REVSTAT

Statistical Journal

vol. 22 - n. 2 - April 2024

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

INDEX

One Parameter Polynomial Exponential Distribution with Binomial Mixture

Authors: MOLAY KUMAR RUIDAS
– Department of Statistics, Triveni Devi Bhalotia College, Kazi Nazrul University,
Raniganj-713 347, India
molayruidas@gmail.com

INDRANI MUKHERJEE
– Department of Statistics, Visva-Bharati University,
Santiniketan-731 235, West Bengal, India
indranimukherjee.stat@gmail.com

MRIGANKA MOULI CHOUDHURY
– Department of Statistics, Visva-Bharati University,
Santiniketan-731 235, West Bengal, India
mmc.uttarpara@gmail.com

SUDHANSU S. MAITI 📧
– Department of Statistics, Visva-Bharati University,
Santiniketan-731 235, West Bengal, India
dssm1@rediffmail.com

SUMANTA ADHYA
– Department of Statistics, West Bengal State University,
Barasat-700 126, India
sumanta.adhya@gmail.com

Abstract:

• A further generalized version of one parameter polynomial exponential distribution with binomial probability mass as a mixture called a Binomial Mixture One Parameter Polynomial Exponential Distribution (BMOPPE) is proposed in the article. The moments and stochastic orderings are studied. Maximum likelihood estimator (MLE) and uniformly minimum variance unbiased estimator (UMVUE) of the probability density function and the cumulative distribution function have been derived and compared in the mean squared error sense. Estimation issues (both MLE and UMVUE) of reliability functions- mission time and stress-strength have been considered, and asymptotic variances of MLEs and variances of UMVUEs have been derived. UMVUEs of the variance of UMVUE of reliability functions have also been derived. Simulation study results have been reported to validate the theoretical findings. Few data sets have been fitted and compared.

📧 Corresponding author

1. INTRODUCTION

The probability density function (PDF) of the Lindley distribution [See, Lindley ([8])] is specified by

$$(1.1) \qquad f_X(x, \theta) = \frac{\theta^2}{1+\theta}(1+x)\exp(-\theta x), \qquad \theta > 0, \ x > 0,$$

and the corresponding cumulative density function (CDF) is given by

$$(1.2) \qquad F_X(x, \theta) = 1 - \frac{1+\theta+\theta x}{1+\theta}\exp(-\theta x), \qquad \theta > 0, \ x > 0.$$

A distribution that is close in form to (1.1) is the well-known exponential distribution given by the PDF

$$f_X(x, \theta) = \theta \exp(-\theta x), \qquad \theta > 0, \ x > 0.$$

Ghitany *et al.* ([5]) showed that in many ways the Lindley distribution is a better model than one based on the exponential distribution. The distribution in (1.1) is a mixture of exponential and gamma distribution with shape parameter 2 and scale parameter θ with mixing proportions $\frac{\theta}{1+\theta}$ and $\frac{1}{1+\theta}$, respectively. So, it is a Bernoulli mixture of gamma distributions of the form $f_X(x; \theta) = \sum_{k=0}^{1} \binom{1}{k}\left(\frac{1}{\theta+1}\right)^k\left(\frac{\theta}{\theta+1}\right)^{1-k} f_{GA}(x; k+1, \theta)$, where, $f_{GA}(x; k+1, \theta) = \frac{\theta^{k+1}}{\Gamma(k+1)}x^k \exp(-\theta x)$ is the gamma PDF with shape parameter $k+1$ and scale parameter θ. We generalize this distribution to binomial mixing with parameter r and $\frac{1}{\theta+1}$ of the form $f_X(x; \theta) = \sum_{k=0}^{r} \binom{r}{k}\left(\frac{1}{\theta+1}\right)^k\left(\frac{\theta}{\theta+1}\right)^{r-k} f_{GA}(x; k+1, \theta)$.

It can be written in more generalized form of the PDF as

$$(1.3) \qquad f_X(x, \theta) = \frac{\sum_{k=0}^{r} a_k p_k h_k(x; \theta)}{\sum_{k=0}^{r} a_k p_k}, \qquad \theta > 0, \ x > 0,$$

where, $p_k = \binom{r}{k}\left(\frac{1}{\theta+1}\right)^k\left(\frac{\theta}{\theta+1}\right)^{r-k}$, $h_k(x; \theta)) = \frac{\theta^{k+1}}{\Gamma(k+1)}x^{(k+1)-1}\exp(-\theta x)$ and a_k's are non-negative constants.

It can also be rewritten as

$$(1.4) \qquad f_X(x, \theta) = h(\theta)p(x)\exp(-\theta x), \qquad \theta > 0, \ x > 0,$$

where, $h(\theta) = \frac{1}{\sum_{k=0}^{r} a_k\binom{r}{k}\frac{1}{\theta^{k+1}}}$, $p(x) = \sum_{k=0}^{r} \frac{a_k}{k!}\binom{r}{k}x^k$.

A random variable X is said to have a Binomial Mixture One Parameter Polynomial Exponential (BMOPPE) with parameter θ, if its probability density function (PDF) is given by

$$(1.5) \qquad f_X(x, \theta) = \frac{\sum_{k=0}^{r} a_k\binom{r}{k}\left(\frac{1}{\theta+1}\right)^k\left(\frac{\theta}{\theta+1}\right)^{r-k}\frac{\theta^{k+1}}{\Gamma(k+1)}x^{(k+1)-1}\exp(-\theta x)}{\sum_{k=0}^{r} a_k\binom{r}{k}\left(\frac{1}{\theta+1}\right)^k\left(\frac{\theta}{\theta+1}\right)^{r-k}}, \qquad \theta > 0, \ x > 0.$$

The CDF of the random variable X is given by

$$(1.6) \qquad F(x) = \frac{\sum_{k=0}^{r} a_k \binom{r}{k} \left(\frac{1}{\theta+1}\right)^k \left(\frac{\theta}{\theta+1}\right)^{r-k} \gamma(k+1, \theta x)}{\sum_{k=0}^{r} a_k \binom{r}{k} \left(\frac{1}{\theta+1}\right)^k \left(\frac{\theta}{\theta+1}\right)^{r-k}}, \qquad \theta > 0, \ x > 0,$$

where $\gamma(s,t) = \frac{1}{\Gamma s} \int_0^t \exp(-x) x^{s-1} dx$ is the lower incomplete gamma function.

The CDF can also be written as

$$(1.7) \qquad F(x) = 1 - \left(\frac{\sum_{k=0}^{r} a_k \binom{r}{k} \frac{1}{\theta^{k+1}} \Gamma(k+1, \theta x)}{\sum_{k=0}^{r} a_k \binom{r}{k} \frac{1}{\theta^{k+1}}} \right), \qquad x, \ \theta > 0,$$

where $\Gamma(m, x) = \frac{1}{\Gamma(m)} \int_x^{\infty} e^{-u} u^{m-1} du$, the upper incomplete gamma function.

(a) For $r = 3$ and $\theta = 0.5$. **(b)** For $r = 4$ and $\theta = 0.5$. **(c)** For $r = 4$ and $\theta = 1$.
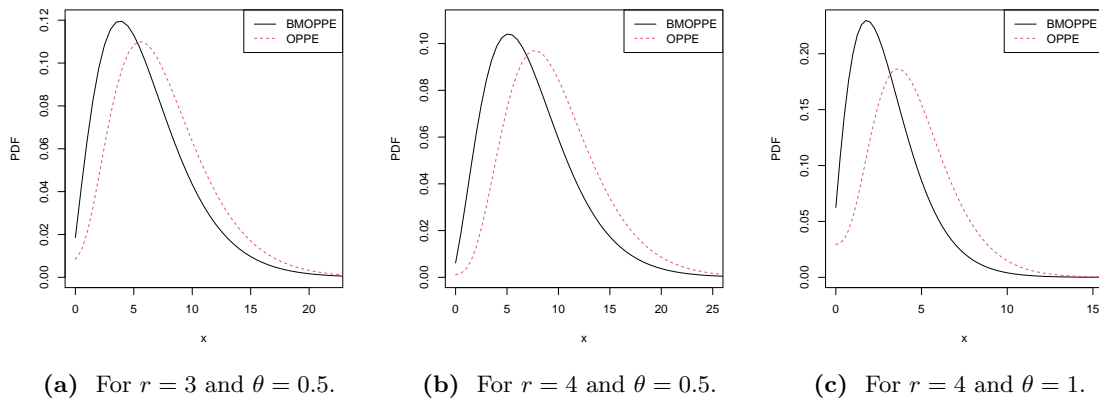
Figure 1: Plot of PDF of BMOPPE and OPPE for different combinations.

Bouchahed and Zeghdoudi ([4]) proposed a new and unified approach to generalizing Lindley's distribution and investigated its properties. Mukherjee *et al.* ([11]) later called it the One Parameter Polynomial Exponential (OPPE) family of distributions and studied the estimation aspect of the PDF and the CDF of the distribution. The natural discrete version of the OPPE called the Natural Discrete One Parameter Polynomial Exponential (NDOPPE) family of distributions is studied by Maiti *et al.* ([9]), and the estimation aspect of the PMF and the CDF is discussed by Mukherjee *et al.* ([10]). The OPPE is a mixture of gamma distributions with some mixing probabilities. In contrast, the BMOPPE is revisited with a different look, and it is also a mixture of gamma distributions with binomial mixing probabilities, which is different from the previous one.

The article is organised as follows. Section 2 discusses different order moments and stochastic orderings of the random variable. In section 3, The maximum likelihood estimator (MLE) and uniformly minimum variance unbiased estimator (UMVUE) of the PDF and the CDF are discussed. The estimators are compared in the mean squared error (MSE) sense. This section also considers the estimation of both mission time and stress-strength reliability functions. Asymptotic variances of MLEs and variances of MVUEs are derived. UMVUEs of variances of UMVUE of reliability functions have been derived. Simulation study results have been reported to verify the theoretical findings in section 4. Three data sets have been analysed for illustration purposes in section 5. Section 6 is for making some concluding remarks.

2.   MOMENTS AND STOCHASTIC ORDERINGS

The s-th raw moments for BMOPPE distribution is

$$(2.1) \qquad \mu'_s = \frac{\sum_{k=0}^{r} a_k \binom{r}{k} \left(\frac{1}{\theta+1}\right)^k \left(\frac{\theta}{\theta+1}\right)^{r-k} \frac{\Gamma(k+s+1)}{\theta^s \Gamma(k+1)}}{\sum_{k=0}^{r} a_k \binom{r}{k} \left(\frac{1}{\theta+1}\right)^k \left(\frac{\theta}{\theta+1}\right)^{r-k}}.$$

The coefficient of skewness and kurtosis measures have been shown in Figure 2. These are shown for $r = 1, 2, 3$ and for different values of θ. The constants of polynomial are taken as $a_i = 1$, $i = 0, 1, 2, 3$. It is noticed that the distribution is positively skewed, and skewness decreases with the increment of the degree of the polynomial. Also, the distribution is leptokurtic, and it becomes long-tailed with the increment of the degree of the polynomial.

(a) Skewness.   **(b)** Kurtosis.
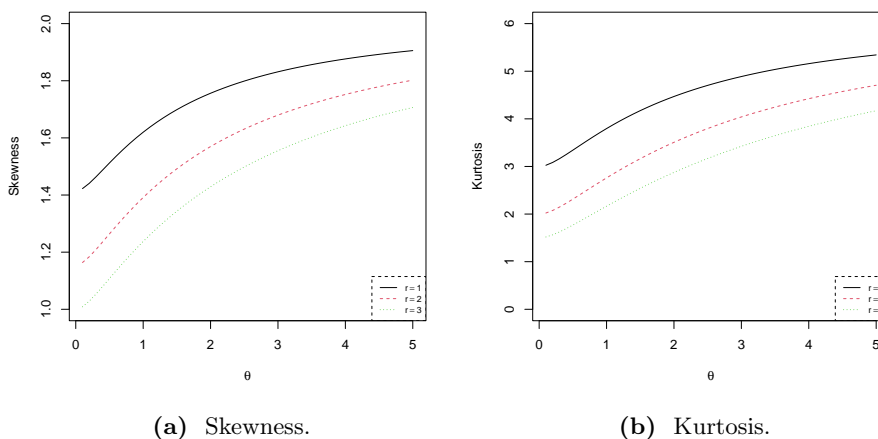
Figure 2: Plot of Skewness and Kurtosis of BMOPPE for different θ and r.

The ordering relations between two BMOPPE random variables have been shown in the following Theorem.

Theorem 2.1. *Let $X_i \sim$ BMOPPE(θ_i), $i = 1, 2$ be two random variables. If $\theta_2 \leq \theta_1$, then $X_1 \prec_{\mathrm{lr}} X_2, X_1 \prec_{\mathrm{hr}} X_2, X_1 \prec_{\mathrm{st}} X_2$ and $X_1 \prec_{\mathrm{cx}} X_2$.*

Proof: Since, Likelihood ratio order $\implies$ Hazard rate order $\implies$ Stochastic order and Convex order $\iff$ Stochastic order, it is sufficient to show that Likelihood ratio order holds.

We have

$$L(x) = \ln\left(\frac{f_{X_1}(x)}{f_{X_2}(x)}\right) = (k+1)\ln\left(\frac{\theta_1}{\theta_2}\right) - (\theta_1 - \theta_2)x.$$

Now,

$$\triangle L(x) = \frac{d}{dx}\left[\ln\left(\frac{f_{X_1}(x)}{f_{X_2}(x)}\right)\right].$$

Clearly, it is evident that $\triangle L(x) \leq 0$, $\forall \theta_2 \leq \theta_1$. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

3. ESTIMATION OF PDF AND CDF AND THEIR APPLICATIONS IN RELIABILITY ESTIMATION

The PDF and the CDF estimates have immense importance in estimating reliability functions (both mission time and stress-strength), entropy functions, Kullback–Leibler divergence measure, Fisher information, cumulative residual entropy, quantile function, hazard rate function, etc. In this section, the MLE and UMVUE of the reliability functions are to be attempted. The asymptotic variances/variances of the estimators and their estimators are to be discussed.

First, we will discuss the MLE and UMVUE of the PDF and the CDF of BMOPPE family of distributions. Let $X_1, X_2, ..., X_n$ be random sample of size n drawn from the BMOPPE distribution in (1.5). The MLE of θ which is denoted as $\widetilde{\theta}$ is obtained by numerically solving the equation

$$(3.1) \qquad \frac{\sum_{k=0}^{r} a_k \binom{r}{k} \frac{k+1}{\theta^{k+2}}}{\sum_{k=0}^{r} a_k \binom{r}{k} \frac{1}{\theta^{k+1}}} - \bar{X} = 0.$$

Using the invariance property of MLE, one can obtain the MLEs of the PDF and that of the CDF by substituting $\widetilde{\theta}$ in (1.5) and (1.6) respectively. Theoretical expressions for the MSE of the MLEs are not available. MSE is to be studied through simulation.

To derive the UMVUE of the PDF and that of the CDF (stated in Theorem 3.2), we will use the following Theorem 3.1 and Lemma 3.1.

Theorem 3.1. *Let $X_1, X_2, ..., X_n$ independently follow* BMOPPE(θ). *Then the distribution of $T = X_1 + X_2 + \cdots + X_n$ is*

$$f(t) = h^n(\theta) \sum_{y_0} \sum_{y_1} \cdots \sum_{y_r} c(n, y_0, y_1, ..., y_r) \exp(-\theta t) \, t^{\sum_{k=0}^{r}(k+1)y_k - 1}, \qquad t > 0,$$

with $y_0 + y_1 + \cdots + y_r = n$ and $c(n, y_0, y_1, ..., y_r) = \frac{n!}{y_0! y_1! \cdots y_r!} \frac{\prod_{k=0}^{r}[a_k \binom{r}{k}]^{y_k}}{\Gamma(\sum_{k=0}^{r}(k+1)y_k)}$.

Proof: Since X_i's are independent and identically distributed, the moment generating function (mgf) of T is

$$M_T(t) = h^n(\theta) \left[\sum_{k=0}^{r} a_k \binom{r}{k} \frac{1}{\theta^{k+1}} \left(1 - \frac{t}{\theta} \right)^{-(k+1)} \right]^n$$

$$= h^n(\theta) \left[a_0 \binom{r}{0} \frac{1}{\theta} \left(1 - \frac{t}{\theta} \right)^{-1} + \cdots + a_r \binom{r}{r} \frac{1}{\theta^{r+1}} \left(1 - \frac{t}{\theta} \right)^{-(r+1)} \right]^n$$

$$= h^n(\theta) \sum_{y_0} \sum_{y_1} \cdots \sum_{y_r} \frac{n!}{y_0! y_1! \cdots y_r!} \prod_{k=0}^{r} \left[a_k \binom{r}{k} \right]^{y_k} \theta^{-\sum_{k=0}^{r}(k+1)y_k} \left(1 - \frac{t}{\theta} \right)^{-\sum_{k=0}^{r}(k+1)y_k}.$$

Hence, the distribution of T is

$$f(t) = h^n(\theta) \sum_{y_0} \sum_{y_1} \cdots \sum_{y_r} \frac{n!}{y_0! y_1! \cdots y_r!} \prod_{k=0}^{r} \left[a_k \binom{r}{k} \right]^{y_k} \theta^{-\sum_{k=0}^{r}(k+1)y_k} f_{GA}\left(t, \sum_{k=0}^{r}(k+1)y_k, \theta\right)$$

$$= h^n(\theta) \sum_{y_0} \sum_{y_1} \cdots \sum_{y_r} c(n, y_0, y_1, ..., y_r) t^{\sum_{k=0}^{r}(k+1)y_k - 1} \exp(-\theta t). \qquad \square$$

Lemma 3.1. *The conditional distribution of X_1 given $T = X_1 + X_2 + \cdots + X_n$ is*

$$f_{X_1|T}(x|t) = \frac{p(x)}{A_n(t)} \sum_{q_0} \sum_{q_1} \cdots \sum_{q_r} c(n-1, q_0, q_1, ..., q_r) (t-x)^{\sum_{k=0}^{r}(k+1)q_k - 1}, \qquad 0 < x < t,$$

where

$$A_n(t) = \sum_{y_0} \sum_{y_1} \cdots \sum_{y_r} c(n, \ y_0, \ y_1, \ ..., \ y_r) t^{\sum_{k=0}^{r}(k+1)y_k - 1},$$

and

$$c(n-1, q_0, q_1, ..., q_r) = \frac{(n-1)!}{q_0! q_1! \cdots q_r!} \prod_{k=0}^{r} \left[a_k \binom{r}{k} \right]^{q_k} \frac{1}{\Gamma(\sum_{k=0}^{r}(k+1)q_k)},$$

with $q_0 + q_1 + q_2 + \cdots + q_r = n - 1$.

Proof: The proof is obviously conducted from

$$f_{X_1|T}(x|t) = \frac{f_{X_1}(x) f(t-x)}{f(t)}$$

$$= \frac{p(x)}{A_n(t)} \sum_{q_0} \sum_{q_1} \cdots \sum_{q_r} c(n-1, q_0, q_1, ..., q_r) (t-x)^{\sum_{k=0}^{r}(k+1)q_k - 1}. \qquad \square$$

Theorem 3.2. *Let $T = t$ be given. Then*

$$(3.2) \quad \widehat{f}(x) = \frac{p(x)}{A_n(t)} \sum_{q_0} \sum_{q_1} \cdots \sum_{q_r} c(n-1, q_0, q_1, ..., q_r) (t-x)^{\sum_{k=0}^{r}(k+1)q_k - 1}, \qquad 0 < x < t,$$

is UMVUE for $f(x)$ and

$$(3.3) \quad \widehat{F}(x) = 1 - \frac{1}{A_n(t)} \sum_{q_0} \sum_{q_1} \cdots \sum_{q_r} c(n-1, q_0, q_1, ..., q_r)$$

$$\times \sum_{k=0}^{r} a_k \binom{r}{k} \frac{1}{\Gamma(k+1)} t^{\sum_{k=0}^{r}(k+1)q_k + k}$$

$$\times B\left((k+1), \sum_{k=0}^{r}(k+1)y_k\right) I_{x/t}\left((k+1), \sum_{k=0}^{r}(k+1)q_k\right), \qquad 0 < x < t,$$

is UMVUE for $F(x)$, where $I_x(\alpha, \beta) = \frac{1}{B(\alpha,\beta)} \int_x^1 u^{\alpha-1}(1-u)^{\beta-1} du$ is an incomplete beta function and $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

Proof: The BMOPPE in (1.4) is a member of one parameter exponential family with $T = \sum_{i=1}^{n} X_i$ as complete sufficient statistic. Therefore, by the use of Lehmann–Scheffe theorem, we get the UMVUE of the PDF from Lemma 3.1.

$$
\begin{aligned}
\widehat{F}(x) &= 1 - \int_x^t \widehat{f}(w)dw \\
&= 1 - \int_x^t \frac{p(w)}{A_n(t)} \sum_{q_0} \sum_{q_1} \cdots \sum_{q_r} c(n-1, q_0, q_1, ..., q_r) \, (t-w)^{\sum_{k=0}^{r}(k+1)q_k - 1} dw \\
&= 1 - \frac{1}{A_n(t)} \sum_{q_0} \sum_{q_1} \cdots \sum_{q_r} c(n-1, q_0, q_1, ..., q_r) \\
&\quad \times \sum_{k=0}^{r} a_k \binom{r}{k} \frac{1}{\Gamma(k+1)} t^{\sum_{k=0}^{r}(k+1)q_k + k} \\
&\quad \times B\left((k+1), \sum_{k=0}^{r}(k+1)y_k\right) I_{x/t}\left((k+1), \sum_{k=0}^{r}(k+1)q_k\right). \qquad \square
\end{aligned}
$$

3.1. Mission Time Reliability

Suppose the life length of a component X is distributed as BMOPPE(θ). Then the reliability of that component for a fixed mission time $t_0 \, (>0)$ is

$$
\begin{aligned}
\bar{F}_X(t_0) &= P(X \geq t_0) \\
&= h(\theta) \sum_{k=0}^{r} \frac{a_k \binom{r}{k}}{\theta^{k+1}} \Gamma(k+1, \, \theta t_0).
\end{aligned}
$$

By using the relation between incomplete gamma function and Poisson probability, $\bar{F}_X(t_0)$ can be written as

$$
\bar{F}_X(t_0) = h(\theta) \sum_{k=0}^{r} \frac{a_k \binom{r}{k}}{\theta^{k+1}} \sum_{j=0}^{k} \frac{e^{-\theta t_0}(\theta t_0)^j}{j!}.
$$

3.1.1. The MLE

The MLE of $\bar{F}_X(t_0)$ based on a random sample of size n, is

$$
\widetilde{\bar{F}}_X(t_0) = h(\widetilde{\theta}) \sum_{k=0}^{r} \frac{a_k \binom{r}{k}}{\widetilde{\theta}^{k+1}} \sum_{j=0}^{k} \frac{e^{-\widetilde{\theta} t_0}(\widetilde{\theta} t_0)^j}{j!},
$$

where $\widetilde{\theta}$ is the solution of the equation (3.1). Since $\text{MSE}(\widetilde{\bar{F}}_X(t_0))$ has no closed form expression, we give the asymptotic distribution of $\widetilde{\bar{F}}_X(t_0)$ by using delta theorem as follows:

$$
\sqrt{n}\left(\widetilde{\bar{F}}_X(t_0) - \bar{F}_X(t_0)\right) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\left[\frac{d\bar{F}_X(t_0)}{d\theta}\right]^2\right),
$$

where

$$I(\theta) = \frac{\left[\sum_{k=0}^{r}\frac{a_k\binom{r}{k}}{\theta^{k+1}}\right]\left[\sum_{k=0}^{r}\frac{a_k\binom{r}{k}(k+2)(k+1)}{\theta^{k+3}}\right] - \left[\sum_{k=0}^{r}\frac{a_k\binom{r}{k}(k+1)}{\theta^{k+2}}\right]^2}{\left[\sum_{k=0}^{r}\frac{a_k\binom{r}{k}}{\theta^{k+1}}\right]^2}$$

is the Fisher information about parameter θ for a single observation X.

3.1.2. The UMVUE

Application of incomplete beta function and binomial probability gives the UMVUE of $\bar{F}_X(t_0)$ as

$$\widehat{\bar{F}}_X(t_0) = \frac{1}{A_n(t)}\sum_{x_0}\sum_{x_1}\cdots\sum_{x_r}{}_{\sum_{i=0}^{r} x_i = n-1} c(n-1,\ x_0,\ x_1,\ ...,\ x_r)$$

$$\times \sum_{k=0}^{r}\frac{a_k\binom{r}{k}}{\Gamma(k+1)}t_0^{\sum_{k=0}^{r}(k+1)x_k+k}B\left((k+1),\ \sum_{k=0}^{r}(k+1)x_k\right)$$

$$\times \sum_{j=0}^{k}\binom{\sum_{l=0}^{r}(l+1)x_l+k}{j}\left(\frac{t_0}{t}\right)^{j}\left(1-\frac{t_0}{t}\right)^{\sum_{l=0}^{r}(l+1)x_l+k-j}.$$

For exponential family of distribution, Blight and Rao ([3]) considered Bhattacharya bounds to calculate the variance of UMVUE of parametric function $\psi(\theta)$. So, variance of $\widehat{\bar{F}}_X(t_0)$ can be expressed as

$$\mathrm{Var}(\widehat{\bar{F}}_X(t_0)) = \sum_{l=1}^{\infty}\frac{[\bar{F}_X^{(l)}(t_0)]^2}{[J_l^*(\theta)]^2},$$

where

$$[J_l^*(\theta)]^2 = [h(\theta)]^n\sum_{j=0}^{l}\sum_{i=0}^{l}(-1)^{i+j}\binom{l}{i}\binom{l}{j}[h^{(l-i)}(\theta)]^n[h^{(l-j)}(\theta)]^n$$

$$\times\sum_{x_0}\sum_{x_1}\cdots\sum_{x_r}{}_{\sum_{i=0}^{r} x_i = n-1} c(n-1,\ x_0,\ x_1,\ ...,\ x_r)$$

$$\times\frac{\Gamma(i+j+\sum_{l=0}^{r}(l+1)x_l)}{\theta^{i+j+\sum_{l=0}^{r}(l+1)x_l}}$$

is the Bhattacharya function and $h^{(i)}(\theta)$ denotes the i-th derivative of h with respect to θ.

Now, for the derivation of UMVUE of $\mathrm{Var}(\widehat{\bar{F}}_X(t_0))$, we consider the representation

$$\mathrm{Var}(\widehat{\bar{F}}_X(t_0)) = E(\widehat{\bar{F}}_X^2(t_0)) - [E(\widehat{\bar{F}}_X(t_0))]^2$$
$$= E(\widehat{\bar{F}}_X^2(t_0)) - \bar{F}_X^2(t_0).$$

If $\widehat{Q}_1(t_0)$ is the UMVUE of $Q_1(t_0) = \bar{F}_X^2(t_0)$, we get the UMVUE of $\mathrm{Var}(\widehat{\bar{F}}_X(t_0))$ as $\widehat{\bar{F}}_X^2(t_0) - \widehat{Q}_1(t_0)$. We start from the fact that

$$I[X_1 \geq t_0, X_2 \geq t_0] = I[X_1 \geq t_0]I[X_2 \geq t_0],$$

which implies that $I[X_1 \geq t_0, X_2 \geq t_0]$ is unbiased for $Q_1(t_0)$. Then, we get the UMVUE by using Rao–Blackwell theorem as

$$\widehat{Q}_1(t_0) = \frac{1}{A_n(t)} \sum_{x_0} \sum_{x_1} \cdots \sum_{x_r} {}_{\sum_{i=0}^r x_i = n-2} c(n-2, \ x_0, \ x_1, \ ..., \ x_r)$$

$$\times \sum_{k=0}^r \sum_{l=0}^r \frac{a_k a_l \binom{r}{k} \binom{r}{l}}{\Gamma(k+1)\Gamma(l+1)} \sum_{i=0}^{\sum_{m=0}^r (m+1)x_m - 1} \frac{(-1)^i \binom{\sum_{m=0}^r (m+1)x_m - 1}{i}}{l+i+1}$$

$$\times \left[t^{\sum_{m=0}^r (m+1)x_m + l} \sum_{u=0}^{\sum_{m=0}^r (m+1)x_m + l} \frac{(-1)^u \binom{\sum_{m=0}^r (m+1)x_m + l}{u}}{t^u(k+u+1)} \left(t^{k+u+1} - t_0^{k+u+1} \right) \right.$$

$$- t_0^{l+i+1} t^{\sum_{m=0}^r (m+1)x_m - i - 1} \sum_{v=0}^{\sum_{m=0}^r (m+1)x_m - i - 1} \frac{(-1)^v \binom{\sum_{m=0}^r (m+1)x_m - i - 1}{v}}{t^v(k+v+1)}$$

$$\left. \times \left(t^{k+v+1} - t_0^{k+v+1} \right) \right].$$

3.2. Stress Strength Reliability

For estimation of stress strength reliability, we assume the stress random variable X follows BMOPPE(θ_1) and strength random variable Y follows BMOPPE(θ_2) and they are independently distributed. Then the expression of stress strength reliability becomes

$$R = P(X < Y)$$
$$= \int_0^\infty \bar{F}_Y(x) f(x) dx$$
$$= h(\theta_1) h(\theta_2) \sum_{k_2=0}^{r_2} \frac{b_{k_2} \binom{r_2}{k_2}}{\theta_2^{k_2+1}} \sum_{k_1=0}^{r_1} \frac{a_{k_1} \binom{r_1}{k_1}}{\theta_1^{k_1+1}} \sum_{j=0}^{k_2} \binom{k_1+j}{j} \left(\frac{\theta_2}{\theta_1+\theta_2} \right)^j \left(\frac{\theta_1}{\theta_1+\theta_2} \right)^{k_1+1}.$$

3.2.1. The MLE of R

Let $(x_1, x_2, ..., x_{n_1})$ and $(y_1, y_2, ..., y_{n_2})$ be independent samples drawn from BMOPPE(θ_1) and BMOPPE(θ_2), respectively. Let the MLEs of θ_1 and θ_2 be $\widetilde{\theta}_1$ and $\widetilde{\theta}_2$, respectively. By putting the values of $\widetilde{\theta}_1$ and $\widetilde{\theta}_2$ in the expression of R, we get $\widetilde{R}$ by its invariance property as

$$\widetilde{R} = h(\widetilde{\theta}_1) h(\widetilde{\theta}_2) \sum_{k_2=0}^{r_2} \frac{b_{k_2} \binom{r_2}{k_2}}{\widetilde{\theta}_2^{k_2+1}} \sum_{k_1=0}^{r_1} \frac{a_{k_1} \binom{r_1}{k_1}}{\widetilde{\theta}_1^{k_1+1}} \sum_{j=0}^{k_2} \binom{k_1+j}{j} \left(\frac{\widetilde{\theta}_2}{\widetilde{\theta}_1+\widetilde{\theta}_2} \right)^j \left(\frac{\widetilde{\theta}_1}{\widetilde{\theta}_1+\widetilde{\theta}_2} \right)^{k_1+1}.$$

Similarly as in mission time reliability, we derive the asymptotic distribution of $\widetilde{R}$ in the following theorem.

Theorem 3.3. *If the ratio $\frac{n_1}{n_2}$ converges to a positive number κ when both n_1 and n_2 tends to infinity, then*

$$\sqrt{n}(\widetilde{R} - R) \xrightarrow{d} N(0, \sigma^2),$$

where $\sigma^2 = \frac{1}{I_1(\theta_1)}[\frac{\partial R}{\partial \theta_1}]^2 + \frac{\kappa}{I_2(\theta_2)}[\frac{\partial R}{\partial \theta_2}]^2$ and $I_1(\theta_1)$ is a Fisher information about parameter θ_1 in a single observation X, $I_2(\theta_2)$ is a Fisher information about θ_2 in a single observation Y.

Proof: Since log-likelihood equations satisfies all regularity conditions of asymptotic normality for MLE, then we have

$$\sqrt{n_j}(\widetilde{\theta}_j - \theta_j) \xrightarrow{d} N(0, [I_j(\theta_j)]^{-1}),$$

where

$$I_1(\theta_1) = \frac{\left[\sum_{k=0}^{r_1} \frac{a_k\binom{r_1}{k}}{\theta_1^{k+1}}\right]\left[\sum_{k=0}^{r_1} \frac{a_k\binom{r_1}{k}(k+2)(k+1)}{\theta_1^{k+3}}\right] - \left[\sum_{k=0}^{r_1} \frac{a_k\binom{r_1}{k}(k+1)}{\theta_1^{k+2}}\right]^2}{\left[\sum_{k=0}^{r_1} \frac{a_k\binom{r_1}{k}}{\theta_1^{k+1}}\right]^2}$$

and

$$I_1(\theta_2) = \frac{\left[\sum_{k=0}^{r_2} \frac{b_k\binom{r_2}{k}}{\theta_2^{k+1}}\right]\left[\sum_{k=0}^{r_2} \frac{b_k\binom{r_2}{k}(k+2)(k+1)}{\theta_2^{k+3}}\right] - \left[\sum_{k=0}^{r_2} \frac{b_k\binom{r_2}{k}(k+1)}{\theta_2^{k+2}}\right]^2}{\left[\sum_{k=0}^{r_2} \frac{b_k\binom{r_2}{k}}{\theta_2^{k+1}}\right]^2}.$$

Again, from the fact of independence of $\widetilde{\theta}_1$ and $\widetilde{\theta}_2$, we get

$$\sqrt{n_1}(\widetilde{\theta}_1 - \theta_1, \widetilde{\theta}_2 - \theta_2) \xrightarrow{d} N_2(0, J(\theta_1, \theta_2)),$$

where

$$J(\theta_1, \theta_2) = \begin{bmatrix} [I_1(\theta_1)]^{-1} & 0 \\ 0 & \kappa[I_2(\theta_2)]^{-1} \end{bmatrix}.$$

Now application of the transformation $R = R(\theta_1, \theta_2)$ together with Delta theorem conclude the proof. □

3.2.2. The UMVUE of R

By using the UMVUE of mission time reliability and the PDF of the BMOPPE distribution, the UMVUE of R can be expressed as

$$\widehat{R} = \int_{x=0}^{\text{Min}(t_1, t_2)} \widehat{\overline{F}}_Y(x)\widehat{f}_X(x)dx,$$

where $t_1 = \sum_{i=1}^{n_1} x_i$ and $t_2 = \sum_{i=1}^{n_2} y_i$, respectively. Evaluation of the integral gives the final form of the UMVUE of R as

$$
\widehat{R} = \frac{1}{A_{n_1}(t_1)A_{n_2}(t_2)} \sum_{q_0} \sum_{q_1} \cdots \sum_{q_{r_1} \sum_{i=0}^{r_1} q_i = n_1 - 1} c(n_1 - 1, \; q_0, \; q_1, \; ..., \; q_{r_1})
$$

$$
\times \; t_1^{\sum_{l=0}^{r_1}(l+1)q_l} \sum_{w_0} \sum_{w_1} \cdots \sum_{w_{r_2} \sum_{i=0}^{r_2} w_i = n_2 - 1} c(n_2 - 1, \; w_0, \; w_1, \; ..., \; w_{r_2})
$$

$$
\times \; t_2^{\sum_{m=0}^{r_2}(m+1)w_m} J_1\left(\sum_{l=0}^{r_1}(l+1)q_l, \; \sum_{m=0}^{r_2}(m+1)w_m \right),
$$

where

$$
J_1(\alpha, \; \beta) = \sum_{k_1=0}^{r_1} \frac{a_{k_1}\binom{r_1}{k_1}t_1^{k_1}}{\Gamma(k_1+1)} \sum_{k_2=0}^{r_2} \frac{b_{k_2}\binom{r_2}{k_2}t_2^{k_2}}{\Gamma(k_2+1)}
$$

$$
\times \sum_{j=0}^{k_2} \sum_{i=0}^{\beta+k_2-j} (-1)^i \left(\frac{t_1}{t_2}\right)^{i+j} \binom{\beta+k_2}{j}\binom{\beta+k_2-j}{i} B(k_2+1, \beta) \, B(i+j+k_1+1, \alpha)
$$

$$
\times \sum_{l=i+j+k_1+1}^{\alpha+i+j+k_1} \binom{\alpha+i+j+k_1}{l} \left(\frac{\text{Min}(t_1,t_2)}{t_1}\right)^l \left(1 - \frac{\text{Min}(t_1,t_2)}{t_1}\right)^{(\alpha+i+j+k_1-l)}.
$$

Under certain regularity conditions, variance of $\widehat{R}$ takes the form [See, Bartoszewicz ([1])] as follows:

$$
\text{Var}(\widehat{R}) = \sum_{k=1}^{\infty} \sum_{j=0}^{k} \left[\frac{\partial^k R}{\partial \theta_1^j \partial \theta_2^{k-j}} \frac{1}{I_j(\theta_1)J_{k-j}(\theta_2)} \right]^2,
$$

where

$$
[I_l(\theta_1)]^2 = [h(\theta_1)]^{n_1} \sum_{j=0}^{l} \sum_{i=0}^{l} (-1)^{i+j} \binom{l}{i}\binom{l}{j} \, [h^{(l-i)}(\theta_1)]^{n_1} \, [h^{(l-j)}(\theta_1)]^{n_1}
$$

$$
\times \sum_{q_0} \sum_{q_1} \cdots \sum_{q_{r_1} \sum_{i=0}^{r_1} q_i = n_1 - 1} c(n_1 - 1, \; q_0, \; q_1, \; ..., \; q_{r_1}) \, \frac{\Gamma(i+j+\sum_{l=0}^{r_1}(l+1)q_l)}{\theta_1^{i+j+\sum_{l=0}^{r_1}(l+1)q_l}},
$$

$$
[J_l(\theta_2)]^2 = [h(\theta_2)]^{n_2} \sum_{j=0}^{l} \sum_{i=0}^{l} (-1)^{i+j} \binom{l}{i}\binom{l}{j} \, [h^{(l-i)}(\theta_2)]^{n_2} \, [h^{(l-j)}(\theta_2)]^{n_2}
$$

$$
\times \sum_{w_0} \sum_{w_1} \cdots \sum_{w_{r_2} \sum_{i=0}^{r_2} w_i = n_2 - 1} c(n_2 - 1, \; w_0, \; w_1, \; ..., \; w_{r_2}) \, \frac{\Gamma(i+j+\sum_{l=0}^{r_2}(l+1)w_l)}{\theta_2^{i+j+\sum_{l=0}^{r_2}(l+1)w_l}}.
$$

As earlier, the following representation

$$
\text{Var}(\widehat{R}) = E(\widehat{R^2}) - [E(\widehat{R})]^2
$$
$$
= E(\widehat{R^2}) - R^2
$$

gives the UMVUE of $\mathrm{Var}(\widehat{R})$ as $\widehat{R^2} - \widehat{Q}_2$, where $\widehat{Q}_2$ is the UMVUE of $Q_2 = R^2$. Similarly, we give the final expression of $\widehat{Q}_2$ in the following equation:

$$
\widehat{Q}_2 = \frac{1}{A_{n_1}(t_1)A_{n_2}(t_2)} \sum_{q_0}\sum_{q_1}\cdots\sum_{q_{r_1}\; \sum_{i=0}^{r_1} q_i = n_1 - 2} c(n_1 - 2,\; q_0,\; q_1,\; ...,\; q_{r_1})
$$

$$
\times \sum_{w_0}\sum_{w_1}\cdots\sum_{w_{r_2}\; \sum_{i=0}^{r_2} w_i = n_2 - 1} c(n_2 - 1,\; w_0,\; w_1,\; ...,\; w_{r_2})
$$

$$
\times \sum_{k_1=0}^{r_1}\sum_{l_1=0}^{r_1} \frac{a_{k_1}a_{l_1}}{\Gamma(k_1+1)}\, \frac{\binom{r_1}{k_1}\binom{r_1}{l_1}}{\Gamma(l_1+1)} \sum_{k_2=0}^{r_2}\sum_{l_2=0}^{r_2} \frac{b_{k_2}b_{l_2}}{\Gamma(k_2+1)}\, \frac{\binom{r_2}{k_2}\binom{r_2}{l_2}}{\Gamma(l_2+1)}
$$

$$
\times \sum_{i=0}^{\sum_{m=0}^{r_2}(m+1)w_m - 1} \frac{(-1)^i \binom{\sum_{m=0}^{r_2}(m+1)w_m - 1}{i}}{l_2 + i + 1}
$$

$$
\times \left[t_2^{\sum_{m=0}^{r_2}(m+1)w_m + l_2} \sum_{u=0}^{\sum_{m=0}^{r_2}(m+1)w_m + l_2} \frac{(-1)^u \binom{\sum_{m=0}^{r_2}(m+1)w_m + l_2}{u}}{t_2^u (k_2 + u + 1)} \right.
$$

$$
\times \left[t_2^u J_2\!\left(k_1,\; l_1,\; \sum_{m=0}^{r_1}(m+1)q_m - 1\right) - J_2\!\left(k_1 + k_1 + u + 1,\; l_1,\; \sum_{m=0}^{r_1}(m+1)q_m - 1\right) \right]
$$

$$
- t_2^{\sum_{m=0}^{r_2}(m+1)w_m - i - 1} \sum_{v=0}^{\sum_{m=0}^{r_2}(m+1)w_m - i - 1} \frac{(-1)^v \binom{\sum_{m=0}^{r_2}(m+1)w_m - i - 1}{v}}{t_2^v (k_2 + v + 1)}
$$

$$
\times \left[t_2^{k_2 + v + 1} J_2\!\left(k_1,\; l_1 + l_2 + i + 1,\; \sum_{m=0}^{r_1}(m+1)q_m - 1\right) \right.
$$

$$
\left.\left. - J_2\!\left(k_1 + k_2 + v + 1,\; l_1 + l_2 + i + 1,\; \sum_{m=0}^{r_1}(m+1)q_m - 1\right) \right]\right],
$$

where

$$
J_2(a,\; b,\; \beta) = \sum_{i=0}^{\beta - 1} (-1)^i \binom{\beta - 1}{i} \frac{t_1^{a+b+\beta+1}}{b + \beta + 1}\, B(a+1,\; b + \beta + 1)
$$

$$
\times \sum_{j=a+1}^{a+b+\beta+1} \binom{a + b + \beta + 1}{j} \left(\frac{\mathrm{Min}(t_1, t_2)}{t_1}\right)^j \left(1 - \frac{\mathrm{Min}(t_1, t_2)}{t_1}\right)^{a+b+\beta+1-j}.
$$

4.   SIMULATION STUDY

Monte Carlo Simulation technique will not be helpful in generating random samples from the BMOPPE distribution, since the equation

$$
F(x) = u, \qquad u \in (0, 1),
$$

cannot explicitly be solved in x. However, since the distribution is a mixture of gamma distributions given in (1.5), one can utilize this fact. For the BMOPPE distribution, the generation of a random sample $X_1, X_2, ..., X_n$ is made using the following algorithm:

1. Generate $U_i \sim \text{Uniform}(0,1), \ i = 1(1)n$.

2. If $\frac{\sum_{k=0}^{j-1} a_k \binom{r}{k} \frac{1}{\theta^k}}{\sum_{k=0}^{r} a_k \binom{r}{k} \frac{1}{\theta^k}} < U_i \leq \frac{\sum_{k=0}^{j} a_k \binom{r}{k} \frac{1}{\theta^k}}{\sum_{k=0}^{r} a_k \binom{r}{k} \frac{1}{\theta^k}}, \ j = 1, 2, ..., r,$ then set $X_i = V_i$, where $V_i \sim$ gamma$(j+1, \theta)$ and if $U_i \leq \frac{a_0}{\sum_{k=0}^{r} a_k \binom{r}{k} \frac{1}{\theta^k}}$, then set $X_i = V_i$, where $V_i \sim \exp(\theta)$.

The graphical representation of mean squared error (MSE) of the MLE and UMVUE of the PDF and the CDF of some BMOPPE distributions for different values of parameter based on simulation data is shown in Figure 3.

---

5. DATA ANALYSIS

In this section, we analyse three real data sets for comparing the performances of the MLE and the UMVUE of the PDF and the CDF. The probability model selection is made based on Akaike's Information Criterion (AIC = -2 log-likelihood $+ 2k$, where k is the number of parameters involved in the model) for each data set. Minimum is the AIC; better is the model fit. For our BMOPPE model, there is only one parameter θ; it is sufficient to select the model with a negative log-likelihood. But since this model is compared with other models with more than one parameter, the analyses are based on AIC. The AIC is calculated first using the MLE of the parameter(s), and the best model is selected. Then the AIC of the chosen model is compared with the AIC of this model calculated using the UMVUE of the PDF. It is mentioned that $a_0, \ a_1, \ ..., \ a_r$ and r are known constants in the model; in practice, these are chosen by trial and error such that AIC is at its minimum.

Data set-I, which is cited from Gross and Clark ([6]) and is given in Table 1, represents the relief times (in minutes) of 20 patients receiving an analgesic. We calculate the AIC of different standard distributions of One Parameter Polynomial Exponential (OPPE) family and few others from literature (presented in Table 4) and it is observed that BMOPPE family with $r = 2, \ a_0 = 0, \ a_1 = 0.1$ and $a_2 = 1$ is a better fit. The negative log-likelihood values of the selected model calculated using the MLE and the UMVUE of the PDF are presented in Table 7. Figures 4(a)–(b) show the histogram, the estimated PDF, and the CDF.

Data set-II represents the number of million revolutions before failure for each of the 23 ball bearings in the life test. It is obtained from Lawless ([7]) and shown in Table 2. For ease of calculation, we divide each observation into the data set by 2. The calculated AIC of different standard distributions of OPPE family and few others have been shown in Table 5 and it is noticed that BMOPPE family with $r = 2, \ a_0 = 0.2, \ a_1 = 0.1$ and $a_2 = 1$ is a better fit. The negative log-likelihood values of the selected model calculated using the MLE and the UMVUE of the PDF are also presented in Table 7. Figures 4(c)–(d) present the corresponding histogram, the estimated PDF, and the CDF.

Data set-III is a collection from Bjerkedal ([2]), and Table 3 displays the survival times (in days) of 72 guinea pigs infected with virulent tubercle bacilli. This data set has been fitted with a distribution of BMOPPE family with $r = 2, \ a_0 = 0.01, \ a_1 = 0.02$ and $a_2 = 4$ and it is found to be a good fit. AIC for this distribution and some other distributions available in the literature are listed in Table 6 that supports our claim. The negative log-likelihood values of the selected model calculated using the MLE and the UMVUE of the PDF are also presented in Table 7. The histogram, the estimated PDF, and the CDF have been shown in Figures 4(e)–(f).
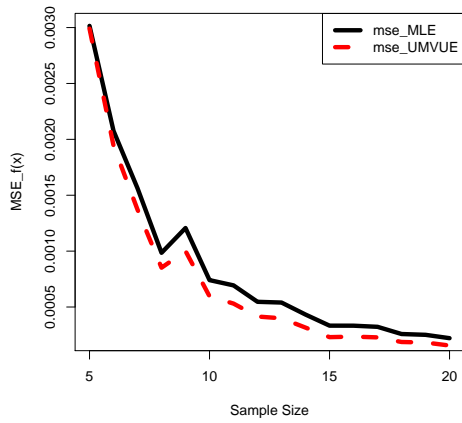
(a) For PDF at $\theta = 1.05$, $x = 2$, $a_0 = 0$, $a_1 = 0.1$, $a_2 = 1$ and $r = 2$.
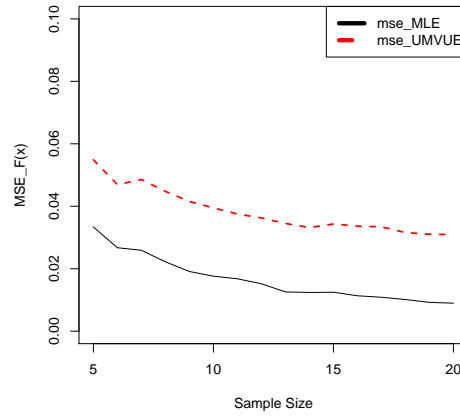
(b) For CDF at $\theta = 1.05$, $x = 2$, $a_0 = 0$, $a_1 = 0.1$, $a_2 = 1$ and $r = 2$.
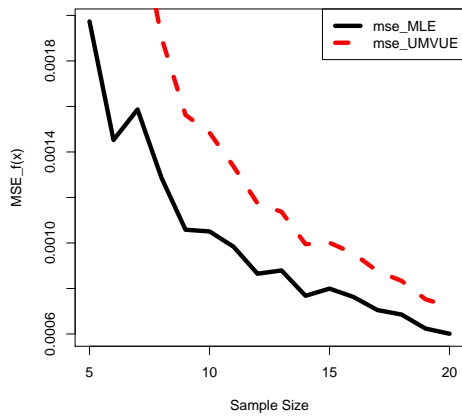
(c) For $\theta = 1.5$, $x = 2$, $a_0 = 0.01$, $a_1 = 0.02$, $a_2 = 4$ and $r = 2$.
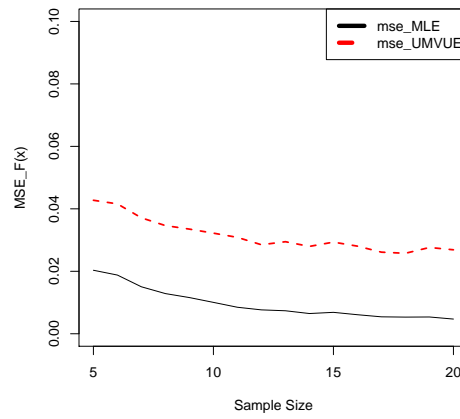
(d) For $\theta = 1.5$, $x = 2$, $a_0 = 0.01$, $a_1 = 0.02$, $a_2 = 4$ and $r = 2$.
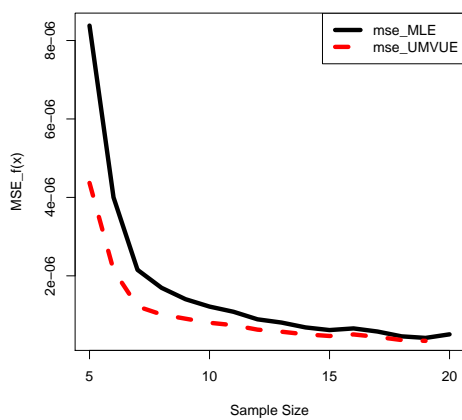
(e) For $\theta = 0.09$, $x = 2$, $a_0 = 0.2$, $a_1 = 0.1$, $a_2 = 1$ and $r = 2$.
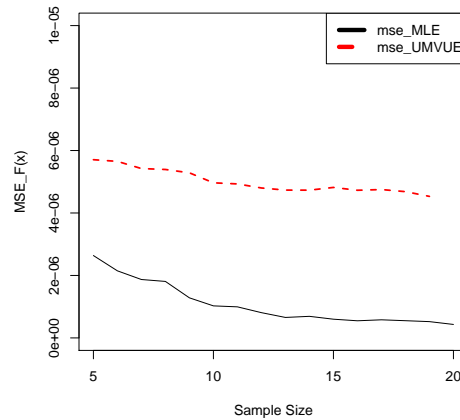
(f) For $\theta = 0.09$, $x = 2$, $a_0 = 0.2$, $a_1 = 0.1$, $a_2 = 1$ and $r = 2$.

Figure 3: Graph of simulated MSE of the MLE and UMVUE of the PDF and the CDF of BMOPPE distribution.
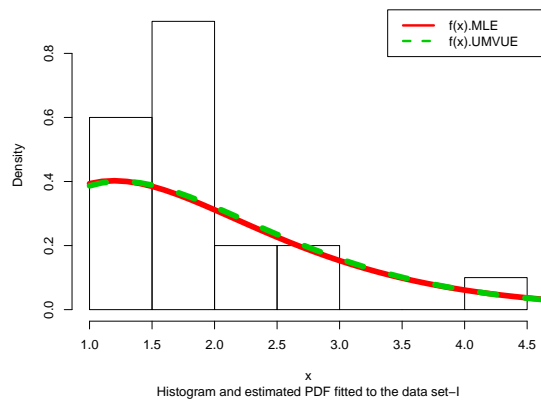
(a) Fitted PDF at $a_0 = 0.2$, $a_1 = 0.1$ and $a_2 = 1$ to the data set-I.
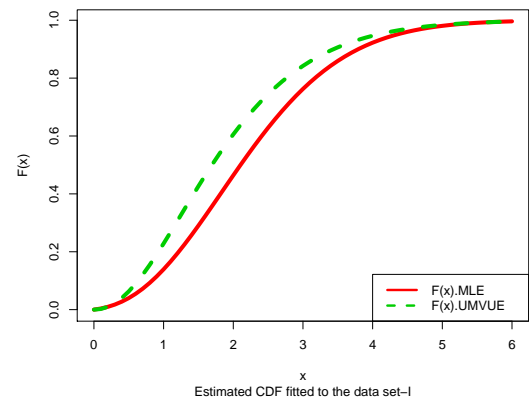
(b) Fitted CDF at $a_0 = 0.2$, $a_1 = 0.1$ and $a_2 = 1$ to the data set-I.
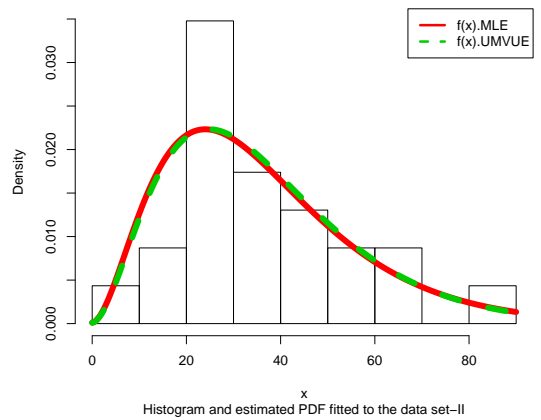
(c) Fitted PDF at $a_0 = 0$, $a_1 = 0.1$ and $a_2 = 1$ to the data set-II.
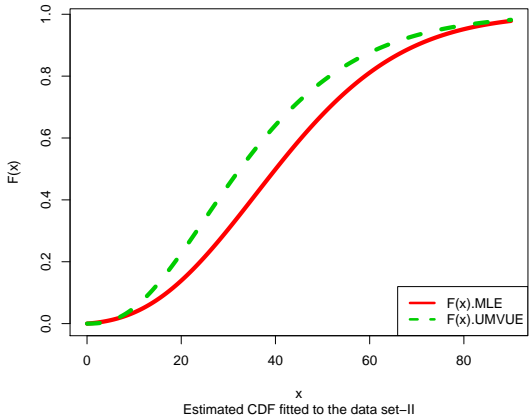
(d) Fitted CDF at $a_0 = 0$, $a_1 = 0.1$ and $a_2 = 1$ to the data set-II.
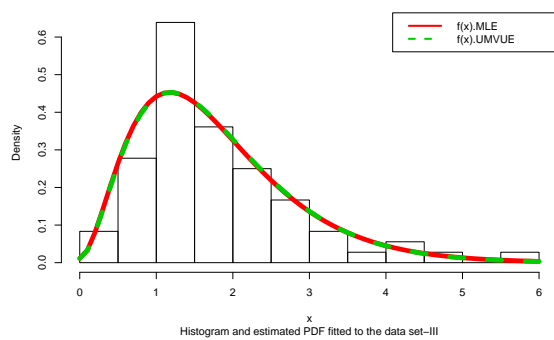
(e) Fitted PDF at $a_0 = 0.01$, $a_1 = 0.02$ and $a_2 = 4$ to the data set-III.
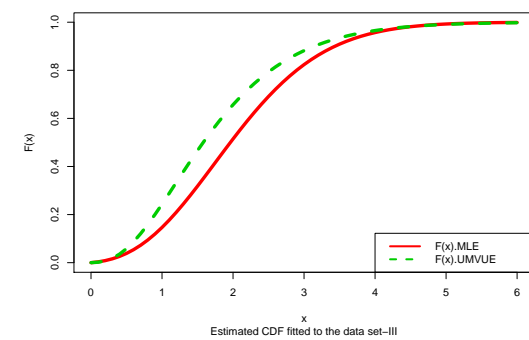
(f) Fitted CDF at $a_0 = 0.01$, $a_1 = 0.02$ and $a_2 = 4$ to the data set-III.

Figure 4: Graph of the estimated PDF and CDF of BMOPPE distribution for different data sets.

Table 1:   Relief times (in minutes) of 20 patients receiving an analgesic.

1.1	1.4	1.3	1.7	1.9	1.8	1.6	2.2	1.7	2.7
4.1	1.8	1.5	1.2	1.4	3	1.7	2.3	1.6	2

Table 2:   The number of million revolutions before failure
for each of the 23 ball bearings in the life tests.

17.88	28.92	33.00	41.52	42.12	45.60	48.80	51.84
51.96	54.12	55.56	67.80	68.44	68.64	68.88	84.12
93.12	98.64	105.12	105.84	127.92	128.04	173.40	

Table 3:   Survival times (in days) of 72 guinea pigs.

0.1	0.33	0.44	0.56	0.59	0.72	0.74	0.77	0.92
0.93	0.96	1	1	1.02	1.05	1.07	1.07	1.08
1.08	1.08	1.09	1.12	1.13	1.15	1.16	1.2	1.21
1.22	1.22	1.24	1.3	1.34	1.36	1.39	1.44	1.46
1.53	1.59	1.6	1.63	1.63	1.68	1.71	1.72	1.76
1.83	1.95	1.96	1.97	2.02	2.13	2.15	2.16	2.22
2.3	2.31	2.4	2.45	2.51	2.53	2.54	2.54	2.78
2.93	3.27	3.42	3.47	3.61	4.02	4.32	4.58	5.55

Table 4:   Model selection criterion for data set-I.

Model	-2 log-likelihood value	AIC
BMOPPE $(a_0 = 0, a_1 = 0.1$ and $a_2 = 1)$	48.10	50.10
Length-biased Lindley	49.70	51.70
Akash	59.52	61.52
Shanker	59.78	61.78
Lindley	60.50	62.50
Moment Exponential	52.32	54.32
Exponential	65.67	67.67

Table 5:   Model selection criterion for data set-II.

Model	-2 log-likelihood value	AIC
BMOPPE $(a_0 = 0.2, a_1 = 0.1$ and $a_2 = 1)$	195.26	197.26
Sujatha	195.38	197.38
Akash	227.06	229.06
Shanker	231.06	233.06
Lindley	231.47	233.47
Gamma	226.04	230.04
Weibull	232.27	236.27
Exponential	242.87	244.87

Table 6:   Model selection criterion for data set-III.

Model	-2 log-likelihood value	AIC
BMOPPE ($a_0 = 0.01$, $a_1 = 0.02$ and $a_2 = 4$)	188.18	190.18
OPPE ($a_0 = 0.01$, $a_1 = 0.02$ and $a_2 = 4$)	188.27	190.27
Lindley	213.85	215.85
New Generalized Lindley	188.36	194.36
Moment Exponential	208.40	210.40
Marshall–Olkin Exponential	206.36	210.36

Table 7:   Negative log-likelihood value using MLE and UMVUE fitted in data set I–III.

Data Set	Model	Negative log-likelihood value	
		MLE	UMVUE
I	BMOPPE ($a_0 = 0$, $a_1 = 0.1$ and $a_2 = 1$)	24.05	23.71
II	BMOPPE ($a_0 = 0.2$, $a_1 = 0.1$ and $a_2 = 1$)	97.63	97.52
III	BMOPPE ($a_0 = 0.01$, $a_1 = 0.02$ and $a_2 = 4$)	94.09	94.09

6.   CONCLUDING REMARKS

The article searches for a more generalised version of Lindley distribution. Starting with the Lindley distribution as a Bernoulli mixture of gamma distributions, a generalised binomial mixture of gamma distributions called the BMOPPE family of distributions has been derived. It is a revisit of the Lindley distribution from a different angle. As a result, the generalised version of the Lindley, the OPPE family of distributions, got mixed with the binomial probabilities, and therefore, the BMOPPE is an improvement. Moments and stochastic orderings are discussed. The process of generation of observations is pointed out, and the results are summarised. Estimations of the PDF and the CDF are discussed. The MLEs and UMVUEs are derived and compared. We have the estimators in biased (i.e., MLE) and unbiased (i.e., UMVUE) classes. Estimators of reliability functions are derived. Asymptotic variances of MLEs and variances of UMVUEs have been derived. The UMVUEs of the variance of UMVUEs of reliability functions have also been derived. These may be helpful in data analysis and comparison. Few data sets have been fitted, and it is found that the proposed distribution fits well in AIC sense. Even though the gain in AIC is minimal compared to the OPPE family of distributions, the BMOPPE is an improvement and is preferred.

ACKNOWLEDGMENTS

REFERENCES

[1] BARTOSZEWICZ, J. (1980). On the convergence of the Bhattacharya bounds in the multiparameter case, *Zastosowania Mathematica*, **16**, 601–608.

[2] BJERKEDAL, T. (1960). Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli, *American Journal of Hygiene*, **72**(1), 130–148.

[3] BLIGHT, B.J.N. and RAO, P.V. (1974). The convergence of the Bhattacharya bounds, *Biometrika*, **61**, 137–142.

[4] BOUCHAHED, L. and ZEGHDOUDI, H. (2018). A new and unified approach in generalizing the Lindley's distribution with applications, *Statistics in Transition*, **19**(1), 61–64.

[5] GHITANY, E.M.; AL-MUTAIRI, D.K.; BALAKRISHNAN, N. and AL-ENEZI, J.L. (2013). Power Lindley distribution and associated inference, *Computational Statistics & Data Analysis*, **64**, 20–33.

[6] GROSS, A.J. and CLARK, V.A. (1975). *Survival distributions: Reliability Applications in the Biometrical Sciences*, John Wiley, New York, USA.

[7] LAWLESS, J.F. (1982). *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, Inc., Hoboken, New Jersey.

[8] LINDLEY, D.V. (1958). Fiducial distributions and Bayes' theorem, *Journal of the Royal Statistical Society – Series B (Methodological)*, **20**(1), 102–107.

[9] MAITI, S.S.; RUIDAS, M.K. and ADHYA, S. (2022). Natural discrete one parameter polynomial exponential family of distributions and the application, *Annals of Data Science*. https://doi.org/10.1007/s40745-022-00422-8

[10] MUKHERJEE, I.; MAITI, S.S. and SHANKER, R. (2021). On estimation of the probability mass function and the cumulative distribution function of a natural discrete one parameter polynomial exponential distribution, *International Journal of Applied Mathematics and Statistics*, **60**(3), 49–61.

[11] MUKHERJEE, I.; MAITI, S.S. and SINGH, V.V. (2023). On estimation of the PDF and the CDF of the one-parameter polynomial exponential family of distributions, *Communications in Statistics – Theory and Methods*, **52**(1), 104–120.

Pitman's Measure of Closeness for Weighted Random Variables

Authors: MOSAYEB AHMADI
    – Department of Mathematics, University of Science and Technology of Mazandaran,
    Behshahr, Iran
    m.ahmadi@mazust.ac.ir , mbahmadi@gmail.com

    G.R. MOHTASHAMI BORZADARAN 🆔 ✉
    – Department of Statistics, Ferdowsi University of Mashhad,
    Mashhad, Iran
    grmohtashami@um.ac.ir

    MOSTAFA RAZMKHAH 🆔
    – Department of Statistics, Ferdowsi University of Mashhad,
    Mashhad, Iran
    razmkhah_m@um.ac.ir

Abstract:

• Statistical inference based on weighted random variables is developed in the sense of Pitman's
measure of closeness. Some general formulas are presented to compute the Pitman closeness of
two weighted random variables to the parameter of interest. A new general weighted model is
also introduced and some properties are investigated. Also, the concept of Pitman's measure of
closeness is used for measuring the nearness of some weighted random variables with respect to
each other. The results are illustrated using some real data sets. Eventually, some conclusions are
stated.

