampling (*IS*) and Metropolis–Hastings (*MH*) methods (see [Metropolis et al., 1953](#), and [Hastings, 1970](#), for the details of the *MH* algorithm).

To implement the *IS* procedure, we rewrite the posterior density function (2.3) as follows:

$$\pi(\theta|\mathbf{x}) = C(\mathbf{x}) \text{gamma}\left(\theta; 2m + a, \frac{1}{x_m} + b\right)h(\theta),$$

where $\text{gamma}(\theta; 2m + a, \frac{1}{x_m} + b)$ is the density of the gamma distribution with shape and rate parameters $2m + a$ and $\frac{1}{x_m} + b$, respectively, $C(\mathbf{x}) = \frac{m(\mathbf{x})\Gamma(2m+a)}{(x_m^{-1}+b)^{2m+a}}$ and

$$h(\theta) = \left\{ (1 + \theta) \prod_{i=1}^{m-1} (\theta(1 + x_i) + x_i) \right\}^{-1}.$$

Now, let $G(\theta|\mathbf{x}) = \text{gamma}(\theta; 2m + a, \frac{1}{x_m} + b)h(\theta)$. Then the Bayes estimate of θ under the SE loss function is given by

$$(2.7) \quad \widehat{\theta}_{SE} = \frac{\int_0^\infty \theta G(\theta|\mathbf{x})d\theta}{\int_0^\infty G(\theta|\mathbf{x})d\theta}.$$

Consider the following algorithm.

Algorithm 2

Step 1: Generate θ from the gamma distribution with shape and rate parameters respectively as $2m + a$ and $\frac{1}{x_m} + b$.

Step 2: Repeat Step 1, N times to obtain the importance sample $\theta_1, \theta_2, \dots, \theta_N$.

The approximate value of (2.7), which is the approximate Bayes estimate of θ under the SE loss function, can be obtained as

$$\hat{\theta}_{SE} = \frac{\sum_{i=1}^N \theta_i h(\theta_i)}{\sum_{i=1}^N h(\theta_i)} = \sum_{i=1}^N \theta_i w_i,$$

where $w_i = \frac{h(\theta_i)}{\sum_{i=1}^N h(\theta_i)}$. Besides, the approximate Bayes estimates of θ under the LE and GE loss functions are given by

$$\hat{\theta}_{LE} = -\frac{1}{c} \ln \left(\sum_{i=1}^N e_i^{-c\theta} w_i \right), \quad \text{and} \quad \hat{\theta}_{GE} = \left(\sum_{i=1}^N \theta_i^{-p} w_i \right)^{-\frac{1}{p}},$$

respectively.

In the sequel, we use the MH algorithm to approximate the Bayes estimates of the parameter of the ILD . Here, we consider the normal distribution as a symmetric proposal distribution. According to [Dey and Pradhan \(2014\)](#), we write the MH algorithm steps as follows:

Algorithm 3

Step 1: Set an initial value $\theta^{(0)}$, we propose to consider the ML estimate of θ as the initial value.

Step 2: For $j = 1, \dots, N'$, repeat the following steps:

- Set $\theta = \theta^{(j-1)}$.
- Following [Dey and Pradhan \(2014\)](#), generate a new candidate parameter value δ from $N(\ln(\theta), \frac{S_{\theta_0}}{[\theta^{(0)}]^2})$, where S_{θ_0} can be obtained using the inverse of the observed Fisher information as follows:

$$S_{\theta_0} = \left\{ \frac{2m}{\theta^2} - \frac{1}{(1+\theta)^2} - \sum_{i=1}^{m-1} \left(\frac{1+x_i}{\theta(1+x_i)+x_i} \right)^2 \right\}^{-1} \Bigg|_{\theta=\theta^{(0)}}.$$

- Set $\theta' = \exp(\delta)$.
 - Calculate $P = \min \left\{ 1, \frac{\pi(\theta'|\mathbf{x})q(\theta|\theta')}{\pi(\theta|\mathbf{x})q(\theta'|\theta)} \right\}$, where $q(x|b)$ is the density of the log-normal distribution with parameters $\ln(b)$ and $\frac{S_{\theta_0}}{[\theta^{(0)}]^2}$.
 - Update $\theta^{(j)} = \theta'$ with probability P , otherwise set $\theta^{(j)} = \theta$.
-

We may discard the first k generated data, where k is the burn-in period. Suppose $\{\theta_l, l = 1, \dots, M\}$ is a sample produced according to Algorithm 3 with $M = N' - k$. Therefore, the approximate Bayes estimates of θ under the SE , LE and GE loss functions are given by

$$\tilde{\theta}_{SE}^* = \frac{1}{M} \sum_{l=1}^M \theta_l, \quad \text{and} \quad \tilde{\theta}_{LE}^* = -\frac{1}{c} \ln \left(\frac{1}{M} \sum_{l=1}^M e^{-c\theta_l} \right),$$

and

$$\tilde{\theta}_{GE}^* = \left(\frac{1}{M} \sum_{l=1}^M \theta_l^{-p} \right)^{-\frac{1}{p}},$$

respectively.

3. PREDICTION of a FUTURE RECORD VALUE

Prediction of future records has been studied by many authors (see for example [Dunsmore, 1983](#); [Berred, 1998](#); and [Ahmadi and Doostparast, 2006](#)). In this section, we study the problem of predicting a future record value, given a sample of observed past record values.

3.1. Maximum likelihood prediction

Suppose that the first m lower record values $X_{L(1)}, \dots, X_{L(m)}$ are available from a population with PDF $f(x; \theta)$ and CDF $F(x; \theta)$. Let $Z = X_{L(n)}$, $n > m$, is an unobserved future record value. Then, the joint PDF of Z and $X_{L(1)}, \dots, X_{L(m)}$ is given by [Basak and Balakrishnan \(2003\)](#), which can also be obtained from the Markovian property of records (see e.g. [Arnold et al., 1998](#)). Here, using the result given by [Basak and Balakrishnan \(2003\)](#) and from (1.1) and (1.2), the logarithm of the predictive likelihood function of the parameter and Z for the $ILLD$ is given by

$$\begin{aligned} \ln L(z, \theta; \mathbf{x}) = & (2m + 2) \ln(\theta) - \ln(1 + \theta) - \ln \Gamma(n - m) + \ln(z + 1) - 3 \ln(z) \\ & - \frac{\theta}{z} + (n - m - 1) \ln \left(\frac{\theta}{z} - \frac{\theta}{x_m} + \ln \left(\frac{z[(1 + \theta)x_m + \theta]}{x_m[(1 + \theta)z + \theta]} \right) \right) \\ (3.1) \quad & - \sum_{i=1}^m \ln(\theta(1 + x_i) + x_i) + \sum_{i=1}^m \ln(1 + x_i) - 2 \sum_{i=1}^m \ln(x_i), \quad z < x_m. \end{aligned}$$

Maximizing (3.1) with respect to θ and z , we could find the ML prediction of Z and the predictive maximum likelihood estimate of θ . Upon differentiating (3.1) partially with respect

to θ and z and equating the results with zero, we have the following equations:

$$\begin{aligned}
 \frac{\partial \ln L(z, \theta; \mathbf{x})}{\partial \theta} &= \frac{2m+2}{\theta} - \frac{1}{1+\theta} - \frac{1}{z} - \sum_{i=1}^m \frac{1+x_i}{\theta(1+x_i)+x_i} \\
 &\quad + \frac{(n-m-1) \left(\frac{1}{z} - \frac{1}{x_m} + \frac{x_m+1}{(1+\theta)x_m+\theta} - \frac{z+1}{(1+\theta)z+\theta} \right)}{\frac{\theta}{z} - \frac{\theta}{x_m} + \ln \left(\frac{z[(1+\theta)x_m+\theta]}{x_m[(1+\theta)z+\theta]} \right)} = 0, \\
 \frac{\partial \ln L(z, \theta; \mathbf{x})}{\partial z} &= \frac{1}{1+z} - \frac{3}{z} + \frac{\theta}{z^2} \\
 &\quad + \frac{(n-m-1) \left(\frac{1}{z} - \frac{\theta}{z^2} - \frac{\theta+1}{(1+\theta)z+\theta} \right)}{\frac{\theta}{z} - \frac{\theta}{x_m} + \ln \left(\frac{z[(1+\theta)x_m+\theta]}{x_m[(1+\theta)z+\theta]} \right)} = 0.
 \end{aligned}
 \tag{3.2}$$

A numerical procedure can help us to find the solutions of the above equations.

One may also find the approximate *ML* (*AML*) prediction of Z by means of solving (3.2) after replacing θ with its *ML* estimate.

3.2. Interval prediction

In this subsection, we study the problem of interval prediction of a future record based on observed past lower record values coming from the *ILD*. Shortest and equal tails intervals have been nicely discussed in [Ferentinos and Karakostas \(2006\)](#). As mentioned earlier, record values satisfy the Markovian property (see e.g. [Arnold et al., 1998](#)), in the sense that the conditional density of $Z = X_{L(n)}$ ($n > m \geq 1$) given the set of the first m lower records $(X_{L(1)}, \dots, X_{L(m)}) = (x_1, \dots, x_m)$ is the same as the conditional density of Z given $X_{L(m)} = x_m$. From (1.1) and (1.2), the conditional *PDF* of Z given x_m for the *ILD* becomes

$$f_Z(z|x_m; \theta) = \frac{\theta^2(1+z)x_m \left(\frac{\theta}{z} - \frac{\theta}{x_m} + \ln \left(\frac{z[(1+\theta)x_m+\theta]}{x_m[(1+\theta)z+\theta]} \right) \right)^{n-m-1}}{(n-m-1)![\theta(1+x_m)+x_m]z^3} e^{-\theta(\frac{1}{z} - \frac{1}{x_m})},
 \tag{3.3}$$

where $z < x_m$.

As a consequence of (3.3), it can be proved that (see Appendix)

$$U = \frac{\theta}{Z} - \frac{\theta}{x_m} + \ln \left(\frac{Z[(1+\theta)x_m+\theta]}{x_m[(1+\theta)Z+\theta]} \right) \Big|_{X_{L(m)} = x_m} \sim \text{Gamma}(n-m, 1),
 \tag{3.4}$$

with the following density:

$$g_U(z) = \frac{1}{\Gamma(n-m)} z^{n-m-1} e^{-z}, \quad z > 0.$$

Then, the highest conditional density (*HCD*) interval for U at the level of $(1-\alpha)$ is in the form of $[c_1, c_2]$ if

$$[c_1, c_2] = \{c : c \geq 0, g_U(c) \geq k\},$$

for some $k > 0$, where $\int_{c_1}^{c_2} g_U(c) dc = 1 - \alpha$.

If $n > m + 1$, then $g_U(c)$ is a unimodal *PDF* whose maximum value is achieved at $v = n - m - 1 > 0$. In this case, c_1 and c_2 are the solutions of the following non-linear equations (see e.g. Casella and Berger, 2002):

$$\int_{c_1}^{c_2} g_U(c)dc = 1 - \alpha, \quad \text{and} \quad g_U(c_1) = g_U(c_2).$$

The above equations can be reexpressed as follows:

$$\gamma(c_2, n - m) - \gamma(c_1, n - m) = 1 - \alpha, \quad \text{and} \quad \frac{c_1}{c_2} = \exp\left(-\frac{c_2 - c_1}{n - m - 1}\right),$$

where $\gamma(c, a) = \frac{1}{\Gamma(a)} \int_0^c x^{a-1} e^{-x} dx$ is the incomplete gamma function.

Thus, a $100(1 - \alpha)\%$ prediction interval (*PI*) of Z based on the above (*HCD*) method is in the form of (L^*, U^*) , where L^* and U^* satisfy the following non-linear equations:

$$\frac{L^*(x_m(1 + \theta) + \theta)}{x_m(L^*(1 + \theta) + \theta)} = \exp\left(-\frac{\theta}{L^*} + \frac{\theta}{x_m} + c_2\right),$$

and

$$\frac{U^*(x_m(1 + \theta) + \theta)}{x_m(U^*(1 + \theta) + \theta)} = \exp\left(-\frac{\theta}{U^*} + \frac{\theta}{x_m} + c_1\right),$$

respectively. If θ is unknown, then it can be replaced by its *MLE*, which leads to a $100(1 - \alpha)\%$ approximate *PI* (*API*) for Z .

Next, we consider the case when $n = m + 1$, where $g_U(c)$ is a decreasing function with $g_U(0) = 1$ and $g_U(\infty) = 0$. So, we find the interval of the form $[0, c_1]$ where c_1 satisfies the following equation:

$$\int_0^{c_1} g_U(c)dc = 1 - \alpha.$$

Therefore, $c_1 = -\ln \alpha$ and a $100(1 - \alpha)\%$ *PI* for Z will be in the form of (L^*, x_m) , where L^* satisfies the following equation:

$$\frac{\alpha L^*(x_m(1 + \theta) + \theta)}{x_m(L^*(1 + \theta) + \theta)} = \exp\left(-\frac{\theta}{L^*} + \frac{\theta}{x_m}\right).$$

3.3. Bayesian prediction

In this subsection, we consider the prediction of a future record based on a Bayesian approach under the *SE*, *LE*, and *GE* loss functions. Suppose that the first m lower records $X_{L(1)}, \dots, X_{L(m)}$ are available from the *ILD* and we wish to predict the n -th lower record $Z = X_{L(n)}$, $n > m$, based on the observed vector \mathbf{x} . From (2.3) and (3.3), the Bayes predictive density of Z given \mathbf{x} is given by

$$\begin{aligned} f_Z(z|\mathbf{x}) &= \int_0^\infty f_Z(z|x_m; \theta) \pi(\theta|\mathbf{x}) d\theta \\ &= \frac{(1+z)x_m m(\mathbf{x})}{z^3 \Gamma(n-m)} \int_0^\infty \left(\frac{\theta}{z} - \frac{\theta}{x_m} + \ln \left(\frac{z[(1+\theta)x_m + \theta]}{x_m[(1+\theta)z + \theta]} \right) \right)^{n-m-1} \\ &\quad \times \frac{\theta^{2m+a+1}}{x_m + \theta(1+x_m)} e^{-\theta(\frac{1}{z}+b)} \left[(1+\theta) \prod_{i=1}^{m-1} (x_i + \theta(1+x_i)) \right]^{-1} d\theta. \end{aligned}$$

In the particular case of $n = m + 1$, the Bayes predictive density function of Z simplifies as

$$f_Z(z|\mathbf{x}) = \frac{(1+z)x_m m(\mathbf{x})}{z^3} \int_0^\infty \frac{\theta^{2m+a+1} e^{-\theta(\frac{1}{z}+b)}}{x_m + \theta(1+x_m)} \left[(1+\theta) \prod_{i=1}^{m-1} (x_i + \theta(1+x_i)) \right]^{-1} d\theta.$$

The Bayesian prediction of the n -th lower record under the SE loss function is given by

$$\hat{Z}_{BS} = \hat{E}(Z|\mathbf{x}) = \int_0^{x_m} z f_Z(z|\mathbf{x}) dz,$$

and the Bayesian predictions of Z under the LE and GE loss functions are

$$\hat{Z}_{BL} = -\frac{1}{c} \ln \hat{E}(e^{-cZ}|\mathbf{x}) = -\frac{1}{c} \ln \left(\int_0^{x_m} e^{-cz} f_Z(z|\mathbf{x}) dz \right),$$

and

$$\hat{Z}_{BG} = [\hat{E}(Z^{-p}|\mathbf{x})]^{-\frac{1}{p}} = \left[\int_0^{x_m} z^{-p} f_Z(z|\mathbf{x}) dz \right]^{-\frac{1}{p}},$$

respectively, provided that the above integrals exist and are finite.

The predictive limits of a $100(1 - \tau)\%$ two-sided PI for the future lower record $Z = X_{L(n)}$ can be obtained by solving the following two equations simultaneously with respect to L^{**} and U^{**} :

$$\int_{L^{**}}^\infty f_Z(z|\mathbf{x}) dz = 1 - \frac{\tau}{2}, \quad \text{and} \quad \int_{U^{**}}^\infty f_Z(z|\mathbf{x}) dz = \frac{\tau}{2}.$$

4. A SIMULATION STUDY

In this section, we performed a simulation study to assess the performance of the point and interval estimators of θ and predictors of a future record value coming from the ILD . With this in mind, in each iteration of the simulation, we generate m lower records from the ILD with parameter θ and then we compute the ML estimate, the approximate Bayes estimates under the SE , LE , and GE loss functions using the TK , IS , and MH methods. The 95% asymptotic CIs , as well as the two bootstrap-type CIs are obtained. In the context of prediction, we compute AML prediction and 95% PI (based on the HCD method) for the $(m + 1)$ -th lower record value. The following setting has been applied: We consider three different values for the number of lower records as $m = 3, 4, 5$ and three different values for the parameter as $\theta = 0.5, 1, 2$. The number of bootstrap repetitions is taken to be $B = 1000$. In the context of the Bayesian estimation, two gamma priors have been applied, Prior 1 with $(a_1, b_1) = (0.2, 1.5)$ and Prior 2 with $(a_2, b_2) = (3, 1)$. Besides, we take $c = -0.2, 0.2$ for the LE loss function and $p = -0.2, 0.2$ for the GE loss function. The results of the simulation study are based on $N = 1000$ iterations.

The assessment of the performances of the point estimators is based on estimated risks (ERs) under the SE , LE , and GE functions, and the evaluation of CIs is based on average length (AL) and coverage probability (CP). Let $\tilde{\theta}$ be an estimator of θ and $\tilde{\theta}_i$ be

the corresponding estimate obtained in the i -th iteration. Then the estimated bias (bias for short) and ER s of $\tilde{\theta}$ under the SE , LE , and GE loss functions are given by

$$(4.1) \quad \text{Bias}(\tilde{\theta}) = \frac{1}{N} \sum_{i=1}^N (\tilde{\theta}_i - \theta),$$

$$(4.2) \quad ER_{SE}(\tilde{\theta}) = \frac{1}{N} \sum_{i=1}^N (\tilde{\theta}_i - \theta)^2,$$

$$ER_{LE}(\tilde{\theta}) = \frac{1}{N} \sum_{i=1}^N \left[e^{c(\tilde{\theta}_i - \theta)} - c(\tilde{\theta}_i - \theta) - 1 \right],$$

$$ER_{GE}(\tilde{\theta}) = \frac{1}{N} \sum_{i=1}^N \left[\left(\frac{\tilde{\theta}_i}{\theta} \right)^p - p \ln \left(\frac{\tilde{\theta}_i}{\theta} \right) - 1 \right],$$

respectively.

Besides, we compute the empirical biases (biases for short) and mean squared prediction errors ($EMSP$ Es) of the AML predictors (which can be formulated similarly as (4.1) and (4.2), respectively) and the AL s and CP s of the interval predictors.

The simulation results related to the point estimation are presented in Tables 1–6. The following abbreviations are used in Tables 1–6: BS (Bayes estimator under the SE loss function), BLc_1 (Bayes estimator under the SE loss function with $c_1 = 0.2$), BLc_2 (Bayes estimator under the SE loss function with $c_2 = -0.2$), BGp_1 (Bayes estimator under the GE loss function with $p_1 = 0.2$) and BGp_2 (Bayes estimator under the GE loss function with $p_2 = -0.2$). It is observed from Tables 1–6 that in all estimation methods, ER s are decreasing with respect to the number of records except for the case under Prior 2 when $\theta = 2$. We also observe that the ER s are close to each other for the TK , IS , and MH methods. Furthermore, for Prior 1, the ER s of the Bayes estimators are less than or equal to those of the ML estimators (a few exceptions exist), whereas for Prior 2 when $\theta = 0.5$, the ML estimators outperform the Bayes estimators in the sense of ER and bias. Prior 1 produces smaller ER s than Prior 2, when $\theta = 0.5$ and 1, which is also true for $\theta = 2$ in the most cases.

The performances of the asymptotic CI s and two different bootstrap CI s ($Boot-B$ and $Boot-P$ methods) are compared in terms of their AL s and CP s in Table 7. Table 7 shows that in all three methods, the AL of the CI decreases as the number of records increases. Besides, in all cases, the CP s of the asymptotic CI s are more than the corresponding CP s of the bootstrap CI s., and the AL s of the asymptotic CI s are less than those of the others. We also observe that the $Boot-B$ CI s perform better than the $Boot-P$ CI s in the sense of CP .

Finally, Table 8 presents the biases and $EMSP$ Es of the AML predictors as well as the AL s and CP s of the API s for the $(m + 1)$ -th lower record value. From Table 8, we observe that for all values of θ , the AL , bias, and $EMSP$ E decrease as the number of records increases.

Table 1: Estimated biases and ERs of point estimators of θ for Prior 1 = (0.2, 1.5) and $\theta = 0.5$.

m	ER	Method																	
		ML			TK						IS						MH		
		BS	BLc_1	BLc_2	BS	BLc_1	BLc_2	BGp_1	BGp_2	BS	BLc_1	BLc_2	BGp_1	BGp_2	BS	BLc_1	BLc_2	BGp_1	BGp_2
3	Bias	0.175	0.064	0.052	0.067	0.067	0.001	0.022	0.068	0.053	0.068	0.068	-0.003	0.018	0.060	0.053	0.068	-0.003	0.018
	ER_{SE}	0.218	0.063	0.057	0.066	0.046	0.046	0.051	0.066	0.057	0.066	0.045	0.050	0.064	0.060	0.069	0.047	0.052	
	$ER_{LE} (c = 0.2)$	0.005	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	
	$ER_{LE} (c = -0.2)$	0.004	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	
	$ER_{GE} (p = 0.2)$	0.006	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	
	$ER_{GE} (p = -0.2)$	0.005	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	
4	Bias	0.132	0.062	0.054	0.065	0.011	0.028	0.065	0.054	0.065	0.010	0.026	0.060	0.060	0.054	0.065	0.009	0.026	
	ER_{SE}	0.097	0.044	0.041	0.046	0.033	0.036	0.046	0.041	0.046	0.033	0.036	0.044	0.044	0.042	0.047	0.033	0.036	
	$ER_{LE} (c = 0.2)$	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	
	$ER_{LE} (c = -0.2)$	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	
	$ER_{GE} (p = 0.2)$	0.004	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	
	$ER_{GE} (p = -0.2)$	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	
5	Bias	0.124	0.070	0.063	0.072	0.027	0.041	0.072	0.063	0.072	0.026	0.040	0.068	0.068	0.064	0.073	0.026	0.040	
	ER_{SE}	0.060	0.031	0.029	0.032	0.023	0.026	0.032	0.029	0.032	0.023	0.026	0.031	0.031	0.030	0.033	0.023	0.026	
	$ER_{LE} (c = 0.2)$	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	
	$ER_{LE} (c = -0.2)$	0.001	0.001	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	
	$ER_{GE} (p = 0.2)$	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.002	0.002	0.002	0.001	0.001	0.002	0.002	0.002	0.002	0.001	
	$ER_{GE} (p = -0.2)$	0.002	0.002	0.001	0.002	0.001	0.001	0.001	0.002	0.002	0.002	0.001	0.001	0.002	0.002	0.002	0.002	0.001	

Table 3: Estimated biases and ERs of point estimators of θ for Prior 1 = (0.2, 1.5) and $\theta = 2$.

m	ER	Method																	
		ML			TK						IS						MH		
		BS	BLc_1	BLc_2	BLc_1	BLc_2	BGp_1	BGp_2	BS	BLc_1	BLc_2	BGp_1	BGp_2	BS	BLc_1	BLc_2	BGp_1	BGp_2	
3	Bias	0.801	-0.588	-0.642	-0.556	-0.751	-0.697	-0.600	-0.640	-0.555	-0.763	-0.709	-0.599	-0.64	-0.555	-0.762	-0.708		
	ER_{SE}	4.919	0.470	0.520	0.449	0.661	0.591	0.482	0.518	0.447	0.677	0.605	0.486	0.522	0.453	0.680	0.609		
	$ER_{LE} (c = 0.2)$	0.341	0.009	0.010	0.008	0.012	0.011	0.009	0.010	0.008	0.013	0.011	0.009	0.010	0.009	0.013	0.011		
	$ER_{LE} (c = -0.2)$	0.061	0.010	0.011	0.009	0.014	0.013	0.010	0.011	0.009	0.014	0.013	0.010	0.011	0.010	0.015	0.013		
	$ER_{GE} (p = 0.2)$	0.006	0.004	0.004	0.004	0.006	0.005	0.004	0.004	0.004	0.006	0.005	0.004	0.004	0.004	0.006	0.005		
	$ER_{GE} (p = -0.2)$	0.006	0.004	0.005	0.004	0.007	0.006	0.004	0.005	0.004	0.007	0.006	0.004	0.005	0.004	0.007	0.006		
4	Bias	0.612	-0.449	-0.497	-0.415	-0.593	-0.545	-0.456	-0.495	-0.414	-0.599	-0.551	-0.457	-0.495	-0.415	-0.599	-0.552		
	ER_{SE}	2.344	0.330	0.360	0.314	0.455	0.409	0.339	0.363	0.317	0.469	0.420	0.339	0.363	0.319	0.466	0.419		
	$ER_{LE} (c = 0.2)$	0.082	0.006	0.007	0.006	0.009	0.008	0.006	0.007	0.006	0.009	0.008	0.006	0.007	0.006	0.009	0.008		
	$ER_{LE} (c = -0.2)$	0.034	0.007	0.008	0.007	0.010	0.009	0.007	0.008	0.007	0.010	0.009	0.007	0.008	0.007	0.010	0.009		
	$ER_{GE} (p = 0.2)$	0.004	0.002	0.003	0.002	0.004	0.003	0.003	0.003	0.003	0.004	0.003	0.003	0.003	0.003	0.004	0.003		
	$ER_{GE} (p = -0.2)$	0.004	0.003	0.003	0.002	0.004	0.003	0.003	0.003	0.002	0.004	0.004	0.003	0.003	0.003	0.004	0.004		
5	Bias	0.719	-0.280	-0.326	-0.243	-0.414	-0.369	-0.283	-0.323	-0.240	-0.414	-0.371	-0.288	-0.328	-0.246	-0.420	-0.376		
	ER_{SE}	1.971	0.220	0.233	0.215	0.290	0.262	0.224	0.235	0.218	0.297	0.269	0.229	0.239	0.223	0.300	0.272		
	$ER_{LE} (c = 0.2)$	0.057	0.004	0.005	0.004	0.006	0.005	0.004	0.004	0.005	0.006	0.005	0.004	0.005	0.004	0.006	0.005		
	$ER_{LE} (c = -0.2)$	0.031	0.005	0.005	0.004	0.006	0.005	0.005	0.005	0.004	0.006	0.006	0.005	0.005	0.005	0.006	0.006		
	$ER_{GE} (p = 0.2)$	0.004	0.001	0.002	0.001	0.002	0.002	0.001	0.001	0.002	0.001	0.002	0.002	0.002	0.001	0.002	0.002		
	$ER_{GE} (p = -0.2)$	0.003	0.002	0.002	0.001	0.002	0.002	0.002	0.002	0.002	0.001	0.002	0.002	0.002	0.001	0.002	0.002		

Table 4: Estimated biases and ERs of point estimators of θ for Prior 2 = (3, 1) and $\theta = 0.5$.

m	ER	Method														
		ML			TK			IS			MH					
		BS	BLc_1	BLc_2	BS	BLc_1	BLc_2	BS	BLc_1	BLc_2	BS	BLc_1	BLc_2	BS	BLc_1	BLc_2
3	Bias	0.175	0.442	0.425	0.454	0.368	0.393	0.44	0.426	0.455	0.367	0.391	0.439	0.425	0.454	0.366
	ER_{SE}	0.218	0.404	0.373	0.432	0.312	0.341	0.401	0.373	0.432	0.308	0.338	0.403	0.376	0.434	0.312
	$ER_{LE} (c = 0.2)$	0.005	0.009	0.008	0.010	0.007	0.007	0.009	0.008	0.010	0.007	0.007	0.009	0.008	0.010	0.007
	$ER_{LE} (c = -0.2)$	0.004	0.007	0.007	0.008	0.006	0.006	0.007	0.007	0.008	0.006	0.006	0.007	0.007	0.008	0.006
	$ER_{GE} (p = 0.2)$	0.006	0.010	0.010	0.010	0.008	0.009	0.010	0.010	0.010	0.008	0.009	0.010	0.010	0.010	0.008
	$ER_{GE} (p = -0.2)$	0.005	0.009	0.008	0.009	0.007	0.008	0.009	0.008	0.009	0.007	0.008	0.009	0.008	0.009	0.007
4	Bias	0.132	0.351	0.341	0.359	0.295	0.314	0.351	0.342	0.361	0.295	0.314	0.350	0.340	0.359	0.293
	ER_{SE}	0.097	0.232	0.218	0.242	0.181	0.197	0.231	0.219	0.243	0.180	0.196	0.231	0.219	0.244	0.180
	$ER_{LE} (c = 0.2)$	0.002	0.005	0.005	0.005	0.004	0.004	0.005	0.005	0.005	0.004	0.004	0.005	0.005	0.005	0.004
	$ER_{LE} (c = -0.2)$	0.002	0.004	0.004	0.005	0.003	0.004	0.004	0.004	0.005	0.003	0.004	0.004	0.004	0.005	0.003
	$ER_{GE} (p = 0.2)$	0.004	0.007	0.007	0.007	0.006	0.006	0.007	0.007	0.007	0.006	0.006	0.007	0.007	0.007	0.006
	$ER_{GE} (p = -0.2)$	0.003	0.006	0.006	0.007	0.005	0.006	0.006	0.006	0.007	0.005	0.006	0.006	0.006	0.007	0.005
5	Bias	0.124	0.311	0.303	0.317	0.264	0.279	0.310	0.303	0.317	0.263	0.279	0.309	0.302	0.317	0.262
	ER_{SE}	0.060	0.157	0.150	0.164	0.123	0.134	0.158	0.151	0.165	0.124	0.135	0.158	0.151	0.165	0.124
	$ER_{LE} (c = 0.2)$	0.001	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
	$ER_{LE} (c = -0.2)$	0.001	0.003	0.003	0.003	0.002	0.003	0.003	0.003	0.003	0.002	0.003	0.003	0.003	0.003	0.002
	$ER_{GE} (p = 0.2)$	0.002	0.006	0.006	0.006	0.005	0.005	0.006	0.006	0.006	0.005	0.005	0.006	0.006	0.006	0.005
	$ER_{GE} (p = -0.2)$	0.002	0.005	0.005	0.005	0.004	0.005	0.005	0.005	0.005	0.004	0.005	0.005	0.005	0.005	0.004

Table 5: Estimated biases and *ERs* of point estimators of θ for Prior 2 = (3, 1) and $\theta = 1$.

<i>m</i>	<i>ER</i>	Method														
		<i>ML</i>			<i>TK</i>			<i>IS</i>			<i>MH</i>					
		<i>BS</i>	<i>BLc₁</i>	<i>BLc₂</i>	<i>BS</i>	<i>BLc₁</i>	<i>BLc₂</i>	<i>BS</i>	<i>BLc₁</i>	<i>BLc₂</i>	<i>BS</i>	<i>BLc₁</i>	<i>BLc₂</i>	<i>BS</i>	<i>BLc₁</i>	<i>BLc₂</i>
3	Bias	0.348	0.646	0.601	0.687	0.516	0.559	0.645	0.604	0.689	0.515	0.558	0.639	0.598	0.684	0.552
	<i>ER_{SE}</i> (<i>c</i> = 0.2)	0.890	0.865	0.757	0.975	0.641	0.712	0.865	0.766	0.983	0.644	0.713	0.851	0.753	0.967	0.700
	<i>ER_{LE}</i> (<i>c</i> = -0.2)	0.023	0.020	0.017	0.022	0.014	0.016	0.020	0.017	0.023	0.014	0.016	0.019	0.017	0.022	0.016
	<i>ER_{LE}</i> (<i>c</i> = -0.2)	0.015	0.015	0.014	0.017	0.012	0.013	0.015	0.014	0.017	0.012	0.013	0.015	0.014	0.017	0.013
	<i>ER_{GE}</i> (<i>p</i> = 0.2)	0.006	0.007	0.006	0.007	0.005	0.006	0.007	0.006	0.007	0.005	0.006	0.007	0.006	0.007	0.006
	<i>ER_{GE}</i> (<i>p</i> = -0.2)	0.005	0.006	0.006	0.007	0.005	0.005	0.006	0.006	0.007	0.005	0.005	0.006	0.006	0.007	0.005
4	Bias	0.264	0.545	0.512	0.574	0.441	0.476	0.544	0.515	0.576	0.441	0.475	0.544	0.514	0.575	0.475
	<i>ER_{SE}</i>	0.428	0.570	0.511	0.629	0.430	0.474	0.574	0.519	0.636	0.437	0.480	0.573	0.517	0.636	0.477
	<i>ER_{LE}</i> (<i>c</i> = 0.2)	0.010	0.013	0.011	0.014	0.009	0.010	0.013	0.011	0.014	0.010	0.011	0.013	0.011	0.014	0.011
	<i>ER_{LE}</i> (<i>c</i> = -0.2)	0.007	0.010	0.009	0.011	0.008	0.009	0.010	0.010	0.012	0.008	0.009	0.010	0.009	0.012	0.009
	<i>ER_{GE}</i> (<i>p</i> = 0.2)	0.003	0.005	0.005	0.005	0.004	0.004	0.005	0.005	0.005	0.005	0.004	0.005	0.005	0.005	0.004
	<i>ER_{GE}</i> (<i>p</i> = -0.2)	0.003	0.005	0.004	0.005	0.004	0.004	0.005	0.004	0.005	0.004	0.004	0.005	0.004	0.005	0.004
5	Bias	0.269	0.518	0.491	0.541	0.428	0.458	0.518	0.494	0.544	0.430	0.459	0.516	0.491	0.541	0.457
	<i>ER_{SE}</i>	0.320	0.475	0.432	0.517	0.365	0.400	0.481	0.442	0.525	0.375	0.408	0.474	0.434	0.518	0.400
	<i>ER_{LE}</i> (<i>c</i> = 0.2)	0.007	0.010	0.009	0.011	0.008	0.009	0.011	0.010	0.012	0.008	0.009	0.010	0.009	0.011	0.009
	<i>ER_{LE}</i> (<i>c</i> = -0.2)	0.006	0.009	0.008	0.009	0.007	0.007	0.009	0.008	0.010	0.007	0.008	0.009	0.008	0.010	0.007
	<i>ER_{GE}</i> (<i>p</i> = 0.2)	0.003	0.005	0.004	0.005	0.004	0.004	0.005	0.004	0.005	0.004	0.004	0.005	0.004	0.005	0.004
	<i>ER_{GE}</i> (<i>p</i> = -0.2)	0.003	0.004	0.004	0.004	0.003	0.004	0.004	0.004	0.004	0.004	0.003	0.004	0.004	0.004	0.004

Table 6: Estimated biases and ERs of point estimators of θ for Prior 2 = (3, 1) and $\theta = 2$.

m	ER	Method														
		ML			TK			IS			MH					
		BS	BLc_1	BLc_2	BS	BLc_1	BLc_2	BS	BLc_1	BLc_2	BS	BLc_1	BLc_2	BS	BLc_1	BLc_2
3	Bias	0.801	0.668	0.558	0.777	0.452	0.524	0.664	0.562	0.780	0.449	0.520	0.662	0.56	0.778	0.446
	ER_{SE}	4.919	1.139	0.895	1.430	0.782	0.890	1.138	0.909	1.442	0.788	0.893	1.135	0.906	1.441	0.782
	$ER_{LE} (c = 0.2)$	0.341	0.026	0.020	0.033	0.017	0.020	0.026	0.020	0.034	0.018	0.020	0.026	0.020	0.034	0.017
	$ER_{LE} (c = -0.2)$	0.061	0.020	0.016	0.025	0.014	0.016	0.020	0.016	0.025	0.014	0.016	0.020	0.016	0.025	0.014
	$ER_{GE} (p = 0.2)$	0.006	0.003	0.003	0.004	0.002	0.003	0.003	0.003	0.003	0.004	0.002	0.003	0.003	0.003	0.004
	$ER_{GE} (p = -0.2)$	0.006	0.003	0.002	0.003	0.002	0.002	0.003	0.002	0.003	0.003	0.002	0.002	0.003	0.002	0.003
4	Bias	0.612	0.630	0.540	0.717	0.447	0.508	0.627	0.544	0.720	0.446	0.507	0.623	0.539	0.717	0.440
	ER_{SE}	2.344	0.941	0.762	1.149	0.668	0.751	0.944	0.776	1.159	0.678	0.758	0.939	0.768	1.156	0.665
	$ER_{LE} (c = 0.2)$	0.082	0.021	0.017	0.027	0.015	0.017	0.021	0.017	0.027	0.015	0.017	0.021	0.017	0.027	0.015
	$ER_{LE} (c = -0.2)$	0.034	0.017	0.014	0.020	0.012	0.014	0.017	0.014	0.020	0.012	0.014	0.017	0.014	0.020	0.012
	$ER_{GE} (p = 0.2)$	0.004	0.003	0.002	0.003	0.002	0.002	0.003	0.002	0.003	0.002	0.002	0.003	0.002	0.003	0.002
	$ER_{GE} (p = -0.2)$	0.004	0.002	0.002	0.003	0.002	0.002	0.002	0.002	0.002	0.003	0.002	0.002	0.002	0.002	0.002
5	Bias	0.719	0.725	0.642	0.807	0.559	0.615	0.723	0.645	0.808	0.559	0.613	0.722	0.643	0.808	0.556
	ER_{SE}	1.971	1.036	0.859	1.236	0.759	0.845	1.036	0.871	1.241	0.771	0.852	1.032	0.864	1.240	0.758
	$ER_{LE} (c = 0.2)$	0.057	0.024	0.019	0.029	0.017	0.019	0.024	0.020	0.029	0.017	0.019	0.024	0.019	0.029	0.017
	$ER_{LE} (c = -0.2)$	0.031	0.018	0.015	0.022	0.014	0.015	0.018	0.016	0.022	0.014	0.015	0.018	0.015	0.022	0.014
	$ER_{GE} (p = 0.2)$	0.004	0.003	0.002	0.003	0.002	0.002	0.003	0.002	0.003	0.002	0.002	0.003	0.002	0.003	0.002
	$ER_{GE} (p = -0.2)$	0.003	0.003	0.002	0.003	0.002	0.002	0.003	0.002	0.002	0.003	0.002	0.003	0.002	0.003	0.002

Table 7: ALs and CPs of 95% CIs of θ .

θ	m		Method		
			Asymptotic	Boot-B	Boot-P
0.5	3	AL	1.222	1.938	1.938
		CP	0.985	0.868	0.840
	4	AL	1.005	1.362	1.362
		CP	0.997	0.902	0.788
	5	AL	0.900	1.113	1.113
		CP	0.999	0.927	0.668
1	3	AL	2.517	4.328	4.328
		CP	0.984	0.863	0.838
	4	AL	2.062	2.956	2.956
		CP	0.999	0.916	0.787
	5	AL	1.872	2.443	2.443
		CP	0.998	0.929	0.678
2	3	AL	5.481	10.002	10.002
		CP	0.982	0.879	0.812
	4	AL	4.431	6.771	6.771
		CP	0.995	0.918	0.772
	5	AL	4.163	5.750	5.750
		CP	0.998	0.942	0.633

Table 8: The estimated biases and EMSPEs of the AML predictors and the ALs and CPs of the 95% APIs based on the HCD methods.

θ	m	AML Predictor		API based on the HCD method	
		Bias	EMSPE	AL	CP
0.5	3	Bias	0.041	AL	0.070
		EMSPE	0.004	CP	0.886
	4	Bias	0.026	AL	0.046
		EMSPE	0.002	CP	0.899
	5	Bias	0.016	AL	0.033
		EMSPE	0.000	CP	0.887
1	3	Bias	0.095	AL	0.156
		EMSPE	0.037	CP	0.887
	4	Bias	0.049	AL	0.098
		EMSPE	0.007	CP	0.917
	5	Bias	0.036	AL	0.071
		EMSPE	0.003	CP	0.887
2	3	Bias	0.225	AL	0.362
		EMSPE	0.226	CP	0.884
	4	Bias	0.130	AL	0.223
		EMSPE	0.055	CP	0.880
	5	Bias	0.090	AL	0.162
		EMSPE	0.026	CP	0.883

5. REAL DATA EXAMPLE

In this section, we use a real data set to illustrate the estimation and prediction procedures for the *ILD*. The data are the monthly rainfall during December recorded at Los Angeles civic center from 2001 to 2016 (see the website of Los Angeles Almanac: www.laalmanac.com/weather/we08aa.htm):

1.38 3.31 1.35 8.77 1.03 0.81 1.73 2.79
2.89 10.23 1.01 2.16 0.20 3.88 0.57 4.55

To assess the suitability of the inverse Lindley distribution for the provided dataset, various statistical tests and criteria were applied, including the Kolmogorov–Smirnov ($K-S$) test, Akaike information criterion (AIC), and Bayesian information criterion (BIC). The fitness results for the *ILD* were compared with those for the inverse xgamma distribution introduced by Yadav *et al.* (2021), with $PDF f(x) = \frac{\theta^2}{x^2(1+\theta)}(1 + \frac{\theta}{2x^2}) \exp(-\frac{\theta}{x})$, the inverse Maxwell distribution introduced by Singh and Srivastava (2014), with $PDF f(x) = \frac{4\theta^{1.5}}{\sqrt{\pi x^4}} \exp(-\frac{\theta}{x^2})$, and the inverse Rayleigh distribution with $PDF f(x) = \frac{2\theta}{x^3} \exp(-\frac{\theta}{x^2})$. The results of the $K-S$ test, AIC , and BIC collectively support the appropriateness of the inverse Lindley distribution for the dataset. Specifically, the $K-S$ test yielded a p -value of 0.8047 for the *ILD*, as opposed to 0.6995 for the inverse xgamma, 0.000065 for the inverse Maxwell, and 0.001386 for the inverse Rayleigh distributions. This indicates that both the inverse Lindley and inverse xgamma distributions are suitable for these data. The AIC and BIC values for the *ILD* were obtained to be 71.9553 and 72.7279, respectively. In contrast, for the inverse xgamma distribution, the AIC and BIC values were computed as 72.7797 and 73.5523, suggesting that the inverse Lindley distribution is more appropriate for modeling this dataset.

From the original data set, we have extracted the first five lower records as follows: 1.38, 1.35, 1.03, 0.81, 0.20. Here, we use the same priors used in the simulation study, which are Prior 1 and Prior 2. We calculated the point and interval estimates for the unknown parameter θ based on the observed five lower records. Besides, we computed the AML prediction and the 95% API for the 6-th lower record value. Table 9 represents our numerical findings.

Table 9: The numerical results of the example.

Point Estimation					
		MLE	TK	IS	MH
Prior 1	SE	1.315	1.090	1.121	1.009
	$LE(c = 0.2)$		1.073	1.108	0.997
	$LE(c = -0.2)$		1.102	1.135	1.021
	$GE(p = -0.2)$		1.013	1.052	0.940
	$GE(p = -0.2)$		1.039	1.074	0.963
Prior 2	SE		1.578	1.588	1.625
	$LE(c = 0.2)$		1.553	1.566	1.604
	$LE(c = -0.2)$		1.600	1.610	1.648
	$GE(p = -0.2)$		1.489	1.508	1.546
	$GE(p = -0.2)$		1.519	1.535	1.572
Interval Estimation					
95% Asymptotic CI		95% $Boot-B$ CI		95% $Boot-P$ CI	
(0.382, 2.248)		(-0.814, 1.679)		(0.951, 3.443)	
Prediction					
AML prediction			95% API		
0.200			(0.133, 0.200)		

6. CONCLUSIONS

The inverse Lindley distribution, introduced by [Sharma *et al.* \(2015\)](#), offers a versatile distribution with an inverted bathtub-shaped hazard rate function. [Sharma *et al.* \(2015\)](#) demonstrated its applicability to real-world data, specifically survival times of head and neck cancer patients. Since its inception, various authors have explored inferential aspects of the inverse Lindley distribution (*ILD*).

This paper focuses on the estimation of the unknown parameter of the *ILD* when the first m record values are available. The classical and Bayesian procedures were employed for parameter estimation, and attention was given to predicting a future record value. The article includes a simulation study and a real data application to illustrate the proposed procedures. A comparative analysis involved the maximum likelihood estimator and different Bayes estimators under squared error, linear-exponential, and general entropy loss functions, considering average empirical biases and associated estimated risks. The asymptotic and two bootstrap-type confidence intervals were assessed for their coverage probabilities and average lengths. Notably, the asymptotic confidence intervals demonstrated shorter lengths and larger coverage probabilities compared to bootstrap confidence intervals. Furthermore, Bayesian methods with small prior variance emerged as more preferable than classical methods.

The exploration extends to the estimation problem for $R = P(X < Y)$, utilizing two sequences of lower record values from two inverse Lindley populations with different parameters. Future work is suggested on inferential challenges for generalizations of the *ILD* based on record data. Additionally, the paper proposes investigating estimation and prediction problems for the *ILD* using alternative data types, such as progressively type I and type II censored data, hybrid censored data, progressively first failure censored data, and more. The authors anticipate reporting findings on some of these topics in future research endeavors. All computations were carried out using the statistical software R ([R Core Team, 2020](#)) and the packages `AdequacyModel` ([Marinho *et al.*, 2013](#)), `LindleyR` ([Mazucheli *et al.*, 2016](#)), `lamW` ([Adler, 2017](#)), and `nleqslv` ([Hasselmann, 2018](#)) therein.

