Probability and Statistics
Concepts and Applications

Derek Beaven

Probability and Statistics: Concepts and Applications

Probability and Statistics: Concepts and Applications

Edited by
Derek Beaven

Contents

Preface

Probability and Statistics are two closely related sub-disciplines of mathematics. Statistics is a mathematical branch that deals with data collection, organization, interpretation, presentation and analysis. There are two main statistical methods used in data analysis - descriptive statistics and inferential statistics. Descriptive statistics summarizes the data from a sample by using indexes like mean and standard deviation, whereas, inferential statistics conclude from data that are subject to random variations. Probability is a measure that quantifies the likelihood that events are going to occur. It measures the quantity as a number between 0 and 1 that respectively indicate the impossibility and certainty of an event. Probability distributions are commonly used for statistical analysis. Both these topics are often studied in conjunction with one another. This book presents researches and studies performed by experts across the globe. It studies, analyzes and upholds the pillars of probability and statistics and their utmost significance in modern times. This book attempts to assist those with a goal of delving into these areas.

The information shared in this book is based on empirical researches made by veterans in this field of study. The elaborative information provided in this book will help the readers further their scope of knowledge leading to advancements in this field.

Finally, I would like to thank my fellow researchers who gave constructive feedback and my family members who supported me at every step of my research.

Editor

A Poisson-Gamma Model for Zero Inflated Rainfall Data

Nelson Christopher Dzupire [ID],[1] **Philip Ngare,**[1,2] **and Leo Odongo**[1,3]

[1]*Pan African University Institute of Basic Sciences, Technology and Innovation, Juja, Kenya*
[2]*University of Nairobi, Nairobi, Kenya*
[3]*Kenyatta University, Nairobi, Kenya*

Correspondence should be addressed to Nelson Christopher Dzupire; ndzupire@cc.ac.mw

Rainfall modeling is significant for prediction and forecasting purposes in agriculture, weather derivatives, hydrology, and risk and disaster preparedness. Normally two models are used to model the rainfall process as a chain dependent process representing the occurrence and intensity of rainfall. Such two models help in understanding the physical features and dynamics of rainfall process. However rainfall data is zero inflated and exhibits overdispersion which is always underestimated by such models. In this study we have modeled the two processes simultaneously as a compound Poisson process. The rainfall events are modeled as a Poisson process while the intensity of each rainfall event is Gamma distributed. We minimize overdispersion by introducing the dispersion parameter in the model implemented through Tweedie distributions. Simulated rainfall data from the model shows a resemblance of the actual rainfall data in terms of seasonal variation, means, variance, and magnitude. The model also provides mechanisms for small but important properties of the rainfall process. The model developed can be used in forecasting and predicting rainfall amounts and occurrences which is important in weather derivatives, agriculture, hydrology, and prediction of drought and flood occurrences.

1. Introduction

Climate variables, in particular, rainfall occurrence and intensity, hugely impact human and physical environment. Knowledge of the frequency of the occurrence and intensity of rainfall events is essential for planning, designing, and management of various water resources system [1]. Specifically rain-fed agriculture is a sensitive sector to weather and crop production is directly dependent on the amount of rainfall and its occurrence. Rainfall modeling has a great impact on crop growth, weather derivatives, hydrological systems, drought, and flood management and crop simulated studies.

Rainfall modeling is also important in pricing of weather derivatives which are financial instruments that are used as a tool for risk management to reduce risk associated with adverse or unexpected weather conditions.

Further as climate change greatly affects the environment there is an urgent need for predicting the variability of rainfall for future periods for different climate change scenarios in order to provide necessary information for high quality climate related impact studies [1].