✉ Corresponding author

1.   INTRODUCTION

There are situations in which the observed data would not be modeled by a proper stochastic model. In these cases, some general models leading to weighted distributions may be used. The concept of weighted distributions has been introduced by Fisher [17] and Rao [35] in connection with modeling statistical data in situation that the standard distribution was not appropriate for their purposes. A formal definition of a weighted distribution for random variable X with density (mass) function $f(x)$ and non-negative weight $w(x)$ is obtained by $g(x) = \frac{w(x)f(x)}{E(w(X))}$, where $E(w(X)) > 0$. The corresponding weighted random variable is denoted by X_w in this paper. For more details about the statistical applications of weighted distributions, we refer the readers to Patil and Rao [31, 32] and Rao [38]. From Gupta and Keating [20] to Bartoszewicz [12] and the references therein, the idea of the weighted distributions is developed to various applications and properties. Saghir *et al.* [44] carried out a brief review of weighted distributions, and investigated the implications of the differing weight models as well as characterizations of these distributions based on a simple relationship between two truncated moments. In recent years, this concept has been applied in many areas of statistics such as biomedical, human study, wildlife populations, electronic, etc. For example, Jiang [21] used the weight function distribution to analyze measurement errors of an electromagnetic flowmeter. It is of great importance to note that the models of ordered data are special cases of weighted distributions, for example the i-th order statistic in sample of size n, denoted by $X_{i:n}$ $1 \leq i \leq n$, from a population with probability density function (pdf) $f(\cdot)$ and cumulative distribution function (cdf) $F(\cdot)$ is a weighted random variable with weight function $w(x) = (F(x))^{i-1}(\bar{F}(x))^{n-i}$. The n-th upper k-record and the n-th lower k-record are also weighted random variables with weight functions $w(x) = \left[-\log \bar{F}(x)\right]^{n-1}(\bar{F}(x))^{k-1}$ and $w(x) = [-\log F(x)]^{n-1}F(x)^{k-1}$, respectively. For more details and literature on order statistics and record values, see for example, Arnold *et al.* [6, 7] and David and Nagaraja [14]. The skew distributions are also another kind of weighted distributions. Recently, Gómez-Déniz *et al.* [19] investigated the properties and applications of a new family of skew distributions. Azzalini [8] has done an overview on the progeny of the skew-normal family.

It is known that the theory of estimation is a fundamental discipline dealing with the specification of probabilistic model in terms of observed data. There are many measures to evaluate the performance of an estimator like mean squared error (MSE), mean absolute deviation, etc. The Pitman's measure of closeness (PMC) introduced by Pitman [33] is another criterion which has been used by several authors to compare the performance of the estimators. Let us first recall the formal definition:

Definition 1.1.   Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two estimators of a common parameter θ. The Pitman's measure of closeness of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is denoted by $\pi(\hat{\theta}_1, \hat{\theta}_2|\theta)$ and defined as

$$(1.1) \qquad \pi(\hat{\theta}_1, \hat{\theta}_2|\theta) = \Pr(|\hat{\theta}_1 - \theta| < |\hat{\theta}_2 - \theta|), \quad \forall \, \theta \in \Theta,$$

where Θ is the parameter space. If $\Pr(\hat{\theta}_1 = \hat{\theta}_2) = 0$ and $\pi(\hat{\theta}_1, \hat{\theta}_2|\theta) \geq 1/2$ for all $\theta \in \Theta$, with strict inequality holding for at least one θ, then $\hat{\theta}_1$ is said to be a Pitman closer estimator than $\hat{\theta}_2$ with respect to θ.

It is obvious that $\pi(\hat{\theta}_1, \hat{\theta}_2 | \theta)$ determines the relative frequency that the estimator $\hat{\theta}_1$ is closer than $\hat{\theta}_2$ to the true but unknown value of the parameter θ. Note that the absolute difference in (1.1) may be replaced by any other loss function, in this case the associated probability is called generalized PMC. Several developments in estimation theory via PMC arguments have been written by Efron [16]. Also, Rao [36, 37] concluded noticeable comments about the necessity of inference based on PMC. Keating [23] derived various aspects of PMC. Keating and Mason [24] provided many practical examples in which the PMC may be more useful than MSE. Sen [45] presented the condition on variance for an estimator being closer than another. Ghosh and Sen [18] have shown under certain conditions a median unbiased estimator of parameter is the Pitman closest within a certain class of estimators. Some characterization results in statistical inference and decision theory with examples have been explained by Lee [27] in view of the PMC for location and scale families. Nayak [30] used the PMC to find best estimators of a location or scale parameter. Many relevant references can be found for example in the monograph by Keating *et al.* [25]. A useful discussion to the question "Is Pitman closeness a reasonable criterion?" with comments and rejoinder about this measure can be found in Robert *et al.* [43]. Kourouklis [26] used the PMC to get an improved estimation.

It is of great importance to note that two various statistics in estimating a common parameter would be considered as two weighted random variables. This motivated us to study the PMC in view of the weighted distributions. Toward this end, the PMC of weighted random variables with respect to any unknown parameter is first investigated. A new general weighted model is also introduced and some properties are investigated. Then, the concept of PMC is used to measure the nearness of some weighted random variables with respect to each other.

The rest of the paper is organized as follows: In Section 2, some general results are given to compute the PMC of the weighted random variables to the parameter of interest. In Section 3, a new general weighted model is introduced and some numerical results and conclusions are presented. A real data set is used in this section to illustrate the proposed procedure. In Section 4, the PMC of two weighted random variables with respect to another one is investigated. Moreover, the results are applied to another real data set. Finally, some conclusions are presented in Section 5.

2. GENERAL RESULTS

Let X be a random variable with pdf $f(\cdot)$ and cdf $F(\cdot)$ which is defined on finite interval $[a, b]$. Assume X_{w_1} and X_{w_2} are two independent weighted random variables of X with weight functions $w_1(\cdot)$ and $w_2(\cdot)$, respectively. Again, it is pointed out that these random variables may be considered as two various estimators for a common parameter θ. Based on (1.1), the PMC of X_{w_1} and X_{w_2} with respect to θ is given by

$$(2.1) \qquad \pi(X_{w_1}, X_{w_2} | \theta) = P(|X_{w_1} - \theta| < |X_{w_2} - \theta|).$$

A general expression for the probability in (2.1) can be written as follows:

$$
\begin{aligned}
\pi(X_{w_1}, X_{w_2}|\theta) &= P(X_{w_1} < X_{w_2}, X_{w_1} + X_{w_2} > 2\theta) + P(X_{w_1} > X_{w_2}, X_{w_1} + X_{w_2} < 2\theta) \\
&= \int_{\max\{a,2\theta-b\}}^{\theta} \int_{2\theta-x}^{b} g_1(x)g_2(y)dydx + \int_{\theta}^{b} \int_{x}^{b} g_1(x)g_2(y)dydx \\
&\quad + \int_{a}^{\theta} \int_{a}^{x} g_1(x)g_2(y)dydx + \int_{\theta}^{\min\{b,2\theta-a\}} \int_{a}^{2\theta-x} g_1(x)g_2(y)dydx \\
&= \int_{\max\{a,2\theta-b\}}^{\theta} \bar{G}_2(2\theta - x)g_1(x)dx + \int_{\theta}^{b} \bar{G}_2(x)g_1(x)dx \\
&\quad + \int_{a}^{\theta} G_2(x)g_1(x)dx + \int_{\theta}^{\min\{b,2\theta-a\}} G_2(2\theta - x)g_1(x)dx,
\end{aligned}
$$

(2.2)

where for $i = 1, 2$, the functions $g_i(x)$ and $G_i(x)$ are respectively the pdf and cdf of X_{w_i}. It can be shown that $\bar{G}_i(x) = 1 - G_i(x) = \frac{B_i(x)}{E(w_i(X))}$, where

$$
B_i(x) = \bar{F}(x)E(w_i(X)|X > x) = \int_{x}^{\infty} w_i(t)f(t)dt.
$$

Therefore, we have

$$
\begin{aligned}
\pi(X_{w_1}, X_{w_2}|\theta) &= \int_{\max\{a,2\theta-b\}}^{\theta} \frac{B_2(2\theta - x)}{E(w_2(X))} \frac{w_1(x)f(x)}{E(w_1(X))}dx + \int_{\theta}^{b} \frac{B_2(x)}{E(w_2(X))} \frac{w_1(x)f(x)}{E(w_1(X))}dx \\
&\quad + \int_{a}^{\theta} \left(1 - \frac{B_2(x)}{E(w_2(X))}\right) \frac{w_1(x)f(x)}{E(w_1(X))}dx \\
&\quad + \int_{\theta}^{\min\{b,2\theta-a\}} \left(1 - \frac{B_2(2\theta - x)}{E(w_2(X))}\right) \frac{w_1(x)f(x)}{E(w_1(X))}dx.
\end{aligned}
$$

(2.3)

In special case of $w_i(x) = \varphi_i\big(F(x)\big)$ $(i = 1, 2)$, by transforming $u = F(x)$, we get

$$
\begin{aligned}
\pi(X_{w_1}, X_{w_2}|\theta) &= \frac{1}{\gamma_1\gamma_2}\left\{ \int_{F(\max\{a,2\theta-b\})}^{F(\theta)} B_2(2\theta - F^{-1}(u))\varphi_1(u)du + \int_{F(\theta)}^{1} B_2(F^{-1}(u))\varphi_1(u)du \right\} \\
&\quad - \frac{1}{\gamma_1\gamma_2}\left\{ \int_{0}^{F(\theta)} B_2(F^{-1}(u))\varphi_1(u)du \right. \\
&\quad \left. + \int_{F(\theta)}^{F(\min\{b,2\theta-a\})} B_2(2\theta - F^{-1}(u))\varphi_1(u)du \right\} \\
&\quad + \frac{1}{\gamma_1}\left\{ \int_{0}^{F(\theta)} \varphi_1(u)du + \int_{F(\theta)}^{F(\min\{b,2\theta-a\})} \varphi_1(u)du \right\},
\end{aligned}
$$

(2.4)

where $\gamma_i = E[\varphi_i(U)]$ $(i = 1, 2)$, such that U is a Uniform$(0, 1)$ random variable. Table 1 shows some special cases of $\varphi_i(u)$ $(i = 1, 2)$, which have been previously studied by several authors.

The definition of the weighted random variables may be extended to a random sample of size n. Let $X_1, X_2, ..., X_n$ be independent and identically distributed (iid) random variables with cdf $F(\cdot)$ and pdf $f(\cdot)$, then a weighted version of this sample can be defined by $(X_1^*, X_2^*, ..., X_n^*)$, which have the joint pdf

(2.5)
$$
h_{X_1^*, X_2^*, ..., X_n^*}(x_1^*, x_2^*, ...x_n^*) = \frac{w(x_1^*, x_2^*, ...x_n^*)}{E[w(X_1^*, X_2^*, ..., X_n^*)]} \prod_{i=1}^{n} f(x_i^*).
$$

Table 1: Some special cases of $\varphi_i(u)$ related to PMC.

References	$\varphi_1(u)$	$\varphi_2(u)$	Descriptions
Ahmadi and Raqab [4]	$u^{i-1}(1-u)^{m-i+1}$	$u^{j-1}(1-u)^{n-j+1}$	Order statistics in two samples
Raqab and Ahmadi [39]	$[-\log(1-u)]^i$	$[-\log(1-u)]^j$	Record values from two sequences
Volterman *et al.* [47]	$\sum_{l=1}^i a_l^R(i)(1-u)^{\gamma_l^R-1}$	$\sum_{l=1}^j a_l^S(j)(1-u)^{\gamma_l^S-1}$	Two-sample progressive type-II censoring
Ahmadi and Mohtashami Borzadaran [5]	$[-\log(1-u)]^m$	$u^{i-1}(1-u)^{n-i+1}$	Record values and order statistics

Notice that the joint pdf in (2.5) is a weighted version of $\prod_{i=1}^n f(x_i^*)$. Now, for the special case of $n=2$, this joint distribution can be expressed as

$$(2.6) \qquad h_{X_1^*,X_2^*}(x,y) = \frac{w(x,y)}{E[w(X_1^*,X_2^*)]} f(x)f(y),$$

where include the joint distribution of two order statistics in a finite random sample or two k-records in a sequence of iid random variables. By assuming $w(x,y)=\varphi(u,v)$, where $u=F(x)$ and $v=F(y)$, the PMC of X_1^* and X_2^* with respect to parameter θ can be determined as follows:

$$\pi(X_1^*,X_2^*|\theta) = \frac{1}{A}\Bigg\{ \int_{F(\max(a,2\theta-b))}^{F(\theta)} \int_{F(2\theta-F^{-1}(u))}^1 \varphi(u,v)\,dv\,du$$

$$+ \int_{F(\theta)}^1 \int_u^1 \varphi(u,v)\,dv\,du + \int_0^{F(\theta)} \int_0^u \varphi(u,v)\,dv\,du$$

$$(2.7) \qquad + \int_{F(\theta)}^{F(\min(b,2\theta-a))} \int_0^{F(2\theta-F^{-1}(u))} \varphi(u,v)\,dv\,du \Bigg\},$$

where $A = E[\varphi(F(X_1^*),F(X_2^*))]$. Some special cases of the weighted model in (2.6) have been previously studied in view of PMC by some authors which are summarized in Table 2.

Table 2: Some special cases of $\varphi(u,v)$ related to PMC.

References	$\varphi(u,v)$	Descriptions
Balakrishnan *et al.* [11]	$u^{m-1}(v-u)^{i-m-1}(1-v)^{n-i}$	Sample median and the i-th order statistic
Balakrishnan *et al.* [9]	$u^{i-1}(v-u)^{j-i-1}(1-v)^{n-i}$	Two order statistics in one sample
Ahmadi and Balakrishnan [1]	$\frac{[-\log(1-u)]^i}{1-u}\left[-\log(\frac{1-v}{1-u})\right]^{j-i-1}$	Two upper records in one sequence
Ahmadi and Balakrishnan [3]	$[-\log(u+1-v)]^{i-1}$	The i-th lower and upper records in one sequence
Ahmadi and Balakrishnan [2]	$[-\log(1-u)]^i\left[-\log(\frac{1-v}{1-u})\right]^{j-i-1}\frac{(1-v)^{k-1}}{1-u}$	Two k-records in one sequence

Remark 2.1. It is worthwhile to note that not only for the special cases presented in Tables 1 and 2, but also the PMC of the general versions of weighted random variables with respect to the parameter of interest can be derived by the use of (2.4) and (2.7).

---

3.   A GENERAL WEIGHTED MODEL

In Section 2, a weighted random variable whose weight function was a function of the baseline cdf was considered in some special cases. Here, we introduce a more general model contains all previous ones. Let X be a continuous random variable with pdf $f(\cdot)$ and cdf $F(\cdot)$. For positive real constants α, β, γ and δ, one may consider a general weighted random variable with the following pdf

$$(3.1) \qquad g_F(x; \lambda) = \frac{1}{M(\lambda)} [F(x)]^{\alpha-1} [\bar{F}(x)]^{\gamma-1} [-\log F(x)]^{\beta-1} [-\log \bar{F}(x)]^{\delta-1} f(x),$$

where $\lambda = (\alpha, \gamma, \beta, \delta)$ and $M(\lambda)$ is the normalizing constant which is given by

$$
\begin{aligned}
M(\lambda) &= M(\alpha, \gamma, \beta, \delta) \\
(3.2) \qquad &= \int_0^1 u^{\alpha-1}(1-u)^{\gamma-1}(-\log u)^{\beta-1}[-\log(1-u)]^{\delta-1} du.
\end{aligned}
$$

The model defined in (3.1) includes many famous families of distributions, such as distribution of order statistics (see David and Nagaraja, [14]), distributions of upper and lower k-records (see, Arnold *et al.*, [6]), Jones model (see Jones, [22]) and proportional (reversed) hazard rate model. This model can be also considered as the pdf of a weighted k-record statistic and also weighted order statistics. In the next two subsection we consider two famous member of the proposed model.

3.1. Results based on records

Let us recall the sequences of upper k-record times, $T_{n,k}$, and upper k-record values, $R_{n,k}$, which are defined as follows: $T_{1,k} = k$ with probability one, $R_{1,k} = X_{1:k}$ and for $n \geq 2$

$$T_{n,k} = \min\{j : j > T_{n-1,k}, \ X_j > X_{T_{n-1,k}-k+1:T_{n-1,k}}\},$$

and the n-th upper k-record value is defined by $R_{n,k} = X_{T_{n,k}-k+1:T_{n,k}}$, for $n \geq 1$. For $k = 1$, the ordinary records are recovered. Then, the pdf of $R_{n,k}$ is given by

$$f_{n,k}(x) = \frac{k^n}{\Gamma(n)} \big[-\log \bar{F}(x)\big]^{n-1} (\bar{F}(x))^{k-1} f(x),$$

where $\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx$ stands for the complete gamma function; see Arnold *et al.* [6] for more details. It is obvious that by taking $\gamma = k$ and $\delta = n$, the pdf in (3.1) can be interpreted as the pdf of a weighted the n-th upper k-record statistic, which is denoted by $R_{n,k}^w$. Similarly, the weighted lower k-records may be defined. The first question arises here is that whether k-records or the weighted version ones are closer to the parameter of interest. By using (2.4)

and performing some algebraic calculations, it can be shown that the PMC of $R_{n,k}^w$ and $R_{n,k}$ with respect to θ is

$$
\begin{aligned}
\pi_{\alpha,\beta}(R_{n,k}^w, R_{n,k}|\theta) &= \frac{1}{M(\alpha,\beta,k,n)} \sum_{j=\beta-1}^{\infty} \sum_{r=0}^{\alpha-1} \binom{\alpha-1}{r}(-1)^r C_j(\beta-1) \\
&\times \Bigg\{ \frac{\Gamma\big(n, -(r+j+k)\log \bar{F}(\max(a, 2\theta - b))\big)}{(r+j+k)^n} \\
&+ \sum_{i=n}^{\infty} \frac{k^i}{i!} \left(\frac{\Gamma(n) - 2\Gamma\big(n, -(r+j+2k)\log \bar{F}(\theta)\big)}{(n+j+2k)^{n+i}} \right. \\
&+ \int_{-\log \bar{F}(\theta)}^{-\log \bar{F}(\min(b, 2\theta-a))} \psi(y; i, j, r)dy \\
&- \left. \int_{-\log \bar{F}(\max(a, 2\theta-b))}^{-\log \bar{F}(\theta)} \psi(y; i, j, r)dy \right) \Bigg\},
\end{aligned}
$$

(3.3)

where $\Gamma(n, a) = \int_a^{\infty} x^{n-1} e^{-x} dx$ stands for the incomplete gamma function, the function $M(\alpha, \beta, k, n)$ is as defined in (3.2), $C_j(n)$ is the coefficient of w^j in the expansion of $(\sum_{i=1}^{\infty} \frac{w^i}{i})^n$, and

$$
\psi(y; i, j, r) = \big(-\log \bar{F}(2\theta - F^{-1}(1 - e^{-y}))\big)^i \big(\bar{F}(2\theta - F^{-1}(1 - e^{-y}))\big)^k y^{n-1} e^{-(r+j+k)y}.
$$

Remark 3.1. It is clear that for other different versions of weight functions, we can express the PMC via similar arguments in (3.3), for which they can be interpreted as applications and specifications of several forms.

In the rest of this section, the population quantile is considered as the parameter of interest θ. We recall that the population quantile ξ_p of order p $(0 < p < 1)$ of the cdf $F(\cdot)$ is defined by $\xi_p = \inf\{x : F(x) \geq p\}$. Balakrishnan *et al.* [11, 10] determined the closest order statistic in a random sample of size n to a specific population quantile and specially studied the PMC of sample median to population median. Ahmadi and Balakrishnan [1] examined the PMC of record statistics to the population quantiles of location-scale family of distributions. Moreover, Ahmadi and Balakrishnan [2] investigated the PMC of k-records and Ahmadi and Balakrishnan [3] obtained the PMC of current records for location-scale families. Similar work was done by Razmkhah and Ahmadi [41] regarding the current k-records. The PMC of upper (lower) records in two independent sequences of iid continuous random variables to population quantiles was studied by Raqab and Ahmadi [39]. Similarly, the PMC of order statistics in a two-sample problem was investigated by Ahmadi and Raqab [4]. Moreover, a comparison study for order statistics and records was performed with some remarks by Ahmadi and Mohtashami Borzadaran [5]. Davies [15] studied some PMC results for type-I hybrid censored data from exponential distribution. Morabbi and Razmkhah [29] used the PMC to get the quantile estimation based on modified ranked set sampling schemes. The weighted versions of the aforementioned papers, for example, weighted order statistics and weighted k-records may be of great interest which are discussed in this paper.

From (3.3), it is seen that the PMC is not distribution-free. So, the baseline distribution has to be specified. Therefore, for as an example, the standard exponential distribution has been considered and the PMC investigated for the population quantiles. With this in mind, let $\{X_i, i \geq 1\}$ be a sequence of iid random variables from standard exponential distribution.

The PMC of $R_{n,k}^w$ and $R_{n,k}$ with respect to the p-th quantile of this distribution, can be simplified as follows:

$$\pi_{\alpha,\beta}(R_{n,k}^w, R_{n,k}|\xi_p) = \frac{1}{M(\alpha,\beta,k,n)} \sum_{j=\beta-1}^{\infty} \sum_{r=0}^{\alpha-1} \binom{\alpha-1}{r}(-1)^r C_j(\beta-1)$$

$$\times \left\{ \frac{\Gamma(n)}{(r+j+k)^n} + \sum_{i=n}^{\infty} \frac{k^i}{i!} \left(\frac{\Gamma(n) - 2\Gamma(n, -(r+j+2k)\log q)}{(r+j+2k)^{n+i}} \right. \right.$$

$$\left. \left. + q^{2k} \int_{-\log q}^{-\log q^2} h(y;i,j,r)dy - q^{2k}\int_0^{-\log q} h(y;i,j,r)dy \right) \right\},$$

(3.4)

where

$$h(y;i,j,r) = \left(-\log(q^2 e^y)\right)^i y^{n-1} e^{-(r+j)y}.$$

Using (3.4), the numerical values of $\pi_{\alpha,\beta}(R_{n,k}^w, R_{n,k}|\xi_p)$ have been computed for $n = 3, 4$, $k = 3, 4$ and some selected values of α, β and p. The results are presented in Table 3.

Table 3: Values of $\pi_{\alpha,\beta}(R_{n,k}^{w,U}, R_{n,k}^U|\xi_p)$ for standard exponential distribution.

n	k	β	α	p								
				0.10	0.25	0.5	0.60	0.75	0.80	0.90	0.95	0.99
3	3	1	2	0.4036	0.4044	0.4439	0.4899	**0.5654**	**0.5819**	**0.5955**	**0.5964**	**0.5964**
			3	0.3310	0.3315	0.3805	0.4540	**0.5989**	**0.6347**	**0.6667**	**0.6689**	**0.6690**
			4	0.2754	0.2757	0.3213	0.4077	**0.6099**	**0.6664**	**0.7204**	**0.7244**	**0.7246**
			5	0.2321	0.2322	0.2702	0.3591	**0.6056**	**0.6827**	**0.7613**	**0.7676**	**0.7679**
		2	1	**0.6789**	**0.6745**	**0.5730**	0.4839	0.3626	0.3397	0.3222	0.3212	0.3211
			2	**0.5748**	**0.5759**	**0.5727**	**0.5437**	0.4675	0.4463	0.4267	0.4253	0.4252
			3	0.4888	0.4902	**0.5363**	**0.5613**	**0.5467**	**0.5319**	**0.5130**	**0.5113**	**0.5112**
			4	0.4184	0.4194	0.4838	**0.5497**	**0.6030**	**0.5985**	**0.5838**	**0.5816**	**0.5815**
			5	0.3609	0.3614	0.4273	**0.5201**	**0.6400**	**0.6491**	**0.6414**	**0.6392**	**0.6391**
4	3	1	2	0.4291	0.4292	0.4371	0.4560	**0.5166**	**0.5392**	**0.5674**	**0.5706**	**0.5709**
			3	0.3720	0.3720	0.3812	0.4088	**0.5149**	**0.5596**	**0.6200**	**0.6275**	**0.6280**
			4	0.3254	0.3254	0.3335	0.3633	**0.5012**	**0.5663**	**0.6609**	**0.6737**	**0.6746**
			5	0.2869	0.2870	0.2933	0.3219	0.4800	**0.5631**	**0.6928**	**0.7116**	**0.7130**
		2	1	**0.6884**	**0.6881**	**0.6554**	**0.5937**	0.4321	0.3790	0.3183	0.3120	0.3116
			2	**0.6123**	**0.6123**	**0.6099**	**0.5914**	0.4988	0.4559	0.3959	0.3883	0.3877
			3	**0.5461**	**0.5462**	**0.5566**	**0.5678**	**0.5440**	**0.5163**	0.4632	0.4546	0.4539
			4	0.4890	0.4890	**0.5032**	**0.5325**	**0.5711**	**0.5618**	**0.5211**	**0.5118**	**0.5110**
			5	0.4397	0.4397	0.4535	0.4921	**0.5835**	**0.5943**	**0.5708**	**0.5613**	**0.5603**
	4	1	2	0.4167	0.4169	0.4456	0.4889	**0.5612**	**0.5745**	**0.5830**	**0.5833**	**0.5833**
			3	0.3498	0.3500	0.3871	0.4577	**0.5992**	**0.6291**	**0.6495**	**0.6501**	**0.6502**
			4	0.2959	0.2960	0.3318	0.4166	**0.6189**	**0.6674**	**0.7028**	**0.7040**	**0.7041**
			5	0.2522	0.2523	0.2830	0.3722	**0.6246**	**0.6927**	**0.7456**	**0.7477**	**0.7477**
		2	1	**0.6601**	**0.6589**	**0.5828**	0.4942	0.3715	0.3518	0.3402	0.3399	0.3399
			2	**0.5715**	**0.5718**	**0.5667**	**0.5370**	0.4599	0.4417	0.4289	0.4285	0.4285
			3	0.4949	0.4953	**0.5269**	**0.5485**	**0.5310**	**0.5179**	**0.5056**	**0.5051**	**0.5050**
			4	0.4297	0.4299	0.4763	**0.5371**	**0.5855**	**0.5809**	**0.5710**	**0.5703**	**0.5703**
			5	0.3743	0.3744	0.4232	**0.5103**	**0.6250**	**0.6317**	**0.6264**	**0.6257**	**0.6257**
		3	1	**0.7803**	**0.7753**	**0.5810**	0.4165	0.2483	0.2289	0.2198	0.2197	0.2197
			2	**0.6988**	**0.6983**	**0.6137**	0.4973	0.3368	0.3136	0.3015	0.3012	0.3012
			3	**0.6226**	**0.6231**	**0.6117**	**0.5510**	0.4180	0.3926	0.3778	0.3774	0.3774
			4	**0.5537**	**0.5542**	**0.5856**	**0.5789**	0.4892	0.4639	0.4468	0.4463	0.4463

The bold numbers in Table 3 are referring to the fact that the weighted k-records are Pitman closer than the usual k-records to the specific population quantile. From this table, it is intuitively observed that for given α and β, there exist real values α_0 and $p_0 \in (0,1)$ such that for $\alpha \le \alpha_0$ and $p \le p_0$, $\pi(R_{n,k}^{U,w}, R_{n,k}^{U}|\xi_p) > 0.5$. Similarly for $\alpha > \alpha_0$, there exist a real value $p_* \in (0,1)$, so that for $p \ge p_*$, $\pi(R_{n,k}^{U,w}, R_{n,k}^{U}|\xi_p) > 0.5$.

3.2. Results based on order statistics

We recall that the pdf of the r-th order statistic from an iid sample of size n from cdf $F(\cdot)$ and pdf $f(\cdot)$ is given by

$$(3.5) \qquad f_{r:n}(x) = k \binom{n}{r} F^{r-1}(x) \bar{F}^{n-r}(x) f(x).$$

Let us denote the weighted order statistic by $X_{r:n}^w$. Here, the PMC of $X_{r:n}^w$ and $X_{r:n}$ with respect to the p-th quantile ξ_p is studied when the baseline distribution is standard exponential distribution. From (2.4) and (3.5), we find

$$
\begin{aligned}
\pi_{\beta,\delta}(X_{r:n}^w, X_{r:n}|\xi_p) &= \frac{1}{M(r,n-r+1,\beta,\delta)} \Bigg\{ \sum_{j=0}^{r-1} \sum_{k=1}^{\infty} (-1)^j \binom{r-1}{j} C_k(\beta-1) \\
&\times \left(\frac{\Gamma(\delta)}{(j+n-r+k+1)} - \frac{1}{M(r,n-r+1,1,1)} \right. \\
&\times \sum_{i=0}^{n-r} \sum_{l=0}^{r+i} \left(\frac{(-1)^{i+l} \binom{n-r}{i}\binom{r+i}{l}(1-p)^{2l}}{r+i} \right. \\
&\times \left. \left. \frac{\Gamma(\delta) - \Gamma\big(\delta, -(j+n-r+k-l+1)\ln(1-p)\big)}{(j+n-r+k-l+1)^\delta} \right) \right) \\
&+ \frac{1}{M(r,n-r+1,1,1)} \sum_{i=0}^{n-r} \sum_{j=0}^{2r+i-1} \sum_{k=1}^{\infty} \left(\frac{(-1)^{i+j} \binom{n-r}{i}\binom{2r+i-1}{j}}{r+i} \right. \\
&\times \left. C_k(\beta-1) \frac{\Gamma(\delta) - 2\Gamma\big(\delta, -(j+n-r+k+1)\ln(1-p)\big)}{(j+n-r+k+1)^\delta} \right) \Bigg\}.
\end{aligned}
$$

(3.6)

For extreme order statistics $X_{1:n}$ and $X_{n:n}$ we have

$$
\begin{aligned}
\pi_{\beta,\delta}(X_{1:n}^w, X_{1:n}|\xi_p) &= \frac{1}{M(1,n,\beta,\delta)} \sum_{k=1}^{\infty} C_k(\beta-1) \Bigg\{ \frac{\Gamma(\delta)}{(n+k)} \\
&+ n \sum_{i=0}^{n-1} \sum_{j=0}^{i+1} \frac{(-1)^{i+j} \binom{n-1}{i}\binom{i+1}{j}}{i+1} \\
&\times \left(\frac{\Gamma(\delta) - 2\Gamma\big(\delta, -(n+k+j)\ln(1-p)\big)}{(n+k+j)^\delta} \right. \\
&\left. - \frac{\Gamma(\delta) - \Gamma\big(\delta, -(n+k-j)\ln(1-p)\big)}{(n+k-j)^\delta} (1-p)^{2j} \right) \Bigg\}
\end{aligned}
$$

and

$$
\begin{aligned}
\pi_{\beta,\delta}(X_{n:n}^{w}, X_{n:n}|\xi_p) \;=\; & \frac{1}{M(n,1,\beta,\delta)} \sum_{k=1}^{\infty} C_k(\beta-1)\Bigg\{ \sum_{j=0}^{n-1}(-1)^j \binom{n-1}{j} \frac{\Gamma(\delta)}{(j+k+1)} \\
& + \sum_{j=0}^{2n-1}(-1)^j \binom{2n-1}{j} \frac{\Gamma(\delta) - 2\Gamma\big(\delta, -(j+k+1)\ln(1-p)\big)}{(j+k+1)^\delta} \\
& - n \sum_{i=0}^{n-r}\sum_{j=0}^{n-1}\sum_{l=0}^{n} \Bigg(\frac{(-1)^{j+l}\binom{n-1}{j}\binom{n}{l}(1-p)^{2l}}{n+i} \\
& \times \frac{\Gamma(\delta) - \Gamma\big(\delta, -(j+k-l+1)\ln(1-p)\big)}{(j+k-l+1)^\delta} \Bigg) \Bigg\},
\end{aligned}
$$

respectively. Using (3.6), the numerical values of $\pi_{\beta,\delta}(X_{r:n}^{w}, X_{r:n}|\xi_p)$ have been computed for some selected values of n, r, β, δ and p. The results are presented in Table 4. The bold numbers in this table are referring to the fact that the weighted order statistics are Pitman closer than the usual order statistics to the specific population quantile.

Table 4:   Values of $\pi_{\beta,\delta}(X_{r:n}^{w}, X_{r:n}|\xi_p)$ for standard exponential distribution.

n	r	β	δ	p				
				0.10	0.25	0.5	0.75	0.90
5	1	2	3	0.2326	**0.5067**	**0.7858**	**0.8018**	**0.8019**
		3	2	**0.5243**	**0.5585**	**0.5299**	**0.5276**	**0.5275**
	3	2	3	0.3628	0.3658	0.4716	**0.6292**	**0.6372**
		3	2	**0.6371**	**0.6360**	**0.5349**	0.3758	0.3629
	4	2	3	0.4301	0.4302	0.4447	**0.5515**	**0.5751**
		3	2	**0.7064**	**0.7063**	**0.6731**	0.4180	0.3005
8	1	2	3	0.2361	**0.6905**	**0.8162**	**0.8169**	**0.8169**
		3	2	**0.5307**	**0.5699**	**0.5636**	**0.5635**	**0.5635**
	7	2	3	0.4833	0.4834	0.4839	**0.5180**	**0.5341**
		3	2	**0.7338**	**0.7337**	**0.7313**	**0.5615**	0.2974
10	2	2	3	0.2741	**0.5501**	**0.7414**	**0.7429**	**0.7434**
		3	2	**0.5131**	**0.5259**	**0.5170**	**0.5024**	0.5000
	5	2	3	0.3758	0.3777	**0.5190**	**0.6237**	**0.6241**
		3	2	**0.5966**	**0.5958**	0.4858	0.4038	0.4033
	7	2	3	0.4216	0.4307	0.4415	**0.5573**	**0.5693**
		3	2	**0.6758**	**0.6547**	**0.6320**	0.3869	0.3454

From Table 4 we observe that, if $\beta \leq \delta$ (or $\beta > \delta$), there exist $p_0 \in (0,1)$ such that for $p \geq p_0$ (or $p \leq p_0$), we get $\pi_{\beta,\delta}(X_{n:n}^{w}, X_{n:n}|\xi_p) \geq 0.5$.

4. PITMAN CLOSENESS AND WEIGHTED RANDOM VARIABLES

Note that in the field of PMC, there is no necessity to restrict our attention to parameters and their estimators. There are situations in which the problem of closeness of some random variables with respect to each others may be of great importance. For instance, if Z is a random variable and X and Y are two other random variables, then for an appropriate measure of distance $d(\cdot, \cdot)$, we may define $GPN(X, Y, Z) = P(d(X, Z) < d(Y, Z))$ to determine the closer random variable between X and Y to Z; see for example, Mendes and Merkle [28]. Here, we focus on the PMC by using the distance $d(X, Z) = |X - Z|$.

Now, consider the situation in which the cdf of a distribution varies in some steps when the time proceeds. Let us show the corresponding random variable to the baseline (initial) distribution by X. Then, the transformed distribution at the i-th temporal step can be characterized by a weighted random variable, such as X_{w_i} $(i \geq 1)$ with pdf $g_i(\cdot)$. So, the PMC of any two transformed distribution to the baseline distribution is discussed in what follows.

Let X_{w_1}, X_{w_2} and X_{w_3} be independent weighted random variables which are considered to describe the transformations on a baseline distribution by pdf $f(\cdot)$ and cdf $F(\cdot)$. Therefore, the pdf of X_{w_i} is given by

$$(4.1) \qquad g_i(x) = \frac{w_i(x)f(x)}{E(w_i(X))}, \quad E(w_i(X)) > 0, \quad i = 1, 2, 3.$$

Here, we generally focus our attention to compute the probability of closeness of X_{w_1} and X_{w_2} with respect to X_{w_3}. Note that in the special case of $w_3(x) = 1$, in fact the closeness of two transformed distributions with respect to the baseline distribution is studied. Using (1.1), we get

$$
\begin{aligned}
\pi(X_{w_1}, X_{w_2} | X_{w_3}) &= P(|X_{w_1} - X_{w_3}| < |X_{w_2} - X_{w_3}|) \\
&= P(-|X_{w_2} - X_{w_3}| < X_{w_1} - X_{w_3} < |X_{w_2} - X_{w_3}|) \\
&= P(X_{w_2} < X_{w_1} < 2X_{w_3} - X_{w_2}, X_{w_2} < X_{w_3}) \\
&\quad + P(2X_{w_3} - X_{w_2} < X_{w_1} < X_{w_2}, X_{w_2} > X_{w_3}).
\end{aligned}
$$

Assuming the random variables are defined on $[a, b]$, we have

$$
\begin{aligned}
\pi(X_{w_1}, X_{w_2} | X_{w_3}) &= \int_a^b \int_a^z \int_y^{2z-y} g_1(x)g_2(y)g_3(z)\,dx\,dy\,dz \\
&\quad + \int_a^b \int_z^b \int_{2z-y}^y g_1(x)g_2(y)g_3(z)\,dx\,dy\,dz \\
&= \frac{1}{\Lambda}\Bigg\{ \int_a^b \int_a^z A_1(y)\bar{F}(y)w_2(y)f(y)w_3(z)f(z)\,dy\,dz \\
&\quad - \int_a^b \int_a^z A_1(2z-y)\bar{F}(2z-y)w_2(y)f(y)w_3(z)f(z)\,dy\,dz \\
&\quad + \int_a^b \int_z^b A_1(2z-y)\bar{F}(2z-y)w_2(y)f(y)w_3(z)f(z)\,dy\,dz \\
(4.2) &\quad - \int_a^b \int_z^b A_1(y)\bar{F}(y)w_2(y)f(y)w_3(z)f(z)\,dy\,dz \Bigg\},
\end{aligned}
$$

where

$$\Lambda = E(w_1(X))E(w_2(X))E(w_3(X))$$

and

$$A_1(x) = E(w_1(X)|X > x) = \frac{1}{\bar{F}(x)} \int_x^b w_1(t)f(t)dt.$$

Let us now suppose that $w_i(x) = \varphi_i(F(x))$, for $i = 1, 2, 3$, then by using transformations $u = F(y)$ and $v = F(z)$, the PMC in (4.2) can be rewritten as follows:

$$\pi(X_{w_1}, X_{w_2}|X_{w_3}) = \frac{1}{\Lambda} \left\{ \int_0^1 \int_0^v [C_1(u,v) - C_2(u,v)]dudv - \int_0^1 \int_v^1 [C_1(u,v) - C_2(u,v)]dudv \right\},$$

where

$$C_1(u,v) = A_1(F^{-1}(u))(1-u)\varphi_2(u)\varphi_3(v)$$

and

$$C_2(u,v) = A_1(2F^{-1}(v) - F^{-1}(u))\bar{F}(2F^{-1}(v) - F^{-1}(u))\varphi_2(u)\varphi_3(v).$$

By selecting various choices of $w_i(x)$, specially those of presented in the previous sections, the above probability of closeness can be derived via tedious calculations.

It is worth to mention that formula (4.2) can also be used to assess the relationship among different statistics. Suppose we have three independent sample from the same distribution. It is known that in the context of nonparametric $X_{[np]:n}$ used as an estimator for ξ_p. Denote the the sample mean of a standard exponential population by $\bar{X}$, and consider the sample quantiles of two independent standard exponential populations by $X_{r:n}$ and $X_{s:n}$, respectively. Then by using (4.2) we have

$$\pi(X_{r:n}, X_{s:n}|\overline{X}) = n^n rs \binom{n}{r}\binom{n}{s} \sum_{i=1}^{r-1}\sum_{j=1}^{s-1}(-1)^{i+j}\binom{r-1}{i}\binom{s-1}{j}$$

$$\times \left\{ \frac{(n-r+i+j)^{-1}}{(2n-r-s+i+j+2)} \left(\frac{1}{n^n} - \frac{1}{(3n-r-s+i+j+3)^n} \right. \right.$$

$$\left. - \frac{1}{(3n-2r+2i+2)^n} + \frac{1}{(4n-3r-s+3i+j+4)^n} \right)$$

$$\left. + \frac{(3n-r-s+i+j+2)^{-n}}{(r-s-i+j)(2n-r-s+i+j+2)} \right\}.$$

5.   APPLICATIONS

To illustrate the performance of the proposed procedure in Sections 4 and 5, we use two real data sets in the following examples.

Example 1 (*Telephone calls*)**.** Table 5 contains the data concerning the times (in minutes) between 48 consecutive telephone calls to a company's switchboard which is presented by Castillo *et al.* [13]. They assumed that the data come from the exponential distribution.

Table 5:   Times (in minutes) between 48 consecutive calls.

1.34	0.14	0.33	1.68	1.86	1.31	0.83	0.33	2.20	0.62	3.20	1.38
0.96	0.28	0.44	0.59	0.25	0.51	1.61	1.85	0.47	0.41	1.46	0.09
2.18	0.07	0.02	0.64	0.28	0.68	1.07	3.25	0.59	2.39	0.27	0.34
2.18	0.41	1.08	0.57	0.35	0.69	0.25	0.57	1.90	0.56	0.09	0.28

By using the definition of k-records, as presented in subsection 3.1, from this data set, the second upper 2-records (values of $R_{2,2}$) have been extracted which are 0.33, 1.86, 1.38, 0.44, 0.51, 1.85, 0.07, 0.68, 2.39, 0.34, 0.57 and 0.57. Note that these data are indeed the second largest observations in the partial samples. Moreover, after observing each data point, the procedure of collecting the next second upper 2-record has been restarted. Moreover, the initial sample maxima to attain the second upper 2-records are observed as

$$1.34, \ 2.2, \ 3.2, \ 0.96, \ 0.59, \ 2.18, \ 0.64, \ 1.07, \ 3.25, \ 2.18, \ 1.08, \ 0.69.$$

The null hypothesis that the above data are coming from a weighted distribution with pdf (3.1) and the parameters $\lambda = (\alpha, \beta, 2, 2)$ (distribution of the second upper 2-record), is checked by using the Kolmogorov–Smirnov (K-S) distances between the empirical distribution functions and the fitted distribution function. The observed MLEs of α and β are 3.5 and 0.8, respectively. Also, the observed value of K-S statistic is 0.2158 and the associated p-value is 0.6311. So, based on these observations, the weighted distribution is adequate for the data regarding the sample maxima to attain the second upper 2-records. That is, one may accept that these data are some observed values of $R_{2,2}^w$.

Using (3.4), the values of $\pi_{3.5,0.8}(R_{2,2}^w, R_{2,2}|\xi_p)$ have been numerically obtained and presented in Table 6 for some choices of p. It is observed that for upper quantiles ($p \geq 0.75$), $R_{2,2}^w$ is Pitman closer to ξ_p than $R_{2,2}$.

Table 6:   Values of $\pi_{3.5,0.8}(R_{2,2}^w, R_{2,2}|\xi_p)$ for standard exponential distribution.

p	0.10	0.25	0.5	0.60	0.75	0.80	0.90	0.95	0.99
$\pi_{3.5,0.8}(R_{2,2}^w, R_{2,2}|\xi_p)$	0.2378	0.2389	0.2935	0.3727	0.5616	0.6341	0.7366	0.7581	0.7621

Example 2 (*Air conditioning system*)**.** In this example we use the data set which consist of the intervals between failures (in hours) of the air conditioning system in three Boeing 720 jet aircrafts. The data are reported in Table 7. See Proschan [34] for a detailed description of the data set. He tested and accepted the hypothesis that the successive failure times were iid exponential for each plane, but with different failure rates.

Table 7:   Intervals between failures of the air conditioning system in three Boeing 720 jet aircraft.

Plane 7909	90,10,60,186,61,49,14,24,56,20,79,84,44,59,29,118,25,156,310,76,26,44,23,62,130,208,70,101,208
Plane 7912	23,261,87,7,120,14,62,47,225,71,246,21,42,20,5,12,120,11,3,14,71,11,14,11,16,90,1,16,52,95
Plane 7913	97,51,11,4,141,18,142,68,77,80,1,16,106,206,82,54,31,216,46,111,39,63,18,191,18,163,24

Let us denote the intervals between failures of the air conditioning system of Planes 7913, 7912 and 7909 by X, X_{w_1} and X_{w_2}, respectively. As well as assumed by Proschan [34], the observed values of X come from exponential distribution. Moreover, it can be deduced that X_{w_1} and X_{w_2} are two different weighted versions of X. More precisely, according to the general weighted pdf presented in (3.1) with baseline exponential distribution, the null hypotheses that the associated data for Planes 7912 and 7909 are coming from

$$g_F(x; 1, 1.75, 1, 1) = 1.75[\bar{F}(x)]^{0.75} f(x)$$

and

$$g_F(x; 1, 1, 1, 1.15) = \frac{1}{\Gamma(1.15)}[-\log \bar{F}(x)]^{0.15} f(x),$$

respectively, are checked by using the K-S distances. The observed p-values are 0.6455 and 0.9123, respectively. So, based on these observations, the mentioned weighted versions of the exponential distribution are adequate for the data. By using (4.2) and doing some algebraic calculations and numerical computations, we get $\pi(X_{w_1}, X_{w_2}|X) = 0.0452$. That is, the intervals between failures of the air conditioning system of Plane 7909 are Pitman closer than those of Plane 7912 with respect to Plane 7913.

6. CONCLUSION

In this paper, the weighted random variables were considered and their closeness to a common parameter was investigated in the sense of PMC. Some general results were derived and a new general weighted model was introduced which subsumes most of the previous works as special cases. Some numerical results and conclusions were presented for exponential distribution in details. It was seen that the weighted upper k-records are Pitman closer than the usual ones to certain population quantiles. To illustrate the proposed procedure, a real data set was used. Furthermore, the concept of PMC was applied for measuring the nearness of some weighted random variables with respect to each other. This procedure was also explained via application to a real data set.

REFERENCES

[1]   AHMADI, J. and BALAKRISHNAN, N. (2009). Pitman closeness of record values to population quantiles, *Statistics and Probability Letters*, **79**, 2037–2044.

[2]   AHMADI, J. and BALAKRISHNAN, N. (2010a). On Pitman's measure of closeness of k-records, *Journal of Statistical Computation and Simulation*, **81**, 497–509.

[3]   AHMADI, J. and BALAKRISHNAN, N. (2010b). Pitman closeness of current records for location-scale families, *Statistics and Probability Letters*, **80**, 1577–1583.

[4]   AHMADI, J. and RAQAB, M.Z. (2013). Comparison of order statistics in two-sample problem in the sense of Pitman closeness, *Statistics*, **47**, 729–743.

[5]   AHMADI, M. and MOHTASHAMI BORZADARAN, G.R. (2014). On the nearness of records to population quantiles with respect to order statistics in the sense of Pitman closeness, *Communications in Statistics – Theory and Methods*, **43**, 4357-4370.

[6]   ARNOLD, B.C.; BALAKRISHNAN, N. and NAGARAJA, H.N. (1998). *Records*, John Wiley & Sons, New York.

[7]   ARNOLD, B.C.; BALAKRISHNAN, N. and NAGARAJA, H.N. (2008). *A First Course in Order Statistics*, SIAM.

[8]   AZZALINI, A. (2022). An overview on the progeny of the skew-normal family — A personal perspective, *Journal of Multivariate Analysis*, **188**, 104851.
`https://doi.org/10.1016/j.jmva.2021.104851`

[9]   BALAKRISHNAN, N.; DAVIES, K. and KEATING, J.P. (2009). Pitman closeness of order statistics to population quantiles, *Communications in Statistics – Simulation and Computation*, **38**, 802–820.

[10]   BALAKRISHNAN, N.; DAVIES, K.; KEATING, J.P. and MASON, R.L. (2010). Simultaneous closeness among order statistics to population quantiles, *Journal of Statistical Planning and Inference*, **140**, 2408–2415.

[11]   BALAKRISHNAN, N.; ILIOPOULOS, G.; KEATING, J.P. and MASON, R.L. (2009). Pitman closeness of sample median to population median, *Statistics and Probability Letters*, **79**, 1759–1766.

[12]   BARTOSZEWICZ, J. (2009). On a representation of weighted distributions, *Statistics and Probability Letters*, **79**, 1690–1694.

[13]   CASTILLO, F.; HADI, A.S.; BALAKRISHNAN, N. and SARABIA, J.M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*, John Wiley & Sons, Inc.

[14]   DAVID, H.A. and NAGARAJA, H.N. (2003). *Order Statistics*, third edition, John Wiley & Sons, Hoboken, New Jersey.

[15]   DAVIES, K.F. (2021). Pitman closeness results for Type-I hybrid censored data from exponential distribution, *Journal of Statistical Computation and Simulation*, **91**, 58–80.

[16]   EFRON, B. (1975). Biased versus unbiased estimation, *Advances in Mathematics*, **16**, 259–277.

[17]   FISHER, R.A. (1934). The effect of methods of ascertainment upon the estimation of frequencies, *Annals of Eugenics*, **6**, 13–25.

[18]   GHOSH, M. and SEN, P.K. (1989). Median unbiasedness and Pitman closeness, *Journal of the American Statistical Association*, **84**, 1089–1091.

[19]   GÓMEZ-DÉNIZ, E.; ARNOLD, B.C.; SARABIA, J.M. and GÓMEZ, H.W. (2021). Properties and applications of a new family of skew distributions, *Mathematics*, **9**, p. 87.

[20]   GUPTA, R.C. and KEATING, J.P. (1986). Relations for the reliability measures under length biased sampling, *Scandinavian Journal of Statistics*, **13**, 49–56.

[21]   Jiang, Y. (2020). Study on weight function distribution of hybrid gas-liquid two-phase flow electromagnetic flowmeter, *Sensors*, **20**, 1431. `https://doi.org/10.3390/s20051431`

[22]   Jones, M.C. (2004). Families of distributions arising from distributions of order statistics (with discussion), *Test*, **13**, 1–43.

[23]   Keating, J.P. (1985). More on Rao's phenomenon, *Sankhya, Series A*, **47**, 18–21.

[24]   Keating, J.P. and Mason, R.L. (1985). Practical relevance of an alternative criterion in estimation, *American Statistician*, **39**, 203–205.

[25]   Keating, J.P.; Mason, R.L. and Sen, P.K. (1993). *Pitman's Measure of Closeness: A Comparison of Statistical Estimators*, Society for Industrial and Applied Mathematics, Philadelphia.

[26]   Kourouklis, S. (1996). Improved estimation under Pitman's measure of closeness, *Annals of Statistics*, **48**, 509–518.

[27]   Lee, C.M.S. (1990). On the characterization of Pitman's measure of nearness, *Statistics and Probability Letters*, **8**, 41–46.

[28]   Mendes, B. and Merkle, M. (2005). Some remarks regarding Pitman closeness, *Univ. Beograd Publ. Elektrotehn. Fak. Ser. Mat.*, **16**, 1–11.

[29]   Morabbi, H. and Razmkhah, M. (2022). Quantile estimation based on modified ranked set sampling schemes using Pitman closeness, *Communications in Statistics – Simulation and Computation*, **51**, 6968–6988. `https://doi.org/10.1080/03610918.2020.1811329`

[30]   Nayak, T.K. (1990). Estimation of location and scale parameters using generalized Pitman nearness criterion, *Journal of Statistical Planning and Inference*, **24**, 259–268.

[31]   Patil, G.P. and Rao, C.R. (1977). *Weighted distributions: a survey of their application*. In "Applications of Statistics" (P.R. Krishnaiah, Ed.), North Holland Publishing Company, 383–405.

[32]   Patil, G.P. and Rao, C.R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families, *Biometrics*, **34**, 179–180.

[33]   Pitman, E.J.G. (1937). The closest estimates of statistical parameters, *Proceedings of The Cambridge Philosophical Society*, **33**, 212–222.

[34]   Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate, *Technometrics*, **5**, 375–383.

[35]   Rao, C.R. (1965). *Weighted distributions arising out of methods of ascertainment*. In "Classical and Contagious Discrete Distributions" (G.P. Patil, Eds.), Calcuta, Pergamon Press and Statistical Publishing Society, 320–332.

[36]   Rao, C.R. (1980). Discussion of Minimum chi-square, not maximum likelihood by J. Berkson, *Annals of Statistics*, **8**, 482–485.

[37]   Rao, C.R. (1981). *Some comments on the minimum mean square error as a criterion of estimation*. In "Statistics and Related Topics" (M. Csorgo, D. Dawson, J. Rao, and A. Saleh, Eds.), North Holland, Amsterdam, 123–143.

[38]   Rao, C.R. (1985). *Weighted distributions arising out of methods of ascertainment*. In "A Celebration of Statistics" (A.C. Atkinson and S.E. Fienberg, Eds.), Springer-Verlag, New York, Chapter 24, pp. 543–569.

[39]   Raqab, M.Z. and Ahmadi, J. (2012). Pitman closeness of record values from two sequences to population quantiles, *Journal of Statistical Planning and Inference*, **142**, 855–862.

[40]   Razmkhah, M. and Ahmadi, J. (2011). Nonparametric prediction intervals for future order statistics in a proportional hazard model, *Communications in Statistics – Theory and Methods*, **40**, 1807–1820.

[41] RAZMKHAH, M. and AHMADI, J. (2013). Pitman closeness of current k-records to population quantiles, *Statistics & Probability Letters*, **83**, 148–156.

[42] RAZMKHAH, M.; AHMADI, J. and KHATIB, B. (2008). Nonparametric confidence intervals and tolerance limits based on minima and maxima, *Communications in Statistics – Theory and Methods*, **37**, 1523–1542.

[43] ROBERT, C.P.; HWANG, J.T.G. and STRAWDERMAN, W.E. (1993). Is Pitman closeness a reasonable criterion?, *Journal of the American Statistical Association*, **88**, 52–76.

[44] SAGHIR, A.; HAMEDANI, G.G.; TAZEEM, S. and KHADIM, A. (2017). Weighted distributions: a brief review, perspective and characterizations, *International Journal of Statistics and Probability*, **6**, 109–131.

[45] SEN, P.K. (1986). Are BAN estimators the Pitman closest ones too?, *Sankhya*, Series A, **48**, 51–58.

[46] SEN, P.K.; KUBOKAWA, T. and SALEH, A.K.M.E. (1989). The Stein paradox in the sense of the Pitman measure of closeness, *Annals of Statistics*, **17**, 1375–1386.

[47] VOLTERMAN, W.; DAVIES, K.F. and BALAKRISHNAN, N. (2013). Simultaneous Pitman closeness of progressively type-II right censored order statistics to population quantiles, *Statistics*, **47**, 439–452.

A Novel Fractional Forecasting Model for Time Dependent Real World Cases

Authors: Ummugulsum Erdinc [ID]
– Department of Mathematics, Aksaray University,
Aksaray, Turkey
ummugulsumerdinc@aksaray.edu.tr

Halis Bilgil [ID] [✉]
– Department of Mathematics, Aksaray University,
Aksaray, Turkey
halis@aksaray.edu.tr

Zafer Ozturk [ID]
– Institute of Science, Nevsehir Haci Bektas Veli University,
Turkey
zaferozturk@aksaray.edu.tr

Abstract:

• The grey modelling in the prediction of time series has been one of the interesting study fields recently due to its efficiency and convenience. Fractional grey models, on the other hand, have become preferable, despite the difficulty in calculations, since they give more effective results than standard models. Difficulties in fractional accumulation and difference calculations have begun to be overcome thanks to new definitions and theorems made in recent years. The new trend in grey modelling is to compose models that are more useful than the previous ones and give results with less error. In this paper, a new grey model derived from the conformable fractional order is defined and it is shown that more effective estimates are made compared to the models created in recent years. The verification of the method is shown with real data set. The results show that the proposed conformable fractional grey model is more effective than the existing models.

[✉] Corresponding author

1.   INTRODUCTION

The grey system is a discipline that studies the problem of uncertainty, which was first proposed by Deng in 1982, plays an important role in the grey system theory [4]. It has been a useful tool in processing uncertain or excursive systems with small samples and limited data set as distinct from the machine learning models, hybrid models, the empirical models, ..., etc. The grey prediction models have been widely and successfully applied to various fields, such as science and technology, energy, environmental problems, economy, health and other fields [24, 35, 36, 7, 1, 33, 5, 18].

Recent studies on grey modelling focus on two main purpose: practicality and prediction accuracy. For these purposes, important studies have been carried out in recent times. Ma and Lui [15] proposed a time-delayed polynomial grey prediction model called TDPGM(1, 1) model, the grey polynomial model with a tuned background coefficient was proposed by Wei *et al.* [20], Cui *et al.* [3] developed a parameter optimization method to improve the ONGM(1, 1, k) model, Bilgil [1] proposed an exponential grey model named EXGM(1, 1). Furthermore, Ma *et al.* [14] developed a novel nonlinear multivariate forecasting grey model based on the Bernoulli equation named NGBMC(1, n), Wang *et al.* [19] introduced a seasonal grey model called SGM(1, 1), Wu *et al.* [25] proposed a new grey model called BNGM(1, 1, t^2) model, Liu and Wu [12] proposed the ANDGM model, Ma [13] proposed kernel-based KARGM(1, 1) model, Li *et al.* [10] developed structure-adaptive intelligent grey forecasting model, Wu *et al.* [26] developed a novel grey Riccati model (GRM), the modified grey prediction model with damping trend factor was proposed by Liu *et al.* [11], the nonlinear grey Bernoulli model with improved parameters, INGBM(1, 1), was proposed by Jiang *et al.* [8].

It is clear that most of the existing grey models are defined with a first-order whitening differential equation. If the original data is a disordered sequence, the characteristic features of the sequence may not be exactly found out by first-order accumulative generation operation (1-AGO) [32]. Moreover, first-order derivative models are ideal memory models, which are not suitable for describing irregular phenomena. As a result of this, the parameters of the model may not be compatible according to the data characteristics of the actual problem for a sequence with large data fluctuation. Therefore, fractional accumulation generating operation and fractional derivative should be introduced into the grey model to overcome this problem [32].

Wu *et al.* developed traditional GM(1, 1) with fractional order accumulated operator named FAGM(1, 1) [23]. Some researchers optimized the FAGM(1, 1) model and reached better prediction accuracy in recent years. Wu *et al.* suggested a fractional FAGMO(1, 1, k) model with linear grey input of time in lieu of constant grey input in the initial FAGM(1, 1) model and optimized it with optimal order and optimal parameters [27]. Besides, Mao *et al.* introduced the fractional grey model FGM(q, 1) [17], a power-driven fractional accumulated grey model named PFAGM is introduced by Zhang *et al.* [34], Yuxiao *et al.* proposed the multivariable Caputo fractional derivative grey model with convolution integral named CFGMC(q, N), Xie *et al.* developed a conformable fractional grey model in opposite direction CFGOM(1, 1) [30].

However, the definition is given by Wu [23] only show us a single situation of fractional order calculus and differencing. From the perspective of computational complexity, frequently

operated descriptions of fractional accumulation and its suitable fractional differencing in the present grey models are not easy to apply, and it causes serious rigours to the more deeply theoretical analysis. Yang and Xue [31] submitted the fractional order calculus, but the exact solution of this kind of model include infinite series, and it is clearly difficult to use and analyse. Furthermore, this kind of complications would also impede the improvement of the fractional order grey models thanks to some new operators [22, 21].

Lately, Khalil *et al.* defined a limit-based fractional derivative in 2014 [9], which is named the conformable fractional derivative. The structure of this new definitions of fractional derivative is simpler than that of other popular fractional derivatives, such as the Caputo derivative and Riemann–Liouville derivative. Then, Ma *et al.* [16] introduced the new useful definitions of the fractional order difference and accumulation based on the conformable fractional derivative in which the computational complexity of accumulation is lower than that of the traditional fractional accumulated operator and they firstly proposed an improved fractional order grey model named CFGM(1, 1). Recently, a continuous grey model named CCFGM based on the conformable fractional derivative was described by Xie *et al.* [29].

In this paper, we introduced the novel exponential conformable fractional grey model (denoted as ECFGM(1, 1) for short) by using the new definitions of conformable fractional difference and conformable fractional accumulation by Ma *et al.* [16]. The structural order of ECFGM(1, 1) is sought out by using the Brute Force algorithm. The effectiveness of ECFGM(1, 1) is validated by real data sets in comparison with other used new grey models and it is seen that the performance of the ECFGM(1, 1) model is very successful.

The rest of this paper is organized as follows: Section 1 includes relevant literature. Some useful properties and definitions of the conformable fractional calculus are given in Section 2. The presentations and modelling mechanism of the ECFGM(1, 1) are introduced in Section 3. In Sections 4 and 5, we present a series of samples to validate ECFGM(1, 1). Finally, the conclusions of this study are given in Sections 6.

2. SOME DEFINITIONS AND PROPERTIES ON CONFORMABLE FRACTIONAL CALCULUS

In this section, some useful definions and properties of the conformable fractional derivative are summerized.

2.1. The conformable fractional derivative

Definition 2.1 (See [16]). If $f\colon [0,\infty) \to R$ is a differentiable function, the conformable fractional derivative of f with $\alpha \in (n, n+1]$ order is defined as

$$(2.1) \qquad T_\alpha(f)(t) = \lim_{\varepsilon \to 0} \frac{f(t + \varepsilon t^{\lceil \alpha \rceil - \alpha}) - f(t)}{\varepsilon} = t^{\lceil \alpha \rceil - \alpha} \frac{df(t)}{dt},$$

where $\lceil \cdot \rceil$ is the ceil function, i.e. the $\lceil \alpha \rceil$ is the smallest integer no larger than α. It is clear that $\lceil \alpha \rceil = 1$ for $\alpha \in (0,1]$. Hence, equation (2.1) can be written as

$$T_\alpha(f)(t) = \lim_{\varepsilon \to 0} \frac{f(t + \varepsilon t^{1-\alpha}) - f(t)}{\varepsilon} = t^{1-\alpha} \frac{df(t)}{dt}$$

for all $t > 0$.

The following theorem gives the properties of the definition (Khalil *et al.* [9]).

Theorem 2.1 (See [9])**.** *If the functions f and g are differentiable, $\alpha \in (0,1]$, then we have:*

1. $T_\alpha(f)(t) = t^{1-\alpha} \frac{df(t)}{dt}$;
2. $T_\alpha(mf + ng) = mT_\alpha(f) + nT_\alpha(g)$ *for all $m, n \in \mathbb{R}$;*
3. $T_\alpha(f.g) = fT_\alpha(g) + gT_\alpha(f)$;
4. $T_\alpha\left(\frac{f}{g}\right) = \frac{gT_\alpha(f) - fT_\alpha(g)}{g^2}$;
5. $T_\alpha(c) = 0$ *for all constant c;*
6. $T_\alpha(t^p) = pt^{p-\alpha}$ *for all $p \in \mathbb{R}$;*
7. $T_\alpha(e^{cx}) = cx^{1-\alpha}e^{cx}$ *for all $c \in \mathbb{R}$.*

Proof: The proof is omitted. $\qquad\square$

2.2. The conformable fractional accumulation and difference

New definitions to calculate the conformable fractional accumulation (CFA) and the conformable fractional difference (CFD) are given by Ma *et al.* [16] as follows.