A. APPENDIX

Here, we want to prove (3.4). From (3.3), the conditional *PDF* of Z given the last observed record x_m for the *ILD*, $f_Z(z) \equiv f_Z(z|x_m; \theta)$, can be rewritten as

$$(A.1) \quad f_Z(z) = \frac{\left(\frac{\theta}{z} - \frac{\theta}{x_m} + \ln\left(\frac{z[(1+\theta)x_m + \theta]}{x_m[(1+\theta)z + \theta]}\right)\right)^{n-m-1}}{(n-m-1)!} \times \frac{\theta^2(1+z)}{z^2[(1+\theta)z + \theta]} \\ \times \frac{x_m[(1+\theta)z + \theta]}{z[(1+\theta)x_m + \theta]} e^{-\theta\left(\frac{1}{z} - \frac{1}{x_m}\right)}.$$

Let

$$(A.2) \quad u = g^*(z) = \frac{\theta}{z} - \frac{\theta}{x_m} + \ln\left(\frac{z[(1+\theta)x_m + \theta]}{x_m[(1+\theta)z + \theta]}\right).$$

Then, the jacobian is obtained to be

$$(A.3) \quad J = \frac{\partial g^*(z)}{\partial z} = -\frac{\theta^2(1+z)}{z^2[(1+\theta)z + \theta]}.$$

In addition, from (A.2), we get

$$(A.4) \quad e^{-u} = \frac{x_m[(1+\theta)z + \theta]}{z[(1+\theta)x_m + \theta]} e^{-\theta\left(\frac{1}{z} - \frac{1}{x_m}\right)}.$$

Note that the *PDF* of U , given in (3.4), can be written as $g_U(u) = \frac{f_Z(g^{*-1}(u))}{|J|}$, where $g^{*-1}(\cdot)$ is the inverse function of $g^*(\cdot)$. So, the result follows from (A.1), (A.2), (A.3), and (A.4).

ACKNOWLEDGMENTS

We would like to thank the referee for his/her valuable comments.

REFERENCES

- Adler, A. (2017). *lamW: Lambert-W function*, R package version 1.3.0.
<https://CRAN.R-project.org/package=lamW>
- Ahmadi, J. and Doostparast, M. (2006). Bayesian estimation and prediction for some life time distributions based on record values. *Statistical Papers*, 47(3):373–392.
- Arnold, B.C., Balakrishnan, N., and Nagaraja, H.N. (1998). *Records*. John Wiley & Sons, New York.
- Asgharzadeh, A., Fallah, A., Ragab, M.Z., and Valiollah, R. (2018). Statistical inference based on Lindley record data. *Statistical Papers*, 59(2):759–779.
- Asgharzadeh, A., Nadarajah, S., and Sharafi, F. (2017). Generalized inverse Lindley distribution with application to Danish fire insurance data. *Communications in Statistics – Theory and Methods*, 46(10):5001–5021.
- Basak, P. and Balakrishnan, N. (2003). *Maximum Likelihood Prediction of Future Record Statistic*. In “Mathematical and Statistical Methods in Reliability”, B.H. Lindqvist and K.A. Doksum, (Eds.), Series on Quality, Reliability and Engineering Statistics: Volume 7, 159–175. World Scientific Publishing, New Jersey.
- Basu, S., Singh, S.K., and Singh, U. (2017). Parameter estimation of inverse Lindley distribution for Type-I censored data. *Computational Statistics*, 32(1):367–385.
- Basu, S., Singh, S.K., and Singh, U. (2019). Estimation of inverse Lindley distribution using product of spacings function for hybrid censored data. *Methodology and Computing in Applied Probability*, 21(4):1377–1394.
- Berred, A.M. (1998). Prediction of record values. *Communications in Statistics – Theory and Methods*, 27(9):2221–2240.
- Calabria, R. and Pulcini, G. (1994). An engineering approach to Bayes estimation for the Weibull distribution. *Microelectronics Reliability*, 34(5):789–802.
- Casella, G. and Berger, R.L. (2002). *Statistical Inference*, 2nd ed.. Duxbury, Pacific Grove.
- Chandler, K.N. (1952). The distribution and frequency of record values. *Journal of the Royal Statistical Society, Series B (Methodological)*, 14(2):220–228.
- Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.
- Dey, S. and Pradhan, B. (2014). Generalized inverted exponential distribution under hybrid censoring. *Statistical Methodology*, 18:101–114.
- Dunsmore, I.R. (1983). The future occurrence of records. *Annals of the Institute of Statistical Mathematics*, 35(2):267–277.
- Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. CMBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia.
- Fallah, A., Asgharzadeh, A., and MirMostafaei, S.M.T.K. (2018). On the Lindley record values and associated inference. *Journal of Statistical Theory and Applications*, 17(4):686–702.
- Ferentinos, K.K. and Karakostas, K.X. (2006). More on shortest and equal tails confidence intervals. *Communications in Statistics – Theory and Methods*, 35(5):821–829.
- Hasselmann, B. (2018). *nleqslv: Solve systems of nonlinear equations*. R package version 3.3.2.
<https://CRAN.R-project.org/package=nleqslv>
- Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–101.

- Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd ed.. John Wiley & Sons, Hoboken.
- Marinho, P.R.D., Bourguignon, M., and Dias, C.R.B. (2013). *AdequacyModel: Adequacy of probabilistic models and generation of pseudo-random numbers*. R package version 1.0.8. <https://CRAN.R-project.org/package=AdequacyModel>
- Mazucheli, J., Fernandes, L.B., and de Oliveira, R.P. (2016). *LindleyR: The Lindley distribution and its modifications*. R package version 1.0.0. <https://CRAN.R-project.org/package=LindleyR>
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). Equations of state calculations by fast computing machine. *The Journal of Chemical Physics*, 21(6):1087–1092.
- Nezvorov, V.B. (1988). Records. *Theory of Probability and Its Applications*, 32(2):201–228.
- R Core Team (2020). *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Resnick, S.I. (1973). Record values and maxima. *The Annals of Probability*, 1(4):650–662.
- Sharma, V.K., Singh, S.K., Singh, U., and Agiwal, V. (2015). The inverse Lindley distribution: a stress-strength reliability model with application to head and neck cancer data. *Journal of Industrial and Production Engineering*, 32(3):162–173.
- Singh, S., Dey, S., and Kumar, D. (2020). Statistical inference based on generalized Lindley record values. *Journal of Applied Statistics*, 47(9):1543–1561.
- Singh, K.L. and Srivastava, R.S. (2014). Inverse Maxwell distribution as a survival model, genesis and parameter estimation. *Research Journal of Mathematical and Statistical Sciences*, 2(7):23–28.
- Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86.
- Varian, H. (1975). *A Bayesian approach to real estate assessment*. In “Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savege”, S.E. Fienberg and A. Zellner (Eds.), 195–208. North-Holland Publishing Company, Amsterdam.
- Yadav, A.S., Maiti, S.S., and Saha, M. (2021). The inverse xgamma distribution: statistical properties and different methods of estimation. *Annals of Data Science*, 8(2):275–293.
- Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, 81(394):446–451.

On the Construction of Third-Order Rotatable Designs in Smaller Runs

Authors: ANKITA VERMA

– ICAR-Indian Agricultural Statistics Research Institute,
New Delhi - 110012
ankiverma95@gmail.com

SEEMA JAGGI

– Agricultural Education Division, ICAR, New Delhi - 110012
seema.jaggi@icar.org.in

ELDHO VARGHESE  

– ICAR-Central Marine Fisheries Research Institute, Kochi - 682018
eldhoiasri@gmail.com, eldho.varghese@icar.org.in

CINI VARGHESE

– ICAR-Indian Agricultural Statistics Research Institute, New Delhi - 110012
cini.varghese@icar.org.in

ARPAN BHOWMIK 

– ICAR-Indian Agricultural Research Institute, Assam
arpan.stat@gmail.com

ANINDITA DATTA

– ICAR-Indian Agricultural Statistics Research Institute, New Delhi - 110012
aninditaiasri@gmail.com

HEMAVATHI M.

– Pandit Jawaharlal Nehru College of Agriculture and Research Institute, Puducherry
hemavathi.phdstat@gmail.com

ASHUTOSH DALAL

– ICAR-Indian Agricultural Statistics Research Institute, New Delhi - 110012
ashutosh.dalal97@gmail.com

Received: June 2023

Revised: January 2024

Accepted: January 2024

Abstract:

- There may be situations in a product/process optimisation where a fitted second-order model fails to adequately represent the significance of the relationship between the input and response variables, ultimately leading to the estimation of parameters based on a higher (or third) order model. Sequential third-order response surface designs are beneficial for such situations, which allow for experimenting with a few additional runs and fitting a third-order model without discarding the first-stage experimental runs and generated responses. Sequential experimentation is useful in practice since it is more economical and requires fewer resources. For symmetric as well as mixed-level factors, methods of construction of sequential third-order design have been proposed in the paper that satisfies the necessary moment matrix requirement and ensure rotatability. Additionally, the proposed designs have smaller runs, making it more cost-effective to attain the best response. A comparison of the proposed design with existing sequential designs is also made on the basis of design size, G-efficiency and prediction capability using FDS and VDG plots.

Keywords:

- *lack of fit; moment matrix; response surface methodology; rotatability; sequential design; third-order model.*

AMS Subject Classification:

- 49A05, 78B26.

1. INTRODUCTION

Response surface methodology (RSM) is a statistical technique used to design, analyse and optimise complex systems by identifying relationships between the input variables and output responses. It involves creating a mathematical model of the system using a set of experimental data points that are collected from a series of well-designed experiments.

The second-order response model is typically very popular for experimenters who wish to estimate the second-order model due to its high efficiency and simple structure. However, sometimes second-order model representations become unrealistic and inadequate due to the lack of fit caused by the presence of third or higher-order relationships in the true response surface model. Thus, regarding various design criteria, the second-order response surface designs might not be precise and efficient enough to express the real model of the systems or processes accurately. A higher-order model, or third-order design, would be helpful in this situation. A thorough explanation of the RSM and a detailed review on Response Surface Designs (RSD) can be found in [Myers \(1971\)](#), [Khuri and Cornell \(1996\)](#), [Box and Draper \(1987, 2007\)](#), [Khuri \(2006\)](#), [Myers *et al.* \(2016\)](#) and [Anderson and Whitcomb \(2016\)](#) with a detailed review available in [Hemavathi *et al.* \(2022b\)](#). RSD can be classified as designs suitable for experimentation either in a sequential or non-sequential manner. In the sequential approach, the design points/runs that fit the response surface model of sequential order are executed in stages without discarding preceding design points, whereas in non-sequential experimentation, the new set of design points is experimented with if the preceding model founds inadequate (significant lack of fit). The design constructed with an equal number of levels for all the factors is called symmetric response surface design, and the design with unequal levels (mixed levels) of factors is asymmetric response surface design ([Hemavathi *et al.*, 2022a](#)).

Sequential and non-sequential response surface designs are both used in experiments to optimise a response surface or explore the optimal conditions for a given factor. Sequential designs rely on the outcome of the previous experiment to guide decisions for the next experiment, which allows for more efficient use of resources and faster convergence to the optimal conditions. Non-sequential designs, however, are less efficient but do not rely on the results of previous experiments. Instead, the design is pre-determined, and all experiments are conducted based on this design.

A first-order design is considered First-Order Rotatable Design (FORD) when it ensures that the variance of the estimated response remains consistent for all points located at equal distances from the design centre. RSM begins with screening and selecting input factors that affect response by experimenting FORD. The process proceeds with the assumption of fitting the first-order with interaction and then a higher-order model by sequential experimentation.

A higher-order model, such as a quadratic, full second-order, or even a cubic model, will be required if significant curvature is discovered. Higher-order designs offer more precise responses to predict more complex systems. Sequential experimentation is more practical since it is more cost-effective and requires fewer resources.

A design earns the label of a Second-Order Rotatable Design (SOR) when it successfully incorporates a second-order model while preserving the property of rotatability. Similarly, when a design fits a third-order model while still maintaining the rotatability property, it is referred to as a Third-Order Rotatable Design (TOR) by Hemavathi *et al.* (2022b). Gardiner *et al.* (1959) studied third-order design in detail and constructed sequential third-order rotatable design for three and four factors. Draper (1960a) constructed a third-order rotatable design (TOR) in 4 dimensions, which requires 96 runs and is a combination of 4 second-order rotatable design (SOR) arrangements. A third-order rotatable design in 3 dimensions, combining two second-order rotatable arrangements constructed by Draper (1960b). Thaker and Das (1961) obtained sequential TOR up to eleven factors. Das and Narasimham (1962) proposed sequential third-order designs that can be used to estimate a complete third-order model in case the second-order model shows a lack of fit. However, these designs have very large run sizes. Third-order rotatable designs, both sequential and non-sequential, up to 15 factors have been obtained with the help of doubly balanced incomplete block designs and complementary BIB designs. Adhikary and Panda (1982) gave mixed-order response surface designs like FORD-SOR, FORD-TOR, and SOR-TOR and discussed their analysis and construction. Huda (1982a,b) obtained some new third-order rotatable designs in 5, 6, 7 and 8 dimensions in sequential set-up. Construction of TORs for factor, $v = 6$ available in Mutiso and Koske (2007) and Mutiso (1998).

Arshad *et al.* (2012) constructed an Augmented Box-Behnken Design (ABBD) using combinations of factorial, axial, and complementary design points. These augmented designs can be used to estimate the parameters of a third-order response surface model. Rashid *et al.* (2017) developed Augmented Fractional Box-Behnken designs (AFBBD) using combinations of fractional Box-Behnken design points, factorial design points, axial design points and complementary design points. Arshad *et al.* (2020) developed sequential third-order designs for the estimation of a complete third-order model in case the second-order model's lack of fit is exhibited. These designs have smaller run sizes as compared to Das-Narasimham designs and are symmetric in each. By merging SORs, Cornelious (2019a,b) created TORs for $v = 4$ factors in 56 points and for $v = 5$ factors in 134 points. Cornelious and Cruyff (2019) presented an illustrative case study of sequential TOR for $v = 4$ in 80 points.

There may be situations where the number of levels for all the factors studied in the experiment are not the same. For fitting second-order response surfaces, Ramchander (1963) obtained asymmetrical response surface designs of type 3×5^v . Draper and Stoneman (1968) studied the number of runs required to fit the response surface model to mixed two-level and three-level factorial designs and mixed two-level and four-level designs. Mehta and Das (1968) demonstrated how an orthogonal transformation might be used to convert a second-order symmetric rotatable design into a second-order asymmetric rotatable design. Dey (1969) discussed techniques for creating partially rotatable second-order asymmetric response designs of the kind $3^v \times 5$. Compared to the conventional rotatable response surface designs for a quadratic response surface, Das *et al.* (1999) provided variously modified and/or rotatable response surface symmetric and asymmetric designs. Some asymmetric third-order designs that are appropriate for sequential experiments were introduced by Hemavathi *et al.* (2022a).

This article presents a method of constructing a series of Sequential Third-Order Rotatable Designs (STORDs) for symmetric and asymmetric levels of factors. The first stage design can be utilized to fit the second-order model, and further, a third-order model may be fitted with the addition of a few more runs without discarding the initial design. The proposed designs are more cost-effective in terms of the number of runs in obtaining the optimal response.

2. RESPONSE SURFACE MODEL

The response surface that depicts the relationship between the response and the factors influencing it is expressed as

$$(2.1) \quad y_u = f(x_{1u}, x_{2u}, \dots, x_{vu}) + e_u,$$

where $u = 1, 2, \dots, N$, y_u is the response obtained from the u^{th} treatment combination, and x_{iu} is the level of the i^{th} ($i = 1, 2, \dots, v$) factor in the u^{th} treatment combination. The function f describes the form in which the response and the input variables are related. e_u is the random error associated with the u^{th} observation that is independently and normally distributed with mean zero and common variance σ^2 . In matrix notation, the relationship can be expressed as:

$$(2.2) \quad \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

where $\mathbf{Y} = (y_1 y_2 \dots y_N)'$ is an $N \times 1$ vector of observations, \mathbf{X} is a $N \times (p+1)$ matrix of independent variables, $\boldsymbol{\beta} = (\beta_0 \beta_1 \dots \beta_p)'$ is a $(p+1) \times 1$ vector of parameters and $\mathbf{e} = (e_1 e_2 \dots e_N)'$ is $N \times 1$ vector of random errors distributed as $N(0, \sigma^2 \mathbf{I}_N)$. For a second-order response surface model with v factors, the function f in 2.1 is of the form:

$$(2.3) \quad f(x_u) = \beta_0 + \sum_{i=1}^v \beta_i x_{iu} + \sum_{i \leq j=1}^v \beta_{ij} x_{iu} x_{ju} \quad u = 1, 2, \dots, N,$$

where $\beta_0, \beta_i, \beta_{ii}$ and β_{ij} are the intercept, linear regression, quadratic and interaction coefficients, respectively. The total number of parameters p in this complete second-order model to be estimated are $\binom{v+2}{2}$. For $v = 2$, the second-order response surface model takes the following form:

$$(2.4) \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + e.$$

For a third-order response surface model with v factors, the function f in 2.1 is of the form:

$$(2.5) \quad f(x_u) = \beta_0 + \sum_{i=1}^v \beta_i x_{iu} + \sum_{i \leq j \leq k=1}^v \beta_{ijk} x_{iu} x_{ju} x_{ku} \quad u = 1, 2, \dots, N,$$

where $\beta_0, \beta_i, \beta_{ii}, \beta_{iii}, \beta_{ij}$ and β_{ijk} are the intercept, linear regression, quadratic, cubic, second and third-order interaction coefficients, respectively. The total number of parameters p in this complete third-order model to be estimated are $\binom{v+3}{3}$. For $v = 3$, the third-order response surface model takes the following form:

$$(2.6) \quad \begin{aligned} y = & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 \\ & + \beta_{33} x_3^2 + \beta_{122} x_1 x_2^2 + \beta_{133} x_1 x_3^2 + \beta_{113} x_1^2 x_3 + \beta_{233} x_2 x_3^2 + \beta_{123} x_1 x_2 x_3 \\ & + \beta_{111} x_1^3 + \beta_{222} x_2^3 + \beta_{333} x_3^3 + e. \end{aligned}$$

The conditions for near-orthogonal estimation of parameters and constancy of variances of all parameters for second and third-order models are given in the appendix. The design point that satisfies the respective conditions is called second-order and third-order rotatable design, respectively.

According to Gardiner's ordering ([Gardiner et al., 1959](#)), the components in the matrix (\mathbf{X}) for the third-order model are written as follows:

$$(2.7) \quad [x_0 \quad x_1^2 \quad x_2^2 \quad \dots \quad x_v^2 \quad x_i \quad x_i^3 \quad x_i x_1^2 \quad x_i x_2^2, \dots, x_i x_v^2 \quad x_1 x_2 \quad x_1 x_3 \dots x_{v-1} x_v \\ x_1 x_2 x_3 \quad x_1 x_2 x_4 \dots x_{v-2} x_{v-1} x_v].$$

Some of the properties that must be considered when choosing a response surface design are given by [Box and Draper \(1975\)](#), which was further emphasized by [Khuri and Cornell \(1996\)](#).

3. SEQUENTIAL THIRD-ORDER RESPONSE SURFACE DESIGNS

This section presents the method of constructing sequential third-order response surface designs for symmetric and asymmetric factor levels. The design is constructed in two stages, denoted by S_1 and S_2 .

3.1. TORDs with Symmetric Factor Levels

In stage 1, 2^v factorial with levels $\pm a$ or a fraction of 2^v factorial is taken along with the $2v$ axial points with levels $\pm \alpha$ where $\alpha = \sqrt{v}$ and an appropriate number of centre points (n_c) to prevent the singularity of the design. The S_1 design is a second-order rotatable design. A formal test for lack of fit is conducted following the estimation of the second-order model. If a significant lack of fit is seen in the second-order model, then it is important to estimate some or all of the third-order elements that could be present. For S_2 , consider a Balanced Incomplete Block (BIB) design with parameters (v, b, r, k, λ) ([Dey, 2010](#)). This design can be expressed in terms of the incidence matrix of order $b \times v$ with the elements 0 and a . By multiplying each of these b combinations by 2^v with levels 1 and -1 or a fraction of it, i.e. $2^k (k < v)$, $b \times 2^k$ design points are obtained. Combining the design points of S_1 and S_2 , the Sequential Third-Order Rotatable Design (STORD) is obtained that consists of v quantitative factors with each factor at levels $\pm a$, $\pm \sqrt{v}$ and 0.

Example 3.1. For $v = 3$, S_1 of the design is created by $(a, a, a) \times \frac{1}{2}2^3, (\alpha, 0, 0) \times 2$, which results in 10 runs and for S_2 , consider a BIB design with parameters $(3, 3, 2, 2, 1)$. The 3×3 incidence matrix is written as

$$N_{3 \times 3} = \begin{bmatrix} a & a & 0 \\ a & 0 & a \\ 0 & a & a \end{bmatrix}.$$

The following 12 design points are obtained by multiplying each combination of the incidence matrix by the 2^2 combinations of +1 and -1:

$$N_{12 \times 3} = \begin{bmatrix} \pm a & \pm a & 0 \\ \pm a & 0 & \pm a \\ 0 & \pm a & \pm a \end{bmatrix}.$$

For $a = 1$, $\alpha = \sqrt{v} = 1.7321$ (Ramchander, 1963) S_1 and S_2 together result in a STORD in 22 runs with the design \mathbf{X} as:

$$\mathbf{X} = \begin{bmatrix} \frac{1}{2}(\pm a & \pm a & \pm a) \\ \pm a & 0 & 0 \\ 0 & \pm a & 0 \\ 0 & 0 & \pm a \\ \pm a & \pm a & 0 \\ \pm a & 0 & \pm a \\ 0 & \pm a & \pm a \end{bmatrix}.$$

In the design for $v = 3$, each factor has 5 levels. The variances of estimated response i.e. $V(\hat{y})$ obtained in S_1 i.e. for a second-order design after adding 3 centre runs are as $0.3333\sigma^2$, $0.85\sigma^2$ and $0.9333\sigma^2$. Further, the variances of estimated response i.e. $V(\hat{y})$ obtained by taking S_1 and S_2 together for a third-order design are $0.8636\sigma^2$, $0.9394\sigma^2$ and $1\sigma^2$. It can be seen that there are three different variances based on the three different input factorial combinations taking into account the distance from the design centre. Thus, the final design is also rotatable.

There are 20 parameters in the third-order model. The variance of these estimated parameters are obtained as follows:

$$\begin{aligned} V(\hat{\beta}_0) &= 1\sigma^2, & V(\hat{\beta}_i) &= 0.9090\sigma^2, & V(\hat{\beta}_{ii}) &= 2\sigma^2, & i &= 1, 2, 3; \\ V(\hat{\beta}_{ijj}) &= V(\hat{\beta}_{ij}) &= 0.3636\sigma^2, & i \neq j &= 1, 2, 3; \\ V(\hat{\beta}_{iii}) &= 6.3636\sigma^2, & i &= 1, 2, 3; \\ V(\hat{\beta}_{123}) &= 0.1818\sigma^2. \end{aligned}$$

It is seen that the parameters of a certain order are estimated with the same variance.

Note: G-efficiency = $\frac{p}{N \times \max(v(x))_{x \in R}}$, where $\max(v(x))_{x \in R}$ is the maximum prediction variance over the design space R. The G-efficiency criterion seeks to maximize a design's ability to predict by reducing the variances of the predicted values.

Table 1 presents a list of STORD for 3 to 9 factors. It includes the number of factors (v), number of runs (N), variances of predicted response $\frac{V(\hat{y})}{\sigma^2}$, design points, and G-efficiency. The variance of the estimated response and the variance of estimated parameters has been obtained using a computer programme developed in SAS IML (Varghese et al., 2022). The proposed STORD has smaller runs as compared to other existing sequential designs and upto seven factors G-efficiency is considerably high.

Table 1: Sequential Third-Order Rotatable Designs for factors ranging from 3 to 9.

v	N	$\frac{V(\hat{y})}{\sigma^2}$	Design Points	G-Efficiency ($a = 1, \alpha = \sqrt{v}$)
3	$22+n_c$	0.8636 0.9394 1.0000	$S_1 : (a, a, a) \times \frac{1}{2}2^3$ $(\alpha, 0, 0) \times 2$ $S_2 : a(3, 3, 2, 2, 1)$	0.9091
4	$48+n_c$	0.6167 0.8417	$S_1 : (a, a, a, a) \times \frac{1}{2}2^4$ $(\alpha, 0, 0, 0) \times 2$ $S_2 : a(4, 6, 3, 2, 1)$	0.8663
5	$66+n_c$	0.0643 0.7892 0.8591 0.9740	$S_1 : (a, a, a, a, a) \times \frac{1}{2}2^5$ $(\alpha, 0, 0, 0, 0) \times 2$ $S_2 : a(5, 10, 4, 2, 1)$	0.8214
6	$124+n_c$	0.0597 0.5891 0.8339	$S_1 : (a, a, a, a, a, a) \times \frac{1}{2}2^6$ $(\alpha, 0, 0, 0, 0, 0) \times 2$ $S_2 : a(6, 10, 5, 3, 2)$	0.7931
7	$162+n_c$	0.0211 0.5369 0.8686 0.9777	$S_1 : (a, a, a, a, a, a, a) \times \frac{1}{2}2^7$ $(\alpha, 0, 0, 0, 0, 0, 0) \times 2$ $S_2 : a(7, 21, 6, 2, 1)$	0.7393
8	$372+n_c$	0.0226 0.3265 0.6172 0.79844	$S_1 : (a, a, a, a, a, a, a, a) \times \frac{1}{2}2^8$ $(\alpha, 0, 0, 0, 0, 0, 0, 0) \times 2$ $S_2 : a(8, 14, 7, 4, 3)$	0.5555
9	$562+n_c$	0.2969 0.4690 0.8012	$S_1 : (a, a, a, a, a, a, a, a, a) \times \frac{1}{2}2^9$ $(\alpha, 0, 0, 0, 0, 0, 0, 0, 0) \times 2$ $S_2 : a(9, 18, 8, 4, 3)$	0.4886

3.2. TORDs with Asymmetric Factor Levels

Let the design matrix $\mathbf{X} = (x_{1u}x_{2u}\dots x_{vu})$, where $x_{iu}, i = 1, 2, \dots, v$ is a vector of order $N \times 1$ given in 3.1, be transformed to \mathbf{Z} through a conformable orthonormal transformation matrix \mathbf{B} as given below:

$$\mathbf{Z} = \mathbf{B}\mathbf{X}.$$

\mathbf{B} is a transformation matrix of order $v \times v$ such that its elements b_{ii} satisfy the relations:

$$\sum_{i=1}^v b^2_{ii} = 1 \quad \forall i = 1, 2, \dots, v; \quad \sum_{i=1}^v b_{ij}b_{kj} = 0 \quad \forall i \neq j \neq k = 1, 2, \dots, v.$$

Adding an appropriate number of centre points (n_c), the design \mathbf{Z} obtained satisfies the conditions for near-orthogonal estimation of parameters and constancy of variances of linear and quadratic parameters for a second-order model and third-order model are given in the Appendix. The design so obtained is a Sequential Asymmetrical Third-Order Rotatable Design (SATORD), which is rotatable. It can be noted that the transformed asymmetrical design \mathbf{Z} 's estimated response and G-efficiency variances are equivalent to those of the analogous symmetrical design \mathbf{X} .

Example 3.2. For $v = 3$, consider the X as given in Example 3.1. Let

$$\mathbf{B} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} & 0 \\ -\frac{4}{5} & \frac{3}{5} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The \mathbf{Z} matrix so obtained has first two factors at 9 levels as $\pm\frac{7}{5}a, \pm\frac{7}{5}a, \pm\frac{3}{5}b, \pm\frac{4}{5}b, 0$ and 3^{rd} factor has 5 levels for S_1 for fitting of a second-order model.

When S_1 and S_2 are taken together for fitting a third-order model, then the first two factors have 13 levels while the third factor has 5 levels. Along with the previously mentioned 9 levels, there are an additional 4 levels of first two factors resulting in SATORD which are $\pm\frac{3}{5}a, \pm\frac{4}{5}a$. The variances of these estimated parameters are obtained as follows:

$$\begin{aligned} V(\hat{\beta}_0) &= 1\sigma^2, V(\hat{\beta}_i) = 0.8182\sigma^2, V(\hat{\beta}_{ii}) = 1.363\sigma^2, i = 1, 2, 3; \\ V(\hat{\beta}_{ijj}) &= V(\hat{\beta}_{ij}) = 0.3636\sigma^2, i \neq j = 1, 2, 3; V(\hat{\beta}_{iii}) = 3\sigma^2, i = 1, 2, 3; \\ V(\hat{\beta}_{123}) &= 0.1818\sigma^2. \end{aligned}$$

It is clear that the parameters of a certain order are estimated with the same variance; specifically, the variances of the linear coefficients, quadratic coefficients, second-order interaction coefficients, cubic coefficients and third-order interaction coefficients are all similar. Additionally, it can be seen that interaction coefficient variances are comparable to those of symmetrical designs. Additionally, the predicted response variances are also the same as those of the symmetrical design.

This design for 3 factors require 22 runs which is less than the design constructed by Hemavathi et al. (2022a) in 46 runs.

A list of sequential asymmetrical third-order rotatable designs is presented in Table 2, along with the factor levels and orthonormal transformation matrices that were used, where

$$\begin{aligned} \mathbf{Q}_0 &= \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{bmatrix}, \mathbf{Q}_1 = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}, \mathbf{Q}_2 = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \end{bmatrix}, \\ \mathbf{Q}_3 &= \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}, \mathbf{Q}_4 = \begin{bmatrix} \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{12}} \\ \frac{1}{2} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{12}} \\ \frac{1}{2} & 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{12}} \\ \frac{1}{2} & 0 & 0 & -\frac{3}{\sqrt{12}} \end{bmatrix}. \end{aligned}$$

The variance of estimated response and G-efficiency are identical to those of the symmetrical designs shown in Table 1. Levels of factors depend on the orthonormal transformation matrix used.

Table 2: SATORD for factors ranging from 3 to 9.

v	N	Design Points ($a = 1$)	Levels	B
3(i)	$22+n_c$	$S_1 : (a, a, a) \times \frac{1}{2}2^3$ $(\alpha, 0, 0) \times 2$ $S_2 : a(3, 3, 2, 2, 1)$	$S_1 : (9, 9, 5)$ $S_1 + S_2 : (13, 13, 5)$	$\begin{bmatrix} Q_0 & \mathbf{0} \\ \mathbf{0}^T & \mathbf{1} \end{bmatrix}$
3(ii)	$22+n_c$	$S_1 : (a, a, a) \times \frac{1}{2}2^3$ $(\alpha, 0, 0) \times 2$ $S_2 : a(3, 3, 2, 2, 1)$	$S_1 : (9, 9, 5)$ $S_1 + S_2 : (13, 13, 5)$	$\begin{bmatrix} Q_1 & \mathbf{0} \\ \mathbf{0}^T & \mathbf{1} \end{bmatrix}$
4(i)	$48+n_c$	$S_1 : (a, a, a, a) \times \frac{1}{2}2^4$ $(\alpha, 0, 0, 0) \times 2$ $S_2 : a(4, 6, 3, 2, 1)$	$S_1 : (9, 9, 5, 5)$ $S_1 + S_2 : (13, 13, 5, 5)$	$\begin{bmatrix} Q_1 & \mathbf{0} \\ \mathbf{0}^T & I_{2 \times 2} \end{bmatrix}$
4(ii)	$48+n_c$	$S_1 : (a, a, a, a) \times \frac{1}{2}2^4$ $(\alpha, 0, 0, 0) \times 2$ $S_2 : a(4, 6, 3, 2, 1)$	$S_1 : (7, 5, 3, 5)$ $S_1 + S_2 : (7, 9, 5, 5)$	$\begin{bmatrix} Q_2 & \mathbf{0} \\ \mathbf{0}^T & \mathbf{1} \end{bmatrix}$
5(i)	$66+n_c$	$S_1 : (a, a, a, a, a) \times \frac{1}{2}2^5$ $(\alpha, 0, 0, 0, 0) \times 2$ $S_2 : a(5, 10, 4, 2, 1)$	$S_1 : (11, 11, 11, 5, 5)$ $S_1 + S_2 : (11, 11, 11, 5, 5)$	$\begin{bmatrix} Q_3 & \mathbf{0} \\ \mathbf{0}^T & I_{2 \times 2} \end{bmatrix}$
5(ii)	$48+n_c$	$S_1 : (a, a, a, a, a) \times \frac{1}{2}2^5$ $(\alpha, 0, 0, 0, 0) \times 2$ $S_2 : a(5, 10, 4, 2, 1)$	$S_1 : (9, 9, 5, 5, 5)$ $S_1 + S_2 : (13, 13, 5, 5, 5)$	$\begin{bmatrix} Q_0 & \mathbf{0} \\ \mathbf{0}^T & I_{3 \times 3} \end{bmatrix}$
6(i)	$124+n_c$	$S_1 : (a, a, a, a, a, a) \times \frac{1}{2}2^6$ $(\alpha, 0, 0, 0, 0, 0) \times 2$ $S_2 : a(6, 10, 5, 3, 2)$	$S_1 : (9, 9, 5, 5, 5, 5)$ $S_1 + S_2 : (13, 13, 5, 5, 5, 5)$	$\begin{bmatrix} Q_0 & \mathbf{0} \\ \mathbf{0}^T & I_{4 \times 4} \end{bmatrix}$
6(ii)	$48+n_c$	$S_1 : (a, a, a, a, a, a) \times \frac{1}{2}2^6$ $(\alpha, 0, 0, 0, 0, 0) \times 2$ $S_2 : a(6, 10, 5, 3, 2)$	$S_1 : (7, 9, 5, 5, 5, 5)$ $S_1 + S_2 : (9, 13, 7, 5, 5, 5)$	$\begin{bmatrix} Q_2 & \mathbf{0} \\ \mathbf{0}^T & I_{3 \times 3} \end{bmatrix}$
7(i)	$162+n_c$	$S_1 : (a, a, a, a, a, a, a) \times \frac{1}{2}2^7$ $(\alpha, 0, 0, 0, 0, 0, 0) \times 2$ $S_2 : a(7, 21, 6, 2, 1)$	$S_1 : (11, 11, 11, 7, 5, 9, 5)$ $S_1 + S_2 : (15, 15, 15, 9, 7, 11, 5)$	$\begin{bmatrix} Q_3 & \mathbf{0}_{3 \times 3} & \mathbf{0} \\ \mathbf{0}_{3 \times 3} & Q_2 & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0}^T & \mathbf{1} \end{bmatrix}$
7(ii)	$162+n_c$	$S_1 : (a, a, a, a, a, a, a) \times \frac{1}{2}2^7$ $(\alpha, 0, 0, 0, 0, 0, 0) \times 2$ $S_2 : a(7, 21, 6, 2, 1)$	$S_1 : (11, 11, 11, 9, 9, 5, 5)$ $S_1 + S_2 : (15, 15, 15, 13, 13, 5, 5)$	$\begin{bmatrix} Q_3 & \mathbf{0}_{3 \times 2} & \mathbf{0} \\ \mathbf{0}_{2 \times 3} & Q_0 & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0}^T & I_{2 \times 2} \end{bmatrix}$
8(i)	$372+n_c$	$S_1 : (a, a, a, a, a, a, a, a) \times \frac{1}{2}2^8$ $(\alpha, 0, 0, 0, 0, 0, 0, 0) \times 2$ $S_2 : a(8, 14, 7, 4, 3)$	$S_1 : (9^2, 5^6)$ $S_1 + S_2 : (13, 5^6)$	$\begin{bmatrix} Q_0 & \mathbf{0} \\ \mathbf{0}^T & I_{6 \times 6} \end{bmatrix}$
8(ii)	$372+n_c$	$S_1 : (a, a, a, a, a, a, a, a) \times \frac{1}{2}2^8$ $(\alpha, 0, 0, 0, 0, 0, 0, 0) \times 2$ $S_2 : a(8, 14, 7, 4, 3)$	$S_1 : (13^5, 5^3)$ $S_1 + S_2 : (11^3, 9^2, 5^3)$	$\begin{bmatrix} Q_3 & \mathbf{0}_{3 \times 2} & \mathbf{0} \\ \mathbf{0}_{2 \times 3} & Q_0 & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0}^T & I_{3 \times 3} \end{bmatrix}$

8(iii)	$372+n_c$	$S_1 : (a, a, a, a, a, a, a, a) \times \frac{1}{2}2^8$ $(\alpha, 0, 0, 0, 0, 0, 0, 0) \times 2$ $S_2 : a(8, 14, 7, 4, 3)$	$S_1 : (7, 7, 5, 7, 7, 5, 5, 5)$ $S_1 + S_2 : (9, 11, 7, 9, 11, 7, 5, 5)$	$\begin{bmatrix} Q_2 & \mathbf{0}_{3 \times 3} & \mathbf{0}^T \\ \mathbf{0}_{3 \times 3} & Q_2 & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0}^T & I_{2 \times 2} \end{bmatrix}$
9(i)	$562+n_c$	$S_1 : (a, a, a, a, a, a, a, a) \times \frac{1}{2}2^9$ $(\alpha, 0, 0, 0, 0, 0, 0, 0) \times 2$ $S_2 : a(9, 18, 8, 4, 3)$	$S_1 : (7, 5, 9, 11, 7, 5, 9, 5, 5)$ $S_1 + S_2 : (11, 7, 13, 17, 9, 7, 13, 5, 5)$	$\begin{bmatrix} Q_4 & \mathbf{0}_{4 \times 3} & \mathbf{0} \\ \mathbf{0}_{3 \times 4} & Q_2 & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0}^T & I_{2 \times 2} \end{bmatrix}$
9(ii)	$562+n_c$	$S_1 : (a, a, a, a, a, a, a, a) \times \frac{1}{2}2^9$ $(\alpha, 0, 0, 0, 0, 0, 0, 0) \times 2$ $S_2 : a(9, 18, 8, 4, 3)$	$S_1 : (9, 9, 5^7)$ $S_1 + S_2 : (13, 13, 5^7)$	$\begin{bmatrix} Q_0 & \mathbf{0} \\ \mathbf{0}^T & I_{7 \times 7} \end{bmatrix}$
9(iii)	$562+n_c$	$S_1 : (a, a, a, a, a, a, a, a) \times \frac{1}{2}2^9$ $(\alpha, 0, 0, 0, 0, 0, 0, 0) \times 2$ $S_2 : a(9, 18, 8, 4, 3)$	$S_1 : (7, 9, 5, 7, 9, 5, 5, 5, 5)$ $S_1 + S_2 : (9, 13, 7, 9, 13, 7, 5, 5, 5)$	$\begin{bmatrix} Q_2 & \mathbf{0}_{3 \times 3} & \mathbf{0} \\ \mathbf{0}_{3 \times 3} & Q_2 & \mathbf{0} \\ \mathbf{0}^T & \mathbf{0}^T & I_{3 \times 3} \end{bmatrix}$

4. COMPARATIVE STUDY

On the basis of the design runs, G-efficiency and prediction ability, a comparison of the proposed design (STORD) with the existing sequential designs, namely, the New Augmented Box-Behnken Design (NABBD) by [Arshad *et al.* \(2020\)](#), the Augmented Fractional Box-Behnken Design (AFBBD) by [Rashid *et al.* \(2017\)](#), the Augmented Box-Behnken Design (ABBD) by [Arshad *et al.* \(2012\)](#) and [Das and Narasimham \(1962\)](#) made sequential design (say, DND).

One of the critical characteristics of a design that must be maintained is run size. When the experimental material is expensive, an experimenter may prefer design of minimum runs, which is cost effective. Table 3 and Figure 1 both compare the runs of all the designs for easier review.

Table 3: Comparison of sequential designs with respect to runs.

Factors	STORD	NABBD	ABBD	AFBBD	DND
3	22	-	26	20	40
4	48	64	48	42	72
5	66	140	82	62	192
6	124	172	172	124	260
7	162	182	182	154	238
8	372	448	464	-	480
9	562	828	842	-	1256

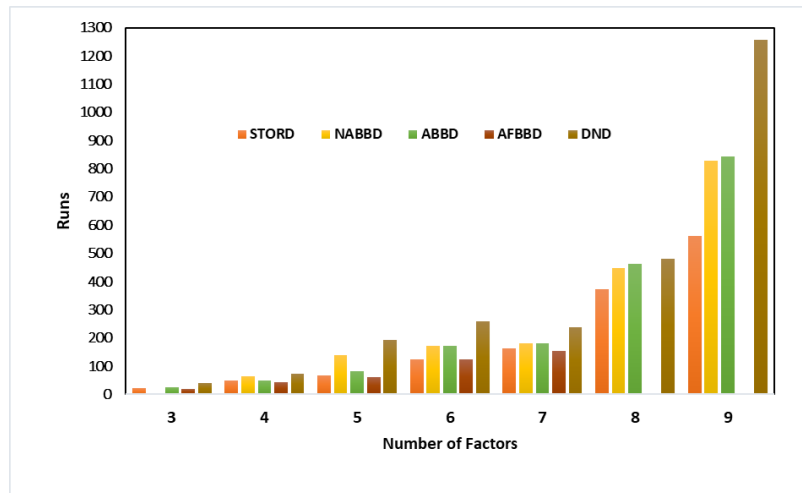


Figure 1: Comparison of STORD and other sequential designs with respect to runs.

The G-efficiency of all the designs are calculated and listed in Table 4 and Figure 2 for better evaluation, is used to compare all of the designs.

Table 4: G-efficiency of STORD, NABBD, ABBD, AFBBB and DND.

Factors	STORD	NABBD	ABBD	AFBBB	DND
3	90.91	-	63.70	83.33	81.18
4	86.63	74.48	60.58	76.09	71.09
5	82.13	90.40	66.57	77.00	30.97
6	79.31	67.71	52.15	76.38	73.18
7	73.93	71.21	60.63	76.43	78.70
8	55.56	77.56	40.06	-	71.91
9	48.86	70.84	15.71	-	57.57

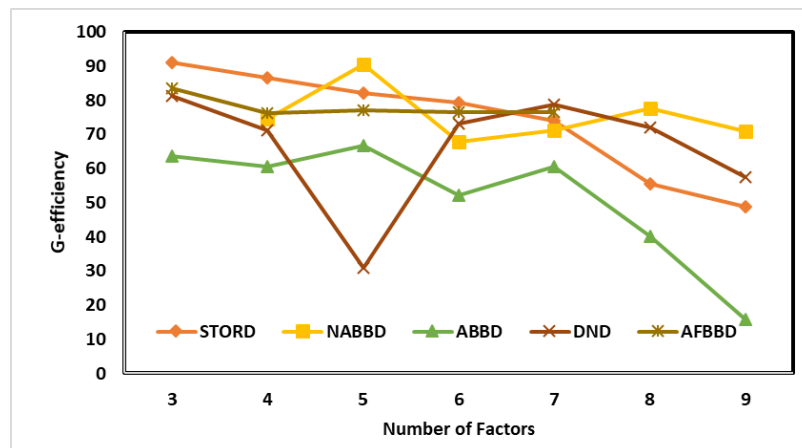


Figure 2: Comparison of STORD and other sequential designs with respect to G-Efficiency.

Table 3 and 4 and Figures 1 and 2 show that STORDs are much smaller than NABBD for each factor from 3 to 9. STORDs perform better for factors 4, 6, and 7 in terms of G-efficiency. In addition, it should be remembered that while STORD is rotatable for all factors, NABBD is not. Run-wise, STORDs are larger than AFBBB. In contrast to STORDs, which are rotatable and have parameter estimates that are nearly orthogonal, AFBBBs do not satisfy the moment matrix criterion, meaning that all parameters are correlated and designs are only partially rotatable. Consider using STORDs if you want responses to be estimated with the same accuracy. In terms of design runs size STORD is smaller than ABBDs, and in terms of G-efficiencies, STORDs are much better than ABBDs. STORDs are substantially smaller than DNDs in terms of runs. STORDs are only effective for factors 3,4,5 and 6 in terms of G-efficiencies. Both designs possess the property of rotatability.

Variance Dispersion Graphs (VDGs) can be used to analyse a response surface design with spherical parts (Giovannitti-Jensen and Myers, 1989). The maximum, lowest and average scaled variance curves represent the predicted value on a hypersphere. Each value is plotted in relation to the circumference of the hypersphere. The degree of rotatability of the Scaled Prediction Variance (SPV) at any specific radius of spheres is indicated by comparing the highest and lowest SPV values across the range of radii.

The Fraction of Design Space (FDS) can be used to examine how well the design's predictions perform across the whole design space (Zahran *et al.*, 2003). The volume of the design region, as well as the maximum, minimum and quantiles of the SPV distribution, are plotted in FDS. The assumption is that an SPV's design is better if it occupies a larger portion of the design space, which is close to the minimum. Additionally, the SPV distribution for that design is more stable the flatter the line is. For the design in Table 3 for $v = 4$, the FDS and VDG for each of the aforementioned designs are plotted and displayed in Figure 3 and Figure 4.

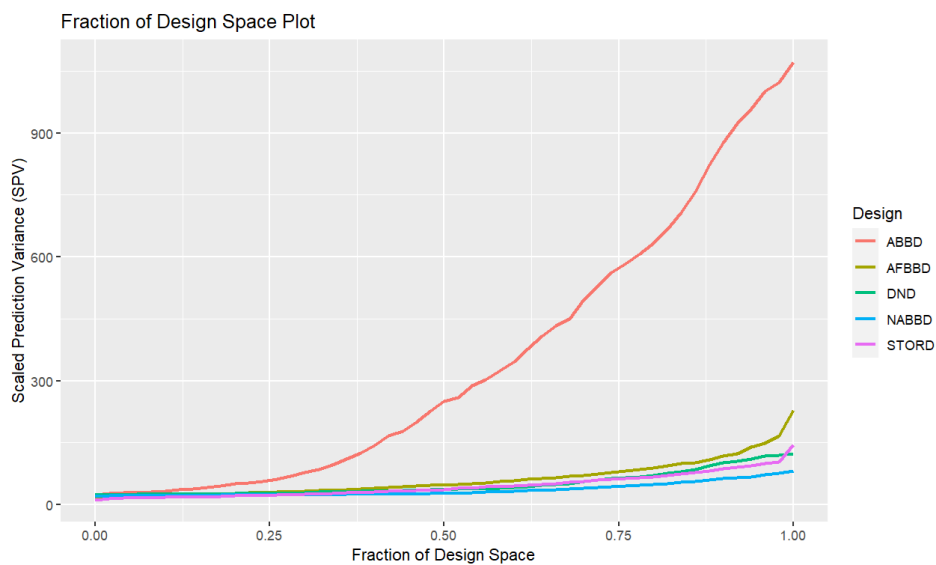


Figure 3: FDS plot for $v = 4$ for different designs.