However modeling precipitation poses a lot of challenges, namely, accurate measurement of precipitation since rainfall data consists of sequences of values which are either zero or some positive numbers (intensity) depending on the depth of accumulation over discrete intervals. In addition factors like wind can affect collection accuracy. Rainfall is localized unlike temperature which is highly correlated across regions; therefore a derivative holder based on rainfall may suffer geographical basis risk in case of pricing weather derivatives. The final challenge is the choice of a proper probability distribution function to describe precipitation data. The statistical property of precipitation is far more complex and a more sophisticated distribution is required [2].

Rainfall has been modeled as a chain dependent process where a two-state Markov chain model represents the occurrence of rainfall and the intensity of rainfall is modeled by fitting a suitable distribution like Gamma [3], exponential, and mixed exponential [1, 4]. These models are easy to

understand and interpret and use maximum likelihood to find the parameters. However models involve many parameters to fully describe the dynamics of rainfall as well as making several assumptions for the process.

Wilks [5] proposed a multisite model for daily precipitation using a combination of two-state Markov process (for the rainfall occurrence) and a mixed exponential distribution (for the precipitation amount). He found that the mixture of exponential distributions offered a much better fit than the commonly used Gamma distribution.

In study of Leobacher and Ngare [3] the precipitation is modeled on a monthly basis by constructing a suitable Markov-Gamma process to take into account seasonal changes of precipitation. It is assumed that rainfall data for different years of the same month is independent and identically distributed. It is assumed that precipitation can be forecast with sufficient accuracy for a month.

Another approach of modeling rainfall is based on the Poisson cluster model where two of the most recognized cluster based models in the stochastic modeling of rainfall are the Newman-Scott Rectangular Pulses model and the Bartlett-Lewis Rectangular Pulse model. These models represent rainfall sequences in time and rainfall fields in space where both the occurrence and depth processes are combined. The difficulty in Poisson cluster models as observed by Onof et al. [6] is the challenge of how many features should be addressed so that the model is still mathematically tractable. In addition the models are best fitted by the method of moments and so requires matching analytic expressions for the statistical properties such as mean and variance.

Carmona and Diko [7] developed a time-homogeneous jump Markov process to describe rainfall dynamics. The rainfall process was assumed to be in form of storms which consists of cells themselves. At a cell arrival time the rainfall process jumps up by a random amount and at extinction time it jumps down by a random amount, both modeled as Poisson process. Each time the rain intensity changes, an exponential increase occurs either upwards or downwards. To preserve nonnegative intensity, the downward jump size is truncated to the current jump size. The Markov jump process also allows for a jump directly to zero corresponding to the state of no rain [8].

In this study the rainfall process is modeled as a single model where the occurrence and intensity of rainfall are simultaneously modeled. The Poisson process models the daily occurrence of rainfall while the intensity is modeled using Gamma distribution as the magnitude of the jumps of the Poisson process. Hence we have a compound Poisson process which is Poisson-Gamma model. The contribution of this study is twofold: a Poisson-Gamma model that simultaneously describes the rainfall occurrence and intensity at once and a suitable model for zero inflated data which reduces overdispersion.

This paper is structured as follows. In Section 2 the Poisson-Gamma model is described and then formulated mathematically while Section 3 presents methods of estimating the parameters of the model. In Section 4 the model is fitted to the data and goodness of fit of the model is evaluated by mean deviance whereas quantile residuals perform the diagnostics check of the model. Simulation and forecasting are carried out in Section 5 and the study concludes in Section 6.

2. Model Formulation

2.1. Model Description. Rainfall comprises discrete and continuous components in that if it does not rain the amount of rainfall is discrete whereas if it rains the amount is continuous. In most research works [3, 4, 9] the rainfall process is presented by use of two separate models: one is for the occurrence and conditioned on the occurrence and another model is developed for the amount of rainfall. Rainfall occurrence is basically modeled as first or higher order Markov chain process and conditioned on this process a distribution is used to fit the precipitation amount. Commonly used distributions are Gamma, exponential, mixture of exponential, Weibull, and so on. These models work based on several assumptions and inclusion of several parameters to capture the observed temporal dependence of the rainfall process. However rainfall data exhibit overdispersion [10] which is caused by various factors like clustering, unaccounted temporal correlation, or the fact that the data is a product of Bernoulli trials with unequal probability of events. The stochastic models developed in this way underestimate the overdispersion of rainfall data which may result in underestimating the risk of low or high seasonal rainfall.