Definition 2.2 (see [16, 28])**.** The conformable fractional difference (CFD) of f with α order is defined as

(2.2)
$$\Delta^\alpha f(k) = k^{1-\alpha} \Delta f(k) = k^{1-\alpha}[f(k) - f(k-1)]$$

for all $k \in N^+$, $\alpha \in (0,1]$, and

(2.3)
$$\Delta^\alpha f(k) = k^{\lceil \alpha \rceil - \alpha} \Delta^{n+1} f(k) = k^{\lceil \alpha \rceil - \alpha} \sum_{j=k-\lceil \alpha \rceil}^{k} (-1)^{k-j} \binom{\lceil \alpha \rceil}{k-j} f(j)$$

for all $k \in N^+$, $\alpha \in (n, n+1]$.

Definition 2.3 (see [16])**.** The conformable fractional accumulation (CFA) of f with α order is defined as

(2.4)
$$\nabla^\alpha f(k) = \nabla\left(\frac{f(k)}{k^{1-\alpha}}\right) = \sum_{j=1}^{k} \frac{f(j)}{j^{1-\alpha}}$$

for all $k \in N^+$, $\alpha \in (0,1]$, and

(2.5)
$$\nabla^\alpha f(k) = \nabla^{n+1}\left(\frac{f(k)}{k^{\lceil \alpha \rceil - \alpha}}\right)$$

for all $k \in N^+$, $\alpha \in (n, n+1]$.

3. PRESENTATION OF EXPONENTIAL CONFORMABLE FRACTIONAL GREY MODEL

In this section, a novel exponential conformable fractional grey model, named ECFGM$(1,1)$, is introduced, which optimizes the classical CFGM$(1,1)$ model with an exponential grey action quantity.

3.1. Formulation of proposed fractional grey model

The original series $X^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n)\right)$ is given. CFA with α order is calculated as follows:

$$(3.1) \qquad X^{(\alpha)} = \left(x^{(\alpha)}(1), x^{(\alpha)}(2), ..., x^{(\alpha)}(n)\right),$$

where

$$(3.2) \qquad x^{(\alpha)}(k) = \nabla^\alpha x^{(0)}(k) = \sum_{i=1}^{k} \begin{bmatrix} \lceil \alpha \rceil \\ k-i \end{bmatrix} \frac{x^{(0)}(i)}{i^{\lceil \alpha \rceil - \alpha}}, \quad \alpha \in \mathbb{R}^+,$$

where $\begin{bmatrix} \lceil \alpha \rceil \\ k-i \end{bmatrix} = \frac{\Gamma(k-i+\lceil \alpha \rceil)}{\Gamma(k-i+1)\Gamma(\lceil \alpha \rceil)} = \frac{(k-i+\lceil \alpha \rceil-1)!}{(k-i)!(\lceil \alpha \rceil-1)!}$ (see [28]).

Definition 3.1. The first-order whitening differential equation of the ECFGM$(1,1)$ is defined as

$$(3.3) \qquad \frac{dx^{(\alpha)}(t)}{dt} + ax^{(\alpha)}(t) = b + ce^{-t},$$

where a is a development coefficient, b is called driving coefficient and ce^{-t} is an exponential grey action quantity. So that, the monotone decreasing term ce^{-t} will suppress the growth of the prediction error.

When $\alpha = 1$, the ECFGM$(1,1)$ model yields the EXGM$(1,1)$ [9]. In addition, the proposed model can be translated to the conventional CFGM$(1,1)$ model for $c = 0$ [23].

3.2. Parameters estimation

Theorem 3.1. *For the computed CFA and the value of fractional order, the system parameters a, b and c of the ECFGM$(1,1)$ satisfy the following equation:*

$$(3.4) \qquad [a, b, c]^\top = (B^\top B)^{-1} B^\top Y,$$

where the matrix B and Y are

$$(3.5) \qquad B = \begin{bmatrix} -0.5(x^{(\alpha)}(2) + x^{(\alpha)}(1)) & 1 & (e-1)e^{-2} \\ -0.5(x^{(\alpha)}(3) + x^{(\alpha)}(2)) & 1 & (e-1)e^{-3} \\ \vdots & \vdots & \vdots \\ -0.5(x^{(\alpha)}(n) + x^{(\alpha)}(n-1)) & 1 & (e-1)e^{-n} \end{bmatrix}, \quad Y = \begin{bmatrix} x^{(\alpha)}(2) - x^{(\alpha)}(1) \\ x^{(\alpha)}(3) - x^{(\alpha)}(2) \\ \vdots \\ x^{(\alpha)}(n) - x^{(\alpha)}(n-1) \end{bmatrix}.$$

Proof: Integrating both sides of the whitening equation (3.3) within the interval $[k-1, k]$ the discrete form of ECFGM$(1,1)$ model is obtained as follows:

$$(3.6) \qquad \int_{k-1}^{k} \frac{dx^{(\alpha)}(t)}{dt} dt + a \int_{k-1}^{k} x^{(\alpha)}(t) dt = \int_{k-1}^{k} (b + ce^{-t}) dt.$$

According to the Newton–Leibniz formula, the first integral of (3.6) can be expressed as

$$(3.7) \qquad \int_{k-1}^{k} \frac{dx^{(\alpha)}(t)}{dt} dt = x^{(\alpha)}(k) - x^{(\alpha)}(k-1).$$

It is clear that the integration term $\int_{k-1}^{k} x^{(\alpha)}(t) dt$ denotes the area between t-axis and the curve $x^{(\alpha)}(t)$ in the interval $[k-1, k]$. Then, using the generalized trapezoid formula as in many recent studies [23, 25, 16, 29, 26, 17, 10, 21], the second integral of (3.6) can be obtained as

$$(3.8) \qquad a \int_{k-1}^{k} x^{(\alpha)}(t) dt = \frac{a}{2} \left(x^{(\alpha)}(k) + x^{(\alpha)}(k-1) \right)$$

and the right side of (3.6) is equal to

$$(3.9) \qquad \int_{k-1}^{k} (b + ce^{-t}) dt = b + c(e^{1-k} - e^{-k}).$$

Substituting equations (3.7)–(3.9) into equation (3.6), it can be written as

$$(3.10) \qquad \left(x^{(\alpha)}(k) - x^{(\alpha)}(k-1) \right) + \frac{a}{2} \left(x^{(\alpha)}(k) + x^{(\alpha)}(k-1) \right) = b + c(e^{1-k} - e^{-k}),$$

where $k = 2, 3, ..., n$.

The linear equations system (3.10) can be written as follows:

$$(3.11) \qquad \begin{aligned} x^{(\alpha)}(2) - x^{(\alpha)}(1) &= -0.5a(x^{(\alpha)}(2) + x^{(\alpha)}(1)) + b + c(e^{-1} - e^{-2}) \\ x^{(\alpha)}(3) - x^{(\alpha)}(2) &= -0.5a(x^{(\alpha)}(3) + x^{(\alpha)}(2)) + b + c(e^{-2} - e^{-3}) \\ &\quad\vdots \quad\vdots \quad\vdots \\ x^{(\alpha)}(n) - x^{(\alpha)}(n-1) &= -0.5a(x^{(\alpha)}(n) + x^{(\alpha)}(n-1)) + b + c(e^{1-n} - e^{-n}) \end{aligned}$$

and system (3.11) can be written as

$$(3.12) \qquad Y = B\rho,$$

where

$$(3.13) \qquad B = \begin{bmatrix} -0.5(x^{(\alpha)}(2) + x^{(\alpha)}(1)) & 1 & (e-1)e^{-2} \\ -0.5(x^{(\alpha)}(3) + x^{(\alpha)}(2)) & 1 & (e-1)e^{-3} \\ \vdots & \vdots & \vdots \\ -0.5(x^{(\alpha)}(n) + x^{(\alpha)}(n-1)) & 1 & (e-1)e^{-n} \end{bmatrix}, \quad Y = \begin{bmatrix} x^{(\alpha)}(2) - x^{(\alpha)}(1) \\ x^{(\alpha)}(3) - x^{(\alpha)}(2) \\ \vdots \\ x^{(\alpha)}(n) - x^{(\alpha)}(n-1) \end{bmatrix}$$

and $\rho = [a, b, c]^{\top}$ in which n is the number of samples used to construct the model.

The parameter estimation of the ECFGM$(1,1)$ model using the least squares method can be obtained. For the estimated value of the parameter sequence ρ, the $x^{(\alpha)}(k) - x^{(\alpha)}(k-1)$ on the left side of the equation (3.11) is replaced with $-0.5a(x^{(\alpha)}(k) + x^{(\alpha)}(k-1)) + b + c(e^{1-k} - e^{-k})$, the error sequence $\epsilon = Y - B\rho$ is obtained. Here,

$$(3.14) \qquad \epsilon = [\epsilon_2, \epsilon_3, ..., \epsilon_n]^\top$$

and ϵ_k represents the error for each equation in the system (3.11) for $k = 2, 3, ..., n$.

Notice, $S(\rho)$ is defined as the sum of squares of errors, which yields

$$
\begin{aligned}
S(\rho) &= \sum_{k=2}^{n} \epsilon_k^2 \\
&= \epsilon^\top \epsilon \\
&= (Y - B\rho)^\top (Y - B\rho) \\
&= (Y^\top - \rho^\top B^\top)(Y - B\rho) \\
&= Y^\top Y - Y^\top B\rho - \rho^\top B^\top Y + \rho^\top B^\top B\rho \\
&= Y^\top Y - 2\rho^\top B^\top Y + \rho^\top B^\top B\rho.
\end{aligned}
$$
(3.15)

The parameter vector $\rho = [a, b, c]^\top$ that minimize $S(\rho)$ satisfies

$$(3.16) \qquad \frac{\partial S}{\partial \rho} = -2B^\top Y + 2B^\top B\rho = 0,$$

so

$$(3.17) \qquad B^\top Y = B^\top B\rho.$$

Thus

$$(3.18) \qquad \rho = (B^\top B)^{-1} B^\top Y,$$

or

$$(3.19) \qquad [a, b, c]^\top = \left(B^\top B\right)^{-1} B^\top Y.$$

Thence the proof is completed by using the least square estimation method. $\qquad \square$

3.3. Response function and restored values

Theorem 3.2. *The discrete form of the response function of* ECFGM$(1,1)$ *model is given as*

$$(3.20) \qquad \hat{x}^{(\alpha)}(k) = \left(x^{(0)}(1) - \frac{b}{a} - \frac{c}{a-1}e^{-1}\right)e^{a(1-k)} + \frac{b}{a} + \frac{c}{a-1}e^{-k},$$

where $k = 2, 3, ..., n$.

Proof: It is clear that the solution of the first order linear *whitening differential equation* (3.3) can be obtained as

$$(3.21) \qquad x^{(\alpha)}(t) = \frac{b}{a} + \frac{c}{a-1}e^{-t} + de^{-at},$$

where d is integral constant. By using the initial condition $x^{(\alpha)}(1) = x^{(0)}(1)$, the constant d can be found as

$$d = \left(x^{(0)}(1) - \frac{b}{a} - \frac{c}{a-1}e^{-1} \right)e^{a}.$$

Therefore the grey prediction model equation (3.21) can be obtained as

$$\hat{x}^{(\alpha)}(t) = \left(x^{(0)}(1) - \frac{b}{a} - \frac{c}{a-1}e^{-1} \right)e^{a(1-t)} + \frac{b}{a} + \frac{c}{a-1}e^{-t}$$

and the discrete form of the response function can be written as

$$\hat{x}^{(\alpha)}(k) = \left(x^{(0)}(1) - \frac{b}{a} - \frac{c}{a-1}e^{-1} \right)e^{a(1-k)} + \frac{b}{a} + \frac{c}{a-1}e^{-k}.$$

The proof is completed. □

Theorem 3.3. *Then, restored values can be given as*

$$(3.22) \qquad \hat{x}^{(0)}(k) = \Delta^{\alpha}\hat{x}^{(\alpha)}(k) = k^{\lceil \alpha \rceil - \alpha}\Delta^{n+1}\hat{x}^{(\alpha)}(k), \quad \alpha \in (n, n+1],$$

where $k = 2, 3, ..., n$.

Proof: From (3.2) it can be seen that $\hat{x}^{(\alpha)}(k) = \nabla^{\alpha}\hat{x}^{(0)}(k)$. If we apply the inverse operator Δ^{α}, it is obtained as

$$\hat{x}^{(0)}(k) = \Delta^{\alpha}\hat{x}^{(\alpha)}(k)$$

and from Definition 2.2 it can be written as

$$\hat{x}^{(0)}(k) = \Delta^{\alpha}\hat{x}^{(\alpha)}(k) = k^{\lceil \alpha \rceil - \alpha}\Delta^{n+1}\hat{x}^{(\alpha)}(k), \quad \alpha \in (n, n+1].$$

This completes the proof. □

It is clear that, the restored values for $\alpha \in (0, 1]$ can be written as

$$(3.23) \qquad \hat{x}^{(0)}(k) = k^{1-\alpha}(\hat{x}^{(\alpha)}(k) - \hat{x}^{(\alpha)}(k-1)).$$

3.4. Evaluative accuracy of the forecasting model

The relative percentage error (RPE) and the mean absolute percentage error (MAPE) are used to evaluate the fitting and predicting performance of ECFGM$(1, 1)$. The lowest MAPE value indicates the best prediction model. They are defined as follows:

$$(3.24) \qquad \mathrm{RPE}(k) = \left| \frac{\hat{x}^{(0)}(k) - x^{(0)}(k)}{x^{(0)}(k)} \right| \times 100\%,$$

$$(3.25) \qquad \mathrm{MAPE} = \frac{1}{n}\sum_{k=1}^{n}\mathrm{RPE}(k),$$

where $x^{(0)}(k)$ is the original series, and $\hat{x}^{(0)}(k)$ is the predicted series.

For a raw sequence with n samples, the metric MAPE_{fit} is defined as the fitting performance metric while MAPE_{pre} is defined as a prediction performance metric. Mathematically, they can be formulated as:

$$(3.26) \qquad \text{MAPE}_{\text{fit}} = \frac{1}{p} \sum_{k=1}^{p} \text{RPE}(k),$$

$$(3.27) \qquad \text{MAPE}_{\text{pre}} = \frac{1}{n-p} \sum_{k=p+1}^{n} \text{RPE}(k),$$

where p represents the number of samples used for fitting a model while the rest of the raw sequence is used to examine the prediction accuracy of the model. The total MAPE is given in (3.25) and it is used to evaluate the whole performance of a model.

3.5. Computation steps

The computation steps of $\text{ECFGM}(1,1)$ model with given sample and α can be summarized as follows:

Step 1: Create a raw data set $\left(x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n)\right)$;

Step 2: Take as $\alpha = 0.01$ to the initial value and designate an initial value of MAPE_{min};

Step 3: Compute the CFA series with α order of the given raw data set by using (3.2);

Step 4: Build the matrix B and Y using (3.5);

Step 5: Calculate the parameters a, b and c using (3.4);

Step 6: Calculate the predicted values and the response function using (3.20) and (3.22);

Step 7: Calculate the mean absolute percentage error (MAPE) using (3.25);

Step 8: If MAPE is greater than MAPE_{min}, set α as $\alpha = \alpha + h$ (where h is the step size), otherwise take as $\text{MAPE}_{\text{min}} = \text{MAPE}$ and go to Step 3; where step 8 is continued until α reaches the predetermined value.

Brute Force is a straightforward approach, which is also known as the Naive algorithm, for solving optimization problems that rely on sheer computing power and trying every possibility rather than advanced techniques to improve efficiency. Unlike some of the other popular swarm intelligence algorithms, Brute Force is applicable to a very wide variety of problems and it is an effective and easy method to find the optimum value in the solution interval.

Despite the convergence speed advantages of other algorithms, it is a disadvantage that they may focus on the local extremum point rather than the global extremum. However, the Brute Force algorithm scans the whole domain, evaluates each point, then calculates the MAPE's on these points and reaches optimum parameters without any delusion.

In this paper, our purpose is to find the optimum parameter α that minimizes the model's mean absolute percentage error (MAPE). Therefore we enumerate all the values in an interval with step 0.01. For suitability, in the next section, α will be generated in $(0, 1]$. In this way, all the α points in the whole interval and the MAPE's at these points are calculated.

Hence α, which give the minimum of the calculated MAPE's, is determined as the optimum parameters. The computational steps mentioned above are employed for all the α points.

From equation (2.5), as the α value approaches 1, the CFA series approaches the 1-AGO series, and as the α value approaches 2, the CFA series approaches the 2-AGO series. Obviously, it can be seen that for each point the CFA value $x^{(\alpha)}(k)$ becomes larger with larger α, and the growing speed also increases with larger α [16].

According to the Theorem 3.3, if $\alpha \in (0,1]$ the following first-order fractional difference must be calculated to evaluate the $\hat{x}^{(0)}(k)$ values:

$$(3.28) \qquad \hat{x}^{(0)}(k) = k^{1-\alpha}(\hat{x}^{(\alpha)}(k) - \hat{x}^{(\alpha)}(k-1)),$$

where $k = 2, 3, ..., n$. This shows that the error in $\hat{x}^{(0)}(k)$ values is due to two points of the CFA series. If $\alpha \in (1,2]$, the following second-order fractional difference must be calculated to evaluate the $\hat{x}^{(0)}(k)$ values:

$$(3.29) \qquad \hat{x}^{(0)}(k) = k^{2-\alpha}\Delta^2\hat{x}^{(\alpha)}(k) = k^{2-\alpha}(\hat{x}^{(\alpha)}(k) - 2\hat{x}^{(\alpha)}(k-1) + \hat{x}^{(\alpha)}(k-2)),$$

where $k = 3, ..., n$. If $k = 2$, $\hat{x}^{(\alpha)}(0)$, $\hat{x}^{(\alpha)}(1)$, and $\hat{x}^{(\alpha)}(2)$ values are needed to determine the value of $\hat{x}^{(0)}(2)$. However, according to equation (3.20), since there is no $\hat{x}^{(\alpha)}(0)$ value in the CFA series, the second-order difference can only be calculated for $k = 3, 4, ..., n$. In this case, although $\hat{x}^{(\alpha)}(1)$ is known as $x^{(0)}(1)$ according to the initial value, the value of $\hat{x}^{(\alpha)}(2)$ can only be calculated with the help of a first-order fractional difference. This will increase the total error rate. According to equation (3.29), the error of restored value is effected by the errors of the CFA series with three points, which will be another factor that increases the error rate. Moreover, 99% optimal α values in the CFGM model are in the range of $[0,1)$. In the FGM model, 72% optimal α values are in the range of $(0,1)$ (for more detailed information, see [16]). In the ECFGM model, 89% optimal α values are in the range of $(0,1)$ and 10% optimal α are obtained at 1.

All these processes are completed in about five seconds by writing a simple FORTRAN code.

4.   VALIDATION OF THE ECFGM(1,1)

In this section, two instructive examples are given to demonstrate the efficacy of the proposed model.

4.1. Example A

In this subsection, a numerical example is presented to show the computational steps of the ECFGM(1, 1) model with the raw data $X^{(0)}(k) = (13.21, 18.82, 26.45, 36.04, 42.34, 51.00, 59.12)$. In this example, we select the raw data as a monotone increasing series.

4.1.1. Selecting the optimal α

By using the Brute Force strategy, calculated MAPEs with α in the interval $(0, 1]$ with step 0.01 are given in Figure 1. It can be seen that the values of MAPE increase when the α moves away from its optimum value. For this example, the optimal α is obtained at $\alpha = 0.14$, and the MAPE is calculated for the optimal α as MAPE $= 0.9843$. It is seen that the optimal α is easily obtained by using the Brute Force strategy. Furthermore, it is clear that the performance of the proposed ECFGM$(1, 1)$ model is respectable.
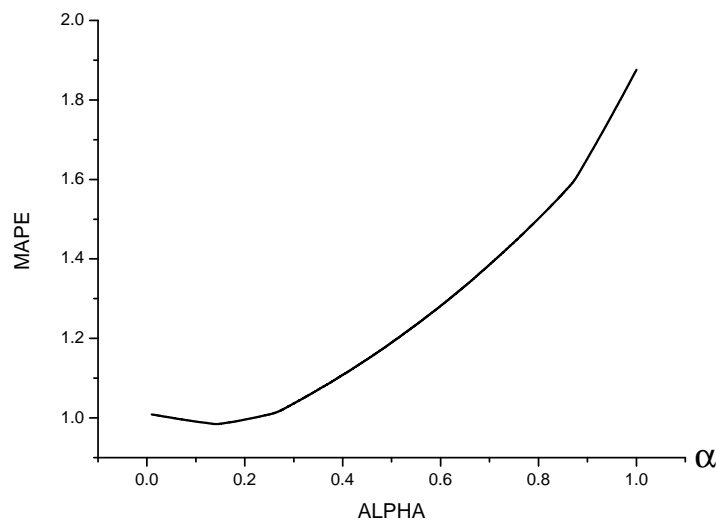
Figure 1: MAPEs of ECFGM model with α in $(0, 1]$ for example A.

4.1.2. Computing the CFA of the original series and modelling the ECFGM(1,1)

Computation of the CFA of the original series is the first step to build the ECFGM$(1, 1)$ model. For the optimum $\alpha = 0.14$, the conformable fractional accumulation series can be obtained using (3.2) as $X^{(0.14)}(k) = (13.2100, 23.5789, 33.8615, 44.8014, 55.4095, 66.3329, 77.4234)$.

The matrices B and Y can be constructed as

$$
B = \begin{bmatrix} -18.3945 & 1 & 0.2325 \\ -28.7202 & 1 & 0.0855 \\ -39.3314 & 1 & 0.0315 \\ -50.1054 & 1 & 0.0116 \\ -60.8712 & 1 & 0.0043 \\ -71.8781 & 1 & 0.0016 \end{bmatrix}, \qquad Y = \begin{bmatrix} 10.3689 \\ 10.2826 \\ 10.9399 \\ 10.6081 \\ 10.9235 \\ 11.0905 \end{bmatrix}.
$$

Then we obtain the parameters $\hat{a}$, $\hat{b}$ and $\hat{c}$ using the least squares solution as

$$[a, b, c]^\top = \left(B^\top B\right)^{-1} B^\top Y$$

and

$$a = -0.0136199261,$$
$$b = 10.0950218700,$$
$$c = -0.0670244100.$$

By substituting the parameters into the response function equation (3.20) we have

(4.1) $$\hat{x}^{(\alpha)}(k) = 754.3807\, e^{-0.0136(1-k)} - 741.1951 + 0.06612\, e^{-k}.$$

Then the restored can be obtained using (4.1) by k from 1 to 7 as

$$\hat{X}^{(0.14)}(k) = (13.2100,\ 23.5396,\ 34.0206,\ 44.6491,\ 55.4247,\ 66.3485,\ 77.4223).$$

Then the restored values can be obtained using the CFD in (3.23) as

$$\hat{X}^{(0)}(k) = (13.2100, 18.7485,\ 26.9607,\ 35.0141,\ 43.0085,\ 51.0018,\ 59.0312).$$

The original raw series $X^{(0)}(k)$ and predicted values $\hat{X}^{(0)}(k)$ are plotted in Figure 2.
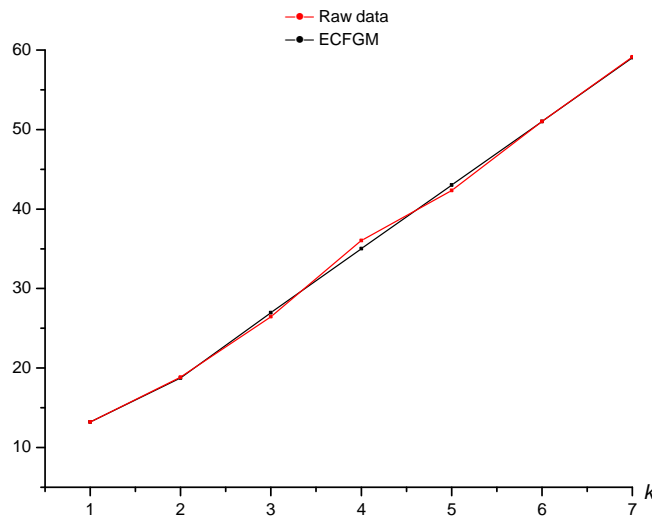
Figure 2: Actual values and forecasting values of example A.

4.2. Example B

Different from example A, the raw data are not a monotone increasing series here. The computational mechanism is similar to example A. In this example, the raw data select as $X^{(0)}(k) = (120.21, 131.83,\ 143.45,\ 150.02, 134.34,\ 121.04,\ 110.15).$

4.2.1. Selecting the optimal α

The calculated MAPEs with α in the interval $(0,1]$ with step 0.01 is given in Figure 3. For this example, the minimum MAPE is calculated at optimal α as $\alpha = 0.89$. The MAPE is calculated for the optimal α as MAPE $= 1.1836$. It is seen that the prediction performance of the proposed ECFGM$(1,1)$ model is successful again.
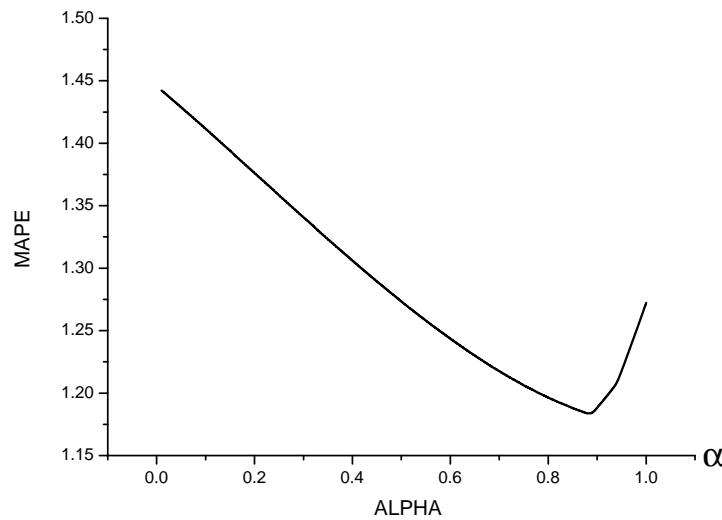
Figure 3: MAPEs of ECFGM model with α in $(0,1]$ for example B.

4.2.2. Computing the CFA of the original series and modelling the ECFGM

Computation of the conformable fractional accumulation series of the original series is the first step to build the ECFGM model. For $\alpha = 0.89$, the CFA series can be obtained using (3.2) as $X^{(0.89)}(k) = (120.2100, 242.3621, 369.4831, 498.2851, 610.8281, 710.2157, 799.1407)$.

The matrices B and Y can be constructed as

$$B = \begin{bmatrix} -181.2861 & 1 & 0.2325 \\ -305.9223 & 1 & 0.0855 \\ -433.8841 & 1 & 0.0315 \\ -554.5566 & 1 & 0.0116 \\ -660.5219 & 1 & 0.0043 \\ -754.6782 & 1 & 0.0016 \end{bmatrix}, \quad Y = \begin{bmatrix} 122.1521 \\ 127.1210 \\ 127.1210 \\ 112.5430 \\ 99.3876 \\ 88.9250 \end{bmatrix}.$$

Then we obtain the parameters $\hat{a}$, $\hat{b}$ and $\hat{c}$ using the least squares solution as

$$[a, b, c]^\top = \left(B^\top B\right)^{-1} B^\top Y$$

and

$$a = 0.1266697621,$$
$$b = 184.88810740,$$
$$c = -174.9875634.$$

By substituting the parameters into the response function equation (3.20) we have

$$\hat{x}^{(\alpha)}(k) = -1413.1086\ e^{0.12667(1-k)} + 1459.6073 + 200.3681\ e^{-k}. \tag{4.2}$$

Then the restored can be obtained using (4.2) by k from 1 to 7 as

$$\hat{X}^{(0.89)}(k) = (120.2100,\ 241.7407,\ 372.7221, 496.9157,\ 609.5692,\ 710.0100,\ 798.9388).$$

Then the restored values can be obtained using the CFD in (3.23) as

$$\hat{X}^{(0)}(k) = (120.2100, 131.1594,\ 147.8062,\ 144.6525,\ 134.4718,\ 122.3227,\ 110.1547).$$

The original raw series $X^{(0)}(k)$ and predicted values $\hat{X}^{(0)}(k)$ are plotted in Figure 4.
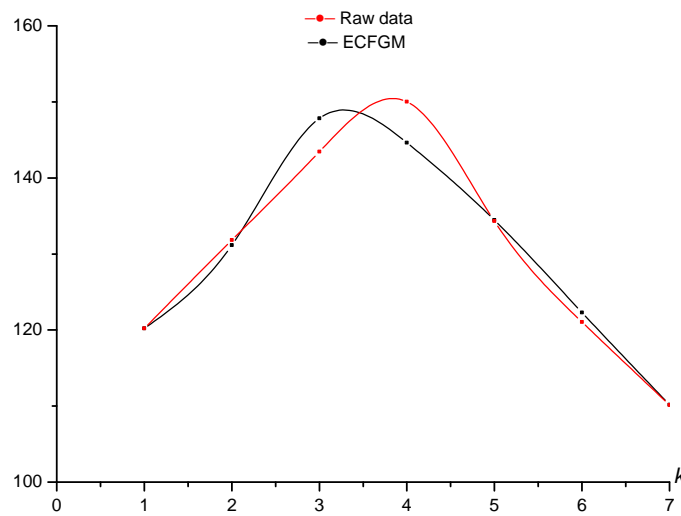
Figure 4: Actual values and forecasting values of example B.

5. VALIDATION OF ECFGM(1,1) WITH A REAL CASE

One of the renewable clean energy sources is wind. Generally, comparing with the many other energy sources, producing energy using wind has fewer effects on the environment. There is no released emission that can pollute the air or water and they do not require water for cooling. Wind turbines may also have the benefit that reduce the amount of power generation from fossil fuels, which outcomes in lower total air pollution and carbon dioxide emissions that could help solve the shortage problem of energy. From prehistoric to today, human beings have used wind energy for sailing, windmills, and wind turbines. Electric generators convert wind energy to electrical energy [16].

One of the biggest countries with large land mass and coastline is China has rich wind resources. With regard to the evaluations by China Meteorological Administration, based on the relatively low height of 10 m above ground, the total theoretical wind power reserves in China 4350 GW, while the technically exploitable wind resources estimated at 297 GW [9].

In this section, we use the novel ECFGM$(1,1)$ model to predict wind energy consumption in China. The data set from [34, 2, 6] is used to test for the efficacy and applicability of the proposed grey model. Furthermore, the ECFGM model compared with the six effective models, including the GM$(1,1)$, EXGM$(1,1)$, FAGM$(1,1)$, FAGMO$(1,1,k)$, PFAGM$(1,1)$ and CFGM$(1,1)$. The response function of the grey models can be given as follows.

The response function of standard GM$(1,1)$ model is obtained as

$$(5.1) \qquad x^{(1)}(k) = 56.12166e^{0.22483(k-1)} - 49.87166.$$

The response function of EXGM$(1,1)$ is

$$(5.2) \qquad x^{(1)}(k) = 67.78567e^{-0.20588(1-k)} - 70.97226.$$

For FAGM$(1,1)$ model, the response function is

$$(5.3) \qquad x^{(0.36872)}(k) = 37.55924e^{0.17072(k-1)} - 31.35924.$$

The response function of FAGMO$(1,1,k)$ is

$$(5.4) \qquad x^{(1.13366)}(k) = 283.1282e^{0.13851(k-1)} - 30.4471k - 246.4811.$$

The response function of PFAGM$(1,1)$ is,

$$(5.5) \qquad x^{(0.17874)}(k) = 19.69290e^{0.17874k} - 14.47094e^{-0.08124(k-1)} - 2.87604.$$

The response function of CFGM$(1,1)$ model is obtained as

$$(5.6) \qquad x^{(0.07)}(k) = 81.63469e^{0.06355(k-1)} - 75.38469.$$

Here, the optimal value of α is obtained as 0.07 by using the Brute Force strategy for the CFGM$(1,1)$.

The response function of ECFGM$(1,1)$ model is obtained as

$$(5.7) \qquad x^{(0.3319)}(k) = 66.64788e^{0.10331(k-1)} + 3.98468e^{-k} - 61.86377.$$

In addition, the mean absolute percentage error (MAPE) is used to assess the prediction accuracy of these grey models. Firstly, we split the raw sequence into two groups to build a model and test the model. The first group, including the consumption from 2009 to 2017, is used to build models for the seven grey models separately. The second group, including wind energy consumption from 2018 to 2020, is used to verify the prediction accuracy of these grey models. In this section we enumerate all the values in the interval $[0, 2]$ with step 0.0001, then use the computational steps presented in Section 3.4 and select the α corresponding to the minimum MAPE_{fit} as the optimal value. Optimum α is found as $\alpha = 0.3319$ by using the Brute Force strategy and values of α and calculated MAPEs are shown in Figure 5.
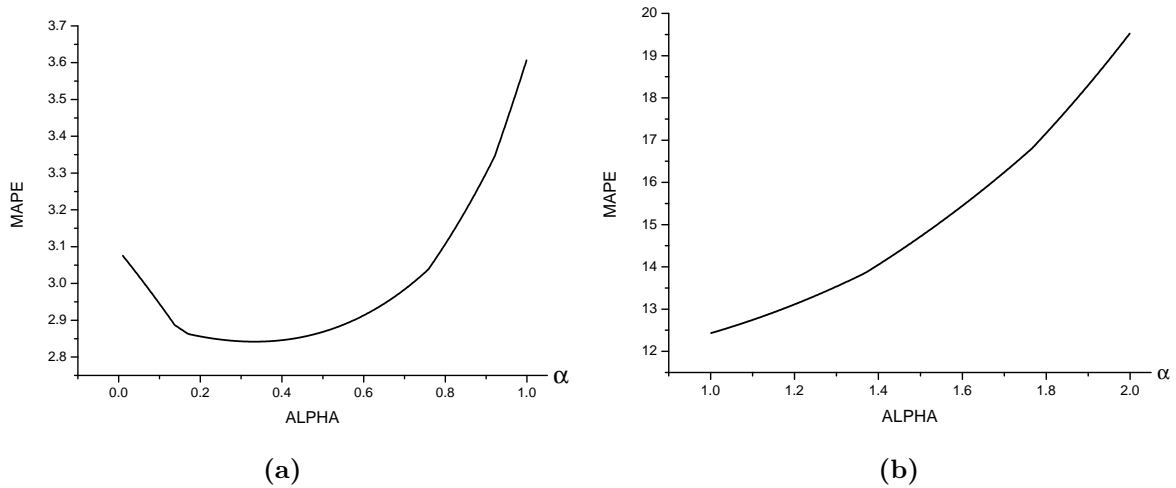
Figure 5: MAPEs of ECFGM model for the real data set. (a) $\alpha \in (0, 1]$; (b) $\alpha \in (1, 2]$.

From Figure 5 it is clear that the values of MAPE increase when the α moves away from its optimum value. The optimal parameters are calculated as $\alpha = 0.3319$, $a = -0.10331$, $b = 6.39111$ and $c = -4.39633$. The prediction results and the mean absolute percentage errors of the recent models are shown in Table 1.

Table 1:   The results generated by the proposed model and other comparative grey models for forecasting values of China's wind energy consumption (million tonnes oil equivalent).

Year	Actual Value	GM(1, 1)	EXGM(1, 1)	FAGM(1, 1)	FAGMO(1, 1, k)	PFAGM	CFGM	ECFGM
2009	6.25	6.2500	6.2500	6.2500	6.2500	6.2500	6.2500	6.2500
2010	10.10	14.1489	9.5314	10.9059	10.7861	10.8294	10.2050	10.0534
2011	15.91	17.7160	16.8448	15.9000	15.9521	15.9001	15.8554	16.0464
2012	21.72	22.1824	22.5845	21.4733	21.7572	21.5803	22.0785	22.2010
2013	31.95	27.7749	28.4425	27.8495	28.3244	28.0628	28.9531	28.8472
2014	35.32	34.7772	35.2004	35.2395	35.7864	35.5571	36.5538	36.2435
2015	42.03	43.5450	43.3416	43.8718	44.2900	44.3005	44.9565	44.5937
2016	53.64	54.5231	53.2845	54.0032	53.9997	54.5677	54.2405	54.0796
2017	66.75	68.2690	65.4787	65.9310	65.1024	66.6817	64.4903	64.8821
MAPE$_{\text{fit}}$		8.4111	3.6099	3.1595	3.1901	3.1104	3.0416	**2.8418**
2018	82.82	85.4804	80.4526	80.0030	77.8110	81.0239	75.7962	77.1931
2019	93.31	107.0310	98.8466	96.6289	92.3686	98.0478	88.2555	91.2232
2020	107.30	134.0148	121.4447	116.2921	109.0538	118.2925	101.9728	107.2063
MAPE$_{\text{pre}}$		14.2714	7.3248	5.1128	2.8971	5.8303	6.2875	**3.0392**
MAPE		9.8762	4.5386	3.6478	3.1169	3.7904	3.8531	**2.8912**

According to the Table 1, the essential conclusions can be drawn as follows:

i. Table 1 reveals that, the seven grey models' MAPE_{fit} values are 8.4111%, 3.6099%, 3.1595%, 3.1901%, 3.1104%, 3.0416% and **2.8418**%, respectively. So that, in the fitting period, the fitting performance of the ECFGM$(1,1)$ model is best.

ii. The MAPE_{pre} values of seven models are calculated as 14.2714%, 7.3248%, 5.1128%, 2.8971%, 5.8303%, 6.2875% and **3.0392**%, respectively. At this point, the FAGMO$(1,1,k)$ model has the smallest MAPE_{pre} value while the ECFGM$(1,1)$ model has the second smallest MAPE_{pre}.

iii. It is observed from Table 1 that for the whole period, the total MAPEs of seven models are 9.8762%, 4.5386%, 3.6478%, 3.1169%, 3.7904%, 3.8531% and **2.8912**%, respectively. Thence, in the whole period, the performance of the ECFGM$(1,1)$ model is the best (see Figure 6).
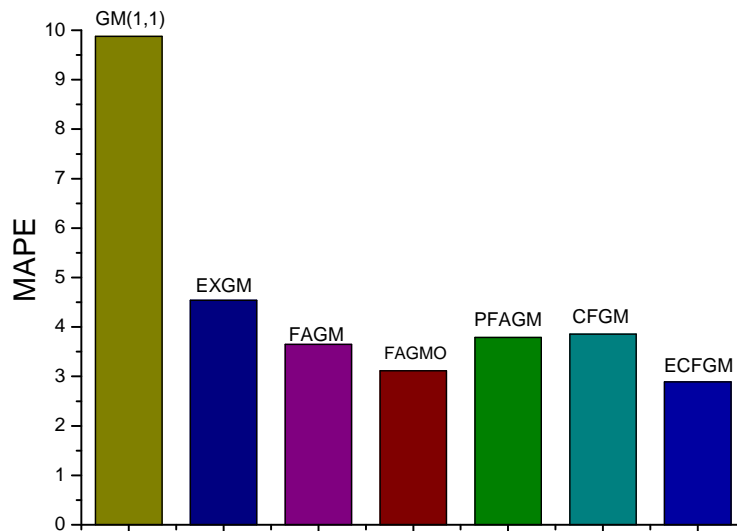
Figure 6: Total MAPE values of the models.

iv. The wind energy consumption of China can be estimated properly with the proposed ECFGM$(1,1)$ model. Hence, the forecasting values of wind energy consumption of China are given in Table 2.

Table 2: Forecasted wind energy consumption of China.

Years	Forecasting values of wind energy consumption
2021	125.4042
2022	146.1105
2023	169.6550
2024	196.4082
2025	226.7862

It is seen from Figure 7 that the actual (red line) and forecasted (black line) values of wind energy consumption of China are matched to each other.
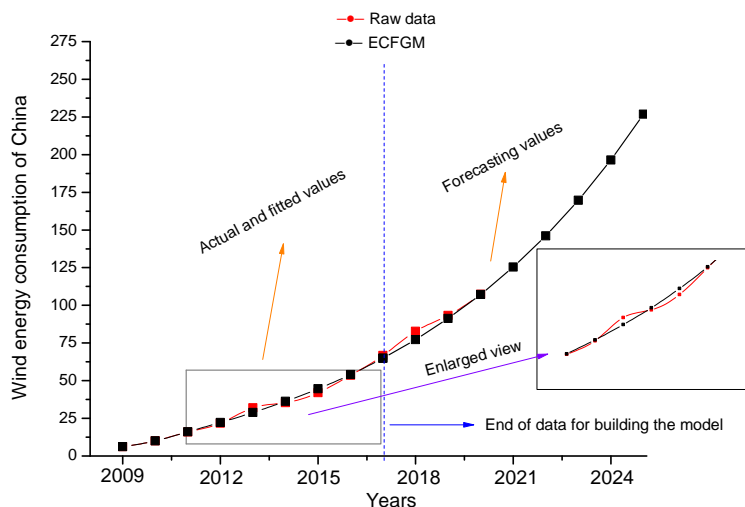
Figure 7: Actual, fitting and forecasting values of wind energy consumption of China.

6.   CONCLUSION

Since the fractional calculations are most important to the grey prediction model, there are many scholars proposing new methods on the fractional grey models. Hence, a novel optimization for the CFGM$(1,1)$ and EXGM$(1,1)$ models have been developed in this study. The results of the numerical examples indicated that the proposed grey prediction model aims to achieve very effective performance. The structural parameters (a, b and c) of the model can be dynamically adjusted according to the real world systems. The optimal value of fractional order, α, is calculated by using the brute-force approach. The proposed ECFGM$(1,1)$ model is suitable for predicting the data sequence with the characteristics of non-homogeneous exponential law. Comparison results indicate, ECFGM$(1,1)$ performs better than those achieved by the other grey models such as GM$(1,1)$, EXGM$(1,1)$, FAGM$(1,1)$, FAGMO$(1,1,k)$, PFAGM$(1,1)$ and CFGM$(1,1)$. However, they can all be employed for estimations.

Wind energy with cleanliness and pollution-free will have a positive attitude on global energy transformation. Because of this, research on more accurate prediction of wind energy consumption is quite important for wind power generation. Therefore, the wind energy consumption of China is predicted successfully by using a novel proposed fractional grey model based on the conformable fractional difference and conformable fractional accumulation in the paper. The forecasting results show that wind energy consumption of China will develop rapidly in recent years, and will reach approximately 200 million tones oil equivalent by 2023.

Using two examples and a case study in Sections 4 and 5, we show that the MAPE of the ECFGM$(1,1)$ model is very low.

The proposed ECFGM$(1,1)$ model may play an important role in enriching the theoretical system of grey forecasting theory and it can be used for other real cases of small sample forecasting in the future. Besides, the combination of other fractional forecasting models, especially for the time series with highly effective, is also an interesting direction for next studies.

ACKNOWLEDGMENTS

Conflict of interest

The author declares no conflict of interest in this paper.

REFERENCES

[1] BILGIL, H. (2021). New grey forecasting model with its application and computer code, *AIMS Mathematics*, **6**, 1497–1514.

[2] CAO, J.; LIU, L.; YANG, L. and XIE, S. (2020). Application of a novel fractional order grey support vector regression model to forecast wind energy consumption in China, *Journal of Advances in Mathematics and Computer Science*, **35**(2), 58–69.

[3] CUI, J.; LIU, S.; ZENG, B. and XIE, N. (2013). A novel grey forecasting model and its optimization, *Appl. Math. Model.*, **37**, 4399–4406.

[4] DENG, J.L. (1982). Control problems of grey systems, *Syst. Control Lett.*, **1**, 288–294.

[5] ENE, S. and OZTURK, N. (2018). Grey modelling based forecasting system for return flow of end-of-life vehicles, *Technol. Forecast. Soc. Change*, **117**, 155–166.

[6] CHINA ENERGY PORTAL (2021). 2020 electricity and other energy statistics (preliminary), *China Energy Portal*.
https://chinaenergyportal.org/2020-electricity-other-energy-statistics-preliminary/

[7] JAVED, S.A. and LIU, S. (2018). Predicting the research output/growth of selected countries: application of even GM(1, 1) and NDGM models, *Scientometrics*, **115**, 395–413.

[8] JIANG, J.; FENG, T. and LIU, C. (2021). An improved nonlinear grey Bernoulli model based on the whale optimization algorithm and its application, *Mathematical Problems in Engineering*, **2021**, 6691724.

[9] KHALIL, R.; AL HORANI, M.; ABDELRAHMAN, Y. and MOHAMMAD, S. (2014). A new definition of fractional derivative, *J. Comput. Appl. Math.*, **264**, 65–70.

[10] LI, S.; MA, X. and YANG, C. (2018). A novel structure-adaptive intelligent grey forecasting model with full-order time power terms and its application, *Computers & Industrial Engineering*, **120**, 53–67.

[11] LIU, L.; CHEN, Y. and WU, L. (2021). The damping accumulated grey model and its application, *Communications in Nonlinear Science and Numerical Simulation*, **95**, 105665.

[12] LIU, L. and WU, L. (2021). Forecasting the renewable energy consumption of the european countries by an adjacent non-homogeneous grey model, *Appl. Math. Modell.*, **89**, 1932–1948.

[13] MA, X. (2016). Research on a novel kernel based grey prediction model and its applications, *Mathematical Problems in Engineering*, **2016**, 1–9.

[14] MA, X.; LIU, Z. and WANG, Y. (2019). Application of a novel nonlinear multivariate grey Bernoulli model to predict the tourist income of China, *J. Comput. Appl. Math.*, **347**, 84–94.

[15] MA, X. and LIU, Z.B. (2017). Application of a novel time-delayed polynomial grey model to predict the natural gas consumption in China, *J. Comput. Appl. Math.*, **324**, 17–24.

[16] MA, X.; WU, W.; ZENG, B.; WANG, Y. and WU, X. (2020). The conformable fractional grey system model, *ISA Transactions*, **96**, 255–271.

[17] MAO, S.; GAO, M.; XIAO, X. and ZHU, M. (2016). A novel fractional grey system model and its application, *Appl. Math. Model*, **40**, 5063–5076.

[18] SAHIN, U. and SAHIN, T. (2020). Forecasting the cumulative number of confirmed cases of COVID-19 in Italy, UK and USA using fractional nonlinear grey Bernoulli model, *Chaos, Solitons & Fract.*, **138**, 109948.

[19] WANG, Z.X.; LI, Q. and PEI, L.L. (2018). A seasonal GM(1,1) model for forecasting the electricity consumption of the primary economic sectors, *Energy*, **154**, 522–534.

[20] WEI, B.; XIE, N. and HU, A. (2018). Optimal solution for novel grey polynomial prediction model, *Applied Mathematical Modelling*, **62**, 717–727.

[21] WEI, M.; LI, Q. and ZENG, B. (2016). Study on fractional order grey reducing generation operator, *Grey Syst Theory Application*, **6**, 80–95.

[22] WU, L.; LIU, S.; YANG, Y.; MA, L. and LIU, H. (2016). Multivariable weakening buffer operator and its application, *Inform. Sci.*, **339**, 98–107.

[23] WU, L.; LIU, S.; YAO, L.; YAN, S. and LIU, D. (2013). Grey system model with the fractional order accumulation, *Commun. Nonlinear Sci. Numer. Simul.*, **18**, 1775–1785.

[24] WU, L.F.; LIU, S.F.; CHEN, D.; YAO, L. and CUI, W. (2014). Using gray model with fractional order accumulation to predict gas emission, *Nat. Hazards*, **71**, 2231–2236.

[25] WU, L.Z.; LI, S.H.; HUANG, R.Q. and XI, Q. (2020). A new grey prediction model and its application to predicting landslide displacement, *Appl. Soft Comput.*, **95**, 106543.

[26] WU, W.; MA, X.; WANG, Y.; CAI, W. and ZENG, B. (2020). Predicting China's energy consumption using a novel grey Riccati model, *Appl. Soft Comput.*, **95**, 106555.

[27] WU, W.; MA, X.; ZENG, B.; WANG, Y. and CAI, W. (2018). Application of the novel fractional grey model FAGMO(1,1,k) to predict China's nuclear energy consumption, *Energy*, **165**, 223–234.

[28] WU, W.; MA, X.; ZHANG, Y.; LI, W. and WANG, Y. (2020). A novel conformable fractional non-homogeneous grey model for forecasting carbon dioxide emissions of BRICS countries, *Sci. Total Environ.*, **707**, 135447.

[29] XIE, W.; CAIXIA, L.; WU, W.; WEIDONG, L. and CHONG, L. (2020). Continuous grey model with conformable fractional derivative, *Chaos, Solitons & Fractals*, **139**, 110285.

[30] XIE, W.; WU, W.Z.; LIU, C. and ZHAO, J. (2020). Forecasting annual electricity consumption in China by employing a conformable fractional grey model in opposite direction, *Energy*, **202**, 117682.

[31] YANG, Y. and XUE, D. (2016). Continuous fractional-order grey model and electricity prediction research based on the observation error feedback, *Energy*, **115**, 722–733.

[32] YUXIAO, K.; SHUHUA, M.; YONGHONG, Z. and HUIMIN, Z. (2020). Fractional derivative multivariable grey model for nonstationary sequence and its application, *Journal of Systems Engineering and Electronics*, **31**, 1009–1018.

[33] ZENG, B.; TAN, Y.; XU, H.; QUAN, J.; WANG, L. and ZHOU, X. (2018). Forecasting the electricity consumption of commercial sector in Hong Kong using a novel grey dynamic prediction model, *J. Grey Syst.*, **30**, 157–172.

[34] ZHANG, P.; MA, X. and SHE, K. (2019). A novel power-driven fractional accumulated grey model and its application in forecasting wind energy consumption of China, *Plos One*, **14**, e0225362.

[35] ZHANG, Y.G.; XU, Y. and WANG, Z.P. (2009). GM(1,1) grey prediction of Lorenz chaotic system, *Chaos, Solitons & Fract.*, **42**, 1003–1009.

[36] ZHOU, W. and HE, J.M. (2013). Generalized GM(1,1) model and its application in forecasting of fuel production, *Appl. Math. Modell.*, **37**, 6234–6243.

Robust Estimation of Component Reliability
Based on System Lifetime Data with Known Signature

Authors: XIAOJIE ZHU
            – Department of Statistical Science, Southern Methodist University,
            Dallas, Texas 75275-0332, U.S.A.
            xiaojiez@mail.smu.edu

            HON KEUNG TONY NG [ID] [✉]
            – Department of Mathematical Sciences, Bentley University,
            Waltham, Massachusetts 02452, U.S.A.
            tng@bentley.edu

            PING SHING CHAN [ID]
            – Department of Statistics, The Chinese University of Hong Kong,
            Shatin, N.T., Hong Kong
            benchan@cuhk.edu.hk

Abstract:

• This paper considers the estimation of component reliability based on system lifetime data with known system signature using the minimum density divergence estimation method. Different estimation procedures based on the minimum density divergence estimation method are proposed. Standard error estimation and interval estimation procedures are also studied. Then, a Monte Carlo simulation study is used to evaluate the performance of those proposed procedures and compare those procedures with the maximum likelihood estimation method under different contaminated models. A numerical example is presented to illustrate the effectiveness of the proposed minimum density divergence estimation method. We have shown that the proposed estimation procedures are robust to contamination and model misspecification. Finally, concluding remarks with some possible future research directions are provided.

✉ Corresponding author

1.   INTRODUCTION

System lifetime data are commonly encountered in industrial and engineering settings. In system reliability studies, engineers are always interested in the lifetime distribution of the system as well as the lifetime distribution of the components which make up the system. We consider here the situation that the lifetimes of an n-component system can be observed but not the lifetime of the components. This situation occurs when putting the individual component on a life testing experiment after the n-component system is built is not possible, or when the distribution of the component lifetimes changes while they are used in a specified system. Suppose the lifetimes of the n components in an n-component system are independent and identically distributed (i.i.d.) random variables, denoted as $X_1, X_2, ..., X_n$, with probability density function (p.d.f.) $f_X(t; \boldsymbol{\theta})$, cumulative distribution function (c.d.f.) $F_X(t; \boldsymbol{\theta})$ and survival function (s.f.) $\bar{F}_X(t; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the parameter vector. We further denote the ordered component lifetimes within an n-component system as $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ with $X_{i:n}$ be the i-th ordered component lifetime. Although the i.i.d. assumption is restrictive, there are many practical situations in which the i.i.d. assumption is applicable. For instance, Bhattacharya and Samaniego [3] discussed some of the practical examples that the i.i.d. assumption is reasonable such as batteries in a lighting device, wafers in a digital computer, and spark plugs in an automobile, and Jin *et al.* [13] discussed that the performance of "Redundant Array of Independent Disks (RAID)" computer hardware with n independent disks can be designed to perform like a k-out-of-n system.

When the component lifetime follows an absolutely continuous distribution, the failure time of an n-component system corresponds to the failure time of one of the n components. We consider the coherent system in which every component is relevant and the system has a monotone structure function [7]. In a coherent system consists of n i.i.d. components, the system structure can be described by the system signature defined as an n-element probability vector $\boldsymbol{s} = (s_1, s_2, ..., s_n)$, where the i-th element is the probability that the i-th ordered component failure causes the failure of the system [24], i.e.,

$$s_i = \Pr(T = X_{i:n}), \qquad i = 1, 2, ..., n.$$

Note that the system signature is only depending on the system structure and hence, it is distribution-free. To illustrate the idea of system signature, we consider the 4-component series-parallel III system with system lifetime $T = \min\{X_1, \max\{X_2, X_3, X_4\}\}$ (Figure 1a). For the 4-component series-parallel III system, there are $4! = 24$ possible arrangements of the component lifetimes. The 24 arrangements and their corresponding system lifetimes are presented in Table 1. From Table 1, we can obtain

$$
\begin{aligned}
s_1 &= \Pr(T = X_{1:4}) = 6/24 = 1/4, \\
s_2 &= \Pr(T = X_{2:4}) = 6/24 = 1/4, \\
s_3 &= \Pr(T = X_{3:4}) = 12/24 = 1/2, \\
\text{and} \quad s_4 &= \Pr(T = X_{4:4}) = 0.
\end{aligned}
$$

Hence, the system signature of the 4-component series-parallel III system is $\boldsymbol{s} = (1/4, 1/4, 1/2, 0)$. Similarly, for the 4-component mixed parallel I system (Figure 1b), the system signature is $\boldsymbol{s} = (0, 1/2, 1/4, 1/4)$.

a. series-parallel III system   **b**. mixed parallel I system
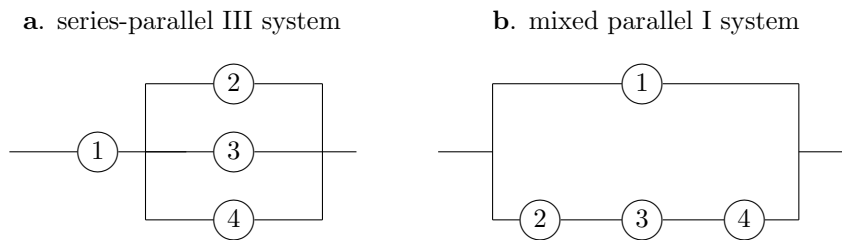
Figure 1: Two 4-component systems for illustration.

Table 1: The 24 possible arrangements of the component lifetime in a 4-component series-parallel III system.

Arrangement	System lifetime T	Arrangement	System lifetime T
$X_1 < X_2 < X_3 < X_4$	$X_{1:4}$	$X_3 < X_1 < X_4 < X_2$	$X_{2:4}$
$X_1 < X_2 < X_4 < X_3$	$X_{1:4}$	$X_3 < X_4 < X_1 < X_2$	$X_{3:4}$
$X_1 < X_4 < X_2 < X_3$	$X_{1:4}$	$X_3 < X_1 < X_2 < X_4$	$X_{2:4}$
$X_1 < X_4 < X_3 < X_2$	$X_{1:4}$	$X_3 < X_4 < X_1 < X_2$	$X_{3:4}$
$X_1 < X_3 < X_2 < X_4$	$X_{1:4}$	$X_3 < X_2 < X_1 < X_4$	$X_{3:4}$
$X_1 < X_3 < X_4 < X_2$	$X_{1:4}$	$X_3 < X_2 < X_4 < X_1$	$X_{3:4}$
$X_2 < X_1 < X_3 < X_4$	$X_{2:4}$	$X_4 < X_1 < X_2 < X_3$	$X_{2:4}$
$X_2 < X_1 < X_4 < X_3$	$X_{2:4}$	$X_4 < X_1 < X_3 < X_2$	$X_{2:4}$
$X_2 < X_3 < X_1 < X_4$	$X_{3:4}$	$X_4 < X_2 < X_1 < X_3$	$X_{3:4}$
$X_2 < X_3 < X_4 < X_1$	$X_{3:4}$	$X_4 < X_2 < X_3 < X_1$	$X_{3:4}$
$X_2 < X_4 < X_1 < X_3$	$X_{3:4}$	$X_4 < X_3 < X_1 < X_2$	$X_{3:4}$
$X_2 < X_4 < X_3 < X_1$	$X_{3:4}$	$X_4 < X_3 < X_2 < X_1$	$X_{3:4}$

Given the system signature $\boldsymbol{s}$, the p.d.f. and s.f. of the system lifetime T of an n-component system can be expressed as

$$(1.1) \qquad f_T(t; \boldsymbol{\theta}) = \sum_{i=1}^{n} s_i \binom{n}{i} i f_X(t; \boldsymbol{\theta}) [F_X(t; \boldsymbol{\theta})]^{i-1} \big[\bar{F}_X(t; \boldsymbol{\theta})\big]^{n-i}$$

$$(1.2) \qquad \text{and} \quad \bar{F}_T(t; \boldsymbol{\theta}) = \sum_{i=1}^{n} s_i \sum_{j=0}^{i-1} [F_X(t; \boldsymbol{\theta})]^{j} \big[\bar{F}_X(t; \boldsymbol{\theta})\big]^{n-j},$$

respectively [15]. Based on system lifetimes with known system signatures, the statistical inference of the component lifetime distribution have been discussed in the literature. Balakrishnan *et al.* [1] developed an exact nonparametric inference for population quantiles and tolerance limits of component lifetime distribution in a system. Balakrishnan *et al.* [2] derived the best linear unbiased estimator (BLUE) for the parameters in the component lifetime distribution. Navarro *et al.* [21] discussed the method of moments, the maximum likelihood method and the least squares methods for system lifetime data under a proportional hazard rate model. Chahkandi *et al.* [8] proposed several nonparametric methods to construct prediction intervals for the lifetime of coherent systems. Zhang *et al.* [28] proposed a regression-based estimation method for the model parameters of component lifetime

distribution based on censored system failure data. Yang *et al.* [26] proposed a stochastic expectation-maximization (EM) algorithm to obtain an approximation of the maximum likelihood estimates (MLEs) of the parameters in component lifetime distribution. Recently, Yang *et al.* [27] and Hermanns *et al.* [12] considered the EM algorithm to obtain the MLEs of the parameters in component lifetime distribution based on system lifetime data when the system structure is unknown. The theory and applications of system signatures are an active research area. For a comprehensive review and bibliometric analysis on system signatures, one can refer to a recent paper by Naqvi *et al.* [19].

In industrial experiments on systems, there are many situations in which systems are removed from experimentation before the occurrence of the failure of the system. Two common reasons for pre-planned censoring are saving the time on tests and reducing the cost associated with the experiment because failure implies the destruction of a system which may be costly [9, 18]. In this paper, we consider Type-II right censoring scheme in which the number of observed failures is pre-specified as r and the experiment is terminated as soon as the r-th ordered system failure is observed. Several studies on the Type-II censored system lifetime data in a system with system signature have been conducted [2, 12, 21, 28, 26, 27].

In the manufacturing industry, defectives could be induced in the manufacturing process due to different reasons such as human error, insufficient quality control, and failure in addressing reliability aspects during the design stage, etc.. As Raina [22] pointed out, zero-defect is an impossible goal to achieve or cost-prohibitive in the manufacturing process. Manufacturing defects often lead to potential outliers or contamination of the lifetime data. When there is outliers exist in observed lifetime data, the performance of the maximum likelihood or other classical estimation methods may be affected and a poor estimate of the component reliability characteristics may be yielded. Note that the maximum likelihood estimation is sensitive to the outliers as each observation contributes equal information to the estimate. Therefore, it is desired to develop parameter estimation procedures that are less sensitive to contaminated observations. Basu *et al.* [4] developed a family of density-based divergences measures with a single parameter α that controls the trade-off between robustness and efficiency, and proposed a procedure for estimating model parameters based on minimizing the density divergence. Basu *et al.* [5] further extended the minimum density divergence procedure to censored survival data with and without contamination, and found that the minimum density divergence estimator (MDE) is superior to the MLE when there is contamination in the censored survival data. Recently, Riani *et al.* [23] developed an alternative minimum density power divergence estimation procedure using the methods of S-estimation. Basak *et al.* [6] proposed a procedure to determinate the optimal density power divergence tuning parameter.

In this paper, we study the robust minimum density divergence estimation method for the system lifetime with and without contamination. In Section 2, we introduced the minimum density power divergence and its application to system lifetime data with known system signatures. We also discuss the estimation of the standard error of the estimate and interval estimation, and we show that the bootstrap method for standard error estimation can be adopted for the MDEs. In Section 3, a numerical example is used to illustrate the proposed MDEs. A Monte Carlo simulation study is presented in Section 4 to study the performance of the proposed methodologies. Finally, some concluding remarks and possible extensions are provided in Section 5.

2. MINIMUM DENSITY DIVERGENCE ESTIMATOR FOR SYSTEM LIFE-TIME DATA

2.1. Minimum density divergence estimator

The density power divergence, proposed by Basu *et al.* [4], describes a family of density-based divergence measures between two p.d.f.s $g(t)$ and $f(t)$ with a single parameter α. Consider that $f(t; \boldsymbol{\theta})$ is a parametric p.d.f. of the fitted model with parameter vector $\boldsymbol{\theta}$ and $g(t)$ is the target p.d.f., the density power divergence between $f(t; \boldsymbol{\theta})$ and $g(t)$ is defined as

$$d_\alpha(g, f) = \int \left[f^{1+\alpha}(t; \boldsymbol{\theta}) - \left(1 + \frac{1}{\alpha} \right) g(t) f^\alpha(t; \boldsymbol{\theta}) + \frac{1}{\alpha} g^{1+\alpha}(t) \right] dt, \quad \alpha > 0$$

(2.1)

and

(2.2)
$$d_0(g, f) = \lim_{\alpha \to 0} d_\alpha(g, f) = \int g(t) \ln \left[\frac{g(t)}{f(t; \boldsymbol{\theta})} \right] dt.$$

$d_\alpha(g, f) = 0$ when $f(t; \boldsymbol{\theta}) = g(t)$. The MDE of the parameter vector $\boldsymbol{\theta}$ can be obtained by minimizing the density power divergence between $f(t; \boldsymbol{\theta})$ and $g(t)$ with respect to (w.r.t.) $\boldsymbol{\theta}$. Since the term $\int \left[(1/\alpha) g^{1+\alpha}(t) \right] dt$ in (2.1) does not depend on the parameter vector $\boldsymbol{\theta}$, the minimum divergence estimator of $\boldsymbol{\theta}$ can be obtained by minimizing

(2.3)
$$\int \left[f^{1+\alpha}(t; \boldsymbol{\theta}) - \left(1 + \frac{1}{\alpha} \right) g(t) f^\alpha(t; \boldsymbol{\theta}) \right] dt$$

w.r.t. $\boldsymbol{\theta}$.