Figure 3 FDS plot illustrates that the slope of the curve which indicates how quickly the design approaches the maximum value of the Scaled Prediction Variance (SPV), with a slope that is closer to horizontal being desirable. ABBD performed worst from centre to periphery, reaches a very high SPV value as compared to other designs. When STORD have steady SPV from the origin to the periphery, NABBD performs best there. DND performs about the same as STORD but slightly poorer in the perimeter. This is once more made much more obvious by the VDG plot in Figure 4. The SPV profiles of the STORD, DND, and NABBD are similar up to a distance of 1.5 from the origin, and as the designs are rotatable, the minimum, maximum, and average SPV curves are the same. While STORD, max, min, and average SPV are not comparable at the perimeter, DND and NABBD have the most stable SPV profiles. The SPV profile of ABBD is most erratic and performs worst at the periphery. AFBBB do not exhibit property of rotatability.

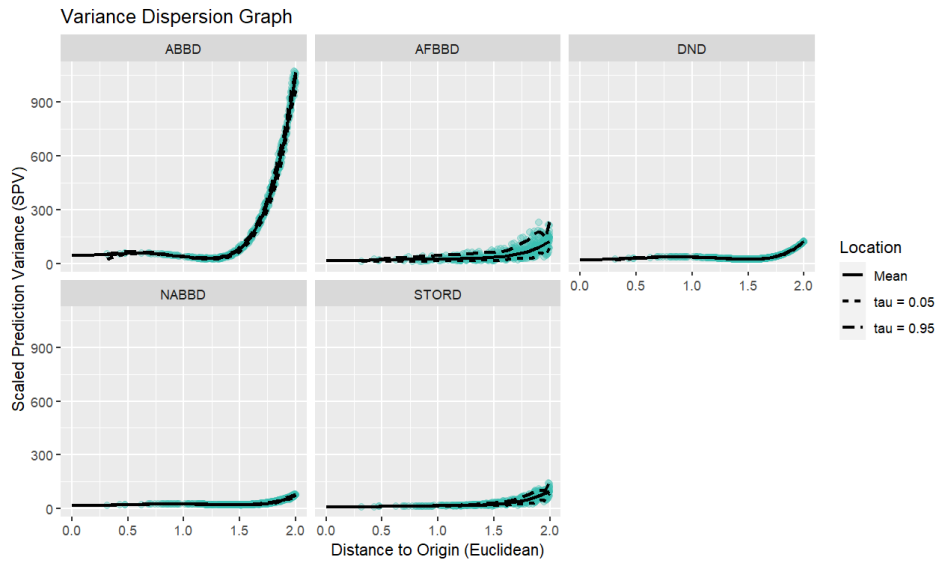


Figure 4: VDG plot for $v = 4$ for different designs.

4.1. Summary and Conclusion

Third-order design in symmetric, as well as asymmetric factor levels suitable for sequential experimentation, is proposed in this article which is termed as Sequential Third-Order Rotatable Design (STORD). Designs are obtained by taking factorial points and axial points in S_1 of the design and BIB design in S_2 of the design. S_1 of the design is second-order rotatable design and satisfies all second-order moment matrix criteria; if there exists a lack of fit of the second-order model, then without discarding S_1 design points, S_2 is augmented in S_1 to form third-order design which is rotatable as well as satisfy all moment matrix criterion. The most crucial feature of the suggested design is that it uses fewer resources and is smaller than all sequential designs that already exist and have all of the desired design characteristics. This characteristic makes the design easy to use and affordable. Using orthonormal transformation, designs with symmetric levels are converted to designs with asymmetric levels resulting

in Sequential Asymmetrical Third-Order Rotatable Design (SATORD), which broadens the applicability of the suggested design. A catalogue of both designs with their G-efficiency, variances of estimated response, number of runs, levels and transformation matrix has been prepared and presented. A comparison of proposed designs with existing designs, have also been made in terms of design size and G-efficiencies. Additionally, using the Fraction of Design Space (FDS) plot and the Variance Dispersion Graph (VDG), a comparison of the designs' ability to predict the effects of four factors has been made.

STORDs satisfy all the properties of good response surface design including good fit of the model, cost effectiveness, sequential build up of the design and rotatability property. AFBBD, ABBD, NABBD do not possess rotatability property and DNDs are not cost effective. Overall, STORDs and SATORDs are best if small run design is more desirable for an experimenter other than rotatability, particularly when the experimental material is expensive, predicted response precision, and uncorrelated parameter estimations are taken into account. The proposed design ensures considerably high G-efficiencies and performs well in terms of prediction capability.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

A. APPENDIX

The conditions for near-orthogonal estimation of parameters and constancy of variances of linear and quadratic parameters for a second-order model ([Hemavathi et al., 2022a](#)) are:

$$\begin{aligned}
\sum_{u=1}^N \prod_{i=1}^2 x_{iu}^{w_i} &= 0 \text{ for } w_i = 0, 1 \text{ or } 3 \text{ and } \sum w_i \leq 4, \\
\sum_{u=1}^N x_{iu}^2 &= N\lambda_2 \quad \forall i = 1, 2, \dots, v, \\
\sum_{u=1}^N x_{iu}^2 x_{ju}^2 &= N\lambda_4 \quad \forall i \neq j = 1, 2, \dots, v, \\
\sum_{u=1}^N x_{iu}^4 &= 3N\lambda_4 \quad \forall i = 1, 2, \dots, v, \\
\sum_{u=1}^N x_{iu}^4 &= 3 \sum_{u=1}^N x_{iu}^2 x_{ju}^2 = 3N\lambda_4 \quad \forall i \neq j = 1, 2, \dots, v, \\
\frac{\lambda_4}{\lambda_2^2} &> \frac{v}{v+2}.
\end{aligned}$$

In addition to the above, the other conditions for a third-order model ([Hemavathi et al., 2022b](#)) are:

$$\begin{aligned}
\sum_{u=1}^N \prod_{i=1}^2 x_{iu}^6 &= 15N\lambda_6 \quad \forall i = 1, 2, \dots, v, \\
\sum_{u=1}^N x_{iu}^2 x_{ju}^2 x_{ku}^2 &= N\lambda_6 \quad \forall i \neq j \neq k = 1, 2, \dots, v, \\
\sum_{u=1}^N x_{iu}^4 x_{ju}^2 &= \dots = \sum_{u=1}^N x_{ju}^2 x_{ku}^4 = 3N\lambda_6 \quad \forall i \neq j \neq k = 1, 2, \dots, v, \\
\sum_{u=1}^N x_{iu}^6 &= 5 \sum_{u=1}^N x_{iu}^2 x_{ju}^4, \\
\sum_{u=1}^N x_{iu}^4 x_{ju}^2 &= 3 \sum_{u=1}^N x_{iu}^2 x_{ju}^2 x_{ku}^2, \\
\frac{\lambda_2 \lambda_6}{\lambda_4^2} &> \frac{v+2}{v+4}.
\end{aligned}$$

where all sum of powers and products up to ≤ 6 are zeros.

ACKNOWLEDGMENTS

We are grateful to the Editor-in-Chief, Co-Editor, Executive Editor and anonymous reviewers for the constructive comments that have led to considerable improvement in the article. We also acknowledge the facilities provided by ICAR-IASRI, New Delhi, The Graduate School, IARI, New Delhi, ICAR-CMFRI, Kochi for conducting the research. Financial assistance received in the form of Junior Research Fellowship (AD) and Lal Bahadur Shastri Young Scientist Award Project (EV) from Indian Council of Agricultural Research, Govt. of India is also duly acknowledged.

REFERENCES

- Adhikary, B. and Panda, R. (1982). On some mixed order response surface designs useful in sequential experiments. *Calcutta Statistical Association Bulletin*, 31(1-2):27–52.
- Anderson, M.J. and Whitcomb, P.J. (2016). *RSM Simplified: Optimising Processes Using Response Surface Methods for Design of Experiments*. Productivity press.
- Arshad, H.M., Ahmad, T., and Akhtar, M. (2020). Some sequential third-order response surface designs. *Communications in Statistics: Simulation and Computation*, 49(7):1872–1885.
- Arshad, H.M., Akhtar, M., and Gilmour, S.G. (2012). Augmented Box–Behnken designs for fitting third-order response surfaces. *Communications in Statistics: Theory and Methods*, 41(23):4225–4239.
- Box, G.E. and Draper, N.R. (1975). Robust designs. *Biometrika*, 62(2):347–352.
- Box, G.E. and Draper, N.R. (1987). *Empirical Model-Building and Response Surfaces*. New York, Wiley.
- Box, G.E. and Draper, N.R. (2007). *Response Surfaces, Mixtures and Ridge Analyses*. New York, Wiley.
- Cornelious, N.O. (2019a). A hypothetical sequential third-order rotatable design in five dimensions. *European International Journal of Science and Technology*, 8(4):40–50.
- Cornelious, N.O. (2019b). A sequential third-order rotatable design in four dimensions. *Current Journal of Applied Science and Technology*, 33(6):1–7.
- Cornelious, N.O. and Cruyff, M.N. (2019). A sequential third-order rotatable design of eighty points in four dimensions with a hypothetical case study. *Asian Journal of Probability and Statistics*, 4:1–9.
- Das, M.N. and Narasimham, V.L. (1962). Construction of rotatable designs through balanced incomplete block designs. *The Annals of Mathematical Statistics*, 33(4):1421–1439.
- Das, M.N., Parsad, R., and Manocha V.P. (1999). Response surface designs, symmetrical and asymmetrical, rotatable and modified. *Statistics and Applications*, 1(1):17–34.
- Dey, A. (2010). *Incomplete Block Designs*. World Scientific, Hackensack, NJ.
- Dey, A. (1969). *Some investigation on response surface designs*. Unpublished Ph.D. thesis, I.A.R.I., New Delhi.
- Draper, N.R. (1960a). A third-order rotatable design in four dimensions. *The Annals of Mathematical Statistics*, 31(4):875–877.

- Draper, N.R. (1960b). Third-order rotatable designs in three dimensions. *The Annals of Mathematical Statistics*, 865–874.
- Draper, N.R. (1962). Third-order rotatable designs in three factors: Analysis. *Technometrics*, 4(2):219–234.
- Draper, N.R. and Stoneman, D.M. (1968). Response surface designs for factors at two and three levels and at two and four levels. *Technometrics*, 10:177–192.
- Gardiner, D.A., Grandage, A.H.E., and Hader, R.J. (1959). Third-order rotatable designs for exploring response surfaces. *The Annals of Mathematical Statistics*, 30(4):1082–1096.
- Giovannitti-Jensen, A. and Myers, R.H. (1989). Graphical assessment of the prediction capability of response surface designs. *Technometrics*, 31(2):159–171.
- Hemavathi, M., Varghese, E., Shekhar, S., and Jaggi, S. (2022a). Sequential asymmetric third-order rotatable designs (SATORDs). *Journal of Applied Statistics*, 49(6):2009–2033.
- Hemavathi, M., Shekhar, S., Varghese, E., Jaggi, S., Sinha, B., and Mandal, N.K. (2022b). Theoretical developments in response surface designs: an informative review and further thoughts. *Communications in Statistics – Theory and Methods*, 51(7):2009–2033.
- Huda, S. (1982a). Some third-order rotatable designs. *Biometrical Journal*, 24(3):257–263.
- Huda, S. (1982b). Some third-order rotatable designs in three dimensions. *Annals of the Institute of Statistical Mathematics*, 34(2):365–371.
- Iwundu, M.P. and Agadaga, G.O. (2021). Sequential third-order response surface designs. *Journal of the Nigerian Statistical Association*, 33.
- Khuri, A.I. (2017). A general overview of response surface methodology. *Biomedical and Biostatistics International Journal*, 5(3):87–93. <https://doi.org/10.15406/bbij.2017.05.00133>
- Khuri, A.I. (2006). A general overview of response surface methodology. *Response Surface Methodology and Related Topics*. World Scientific Publishing, Singapore.
- Khuri, A.I. and Cornell, J.A. (1996). *Response Surfaces, Designs and Analyses*. 2nd Ed., Dekker, New York.
- Koske, J.K., Kosgei, M.K., and Mutiso, J.M. (2011). A new third order rotatable design in five dimensions through balanced incomplete block designs. *Journal of Agriculture, Science and Technology*, 13(1):157–163.
- Mehta, J.S. and Das, M.N. (1968). Asymmetric rotatable designs and orthogonal transformations. *Technometrics*, 10(2):313–322.
- Mutiso, J.M. (1998). *Specific and sequential second and third-order rotatable designs in k dimensions*. Ph. D. Thesis, Moi University.
- Mutiso, J.M. and Koske, J.K. (2007). Some third-order rotatable designs in six dimensions. *Journal of Agriculture, Science and Technology*, 9(1):78–87.
- Myers, R.H. (1971). *Response Surface Methodology*. Allyn and Bacon, Inc., Boston, MA.
- Myers, R.H., Montgomery, D.C., and Anderson-Cook, C.M. (2016). *Response Surface Methodology: Process and Product Optimisation Using Designed Experiments*. John Wiley and Sons.
- Ramchander, P.R. (1963). *Asymmetrical response surface designs*. Diploma Thesis, IARI, New Delhi.
- Rashid, F., Akram, M., Akbar, A., and Javed, A. (2017). Some new augmented fractional Box–Behnken designs. *Communications in Statistics: Theory and Methods*, 46(4):2007–2012.
- Thaker, P.J. and Das, M.N. (1961). Sequential third-order rotatable designs up to eleven factors. *Journal of the Indian Society of Agricultural Statistics*, 13:218–231.
- Varghese, E., Hemavathi, M., Dalal, A., Bhowmik, A., and Jaggi, S. (2022). SAS macro for checking the rotatability of response surface design. <https://krishi.icar.gov.in/jspui/handle/123456789/71141>.
- Zahrán, A.R., Anderson-Cook, C.M., and Myers, R.H. (2003). Fraction of design space to assess prediction capability of response surface designs. *Journal of Quality Technology*, 35(4):377–386.

A New General Class of Ridge-type Estimator in Linear Regression Models

Authors: KADRI ULAŞ AKAY  

– Department of Mathematics, Science Faculty, University of Istanbul
Istanbul, Türkiye
kulas@istanbul.edu.tr

ESRA ERTAN 

– Department of Mathematics, Science Faculty, University of Istanbul
Istanbul, Türkiye
eertan@istanbul.edu.tr

ALI ERKOÇ 

– Department of Statistics, Faculty of Arts and Sciences,
Mimar Sinan Fine Arts University, Istanbul, Türkiye
ali.erkoc@msgsu.edu.tr

FERHAT TAŞ 

– Department of Computer Science, Science Faculty, University of Istanbul
Istanbul, Türkiye
tasf@istanbul.edu.tr

Received: March 2023

Revised: January 2024

Accepted: January 2024

Abstract:

- In linear regression models, researchers have developed new biased estimators to mitigate the effects of multicollinearity instead of using the Ordinary Least Squares (OLS) estimator, which is affected by multicollinearity. In this study, we define a general class of estimators called Ridge-type estimators (RTE). The superiority of RTE over other biased estimators is investigated under the matrix mean square error criterion. In addition, two separate Monte Carlo simulation studies are conducted to compare the performance of the considered biased estimators. A numerical example is given to demonstrate the superiority of the proposed estimator over other biased estimators.

Keywords:

- *biased regression; Liu Estimator; Liu-type estimator; multicollinearity; Ridge Estimator.*

AMS Subject Classification:

- 62J07, 62F10.

1. INTRODUCTION

Regression analysis is widely used in many disciplines, including business, engineering, agriculture, and economics, to describe the statistical relationship between explanatory and response variables by using a model. The linear regression model, which assumes that the response variable is normally distributed, is one of the most commonly used statistical models. Let us consider the following linear regression model:

$$(1.1) \quad Y = X\beta + \varepsilon,$$

where Y is an $n \times 1$ vector of dependent variables, X is an $n \times p$ full column rank matrix of n observations on p independent explanatory variables, β is a $p \times 1$ vector of unknown parameters, and ε is an $n \times 1$ vector of random errors which are distributed as Normal with the mean vector 0 and the covariance matrix $\sigma^2 I$. The Ordinary Least Squares (OLS) estimator of β is given by

$$(1.2) \quad \hat{\beta}_{OLS} = (X'X)^{-1}X'Y.$$

In addition, the covariance matrix of $\hat{\beta}_{OLS}$ is obtained as $cov(\hat{\beta}_{OLS}) = \sigma^2(X'X)^{-1}$. In linear regression models, computational difficulties arise when the independent variables are collinear. The problem of multicollinearity occurs when one or more variables can be expressed as an exact or almost linear combination of the others in the data set. Multicollinearity will also provide statistical challenges if the problem aims to estimate parameters. There are many criteria to determine multicollinearity. Multicollinearity causes the diagonal elements of $(X'X)^{-1}$ to inflate, which implies that the estimated variance of $\hat{\beta}_{OLS}$ will be large. In addition, the coefficients of the OLS estimator may have wrong signs and large variances and be statistically insignificant. For such cases, alternative biased estimators have been proposed by many researchers to overcome the problems caused by the presence of multicollinearity. Issues related to these proposed biased estimators in linear regression models have been investigated and discussed in the literature by many researchers (Stein, 1956; Hoerl and Kennard, 1970; Liu, 1993, 2003; Kibria, 2003; Özkale and Kaçiranlar, 2007; Sakalhoğlu and Kaçiranlar, 2008; Yang and Chang, 2010; Kurnaz and Akay, 2015, 2018; Qasim *et al.*, 2020; Lukman *et al.*, 2019; Lukman *et al.*, 2020; Aslam and Ahmad, 2022; Zeinal and Azmoun Zavie Kivi, 2023; Üstündağ *et al.*, 2021; Ahmad and Aslam, 2022; Babar and Chand, 2022; Dawoud *et al.*, 2022; Qasim *et al.*, 2022; Shewa and Ugwuowo, 2023; Idowu *et al.*, 2023). The Ridge Estimator (RE), proposed by Hoerl and Kennard (1970), is the most significant of these estimators. The RE is defined by

$$(1.3) \quad \hat{\beta}_{RE} = (X'X + kI)^{-1}X'Y, \quad k > 0,$$

where k is a biasing parameter. On the other hand, Liu (1993) proposed the Liu Estimator (LE) combining the advantages of RE and Stein estimator. The Stein estimator was defined by Stein (1956) as follows $\hat{\beta}_S = c\hat{\beta}_{OLS}$ where $0 < c < 1$. The LE is defined as follows:

$$(1.4) \quad \hat{\beta}_{LE} = (X'X + I)^{-1}(X'Y + d\hat{\beta}_{OLS}), \quad 0 < d < 1,$$

where d is a biasing parameter. On the other hand, Lukman *et al.* (2020) noted that the estimates of the parameter d in LE are usually negative. To overcome this, model (1.1) is augmented with $-d_{ML}\hat{\beta}_{OLS} = \beta + \varepsilon'$ and then the OLS method is used.

The resulting estimator is called the Modified One-Parameter Liu (ML) Estimator and is defined as follows:

$$(1.5) \quad \hat{\beta}_{ML} = (X'X + I)^{-1}(X'X - d_{ML}I)\hat{\beta}_{OLS}, \quad 0 < d_{ML} < 1,$$

where d_{ML} is a biasing parameter. According to [Lukman et al. \(2020\)](#), this modification provides a positive value of the biasing parameter d_{ML} . However, although RE and LE are often preferred in the presence of collinearity in linear regression models, these estimators have some drawbacks. Researchers have developed estimators with two biasing parameters k and d to cover both RE and LE. For example, [Liu \(2003\)](#) introduced an estimator that is based on k and d as follows:

$$(1.6) \quad \hat{\beta}_{LTE} = (X'X + kI)^{-1}(X'Y - d\hat{\beta}^*), \quad k > 0, \quad -\infty < d < \infty,$$

where $\hat{\beta}^*$ can be any estimator of β . This estimator, which is called the Liu-type estimator, is obtained by augmenting $(-d/k^{1/2})\hat{\beta}^* = k^{1/2}\beta + \varepsilon'$ to (1.1) and then using the OLS method ([Liu, 2003](#)). As an alternative, [Özkale and Kaçiranlar \(2007\)](#) introduced a Two-Parameter Estimator (TPE) as follows:

$$(1.7) \quad \hat{\beta}_{TPE} = (X'X + kI)^{-1}(X'Y + kd\hat{\beta}_{OLS}), \quad k > 0, \quad 0 < d < 1,$$

where k and d are two biasing parameters. The TPE is a general estimator that includes the OLS, RE, and LE as special cases. As an alternative to the estimators introduced so far, [Kurnaz and Akay \(2015\)](#) proposed a general Liu-type estimator that includes estimators given by (1.2), (1.3), (1.4), (1.5), (1.6), and (1.7) estimators as special cases. The new Liu-type estimator is defined as follows:

$$(1.8) \quad \hat{\beta}_{NLTE} = (X'X + kI)^{-1}(X'Y + f(k)\hat{\beta}^*), \quad k > 0,$$

where $\hat{\beta}^*$ is any estimator of β , and $f(k)$ is a continuous function of the biasing parameter k . Similarly, NLTE is obtained by augmenting $\frac{f(k)}{k^{1/2}}\hat{\beta}^* = k^{1/2}\beta + \varepsilon'$ to (1.1) and then using the OLS method. For example, if $f(k) = -k$ and $\hat{\beta}^* = \hat{\beta}_{OLS}$, the KL estimator given by [Kibria and Lukman \(2020\)](#) is obtained. The KL estimator, which is a special case of the estimator (1.8), is defined as follows:

$$(1.9) \quad \hat{\beta}_{KL} = (X'X + kI)^{-1}(X'X - kI)\hat{\beta}_{OLS}, \quad k > 0,$$

where k is the biasing parameter. On the other hand, [Qasim et al. \(2022\)](#) proposed a Two-Step Shrinkage (TSS) estimator in the presence of multicollinearity as follows:

$$(1.10) \quad \hat{\beta}_{TSS} = (X'X + kI)^{-1}(X'X - kdI)\hat{\beta}_{OLS}, \quad k > 0, \quad 0 \leq d < 1,$$

where k and d are two biasing parameters. Note that this estimator given in (1.10) can be obtained by taking $f(k) = -kd$ and $\hat{\beta}^* = \hat{\beta}_{OLS}$ in (1.8). Furthermore, [Sakalhoğlu and Kaçiranlar \(2008\)](#) proposed another biased estimator based on RE which is given by

$$(1.11) \quad \hat{\beta}_{SK}(k, d) = (X'X + I)^{-1}(X'Y + d\hat{\beta}_{RE}), \quad k > 0, \quad -\infty < d < \infty,$$

where k and d are two biasing parameters. The estimator given in (1.11) is called a k - d class estimator and is a general estimator that includes the OLS, RE, and LEs as special cases ([Sakalhoğlu and Kaçiranlar, 2008](#)). The k - d class estimator is obtained by augmenting the

equation $d\hat{\beta}_{RE} = \beta + \varepsilon'$ to (1.1) and using the OLS method, too. Also, Yang and Chang (2010) proposed a new biased estimator based on RE as follows:

$$(1.12) \quad \hat{\beta}_{YC}(k, d) = (X'X + I)^{-1}(X'X + dI)\hat{\beta}_{RE}, \quad k > 0, \quad 0 < d < 1,$$

where k and d are two biasing parameters. The estimator given in (1.12) is obtained by augmenting $(d - k)\hat{\beta}_{RE} = \beta + \varepsilon'$ to (1.1) and using the OLS method. In addition, the YC estimator is a general estimator that includes OLS, RE, and LE as special cases. Ahmad and Aslam (2022) proposed another biased estimator similar to the YC estimator. Instead of $\hat{\beta}_{RE}$ in (1.12), they used the estimator proposed by Dorugade (2014). This estimator, called Modified New Two-Parameter Estimator (MNTPE), is given as follows:

$$(1.13) \quad \hat{\beta}_{MNTPE} = (X'X + I)^{-1}(X'X + dI)(X'X + kdI)^{-1}X'Y, \quad k > 0, \quad 0 < d < 1,$$

where k and d are two biasing parameters. Dawoud *et al.* (2022) proposed another biased estimator with the biasing parameters k and d similar to the YC estimator. They also used a similar approach applied by Ahmad and Aslam (2022). Instead of OLS in (1.7), defined by Özkale and Kaçiranlar (2007), they preferred to use the KL estimator. They defined this estimator, called the NBR estimator, as follows:

$$(1.14) \quad \hat{\beta}_{NBR} = (X'X + kI)^{-1}(X'X + kdI)(X'X + kI)^{-1}(X'X - kI)\hat{\beta}_{OLS},$$

where $k > 0$ and $0 < d < 1$ are two biasing parameters. On the other hand, Shewa and Ugwuowo (2023) proposed another biased estimator based on the KL estimator. Following the modification by Aladeitan *et al.* (2021), they proposed a new estimator called KL-MRT as follows:

$$(1.15) \quad \hat{\beta}_{KLMRT} = (X'X + kI)^{-1}(X'X - kI)(X'X + k(1 + d)I)^{-1}X'Y, \quad k \geq 0, \quad d \geq 0,$$

where k and d are two biasing parameters. On the other hand, Idowu *et al.* (2023) made a modification to the LE given by (1.4). Instead of the OLS estimator utilized in LE, they used the KL estimator given by (1.9). Their estimator called LKL is defined as follows:

$$(1.16) \quad \hat{\beta}_{LKL} = (X'X + I)^{-1}(X'X + dI)(X'X + kI)^{-1}(X'X - kI)\hat{\beta}_{OLS}, \quad k > 0, \quad 0 < d < 1,$$

where k and d are two biasing parameters. The estimators with two biasing parameters k and d have been generally developed based on RE, LE, and LTE. In particular, these estimators depend on the OLS estimator, and a more powerful estimator is preferred over the OLS estimator to minimize the effects of multicollinearity. In addition to these modifications to reduce the effects of multicollinearity, it is also necessary to consider the optimal performance of the proposed estimator. From another point of view, as the number of biasing parameters included in the estimator increases, it becomes more difficult to assess the optimal performance of the estimator because the performance of biased estimators is affected by the selection of the biasing parameter. In general, the estimates of the biasing parameters are obtained in such a way that the scalar mean square error function is minimized. Since the mean square error function is a nonlinear function of the biasing parameters, the estimates of these biasing parameters can be approximately obtained. There are many studies focusing on this issue in the literature (Hoerl and Kennard, 1970; Liu, 2003; Kibria, 2003; Yang and Chang, 2010; Sakallıoğlu and Kaçiranlar, 2008; Shukur *et al.*, 2008; Lukman *et al.*, 2020; Ahmad and Aslam, 2022; Dawoud *et al.*, 2022; Qasim *et al.*, 2020; Shewa and Ugwuowo, 2023; Idowu *et al.*, 2023).

On the other hand, estimators with two biasing parameters k and d have attracted the attention of many researchers in recent years. However, the most important problem for these estimators is that the number of these biasing parameters is large and it is also very difficult to find their optimal estimates. Although many iterative techniques have been proposed to find the optimal estimates of these biasing parameters, it is a complex process to obtain these estimates. In these cases, one of the biasing parameters can be estimated depending on the other biasing parameters or vice versa (Liu, 2003; Özkale and Kaçiranlar, 2007; Sakalhoğlu and Kaçiranlar, 2008; Yang and Chang, 2010; Ahmad and Aslam, 2022; Dawoud *et al.*, 2022; Qasim *et al.*, 2022; Shewa and Ugwuowo, 2023; Idowu *et al.*, 2023). Therefore, it can be considered that there is an unknown functional relationship between these two biasing parameters k and d . In the literature, there are some studies that examine the applications of this consideration in various other statistical models (Ertan and Akay, 2022, Akay and Ertan, 2022, and Erkoç *et al.*, 2023).

The purpose of this paper is to examine the performance of the estimator to be obtained under the hypothesis of an unknown functional relationship between these two biasing parameters k and d . In this context, we first develop a new hybrid estimator that combines the advantages of LE and RE. Then, we try to find the optimal functional relationship between the biasing parameters. With the help of the functional structure used in this hybrid estimator, it is expected that the estimated model parameter values will not be affected at large biasing parameter values k . In addition to this feature, the proposed hybrid estimator can be defined to include the estimators given by (1.2), (1.3), (1.4), (1.5), (1.11), and (1.12) estimators as special cases. In other words, it can be said that the proposed estimator forms a general class of estimators like the estimator given in (1.8). In addition, a comprehensive comparison of these two proposed classes of estimators was carried out using simulation studies.

The article is organized as follows: In Section 2, the proposed biased estimator is introduced and some properties are given. In Section 3, a general theorem is given to compare RTE and NLTE in the sense of the matrix mean square error. In Section 4, alternative approaches to determine the functional relationship between the biasing parameters are presented. Two Monte Carlo simulation studies are designed to evaluate the performances of the considered estimators in Section 5. In Section 6, the performance evaluation of all considered estimators is given in the Portland cement data. Finally, the conclusion of the study is given in Section 7.

2. A NEW GENERAL RIDGE-TYPE ESTIMATOR

To mitigate the effect of multicollinearity, researchers have made efforts to develop alternative estimators for linear regression models instead of the OLS, which are affected by collinearity between variables. Especially when the estimators given by (1.11) and (1.12) are examined, it is observed that RE, which is more resistant to collinearity effects, is used instead of the OLS estimator. However, a major disadvantage of RE is that it can result in small parameter estimates at large values of the biasing parameter k . To overcome this problem, researchers have developed hybrid estimators that combine the advantages of RE and Liu Estimators (Sakalhoğlu and Kaçiranlar, 2008; Yang and Chang, 2010). In order

to collect these estimators under a general class with the help of an unknown functional relationship that can be among the biasing parameters, we can define the new Ridge-type estimator (RTE) for β as follows:

$$(2.1) \quad \hat{\beta}_{RTE}(k) = (X'X + I)^{-1}(X'X + g(k)I)(X'X + kI)^{-1}X'Y, \quad k > 0,$$

where $g(k)$ is a continuous function of the biasing parameter k . We can obtain the estimator given in (2.1) by augmenting $(g(k) - k)\hat{\beta}_{RE} = \beta + \varepsilon'$ to model (1.1) and using the OLS method. The advantage of RTE over other estimators is that the $g(k)$ function helps us determine the optimal estimator. When we select $g(k)$ as a linear function of the biasing parameter, such as $g(k) = ak + b$ where $a, b \in R$, RTE is a general estimator that includes other biased estimators as follows: $\hat{\beta}_{RTE} = \hat{\beta}_{OLS}$ for $g(0) = 1$ where $k = 0$ and $b = 1$. $\hat{\beta}_{RTE} = \hat{\beta}_{RE}$ for $g(k) = 1$ where $a = 0$ and $b = 1$. $\hat{\beta}_{RTE} = \hat{\beta}_{LE}$ for $g(0) = b$ where b corresponds to the biasing parameter d . $\hat{\beta}_{RTE} = \hat{\beta}_{ML}$ for $g(0) = -b$ where b corresponds to the biasing parameter d_{ML} . $\hat{\beta}_{RTE} = \hat{\beta}_{YC}(k, d)$ for $g(k) = b$ where $a = 0$ and the b corresponds to the biasing parameter d . $\hat{\beta}_{RTE} = \hat{\beta}_{SK}(k, d)$ for $g(k) = k + b$ where $a = 1$ and b corresponds to the biasing parameter d . Note that the proposed estimator given in (2.1) is different from the biased estimator given in (1.7). That is, when we use $\hat{\beta}_{RE}$ instead of $\hat{\beta}^*$ in (1.8), $\hat{\beta}_{NLTE}$ does not correspond to the estimator in (2.1). Also, if $\hat{\beta}_{RE}$ is used instead of $\hat{\beta}^*$ in (1.8), the obtained estimator does not exactly correspond to the estimators proposed by Yang and Chang (2010) and Sakalhoğlu and Kaçiranlar (2008), respectively. We rewrite the model given in (1.1) in canonical form:

$$(2.2) \quad Y = Z\alpha + \varepsilon,$$

where $Z = XQ$, $\alpha = Q'\beta$, and Q is the orthogonal matrix whose columns constitute the eigenvectors of $X'X$. Then $Z'Z = Q'X'XQ = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$ where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ are the ordered eigenvalues of $X'X$. For the model (2.2), we can rewrite the above estimators in canonical form as follows:

$$(2.3) \quad \hat{\alpha}_{NLTE} = (\Lambda + kI)^{-1}(\Lambda + f(k)I)\hat{\alpha}_{OLS} = (\Lambda + kI)^{-1}(\Lambda + f(k)I)\Lambda^{-1}Z'Y = A_1Y,$$

$$(2.4) \quad \hat{\alpha}_{RTE} = (\Lambda + I)^{-1}(\Lambda + g(k)I)(\Lambda + kI)^{-1}Z'Y = A_2Y,$$

where the other existing estimators can be obtained based on the appropriate selection of $f(k)$ and $g(k)$. As known, Matrix Mean Squared Error (MMSE) and Scalar Mean Squared Error (SMSE) are the two most common methods used to detect the superiority of the estimators to each other. The MMSE and SMSE of an estimator $\tilde{\beta}$ is defined as:

$$(2.5) \quad \begin{aligned} MMSE(\tilde{\beta}) &= \text{var}(\tilde{\beta}) + [\text{bias}(\tilde{\beta})][\text{bias}(\tilde{\beta})]', \\ SMSE(\tilde{\beta}) &= \text{tr}(MMSE(\tilde{\beta})) = \text{tr}(\text{var}(\tilde{\beta})) + \text{bias}(\tilde{\beta})' \text{bias}(\tilde{\beta}), \end{aligned}$$

where $\text{var}(\tilde{\beta})$ is the variance-covariance matrix and $\text{bias}(\tilde{\beta}) = E(\tilde{\beta}) - \beta$ is the biasing vector. Let $\tilde{\beta}_1$ and $\tilde{\beta}_2$ be any two estimators of parameter β . Then, $\tilde{\beta}_2$ is superior to $\tilde{\beta}_1$ with respect to the MMSE criterion if and only if $MMSE(\tilde{\beta}_1) - MMSE(\tilde{\beta}_2)$ is a positive definite (pd) matrix. If $MMSE(\tilde{\beta}_1) - MMSE(\tilde{\beta}_2)$ is a non-negative definite matrix, then $SMSE(\tilde{\beta}_1) - SMSE(\tilde{\beta}_2) \geq 0$. But, the reverse is not always true (Theobald, 1974).

Because of the relation of $\alpha = Q'\beta$; $\hat{\beta}_{OLS}$, $\hat{\beta}_{RE}$, $\hat{\beta}_{LE}$, $\hat{\beta}_{NLTE}$, $\hat{\beta}_{SK}(k, d)$, $\hat{\beta}_{YC}(k, d)$ and $\hat{\beta}_{RTE}(k)$ have the same mean squared error values as $\hat{\alpha}_{OLS}$, $\hat{\alpha}_{RE}$, $\hat{\alpha}_{LE}$, $\hat{\alpha}_{NLTE}$, $\hat{\alpha}_{SK}(k, d)$, $\hat{\alpha}_{YC}(k, d)$ and $\hat{\alpha}_{RTE}(k)$, respectively. To compare the biased estimators mentioned above in terms of MMSE, we use the following theorems:

Theorem 2.1 (Farebrother, 1976). *Let A be a positive definite matrix, namely $A > 0$, and c be a nonzero vector. Then, $A - cc'$ is a positive definite matrix iff $c'A^{-1}c \leq 1$.*

Theorem 2.2 (Trenkler and Toutenburg, 1990). *Let $\tilde{\beta}_l = B_l Y$, $l = 1, 2$ be two homogeneous linear estimators of β and C be a positive definite matrix, where $C = B_1 B_1' - B_2 B_2'$. Then $MMSE(\tilde{\beta}_1) - MMSE(\tilde{\beta}_2) > 0$ if and only if $bias(\tilde{\beta}_2)' \left(\sigma^2 C + bias(\tilde{\beta}_1) bias(\tilde{\beta}_1)' \right)^{-1} bias(\tilde{\beta}_2) < 1$.*

3. THE SUPERIORITY OF THE PROPOSED RIDGE-TYPE ESTIMATOR

In this section, we give a general theorem to compare RTE and NLTE in the sense of MMSE. With this general theorem, it is possible to compare the above-mentioned estimators obtained by choosing different $g(k)$ and $f(k)$ functions in terms of MMSE sense. As a result of this comparison, the superiority of RTE over OLS, RE, LE, LTE, TPE, ML, TSS, and KL estimators is determined. Similarly, to determine the superiority of the RTE over the MNTP, NBR, KLMRT, and LKL estimators, the constraints on the function $g(k)$ are given.

3.1. The comparison between the RTE and the NLTE estimator

Firstly, we can compute the MMSE of $\hat{\alpha}_{NLTE} = A_1 Y$ and $\hat{\alpha}_{RTE} = A_2 Y$ as follows:

$$\begin{aligned} MMSE(\hat{\alpha}_{NLTE}) &= \sigma^2 A_1 A_1' + (A_1 Z - I) \alpha \alpha' (A_1 Z - I) \\ &= \sigma^2 (\Lambda + kI)^{-1} (\Lambda + f(k)I) \Lambda^{-1} (\Lambda + f(k)I) (\Lambda + kI)^{-1} \\ &\quad + (f(k) - k)^2 (\Lambda + kI)^{-1} \alpha \alpha' (\Lambda + kI)^{-1}, \\ MMSE(\hat{\alpha}_{RTE}) &= \sigma^2 A_2 A_2' + (A_2 Z - I) \alpha \alpha' (A_2 Z - I) \\ &= \sigma^2 (\Lambda + I)^{-1} (\Lambda + g(k)I) (\Lambda + kI)^{-1} \Lambda (\Lambda + kI)^{-1} (\Lambda + g(k)I) (\Lambda + I)^{-1} \\ &\quad + ((g(k) - k - 1)\Lambda - kI) (\Lambda + I)^{-1} (\Lambda + kI)^{-1} \alpha \alpha' (\Lambda + kI)^{-1} (\Lambda + I)^{-1} ((g(k) - k - 1)\Lambda - kI). \end{aligned}$$

Then, we can give the following theorem:

Theorem 3.1. *Let be $k > 0$ and $-\lambda_j - \frac{|\lambda_j + f(k)|(\lambda_j + 1)}{\lambda_j} < g(k) < -\lambda_j + \frac{|\lambda_j + f(k)|(\lambda_j + 1)}{\lambda_j}$ where $j = 1, 2, \dots, p + 1$. Then, $MMSE(\hat{\alpha}_{NLTE}) - MMSE(\hat{\alpha}_{RTE}) > 0$ if and only if*

$$(3.1) \quad bias(\hat{\alpha}_{RTE})' \left[\sigma^2 (A_1 A_1' - A_2 A_2') + bias(\hat{\alpha}_{NLTE}) bias(\hat{\alpha}_{NLTE})' \right]^{-1} bias(\hat{\alpha}_{RTE}) < 1,$$

where $\hat{\alpha}_{RTE}$ and $\hat{\alpha}_{NLTE}$ are two estimators for α and $bias(\hat{\alpha}_{NLTE}) = (f(k) - k)(\Lambda + kI)^{-1} \alpha$.

Proof: Using (2.3) and (2.4), we obtain

$$\begin{aligned} \text{cov}(\hat{\alpha}_{NLTE}) - \text{cov}(\hat{\alpha}_{RTE}) &= \sigma^2[A_1A_1' - A_2A_2'] \\ &= \sigma^2[(\Lambda + kI)^{-1}(\Lambda + f(k)I)\Lambda^{-1}(\Lambda + f(k)I)(\Lambda + kI)^{-1} \\ &\quad - (\Lambda + I)^{-1}(\Lambda + g(k)I)(\Lambda + kI)^{-1}\Lambda(\Lambda + kI)^{-1}(\Lambda + g(k)I)(\Lambda + I)^{-1}] \\ &= \sigma^2 \text{diag} \left\{ \frac{(\lambda_j + f(k))^2}{\lambda_j(\lambda_j + k)^2} - \frac{(\lambda_j + g(k))^2\lambda_j}{(\lambda_j + 1)^2(\lambda_j + k)^2} \right\}_{j=1}^{p+1}. \end{aligned}$$

We observe that $A_1A_1' - A_2A_2' > 0$ if and only if $(\lambda_j + 1)^2(\lambda_j + f(k))^2 - \lambda_j^2(\lambda_j + g(k))^2 > 0$. If this inequality is rearranged for $g(k)$ function, we can obtain $-\lambda_j - \frac{|\lambda_j + f(k)|(\lambda_j + 1)}{\lambda_j} < g(k) < -\lambda_j + \frac{|\lambda_j + f(k)|(\lambda_j + 1)}{\lambda_j}$ where $j = 1, 2, \dots, p + 1$. That is, the RTE is superior to NLTE when $g(k)$ function is selected as $-\lambda_j - \frac{|\lambda_j + f(k)|(\lambda_j + 1)}{\lambda_j} < g(k) < -\lambda_j + \frac{|\lambda_j + f(k)|(\lambda_j + 1)}{\lambda_j}$ where $j = 1, 2, \dots, p + 1$. Therefore, $A_1A_1' - A_2A_2'$ is the pd matrix. By Theorem 2.2, the proof is complete. \square

3.2. The comparison between the RTE and the MNTP estimator

The MMSE of $\hat{\alpha}_{MNTP} = (\Lambda + I)^{-1}(\Lambda + dI)(\Lambda + kdI)^{-1}Z'Y = A_3Y$ estimator is

$$MMSE(\hat{\alpha}_{MNTP}) = \sigma^2 A_3 A_3' + (A_3 Z - I) \alpha \alpha' (A_3 Z - I).$$

We use the MMSE difference given below to compare the MNTP and the RTE:

$$\begin{aligned} MMSE(\hat{\alpha}_{MNTP}) - MMSE(\hat{\alpha}_{RTE}) &= \sigma^2[A_3A_3' - A_2A_2'] + (A_3Z - I)\alpha\alpha'(A_3Z - I) \\ &\quad - (A_2Z - I)\alpha\alpha'(A_2Z - I). \end{aligned}$$

Then, we give the following theorem:

Theorem 3.2. Let be $k > 0$, $0 < d < 1$ and $-\lambda_j - \frac{(\lambda_j + d)(\lambda_j + k)}{(\lambda_j + kd)} < g(k) < -\lambda_j + \frac{(\lambda_j + d)(\lambda_j + k)}{(\lambda_j + kd)}$ where $j = 1, 2, \dots, p + 1$. Then, $MMSE(\hat{\alpha}_{MNTP}) - MMSE(\hat{\alpha}_{RTE}) > 0$ if and only if

$$(3.2) \quad \text{bias}(\hat{\alpha}_{RTE})' [\sigma^2(A_3A_3' - A_2A_2') + (A_3Z - I)\alpha\alpha'(A_3Z - I)]^{-1} \text{bias}(\hat{\alpha}_{RTE}) < 1,$$

where $\hat{\alpha}_{RTE}$ and $\hat{\alpha}_{MNTP}$ are two linear estimators for the parameter α .

Proof: We can obtain

$$\begin{aligned} \text{cov}(\hat{\alpha}_{MNTP}) - \text{cov}(\hat{\alpha}_{RTE}) &= \sigma^2[A_3A_3' - A_2A_2'] \\ &= \sigma^2[(\Lambda + I)^{-1}(\Lambda + dI)(\Lambda + kdI)^{-1}\Lambda(\Lambda + kdI)^{-1}(\Lambda + dI)(\Lambda + I)^{-1} \\ &\quad - (\Lambda + I)^{-1}(\Lambda + g(k)I)(\Lambda + kI)^{-1}\Lambda(\Lambda + kI)^{-1}(\Lambda + g(k)I)(\Lambda + I)^{-1}] \\ &= \sigma^2 \text{diag} \left\{ \frac{(\lambda_j + d)^2\lambda_i}{(\lambda_j + 1)^2(\lambda_j + kd)^2} - \frac{(\lambda_j + g(k))^2\lambda_j}{(\lambda_j + 1)^2(\lambda_j + k)^2} \right\}_{j=1}^p. \end{aligned}$$

We can observe that $A_3A_3' - A_2A_2' > 0$ if and only if $(\lambda_j + d)^2(\lambda_j + k)^2 - (\lambda_j + g(k))^2 \cdot (\lambda_j + kd)^2 > 0$. The RTE is superior to the MNTP estimator when $g(k)$ function is selected as $-\lambda_j - \frac{(\lambda_j + d)(\lambda_j + k)}{(\lambda_j + kd)} < g(k) < -\lambda_j + \frac{(\lambda_j + d)(\lambda_j + k)}{(\lambda_j + kd)}$ where $j = 1, 2, \dots, p + 1$. Therefore, $A_3A_3' - A_2A_2'$ is the pd matrix. By Theorem 2.2, the proof is complete. \square

3.3. The comparison between the RTE and the NBR estimator

The $MMSE$ of $\hat{\alpha}_{NBR} = (\Lambda + kI)^{-1}(\Lambda + kdI)(\Lambda + kI)^{-1}(\Lambda - kI)\Lambda^{-1}Z'Y = A_4Y$ estimator is

$$MMSE(\hat{\alpha}_{NBR}) = \sigma^2 A_4 A_4' + (A_4 Z - I)\alpha\alpha'(A_4 Z - I).$$

We use the $MMSE$ difference given below to compare the NBR and the RTE:

$$MMSE(\hat{\alpha}_{NBR}) - MMSE(\hat{\alpha}_{RTE}) = \sigma^2[A_4 A_4' - A_2 A_2'] + (A_4 Z - I)\alpha\alpha'(A_4 Z - I) - (A_2 Z - I)\alpha\alpha'(A_2 Z - I).$$

Then, we can give the following theorem:

Theorem 3.3. Let be $k > 0$, $0 < d < 1$ and $-\frac{(\lambda_j + kd)(\lambda_j + 1)|\lambda_j - k|}{\lambda_j(\lambda_j + k)} - \lambda_j < g(k) < \frac{(\lambda_j + kd)(\lambda_j + 1)|\lambda_j - k|}{\lambda_j(\lambda_j + k)} - \lambda_j$, where $j = 1, 2, \dots, p+1$. Then, $MMSE(\hat{\alpha}_{NBR}) - MMSE(\hat{\alpha}_{RTE}) > 0$ if and only if

$$(3.3) \quad bias(\hat{\alpha}_{RTE})' [\sigma^2(A_4 A_4' - A_2 A_2') + (A_4 Z - I)\alpha\alpha'(A_4 Z - I)]^{-1} bias(\hat{\alpha}_{RTE}) < 1,$$

where $\hat{\alpha}_{RTE}$ and $\hat{\alpha}_{NBR}$ are two linear estimators for α parameter.