Our interest in this research is to simultaneously model the occurrence and intensity of rainfall in one model. We would model the rainfall process by using a Poisson-Gamma probability distribution which is flexible to model the exact zeros and the amount of rainfall together.

Rainfall is modeled as a compound Poisson process which is a Lévy process with Gamma distributed jumps. This is motivated by the sudden changes of rainfall amount from zero to a large positive value following each rainfall event which are modeled as pure jumps of the compound Poisson process.

We assume rainfall arrives in forms of storms following a Poisson process, and at each arrival time the current intensity increases by a random amount based on Gamma distribution. The jumps of the driving process represent the arrival of the storm events generating a jump size of random size. Each storm comprises cells that also arrive following another Poisson process.

The Poisson cluster processes gives an appropriate tool as rainfall data indicating presence of clusters of rainfall cells. As observed by Onof et al. [6] use of Gamma distributed variables for cell depth improves the reproduction of extreme values.

Lord [11] used the Poisson-Gamma compound process to model the motor vehicle crashes where they examined the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Wang [12] proposed a Poisson-Gamma compound approach for species richness estimation.

2.2. Mathematical Formulation. Let N_t be total number of rainfall event per day following a Poisson process such that

$$P(N_t = n) = e^{-\lambda} \frac{\lambda^n}{n!}, \quad \forall n \in \mathbb{N},$$

$$N_t = \sum_{t \geq 1} 1_{[t,\infty)}(t).$$

(1)

The amount of rainfall is the total sum of the jumps of each rainfall event, say $(y_i)_{i \geq 1}$, assumed to be identically and independently Gamma distributed and independent of the times of the occurrence of rainfall:

$$L(t) = \begin{cases} \sum_{i=1}^{N_t} y_i & N_t = 1, 2, 3, \dots \\ 0 & N_t = 0, \end{cases}$$

(2)

such that $y_i \sim \text{Gamma}(\alpha, P)$ is with probability density function

$$f(y) = \begin{cases} \dfrac{\alpha^P y^{P-1} e^{-\alpha y}}{\Gamma(P)} & y > 0, \\ 0 & y \leq 0. \end{cases}$$

(3)

Lemma 1. *The compound Poisson process (2) has a cumulant function*

$$\psi(s, t, x) = \lambda t \left(e^{M_Y(x)} - 1 \right),$$

(4)

for $0 \leq s < t$ and $x \in \mathbb{R}$, where $M_Y(x)$ is the moment generating function of the Gamma distribution.

Proof. The moment generating function $\Phi(s)$ of $L(s)$ is given by

$$M_L(s) = \mathbf{E}\left(e^{sL(t)}\right)$$

$$= \sum_{j=0}^{\infty} \mathbf{E}\left(e^{sL(t)} \mid N(t) = j\right) P(N(t) = j)$$

$$= \sum_{j=0}^{\infty} \mathbf{E}\left(e^{s(L(1)+L(2)+\cdots+L(j))} \mid N(t) = j\right) P(N(t) = j)$$

$$= \sum_{j=0}^{\infty} \mathbf{E}\left(e^{s(L(1)+L(2)+\cdots+L(j))}\right) P(N(t) = j)$$

(5)

because of independence of L and $N(t)$

$$= \sum_{j=0}^{\infty} (M_Y(s))^j e^{-\lambda t} \frac{(\lambda t)^j}{j!} = e^{-\lambda t} \sum_{j=0}^{\infty} (M_Y(s))^j \frac{(\lambda t)^j}{j!}$$

$$= e^{-\lambda t + M_Y(s)\lambda t}.$$

So the cumulant of L is

$$\ln M_L(s) = \lambda (M_Y(s) - 1) = \lambda \left[(1 - \alpha x)^{-P} - 1 \right].$$

(6)

$\square$

If we observe the occurrence of rainfall for n periods, then we have the sequence $\{L_i\}_{i=1}^n$ which is independent and identically distributed.