The density power divergence reduces to the Kullack–Leibler divergence [16] when $\alpha = 0$, and is the mean squared error when $\alpha = 1$. Hence, the minimum density power divergence procedure is degenerated into the maximum likelihood method when $\alpha = 0$, and becomes the minimization of the mean squared error when $\alpha = 1$. The parameter α in (2.1) controls the trade-off between robustness and efficiency of the minimum divergence estimator [4, 5]. It has been shown that the typical value of α is in between 0 and 1 and the estimation procedure becomes less efficient as α increases [4]. Hence, in this paper, we consider the value of α in $(0, 1)$.

Basu *et al.* [5] proposed a method for using the empirical c.d.f. $\hat{G}_n$ to estimate the target distribution G to obtain

$$\int \left[f^{1+\alpha}(t; \boldsymbol{\theta}) - (1 + 1/\alpha) g(t) f^\alpha(t; \boldsymbol{\theta}) \right] dt = \int f^{1+\alpha}(t; \boldsymbol{\theta}) dt - \int (1 + 1/\alpha) f^\alpha(t; \boldsymbol{\theta}) dG(t)$$

$$\approx \int f^{1+\alpha}(t; \boldsymbol{\theta}) dt - \int (1 + 1/\alpha) f^\alpha(t; \boldsymbol{\theta}) d\hat{G}(t).$$

Suppose that in a life testing experiment with m independent n-component systems and a Type-II censored system lifetime data $T_{1:m} < T_{2:m} < \cdots < T_{r:m}$ $(r < m)$ is observed, the empirical c.d.f. of the system lifetime, $\hat{G}_T(t)$, can be obtained by using the Kaplan–Meier (K-M)

estimator of the survival function $\hat{S}_T(t) = 1 - \hat{G}_T(t)$ [14] based on the Type-II censored system lifetime data. Then, the MDE of $\boldsymbol{\theta}$ can be obtained at the system level by minimizing

$$(2.4) \qquad \hat{d}_\alpha(g, f) = \int f_T^{1+\alpha}(t; \boldsymbol{\theta}) dt - \int \left(1 + \frac{1}{\alpha}\right) f_T^\alpha(t; \boldsymbol{\theta}) d\hat{G}_T(t)$$

w.r.t. $\boldsymbol{\theta}$. As this minimization is carried out at the system lifetime level, this estimator is named as the MDE at system lifetime level, denoted as MDE_S.

In addition to the MDE at system lifetime level, the MDE can be considered at the component level. Based on the K-M estimator of the survival function of the system lifetime $\hat{S}_T(t)$, a nonparametric empirical distribution of the component lifetime distribution $\hat{G}_X(t)$ can be obtained based on the relationship between F_T and F_X in (1.2) as

$$\hat{S}_T(t) = \sum_{i=1}^n s_i \sum_{j=0}^{i-1} \left[\hat{G}_X(t)\right]^j \left[1 - \hat{G}_X(t)\right]^{n-j}.$$

Then, the model parameter $\boldsymbol{\theta}$ can be estimated by minimizing the density power divergence at component lifetime distribution

$$(2.5) \qquad \hat{d}_\alpha(g, f) = \int f_X^{1+\alpha}(t; \boldsymbol{\theta}) dt - \int \left(1 + \frac{1}{\alpha}\right) f_X^\alpha(t; \boldsymbol{\theta}) d\hat{G}_X(t).$$

Since the MDE is obtained based on component lifetime distribution, we refer to the estimator obtained by minimizing (2.5) as the MDE at the component lifetime level, denoted as MDE_C.

Instead of estimating the c.d.f. nonparametrically, we consider the nonparametric kernel density estimator to estimate the p.d.f. of system lifetime $g_T(t)$ [25]. With the observed Type-II censored system lifetime data, the p.d.f. of system lifetime can be estimated using the Gaussian kernel density estimator, denoted as $\hat{g}_T(t)$. Then, the density power divergence function can be expressed as

$$(2.6) \qquad \hat{d}_\alpha(g, f) = \int f_T^{1+\alpha}(t; \boldsymbol{\theta}) dt - \int \left(1 + \frac{1}{\alpha}\right) f_T^\alpha(t; \boldsymbol{\theta}) \hat{g}_T(t) dt.$$

A MDE of $\boldsymbol{\theta}$ can be obtained by minimizing the density power divergence in (2.6) with the estimated kernel density $\hat{g}_T(t)$ w.r.t. $\boldsymbol{\theta}$. We name the MDE obtained by minimizing (2.6) as the MDE with estimated p.d.f., denoted as MDE_P.

For comparative purposes, we also consider the MLE of $\boldsymbol{\theta}$ based on Type-II censored system lifetime data. The log-likelihood function based on the observed Type-II censored system lifetime data $t_{1:m} < t_{2:m} < \cdots < t_{r:m}$ is

$$(2.7) \qquad \ln L(\boldsymbol{\theta}|t_{1:m}, t_{2:m}...t_{r:m}) = \sum_{k=1}^r \ln f_T(t_{k:m}; \boldsymbol{\theta}) + (m-r) \ln \bar{F}_T(t_{r:m}; \boldsymbol{\theta}),$$

where $r \leq m$ is the number of observed system failures and m is the total number of systems on the test. The MLE of $\boldsymbol{\theta}$ can be obtained by maximizing the log-likelihood function in (2.7) w.r.t. $\boldsymbol{\theta}$.

2.2. Standard error estimation and confidence intervals

2.2.1. Based on the theoretical results from Basu *et al.* [4]

For the MDE, Theorem 2.2 in [4] proved that under some regularity conditions, the MDE of the parameter $\boldsymbol{\theta}$ (denoted as $\hat{\boldsymbol{\theta}}$) is a consistent estimator for $\boldsymbol{\theta}$, and $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically multivariate normally distributed with zero mean and variance-covariance matrix $J^{-1}KJ^{-1}$, where

$$
\begin{aligned}
J = & \int \boldsymbol{u}_\theta(t)\boldsymbol{u}_\theta^\top(t)f^{1+\alpha}(t,\boldsymbol{\theta})dt \\
& + \int \left[\boldsymbol{i}_\theta(t) - \alpha\boldsymbol{u}_\theta(t)\boldsymbol{u}_\theta^\top(t) \right][g(t) - f(t;\boldsymbol{\theta})]f^\alpha(t;\boldsymbol{\theta})dt
\end{aligned}
$$
(2.8)

and

$$
\begin{aligned}
K = & \int \boldsymbol{u}_\theta(t)\boldsymbol{u}_\theta^\top(t)f^{2\alpha}(t;\boldsymbol{\theta})g(t)dt \\
& - \left[\int \boldsymbol{u}_\theta f^\alpha(t;\boldsymbol{\theta})g(t)dt \right]\left[\int \boldsymbol{u}_\theta f^\alpha(t;\boldsymbol{\theta})g(t)dt \right]^\top,
\end{aligned}
$$
(2.9)

with $\boldsymbol{u}_\theta(t) = \partial \ln f(t;\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, and $\boldsymbol{i}_\theta(t) = -\partial\boldsymbol{u}_\theta(t)/\partial\boldsymbol{\theta}$. Basu *et al.* [5] further proved that the asymptotic property of the MDE holds for censored survival data as well. Based on these results, the variance of the MDE can be approximated by discretizing the integrals in (2.8) and (2.9) with the nonparametric estimated c.d.f. $\hat{G}(t)$ or the nonparametric estimated p.d.f. $\hat{g}(t)$. Considering the estimator MDE$_S$, $\boldsymbol{\theta}_S$, the standard error of the MDE$_S$ can be approximated as

$$
\widehat{\mathrm{SE}}_A(\boldsymbol{\theta}_S) = \sqrt{\hat{J}_S^{-1}\hat{K}_S\hat{J}_S^{-1}/n},
$$
(2.10)

where

$$
\begin{aligned}
\hat{J}_S = & \int \left[(1+\alpha)\boldsymbol{u}_{\hat{\theta}_S}(t)\boldsymbol{u}_{\hat{\theta}_S}^\top(t) - \boldsymbol{i}_{\hat{\theta}_S}(t) \right]f_T^{1+\alpha}(t,\hat{\boldsymbol{\theta}}_S)dt \\
& + \int \left[i_{\hat{\theta}_S}(t) - \alpha\boldsymbol{u}_{\hat{\theta}_S}(t)\boldsymbol{u}_{\hat{\theta}_S}^\top(t) \right]f_T^\alpha(t;\hat{\boldsymbol{\theta}}_S)d\hat{G}_T(t)
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{K}_S = & \int \boldsymbol{u}_{\hat{\theta}_S}(t)\boldsymbol{u}_{\hat{\theta}_S}^\top(t)f_T^{2\alpha}(t;\hat{\boldsymbol{\theta}}_S)d\hat{G}_T \\
& - \left[\int \boldsymbol{u}_{\hat{\theta}_S}(t)f_T^\alpha(t;\hat{\boldsymbol{\theta}}_S)d\hat{G}_T \right]\left[\int \boldsymbol{u}_{\hat{\theta}_S}f_T^\alpha(t;\hat{\boldsymbol{\theta}}_S)d\hat{G}_T \right]^\top,
\end{aligned}
$$

where

$$
\boldsymbol{u}_{\hat{\theta}_S}(t) = \frac{1}{f_T(t;\boldsymbol{\theta})}\frac{\partial f_T(t;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_S} \quad \text{and} \quad \boldsymbol{i}_{\hat{\theta}_S}(t) = -\frac{\partial\boldsymbol{u}_\theta(t)}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_S}.
$$

The variance-covariance matrices of the estimators MDE$_C$ and MDE$_P$, $\hat{\boldsymbol{\theta}}_C$ and $\hat{\boldsymbol{\theta}}_P$, can be obtained in a similar manner.

2.2.2. Based on Fisher information matrix of the MLE

In our preliminary study (results are not presented here), we found that the performance of the standard error estimation based on the theoretical results in [5] may not be satisfactory. Therefore, we consider different ways to approximate the standard error of the MDE proposed in this paper. Based on our observations in the preliminary study, the standard error of the MLE and the standard error of the MDE is in the same order of magnitude, especially when the value of α is close to 0. Hence, we consider a standard error estimation method based on the Fisher information matrix similar to using the inverse of observed Fisher information matrix in estimating the standard error of MLE. For the MLE of $\boldsymbol{\theta}$, the asymptotic variance-covariance matrix of the MLE can be approximated by the inverse of the observed Fisher information matrix, i.e.,

$$\widehat{\mathrm{SE}}_F(\hat{\boldsymbol{\theta}}) = \sqrt{\widehat{\mathrm{Var}(\hat{\boldsymbol{\theta}})}} = \sqrt{\mathrm{diag}\left(\left[-\frac{\partial^2 \ln L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right]^{-1}\right)},$$

where $\mathrm{diag}(A)$ denotes the diagonal elements of matrix A. According to the asymptotic theory of the MLE, the sampling distribution of $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically multivariate normally distributed with mean zero and variance $\mathrm{Var}(\hat{\boldsymbol{\theta}})$. When $\alpha = 0$, the MDE is equivalent to the MLE. Here, we propose to approximate the variance of the MDEs by inverting the observed Fisher information by substituting $\boldsymbol{\theta}$ with its MDE.

2.2.3. Based on bootstrap method

As we expected, when the value of α is far from zero, the performance of the approximation based on the Fisher information matrix may not fulfilling the expectations, therefore, we also consider approximating the standard error of the MDE based on the bootstrap method. Given the estimated parameters, parametric bootstrap samples of system lifetimes are generated with the corresponding censoring proportion. For each bootstrap sample, the MDE is obtained as a bootstrap MDE. Based on B bootstrap MDEs, we compute the standard deviation of those bootstrap MDEs as an approximation of the standard error of the MDE. For instance, consider the MDE based on system-level data, suppose we have B bootstrap samples and the B bootstrap MDEs are $\hat{\boldsymbol{\theta}}_S^{(1)}, \hat{\boldsymbol{\theta}}_S^{(2)}, ..., \hat{\boldsymbol{\theta}}_S^{(B)}$, the standard error of the estimator $\hat{\boldsymbol{\theta}}_S$ can be approximated as

$$(2.11) \qquad \widehat{\mathrm{SE}}_B(\hat{\boldsymbol{\theta}}_S) = \sqrt{\frac{1}{B}\sum_{b=1}^{B}(\hat{\boldsymbol{\theta}}_S^{(b)} - \bar{\hat{\boldsymbol{\theta}}}_S)^2},$$

where $\bar{\hat{\boldsymbol{\theta}}}_S = \sum_{b=1}^{B}\hat{\boldsymbol{\theta}}_S^{(b)}/B$. The size of bootstrap samples needed will be discussed in Section 3 based on a Monte Carlo simulation study.

After obtaining the standard error estimate based on the methods described in Sections 2.2.1, 2.2.2, and 2.2.3, a two-sided $100(1-\alpha)\%$ normal approximated confidence interval of the k-th element of the parameter vector $\boldsymbol{\theta}$ can be obtained as

$$[\theta_{kl}, \theta_{ku}] = \left[\hat{\theta}_k - z_{1-\alpha/2}\widehat{\mathrm{SE}}(\hat{\theta}_k), \hat{\theta}_k + z_{1-\alpha/2}\widehat{\mathrm{SE}}(\hat{\theta}_k)\right],$$

where z_q is the q-th upper percentile of the standard normal distribution. The performance of the standard error estimation methods and the corresponding confidence intervals will be evaluated via a Monte Carlo simulation study in Section 3.

3.   MONTE CARLO SIMULATION STUDIES

In this section, Monte Carlo simulation studies are used to evaluate the performance of the proposed estimation methods for different system structures, different sample sizes with different censoring rates, different underlying distributions, and different values of α for the MDE and different contamination proportions. Based on our preliminary study, since similar observations are obtained based on different sample sizes, different system structures, and different underlying distributions, for the sake of simplicity, we only present the simulation results for the 4-component series-parallel III system (namely System I) and the 4-component mixed parallel I system (namely System II) in Figure 1 for sample size $m = 50$ (with different censoring rate) and the component lifetime X follows the two-parameter Weibull distribution with p.d.f.

$$(3.1) \qquad f_X(x; a, b) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1} \exp\left[-\left(\frac{x}{a}\right)^b\right], \quad x > 0,$$

where a is the scale parameter and b is the shape parameter (denoted as Weibull(a, b)). The Weibull distribution is considered here as it is one of the commonly used probability models in lifetime data analysis which can be used to model items with increasing, constant, and decreasing failure rates [17, 18]. Moreover, many other commonly used probability distributions such as the exponential distribution and the Rayleigh distribution are special cases of the Weibull distribution. We consider the scale parameter $a = 3$ or $a = 9$ and the shape parameter to be 2 ($b = 2$). For the case that the contaminates have a longer lifetime than the true distribution on average (namely the longer-life contamination model), the Weibull$(3, 2)$ distribution with a mean lifetime of 2.6587 is the true distribution, and the Weibull$(9, 2)$ distribution with mean lifetime 7.9760 is the contaminated distribution. Similarly, for the case that the contaminates have a shorter lifetime than the true distribution on average (namely the shorter-life contamination model), the Weibull$(9, 2)$ distribution is the true distribution and the Weibull$(3, 2)$ distribution is the contaminated distribution. We also consider other parameter settings; however, for the sake of brevity, we only present the results for Weibull$(9, 2)$ and Weibull$(3, 2)$ here. In the simulation study, the contamination proportion is set to be $0\%, 5\%, 10\%$ and 15%, the Type-II censoring rate $(1 - r/m)$ is set to be 0% and 5% (i.e., no censoring and $r = 0.95m$, respectively). The power parameter α in the MDE method is set to be $0.01, 0.1, 0.25, 0.5, 0.75$ and 0.9.

3.1. Results for point estimation

To evaluate the performance of the proposed estimation procedures for point estimation, the three proposed MDEs — MDE_S, MDE_C and MDE_P — are compared with the MLE in terms of their mean squared errors (MSEs) for estimating the mean component lifetime, i.e., $a\Gamma(1 + 1/b)$, where $\Gamma(\cdot)$ is the gamma function. Specifically, in the ℓ-th simulation, we first

estimate the parameter $\boldsymbol{\theta} = (a, b)$ based on different methods, denoted as $\hat{\boldsymbol{\theta}}_{(\ell)} = (\hat{a}_{(\ell)}, \hat{b}_{(\ell)})$ for Weibull(a, b) distribution, and then the estimated mean component lifetime is computed as $\hat{a}_{(\ell)}\Gamma(1 + 1/\hat{b}_{(\ell)})$. The MSE of an estimator is computed as

$$\frac{1}{L} \sum_{\ell=1}^{L} \left[\hat{a}_{(\ell)}\Gamma(1 + 1/\hat{b}_{(\ell)}) - a\Gamma(1 + 1/b)\right]^2.$$

The simulation results in this subsection are computed based on 10000 realizations ($L = 10000$). For comparative purposes, we define the relative efficiency of the MDE to MLE as

$$\mathrm{RE}_{\mathrm{MDE}} = \frac{\mathrm{MSE(MLE)}}{\mathrm{MSE(MDE)}}.$$

The value of relative efficiency greater than 1 indicates that the performance of the MDE is better than the MLE. The relative efficiency for different censoring rates, different contamination proportions, and different values of α for the combinations of System I and System II, and the longer-life contamination model and shorter-life contamination model, are plotted in Figures 2–5. From Figures 2–5, we observe that the performance of the MDE_C is the worst among the three proposed MDEs as the relative efficiency is below 1 in many cases. Therefore, we focus the discussion of the results below on the MDE_S and MDE_P.

In Figures 2 and 4, the relative efficiency of MDE_S, MDE_C and MDE_P for System I and System II with longer-life contamination model are presented, respectively. We can observe that MDE_S and MDE_P have similar performance for System I and System II. When there is no contamination (dashed lines with triangles in Figures 2 and 4), the relative efficiency is less than 1 for MDE_S and MDE_P, which indicates that the MLE performs better than MDE_S and MDE_P in terms of MSEs. When the contamination rate increases, the relative efficiency increases and becomes larger than 1 for MDE_S and MDE_P. Moreover, we observe that the performances of MDE_S and MDE_P improve when α gets closer to 1. These observations are consistent in both no censoring case (Figures 2 and 4(a)–(c)) and the 5% censoring case (Figures 2 and 4(d)–(f)). However, in the longer-life contamination model, the relative efficiency in the censoring case is smaller than those in the complete sample case. This indicates that Type-II censoring reduces the influence of the contamination in estimating the parameters. It is likely that the contaminated observations with a longer life are censored in the Type-II censoring scheme. For example, the relative efficiency of the MDE_S with $\alpha = 0.9$ is close to 15 when the contamination rate is 15% with no censoring, while the relative efficiency of the MDE_S with $\alpha = 0.9$ reduces to 10 when the contamination rate of 15% with 5% censoring.

In Figures 3 and 5, the relative efficiency of MDE_S, MDE_C and MDE_P for System I and System II with shorter-life contamination model are presented. We can observe that MDE_S and MDE_P have similar performance for System I and System II. In contrast to the longer-life contamination model, MDE_S and MDE_P have different performances in the shorter-life contamination model. In the complete sample case, the MDE_S and MDE_P have relative efficiency greater than 1 when the contamination rate is over 10% in most cases (Figures 3 and 5 (a) and (c)). In the Type-II censoring with 5% censoring case, the MDE_S has relative efficiency greater than 1 when the contamination rate is 15% and the value of α is close to 1 (Figures 3 and 5 (d)), while the MDE_P has relative efficiency less than 1 in most cases (Figures 3 and 5 (f)).
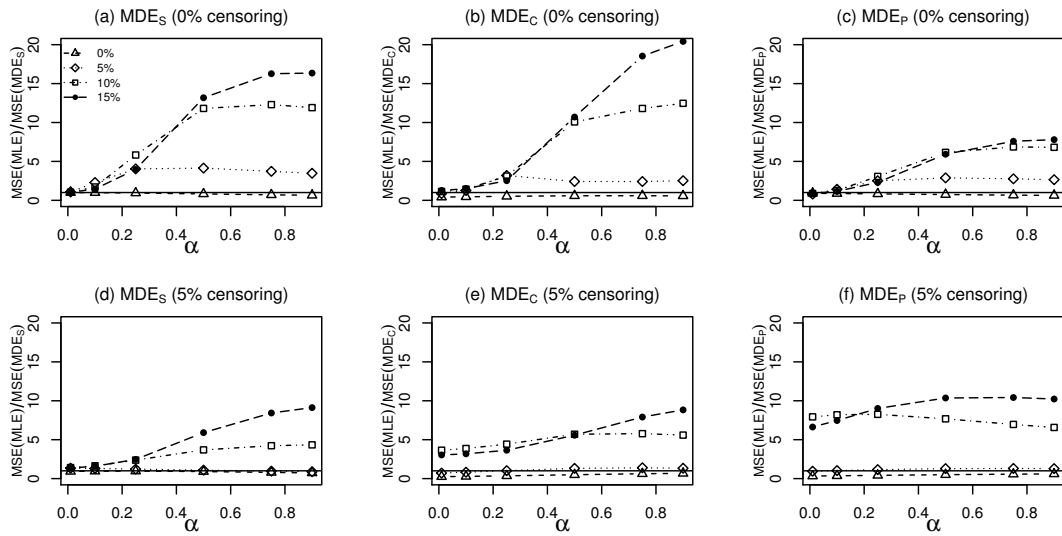
Figure 2: Relative efficiency of estimated mean component lifetime for System I with longer-life contamination model.
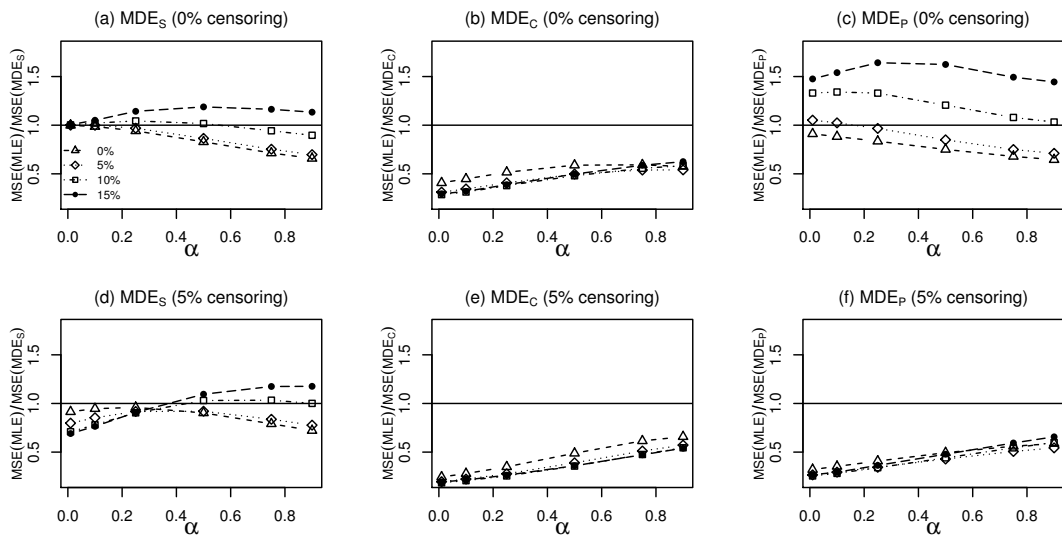
Figure 3: Relative efficiency of estimated mean component lifetime for System I with shorter-life contamination model.

In summary, the proposed estimator MDE_S has a better performance compared to MDE_C and MDE_P and it shows an advantage over the MLE when there is contamination present in the data. Moreover, the performance of MDE_S is not much worse than the MLE even when there is no contamination or with a low contamination rate (i.e., relative efficiency less than but close to 1). In the contamination cases, the value of α closer to 1 for the MDE_S has better performance. Therefore, we recommend the use of MDE_S, especially when it is suspected that there is contamination exists in the data. Based on these simulation results and for the simplicity sake, we consider the MDE_S but not the MDE_C and MDE_P in the subsequent study of the performance of standard error estimation and interval estimation.
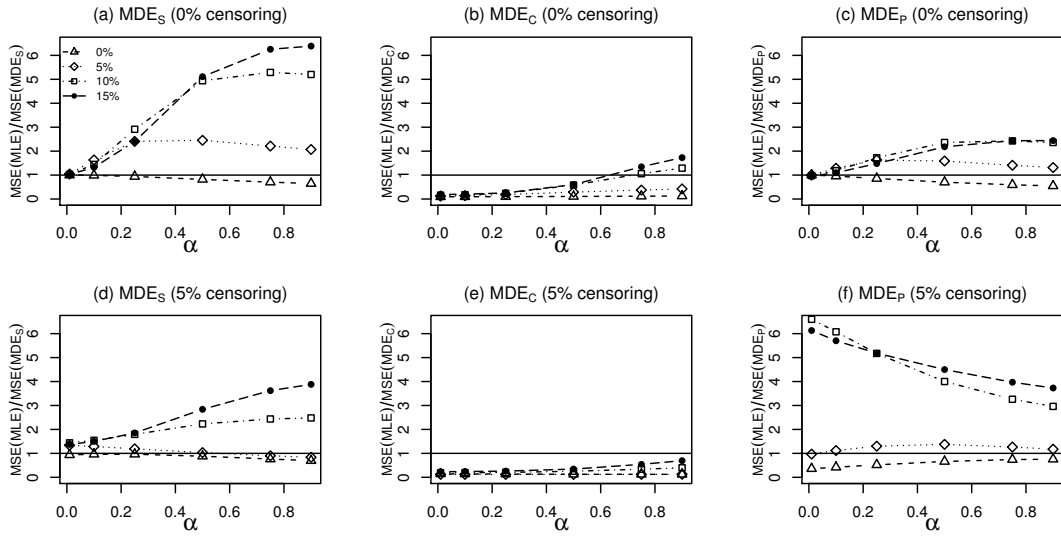
Figure 4: Relative efficiency of estimated mean component lifetime for
System II with longer-life contamination model.
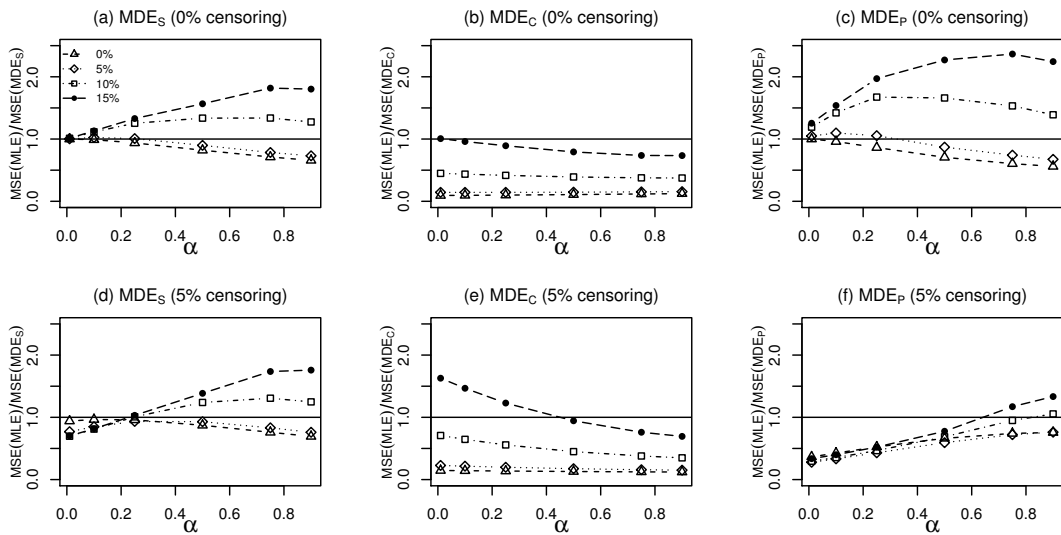
Figure 5: Relative efficiency of estimated mean component lifetime for
system II with shorter-life contamination model.

3.2. Results for standard error estimation and interval estimation

3.2.1. Determining a suitable bootstrap size for standard error estimation

To determine the required bootstrap size B for the standard error estimation for MDE described in Section 2.2.3, following Efron and Tibshirani [10], we consider evaluating the coefficient of variation of the standard error estimates to obtain a reasonable value of the number of bootstrap replicates. We consider the coefficient of variation of the standard

error estimates, which is computed as the ratio of the variance of the bootstrap estimate of standard error $\widehat{\text{SE}}_B$ to the expectation of $\widehat{\text{SE}}_B$ with different bootstrap size B. The variability of bootstrap estimates can be evaluated by using the coefficient of variation and a suitable value of B is a value such that the variability does not change significantly after increasing the value of B.

A Monte Carlo simulation is carried out to evaluate the coefficient of variation for different bootstrap sizes B in order to determine the proper number of bootstrap replications. We simulate 200 samples of $m = 50$ system lifetimes based on System I (4-component series-parallel III system) with true underlying component lifetime distribution Weibull$(3, 2)$, no contamination, and no censoring. For each simulation, given a bootstrap replication number B, the bootstrap standard error estimate of MDE_S is calculated, denoted as $\widehat{\text{SE}}_B$. Then, with the 200 bootstrap standard error estimates $\widehat{\text{SE}}_B^{(1)}, \widehat{\text{SE}}_B^{(2)}, ..., \widehat{\text{SE}}_B^{(200)}$, the simulated coefficient of variation is computed as:

$$\widehat{CV}(\widehat{\text{SE}}_B) = \frac{\widehat{\text{Var}}(\widehat{\text{SE}}_B)}{\hat{E}(\widehat{\text{SE}}_B)},$$

where

$$\hat{E}(\widehat{\text{SE}}_B) = \frac{1}{B}\sum_{i=1}^{200}\widehat{\text{SE}}_B^{(i)},$$

$$\text{and } \widehat{\text{Var}}(\widehat{\text{SE}}_B) = \frac{1}{B}\sum_{i=1}^{200}(\widehat{\text{SE}}_B^{(i)} - \hat{E}(\widehat{\text{SE}}_B))^2.$$

Figure 6 presented the simulated coefficient of variation of the standard error of MDE_S. From Figure 6, we observe that when the bootstrap size B gets above 250, a further increase in the bootstrap size does not bring a substantial reduction in the variation. Hence, we consider the number of bootstrap replications $B = 250$ in the Monte Carlo simulation study for evaluating the performance of confidence intervals.
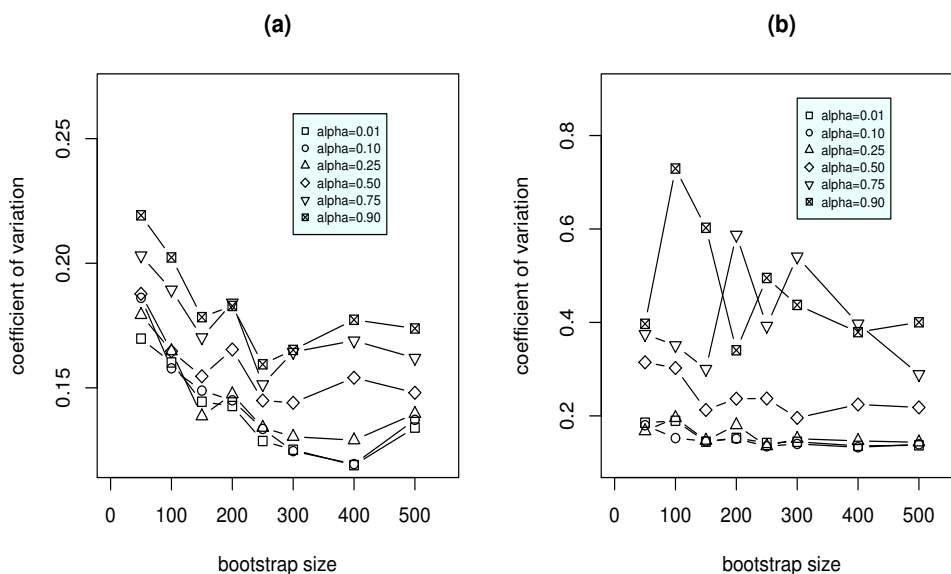
Figure 6: Coefficients of variation of $\widehat{\text{SE}}_B$ for (a) shape parameter and (b) scale parameter as functions of the number of bootstrap samples B.

3.2.2. Performance of standard error estimates

To evaluate the performance of the three standard error estimation methods for MDE presented in Section 2.2, we compare the simulated standard errors of the MDE based on the system-level data, MDE_S, and the averaged values of the standard error estimates based on the theoretical results from Basu *et al.* [4] (i.e., $\widehat{\text{SE}}_A$), based on observed Fisher information matrix (i.e., $\widehat{\text{SE}}_F$), and based on bootstrap method (i.e., $\widehat{\text{SE}}_B$) with bootstrap size $B = 250$. We simulate 1000 samples of $m = 50$ system lifetimes based on System 1 (4-component series-parallel III system) with true underlying component lifetime distribution Weibull$(3, 2)$, no contamination, and no censoring. The simulation results are presented in Table 2.

Table 2: Simulated standard errors the MDE_S and the averaged standard error estimates based on the theoretical results from [4] ($\widehat{\text{SE}}_A$), based on observed Fisher information matrix ($\widehat{\text{SE}}_F$), and based on bootstrap method ($\widehat{\text{SE}}_B$) with bootstrap size $B = 250$.

	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 0.9$
Simulated $\widehat{\text{SE}}(\hat{a})$	0.179	0.180	0.184	0.196	0.210	0.219
Average $\widehat{\text{SE}}_A(\hat{a})$	0.013	0.011	0.008	0.006	0.005	0.005
Average $\widehat{\text{SE}}_F(\hat{a})$	0.214	0.206	0.190	0.172	0.165	0.164
Average $\widehat{\text{SE}}_B(\hat{a})$	0.205	0.210	0.197	0.183	0.187	0.189
Simulated $\widehat{\text{SE}}(\hat{b})$	0.247	0.248	0.257	0.290	0.326	0.336
Average $\widehat{\text{SE}}_A(\hat{b})$	0.009	0.007	0.005	0.003	0.003	0.002
Average $\widehat{\text{SE}}_F(\hat{b})$	0.203	0.207	0.215	0.232	0.244	0.248
Average $\widehat{\text{SE}}_B(\hat{b})$	0.238	0.230	0.255	0.322	0.375	0.389

From Table 2, we observe that the standard error estimates based on the theoretical results from Basu *et al.* [4] can seriously underestimate the standard error of MDE_S, while the standard error estimates based on observed Fisher information matrix provide a reasonable approximation to the standard errors of MDE_S when α is close to 0. Overall, among the three standard error estimation methods for MDE, the bootstrap method with bootstrap size $B = 250$ provides a reasonable approximation to standard error of the MDE_S for all the values of α considered here. Therefore, in the following simulation study for confidence intervals, we use the standard error estimates based on the bootstrap method.

3.2.3. Performance of confidence intervals

In this subsection, the simulated coverage probabilities and the average widths of 95% confidence intervals of the Weibull parameters a and b for the MLE and the MDE based on system-level data (MDE_S) with different values of α are compared. The two systems (System I and System II) and the longer-life and shorter-life contamination models described in Section 3 are considered here. Specifically, a two-sided $100(1 - \alpha)\%$ normal approximated

confidence interval of a is constructed as

$$[a_l, a_u] = \left[\hat{a} - z_{1-\alpha/2}\widehat{\mathrm{SE}}(\hat{a}), \hat{a} + z_{1-\alpha/2}\widehat{\mathrm{SE}}(\hat{a})\right],$$

where the estimated standard error $\widehat{\mathrm{SE}}(\hat{a})$ is obtained based on the bootstrap method. Similarly, a two-sided $100(1-\alpha)\%$ normal approximated confidence interval of b is constructed as

$$[b_l, b_u] = \left[\hat{b} - z_{1-\alpha/2}\widehat{\mathrm{SE}}(\hat{b}), \hat{b} + z_{1-\alpha/2}\widehat{\mathrm{SE}}(\hat{b})\right],$$

where the estimated standard error $\widehat{\mathrm{SE}}(\hat{a})$ is obtained based on the bootstrap method. The simulated coverage probability (CP) is computed as the proportion of cases that the true value of the parameter falls within the confidence interval, and the average width (AW) is computed as $2z_{1-\alpha/2}\widehat{\mathrm{SE}}(\hat{a})$ and $2z_{1-\alpha/2}\widehat{\mathrm{SE}}(\hat{b})$ for parameters a and b, respectively. The simulation results are presented in Tables 3–4.

Table 3: Simulated coverage probabilities (in %) and average widths of confidence intervals of the scale parameter computed based on MLE and MDE$_S$ with different values of α under the longer-life and shorter-life contamination models for System I.

Coverage Probability	Longer-life Contamination Model								Shorter-life Contamination Model							
	No censoring				5% censoring				No censoring				5% censoring			
Contamination Proportion	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%
MLE	93.9	88.1	44.9	18.0	93.3	92.0	58.0	24.5	93.7	92.4	86.3	80.7	93.6	92.3	87.8	82.0
MDE$_S$ ($\alpha=0.01$)	93.9	92.3	59.1	31.7	87.9	93.1	79.5	50.6	93.6	92.2	86.0	80.0	88.6	84.7	75.3	67.5
MDE$_S$ ($\alpha=0.10$)	93.9	94.7	73.5	45.9	89.7	93.2	81.3	55.5	93.7	92.5	87.1	81.6	89.8	87.0	79.6	72.1
MDE$_S$ ($\alpha=0.25$)	93.9	95.2	93.1	82.6	91.3	93.2	86.1	70.2	93.3	91.8	83.3	73.2	90.9	89.2	82.8	73.6
MDE$_S$ ($\alpha=0.50$)	93.3	94.6	95.9	95.8	92.7	93.3	92.6	91.5	92.6	91.4	82.8	71.3	92.0	91.2	85.2	75.7
MDE$_S$ ($\alpha=0.75$)	92.9	94.0	95.4	95.9	93.1	93.6	94.8	95.5	92.1	90.5	81.4	68.8	92.2	91.4	85.2	74.8
MDE$_S$ ($\alpha=0.90$)	92.3	93.8	95.3	95.7	93.2	93.6	95.0	95.9	91.6	90.1	80.2	67.6	92.2	91.3	84.6	73.8

Average Width	Longer-life Contamination Model								Shorter-life Contamination Model							
	No censoring				5% censoring				No censoring				5% censoring			
Contamination Proportion	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%
MLE	0.69	0.99	1.32	1.50	0.70	0.77	1.06	1.33	2.06	2.14	2.24	2.30	2.09	2.18	2.30	2.37
MDE$_S$ ($\alpha=0.01$)	0.69	1.08	1.45	1.64	0.63	0.73	1.19	1.45	2.06	2.14	2.22	2.26	1.89	1.97	2.06	2.10
MDE$_S$ ($\alpha=0.10$)	0.69	0.93	1.33	1.54	0.64	0.74	1.14	1.41	2.08	2.17	2.27	2.32	1.92	2.02	2.12	2.18
MDE$_S$ ($\alpha=0.25$)	0.71	0.82	1.11	1.35	0.66	0.76	1.06	1.31	2.12	2.23	2.34	2.38	1.99	2.10	2.22	2.27
MDE$_S$ ($\alpha=0.50$)	0.75	0.82	0.98	1.15	0.71	0.80	1.04	1.23	2.24	2.38	2.52	2.57	2.12	2.25	2.40	2.46
MDE$_S$ ($\alpha=0.75$)	0.80	0.86	0.98	1.11	0.76	0.87	1.08	1.22	2.39	2.53	2.68	2.75	2.27	2.42	2.58	2.65
MDE$_S$ ($\alpha=0.90$)	0.83	0.89	1.00	1.11	0.80	0.91	1.10	1.21	2.47	2.61	2.78	2.85	2.38	2.52	2.70	2.77

From Tables 3–4, we observe that when there is no contamination, the confidence intervals based on MLEs give coverage probabilities close to the nominal 95% for both scale and shape parameters. Compared with MDE$_S$, the confidence intervals based on MLEs give the highest coverage probabilities and the smallest average widths when there is no contamination (i.e., contamination rate is 0). However, when the contamination rate increases, the

coverage probabilities of the confidence intervals based on MLEs decrease for both scale and shape parameters and the average widths increase for the scale parameter but decrease for the shape parameter. These observations are consistent in both the longer-life and shorter-life contamination models with and without censoring for System I and System II. We also observe that the coverage probabilities of the confidence intervals based on MLEs are sensitive to the contamination rate and the type of contamination. In the longer-life contamination model, the coverage probabilities for both scale and shape parameters drop dramatically when the contamination rate increases. For example, in Table 3, when the contamination rate is 15% with no censoring, the simulated coverage probabilities of the confidence intervals based on MLE are only 18% for the scale parameter and 2% for the shape parameter under the longer-life contamination model, while the simulated coverage probabilities of the confidence intervals based on MLE are 80.7% for the scale parameter and 68.7% for the shape parameter under the shorter-life contamination model.

Table 4:　Simulated coverage probabilities and average widths of confidence intervals of the shape parameter computed based on MLE and MDE_S with different values of α under the longer-life and shorter-life contamination models for System I.

Coverage Probability	Longer-life Contamination Model								Shorter-life Contamination Model							
	No censoring				5% censoring				No censoring				5% censoring			
Contamination Proportion	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%
MLE	95.2	36.8	6.5	2.0	95.0	93.0	49.8	20.4	95.2	93.7	82.0	68.7	95.3	93.3	81.6	68.7
MDE_S ($\alpha=0.01$)	92.4	42.3	4.4	1.0	88.7	91.7	73.0	33.8	92.4	92.1	74.8	52.6	88.9	91.9	91.4	82.0
MDE_S ($\alpha=0.10$)	92.4	70.0	10.3	2.3	89.3	92.2	72.6	33.2	92.5	91.4	72.2	47.7	89.7	92.3	89.4	77.2
MDE_S ($\alpha=0.25$)	92.5	91.8	70.3	37.5	89.9	92.6	77.7	40.4	92.2	90.0	65.4	37.6	90.3	91.8	83.7	64.2
MDE_S ($\alpha=0.50$)	92.5	93.0	91.6	87.5	91.1	92.8	88.8	80.0	91.9	89.8	62.2	31.9	91.2	90.8	75.8	49.7
MDE_S ($\alpha=0.75$)	92.3	93.0	92.8	91.1	92.1	93.4	91.6	89.7	91.4	88.2	60.3	28.9	91.6	89.5	70.0	41.6
MDE_S ($\alpha=0.90$)	92.3	93.0	92.6	91.6	92.4	93.3	92.3	90.8	91.2	87.3	59.9	28.7	91.4	88.7	67.6	39.0

Average Width	Longer-life Contamination Model								Shorter-life Contamination Model							
	No censoring				5% censoring				No censoring				5% censoring			
Contamination Proportion	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%	0%	5%	10%	15%
MLE	0.95	0.68	0.56	0.53	1.02	0.96	0.77	0.66	0.95	0.90	0.84	0.80	1.02	0.96	0.89	0.85
MDE_S ($\alpha=0.01$)	1.01	0.78	0.57	0.54	1.25	1.12	0.98	0.79	1.01	0.86	0.71	0.64	1.25	1.04	0.85	0.76
MDE_S ($\alpha=0.10$)	1.01	0.84	0.60	0.55	1.24	1.11	0.94	0.77	1.01	0.86	0.70	0.62	1.24	1.04	0.84	0.74
MDE_S ($\alpha=0.25$)	1.05	1.03	0.89	0.74	1.25	1.10	0.91	0.76	1.05	0.89	0.71	0.62	1.25	1.06	0.83	0.73
MDE_S ($\alpha=0.50$)	1.20	1.18	1.14	1.09	1.33	1.17	1.02	1.00	1.20	1.01	0.75	0.63	1.31	1.12	0.84	0.70
MDE_S ($\alpha=0.75$)	1.41	1.37	1.31	1.24	1.49	1.34	1.21	1.19	1.39	1.17	0.82	0.65	1.45	1.23	0.90	0.71
MDE_S ($\alpha=0.90$)	1.51	1.47	1.39	1.32	1.57	1.46	1.31	1.28	1.54	1.28	0.88	0.68	1.55	1.34	0.94	0.73

Similar to the MLEs, the coverage probabilities of the confidence intervals based on MDE_S are also sensitive to the contamination rate and the type of the contamination. For the longer-life contamination model, the coverage probabilities of confidence intervals of the scale parameter based on MDE_S with α close to 1 are closer to the nominal levels when the contamination rate is high. However, for the shorter-life contamination model, the coverage probabilities of the confidence intervals of the scale parameter based on MDE_S with α close to 1 are far away from the nominal level when the contamination rate is high (see Table 3).

The confidence intervals based on MDE_S have better coverage probabilities than the confidence intervals based on MLE in the longer-life contamination model when the contamination rate is high. For example, when the contamination rate is 15% under the longer-life contamination model, the coverage probabilities of the confidence intervals based on MDE_S with α close to 1 can still maintain at 95.7% for the scale parameter (Table 3) and 91.6% for shape parameter (Table 4), while the coverage probabilities of the confidence intervals of based on MLE are down to 18% for scale parameter and 2% for shape parameter.

In general, for the longer-life contamination model, compared to the confidence intervals based on MLE, the confidence intervals based on MDE_S have higher coverage probabilities and larger average widths (Tables 3 and 4). Nevertheless, for the shorter-life contamination model, compared to the confidence intervals based on MLE, the confidence intervals based on MDE_S have lower coverage probabilities.

4.   ILLUSTRATIVE EXAMPLE

In this section, a numerical example based on the system lifetime data of the 4-component series-parallel III system with Weibull component lifetimes is used to illustrate the estimation methods proposed in this paper. The system lifetime data was originally presented in [2] and further analyzed by [26]. The data are 10 system lifetimes from the 4-component system with system signature $\boldsymbol{s} = (1/4, 1/4, 1/2, 0)$ with component lifetime follows Weibull$(3, 2)$:

$0.72717, 1.02050, 1.38633, 1.61244, 1.70590, 1.76789, 2.6786, 3.02676, 3.25943, 3.78497.$

To illustrate the effect of contamination in the statistical inference procedures, we simulated an observation from the Weibull$(9, 2)$ to replace one of the observations in the original data set. Specifically, the observation 1.76789 is replaced by 5.48619. The contaminated data set is as follows.

$0.72717, 1.02050, 1.38633, 1.61244, 1.70590, 5.48619, 2.6786, 3.02676, 3.25943, 3.78497.$

Based on the original and the contaminated data sets, the MLE and the three proposed MDEs of the Weibull parameters a and b and the corresponding confidence intervals are presented in Tables 5 and 6.

For point estimation, form Tables 5 and 6, the MLE, MDE_S, MDE_C and MDE_P with different values of α provide similar point estimates of the parameters a and b. By comparing the estimates obtained from the data sets with and without contamination, the difference between MDEs (especially α close to 1) obtained from the data sets with and without contamination is smaller than the difference between MLEs obtained from the data sets with and without contamination in general. For example, the MLE of a is 2.695 for the data set without contamination and the MLE of a is 3.249 for the data set with contamination which has a difference 0.554, while the MDE_S with $\alpha = 0.9$ is 2.691 for the data set without contamination and the MDE_S with $\alpha = 0.9$ is 3.105 for the data set with contamination, which has a difference 0.414.

Table 5:   Point and interval estimates for Weibull parameters
for the original data set presented in Section 4.

Estimator		$\hat{a}$	95% CI based on $\widehat{SE}_A(\hat{a})$	95% CI based on $\widehat{SE}_F(\hat{a})$	95% CI based on $\widehat{SE}_B(\hat{a})$
MLE		2.695		(1.978, 3.412)	(1.980, 3.410)
MDE_S	$\alpha = 0.01$	2.696	(2.490, 2.902)	(1.976, 3.416)	(2.014, 3.378)
	$\alpha = 0.10$	2.700	(2.504, 2.896)	(1.961, 3.439)	(1.967, 3.433)
	$\alpha = 0.25$	2.706	(2.531, 2.881)	(1.937, 3.475)	(1.922, 3.490)
	$\alpha = 0.50$	2.710	(2.495, 2.925)	(1.900, 3.520)	(1.780, 3.640)
	$\alpha = 0.75$	2.703	(2.507, 2.899)	(1.867, 3.539)	(1.731, 3.675)
	$\alpha = 0.90$	2.691	(2.495, 2.887)	(1.850, 3.532)	(1.667, 3.715)
MDE_C	$\alpha = 0.01$	2.617	(2.421, 2.813)	(2.003, 3.231)	(2.019, 3.215)
	$\alpha = 0.10$	2.628	(2.442, 2.814)	(2.002, 3.254)	(2.027, 3.229)
	$\alpha = 0.25$	2.647	(2.483, 2.811)	(1.997, 3.297)	(2.024, 3.270)
	$\alpha = 0.50$	2.677	(2.538, 2.816)	(1.987, 3.367)	(1.976, 3.378)
	$\alpha = 0.75$	2.700	(2.576, 2.824)	(1.972, 3.428)	(1.909, 3.491)
	$\alpha = 0.90$	2.709	(2.585, 2.833)	(1.960, 3.458)	(1.896, 3.522)
MDE_P	$\alpha = 0.01$	2.769	(2.453, 3.085)	(1.892, 3.646)	(1.937, 3.601)
	$\alpha = 0.10$	2.782	(2.505, 3.059)	(1.884, 3.680)	(1.877, 3.687)
	$\alpha = 0.25$	2.802	(2.579, 3.025)	(1.872, 3.732)	(1.866, 3.738)
	$\alpha = 0.50$	2.829	(2.665, 2.993)	(1.851, 3.807)	(1.853, 3.805)
	$\alpha = 0.75$	2.848	(2.633, 3.063)	(1.835, 3.861)	(1.815, 3.881)
	$\alpha = 0.90$	2.855	(2.649, 3.061)	(1.827, 3.883)	(1.780, 3.930)

Estimator		$\hat{b}$	95% CI based on $\widehat{SE}_A(\hat{b})$	95% CI based on $\widehat{SE}_F(\hat{b})$	95% CI based on $\widehat{SE}_B(\hat{b})$
MLE		2.004		(0.945, 3.063)	(0.566, 3.442)
MDE_S	$\alpha = 0.01$	1.999	(1.847, 2.151)	(0.942, 3.056)	(0.668, 3.330)
	$\alpha = 0.10$	1.946	(1.794, 2.098)	(0.916, 2.976)	(0.430, 3.462)
	$\alpha = 0.25$	1.872	(1.733, 2.011)	(0.878, 2.866)	(0.275, 3.469)
	$\alpha = 0.50$	1.782	(1.630, 1.934)	(0.832, 2.732)	(0.000, 3.705)
	$\alpha = 0.75$	1.718	(1.594, 1.842)	(0.799, 2.637)	(0.000, 5.163)
	$\alpha = 0.90$	1.690	(1.566, 1.814)	(0.788, 2.592)	(0.000, 4.666)
MDE_C	$\alpha = 0.01$	2.340	(2.188, 2.492)	(1.079, 3.601)	(0.310, 4.370)
	$\alpha = 0.10$	2.276	(2.124, 2.428)	(1.058, 3.494)	(0.349, 4.203)
	$\alpha = 0.25$	2.184	(2.045, 2.323)	(1.024, 3.344)	(0.257, 4.111)
	$\alpha = 0.50$	2.065	(1.941, 2.189)	(0.974, 3.156)	(0.000, 4.671)
	$\alpha = 0.75$	1.978	(1.854, 2.102)	(0.932, 3.024)	(0.000, 9.267)
	$\alpha = 0.90$	1.937	(1.813, 2.061)	(0.911, 2.963)	(0.000, 6.431)
MDE_P	$\alpha = 0.01$	1.732	(1.462, 2.002)	(0.800, 2.664)	(0.715, 2.749)
	$\alpha = 0.10$	1.713	(1.473, 1.953)	(0.787, 2.639)	(0.775, 2.651)
	$\alpha = 0.25$	1.688	(1.536, 1.840)	(0.773, 2.603)	(0.373, 3.003)
	$\alpha = 0.50$	1.653	(1.501, 1.805)	(0.748, 2.558)	(0.403, 2.903)
	$\alpha = 0.75$	1.627	(1.503, 1.751)	(0.731, 2.523)	(0.279, 2.975)
	$\alpha = 0.90$	1.615	(1.491, 1.739)	(0.723, 2.507)	(0.028, 3.202)

For interval estimation, in both with and without contamination cases (Tables 5 and 6), the confidence intervals for the scale parameter based on the observed Fisher information matrix are very close to the one obtained from the bootstrap method. However, the confidence intervals for the shape parameter based on the bootstrap method is wider than those based on the observed Fisher information matrix. The confidence intervals using MDEs with stan-

dard error estimates based on the theoretical results are much narrower than the confidence intervals with standard error estimates based on the observed Fisher information matrix and based on the bootstrap method. This observation agrees with the results in the Monte Carlo simulation that the standard error estimates based on the theoretical results are likely to underestimate the standard errors of the MDEs.

Table 6: Point and interval estimates for Weibull parameters for the contaminated data set presented in Section 4.

Estimator		$\hat{a}$	95% CI based on $\widehat{\mathrm{SE}}_A(\hat{a})$	95% CI based on $\widehat{\mathrm{SE}}_F(\hat{a})$	95% CI based on $\widehat{\mathrm{SE}}_B(\hat{a})$
MLE		3.249		(2.172, 4.326)	(2.192, 4.306)
MDE_S	$\alpha = 0.01$	3.248	(2.851, 3.645)	(2.169, 4.327)	(2.194, 4.302)
	$\alpha = 0.10$	3.235	(2.853, 3.617)	(2.154, 4.316)	(2.183, 4.287)
	$\alpha = 0.25$	3.210	(2.859, 3.561)	(2.131, 4.289)	(2.129, 4.291)
	$\alpha = 0.50$	3.165	(2.861, 3.469)	(2.100, 4.230)	(2.010, 4.320)
	$\alpha = 0.75$	3.124	(2.884, 3.364)	(2.072, 4.176)	(1.935, 4.313)
	$\alpha = 0.90$	3.105	(2.941, 3.269)	(2.057, 4.153)	(1.855, 4.355)
MDE_C	$\alpha = 0.01$	3.203	(2.887, 3.519)	(2.261, 4.145)	(2.329, 4.077)
	$\alpha = 0.10$	3.207	(2.910, 3.504)	(2.247, 4.167)	(2.292, 4.122)
	$\alpha = 0.25$	3.213	(2.965, 3.461)	(2.225, 4.201)	(2.269, 4.157)
	$\alpha = 0.50$	3.216	(3.030, 3.402)	(2.194, 4.238)	(2.168, 4.264)
	$\alpha = 0.75$	3.206	(2.983, 3.429)	(2.165, 4.247)	(2.090, 4.322)
	$\alpha = 0.90$	3.192	(2.969, 3.415)	(2.148, 4.236)	(2.036, 4.348)
MDE_P	$\alpha = 0.01$	3.368	(2.952, 3.784)	(2.096, 4.640)	(2.168, 4.568)
	$\alpha = 0.10$	3.379	(2.973, 3.785)	(2.089, 4.669)	(2.076, 4.682)
	$\alpha = 0.25$	3.392	(3.005, 3.779)	(2.080, 4.704)	(2.071, 4.713)
	$\alpha = 0.50$	3.406	(3.050, 3.762)	(2.069, 4.743)	(2.008, 4.804)
	$\alpha = 0.75$	3.414	(3.080, 3.748)	(2.065, 4.763)	(1.912, 4.916)
	$\alpha = 0.90$	3.416	(3.100, 3.732)	(2.062, 4.770)	(1.685, 5.147)

Estimator		$\hat{b}$	95% CI based on $\widehat{\mathrm{SE}}_A(\hat{b})$	95% CI based on $\widehat{\mathrm{SE}}_F(\hat{b})$	95% CI based on $\widehat{\mathrm{SE}}_B(\hat{b})$
MLE		1.607		(0.782, 2.432)	(0.485, 2.729)
MDE_S	$\alpha = 0.01$	1.604	(1.300, 1.908)	(0.779, 2.429)	(0.404, 2.804)
	$\alpha = 0.10$	1.588	(1.284, 1.892)	(0.770, 2.406)	(0.476, 2.700)
	$\alpha = 0.25$	1.569	(1.272, 1.866)	(0.761, 2.377)	(0.319, 2.819)
	$\alpha = 0.50$	1.550	(1.287, 1.813)	(0.751, 2.349)	(0.000, 3.695)
	$\alpha = 0.75$	1.535	(1.411, 1.659)	(0.741, 2.329)	(0.000, 4.102)
	$\alpha = 0.90$	1.525	(1.386, 1.664)	(0.736, 2.314)	(0.000, 3.997)
MDE_C	$\alpha = 0.01$	1.825	(1.593, 2.057)	(0.899, 2.751)	(0.206, 3.444)
	$\alpha = 0.10$	1.786	(1.571, 2.001)	(0.879, 2.693)	(0.382, 3.190)
	$\alpha = 0.25$	1.732	(1.580, 1.884)	(0.851, 2.613)	(0.162, 3.302)
	$\alpha = 0.50$	1.668	(1.504, 1.832)	(0.816, 2.520)	(0.000, 8.867)
	$\alpha = 0.75$	1.625	(1.501, 1.749)	(0.791, 2.459)	(0.000, 4.849)
	$\alpha = 0.90$	1.606	(1.482, 1.730)	(0.781, 2.431)	(0.000, 5.860)
MDE_P	$\alpha = 0.01$	1.464	(1.167, 1.761)	(0.697, 2.231)	(0.733, 2.195)
	$\alpha = 0.10$	1.455	(1.158, 1.752)	(0.691, 2.219)	(0.626, 2.284)
	$\alpha = 0.25$	1.444	(1.147, 1.741)	(0.685, 2.203)	(0.533, 2.355)
	$\alpha = 0.50$	1.433	(1.149, 1.717)	(0.676, 2.190)	(0.432, 2.434)
	$\alpha = 0.75$	1.428	(1.158, 1.698)	(0.674, 2.182)	(0.292, 2.564)
	$\alpha = 0.90$	1.426	(1.170, 1.682)	(0.672, 2.180)	(0.021, 2.831)

5.   CONCLUDING REMARKS

In this paper, we study the robust estimation method for the model parameters in the component lifetime distribution based on system lifetime data with known system structure. The minimum density power divergence estimation method is considered and three different MDEs are proposed. Standard error estimation and interval estimation procedures based on the MDEs are also studied. The three proposed estimation procedures are compared to the maximum likelihood estimation method via a Monte Carlo simulation study. It is shown that the minimum density power divergence estimation method based on system-level data can provide better performance in both point and interval estimation when there is longer-life contamination in the data. We have also shown that the standard error estimates based on the bootstrap method can be adopted for estimating the standard errors of the MDEs.

From our simulation study, for point estimation, the MLE outperforms the MDEs when there is no contamination in the data. However, we observe that the system-level MDE, MDE_S, is a robust estimation procedure than the MLE when there is contamination in the data. For interval estimation, we observe that the contaminated data considerably affect the coverage probabilities of the confidence intervals based on MLE and MDE_S. The confidence intervals based on MDE_S perform better than those based on MLE for contaminated data, especially when the contamination rate is high (say, 10% or 15%) in the longer-life contamination model. For contamination data with longer lifetimes, MDE_S with large value of α ($\alpha = 0.75$ or 0.9) is recommended. For contamination data with shorter lifetimes, MDE_S with small value of α ($\alpha = 0.01$ or 0.1) is recommended. Since the choice of the value α for the MDE_S affects the results for the interval estimation, it is interesting to study the choice of the value of α in the system-level minimum divergence estimator MDE_S. In practice, the sample size m, the system signature $\boldsymbol{s}$, and the censoring proportion are known, but the underlying component lifetime distribution and the contamination rate are usually unknown. The performance of the estimators with different values of α can be studied under different underlying component lifetime distributions and contamination rates via simulation, and then a reasonable range of the value of α can be obtained.

For future research, a systematic way to choose the value of α for the MDE can be studied. On the other hand, since the simulated coverage probabilities of the confidence intervals based on the estimators studied in this paper can be much lower than the nominal level when there is contamination in the data, it is desired to develop better standard error estimation methods and confidence interval estimation methods which can provide better coverage probabilities when the contamination rates. The current work can be extended to the situation when the lifetime of the unit may be affected by one or more factors/explanatory variables (such as temperature, voltage, load, etc.). For example, consider a Weibull regression model in which K covariates $\boldsymbol{z} = (z_1, z_2, ..., z_K)$ affect the scale parameter a in (3.1), then, we have a parametric proportional hazard model for the lifetime X

$$f_X(x; \boldsymbol{\theta}) = f_X(x; a(\boldsymbol{z}), b),$$

where

$$a(\boldsymbol{z}) = \exp(\nu_0 + \nu_1 z_1 + \nu_2 z_2 + \cdots + \nu_K z_K)$$

and the parameter vector is $\boldsymbol{\theta} = (\nu_0, \nu_1, ..., \nu_K, b)$. The proposed minimum density divergence estimation method can be applied to estimate the parameter vector $\boldsymbol{\theta} = (\nu_0, \nu_1, ..., \nu_K, b)$.

On the other hand, we assume that the system signature is known in this paper, however, for some black box systems, we may not have any knowledge on the system structures. Following the work by Yang *et al.* [27], one can develop robust procedures for estimating the parameters of the component lifetime distribution and for identifying the system structure based on system-level data simultaneously by assuming the system is a coherent system. We are currently working on these extensions and we hope to report the findings in future work.

ACKNOWLEDGMENTS

REFERENCES

[1] BALAKRISHNAN, N.; NG, H.K.T. and NAVARRO, J. (2011). Exact nonparametric inference for component lifetime distribution based on lifetime data from systems with known signatures, *Journal of Nonparametric Statistics*, **23**, 741–752.

[2] BALAKRISHNAN, N.; NG, H.K.T. and NAVARRO, J. (2011). Linear inference for Type-II censored lifetime data of reliability systems with known signature, *IEEE Transactions on Reliability*, **60**, 426–440.

[3] BHATTACHARYA, D. and SAMANIEGO, F.J. (2010). Estimating component characteristics from system failure-time data, *Naval Research Logistics*, **57**, 380–389.

[4] BASU, A.; HARRIS, I.R.; HJORT, N.L. and JONES, M.C. (1998). Robust and efficient estimation by minimising a density power divergence, *Biometrika*, **3**, 549–559.

[5] BASU, S.; BASU, A. and JONES, M.C. (2006). Robust and efficient parametric estimation for censored survival data, *Annals of the Institute of Statistical Mathematics*, **58**, 341–355.

[6] BASAK, S.; BASU, A. and JONES, M.C. (2021). On the 'optimal' density power divergence tuning parameter, *Journal of Applied Statistics*, **48**(3), 536–556.

[7] BOLAND, P.J. and SAMANIEGO, F.J. (2004). *The signature of a coherent system and its application in reliability.* In "Mathematical Reliability: An Expository Perspective" (R. Soyer, T.A. Mazzuchi and N.D. Singpurwalla, Eds.), International Series in Operations Research & Management Science, vol. 67, Springer, Boston, MA, 3–30.

[8] CHAHKANDI, M.; AHMADI, J. and BARATPOUR, S. (2014). Non-parametric prediction intervals for the lifetime of coherent systems, *Statistical Papers*, **55**, 1019–1034.

[9] COHEN, A.C. (1991). *Truncated and Censored Samples: Theory and Applications*, Marcel Dekker, New York.

[10] EFRON, B. and TIBSHIRANI, R.J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.

[11] FALLAH, A.; ASGHARZADEH, A. and NG, H.K.T. (2021). Statistical inference for component lifetime distribution from coherent system lifetimes under a proportional reversed hazard model, *Communications in Statistics – Theory and Methods*, **50**, 3809–3833.