Proof: We can obtain

$$\begin{aligned} cov(\hat{\alpha}_{NBR}) - cov(\hat{\alpha}_{RTE}) &= \sigma^2[A_4 A_4' - A_2 A_2'] \\ &= \sigma^2[(\Lambda + kI)^{-1}(\Lambda + kdI)(\Lambda + kI)^{-1}(\Lambda - kI)\Lambda^{-1}(\Lambda - kI)(\Lambda + kI)^{-1}(\Lambda + kdI)(\Lambda + kI)^{-1} \\ &\quad - (\Lambda + I)^{-1}(\Lambda + g(k)I)(\Lambda + kI)^{-1}\Lambda(\Lambda + kI)^{-1}(\Lambda + g(k)I)(\Lambda + I)^{-1}] \\ &= \sigma^2 \text{diag} \left\{ \frac{(\lambda_j + kd)^2(\lambda_j - k)^2}{\lambda_j(\lambda_j + k)^4} - \frac{(\lambda_j + g(k))^2\lambda_j}{(\lambda_j + 1)^2(\lambda_j + k)^2} \right\}_{j=1}^{p+1}. \end{aligned}$$

We can observe that $A_4 A_4' - A_2 A_2' > 0$ if and only if $(\lambda_j + kd)^2(\lambda_j - k)^2(\lambda_j + 1)^2 - (\lambda_j + k)^2(\lambda_j + g(k))^2\lambda_j^2 > 0$. From the solution of this inequality with respect to the function $g(k)$ we can derive the following condition: $-\frac{(\lambda_j + kd)(\lambda_j + 1)|\lambda_j - k|}{\lambda_j(\lambda_j + k)} - \lambda_j < g(k) < \frac{(\lambda_j + kd)(\lambda_j + 1)|\lambda_j - k|}{\lambda_j(\lambda_j + k)} - \lambda_j$, where $j = 1, 2, \dots, p+1$, $k > 0$, $0 < d < 1$. RTE outperforms the NBR estimator in terms of $MMSE$ if the function $g(k)$ is determined in a way that satisfies the condition given above. Therefore, $A_4 A_4' - A_2 A_2'$ is the pd matrix. By Theorem 2.2, the proof is complete. \square

3.4. The comparison between the RTE and the KLMRT estimator

The $MMSE$ of $\hat{\alpha}_{KLMRT} = (\Lambda + kI)^{-1}(\Lambda - kI)(\Lambda + k(1 + d)I)^{-1}Z'Y = A_5Y$ is

$$MMSE(\hat{\alpha}_{KLMRT}) = \sigma^2 A_5 A_5' + (A_5 Z - I)\alpha\alpha'(A_5 Z - I).$$

We get attention to the $MMSE$ difference given below to compare the KLMRT and the RTE:

$$MMSE(\hat{\alpha}_{KLMRT}) - MMSE(\hat{\alpha}_{RTE}) = \sigma^2[A_5 A_5' - A_2 A_2'] + (A_5 Z - I)\alpha\alpha'(A_5 Z - I) - (A_2 Z - I)\alpha\alpha'(A_2 Z - I).$$

Then, we can give the following theorem:

Theorem 3.4. Let be $k > 0$, $d > 0$ and $-\lambda_j - \sqrt{\frac{(\lambda_j - k)^2(\lambda_j + 1)^2}{\lambda_j(\lambda_j + k(1+d))^2}} < g(k) < -\lambda_j + \sqrt{\frac{(\lambda_j - k)^2(\lambda_j + 1)^2}{\lambda_j(\lambda_j + k(1+d))^2}}$, $j = 1, 2, \dots, p + 1$. Then, $MMSE(\hat{\alpha}_{KLMRT}) - MMSE(\hat{\alpha}_{RTE}) > 0$ if and only if

$$(3.4) \quad bias(\hat{\alpha}_{RTE})' [\sigma^2(A_5A_5' - A_1A_1') + (A_5Z - I)\alpha\alpha'(A_5Z - I)]^{-1} bias(\hat{\alpha}_{RTE}) < 1,$$

where $\hat{\alpha}_{RTE}$ and $\hat{\alpha}_{KLMRT}$ are two linear estimators for the parameter α .

Proof: We can obtain

$$\begin{aligned} cov(\hat{\alpha}_{KLMRT}) - cov(\hat{\alpha}_{RTE}) &= \sigma^2[A_5A_5' - A_2A_2'] \\ &= \sigma^2 \left[(\Lambda + kI)^{-1}(\Lambda - kI)(\Lambda + k(1+d)I)^{-1}\Lambda(\Lambda + k(1+d)I)^{-1}(\Lambda - kI)(\Lambda + kI)^{-1} \right. \\ &\quad \left. - (\Lambda + I)^{-1}(\Lambda + g(k)I)(\Lambda + kI)^{-1}\Lambda(\Lambda + kI)^{-1}(\Lambda + g(k)I)(\Lambda + I)^{-1} \right] \\ &= \sigma^2 \text{diag} \left\{ \frac{(\lambda_j - k)^2}{(\lambda_j + k)^2(\lambda_j + k(1+d))^2} - \frac{(\lambda_j + g(k))^2\lambda_j}{(\lambda_j + 1)^2(\lambda_j + k)^2} \right\}_{j=1}^{p+1}. \end{aligned}$$

We observe that $A_5A_5' - A_2A_2' > 0$ if and only if $(\lambda_j - k)^2(\lambda_j + 1)^2 - (\lambda_j + g(k))^2(\lambda_j + k(1+d))^2\lambda_j > 0$. So, the RTE is superior to the KLMRT estimator when $g(k)$ function is selected as $-\lambda_j - \sqrt{\frac{(\lambda_j - k)^2(\lambda_j + 1)^2}{(\lambda_j + k(1+d))^2\lambda_j}} < g(k) < -\lambda_j + \sqrt{\frac{(\lambda_j - k)^2(\lambda_j + 1)^2}{(\lambda_j + k(1+d))^2\lambda_j}}$, $j = 1, 2, \dots, p + 1$. Therefore, $A_5A_5' - A_2A_2'$ is the pd matrix. By Theorem 2.2, the proof is complete. \square

3.5. The comparison between the RTE and the LKL estimator

The $MMSE$ of $\hat{\alpha}_{LKL} = (\Lambda + kI)^{-1}(\Lambda + dI)(\Lambda + kI)^{-1}(\Lambda - kI)\Lambda^{-1}Z'Y = A_6Y$ estimator is

$$MMSE(\hat{\alpha}_{LKL}) = \sigma^2 A_6A_6' + (A_6Z - I)\alpha\alpha'(A_6Z - I).$$

We use the $MMSE$ difference given below to compare the LKL estimator and RTE,

$$MMSE(\hat{\alpha}_{LKL}) - MMSE(\hat{\alpha}_{RTE}) = \sigma^2[A_6A_6' - A_2A_2'] + (A_6Z - I)\alpha\alpha'(A_6Z - I) - (A_2Z - I)\alpha\alpha'(A_2Z - I).$$

Then, we give the following theorem:

Theorem 3.5. Let be $k > 0$, $0 < d < 1$ and $-\lambda_j - \frac{|\lambda_j - k|(\lambda_j + d)}{\lambda_j} < g(k) < -\lambda_j + \frac{|\lambda_j - k|(\lambda_j + d)}{\lambda_j}$ where $j = 1, 2, \dots, p + 1$. Then, $MMSE(\hat{\alpha}_{LKL}) - MMSE(\hat{\alpha}_{RTE}) > 0$ if and only if

$$(3.5) \quad bias(\hat{\alpha}_{RTE})' [\sigma^2(A_6A_6' - A_2A_2') + (A_6Z - I)\alpha\alpha'(A_6Z - I)]^{-1} bias(\hat{\alpha}_{RTE}) < 1,$$

where $\hat{\alpha}_{RTE}$ and $\hat{\alpha}_{LKL}$ are two linear estimators for the parameter α .

Proof: We can obtain

$$\begin{aligned} cov(\hat{\alpha}_{LKL}) - cov(\hat{\alpha}_{RTE}) &= \sigma^2[A_6A_6' - A_2A_2'] \\ &= \sigma^2 \left[(\Lambda + I)^{-1}(\Lambda + dI)(\Lambda + kI)^{-1}(\Lambda - kI)\Lambda^{-1}(\Lambda - kI)(\Lambda + kI)^{-1}(\Lambda + dI)(\Lambda + I)^{-1} \right. \\ &\quad \left. - (\Lambda + I)^{-1}(\Lambda + g(k)I)(\Lambda + kI)^{-1}\Lambda(\Lambda + kI)^{-1}(\Lambda + g(k)I)(\Lambda + I)^{-1} \right] \\ &= \sigma^2 \text{diag} \left\{ \frac{(\lambda_j + d)^2(\lambda_j - k)^2}{\lambda_j(\lambda_j + 1)^2(\lambda_j + k)^2} - \frac{(\lambda_j + g(k))^2\lambda_j}{(\lambda_j + 1)^2(\lambda_j + k)^2} \right\}_{j=1}^{p+1}. \end{aligned}$$

We can observe that $A_6A_6' - A_2A_2' > 0$ if and only if $(\lambda_j + d)^2(\lambda_j - k)^2 - (\lambda_j + g(k))^2\lambda_j^2 > 0$ where $j = 1, 2, \dots, p + 1$. The RTE is superior to the MNTP estimator when $g(k)$ function is selected as $-\lambda_j - \frac{|\lambda_j - k|(\lambda_j + d)}{\lambda_j} < g(k) < -\lambda_j + \frac{|\lambda_j - k|(\lambda_j + d)}{\lambda_j}$ where $j = 1, 2, \dots, p + 1$. Therefore, $A_6A_6' - A_2A_2'$ is the pd matrix. By Theorem 2.2, the proof is complete. \square

4. DETERMINATION OF $g(k)$ FUNCTION

Determining the optimal estimate of the biasing parameter is very important because it is associated with the performance of the biased estimator. For practitioners, this is a complex process. This process becomes even more complicated for a biased estimator with biasing parameters k and d . Many different techniques have been proposed by many researchers to estimate the biasing parameter(s) (Hoerl and Kennard, 1970; Liu, 1993, 2003; Kibria, 2003; Yang and Chang, 2010; Sakalioğlu and Kaçiranlar, 2008; Shukur *et al.*, 2008; Ahmad and Aslam, 2022; Dawoud *et al.*, 2022; Qasim *et al.*, 2022; Shewa and Ugwuowo, 2023).

The main advantage of RTE over the estimators with two biasing parameters k and d is that there is a functional relationship between the biasing parameters. The performance of the proposed RTE is based on $g(k)$, and therefore the single biasing parameter is k . Note that different choices of the $g(k)$ function lead to different estimators. To find the optimal $g(k)$ function, let's take the derivative of $SMSE(\hat{\alpha}_{RTE})$ depending on k . The $SMSE(\hat{\alpha}_{RTE})$ is calculated using (2.4) as follows:

$$(4.1) \quad SMSE(\hat{\alpha}_{RTE}) = \sigma^2 \sum_{j=1}^{p+1} \frac{(\lambda_j + g(k))^2 \lambda_j}{(\lambda_j + 1)^2 (\lambda_j + k)^2} + \sum_{j=1}^{p+1} \frac{((g(k) - k - 1)\lambda_j - k)^2 \alpha_j^2}{(\lambda_j + 1)^2 (\lambda_j + k)^2}.$$

Note that Equation (4.1) is a function of the k parameter; that is, $h(k) = SMSE(\hat{\alpha}_{RTE})$. We can find $h'(k)$ as follows:

$$h'(k) = \sum_{j=1}^{p+1} \frac{2[\lambda_j(\lambda_j - g'(k)\lambda_j - g'(k)k + g(k))][\alpha_j^2((k + 1 - g(k))\lambda_j + k) - \sigma^2(\lambda_j + g(k))]}{(\lambda_j + 1)^2(\lambda_j + k)^3}.$$

In case $h'(k) = 0$, there are two scenarios:

Fact 1. $\lambda_j(\lambda_j - g'(k)\lambda_j - g'(k)k + g(k)) = 0$ differential equation is found. Then, we have

$$(4.2) \quad g(k) = ck + (c - 1)\lambda_j,$$

where c is the constant of integration.

Fact 2. $\alpha_j^2((k + 1 - g(k))\lambda_j + k) - \sigma^2(\lambda_j + g(k)) = 0$ equation is found. Here, $g(k)$ is obtained as follows:

$$(4.3) \quad g(k) = \frac{(1 + \lambda_j)\alpha_j^2}{\sigma^2 + \lambda_j\alpha_j^2}k + \frac{(\alpha_j^2 - \sigma^2)}{\sigma^2 + \lambda_j\alpha_j^2}\lambda_j \quad \text{or} \quad g(k) = \frac{(1 + \lambda_j)\alpha_j^2}{\sigma^2 + \lambda_j\alpha_j^2}k + \left(\frac{(1 + \lambda_j)\alpha_j^2}{\sigma^2 + \lambda_j\alpha_j^2} - 1 \right)\lambda_j,$$

where $j = 1, 2, \dots, p + 1$. Based on the first and second facts, it can be said that the selection of $g(k)$ as a linear function of the biasing parameter k is appropriate. Also, $g(k)$ which is obtained in Fact 2, is a solution of the differential equation, which is obtained in Fact 1. Here, depending on the functions obtained in Fact 1 and Fact 2, we can observe the following results: Firstly, note that $g(k)$ given in (4.2) and (4.3) makes $SMSE(\hat{\alpha}_{RTE})$ function approximately minimum for a given j value. So, the determination of $g(k)$ depends on the eigenvalues of $X'X$, the unknown α parameter, and the estimate of the biasing parameter k . In other words, many $g(k)$ functions can be determined depending on the functional relationship given in (4.2) and (4.3). For example, the following functional relationships can be given to determine $g(k)$ in this sense:

$$(4.4) \quad g(k) = \frac{(1 + \lambda_{\min})\alpha_{\min}^2}{\hat{\sigma}^2 + \lambda_{\max}\alpha_{\max}^2}k + \left(\frac{(1 + \lambda_{\min})\alpha_{\min}^2}{\hat{\sigma}^2 + \lambda_{\max}\alpha_{\max}^2} - 1 \right)\lambda_{\min},$$

$$(4.5) \quad g(k) = \frac{(1 + \lambda_{\min})\alpha_{\min}^2}{\hat{\sigma}^2 + \lambda_{\max}\alpha_{\max}^2}k + \min\left(\frac{\alpha^2 - \hat{\sigma}^2}{\hat{\sigma}^2 + \lambda_{\max}\alpha_{\max}^2} \right)\lambda_{\min},$$

$$(4.6) \quad g(k) = \frac{(1 + \lambda_{\min})\alpha_{\min}^2}{\hat{\sigma}^2 + \lambda_{\max}\alpha_{\max}^2}k + \min\left(\frac{\alpha^2 - \hat{\sigma}^2}{\hat{\sigma}^2 + \lambda\alpha^2} \right)\lambda_{\min},$$

where α_{\min}^2 and α_{\max}^2 are defined as the minimum and maximum value of α_j^2 , $j = 1, 2, \dots, p + 1$, respectively. Similarly, λ_{\min} and λ_{\max} indicate the minimum and maximum values of the eigenvalues of $X'\hat{W}X$, respectively.

In this study, to determine the optimal $g(k)$ function, we examined only the first-degree polynomial functions such as those given in equations from (4.4) to (4.6). Note that it is clear that $g(k)$ can be selected as any continuous function of k . However, the proposed estimator depends on a single biasing parameter k . In this case, we should use an appropriate estimator of k to control the conditioning of the $X'X$ matrix. Since the proposed estimator depends on a parameter k , a suitable estimator of k can be used, as given in Kibria (2003). In addition to the previously proposed estimators, we can use the following estimators to estimate k : $\hat{k}_{RTE} = \frac{p\sigma^2\alpha_{\min}^2}{n}$, $\hat{k}_{RTE} = \frac{p\sigma^2\lambda_{\min}}{n\lambda_{\max}}$, $\hat{k}_{RTE} = \frac{\sigma^2}{n\sum_{j=1}^{p+1}\lambda_j\alpha_j^2}$, $\hat{k}_{RTE} = \frac{p\sigma^2}{n\alpha_{\max}^2}$, $\hat{k}_{RTE} = \frac{p\sigma^2\min(\lambda_j\alpha_j^2)}{n\max(\lambda_j\alpha_j^2)}$, $\hat{k}_{RTE} = \frac{\lambda_{\max} + \lambda_{\min}}{p}$, $\hat{k}_{RTE} = \frac{\lambda_{\max} - \lambda_{\min}}{p}$ where $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p - 1}$.

5. THE MONTE CARLO SIMULATION STUDIES

In this section, we designed two separate Monte Carlo simulations to examine the performance of the proposed biased estimator relative to other existing estimators in linear regression models. In the first design, we investigated the effects of sample size (n), the degree of the collinearity (ρ), the number of explanatory variables (p), and the variance (σ^2) on the performances of OLS, RE, LE, LTE, SK, YC, MNTP, NBR, ML, TSS, KLMRT, LKL estimators and RTEs. In the second simulation design, we examined RTE and NLTE performances for each of n , p , ρ , and σ^2 values at certain values of k . For both simulation designs, we generate the explanatory variables by following [McDonald and Galarneau \(1975\)](#) and [Kibria \(2003\)](#) as

$$(5.1) \quad x_{ij} = (1 - \rho^2)^{1/2} w_{ij} + \rho w_{ip+1}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p,$$

where w_{ij} are independent standard normal pseudo-random numbers and ρ is specified in such a way that the correlation between any two variables is given by ρ^2 . These variables are standardized such that $X'X$ is a correlation matrix. Four different sets of correlations are investigated corresponding to $\rho = 0.8, 0.9$ and 0.99 . The response variable is generated by

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

where $\varepsilon_i \sim N(0, \sigma^2)$ and β_0 considered to be zero. For different comparisons of the error term, the value of σ^2 is considered to be 0.5, 1, 5, and 10. For each set of explanatory variables, the real parameter vector β is chosen as the normalized eigenvector corresponding to the largest eigenvalue of $X'X$ so that $\beta'\beta = 1$. The sample size n is taken to be 50, 100, and 200. The number of explanatory variables is chosen as $p = 2, 4, 8$, and 12.

In the simulation and application sections, the estimates of the biasing parameters for RE, LE, LTE, SK, YC, MNTP, NBR, ML, TSS, KLMRT, and LKL are chosen based on the best estimators suggested in the literature ([Kibria, 2003](#); [Liu, 2003](#); [Qasim et al., 2020](#); [Sakalhoğlu and Kaçranlar, 2008](#); [Yang and Chang, 2010](#); [Ahmad and Aslam, 2022](#); [Dawoud et al., 2022](#); [Idowu et al., 2023](#); [Lukman et al., 2020](#); [Qasim et al., 2022](#); [Shewa and Ugwuowo, 2023](#)).

To estimate the biasing parameter k in RE, [Kibria \(2003\)](#) proposed the best estimates of k as follows, $\hat{k}_{RE} = \frac{\hat{\sigma}^2}{(\prod_{j=1}^{p+1} \hat{\alpha}_j^2)^{\frac{1}{p+1}}}$ where $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-p-1}$. Based on the results given by [Qasim et al. \(2020\)](#), we use the best estimation of d in LE as

$$\hat{d}_{LE} = \max \left(0, \min \left(\frac{\hat{\alpha}_j^2 - \hat{\sigma}^2}{\max \left(\frac{\hat{\sigma}^2}{\lambda_j} + \hat{\alpha}_{\max}^2 \right)} \right) \right).$$

On the other hand, k_{LTE} and d_{LTE} in LTE are estimated by using the methods given by [Liu \(2003\)](#). [Sakalhoğlu and Kaçranlar \(2008\)](#) and [Yang and Chang \(2010\)](#) did not provide a specific technique for estimating the biasing parameters k and d for SK and YC estimators, respectively. Therefore, we used \hat{k}_{RE} as an estimate of k for the SK estimator. Also, the estimate of the biasing parameter d was determined in such a way that $SMSE(\hat{\alpha}_{SK})$ was minimized. Moreover, we used two methods proposed by [Huang and Yang \(2014\)](#) to estimate

the biasing parameters of the YC estimator. Huang and Yang (2014) proposed two methods. We referred to these methods as (K1, D1) and (K2, D2) (Huang and Yang, 2014). We used these methods by adapting them for the YC estimator in linear regression models. As a result, the estimator obtained with (K1, D1) indicated YC I, and the estimator obtained with (K2, D2) indicated YC II. Moreover, for the MNTP, NBR, ML, TSS, KLMRT, and LKL estimators, the iterative techniques from the relevant papers are used together with the optimal biasing parameters. Since there are many combinations to determine k and $g(k)$ functions in RTE, we only report the simulation results for the following k estimates and $g(k)$ functions:

$$\begin{aligned} \text{RTE I: } \hat{k}_{RTE\text{ I}} &= \frac{p\sigma^2\alpha_{\min}^2}{n} \text{ and } g(k) = \frac{(1+\lambda_{\min})\alpha_{\min}^2}{\hat{\sigma}^2+\lambda_{\max}\alpha_{\max}^2}k + \left(\frac{(1+\lambda_{\min})\alpha_{\min}^2}{\hat{\sigma}^2+\lambda_{\max}\alpha_{\max}^2} - 1\right)\lambda_{\min}, \\ \text{RTE II: } \hat{k}_{RTE\text{ II}} &= \frac{p\sigma^2 \min(\lambda_j \alpha_j^2)}{n} \text{ and } g(k) = \frac{(1+\lambda_{\min})\alpha_{\min}^2}{\hat{\sigma}^2+\lambda_{\max}\alpha_{\max}^2}k + \min\left(\frac{\alpha^2-\hat{\sigma}^2}{\hat{\sigma}^2+\lambda_{\max}\alpha_{\max}^2}\right)\lambda_{\min}, \\ \text{RTE III: } \hat{k}_{RTE\text{ III}} &= \frac{p\sigma^2 \min(\lambda_j \alpha_j^2)}{n \max(\lambda_j \alpha_j^2)} \text{ and } g(k) = \frac{(1+\lambda_{\min})\alpha_{\min}^2}{\hat{\sigma}^2+\lambda_{\max}\alpha_{\max}^2}k + \min\left(\frac{\alpha_j^2-\hat{\sigma}^2}{\hat{\sigma}^2+\lambda_j \alpha_j^2}\right)\lambda_{\min}, \\ \text{RTE IV: } \hat{k}_{RTE\text{ IV}} &= \frac{\lambda_{\max}+\lambda_{\min}}{p} \text{ and } g(k) = \frac{(1+\lambda_{\min})\alpha_{\min}^2}{n(\hat{\sigma}^2+\lambda_{\max}\alpha_{\max}^2)}k + \left(\frac{(1+\lambda_{\min})\alpha_{\min}^2}{n(\hat{\sigma}^2+\lambda_{\max}\alpha_{\max}^2)} - 1\right)\lambda_{\min}. \end{aligned}$$

The performance of the estimated MSEs (EMSEs) is used as the basis for comparison of the proposed estimators calculated for an estimator $\hat{\beta}$ of β as follows:

$$(5.2) \quad EMSE(\hat{\beta}) = \frac{1}{N} \sum_{r=1}^N (\hat{\beta}_r - \beta)' (\hat{\beta}_r - \beta),$$

where $(\hat{\beta}_r - \beta)$ is the difference between the estimated and true parameter vectors at r -th replication, and N is the number of replications. For each case of n , p , σ^2 , and ρ , the experiment was replicated 2000 times by generating response variables using R programming. The results are given in Tables 1–4 where the bold numbers show the smallest EMSE values. In addition, the signs (*), (**), and (***) show the first, second, and third smallest EMSE values in each row, respectively.

Based on Tables 1–4, we can conclude that the degree of correlation, number of explanatory variables, sample size, and variance have different effects on all estimators in the simulation. Several observations can be obtained as follows:

1. When the number of observations n and σ^2 are kept constant, it is observed that as the number of variables increased, generally, the EMSE values of the estimators tend to increase for models with low correlation variables and to decrease for models with high correlation. However, it is seen that in the increasing trend of EMSE values, the slopes of the proposed estimators RTE I, RTE II, RTE III, and RTE IV are much lower than the other existing estimators.
2. It is observed that when the number of variables p , n , and σ^2 are kept constant, as the correlations of the variables increase, the EMSE values of all estimators tend to decrease in general. However, the RTE I is not as dramatically affected by the increase in the correlation between the independent variables compared to the other existing estimators. Based on this situation, it can be concluded that RTE I has a robust structure depending on low or high correlation.

Table 1: The EMSE values of the estimators for the model when $p = 2$.

σ^2	n	ρ	OLS	RE	LE	YC I	YC II	SK	LTE	MNTP	NBR	ML	TSS	KLMRT	LKL	RTE I	RTE II	RTE III	RTE IV
0.5	50	0.8	1.618	0.777	0.459	1.005	0.945	0.59	0.936	0.562	0.952	0.551	0.977	1.130	0.957	0.336*	0.38**	0.417***	0.525
1	50	0.8	3.362	1.189	0.733	1.897	1.794	0.963	1.905	1.028	1.842	0.727	1.867	1.141	1.843	0.415*	0.568***	0.592	0.546**
5	50	0.8	15.906	3.584	2.601	8.143	7.851	3.405	9.008	4.392	8.381	2.133	8.154	1.204***	8.237	0.947**	1.422	1.641	0.811*
10	50	0.8	33.336	7.192	5.255	17.225	16.474	6.966	19.286	9.123	17.965	3.993	17.255	1.332***	17.46	1.252**	1.928	2.93	1.058*
0.5	50	0.9	2.411	0.839	0.420	1.321	1.276	0.569	1.258	0.634	1.256	0.757	1.285	1.020	1.257	0.284*	0.325**	0.346***	0.454
1	50	0.9	5.001	1.314	0.619	2.626	2.56	0.93	2.636	1.212	2.579	1.271	2.600	1.030	2.579	0.372*	0.472**	0.491	0.49***
5	50	0.9	24.621	4.372	3.220	12.272	12.015	3.401	13.022	5.519	12.619	4.99	12.312	1.113***	12.478	0.882**	1.225	1.502	0.781*
10	50	0.9	50.032	7.954	4.370	24.739	24.087	6.421	26.396	11.049	25.349	9.838	24.845	1.224***	25.236	1.161**	1.729	2.549	1.127*
0.5	50	0.99	32.129	2.127	0.253	15.181	15.201	0.513	15.179	4.105	15.178	14.331	15.18	0.369	15.179	0.229*	0.23**	0.232***	0.364
1	50	0.99	64.220	2.882	0.370	30.653	30.609	0.689	30.675	8.207	30.674	28.759	30.659	0.434	30.675	0.329*	0.34**	0.344***	0.414
5	50	0.99	330.294	8.046	1.259	157.766	157.492	1.895	158.013	41.855	158.007	147.428	157.838	0.942***	158.013	0.781**	0.958	1.132	0.779*
10	50	0.99	688.415	16.812	2.472	338.238	337.655	4.038	338.758	89.554	338.749	315.882	338.397	1.614	338.759	0.978*	1.35***	2.115	1.259**
0.5	100	0.8	1.656	0.750	0.456	0.994	0.938	0.567	0.927	0.552	0.942	0.551	0.965	1.119	0.943	0.324*	0.372**	0.403***	0.513
1	100	0.8	3.426	1.145	0.699	1.925	1.822	0.917	1.927	1.033	1.865	0.763	1.894	1.129	1.868	0.412*	0.571***	0.573	0.544**
5	100	0.8	16.554	3.701	2.605	8.336	7.987	3.455	9.219	4.471	8.596	2.288	8.359	1.199***	8.441	0.992**	1.567	1.743	0.798*
10	100	0.8	31.583	6.382	4.756	15.488	14.833	6.19	17.442	8.262	16.186	3.928	15.529	1.304**	15.755	1.312***	2.087	2.85	1.023*
0.5	100	0.9	2.296	0.824	0.424	1.26	1.209	0.572	1.2	0.616	1.195	0.704	1.223	1.039	1.197	0.280*	0.329**	0.35***	0.454
1	100	0.9	4.828	1.302	0.642	2.537	2.434	0.941	2.561	1.195	2.487	1.147	2.505	1.051	2.485	0.365*	0.498***	0.509	0.477**
5	100	0.9	23.645	4.422	2.415	11.858	11.605	3.531	12.601	5.447	12.287	4.528	11.896	1.131***	12.04	0.961**	1.404	1.627	0.792*
10	100	0.9	45.931	7.982	4.472	22.675	22.214	6.535	24.381	10.456	23.447	8.439	22.797	1.231**	23.166	1.3***	1.923	2.817	1.149*
0.5	100	0.99	28.916	2.003	0.249	13.852	13.85	0.496	13.852	3.779	13.85	12.98	13.851	0.39	13.852	0.225*	0.227**	0.23***	0.364
1	100	0.99	58.151	2.905	0.373	28.054	28.014	0.695	28.078	7.579	28.075	26.089	28.061	0.45	28.077	0.332*	0.341**	0.346***	0.414
5	100	0.99	295.883	8.379	1.393	138.579	138.274	2.152	138.815	37.168	138.805	128.047	138.651	0.963***	138.812	0.908**	1.061	1.256	0.827*
10	100	0.99	566.008	14.678	2.539	264.646	264.037	3.807	265.197	70.921	265.174	244.384	264.817	1.522	265.187	1.092*	1.477***	2.23	1.268**
0.5	200	0.8	1.946	0.792	0.445	1.124	1.062	0.575	1.059	0.588	1.064	0.606	1.091	1.086	1.067	0.306*	0.362**	0.385***	0.486
1	200	0.8	3.819	1.215	0.683	2.069	1.974	0.937	2.087	1.07	2.019	0.882	2.041	1.094	2.018	0.406*	0.558***	0.563	0.525**
5	200	0.8	18.927	3.905	2.421	9.432	9.118	3.396	10.251	4.719	9.67	2.976	9.456	1.178***	9.555	0.97**	1.575	1.681	0.747*
10	200	0.8	39.471	7.575	4.818	19.579	19.033	6.747	21.442	9.769	20.203	5.94	19.674	1.276**	19.969	1.498***	2.39	3.157	1.123*
0.5	200	0.9	2.647	0.860	0.403	1.416	1.374	0.573	1.365	0.651	1.358	0.823	1.381	0.993	1.355	0.264*	0.31**	0.328***	0.432
1	200	0.9	5.335	1.278	0.626	2.699	2.627	0.876	2.717	1.222	2.653	1.365	2.672	1.003	2.653	0.372*	0.493	0.492***	0.476**
5	200	0.9	27.563	4.503	2.171	13.77	13.507	3.337	14.429	5.898	14.152	6.125	13.814	1.091***	13.98	0.986**	1.432	1.553	0.777*
10	200	0.9	52.921	7.628	4.079	25.52	25.09	5.905	27.057	12.545	26.533	11.137	25.669	1.193**	26.081	1.443***	2.109	2.727	1.145*
0.5	200	0.99	30.667	2.026	0.245	14.453	14.447	0.477	14.454	3.924	14.453	13.578	14.453	0.376	14.454	0.219*	0.223**	0.226***	0.358
1	200	0.99	62.054	2.855	0.38	29.61	29.564	0.664	29.631	7.973	29.629	27.617	29.617	0.449	29.631	0.343*	0.352**	0.356***	0.418
5	200	0.99	307.111	8.671	1.345	147.727	147.469	2.187	147.983	39.423	147.973	137.132	147.806	0.958**	147.98	0.997***	1.133	1.237	0.810*
10	200	0.99	628.440	15.042	2.577	298.862	298.277	3.799	299.384	79.68	299.37	277.161	299.021	1.597***	299.38	1.309**	1.694	2.32	1.294*

Table 2: The EMSE values of the estimators for the model when $p = 4$.

σ^2	n	ρ	OLS	RE	LE	YC I	YC II	SK	LTE	MNTP	NBR	ML	TSS	KLMRT	LKL	RTE I	RTE II	RTE III	RTE IV
0.5	50	0.8	5.159	0.993	0.444	2.857	2.038	0.677	3.247	0.689	2.811	0.367	2.811	0.693	2.811	0.211*	0.402	0.313***	0.249**
1	50	0.8	10.701	1.722	0.833	5.873	4.312	1.26	6.777	1.342	5.87	0.656	5.867	0.749	5.875	0.332**	0.722	0.532***	0.308*
5	50	0.8	53.403	5.821	3.717	29.135	21.639	5.221	33.801	6.355	29.374	2.783	29.298	1.183***	29.402	0.975**	2.158	1.801	0.745*
10	50	0.8	110.123	11.571	7.527	60.821	45.317	10.557	69.948	13.014	61.312	5.711	61.159	1.733***	61.359	1.411**	3.006	3.29	1.315*
0.5	50	0.9	9.815	1.608	0.365	5.404	3.867	0.767	4.329	0.761	5.378	0.546	5.379	0.441	5.379	0.151*	0.307	0.26***	0.216**
1	50	0.9	19.185	2.462	0.629	10.326	7.448	1.244	8.3	1.42	10.329	0.967	10.329	0.544	10.333	0.24*	0.519	0.405***	0.274**
5	50	0.9	96.371	8.474	2.754	51.882	37.645	4.814	41.907	6.887	52.004	4.367	51.988	1.341	52.025	0.732**	1.521	1.293***	0.675*
10	50	0.9	192.27	15.905	5.521	103.568	74.457	9.336	83.282	13.76	103.826	8.783	103.79	2.362	103.867	1.088*	2.208***	2.387	1.226**
0.5	50	0.99	129.096	8.632	0.15	69.259	50.331	1.215	47.186	3.364	69.487	9.151	69.475	0.492	69.588	0.121*	0.132**	0.135***	0.184
1	50	0.99	257.941	13.235	0.249	138.182	100.785	1.66	93.728	6.657	138.569	16.563	138.543	0.989	138.765	0.197*	0.219**	0.223***	0.238
5	50	0.99	1276.487	42.965	1.042	689.326	503.699	4.917	470.116	33.16	691.244	85.176	691.176	4.961	692.19	0.591*	0.806***	0.88	0.646**
10	50	0.99	2583.039	78.338	1.945	1377.057	1020.014	8.466	941.009	66.43	1382.205	198.231	1382.19	10.056	1384.479	0.791*	1.217***	1.578	1.096**
0.5	100	0.8	4.942	0.966	0.449	2.733	1.927	0.669	2.912	0.689	2.683	0.312	2.685	0.704	2.682	0.196*	0.409	0.305***	0.259**
1	100	0.8	10.159	1.623	0.812	5.641	3.995	1.224	6.05	1.325	5.614	0.503	5.616	0.752	5.616	0.285*	0.734	0.483***	0.3**
5	100	0.8	49.748	5.627	3.746	26.716	19.237	5.213	29.431	6.134	26.839	2.075	26.805	1.193***	26.848	0.947**	2.466	1.702	0.745*
10	100	0.8	104.954	11.041	7.567	57.643	42.168	10.48	62.764	12.824	57.906	4.439	57.834	1.744***	57.924	1.402**	3.569	3.083	1.252*
0.5	100	0.9	8.317	1.417	0.391	4.597	3.225	0.741	3.29	0.735	4.56	0.439	4.562	0.491	4.559	0.154*	0.336	0.273***	0.228**
1	100	0.9	16.407	2.225	0.669	8.929	6.365	1.226	6.46	1.375	8.917	0.759	8.918	0.577	8.918	0.229*	0.576	0.419***	0.276**
5	100	0.9	79.538	7.14	2.886	42.363	29.83	4.609	30.384	6.337	42.427	3.251	42.41	1.251***	42.432	0.749**	1.788	1.289	0.657*
10	100	0.9	160.483	13.853	5.774	85.528	61.062	9.235	61.882	12.815	85.688	6.445	85.656	2.101***	85.699	1.132*	2.667	2.344	1.137**
0.5	100	0.99	80.933	6.319	0.167	43.118	30.476	1.063	28.578	2.224	43.129	5.561	43.129	0.404	43.133	0.123*	0.143**	0.144***	0.186
1	100	0.99	163.625	10.569	0.27	87.433	62.725	1.687	58.019	4.473	87.446	11.806	87.445	0.792	87.451	0.199*	0.232**	0.235***	0.237
5	100	0.99	815.854	31.082	1.087	435.566	311.746	4.172	287.582	22.013	435.567	51.586	435.549	3.984	435.571	0.637**	0.816***	0.857	0.62*
10	100	0.99	1588.408	53.5	2.193	841.228	599.104	7.264	555.067	42.624	841.296	99.113	841.281	7.959	841.328	0.974*	1.44***	1.703	1.166**
0.5	200	0.8	4.435	0.961	0.505	2.57	1.778	0.731	2.198	0.734	2.514	0.283**	2.515	0.804	2.512	0.205*	0.465	0.326	0.29***
1	200	0.8	8.575	1.487	0.883	4.795	3.385	1.247	4.218	1.327	4.751	0.39***	4.754	0.843	4.75	0.272*	0.805	0.474	0.322**
5	200	0.8	41.602	4.843	3.951	22.222	15.839	5.156	20.055	5.986	22.294	1.453	22.276	1.205***	22.294	0.84**	2.932	1.49	0.688*
10	200	0.8	83.286	9.97	7.961	45.071	32.168	10.666	40.72	12.049	45.241	2.925	45.19	1.632***	45.241	1.367**	4.429	2.924	1.166*
0.5	200	0.9	6.861	1.194	0.401	3.71	2.575	0.686	2.793	0.681	3.671	0.354	3.673	0.542	3.67	0.16*	0.353	0.272***	0.234**
1	200	0.9	14.134	1.993	0.707	7.616	5.371	1.203	5.819	1.325	7.596	0.615	7.597	0.607	7.597	0.231*	0.634	0.429***	0.271**
5	200	0.9	68.251	6.538	3.071	36.202	25.71	4.705	27.842	6.137	36.266	2.665	36.252	1.192***	36.272	0.785**	2.207	1.365	0.644*
10	200	0.9	140.725	12.843	6.247	76.495	54.611	9.643	58.804	12.788	76.662	5.423	76.628	1.992***	76.671	1.297**	3.412	2.521	1.149*
0.5	200	0.99	74.461	5.788	0.172	39.196	27.954	0.948	26.337	2.069	39.24	4.935	39.238	0.388	39.257	0.126*	0.15**	0.151***	0.19
1	200	0.99	144.674	8.872	0.285	75.511	53.939	1.348	50.165	3.939	75.598	9.06	75.592	0.757	75.632	0.211*	0.253	0.251***	0.241**
5	200	0.99	739.499	26.722	1.173	390.456	277.177	3.639	259.167	20.088	390.762	47.168	390.74	3.788	390.9	0.755**	0.927***	0.943	0.678*
10	200	0.99	1483.401	51.917	2.274	784.535	559.662	7.328	523.261	40.447	785.653	89.251	785.68	7.598	786.074	1.093*	1.575***	1.77	1.17**

Table 3: The EMSE values of the estimators for the model when $p = 8$.

σ^2	n	ρ	OLS	RE	LE	YC I	YC II	SK	LTE	MNTP	NBR	ML	TSS	KLMRT	LKL	RTE I	RTE II	RTE III	RTE IV
0.5	50	0.8	10.963	1.281	0.767	6.713	3.627	1.054	5.994	1.039	6.718	0.275***	6.719	0.449	6.723	0.256**	0.745	0.335	0.155*
1	50	0.8	21.551	2.237	1.497	13.031	6.972	1.988	11.671	1.998	13.092	0.476***	13.094	0.587	13.106	0.459**	1.412	0.62	0.231*
5	50	0.8	109.743	8.648	7.435	66.498	35.953	9.338	59.944	9.915	66.9	2.3	66.914	1.729***	66.974	1.556**	4.921	2.653	0.833*
10	50	0.8	221.218	16.724	14.939	133.789	73.022	18.632	120.621	19.83	134.668	4.661	134.708	3.205***	134.825	2.373**	7.15	5.021	1.638*
0.5	50	0.9	23.039	2.231	0.534	14.013	7.557	0.987	9.93	0.913	14.085	0.344	14.089	0.36	14.103	0.191**	0.514	0.281***	0.107*
1	50	0.9	44.473	3.946	1.024	26.587	14.128	1.832	18.779	1.737	26.74	0.669	26.746	0.611	26.777	0.326**	0.936	0.524***	0.154*
5	50	0.9	222.946	14.449	5.068	133.78	70.271	7.788	93.708	8.649	134.549	3.008	134.584	2.744	134.737	1.047**	3.016	2.139***	0.553*
10	50	0.9	440.11	26.692	9.859	262.906	141.258	14.853	187.16	16.969	264.814	6.374	264.928	5.308	265.268	1.524**	4.074	3.866***	1.067*
0.5	50	0.99	179.643	13.262	0.152	107.686	56.714	1.6	57.125	1.408	108.098	1.637	108.137	1.518	108.235	0.17*	0.128	0.116***	0.084**
1	50	0.99	347.312	20.067	0.265	205.644	105.842	2.066	106.649	2.631	206.557	2.669	206.647	2.94	206.837	0.117*	0.212	0.199***	0.121**
5	50	0.99	1780.847	77.538	1.257	1079.538	568.707	7.628	573.19	13.93	1084.237	14.127	1084.676	15.133	1085.599	0.428*	0.694***	0.751	0.439**
10	50	0.99	3574.818	157.409	2.559	2164.081	1139.681	16.077	1146.738	27.97	2172.414	32.092	2173.152	30.483	2174.92	0.688*	1.154***	1.409	0.855**
0.5	100	0.8	10.603	1.201	0.769	6.449	3.368	1.021	5.482	1.018	6.442	0.247***	6.444	0.453	6.446	0.225**	0.75	0.297	0.153*
1	100	0.8	21.435	2.203	1.519	12.894	6.802	1.992	11.05	1.991	12.952	0.46***	12.958	0.601	12.964	0.405**	1.456	0.563	0.226*
5	100	0.8	106.783	8.475	7.429	64.153	34.314	9.276	55.277	9.84	64.524	2.115	64.543	1.718***	64.587	1.477**	5.59	2.35	0.75*
10	100	0.8	216.062	16.232	14.804	131.28	70.792	18.483	112.891	19.669	131.983	4.288	132.023	3.122***	132.097	2.279**	8.422	4.334	1.346*
0.5	100	0.9	19.838	1.905	0.54	11.856	6.091	0.919	7.999	0.868	11.895	0.247***	11.898	0.355	11.906	0.161**	0.524	0.247***	0.106*
1	100	0.9	40.778	3.66	1.077	24.838	13.024	1.849	16.825	1.769	24.943	0.473	24.953	0.616	24.967	0.29**	1.022	0.472***	0.153*
5	100	0.9	199.467	13.852	5.209	119.483	61.984	8.02	80.836	8.575	120.02	2.188	120.07	2.647	120.132	0.989**	3.57	1.921***	0.528*
10	100	0.9	394.458	23.96	10.242	235.022	121.152	14.934	159.1	16.842	236.083	4.328	236.181	5.143	236.322	1.541**	5.221	3.46***	0.972*
0.5	100	0.99	185.186	12.524	0.148	111.087	56.641	1.328	56.856	1.397	111.356	1.434	111.382	1.616	111.434	0.073*	0.13	0.115***	0.088**
1	100	0.99	366.431	21.167	0.267	217.66	110.366	2.164	110.517	2.718	218.195	2.965	218.259	3.217	218.362	0.127*	0.228	0.203***	0.132**
5	100	0.99	1830.505	81.447	1.191	1088.923	556.968	7.795	558.111	13.573	1092.375	15.336	1092.759	15.972	1093.342	0.449**	0.722***	0.731	0.436*
10	100	0.99	3751.459	153.338	2.384	2246.778	1156.604	14.89	1159.543	28.067	2252.611	31.294	2253.289	32.708	2254.351	0.719*	1.192***	1.316	0.814**
0.5	200	0.8	10.66	1.246	0.761	6.451	3.304	1.037	4.71	1.012	6.427	0.157***	6.427	0.441	6.427	0.146**	0.74	0.222	0.137*
1	200	0.8	21.333	2.212	1.476	12.979	6.666	1.989	9.429	1.961	12.987	0.263***	12.99	0.576	12.991	0.235**	1.423	0.381	0.173*
5	200	0.8	104.209	7.95	7.137	62.972	31.982	8.91	45.832	9.391	63.039	1.159***	63.042	1.785	63.052	0.935**	5.852	1.549	0.578*
10	200	0.8	210.074	14.973	14.4	126.321	64.277	17.708	92.213	18.896	126.483	2.272***	126.495	3.333	126.513	1.627**	9.579	2.925	1.094*
0.5	200	0.9	19.286	1.941	0.543	11.538	5.922	0.951	7.366	0.871	11.559	0.196***	11.562	0.36	11.565	0.135**	0.529	0.223	0.104*
1	200	0.9	38.292	3.33	1.063	22.818	11.587	1.748	14.491	1.704	22.868	0.367***	22.874	0.618	22.88	0.243**	1.025	0.409	0.151*
5	200	0.9	192.914	13.086	5.232	116.524	59.091	7.861	73.792	8.529	116.721	1.808	116.737	2.702	116.761	0.944**	4.024	1.725***	0.537*
10	200	0.9	388.286	25.075	10.535	233.444	119.683	15.599	148.703	17.191	233.989	3.688	234.033	5.326	234.104	1.566**	6.184	3.171***	0.987*
0.5	200	0.99	214.867	15.351	0.134	129.551	66.368	1.637	66.195	1.551	129.674	1.906	129.689	1.913	129.715	0.064*	0.217	0.101***	0.082**
1	200	0.99	422.703	24.4	0.244	254.318	128.303	2.299	127.823	3.003	254.639	3.046	254.689	3.796	254.717	0.122*	0.116	0.188***	0.131**
5	200	0.99	2110.968	96.261	1.094	1258.604	642.555	8.7	637.585	14.92	1260.247	19.688	1260.535	18.974	1260.777	0.465**	0.723	0.683***	0.433*
10	200	0.99	4251.352	171.417	2.109	2543.115	1292.978	14.78	1287.173	29.982	2546.31	35.578	2546.865	38.132	2547.238	0.731*	1.179***	1.198	0.761**

Table 4: The EMSE values of the estimators for the model when $p = 12$.