If on a particular day there is no rainfall that occurred, then

$$P(L = 0) = \exp(-\lambda) \frac{(\lambda)^0}{0!} = \exp(-\lambda) = p_0.$$

(7)

Therefore the process has a point mass at 0 which implies that it is not entirely continuous random variable.

Lemma 2. *The probability density function of L in (2) is*

$$f_\theta(L) = \exp(-\lambda) \delta(L) + \exp(-\lambda - \alpha L) L^{-1} r_P \left(\nu L^P \right),$$

(8)

where $\delta_0(L)$ is a dirac function at zero.

Proof. Let $q_0 = 1 - p_0$ be the probability that it rained. Hence for $L_i > 0$ we have

$$f_\theta^+(L) = \sum_{i=1}^{\infty} \frac{p_i}{q_0} \left(\frac{\alpha^{iP} L^{iP-1} \exp(-\alpha L)}{\Gamma(ip)} \right)$$

$$\text{where } p_i = \exp(-\lambda) \frac{(\lambda)^i}{i!}$$

$$= \frac{1}{q_0} \left[\sum_{i=1}^{\infty} p_i \exp(-\alpha L) \frac{\alpha^{iP} L^{iP-1}}{\Gamma(iP)} \right]$$

$$= \frac{1}{q_0} \left[\exp(-\alpha L) \sum_{i=1}^{\infty} p_i \frac{\alpha^{iP} L^{iP-1}}{\Gamma(iP)} \right]$$

$$= \frac{1}{q_0} \left[\exp(-\alpha L) \sum_{i=1}^{\infty} \left(\exp(-\lambda) \frac{(\lambda)^i}{i!} \right) \frac{\alpha^{iP} L^{iP-1}}{\Gamma(iP)} \right]$$

(9)

$$= \frac{\exp(-\lambda)}{q_0} \left[\exp(-\alpha L) \sum_{i=1}^{\infty} \left(\frac{(\lambda)^i}{i!} \right) \frac{\alpha^{ip} L^{iP-1}}{\Gamma(ip)} \right]$$

$$= \frac{\exp(-\lambda)}{q_0} \exp(-\alpha L) \left[\sum_{i=1}^{\infty} \frac{(\lambda)^i (\alpha L)^{ip}}{Li!\Gamma(iP)} \right]$$

$$= \frac{L^{-1} \exp(-\alpha L)}{(\exp(\lambda) - 1)} \sum_{i=1}^{\infty} \frac{\lambda^P \alpha^P L^P}{i!\Gamma(iP)}.$$

If we let $\nu = \lambda \alpha^P$ and $r_p(\nu L^P) = \sum_{i=1}^{\infty}(\nu L^P / i!\Gamma(iP))$, then we have

$$f_\theta^+(L) = \frac{L^{-1} \exp(-\alpha L)}{(\exp(\lambda) - 1)} r_P \left(\nu L^P \right).$$

(10)

We can express the probability density function $f_\theta(L)$ in terms of a Dirac function as

$$f_\theta(L) = p_0 \delta_0(L) + q_0 f_\theta^+(L)$$

$$= \exp(-\lambda) \delta_0(L)$$

$$+ \left[\frac{q_0}{(\exp(\lambda) - 1)} \right] L^{-1} \exp(-\alpha L) r_P(\nu L^P) \quad (11)$$

$$= \exp(-\lambda) \delta_0(L)$$

$$+ \exp(-\lambda - \alpha L) L^{-1} r_P(\nu L^P).$$

$\square$

Consider a random sample of size n of L_i with the probability density function

$$f_\theta(L) = \exp(-\lambda) \delta(L) + \exp(-\lambda - \alpha L) L^{-1} r_p(\nu L^P). \quad (12)$$

If we assume that there are m positive values $L_1, L_2, \ldots, L_m$, then there are $M = n - m$ zeros where $m > 0$.