[12] HERMANNS, M.; CRAMER, E. and NG, H.K.T. (2020). EM algorithms for ordered and censored system lifetime data under a proportional hazard rate model, *Journal of Statistical Computation and Simulation*, **90**, 3301–3337.

[13] JIN, Y.; HALL, P.G.; JIANG, J. and SAMANIEGO, F.J. (2017). Estimating component reliability based on failure time data from a system of unknown design, *Statistica Sinica*, **27**, 479–499.

[14] KAPLAN, E.L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, **53**, 457–481.

[15] KOCHAR, S.; MUKERJEE, H. and SAMANIEGO, F.J. (1999). The signature of a coherent system and its application to comparison among systems, *Naval Research Logistics*, **46**, 507–523.

[16] KULLBACK, S. and LEIBLER, R.A. (1951). On information and sufficiency, *Annals of Mathematical Statistics*, **22**, 79–86.

[17] LAWLESS, J.F. (2003). *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York.

[18] MEEKER, M.Q. and ESCOBAR, L.A. (1998). *Statistical Methods for Reliability Data*, John Wiley & Sons, New York.

[19] NAQVI, S.; CHAN, P.S. and MISHRA, D.B. (2022). System signatures: a review and bibliometric analysis, *Communications in Statistics – Theory and Methods*, **51**, 1993–2008.

[20] NAVARRO, J.; RUIZ, J.M. and SANDOVAL, C.J. (2007). Properties of coherent systems with dependent components, *Communications in Statistics – Theory and Methods*, **36**, 175–191.

[21] NG, H.K.T.; NAVARRO, J. and BALAKRISHNAN, N. (2012). Parametric inference from system lifetime data under a proportional hazard rate model, *Metrika*, **75**, 367–388.

[22] RAINA, R. (2008). *Achieving zero-defects for automotive applications.* In "2008 IEEE International Test Conference", pp. 1–10, IEEE.

[23] RIANI, M.; ATKINSON, A.C.; CORBELLINI, A. and PERROTTA, D. (2020). Robust regression with density power divergence: Theory, Comparisons, and Data Analysis, *Entropy*, **22**, 399–415.

[24] SAMANIEGO, F.J. (2007). *System Signatures and Their Applications in Engineering Reliability*, Springer, New York.

[25] SHEATHER, S.J. and JONES, M.C. (1991). A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society – Series B*, **53**, 683–690.

[26] YANG, Y.; NG, H.K.T. and BALAKRISHNAN, N. (2016). A stochastic expectation-maximization algorithm for the analysis of system lifetime data with known signature, *Computational Statistics*, **31**, 609–641.

[27] YANG, Y.; NG, H.K.T. and BALAKRISHNAN, N. (2019). Expectation-maximization algorithm for system-based lifetime data with unknown system structure, *AStA Advances in Statistical Analysis*, **103**, 69–98.

[28] ZHANG, J.; NG, H.K.T. and BALAKRISHNAN, N. (2015). Statistical inference of component lifetimes with location-scale distribution from censored system failure data with known signature, *IEEE Transactions on Reliability*, **64**, 613–626.

Birnbaum–Saunders Semi-Parametric Additive Modeling: Estimation, Smoothing, Diagnostics, and Application

Authors: ESTEBAN CÁRCAMO
– Institute of Statistics, Universidad de Valparaíso, Chile
esteban.carcamo@alumnos.uv.cl

CAROLINA MARCHANT [iD]
– Faculty of Basic Sciences, Universidad Católica del Maule, Talca, Chile
– ANID – Millennium Science Initiative Program – Millennium Nucleus Center for the Discovery of Structures in Complex Data, Santiago, Chile
carolina.marchant.fuentes@gmail.com

GERMÁN IBACACHE-PULGAR [iD]
– Institute of Statistics, Universidad de Valparaíso, Valparaíso, Chile
– Interdisciplinary Center for Atmospheric and Astro-Statistical Studies, Universidad de Valparaíso, Valparaíso, Chile
gibacachepulgar@gmail.com

VÍCTOR LEIVA [iD] [✉]
– School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
victorleivasanchez@gmail.com

Abstract:

• Inclusion of nonparametric functions enhances the modeling when accommodating non-linear effects of covariates. Semi-parametric models have been successfully used for describing non-linear structures by means of parametric and nonparametric components. In this work, we formulate a semi-parametric additive regression model based on a Birnbaum–Saunders distribution and carry out influence diagnostics for such a model. This semi-parametric structure permits us to model the mean and variance simultaneously. We employ a back-fitting algorithm to get the penalized maximum likelihood estimates by utilizing cubic smoothing splines. We derive methods of local influence by calculating the normal curvatures under different perturbation schemes. The obtained results are computationally implemented in the R software so that diverse users have available this model computationally to be applied in practice. Finally, an application of the proposed model with real data from one of the most polluted cities in the world is presented.

[✉] Corresponding author

1.   INTRODUCTION

Semi-parametric structures are a powerful tool in statistical modeling when incorporating covariates that can contribute parametrically or nonparametrically in the model [20, 21]. The Birnbaum–Saunders (BS) distribution has received considerable attention in recent years, due to its theoretical arguments associated with cumulative damage processes, its properties, and its relation with the normal distribution. Specifically, the amount of cumulative damage that allows the BS distribution to be generated is assumed to follow a normal distribution. The BS model corresponds to a unimodal, positively skewed, two-parameter distribution and with support in the positive real numbers. The BS distribution has its genesis from engineering, but it has been also mathematically justified and applied to solve environmental problems [7, 8, 28, 31, 35, 40, 48].

Extensive work has been done on different aspects with regard to the BS distribution during the last two decades. The main addressed topics are related to mathematical and statistical properties of this distribution, point and interval estimation with classical or Bayesian methods, hypothesis testing, log-linear and non-linear regression models, random number generators, development of packages in the R software, generalizations and reparameterizations, multivariate versions, and extensions to temporal and spatial modeling; for a recent and detailed review on these issues, see [3, 30].

A new parameterization of the BS (RBS) distribution stated in [46, 47] is based on its mean and precision parameters. By using this parametrization, a type of generalized linear model (GLM) was formulated in [27] based on the BS distribution although this does not belong to the exponential family. Such a formulation permits us to model the mean and variance simultaneously, maintaining the original scale of the data with no transformations on the modeled variable, because it is well known that transformations reduce interpretability. Statistical modeling based on BS distributions has considerably attracted the attention of a number of researchers; see, for example, [10, 32, 42, 49]. To the best of our knowledge, no semi-parametric models based on the RBS distribution have been derived until now in the literature.

Diagnostic analytics is an important step to be considered in all data modeling. Diagnostics can be conducted by goodness-of-fit techniques, residual analysis, and global/local influence methods. Goodness-of-fit methods allow us to assess the adequacy of a model to a data set. Residual analysis is a helpful tool for evaluating the fit of a statistical model. Several types of residuals for BS regressions were proposed in [27, 29], stating by simulations which of them has better performance. Global influence techniques remove cases and evaluate their effect on the fitted model. Local influence was proposed in [9] and permits us to detect the effect of perturbations on the estimates of model parameters. Local influence methods for RBS regression models were derived in [27] by calculating the normal curvatures under different perturbation schemes. To the best of our knowledge, no diagnostic analytics in semi-parametric additive models (SAM) based on BS distributions have been considered in the literature at present.

Our main objective is to formulate a novel RBS semi-parametric additive regression model (RBS-SAM). This semi-parametric model enables us to describe the mean and variance

simultaneously. Our secondary objectives are:

- (**i**) to estimate the model parameters with the maximum penalized likelihood (MPL) method and a back-fitting algorithm;
- (**ii**) to describe the nonparametric structure with cubic smoothing splines;
- (**iii**) to derive local influence for model diagnostics by calculating the normal curvatures under different perturbations;
- (**iv**) to implement the obtained results in the R software;
- (**v**) to apply the results to real data associated with pollutant contents in Santiago of Chile — one of the most polluted cities in the world [33, 8, 40].

These data were secured from the website of the Chilean Ministry of Environment. Our numerical illustrations were developed with the R software [41].

The application presented in this study is motivated by the fact that inclusion of nonparametric functions greatly enhances the modeling when accommodating non-linear effects of covariates [21, 20]. These covariates correspond in our case to contents of pollutants and meteorological variables as atmospheric pressure, precipitation, relative humidity, temperature, and wind speed [40]. Semi-parametric estructures have been successfully used to model non-linear components [23]. In addition, the application considered in the present study is supported by theoretical arguments that permit us to justify the use of BS distributions for describing environmental data. This argumentation was formalized in [28] employing a novel mathematical model for environmental sciences based on physical laws.

This article is organized as follows. Section 2 provides an overview of BS distributions, introduces the RBS-SAM, and considers a penalized log-likelihood function for parameter estimation. In Section 3, we obtain the MPL estimators, use a back-fitting algorithm, derive the penalized score vector and Hessian matrix, determine the degrees of freedom, select the smoothing parameter, and conduct inference for the corresponding parameters. In Section 4, the main concepts of local influence and the derivation of normal curvatures for some perturbation schemes are presented. Section 5 introduces the empirical application of the proposed model to an environmental data set. In Section 6, some concluding remarks and ideas for future research are given.

2. PARAMETRIC SPECIFICATION AND NONPARAMETRIC COMPONENT

In this section, an overview of BS distributions, the RBS-SAM structure, and the corresponding penalized log-likelihood function for parameter estimation are stated.

2.1. The BS distribution and related models

Next, we present the traditional BS distribution and related models. If a random variable Y^* is BS distributed with shape ($\alpha > 0$) and scale ($\beta > 0$) parameters, we employ the notation $Y^* \sim \mathrm{BS}(\alpha, \beta)$. Note that $Z = (1/\alpha)((Y^*/\beta)^{1/2} - (\beta/Y^*)^{1/2}) \sim \mathrm{N}(0, 1)$ if $Y^* \sim \mathrm{BS}(\alpha, \beta)$.

Hence, each BS distributed random variable Y^* may be obtained as a transformation of a standard normal distributed random variable Z by means of

$$(2.1) \qquad Y^* = \beta \left(\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2} \right)^2 + 1} \right)^2.$$

Note from the formula stated in (2.1) that the quantile function or $100q$-th quantile of the BS distribution is directly generated as

$$(2.2) \qquad y^*(q; \alpha, \beta) = \beta \left(\frac{\alpha z(q)}{2} + \sqrt{\left(\frac{\alpha z(q)}{2} \right)^2 + 1} \right)^2, \quad 0 < q \le 1,$$

where $z(q)$ is the $100q$-th quantile of the standard normal distribution. We observe that, if $q = 0.5$, then $z(0.5) = 0$, which corresponds to the mean, median and mode of $Z \sim \mathrm{N}(0,1)$. Thus, from the quantile function defined in (2.2), we have that $y^*(0.5; \alpha, \beta) = \beta$, which confirms that the scale parameter $\beta > 0$ is also the BS median. The expression given in (2.1) is used for generating random numbers of the BS distribution and also for deriving goodness-of-fit tools associated with this distribution.

Let $Y^* \sim \mathrm{BS}(\alpha, \beta)$. Then, the probability density function of Y^* is stated as

$$(2.3) \quad f_{Y^*}(y; \alpha, \beta) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2\alpha^2} \left(\frac{y}{\beta} + \frac{\beta}{y} - 2 \right) \right) \frac{1}{2\alpha\beta} \left(\left(\frac{y}{\beta} \right)^{-1/2} + \left(\frac{y}{\beta} \right)^{-3/2} \right),$$

for $y, \alpha, \beta > 0$. Properties of $Y^* \sim \mathrm{BS}(\alpha, \beta)$ are: $bY^* \sim \mathrm{BS}(b\alpha, \beta)$, with $b > 0$; $1/Y^* \sim \mathrm{BS}(\alpha, 1/\beta)$. Also, the mean and variance of Y^* are $\mathrm{E}(Y^*) = \beta(1 + \alpha^2/2)$ and $\mathrm{Var}(Y^*) = \beta^2 \alpha^2 (1 + 5\alpha^2/4)$.

It is worth noting that diverse models related to the BS distribution have been derived. The probability density function formulated in (2.3) corresponds to the traditional BS distribution and this is quite flexible. However, new versions of the BS distribution have been proposed with different properties and shapes even more flexible. For some recent works on these new versions and their mathematical and statistical features, the interested reader is referred to [2] for a mixture Birnbaum–Saunders distribution with applications in biomedicine; to [1] for skew BS distributions with illustrations in fatigue of materials; to [5] for a transmuted BS distribution with examples in problems of engineering and medicine; to [36, 37, 38, 45] for unit and quantile BS distributions considering applications in economy, medicine and politics; to [43] for a bimodal BS distribution with illustrations using environmental and medical data; and to [13] for a BS-gamma distributed claim size applied to insurance.

2.2. RBS distribution

The RBS distribution is a novel model related to the traditional BS case, based on the parameters $\mu, \delta > 0$, where μ is a scale parameter and the distribution mean, whereas δ is a shape and precision parameter. In this case, the notation $Y \sim \mathrm{RBS}(\mu, \delta)$ is used. Based on this reformulation of the BS distribution, the probability density function of the random variable $Y \sim \mathrm{RBS}(\mu, \delta)$ is given by

$$(2.4) \qquad f_Y(y; \mu, \delta) = \frac{\exp(\delta/2)\sqrt{\delta+1}}{4\, y^{3/2}\, \sqrt{\pi\mu}} \left(y + \frac{\delta\mu}{\delta+1} \right) \exp \left(-\frac{\delta}{4} \left(\frac{y(\delta+1)}{\delta\mu} + \frac{\delta\mu}{y(\delta+1)} \right) \right),$$

where $y, \mu, \delta > 0$. The mean and variance of Y are stated as $\mathrm{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \mu^2/\phi$, respectively, with $\phi = (\delta + 1)^2/(2\delta + 5)$ so that, as mentioned, δ is a precision parameter. Note that, for fixed values of μ, $\mathrm{Var}(Y) \to 0$ as $\delta \to \infty$, but if $\delta \to 0$, then $\mathrm{Var}(Y) \to 5\mu^2$. In addition, $\mathrm{Var}(Y)$ is similar to the variance of the gamma distribution, which has a quadratic relation with its mean. It is also possible to show that $bY \sim \mathrm{RBS}(b\,\mu, \delta)$, with $b > 0$, and $1/Y \sim \mathrm{RBS}(\mu^*, \delta)$, where $\mu^* = (\delta + 1)/(\delta\mu)$.

2.3. Modeling

Let $\boldsymbol{Y} = (Y_1, ..., Y_n)^\top$ be independent random variables, where $Y_i \sim \mathrm{RBS}(\mu_i, \delta)$, for $i \in \{1, ..., n\}$, and $\boldsymbol{y} = (y_1, ..., y_n)^\top$ are the corresponding observations of $\boldsymbol{Y}$. Then, we define the RBS-SAM structure based on (2.4) by the systematic component expressed as

$$(2.5) \qquad h(\mu_i) = \eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + f_1(t_{1_i}) + \cdots + f_s(t_{s_i}), \quad i \in \{1, ..., n\},$$

or, equivalently,

$$(2.6) \qquad h(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \boldsymbol{n}_{1_i}^\top \boldsymbol{f}_1 + \cdots + \boldsymbol{n}_{s_i}^\top \boldsymbol{f}_s,$$

where $\boldsymbol{x}_i^\top = (1, x_{i_2}, ..., x_{i_p})$ is the i-th row of observed values for the covariates matrix $\boldsymbol{X}$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\top$ is a vector of regression parameters to be estimated. In addition, $\boldsymbol{n}_{k_i}^\top$ denotes the i-th row of the incidence matrix $\boldsymbol{N}_k$, whose (i, l)-th element is equal to the indicator function $\mathbb{1}\,(t_{k_i} = t_{k_l}^0)$; $\boldsymbol{f}_k = (\xi_{k_1}, ..., \xi_{k_{r_k}})^\top$ is an $r_k \times 1$ vector such that $\xi_{k_j} = f_k(t_{k_j}^0)$, where f_k is an arbitrary unidimensional (scalar) smooth function that quantifies the effect of the k-th covariate t_k for $k \in \{1, ..., s\}$ (note that f_k is different from the vector $\boldsymbol{f}_k$ above defined); and $t_{k_j}^0$, for $l \in \{1, ..., r_k\}$, are the distinct and ordered values of the covariate t_k. Note that the model stated in (2.5) is formed by both parametric and nonparametric components. In effect, for $p < n$, $\boldsymbol{x}_i^\top \boldsymbol{\beta}$ is the parametric component of the model and $\boldsymbol{n}_{1_i}^\top \boldsymbol{f}_1 + \cdots + \boldsymbol{n}_{s_i}^\top \boldsymbol{f}_s$ is the corresponding nonparametric component. In matrix terms, (2.6) can be written as $h(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{N}_1 \boldsymbol{f}_1 + \cdots + \boldsymbol{N}_s \boldsymbol{f}_s$, where $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)^\top$ and $\mu_i = h^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\beta} + \boldsymbol{n}_{1_i}^\top \boldsymbol{f}_1 + \cdots + \boldsymbol{n}_{s_i}^\top \boldsymbol{f}_s)$, for $i \in \{1, ..., n\}$, with h^{-1} being the inverse of the link function $h\colon \mathbb{R} \rightsquigarrow \mathbb{R}^+$, which is strictly monotone, positive, and at least twice differentiable; for example, $h(\mu) = \log(\mu)$ or $h(\mu) = \sqrt{\mu}$.

Formally, in the RBS-SAM, we have that $\mathrm{Var}(Y_i)$ is a function of μ_i and, consequently, of the covariates $\boldsymbol{x}_i$. Then, because we are modeling the mean based on a particular structure, we are also modeling the variance due to $\mathrm{Var}(Y_i) = \mu_i^2/\phi$. Therefore, problems where a non-constant variance is present could be analyzed by using this model as well.

2.4. Penalized likelihood function

The RBS-SAM log-likelihood function defined in (2.5) for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{f}_1^\top, ..., \boldsymbol{f}_s^\top, \delta)^\top$ is given by

$$(2.7) \qquad \ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\mu_i, \delta; y_i),$$

where

$$\ell_i(\mu_i, \delta; y_i) = \frac{\delta}{2} - \frac{\log(16\pi)}{2} - \frac{1}{2}\log\left(\frac{(\delta+1)y_i^3\mu_i}{(\delta\,y_i + y_i + \delta\mu_i)^2}\right) - \frac{y_i(\delta+1)}{4\mu_i} - \frac{\delta^2\mu_i}{4(\delta+1)y_i}.$$

To avoid the problem of overfitting and non-identification of the parameter $\boldsymbol{\beta}$, we can incorporate a penalty term in the log-likelihood function of the model, denoted here by $J(f_k)$, on each smooth function f_k, with f_k belonging to the Sobolev function space defined as $\mathcal{W}_2^{(2)} \equiv \{f_k\colon f_k^{(1)}, f_k^{(2)} \in \mathcal{L}^2[a_k, b_k]\}$, where f_k is an absolutely continuous function and $f_k^{(2)}(t_k) = \mathrm{d}^2 f_k(t_k)/\mathrm{d}t_k^2$. Therefore, in our case, the log-likelihood function given in (2.7) is now expressed as a penalized function stated by

$$(2.8) \qquad\qquad \ell_{\mathrm{p}}(\boldsymbol{\theta}, \lambda_1, ..., \lambda_s) = \ell(\boldsymbol{\theta}) + \sum_{k=1}^s \lambda_k^* J(f_k),$$

where λ_k is a constant that depends on the smoothing parameter $\lambda_k \geq 0$ that controls the trade-off between goodness-of-fit and the selected smoothness functions.

Different types of penalties have been proposed depending on the method to fit the nonparametric curves. We consider, as a measure of the curvature of the functions, the formula established as

$$(2.9) \qquad\qquad J(f_k) = \int_{a_k}^{b_k} f_k^{(2)}(t_k)^2 \mathrm{d}t_k.$$

The first term on the right side of (2.8) measures the goodness of fit, whereas the second term, defined by (2.9), penalizes the roughness of each f_k with a fixed parameter λ_k. In this case, the selection of f_k leads to a cubic spline with knots at the points $t_{k_l}^0$, that is, this is a third degree polynomial partitioned on each interval $[t_{k_l}, t_{k_{l+1}}]$, for $l \in \{1, ..., r_k - 1\}$. According to [14], $J(f_k)$ can be written as $J(f_k) = \boldsymbol{f}_k^\top \boldsymbol{K}_k \boldsymbol{f}_k$, where $\boldsymbol{K}_k$ is an $r_k \times r_k$ non-negative-definite matrix that depends only on the knots t_k^0. Then, if we consider $\lambda_k^* = -\lambda_k/2$, the penalized log-likelihood function given in (2.8) may be expressed as

$$(2.10) \qquad\qquad \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \ell(\boldsymbol{\theta}) - \sum_{k=1}^s \frac{\lambda_k}{2} \boldsymbol{f}_k^\top \boldsymbol{K}_k \boldsymbol{f}_k,$$

where $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_s)^\top$ denotes an $s \times 1$ vector of smoothing parameters. Note that an essential aspect of the semi-parametric modeling process is related to the selection of smoothing parameters. In the literature there are several efficient methods of selection, among which cross validation, generalized cross validation, Akaike information criterion (AIC) and mean average squared error can be mentioned.

3.   PARAMETERS ESTIMATION AND INFERENCE

In this section, we discuss the process of obtaining MPL estimators, as well as the derivation of a back-fitting algorithm, the penalized score vector, the penalized Hessian matrix, the determination of degrees of freedom, the selection of the smoothing parameter, and the corresponding statistical inference.

3.1. Penalized score function

Assuming that the function given in (2.10) is regular with respect to $\boldsymbol{\beta}, \boldsymbol{f}_1, ..., \boldsymbol{f}_s$ and δ, we have that the penalized score function vector of $\boldsymbol{\theta}$ is stated as

$$\boldsymbol{U}_{\mathrm{p}}(\boldsymbol{\theta}) = \frac{\partial \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \boldsymbol{U}_{\mathrm{p}}^{\boldsymbol{\beta}}(\boldsymbol{\theta}) \\ \boldsymbol{U}_{\mathrm{p}}^{\boldsymbol{f}_1}(\boldsymbol{\theta}) \\ \vdots \\ \boldsymbol{U}_{\mathrm{p}}^{\boldsymbol{f}_s}(\boldsymbol{\theta}) \\ U_{\mathrm{p}}^{\delta}(\boldsymbol{\theta}) \end{pmatrix},$$

whose elements of the vector have the form $\boldsymbol{U}_{\mathrm{p}}^{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \boldsymbol{X}^{\top} \boldsymbol{D}_a \boldsymbol{z}, \boldsymbol{U}_{\mathrm{p}}^{\boldsymbol{f}_k}(\boldsymbol{\theta}) = \boldsymbol{N}_k^{\top} \boldsymbol{D}_a \boldsymbol{z} - \lambda_k \boldsymbol{K}_k \boldsymbol{f}_k$, for $k \in \{1, ..., s\}$, and $U_{\mathrm{p}}^{\delta}(\boldsymbol{\theta}) = \mathrm{tr}(\boldsymbol{D}_b)$, where $\boldsymbol{D}_a = \mathrm{diag}\{a_1, ..., a_n\}$ and $\boldsymbol{D}_b = \mathrm{diag}\{b_1, ..., b_n\}$ are $n \times n$ matrices, while $\boldsymbol{z} = (z_1, ..., z_n)^{\top}$, with

$$
\begin{aligned}
z_i &= -\frac{1}{2\mu_i} + \frac{\delta}{(\delta\, y_i + y_i + \delta\mu_i)} + \frac{y_i(\delta + 1)}{4\mu_i^2} - \frac{\delta^2}{4y_i(\delta + 1)}, \\
(3.1) \qquad b_i &= \frac{1}{2} - \frac{1}{2(\delta + 1)} + \frac{(y_i + \mu_i)}{(\delta\, y_i + y_i + \delta\mu_i)} - \frac{y_i}{4\mu_i} - \frac{\delta(\delta + 2)\mu_i}{4(\delta + 1)^2 y_i}, \quad a_i = \frac{1}{h'(\mu_i)},
\end{aligned}
$$

where h' is the derivative of h. In the RBS-SAM context, the MPL estimators of the model parameters cannot be obtained in an explicit form and need to be calculated by solving a non-linear equation. The Fisher scoring method was used in [27] to estimate the RBS regression parameters. We propose to adjust the RBS-SAM by combining the proposals given in [21, 27] to jointly determine the regression coefficients, the smooth functions, and the precision parameter based on penalized likelihood criterion.

3.2. Penalized Hessian and information matrix

Let $\ddot{\boldsymbol{\ell}}_{\mathrm{p}}(\boldsymbol{\theta})$ be the $p^* \times p^*$ penalized Hessian matrix with its (j^*, l^*)-th element being given by $\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda})/\partial\theta_{j^*}\theta_{l^*}$, for $j^*, l^* \in \{1, ..., p^*\}$ and $p^* = 1 + p + \sum_{k=1}^{s} r_k$. After algebraic manipulation, we find that the corresponding penalized Hessian matrix has the form

$$(3.2) \qquad \ddot{\boldsymbol{\ell}}_{\mathrm{p}}(\boldsymbol{\theta}) = \frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\top}} = \begin{pmatrix} \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) & \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_1}(\boldsymbol{\theta}) & \cdots & \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_s}(\boldsymbol{\theta}) & \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{\beta}\delta}(\boldsymbol{\theta}) \\ \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_1^{\top}}(\boldsymbol{\theta}) & \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{f}_1\boldsymbol{f}_1}(\boldsymbol{\theta}) & \cdots & \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{f}_1\boldsymbol{f}_s}(\boldsymbol{\theta}) & \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{f}_1\delta}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_s^{\top}}(\boldsymbol{\theta}) & \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{f}_1\boldsymbol{f}_s^{\top}}(\boldsymbol{\theta}) & \cdots & \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{f}_s\boldsymbol{f}_s}(\boldsymbol{\theta}) & \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{f}_s\delta}(\boldsymbol{\theta}) \\ \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{\beta}\delta}(\boldsymbol{\theta}) & \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{f}_1\delta}(\boldsymbol{\theta}) & \cdots & \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{f}_s\delta}(\boldsymbol{\theta}) & \ddot{\ell}_{\mathrm{p}}^{\delta\delta}(\boldsymbol{\theta}) \end{pmatrix},$$

whose elements of the matrix can be written as

$$
\begin{aligned}
\ddot{\ell}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \boldsymbol{X}^{\top} \boldsymbol{D}_c \boldsymbol{X}, \quad \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{\beta}\boldsymbol{f}_k}(\boldsymbol{\theta}) = \boldsymbol{X}^{\top} \boldsymbol{D}_c \boldsymbol{N}_k, \quad k \in \{1, ..., s\}, \\
\ddot{\ell}_{\mathrm{p}}^{\boldsymbol{\beta}\delta}(\boldsymbol{\theta}) &= \boldsymbol{X}^{\top} \boldsymbol{D}_a \boldsymbol{m}, \quad \ddot{\ell}_{\mathrm{p}}^{\boldsymbol{f}_k\delta}(\boldsymbol{\theta}) = \boldsymbol{N}_k^{\top} \boldsymbol{D}_a \boldsymbol{m}, \quad \ddot{\ell}_{\mathrm{p}}^{\delta\delta}(\boldsymbol{\theta}) = \mathrm{tr}(\boldsymbol{D}_d), \\
\ddot{\ell}_{\mathrm{p}}^{\boldsymbol{f}_k\boldsymbol{f}_{k'}}(\boldsymbol{\theta}) &= \begin{cases} -\boldsymbol{N}_k^{\top} \boldsymbol{D}_c \boldsymbol{N}_k - \lambda_k \boldsymbol{K}_k, & k = k', \\ \boldsymbol{N}_k^{\top} \boldsymbol{D}_c \boldsymbol{N}_{k'}, & k \neq k', \end{cases}
\end{aligned}
$$

where $\boldsymbol{D}_a = \mathrm{diag}\{a_1, ..., a_n\}$, $\boldsymbol{D}_c = \mathrm{diag}\{c_1, ..., c_n\}$, $\boldsymbol{D}_d = \mathrm{diag}\{d_1, ..., d_n\}$, and $\boldsymbol{m} = (m_1, ..., m_n)^\top$, with a_i being stated as in (3.1) and

$$c_i = \frac{\partial^2 \ell_i(\mu_i, \delta)}{\partial \mu_i^2} \left(\frac{\mathrm{d}\mu_i}{\mathrm{d}\eta_i} \right)^2 + \frac{\partial \ell_i(\mu_i, \delta)}{\partial \mu_i} \left(\frac{\partial}{\partial \mu_i} \frac{\mathrm{d}\mu_i}{\mathrm{d}\eta_i} \right) \frac{\mathrm{d}\mu_i}{\mathrm{d}\eta_i},$$

$$d_i = \frac{1}{2(\delta + 1)^2} - \frac{(y_i + \mu_i)^2}{(\delta\, y_i + y_i + \delta\mu_i)^2} - \frac{\mu_i}{2(\delta + 1)^3 y_i},$$

$$m_i = \frac{y_i}{(\delta\, y_i + y_i + \delta\mu_i)^2} + \frac{y_i}{4\mu_i^2} - \frac{\delta(\delta + 2)}{4(\delta + 1)^2 y_i}, \quad i \in \{1, ..., n\}.$$

In addition, calculating the expectation of the matrix $-\ddot{\boldsymbol{\ell}}_\mathrm{p}(\boldsymbol{\theta})$ given in (3.2), we obtain the $p^* \times p^*$ penalized expected information matrix expressed as

$$(3.3) \qquad \boldsymbol{\mathcal{I}}_\mathrm{p}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{\beta f}_1}(\boldsymbol{\theta}) & \cdots & \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{\beta f}_s}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{\beta\delta}}(\boldsymbol{\theta}) \\ \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{\beta f}_1^\top}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{f}_1 \boldsymbol{f}_1}(\boldsymbol{\theta}) & \cdots & \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{f}_1 \boldsymbol{f}_s}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{f}_1 \delta}(\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{\beta f}_s^\top}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{f}_1 \boldsymbol{f}_s^\top}(\boldsymbol{\theta}) & \cdots & \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{f}_s \boldsymbol{f}_s}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{f}_s \delta}(\boldsymbol{\theta}) \\ \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{\beta\delta}}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{f}_1 \delta}(\boldsymbol{\theta}) & \cdots & \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{f}_s \delta}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{I}}_\mathrm{p}^{\delta\delta}(\boldsymbol{\theta}) \end{pmatrix},$$

where each element of the matrix can be written as

$$\boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) = \boldsymbol{X}^\top \boldsymbol{D}_v \boldsymbol{X}, \quad \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{\beta f}_k}(\boldsymbol{\theta}) = \boldsymbol{X}^\top \boldsymbol{D}_v \boldsymbol{N}_k, \quad k \in \{1, ..., s\},$$

$$\boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{\beta\delta}}(\boldsymbol{\theta}) = \boldsymbol{X}^\top \boldsymbol{D}_a \boldsymbol{s}, \quad \boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{f}_k \delta}(\boldsymbol{\theta}) = \boldsymbol{N}^\top \boldsymbol{D}_a \boldsymbol{s}, \quad \mathcal{I}_\mathrm{p}^{\delta\delta}(\boldsymbol{\theta}) = \mathrm{tr}(\boldsymbol{D}_u),$$

$$\boldsymbol{\mathcal{I}}_\mathrm{p}^{\boldsymbol{f}_k \boldsymbol{f}_{k'}}(\boldsymbol{\theta}) = \begin{cases} \boldsymbol{N}_k^\top \boldsymbol{D}_v \boldsymbol{N}_k + \lambda_k \boldsymbol{K}_k, & k = k', \\ \\ \boldsymbol{N}_k^\top \boldsymbol{D}_v \boldsymbol{N}_{k'} & k \neq k', \end{cases}$$

with $\boldsymbol{D}_v = \mathrm{diag}\{v_1, ..., v_n\}$, $\boldsymbol{D}_u = \mathrm{diag}\{u_1, ..., u_n\}$ and $\boldsymbol{s} = (s_1, ..., s_n)^\top$, considering

$$v_i = \frac{\delta a_i^2}{2\mu_i^2} + \frac{\delta^2 a_i^2}{(\delta + 1)^2} \mathcal{J}(\boldsymbol{\theta}), \quad u_i = \frac{(\delta^2 + 3\delta + 1)}{2\delta^2(\delta + 1)^2} + \frac{\mu_i^2}{(\delta + 1)^4} \mathcal{J}(\boldsymbol{\theta}), \quad s_i = \frac{1}{2\mu_i(\delta + 1)} + \frac{\delta\mu_i}{(\delta + 1)^3} \mathcal{J}(\boldsymbol{\theta}),$$

for $i \in \{1, ..., n\}$, where

$$\mathcal{J}(\boldsymbol{\theta}) = \mathrm{E}\left(\left(Y + \frac{\mu\delta}{(\delta + 1)} \right)^{-2} \right)$$

$$= \int_0^\infty \frac{\sqrt{\delta + 1}\, \exp(\delta/2)}{4\sqrt{\pi\mu}\, y^{3/2}} \left(y + \frac{\delta\mu}{\delta + 1} \right)^{-2} \exp\left(-\frac{\delta}{4} \left(\frac{(\delta + 1)y}{\delta\mu} + \frac{\delta\mu}{(\delta + 1)y} \right) \right) \mathrm{d}y.$$

Note that, in the RBS-SAM, the property of orthogonality between parameters vectors $(\boldsymbol{\beta}, \boldsymbol{f}_k)$ and δ is not verified, unlike what is observed in other models, as the class of GLM, among others. In general, the orthogonality property simplifies the estimation process, in the sense that it allows the parameters to be estimated separately. More details about this issue in the semi-parametric context can be found in [21].

3.3. Finding the solution in practice: Iterative process

To estimate the model parameters by the MPL method, we solve the equation $\boldsymbol{U}_\mathrm{p}(\boldsymbol{\theta}) = \boldsymbol{0}$. However, as mentioned, no closed-form expressions for the MPL estimate of $\boldsymbol{\theta}$ are available.

Then, an iterative method for non-linear optimization is needed, such as the Fisher scoring or Newton or quasi-Newton algorithms, where the Broyden–Fletcher–Goldfarb–Shanno method provides often good results. Considering that the matrix $-\ddot{\ell}_{\mathrm{p}}(\boldsymbol{\theta})$ can be non-positive definite, we suggest replacing it with the matrix $-\mathcal{I}_{\mathrm{p}}(\boldsymbol{\theta})$ and using the Fisher scoring method. Then, the algorithm for estimating $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} + (\mathcal{I}_{\mathrm{p}}(\boldsymbol{\theta})^{-1})^{(m)} \boldsymbol{U}_{\mathrm{p}}(\boldsymbol{\theta})^{(m)}, \quad m \in \{0, 1, \ldots\},$$

which is equivalent to solving the matrix equation

(3.4)
$$\begin{pmatrix} \boldsymbol{X}^\top \boldsymbol{D}_v \boldsymbol{X} & \boldsymbol{X}^\top \boldsymbol{D}_v \boldsymbol{N}_1 & \cdots & \boldsymbol{X}^\top \boldsymbol{D}_v \boldsymbol{N}_s & \boldsymbol{X}^\top \boldsymbol{D}_a \boldsymbol{s} \\ \boldsymbol{N}_1^\top \boldsymbol{D}_v \boldsymbol{X} & \boldsymbol{N}_1^\top \boldsymbol{D}_v \boldsymbol{N}_1 + \lambda_1 \boldsymbol{K}_1 & \cdots & \boldsymbol{N}_1^\top \boldsymbol{D}_v \boldsymbol{N}_s & \boldsymbol{N}_1^\top \boldsymbol{D}_a \boldsymbol{s} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{N}_s^\top \boldsymbol{D}_v \boldsymbol{X} & \boldsymbol{N}_s^\top \boldsymbol{D}_v \boldsymbol{N}_1 & \cdots & \boldsymbol{N}_s^\top \boldsymbol{D}_v \boldsymbol{N}_s + \lambda_s \boldsymbol{K}_s & \boldsymbol{N}_s^\top \boldsymbol{D}_a \boldsymbol{s} \\ \boldsymbol{s}^\top \boldsymbol{D}_a \boldsymbol{X} & \boldsymbol{s}^\top \boldsymbol{D}_a \boldsymbol{N}_1 & \cdots & \boldsymbol{s}^\top \boldsymbol{D}_a \boldsymbol{N}_s & \mathrm{tr}(\boldsymbol{D}_u) \end{pmatrix}^{(m)} \begin{pmatrix} \boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(m+1,m)} \\ \boldsymbol{\Delta}_{\boldsymbol{f}_1}^{(m+1,m)} \\ \vdots \\ \boldsymbol{\Delta}_{\boldsymbol{f}_s}^{(m+1,m)} \\ \Delta_{\delta}^{(m+1,m)} \end{pmatrix}$$

$$= \begin{pmatrix} \boldsymbol{X}^\top \boldsymbol{D}_a \boldsymbol{z} \\ \boldsymbol{N}_1^\top \boldsymbol{D}_a \boldsymbol{z} - \lambda_1 \boldsymbol{K}_1 \boldsymbol{f}_1 \\ \vdots \\ \boldsymbol{N}_s^\top \boldsymbol{D}_a \boldsymbol{z} - \lambda_s \boldsymbol{K}_s \boldsymbol{f}_s \\ \mathrm{tr}(\boldsymbol{D}_b) \end{pmatrix}^{(m)}$$

where $\boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(m+1,m)} = \boldsymbol{\beta}^{(m+1)} - \boldsymbol{\beta}^{(m)}$, $\boldsymbol{\Delta}_{\boldsymbol{f}_k}^{(m+1,m)} = \boldsymbol{f}_k^{(m+1)} - \boldsymbol{f}_k^{(m)}$ and $\Delta_{\delta}^{(m+1,m)} = \delta^{(m+1)} - \delta^{(m)}$. Then, after algebraic manipulation, we obtain expressions for the iterative solutions stated as

$$\boldsymbol{\beta}^{(m+1)} = (\boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \left(\boldsymbol{\psi}_{\boldsymbol{\beta}}^{(m)} - \boldsymbol{D}_{v,a}^{(m)} \boldsymbol{s} \Delta_{\delta}^{(m+1,m)} - \sum_{k=1}^{s} \boldsymbol{N}_k \boldsymbol{\Delta}_{\boldsymbol{f}_k}^{(m+1,m)} \right),$$

$$\boldsymbol{f}_k^{(m+1)} = (\boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)} \boldsymbol{N}_k + \lambda_k \boldsymbol{K})^{-1} \boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)}$$
$$\times \left(\boldsymbol{\psi}_{\boldsymbol{f}_l}^{(m)} - \boldsymbol{D}_{v,a}^{(m)} \boldsymbol{s} \Delta_{\delta}^{(m+1,m)} - \boldsymbol{X} \boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(m+1,m)} - \sum_{k=1,k\neq\ell}^{s} \boldsymbol{N}_k \boldsymbol{\Delta}_{\boldsymbol{f}_k}^{(m+1,m)} \right), \quad k \in \{1, \ldots, s\},$$

$$\delta^{(m+1)} = \mathrm{tr}^{-1}(\boldsymbol{D}_u^{(m)})$$
$$\times \left(\mathrm{tr}(\boldsymbol{D}_b^{(m)}) + \mathrm{tr}(\boldsymbol{D}_u^{(m)}) \delta^{(m)} - \boldsymbol{s}^\top \boldsymbol{D}_a^{(m)} \boldsymbol{X} \boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(m+1,m)} - \boldsymbol{s}^\top \boldsymbol{D}_a^{(m)} \sum_{k=1}^{s} \boldsymbol{N}_k \boldsymbol{\Delta}_{\boldsymbol{f}_k}^{(m+1,m)} \right),$$

where $\boldsymbol{\psi}_{\boldsymbol{\beta}}^{(m)} = \boldsymbol{D}_{v,a}^{(m)} \boldsymbol{z}^{(m)} + \boldsymbol{X}\boldsymbol{\beta}^{(m)}$ and $\boldsymbol{\psi}_{\boldsymbol{f}_l}^{(m)} = \boldsymbol{D}_{v,a}^{(m)} \boldsymbol{z}^{(m)} + \boldsymbol{N}_l \boldsymbol{f}_l^{(m)}$, with $\boldsymbol{D}_{v,a}^{(m)} = \boldsymbol{D}_v^{(m)-1} \boldsymbol{D}_a^{(m)}$.

In general, the system of equations given in (3.4) is consistent and the back-fitting algorithm stated as in (3.5) converges to a solution for any starting values if the $\boldsymbol{D}_v$ weight matrix is symmetric and positive-definite. Additionally, this solution is unique under no concurvity in the data [4, 21].

When δ is known, it is possible to obtain simplified expressions for the iterative solutions of $\boldsymbol{\beta}^{(m+1)}$ and $\boldsymbol{f}_k^{(m+1)}$. Indeed, after some algebraic manipulation, we get that

$$\boldsymbol{\beta}^{(m+1)} = (\boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \Big(\boldsymbol{\psi}_{\boldsymbol{\beta}}^{(m)} - \sum_{k=1}^{s} \boldsymbol{N}_k \boldsymbol{\Delta}_{\boldsymbol{f}_k}^{(m+1,m)} \Big),$$

$$\boldsymbol{f}_k^{(m+1)} = (\boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)} \boldsymbol{N}_k + \lambda_k \boldsymbol{K})^{-1} \boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)} \Big(\boldsymbol{\psi}_{\boldsymbol{f}_l}^{(m)} - \boldsymbol{X} \boldsymbol{\Delta}_{\boldsymbol{\beta}}^{(m+1,m)} - \sum_{k=1,k\neq\ell}^{s} \boldsymbol{N}_k \boldsymbol{\Delta}_{\boldsymbol{f}_k}^{(m+1,m)} \Big),$$

for $k \in \{1, ..., s\}$, or, equivalently,

$$\boldsymbol{\beta}^{(m+1)} = (\boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \Big(\boldsymbol{r}_{v,a}^{(m)} - \sum_{k=1}^{s} \boldsymbol{N}_k \boldsymbol{f}_k^{(m+1)} \Big),$$

$$\boldsymbol{f}_k^{(m+1)} = (\boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)} \boldsymbol{N}_k + \lambda_k \boldsymbol{K})^{-1} \boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)} \Big(\boldsymbol{r}_{v,a}^{(m)} - \boldsymbol{X} \boldsymbol{\beta}^{(m+1)} - \sum_{k=1,k\neq\ell}^{s} \boldsymbol{N}_k \boldsymbol{f}_k^{(m+1)} \Big),$$

where $\boldsymbol{r}_{v,a}^{(m)} = \boldsymbol{D}_{(v,a)}^{(m)} \boldsymbol{z}^{(m)} + \boldsymbol{\eta}^{(m)}$, with $\boldsymbol{\eta}^{(m)} = \boldsymbol{X}\boldsymbol{\beta}^{(m)} + \sum_{k=1}^{s} \boldsymbol{N}_k \boldsymbol{f}_k^{(m)}$. It is possible to prove that these expressions correspond to the weighted back-fitting (Gauss–Seidel) iterations considering $\boldsymbol{r}_{v,a}^{(m)}$ as a dependent modified variable and $\boldsymbol{D}_v$ as a matrix of weights that changes with each iteration of the process. A general formula for these iterations is stated as

$$(3.5) \qquad \boldsymbol{f}_l^{(m+1)} = \boldsymbol{S}_l^{(m)} \Big(\boldsymbol{r}_{v,a}^{(m)} - \sum_{k=0,k\neq l}^{s} \boldsymbol{N}_k \boldsymbol{f}_k^{(m+1)} \Big), \quad l \in \{0,1,...,s\},$$

where $\boldsymbol{r}_{v,a}^{(m)} = \boldsymbol{D}_{v,a}^{(m)} \boldsymbol{z}^{(m)} + \boldsymbol{\eta}^{(m)}$, with $\boldsymbol{\eta}^{(m)} = \sum_{k=0}^{s} \boldsymbol{N}_k \boldsymbol{f}_k^{(m)}$, $\boldsymbol{N}_0 = \boldsymbol{X}$, $\boldsymbol{f}_0 = \boldsymbol{\beta}$, and

$$\boldsymbol{S}_0^{(m)} = (\boldsymbol{N}_0^\top \boldsymbol{D}_v^{(m)} \boldsymbol{N}_0)^{-1} \boldsymbol{N}_0^\top \boldsymbol{D}_v^{(m)}, \boldsymbol{S}_k^{(m)} = (\boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)} \boldsymbol{N}_k + \lambda_k \boldsymbol{K}_k)^{-1} \boldsymbol{N}_k^\top \boldsymbol{D}_v^{(m)}, \quad k \in \{1, ..., s\}.$$

Note that if the nonparametric component is not present in the model formulated in (2.5), the iterative process presented in (3.5) reduces to the expressions proposed in [27], that is, to $\boldsymbol{\beta}^{(m+1)} = (\boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{D}_v^{(m)} \boldsymbol{z}_{\boldsymbol{\beta}}^{(m)}$, where $\boldsymbol{z}_{\boldsymbol{\beta}}^{(m)} = \boldsymbol{D}_{v,a}^{(m)} \boldsymbol{z}^{(m)} + \boldsymbol{X}\boldsymbol{\beta}^{(m)}$. A similar formula can be obtained for the case where the parametric component is absent in the model and only a smooth function is considered as linear predictor. Specifically, the iterative process would reduce to $\boldsymbol{f}^{(m+1)} = (\boldsymbol{N}^\top \boldsymbol{D}_v^{(m)} \boldsymbol{N})^{-1} \boldsymbol{N}^\top \boldsymbol{D}_v^{(m)} \boldsymbol{z}_{\boldsymbol{f}}^{(m)}$, where $\boldsymbol{z}_f^{(m)} = \boldsymbol{D}_{v,a}^{(m)} \boldsymbol{z}^{(m)} + \boldsymbol{N}\boldsymbol{f}^{(m)}$. Thus, at convergence, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{f}}$ can be interpreted as least squares estimators of $\boldsymbol{D}_{\widehat{v}}^{1/2} \widehat{\boldsymbol{z}}_{\boldsymbol{\beta}}$ and $\boldsymbol{D}_{\widehat{v}}^{1/2} \widehat{\boldsymbol{z}}_f$ against the columns of $\boldsymbol{D}_{\widehat{v}}^{1/2} \boldsymbol{X}$ and $\boldsymbol{D}_{\widehat{v}}^{1/2} \boldsymbol{N}$, respectively, where $\widehat{\boldsymbol{z}}_{\boldsymbol{\beta}} = \boldsymbol{D}_{\widehat{v},\widehat{a}} \widehat{\boldsymbol{z}} + \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{z}}_{\boldsymbol{f}} = \boldsymbol{D}_{\widehat{v},\widehat{a}} \widehat{\boldsymbol{z}} + \boldsymbol{N}\widehat{\boldsymbol{f}}$.

3.4. Approximate standard errors

Next, we consider the problem of estimating the variance-covariance matrix of the MPL estimator $\widehat{\boldsymbol{\theta}}$. Considering the fact that we obtain the MPL estimate of $\boldsymbol{\theta}$ through the Fisher scoring algorithm, it is reasonable to derive the corresponding variance-covariance matrix by using the inverse of the penalized Fisher information matrix [20, 50, 51]. To compute the

inverse matrix of $\mathcal{I}_{\mathrm{p}}(\boldsymbol{\theta})$ given in (3.3), consider

$$
\mathcal{I}_{\mathrm{p}}^{11} = \begin{pmatrix}
\mathcal{I}_{\mathrm{p}}^{\boldsymbol{\beta\beta}}(\boldsymbol{\theta}) & \mathcal{I}_{\mathrm{p}}^{\boldsymbol{\beta f}_1}(\boldsymbol{\theta}) & \cdots & \mathcal{I}_{\mathrm{p}}^{\boldsymbol{\beta f}_s}(\boldsymbol{\theta}) \\
\mathcal{I}_{\mathrm{p}}^{\boldsymbol{\beta f}_1^\top}(\boldsymbol{\theta}) & \mathcal{I}_{\mathrm{p}}^{\boldsymbol{f}_1 \boldsymbol{f}_1}(\boldsymbol{\theta}) & \cdots & \mathcal{I}_{\mathrm{p}}^{\boldsymbol{f}_1 \boldsymbol{f}_s}(\boldsymbol{\theta}) \\
\vdots & \vdots & \ddots & \vdots \\
\mathcal{I}_{\mathrm{p}}^{\boldsymbol{\beta f}_s^\top}(\boldsymbol{\theta}) & \mathcal{I}_{\mathrm{p}}^{\boldsymbol{f}_1 \boldsymbol{f}_s^\top}(\boldsymbol{\theta}) & \cdots & \mathcal{I}_{\mathrm{p}}^{\boldsymbol{f}_s \boldsymbol{f}_s}(\boldsymbol{\theta})
\end{pmatrix}, \quad
\mathcal{I}_{\mathrm{p}}^{12} = \begin{pmatrix}
\mathcal{I}_{\mathrm{p}}^{\boldsymbol{\beta}\delta}(\boldsymbol{\theta}) \\
\mathcal{I}_{\mathrm{p}}^{\boldsymbol{f}_1 \delta}(\boldsymbol{\theta}) \\
\vdots \\
\mathcal{I}_{\mathrm{p}}^{\boldsymbol{f}_s \delta}(\boldsymbol{\theta})
\end{pmatrix}, \quad
\mathcal{I}_{\mathrm{p}}^{22} = \mathcal{I}_{\mathrm{p}}^{\delta\delta}.
$$

Thus, the matrix $\mathcal{I}_{\mathrm{p}}(\boldsymbol{\theta})$ can be written as

$$(3.6) \qquad \mathcal{I}_{\mathrm{p}}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}_{\mathrm{p}}^{11} & \mathcal{I}_{\mathrm{p}}^{12} \\ \mathcal{I}_{\mathrm{p}}^{12^\top} & \mathcal{I}_{\mathrm{p}}^{22} \end{pmatrix}.$$

Assuming that all the necessary inverses exist, some algebraic manipulations on the expression stated in (3.6) show that the inverse matrix of $\mathcal{I}_{\mathrm{p}}(\boldsymbol{\theta})$ assumes a block form given by

$$
\mathcal{I}_{\mathrm{p}}^{-1}(\boldsymbol{\theta}) = \begin{pmatrix}
\mathcal{I}_{\mathrm{p}}^{11.1} & -\mathcal{I}_{\mathrm{p}}^{11.1} \mathcal{I}_{\mathrm{p}}^{12} \mathcal{I}_{\mathrm{p}}^{22^{-1}} \\
-\mathcal{I}_{\mathrm{p}}^{22^{-1}} \mathcal{I}_{\mathrm{p}}^{12^\top} \mathcal{I}_{\mathrm{p}}^{11.1} & \mathcal{I}_{\mathrm{p}}^{22.1}
\end{pmatrix},
$$

where $\mathcal{I}_{\mathrm{p}}^{11.1} = (\mathcal{I}_{\mathrm{p}}^{11} - \mathcal{I}_{\mathrm{p}}^{12} \mathcal{I}_{\mathrm{p}}^{22^{-1}} \mathcal{I}_{\mathrm{p}}^{12^\top})^{-1}$ and $\mathcal{I}_{\mathrm{p}}^{22.1} = \mathcal{I}_{\mathrm{p}}^{22^{-1}} + \mathcal{I}_{\mathrm{p}}^{22^{-1}} \mathcal{I}_{\mathrm{p}}^{12^\top} \mathcal{I}_{\mathrm{p}}^{11.1} \mathcal{I}_{\mathrm{p}}^{12} \mathcal{I}_{\mathrm{p}}^{22^{-1}}$. Therefore, the asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\theta}}$ is defined as

$$(3.7) \qquad \widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\theta}}) \approx \mathcal{I}_{\mathrm{p}}^{-1}(\boldsymbol{\theta})\big|_{\widehat{\boldsymbol{\theta}}}.$$

In particular, we have that $\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{f}}_1, ..., \widehat{\boldsymbol{f}}_s) \approx \mathcal{I}_{\mathrm{p}}^{11.1}\big|_{\widehat{\boldsymbol{\theta}}}$ and $\widehat{\mathrm{Cov}}(\widehat{\delta}) \approx \mathcal{I}_{\mathrm{p}}^{22.1}\big|_{\widehat{\boldsymbol{\theta}}}$. To approximate the pointwise standard errors (SE) bands for nonparametric functions f_k, one can evaluate the accuracy of the estimators $\widehat{f}_k$ for different locations within the range of interest [17]. In our case, these SE bands are constructed using the corresponding diagonal elements of the matrix $\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{f}}_1, ..., \widehat{\boldsymbol{f}}_s)$ as estimators of the SE of f_k. Indeed, we can consider as approximate pointwise SE bands the expression defined as $\mathrm{SE}(f_k(t_l^0)) = \widehat{f}_k(t_l^0) \pm 2(\widehat{\mathrm{Var}}(\widehat{f}_k(t_l^0)))^{1/2}$, where $\mathrm{Var}(\widehat{f}_k(t_l))$ is the l-th principal diagonal element of the matrix stated in (3.7), for $l \in \{1, ..., r\}$. Note that t_l^0 corresponds to the knots associated with each variable whose contribution to the model is nonparametric.

Observe that the asymptotic approach based on the penalized Fisher information matrix to approximate the variance-covariance matrix of $\widehat{\boldsymbol{\theta}}$ has been used by several authors, and has resulted in an efficient tool to approximate the SEs of $\widehat{\boldsymbol{\theta}}$ and construct SE bands for smooth functions [22, 23, 50].

3.5. Degrees of freedom

Next, we define the degrees of freedom associated with the parametric and nonparametric components. In the latter case, they correspond to the number of effective parameters that are being considered in the nonparametric modeling, and that are used in the selection of the smoothing parameters. Our definition of degrees of freedom is based on the convergence of the iterative process given in (3.5) to select $\widehat{f}_l$, for $l \in \{0, 1, ..., s\}$. Indeed, fixing δ

and λ_k, we obtain $\widehat{\boldsymbol{f}}_l = \widehat{\boldsymbol{S}}_l \widehat{\boldsymbol{r}}_{v,a}^*$, for $l \in \{0, 1, ..., s\}$, where $\widehat{\boldsymbol{r}}_{v,a}^* = \widehat{\boldsymbol{r}}_{v,a} - \sum_{k=0,k\neq l}^s \boldsymbol{N}_k \widehat{\boldsymbol{f}}_k$, with $\widehat{\boldsymbol{r}}_{v,a} = \widehat{\boldsymbol{D}}_{v,a} \widehat{\boldsymbol{z}} + \widehat{\boldsymbol{\eta}}$, $\widehat{\boldsymbol{\eta}} = \sum_{k=0}^s \boldsymbol{N}_k \widehat{\boldsymbol{f}}_k$ and

$$\begin{aligned} \widehat{\boldsymbol{S}}_0 &= (\boldsymbol{N}_0^\top \widehat{\boldsymbol{D}}_v \boldsymbol{N}_0)^{-1} \boldsymbol{N}_0^\top \widehat{\boldsymbol{D}}_v, \\ \widehat{\boldsymbol{S}}_k &= (\boldsymbol{N}_k^\top \widehat{\boldsymbol{D}}_v \boldsymbol{N}_k + \lambda_k \boldsymbol{K}_k)^{-1} \boldsymbol{N}_k^\top \widehat{\boldsymbol{D}}_v, \quad k \in \{1, ..., s\}. \end{aligned}$$

In the literature concerning to additive models, there are different definitions for the degrees of freedom, depending on the context in which they are used [6]. In our case, the degrees of freedom associated with the parametric component, $\boldsymbol{N}_0 \widehat{\boldsymbol{f}}_0$ namely, are defined as

$$\mathrm{df}_{\boldsymbol{N}_0} = \mathrm{tr}\big\{ \boldsymbol{N}_0 (\boldsymbol{N}_0^\top \widehat{\boldsymbol{D}}_v \boldsymbol{N}_0)^{-1} \boldsymbol{N}_0^\top \widehat{\boldsymbol{D}}_v \big\} = p,$$

where p is the rank of $\boldsymbol{N}_0 = \boldsymbol{X}$. In addition, the degrees of freedom for the nonparametric component, $\boldsymbol{N}_k \widehat{\boldsymbol{f}}_k$ namely, are stated as

$$(3.8) \qquad \mathrm{df}(\lambda_k) = \mathrm{tr}\big\{ \boldsymbol{N}_k (\boldsymbol{N}_k^\top \widehat{\boldsymbol{D}}_v \boldsymbol{N}_k + \lambda_k \boldsymbol{K}_k)^{-1} \boldsymbol{N}_k^\top \widehat{\boldsymbol{D}}_v \big\},$$

which measure the individual effect contribution of the k-th nonparametric component.

3.6. Selection of smoothing parameters

Note that, based on [23], we consider a grid of values for the smoothing parameter and select one of them for which the AIC is minimized. Alternatively, we can consider the inverse relationship between the degrees of freedom and the smoothing parameter [23], and select the value of this parameter associated with a specific value of the degrees of freedom.

In the previous sections, the smoothing parameters λ_k were assumed fixed. However, in practice situations, the smoothing parameters should be selected from the data. A criterion to select the smoothing parameters based on the AIC is described below. Before proposing a criterion to select the smoothing parameters, remember that, under the RBS-SAM, we have a total of $1 + p + \mathrm{df}(\boldsymbol{\lambda})$ parameters to be determined, with $\mathrm{df}(\boldsymbol{\lambda}) = \sum_{k=1}^s \mathrm{df}(\lambda_k)$ denoting the number of effective parameters involved in the modeling of the smooth functions. In this case, the AIC (or alternatively the Bayesian information criterion) can be used for selecting the smoothing parameters λ_k. The idea is to minimize a function with respect to $\boldsymbol{\lambda}$ formulated as

$$\mathrm{AIC}(\boldsymbol{\lambda}) = -2\ell_\mathrm{p}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + 2(1 + p + \mathrm{df}(\boldsymbol{\lambda})),$$

where $\ell_\mathrm{p}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda})$ denotes the penalized log-likelihood function evaluated at $\widehat{\boldsymbol{\theta}}$ for a fixed $\boldsymbol{\lambda}$. A grid (surface) for different values of $\boldsymbol{\lambda}$ and its corresponding $\mathrm{AIC}(\boldsymbol{\lambda})$ are helpful to choose the suitable smoothing parameters.

Other approach to selecting the smoothing parameters is when the degrees of freedom given in (3.8) depends only on λ_k and, therefore, the corresponding smoothing parameter can be specified. In other words, we specify an objective $\mathrm{df}(\lambda_k)$ for a function and then find the value λ_k that achieves this objective. Such an approach has been used by a number of authors [6, 18, 22, 44].

4.   DIAGNOSTIC ANALYTICS

In this section, we derive the local influence method for different perturbation schemes to assess the potential influence of some observations on the RBS-SAM. These schemes are the case-weight, response and precision parameter perturbations.

4.1. The local influence method

Let $\boldsymbol{\omega} = (\omega_1, ..., \omega_n)^\top$ be an $n \times 1$ vector of perturbations restricted to some open subset $\Omega \in \mathbb{R}^n$ and $\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})$ be the logarithm of the perturbed penalized likelihood function. It is assumed that exists $\boldsymbol{\omega}_0 \in \Omega$, a vector of no perturbation, such that $\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega}_0) = \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda})$. To assess the influence of small perturbations on the MPL estimate $\widehat{\boldsymbol{\theta}}$, we can consider the likelihood displacement stated as

$$\mathrm{LD}(\boldsymbol{\omega}) = 2\Big(\ell_{\mathrm{p}}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) - \ell_{\mathrm{p}}(\widehat{\boldsymbol{\theta}}_\omega, \boldsymbol{\lambda})\Big) \geq 0,$$

where $\widehat{\boldsymbol{\theta}}_\omega$ is the MPL estimate under $\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})$. The measure $\mathrm{LD}(\boldsymbol{\omega})$ is helpful for assessing the distance between $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_\omega$. In [9], it was suggested to study the local behavior of $\mathrm{LD}(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$. The procedure consists of selecting a unit direction $\boldsymbol{d} \in \Omega$, with $\|\boldsymbol{d}\| = 1$, and then to consider the plot of $\mathrm{LD}(\boldsymbol{\omega}_0 + a\boldsymbol{d})$ against a, where $a \in \mathbb{R}$. This plot is called lifted line. Each lifted line may be characterized by considering the normal curvature $C_d(\boldsymbol{\theta})$ around $a = 0$. The suggestion is to assume the direction $\boldsymbol{d} = \boldsymbol{d}_{\max}$ corresponding to the largest curvature $C_{d_{\max}}(\boldsymbol{\theta})$. The index plot of $\boldsymbol{d}_{\max}$ can identify those cases that, under small perturbations, provoke an importante potential influence on $\mathrm{LD}(\boldsymbol{\omega})$. According to [9], the normal curvature at the unitary direction $\boldsymbol{d}$ is expressed as

$$C_d(\boldsymbol{\theta}) = -2\big(\boldsymbol{d}^\top \boldsymbol{\Delta}_{\mathrm{p}}^\top \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{-1} \boldsymbol{\Delta}_{\mathrm{p}} \boldsymbol{d}\big),$$

with

$$\ddot{\boldsymbol{\ell}}_{\mathrm{p}} = \frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\bigg|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}}, \quad \boldsymbol{\Delta}_{\mathrm{p}} = \frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^\top}\bigg|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \, \boldsymbol{\omega} = \boldsymbol{\omega}_0}.$$

Note that $-\ddot{\boldsymbol{\ell}}_{\mathrm{p}}$ is the penalized observed information matrix evaluated at $\widehat{\boldsymbol{\theta}}$ (see Subsection 3.2) and $\boldsymbol{\Delta}_{\mathrm{p}}$ is the penalized perturbation matrix evaluated at $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega}_0$. Observe that $C_d(\boldsymbol{\theta})$ denotes the local influence on the estimate $\widehat{\boldsymbol{\theta}}$ after perturbing the model or data. In [11], it was proposed to study the normal curvature at the direction $\boldsymbol{d} = \boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is an $n \times 1$ vector with a one at the i-th position and zeros at the remaining positions. Thus, the normal curvature, called total local influence of the i-th case, assumes the form $C_{\boldsymbol{e}_i}(\boldsymbol{\theta}) = 2|c_{ii}|$, for $i \in \{1, ..., n\}$, where c_{ii} is the i-th principal diagonal element of the matrix $\boldsymbol{C} = \boldsymbol{\Delta}_{\mathrm{p}}^\top \ddot{\boldsymbol{\ell}}_{\mathrm{p}}^{-1} \boldsymbol{\Delta}_{\mathrm{p}}$.

The local influence results presented in the following sections are extensions of the work presented in [27] but for the case where the linear predictor includes a parametric term and also an additive term of smooth functions.