σ^2	n	ρ	OLS	RE	LE	YC I	YC II	SK	LTE	MNTP	NBR	ML	TSS	KLMRT	LKL	RTE I	RTE II	RTE III	RTE IV
0.5	50	0.8	16.062	1.395	1.141	10.29	4.98	1.404	7.764	1.392	10.353	0.349**	10.359	0.395	10.364	0.374***	1.12	0.419	0.151*
1	50	0.8	32.244	2.652	2.299	20.661	10.092	2.815	15.728	2.776	20.801	0.706***	20.813	0.604**	20.822	0.726	2.217	0.833	0.268*
5	50	0.8	161.578	10.882	11.419	103.893	50.303	13.458	78.631	13.756	104.593	3.381	104.648	2.228**	104.703	2.636***	8.36	3.948	1.155*
10	50	0.8	326.405	21.24	23.154	209.722	101.952	27.231	158.963	28.099	211.033	6.844	211.162	4.288***	211.24	4.026**	12.52	7.808	2.273*
0.5	50	0.9	46.094	3.415	0.671	29.253	14.036	1.202	17.017	1.062	29.495	0.406	29.513	0.51	29.545	0.296**	0.655	0.371***	0.071*
1	50	0.9	91.234	6.28	1.334	57.947	27.806	2.321	33.651	2.113	58.388	0.8	58.418	0.994	58.487	0.528**	1.252	0.731***	0.12*
5	50	0.9	449.125	23.934	6.508	282.055	133.922	10.008	163.335	10.289	284.821	3.602	285.028	4.819	285.392	1.557**	4.145	3.23***	0.49*
10	50	0.9	914.791	47.537	13.168	579.627	277.278	20.252	336.158	20.96	584.434	7.618	584.874	9.782	585.492	2.174**	5.7***	6.221	0.97*
0.5	50	0.99	393.849	22.08	0.145	248.962	118.37	1.703	118.852	1.37	251.029	0.991	251.199	3.376	251.652	0.061**	0.129	0.113***	0.052*
1	50	0.99	785.605	41.203	0.277	497.184	233.947	3.125	234.915	2.713	501.477	1.915	501.841	6.734	502.702	0.099**	0.223	0.213***	0.083*
5	50	0.99	4028.641	168.722	1.374	2569.958	1228.338	11.144	1234.773	14.095	2590.424	9.587	2591.946	34.304	2596.141	0.357**	0.675***	0.869	0.34*
10	50	0.99	7958.812	316.947	2.66	5062.842	2399.464	21.361	2410.877	27.608	5101.034	21.861	5103.976	68.094	5112.183	0.542*	0.993***	1.485	0.618**
0.5	100	0.8	18.144	1.547	1.053	11.42	5.267	1.351	7.965	1.284	11.481	0.262**	11.49	0.381	11.494	0.33***	1.039	0.383	0.116*
1	100	0.8	36.618	2.853	2.093	23.255	10.765	2.643	16.115	2.558	23.377	0.486**	23.389	0.626	23.398	0.62***	2.042	0.731	0.184*
5	100	0.8	185.297	12.147	10.432	118.112	55.017	12.819	82.06	12.789	118.806	2.374***	118.877	2.643	118.916	2.327**	8.265	3.457	0.762*
10	100	0.8	362.361	21.715	20.624	230.075	105.134	24.746	158.979	25.043	231.241	4.573***	231.388	5.106	231.472	3.641**	12.85	6.683	1.501*
0.5	100	0.9	39.29	3.1	0.701	25.17	11.743	1.229	13.902	1.073	25.29	0.247***	25.304	0.516	25.31	0.208**	0.691	0.278	0.068*
1	100	0.9	76.287	5.474	1.369	48.491	22.418	2.285	26.615	2.084	48.74	0.439***	48.772	0.983	48.788	0.38**	1.322	0.527	0.11*
5	100	0.9	385.046	22.913	6.852	245.131	113.562	10.703	134.636	10.458	246.162	2.336***	246.286	4.836	246.355	1.241**	4.872	2.419	0.425*
10	100	0.9	769.431	41.92	13.707	489.421	225.579	20.456	268.219	20.896	491.508	4.471***	491.789	9.643	491.918	1.919**	7.268	4.591	0.818*
0.5	100	0.99	259.13	16.191	0.175	165.169	75.718	1.446	76.167	1.105	165.716	0.526	165.806	2.779	165.848	0.065**	0.163	0.116***	0.057*
1	100	0.99	512.278	28.695	0.331	324.623	146.629	2.456	147.451	2.151	325.761	0.935	325.945	5.512	326.036	0.106**	0.291	0.216***	0.087*
5	100	0.99	2552.714	118.048	1.607	1625.383	741.581	10.129	745.285	10.783	1630.484	5.271	1631.261	27.471	1631.832	0.388**	0.898	0.861***	0.344*
10	100	0.99	5095.394	224.208	3.216	3219.162	1464.12	18.544	1470.211	21.468	3228.083	10.415	3229.43	55.372	3230.382	0.619*	1.324***	1.534	0.64**
0.5	200	0.8	16.075	1.49	1.133	10.235	4.72	1.432	6.839	1.358	10.249	0.197**	10.251	0.398	10.252	0.249***	1.116	0.303	0.115*
1	200	0.8	32.374	2.655	2.25	20.598	9.554	2.777	13.78	2.697	20.667	0.368**	20.674	0.625	20.677	0.466***	2.203	0.568	0.18*
5	200	0.8	160.754	10.983	11.162	102.411	47.202	13.345	68.11	13.364	102.741	1.726**	102.778	2.459	102.79	1.953***	9.646	2.701	0.735*
10	200	0.8	325.34	20.966	22.659	208.12	95.943	26.76	138.599	27.102	208.834	3.547***	208.919	4.824	208.942	3.361**	16.278	5.31	1.441*
0.5	200	0.9	33.743	2.914	0.758	21.586	9.878	1.305	11.145	1.101	21.628	0.147***	21.635	0.499	21.636	0.131**	0.75	0.2	0.066*
1	200	0.9	65.092	4.867	1.458	41.336	18.445	2.343	20.878	2.107	41.427	0.283***	41.441	0.93	41.444	0.237**	1.427	0.372	0.1*
5	200	0.9	332.328	20.235	7.389	213.16	96.096	10.901	108.732	10.766	213.505	1.327***	213.558	4.607	213.566	0.923**	5.97	1.634	0.393*
10	200	0.9	657.186	37.248	14.656	416.99	187.992	21.132	212.632	21.274	417.699	2.642***	417.797	9.129	417.828	1.527**	9.192	3.11	0.784*
0.5	200	0.99	292.039	18.556	0.155	185.59	82.558	1.583	82.718	1.126	185.791	0.571	185.835	3.303	185.837	0.049*	0.141	0.095***	0.054**
1	200	0.99	590.23	33.715	0.287	375.064	166.974	2.759	167.187	2.266	375.468	1.2	375.545	6.68	375.565	0.083**	0.259	0.171***	0.080*
5	200	0.99	2921.739	136.846	1.377	1859.166	827.946	10.449	829.533	11.213	1861.452	6.048	1861.882	33.06	1861.937	0.349**	0.838	0.652***	0.323*
10	200	0.99	5901.947	255.942	2.752	3751.628	1685.386	19.081	1689.929	22.74	3756.582	12.904	3757.596	66.628	3757.709	0.614*	1.303	1.154***	0.622**

3. When the correlations ρ , n , and p are kept constant, the increase in the variance leads to an increase in the EMSE values of all estimators. However, in terms of EMSE values, the increases in all proposed estimators are smaller compared to the increases in other available estimators.
4. It is seen that when the number of variables p , ρ , and σ^2 are kept constant, the EMSE values of the proposed estimators are lower than the EMSE values of the existing estimators in $n = 50, 100, 200$. However, it is observed that there is no significant systematic change in the EMSE values of all estimators as the number of observations increases. As a result, it can be said that compared to ρ and σ^2 , the number of observations has a relatively small effect on EMSE values.

In all scenarios examined, it is observed that all our proposed estimators are significantly superior to existing estimators: OLS, RE, LE, LTE, SK, YC I, YC II, MNTP, NBR, ML, TSS, KLMRT, and LKL. However, even if the estimators RTE I and RTE IV are better than other estimators accessible in all cases, they behave differently in each scenario. In general, RTE I has the best EMSE value in models with few variables and low variance. In contrast, RTE IV has a smaller EMSE value in models with large variance. When the number of variables increased, RTE IV generally gave better results in all scenarios.

In the second simulation scheme, we only investigated the performances of RTE and NLTE for each n , p , ρ , and σ^2 . The purpose of this simulation is to examine the performances of NLTE and RTE at various values of the biasing parameter k depending on EMSE values given in (5.2). There are many $f(k)$ and $g(k)$ functions that can be considered to evaluate the performances of these two classes of estimators. The biasing parameter k is not estimated in the second simulation scheme. Only the EMSE values obtained by increasing k values in the range $[0, 1]$ by 0.05 are compared. In order to compare the performances of these two estimators under some situations as an example, the following estimators with $f(k)$ and $g(k)$ functions are taken:

$$\hat{\beta}_{NLTE} = (X'X + kI)^{-1}(X'X + f(k)I)\hat{\beta}_{OLS} \text{ where } f(k) = \frac{\lambda_{\min}\alpha_{\min}^2}{1+\lambda_{\max}\alpha_{\max}^2}k + \left(\frac{\lambda_{\min}\alpha_{\min}^2}{1+\lambda_{\max}\alpha_{\max}^2} - 1\right)\lambda_{\min},$$

$$\hat{\beta}_{NLTE(RE)} = (X'X + kI)^{-1}(X'X + (k + f(k))I)\hat{\beta}_{RE} \text{ where } f(k) = \frac{\alpha_{\min}^2(k + \lambda_{\min})^2}{1 + \lambda_{\max}\alpha_{\max}^2} - (k + \lambda_{\min}),$$

$$\hat{\beta}_{RTE} = (X'X + I)^{-1}(X'X + g(k)I)\hat{\beta}_{RE} \text{ where } g(k) = \frac{(1 + \lambda_{\min})\alpha_{\min}^2}{\sigma^2 + \lambda_{\max}\alpha_{\max}^2}k + \left(\frac{(1 + \lambda_{\min})\alpha_{\min}^2}{\sigma^2 + \lambda_{\max}\alpha_{\max}^2} - 1\right)\lambda_{\min}.$$

Note that, when we use $\hat{\beta}_{RE}$ instead of $\hat{\beta}^*$ in $\hat{\beta}_{NLTE}$, the obtained estimator is shown as $\hat{\beta}_{NLTE(RE)}$. Also, $f(k)$ functions used in $\hat{\beta}_{NLTE}$ and $\hat{\beta}_{NLTE(RE)}$ were determined in accordance with the rules given by Kurnaz and Akay (2015). We only consider the cases $\rho = 0.9, 0.99$, $n = 50, 200$, and $p = 4, 8, 12$, and $\sigma^2 = 1, 10$. Depending on these n , ρ , p , and σ^2 values, the explanatory variables are generated according to equation (5.1). The simulation is repeated 2000 times for each k value. The results are collectively presented graphically in Figures 1 and 2.

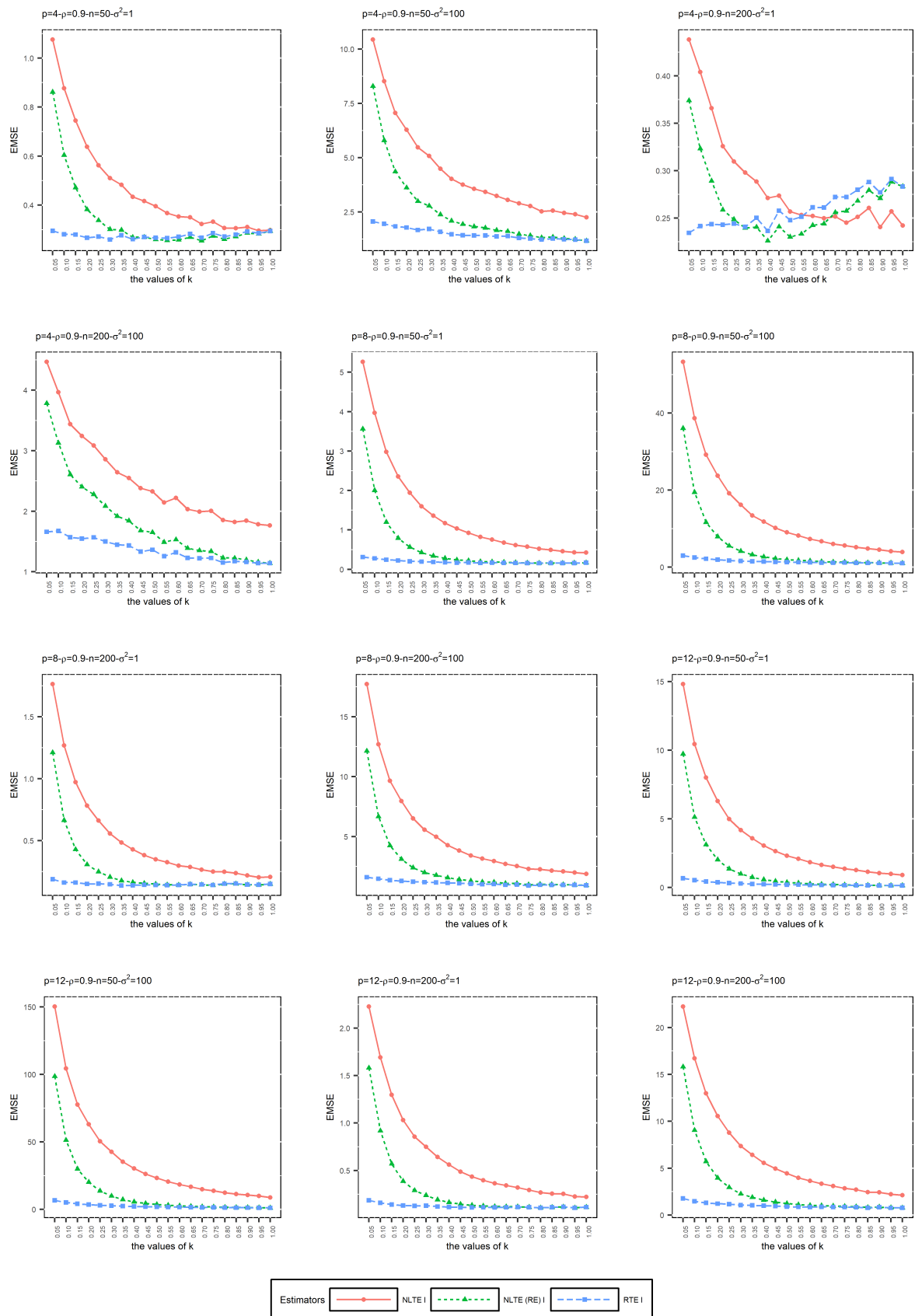


Figure 1: The EMSE values of NLTE I, NLTE(RE) I, RTE I as a function of k where $\rho = 0.9$.

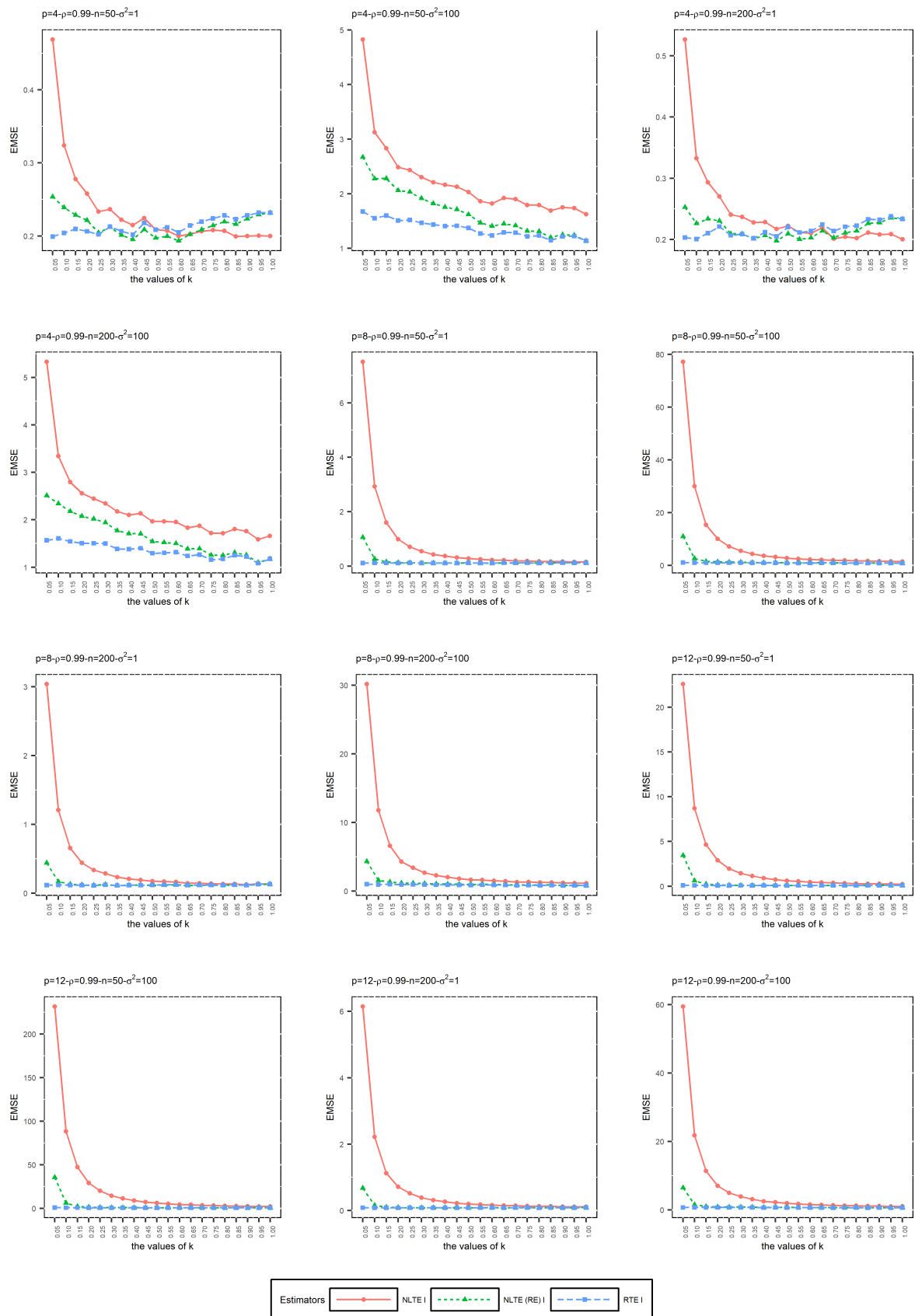


Figure 2: The EMSE values of NLTE I, NLTE(RE) I, RTE I as a function of k where $\rho = 0.99$.

Based on Figures 1–2, we can interpret the results as follows depending on each (n, ρ, p, σ^2) :

- 1) At small values of the biasing parameter k , $\hat{\beta}_{RTE}$ outperforms $\hat{\beta}_{NLTE}$ and $\hat{\beta}_{NLTE(RE)}$. Although both $\hat{\beta}_{RTE}$ and $\hat{\beta}_{NLTE(RE)}$ include the $\hat{\beta}_{RE}$, the performance of $\hat{\beta}_{NLTE(RE)}$ is quite poor compared to $\hat{\beta}_{RTE}$ at small values of the biasing parameter.
- 2) For $p = 4$ and $\rho = 0.9$, $\hat{\beta}_{NLTE(RE)}$ exhibits quite different behavior from $\hat{\beta}_{NLTE}$ and $\hat{\beta}_{RTE}$. If the value of the biasing parameter and the number of explanatory variables increases, $\hat{\beta}_{NLTE}$, $\hat{\beta}_{NLTE(RE)}$, and $\hat{\beta}_{RTE}$ show almost the same behaviors. In general, $\hat{\beta}_{RTE}$ exhibits a more consistent behavior at different values of the biasing parameter k .

Based on the results of the second simulation design, we can recommend $\hat{\beta}_{RTE}$ to the researchers because it is a more consistent estimator than $\hat{\beta}_{NLTE}$ and $\hat{\beta}_{NLTE(RE)}$ for the considered conditions. In general, the performances of these estimators depend on $f(k)$ and $g(k)$ functions. In practice, we need to replace these functions with functional relationships that can occur between the biasing parameters. Therefore, it should be kept in mind that the results of graphical findings may change.

6. NUMERICAL EXAMPLE

In this section, we reconsider the Portland cement data that was analyzed by Hald (1952), Liu (1993), Sakalhoğlu and Kaçiranlar (2008), Yang and Chang (2010), and Kurnaz and Akay (2018). In this data set, the following four compounds are independent variables: tricalcium aluminate (x_1), tetracalcium silicate (x_2), tetracalcium alumino ferrite (x_3), and dicalcium silicate (x_4). The dependent variable y is the heat evolved in calories per gram of cement. We fit a linear regression model with intercept to the data by adding a column of ones to the matrix X . Then, the eigenvalues of $X'X$ are $\lambda_1 = 44676.2059$, $\lambda_2 = 5965.4221$, $\lambda_3 = 809.9521$, $\lambda_4 = 105.4187$ and $\lambda_5 = 0.0012$. The condition number is approximately 3.66×10^7 , therefore the matrix X is quite ill-conditioned. The numerical results are summarized in Table 5 to compare RTEs with other estimators. In addition, three different $f(k)$ functions for both $\hat{\beta}_{NLTE}$ and $\hat{\beta}_{NLTE(RE)}$ are given in Table 5. Since there are many combinations to determine k and $f(k)$ functions in NLTE and NLTE(RE), we use the following k estimators and $f(k)$ functions based on the Kibria (2003) and Kurnaz and Akay (2015). Note that the function $f(k)$ that minimizes $SMSE(\hat{\beta}_{NLTE(RE)})$ is a quadratic function.

$$\text{NLTE I: } \hat{k}_{\text{NLTE I}} = \frac{\sigma^2}{\sum_{j=1}^p \alpha_j^2} \text{ and } f(k) = \frac{\lambda_{\min} \alpha_{\min}^2}{\hat{\sigma}^2 + \lambda_{\max} \alpha_{\max}^2} k + \left(\frac{\lambda_{\min} \alpha_{\min}^2}{\hat{\sigma}^2 + \lambda_{\max} \alpha_{\max}^2} - 1 \right) \lambda_{\min},$$

$$\text{NLTE II: } \hat{k}_{\text{NLTE II}} = \frac{p\sigma^2}{n \sum_{j=1}^p \alpha_j^2} \text{ and } f(k) = \frac{\lambda_{\min} \alpha_{\min}^2}{\hat{\sigma}^2 + \lambda_{\max} \alpha_{\max}^2} k - \frac{\hat{\sigma}^2 \lambda_{\min}}{\hat{\sigma}^2 + \lambda_{\max} \alpha_{\max}^2},$$

$$\text{NLTE III: } \hat{k}_{\text{NLTE III}} = \frac{\hat{\sigma}^2}{(\prod_{j=1}^p \hat{\alpha}_j^2)^{\frac{1}{p}}} \text{ and } f(k) = \frac{\lambda_{\min} \alpha_{\min}^2}{\hat{\sigma}^2 + \lambda_{\max} \alpha_{\max}^2} k - \min \left(\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \lambda_j \alpha_j^2} \right) \lambda_{\min},$$

$$\text{NLTE(RE) I: } \hat{k}_{\text{NLTE(RE) I}} = \frac{\sigma^2}{\sum_{j=1}^p \alpha_j^2} \text{ and } f(k) = \frac{\alpha_{\min}^2}{\hat{\sigma}^2 + \lambda_{\max} \alpha_{\max}^2} (k + \lambda_{\min})^2 - (k + \lambda_{\min}),$$

$$\text{NLTE(RE) II: } \hat{k}_{\text{NLTE(RE) II}} = \frac{p\sigma^2}{n \sum_{j=1}^p \alpha_j^2} \text{ and } f(k) = \frac{\alpha_{\max}^2}{\hat{\sigma}^2 + \lambda_{\max} \alpha_{\max}^2} (k + \lambda_{\min})^2 - (k + \lambda_{\min}),$$

$$\text{NLTE(RE) III: } \hat{k}_{\text{NLTE(RE) III}} = \frac{\hat{\sigma}^2}{(\prod_{j=1}^p \hat{\alpha}_j^2)^{\frac{1}{p}}} \text{ and } f(k) = \frac{\alpha_{\min}^2}{\max(\hat{\sigma}^2 + \lambda_i \alpha_i^2)} (k + \lambda_{\min})^2 - (k + \lambda_{\min}).$$

Table 5: The estimated parameter values and the estimated variance values of the estimators.

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$var(\hat{\beta})$	$SMSE(\hat{\beta})$
$\hat{\beta}_{OLS}$	62.4054	1.5511	0.5102	0.1019	-0.1441	4912.0902	
$\hat{\beta}_{RE} (\hat{k}_{RE} = 1.4250)$	0.1003	2.1725	1.1568	0.7435	0.4882	0.067330	0.067346
$\hat{\beta}_{LE} (\hat{d}_{LE} = 0)$	0.1230	2.1781	1.1552	0.7473	0.4871	0.071467	0.071479
$\hat{\beta}_{LTE} (\hat{k}_{LTE} = 451.2736, \hat{d}_{LTE} = -199.5073)$	27.6065	1.1641	1.0097	0.0891	0.2955	960.089121	960.375122
$\hat{\beta}_{SK} (\hat{k}_{SK} = 1.4250, \hat{d}_{SK} = 493.7504)$	26.4790	8.5996	-0.6618	5.2740	-0.7883	878.099704	879.220053
$\hat{\beta}_{YCI} (\hat{K}_1 = 0.0015, \hat{D}_1 = 0.9992)$	27.6068	1.9090	0.8688	0.4680	0.2075	959.502978	959.502981
$\hat{\beta}_{YCI} (\hat{K}_2 = 0.0008, \hat{D}_2 = 0.7206)$	27.6067	1.9052	0.8697	0.4653	0.2080	959.502677	959.502680
$\hat{\beta}_{MNTP} (\hat{k}_{MNTP} = 1.3761 \times 10^{-6}, \hat{d}_{MNTP} = 0.0883)$	5.6197	2.1227	1.0983	0.6904	0.4314	39.291526	39.291535
$\hat{\beta}_{NBR} (\hat{k}_{NBR} = 0.3388 \times 10^{-3}, \hat{d}_{NBR} = 0.0015)$	27.6068	1.9091	0.8688	0.468	0.2075	959.502980	959.502982
$\hat{\beta}_{ML} (\hat{d}_{ML} = 0.4426)$	-27.4454	2.4556	1.4408	1.033	0.7665	954.838327	954.838351
$\hat{\beta}_{TSS} (\hat{k}_{TSS} = 0.5509 \times 10^{-3}, \hat{d}_{TSS} = 0.7920)$	27.6068	1.9091	0.8688	0.468	0.2075	959.502980	959.502982
$\hat{\beta}_{KLMRT} (\hat{k}_{KLMRT} = 348.2785, \hat{d}_{KLMRT} = 0.4420)$	0.0244	0.2284	1.4622	0.0276	0.6772	0.001739	6.4125088
$\hat{\beta}_{LKL} (\hat{k}_{LKL} = 0.4714 \times 10^{-3}, \hat{d}_{LKL} = 1)$	27.6068	1.9091	0.8688	0.468	0.2075	959.502980	959.502982
$\hat{\beta}_{NLTE I} (f(k) = 3.0866 \times 10^{-13}k - 0.0012)$ $\hat{k}_{NLTE I} = 0.0015$	0.0473	2.1925	1.1528	0.7580	0.4858	0.065275	0.065282
$\hat{\beta}_{NLTE II} (f(k) = 3.0866 \times 10^{-13}k - 4.1930 \times 10^{-11})$ $\hat{k}_{NLTE II} = 0.0006$	42.0456	1.7605	0.7200	0.3161	0.0616	2228.180975	2228.180976
$\hat{\beta}_{NLTE III} (f(k) = 3.0866 \times 10^{-13}k - 6.0859 \times 10^{-8})$ $\hat{k}_{NLTE III} = 1.4250$	0.1003	2.1725	1.1568	0.7435	0.4882	0.067330	0.067346
$\hat{\beta}_{NLTE(RE) I} (f(k) = 2.5341 \times 10^{-10}(k + 0.0012)^2 - (k + 0.0012))$ $\hat{k}_{NLTE(RE) I} = 0.0015$	0.0473	2.1925	1.1529	0.758	0.4858	0.065273	0.065280
$\hat{\beta}_{NLTE(RE) II} (f(k) = 2.2383 \times 10^{-5}(k + 0.0012)^2 - (k + 0.0012))$ $\hat{k}_{NLTE(RE) II} = 0.0006$	0.0473	2.1925	1.1528	0.758	0.4858	0.065275	0.065282
$\hat{\beta}_{NLTE(RE) III} (f(k) = 3.6781 \times 10^{-6}(k + 0.0012)^2 - (k + 0.0012))$ $\hat{k}_{NLTE(RE) III} = 1.4250$	0.0468	2.1538	1.1618	0.7302	0.4917	0.062256	0.062300
$\hat{\beta}_{RTE I} (g(k) = 2.5372 \times 10^{-10}k - 0.0012)$ $\hat{k}_{RTE I} = 0.1013$	0.0471	2.1774	1.1563	0.7471	0.4881	0.064088	0.064099
$\hat{\beta}_{RTE II} (g(k) = 2.5372 \times 10^{-10}k - 4.1621 \times 10^{-11})$ $\hat{k}_{RTE II} = 10.9035$	0.0452	2.0425	1.1873	0.6513	0.5083	0.054048	0.054708
$\hat{\beta}_{RTE III} (g(k) = 2.5372 \times 10^{-10}k - 0.2692 \times 10^{-4})$ $\hat{k}_{RTE III} = 0.9106 \times 10^{-4}$	0.1161	2.1781	1.1553	0.7474	0.4872	0.070216	0.070227
$\hat{\beta}_{RTE IV} (g(k) = 0.1952 \times 10^{-10}k - 0.0122)$ $\hat{k}_{RTE IV} = 8935.2414$	0.0231	0.2431	1.1857	0.2203	0.6003	0.000312	255.273061

In addition, the bootstrap sampling method was used to obtain the actual parameter values to be used instead of the α parameter. Therefore, 10000 bootstrap samples were created and the parameter estimates associated with the estimators were calculated for each of the samples. The mean of the OLS estimates is considered as an estimate of α . The calculated SMSE values are given in Table 5. As seen in Table 5, the estimated variance values and the SMSE values of RTE I, RTE II, RTE III, and RTE IV under the proposed $g(k)$ functions with k estimates can yield appropriate results compared to other existing estimators. To compare the estimators under the MMSE sense, $\hat{\alpha}_{OLS}$ is used in place of the unknown parameter α . Here, the eigenvalues of the matrices obtained with the MMSE differences are taken into account. That is, if any of the eigenvalues is less than or equal to tolerance, then the MMSE difference is not pd. Otherwise, the MMSE difference is pd. The R Programming is used with tolerance 10^{-10} to find whether MMSE differences are pd or not. To illustrate Theorem 3.1, the function $f(k)$ is taken as $f(k) = 3.0866 \times 10^{-13}k - 0.0012$ by using NLTE I. Also, the $g(k)$ is obtained as $g(k) = 2.5372 \times 10^{-10}k - 0.0012$ in RTE I using (4.4).

In this case, $cov(\hat{\beta}_{NLTE I}) - cov(\hat{\beta}_{RTE I})$ is the pd matrix for $k > 0$. But, the criterion (3.1) given in Theorem 3.1 is not held. On the other hand, if functions $g(k)$ and $f(k)$ are arbitrarily taken as $f(k) = 0.5k - 0.05$ and $g(k) = 0.6k - 0.05$, $cov(\hat{\beta}_{NLTE}) - cov(\hat{\beta}_{RTE})$ is pd matrix for $0 < k \leq 0.09754$ or $k \geq 0.09758$. Also, k values which provide (3.1) criterion are $0 < k < 0.0479$. Consequently, $MMSE(\hat{\beta}_{NLTE}) - MMSE(\hat{\beta}_{RTE})$ is the pd matrix where $0 < k < 0.0479$.

7. CONCLUSION

In this study, a new general biased estimator called RTE is proposed as an alternative to other existing biased estimators used in the presence of multicollinearity. The RTE is a general estimator that includes other biased estimators, such as the OLS, RE, LE, ML, YC, and SK estimators as special cases. The RTE is based on a functional relationship $g(k)$ between the biasing parameters, which would provide an alternative method for overcoming multicollinearity. In this study, we investigated several rules for determining the optimal function $g(k)$. The performance of these functions is analyzed using different k estimators. Results revealed that the estimators obtained with these $g(k)$ functions outperformed the other existing estimators under the examined conditions. In particular, RTE I has the best EMSE value in models with few variables and low variance. On the other hand, RTE IV has a small EMSE value in high-variance models. When the number of variables increased, RTE IV generally gave better results in all scenarios. Besides, a general simulation study is performed to compare RTE and NLTE. In the cases we have considered, it has been observed that RTE performs well when the biasing parameter k is small values. Although RTE and NLTE(RE) are both dependent on RE, the main advantage of RTE over NLTE(RE) is that it can minimize the SMSE function with the help of a simpler function. Additionally, Portland data is also considered to illustrate the advantage of RTEs in the linear regression models. Since NLTE and RTE are two general classes of biased estimators, a comparison of these classes is given in Portland data from various perspectives. Finally, based on the results of the simulations and application, it can be recommended that the RTE can be used when there is multicollinearity in the linear regression models.

ACKNOWLEDGMENTS



We wish to thank the referee and the editor for their constructive comments to improve the quality of this paper. Dr. Kadri Ulas AKAY was supported by the Scientific Research Projects Coordination Unit of Istanbul University. Project number UDP-45745.


REFERENCES

- Ahmad, S. and Aslam, M. (2022). Another proposal about the new two-parameter estimator for linear regression model with correlated regressors. *Communications in Statistics – Simulation and Computation*, 51(6):3054–3072.
- Akay, K. U. and Ertan, E. (2022). A new Liu-type estimator in Poisson regression models. *Hacettepe Journal of Mathematics and Statistics*, 51(5):1484–1503.
- Aladeitan, B.B. Adebimpe, O., Lukman, A.F., Oludoun, O., and Abiodun, O.E. (2021). Modified Kibria-Lukman (MKL) estimator for the Poisson Regression Model: application and simulation. *F1000Research*, 10:548. doi: 10.12688/f1000research.53987.2.
- Aslam, M. and Ahmad, S. (2022). The modified Liu-ridge-type estimator: a new class of biased estimators to address multicollinearity. *Communications in Statistics – Simulation and Computation*, 51(11):6591–6609.
- Babar, I. and Chand, S. (2022). Weighted ridge and Liu estimators for linear regression model. *Concurrency and Computation: Practice and Experience*, 34(27):e7343.
- Dawoud, I., Lukman, A.F., and Haadi, A.R. (2022). A new biased regression estimator: Theory, simulation and application. *Scientific African*, 15:e01100.
- Dorugade, A.V. (2014). A modified two-parameter estimator in linear regression. *Statistics in Transition New Series*, 15(1):23–36. doi:10.21307/stattrans-2019-021.
- Erkoç, A., Ertan, E., Algamal, Z. Y. and Akay, K. U. (2023). The beta Liu-type estimator: simulation and application. *Hacettepe Journal of Mathematics and Statistics*, 52(3):828–840.
- Ertan, E. and Akay, K.U. (2022). A new Liu-type estimator in binary logistic regression models. *Communications in Statistics – Theory and Methods*, 51(13):4370–4394.
- Farebrother, R. W. (1976). Further results on the mean square error of ridge regression. *The Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 28:248–250.
- Hald, A. (1952). *Statistical Theory With Engineering Applications*. Wiley, New York.
- Hoerl, A.E. and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.
- Huang, J. and Yang, H. (2014). A two-parameter estimator in the negative binomial regression model. *Journal of Statistical Computation and Simulation*, 84(1):124–134.
- Idowu, J.I., Oladapo, O. J., Owolabi, A. T., Ayinde, K., and Akinmoju, O. (2023). Combating multicollinearity: a new two-parameter approach. *Nicel Bilimler Dergisi*, 5(1):90–100.
- Kibria, B.M.G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics – Simulation and Computation*, 32(2):419–435.
- Kibria, B. M. G. and Lukman, A. F. (2020). A new ridge-type estimator for the linear regression model: Simulations and applications. *Scientifica*, 2020:9758378. doi:10.1155/2020/9758378.
- Kurnaz, F. S. and Akay, K. U. (2015). A new Liu-type estimator. *Statistical Papers*, 56:495–517.
- Kurnaz, F. S. and Akay, K. U. (2018). Matrix mean squared error comparisons of some biased estimators with two biasing parameters. *Communications in Statistics – Theory and Methods*, 47(8):2022–2035.
- Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics – Theory and Methods*, 22(2):393–402.
- Liu, K. (2003). Using Liu-type estimator to combat collinearity. *Communications in Statistics – Theory and Methods*, 32(5):1009–1020.

- Lukman, A. F., Ayinde, K., Siok Kun, S., and Adewuyi, E. T. (2019). A modified new two-parameter estimator in a linear regression model. *Modelling and Simulation in Engineering*, vol. 2019, Article ID 6342702.
- Lukman, A. F., Kibria, B. M. G., Ayinde, K., and Jegede, S. L. (2020). Modified one-parameter Liu estimator for the linear regression model. *Modelling and Simulation in Engineering*, vol. 2020, Article ID 9574304.
- McDonald, G. C. and Galarneau, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, 70(350):407–416.
- Özkale, M. R. and Kaçiranlar, S. (2007). The restricted and unrestricted two-parameter estimators. *Communications in Statistics – Theory and Methods*, 36(10):2707–2725.
- Qasim, M., Amin, M., and Omer, T. (2020). Performance of some new Liu parameters for the linear regression model. *Communications in Statistics – Theory and Methods*, 49(17):4178–4196.
- Qasim, M., Månsson, K., Sjölander, P., and Kibria, B. M. G. (2022). A new class of efficient and debiased two-step shrinkage estimators: method and application. *Journal of Applied Statistics*, 49(16):4181–4205.
- Sakalhoğlu, S. and Kaçiranlar, S. (2008). A new biased estimator based on ridge estimation. *Statistical Papers*, 49:4178–4196.
- Shewa, G. A. and Ugwuowo, F. I. (2023). A new hybrid estimator for linear regression model analysis: Computations and simulations. *Scientific African*, 19:e01441.
- Shukur, G., Månsson, K., and Sjölander, P. (2008). Developing Interaction Shrinkage Parameters for the Liu Estimator with an Application to the Electricity Retail Market. *Computational Economics*, 46(4):539–550.
- Stein, C. (1956). *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*. In “Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability”, vol. 1, 197–206.
- Theobald, C. M. (1974). Generalizations of mean square error applied to ridge regression. *The Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 36:103–106.
- Trenkler, G. and Toutenburg, H. (1990). Mean squared error matrix comparisons between biased estimator — an overview of recent results. *Statistical Papers*, 31:165–179.
- Üstündağ, Ş. G., Toker, S., and Özbay, N. (2021). Defining a two-parameter estimator: a mathematical programming evidence. *Journal of Statistical Computation and Simulation*, 91(11):2133–2152.
- Yang, H. and Chang, X. (2010). A new two-parameter estimator in linear regression. *Communications in Statistics – Theory and Methods*, 39(6):923–934.
- Zeinal, A. and Azmoun Zavie Kivi, M. R. (2023). The generalized new two-type parameter estimator in linear regression model. *Communications in Statistics – Simulation and Computation*, 52(1):98–109.

Comparison and Equality of Bajraktarević-type ψ -estimators

Authors: MÁTYÁS BARCZY  
– HUN-REN-SZTE Analysis and Applications Research Group,
Bolyai Institute, University of Szeged,
Aradi vértanúk tere 1, H-6720 Szeged, Hungary
barczy@math.u-szeged.hu

ZSOLT PÁLES 
– Institute of Mathematics,
University of Debrecen,
Pf. 400, H-4002 Debrecen, Hungary
pales@science.unideb.hu

Received: November 2023

Revised: January 2024

Accepted: January 2024

Abstract:

- We solve the comparison problem for Bajraktarević-type ψ -estimators introduced by Barczy and Páles in 2022. Namely, we derive several necessary and sufficient conditions under which a Bajraktarević-type ψ -estimator is less than or equal to another Bajraktarević-type ψ -estimator for any sample. We also solve the corresponding equality problem. As an important particular case, we obtain the solutions of the two problems in question for quasi-arithmetic-type ψ -estimators.

Keywords:

- *ψ -estimator; Z-estimator; comparison of estimators; equality of estimators; Bajraktarević-type estimator; quasi-arithmetic-type estimator.*

AMS Subject Classification:

- 62F10, 62D99, 26E60.

1. INTRODUCTION

The ψ -estimators (also called Z -estimators) have been playing an important role in statistics since the 1960's. In what follows, let (X, \mathcal{X}) be a measurable space, Θ be a Borel subset of \mathbb{R} , and $\psi : X \times \Theta \rightarrow \mathbb{R}$ be a function which is measurable in its second variable with respect to the sigma-algebra \mathcal{X} . Let $(\xi_n)_{n \geq 1}$ be a sequence of independent identically distributed (i.i.d.) X -valued random variables such that the distribution of ξ_1 depends on an unknown parameter $\vartheta \in \Theta$. For each $n \geq 1$, Huber (1964, 1967) among others introduced an important estimator of ϑ based on the observations ξ_1, \dots, ξ_n as a solution $\widehat{\vartheta}_{n,\psi}(\xi_1, \dots, \xi_n)$ of the equation (with respect to the unknown parameter):

$$\sum_{i=1}^n \psi(\xi_i, t) = 0, \quad t \in \Theta,$$

provided that such a solution exists. In the statistical literature, one calls the random variable $\widehat{\vartheta}_{n,\psi}(\xi_1, \dots, \xi_n)$ a ψ -estimator of the unknown parameter $\vartheta \in \Theta$ based on the i.i.d. observations ξ_1, \dots, ξ_n , while other authors call it a Z -estimator (the letter Z refers to “zero”). In fact, ψ -estimators are special M -estimators (where the letter M refers to “maximum likelihood-type”) that were also introduced by Huber (1964, 1967). For a detailed exposition of M -estimators and ψ -estimators, see, e.g., Kosorok (2008, Sections 2.2.5 and 13) or van der Vaart (1998, Section 5).

Results on the comparison and the equality of ψ -estimators are quite rare in the literature. These two problems can be formulated as follows: given $\psi, \varphi : X \times \Theta \rightarrow \mathbb{R}$ (with the properties described above), we are interested in finding necessary as well as sufficient conditions for the inequality $\widehat{\vartheta}_{n,\psi} \leq \widehat{\vartheta}_{n,\varphi}$ and for the equality $\widehat{\vartheta}_{n,\psi} = \widehat{\vartheta}_{n,\varphi}$ to be valid for all $n \geq 1$, respectively. In Barczy and Páles (2023, Section 3), we derived such conditions in case of generalized ψ -estimators, introduced in Barczy and Páles (2025) (see also Definition 1.2), which are generalizations of ψ -estimators recalled above. In general linear models, many authors investigated the equality of ordinary least squares estimator (OLSE) and best linear unbiased estimator (BLUE) of the regression parameters. For a detailed review of the literature, see Section 2 in our arXiv version Barczy and Páles (2023).

In this paper, we study the comparison and equality problems for a subclass of generalized ψ -estimators, namely for Bajraktarević-type ψ -estimators introduced in Barczy and Páles (2025) (see also (2.2)). The statistical applications of Bajraktarević-type ψ -estimators has not been explored yet, it can be a topic of a future research. Here we point out an important field in practice, where these types of estimators may be indeed useful. Mukhopadhyay *et al.* (2021) found that in the presence of outliers in the data, more precisely, when the data are generated by a mixture population involving a major (dominating) component and a minor (outlying) component, the power mean (also called generalized mean) estimates the mean of the dominating population more accurately compared to the usual maximum likelihood estimator. Thus the class of power means offers an alternative way for estimating the target mean parameter without invoking the complications of sophisticated robust techniques. Power means are special Bajraktarević means, that can be considered as special Bajraktarević-type ψ -estimators. This can indicate some potential of Bajraktarević-type ψ -estimators as well in estimation of parameters for data coming from a mixture population.

By applying the general results from Section 3 of our arXiv version [Barczy and Páles \(2023\)](#) (recalled below as well), we solve the two problems in question. As one will see, our results are not easy and not immediate consequences of the general theory developed in Section 3 of [Barczy and Páles \(2023\)](#).

In Section 2, we recall the notion of Bajraktarević-type ψ -estimators and then we present the solution of the comparison problem for them (cf. Theorems 2.1 and 2.2). In Section 3, we characterize the equality of Bajraktarević-type ψ -estimators (cf. Theorems 3.1 and 3.2). We note that, surprisingly, in the heart of the proof of the equality problem for Bajraktarević-type estimators a result about Schwarzian derivative and rational functions come into play, see Lemma 3.1. We can also characterize the equality of quasarithmetic-type ψ -estimators, see Corollary 3.1. In Proposition 3.1, we derive a necessary and sufficient condition in order that two strictly increasing functions defined on a nondegenerate open interval be the Möbius transforms of each other. The results of the present paper are those contained in Section 4 in our arXiv version [Barczy and Páles \(2023\)](#).

In the rest of this section, we introduce the basic notations and concepts that are used throughout the paper, and we recall some results from Section 3 of our arXiv version [Barczy and Páles \(2023\)](#) that we need to use through the proofs.