We observe that $m \sim Bi(n, 1 - \exp(-\lambda))$ and $p(m = 0) = \exp(-n\lambda)$; hence the likelihood function is

$$L = \binom{n}{m} p_0^{n-m} q_0^m \prod_{i=1}^{m} f_\theta^+(L_i) \quad (13)$$

and the log-likelihood for $\theta = (\lambda, \alpha, p)$ is

$$\log L(\theta; L_1, L_2, \ldots, L_n)$$

$$= \log \left(\binom{n}{m} p_0^{n-m} q_0^m \prod_{i=1}^{m} f_\theta^+(L_i) \right)$$

$$= \log \left(\binom{n}{m} e^{-\lambda n + \lambda m} \left(1 - e^{-\lambda}\right)^m \prod_{i=1}^{m} e^{-\lambda - \alpha L_i} \frac{1}{L_i} \right.$$

$$\left. \cdot \sum_{j=1}^{\infty} \frac{\left(\lambda \alpha^p L_{ij}^p\right)^j}{j! \Gamma(jp)} \right) = \log \binom{n}{m} + \lambda(m-n) + m \quad (14)$$

$$\cdot \log \left(1 - e^{-\lambda}\right) + \sum_{i=1}^{m} -\lambda - \alpha L_i - \log L_i$$

$$+ \log \sum_{i=1}^{m} \sum_{j=1}^{\infty} \frac{\left(\lambda \alpha^p L_{ij}^p\right)^j}{j! \Gamma(jp)}.$$

Now for $\hat{\lambda}$ we have

$$\frac{\partial \log L(\theta; L_1, L_2, \ldots, L_n)}{\partial \lambda}$$

$$= m - n + \frac{m}{1 - e^{-\lambda}} + (-1)^m$$

$$+ \frac{1}{\lambda} \sum_{i=1}^{m} \sum_{j=1}^{\infty} i \frac{\partial \log L(\theta; L_1, L_2, \ldots, L_n)}{\partial \lambda} = 0 \implies \quad (15)$$

$$m - n + \frac{m}{1 - e^{-\lambda}} + (-1)^m + \frac{1}{\lambda} \sum_{i=1}^{m} \sum_{j=1}^{\infty} i = 0.$$

We can observe from the above evaluation that λ can not be expressed in closed form; similar derivation also shows that α as well can not be expressed in closed form. Therefore we can only estimate λ and α using numerical methods. Withers and Nadarajah [13] also observed that the probability density function can not be expressed in closed form and therefore it is difficult to find the analytic form of the estimators. So we will express the probability density function in terms of exponential dispersion models as described below.

Definition 3 (see [14]). A probability density function of the form

$$f(y; \theta, \Theta) = a(y, \Theta) \exp \left\{ \frac{1}{\Theta} [y\theta - k(\theta)] \right\} \quad (16)$$

for suitable functions $k()$ and $a()$ is called an exponential dispersion model.

$\Theta > 0$ is the dispersion parameter. The function $k(\theta)$ is the cumulant of the exponential dispersion model; since $\Theta = 1$, then $k'()$ are the successive cumulants of the distribution [15]. The exponential dispersion models were first introduced by Fisher in 1922.