4.2. Case-weight perturbation

The perturbation of case-weight is considered to identify observations with high contribution to the likelihood function and that can exercise strong influence on the MPL estimates. Let us to consider the attributed weights for the cases in the penalized log-likelihood function as

$$\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega}) = \sum_{i=1}^{n} \omega_i \ell_i(\boldsymbol{\theta}) - \sum_{k=1}^{s} \frac{\lambda_k}{2} \boldsymbol{f}_k^{\top} \boldsymbol{K}_k \boldsymbol{f}_k,$$

where $\boldsymbol{\omega} = (\omega_1, ..., \omega_n)^{\top}$ is the vector of weights, with $0 \le \omega_i \le 1$, for $i \in \{1, ..., n\}$, and $\boldsymbol{\omega}_0 = (1, ..., 1)^{\top}$ denotes the vector of no perturbation. Differentiating $\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})$ with respect to the elements of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$, we obtain

$$\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}}\bigg|_{\theta=\widehat{\theta}, \, \omega=\omega_0} = \boldsymbol{X}^{\top} \widehat{\boldsymbol{D}}_a \widehat{\boldsymbol{D}}_{\boldsymbol{z}},$$

$$\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \boldsymbol{f}_k \partial \boldsymbol{\omega}}\bigg|_{\theta=\widehat{\theta}, \, \omega=\omega_0} = \boldsymbol{N}_k^{\top} \widehat{\boldsymbol{D}}_a \widehat{\boldsymbol{D}}_{\boldsymbol{z}}, \quad k \in \{1, ..., s\},$$

$$\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \delta \partial \boldsymbol{\omega}}\bigg|_{\theta=\widehat{\theta}, \omega=\omega_0} = \widehat{\boldsymbol{b}},$$

for $i \in \{1, ..., n\}$, with $\boldsymbol{D}_a$ and $\boldsymbol{b}$ being defined in previous sections, whereas that $\boldsymbol{D}_{\boldsymbol{z}} = \mathrm{diag}\{z_1, ..., z_n\}$.

4.3. Response perturbation

According to [27], the additive perturbation on the i-th response is given by $y_{i\omega_i} = y_i + \omega_i s(y_i)$, where $s(y_i) = (\widehat{\mu}_i^2/\widehat{\phi})^{1/2}$ and $\omega_i \in \mathbb{R}$, for $i \in \{1, ..., n\}$. Then, the perturbed penalized log-likelihood function is constructed from the expression defined in (2.10) with y_i being replaced by $y_{i\omega}$, that is,

$$\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega}) = \ell(\boldsymbol{\theta}; \boldsymbol{\omega}) - \sum_{k=1}^{s} \frac{\lambda_k}{2} \boldsymbol{f}_k^{\top} \boldsymbol{K}_k \boldsymbol{f}_k,$$

where ℓ is given in (2.7) with $y_{i\omega_i}$ in the place of y_i, for $i \in \{1, ..., n\}$,. Here, the vector of no perturbation is stated as $\boldsymbol{\omega}_0 = (0, ..., 0)^{\top}$. Differentiating $\ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})$ with respect to the elements of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$, we obtain, after some algebraic manipulation, that

$$\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}}\bigg|_{\theta=\widehat{\theta}, \, \omega=\omega_0} = \boldsymbol{X}^{\top} \widehat{\boldsymbol{D}}_a \widehat{\boldsymbol{D}}_{\psi} \widehat{\boldsymbol{D}}_{\vartheta}, \quad \frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \delta \partial \boldsymbol{\omega}}\bigg|_{\theta=\widehat{\theta}, \, \omega=\omega_0} = \widehat{\boldsymbol{\tau}}^{\top} \boldsymbol{D}_{\vartheta},$$

$$\frac{\partial^2 \ell_{\mathrm{p}}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \boldsymbol{f}_k \partial \boldsymbol{\omega}}\bigg|_{\theta=\widehat{\theta}, \, \omega=\omega_0} = \boldsymbol{N}_k^{\top} \widehat{\boldsymbol{D}}_a \widehat{\boldsymbol{D}}_{\psi} \widehat{\boldsymbol{D}}_{\vartheta}, \quad k \in \{1, ..., s\},$$

where $\widehat{\boldsymbol{D}}_{\vartheta} = \mathrm{diag}\{\widehat{\vartheta}_1, ..., \widehat{\vartheta}_n\}$, $\widehat{\boldsymbol{D}}_{\psi} = \mathrm{diag}\{\widehat{\psi}_1, ..., \widehat{\psi}_n\}$, and $\widehat{\boldsymbol{D}}_{\tau} = \mathrm{diag}\{\widehat{\tau}_1, ..., \widehat{\tau}_n\}$, with $\widehat{\vartheta}_i = s(y_i)$,

$$\widehat{\psi}_i = -\frac{\widehat{\delta}(\widehat{\delta}+1)}{(\widehat{\delta}y_i + y_i + \widehat{\delta}\widehat{\mu}_i)^2} + \frac{(\widehat{\delta}+1)}{4\widehat{\mu}_i^2} + \frac{\widehat{\delta}^2}{4(\widehat{\delta}+1)y_i^2},$$

$$\widehat{\tau}_i = -\frac{\widehat{\mu}_i}{(\widehat{\delta}y_i + y_i + \widehat{\delta}\widehat{\mu}_i)^2} - \frac{1}{4\widehat{\mu}_i} + \frac{\widehat{\delta}(\widehat{\delta}+2)\widehat{\mu}_i}{4y_i^2(\widehat{\delta}+1)^2}, \quad i \in \{1, ..., n\}.$$

4.4. Perturbation on the precision parameter

Perturbation of the precision parameter is used for evaluating the sensitivity of the MPL estimate to small modifications of δ. Initially, the RBS-SAM assumes that the precision parameter is constant across data. However, under the perturbed model, the precision parameter is non-constant across cases, this is, $Y_i \sim \mathrm{RBS}(\mu_i, \delta_i)$, where $\delta_i = \delta/\omega_i$, with $\omega_i > 0$, for $i \in \{1, ..., n\}$. Under this perturbation, the vector of no perturbation is given by $\boldsymbol{\omega}_0 = (1, ..., 1)^\top$. Then, the perturbed penalized log-likelihood function is constructed from the expression given in (2.10) with δ being replaced by δ_i. Taking derivatives of $\ell_\mathrm{p}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})$ with respect to the elements of $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$, we obtain, after some algebraic manipulation, that

$$\left. \frac{\partial^2 \ell_\mathrm{p}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}} \right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = \boldsymbol{X}^\top \widehat{\boldsymbol{D}}_a \widehat{\boldsymbol{D}}_\varpi,$$

$$\left. \frac{\partial^2 \ell_\mathrm{p}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \boldsymbol{f}_k \partial \boldsymbol{\omega}} \right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = \boldsymbol{N}_k^\top \widehat{\boldsymbol{D}}_a \widehat{\boldsymbol{D}}_\varpi, \quad k \in \{1, ..., s\},$$

$$\left. \frac{\partial^2 \ell_\mathrm{p}(\boldsymbol{\theta}, \boldsymbol{\lambda}; \boldsymbol{\omega})}{\partial \delta \partial \boldsymbol{\omega}} \right|_{\theta=\widehat{\theta},\ \omega=\omega_0} = \widehat{\boldsymbol{\varphi}}^\top,$$

where $\widehat{\boldsymbol{D}}_\varpi = \mathrm{diag}\{\widehat{\varpi}_1, ..., \widehat{\varpi}_n\}$ and $\boldsymbol{\varphi} = (\widehat{\varphi}_1, ..., \widehat{\varphi}_n)^\top$, with

$$\varpi_i = -\frac{\widehat{\delta} y_i}{(\widehat{\delta} y_i + y_i + \widehat{\delta} \widehat{\mu}_i)^2} - \frac{\widehat{\delta} y_i}{4 \widehat{\mu}_i^2} + \frac{\widehat{\delta}^2(\widehat{\delta}+2)}{4 y_i(\widehat{\delta}+1)^2}, \quad i \in \{1, ..., n\},$$

$$\widehat{\varphi}_i = -\frac{1}{2} + \frac{1}{2(\widehat{\delta}+1)^2} - \frac{y_i(y_i + \widehat{\mu}_i)}{(\widehat{\delta} y_i + y_i + \widehat{\delta} \widehat{\mu}_i)^2} + \frac{y_i}{4 \widehat{\mu}_i} + \frac{\widehat{\delta}^2 \widehat{\mu}_i(\widehat{\delta}+3)}{4 y_i(\widehat{\delta}+1)^3} + \frac{\widehat{\delta} \widehat{\mu}_i}{y_i(\widehat{\delta}+1)^3}.$$

5. APPLICATION TO REAL POLLUTION DATA

In this section, we provide the empirical application of the proposed model to environmental data obtained from the website of the Chilean Ministry of Environment using the R software. This application is motivated, as mentioned, by the fact that inclusion of nonparametric functions greatly enhances the modeling when accommodating non-linear effects of covariates. These covariates correspond in our case to contents of pollutants and meteorological variables as atmospheric pressure, precipitation, relative humidity, temperature, and wind speed, with the response variable being the particulate matter (PM) content.

5.1. Data and definition of the problem

PM pollution is one of the main global urban environmental problems, affecting human health and life quality. According to the World Health Organization (WHO), nine out of every ten people on the planet breathe air that contains high levels of pollutants and seven million people die every year due to this cause (www.who.int). PM is classified according to its diameter, because particle size determines sites of deposition within the respiratory tract.

Coarser particles (those with a diameter over 10 micrometers–μm–) do not penetrate into airways. Instead, these particles are deposited in the upper respiratory tract and are cleared by cilia action. Inhalable particles measuring less than 10 μm are called PM10, whereas those smaller than 2.5 μm are called PM2.5. As size decreases, there is a higher possibility for PM to penetrate deeper into smaller alveoli and airways. Specifically, various effects are produced from exposure to PM, but the nature of those induced effects vary according to the PM composition. Indeed, there is evidence of an increase in the risk of cardiovascular diseases and mortality from exposure to PM2.5, which occurs even after short time periods [8].

In Chile, according to the Ministry of the Environment, 3,494 people died prematurely due to critical air levels during 2017 mainly due to extreme contents of PM2.5, and nine million inhabitants are exposed to levels of pollution that exceeds the air quality standards (`https://bit.ly/2u40gDq`). Santiago of Chile is one of the most polluted cities in the world in terms of PM2.5 and PM10, because of a combination of anthropogenic, meteorological and topographic factors [31]. Several studies indicate that inhabitants of Santiago are under risk, because of the city's poor air quality. As urban air quality declines, the risk of stroke, heart disease, lung cancer, and chronic and acute respiratory diseases, including asthma, increases for the people who live in cities with high air pollution levels. Specifically for Santiago, several investigations provide evidence that exposure to air pollutants produces a risk to the inhabitants of this city; see [34] and references therein.

Periodic episodes of extreme levels of air pollution sometimes occur for certain atmospheric contaminants. Such episodes and their associated high contents vary with geographical and meteorological fluctuations, depending on changes in both source and type of emissions. Because of these variations, PM contents are treated as non-negative random variables that can be modeled by statistical distributions. Frequently, these distributions are asymmetrical and present positive skewness [31]. The current official methodology employed by the Chilean authority in Santiago to predict PM10 contents is based on a multiple regression model using contents of atmospheric pollutants and meteorological variables as covariates [39]. It helps to forecast the maximum value of the 24-hour average content of PM10 in μg/normalized cubic meters (Nm^3) for the period from 00:00 to 24:00 hours of the next day. Furthermore, in 2015, through Supreme Decree number 15/2015 and resolution number 9664/2015, it was instructed by the Chilean Ministry of Health to declare sanitary alert employing also PM2.5 contents.

5.2. Descriptive data analysis

We consider the environmental data set related to air pollution. In particular, the data provided by the National Air Quality Information System (`https://sinca.mma.gob.cl`) corresponding to air pollution in the commune of Pudahuel in Santiago during the critical episodes management (CEM) period (01 April 2019 to 31 August 2019). In this application, we are interested on detecting the association of polluting contents with meteorological variables by utilizing the RBS-SAM. For motivating the semi-parametric models, we consider PM2.5 as response variable and two covariates, PM10 contents and WIND (speed wind in meters/second). In the CEM period, we work with a total of 153 observations.

Table 1 provides a descriptive summary of PM2.5, which includes mean ($\bar{y}$), median (MD), standard deviation (SD), coefficient of variation (CV), coefficient of skewness (CS), coefficient of kurtosis (CK), minimum ($y_{(1)}$), maximum ($y_{(n)}$), and the total of observations (n). Figure 1 contains the histogram and boxplot of PM2.5 contents. The primary air quality regulation for PM2.5 is 50μg/Nm^3, as 24-hour level.

Table 1: Descriptive statistics for the response variable PM2.5 in the CEM period with data from Santiago, Chile.

$\bar{y}$	MD	SD	CV	CS	CK	$y_{(1)}$	$y_{(n)}$	n
42.27	37	21.51	0.51	0.89	3.26	8	105	153

According to Table 1, the primary regulation is exceeded for the response; see, for example, the maximum contents registered in the CEM period. Note that CS $= 0.89$, indicating a slight asymmetry in the data of the response variable, and its CK $= 3.26$, indicating a probability density function with heavy tails in relation to the normal distribution. In addition, from the histogram displayed in Figure 1, we note that the values of PM2.5 have an empirical distribution that is positively skewed, while from Figure 1(b), we identify case #63 as an atypical observation in the boxplot. Consequently, from Table 1 and Figure 1, we propose that the RBS-SAM may be suitable for describing the mean of the data, the non-constant variance and asymmetry detected in the distribution of these data.

(a)   (b)
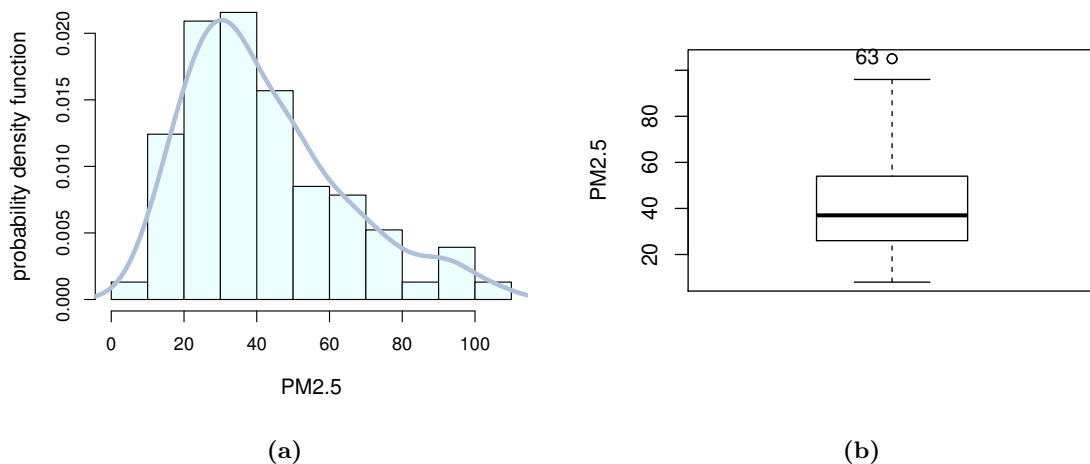
Figure 1: Histogram (a) and boxplot (b) for the response variable PM2.5 in CEM period with data from Santiago, Chile.

Figure 2 contains the scatter plots between the response variable and each covariate. In Figure 2(a), we see that the linear correlation between PM2.5 and the covariate PM10 seems to be positive, which is supported by the correlation coefficient between both variables (0.9261). In addition, we observe that the variability of PM2.5 tends to increase as the PM10 values increase, which could be an indication of a non-constant variance in the data.

Figure 2(a) also shows evidence that the straight line do not go through the origin, so that considering the intercept in the parametric component of the selected model could be helpful. Figure 2(b) indicates that the relationship between PM2.5 and WIND seems to be non-linear.

(a)   (b)
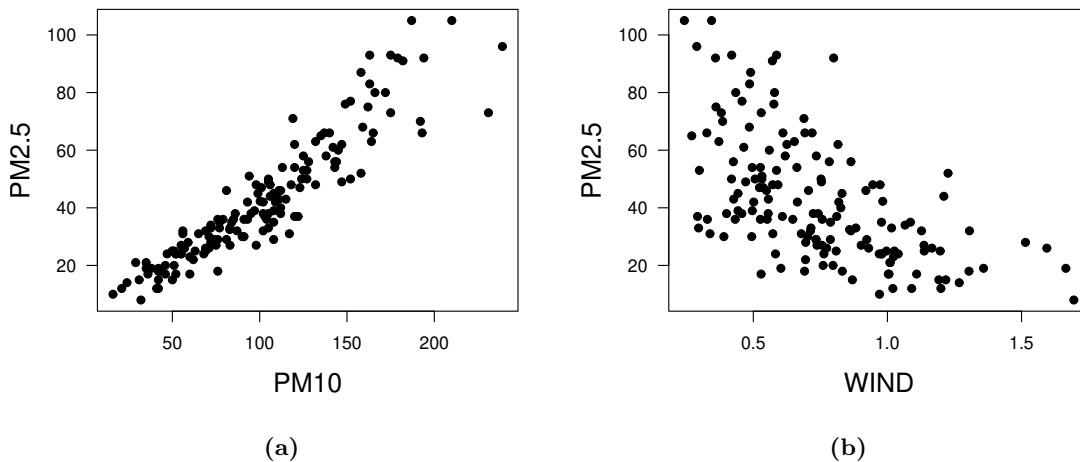
Figure 2: Scatter plots of PM2.5 versus PM10 (a) and PM2.5 versus WIND (b) with data from Santiago, Chile.

After this descriptive data analysis, we can see that the response variable presents characteristics of the RBS distribution, such as positive asymmetry. In addition, as mentioned, the BS distribution has mathematical arguments that allows us to justify its use when describing environmental data. Furthermore, from the scatter plots, note that the covariates contribute linearly and non-linearly to the response variable. Therefore, a semi-parametric structure may be proposed to model PM2.5 contents as a function of PM10 and WIND, contributing parametrically and nonparametrically to the model, respectively.

5.3. Model fitting

The trends described in the exploratory data analysis suggest an RBS-SAM among PM2.5 and the covariates, assuming the identity link function, that is,

$$h(\mu_i) = \mu_i = \beta_1 + x_i\beta_2 + f(t_i), \quad i \in \{1, ..., 153\},$$

where x_i and t_i denotes the values of PM10 and WIND from the i-th case, respectively, $(\beta_1, \beta_2)^\top$ is a parameter vector, and f is a smooth function. We apply the procedure described in Subsection 3.6 to select the smoothing parameter, which results to be $\lambda = 0.0001$.

To estimate the parameters of the parametric component of the model, we maximize the penalized log-likelihood function as described in Subsection 3.3, obtaining $\widehat{\beta_1} = -30.013$, $\widehat{\beta_2} = 0.387$ and $\widehat{\delta} = 60.991$, whose corresponding SEs are 0.000095, 0.0068 and 0.0000011, respectively. To assess that the RBS-SAM is adequate to describe the mean of the response variable, we verify the assumptions established for the model. First, note that β_1 and β_2 are highly significant at 5%, since both empirical p-values (omitted here) are close to zero, as expected by the exploratory data analysis that we performed. Thus, the parametric component of the selected model seems to be adequate.

5.4. Residual analysis

Figure 3 shows the plot of the partial residuals defined as

$$r_i^{(\text{p})} = \widehat{\mu}_i - (\widehat{\beta}_1 + x_i \widehat{\beta}_2), \quad i \in \{1, ..., n\}.$$

We note that the effect of the WIND covariate seems to be non-linear on the mean of the PM2.5 response variable, which indicates that it is reasonable to quantify such an effect through a smooth function.
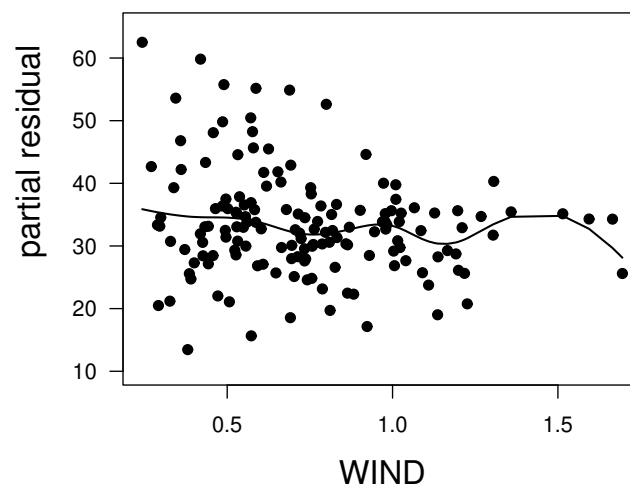
Figure 3: Plot of partial residuals versus WIND covariate with the estimated smooth function overlapped with data from Santiago, Chile.

Figure 4 contains the theoretical quantile versus empirical quantile (QQ) plot (a), histogram (b) and index plot (c) for the standardized residuals defined as

$$r_i^{(\text{s})} = \frac{y_i - \mu_i}{\sqrt{\widehat{\text{Var}}(Y_i)}} = \frac{\widehat{\phi}^{1/2}(y_i - \widehat{\mu}_i)}{\sqrt{\widehat{\mu}_i^2}}, \quad i \in \{1, ..., n\},$$

with ϕ and μ_i being stated in Subsections 2.2, for the model formulated in (2.5).

To verify the distributional assumption established in the model, we consider Figure 4(a). This figure does not show inusual features, so that the response variable seem to be well described by the RBS-SAM. In addition, the independence assumption is also verified. In Figure 4(b), we see a considerable symmetry of the standardized residuals. From Figure 4(c), we note that the standardized residuals take values in the interval $[-2.0, 2.0]$ mostly, except for some values that are outside the bands defined as two times the standard deviation for each residual.

For comparative purposes, we also consider a generalized additive model (GAM) [15, 16] assuming a Gaussian family with link function equal to identity to describe the data set of

pollution. The estimates of the parametric part of the model are $\widehat{\beta}_1 = 0.668$ and $\widehat{\beta}_2 = 0.428$, whose estimated SEs are 1.956 and 0.018, respectively, with a total of three degrees of freedom associated with this fit.
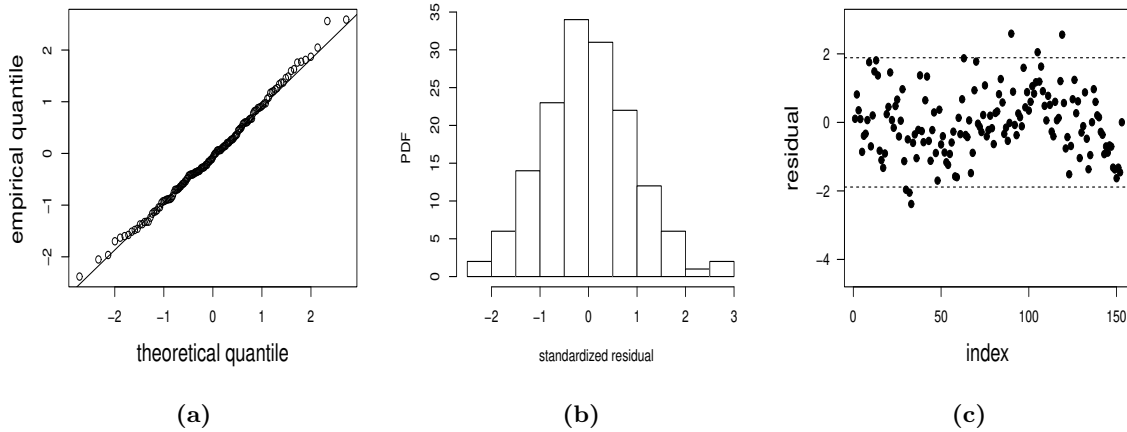
Figure 4: QQ plot (a), histogram (b), and index plot (c) for the standardized residuals of RBS-SAM with data from Santiago, Chile.

Figure 5 shows a residual analysis of the model. Observe the behavior of the tails of the distribution in Figure 5(a) and 5b. In addition, in Figure 5(c), note that there are several points outside the bands defined as twice the standard deviation of the residuals. The corresponding AIC of the fitted model is 1081.612. According to the AIC, the RBS-SAM presents a better fit, since the value of the AIC is 1032.003, with this value being less than the value obtained for the adjusted GAM.
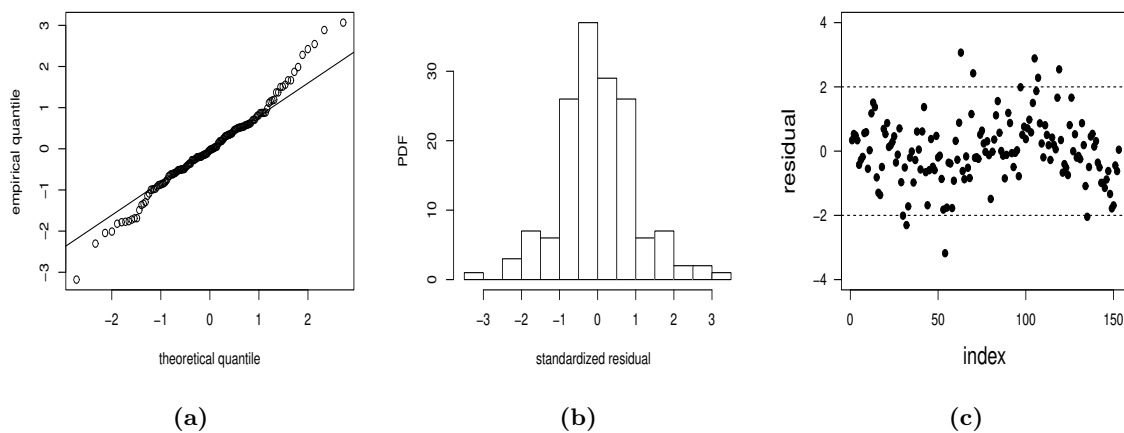
Figure 5: QQ-plot (a), histogram (b), and index plot (c) for the standardized residuals of GAM with data from Santiago, Chile.

5.5. Local influence analysis

Figures 6 (a)–(c) present the index plots of C_i for the case-weight scheme. Considering Figure 6 (a)–(b), we notice that cases #9 and #151 have the highest potential influence on $\widehat{\boldsymbol{\beta}}$ and $\widehat{\delta}$, whereas case #63 is more influential on $\widehat{\boldsymbol{f}}$; see Figure 6(c).

Figures 6 (d)–(f) display the index plot of C_i for the response perturbation scheme. We observe that case #151 has a small influence on $\widehat{\boldsymbol{\beta}}$ and $\widehat{\delta}$, whereas that cases #74 and #151 have the highest potential influence on $\widehat{\boldsymbol{f}}$.

Figures 6 (g)–(i) show the index plot of C_i for the precision perturbation scheme. We see that cases #9 and #151 have the highest potential influence on $\widehat{\boldsymbol{\beta}}$ and $\widehat{\delta}$, whereas case #63 is the more influential on $\widehat{\boldsymbol{f}}$; see Figure 6(i). Note that cases #9, #63, #74 and #151 correspond to 09 April, 02 June, 13 June, and 29 August 2019, respectively.

We now analyze how the MPL estimates change when the cases detected as potentially influential are removed. To perform this analysis, the set of case(s) {9}, {63}, {151}, {9, 63}, {9, 151}, {63, 151}, and {9, 63, 151} are removed and the estimates of the parameters are calculated again. Table 2 provide the relative changes (RC) in the parameter estimates, in their corresponding estimated SEs, and the associated *p*-value. These changes are calculated from $\text{RC}_{\widehat{\beta}_{j(i)}} = |(\widehat{\beta}_j - \widehat{\beta}_{j(i)})/\widehat{\beta}_j| \times 100\%$ and $\text{RC}_{\text{SE}(\widehat{\beta}_j)_{(i)}} = |(\text{SE}(\widehat{\beta}_j) - \text{SE}(\widehat{\beta}_j)_{(i)})/\text{SE}(\widehat{\beta}_j)| \times 100\%$, where $\widehat{\beta}_{j(i)}$ and $\text{SE}(\widehat{\beta}_j)_{(i)}$ denote the MPL estimates of β_j and their corresponding SEs, obtained after extracting case i, for $j \in \{1, 2\}$ and $i \in \{1, ..., 153\}$. Table 2 reports that the highest values of RCs are associated with β_1, in particular, for cases {9} and {9, 63}, corresponding to the most influential observations on the parametric, nonparametric and precision components. Note that the influential cases on the parametric component are not necessarily the same on the nonparametric component, as reported in [24]. In addition, note that the significance of the parameters at 5% does not change since the *p*-values remain below 0.01. Table 2 reports that the diagnostic measures derived here identify potentially influential cases, being them cases {9, 63}, which affect the inference of the model, but not their significance. These cases correspond to the dates: 09 April 2019 and 02 June 2019 of the CEM period. Note that, for case #9, a low PM2.5 content was recorded, close to the minimum, as well as for PM10, while the wind speed was high and close to the maximum recorded. Regarding case #63, this was the day with maximum PM2.5 content recorded throughout the CEC period, while on this day the wind speed recorded the minimum measurement. In addition, for case #151, the PM2.5 content registered its minimum value in the CEC period and the PM10 content was very close to the minimum value recorded, while the wind speed reached its maximum value. Considering the above results, the maximum or minimum PM2.5 content for cases #9, #63 and #151 are strongly related to wind speed, since on those days this speed reached extreme measurements. In summary, the diagnostic analytics based on the local influence method and residuals confirm that the RBS-SAM presented in Section 2 is suitable for modeling environmental data, even if there are outliers and potentially influential observations.
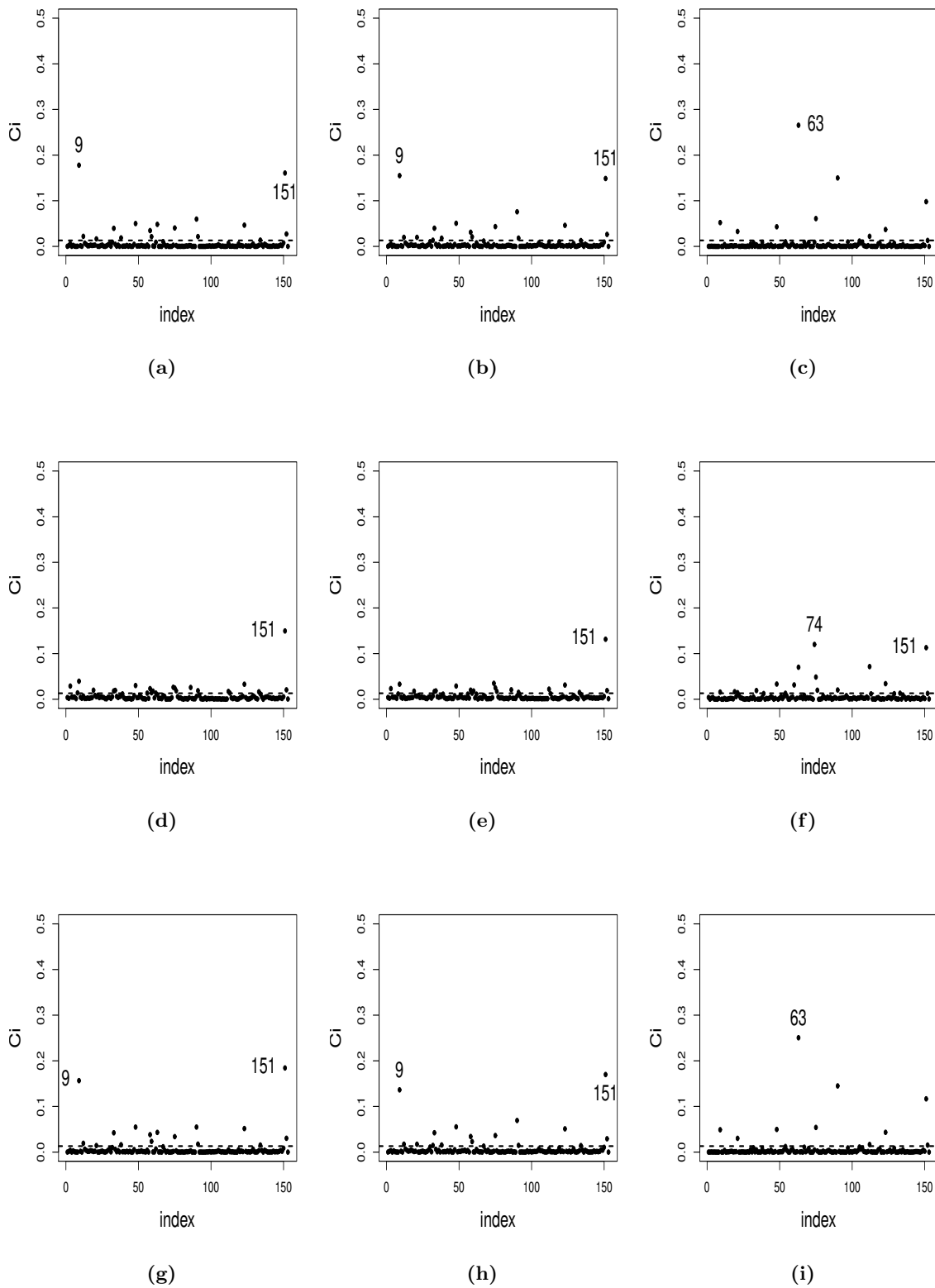
Figure 6: Index plots of C_i for $\boldsymbol{\beta}$ (a,d,g), δ (b,e,h) and $\boldsymbol{f}$ (c,f,i) under case-weight (a,b,c); response (d,e,f); and precision (g,h,i) perturbations with data from Santiago, Chile.

Table 2: RCs (in %) in MPL estimates and in the corresponding estimated SEs for the indicated removed case(s), and respective *p*-values using environmental data and the RBS-SAM with data from Santiago, Chile.

Removed case(s)	RCs in the estimate of	β_1	β_2	δ
None	Parameter	—	—	—
	SE	—	—	—
	(*p*-value)	< 0.01	< 0.01	—
{9}	Parameter	8.52	0.14	1.18
	SE	0.15	1.55	7.85
	(*p*-value)	< 0.01	< 0.01	—
{63}	Parameter	0.47	0.37	0.94
	SE	0.26	1.57	7.94
	(*p*-value)	< 0.01	< 0.01	—
{151}	Parameter	1.28	0.27	1.52
	SE	1.09	2.17	9.23
	(*p*-value)	< 0.01	< 0.01	—
{9, 63}	Parameter	7.29	0.41	2.14
	SE	0.21	1.67	7.75
	(*p*-value)	< 0.01	< 0.01	—
{9, 151}	Parameter	2.84	0.33	1.12
	SE	1.33	2.40	9.37
	(*p*-value)	< 0.01	< 0.01	—
{63, 151}	Parameter	0.92	0.84	2.49
	SE	1.03	2.29	9.13
	(*p*-value)	< 0.01	< 0.01	—
{9, 63, 151}	Parameter	2.13	0.90	2.10
	SE	1.25	2.52	9.27
	(*p*-value)	< 0.01	< 0.01	—

6. CONCLUDING REMARKS

In this study, we proposed and derived a reparametrized Birnbaum–Saunders semi-parametric additive model. This model allowed us to describe the mean of a random variable whose dispersion is non-constant through covariates. These covariates can contribute parametrically or nonparametrically in the model. Furthermore, it was possible to maintain the original scale of the data, since performing transformations on the modeled variable can reduce interpretability. We used a back-fitting algorithm to obtain the maximum penalized likelihood estimates by using the cubic smoothing splines.

To find potentially influential observations, we derived local influence techniques for the proposed model under case-weight, response variable and precision parameter perturbations. We apply the proposed model to real pollution data, verifying that the reparametrized Birnbaum–Saunders semi-parametric additive model is adequate to work with this type of data.

As future research, a package in R will be implemented so that various users have this model computationally to their disposal to be applied in practice. Currently, the R codes are available from the authors upon request. Also, the reparametrized Birnbaum–Saunders semi-parametric model can be extended to the reparametrized Birnbaum–Saunders semi-parametric case with varying precision, thin-plate spline, and partially varying-coefficient models. In addition, the use of functional, latent, spatial, and temporal structures, as well as reduction of dimensionality employing partial least squares regression, and mixture models that can facilitate the parameters estimation, are aspects to be considered in further studies with optimism because they have been explored in the Birnbaum–Saunders parametric case [12, 19, 25, 26, 35].

ACKNOWLEDGMENTS

REFERENCES

[1]   ARRUE, J.; ARELLANO, R.; GOMEZ, H.W. and LEIVA, V. (2020). On a new type of Birnbaum–Saunders models and its inference and application to fatigue data, *Journal of Applied Statistics*, **47**, 2690–2710.

[2]   BALAKRISHNAN, N.; GUPTA, R.; KUNDU, D.; LEIVA, V. and SANHUEZA, A. (2011). On some mixture models based on the Birnbaum–Saunders distribution and associated inference, *Journal of Statistical Planning and Inference*, **141**, 2175–2190.

[3]   BALAKRISHNAN, N. and KUNDU, D. (2019). Birnbaum–Saunders distribution: a review of models, analysis, and applications, *Applied Stochastic Models in Business and Industry*, **35**, 4–49.

[4]   BERHANE, K. and TIBSHIRANI, J. (1998). Generalized additive models for longitudinal data, *The Canadian Journal of Statistics*, **26**, 517–535.

[5]   BOURGUIGNON, M.; LEAO, J.; LEIVA, V. and SANTOS-NETO, M. (2017). The transmuted Birnbaum–Saunders distribution, *REVSTAT – Statistical Journal*, **15**, 601–628.

[6]   BUJA, A.; HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models, *Annals of Statistics*, **17**, 453–555.

[7]   CARRASCO, J.M.F.; FIGUEROA-ZÚÑIGA, J.; LEIVA, V.; RIQUELME, M. and AYKROYD, R.G. (2020). An errors-in-variables model based on the Birnbaum–Saunders and its diagnostics with an application to earthquake data. *Stochastic Environmental Research and Risk Assessment*, **34**, 369–380.

[8]   CAVIERES, M.F.; LEIVA, V.; MARCHANT, C. and ROJAS, F. (2021). A methodology for data-driven decision making in the monitoring of particulate matter environmental contamination in Santiago of Chile, *Reviews of Environmental Contamination and Toxicology*, **250**, 45–67.

[9]   COOK, R.D. (1986). Assessment of local influence (with discussion), *Journal of the Royal Statistical Society B*, **48**, 133–169.

[10] DASILVA, A.; DIAS, R.; LEIVA, V.; MARCHANT, C. and SAULO, H. (2020). Birnbaum–Saunders regression models: a comparative evaluation of three approaches, *Journal of Statistical Computation and Simulation*, **90**, 2552–2570.

[11] ESCOBAR, L. and MEEKER, W. (1992). Assessing influence in regression analysis with censored data, *Biometrics*, **48**, 507–28.

[12] GARCIA-PAPANI, F.; LEIVA, V.; URIBE-OPAZO, M.A. and AYKROYD, R.G. (2018). Birnbaum–Saunders spatial regression models: diagnostics and application to chemical data, *Chemometrics and Intelligent Laboratory Systems*, **17**, 114–128.

[13] GOMEZ-DENIZ, E.; LEIVA, V.; CALDERIN-OJEDA, E. and CHESNEAU, C. (2022). A novel claim size distribution based on a Birnbaum–Saunders and gamma mixture capturing extreme values in insurance: estimation, regression, and applications, *Computational and Applied Mathematics*, **41**, 171.

[14] GREEN, P. and SILVERMAN, B. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty*, Chapman and Hall, New York, USA.

[15] HASTIE, T. and TIBSHIRANI, R. (1986). Generalized additive models, *Statistical Science*, **1**, 293–310.

[16] HASTIE, T. and TIBSHIRANI, R. (1987). Generalized additive models: some applications, *Journal of the American Statistical Association*, **82**, 371–386.

[17] HASTIE, T. and TIBSHIRANI, R. (1990). *Generalized Additive Models*, Chapman and Hall, New York, USA.

[18] HASTIE, T. and TIBSHIRANI, R. (1993). Varying-coefficient models, *Journal of the Royal Statistical Society B*, **55**, 757–796.

[19] HUERTA, M.; LEIVA, V.; LIU, S.; RODRIGUEZ, M. and VILLEGAS, D. (2019). On a partial least squares regression model for asymmetric data with a chemical application in mining, *Chemometrics and Intelligent Laboratory Systems*, **190**, 55–68.

[20] IBACACHE-PULGAR, G.; PAULA, G.A. and CYSNEIROS, F.J.A. (2013). Semi-parametric additive models under symmetric distributions, *TEST*, **22**, 103–121.

[21] IBACACHE-PULGAR, G.; PAULA, G.A. and GALEA, M. (2012). Influence diagnostics for elliptical semi-parametric mixed models, *Statistical Modeling*, **12**, 165–193.

[22] IBACACHE-PULGAR, G. and REYES, S. (2018). Local influence for elliptical partially varying-coefficient model, *Statistical Modeling*, **18**, 149–174.

[23] IBACACHE-PULGAR, G.; FIGUEROA-ZÚÑIGA, J. and MARCHANT, C. (2021). Semi-parametric additive beta regression models: inference and local influence diagnostics, *REVSTAT – Statistical Journal*, **19**, 255–274.

[24] KIM, C.; PARK, B.U. and KIM, W. (2002). Influence diagnostics in semi-parametric regression models, *Statistics and Probability Letters*, **60**, 49–58.

[25] KOTZ, S.; LEIVA, V. and SANHUEZA, A. (2010). Two new mixture models related to the inverse Gaussian distribution, *Methodology and Computing in Applied Probability*, **12**, 199–212.

[26] LEAO, J.; LEIVA, V.; SAULO, H. and TOMAZELLA, V. (2018). Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data, *Statistics in Medicine*, **37**, 4421–4440.

[27] LEIVA, V.; SANTOS-NETO, M.; CYSNEIROS, F.J.A. and BARROS, M. (2014). Birnbaum–Saunders statistical modeling: a new approach, *Statistical Modeling*, **14**, 21–48.

[28] LEIVA, V.; MARCHANT, C.; RUGGERI, F. and SAULO, H. (2015). A criterion for environmental assessment using Birnbaum–Saunders attribute control, *Environmetrics*, **26**, 463–476.

[29] LEIVA, V.; FERREIRA, M.; GOMES, M.I. and LILLO, C. (2016). Extreme value Birnbaum–Saunders regression models applied to environmental data, *Stochastic Environmental Research and Risk Assessment*, **30**, 1045–1058.

[30] LEIVA, V.; CASTRO, C.; VILA, R. and SAULO, H. (2024). Unveiling patterns and trends in research on cumulative damage models for statistical and reliability analyses: Bibliometric and thematic explorations with data analytics, *Chilean Journal of Statistics*, **15**, 81–109.

[31] MARCHANT, C.; LEIVA, V.; CAVIERES, M. and SANHUEZA, A. (2013). Air contaminant statistical distributions with application to PM10 in Santiago, Chile, *Reviews of Environmental Contamination and Toxicology*, **223**, 1–31.

[32] MARCHANT, C.; LEIVA, V. and CYSNEIROS, F.J.A. (2016). A multivariate log-linear model for Birnbaum–Saunders distributions, *IEEE Transactions on Reliability*, **65**, 816–827.

[33] MARCHANT, C.; LEIVA, V.; CYSNEIROS, F.J.A. and LIU, S. (2018). Robust multivariate control charts based on Birnbaum–Saunders distributions, *Journal of Statistical Computation and Simulation*, **88**, 182–202.

[34] MARCHANT, C.; LEIVA, V.; CHRISTAKOS, G. and CAVIERES, M.A. (2019). Monitoring urban environmental pollution by bivariate control charts: new methodology and case study in Santiago, Chile, *Environmetrics*, **30**, e2551.

[35] MARTINEZ, S.; GIRALDO, R. and LEIVA, V. (2019). Birnbaum–Saunders functional regression models for spatial data, *Stochastic Environmental Research and Risk Assessment*, **33**, 1765–1780.

[36] MAZUCHELI, J.; MENEZES, A.F.B. and DEY, S. (2018). The unit-Birnbaum–Saunders distribution with applications, *Chilean Journal of Statistics*, **9**(1), 47–57.

[37] MAZUCHELI, M.; LEIVA, V.; ALVES, B. and MENEZES, A.F.B. (2021). A new quantile regression for modeling bounded data under a unit Birnbaum–Saunders distribution with applications in medicine and politics, *Symmetry*, **13**, 682.

[38] MAZUCHELI, M.; ALVES, B.; MENEZES, A.F.B. and LEIVA, V. (2022). An overview on parametric quantile regression models and their computational implementation with applications to biomedical problems including COVID-19 data, *Computer Methods and Programs in Biomedicine*, **221**, 106816.

[39] MORALES, R.G.E.; LLANOS, A.; MERINO, M. and GONZALEZ-ROJAS, C.H. (2012). A semi-empirical method of PM-10 atmospheric pollution forecast at Santiago de Chile city, *Nature, Environment and Pollution Technology*, **11**, 181–186.

[40] PUENTES, R.; MARCHANT, C.; LEIVA, V.; FIGUEROA-ZÚÑIGA, J. and RUGGERI, F. (2021). Predicting PM2.5 and PM10 levels during critical episodes management in Santiago, Chile, with a bivariate Birnbaum–Saunders log-linear model, *Mathematics*, **9**, 645.

[41] R DEVELOPMENT CORE TEAM (2021). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

[42] RIECK, J.R. and NEDELMAN, J.R. (1991). A log–linear model for the Birnbaum–Saunders distribution, *Technometrics*, **33**, 51–60.

[43] REYES, J.; ARRUE, J.; LEIVA, V. and MARTIN-BARREIRO, C. (2021). A new Birnbaum–Saunders distribution and its mathematical features applied to bimodal real-world data from environment and medicine, *Mathematics*, **9**, 1891.

[44] RIGBY, R.A. and STASINOPOULOS, D.M. (2005). Generalized additive models for location, scale and shape, *Applied Statistics*, **54**, 507–554.

[45] Sanchez, L.; Leiva, V.; Galea, M. and Saulo, H. (2021). Birnbaum–Saunders quantile regression and its diagnostics with application to economic data, *Applied Stochastic Models in Business and Industry*, **3**, 53–73.

[46] Santos–Neto, M.; Cysneiros, F.J.A.; Leiva, V. and Ahmed, S.E. (2012). On new parametrizations of the Birnbaum–Saunders distribution, *Pakistan Journal of Statistics*, **1**, 1–26.

[47] Santos-Neto, M.; Cysneiros, F.J.A.; Leiva, V. and Barros, M. (2014). On a reparameterized Birnbaum–Saunders distribution and its moments, estimation and applications, *REVSTAT – Statistical Journal*, **12**, 247–272.

[48] Saulo, H.; Souza, R.; Vila, R.; Leiva, V. and Aykroyd, R.G. (2021). Modeling mortality based on pollution and temperature using a new Birnbaum–Saunders autoregressive moving average structure with regressors and related-sensors data, *Sensors*, **21**, 6518.

[49] Saulo, H.; Leao, J.; Leiva, V. and Aykroyd, R.G. (2019). Birnbaum–Saunders autoregressive conditional duration models applied to high-frequency financial data, *Statistical Papers*, **60**, 1605–1629.

[50] Segal, M.R. and Bacchetti, P. (1994). Variances for maximum penalized likelihood estimates obtained via the EM algorithm, *Journal of the Royal Statistical Society B*, **56**, 345–352.

[51] Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline, *Journal of the Royal Statistical Society B*, **45**, 133–150.

Estimation and Prediction for the Half-Normal Distribution Based on Progressively Type-II Censored Samples

Authors: SÜMEYRA SERT [ID] [✉]
– Department of Statistics, Selçuk University,
Konya, Turkey
sumeyra.sert@selcuk.edu.tr

İHAB A.S. ABUSAIF [ID]
– Department of Statistics, Selçuk University,
Konya, Turkey
censtat@gmail.com

ERTAN AKGENÇ [ID]
– Department of Statistics, Selçuk University,
Konya, Turkey
statistician.ertan@gmail.com

KADIR KARAKAYA [ID]
– Department of Statistics, Selçuk University,
Konya, Turkey
kkarakaya@selcuk.edu.tr

COŞKUN KUŞ [ID]
– Department of Statistics, Selçuk University,
Konya, Turkey
coskun@selcuk.edu.tr

Abstract:

• In this paper, estimation and prediction problems are discussed for the half-normal distribution under a progressively Type-II censoring scheme. This study focuses on two statistical inferential problems. In the first part of the study, several point estimators and confidence intervals are obtained for the scale parameter of the half-normal distribution. In the second part, several predictors and predictive intervals are derived for the removed failure times. A Monte Carlo simulation study is performed to discuss the mean squared error (mean squared prediction errors) and bias of estimates (predictors). The coverage probabilities and average length of the confidence and predictive intervals are simulated and a numerical example is provided.

✉ Corresponding author

1.   INTRODUCTION

In the statistical literature, numerous distributions exist with two or more parameters. However, multi-parameter distributions can have problems with estimation and prediction due to non-identification. Therefore, in practice it is sometimes convenient to work with one-parameter distributions. One of the most popular single-parameter distributions is the half-normal (HN) distribution. In a recent study, Huang and Roth [10] demonstrated that the HN distribution is not only used for lifetime data but also in pragmatic randomized trials proving the convenience of the HN distribution for modeling different real data sets. The probability density function (pdf) and cumulative distribution function (cdf) of the HN distribution are given, respectively, by

$$f(x;\theta) = \frac{2}{\Gamma(\frac{1}{2})\theta} \exp\left\{-\left(\frac{x}{\theta}\right)^2\right\}, \ x > 0,$$

and

(1.1)
$$F(x;\theta) = 1 - \frac{\Gamma\left(\frac{1}{2}, \left(\frac{x}{\theta}\right)^2\right)}{\Gamma\left(\frac{1}{2}\right)},$$

where $\theta > 0$ is a scale parameter and $\Gamma(\alpha, z)$ is the upper incomplete gamma function defined by

$$\Gamma(\alpha, z) = \int_z^\infty e^{-y} y^{\alpha-1} dy, \ \alpha, z > 0.$$

The distribution given with a cdf (1.1) will be denoted by $\text{HN}(\theta)$ for the remainder of this study. Shanker *et al.* [23] derived some statistical properties of the HN distribution (they called it Quasi-Exponential), such as moments, hazard and the hazard rate function, survival function and mean residual function. The maximum likelihood estimator (MLE) of the scale parameter is also studied. They examined the real data modeling capability of the HN distribution using lifetime data from biomedical science.

In reliability applications, the data is generally collected under some censoring schemes when the lifetime of the products is too long. One of the most popular schemes is progressive censoring. It should be pointed out that progressive censoring is not only used for reliability applications but also quite common in clinical trials due to the staggered entry. We refer readers to [15], [16], [20] and [21] for progressive censoring with staggered entry. In this paper, we consider the point and interval estimation (prediction) of the HN distribution under the progressive censoring scheme. A progressively Type-II censoring scheme is well-discussed by Balakrishnan and Aggarwala [4]. Progressively Type-II censored samples can be explained as follows: Let n units are put on a life test. When the first failure is observed, randomly selected r_1 of the $n - 1$ surviving units are removed (withdrawn or censored) from the test. When the second failure occurs, randomly selected r_2 of the $n - 2 - r_1$ surviving units are removed from the test and this process is repeated. At the time of the m-th failure, the remaining $r_m = n - m - r_1 - \cdots - r_{m-1}$ surviving units are removed randomly from the test. It can be easily observed that $n = m + \sum_{i=1}^{m} r_i$. The progressively Type-II censored (PCII) failure times are denoted by $X_{1:m:n}^{\mathbf{r}} < X_{2:m:n}^{\mathbf{r}} < \cdots < X_{m:m:n}^{\mathbf{r}}$.

There are several studies discussing the estimation and prediction problem based on the PCII sample. Seo and Kang [22] discussed the problem of point and interval estimation for the scaled half-logistic distribution and proposed a method to estimate the scale parameter using the pivotal quantity method under PCII samples. They also tackled the problem of estimation and prediction for the two-parameter half-logistic distribution. Ma and Gui [17] used the pivotal quantity method to derive the estimator for the inverse Rayleigh distribution based on general PCII samples. They also derived an explicit estimator of the scale parameter by the approximation of the likelihood equation using Taylor expansion. Khan [11] studied on the predictive inference for the HN distribution under the Type II censoring scheme. In a more recent study, Sindhu and Hussain [24] used Bayesian methods and made predictive inference on the HN distribution for left-censored data. Asgharzadeh and Valiollahi [3] studied prediction intervals for the PCII from proportional hazard rate models. El-Din and Shafay [18] derived one-sample and two-sample Bayesian prediction intervals based on PCII using Exponential, Pareto, Weibull and Burr Type X-II models. Dey *et al.* [5] discussed the parameter estimation problem for generalized inverted exponential distribution under PCII. The studies conducted by Wang *et al.* [26], Hemmati [9] and Kinaci *et al.* [12] are also examples of studies on deriving exact confidence intervals under PCII. Recently, Ahmadi *et al.* [2] studied statistical inference for the two-parameter generalized half-normal distribution based on a PCII sample.

In this study, the point estimation, interval estimation and prediction intervals are discussed for the HN model under PCII. This paper is organized as follows: In Section 2, the maximum likelihood and an approximate maximum likelihood estimation are discussed. In Section 3, pivotal type estimation is studied with an approximate version. Interval estimation is also discussed through MLE, likelihood ratio statistic and a pivotal quantity in Section 4. In Section 5, the prediction of the removed failure times is discussed. The predictive intervals are derived in Section 6. In Section 7, a simulation study is performed to observe the behavior of the point and interval estimates. A simulation study is also conducted to compare the predictors and predictive intervals. In Section 8, a numerical example is presented for illustration. The concluding remarks are given in Section 9.

2. MLE AND APPROXIMATE MLE ESTIMATION

Let $X^{\mathbf{r}}_{1:m:n} < X^{\mathbf{r}}_{2:m:n} < \cdots < X^{\mathbf{r}}_{m:m:n}$ be the progressively censored order statistics from HN(θ). Then the log-likelihood function can be written by

$$(2.1) \qquad \ell(\theta) \propto -m \log(\theta) - \sum_{i=1}^{m} \left(\frac{x_i}{\theta}\right)^2 + \sum_{i=1}^{m} r_i \log\left(\Gamma\left(\frac{1}{2}, \left(\frac{x_i}{\theta}\right)^2\right)\right),$$

where x_i is a realization of $X^{\mathbf{r}}_{i:m:n}$ for $i = 1, 2, ..., m$. The log-likelihood function (2.1) is non-linear in parameter θ and MLE can not be obtained, explicitly. Therefore, nonlinear optimization methods such as Nelder-Mead or BFGS should be applied to get the MLE of the scale parameter θ. Initial point selection is an important problem in nonlinear optimization methods. An arbitrary initial point may lead us to misinterpretation. Therefore, the analytically obtained approximate MLE (AMLE) estimator, which does not require a searching

method, will be discussed below. Let us consider a first-order likelihood equation

$$(2.2) \qquad \frac{d\ell(\theta)}{d\theta} = -\frac{m}{\theta} + \frac{2}{\theta^3} \sum_{i=1}^{m} x_i^2 + \frac{2}{\theta^2} \sum_{i=1}^{m} \frac{r_i x_i \exp\left(\left(\frac{x_i}{\theta}\right)^2\right)}{\Gamma\left(\frac{1}{2}, \left(\frac{x_i}{\theta}\right)^2\right)} = 0.$$

We consider the random variable $Z = X/\theta$, it is easy to know that Z has the standard HN distribution (with $\theta=1$) since the θ is a scale parameter. After some algebra, the Eq (2.2) can be re-written by

$$(2.3) \qquad -\frac{m}{\theta} + \frac{2}{\theta^3} \sum_{i=1}^{m} x_i^2 + \frac{2}{\theta^2} \sum_{i=1}^{m} r_i x_i \frac{\exp\left(-z_i^2\right)}{\Gamma\left(\frac{1}{2}, z_i^2\right)} = 0,$$

where z_i is a realization of $Z_{i:m:n}^{\mathbf{r}} = X_{i:m:n}^{\mathbf{r}}/\theta$ which is progressively censored order statistic from standard HN distribution for $i = 1, 2, ..., m$.

Since equation (2.3) can not be solved analytically, we approximate the tricky part $\frac{\exp\left(-z_i^2\right)}{\Gamma\left(\frac{1}{2}, z_i^2\right)}$ by expanding it in Taylor series around $v_i = E(Z_{i:m:n}^{\mathbf{r}})$. By the way, we can write

$$(2.4) \qquad G(Z_{i:m:n}^{\mathbf{r}}) = U_{i:m:n}^{\mathbf{r}}$$

by using the probability integral transformation, where $Z_{i:m:n}^{\mathbf{r}}$ is the i-th the progressively censored order statistic from standard HN distribution with cdf

$$(2.5) \qquad G(z; \theta) = 1 - \frac{\Gamma\left(\frac{1}{2}, z^2\right)}{\Gamma\left(\frac{1}{2}\right)}, z > 0,$$

and $U_{i:m:n}^{\mathbf{r}}$ is the standard uniform progressively censored order statistic.

According to Balakrishnan and Aggarwala [4], and using transformation (2.4) one can write

$$(2.6) \qquad v_i = E(Z_{i:m:n}^{\mathbf{r}}) \approx G^{-1}(p_i),$$

where

$$(2.7) \qquad p_i = E(U_{i:m:n}^{\mathbf{r}}) = 1 - \prod_{j=m-i+1}^{m} \frac{j + r_{m-j+1} + \cdots + r_m}{j + 1 + r_{m-j+1} + \cdots + r_m}.$$

Using equations (2.5)–(2.7), v_i can be determined by solving the equation

$$(2.8) \qquad \Gamma\left(\frac{1}{2}, v_i^2\right) = \Gamma\left(\frac{1}{2}\right)(1 - p_i), \ i = 1, 2, ..., m.$$

Let us turn back to (2.3) and we now focus on the Taylor expansion on the part

$$H(z_i) = \frac{\exp\left(-z_i^2\right)}{\Gamma\left(\frac{1}{2}, z_i^2\right)}.$$

Let h be the first-order derivative of the function H which is given by

$$h(z_i) = -\frac{2z_i \exp\left(-z_i^2\right)}{\Gamma\left(\frac{1}{2}, z_i^2\right)} + \frac{\exp\left(-z_i^2\right)\exp(-z_i)}{\Gamma^2\left(\frac{1}{2}, z_i^2\right)\sqrt{z_i}}.$$

Then

$$H(z_i) \approx H(v_i) + (z_i - v_i)h(v_i)$$

(2.9)
$$= A_i + B_i z_i,$$

where

(2.10)
$$A_i = H(v_i) - v_i h(v_i)$$
$$= \frac{\Gamma\left(\frac{1}{2}, v_i^2\right)\left(1 - 2v_i^2\right)\exp\left(v_i^2\right) - 2v_i}{\Gamma^2\left(\frac{1}{2}, v_i^2\right)}$$

and

(2.11)
$$B_i = h(v_i)$$
$$= \frac{2v_i \exp\left(v_i^2\right)}{\Gamma\left(\frac{1}{2}, v_i^2\right)} + \frac{2}{\Gamma^2\left(\frac{1}{2}, v_i^2\right)}.$$

Eventually, using (2.9) in (2.3), we can reach to the approximate likelihood equation

$$-\frac{m}{\theta} + \frac{2}{\theta^3}\sum_{i=1}^{m} x_i^2 + \frac{2}{\theta^2}\sum_{i=1}^{m} r_i x_i\left(A_i + B_i\frac{x_i}{\theta}\right) = 0.$$

After some algebra, AMLE of the parameter θ can be obtained by

(2.12)
$$\widetilde{\theta} = \frac{c + \sqrt{c^2 + 4mb + 4md}}{2m},$$

where

$$b = 2\sum_{i=1}^{m} (X_{i:m:n}^{\mathbf{r}})^2,$$

$$c = 2\sum_{i=1}^{m} r_i X_{i:m:n}^{\mathbf{r}} A_i,$$

and

$$d = 2\sum_{i=1}^{m} r_i (X_{i:m:n}^{\mathbf{r}})^2 B_i.$$

It is noticed that the estimate (2.12) is not novel but it is another form of the MMLE given in Ahmadi *et al.* [2]. It is important here that the following revision brings us to the original ML estimate of θ without any searching methods.

Revised estimates: Following a suggestion by Lee *et al.* [13], we now calculate A_i and B_i in (2.10)–(2.11) with replacing v_i by

(2.13)
$$v_i = \frac{x_{i,}}{\widetilde{\theta}}, \ 1 \leq i \leq m.$$

and calculate the revised estimate $\widetilde{\theta}_{\text{revised}}$ by using (2.12). This process should be repeated a few times until the coefficients stabilize sufficiently enough. The flowchart is given in Figure 1 to illustrate the revising process.
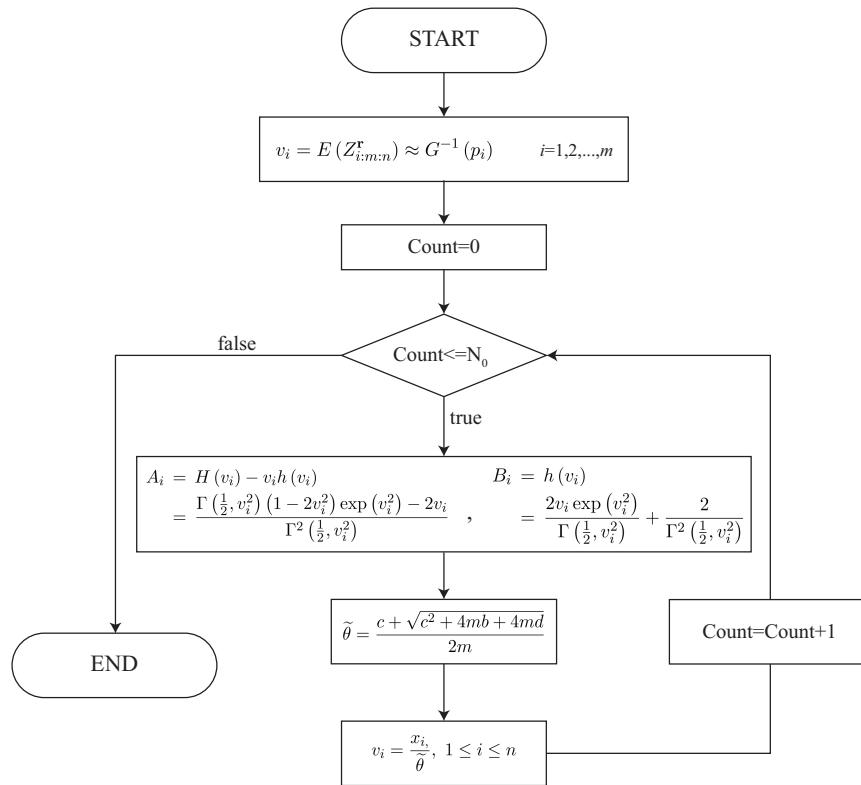
Figure 1: Flow chart for revised estimates.