Throughout this paper, we fix the following notations: the symbols \mathbb{N} , \mathbb{Z}_+ , \mathbb{Q} , \mathbb{R} , \mathbb{R}_+ , \mathbb{R}_{++} , and \mathbb{R}_{--} will stand for the sets of positive integers, non-negative integers, rational numbers, real numbers, non-negative real numbers, positive real numbers, and negative real numbers, respectively. For a subset $S \subseteq \mathbb{R}$, the convex hull of S (which is the smallest interval containing S) is denoted by $\text{conv}(S)$. A real interval will be called nondegenerate if it contains at least two distinct points. For each $n \in \mathbb{N}$, let us also introduce the set $\Lambda_n := \mathbb{R}_+^n \setminus \{(0, \dots, 0)\}$. All the random variables are defined on an appropriate probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Throughout this paper, let X be a nonempty set, Θ be a nondegenerate open interval of \mathbb{R} and let $\Psi(X, \Theta)$ denote the class of real-valued functions $\psi : X \times \Theta \rightarrow \mathbb{R}$ such that, for all $x \in X$, there exist $t_+, t_- \in \Theta$ such that $t_+ < t_-$ and $\psi(x, t_+) > 0 > \psi(x, t_-)$. Roughly speaking, a function $\psi \in \Psi(X, \Theta)$ satisfies the following property: for all $x \in X$, the function $t \mapsto \psi(x, t)$ changes sign on the interval Θ at least once.

Definition 1.1. For a function $f : \Theta \rightarrow \mathbb{R}$, consider the following three level sets:

$$\begin{aligned}\Theta_{f>0} &:= \{t \in \Theta : f(t) > 0\}, \\ \Theta_{f=0} &:= \{t \in \Theta : f(t) = 0\}, \\ \Theta_{f<0} &:= \{t \in \Theta : f(t) < 0\}.\end{aligned}$$

We say that $\vartheta \in \Theta$ is a *point of sign change (of decreasing-type) for f* if

$$f(t) > 0 \quad \text{for } t < \vartheta, \quad \text{and} \quad f(t) < 0 \quad \text{for } t > \vartheta.$$

Note that there can exist at most one element $\vartheta \in \Theta$ which is a point of sign change for f . Further, if f is continuous at a point ϑ of sign change, then $f(\vartheta) = 0$.

Definition 1.2. We say that a function $\psi \in \Psi(X, \Theta)$:

- (i) Possesses the property [C] (briefly, ψ is a C-function) if it is continuous in its second variable, i.e., if, for all $x \in X$, the mapping $\Theta \ni t \mapsto \psi(x, t)$ is continuous.
- (ii) Possesses the property $[T_n]$ (briefly, ψ is a T_n -function) for some $n \in \mathbb{N}$ if there exists a mapping $\vartheta_{n,\psi} : X^n \rightarrow \Theta$ such that, for all $\mathbf{x} = (x_1, \dots, x_n) \in X^n$ and $t \in \Theta$,

$$\psi_{\mathbf{x}}(t) = \sum_{i=1}^n \psi(x_i, t) \begin{cases} > 0 & \text{if } t < \vartheta_{n,\psi}(\mathbf{x}), \\ < 0 & \text{if } t > \vartheta_{n,\psi}(\mathbf{x}), \end{cases}$$

that is, for all $\mathbf{x} \in X^n$, the value $\vartheta_{n,\psi}(\mathbf{x})$ is a point of sign change for the function $\psi_{\mathbf{x}}$. If there is no confusion, instead of $\vartheta_{n,\psi}$ we simply write ϑ_n . We may call $\vartheta_{n,\psi}(\mathbf{x})$ as a generalized ψ -estimator for some unknown parameter in Θ based on the realization $\mathbf{x} = (x_1, \dots, x_n) \in X^n$. If, for each $n \in \mathbb{N}$, ψ is a T_n -function, then we say that ψ possesses the property [T] (briefly, ψ is a T-function).

- (iii) Possesses the property $[Z_n]$ (briefly, ψ is a Z_n -function) for some $n \in \mathbb{N}$ if it is a T_n -function and

$$\psi_{\mathbf{x}}(\vartheta_{n,\psi}(\mathbf{x})) = \sum_{i=1}^n \psi(x_i, \vartheta_{n,\psi}(\mathbf{x})) = 0 \quad \text{for all } \mathbf{x} = (x_1, \dots, x_n) \in X^n.$$

If, for each $n \in \mathbb{N}$, ψ is a Z_n -function, then we say that ψ possesses the property [Z] (briefly, ψ is a Z-function).

- (iv) Possesses the property $[T_n^\lambda]$ for some $n \in \mathbb{N}$ and $\lambda = (\lambda_1, \dots, \lambda_n) \in \Lambda_n$ (briefly, ψ is a T_n^λ -function) if there exists a mapping $\vartheta_{n,\psi}^\lambda : X^n \rightarrow \Theta$ such that, for all $\mathbf{x} = (x_1, \dots, x_n) \in X^n$ and $t \in \Theta$,

$$\psi_{\mathbf{x},\lambda}(t) = \sum_{i=1}^n \lambda_i \psi(x_i, t) \begin{cases} > 0 & \text{if } t < \vartheta_{n,\psi}^\lambda(\mathbf{x}), \\ < 0 & \text{if } t > \vartheta_{n,\psi}^\lambda(\mathbf{x}), \end{cases}$$

that is, for all $\mathbf{x} \in X^n$, the value $\vartheta_{n,\psi}^\lambda(\mathbf{x})$ is a point of sign change for the function $\psi_{\mathbf{x},\lambda}$. If there is no confusion, instead of $\vartheta_{n,\psi}^\lambda$ we simply write ϑ_n^λ . We may call $\vartheta_{n,\psi}^\lambda(\mathbf{x})$ as a weighted generalized ψ -estimator for some unknown parameter in Θ based on the realization $\mathbf{x} = (x_1, \dots, x_n) \in X^n$ and weights $(\lambda_1, \dots, \lambda_n) \in \Lambda_n$.

It can be seen that if ψ is continuous in its second variable and, for some $n \in \mathbb{N}$, it is a T_n -function, then it also a Z_n -function.

Given $q \in \mathbb{N}$ and properties

$$[P_1], \dots, [P_q] \in \{[C], [T], [Z]\} \cup \{[T_n], [Z_n] : n \in \mathbb{N}\} \cup \{[T_n^\lambda] : n \in \mathbb{N}, \lambda \in \Lambda_n\},$$

the subclass of $\Psi(X, \Theta)$ consisting of elements possessing the properties $[P_1], \dots, [P_q]$ will be denoted by $\Psi[P_1, \dots, P_q](X, \Theta)$, i.e.,

$$\Psi[P_1, \dots, P_q](X, \Theta) := \bigcap_{i=1}^q \Psi[P_i](X, \Theta).$$

For a function $\psi \in \Psi[T_1](X, \Theta)$, we introduce the notation

$$(1.1) \quad \Theta_\psi := \{t \in \Theta \mid \exists x, y \in X : \vartheta_{1,\psi}(x) < t < \vartheta_{1,\psi}(y)\}.$$

Then Θ_ψ is open, and it coincides with the interior of the convex hull of $\vartheta_{1,\psi}(X)$, see Section 3 of our arXiv version [Barczy and Páles \(2023\)](#). Consequently, Θ_ψ is an open (possibly degenerate) subinterval of Θ .

Next, we recall a result about the comparison of generalized ψ -estimators due to [Barczy and Páles \(2023, Theorem 3.1\)](#).

Theorem 1.1. *Let $\psi \in \Psi[T, Z_1](X, \Theta)$ and $\varphi \in \Psi[Z](X, \Theta)$. Then the following assertions are equivalent to each other:*

(i) *The inequality*

$$(1.2) \quad \vartheta_{n,\psi}(x_1, \dots, x_n) \leq \vartheta_{n,\varphi}(x_1, \dots, x_n)$$

holds for each $n \in \mathbb{N}$ and $x_1, \dots, x_n \in X$.

(ii) *The inequality*

$$\vartheta_{k+m,\psi}(\underbrace{x, \dots, x}_k, \underbrace{y, \dots, y}_m) \leq \vartheta_{k+m,\varphi}(\underbrace{x, \dots, x}_k, \underbrace{y, \dots, y}_m)$$

holds for each $k, m \in \mathbb{N}$ and $x, y \in X$.

(iii) *For all $x \in X$, we have $\vartheta_{1,\psi}(x) \leq \vartheta_{1,\varphi}(x)$, and the inequality*

$$\psi(x, t)\varphi(y, t) \leq \psi(y, t)\varphi(x, t)$$

is valid for all $t \in \Theta$ and for all $x, y \in X$ with $\vartheta_{1,\varphi}(x) < t < \vartheta_{1,\varphi}(y)$.

(iv) *For all $x \in X$, we have $\vartheta_{1,\psi}(x) \leq \vartheta_{1,\varphi}(x)$, and there exists a nonnegative function $p : \Theta_\varphi \rightarrow \mathbb{R}_+$ such that*

$$\psi(z, t) \leq p(t)\varphi(z, t), \quad z \in X, t \in \Theta_\varphi.$$

Under some additional regularity assumptions on ψ and φ , [Barczy and Páles \(2023, Theorem 3.4\)](#) derived another set of conditions that is equivalent to (1.2). We also recall this result below.

Theorem 1.2. *Let $\psi, \varphi \in \Psi[C, Z](X, \Theta)$. Assume that $\vartheta_{1,\psi} = \vartheta_{1,\varphi} =: \vartheta_1$ on X , $\vartheta_1(X) = \Theta$, and, for all $x \in X$, the maps*

$$\Theta \ni t \mapsto \psi(x, t) \quad \text{and} \quad \Theta \ni t \mapsto \varphi(x, t)$$

are differentiable at $\vartheta_1(x)$ with a non-vanishing derivative. Then any of the equivalent assertions (i), (ii), (iii) and (iv) of Theorem 1.1 is equivalent to the following one:

(v) *For all $x, y \in X$, we have*

$$-\frac{\psi(y, \vartheta_1(x))}{\partial_2 \psi(x, \vartheta_1(x))} \leq -\frac{\varphi(y, \vartheta_1(x))}{\partial_2 \varphi(x, \vartheta_1(x))}.$$

Finally, we recall a result about the equality of generalized ψ -estimators due to [Barczy and Páles \(2023, Theorem 3.5\)](#).

Theorem 1.3. Let $\psi \in \Psi[T, Z_1](X, \Theta)$ and $\varphi \in \Psi[Z](X, \Theta)$. Assume that $\vartheta_{1,\psi} = \vartheta_{1,\varphi}$ on X . Then $\Theta_\psi = \Theta_\varphi$ and the following assertions are equivalent:

(i) The equality

$$\vartheta_{n,\psi}(x_1, \dots, x_n) = \vartheta_{n,\varphi}(x_1, \dots, x_n)$$

holds for each $n \in \mathbb{N}$ and $x_1, \dots, x_n \in X$.

(ii) The equality

$$\vartheta_{k+m,\psi}(\underbrace{x, \dots, x}_k, \underbrace{y, \dots, y}_m) = \vartheta_{k+m,\varphi}(\underbrace{x, \dots, x}_k, \underbrace{y, \dots, y}_m)$$

holds for each $k, m \in \mathbb{N}$ and $x, y \in X$.

(iii) There exists a positive function $p : \Theta_\varphi \rightarrow (0, \infty)$ such that

$$\psi(z, t) = p(t)\varphi(z, t), \quad z \in X, t \in \Theta_\varphi.$$

2. COMPARISON OF BAJRAKTAREVIĆ-TYPE ψ -ESTIMATORS

In this section we apply Theorems 1.1 and 1.2 for solving the comparison problem of Bajraktarević-type estimators that are special generalized ψ -estimators.

First, we recall the notions of Bajraktarević-type functions and then Bajraktarević-type estimators. Let $f : \Theta \rightarrow \mathbb{R}$ be strictly increasing, $p : X \rightarrow \mathbb{R}_{++}$ and $F : X \rightarrow \text{conv}(f(\Theta))$. In terms of these functions, define $\psi : X \times \Theta \rightarrow \mathbb{R}$ by

$$(2.1) \quad \psi(x, t) := p(x)(F(x) - f(t)), \quad x \in X, t \in \Theta.$$

By Lemma 1 in Grünwald and Páles (2020), there exists a uniquely determined monotone function $g : \text{conv}(f(\Theta)) \rightarrow \Theta$ such that g is the left inverse of f , i.e.,

$$(g \circ f)(t) = t, \quad t \in \Theta.$$

Furthermore, g is monotone in the same sense as f (i.e., f is monotone increasing), is continuous, and the following relation holds:

$$(f \circ g)(y) = y, \quad y \in f(\Theta).$$

The function $g : \text{conv}(f(\Theta)) \rightarrow \Theta$ is called the *generalized left inverse of the strictly increasing function $f : \Theta \rightarrow \mathbb{R}$* and is denoted by $f^{(-1)}$. It is clear that the restriction of $f^{(-1)}$ to $f(\Theta)$ is the inverse of f in the standard sense, which is also strictly increasing. Therefore, $f^{(-1)}$ is the continuous and monotone extension of the inverse of f to the smallest interval containing the range of f , that is, to the convex hull of $f(\Theta)$.

Recall also that, by Proposition 6 in Barczy and Páles (2025), under the above assumptions, ψ is a T_n^λ -function for each $n \in \mathbb{N}$ and $\lambda \in \Lambda_n$, and

$$(2.2) \quad \vartheta_{n,\psi}^\lambda(\mathbf{x}) = f^{(-1)}\left(\frac{\lambda_1 p(x_1)F(x_1) + \dots + \lambda_n p(x_n)F(x_n)}{\lambda_1 p(x_1) + \dots + \lambda_n p(x_n)}\right)$$

for each $n \in \mathbb{N}$, $\mathbf{x} = (x_1, \dots, x_n) \in X^n$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n) \in \Lambda_n$. In particular, $\vartheta_{1,\psi} = f^{(-1)} \circ F$ holds. The value $\vartheta_{n,\psi}^{\boldsymbol{\lambda}}(\mathbf{x})$ given by (2.2) can be called a Bajraktarević-type ψ -estimator of some unknown parameter in Θ based on the realization $\mathbf{x} = (x_1, \dots, x_n) \in X^n$ and weights $(\lambda_1, \dots, \lambda_n) \in \Lambda_n$ corresponding to the Bajraktarević-type function given by (2.1). In particular, if $p = 1$ is a constant function in (2.1), then we speak about a quasi-arithmetic-type ψ -estimator.

As a first result, we give a necessary and sufficient condition in order that $\Theta_\psi = \emptyset$ hold, where Θ_ψ is given by (1.1).

Lemma 2.1. *Let $f : \Theta \rightarrow \mathbb{R}$ be a strictly increasing function, $p : X \rightarrow \mathbb{R}_{++}$, $F : X \rightarrow \text{conv}(f(\Theta))$, and define $\psi : X \times \Theta \rightarrow \mathbb{R}$ by (2.1). Then $\Theta_\psi = \emptyset$ holds if and only if there exists $t_0 \in \Theta$ such that the range $F(X)$ of F is contained in $[f(t_0 - 0), f(t_0 + 0)]$, where $f(t_0 - 0)$ and $f(t_0 + 0)$ denote the left and right hand limits of f at t_0 , respectively.*

Proof: First, let us suppose that there exists $t_0 \in \Theta$ such that $F(X) \subseteq J_f(t_0) := [f(t_0 - 0), f(t_0 + 0)]$. Then, using that f is strictly increasing, for all $x \in X$ and $t', t'' \in \Theta$ with $t' < t_0 < t''$, we have that $f(t') < f(t_0 - 0) \leq F(x) \leq f(t_0 + 0) < f(t'')$, and therefore, taking into account that p is positive, for all $x \in X$, we get

$$p(x)(F(x) - f(t)) \begin{cases} > 0 & \text{if } t < t_0, t \in \Theta, \\ < 0 & \text{if } t > t_0, t \in \Theta. \end{cases}$$

Hence, $\vartheta_{1,\psi}(x) = t_0$ for all $x \in X$. This yields that $\Theta_\psi = \emptyset$.

To prove the converse statement, we check that if there does not exist $t_0 \in \Theta$ such that $F(X) \subseteq J_f(t_0)$, then $\Theta_\psi \neq \emptyset$. Since f is strictly increasing, we have

$$\text{conv}(f(\Theta)) = \bigcup_{t \in \Theta} J_f(t),$$

and $J_f(t') \cap J_f(t'') = \emptyset$ for all $t', t'' \in \Theta$ with $t' \neq t''$. Using also that $F(X) \subseteq \text{conv}(f(\Theta))$, there exist $t_1, t_2 \in \Theta$ with $t_1 < t_2$ such that $F(X) \cap J_f(t_i) \neq \emptyset$, $i = 1, 2$. Hence there exist $x_1, x_2 \in X$ such that $F(x_i) \in J_f(t_i)$, $i = 1, 2$. Consequently, similarly as before, we have

$$p(x_i)(F(x_i) - f(t)) \begin{cases} > 0 & \text{if } t < t_i, t \in \Theta, \\ < 0 & \text{if } t > t_i, t \in \Theta, \end{cases} \quad i = 1, 2,$$

yielding that $\vartheta_{1,\psi}(x_i) = t_i$, $i = 1, 2$. Therefore, $(t_1, t_2) \subseteq \Theta_\psi$, showing that Θ_ψ is not empty, as expected. \square

The next lemma is connected to Theorem 1 in Barczy and Páles (2025), where we derived several implications between the property $[T]$ of a function $\Psi(X, \Theta)$ and the monotonicity properties of the map (2.3) defined in the next lemma.

Lemma 2.2. *Let $f : \Theta \rightarrow \mathbb{R}$ be a strictly increasing function, $p : X \rightarrow \mathbb{R}_{++}$, $F : X \rightarrow \text{conv}(f(\Theta))$, and define $\psi : X \times \Theta \rightarrow \mathbb{R}$ by (2.1). Then, for all $x, y \in X$ with $\vartheta_{1,\psi}(x) < \vartheta_{1,\psi}(y)$, the map*

$$(2.3) \quad (\vartheta_{1,\psi}(x), \vartheta_{1,\psi}(y)) \ni u \mapsto -\frac{\psi(x, u)}{\psi(y, u)}$$

is positive and strictly increasing.

Proof: Let $x, y \in X$ with $\vartheta_{1,\psi}(x) < \vartheta_{1,\psi}(y)$ and let $u \in (\vartheta_{1,\psi}(x), \vartheta_{1,\psi}(y))$ be arbitrary. Then $\psi(x, u) < 0 < \psi(y, u)$, which proves that the map (2.3) is positive valued. To see the strict monotonicity property, let additionally $v \in (\vartheta_{1,\psi}(x), \vartheta_{1,\psi}(y))$ be arbitrary with $u < v$. Then $\psi(x, v) < 0 < \psi(y, v)$, which implies that $F(x) < f(v) < F(y)$. Thus $F(x) < F(y)$ and, by the strict monotonicity of f , we also have $f(u) < f(v)$. Therefore

$$(F(y) - F(x))(f(v) - f(u)) > 0,$$

which is equivalent to the inequality

$$(F(x) - f(u))(F(y) - f(v)) > (F(y) - f(u))(F(x) - f(v)).$$

Multiplying this inequality by $\frac{-p(x)}{p(y)(F(y)-f(u))(F(y)-f(v))} < 0$ side by side, it follows that

$$-\frac{\psi(x, u)}{\psi(y, u)} = -\frac{p(x)(F(x) - f(u))}{p(y)(F(y) - f(u))} < -\frac{p(x)(F(x) - f(v))}{p(y)(F(y) - f(v))} = -\frac{\psi(x, v)}{\psi(y, v)}.$$

This completes the proof of the strict increasingness of the map (2.3). We note that the statement also follows from the proof of Proposition 6 in Barczy and Páles (2025). \square

In the following result, we describe sufficient conditions which imply that the function ψ defined by (2.1) possesses the property $[Z_1]$ and $[Z]$, respectively.

Lemma 2.3. *Let $f : \Theta \rightarrow \mathbb{R}$ be a strictly increasing function, $p : X \rightarrow \mathbb{R}_{++}$, $F : X \rightarrow f(\Theta)$ and define $\psi : X \times \Theta \rightarrow \mathbb{R}$ by (2.1). Then ψ has the property $[Z_1]$ and*

$$(2.4) \quad \Theta_\psi = \{t \in \Theta \mid \exists x, y \in X : F(x) < f(t) < F(y)\}.$$

If, in addition, $\text{conv}(F(X)) \subseteq f(\Theta)$, then ψ has the property $[Z]$ as well.

Proof: Since $F(X) \subseteq f(\Theta)$, the restriction of $f^{(-1)}$ onto $F(X)$ is the strictly increasing inverse of f in the standard sense restricted to $F(X)$. Thus, for all $x \in X$, we have

$$\begin{aligned} \psi(x, \vartheta_{1,\psi}(x)) &= p(x)(F(x) - f(\vartheta_{1,\psi}(x))) = p(x)(F(x) - f(f^{(-1)}(F(x)))) \\ &= p(x)(F(x) - F(x)) = 0, \end{aligned}$$

yielding that ψ has the property $[Z_1]$. Furthermore, using also that $(f^{(-1)} \circ f)(t) = t$, $t \in \Theta$, we get that

$$\begin{aligned} \Theta_\psi &= \{t \in \Theta \mid \exists x, y \in X : \vartheta_{1,\psi}(x) < t < \vartheta_{1,\psi}(y)\} \\ &= \{t \in \Theta \mid \exists x, y \in X : f^{(-1)}(F(x)) < t < f^{(-1)}(F(y))\} \\ &= \{t \in \Theta \mid \exists x, y \in X : (f \circ f^{(-1)})(F(x)) < f(t) < (f \circ f^{(-1)})(F(y))\} \\ &= \{t \in \Theta \mid \exists x, y \in X : F(x) < f(t) < F(y)\}, \end{aligned}$$

as desired.

To prove the last assertion, let us assume that $\text{conv}(F(X)) \subseteq f(\Theta)$. Then, by (2.2), for each $n \in \mathbb{N}$ and $\mathbf{x} = (x_1, \dots, x_n) \in X^n$, we have

$$\begin{aligned} \sum_{i=1}^n \psi(x_i, \vartheta_{n,\psi}(\mathbf{x})) &= \sum_{i=1}^n p(x_i)(F(x_i) - f(\vartheta_{n,\psi}(\mathbf{x}))) \\ &= \sum_{i=1}^n p(x_i)F(x_i) - \sum_{i=1}^n p(x_i)f\left(f^{(-1)}\left(\sum_{j=1}^n \frac{p(x_j)}{p(x_1) + \dots + p(x_n)}F(x_j)\right)\right) \\ &= \sum_{i=1}^n p(x_i)F(x_i) - \sum_{i=1}^n p(x_i) \sum_{j=1}^n \frac{p(x_j)}{p(x_1) + \dots + p(x_n)}F(x_j) = 0, \end{aligned}$$

where we used that $\sum_{j=1}^n \frac{p(x_j)}{p(x_1) + \dots + p(x_n)}F(x_j) \in \text{conv}(F(X)) \subseteq f(\Theta)$. \square

In the next remark, we point out the fact that (2.4) does not hold in general, showing that the assumption $F(X) \subseteq f(\Theta)$ in Lemma 2.3 is indispensable.

Remark 2.1. Let $X := \{x_1, x_2\}$, $\Theta := \mathbb{R}$, let $p : X \rightarrow \mathbb{R}_{++}$, $f : \mathbb{R} \rightarrow \mathbb{R}$ be defined by

$$f(t) := \begin{cases} t & \text{if } t \leq 1, \\ t + 1 & \text{if } 1 < t \leq 2, \\ t + 2 & \text{if } t > 2, \end{cases}$$

and $F : X \rightarrow \mathbb{R}$ be such that $F(x_1) := 1$ and $F(x_2) := 3, 5$. Then $\text{conv}(f(\Theta)) = \mathbb{R}$, and

$$p(x_1)(F(x_1) - f(t)) \begin{cases} > 0 & \text{if } t < 1, \\ = 0 & \text{if } t = 1, \\ < 0 & \text{if } t > 1, \end{cases}$$

and

$$p(x_2)(F(x_2) - f(t)) \begin{cases} > 0 & \text{if } t \leq 2, \\ < 0 & \text{if } t > 2, \end{cases}$$

yielding that $\vartheta_{1,\psi}(x_j) = j$, $j = 1, 2$. Hence $\Theta_\psi = (1, 2)$. However,

$$\{t \in \Theta \mid \exists x, y \in X : F(x) < f(t) < F(y)\} = (1, 2],$$

which does not coincide with $\Theta_\psi = (1, 2)$. Note that $F(X) \subseteq f(\Theta)$ is also not valid, and ψ does not have the property $[Z_1]$, since $\psi(x_2, \vartheta_{1,\psi}(x_2)) = \psi(x_2, 2) = p(x_2)(F(x_2) - f(2)) = \frac{p(x_2)}{2} > 0$. \square

Below, we solve the comparison problem for Bajraktarević-type estimators.

Theorem 2.1. Let $f, g : \Theta \rightarrow \mathbb{R}$ be strictly increasing functions, $p, q : X \rightarrow \mathbb{R}_{++}$, $F : X \rightarrow f(\Theta)$, $G : X \rightarrow g(\Theta)$, and suppose that $\text{conv}(G(X)) \subseteq g(\Theta)$. Let $\psi : X \times \Theta \rightarrow \mathbb{R}$ and $\varphi : X \times \Theta \rightarrow \mathbb{R}$ be given by

$$(2.5) \quad \begin{aligned} \psi(x, t) &:= p(x)(F(x) - f(t)), & x \in X, t \in \Theta, \\ \varphi(x, t) &:= q(x)(G(x) - g(t)), & x \in X, t \in \Theta. \end{aligned}$$

Assume that $(f^{(-1)} \circ F)(x) \leq (g^{(-1)} \circ G)(x)$, $x \in X$. Then $\vartheta_{n,\psi}(\mathbf{x}) \leq \vartheta_{n,\varphi}(\mathbf{x})$ holds for each $n \in \mathbb{N}$ and $\mathbf{x} \in X^n$ if and only if the inequality

$$(2.6) \quad \frac{q(y)}{q(x)} \cdot \frac{G(y) - g(t)}{G(x) - g(t)} \leq \frac{p(y)}{p(x)} \cdot \frac{F(y) - f(t)}{F(x) - f(t)}$$

is valid for all $t \in \Theta$ and for all $x, y \in X$ with $G(x) < g(t) < G(y)$.

Proof: By Proposition 6 in Barczy and Páles (2025) and Lemma 2.3, ψ has the properties [T] and $[Z_1]$, and φ has the property [Z]. Using Theorem 1.1, we have that $\vartheta_{n,\psi}(\mathbf{x}) \leq \vartheta_{n,\varphi}(\mathbf{x})$ holds for each $n \in \mathbb{N}$ and $\mathbf{x} \in X^n$ if and only if the inequality

$$\psi(x, t)\varphi(y, t) \leq \psi(y, t)\varphi(x, t)$$

is valid for all $t \in \Theta$ and for all $x, y \in X$ with $\vartheta_{1,\varphi}(x) < t < \vartheta_{1,\varphi}(y)$.

Using $G(X) \subseteq g(\Theta)$, Lemma 2.3 yields that

$$\Theta_\varphi = \{t \in \Theta \mid \exists x, y \in X : G(x) < g(t) < G(y)\},$$

and for all $t \in \Theta$ and $x, y \in X$, the inequality $\vartheta_{1,\varphi}(x) < t < \vartheta_{1,\varphi}(y)$ holds if and only if $G(x) < g(t) < G(y)$. Consequently, $\vartheta_{n,\psi}(\mathbf{x}) \leq \vartheta_{n,\varphi}(\mathbf{x})$ holds for each $n \in \mathbb{N}$ and $\mathbf{x} \in X^n$ if and only if

$$(2.7) \quad p(x)(F(x) - f(t))q(y)(G(y) - g(t)) \leq p(y)(F(y) - f(t))q(x)(G(x) - g(t))$$

is valid for all $t \in \Theta$ and for all $x, y \in X$ with $G(x) < g(t) < G(y)$. Using that f is strictly increasing, $F(X) \subseteq f(\Theta)$, and $g^{(-1)}$ restricted to $g(\Theta)$ is strictly increasing, for all $t \in \Theta$ and for all $x \in X$ with $G(x) < g(t)$, we have that

$$\begin{aligned} F(x) &= (f \circ f^{(-1)})(F(x)) = f(f^{(-1)}(F(x))) \\ &\leq f(g^{(-1)}(G(x))) < f(g^{(-1)}(g(t))) = f(t). \end{aligned}$$

Consequently, $G(x) - g(t) < 0$ and $F(x) - f(t) < 0$ in the inequality (2.7), and hence by rearranging it, the assertion follows. \square

In the next result, under some additional regularity assumptions on f and g , we derive another set of conditions that is equivalent to (2.6).

Theorem 2.2. Let $f, g : \Theta \rightarrow \mathbb{R}$ be strictly increasing functions, $p, q : X \rightarrow \mathbb{R}_{++}$, $F : X \rightarrow f(\Theta)$, $G : X \rightarrow g(\Theta)$, and suppose that $F(X) = f(\Theta)$ and $\text{conv}(G(X)) \subseteq g(\Theta)$. Let $\psi : X \times \Theta \rightarrow \mathbb{R}$ and $\varphi : X \times \Theta \rightarrow \mathbb{R}$ be given by (2.5). Assume that $(f^{(-1)} \circ F)(x) = (g^{(-1)} \circ G)(x) =: \vartheta_1(x)$, $x \in X$, and that f and g are differentiable at $\vartheta_1(x)$ for all $x \in X$ with non-vanishing (and hence positive) derivatives. Then $\vartheta_{n,\psi}(\mathbf{x}) \leq \vartheta_{n,\varphi}(\mathbf{x})$ holds for each $n \in \mathbb{N}$ and $\mathbf{x} \in X^n$ if and only if the inequality

$$\frac{p(y)}{p(x)} \cdot \frac{F(y) - F(x)}{f'(f^{(-1)}(F(x)))} \leq \frac{q(y)}{q(x)} \cdot \frac{G(y) - G(x)}{g'(g^{(-1)}(G(x)))}$$

is valid for all $x, y \in X$.

Proof: Since $F(X) = f(\Theta)$, we have that

$$\vartheta_1(X) = (f^{(-1)} \circ F)(X) = f^{(-1)}(F(X)) = f^{(-1)}(f(\Theta)) = \Theta.$$

Further, for all $x \in X$, we have

$$\begin{aligned}\partial_2\psi(x, \vartheta_1(x)) &= -p(x)f'(\vartheta_1(x)) = -p(x)f'(f^{(-1)}(F(x))), \\ \partial_2\varphi(x, \vartheta_1(x)) &= -q(x)g'(\vartheta_1(x)) = -q(x)g'(g^{(-1)}(G(x))).\end{aligned}$$

Consequently, using Theorem 1.2, we have that $\vartheta_{n,\psi}(\mathbf{x}) \leq \vartheta_{n,\varphi}(\mathbf{x})$ holds for each $n \in \mathbb{N}$ and $\mathbf{x} \in X^n$ if and only if

$$\frac{p(y)}{p(x)} \cdot \frac{F(y) - f(\vartheta_1(x))}{f'(\vartheta_1(x))} \leq \frac{q(y)}{q(x)} \cdot \frac{G(y) - g(\vartheta_1(x))}{g'(\vartheta_1(x))}$$

is valid for all $x, y \in X$, which yields the statement, since

$$f(\vartheta_1(x)) = f(f^{(-1)}(F(x))) = (f \circ f^{(-1)})(F(x)) = F(x), \quad x \in X,$$

due to the condition $F(X) \subseteq f(\Theta)$, and similarly, we also have $g(\vartheta_1(x)) = G(x)$, $x \in X$. \square

In the next result, among others, we point out that, under the assumptions of Theorem 2.2, the function g is continuous.

Lemma 2.4. *Let $f, g : \Theta \rightarrow \mathbb{R}$ be strictly increasing functions, $F : X \rightarrow f(\Theta)$, $G : X \rightarrow g(\Theta)$, and suppose that $F(X) = f(\Theta)$ and $\text{conv}(G(X)) \subseteq g(\Theta)$. Assume that $(f^{(-1)} \circ F)(x) = (g^{(-1)} \circ G)(x)$, $x \in X$. Then:*

- (i) $G(X) = g(\Theta)$ and this set is convex.
- (ii) g is continuous, and $G(x) = (g \circ f^{(-1)})(F(x))$, $x \in X$.

Proof: (i). Since $F(X) = f(\Theta)$, we have that

$$(f^{(-1)} \circ F)(X) = f^{(-1)}(F(X)) = f^{(-1)}(f(\Theta)) = \Theta,$$

which implies that $(g^{(-1)} \circ G)(X) = (f^{(-1)} \circ F)(X) = \Theta$ holds as well. Hence, using that $G(X) \subseteq g(\Theta)$, we have that

$$g(\Theta) = g((g^{(-1)} \circ G)(X)) = (g \circ g^{(-1)})(G(X)) = G(X).$$

Consequently, $\text{conv}(g(\Theta)) = \text{conv}(G(X))$, and, since $\text{conv}(G(X)) \subseteq g(\Theta)$, we obtain that $g(\Theta) \subseteq \text{conv}(g(\Theta)) = \text{conv}(G(X)) \subseteq g(\Theta)$, yielding that $g(\Theta) = \text{conv}(G(X))$ is a convex set.

(ii). Since g is strictly increasing and its range $g(\Theta)$ is convex (see part (i)), we can see that g is continuous as well. Finally, note that the equality $(f^{(-1)} \circ F)(x) = (g^{(-1)} \circ G)(x)$, $x \in X$, implies that $G(x) = (g \circ f^{(-1)})(F(x))$, $x \in X$, since $G(X) \subseteq g(\Theta)$. \square

3. EQUALITY OF BAJRAKTAREVIĆ-TYPE ψ -ESTIMATORS

In this section, we apply Theorem 1.3 for solving the equality problem for Bajraktarević-type estimators. In the proof, we will use a result on the Schwarzian derivative of a function. Given a nondegenerate open interval $I \subseteq \mathbb{R}$, for a three times differentiable function $h : I \rightarrow \mathbb{R}$ with a nonvanishing first derivative, its Schwarzian derivative $S_h : I \rightarrow \mathbb{R}$ is defined by

$$S_h(x) := \frac{h'''(x)}{h'(x)} - \frac{3}{2} \left(\frac{h''(x)}{h'(x)} \right)^2, \quad x \in I.$$

The following result is well-known, see, e.g., Grünwald and Páles (2022, Corollary 3).

Lemma 3.1. *Let $I \subseteq \mathbb{R}$ be a nondegenerate open interval, and $h : I \rightarrow \mathbb{R}$ be a three times differentiable function such that h' does not vanish on I . Then $S_h(x) = 0$, $x \in I$, holds if and only if there exist four constants $a, b, c, d \in \mathbb{R}$ with $ad \neq bc$ and $0 \notin cI + d$ such that*

$$h(x) = \frac{ax + b}{cx + d}, \quad x \in I.$$

In the proof of the subsequent theorems, the following auxiliary result plays an important role.

Lemma 3.2. *Let I be a nondegenerate open interval of \mathbb{R} . Let $f, g : I \rightarrow \mathbb{R}$ be strictly increasing functions such that there exist four constants $a, b, c, d \in \mathbb{R}$ with $0 \notin cf(I) + d$ and*

$$g(t) = \frac{af(t) + b}{cf(t) + d}, \quad t \in I.$$

Then $ad > bc$ and $cf + d$ is either everywhere positive or everywhere negative on I .

Proof: The condition $0 \notin cf(I) + d$ yields that $(c, d) \neq (0, 0)$. If $ad = bc$, then there exists $\lambda \in \mathbb{R}$ such that $(a, b) = \lambda(c, d)$. In this case, we get that $g(t) = \lambda$ for all $t \in I$, which contradicts the strict monotonicity of g . Thus $ad \neq bc$ must hold.

Next, we check that $cf + d$ is either everywhere positive or everywhere negative on I . On the contrary, if $cf + d$ changes sign in I , then c cannot be zero and hence $cf + d$ is also strictly monotone. Therefore, using also that I is a nondegenerate open interval, there exists a unique point $\tau \in I$ such that $cf(t) + d > (<) 0$ for all $t < \tau$, $t \in I$, and $cf(t) + d < (>) 0$ for all $t > \tau$, $t \in I$. Let $t < \tau < r < s$ be arbitrarily fixed elements of I . Then $(cf(t) + d)(cf(r) + d) < 0$ and $(cf(r) + d)(cf(s) + d) > 0$. Consequently, using the strict increasingness of g , the inequalities

$$\frac{af(t) + b}{cf(t) + d} = g(t) < g(r) = \frac{af(r) + b}{cf(r) + d}$$

and

$$\frac{af(r) + b}{cf(r) + d} = g(r) < g(s) = \frac{af(s) + b}{cf(s) + d}$$

imply that

$$\begin{aligned} (af(t) + b)(cf(r) + d) &> (af(r) + b)(cf(t) + d), \\ (af(r) + b)(cf(s) + d) &< (af(s) + b)(cf(r) + d), \end{aligned}$$

or equivalently,

$$(3.1) \quad 0 > (ad - bc)(f(r) - f(t)) \quad \text{and} \quad 0 < (ad - bc)(f(s) - f(r)).$$

Since f is also strictly increasing, we have that $f(t) < f(r) < f(s)$. Therefore, $(ad - bc)(f(r) - f(t))$ and $(ad - bc)(f(s) - f(r))$ should have the same signs. This together with the inequalities (3.1) lead us to a contradiction.

Finally, to show that $ad > bc$, let $r, s \in I$ with $r < s$ be fixed. Then, using that $cf + d$ does not change sign in I , we have $(cf(r) + d)(cf(s) + d) > 0$, and therefore the inequality $g(r) < g(s)$, in the same way as above, implies $0 < (ad - bc)(f(s) - f(r))$. This, in view of the strict increasingness of f yields that $ad - bc > 0$. \square

Theorem 3.1. *Let $f, g : \Theta \rightarrow \mathbb{R}$ be strictly increasing functions such that f is continuous, $p, q : X \rightarrow \mathbb{R}_{++}$, $F : X \rightarrow f(\Theta)$, $G : X \rightarrow g(\Theta)$, and suppose that $F(X) = f(\Theta)$ and $\text{conv}(G(X)) \subseteq g(\Theta)$. Let $\psi : X \times \Theta \rightarrow \mathbb{R}$ and $\varphi : X \times \Theta \rightarrow \mathbb{R}$ be given by (2.5). If $\vartheta_{n,\psi}(\mathbf{x}) = \vartheta_{n,\varphi}(\mathbf{x})$ holds for each $n \in \mathbb{N}$ and $\mathbf{x} \in X^n$, then there exist four constants $a, b, c, d \in \mathbb{R}$ with $ad \neq bc$ and $0 \notin cf(\Theta) + d$ such that*

$$(3.2) \quad \begin{aligned} g(t) &= \frac{af(t) + b}{cf(t) + d}, & t \in \Theta, \\ G(x) &= \frac{aF(x) + b}{cF(x) + d}, & x \in X, \\ q(x) &= (cF(x) + d)p(x), & x \in X. \end{aligned}$$

Proof: Since f is strictly increasing and continuous and Θ is a non-degenerate open interval, we have $f(\Theta)$ is a non-degenerate open interval. Hence $\text{conv}(f(\Theta)) = f(\Theta) = F(X)$, and then, as a consequence of Lemma 2.3, we get $\psi \in \Psi[Z](X, \Theta)$. Further, since $\text{conv}(G(X)) \subseteq g(\Theta)$, Lemma 2.3 also yields that $\varphi \in \Psi[Z](X, \Theta)$.

We first verify that $\Theta_\psi = \Theta$. The inclusion $\Theta_\psi \subseteq \Theta$ is trivial. To prove the reversed one, let $t \in \Theta$ be arbitrary. Now choose $r, s \in \Theta$ such that $r < t < s$. Using that $F(X) = f(\Theta)$, we can find $x, y \in X$ such that $f(r) = F(x)$ and $f(s) = F(y)$. Since f is strictly increasing, it follows that $F(x) = f(r) < f(t) < f(s) = F(y)$, showing that t belongs to the set

$$\{t \in \Theta \mid \exists x, y \in X : F(x) < f(t) < F(y)\},$$

which, according to (2.4), equals Θ_ψ .

Assume that $\vartheta_{n,\psi}(\mathbf{x}) = \vartheta_{n,\varphi}(\mathbf{x})$ holds for each $n \in \mathbb{N}$ and $\mathbf{x} \in X^n$. Then, in the case $n = 1$, this equality and (2.2) imply that $(f^{(-1)} \circ F)(x) = \vartheta_{1,\psi}(x) = \vartheta_{1,\varphi}(x) = (g^{(-1)} \circ G)(x)$, $x \in X$. Hence, according to Lemma 2.4, we get that $G(X) = \text{conv}(G(X)) = g(\Theta)$ is a convex set and g is continuous. Then, similarly as we derived $\Theta_\psi = \Theta$, we have that $\Theta_\varphi = \Theta$ holds as well.

For all $x, y \in X$ with $G(x) < G(y)$, let us introduce the notation

$$\Theta_{x,y} := \{t \in \Theta \mid G(x) < g(t) < G(y)\}.$$

Using that $F(X) \subseteq f(\Theta)$, $G(X) \subseteq g(\Theta)$, and that the restrictions of $f^{(-1)}$ and $g^{(-1)}$ to $f(\Theta)$ and $g(\Theta)$ are the strictly increasing inverses of f and g in the standard sense, respectively, for all $x, y \in X$ with $G(x) < G(y)$, we have

$$\begin{aligned}
 \Theta_{x,y} &= \{t \in \Theta \mid g^{(-1)}(G(x)) < t < g^{(-1)}(G(y))\} \\
 (3.3) \quad &= \{t \in \Theta \mid f^{(-1)}(F(x)) < t < f^{(-1)}(F(y))\} \\
 &= \{t \in \Theta \mid F(x) < f(t) < F(y)\}.
 \end{aligned}$$

The previous argument also shows that for all $x, y \in X$, we get $G(x) < G(y)$ if and only if $F(x) < F(y)$, and the set $\Theta_{x,y}$ is a nonempty open interval for all $x, y \in X$ with $G(x) < G(y)$, since it is the intersection of the open intervals $(g^{(-1)}(G(x)), g^{(-1)}(G(y)))$ and Θ .

In view of (3.3), Lemma 2.3 and the equality $\Theta_\psi = \Theta$ we can see that

$$\begin{aligned}
 \Theta_\varphi &= \bigcup_{\{x,y \in X: G(x) < G(y)\}} \Theta_{x,y} \\
 (3.4) \quad &= \bigcup_{\{x,y \in X: F(x) < F(y)\}} \{t \in \Theta \mid F(x) < f(t) < F(y)\} = \Theta_\psi = \Theta.
 \end{aligned}$$

Using that the image under f of a union of subsets is the union of the images under f of the given subsets, (3.4) immediately yields that

$$(3.5) \quad \bigcup_{\{x,y \in X: G(x) < G(y)\}} f(\Theta_{x,y}) = f(\Theta),$$

which is an open interval, since Θ is a nonempty open interval and f is strictly increasing and continuous.

Using Theorem 2.1, we have that

$$(3.6) \quad \frac{q(y)}{q(x)} \cdot \frac{G(y) - g(t)}{G(x) - g(t)} = \frac{p(y)}{p(x)} \cdot \frac{F(y) - f(t)}{F(x) - f(t)}$$

holds for all $t \in \Theta$ and for all $x, y \in X$ with $G(x) < g(t) < G(y)$, or equivalently, (3.6) holds for all $x, y \in X$ with $G(x) < G(y)$ and for all $t \in \Theta_{x,y}$.

One can readily check that for all $x, y \in X$ with $G(x) < G(y)$ and for all $t \in \Theta_{x,y}$, the equality (3.6) is equivalent to any of the following three equalities:

$$\begin{aligned}
 &p(x)(F(x) - f(t))q(y)(G(y) - g(t)) = p(y)(F(y) - f(t))q(x)(G(x) - g(t)), \\
 (3.7) \quad &\left(p(y)(F(y) - f(t))q(x) - p(x)(F(x) - f(t))q(y) \right) g(t) \\
 &= p(y)(F(y) - f(t))q(x)G(x) - p(x)(F(x) - f(t))q(y)G(y), \\
 &(c_{x,y}f(t) + d_{x,y})g(t) = a_{x,y}f(t) + b_{x,y},
 \end{aligned}$$

where

$$\begin{aligned}
 a_{x,y} &:= p(x)q(y)G(y) - p(y)q(x)G(x), \\
 b_{x,y} &:= p(y)q(x)F(y)G(x) - p(x)q(y)F(x)G(y), \\
 c_{x,y} &:= p(x)q(y) - p(y)q(x), \\
 d_{x,y} &:= p(y)q(x)F(y) - p(x)q(y)F(x).
 \end{aligned}$$

Here, due to $G(x) \neq G(y)$, we have that $(a_{x,y}, c_{x,y}) \neq (0, 0)$ and $(b_{x,y}, d_{x,y}) \neq (0, 0)$ hold. Substituting $s := f(t)$ (i.e., $t = f^{(-1)}(s)$) in the third equality in (3.7), it follows that

$$(3.8) \quad (c_{x,y}s + d_{x,y})(g \circ f^{(-1)})(s) = a_{x,y}s + b_{x,y}$$

for all $x, y \in X$ with $G(x) < G(y)$ and for all $s \in f(\Theta_{x,y})$.