If we let $L_i = \log f(y_i; \theta_i, \Theta)$ as a contribution of y_i to the likelihood function $L = \sum_i L_i$, then

$$L_i = \frac{1}{\Theta} [y_i \theta - k(\theta_i)] + \log a(y, \Theta),$$

$$\frac{\partial L_i}{\partial \theta_i} = \frac{1}{\Theta} (y_i - k'(\theta_i)), \quad (17)$$

$$\frac{\partial^2 L_i}{\partial \theta_i^2} = -\frac{1}{\Theta} k''(\theta_i).$$

However we expect that $\mathbf{E}(\partial L_i / \partial \theta_i) = 0$ and $-\mathbf{E}(\partial^2 L_i / \partial \theta_i^2) = \mathbf{E}(\partial L_i / \partial \theta_i)^2$ so that

$$\mathbf{E} \left(\frac{1}{\Theta} (y_i - k'(\theta_i)) \right) = 0,$$

$$\frac{1}{\Theta} (\mathbf{E}(y_i) - k'(\theta_i)) = 0, \quad (18)$$

$$\mathbf{E}(y_i) = k'(\theta_i).$$

Furthermore

$$-\mathbf{E} \left(\frac{\partial^2 L_i}{\partial \theta_i^2} \right) = \mathbf{E} \left(\frac{\partial L_i}{\partial \theta_i} \right)^2,$$

$$-\mathbf{E} \left(-\frac{1}{\Theta} k''(\theta_i) \right) = \mathbf{E} \left(\frac{1}{\Theta} (y_i - k'(\theta_i)) \right)^2, \quad (19)$$

$$\frac{k''(\theta_i)}{\Theta} = \frac{\text{Var}(y_i)}{\Theta^2},$$

$$\text{Var}(y_i) = \Theta k''(\theta_i).$$

Therefore the mean of the distribution is $\mathbf{E}[Y] = \mu = dk(\theta)/d\theta$ and the variance is $\text{Var}(Y) = \Theta(d^2 k(\theta)/d\theta^2)$.

The relationship $\mu = dk(\theta)/d\theta$ is invertible so that θ can be expressed as a function of μ; as such we have $\text{Var}(Y) = \Theta V(\mu)$, where $V(\mu)$ is called a variance function.

Definition 4. The family of exponential dispersion models, whose variance functions are of the form $V(\mu) = \mu^p$ for $p \in (-\infty, 0] \cup [1, \infty)$, are called Tweedie family distributions.

Examples are as follows: for $p = 0$ then we have a normal distribution, $p = 1$, and $\Theta = 1$; it is a Poisson distribution, and Gamma distribution for $p = 2$, while when $p = 3$ it is Gaussian inverse distribution. Tweedie densities can not be expressed in closed form (apart from the examples above) but can instead be identified by their cumulants generating functions.

From $\text{Var}(Y) = \Theta(d^2k(\theta)/d\theta^2)$, then for Tweedie family distribution we have

$$\text{Var}(Y) = \Theta \frac{d^2 k(\theta)}{d\theta^2} = \Theta V(\mu) = \Theta \mu^p. \qquad (20)$$

Hence we can solve for μ and $k(\theta)$ as follows:

$$\mu = \frac{dk(\theta)}{d\theta},$$

$$\frac{d\mu}{d\theta} = \mu^p \implies$$

$$\int \frac{d\mu}{\mu^p} = \int d\theta, \qquad (21)$$

$$\theta = \begin{cases} \dfrac{\mu^{1-p}}{1-p} & p \neq 1, \\ \log \mu & p = 1 \end{cases}$$

by equating the constants of integration above to zero.

For $p \neq 1$ we have $\mu = [(1-p)\theta]^{1/(1-p)}$ so that

$$\int dk(\theta) = \int [(1-p)\theta]^{1/(1-p)} d\theta,$$

$$k(\theta) = \frac{[(1-p)\theta]^{(2-p)/(1-p)}}{2-p} = \frac{\mu^{(2-p)/(1-p)}}{2-p}, \qquad (22)$$

$$p \neq 2.$$

Proposition 5. *The cumulant generating function of a Tweedie distribution for $1 < p < 2$ is*

$$\log M_Y(t)$$

$$= \frac{1}{\Theta} \frac{\mu^{2-p}}{p-1} \left[\left(1 + t\Theta(1-p)\mu^{p-1}\right)^{(2-p)/(1-p)} - 1 \right]. \qquad (23)$$