3. A PIVOTAL QUANTITY ESTIMATION

In the previous section, AMLE is obtained for the scale parameter θ. In this section, the pivotal quantity type inference is discussed. This method is adopted from the results in Ma and Gui [17]. Let $X^{\mathbf{r}}_{1:m:n} < X^{\mathbf{r}}_{2:m:n} < \cdots < X^{\mathbf{r}}_{m:m:n}$ be a PCII sample from the HN(θ). Let

$$Y^{\mathbf{r}}_{i:m:n} = -\log\left(\frac{\Gamma\left(\frac{1}{2}, \left(\frac{X^{\mathbf{r}}_{i:m:n}}{\theta}\right)^2\right)}{\Gamma\left(\frac{1}{2}\right)}\right), \ i = 1, 2, ..., m.$$

It can be easily seen that, $Y^{\mathbf{r}}_{1:m:n} < Y^{\mathbf{r}}_{2:m:n} < \cdots < Y^{\mathbf{r}}_{m:m:n}$ are PCII samples from a standard exponential distribution. Let us consider the following transformations:

$$S_1 = nY^{\mathbf{r}}_{1:m:n},$$

$$S_i = \left[n - \sum_{j=1}^{i-1}(r_j + 1)\right]\left(Y^{\mathbf{r}}_{i:m:n} - Y^{\mathbf{r}}_{i-1:m:n}\right), \ i = 2, ..., m.$$

According to Thomas and Wilson [25], $S_1, S_2, ..., S_m$ are also independent and identically distributed from a standard exponential distribution. It is well-known that the pivotal quantity

$$W(\theta) = 2\sum_{i=1}^{m} S_i$$

$$= 2\sum_{i=1}^{m}(r_i+1)Y_{i:m:n}^{\mathbf{r}}$$

(3.1)
$$= -2\sum_{i=1}^{m}(r_i+1)\log\left(\frac{\Gamma\left(\frac{1}{2}, \left(\frac{X_{i:m:n}^{\mathbf{r}}}{\theta}\right)^2\right)}{\Gamma\left(\frac{1}{2}\right)}\right)$$

is distributed χ^2 distribution with $2m$ degrees of freedom.

Ma and Gui [17] pointed out that the pivotal quantity $W(\theta)/(2m+2)$ converges to one in probability as $m \to \infty$. In this case, the pivotal type estimate θ^* of θ can be proposed by solving the equation

(3.2)
$$\sum_{i=1}^{m}(r_i+1)\left[-\log\left(\frac{\Gamma\left(\frac{1}{2}, \left(\frac{X_{i:m:n}}{\theta}\right)^2\right)}{\Gamma\left(\frac{1}{2}\right)}\right)\right] = m+1.$$

However, (3.2) does not allow for an explicit solution for θ. Since one-dimensional searching method can be used to get the pivotal type estimate of θ, the approximation method discussed in the previous section can also be applied to solve the (3.2). Let us start the re-write (3.2) by

(3.3)
$$\sum_{i=1}^{m}(r_i+1)\left\{-\log\left(\frac{\Gamma\left(\frac{1}{2}, z_i^2\right)}{\Gamma\left(\frac{1}{2}\right)}\right)\right\} = m+1,$$

where z_i is a realization of $Z_{i:m:n}^{\mathbf{r}} = X_{i:m:n}^{\mathbf{r}}/\theta$ which is standard progressively censored order statistic from HN (1) for $i = 1, 2, ..., m$.

We expand the tricky part

$$K(z_i) = -\log\left(\frac{\Gamma\left(\frac{1}{2}, z_i^2\right)}{\Gamma\left(\frac{1}{2}\right)}\right)$$

around the point $v_i = E(Z_{i:m:n}^{\mathbf{r}})$ which is already defined in (2.8). Let k denotes the first-order derivative of K and it is given by

$$k(z_i) = \frac{2\exp\left(-z_i^2\right)}{\Gamma\left(\frac{1}{2}, z_i^2\right)}.$$

Hence, we can write

$$K(z_i) \approx K(v_i) + (z_i - v_i)k(v_i)$$
$$= C_i + D_i z_i, \ i = 1, 2, ..., m,$$

where

(3.4)
$$C_i = K(v_i) - v_i k(v_i)$$
$$= -\log\left(\frac{\Gamma\left(\frac{1}{2}, v_i^2\right)}{\Gamma\left(\frac{1}{2}\right)}\right) - \frac{2v_i\exp\left(-v_i^2\right)}{\Gamma\left(\frac{1}{2}, v_i^2\right)},$$

and

$$(3.5) \qquad \begin{aligned} D_i &= k(v_i) \\ &= \frac{2\exp(-v_i^2)}{\Gamma\left(\frac{1}{2}, v_i^2\right)}. \end{aligned}$$

Hence, the left-hand side of (3.3) can be approximated by

$$\sum_{i=1}^{m}(r_i+1)\left(C_i + D_i\frac{X_{i:m:n}^{\mathbf{r}}}{\theta}\right) = m+1$$

and the approximate pivotal quantity type estimate is obtained by

$$(3.6) \qquad \theta^{*(a)} = \frac{\displaystyle\sum_{i=1}^{m}(r_i+1)X_{i:m:n}^{\mathbf{r}}D_i}{m+1-\displaystyle\sum_{i=1}^{m}(r_i+1)C_i}.$$

Revised estimates: We now use the method proposed by Lee *et al.* [13], and calculate C_i and D_i in the (3.4)–(3.5) by replacing v_i by

$$v_i = \frac{x_{i,}}{\theta^*}, \ \ 1 \le i \le m.$$

and calculate the revised estimate $\theta^{*(a)}_{\text{revised}}$ by using (3.6). This process should also be repeated a few times until the coefficients stabilize sufficiently enough.

---

4. INTERVAL ESTIMATIONS

In this section, we discuss the confidence interval estimation of the parameter θ based on progressively censored data $X_{1:m:n}^{\mathbf{r}} < X_{2:m:n}^{\mathbf{r}} < \cdots < X_{m:m:n}^{\mathbf{r}}$ from the HN(θ). In the ML theory, it is well-known that

$$\hat{\theta} \approx \text{AN}\left(\theta, I^{-1}(\theta)\right),$$

where

$$I(\theta) = -E\left(\frac{d^2}{d\theta^2}\ell(\theta)\right)$$

is the Fisher Information. It can be estimated by

$$I\left(\hat{\theta}\right) = -\left(\frac{d^2}{d\theta^2}l(\theta)\right)\Big|_{\hat{\theta}}$$

and standard error of $\hat{\theta}$ is estimated by $\text{SE}\left(\hat{\theta}\right) = \sqrt{I^{-1}\left(\hat{\theta}\right)}$. Then we can write an approximate $(1-\alpha)100\%$ confidence interval for θ as follows:

$$(4.1) \qquad \left(\hat{\theta} - z_{1-\frac{\alpha}{2}}\,\text{SE}\left(\hat{\theta}\right), \hat{\theta} + z_{1-\frac{\alpha}{2}}\,\text{SE}\left(\hat{\theta}\right)\right),$$

where z_a is the a-th quantile of the standard normal distribution.

Let us define

$$W(\theta) = -2\sum_{i=1}^{m}(r_i + 1)Q(\theta), \ \theta > 0,$$

where

$$Q(\theta) = \log\left(\frac{\Gamma\left(\frac{1}{2}, \left(\frac{X_{i:m:n}^{\mathbf{r}}}{\theta}\right)^2\right)}{\Gamma\left(\frac{1}{2}\right)}\right).$$

It is well-known that the pivot $W(\theta)$ is distributed as χ^2 with $2m$ degrees of freedom. The following lemma is given before introducing a new confidence interval (CI) for the parameter θ.

Lemma 4.1. *Suppose that $0 < a_1 < a_2 < \cdots < a_m < \infty$. Then, $W(\theta)$ is strictly decreasing in θ for any $\theta > 0$. Furthermore, if $t > 0$, the equation $W(\theta) = t$ has a unique solution for any $\theta > 0$.*

Proof: Let us consider the first-order derivative of $Q(\theta)$ in θ which is given by

$$\frac{dQ(\theta)}{d\theta} = -\frac{2\exp(-\frac{a_i^2}{\theta^2})x}{\sqrt{\pi}\theta^2\left(\mathrm{erf}(x)\left(\frac{a_i}{\theta}\right) - 1\right)},$$

where $\mathrm{erf}(\cdot)$ is a well-known error function and it is defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp\left(-t^2\right)dt.$$

Since the $\mathrm{erf}(x)(x) \in [0,1)$, it is observed that $\frac{dQ(\theta)}{d\theta} > 0$. This indicates that $\frac{dW(\theta)}{d\theta} < 0$ and $W(\theta)$ is decreasing function in θ. Furthermore, $\lim_{\theta\to 0}W(\theta) = \infty$ and $\lim_{\theta\to\infty}W(\theta) = 0$. Thus, if $t > 0$, $W(\theta) = t$ has a unique solution for any $\theta > 0$. □

Let $\chi^2_{(a)2m}$ denotes the a-th quantile of the χ^2 distribution with $2m$ degrees of freedom. The following theorem gives an exact CI for parameter θ.

Theorem 4.1. *A $(1-\alpha)100\%$ exact CI for θ is constructed by*

(4.2)
$$\left(W^{-1}\left(\chi^2_{(1-\alpha/2)2m}\right), W^{-1}\left(\chi^2_{(\alpha/2)2m}\right)\right),$$

where $W^{-1}(t)$ is the solution of equation $W(\theta) = t$.

Proof: The proof follows from Lemma 4.1 and the fact $W(\theta) \sim \chi^2_{(2m)}$ □

Corollary 4.1. *An approximately $(1-\alpha)100\%$ CI for θ is constructed by*

(4.3)
$$\left(\frac{\sum_{i=1}^{m}(r_i+1)D_iX_{i:m:n}^{\mathbf{r}}}{\frac{\chi^2_{(1-\alpha/2)2m}}{2} - \sum_{i=1}^{m}(r_i+1)C_i}, \frac{\sum_{i=1}^{m}(r_i+1)D_iX_{i:m:n}^{\mathbf{r}}}{\frac{\chi^2_{(\alpha/2)2m}}{2} - \sum_{i=1}^{m}(r_i+1)C_i}\right),$$

where C_i and D_i are defined as in (3.4) and (3.5) respectively.

Proof: The proof is analogous to that of Theorem 4.1 in Ma and Gui [17]. □

By the way, there is another method called the uncorrected likelihood ratio (ULR) interval, which has desirable properties. The ULR intervals are discussed in Doganaksoy and Schmee [6] and Doganaksoy [7]. The ULR interval can be described as follows: Under some mild regularity conditions, if the θ is the true parameter, then the likelihood ratio statistic $\Lambda = -2\left(\ell(\theta) - \ell\left(\hat{\theta}\right)\right)$ is distributed as χ^2 with degrees of freedom 1, where ℓ is the log-likelihood function as in (2.1) and $\hat{\theta}$ is the MLE of θ. The test statistic Λ can be used for testing $H_0 : \theta = \theta_0$ against $H_0 : \theta \neq \theta_0$ with critical region $\Lambda > \chi^2_{(1)}(1 - \alpha)$. Then, we conclude that the ULR confidence interval for θ readily arises a nice set $\left\{\theta : \Lambda < \chi^2_{(1)}(\alpha)\right\}$. Using this fact, $100(1 - \alpha)\%$ ULR CI limits

(4.4)
$$(\theta_L, \theta_U)$$

that satisfy

(4.5)
$$-2\left(\ell(\theta) - \ell\left(\hat{\theta}\right)\right) - \chi^2_{(1)}(1 - \alpha) = 0$$

with $\theta_L < \hat{\theta}$ and $\theta_U > \hat{\theta}$.

It is noticed by Fraser [8] that the ULR and asymptotically normal (AN) CIs are asymptotically equivalent. The ULR CIs are transformation invariant, unlike the AN method. Furthermore, the ULR CIs always produce limits inside of the parameter space.

In the following section, the prediction problem is discussed for the removed failure times within the PCII scheme.

5.   PREDICTION

Let $X^{\mathbf{r}}_{1:m:n} < X^{\mathbf{r}}_{2:m:n} < \cdots < X^{\mathbf{r}}_{m:m:n}$ be progressively censored sample from the $\mathrm{HN}(\theta)$ distribution and $Y = Y_{j:r_k}$ denotes the j-th order statistics related to the removed sample of size r_k at the progressive stage k. Using the theorem in Ng *et al.* [19], the conditional pdf of $Y|X^{\mathbf{r}}_{k:m:n}$ can be written by

$$f_{Y|X^{\mathbf{r}}_{k:m:n}}(y|x_k) = \frac{r_k!}{(j-1)!(r_k-j)!} f(y)(F(y) - F(x_k))^{j-1}$$
$$\times (1 - F(y))^{r_k - j}(1 - F(x_k))^{-r_k} \quad (y > x_k),$$

where y and x_k are the realizations of Y and $X^{\mathbf{r}}_{j:m:n}$. Then, the predictive log-likelihood function is given by

$$\ell(y, \theta) \propto \ell(\theta) - \log(\theta) - \left(\frac{y}{\theta}\right)^2$$
$$+ (j-1)\log\left\{\Gamma\left(\frac{1}{2}, \left(\frac{x_k}{\theta}\right)^2\right) - \Gamma\left(\frac{1}{2}, \left(\frac{y}{\theta}\right)^2\right)\right\}$$
$$+ (r_k - j)\log\left\{\Gamma\left(\frac{1}{2}, \left(\frac{y}{\theta}\right)^2\right)\right\}$$

(5.1)
$$- r_k \log\left\{\Gamma\left(\frac{1}{2}, \left(\frac{x_k}{\theta}\right)^2\right)\right\}.$$

The predictive log-likelihood (5.1) is non-linear in y and parameter θ, and they can not be obtained explicitly. Therefore nonlinear optimization methods such as Nelder-Mead or BFGS should be applied to get the maximum likelihood predictor (MLP) of Y and predictive maximum likelihood estimate (PMLE) of the scale parameter θ. The MLP and PMLE are denoted by $\hat{Y}$ and $\hat{\theta}_P$, respectively. It is noted that the MLP of Y is the same as $X_{k:m:n}^{\mathbf{r}}$ for $j = 1$.

From Lemma 3.1 in Seo and Kang [22], the pivot

$$(5.2) \qquad W_\theta(Y) = \frac{1 - F(Y)}{1 - F(x_k)} = \frac{\Gamma\left(\frac{1}{2}, \left(\frac{Y}{\theta}\right)^2\right)}{\Gamma\left(\frac{1}{2}, \left(\frac{x_k}{\theta}\right)^2\right)}$$

has beta distribution with parameters $r_k - j + 1$ and j.

When the parameter θ is known or given, we can obtain a predictor for Y by solving the following equation:

$$(5.3) \qquad W_\theta(Y) = \frac{\Gamma\left(\frac{1}{2}, \left(\frac{Y}{\theta}\right)^2\right)}{\Gamma\left(\frac{1}{2}, \left(\frac{x_k}{\theta}\right)^2\right)} \approx E(B_{r_k-j+1,j}) = \frac{r_k - j + 1}{r_k + 1},$$

where $B_{r_k-j+1,j}$ beta random variable with parameters $r_k - j + 1$ and j. This predictor is denoted by $\hat{Y}_2$. If θ is unknown, we can use $\hat{\theta}_P$ for θ in (5.3), then a new predictor of Y can be obtained by the solution of the equation

$$W_{\hat{\theta}_P}(Y) = \frac{r_k - j + 1}{r_k + 1},$$

where $\hat{\theta}_P$ is PMLE of θ and this predictor is denoted by $\hat{Y}_3$.

Presently, we aim to obtain some predictors which give explicit predictions. Let us define

$$p^* = E(U_{j:r_k} | U_{k:m:n} > F(x_k))$$

$$= \frac{r_k!(1 - F(x_k))^{-r_k}}{(j-1)!(r_k-j)!} \int_{F(x_k)}^{1} y(y - F(x_k))^{j-1}(1-y)^{r_k-j} dy$$

$$= \frac{r_k!(1 - F(x_k))^{-r_k}}{(j-1)!(r_k-j)!} \sum_{i_1=0}^{j-1} \sum_{i_2=0}^{r_k-j} \binom{j-1}{i_i} \binom{r_k-j}{i_2}$$

$$(5.4) \qquad \times (-1)^{i_1+i_2} F^{i_1}(x_k) \left(\frac{1 - F^{j-i_1+i_2+1}(x_k)}{j - i_1 + i_2 + 1} \right),$$

where $U_{k:m:n}$ is the k-th standard uniform progressive censored statistic and $U_{j:r_k}$ is the standard uniform j-th ordinary order statistics related to removed sample of size r_k at the stage k. Using the same methodology in Section 3, we can also write

$$(5.5) \qquad N_\theta(Y) = -\log\left(\frac{\Gamma\left(\frac{1}{2}, \left(\frac{Y}{\theta}\right)^2\right)}{\Gamma\left(\frac{1}{2}\right)} \right) \approx L + M\frac{Y}{\theta},$$

where

$$L = -\log\left(\frac{\Gamma\left(\frac{1}{2}, \xi^2\right)}{\Gamma\left(\frac{1}{2}\right)}\right) - \frac{2\xi \exp\left(-\xi^2\right)}{\Gamma\left(\frac{1}{2}, \xi^2\right)},$$

$$M = \frac{2 \exp\left(-\xi^2\right)}{\Gamma\left(\frac{1}{2}, \xi^2\right)},$$

and

$$\xi = E(Y_{j:R_k}) \approx G^{-1}(p^*).$$

Using (2.5) and (5.4), ξ can be determined by solving the following equation:

$$\Gamma\left(\frac{1}{2}, \xi^2\right) = \Gamma\left(\frac{1}{2}\right)(1 - p^*).$$

We are now ready to give a new explicit predictor. Using (5.5) in (5.3), a new predictor of Y is given by

$$\hat{Y}_4 = \frac{\theta\left(-\log\left(\frac{\Gamma\left(\frac{1}{2}, \left(\frac{x_k}{\theta}\right)^2\right)}{\Gamma\left(\frac{1}{2}\right)} \frac{r_k - j + 1}{r_k + 1}\right) - L\right)}{M},$$

where θ is given or known. If the θ is unknown, another predictor of Y can be defined as

$$\hat{Y}_5 = \frac{\hat{\theta}_P\left(-\log\left(\frac{\Gamma\left(\frac{1}{2}, \left(\frac{x_k}{\hat{\theta}_P}\right)^2\right)}{\Gamma\left(\frac{1}{2}\right)} \frac{r_k - j + 1}{r_k + 1}\right) - L\right)}{M}.$$

In the following section, the predictive intervals are discussed for failure times of progressively removed units.

6.   PREDICTIVE INTERVALS

In this section, we discuss the predictive intervals (PIs) for Y. Using maximum likelihood theory, an approximately predictive interval for Y can be written by

(6.1) $\left(\hat{Y} - z_{1-\frac{\alpha}{2}} \operatorname{SE}\left(\hat{Y}\right), \hat{Y} + z_{1-\frac{\alpha}{2}} \operatorname{SE}\left(\hat{Y}\right)\right),$

where $\operatorname{SE}\left(\hat{Y}\right)$ can be found in a similar way in Section 4 by using the negative Hessian matrix of predictive log-likelihood function (5.1).

Let us consider the pivot (5.2) to construct a new PI for Y. For this purpose, we need the following lemma.

Lemma 6.1.   *Suppose that $0 < a_1 < a_2 < \cdots < a_m < \infty$. Then, $W_\theta(y)$ is strictly decreasing in y for any $y > 0$. Furthermore, if $t > 0$, the equation $W_\theta(y) = t$ has a unique solution for any $y > 0$.*

Proof: The proof is similar to the proof of Lemma 4.1 and it is omitted.   □

Let $\beta_{(a)}$ be the a-th quantile of the beta distribution with parameters $r_k - j + 1$ and j. Then the following theorem gives an exact PI of Y.

Theorem 6.1. *An exact* $100(1-\alpha)\%$ *predictive interval for* Y *can be constructed by*

$$(6.2) \qquad \left(W_\theta^{-1}\big(\beta_{(1-\alpha/2)}\big), W_\theta^{-1}\big(\beta_{(\alpha/2)}\big)\right),$$

where scale parameter θ *is known.*

Proof: The proof follows by using Lemma 6.1 and Lemma 3.1 in Seo and Kang [22]. $\qquad\square$

Corollary 6.1. *When the scale parameter* θ *is unknown, an approximately* $100(1-\alpha)\%$ *predictive interval for* Y *can be given by*

$$(6.3) \qquad \left(W_{\hat\theta_P}^{-1}\big(\beta_{(1-\alpha/2)}\big), W_{\hat\theta_P}^{-1}\big(\beta_{(\alpha/2)}\big)\right),$$

where $\hat\theta_P$ *is PMLE of* θ.

However, PIs in (6.2)–(6.3) cannot be obtained explicitly. In the following, we provide an explicit solution for the PI bounds. Using the pivot in (5.2), we have

$$
\begin{aligned}
1 - \alpha &\approx P\big(\beta_{(\alpha/2)} < W_\theta(Y) < \beta_{(1-\alpha/2)}\big)\\
&= P\left(\Gamma\left(\frac{1}{2}, \left(\frac{x_k}{\theta}\right)^2\right)\beta_{(\alpha/2)} < \Gamma\left(\frac{1}{2}, \left(\frac{Y}{\theta}\right)^2\right) < \Gamma\left(\frac{1}{2}, \left(\frac{x_k}{\theta}\right)^2\right)\beta_{(1-\alpha/2)}\right)\\
(6.4) \quad &= P\left(-\log\left(\frac{\Gamma\left(\frac{1}{2}, \left(\frac{x_k}{\theta}\right)^2\right)\beta_{(1-\alpha/2)}}{\Gamma\left(\frac{1}{2}\right)}\right) < N_\theta(Y) < -\log\left(\frac{\Gamma\left(\frac{1}{2}, \left(\frac{x_k}{\theta}\right)^2\right)\beta_{(\alpha/2)}}{\Gamma\left(\frac{1}{2}\right)}\right)\right).
\end{aligned}
$$

By substituting (5.5) in (6.4), we have the following corollaries.

Corollary 6.2. *An approximately* $100(1-\alpha)\%$ *predictive interval for* Y *can be constructed by*

$$(6.5) \qquad \left(\frac{\theta\left(-\log\left(\frac{\Gamma\left(\frac{1}{2}, \left(\frac{x_k}{\theta}\right)^2\right)\beta_{(1-\alpha/2)}}{\Gamma\left(\frac{1}{2}\right)}\right) - L\right)}{M}, \frac{\theta\left(-\log\left(\frac{\Gamma\left(\frac{1}{2}, \left(\frac{x_k}{\theta}\right)^2\right)\beta_{(\alpha/2)}}{\Gamma\left(\frac{1}{2}\right)}\right) - L\right)}{M}\right),$$

where the scale parameter θ *is known.*

Corollary 6.3. *When the scale parameter θ is unknown, an approximately $100(1-\alpha)\%$ predictive interval for Y can be constructed by*

$$(6.6) \quad \left(\frac{\hat{\theta}_P \left(-\log\left(\frac{\Gamma\left(\frac{1}{2}, \left(\frac{x_k}{\hat{\theta}_P}\right)^2\right)\beta_{(1-\alpha/2)}}{\Gamma\left(\frac{1}{2}\right)} \right) - L \right)}{M}, \frac{\hat{\theta}_P \left(-\log\left(\frac{\Gamma\left(\frac{1}{2}, \left(\frac{x_k}{\hat{\theta}_P}\right)^2\right)\beta_{(\alpha/2)}}{\Gamma\left(\frac{1}{2}\right)} \right) - L \right)}{M} \right).$$

In the following section, all estimation and prediction methods are compared through the Monte Carlo simulation.

7.   SIMULATION STUDY

In this section, we perform a simulation study to observe the performance of estimators, predictors, confidence intervals and predictive intervals discussed in Sections 2-6. Several censoring schemes are used in this study. 5000 trials are used in the simulation. The bias, variance and mean squared errors (MSEs) of the estimates $\hat{\theta}, \widetilde{\theta}, \widetilde{\theta}_{\text{revised}}, \theta^*, \theta^{*(a)}$ and $\theta^{*(a)}_{\text{revised}}$ are simulated. 100 iteration is performed to reach all the revised estimates. The coverage probabilities (CPs) and average lengths (ALs) of the CIs given in (4.1)–(4.4) are calculated. Furthermore, the bias and mean squared prediction errors (MSPEs) of the predictors $\hat{Y}_1$, $\hat{Y}_2$, $\hat{Y}_3$, $\hat{Y}_4$ and $\hat{Y}_5$ are simulated. The CPs and ALs of the PIs given in (6.1), (6.2), (6.3), (6.5) and (6.6) are also calculated. The nominal value is fixed at $\alpha = 0.05$ for all CIs and PIs.

In the following tables, the CIs has given in (4.1)–(4.4) are denoted by CI1, CI2, CI3, CI4 and CI5 respectively. The PIs given in (6.1), (6.2), (6.3), (6.5) and (6.6) are denoted by PI1, PI2, PI3, PI4 and PI5. A little part of the simulation results are presented in Tables 1–4, and the rest of the tables are included in the supplementary file.

According to Table 1 and the rest of the results in the supplementary file, it is concluded that $\hat{\theta}$ and $\widetilde{\theta}_{\text{revised}}$ have identical MSEs and bias. This result shows that the revised AMLE tends to MLE. θ^*, $\theta^{*(a)}$ and $\theta^{*(a)}_{\text{revised}}$ are worse than the others with a slight difference in terms of MSEs. It is observed that the censoring made at the first stage is a better choice to get low MSEs. It is also observed from Table 2 that the CI2, CI3 and CI4 have desired CPs even if small sample cases. However, the CPs of CI1 are not at the desired level for a small sample case but it reaches to the nominal value for the moderate size of m. It should be pointed out that CI4 (ULR) has the smallest average length in almost all censoring schemes.

According to Tables 3–4 and the rest of the results in the supplementary file, it is concluded that as j increases, the MSPEs of all predictors increases whereas the MSPEs of predictors decrease as k increases. The MSPEs of $\hat{Y}_2$ and $\hat{Y}_4$ are the same where $\hat{Y}_2$ is obtained by a numerical method and $\hat{Y}_4$ is obtained explicitly. Then we concluded that $\hat{Y}_4$ should be used to predict the Y instead of $\hat{Y}_2$ when the the θ is known. The MSPEs of $\hat{Y}_3$ and $\hat{Y}_5$ are almost the same where $\hat{Y}_3$ is obtained by a numerical method and $\hat{Y}_5$ is obtained explicitly.

It is concluded that $\hat{Y}_5$ should be used to predict the Y instead of $\hat{Y}_3$ when the θ is unknown. The $\hat{Y}_5$ has better MSPEs than the $\hat{Y}_1$ has when the small values of j. This pattern is reversed for the large values of j.

Table 1: MSEs, bias (in parenthesis) and variances for estimates of the scale parameter ($\hat{\theta}=1$).

n	m	Censoring Schemes	$\hat{\theta}$	$\tilde{\theta}$	$\tilde{\theta}_{\text{revised}}$	θ^*	$\theta^{*(a)}$	$\theta^{*(a)}_{\text{revised}}$
10	10	(10^*0)	0.0487 (-0.0253) 0.0480	0.0487 (-0.0253) 0.0480	0.0487 (-0.0253) 0.0480	0.0565 (-0.1347) 0.0384	0.0518 (-0.0965) 0.0425	0.0498 (-0.0799) 0.0435
20	10	(10^*1)	0.0583 (-0.0236) 0.0577	0.0621 (0.0021) 0.0621	0.0583 (-0.0236) 0.0577	0.0654 (-0.1437) 0.0448	0.0596 (-0.0955) 0.0505	0.0584 (-0.0838) 0.0513
	10	$(5, 8^*0, 5)$	0.0584 (-0.0238) 0.0579	0.0600 (-0.0094) 0.0599	0.0584 (-0.0238) 0.0579	0.0655 (-0.1440) 0.0447	0.0588 (-0.0902) 0.0507	0.0583 (-0.0841) 0.0513
	10	$(5, 5, 8^*0)$	0.0507 (-0.0252) 0.0501	0.0508 (-0.0230) 0.0503	0.0507 (-0.0252) 0.0501	0.0586 (-0.1368) 0.0399	0.0541 (-0.0986) 0.0443	0.0519 (-0.0807) 0.0454
	10	$(8^*0, 5, 5)$	0.0689 (-0.0154) 0.0687	0.07001 (-0.0077) 0.0700	0.0689 (-0.0154) 0.0687	0.0724 (-0.1424) 0.0522	0.0667 (-0.0848) 0.0595	0.0664 (-0.0799) 0.0601
	10	$(4^*0, 5, 5, 4^*0)$	0.0560 (-0.0237) 0.0554	0.0562 (-0.0205) 0.0558	0.0560 (-0.0237) 0.0554	0.0634 (-0.1415) 0.0434	0.0591 (-0.1014) 0.0488	0.0565 (-0.0825) 0.0497
	20	(20^*0)	0.0244 (-0.0119) 0.0243	0.0244 (-0.0119) 0.0243	0.0244 (-0.0119) 0.0243	0.0270 (-0.0716) 0.0219	0.0257 (-0.0496) 0.0232	0.0250 (-0.0404) 0.0234
40	20	(20^*1)	0.0295 (-0.0113) 0.0293	0.0326 (0.0137) 0.0325	0.0295 (-0.0113) 0.0293	0.0319 (-0.0770) 0.0260	0.0303 (-0.0497) 0.0278	0.0298 (-0.0429) 0.0280
	20	$(10, 18^*0, 10)$	0.0296 (-0.0116) 0.0294	0.0301 (-0.0028) 0.0301	0.0296 (-0.0116) 0.0294	0.0319 (-0.0772) 0.0259	0.0299 (-0.0457) 0.0278	0.0298 (-0.0431) 0.0279
	20	$(10, 10, 18^*0)$	0.0249 (-0.0119) 0.0248	0.0250 (-0.0109) 0.0249	0.0249 (-0.0119) 0.0248	0.0276 (-0.0722) 0.0223	0.0263 (-0.0502) 0.0237	0.0256 (-0.0406) 0.0239
	20	$(18^*0, 10, 10)$	0.0324 (-0.0106) 0.0323	0.0326 (-0.0068) 0.0325	0.0324 (-0.0106) 0.0323	0.0345 (-0.0798) 0.0281	0.0324 (-0.0460) 0.0303	0.0323 (-0.0442) 0.0304
	20	$(9^*0, 10, 10, 9^*0)$	0.0281 (-0.0112) 0.0279	0.0282 (-0.0095) 0.0281	0.0281 (-0.0112) 0.0279	0.0307 (-0.0755) 0.0250	0.0296 (-0.0529) 0.0268	0.0287 (-0.0420) 0.0269
50	50	(50^*0)	0.0095 (-0.0039) 0.0095	0.0095 (-0.0039) 0.0095	0.0095 (-0.0039) 0.0095	0.0103 (-0.0300) 0.0094	0.0101 (-0.0206) 0.0097	0.0100 (-0.0165) 0.0097
100	50	(50^*1)	0.0116 (-0.0038) 0.0116	0.0143 (0.0178) 0.0140	0.0116 (-0.0038) 0.0116	0.0123 (-0.0324) 0.0113	0.0121 (-0.0209) 0.0117	0.0120 (-0.0175) 0.0117
	50	$(25, 48^*0, 25)$	0.0116 (-0.0040) 0.0116	0.0117 (0.0001) 0.0117	0.0116 (-0.0040) 0.0116	0.01236 (-0.0325) 0.0113	0.0119 (-0.0186) 0.0116	0.0119 (-0.0177) 0.0116
	50	$(25, 25, 48^*0)$	0.0096 (-0.0039) 0.0096	0.0096 (-0.0036) 0.0096	0.0096 (-0.0039) 0.0096	0.0104 (-0.0301) 0.0095	0.0102 (-0.0207) 0.0098	0.0100 (-0.0165) 0.0098
	50	$(48^*0, 25, 25)$	0.0129 (-0.0037) 0.0129	0.0129 (-0.0021) 0.0129	0.0129 (-0.0037) 0.0129	0.0135 (-0.0337) 0.0124	0.0131 (-0.0187) 0.0128	0.0131 (-0.0182) 0.0128
	50	$(24^*0, 25, 25, 24^*0)$	0.0110 (-0.0037) 0.0110	0.0110 (-0.0031) 0.0110	0.0110 (-0.0037) 0.0110	0.0118 (-0.0316) 0.0108	0.0117 (-0.0222) 0.0112	0.0114 (-0.0171) 0.0112

Table 2: CPs and ALs of CI for the scale parameter $\hat{\theta}$ at the 0.95 confidence level ($\hat{\theta}=1$).

n	m	Censoring Schemes	CPs				ALs			
			CI1	CI2	CI3	CI4	CI1	CI2	CI3	CI4
10	10	(10^*0)	0.8988	0.9504	0.9516	0.9464	0.8530	1.0435	0.9810	0.9613
20	10	(10^*1)	0.9014	0.9542	0.9518	0.9512	0.9427	1.1611	1.0788	1.0739
	10	$(5, 8^*0, 5)$	0.8918	0.9514	0.9438	0.9452	0.9417	1.1570	1.0658	1.0731
	10	$(5, 5, 8^*0)$	0.9052	0.9512	0.9540	0.9478	0.8798	1.0793	1.0167	0.9937
	10	$(8^*0, 5, 5)$	0.9010	0.9492	0.9462	0.9460	0.9852	1.2102	1.1193	1.1263
	10	$(4^*0, 5, 5, 4^*0)$	0.8976	0.9490	0.9518	0.9478	0.9236	1.1426	1.0762	1.0492
	20	(20^*0)	0.9208	0.9514	0.9488	0.9448	0.6115	0.6791	0.6634	0.6488
40	20	(20^*1)	0.9234	0.9486	0.9490	0.9464	0.6741	0.7504	0.7260	0.7191
	20	$(10, 18^*0, 10)$	0.9266	0.9518	0.9496	0.9510	0.6737	0.7479	0.7188	0.7188
	20	$(10, 10, 18^*0)$	0.9292	0.9488	0.9524	0.9468	0.6194	0.6899	0.6738	0.6576
	20	$(18^*0, 10, 10)$	0.9226	0.9488	0.9488	0.9490	0.7078	0.7852	0.7561	0.7566
	20	$(9^*0, 10, 10, 9^*0)$	0.9252	0.9452	0.9472	0.9454	0.6584	0.7359	0.7196	0.7013
50	50	(50^*0)	0.9368	0.9580	0.9550	0.9530	0.3895	0.4088	0.4068	0.3988
100	50	(50^*1)	0.9374	0.9490	0.9506	0.9490	0.4295	0.4509	0.4458	0.4407
	50	$(25, 48^*0, 25)$	0.9422	0.9508	0.9510	0.9518	0.4306	0.4507	0.4437	0.4418
	50	$(25, 25, 48^*0)$	0.9450	0.9510	0.9546	0.9498	0.3928	0.4128	0.4106	0.4022
	50	$(48^*0, 25, 25)$	0.9346	0.9482	0.9484	0.9476	0.4504	0.4707	0.4637	0.4625
	50	$(24^*0, 25, 25, 24^*0)$	0.9352	0.9500	0.9494	0.9466	0.4167	0.4393	0.4372	0.4273

The CPs of PI3 and PI5 are at the nominal level for small values of j. When j increases, CPs of PI3 and PI5 decrease from the nominal level 0.95. It should be pointed out that the as j increases, CPs are may decrease to 0.88, 0.91 and 0.93 for $m = 20, 40$ and 100, respectively. Fortunately, increasing m overcomes the low CPs problem. Also, the PI2 and PI4 keep the nominal level since the PI2 is an exact predictive interval and PI4 is an approximated version of PI2 with a Taylor expansion. The CPs of PI1 increase first to the nominal level and then decrease when the j increases. We conclude that PI3 should be used for small and large values of j, respectively. PI1 can be used for moderate values of j. That is, if $r_{10} = 10$, PI1 may be used to construct the PI of Y, for $j = 6, 7, 8$, otherwise, PI3 should be used instead of PI1. As j increases, the ALs of all predictive intervals increase whereas they get smaller for large values of k. PI1 and PI3 have almost the same AL when they have the same CPs. PI2 (PI3) has smaller ALs than PI4 (PI5) for all cases discussed here.

Table 3: MSEs, bias (in parenthesis) for $\hat{Y}_i$ and CPs [ ], ALs [ ] for PI_i
when $\hat{\theta}=1$, $m=20$, $\mathbf{r}=(5,5,...,5)$.

CSs	k	j		$\hat{Y}_1$ [PI1]	$\hat{Y}_2$ [PI2]	$\hat{Y}_3$ [PI3]	$\hat{Y}_4$[PI4]	$\hat{Y}_5$ [PI5]
(5*m)	k=5	j=2		0.0437 (−0.1042)	0.0329 (−0.0070)	0.0365 (−0.0109)	0.0329 (−0.0070)	0.0365 (−0.0109)
			CPs	[0.8898]	[0.9518]	[0.9324]	[0.9296]	[0.9118]
			ALs	[0.6862]	[0.6940]	[0.6848]	[0.7609]	[0.7503]
		j=3		0.0631 (−0.0940)	0.0544 (−0.0273)	0.0642 (−0.0307)	0.0544 (−0.0273)	0.0642 (−0.0307)
			CPs	[0.9184]	[0.9478]	[0.9124]	[0.9302]	[0.9010]
			ALs	[0.8560]	[0.8778]	[0.8713]	[0.9589]	[0.9513]
		j=4		0.0907 (−0.0879)	0.0837 (−0.0502)	0.1014 (−0.0555)	0.0837 (−0.0502)	0.1014 (−0.0555)
			CPs	[0.9262]	[0.9486]	[0.9078]	[0.9348]	[0.9086]
			ALs	[1.1337]	[1.0959]	[1.0871]	[1.2107]	[1.2005]
		j=5		0.1785 (−0.1140)	0.1737 (−0.1392)	0.2122 (−0.1444)	0.1737 (−0.1392)	0.2123 (−0.1444)
			CPs	[0.9244]	[0.9490]	[0.9012]	[0.9458]	[0.9172]
			ALs	[1.7537]	[1.5142]	[1.5060]	[1.7623]	[1.7522]
	k=10	j=2		0.0429 (−0.1079)	0.0313 (−0.0127)	0.0350 (−0.0144)	0.0313 (−0.0127)	0.0350 (−0.0143)
			CPs	[0.8786]	[0.9538]	[0.9324]	[0.9358]	[0.9170]
			ALs	[0.6404]	[0.6712]	[0.6673]	[0.7336]	[0.7279]
		j=3		0.0592 (−0.0947)	0.0505 (−0.0276)	0.0598 (−0.0313)	0.0505 (−0.0276)	0.0598 (−0.0312)
			CPs	[0.9156]	[0.9476]	[0.9184]	[0.9374]	[0.9122]
			ALs	[0.8970]	[0.8548]	[0.8480]	[0.9322]	[0.9232]
		j=4		0.0913 (−0.0989)	0.0833 (−0.0594)	0.1022 (−0.0660)	0.0833 (−0.0594)	0.1022 (−0.0660)
			CPs	[0.9230]	[0.9480]	[0.9072]	[0.9398]	[0.9082]
			ALs	[1.0754]	[1.0738]	[1.0630]	[1.1846]	[1.1713]
		j=5		0.1719 (−0.1189)	0.1664 (−0.1390)	0.2071 (−0.1518)	0.1664 (−0.1390)	0.2071 (−0.1517)
			CPs	[0.9294]	[0.9516]	[0.9120]	[0.9454]	[0.9170]
			ALs	[1.3226]	[1.4936]	[1.4737]	[1.7355]	[1.7109]
	k=15	j=2		0.0389 (−0.1030)	0.0283 (−0.0105)	0.0319 (−0.0130)	0.0283 (−0.0105)	0.0319 (−0.0217)
			CPs	[0.8746]	[0.9480]	[0.9266]	[0.9301]	[0.9082]
			ALs	[0.6047]	[0.6382]	[0.6325]	[0.6942]	[0.6857]
		j=3		0.0561 (−0.0936)	0.0475 (−0.0265)	0.0567 (−0.0306)	0.0475 (−0.0265)	0.0567 (−0.0303)
			CPs	[0.9132]	[0.9476]	[0.9140]	[0.9336]	[0.9042]
			ALs	[0.7431]	[0.8203]	[0.8129]	[0.8919]	[0.8813]
		j=4		0.0860 (−0.0978)	0.0779 (−0.0579)	0.0962 (−0.0626)	0.0779 (−0.0579)	0.0962 (−0.0624)
			CPs	[0.9248]	[0.9482]	[0.9106]	[0.9410]	[0.9086]
			ALs	[1.0573]	[1.0395]	[1.0321]	[1.1441]	[1.1333]
		j=5		0.1550 (−0.0988)	0.1508 (−0.1202)	0.1866 (−0.1280)	0.1508 (−0.1202)	0.1866 (−0.1278)
			CPs	[0.9336]	[0.9514]	[0.9112]	[0.9444]	[0.9160]
			ALs	[1.4495]	[1.4608]	[1.4490]	[1.6924]	[1.6756]

Table 4:   MSEs, bias (in parenthesis) for $\hat{Y}_i$ and CPs [ ], ALs [ ] for PI$_i$
when $\hat{\theta}=1$, $m=20$, $\mathbf{r}=(10,10,...,10)$.

CSs	k	j		$\hat{Y}_1$ [PI1]	$\hat{Y}_2$ [PI2]	$\hat{Y}_3$ [PI3]	$\hat{Y}_4$[PI4]	$\hat{Y}_5$ [PI5]
(10*m)	k=5	j=2		0.0151 (−0.0671)	0.0105 (−0.0023)	0.0172 (−0.0018)	0.0105 (−0.0013)	0.0117 (−0.0018)
			CPs	[0.8540]	[0.9528]	[0.9360]	[0.9404]	[0.9220]
			ALs	[0.3494]	[0.3881]	[0.3893]	[0.4114]	[0.4124]
		j=3		0.0193 (−0.0589)	0.0159 (−0.0048)	0.0184 (−0.0061)	0.0159 (−0.0048)	0.0184 (−0.0060)
			CPs	[0.9040]	[0.9496]	[0.9218]	[0.9354]	[0.9130]
			ALs	[0.4719]	[0.4746]	[0.4721]	[0.5005]	[0.4976]
		j=4		0.0224 (−0.0509)	0.0199 (−0.0055)	0.0245 (−0.0175)	0.0199 (−0.0055)	0.0245 (−0.0075)
			CPs	[0.9310]	[0.9488]	[0.9170]	[0.9418]	[0.9110]
			ALs	[0.5689]	[0.5484]	[0.5450]	[0.5763]	[0.5725]
		j=5		0.0284 (−0.0488)	0.0261 (−0.0111)	0.0338 (−0.0130)	0.0261 (−0.0111)	0.0338 (−0.0130)
			CPs	[0.9344]	[0.9438]	[0.8986]	[0.9356]	[0.8926]
			ALs	[0.6365]	[0.6174]	[0.6147]	[0.6479]	[0.6447]
		j=6		0.0343 (−0.0478)	0.0320 (−0.0169)	0.0439 (−0.0222)	0.0320 (−0.0169)	0.0439 (−0.0222)
			CPs	[0.9344]	[0.9500]	[0.8900]	[0.9420]	[0.8898]
			ALs	[0.6699]	[0.6886]	[0.6815]	[0.7229]	[0.7152]
		j=7		0.0409 (−0.0465)	0.0389 (−0.0230)	0.5556 (−0.0293)	0.0389 (−0.0230)	0.05556 (−0.0293)
			CPs	[0.9426]	[0.9512]	[0.8878]	[0.9414]	[0.8894]
			ALs	[0.7445]	[0.7692]	[0.7616]	[0.8104]	[0.8021]
		j=8		0.0539 (−0.0536)	0.0516 (−0.0407)	0.0780 (−0.0420)	0.0516 (−0.0407)	0.0780 (−0.0420)
			CPs	[0.9346]	[0.9512]	[0.8764]	[0.9482]	[0.8766]
			ALs	[0.8408]	[0.8733]	[0.8719]	[0.9276]	[0.9257]
		j=9		0.0754 (−0.0553)	0.0731 (−0.0594)	0.1128 (−0.0641)	0.0731 (−0.0594)	0.1128 (−0.0641)
			CPs	[0.9420]	[0.9526]	[0.8710]	[0.9432]	[0.8798]
			ALs	[1.0038]	[1.0361]	[1.0310]	[1.1206]	[1.1148]
		j=10		0.1513 (−0.0698)	0.1538 (−0.1347)	0.2165 (−0.1382)	0.1538 (−0.1347)	0.2170 (−0.1379)
			CPs	[0.9348]	[0.9480]	[0.8738]	[0.9482]	[0.8800]
			ALs	[1.5188]	[1.4091]	[1.4053]	[1.6079]	[1.6033]
	k=10	j=2		0.0137 (−0.0634)	0.0097(0.0001)	0.0108 (−0.0011)	0.0097(0.0001)	0.0108 (−0.0010)
			CPs	[0.8544]	[0.9542]	[0.9370]	[0.9398]	[0.9258]
			ALs	[0.3423]	[0.3792]	[0.3762]	[0.4013]	[0.3975]
		j=3		0.0182 (−0.0604)	0.0146 (−0.0070)	0.0172 (−0.0081)	0.0146 (−0.0070)	0.0172 (−0.0080)
			CPs	[0.9050]	[0.9540]	[0.9246]	[0.9454]	[0.9162]
			ALs	[0.4557]	[0.4647]	[0.4626]	[0.4895]	[0.4866]
		j=4		0.0222 (−0.0539)	0.0193 (−0.0089)	0.0244 (−0.0102)	0.0193 (−0.0089)	0.0244 (−0.0101)
			CPs	[0.9266]	[0.9542]	[0.9162]	[0.9468]	[0.9138]
			ALs	[0.5367]	[0.5381]	[0.5359]	[0.5651]	[0.5621]
		j=5		0.0284 (−0.0509)	0.0258 (−0.0132)	0.0343 (−0.0142)	0.0258 (−0.0132)	0.0343 (−0.0142)
			CPs	[0.9328]	[0.9462]	[0.9016]	[0.9394]	[0.8958]
			ALs	[0.5935]	[0.6073]	[0.6059]	[0.6369]	[0.6346]
		j=6		0.0333 (−0.0476)	0.0310 (−0.0164)	0.0435 (−0.0217)	0.0310 (−0.0164)	0.0435 (−0.0217)
			CPs	[0.9366]	[0.9524]	[0.8880]	[0.9416]	[0.8860]
			ALs	[0.7046]	[0.6787]	[0.6716]	[0.7122]	[0.7043]
		j=7		0.0401 (−0.0466)	0.0379 (−0.0232)	0.0553 (−0.0268)	0.0379 (−0.0232)	0.0553 (−0.0268)
			CPs	[0.9422]	[0.9546]	[0.8890]	[0.9462]	[0.8856]
			ALs	[0.7768]	[0.7597]	[0.7554]	[0.8000]	[0.7949]
		j=8		0.0535 (−0.0557)	(0.0511)-0.0418	(0.0769)-0.0458	0.0511 (−0.0418)	0.0769 (−0.0458)
			CPs	[0.9348]	[0.9554]	[0.8762]	[0.9512]	[8784]
			ALs	[0.8871]	[0.8643]	[0.8598]	[0.9176]	[0.9121]
		j=9		0.0751 (−0.0599)	0.0736 (−0.0635)	0.1106 (−0.0687)	0.0736 (−0.0635)	0.1106 (−0.0687)
			CPs	[0.9340]	[0.9498]	[0.8758]	[0.9474]	[0.8808]
			ALs	[1.1053]	[1.0274]	[1.0217]	[1.1106]	[1.1037]
		j=10		0.1401 (−0.0639)	0.1413 (−0.1266)	0.2050 (−0.1345)	0.1413 (−0.1266)	0.2054 (−0.1343)
			CPs	[0.9372]	[0.9528]	[0.8748]	[0.9482]	[0.8874]
			ALs	[1.3007]	[1.4006]	[1.3915]	[1.5969]	[1.5859]
	k=15	j=2		0.0129 (−0.0611)	0.0092(0.0004)	0.0103 (−0.0005)	0.0092(0.0004)	0.0103 (−0.0004)
			CPs	[0.8506]	[0.9516]	[0.9282]	[0.9340]	[0.9168]
			ALs	[0.3302]	[0.3653]	[0.3629]	[0.3856]	[0.3819]
		j=3		0.0174 (−0.0580)	0.0140 (−0.0056)	0.0166 (−0.0079)	0.0140 (−0.0056)	0.0166 (−0.0077)
			CPs	[0.8992]	[0.9476]	[0.9184]	[0.9394]	[0.9144]
			ALs	[0.4316]	[0.4496]	[0.4451]	[0.4727]	[0.4670]
		j=4		0.0221 (−0.0569)	0.0189 (−0.0122)	0.0237 (−0.0144)	0.0189 (−0.0122)	0.0237 (−0.0143)
			CPs	[0.9192]	[0.9444]	[0.9056]	[0.9380]	[0.9044]
			ALs	[0.5284]	[0.5224]	[0.5187]	[0.5479]	[0.5427]
		j=5		0.0253 (−0.0499)	0.0229 (−0.0124)	0.0300 (−0.0142)	0.0229 (−0.0124)	0.0300 (−0.0141)
			CPs	[0.9348]	[0.9536]	[0.9032]	[0.9438]	[0.9034]
			ALs	[0.6051]	[0.5891]	[0.5891]	[0.6198]	[0.6159]
		j=6		0.0322 (−0.0528)	0.0295 (−0.0216)	0.0418 (−0.0256)	0.0295 (−0.0216)	0.0418 (−0.0255)
			CPs	[0.9328]	[0.9528]	[0.8858]	[0.9464]	[0.8888]
			ALs	[0.6727]	[0.6632]	[0.6579]	[0.6955]	[0.6887]
		j=7		0.0400 (−0.0496)	0.0378 (−0.0254)	0.0549 (−0.0329)	0.0378 (−0.0254)	0.0549 (−0.0328)
			CPs	[0.9356]	[0.9474]	[0.8764]	[0.9410]	[0.8788]
			ALs	[0.7739]	[0.7448]	[0.7358]	[0.7839]	[0.7734]
		j=8		0.0497 (−0.0466)	0.0479 (−0.0329)	0.0718 (−0.0364)	0.0479 (−0.0329)	0.0718 (−0.0363)
			CPs	[0.9424]	[0.9486]	[0.8852]	[0.9432]	[0.8864]
			ALs	[0.8077]	[0.8496]	[0.8458]	[0.9014]	[0.8961]
		j=9		0.0734 (−0.0581)	0.0715 (−0.0619)	0.1099 (−0.0646)	0.0715 (−0.0619)	0.1099 (−0.0645)
			CPs	[0.9378]	[0.9514]	[0.8686]	[0.9474]	[0.8682]
			ALs	[0.9801]	[1.0135]	[1.0107]	[1.0947]	[1.0902]
		j=10		0.1332 (−0.0635)	0.1367 (−0.1266)	0.1952 (−0.1327)	0.1367 (−0.1266)	0.1956 (−0.1324)
			CPs	[0.9426]	[0.9556]	[0.8858]	[0.9504]	[0.8916]
			ALs	[1.4331]	[1.3869]	[1.3801]	[1.5792]	[1.5699]

8.   ILLUSTRATIVE EXAMPLE

The real dataset represents the survival times of 121 patients with breast cancer obtained from a large hospital in a period from 1929 to 1938 in Lee [14]. For the complete data, the parameter estimation of the θ is obtained by ML methodology. The likelihood values, MLE of θ with the standard error, some goodness of fit tests such as Anderson-Darling (A), Cramer-von Mises (W) and Kolmogrov-Smirnov(K) test statistic, and corresponding p values (in parentheses) are given in Table 5. HN plot is also presented in Figure 2 which indicates the possibility to model for breast cancer data. In addition, since the p-values for the goodness of fit tests reported in Table 5 are greater than 0.05, the hypothesis that the data comes from the HN distribution cannot be rejected at a significance level 0.05.

Table 5:   Some results for survival times of 121 patients data when complete data.

ℓ	A	W	K	$\hat{\theta}$	SE
-579.4310	0.6606 (0.5922)	0.0896 (0.6388)	0.0548 (0.8596)	82.2258	5.2856

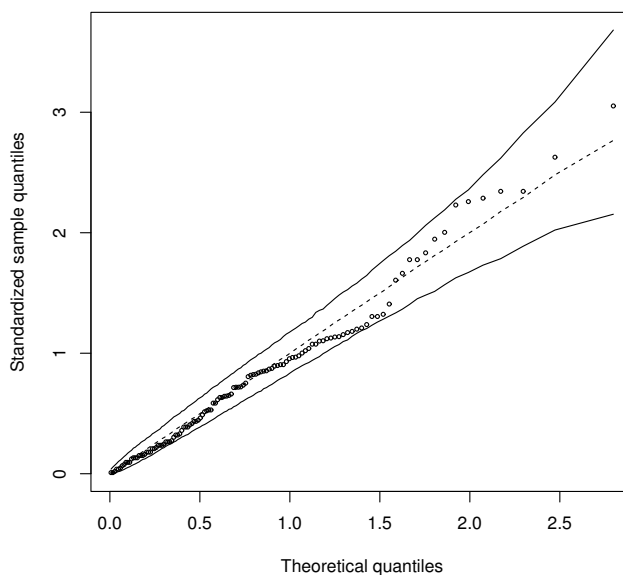**Figure 2**:   Half-Normal plot for the real data.

Let us censoring the complete data with scheme $r = \begin{pmatrix} \overbrace{109*0}^{109 \text{ times “0”}} ,5,5 \end{pmatrix}$. Then the progressively censored data is produced by: 0.3, 0.3, 4.0, 5.0, 5.6, 6.2, 6.3, 6.6, 6.8, 7.4, 7.5, 8.4, 8.4, 10.3, 11.0, 11.8, 12.2, 12.3, 13.5, 14.4, 14.4, 14.8, 15.5, 15.7, 16.2, 16.3, 16.5, 16.8, 17.2, 17.3, 17.5, 17.9, 19.8, 20.4, 20.9, 21.0, 21.0, 21.1, 23.0, 23.4, 23.6, 24.0, 24.0, 27.9, 28.2, 29.1, 30.0, 31.0, 31.0, 32.0, 35.0, 35.0, 37.0, 37.0, 37.0, 38.0, 38.0, 38.0, 39.0, 39.0, 40.0, 40.0, 40.0, 41.0, 41.0, 41.0, 42.0, 43.0, 43.0, 43.0, 44.0, 45.0, 45.0, 46.0, 46.0, 47.0, 48.0, 49.0, 51.0, 51.0, 51.0, 52.0, 54.0, 55.0, 56.0, 57.0, 58.0, 59.0, 60.0, 60.0, 60.0, 61.0, 62.0, 65.0, 65.0, 67.0, 67.0, 68.0, 69.0, 78.0, 80.0, 83.0, 88.0, 89.0, 90.0, 93.0, 96.0, 103.0, 105.0, 109.0, 129.0.

Using this progressively censored data, $\hat{\theta}$, $\widetilde{\theta}$, $\widetilde{\theta}_{\text{revised}}$, θ^*, $\theta^{*(a)}$ and $\theta^{*(a)}_{\text{revised}}$ give the estimates 87.1066, 97.7986, 87.1079, 84.9575, 85.1070 and 85.5067 respectively. CI1, CI2,CI3 and CI4 are also calculated as (75.3322, 98.8810), (75.7754, 99.3832), (74.9352, 98.5545) and (76.5309, 97.1066). Using this progressively censored data, predictions and predictive intervals are given in Table 6.

Table 6: PMLEs, MLPs and PIs for the first real data.

k	j	Y_j	$\hat{\theta}_P$	$\hat{Y}_1$	$\hat{Y}_3$	$\hat{Y}_5$	PI1	PI3	PI5
110	2	117	88.1947	117.5966	120.3328	121.6912	(101.1574, 134.0358)	(110.5625, 141.9586)	(110.9140, 147.9162)
	3	125	88.4368	128.4170	128.0046	129.1463	(102.7937, 154.0403)	(113.5498, 157.2334)	(113.6089, 164.7992)
	4	129	88.7842	140.7604	138.3815	139.2926	(106.9345, 174.5862)	(118.4837, 179.0106)	(118.5047, 189.3842)
	5	139	89.3816	160.4515	155.1442	155.7899	(112.6896, 208.2134)	(127.2086, 222.6391)	(127.6536, 241.9852)
111	2	111	88.7016	136.7600	139.1889	140.9083	(121.8496, 151.6704)	(130.3967, 158.9649)	(131.0636, 164.8641)
	3	126	88.9844	146.1106	146.1730	147.6439	(123.2522, 168.9689)	(133.0780, 173.1843)	(133.3209, 180.5102)
	4	127	89.3725	158.1813	155.712	156.9107	(126.6520, 189.7105)	(137.5327, 193.7159)	(137.5481, 203.5679)
	5	154	89.9751	175.3795	171.2774	172.1548	(131.3465, 219.4126)	(145.4661, 235.4339)	(145.6123, 253.4672)

9. CONCLUDING REMARKS

This study addresses the problem of estimating the scale parameter of HN distribution under progressively Type-II censoring. An approximate maximum likelihood, pivotal type and approximate pivotal type estimators (predictors) are derived and confidence intervals (predictors) are constructed for the scale parameter. The performance of the derived estimators (predictors) is compared under different censoring schemes. In addition, a numerical example is presented. It is concluded that under progressive Type-II censoring, the above-mentioned estimators (predictors) and CIs (PIs) can be used in HN distribution and they are competitive with the MLEs (PMLEs). Furthermore, some of our estimators (predictors) are explicitly obtained unlike MLEs (PMLEs). Considering several methods and proving their superior performances with different criterias, this study contributes to estimation and prediction problems in different ways.

REFERENCES

[1] AARSET, M.V. (1987). How to identify a bathtub hazard rate, *IEEE Transactions on Reliability*, **36**(1), 106–108.

[2] AHMADI, K.; REZAEI, M. and YOUSEFZADEH, F. (2015). Estimation for the generalized half-normal distribution based on progressive type-II censoring, *Journal of Statistical Computation and Simulation*, **85**(6), 1128–1150.

[3] ASGHARZADEH, A. and VALIOLLAHI, R. (2010). Prediction intervals for proportional hazard rate models based on progressively Type II censored samples, *Communications for Statistical Applications and Methods*, **17**(1), 99–106.

[4] BALAKRISHNAN, N. and AGGARWALA, R. (2000). *Progressive Censoring: Theory, Methods, and Applications*, Springer Science & Business Media.

[5] DEY, S.; SINGH, S.; TRIPATHI, Y.M. and ASGHARZADEH, A. (2016). Estimation and prediction for a progressively censored generalized inverted exponential distribution, *Statistical Methodology*, **32**, 185–202.

[6]   DOGANAKSOY, N. and SCHMEE, J. (1993). Comparisons of approximate confidence intervals for distributions used in life-data analysis, *Technometrics*, **35**(2), 175–184.

[7]   DOGANAKSOY, N. (2020). A simplified formulation of likelihood ratio confidence intervals using a novel property, *Technometrics*, **63**, 127–135.

[8]   FRASER, D.A.S. (1976). *Probability and Statistics: Theory and Applications*, North Scituate MA, Duxbury Press.

[9]   HEMMATI, F. and KHORRAM, E. (2017). On adaptive progressively Type-II censored competing risks data, *Communications in Statistics – Simulation and Computation*, **46**(6), 4671–4693.

[10] HUANG, J. and ROTH, D.L. (2021). Using the half normal distribution to quantify covariate balance in cluster-randomized pragmatic trials, *Trials*, **22**(1), 1–11.

[11] KHAN, H.M.R. (2013). Predictive inference from the half-normal model given a type II censored sample, *Communications in Statistics – Theory and Methods*, **42**(1), 42–55.

[12] KINACI, I.; WU, S.J. and KUS, C. (2019). Confidence intervals and regions for the generalized inverted exponential distribution based on progressively censored and upper records data, *REVSTAT – Statistical Journal*, **17**(4), 429–448.

[13] LEE, K.R.; KAPADIA, C.H. and BROCK, D.B. (1980). On estimating the scale parameter of the Rayleigh distribution from doubly censored samples, *Springer*, **21**(1), 14–29.

[14] LEE, E.T. (1992). *Statistical Methods for Survival Data Analysis*, John Wiley & Sons, New York.

[15] MAJUMDAR, H. and SEN, P.K. (1978a). Nonparametric testing for simple regression under progressive censoring with staggering entry and random withdrawal, *Communication in Statistics – Theory and Methods*, **7**(4), 349–371.

[16] MAJUMDAR, H. and SEN, P.K. (1978b). Rank order tests for multiple regression under progressive censoring, *Journal of Multivariate Analysis*, **8**(1), 73–95.

[17] MA, Y. and GUI, W. (2019). Pivotal inference for the inverse Rayleigh distribution based on general progressively Type-II censored samples, *Journal of Applied Statistics*, **46**(5), 771–797.

[18] MOHIE EL-DIN, M.M. and SHAFAY, A.R. (2013). One-and two-sample Bayesian prediction intervals based on progressively Type-II censored data, *Statistical Papers*, **54**(2), 287–307.

[19] NG, H.K.T.; CHAN, P.S. and BALAKRISHNAN, N. (2002). Estimation of parameters from progressively censored data using EM algorithm, *Computational Statistics & Data Analysis*, **39**(4), 371–386.

[20] NG, H.K.T. (2021). *Progressively Censored Data Analysis*, Wiley StatsRef: Statistics Reference Online, 1–10.

[21] SEN, P.K. (2004). Progressively Censored Data Analysis, *Encyclopedia of Statistical Sciences*.

[22] SEO, J.I. and KANG, S.B. (2017). Inference for the two-parameter half-logistic distribution using pivotal quantities under progressively type-II censoring schemes, *Communications in Statistics – Simulation and Computation*, **46**(7), 5462–5478.

[23] SHANKER, R.; SHUKLA, K.K.; SHARMA, S. and SHANKER, R. (2018). A quasi exponential distribution, *Biometrics and Biostatistics International Journal*, **7**(2), 90–94.

[24] SINDHU, T.N. and HUSSAIN, Z. (2020). Predictive inference and parameter estimation from the half-normal distribution for the left censored data, *Annals of Data Science*, 1–15.

[25] THOMAS, D.R. and WILSON, W.M. (1972). Linear order statistic estimation for the two-parameter Weibull and extreme-value distributions from type II progressively censored samples, *Technometrics*, **14**(3), 679–691.

[26] WANG, B.X.; YU, K. and SHENG, Z. (2014). New inference for constant-stress accelerated life tests with Weibull distribution and progressively type-II censoring, *IEEE Transactions on Reliability*, **63**(3), 807–815.

Computational Approach Test using Likelihood Based Tests for the Equality of Inverse Gaussian Means

Authors: GAMZE GÜVEN [iD] ✉
         – Department of Statistics, Eskisehir Osmangazi University,
            Turkey
            gamzeguven@ogu.edu.tr

         HATICE ŞAMKAR [iD]
         – Department of Statistics, Eskisehir Osmangazi University,
            Turkey
            hfidan@ogu.edu.tr

         FIKRI GÖKPINAR [iD]
         – Department of Statistics, Gazi University,
            Turkey
            fikri@gazi.edu.tr

Abstract:

• In this study, we propose three different test procedures by plugging the Wald (W), score (S) and likelihood ratio (LR) statistics into the computational approach test (CAT) to test the equality of inverse Gaussian means when the scale parameters are unknown and arbitrary. Restricted maximum likelihood (RML) estimators are used in developing the proposed test procedures. Since the RML estimators cannot be derived in closed-form, the bisection method is used to obtain the numerical solutions. The motivation behind using the CAT procedure is that it can easily be implemented to hypothesis testing problems without knowing the sampling distributions of the test statistics. The proposed and existing procedures are compared in terms of type I error rates and powers via an extensive Monte Carlo simulation study. In addition, a real data set is analyzed for illustration.

✉ Corresponding author

---

1.   INTRODUCTION

The inverse Gaussian (IG) distribution was first introduced by Schrödinger [12] and Smoluchowsky [18] as the probability distribution of the first passage time in Brownian motion. This distribution was named as IG by Tweedie [17] since there exists inverse relationship between the cumulant generating function of the first passage time distribution and that of the normal distribution. The IG distribution is also known as Wald's distribution, especially in Russian literature.