Next, we check that $c_{x,y}s + d_{x,y} \neq 0$ for all $x, y \in X$ with $G(x) < G(y)$ and for all $s \in f(\Theta_{x,y})$. If $c_{x,y}s + d_{x,y} = 0$ and $c_{x,y} = 0$ were true, then $d_{x,y} = 0$, $a_{x,y} \neq 0$, $b_{x,y} \neq 0$ and $a_{x,y}s + b_{x,y} = 0$ (following from (3.8)). This leads us to a contradiction, since $c_{x,y} = d_{x,y} = 0$ implies that $p(x)q(y) = p(y)q(x)$ and $F(x) = F(y)$, which cannot happen due to $F(x) < F(y)$. If $c_{x,y}s + d_{x,y} = 0$ and $c_{x,y} \neq 0$ were true, then $s = -\frac{d_{x,y}}{c_{x,y}}$ and $a_{x,y}s + b_{x,y} = 0$, yielding that $c_{x,y}b_{x,y} - d_{x,y}a_{x,y} = 0$. This leads us to a contradiction, since an easy calculation shows that

$$(3.9) \quad c_{x,y}b_{x,y} - d_{x,y}a_{x,y} = p(x)p(y)q(x)q(y)(F(x) - F(y))(G(y) - G(x)),$$

which cannot be 0 for any $x, y \in X$ with $G(x) < G(y)$.

Consequently,

$$(3.10) \quad (g \circ f^{(-1)})(s) = \frac{a_{x,y}s + b_{x,y}}{c_{x,y}s + d_{x,y}}$$

for all $x, y \in X$ with $G(x) < G(y)$ and for all $s \in f(\Theta_{x,y})$.

We can apply Lemma 3.1 to the function $h := g \circ f^{(-1)}$ defined on the nondegenerate open interval $I := f(\Theta)$, since (3.10) implies that h is three times differentiable on I such that h' does not vanish on I . Indeed, using (3.10), we have that

$$h'(s) = \frac{a_{x,y}d_{x,y} - b_{x,y}c_{x,y}}{(c_{x,y}s + d_{x,y})^2} \neq 0, \quad s \in f(\Theta_{x,y})$$

for all $x, y \in X$ with $G(x) < G(y)$, where we used (3.9). Hence, as a consequence of (3.5), we have $h'(s) \neq 0$, $s \in f(\Theta)$. Taking into account that $f(\Theta_{x,y})$ is a nondegenerate open interval for all $x, y \in X$ with $G(x) < G(y)$, Lemma 3.1 and (3.10) imply that $S_h(s) = 0$, $s \in f(\Theta)$. Consequently, using again Lemma 3.1, there exist four constants $a^*, b^*, c^*, d^* \in \mathbb{R}$ with $a^*d^* \neq b^*c^*$ and $0 \notin c^*f(\Theta) + d^*$ such that

$$(3.11) \quad h(s) = (g \circ f^{(-1)})(s) = \frac{a^*s + b^*}{c^*s + d^*}, \quad s \in f(\Theta).$$

By substituting $s := f(t)$, where $t \in \Theta$, it follows that

$$g(t) = \frac{a^*f(t) + b^*}{c^*f(t) + d^*}, \quad t \in \Theta,$$

as desired. Using (3.11), the assumptions $f^{(-1)} \circ F = g^{(-1)} \circ G$ and $G(X) \subseteq g(\Theta)$, we get that

$$G(x) = g((f^{(-1)} \circ F)(x)) = (g \circ f^{(-1)})(F(x)) = \frac{a^*F(x) + b^*}{c^*F(x) + d^*}, \quad x \in X,$$

where $0 \notin c^*F(X) + d^*$, since $F(X) = f(\Theta)$ and $0 \notin c^*f(\Theta) + d^*$.

By (3.6) and taking into account the forms of G and g , we get

$$\frac{q(y)}{q(x)} \cdot \frac{\frac{a^*F(y) + b^*}{c^*F(y) + d^*} - \frac{a^*f(t) + b^*}{c^*f(t) + d^*}}{\frac{a^*F(x) + b^*}{c^*F(x) + d^*} - \frac{a^*f(t) + b^*}{c^*f(t) + d^*}} = \frac{p(y)}{p(x)} \cdot \frac{F(y) - f(t)}{F(x) - f(t)}$$

holds for all $x, y \in X$ with $G(x) < G(y)$ and for all $t \in \Theta_{x,y}$. Using that $a^*d^* - b^*c^* \neq 0$, after some algebraic calculations, we obtain that

$$\frac{q(y)}{p(y)} = \frac{c^*F(y) + d^*}{c^*F(x) + d^*} \cdot \frac{q(x)}{p(x)}$$

holds for all $x, y \in X$ with $G(x) < G(y)$, or equivalently,

$$\frac{\left(\frac{q}{p}\right)(y)}{c^*F(y) + d^*} = \frac{\left(\frac{q}{p}\right)(x)}{c^*F(x) + d^*}$$

holds for all $x, y \in X$ with $G(x) < G(y)$. Since q/p is positive, it follows that there exists a constant $k \in \mathbb{R} \setminus \{0\}$ such that

$$q(x) = k(c^*F(x) + d^*)p(x), \quad x \in X.$$

The statement of the proposition now holds with the choices $a := ka^*$, $b := kb^*$, $c := kc^*$ and $d := kd^*$. \square

We note that in the proof of Theorem 3.1, the assumption that f is continuous is used for deriving that $f(\Theta)$ is an open interval, which is essential when we apply Lemma 3.1. Note also that in the proof of Theorem 3.1 it turned out that g is continuous as well, however, we did not utilize this property in the proof. Nonetheless, the continuity of g also follows from the result itself, since f is continuous and $g(t) = (af(t) + b)/(cf(t) + d)$, $t \in \Theta$.

Next, we will provide a set of sufficient conditions on f , g , F and G in order that $\vartheta_{n,\psi}(\mathbf{x}) = \vartheta_{n,\varphi}(\mathbf{x})$ hold for each $n \in \mathbb{N}$ and $\mathbf{x} \in X^n$.

Theorem 3.2. *Let $f, g : \Theta \rightarrow \mathbb{R}$ be strictly increasing functions, $p, q : X \rightarrow \mathbb{R}_{++}$, $F : X \rightarrow \text{conv}(f(\Theta))$, and $G : X \rightarrow \text{conv}(g(\Theta))$. Let $\psi : X \times \Theta \rightarrow \mathbb{R}$ and $\varphi : X \times \Theta \rightarrow \mathbb{R}$ be given by (2.5). If there exist four constants $a, b, c, d \in \mathbb{R}$ with $0 \notin cf(\Theta) + d$ such that (3.2) holds, then $\vartheta_{n,\psi}(\mathbf{x}) = \vartheta_{n,\varphi}(\mathbf{x})$ holds for each $n \in \mathbb{N}$ and $\mathbf{x} \in X^n$.*

Proof: Since p and q are strictly positive functions, as a consequence of the equality $q = (cF + d)p$, we get that $cF + d$ is positive on X . Further, for all $x \in X$ and $t \in \Theta$, we obtain

$$\begin{aligned} (3.12) \quad \varphi(x, t) &= q(x)(G(x) - g(t)) = (cF(x) + d)p(x) \left(\frac{aF(x) + b}{cF(x) + d} - \frac{af(t) + b}{cf(t) + d} \right) \\ &= p(x) \frac{(ad - bc)(F(x) - f(t))}{cf(t) + d} = \frac{ad - bc}{cf(t) + d} \psi(x, t). \end{aligned}$$

Using Lemma 3.2, we have that $ad > bc$, and $cf + d$ is either everywhere positive or everywhere negative on Θ . We show that the latter property cannot hold. To the contrary, assume that $cf + d$ is everywhere negative on Θ , i.e., $cf(\Theta) + d \subseteq \mathbb{R}_{--}$. Then $c \cdot \text{conv}(f(\Theta)) + d = \text{conv}(cf(\Theta) + d) \subseteq \mathbb{R}_{--}$. Using that $F(X) \subseteq \text{conv}(f(\Theta))$, this implies that

$$c \cdot F(X) + d \subseteq c \cdot \text{conv}(f(\Theta)) + d \subseteq \mathbb{R}_{--},$$

which contradicts the positivity of $cF + d$ on X . Consequently, $cf + d$ must be everywhere positive on Θ .

To prove the equality $\vartheta_{n,\psi} = \vartheta_{n,\varphi}$ on X^n , let $n \in \mathbb{N}$ and $\mathbf{x} = (x_1, \dots, x_n) \in X^n$ be arbitrary. Then, by (3.12), we have

$$\sum_{i=1}^n \varphi(x_i, t) = \frac{ad - bc}{cf(t) + d} \sum_{i=1}^n \psi(x_i, t), \quad t \in \Theta.$$

Since $(ad - bc)/(cf + d)$ is positive everywhere on Θ . This implies that

$$\text{sign} \left(\sum_{i=1}^n \varphi(x_i, t) \right) = \text{sign} \left(\sum_{i=1}^n \psi(x_i, t) \right), \quad t \in \Theta.$$

Hence the unique points of sign change of the functions

$$\Theta \ni t \mapsto \sum_{i=1}^n \varphi(x_i, t) \quad \text{and} \quad \Theta \ni t \mapsto \sum_{i=1}^n \psi(x_i, t)$$

are equal to each other, which implies the equality $\vartheta_{n,\psi}(\mathbf{x}) = \vartheta_{n,\varphi}(\mathbf{x})$, as desired. \square

Next, we give an equivalent form of the first equality in (3.2). Roughly speaking, we derive a necessary and sufficient condition in order that two strictly increasing functions defined on a nondegenerate open interval be the Möbius transforms of each other.

Proposition 3.1. *Let I be a nondegenerate open interval of \mathbb{R} . Let $f, g : I \rightarrow \mathbb{R}$ be strictly increasing functions. The following two statements are equivalent:*

- (i) *There exist four constants $a, b, c, d \in \mathbb{R}$ with $0 \notin cf(I) + d$ and*

$$(3.13) \quad g(t) = \frac{af(t) + b}{cf(t) + d}, \quad t \in I.$$

- (ii) *For all $t_1, t_2, t_3, t_4 \in I$, we have*

$$(3.14) \quad \begin{vmatrix} 1 & 1 & 1 & 1 \\ f(t_1) & f(t_2) & f(t_3) & f(t_4) \\ g(t_1) & g(t_2) & g(t_3) & g(t_4) \\ f(t_1)g(t_1) & f(t_2)g(t_2) & f(t_3)g(t_3) & f(t_4)g(t_4) \end{vmatrix} = 0.$$

Proof: (i) \implies (ii). Let us suppose that there exist four real constants $a, b, c, d \in \mathbb{R}$ such that $0 \notin cf(I) + d$ and (3.13) hold. By Lemma 3.2, we have $ad > bc$. Further, $(cf(t) + d)g(t) = af(t) + b$, $t \in I$, yielding that

$$1 \cdot b + f(t) \cdot a + g(t) \cdot (-d) + f(t)g(t) \cdot (-c) = 0, \quad t \in I.$$

In particular, for all $t_1, t_2, t_3, t_4 \in I$, we have that

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ f(t_1) & f(t_2) & f(t_3) & f(t_4) \\ g(t_1) & g(t_2) & g(t_3) & g(t_4) \\ f(t_1)g(t_1) & f(t_2)g(t_2) & f(t_3)g(t_3) & f(t_4)g(t_4) \end{bmatrix}^\top \cdot \begin{bmatrix} b \\ a \\ -d \\ -c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

As a consequence of the inequality $ad > bc$, we have that $(b, a, -d, -c) \neq (0, 0, 0, 0)$, which shows that $(b, a, -d, -c)$ is a nontrivial solution to the above homogeneous system of linear equations. Hence we obtain that (3.14) must hold for all $t_1, t_2, t_3, t_4 \in I$.

(ii) \implies (i). Let $t_3 < t_4$ be fixed elements of I . By the strict monotonicity of f , the vectors $(1, f(t_3))$ and $(1, f(t_4))$ are linearly independent. Assume first that, for all $t \in I$,

$$(3.15) \quad \begin{vmatrix} 1 & 1 & 1 \\ f(t) & f(t_3) & f(t_4) \\ g(t) & g(t_3) & g(t_4) \end{vmatrix} = 0$$

holds. Expanding the determinant along its first column, for all $t \in I$, we get

$$b + af(t) - dg(t) = 0$$

where

$$a := - \begin{vmatrix} 1 & 1 \\ g(t_3) & g(t_4) \end{vmatrix}, \quad b := \begin{vmatrix} f(t_3) & f(t_4) \\ g(t_3) & g(t_4) \end{vmatrix}, \quad d := - \begin{vmatrix} 1 & 1 \\ f(t_3) & f(t_4) \end{vmatrix} \neq 0.$$

Therefore, (3.13) holds with $c = 0$ and we also have that $0 \notin cf(I) + d = \{d\}$.

Now we consider the case when (3.15) is not valid for all $t \in I$, that is, there exists $t_2 \in I$ such that (3.15) does not hold for $t = t_2$. Then, by (3.14), for all $t \in I$,

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ f(t) & f(t_2) & f(t_3) & f(t_4) \\ g(t) & g(t_2) & g(t_3) & g(t_4) \\ f(t)g(t) & f(t_2)g(t_2) & f(t_3)g(t_3) & f(t_4)g(t_4) \end{vmatrix} = 0.$$

Expanding the determinant along its first column, for all $t \in I$, we get

$$(3.16) \quad b + af(t) - dg(t) - cf(t)g(t) = 0,$$

where

$$\begin{aligned} a &:= - \begin{vmatrix} 1 & 1 & 1 \\ g(t_2) & g(t_3) & g(t_4) \\ f(t_2)g(t_2) & f(t_3)g(t_3) & f(t_4)g(t_4) \end{vmatrix}, \\ b &:= \begin{vmatrix} f(t_2) & f(t_3) & f(t_4) \\ g(t_2) & g(t_3) & g(t_4) \\ f(t_2)g(t_2) & f(t_3)g(t_3) & f(t_4)g(t_4) \end{vmatrix}, \\ c &:= \begin{vmatrix} 1 & 1 & 1 \\ f(t_2) & f(t_3) & f(t_4) \\ g(t_2) & g(t_3) & g(t_4) \end{vmatrix} \neq 0, \\ d &:= - \begin{vmatrix} 1 & 1 & 1 \\ f(t_2) & f(t_3) & f(t_4) \\ f(t_2)g(t_2) & f(t_3)g(t_3) & f(t_4)g(t_4) \end{vmatrix}. \end{aligned}$$

Since $c \neq 0$, we have that $cf + d$ is strictly monotone. We now prove that $cf + d$ does not vanish on I . Assume, on the contrary, that for some $t_1 \in I$, we have that $cf(t_1) + d = 0$. Then, $cf(t) + d \neq 0$ for $t \in I \setminus \{t_0\}$, and, by (3.16), we get that $af(t_1) + b = 0$. This implies that $ad = bc$. Therefore, applying (3.16) for $t \in I \setminus \{t_1\}$, we obtain

$$g(t) = \frac{af(t) + b}{cf(t) + d} = \frac{af(t) + \frac{ad}{c}}{cf(t) + d} = \frac{a}{c},$$

which contradicts the strict monotonicity of g . \square

As a consequence of Theorem 3.1, we can characterize the equality of quasiarithmetic-type ψ -estimators.

Corollary 3.1. *Let $f, g : \Theta \rightarrow \mathbb{R}$ be strictly increasing functions, $F : X \rightarrow \text{conv}(f(\Theta))$, and $G : X \rightarrow \text{conv}(g(\Theta))$. Let $\psi : X \times \Theta \rightarrow \mathbb{R}$ and $\varphi : X \times \Theta \rightarrow \mathbb{R}$ be given by*

$$\psi(x, t) := F(x) - f(t), \quad \varphi(x, t) := G(x) - g(t), \quad x \in X, t \in \Theta.$$

The following two assertions hold:

(i) *If there exist two constants $a, b \in \mathbb{R}$ with $a \neq 0$ such that*

$$(3.17) \quad g(t) = af(t) + b, \quad t \in \Theta, \quad \text{and} \quad G(x) = aF(x) + b, \quad x \in X,$$

then $\vartheta_{n,\psi}(\mathbf{x}) = \vartheta_{n,\varphi}(\mathbf{x})$ holds for each $n \in \mathbb{N}$ and $\mathbf{x} \in X^n$.

(ii) *In addition, suppose that f is continuous, $F(X) = f(\Theta)$ and $\text{conv}(G(X)) \subseteq g(\Theta)$. If $\vartheta_{n,\psi}(\mathbf{x}) = \vartheta_{n,\varphi}(\mathbf{x})$ holds for each $n \in \mathbb{N}$ and $\mathbf{x} \in X^n$, then there exist two constants $a, b \in \mathbb{R}$ with $a \neq 0$ such that (3.17) holds.*

Proof: (i). Let us assume that there exist two constants $a, b \in \mathbb{R}$ with $a \neq 0$ such that (3.17) holds. By choosing $c := 0$, $d := 1$ and $p(x) := q(x) := 1$, $x \in X$, we have $cf(\Theta) + d = 1$, and hence $0 \notin cf(\Theta) + d$. Further, (3.2) is satisfied as well. Consequently, Theorem 3.2 yields that $\vartheta_{n,\psi}(\mathbf{x}) = \vartheta_{n,\varphi}(\mathbf{x})$ for each $n \in \mathbb{N}$ and $\mathbf{x} \in X^n$.

(ii). One can apply Theorem 3.1 with the given functions f, g, F and G and by choosing $p(x) := q(x) := 1$, $x \in X$. Then we obtain that there exist four constants $a, b, c, d \in \mathbb{R}$ with $ad \neq bc$ and $0 \notin cf(\Theta) + d$ such that (3.2) holds, i.e.,

$$\begin{aligned} g(t) &= \frac{af(t) + b}{cf(t) + d}, & t \in \Theta, \\ G(x) &= \frac{aF(x) + b}{cF(x) + d}, & x \in X, \\ q(x) &= (cF(x) + d)p(x), & x \in X. \end{aligned}$$

Since $p = q = 1$, we get $cF(x) + d = 1$, $x \in X$, and hence $G(x) = aF(x) + b$, $x \in X$. Consequently, in order to prove the statement, it is enough to verify that $cf(t) + d = 1$, $t \in \Theta$. We check that $c = 0$ and hence $d = 1$. Since Θ is a nondegenerate open interval of \mathbb{R} and f is strictly increasing and continuous, we have that $f(\Theta)$ is a nondegenerate interval of \mathbb{R} . Hence, using that $F(X) = f(\Theta)$, the range $F(X)$ of F contains at least two distinct elements, and consequently, there exist $x_1, x_2 \in X$ such that $F(x_1) \neq F(x_2)$. Since $cF(x_1) + d = 1$ and $cF(x_2) + d = 1$, we have $c(F(x_1) - F(x_2)) = 0$, yielding that $c = 0$, as desired. \square


ACKNOWLEDGMENTS

Mátyás Barczy was supported by the project TKP2021-NVA-09. Project no. TKP2021-NVA-09 has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NVA funding scheme. Zsolt Páles is supported by the K-134191 NKFIH Grant. We also acknowledge the valuable suggestions from the referee.

REFERENCES

- Barczy, M. and Páles, Zs. (2023). Comparison and equality of generalized ψ -estimators. *Arxiv: 2309.04773*. <https://arxiv.org/abs/2309.04773>
- Barczy, M. and Páles, Zs. (2025). Existence and uniqueness of weighted generalized ψ -estimators. *Lithuanian Mathematical Journal*, 65(1):14–49.
- Grünwald, R. and Páles, Zs. (2020). On the equality problem of generalized Bajraktarević means. *Aequationes Mathematicae*, 94(4):651–677.
- Grünwald, R. and Páles, Zs. (2022). On the invariance of the arithmetic mean with respect to generalized Bajraktarević means. *Acta Mathematica Hungarica*, 166(2):594–613.
- Huber, P.J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101.
- Huber, P.J. (1967). *The behavior of maximum likelihood estimates under nonstandard conditions*. In “Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability” (Berkeley, California, 1965/66), Volume I: Statistics (L.M. Le Cam and J. Neyman, Eds.), University of California Press, Berkeley, 221–233.
- Kosorok, M.R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York.
- Mukhopadhyay, S., Das, A.J., Basu, A., Chatterjee, A., and Bhattacharya, S. (2021). Does the generalized mean have the potential to control outliers? *Communications in Statistics – Theory and Methods*, 50(8):1709–1727.
- van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

A Robust Variable Screening Approach with Application to Gene Expression Data

Author: MOHAMMAD KAZEMI 
– Department of Statistics, Faculty of Mathematical Sciences,
University of Guilan,
Rasht, Iran
m.kazemi@guilan.ac.ir, m.kazemie64@yahoo.com

Received: April 2023

Revised: February 2024

Accepted: February 2024

Abstract:

- The presence of outlier observations may lead to misleading results in variable screening problems. To address this issue, this paper presents a new robust variable screening method using the L_1 loss and the Huber loss. As an extension, we also develop an effective iterative procedure to improve the finite sample performance of the presented method. The effectiveness of the proposed methods is illustrated through simulation studies and real data analysis to show their capabilities. Numerical studies show that the proposed methods work well with ultrahigh-dimensional data sets, which may contain outliers, and perform better than some competing methods.

Keywords:

- *gene selection; independence screening; NP dimensionality; outliers; sparsity.*

AMS Subject Classification:

- 62F07, 62F35.

1. INTRODUCTION

Recent research in statistical science has focused on developing effective and useful techniques for analyzing gene expression data. In such ultrahigh-dimensional data, the number of genes is usually in the order of thousands or millions and exponentially larger than the available cases or subjects. For instance, thousands of gene expression profiles can be used in disease classification; millions of single-nucleotide polymorphisms are available for genome-wide association studies between genotypes and phenotypes. If we are to relate these ultrahigh dimensional genes to a response variable in a regression set-up, we need to perform variable selection.

A relevant family of methods for prediction of the response based on the high dimensional gene expression data are penalized linear regression models. Consider the linear regression model with response variable Y and p explanatory variables (e.g. gene expressions) as predictors. Given the responses y_1, \dots, y_n from n independent samples and the corresponding predictor values, say $x_{ij}, i = 1, \dots, n$ for the j -th covariate for $j = 1, \dots, p$, this model can be written in matrix form as

$$(1.1) \quad y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i,$$

where $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ and ε_i s are independent following $N(0, \sigma^2)$ for $i = 1, \dots, n$. The model parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$ and σ^2 need to be estimated from the data. In the ultrahigh dimensional case with $p \gg n$, we need to assume sparsity of the regression coefficient $\boldsymbol{\beta}$ to achieve identifiability of the estimators, i.e., we assume that only a few of the components of $\boldsymbol{\beta}$ are non-zero. Under the sparsity assumption, estimation of the model parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$ is performed through penalized estimation procedures with appropriate penalties which can successfully recover all and only the truly important variables asymptotically with probability tending to one. There are plenty of such penalized regression procedures available in recent literature, starting from the LASSO (Tibshirani, 1996) and its refinements (Zhang and Huang, 2008; Zou, 2006) to more advanced procedures based on penalties like SCAD (Fan and Li, 2001) or MCP (Zhang, 2010) and many more, which work well in moderately high dimensions. However, a common problem with these methods in ultrahigh dimensional set-ups is their computational cost and numerical issues, which has led to development of simpler variable screening methods at the initial stage to reduce the number of genetic predictors from the order of potentially millions to an order of a few hundred (often lesser than the sample size as well) and then apply an appropriate penalization method to obtain final model estimates from the reduced set of covariates. The most popular method for such screening purposes is the Sure Independence Screening (SIS) proposed by Fan and Lv (2008) which has a simple interpretation and theoretical guarantees along with fast computation. Even with its simple structure, the method yet enjoys the model selection oracle property under ultrahigh-dimensional set-ups where $\log(p) = O(n^\zeta)$ for some $0 < \zeta < 1$. An iterative extension, ISIS, is also proposed in Fan and Lv (2008) to tackle the issue of collinearity among covariates. The SIS and ISIS are routinely being applied in ultrahigh dimensional applications and have also been extended to more complex models (Fan *et al.*, 2009, 2011, 2014; Fan and Song, 2010; Kazemi *et al.*, 2018, 2019; Kazemi, 2020, 2024; Li *et al.*, 2012; Zhong and Zhu, 2015; for instance). Each of those proposals focuses on a specific model, and its performance is based upon the belief that the imposed working model is close to the true model. Zhu *et al.* (2011) proposed a sure independent ranking and screening procedure which avoids the specification of a particular model structure.

Motivated by this work, [He et al. \(2013\)](#) proposed a framework called quantile-adaptive model-free screening. However, one major drawback of the SIS or ISIS is their non-robust nature against data contamination as indicated already in the discussion of the original paper itself. This issue can be crucial when applying the method for screening of important genes from large-scale genomic data, which are often prone to at least a few outliers.

In this paper, we develop a new robust screening procedure, in the context of ultrahigh dimensional linear regression, where the number of covariates p may grow exponentially with the sample size n , using robust loss functions such as the L_1 loss and the Huber loss ([Huber, 1964](#)). A robust version of ISIS along the same lines will also be discussed to tackle the correlations among covariates. The suggested methods will be applied to the riboflavin data example.

The plan of the paper is as follows. In [Section 2](#), a brief description of the new screening method is presented. We also introduce an iterative approach to enhance the finite sample performance of the proposed screening procedure. In [Section 3](#), simulation studies are carried out to assess the performance of the suggested approaches, and the riboflavin data set is analyzed.

2. METHODOLOGY

In this section, we develop a marginal utility for robust variable screening based on the L_1 loss and the Huber loss ([Huber, 1964](#)) to reduce the dimensionality.

Suppose that we are interested in exploring the relationship between $\mathbf{X} = (X_1, \dots, X_p)^\top$ and Y . A general robust framework is to minimize an objective function

$$(2.1) \quad Q(\beta_0, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n L(Y_i, \beta_0 + \mathbf{X}_i^\top \boldsymbol{\beta}),$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ is a parameter vector, and $L(\cdot, \cdot)$ is a robust loss function such as the L_1 loss, $L(Y_i, \beta_0 + \mathbf{X}_i^\top \boldsymbol{\beta}) = |Y_i - \beta_0 - \mathbf{X}_i^\top \boldsymbol{\beta}|$ or the Huber loss, defined as

$$(2.2) \quad L_\delta(Y_i, \beta_0 + \mathbf{X}_i^\top \boldsymbol{\beta}) = \begin{cases} \frac{1}{2}(Y_i - \beta_0 - \mathbf{X}_i^\top \boldsymbol{\beta})^2 & \text{if } |Y_i - \beta_0 - \mathbf{X}_i^\top \boldsymbol{\beta}| \leq \delta \\ \delta(|Y_i - \beta_0 - \mathbf{X}_i^\top \boldsymbol{\beta}| - \frac{1}{2}\delta) & \text{if } |Y_i - \beta_0 - \mathbf{X}_i^\top \boldsymbol{\beta}| > \delta \end{cases},$$

where δ is a parameter that controls the robustness level, and a smaller value of δ usually leads to more robust estimation. In implementation, we consider $\delta = 1.345$, the value commonly used in robust regression that produces 95% efficiency for normal errors (see [Huber and Ronchetti, 2009](#)). Here, $L(\cdot, \cdot)$ can be regarded as the loss of using $\beta_0 + \mathbf{X}_i^\top \boldsymbol{\beta}$ to predict Y_i .

We consider the problem of robust variable screening in ultrahigh dimensional data. The goal is to rapidly reduce the number of the predictors from p to a moderate scale via a computationally convenient procedure. Consider the marginal utility of the j -th predictor as

$$(2.3) \quad L_j = \min_{\beta_0, \beta_j} \frac{1}{n} \sum_{i=1}^n L(Y_i, \beta_0 + X_{ij}\beta_j), \quad j = 1, \dots, p,$$

which minimizes the loss function. The idea of SIS in this framework is to compute the vector of marginal utilities $\mathbf{L} = (L_1, \dots, L_p)^\top$ and rank the predictors according to the marginal utilities: the smaller the more important. Note that in order to compute L_j , we need only fit a model with two parameters, β_0 and β_j , so computing the vector \mathbf{L} can be done very quickly and stably, even for an ultrahigh dimensional problem.

The predictor X_j is selected by SIS if L_j is one of the d smallest components of \mathbf{L} . Fan and Lv (2008) suggested setting $d = \lceil n / \log(n) \rceil$, where $\lceil a \rceil$ refers to the integer part of a . Further, Zhu et al. (2011) proposed a combination of hard and soft thresholding strategies to obtain the cutoff point that separates the active and inactive predictors. We refer to the screening procedures described above as LAD-SIS and Huber-SIS, corresponding to the L_1 and the Huber loss functions, which means that the LAD loss and the Huber loss are applied to screen the truly important predictors.

Suppose that d predictors are selected in the screening step. Now we further knock out unimportant predictors among them using a more refined penalized method, as we now describe. By reordering the predictors if necessary, we may assume without loss of generality that X_1, \dots, X_d are the variables recruited by screening. We let $\mathbf{X}_{i,d} = (X_{i1}, \dots, X_{id})^\top$ and redefine $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_d)^\top$. In the penalized robust approach, we seek to minimize

$$(2.4) \quad \min_{\beta_0, \beta_j} \frac{1}{n} \sum_{i=1}^n L(Y_i, \beta_0 + \mathbf{X}_{i,d}^\top \boldsymbol{\beta}) + \lambda \sum_{j=1}^d |\beta_j|,$$

where $\lambda > 0$ is a regularization parameter, which may be chosen by five-fold cross-validation, for example. This two-stage method is summarized in Algorithm 1.

Algorithm 1 Robust SIS

Input: Data matrix \mathbf{X} , vector of responses \mathbf{Y} , model size d

Steps:

- **Step 1**(*screening*): For each $j = 1, \dots, p$, compute the marginal utility of the j -th predictor as

$$L_j = \min_{\beta_0, \beta_j} \frac{1}{n} \sum_{i=1}^n L(Y_i, \beta_0 + X_{ij}\beta_j),$$

where $L(\cdot, \cdot)$ is either the L_1 loss or the Huber loss. The screened sub-model $\hat{\mathcal{M}}_d$ is the indices of the d smallest entries of the marginal utilities $\mathbf{L} = (L_1, \dots, L_p)$.

- **Step 2**(*post-screening variable selection*): Apply a robust penalized regression model, either LAD-LASSO or Huber-LASSO, to the screened variables $\mathbf{X}_{\hat{\mathcal{M}}_d} = \{X_j : j \in \hat{\mathcal{M}}_d\}$ to obtain an estimated coefficient vector, say $\hat{\boldsymbol{\beta}}_d = (\hat{\beta}_{d0}, \hat{\beta}_{dr_1}, \dots, \hat{\beta}_{dr_d})^\top$.

Output: The final estimated model $\hat{\mathcal{M}} = \{1 \leq k \leq d, \hat{\beta}_{dr_k} \neq 0\}$ along with the parameter estimates $\hat{\boldsymbol{\beta}}_d$.

The key idea of Algorithm 1 is to use different marginal utilities to screen predictor variables. As such, it can suffer from the same potential issues as the usual SIS. First, some unimportant predictors that are highly correlated with the important ones can have higher

marginal utilities and thus priority to be selected than other important ones that are relatively weakly related to the response. Second, some important predictors that are jointly correlated but marginally uncorrelated with the response can be missed after screening. Such cases occur mostly due to strong correlation between the important and unimportant predictors. To address these issues, we next briefly discuss an iterative extension of Algorithm 1 that enables us to exploit more fully the joint information among the predictors.

Our iterative extension is motivated by the idea of two-scale learning with the iterative SIS (ISIS) in Fan and Lv (2008) and Fan *et al.* (2009). It works as follows by applying large-scale screening and moderate-scale selection in an iterative fashion. First, apply LAD-SIS (or, Huber-SIS) to the original sample $(\mathbf{X}_i, Y_i)_{i=1}^n$ to select k_1 variables with index set $\mathcal{A}^{(0)} = \{i_1, \dots, i_{k_1}\}$, and then employ the LAD-LASSO (or, Huber-LASSO) to obtain a subset $\hat{\mathcal{M}}^{(0)}$ of these indices. In the second step, we compute the residuals from the fitted regression model of the response \mathbf{Y} on the selected predictors in $\hat{\mathcal{M}}^{(0)}$. The LAD-SIS (or, Huber-SIS) screening is again applied taking these residuals as our new response to select another k_2 predictors from the pool of predictors with index set $\hat{\mathcal{M}}_c^{(0)} = \{1, \dots, p\} \setminus \hat{\mathcal{M}}^{(0)}$; let us denote the index set of these k_2 selected predictors as $\mathcal{A}^{(1)}$. Then apply the LAD-LASSO (or, Huber LASSO) to predictors with indices in $\mathcal{I} = \hat{\mathcal{M}}^{(0)} \cup \mathcal{A}^{(1)}$ to obtain a set $\hat{\mathcal{M}}^{(1)}$ of active indices.

Algorithm 2 Robust ISIS

Input: Data matrix \mathbf{X} , vector of responses \mathbf{Y} , model size d .

Steps:

1. Apply LAD-SIS (or, Huber-SIS) to pick a set $\mathcal{A}^{(0)}$ of indices of size $k_1 = \lceil 2d/3 \rceil$, and then employ the LAD-LASSO regression (or, Huber-LASSO) to select a subset $\hat{\mathcal{M}}^{(0)}$ of these indices.
2. Set $t = 1$.
3. For each $j \in \hat{\mathcal{M}}_c^{(t-1)} = \{1, \dots, p\} \setminus \hat{\mathcal{M}}^{(t-1)}$, compute

$$L_j^{(2)} = \min_{\beta_0, \beta_j} \frac{1}{n} \sum_{i=1}^n L(r_i^{(t-1)}, \beta_0 + \beta_j X_{ij}),$$

where $r_i^{(t-1)} = Y_i - \mathbf{x}_{i, \hat{\mathcal{M}}^{(t-1)}}^\top \hat{\boldsymbol{\beta}}_{\hat{\mathcal{M}}^{(t-1)}}$ is the residual from the previous step of fitting and $\mathbf{x}_{i, \hat{\mathcal{M}}^{(t-1)}}$ is the sub-vector of \mathbf{x}_i consisting of those elements in $\hat{\mathcal{M}}^{(t-1)}$. After ordering $\{L_j^{(2)} : j \in \hat{\mathcal{M}}_c^{(t-1)}\}$, we form the set $\mathcal{A}^{(t)}$ consisting of the indices corresponding to the smallest $k_2^{(t)} = d - |\hat{\mathcal{M}}^{(t-1)}|$ elements.

4. Apply the LAD-LASSO (or, Huber-LASSO) regression to variables with indices in $\mathcal{I} = \hat{\mathcal{M}}^{(t-1)} \cup \mathcal{A}^{(t)}$ to obtain the indices of the coefficients $\hat{\boldsymbol{\beta}}_d = (\hat{\beta}_{d0}, \hat{\beta}_{dr_1}, \dots, \hat{\beta}_{dr_d})^\top$ that are non-zero yield a new estimated set, $\hat{\mathcal{M}}^{(t)}$, of active indices.
5. If obtained a set of indices $\hat{\mathcal{M}}^{(t)}$, which either has reached the size d , or satisfies $\hat{\mathcal{M}}^{(t)} = \hat{\mathcal{M}}^{(t-1)}$, break and go to output stage. Otherwise go to Step 6.
6. Change t to $t + 1$ and go to Step 3.

Output: The final estimated model $\hat{\mathcal{M}} = \{1 \leq k \leq d, \hat{\beta}_{dr_k} \neq 0\}$ along with the parameter estimates $\hat{\boldsymbol{\beta}}_d$.

We further proceed repeating these steps to generate the index sets $\hat{\mathcal{M}}^{(1)}, \dots, \hat{\mathcal{M}}^{(l)}$ of selected predictors in the subsequent stages till we reach our target model size, say d , i.e., till the smallest l for which $|\hat{\mathcal{M}}^{(l)}| = d$. Considering its similarity with the ISIS, we refer to this robust iterative variable screening procedure as LAD-ISIS or Huber-ISIS, corresponding to the L_1 and the Huber loss functions, which is presented schematically in Algorithm 2.

Note that in Step 3 of Algorithm 2, $L_j^{(2)}$ can be interpreted as the additional contribution of predictor X_j given the existence of predictors in $\hat{\mathcal{M}}^{(t-1)}$. In our implementation, we chose $k_1 = \lceil 2d/3 \rceil$, and thereafter at the t -th iteration, we took $k_2^{(t)} = d - |\hat{\mathcal{M}}^{(t-1)}|$. This ensures that the iterated versions of the Robust SIS method takes at least two iterations to terminate.

3. NUMERICAL STUDIES

In this section, we consider some numerical experiments to illustrate the usefulness of the suggested methods in the linear regression model. We first analyze three simulated data sets for more illustrative purposes. Then, we analyze the performance of the proposed screening procedures in a real-world example related to the riboflavin production.

3.1. Monte-Carlo simulation

In this subsection, three simulation examples including different models with various scenarios have been conducted to assess the finite sample performance of the proposed methods. The first example is allocated to our proposed non iterative independence screening procedures, while in the second example, the aim is to examine the influence of the percentage of outliers as well as the impact of sample size on the performance of the proposed methods. In the third example, the iterative ISIS procedure is applied to improve the proposed SIS methods in the situation where SIS fails. The performance of the robust alternatives are compared with the existing competitors, such as DC-SIS (Li *et al.*, 2012), SIRS (Zhu *et al.*, 2011), (I)SIS (Fan and Lv, 2008) and NIS (Fan *et al.*, 2011).

To evaluate the performance of the proposed methods, three criteria are considered. The first criterion is the minimum model size (denoted by M), that is the smallest number of predictors needed to ensure that all the important predictors are selected. To get better inference, the 5%, 25%, 50%, 75% and 95% quantiles of M out of 500 replications were also presented. The second criterion (denoted by P_j) is the empirical probability that the important predictor X_j is selected, when the threshold $d = 2\lceil n/\log(n) \rceil$ is adopted. The last criterion is the proportion (denoted by S) that all important predictors are selected for a given model size in 500 replications, when the threshold $d = 2\lceil n/\log(n) \rceil$ is adopted.

Note that the first criterion does not need to specify a threshold. The more reliable screening procedure, the closer M value to the number of important predictors and also the closer S and P_j value to 1. In an ideal situation, both S and P_j are equal to one.

Example 3.1. Consider the following linear model:

$$Y = c\beta^\top X + \sigma\varepsilon,$$

where $\beta = (0, 1, 0.8, 0.6, 0.4, 0.2, 0, \dots, 0)^\top$ takes grid values and $\sigma^2 = 6.83$. This model is adapted from [Zhu et al. \(2011\)](#). We varied the constant c to control the signal-to-noise ratio. In this example, X_1, X_2, X_3, X_4, X_5 are important predictors and remaining ones (X_6, \dots, X_p) are not relevant.

We choose $c = 1$ and 2 , with the corresponding $R^2 = 50\%$ and 80% . The vector of covariates $X = (X_1, \dots, X_p)$ was generated from the multivariate normal distribution with mean 0 and the covariance matrix $\Sigma = (\sigma_{ij})_{p \times p}$ with $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.8^{|i-j|}$ for $i \neq j$. We set the sample size $n = 200$ and the total number of predictors $p = 2000$ considering two error ε distributions, $N(0, 1)$ and $t(1)$ (t-student), and then repeat each scenario 500 times. The results are given in [Table 1](#).

Table 1: Five quantiles of minimum model size M , the empirical probability P_j and the proportion of S in [Example 3.1](#).

ε	c	Method	M					P_j					S	
			5%	25%	50%	75%	95%	1	2	3	4	5		
N	1	DC-SIS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		SIS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		NIS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		SIRS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		LAD-SIS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Huber-SIS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	DC-SIS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		SIS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		SIRS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		LAD-SIS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Huber-SIS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
t(1)	1	DC-SIS	5	5	5	5	54.50	0.975	0.995	0.990	0.970	0.940	0.940	0.940
		SIS	5.00	105.50	722.50	1584.75	1944.05	0.450	0.435	0.405	0.310	0.260	0.195	0.195
		NIS	11.95	218.50	846.50	1602.00	1933.85	0.245	0.275	0.265	0.205	0.125	0.095	0.095
		SIRS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		LAD-SIS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Huber-SIS	5	5	5	5	6	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	2	DC-SIS	5	5	5	5	5	0.995	0.995	0.995	0.995	0.990	0.990	0.990
		SIS	5	5	41.50	726.25	1872.90	0.650	0.685	0.640	0.620	0.555	0.485	0.485
		NIS	5	14.75	157	1017.50	1901.20	0.510	0.550	0.510	0.480	0.385	0.335	0.335
		SIRS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		LAD-SIS	5	5	5	5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		Huber-SIS	5	5	5	5	6	1.000	1.000	1.000	1.000	1.000	1.000	1.000

From [Table 1](#), it can be seen that when the random error has a normal distribution, all six screening methods perform quit well, with the proportion S equal to 1. However SIS and NIS break down for the heavy-tailed error distribution $t(1)$, while other methods continue to perform well. As expected, when the error distribution is $t(1)$, the proportions S for the LAD-SIS and Huber-SIS are equal to 1, which support the assertion that the LAD-SIS and Huber-SIS process the sure screening property. In addition, the SIRS also performs very well.

This is because the SIRS is robust to the outliers since it only uses the ranks of the observed response values.

Example 3.2. In this example, we consider the impact of both the percentage of outliers and the sample size on the performance of the proposed methods. To this end, we introduce various outlier percentages to assess how the proposed robust methods perform under these different conditions. We also vary the sample size from 50 to 200 for the fixed dimension p , and generate the predictors in the same way as Example 1. The observations for the response variable are determined by

$$Y = \beta^\top X + \varepsilon,$$

where $\beta = (0, 1, 1, 1, 1, 1, 0, \dots, 0)^\top$ and the noise ε is independent of the predictors, and is generated from standard normal distribution. We have investigated various types of contamination schemes, all of which have produced similar results. Hence, for the sake of brevity, we present the results for one particular contamination scheme where the responses are contaminated by replacing its value y by $y - 30$; This choice is arbitrary but simulates a situation of response contamination that arises quite frequently in practice. The contamination proportion is taken as such as 5%, 10%, and 20%, resulting in mild, moderate and heavy contaminations, respectively. For each simulation set-up, we have applied the proposed screening procedures to select the important predictors. The process is replicated 500 times to report some performance measures including the proportion S, when the model size $d = 2\lceil n/\log(n) \rceil$ is chosen; median of the number of true positives selected (TP); median of the minimum model size required to select all four important covariates (M). The simulation results are summarized in Table 2.

From Table 2, we can draw the following conclusions:

- a) For each values of (n, p) , as the fraction of contamination increases, the performance of all six methods in accurately detecting the true model diminishes.
- b) When the sample size is very small ($n = 50$) and the data are contaminated with 20% outliers, values of zero for both S and T indicate that SIS, NIS and DC-SIS can only select important variables by chance. In contrast, SIRS, LAD-SIS and Huber-SIS exhibit better performance in such scenarios.
- c) For larger samples, as the percentage of outlier data increases up to 20%, it becomes evident that SIS, NIS, and DC-SIS are less effective. Conversely, SIRS, LAD-SIS and Huber-SIS demonstrate highly satisfactory performance in identifying important predictors.
- d) In all scenarios, SIRS, LAD-SIS and Huber-SIS consistently outperform the other methods. Specifically, SIRS and LAD-SIS, followed by Huber-SIS, exhibit the best performance.
- e) For each fixed value of p and fixed percentage of contamination, as expected, an increase in sample size leads to an improvement in the performance of all six methods.

Table 2: Median of the minimum model size (M), median of the true positive (TP) and the proportion of S in Example 3.2, considering the percentage of contamination 5%, 10% and 20%.

p	n	Method	5%			10%			20%		
			M	TP	S	M	TP	S	M	TP	S
1000	50	DC-SIS	10	5	0.730	70	4	0.210	197	1	0.005
		SIS	231.5	2	0.050	552	1	0.005	734	0	0.000
		NIS	337	1	0.000	688	0	0.000	758.5	0	0.000
		SIRS	7	5	0.830	12	5	0.655	40.5	4	0.355
		LAD-SIS	6	5	0.885	10.5	5	0.700	30.5	4	0.400
		Huber-SIS	15	5	0.670	26.5	4	0.470	66	3.5	0.120
	100	DC-SIS	5	5	1.000	9	5	0.915	52.5	4	0.390
		SIS	61	4	0.410	193	3	0.170	396.5	2	0.040
		NIS	145	4	0.190	348.5	2	0.050	544.5	1	0.010
		SIRS	5	5	1.000	5	5	1.000	6	5	0.960
		LAD-SIS	5	5	1.000	5	5	0.995	6	5	0.940
		Huber-SIS	5	5	0.985	6	5	0.975	11.5	5	0.915
	200	DC-SIS	5	5	1.000	5	5	1.000	9	5	0.995
		SIS	7	5	0.945	40	5	0.645	145	4	0.310
		NIS	20	5	0.840	108	4	0.390	302	3	0.120
		SIRS	5	5	1.000	5	5	1.000	5	5	1.000
		LAD-SIS	5	5	1.000	5	5	1.000	5	5	1.000
		Huber-SIS	5	5	1.000	5	5	1.000	5	5	1.000
2000	50	DC-SIS	16	5	0.615	118	3	0.110	361.5	0	0.000
		SIS	343.5	2	0.040	1040.5	0.5	0.000	1361.5	0	0.000
		NIS	577.5	0	0.000	1225.5	0	0.000	1520.5	0	0.000
		SIRS	8	5	0.760	21	5	0.525	71	4	0.315
		LAD-SIS	7	5	0.795	16	5	0.620	58.25	4	0.325
		Huber-SIS	24	5	0.520	41.5	4	0.265	133.5	3	0.055
	100	DC-SIS	5	5	0.980	12	5	0.830	88.5	4	0.160
		SIS	90.5	4	0.320	465	3	0.060	885	1	0.020
		NIS	277.5	3	0.090	700	1	0.005	1195.5	1	0.000
		SIRS	5	5	1.000	5	5	0.995	6	5	0.910
		LAD-SIS	5	5	1.000	5	5	0.995	7	5	0.920
		Huber-SIS	5	5	1.000	5	5	0.975	15	5	0.805
	200	DC-SIS	5	5	1.000	5	5	1.000	11.5	5	0.975
		SIS	8	5	0.920	48	5	0.600	258	4	0.235
		NIS	30	5	0.715	160	4	0.285	499.5	3	0.090
		SIRS	5	5	1.000	5	5	1.000	5	5	1.000
		LAD-SIS	5	5	1.000	5	5	1.000	5	5	1.000
		Huber-SIS	5	5	1.000	5	5	1.000	5	5	0.995

Example 3.3. In this example, we compare the empirical performance of the LAD-ISIS and Huber-ISIS with SIS, ISIS, DC-SIS and SIRS in a linear model with weak signal-to-noise ratio, which has the form of

$$(3.1) \quad Y = 2.5X_1 + 2.5X_2 + 2.5X_3 - 7.5\sqrt{\rho}X_4 + \varepsilon.$$

This model was first considered by [Zhong and Zhu \(2015\)](#). We adopted exactly the same settings as in [Zhong and Zhu \(2015\)](#). In this model, $\mathbf{X} = (X_1, \dots, X_{1000})^\top$, each X_k is generated from a normal distribution with zero mean and unit variance. All X_k s except X_4 are equally correlated with the Pearson correlation coefficient ρ , while X_4 has the Pearson correlation $\sqrt{\rho}$ with all other $p - 1$ predictors. We draw ε independently from $N(0, 1)$ and $t(1)$ (t-student).