The probability density function (pdf) of the $IG(\mu, \lambda)$ distribution is defined as

$$(1.1) \qquad f(x; \mu, \lambda) = \sqrt{\frac{\lambda}{2x^3\pi}} exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right), \quad x > 0, \quad \mu > 0, \quad \lambda > 0,$$

where λ is a scale parameter and μ is the mean of the distribution.

The $IG(\mu, \lambda)$ distribution has the following attractive properties (see, Tian [16]; Shi and Lv [14]; Seshadri [13]):

- it is closed under convolution;
- it is suitable to model positively-skewed data sets;
- it is the only distribution that shares many elegant properties with Gaussian models among the distributions used to model positively skewed data, see for example Tian [16].

Due to the aforementioned properties, the IG distribution has been widely used in various fields such as cardiology, pharmacokinetics, linguistics, employment services, mathematical finance, demography, hydrology, management sciences, etc. For example, Chhikara and Folks [1] considered the use of the IG distribution for a lifetime model. Folks and Chhikara [5] reviewed the development of the IG distribution and of statistical methods based on it (Chhikara and Folks [2]). Doksum and Hbyland [3] developed models for variable-stress-accelerated life testing experiments based on Wiener processes and the IG distribution. Takagi *et al.* [15] calculated the percentiles of the IG distribution and considered the application of IG to occupational exposure data. Durham and Padgett [4] used IG models to develop a new general method based on cumulative damage for describing the failure of a system. Mudholkar and Tian [10] presented an entropy characterization of the IG family. Kara *et al.* [7] considered the statistical inference problem for the geometric process when the distribution of the first occurrence time is IG. Punzo [11] considered models based on the IG distribution for the problem of fitting the distribution of insurance and economic data.

Testing the equality of means of k independent IG populations is one of the common problems in statistics. The null and alternative hypotheses for this problem are defined as

$$(1.2) \qquad H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \quad \text{and} \quad H_1 : \text{not all } \mu_i\text{'s are equal},$$

respectively. Here, μ_i $(i = 1, ..., k)$ is the mean of the i-th population.

To test the null hypothesis against the alternative, analysis of reciprocals (ANORE) F test is used under the assumption of homogeneity of scale parameters. This assumption is

not always valid in most of the real-life problems, therefore, in recent years, there have been many studies testing the equality of several IG means under the assumption of heterogeneity of scale parameters. For example, Tian [16] developed a test based on the generalized p-value (GP) approach to test the equality of several IG means. Ma and Tian [9] proposed a parametric bootstrap (PB) approach and compared it with the GP approach in terms of the type I error rates. Shi and Lv [14] defined a new generalized pivot quantity and the generalized p-value based on this pivot. Gökpınar *et al.* [6] proposed a computational approach test (CAT) using a simple test statistic to assess the equality of several IG means under heterogeneity of scale parameters. Different from Gökpınar *et al.* [6], in this study, we propose three different test procedures by plugging the Wald (W), score (S) and likelihood ratio (LR) test statistics into CAT. W, S and LR test statistics are asymptotically equivalent. However, they differ in small samples and therefore their type I errors are different from the nominal level for small samples in general. Consequently, in this study we incorporate W, S and LR test statistics into CAT to improve their performances in terms of Type I error rates. Restricted maximum likelihood (RML) estimators are used in developing the proposed test procedures. In this study, in contrast to Ma and Tian [9] who derived RML estimators of the parameters by using expectation-maximization (EM) algorithm, the bisection method is used to obtain the numerical solutions for the RML estimators by maximizing the profile likelihood function under the null hypothesis.

This article is organized as follows. In Section 2, maximum likelihood (ML) and RML estimators for the parameters of the IG distribution are obtained. In Section 3, the PB approach proposed by Ma and Tian [9] and GP approach proposed by Tian [16] are briefly reviewed. In Section 4, W, S and LR test statistics are defined and the CAT procedure is explained for testing the equality of IG means under heterogeneity of scale parameters. In Section 5, simulated Type I error rates and powers of the existing and proposed tests are presented. Proposed tests are illustrated by using a real data set in Section 6. Concluding remarks are given in Section 7.

2. MAXIMUM LIKELIHOOD AND RESTRICTED MAXIMUM LIKELIHOOD ESTIMATORS

Let X_{ij}, $i = 1, ..., k; j = 1, ..., n_i$ be k independent random samples from $\mathrm{IG}(\mu_i, \lambda_i)$. Here, X_{ij} represents the j-th observation of the i-th population and n_i denotes the number of observations in the i-th population. Then the log-likelihood function ($\ln L_1$) under the unrestricted model (H_1) is obtained as follows:

$$(2.1) \qquad \ln L_1 = \sum_{i=1}^{k} \frac{n_i}{2} (\ln \lambda_i - \ln(2\pi)) - \frac{3}{2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \ln X_{ij} - \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{\lambda_i (X_{ij} - \mu_i)^2}{2\mu_i^2 X_{ij}}.$$

To obtain the ML estimators of the parameters μ_i and λ_i, we take the derivatives of $\ln L_1$ with respect to the unknown parameters and then equate them to zero as follows:

$$(2.2) \qquad \frac{\partial \ln L_1}{\partial \mu_i} = \sum_{j=1}^{n_i} \frac{\lambda_i (X_{ij} - \mu_i)}{\mu_i^3} = 0$$

and

$$(2.3) \qquad \frac{\partial \ln L_1}{\partial \lambda_i} = \frac{n_i}{\lambda_i} - \sum_{j=1}^{n_i} \frac{(X_{ij} - \mu_i)^2}{\mu_i^2 X_{ij}} = 0.$$

Solutions of these likelihood equations in (2.2) and (2.3) are the ML estimators of the parameters μ_i and λ_i. They are obtained as

$$(2.4) \qquad \hat{\mu}_i = \bar{X}_i, \quad i = 1, ..., k,$$

and

$$(2.5) \qquad \hat{\lambda}_i = \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \left(1/X_{ij} - 1/\bar{X}_i \right) \right)^{-1} = \left(\overline{X_i^{-1}} - (\bar{X}_i)^{-1} \right)^{-1}, \quad i = 1, ..., k,$$

respectively. Here, $\overline{X_i^{-1}} = (1/n_i) \sum_{j=1}^{n_i} (1/X_{ij})$. Similarly, under the restricted model (H_0), the log-likelihood function ($\ln L_0$) is obtained as

$$(2.6) \qquad \ln L_0 = \sum_{i=1}^{k} \frac{n_i}{2} (\ln \lambda_i - \ln(2\pi)) - \frac{3}{2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \ln X_{ij} - \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{\lambda_i (X_{ij} - \mu)^2}{2\mu^2 X_{ij}}.$$

By taking derivatives of $\ln L_0$ with respect to the parameters μ and λ_i and equating them to zero, the following likelihood equations are

$$(2.7) \qquad \frac{\partial \ln L_0}{\partial \mu} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(\frac{\lambda_i X_{ij}}{\mu^3} - \frac{\lambda_i}{\mu^2} \right) = 0$$

and

$$(2.8) \qquad \frac{\partial \ln L_0}{\partial \lambda_i} = \frac{n_i}{2\lambda_i} - \sum_{j=1}^{n_i} \frac{(X_{ij} - \mu)^2}{2\mu^2 X_{ij}} = 0,$$

respectively. Solutions of likelihood equations in (2.7) and (2.8) are called as RML estimators. RML estimators of the parameters λ_i and μ are

$$(2.9) \qquad \tilde{\lambda}_i = \frac{n_i \mu^2}{\sum_{j=1}^{n_i} \frac{(X_{ij} - \mu)^2}{X_{ij}}}, \quad i = 1, ..., k,$$

and

$$(2.10) \qquad \tilde{\mu} = \frac{\sum_{i=1}^{k} n_i \lambda_i \bar{X}_i}{\sum_{i=1}^{k} n_i \lambda_i},$$

respectively.

It is obvious from equations (2.9) and (2.10) that the estimators of the unknown parameters have no explicit analytical solutions since there exists μ in the estimator of the parameter λ_i and vice versa. Therefore, the profile likelihood method is used to eliminate the effect of nuisance parameters $(\lambda_1, ..., \lambda_k)$ and consequently to estimate the parameter μ.

The profile log-likelihood function ($\ln L^*$) is obtained by replacing λ_i in equation (2.6) with $\tilde{\lambda}_i$ given in (2.9) as

$$(2.11) \qquad \ln L^* = \frac{1}{2} \sum_{i=1}^{k} n_i \left(\ln(n_i \mu^2) - \ln\left(\sum_{j=1}^{n_i} (X_{ij} - \mu)^2 / X_{ij} \right) \right) + \text{constant}.$$

Then we take the derivative of the profile log-likelihood function with respect to the parameter μ as follows:

$$(2.12) \qquad \frac{\partial \ln L^*}{\partial \mu} = \sum_{i=1}^{k} n_i \left(\frac{1}{\mu} + \frac{\sum_{j=1}^{n_i} (X_{ij} - \mu)/X_{ij}}{\sum_{j=1}^{n_i} (X_{ij} - \mu)^2 / X_{ij}} \right).$$

After rearranging the (2.12) and equating it to zero, we obtain the following equality

$$(2.13) \qquad \sum_{i=1}^{k} \frac{n_i(\bar{X}_i - \mu)}{\mu(\bar{X}_i - 2\mu + \mu^2 \overline{X_i^{-1}})} = 0.$$

By solving (2.13), the restricted profile maximum likelihood (RPML) estimator of μ denoted by $\tilde{\mu}^*$ is obtained. Finally, by incorporating $\tilde{\mu}^*$ into (2.9), we obtain the RPML estimator of λ_i as

$$(2.14) \qquad \tilde{\lambda}_i^* = \frac{n_i(\tilde{\mu}^*)^2}{\sum_{j=1}^{n_i} \frac{(X_{ij} - \tilde{\mu}^*)^2}{X_{ij}}}, \quad i = 1, ..., k.$$

Unfortunately, the RPML estimator of μ has no closed form. Therefore, the bisection method is used to obtain a numerical solution for the estimate value of the parameter μ. It requires an interval in which a root of the given equation must lie. Since the parameter μ is the common mean, the root of the given equation, i.e., the estimate of the parameter μ, must lie between the smallest and the largest group means. Hence, the bisection method always converges to a root and the RPML estimate of the parameter μ is obtained.

3.   REVIEWING SOME EXISTING TESTS

In this section, we briefly describe the GP and PB approaches proposed by Tian[16] and Ma and Tian [9], respectively.

3.1. Generalized p-value (GP) approach

Tian [16] calculated the p-value based on the GP approach using the following algorithm.

Step 1: $R_{\lambda_i} = \lambda_i V_i / v_i \sim \chi^2_{(n_i-1)} / v_i$, $i = 1, ..., k$, which is a generalized pivot for λ_i is generated and $R_{\hat{\mu}} = \dfrac{\sum\limits_{i=1}^{k} n_i R_{\lambda_i} \bar{x}_i}{\sum\limits_{i=1}^{k} n_i R_{\lambda_i}}$ is calculated. Here, $V_i = \sum\limits_{j=1}^{n_i} (X_{ij}^{-1} - \bar{X}_i^{-1})$, v_i is the observed value of V_i and $\bar{x}_i = \sum\limits_{j=1}^{n_i} x_{ij} / n_i$.

Step 2: $q = \sum\limits_{i=1}^{k} n_i R_{\lambda_i} (\bar{x}_i^{-1} - R_{\hat{\mu}}^{-1})$ is calculated.

Step 3: $Q \sim \chi^2_{(k-1)}$ is generated.

Step 4: Step 2 and Step 3 are repeated m times.

Step 5: Monte Carlo estimate of the p-value for testing (1.2) is calculated as $\hat{p} = \#(Q \geq q)/m$.

Step 6: If $\hat{p} < \alpha$, then H_0 given in (1.2) is rejected.

3.2. Parametric Bootstrap (PB) approach

Ma and Tian [9] calculated the p-value based on the PB approach using the following algorithm.

Step 1: For a given data set x_{ij}, v_i and $\bar{x}_i$ are calculated. Here, v_i is the observed value of $V_i = \sum\limits_{j=1}^{n_i} (X_{ij}^{-1} - \bar{X}_i^{-1})$ and $\bar{X}_i = \sum\limits_{j=1}^{n_i} X_{ij} / n_i$, $i = 1, ..., k$; $j = 1, ..., n_i$. Then RML estimates $\tilde{\mu}, \tilde{\lambda}_1, \tilde{\lambda}_2, ..., \tilde{\lambda}_k$ are calculated via EM algorithm. Based on these estimate values $Q_{B0} = \sum\limits_{i=1}^{k} n_i \tilde{\lambda}_i (\bar{x}_i^{-1} - \tilde{\mu}^{-1})$ is computed.

Step 2: $\bar{X}_{Bi} \sim IG\left(\sum\limits_{i=1}^{k} n_i \hat{\lambda}_i \bar{x}_i / \sum\limits_{i=1}^{k} n_i \hat{\lambda}_i, n_i \hat{\lambda}_i \right)$ and $\lambda_{Bi} \sim \chi^2_{(n_i-1)} / v_i, i = 1, ..., k$ are generated independently. Here $\hat{\lambda}_i = n_i / v_i$.

Step 3: $Q_B = \sum\limits_{i=1}^{k} n_i \lambda_{Bi} \left(\dfrac{1}{\bar{X}_{Bi}} - \dfrac{1}{\hat{\mu}_B} \right)$ is computed. Here, $\hat{\mu}_B = \dfrac{\sum\limits_{i=1}^{k} n_i \lambda_{Bi} \bar{X}_{Bi}}{\sum\limits_{i=1}^{k} n_i \lambda_{Bi}}$.

Step 4: Step 2 and Step 3 are repeated m times and $Q_B^{(l)}$, $l = 1, ..., m$, are obtained.

Step 5: Monte Carlo estimate of the p-value for testing (1.2) is calculated as $\hat{p} = \#(Q_B^{(l)} \geq Q_{B0})/m$.

Step 6: If $\hat{p} < \alpha$, then H_0 given in (1.2) is rejected.

4. THE PROPOSED CAT PROCEDURE USING LIKELIHOOD BASED TEST STATISTICS

In this section, the likelihood based test statistics W, S and LR are derived for testing the equality of IG means and then they are plugged into the CAT procedure.

4.1. Wald (W) test statistic

The test statistic W is defined as

(4.1) $$W = (\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}^*)' \left(I^{\boldsymbol{\mu},\boldsymbol{\mu}}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\lambda}}) \right)^{-1} (\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}^*).$$

Here, $\hat{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\mu}}^*$ are the ML and RPML estimators of mean vector $\boldsymbol{\mu} = (\mu_1, ..., \mu_k)$, respectively. Also,

$$(\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}^*)' = (\hat{\mu}_1 - \tilde{\mu}^*, .., \hat{\mu}_k - \tilde{\mu}^*)$$

and

$$I^{\boldsymbol{\mu},\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \left(I_{\boldsymbol{\mu},\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) - I_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) I_{\boldsymbol{\lambda},\boldsymbol{\lambda}}^{-1}(\boldsymbol{\mu}, \boldsymbol{\lambda}) I_{\boldsymbol{\lambda},\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) \right)^{-1}.$$

To obtain (4.1) the following steps are considered:

(i) By taking the second partial derivatives of the $\ln L_1$ with respect to the parameters, expected information matrix

(4.2) $$I = \begin{bmatrix} I_{\boldsymbol{\mu},\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) & I_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) \\ I_{\boldsymbol{\lambda},\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) & I_{\boldsymbol{\lambda},\boldsymbol{\lambda}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) \end{bmatrix}$$

is obtained. Here,

(4.3) $$I_{\boldsymbol{\mu},\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \begin{bmatrix} \frac{n_1 \lambda_1}{\mu_1^3} & 0 & \cdots & 0 \\ 0 & \frac{n_2 \lambda_2}{\mu_2^3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{n_k \lambda_k}{\mu_k^3} \end{bmatrix},$$

(4.4) $$I_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = I_{\boldsymbol{\lambda},\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

and

(4.5) $$I_{\boldsymbol{\lambda},\boldsymbol{\lambda}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \begin{bmatrix} \frac{n_1}{2\lambda_1^2} & 0 & \cdots & 0 \\ 0 & \frac{n_2}{2\lambda_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{n_k}{2\lambda_k^2} \end{bmatrix}.$$

(ii) Using (4.3), (4.4) and (4.5),

(4.6) $$I^{\boldsymbol{\mu},\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \begin{bmatrix} \frac{\mu_1^3}{n_1 \lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{\mu_2^3}{n_2 \lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\mu_k^3}{n_k \lambda_k} \end{bmatrix}$$

and

$$(4.7) \qquad (I^{\boldsymbol{\mu},\boldsymbol{\mu}}(\boldsymbol{\mu},\boldsymbol{\lambda}))^{-1} = \begin{bmatrix} \frac{n_1\lambda_1}{\mu_1^3} & 0 & \cdots & 0 \\ 0 & \frac{n_2\lambda_2}{\mu_2^3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{n_k\lambda_k}{\mu_k^3} \end{bmatrix}$$

are obtained.

(iii) By substituting $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\lambda}})$ into the (4.7), we get

$$(4.8) \qquad \left(I^{\boldsymbol{\mu},\boldsymbol{\mu}}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\lambda}})\right)^{-1} = \text{diag}(n_i\hat{\lambda}_i/\hat{\mu}_i^3) \quad i = 1, ..., k.$$

(iv) The test statistic W is derived as follows:

$$(4.9) \qquad W = \sum_{i=1}^{k} \frac{n_i\hat{\lambda}_i(\hat{\mu}_i - \tilde{\mu}^*)^2}{\hat{\mu}_i^3}.$$

4.2. Score (S) test statistic

The test statistic S is defined as

$$(4.10) \qquad S = \boldsymbol{U}'_{\boldsymbol{\mu}}(\tilde{\boldsymbol{\mu}}^*, \tilde{\boldsymbol{\lambda}}^*)\left(I^{\boldsymbol{\mu},\boldsymbol{\mu}}(\tilde{\boldsymbol{\mu}}^*, \tilde{\boldsymbol{\lambda}}^*)\right)\boldsymbol{U}_{\boldsymbol{\mu}}(\tilde{\boldsymbol{\mu}}^*, \tilde{\boldsymbol{\lambda}}^*).$$

Here, $\tilde{\boldsymbol{\mu}}^*$ and $\tilde{\boldsymbol{\lambda}}^*$ are the RPML estimators of the vector of means $\boldsymbol{\mu}=(\mu_1, ..., \mu_k)$ and the vector of scale parameters $\boldsymbol{\lambda}=(\lambda_1, ..., \lambda_k)$. $\boldsymbol{U}_{\boldsymbol{\mu}}(\tilde{\boldsymbol{\mu}}^*, \tilde{\boldsymbol{\lambda}}^*)$ is the value of the vector of score function $\boldsymbol{U}_{\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\lambda})$ at the point $(\tilde{\boldsymbol{\mu}}^*, \tilde{\boldsymbol{\lambda}}^*)=(\tilde{\mu}^*, ..., \tilde{\mu}^*, \tilde{\lambda}_1^*, ..., \tilde{\lambda}_k^*)$. To obtain the (4.10), the following steps are considered.

(i) By taking the first partial derivatives of the $\ln L_1$ with respect to the parameters

$$(4.11) \qquad \boldsymbol{U}'_{\boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \left[\frac{n_1\lambda_1(\bar{X}_1 - \mu_1)}{\mu_1^3}, ..., \frac{n_k\lambda_k(\bar{X}_k - \mu_k)}{\mu_k^3}\right]$$

is obtained.

(ii) By substituting $(\tilde{\boldsymbol{\mu}}^*, \tilde{\boldsymbol{\lambda}}^*)$ into the (4.11), we get

$$(4.12) \qquad \boldsymbol{U}'_{\boldsymbol{\mu}}(\tilde{\boldsymbol{\mu}}^*, \tilde{\boldsymbol{\lambda}}^*) = \left[\frac{n_1\tilde{\lambda}_1^*(\bar{X}_1 - \tilde{\mu}^*)}{\tilde{\mu}^{*3}}, ..., \frac{n_k\tilde{\lambda}_k^*(\bar{X}_k - \tilde{\mu}^*)}{\tilde{\mu}^{*3}}\right].$$

(iii) By substituting $(\tilde{\boldsymbol{\mu}}^*, \tilde{\boldsymbol{\lambda}}^*)$ into the (4.6), we get

$$(4.13) \qquad \left(I^{\boldsymbol{\mu},\boldsymbol{\mu}}(\tilde{\boldsymbol{\mu}}^*, \tilde{\boldsymbol{\lambda}}^*)\right) = \text{diag}(\tilde{\mu}^{*3}/n_i\tilde{\lambda}_i^*), \quad i = 1, ..., k.$$

(iv) The test statistic S is derived as follows:

$$(4.14) \qquad S = \sum_{i=1}^{k} \frac{n_i\tilde{\lambda}_i^*(\hat{\mu}_i - \tilde{\mu}^*)^2}{\tilde{\mu}^{*3}}.$$

4.3. Likelihood ratio (LR) test statistic

The test statistic LR is derived as

$$LR = 2[\ln L_1(\hat{\mu}_1, ..., \hat{\mu}_k; \hat{\lambda}_1, ..., \hat{\lambda}_k) - \ln L_0(\tilde{\mu}^*, ..., \tilde{\mu}^*; \tilde{\lambda}_1^*, ..., \tilde{\lambda}_k^*)]$$

(4.15)
$$= \sum_{i=1}^{k} n_i \ln \frac{\hat{\lambda}_i}{\tilde{\lambda}_i^*}.$$

It should be noted that W, S and LR test statistics are approximated by a chi-square distribution with $(k-1)$ degrees of freedom under H_0 and their values are asymptotically equivalent to each other. However, the mentioned approximation to the distributions of the test statistics may not be accurate for small sample sizes. This problem is also valid even for moderately large sample sizes. To eliminate this problem, likelihood-based test statistics, which are the most appropriate for PB methods can be used. The CAT procedure is a special case of PB methods; therefore, we incorporate the likelihood based test statistics W, S and LR into the CAT and call these tests $CATW$, $CATS$ and $CATLR$, respectively.

The algorithm of the CAT procedure based on T (T is any of W, S or LR) test statistic is as follows:

Step 1: The observed value of the test statistic T, i.e., T_0, is calculated.

Step 2: RPML estimators $\tilde{\mu}^*$ and $\tilde{\lambda}_i^*$, $i = 1, ..., k$, are computed.

Step 3: Artificial samples X_{ij}, $i = 1, ..., k; j = 1, ..., n_i$ from $IG(\tilde{\mu}^*, \tilde{\lambda}_i^*)$ are generated under H_0.

Step 4: For a large number of times, say m, step 3 is repeated. For each of the replicated samples, the values of the test statistic $T^{(l)}, l = 1, ..., m$, are calculated.

Step 5: Monte Carlo estimate of the p-value for testing (1.2) is calculated as $\hat{p} = \#(T^{(l)} \geq T_0)/m$.

Step 6: If $\hat{p} < \alpha$, then H_0 given in (1.2) is rejected.

---

5. SIMULATION STUDY

In this section, the performances of the proposed tests $CATW$, $CATS$ and $CATLR$ are compared with the GP approach proposed by Tian [16], the PB approach proposed by Ma and Tian [9] and the CAT approach proposed by Gökpınar *et al.* [6] with respect to the estimated Type I error rate and power criteria via the MATLAB environment under the specified nominal level $\alpha = 0.050$. In comparing the performances of the proposed tests and the existing tests, 5,000 random samples are generated from the $IG(\mu_i, \lambda_i)$, $i = 1, ..., k$ distribution and $m = 5,000$ Monte Carlo runs are used for each of the samples. In the simulation study the following setup is used.

	Sample Sizes		Parameter Values	
	(n_1, n_2, n_3)		$(\lambda_1, \lambda_2, \lambda_3)$	(μ_1, μ_2, μ_3)
$k=3$	(8, 8, 8) (15, 15, 15) (30, 30, 30) (8, 10, 12) (10, 20, 30)		(10, 11, 12) (10, 12, 14) (10, 15, 20)	(2, 2.25, 2.5) (2, 2.5, 3) (2, 3, 4)
	$(n_1, n_2, n_3, n_4, n_5)$		$(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$	$(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$
$k=5$	(8, 8, 8, 8, 8) (15, 15, 15, 15, 15) (30, 30, 30, 30, 30) (8, 8, 10, 12, 12) (10, 10, 20, 30, 30)		(10, 10, 11, 12, 12) (10, 10, 12, 14, 14) (10, 10, 15, 20, 20)	(2, 2, 2.25, 2.5, 2.5) (2, 2, 2.5, 3, 3) (2, 2, 3, 4, 4)
	$(n_1, n_2, n_3, n_4, n_5, n_6, n_7)$		$(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7)$	$(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7)$
$k=7$	(8, 8, 8, 8, 8, 8, 8) (15, 15, 15, 15, 15, 15, 15) (30, 30, 30, 30, 30, 30, 30) (8, 8, 10, 10, 10, 12, 12) (10, 10, 20, 20, 20, 30, 30)		(10, 10, 11, 11, 11, 12, 12) (10, 10, 12, 12, 12, 14, 14) (10, 10, 15, 15, 15, 20, 20)	(2, 2, 2.25, 2.25, 2.25, 3, 3) (2, 2, 2.5, 2.5, 2.5, 3, 3) (2, 2, 3, 3, 3, 4, 4)

The leftmost column spans all rows with "Number of Groups".

The estimated type I error rates for the proposed and the existing tests are presented in Tables 1–3.

Table 1: Simulated Type I error rates for the $CATW$, $CATS$, $CATLR$, GP, PB and CAT tests when $k=3$.

$(\lambda_1, \lambda_2, \lambda_3)$	$CATW$	$CATS$	$CATLR$	GP	PB	CAT
$(n_1, n_2, n_3) = (8, 8, 8)$						
(10, 11, 12)	0.0454	0.0486	0.0460	0.0400	0.0616	0.0382
(10, 12, 14)	0.0456	0.0454	0.0476	0.0404	0.0606	0.0398
(10, 15, 20)	0.0448	0.0440	0.0432	0.0374	0.0602	0.0358
$(n_1, n_2, n_3) = (15, 15, 15)$						
(10, 11, 12)	0.0558	0.0558	0.0554	0.0524	0.0636	0.0522
(10, 12, 14)	0.0554	0.0582	0.0558	0.0528	0.0658	0.0506
(10, 15, 20)	0.0482	0.0504	0.0492	0.0444	0.0588	0.0460
$(n_1, n_2, n_3) = (30, 30, 30)$						
(10, 11, 12)	0.0466	0.0464	0.0456	0.0458	0.0496	0.0456
(10, 12, 14)	0.0494	0.0498	0.0496	0.0468	0.0514	0.0504
(10, 15, 20)	0.0464	0.0448	0.0442	0.0436	0.0504	0.0442
$(n_1, n_2, n_3) = (8, 10, 12)$						
(10, 11, 12)	0.0482	0.0514	0.0496	0.0460	0.0654	0.0446
(10, 12, 14)	0.0462	0.0478	0.0464	0.0442	0.0636	0.0380
(10, 15, 20)	0.0500	0.0472	0.0498	0.0466	0.0600	0.0398
$(n_1, n_2, n_3) = (10, 20, 30)$						
(10, 11, 12)	0.0522	0.0522	0.0504	0.0494	0.0584	0.0454
(10, 12, 14)	0.0534	0.0546	0.0534	0.0522	0.0592	0.0498
(10, 15, 20)	0.0524	0.0534	0.0552	0.0542	0.0584	0.0452

Table 2: Simulated Type I error rates for the $CATW$, $CATS$, $CATLR$, GP, PB and CAT tests when k=5.

$(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$	$CATW$	$CATS$	$CATLR$	GP	PB	CAT
$(n_1, n_2, n_3, n_4, n_5) = (8, 8, 8, 8, 8)$						
$(10, 10, 11, 12, 12)$	0.0472	0.0486	0.0484	0.0676	0.0480	0.0322
$(10, 10, 12, 14, 14)$	0.0480	0.0486	0.0458	0.0614	0.0498	0.0310
$(10, 10, 15, 20, 20)$	0.0498	0.0514	0.0490	0.0610	0.0508	0.0338
$(n_1, n_2, n_3, n_4, n_5) = (15, 15, 15, 15, 15)$						
$(10, 10, 11, 12, 12)$	0.0536	0.0508	0.0506	0.0522	0.0524	0.0400
$(10, 10, 12, 14, 14)$	0.0460	0.0468	0.0450	0.0502	0.0470	0.0416
$(10, 10, 15, 20, 20)$	0.0530	0.0568	0.0564	0.0620	0.0572	0.0460
$(n_1, n_2, n_3, n_4, n_5) = (30, 30, 30, 30, 30)$						
$(10, 10, 11, 12, 12)$	0.0530	0.0546	0.0548	0.0570	0.0556	0.0496
$(10, 10, 12, 14, 14)$	0.0508	0.0482	0.0486	0.0510	0.0494	0.0452
$(10, 10, 15, 20, 20)$	0.0500	0.0550	0.0536	0.0556	0.0528	0.0466
$(n_1, n_2, n_3, n_4, n_5) = (8, 8, 10, 12, 12)$						
$(10, 10, 11, 12, 12)$	0.0502	0.0510	0.0508	0.0600	0.0538	0.0350
$(10, 10, 12, 14, 14)$	0.0466	0.0494	0.0496	0.0602	0.0500	0.0376
$(10, 10, 15, 20, 20)$	0.0482	0.0472	0.0472	0.0602	0.0468	0.0332
$(n_1, n_2, n_3, n_4, n_5) = (10, 10, 20, 30, 30)$						
$(10, 10, 11, 12, 12)$	0.0516	0.0522	0.0500	0.0602	0.0492	0.0408
$(10, 10, 12, 14, 14)$	0.0516	0.0522	0.0500	0.0612	0.0492	0.0408
$(10, 10, 15, 20, 20)$	0.0486	0.0480	0.0496	0.0634	0.0472	0.0364

Table 3: Simulated Type I error rates for the $CATW$, $CATS$, $CATLR$, GP, PB and CAT tests when k=7.

$(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7)$	$CATW$	$CATS$	$CATLR$	GP	PB	CAT
$(n_1, n_2, n_3, n_4, n_5, n_6, n_7) = (8, 8, 8, 8, 8, 8, 8)$						
$(10, 10, 11, 11, 11, 12, 12)$	0.0470	0.0510	0.0518	0.0774	0.0482	0.0298
$(10, 10, 12, 12, 12, 14, 14)$	0.0482	0.0482	0.0468	0.0716	0.0422	0.0272
$(10, 10, 15, 15, 15, 20, 20)$	0.0474	0.0550	0.0530	0.0770	0.0484	0.0350
$(n_1, n_2, n_3, n_4, n_5, n_6, n_7) = (15, 15, 15, 15, 15, 15, 15)$						
$(10, 10, 11, 11, 11, 12, 12)$	0.0518	0.0522	0.0526	0.0646	0.0516	0.0440
$(10, 10, 12, 12, 12, 14, 14)$	0.0506	0.0526	0.0534	0.0636	0.0516	0.0388
$(10, 10, 15, 15, 15, 20, 20)$	0.0488	0.0518	0.0498	0.0636	0.0492	0.0398
$(n_1, n_2, n_3, n_4, n_5, n_6, n_7) = (30, 30, 30, 30, 30, 30, 30)$						
$(10, 10, 11, 11, 11, 12, 12)$	0.0500	0.0536	0.0516	0.0610	0.0524	0.0462
$(10, 10, 12, 12, 12, 14, 14)$	0.0518	0.0554	0.0562	0.0620	0.0568	0.0494
$(10, 10, 15, 15, 15, 20, 20)$	0.0500	0.0532	0.0524	0.0624	0.0518	0.0454
$(n_1, n_2, n_3, n_4, n_5, n_6, n_7) = (8, 8, 10, 10, 10, 12, 12)$						
$(10, 10, 11, 11, 11, 12, 12)$	0.0434	0.0436	0.0442	0.0656	0.0436	0.0306
$(10, 10, 12, 12, 12, 14, 14)$	0.0486	0.0486	0.0508	0.0708	0.0444	0.0356
$(10, 10, 15, 15, 15, 20, 20)$	0.0516	0.0488	0.0502	0.0742	0.0458	0.0350
$(n_1, n_2, n_3, n_4, n_5, n_6, n_7) = (10, 10, 20, 20, 20, 30, 30)$						
$(10, 10, 11, 11, 11, 12, 12)$	0.0502	0.0576	0.0532	0.0670	0.0534	0.0442
$(10, 10, 12, 12, 12, 14, 14)$	0.0518	0.0500	0.0506	0.0678	0.0474	0.0408
$(10, 10, 15, 15, 15, 20, 20)$	0.0476	0.0484	0.0462	0.0604	0.0472	0.0400

Table 4:   Simulated power values for the $CATW$, $CATS$, $CATLR$, GP,
PB and CAT tests when k=3.

$(\lambda_1, \lambda_2, \lambda_3)$	(μ_1, μ_2, μ_3)	$CATW$	$CATS$	$CATLR$	GP	PB	CAT
			$(n_1, n_2, n_3) = (8,8,8)$				
	$(2, 2.25, 2.5)$	0.1018	0.1028	0.1194	0.1066	***	0.1006
$(10, 11, 12)$	$(2, 2.5, 3)$	0.2120	0.1896	0.2392	0.2200	***	0.2276
	$(2, 3, 4)$	0.5226	0.3314	0.5486	0.5318	***	0.5660
	$(2, 2.25, 2.5)$	0.1096	0.0940	0.1128	0.1034	***	0.1036
$(10, 12, 14)$	$(2, 2.5, 3)$	0.2392	0.1940	0.2600	0.2402	***	0.2510
	$(2, 3, 4)$	0.5608	0.3254	0.5610	0.5430	***	0.5800
	$(2, 2.25, 2.5)$	0.1378	0.0914	0.1260	0.1138	***	0.1160
$(10, 15, 20)$	$(2, 2.5, 3)$	0.3110	0.1958	0.2952	0.2786	***	0.2846
	$(2, 3, 4)$	0.6432	0.3352	0.6238	0.6108	***	0.6222
			$(n_1, n_2, n_3) = (15,15,15)$				
	$(2, 2.25, 2.5)$	0.1806	0.1616	0.1800	0.1750	***	0.1848
$(10, 11, 12)$	$(2, 2.5, 3)$	0.4906	0.4180	0.4814	0.4694	***	0.4908
	$(2, 3, 4)$	0.9016	0.7620	0.8864	0.8834	***	0.9054
	$(2, 2.25, 2.5)$	0.2076	0.1840	0.2062	0.1968	***	0.2020
$(10, 12, 14)$	$(2, 2.5, 3)$	0.5220	0.4462	0.5156	0.5068	***	0.5308
	$(2, 3, 4)$	0.9174	0.7596	0.9010	0.8974	***	0.9228
	$(2, 2.25, 2.5)$	0.2528	0.1918	0.2300	0.2228	0.2546	0.2290
$(10, 15, 20)$	$(2, 2.5, 3)$	0.6060	0.4696	0.5748	0.5678	0.5966	0.5962
	$(2, 3, 4)$	0.9546	0.7968	0.9372	0.9332	0.9414	0.9510
			$(n_1, n_2, n_3) = (30,30,30)$				
	$(2, 2.25, 2.5)$	0.3708	0.3466	0.3656	0.3622	0.3792	0.3734
$(10, 11, 12)$	$(2, 2.5, 3)$	0.8344	0.8036	0.8292	0.8258	0.8410	0.8430
	$(2, 3, 4)$	0.9982	0.9964	0.9986	0.9982	0.9984	0.9982
	$(2, 2.25, 2.5)$	0.3984	0.3564	0.3796	0.3778	0.3966	0.3882
$(10, 12, 14)$	$(2, 2.5, 3)$	0.8678	0.8272	0.8534	0.8502	0.8610	0.8656
	$(2, 3, 4)$	0.9990	0.9952	0.9986	0.9980	0.9986	0.9986
	$(2, 2.25, 2.5)$	0.4872	0.4090	0.4462	0.4420	0.4622	0.4570
$(10, 15, 20)$	$(2, 2.5, 3)$	0.9174	0.8720	0.9024	0.8976	0.9080	0.9080
	$(2, 3, 4)$	0.9998	0.9980	0.9996	0.9998	0.9998	0.9998
			$(n_1, n_2, n_3) = (8,10,12)$				
	$(2, 2.25, 2.5)$	0.1402	0.0964	0.1302	0.1218	***	0.1188
$(10, 11, 12)$	$(2, 2.5, 3)$	0.3066	0.1904	0.2808	0.2668	***	0.2670
	$(2, 3, 4)$	0.6680	0.3410	0.6166	0.6032	***	0.6240
	$(2, 2.25, 2.5)$	0.1492	0.0912	0.1302	0.1240	***	0.1164
$(10, 12, 14)$	$(2, 2.5, 3)$	0.3560	0.2136	0.3212	0.3082	***	0.3094
	$(2, 3, 4)$	0.7032	0.3448	0.6480	0.6402	***	0.6564
	$(2, 2.25, 2.5)$	0.1782	0.0898	0.1392	0.1322	***	0.1262
$(10, 15, 20)$	$(2, 2.5, 3)$	0.3900	0.2064	0.3408	0.3304	***	0.3282
	$(2, 3, 4)$	0.7658	0.3682	0.6958	0.6892	***	0.7072
			$(n_1, n_2, n_3) = (10,20,30)$				
	$(2, 2.25, 2.5)$	0.2448	0.1264	0.1912	0.1878	0.1958	0.1778
$(10, 11, 12)$	$(2, 2.5, 3)$	0.5624	0.3250	0.4854	0.4824	0.4842	0.4698
	$(2, 3, 4)$	0.9112	0.5672	0.8634	0.8592	0.8502	0.8644
	$(2, 2.25, 2.5)$	0.2514	0.1406	0.2032	0.2016	0.2176	0.1886
$(10, 12, 14)$	$(2, 2.5, 3)$	0.5908	0.3352	0.5062	0.5016	0.5132	0.4956
	$(2, 3, 4)$	0.9360	0.5970	0.8832	0.8796	0.8690	0.8874
	$(2, 2.25, 2.5)$	0.2852	0.1472	0.2222	0.2244	0.2364	0.2040
$(10, 15, 20)$	$(2, 2.5, 3)$	0.6412	0.3972	0.5624	0.5626	0.5684	0.5394
	$(2, 3, 4)$	0.9564	0.6716	0.9162	0.9164	0.9066	0.9128

***: The estimated type I error rates of the tests which are greater than or equal to 0.060.

Table 5: Simulated power values for the $CATW$, $CATS$, $CATLR$, GP, PB and CAT tests when k=5.

$(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$	$(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$	$CATW$	$CATS$	$CATLR$	GP	PB	CAT
	$(n_1, n_2, n_3, n_4, n_5) = (8, 8, 8, 8, 8)$						
	$(2, 2, 2.25, 2.5, 2.5)$	0.1126	0.1266	0.1424	***	0.1462	0.1092
$(10, 10, 11, 12, 12)$	$(2, 2, 2.5, 3, 3)$	0.2800	0.2886	0.3530	***	0.3648	0.3104
	$(2, 2, 3, 4, 4)$	0.6882	0.5216	0.7842	***	0.8000	0.7744
	$(2, 2, 2.25, 2.5, 2.5)$	0.1300	0.1106	0.1414	***	0.1428	0.1126
$(10, 10, 12, 14, 14)$	$(2, 2, 2.5, 3, 3)$	0.3320	0.2964	0.3828	***	0.4018	0.3490
	$(2, 2, 3, 4, 4)$	0.7332	0.5032	0.7970	***	0.8110	0.7900
	$(2, 2, 2.25, 2.5, 2.5)$	0.1698	0.1172	0.1614	***	0.1704	0.1304
$(10, 10, 15, 20, 20)$	$(2, 2, 2.5, 3, 3)$	0.4120	0.2690	0.4224	***	0.4414	0.3868
	$(2, 2, 3, 4, 4)$	0.8282	0.4770	0.8436	***	0.8526	0.8390
	$(n_1, n_2, n_3, n_4, n_5) = (15, 15, 15, 15, 15)$						
	$(2, 2, 2.25, 2.5, 2.5)$	0.2640	0.2408	0.2600	0.2734	0.2642	0.2522
$(10, 10, 11, 12, 12)$	$(2, 2, 2.5, 3, 3)$	0.6962	0.6596	0.7048	0.7216	0.7186	0.7066
	$(2, 2, 3, 4, 4)$	0.9816	0.9676	0.9890	0.9900	0.9900	0.9906
	$(2, 2, 2.25, 2.5, 2.5)$	0.2790	0.2460	0.2778	0.2974	0.2876	0.2714
$(10, 10, 12, 14, 14)$	$(2, 2, 2.5, 3, 3)$	0.7140	0.6614	0.7232	0.7384	0.7310	0.7312
	$(2, 2, 3, 4, 4)$	0.9918	0.9602	0.9904	0.9922	0.9914	0.9936
	$(2, 2, 2.25, 2.5, 2.5)$	0.3378	0.2542	0.3128	***	0.3220	0.2982
$(10, 10, 15, 20, 20)$	$(2, 2, 2.5, 3, 3)$	0.8102	0.6916	0.7866	***	0.7998	0.7896
	$(2, 2, 3, 4, 4)$	0.9950	0.9682	0.9944	***	0.9954	0.9960
	$(n_1, n_2, n_3, n_4, n_5) = (30, 30, 30, 30, 30)$						
	$(2, 2, 2.25, 2.5, 2.5)$	0.5174	0.5080	0.5218	0.5356	0.5324	0.5226
$(10, 10, 11, 12, 12)$	$(2, 2, 2.5, 3, 3)$	0.9670	0.9626	0.9698	0.9708	0.9702	0.9714
	$(2, 2, 3, 4, 4)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$(2, 2, 2.25, 2.5, 2.5)$	0.5746	0.5482	0.5730	0.5822	0.5812	0.5740
$(10, 10, 12, 14, 14)$	$(2, 2, 2.5, 3, 3)$	0.9810	0.9728	0.9804	0.9804	0.9810	0.9808
	$(2, 2, 3, 4, 4)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$(2, 2, 2.25, 2.5, 2.5)$	0.6656	0.5830	0.6274	0.6366	0.6332	0.6196
$(10, 10, 15, 20, 20)$	$(2, 2, 2.5, 3, 3)$	0.9922	0.9832	0.9906	0.9912	0.9914	0.9896
	$(2, 2, 3, 4, 4)$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$(n_1, n_2, n_3, n_4, n_5) = (8, 8, 10, 12, 12)$						
	$(2, 2, 2.25, 2.5, 2.5)$	0.1718	0.1184	0.1612	***	0.1616	0.1292
$(10, 10, 11, 12, 12)$	$(2, 2, 2.5, 3, 3)$	0.4234	0.2792	0.4190	***	0.4180	0.3956
	$(2, 2, 3, 4, 4)$	0.8606	0.4846	0.8500	***	0.8470	0.8520
	$(2, 2, 2.25, 2.5, 2.5)$	0.1942	0.1210	0.1726	***	0.1704	0.1420
$(10, 10, 12, 14, 14)$	$(2, 2, 2.5, 3, 3)$	0.4752	0.2800	0.4500	***	0.4508	0.4248
	$(2, 2, 3, 4, 4)$	0.8762	0.4718	0.8636	***	0.8566	0.8616
	$(2, 2, 2.25, 2.5, 2.5)$	0.2184	0.1098	0.1780	***	0.1708	0.1510
$(10, 10, 15, 20, 20)$	$(2, 2, 2.5, 3, 3)$	0.5506	0.2850	0.5000	***	0.4940	0.4550
	$(2, 2, 3, 4, 4)$	0.9298	0.4582	0.9026	***	0.8962	0.9020
	$(n_1, n_2, n_3, n_4, n_5) = (10, 10, 20, 30, 30)$						
	$(2, 2, 2.25, 2.5, 2.5)$	0.3274	0.1516	0.2532	***	0.2422	0.2304
$(10, 10, 11, 12, 12)$	$(2, 2, 2.5, 3, 3)$	0.7514	0.4418	0.6770	***	0.6580	0.6558
	$(2, 2, 3, 4, 4)$	0.9912	0.7222	0.9792	***	0.9740	0.9822
	$(2, 2, 2.25, 2.5, 2.5)$	0.3394	0.1642	0.2612	***	0.2502	0.2418
$(10, 10, 12, 14, 14)$	$(2, 2, 2.5, 3, 3)$	0.7550	0.4274	0.6740	***	0.6594	0.6650
	$(2, 2, 3, 4, 4)$	0.9930	0.7252	0.9848	***	0.9788	0.9870
	$(2, 2, 2.25, 2.5, 2.5)$	0.3688	0.1674	0.2928	***	0.2798	0.2680
$(10, 10, 15, 20, 20)$	$(2, 2, 2.5, 3, 3)$	0.8206	0.4834	0.7488	***	0.7340	0.7312
	$(2, 2, 3, 4, 4)$	0.9944	0.7646	0.9896	***	0.9852	0.9900

***: The estimated type I error rates of the tests which are greater than or equal to 0.060.

Table 6: Simulated power values for the $CATW$, $CATS$, $CATLR$, GP, PB and CAT tests when $k=7$.

$(\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5,\lambda_6,\lambda_7)$	$(\mu_1,\mu_2,\mu_3,\mu_4,\mu_5,\mu_6,\mu_7)$	$CATW$	$CATS$	$CATLR$	GP	PB	CAT
$(n_1,n_2,n_3,n_4,n_5,n_6,n_7)=(8,8,8,8,8,8,8)$
$(10,10,11,11,11,12,12)$	$(2,2,2.25,2.25,2.25,3,3)$	0.1082	0.1154	0.1292	***	0.1246	0.0954
	$(2,2,2.5,2.5,2.5,3,3)$	0.2386	0.2548	0.2934	***	0.2856	0.2668
	$(2,2,3,3,3,4,4)$	0.5970	0.4976	0.6876	***	0.6808	0.6956
$(10,10,12,12,12,14,14)$	$(2,2,2.25,2.25,2.25,3,3)$	0.1122	0.1064	0.1234	***	0.1176	0.0960
	$(2,2,2.5,2.5,2.5,3,3)$	0.2570	0.2582	0.3110	***	0.3018	0.2844
	$(2,2,3,3,3,4,4)$	0.6514	0.5010	0.7162	***	0.7042	0.7330
$(10,10,15,15,15,20,20)$	$(2,2,2.25,2.25,2.25,3,3)$	0.1412	0.1082	0.1392	***	0.1312	0.1038
	$(2,2,2.5,2.5,2.5,3,3)$	0.3246	0.2614	0.3580	***	0.3540	0.3258
	$(2,2,3,3,3,4,4)$	0.7482	0.5286	0.7824	***	0.7782	0.8010
$(n_1,n_2,n_3,n_4,n_5,n_6,n_7)=(15,15,15,15,15,15,15)$
$(10,10,11,11,11,12,12)$	$(2,2,2.25,2.25,2.25,3,3)$	0.1898	0.1958	0.2112	***	0.2044	0.1914
	$(2,2,2.5,2.5,2.5,3,3)$	0.5696	0.5600	0.6168	***	0.6128	0.6204
	$(2,2,3,3,3,4,4)$	0.9702	0.9168	0.9752	***	0.9746	0.9832
$(10,10,12,12,12,14,14)$	$(2,2,2.25,2.25,2.25,3,3)$	0.2094	0.1966	0.2176	***	0.2210	0.068
	$(2,2,2.5,2.5,2.5,3,3)$	0.6228	0.5796	0.6476	***	0.6504	0.6588
	$(2,2,3,3,3,4,4)$	0.9804	0.9308	0.9822	***	0.9826	0.9870
$(10,10,15,15,15,20,20)$	$(2,2,2.25,2.25,2.25,3,3)$	0.2812	0.2262	0.2672	***	0.2682	0.2462
	$(2,2,2.5,2.5,2.5,3,3)$	0.7222	0.6422	0.7256	***	0.7324	0.7340
	$(2,2,3,3,3,4,4)$	0.9928	0.9510	0.9896	***	0.9896	0.9946
$(n_1,n_2,n_3,n_4,n_5,n_6,n_7)=(30,30,30,30,30,30,30)$
$(10,10,11,11,11,12,12)$	$(2,2,2.25,2.25,2.25,3,3)$	0.4520	0.4534	0.4722	***	0.4678	0.4692
	$(2,2,2.5,2.5,2.5,3,3)$	0.9400	0.9262	0.9404	***	0.9410	0.9474
	$(2,2,3,3,3,4,4)$	1.0000	0.9998	1.0000	***	1.0000	1.0000
$(10,10,12,12,12,14,14)$	$(2,2,2.25,2.25,2.25,3,3)$	0.4894	0.4600	0.4892	***	0.4876	0.4892
	$(2,2,2.5,2.5,2.5,3,3)$	0.9640	0.9464	0.9604	***	0.9618	0.9652
	$(2,2,3,3,3,4,4)$	1.0000	0.9998	1.0000	***	1.0000	1.0000
$(10,10,15,15,15,20,20)$	$(2,2,2.25,2.25,2.25,3,3)$	0.5968	0.5318	0.5752	***	0.5760	0.5670
	$(2,2,2.5,2.5,2.5,3,3)$	0.9790	0.9656	0.9752	***	0.9750	0.9794
	$(2,2,3,3,3,4,4)$	1.0000	1.0000	1.0000	***	1.0000	1.0000
$(n_1,n_2,n_3,n_4,n_5,n_6,n_7)=(8,8,10,10,10,12,12)$
$(10,10,11,11,11,12,12)$	$(2,2,2.25,2.25,2.25,3,3)$	0.1366	0.1086	0.1316	***	0.1280	0.1112
	$(2,2,2.5,2.5,2.5,3,3)$	0.3630	0.7380	0.3652	***	0.3506	0.3298
	$(2,2,3,3,3,4,4)$	0.7752	0.5306	0.7788	***	0.7572	0.7912
$(10,10,12,12,12,14,14)$	$(2,2,2.25,2.25,2.25,3,3)$	0.1526	0.1020	0.1320	***	0.1288	0.1114
	$(2,2,2.5,2.5,2.5,3,3)$	0.3752	0.2830	0.3836	***	0.3702	0.3524
	$(2,2,3,3,3,4,4)$	0.8216	0.5490	0.8094	***	0.7850	0.8248
$(10,10,15,15,15,20,20)$	$(2,2,2.25,2.25,2.25,3,3)$	0.1860	0.1162	0.1634	***	0.1534	0.1378
	$(2,2,2.5,2.5,2.5,3,3)$	0.4652	0.3202	0.4510	***	0.4424	0.4148
	$(2,2,3,3,3,4,4)$	0.8720	0.6136	0.8608	***	0.8484	0.8690
$(n_1,n_2,n_3,n_4,n_5,n_6,n_7)=(10,10,20,20,20,30,30)$
$(10,10,11,11,11,12,12)$	$(2,2,2.25,2.25,2.25,3,3)$	0.2714	0.1678	0.2230	***	0.2140	0.2008
	$(2,2,2.5,2.5,2.5,3,3)$	0.6850	0.4904	0.6396	***	0.6248	0.6198
	$(2,2,3,3,3,4,4)$	0.9818	0.8286	0.9698	***	0.9624	0.9722
$(10,10,12,12,12,14,14)$	$(2,2,2.25,2.25,2.25,3,3)$	0.2916	0.1768	0.2434	***	0.2352	0.2214
	$(2,2,2.5,2.5,2.5,3,3)$	0.7168	0.5176	0.6736	***	0.6576	0.6534
	$(2,2,3,3,3,4,4)$	0.9888	0.8438	0.9780	***	0.9680	0.9824
$(10,10,15,15,15,20,20)$	$(2,2,2.25,2.25,2.25,3,3)$	0.3482	0.2162	0.2938	***	0.2826	0.2522
	$(2,2,2.5,2.5,2.5,3,3)$	0.7792	0.6010	0.7482	***	0.7330	0.7146
	$(2,2,3,3,3,4,4)$	0.9966	0.9236	0.9904	***	0.9864	0.9902

***: The estimated type I error rates of the tests which are greater than or equal to 0.060.

It can be seen from Table 1 that the estimated type I error rates of the tests are close to the nominal level, except for the PB approach when the number of observations in the populations are equal to $(n_1, n_2, n_3) = (8, 8, 8)$, $(8, 10, 12)$ and $(15, 15, 15)$. It can be seen from Table 2 that if the sample sizes are small and/or different the estimated type I error rates for the GP approach are larger than the nominal level. As seen from Table 3 that the estimated type I error rates for the GP approach are always greater than 0.060. It should be noted that the estimated type I error rates of the $CATW$, $CATS$, $CATLR$, PB and CAT approaches are close to the nominal level $\alpha = 0.050$ in all cases when the number of groups is moderate to large. The estimated powers of the proposed and the existing tests are presented in Tables 4–6. It should be noted that in Tables 4–6 the estimated type I error rates of the tests which are greater or equal to 0.060 are denoted by ***.

It can be seen from Table 4 that $CATS$ shows the worst performance in all cases. In addition, the $CATW$ outperforms the other tests when the sample sizes are unequal. It can be seen from Table 5 that the performances of the $CATLR$ and PB tests are similar when the sample sizes are small and equal. The $CATW$ test is more powerful than the other tests when the sample sizes are different. Also, $CATW$ shows the best performance when the sample sizes are moderately large and unequal while $CATS$ shows the worst performance. Since the simulation results given in Table 6 show a similar behavior as in Table 5, the same conclusions are drawn from Table 6. Therefore, we do not repeat them for the sake of brevity.

6. REAL DATA EXAMPLE

In this section, the fatigue life data taken from Leiva *et al.* [8] is analyzed to illustrate the implementation of the proposed and the existing tests; see Table 7. It contains fatigue life (T) of 6061-T6 aluminum pieces which were cut parallel to the direction of rolling and oscillating at 18 cycles/s at maximum stress levels of $x_1 = 2.1$, $x_2 = 2.6$ and $x_3 = 3.1$ psi ($\times 10^4$); see Leiva *et al.* [8] for more detailed information. These stress levels denote the groups whose means are to be compared and the sample sizes of them are $n_1 = 101$, $n_2 = 102$ and $n_3 = 101$, respectively.

Firstly we calculate the values of Lilliefors goodness of fit test and the corresponding p-value for each of the three groups to test the assumption that the data are from an IG distribution; see Table 8.

It can be seen from Table 8 that the IG distribution provides a good fit for each group of the fatigue life data, since the corresponding p-values are all greater than the nominal level $\alpha = 0.050$. Then, we compute the ML estimates $(\hat{\mu}_i, \hat{\lambda}_i)$ of the parameters (μ_i, λ_i) and the RML estimates $(\tilde{\mu}, \tilde{\lambda}_i)$ of the parameters (μ, λ_i), $i = 1, 2, 3$, by using the equalities given in Section 2; see Table 9.

The values of the W, S and LR test statistics based on the ML and RML estimates are calculated as 300141.0259, 41.2081 and 645.9193 using the (4.9), (4.14) and (4.15), respectively. In addition, to test the hypothesis of equality of means, the p-values for the existing tests GP and PB and the proposed tests $CATW$, $CATS$ and $CATLR$ are calculated using the algorithms given in Sections 3 and 4, respectively. The p-value for the CAT is calculated

from the algorithm given in Gökpınar *et al.* [6]. $m=5{,}000$ Monte Carlo runs are used in calculating the p-values for the fatigue life data. p-values for the proposed and the existing tests are obtained to be very close to zero. Therefore, the hypothesis of equality of means is rejected for all tests at the significance level $\alpha = 0.050$. It should be noted that all the calculations are made using MATLAB software.

Table 7: Fatigue life of aluminum pieces submitted to the maximum indicated stress level.

stress levels (psi)	
I: 2.1×10^4	370, 706, 716, 746, 785, 797, 844, 855, 858, 886, 886, 930, 960, 988, 999, 1000, 1010, 1016, 1018, 1020, 1055, 1085, 1102, 1102, 1108, 1115, 1120, 1134, 1140, 1199, 1200, 1200, 1203, 1222, 1235, 1238, 1252, 1258, 1262, 1269, 1270, 1290, 1293, 1300, 1310, 1313, 1315, 1330, 1355, 1390, 1416, 1419, 1420, 1420, 1450, 1452, 1475, 1478, 1481, 1485, 1502, 1505, 1513, 1522, 1522, 1530, 1540, 1560, 1567, 1578, 1594, 1602, 1604, 1608, 1630, 1642, 1674, 1730, 1750, 1750, 1763, 1768, 1781, 1782, 1792, 1820, 1868, 1881, 1890, 1893, 1895, 1910, 1923, 1924, 1945, 2023, 2100, 2130, 2215, 2268, 2440
II: 2.6×10^4	233, 258, 268, 276, 290, 310, 312, 315, 318, 321, 321, 329, 335, 336, 338, 338, 342, 342, 342, 344, 349, 350, 350, 351, 351, 352, 352, 356, 358, 358, 360, 362, 363, 366, 367, 370, 370, 372, 372, 374, 375, 376, 379, 379, 380, 382, 389, 389, 395, 396, 400, 400, 400, 403, 404, 406, 408, 408, 410, 412, 414, 416, 416, 416, 420, 422, 423, 426, 428, 432, 432, 433, 433, 437, 438, 439, 439, 443, 445, 445, 452, 456, 456, 460, 464, 466, 468, 470, 470, 473, 474, 476, 476, 486, 488, 489, 490, 491, 503, 517, 540, 560
III: 3.1×10^4	70, 90, 96, 97, 99, 100, 103, 104, 104, 105, 107, 108, 108, 108, 109, 109, 112, 112, 113, 114, 114, 114, 116, 119, 120, 120, 120, 121, 121, 123, 124, 124, 124, 124, 124, 128, 128, 129, 129, 130, 130, 130, 131, 131, 131, 131, 131, 132, 132, 132, 133, 134, 134, 134, 134, 134, 136, 136, 137, 138, 138, 138, 139, 139, 141, 141, 142, 142, 142, 142, 142, 142, 144, 144, 145, 146, 148, 148, 149, 151, 151, 152, 155, 156, 157, 157, 157, 157, 158, 159, 162, 163, 163, 164, 166, 166, 168, 170, 174, 196, 212

Table 8: Lilliefors goodness of fit test for the stress levels (groups) and the corresponding p-values.

	stress levels (psi)	I: 2.1×10^4	II: 2.6×10^4	III: 3.1×10^4
Lilliefors test	Test statistic	0.0699	0.0452	0.0752
	p-value	0.6011	0.9483	0.5154

Table 9: The ML and RML estimates for the parameters μ_i, μ and λ_i.

stress levels	ML		RML
I	$\hat{\mu}_1 = 1400.8$	$\hat{\lambda}_1 = 14222.3$	$\tilde{\lambda}_1 = 13876.9$
II	$\hat{\mu}_2 = 397.9$	$\hat{\lambda}_2 = 15165.8$	$\tilde{\lambda}_2 = 766.8$
III	$\hat{\mu}_3 = 133.7$	$\hat{\lambda}_3 = 4573.4$	$\tilde{\lambda}_3 = 159.4$
common mean			$\tilde{\mu} = 1334.8$

7.   CONCLUSION

In this study, Wald, score and likelihood ratio test statistics are defined for the problem of testing the equality of IG means. Then they are plugged into the computational approach test procedure. The proposed tests are compared with the existing tests according to the estimated type I error rate and power criteria via a Monte Carlo simulation study. The estimated type I error rates for the proposed tests are close to the nominal level $\alpha = 0.050$ in all cases considered in the study. The computational approach test based on Wald test statistic appears to be more powerful than the other tests especially when the sample sizes are not equal among groups.

ACKNOWLEDGMENTS

REFERENCES

[1]   CHHIKARA, R.S. and FOLKS, J.L. (1977). The inverse Gaussian distribution as a lifetime model, *Technometrics*, **19**(4), 461–468.

[2]   CHHIKARA, R.S. and FOLKS, J.L. (1989). *The Inverse Gaussian Distribution*, Dekker, New York.

[3]   DOKSUM, K.A. and HBYLAND, A. (1992). Models for variable-stress accelerated life testing experiments based on wener processes and the inverse gaussian distribution, *Technometrics*, **34**(1), 74–82.

[4]   DURHAM, S.D. and PADGETT, W.J. (1997). Cumulative damage models for system failure with application to carbon fibers and composites, *Technometrics*, **39**(1), 34–44.

[5]   FOLKS, J.L. and CHHIKARA, R.S. (1978). The inverse Gaussian distribution and its statistical application-a review, *Journal of the Royal Statistical Society: Series B (Methodological)*, **40**(3), 263–275.

[6]   GOKPINAR, E.Y.; POLAT, E.; GOKPINAR, F. and GUNAY, S. (2013). A new computational approach for testing equality of inverse Gaussian means under heterogeneity, *Hacettepe Journal of Mathematics and Statistics*, **42**(5), 581–590.

[7]   KARA, M.; AYDOGDU, H. and TURKSEN, O. (2015). Statistical inference for geometric process with the inverse Gaussian distribution, *Journal of Statistical Computation and Simulation*, **85**(16), 3206–3215.

[8]   LEIVA, V.; ROJAS, E.; GALEA, M. and SANHUEZA, A. (2014). Diagnostics in Birnbaum–Saunders accelerated life models with an application to fatigue data, *Applied Stochastic Models in Business and Industry*, **30**(2), 115–131.

[9]   MA, C.X. and TIAN, L. (2009). A parametric bootstrap approach for testing equality of inverse Gaussian means under heterogeneity, *Communications in Statistics – Simulation and Computation*, **38**(6), 1153–1160.

[10]   Mudholkar, G.S. and Tian, L. (2002). An entropy characterization of the inverse Gaussian distribution and related goodness-of-fit test, *Journal of Statistical Planning and Inference*, **102**(2), 211–221.

[11]   Punzo, A. (2019). A new look at the inverse Gaussian distribution with applications to insurance and economic data, *Journal of Applied Statistics*, **46**(7), 1260–1287.

[12]   Schrodinger, E. (1915). Zur theorie der fall-und steigversuche an teilchen mit brownscher bewegung, *Physikalische Zeitschrift*, **16**, 289–295.

[13]   Seshadri, V. (2012). *The Inverse Gaussian Distribution: Statistical Theory and Applications* (Vol. 137), Springer Science & Business Media.

[14]   Shi, J.H. and Lv, J.L. (2012). A new generalized p-value for testing equality of inverse Gaussian means under heterogeneity, *Statistics & Probability Letters*, **82**(1), 96–102.

[15]   Takagi, K.; Kumagai, S.; Matsunaga, I. and Kusaka, Y. (1997). Application of inverse Gaussian distribution to occupational exposure data, *The Annals of Occupational Hygiene*, **41**(5), 505–514.

[16]   Tian, L. (2006). Testing equality of inverse Gaussian means under heterogeneity, based on generalized test variable, *Computational Statistics & Data Analysis*, **51**(2), 1156–1162.

[17]   Tweedie, M.C. (1945). Inverse statistical variates, *Nature*, **155**(3937), 453–453.

[18]   Von Smoluchowski, M. (1915). Notiz uiber die Berechnung der Brownschen Molekular-bewegung bei der Ehrenhaft-Millikanschen Versuchsanordning, *Physikalische Zeitschrift*, **16**, 318–321.

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