We set the sample size $n = 200$ and $d = 2\lceil n/\log n \rceil = 74$ for the LAD-SIS and Huber-SIS procedures with LASSO penalty function. Then, we repeat the simulations 500 times and summarize the results in Table 3.

Table 3: The proportions of P_j and S given the model size $d = 2\lceil n/\log n \rceil$ in Example 3.3.

ρ	Error	$t(1)$					$N(0,1)$				
		P_j				S	P_j				S
		1	2	3	4		1	2	3	4	
0.2	SIS	1.000	1.000	1.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000
	ISIS	1.000	1.000	1.000	0.890	0.890	1.000	1.000	1.000	1.000	1.000
	SIRS	1.000	1.000	1.000	0.025	0.025	1.000	1.000	1.000	0.030	0.030
	DC-SIS	1.000	1.000	1.000	0.040	0.040	1.000	1.000	1.000	0.020	0.020
	LAD-SIS	1.000	1.000	1.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000
	LAD-ISIS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Huber-SIS	1.000	0.990	0.995	0.000	0.000	0.985	0.990	0.995	0.000	0.000
	Huber-ISIS	1.000	1.000	1.000	0.995	0.995	1.000	1.000	1.000	0.995	0.995
0.5	SIS	1.000	0.995	1.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000
	ISIS	1.000	1.000	1.000	0.550	0.550	1.000	1.000	1.000	0.995	0.995
	SIRS	1.000	0.995	1.000	0.000	0.000	1.000	0.990	1.000	0.010	0.010
	DC-SIS	1.000	1.000	0.995	0.000	0.000	1.000	0.980	0.990	0.010	0.010
	LAD-SIS	0.985	0.965	0.980	0.000	0.000	0.995	0.990	0.995	0.000	0.000
	LAD-ISIS	1.000	1.000	1.000	0.990	0.990	1.000	1.000	1.000	1.000	1.000
	Huber-SIS	0.925	0.945	0.950	0.000	0.000	0.945	0.955	0.960	0.000	0.000
	Huber-ISIS	0.995	1.000	1.000	0.970	0.965	0.995	1.000	0.995	0.985	0.975
0.8	SIS	0.855	0.845	0.845	0.000	0.000	0.940	0.915	0.935	0.000	0.000
	ISIS	1.000	1.000	1.000	0.625	0.625	0.995	0.995	0.995	0.245	0.245
	SIRS	0.910	0.910	0.920	0.000	0.000	1.000	0.990	1.000	0.010	0.010
	DC-SIS	0.890	0.900	0.925	0.000	0.000	1.000	0.980	0.990	0.010	0.010
	LAD-SIS	0.790	0.770	0.805	0.000	0.000	0.845	0.845	0.860	0.000	0.000
	LAD-ISIS	0.995	0.990	0.995	0.930	0.910	0.995	1.000	1.000	0.935	0.930
	Huber-SIS	0.730	0.715	0.730	0.000	0.000	0.790	0.785	0.780	0.000	0.000
	Huber-ISIS	1.000	0.990	1.000	0.930	0.920	0.995	1.000	0.990	0.910	0.895

In this example, X_4 is jointly important but marginally independent to the response Y , so the marginal screening methods (SIS, DC-SIS, SIRS, LAD-SIS and Huber-SIS) can work badly and hardly detect important predictor X_4 . According to Table 3, when the error distribution is normal and correlations among predictors are not strong, i.e. $\rho = 0.2, 0.5$, ISIS selects X_4 with high empirical probability; in other scenarios, ISIS does not perform well to detect the marginal signal of X_4 . In contrast, the proposed LAD-ISIS and Huber-ISIS are able to select X_4 effectively for all different cases in both error distributions. For example, when $\rho = 0.5$ and error distribution is $t(1)$, LAD-ISIS and Huber-ISIS can select all truly important predictors in the model with the empirical probability 99% and 96.5%, respectively, while the ISIS only has 55%. Similarly, when $\rho = 0.8$, the LAD-ISIS and Huber-ISIS can detect all truly important predictors with the empirical probability 91% and 92%, respectively, while the ISIS only has 62.5%. Thus, due to model misspecification, ISIS can not perform as well as in the Example 3.1. However, our LAD-ISIS and Huber-ISIS are still able to identify all important predictors with an overwhelming probability. These once again confirm the capabilities of LAD-ISIS and Huber-ISIS.

3.2. Application to riboflavin production data set

To support our assertions, we consider the data set about riboflavin (vitamin B2) production in *Bacillus subtilis* (Lee *et al.*, 2001; Zamboni *et al.*, 2005), which can be found in R package “hdi”. There is a single real valued response variable which is the logarithm of the riboflavin production rate and $p = 4088$ explanatory variables measuring the logarithm of the expression level of 4088 genes. There is one rather homogeneous data set from $n = 71$ samples that were hybridized repeatedly during a fed batch fermentation process where different engineered strains and strains grown under different fermentation conditions were analyzed.

Figure 1 shows the normal Q-Q plot based on the LASSO regression for the riboflavin production data set. Also, the bivariate boxplot for some of the selected genes of this data is depicted in Figure 2. The bivariate boxplot is a two-dimensional analogue of the boxplot for univariate data. This diagram is based on calculating robust measures of location, scale, and correlation; it consists essentially a pair of concentric ellipses, one of which (the hinge) includes 50% of the data and the other (called the fence) delineates potentially troublesome outliers. In addition, robust regression lines of both response on predictor and vice versa are shown, with their intersection showing the bivariate location estimator. The acute (large) angle between the regression lines will be small (large) for a large (small) absolute value of correlations. Figures 1 and 2 clearly reveal that the data contains some outliers. Now, if we are screening the genes via correlation with response in SIS or ISIS, these outliers will have an erroneous effects.

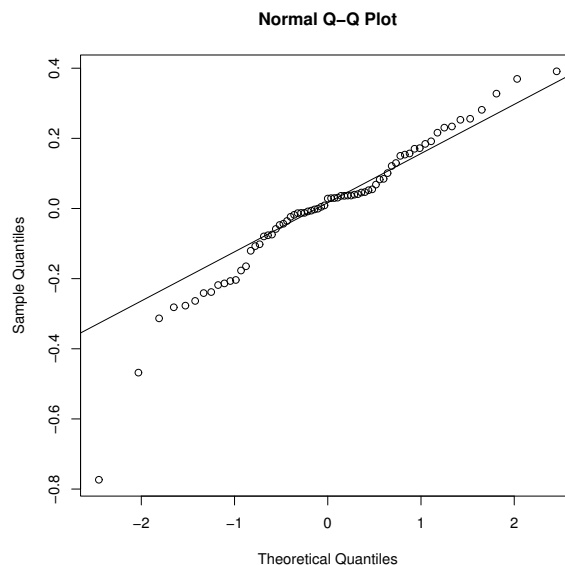


Figure 1: Q-Q plot based on the LASSO regression for the riboflavin production data set.

To analyze this data set, we first apply LAD-(I)SIS and Huber-(I)SIS robust procedures to shrink the dimension from $p = 4088$ down to $d = 2\lceil n/\log(n) \rceil = 32$ genes. After the variable screening, we fit the data by the penalized robust methods such as LAD-LASSO and Huber-LASSO models with the selected genes. On the other hand, LASSO is applied directly to $p = 4088$ genes without screening procedure. For the purpose of comparison, we also implement the (I)SIS with LASSO penalty to select most relevant genes. We compared their performance

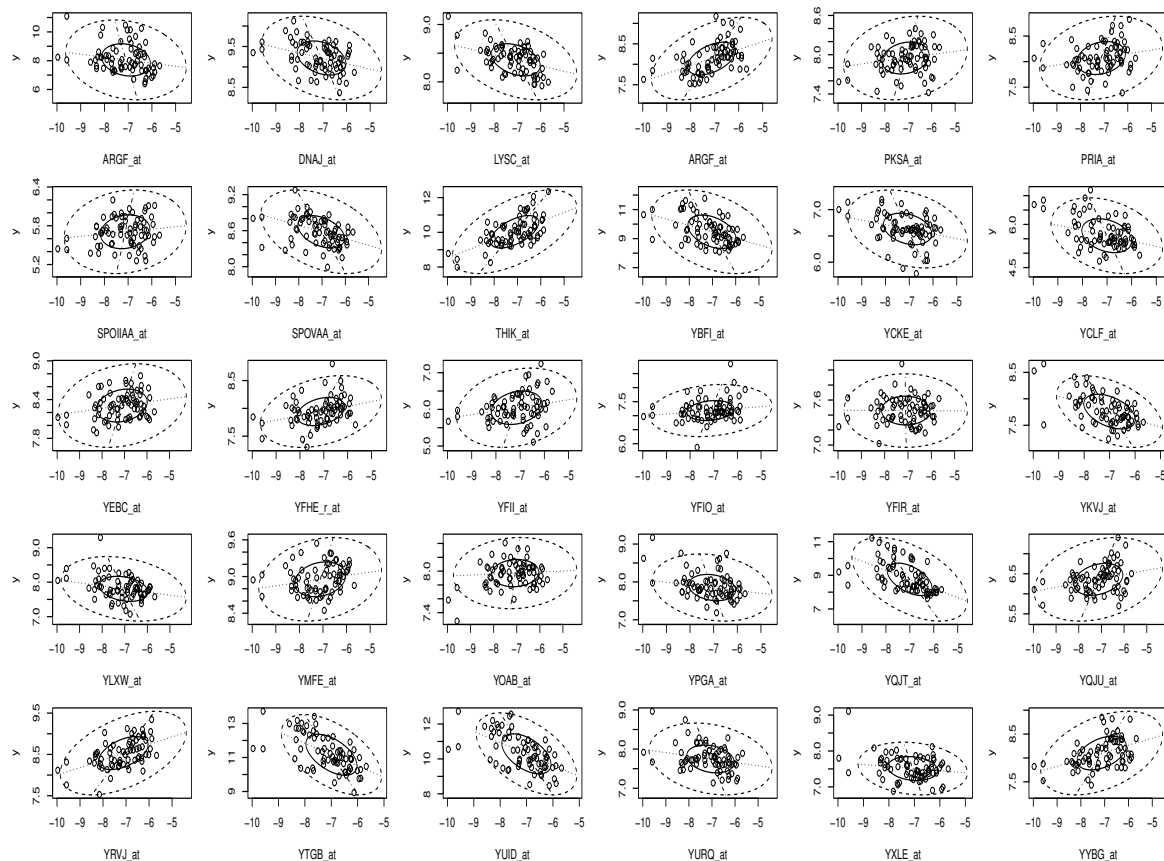


Figure 2: Bivariate boxplot of the riboflavin production data set for some of the effective genes.

in terms of the adjusted R^2 which is defined as $R^2_{adj} = 1 - [(n - 1)/(n - k - 1)](1 - R^2)$, where k is the number of predictors in the model, excluding the intercept, and R^2 is coefficient of determination defined as $R^2 = 1 - \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 / \sum_{i=1}^n (Y_i - \bar{Y})^2$. The results are displayed in Table 4, in which the column labelled “Genes” stands for the number of the genes selected and the column of “ R^2_{adj} ” for the adjusted R^2 .

Table 4: Riboflavin production data analysis results.

Model	Genes	R^2_{adj}	CV	Corr
LASSO	42	90.45%	0.3973	0.730
SIS	11	76.90%	0.4461	0.687
ISIS	13	78.83%	0.3566	0.760
LAD-SIS	22	73.12%	0.3004	0.800
LAD-ISIS	21	80.06%	0.2309	0.852
Huber-SIS	20	73.99%	0.3362	0.779
Huber-ISIS	22	78.63%	0.3175	0.799

Next, to measure the prediction accuracy of proposed estimators, the leave-one-out cross-validation (CV) criterion was used, which is defined by

$$(3.2) \quad CV = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2,$$

where \hat{Y}_i is the predicted value of response variable where i -th observation left out of the estimation of the parameters β . We also compute the correlation between the true and predicted response values \hat{Y}_i obtained from leave-one-out cross-validation.

From the values of R_{adj}^2 , we can infer that although LASSO model can explain 90.45% of the total variation of the logarithm of the riboflavin production rate, it does not perform as well as the screening competitors, because the two-step screening procedures use much fewer genes while giving smaller CV error and higher correlation between the true and predicted values of response variable.

In addition, it can be seen that the iterative screening procedures (ISIS, LAD-ISIS and Huber-ISIS) outperform the corresponding noniterative screening methods (SIS, LAD-SIS and Huber-SIS) with the larger R_{adj}^2 and $\text{Corr}(Y, \hat{Y})$, and smaller CV error, indicating that the iterative procedures identify some genes missed by the noniterative screening methods. Those important genes missed by the SIS (or, LAD-SIS and Huber-SIS) either may be closely marginally independent of the response or have relatively weaker marginal signals than some unimportant genes which are highly correlated with some strong active genes.

Moreover, the number of genes selected by SIS and ISIS are fewer than those of the robust alternatives, while because of the existence of outliers in the data set, it can be seen that CV error and R_{adj}^2 of the robust type methods are more acceptable than those of the non-robust type screening procedures.

In sum, we can clearly conclude from Table 4 that the LAD-ISIS performs the best with the smallest CV error 0.2309, the largest $\text{Corr}(Y, \hat{Y})$ value 0.852, the largest adjusted R^2 value 80.06% (compared to the other robust approaches), indicating that the LAD-ISIS is the best method for riboflavin production data set analysis.

CONCLUSIONS

In this paper, we have studied a robust variable screening methodology for ultrahigh dimensional data using robust loss functions. This technique uses the L_1 loss and the Huber loss which we refer to the LAD-SIS and Huber-SIS. We examined the finite sample performance of the proposed procedures via Monte Carlo studies, and illustrated the proposed methodology through the riboflavin production data set. In our numerical studies, both proposed robust methods (LAD-SIS and Huber-SIS) and SIRS perform equally well and behave better than SIS and DC-SIS in presence of outliers.

Similar to the SIS, the proposed technique may fail to identify some important predictors that are marginally independent of the response. Motivated by this, we introduced an iterative robust sure independence screening procedure. We examined its finite-sample performance via intensive simulations. The simulation results indicate that the iterative robust approach can significantly improve the LAD-SIS and Huber-SIS in the presence of truly important predictors that are marginally independent of the response and unimportant predictors that have relatively stronger marginal signals than some important predictors. Our empirical results indicate that the LAD-ISIS is the best approach among the selected competitors. We used only the LASSO penalty but other penalties such as adaptive LASSO (Zou, 2006), SCAD (Fan and Li, 2001) and MCP (Zhang, 2010) could also be applied.

ACKNOWLEDGMENTS

The author is very grateful to the referee and editor for their valuable suggestions which significantly improved the presentation of paper and led to put more details.

REFERENCES

- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(465):1348–1360.
- Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 70(5):849–911.
- Fan, J., Feng, Y., and Song, R. (2011). Nonparametric independence screening in sparse ultrahigh-dimensional additive models. *Journal of the American Statistical Association*, 106(494):544–557.
- Fan, J., Ma, Y.B., and Dai, W. (2014). Nonparametric independence screening in sparse ultrahigh-dimensional varying coefficient models. *Journal of the American Statistical Association*, 109(507):1270–1284.
- Fan, J., Samworth, R., and Wu, Y. (2009). Ultrahigh dimensional feature selection: beyond the linear model. *The Journal of Machine Learning Research*, 10:2013–2038.
- Fan, J. and Song, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *The Annals of Statistics*, 38(6):3567–3604.
- He, X., Wang, L., and Hong, H. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *The Annals of Statistics*, 41(1):342–369.
- Huber, P. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101.
- Huber, P.J. and Ronchetti, E.M. (2009). *Robust Statistics*. Wiley, Hoboken, NJ, USA.
- Kazemi, M. (2020). Partial correlation screening for varying coefficient models. *Journal of Mathematical Modeling*, 8(4):363–376.
- Kazemi, M. (2024). Support vector machine in ultrahigh dimensional feature space. *Journal of Statistical Computation and Simulation*, 94(3):517–535.
- Kazemi, M., Shahsavani, D., and Arashi, M. (2018). Variable selection and structure identification for ultrahigh-dimensional partially linear additive models with application to cardiomyopathy microarray data. *Statistics, Optimization and Information Computing*, 6(3):373–382.
- Kazemi, M., Shahsavani, D., and Arashi, M. (2019). A sure independence screening procedure for ultra-high dimensional partially linear additive models. *Journal of Applied Statistics*, 46(8):1385–1403.
- Lee, J.M., Zhang, S., Saha, S., Anna, S.S., Jiang, C., and Perkins, J. (2001). RNA expression analysis using an antisense *Bacillus subtilis* genome array. *Journal of Bacteriology*, 183(24):7371–7380.
- Li, R.Z., Zhong, W., and Zhu, L.P. (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association*, 107(499):1129–1139.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 58(1):267–288.

- Zamboni, N., Fischer, E., Muffler, A., Wyss, M., Hohmann, HP., and Sauer, U. (2005). Transient expression and flux changes during a shift from high to low riboflavin production in continuous cultures of *Bacillus subtilis*. *Biotechnology and Bioengineering*, 89(2):219–232.
- Zhang, C.H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of statistics*, 38(2):894–942.
- Zhang, CH. and Huang, J. (2008). The sparsity and bias of the Lasso selection in high-dimensional linear regression. *Annals of Statistics*, 36(4):1567–1594.
- Zhong, W. and Zhu, L. (2015). An iterative approach to distance correlation-based sure independence screening. *Journal of Statistical Computation and Simulation*, 85(11):2331–2345.
- Zhu, L.P., Li, L., Li, R., and Zhu, L.X. (2011). Model-free feature screening for ultrahigh-dimensional data. *Journal of the American Statistical Association*, 106(496):1464–1475.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.

E-value Formulae for the Odds Ratio

Author: KALLE LEPPÄLÄ 
– The Organismal and Evolutionary Biology Research Programme,
University of Helsinki,
Finland
kalle.m.leppala@gmail.com

Received: December 2023

Revised: April 2024

Accepted: April 2024

Abstract:

- The *E*-value is the minimum strength of association that a potential confounder needs to have with either the exposure or the outcome, if it were to alone explain an observed association between the two. Originally defined in terms of the risk ratio (relative risk), this short note explores the *E*-value formulae where one or several of the three relationships is measured using the odds ratio instead. A rigorous but elegant alternative is presented to the square root approximation that has been proposed earlier as a convenient but wrong way to combine odds ratios and *E*-values.

Keywords:

- *E*-value; odds ratio; sensitivity analysis.

AMS Subject Classification:

- 62D20, 92D30.

1. INTRODUCTION

This work concerns two measures of association between binary random variables X and Y , the risk ratio (relative risk) and the odds ratio, respectively defined as

$$\text{RR}_{XY} = \frac{\mathbb{P}(Y|X)}{\mathbb{P}(Y|\bar{X})}, \quad \text{OR}_{XY} = \frac{\mathbb{P}(Y|X)\mathbb{P}(\bar{Y}|\bar{X})}{\mathbb{P}(Y|\bar{X})\mathbb{P}(\bar{Y}|X)}.$$

Let X and Y be independent conditional on a third binary random variable Z , which is not negatively associated with either X or Y :

$$(1.1) \quad (X \perp\!\!\!\perp Y) | Z, \quad \begin{cases} \text{RR}_{XZ} \geq 1, \\ \text{RR}_{ZY} \geq 1. \end{cases}$$

The second assumption is not restrictive as we can replace a variable with its opposite if necessary, and that consequently $\text{RR}_{XY} \geq 1$ as well. In terms of graphical causal models (1.1) means that Z d -separates X and Y , which is the case for example when X and Y have no direct causal relationship, Z is their common cause or a mediator (but not a common consequence), and no other variables play a role.

The classical Cornfield inequalities by [Cornfield *et al.* \(1959\)](#) and [Schlesselman \(1978\)](#) now state that

$$\min\{\text{RR}_{XZ}, \text{RR}_{ZY}\} \geq \text{RR}_{XY}.$$

The causal interpretation is as follows: if and observed association between X and Y was entirely due to a confounder or a mediator Z , then *both* the risk ratio between X and Z and between Z and Y must be at least as big as the risk ratio between X and Y . The first development beyond the classical Cornfield inequalities was by [Lee and Wang \(2008\)](#), who instead of the minimum bounded a quantity called *confounding rate ratio* to assess sensitivity to stratification. A different but natural perspective is to swap the minimum into a maximum. Now *either* the risk ratio between X and Z *or* between Z and Y must be at least the so-called E -value $\text{RR}_{XY} + \sqrt{\text{RR}_{XY}(\text{RR}_{XY} - 1)}$, as formalized in the following theorem of VanderWeele and Ding.

Theorem 1.1 ([VanderWeele and Ding, 2017](#)). *Assuming (1.1),*

$$\max\{\text{RR}_{XZ}, \text{RR}_{ZY}\} \geq \text{RR}_{XY} + \sqrt{\text{RR}_{XY}(\text{RR}_{XY} - 1)}.$$

VanderWeele and Ding also suggest ([VanderWeele and Ding, 2017](#)) that it should become a convention in all observational science to report the E -value or some comparable form of sensitivity analysis. Another E -value formula, although not dubbed as such, has been published in the form of the following theorem by Lee.

Theorem 1.2 ([Lee, 2011](#)). *Assuming (1.1),*

$$\max\{\text{OR}_{XZ}, \text{RR}_{ZY}\} \geq \Omega + \sqrt{(\Omega + 1)(\Omega - 1)},$$

where $\Omega = 1 + 2(\text{RR}_{XY} - 1)$.

This time one of the three relationships is measured by the odds ratio instead of the risk ratio. On that same note, if we want to bound $\max\{\text{RR}_{XZ}, \text{OR}_{ZY}\}$ or $\max\{\text{OR}_{XZ}, \text{OR}_{ZY}\}$ with a function of RR_{XY} we can still use Theorems 1.1 and 1.2, respectively, because $\text{RR}_{ZY} \leq \text{OR}_{ZY}$. The bounds are optimal.

Above two out of the three relationships were allowed to be measured by the odds ratio instead of the risk ratio, while RR_{XY} remained fixed. The purpose of this short note is to explore the cases where the association between X and Y is quantified with OR_{XY} . The odds ratio arises naturally in the context of logistic regression or case-control design, and sensitivity analysis results tailored for odds ratio should be preferred over unnecessary conversions to risk ratios. In a recent paper (Leppälä, 2023), the author published an odds ratio analogue

$$\min\{\text{OR}_{XZ}, \text{OR}_{ZY}\} \geq \text{OR}_{XY}$$

of the classical Cornfield inequalities and completed the collection of joint bounding formulae of the type (4.1) and (4.2) by Ding and VanderWeele (2016) and Lee (2011), respectively. Such joint bounding formulae are easily inverted into E -value formulae. Suppose $t \leq f(r, s)$, where t , r and s are the relevant risk- or odds ratios in question. Provided that f is increasing with respect to both r and s in the domain $r > 1$, $s > 1$, we can bound $t \leq f(\max(r, s), \max(r, s)) = F(\max(r, s))$ and then solve $\max(r, s) \geq F^{-1}(t)$. The author however failed to seize the opportunity to supplement the joint bounding formulae in Leppälä (2023) with the corresponding E -value formulae—this omission is fixed in Section 2. However, some E -value formulae remain impossible, such as bounding of $\max\{\text{RR}_{XZ}, \text{RR}_{ZY}\}$ with a function of OR_{XY} , because not all combinations of risk ratios and odds ratios allow a joint bounding formula (Leppälä, 2023). In this instance VanderWeele recommends (VanderWeele and Ding, 2017) approximating OR_{XY} with either RR_{XY} or RR_{XY}^2 (VanderWeele, 2017) depending on whether the outcome Y is rare or common, respectively. Section 3 proposes a rigorous but elegant alternative to the approximation, in terms of a non-zero parameter α controlling the least and largest of the four elements in the probability mass function of Y given X (incidentally also quantifying how much error the square approximation would introduce). Finally, Section 4 takes a new look at the joint bounding formulae missing from the author’s previous work due to them not existing in general form (Leppälä, 2023), and presents restricted versions of them that use the parameter α . The results of Section 3 could also be derived from these new joint bounding formulae of Section 4 using the inversion procedure described above.

2. E-VALUE FORMULAE FOR THE ODDS RATIO

Theorem 2.1. Assuming (1.1),

$$\max\{\text{OR}_{XZ}, \text{OR}_{ZY}\} \geq \Omega + \sqrt{(\Omega + 1)(\Omega - 1)},$$

where $\Omega = 1 + 2(\text{OR}_{XY} - 1)$.

Proof: The statement is similar to that of Theorem 1.2 but we present a proof nevertheless. Under assumption (1.1) the joint bounding formula

$$\text{OR}_{XY} \leq \left(\frac{\sqrt{\text{OR}_{XZ}\text{OR}_{ZY}} + 1}{\sqrt{\text{OR}_{XZ}} + \sqrt{\text{OR}_{ZY}}} \right)^2$$

holds by another theorem of the author (Leppälä, 2023). Denote $M = \max\{\text{OR}_{XZ}, \text{OR}_{ZY}\}$. Define and differentiate

$$f(r) = \left(\frac{rs + 1}{r + s} \right)^2, \quad f_r(r) = \frac{2(rs + 1)(s^2 - 1)}{(r + s)^3} > 0,$$

where $s > 1$. We therefore get

$$(2.1) \quad \text{OR}_{XY} \leq \left(\frac{M + 1}{2\sqrt{M}} \right)^2 \Leftrightarrow \text{OR}_{XY} - 1 \leq (\sqrt{M} - \sqrt{\text{OR}_{XY}})^2.$$

By the odds ratio versions (Leppälä, 2023) of the Cornfield inequalities we have $M \geq \text{OR}_{XY}$, and thereby also $|\sqrt{M} - \sqrt{\text{OR}_{XY}}| = \sqrt{M} - \sqrt{\text{OR}_{XY}}$. Taking a square root of both sides of (2.1) and solving for \sqrt{M} yields

$$\sqrt{M} \geq \sqrt{\text{OR}_{XY}} + \sqrt{\text{OR}_{XY} - 1},$$

from which the claim follows by squaring. \square

Theorem 2.2. Assuming (1.1),

$$\max\{\text{RR}_{XZ}, \text{OR}_{ZY}\} \geq \Omega + \sqrt{(\Omega + 1)(\Omega - 1)},$$

where $\Omega = 1 + (\text{OR}_{XY} - 1)/2$

Proof: Under assumption (1.1) the joint bounding formula

$$\text{OR}_{XY} \leq \frac{1 + (\text{RR}_{XZ} - 1)\text{OR}_{ZY}}{\text{RR}_{XZ}} = 1 + \frac{(\text{RR}_{XZ} - 1)(\text{OR}_{ZY} - 1)}{\text{RR}_{XZ}}$$

holds by another theorem of the author (Leppälä, 2023). Denote $M = \max\{\text{RR}_{XZ}, \text{OR}_{ZY}\}$. Define and differentiate

$$f(r, s) = \frac{1 + (r - 1)s}{r}, \quad f_r(r, s) = \frac{s(r - 1)}{r^2} > 0, \quad f_s(r, s) = \frac{r - 1}{r} > 0,$$

where $r > 1$. Now

$$(2.2) \quad \text{OR}_{XY} \leq \frac{M^2 - M + 1}{M} \Leftrightarrow (\Omega + 1)(\Omega - 1) \leq (M - \Omega)^2.$$

If it was true that $M < \Omega$, then

$$2M - 1 < 2\Omega - 1 = \text{OR}_{XY} \leq \frac{M^2 - M + 1}{M} \Leftrightarrow M^2 < 1,$$

which is a contradiction. Therefore $|M - \Omega| = M - \Omega$, and the theorem is proved by taking a square root and adding Ω to both sides of (2.2). \square

Due to the symmetry of the odds ratio, Theorem 2.2 can also be used to bound $\max\{\text{OR}_{XZ}, \text{RR}_{YZ}\}$.

3. *E-VALUE FORMULAE UNDER AN EXTRA ASSUMPTION*

As pointed out earlier (Leppälä, 2023), OR_{XY} cannot be bounded from above by a joint function of RR_{XZ} and RR_{ZY} , or of OR_{XZ} and RR_{ZY} . Consequently, neither $\max\{RR_{XZ}, RR_{ZY}\}$ nor $\max\{OR_{XZ}, RR_{ZY}\}$ can be bounded from below by a function of OR_{XY} —joint bounding will be returned to in Section 4. Both the risk ratio and the odds ratio are nevertheless common measures of association, and the desire to do sensitivity analysis doesn't vanish when relationship between X and Y is quantified by the odds ratio. At this point we're aware of the option to use Theorem 2.1, but if one still prefers to measure the associations between X and Z and between Z and Y using the risk ratio, some auxiliary parameter is needed. Assume that the conditional probabilities of Y or \bar{Y} given X or \bar{X} are not close to zero or one, or precisely for some $\alpha > 0$

$$(3.1) \quad \begin{cases} \min\{\mathbb{P}(Y|X), \mathbb{P}(Y|\bar{X}), \mathbb{P}(\bar{Y}|X), \mathbb{P}(\bar{Y}|\bar{X})\} \geq \alpha, \\ \max\{\mathbb{P}(Y|X), \mathbb{P}(Y|\bar{X}), \mathbb{P}(\bar{Y}|X), \mathbb{P}(\bar{Y}|\bar{X})\} \leq 1 - \alpha. \end{cases}$$

Now we can bound the odds ratio with the risk ratio by scaling the portion exceeding one using the parameter α , and as corollaries derive the required *E-value* formulae.

Lemma 3.1. *Assuming (3.1),*

$$OR_{XY} \leq 1 + \frac{RR_{XY} - 1}{\alpha}.$$

Proof: Denote $\mathbb{P}(Y|\bar{X}) = r$ so that

$$\mathbb{P}(Y|X) = RR_{XY}r, \quad 0 < r < \frac{1}{RR_{XY}}.$$

By definition

$$OR_{XY} = \frac{\mathbb{P}(Y|X)\mathbb{P}(\bar{Y}|\bar{X})}{\mathbb{P}(Y|\bar{X})\mathbb{P}(\bar{Y}|X)} = \frac{RR_{XY}(1-r)}{1-RR_{XY}r} = f(r),$$

and differentiating gives

$$f_r(r) = \frac{RR_{XY}(RR_{XY} - 1)}{(1 - RR_{XY}r)^2} > 0.$$

As r approaches $1/RR_{XY}$, the value $f(r)$ tends to infinity, but with the boundaries (3.1) $\mathbb{P}(\bar{Y}|X) = 1 - RR_{XY}r$ actually hits α first at the point $r = (1 - \alpha)/RR_{XY}$. This happens before $\mathbb{P}(\bar{Y}|\bar{X}) = 1 - r$ hits α at $r = 1 - \alpha$. Under (3.1), the odds ratio is therefore bounded from above by $f((1 - \alpha)/RR_{XY})$. \square

Theorem 3.1. *Assuming (1.1) and (3.1),*

$$\max\{RR_{XZ}, RR_{ZY}\} \geq \Omega + \sqrt{\Omega(\Omega - 1)},$$

where $\Omega = 1 + \alpha(OR_{XY} - 1)$.

Proof: Assuming (3.1), $RR_{XY} \geq \Omega$ by Lemma 3.1. Assuming (1.1), we can just plug this to the *E-value* formula of VanderWeele & Ding in Theorem 1.1. \square

Theorem 3.2. Assuming (1.1) and (3.1),

$$\max\{\text{OR}_{XZ}, \text{RR}_{ZY}\} \geq \Omega + \sqrt{(\Omega + 1)(\Omega - 1)},$$

where $\Omega = 1 + 2\alpha(\text{OR}_{XY} - 1)$.

Proof: Assuming (3.1), $\text{RR}_{XY} \geq \Omega$ by Lemma 3.1. Assuming (1.1), we can just plug this to the E -value formula of Lee in Theorem 1.2. \square

VanderWeele and Ding have earlier suggested (VanderWeele and Ding, 2017) that for rare outcomes the original E -value formula of Theorem 1.1 is applicable for the odds ratio as well, and that for common outcomes it becomes applicable after approximating $\sqrt{\text{OR}_{XY}} \approx \text{RR}_{XY}$ (VanderWeele, 2017). The rare–common distinction concerning the outcome Y is somewhat misleading as the error factor introduced by the approximation can be arbitrarily large even if the prevalence of Y is one half, and in fact approximation was mathematically motivated using the assumption (3.1) instead. VanderWeele implicitly derives (VanderWeele, 2017) the upper bound for the introduced error as

$$(3.2) \quad \text{OR}_{XY} \leq \frac{\text{RR}_{XY}^2}{4\alpha(1 - \alpha)},$$

and so it would be more appropriate to plug in $\sqrt{4\alpha(1 - \alpha)\text{OR}_{XY}}$ to Theorem 1.1 than the raw square root approximation. Compared to this strictly correct version of the square root approximation Theorem 3.1 gives a stronger statement, and the new E -value formula can be seen as quite naturally scaling the portion exceeding one. The comparison is visualized in the figure below.

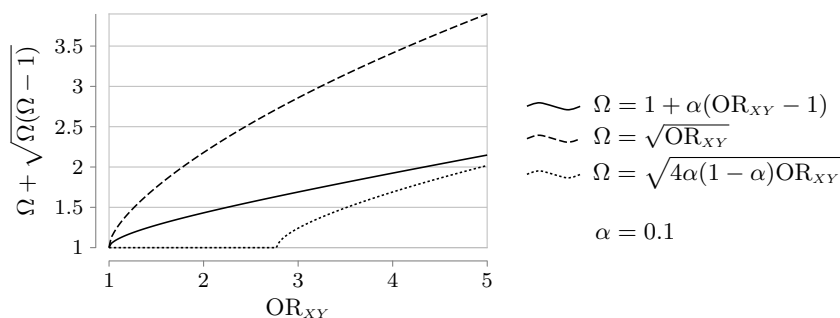


Figure 1: Visualization of different lower bounds of $\max\{\text{RR}_{XZ}, \text{RR}_{ZY}\}$. The solid line is Theorem 3.1. The dashed line is the currently recommended (VanderWeele and Ding, 2017) raw square root approximation; plugging $\sqrt{\text{OR}_{XY}}$ in place of RR_{XY} in Theorem 1.1. It has the drawback that it's not actually a bound, and as can be observed the error introduced by approximation can be big. The dotted line corrects the square root approximation using inequality (3.2); it's rigorous but weaker than the solid line bound given by Theorem 3.1.

4. JOINT BOUNDING FORMULAE UNDER AN EXTRA ASSUMPTION

The last two theorems are joint bounding formulae that only became possible under the newly introduced assumption (3.1), included here for completeness sake.

Theorem 4.1. Assuming (1.1) and (3.1),

$$\text{OR}_{XY} \leq 1 + \frac{(\text{RR}_{XZ} - 1)(\text{RR}_{ZY} - 1)}{\alpha(\text{RR}_{XZ} + \text{RR}_{ZY} - 1)}.$$

Proof: Assuming (1.1),

$$(4.1) \quad \text{RR}_{XY} \leq \frac{\text{RR}_{XZ}\text{RR}_{ZY}}{\text{RR}_{XZ} + \text{RR}_{ZY} - 1} = 1 + \frac{(\text{RR}_{XZ} - 1)(\text{RR}_{ZY} - 1)}{\text{RR}_{XZ} + \text{RR}_{ZY} - 1}$$

by a theorem of [Ding and VanderWeele \(2016\)](#); the second form is obtained with simple algebraic manipulation. Assuming (3.1), we can plug (4.1) into the bound of Lemma 3.1 and simplify. \square

Theorem 4.2. Assuming (1.1) and (3.1),

$$\text{OR}_{XY} \leq 1 + \frac{(\text{OR}_{XZ} - 1)(\text{RR}_{ZY} - 1)}{\alpha(\sqrt{\text{OR}_{XZ}} + \sqrt{\text{RR}_{ZY}})^2}.$$

Proof: Assuming (1.1),

$$(4.2) \quad \text{RR}_{XY} \leq \left(\frac{\sqrt{\text{OR}_{XZ}\text{RR}_{ZY}} + 1}{\sqrt{\text{OR}_{XZ}} + \sqrt{\text{RR}_{ZY}}} \right)^2 = 1 + \frac{(\text{RR}_{XZ} - 1)(\text{RR}_{ZY} - 1)}{(\sqrt{\text{OR}_{XZ}} + \sqrt{\text{RR}_{ZY}})^2}$$

by a theorem of [Lee \(2011\)](#); again the second form is but algebraic manipulation. Assuming (3.1), we can plug (4.2) into the bound of Lemma 3.1 and simplify. \square

It's easy if somewhat laborious to use the law of total probability and differential calculus in the same manner as employed in [Leppälä \(2023\)](#) to show that the bounds in Theorems 4.1 and 4.2 are sharp given the bounds (3.1). Likewise, using Theorems 4.1 and 4.2 as a starting point instead of Lemma 3.1 wouldn't improve the bounds in Theorems 3.1 and 3.2.

REFERENCES

- Cornfield, J. *et al.* (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *J. Natl. Cancer Inst.*, 22(1):173–203.
- Ding, P. and VanderWeele, T. (2016). Sensitivity analysis without assumptions. *Epidemiology (Cambridge, Mass.)*, 27(3):368–377.
- Lee, W.C. (2011). Bounding the bias of unmeasured factors with confounding and effect-modifying potentials. *Stat. Med.*, 30(9):1007–1017.
- Lee, W.C. and Wang, L.-Y. (2008). Simple formulas for gauging the potential impacts of population stratification bias. *Am. J. Epidemiol.*, 167(1):86–89.
- Leppälä, K. (2023). Sensitivity analysis on odds ratios. *Am. J. Epidemiol.*, 192(9):1882–1886.
- Schlesselman, J.J. (1978). Assessing effects of confounding variables. *Am. J. Epidemiol.*, 108(1):3–8.
- VanderWeele, T.J. (2017). On a square-root transformation of the odds ratio for a common outcome. *Epidemiology*, 28(6):e58–e60.
- VanderWeele, T.J. and Ding, P. (2017). Sensitivity analysis in observational research: introducing the *E*-value. *Ann. Intern. Med.*, 167(4):268–274.

REVSTAT — Statistical journal

AIMS AND SCOPE

The aim of REVSTAT — Statistical Journal is to publish articles of high scientific content, developing Statistical Science focused on innovative theory, methods, and applications in different areas of knowledge. Important survey/review contributing to Probability and Statistics advancement is also welcome.

BACKGROUND

Statistics Portugal started in 1996 the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, a quarterly publication whose goal was the publication of papers containing original research results, and application studies, namely in the economic, social, and demographic fields. Statistics Portugal was aware of how vital statistical culture is in understanding most phenomena in the present-day world, and of its responsibilities in disseminating statistical knowledge.

In 1998 it was decided to publish papers in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work. At the time, the editorial board was mainly composed by Portuguese university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal. In 2001, the *Revista de Estatística* published a three volumes special issue containing extended abstracts of the invited and contributed papers presented at the 23rd European Meeting of Statisticians (EMS). During the EMS 2001, its editor-in-chief invited several international participants to join the editorial staff.

In 2003 the name changed to REVSTAT — Statistical Journal, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

EDITORIAL POLICY

REVSTAT — Statistical Journal is an open access peer-reviewed journal published quarterly, in English, by Statistics Portugal.

The editorial policy of REVSTAT is mainly placed on the originality and importance of the research. The whole submission and review processes for REVSTAT are conducted exclusively online on the journal's webpage revstat.ine.pt based in Open Journal System (OJS). The only working language allowed is English. Authors intending to submit any work must register, login and follow the guidelines.

There are no fees for publishing accepted manuscripts that will be made available in open access.

All articles consistent with REVSTAT aims and scope will undergo scientific evaluation by at least two reviewers, one from the Editorial Board and another external. Authors can suggest an editor or reviewer who is expert on the paper subject providing her/his complete information, namely: name, affiliation, email and, if possible, personal URL or ORCID number.

All published works are Open Access (CC BY 4.0) which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. Also, in the context of archiving policy, REVSTAT is a *blue* journal welcoming authors to deposit their works in other scientific repositories regarding the use of the published edition and providing its source.

Journal prints may be ordered at expenses of the author(s), and prior to publication.

ABSTRACT AND INDEXING SERVICES

REVSTAT — *Journal Citation Reports - JCR (Clarivate); DOAJ-Directory of Open Access Journals; Current Index to Statistics; Google Scholar; Mathematical Reviews® (MathSciNet®); Zentralblatt für Mathematic; Scimago Journal & Country Rank; Scopus*

AUTHOR GUIDELINES

The whole submission and review processes for REVSTAT are conducted exclusively online on the journal's webpage <https://revstat.ine.pt/> based in Open Journal System (OJS). Authors intending to submit any work must *register, login* and follow the indications choosing *Submissions*.

REVSTAT — **Statistical Journal** adopts the COPE guidelines on publication ethics.

Work presentation

- the only working language is English;
- the first page should include the name, ORCID iD (optional), Institution, country, and email-address of the author(s);
- a summary of fewer than one hundred words, followed by a maximum of six keywords and the MSC 2020 subject classification should be included also in the first page;
- manuscripts should be typed only in black, in double-spacing, with a left margin of at least 3 cm, with numbered lines, and a maximum of 25 pages, in .pdf format;
- the title should be with no more than 120 characters (with spaces);
- figures must be a minimum of 300dpi and will be reproduced online as in the original work, however, authors should take into account that the printed version is always in black and grey tones;
- authors are encouraged to submit articles using LaTeX which macros are available at *REVSTAT style*;
- citations in text should be included in the text by name and year in parentheses, as in the following examples: § article title in lowercase (Author 1980); § This theorem was proved later by AuthorB and AuthorC (1990); § This subject has been widely addressed (AuthorA 1990; AuthorB et al. 1995; AuthorA and AuthorB 1998);
- references should be listed in alphabetical order of the author's scientific surname at the end of the article;
- acknowledgments of people, grants or funds should be placed in a short section before the References title page. Note that religious beliefs, ethnic background, citizenship and political orientations of the author(s) are not allowed in the text;
- authors are welcome to suggest one of the Editors or Associate Editors or yet other reviewer expert on the subject providing a complete information, namely: name, affiliation, email, and personal URL or ORCID number in the Comments for the Editor (submission form).

ACCEPTED PAPERS

After final revision and acceptance of an article for publication, authors are requested to provide the corresponding LaTeX file, as in REVSTAT style.

Supplementary files may be included and submitted separately in .tiff, .gif, .jpg, .png, .eps, .ps or .pdf format. These supplementary files may be published online along with an article, containing data, programming code, extra figures, or extra proofs, etc; however, REVSTAT is not responsible for any supporting information supplied by the author(s).

COPYRIGHT NOTICE

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, Statistics Portugal, to ensure the widest possible dissemination of information.

According to REVSTAT's archiving policy, after assigning the copyright form, authors may cite and use limited excerpts (figures, tables, etc.) of their works accepted/published in REVSTAT in other publications and may deposit only the published edition in scientific repositories providing its source as REVSTAT while the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.

EDITORIAL BOARD 2025-2026

Editor-in-Chief

Manuel SCOTTO, University of Lisbon, Portugal

Co-Editor

Cláudia NUNES, University of Lisbon, Portugal

Associate Editors

Abdelhakim AKNOUCHE, Qassim University, Saudi Arabia

Andrés ALONSO, Carlos III University of Madrid, Spain

Barry ARNOLD, University of California, United States

Wagner BARRETO-SOUZA, University College Dublin, Ireland

Francisco BLASQUES, VU Amsterdam, The Netherlands

Paula BRITO, University of Porto, Portugal

Rui CASTRO, Eindhoven University of Technology, The Netherlands

Valérie CHAVEZ-DEMOULIN, University of Lausanne, Switzerland

David CONESA, University of Valencia, Spain

Charmaine DEAN, University of Waterloo, Canada

Fernanda FIGUEIREDO, University of Porto, Portugal

Jorge Milhazes FREITAS, University of Porto, Portugal

Stéphane GIRARD, Inria Grenoble Rhône-Alpes, France

Sónia GOUVEIA, University of Aveiro, Portugal

Victor LEIVA, Pontificia Universidad Católica de Valparaíso, Chile

Artur LEMONTE, Federal University of Rio Grande do Norte, Brazil

Shuangzhe LIU, University of Canberra, Australia

Raquel MENEZES, University of Minho, Portugal

Fernando MOURA, Federal University of Rio de Janeiro, Brazil

Cláudia NEVES, King's College London, England

John NOLAN, American University, United States

Carlos OLIVEIRA, Norwegian University of Science and Technology, Norway

Paulo Eduardo OLIVEIRA, University of Coimbra, Portugal

Pedro OLIVEIRA, University of Porto, Portugal

Rosário OLIVEIRA, University of Lisbon, Portugal

Gilbert SAPORTA, Conservatoire National des Arts et Métiers, France

Alexandra M. SCHMIDT, McGill University, Canada

Lisete SOUSA, University of Lisbon, Portugal

Jacobo de UÑA-ÁLVAREZ, University of Vigo, Spain

Christian WEIB, Helmut Schmidt University, Germany

Executive Editor

Olga BESSA MENDES, Statistics Portugal