Proof. From (16) the moment generating function is given by

$$M_Y(t) = \int \exp(ty) a(y, \Theta) \exp\left\{\frac{1}{\Theta}[y\theta - k(\theta)]\right\} dy$$

$$= \int a(y, \Theta) \exp\left\{\frac{1}{\Theta}[y(\theta + t\Theta) - k(\theta)]\right\} dy$$

$$= \int a(y, \Theta) \exp\left(\frac{y(\theta + t\Theta) - k(\theta)}{\Theta}\right.$$

$$+ \frac{k(\theta + t\Theta) - k(\theta)}{\Theta}\right) dy = \int a(y, \Theta)$$

$$\cdot \exp\left(\frac{y(\theta + t\Theta) - k(\theta + t\Theta)}{\Theta}\right.$$

$$+ \frac{k(\theta + t\Theta) - k(\theta + t\Theta)}{\Theta}\right) dy = \int a(y, \Theta) \qquad (24)$$

$$\cdot \exp\left(\frac{y(\theta + t\Theta) - k(\theta + t\Theta)}{\Theta}\right)$$

$$\cdot \exp\left(\frac{k(\theta + t\Theta) - k(\theta + t\Theta)}{\Theta}\right) dy$$

$$= \exp\left(\frac{k(\theta + t\Theta) - k(\theta + t\Theta)}{\Theta}\right) \int a(y, \Theta)$$

$$\cdot \exp\left(\frac{y(\theta + t\Theta) - k(\theta + t\Theta)}{\Theta}\right) dy$$

$$= \exp\left\{\frac{1}{\Theta}[k(\theta + t\Theta) - k(\theta)]\right\}.$$

Hence cumulant generating function is

$$\log M_Y(t) = \frac{1}{\Theta}[k(\theta + t\Theta) - k(\theta)]. \qquad (25)$$

For $1 < p < 2$ we substitute θ and $k(\theta)$ to have

$$\log M_Y(t)$$

$$= \frac{1}{\Theta} \frac{\mu^{2-p}}{p-1} \left[\left(1 + t\Theta(1-p)\mu^{p-1}\right)^{(2-p)/(1-p)} - 1 \right]. \qquad (26)$$

$\square$

By comparing the cumulant generating functions in Lemma 1 and Proposition 5 the compound Poisson process can be thought of as Tweedie distribution with parameters (λ, α, P) expressed as follows:

$$\lambda = \frac{\mu^{2-p}}{\Theta(2-p)},$$

$$\alpha = \Theta(p-1)\mu^{p-1}, \qquad (27)$$

$$P = \frac{2-p}{p-1}.$$

The requirement that the Gamma shape parameter P be positive implies that only Tweedie distributions between $1 < p < 2$ can represent the Poisson-Gamma compound process. In addition, for $\lambda > 0$, $\alpha > 0$ implies $\mu > 0$ and $\Theta > 0$.

Proposition 6. *Based on Tweedie distribution, the probability of receiving no rainfall at all is*

$$P(L = 0) = \exp\left[-\frac{\mu^{2-p}}{\Theta(2-p)}\right] \quad (28)$$

and the probability of having a rainfall event is

$$P(L > 0)$$

$$= W(\lambda, \alpha, L, P) \exp\left[\frac{L}{(1-p)\mu^{p-1}} - \frac{\mu^{2-p}}{2-p}\right], \quad (29)$$

where

$$W(\lambda, \alpha, L, P) = \sum_{j=1}^{\infty} \frac{\lambda^j (\alpha L)^{jP} e^{-\lambda}}{j! \Gamma(jP)}. \quad (30)$$

Proof. This follows by directly substituting the values of λ and $\theta, k(\theta)$ into (16). □

The function $W(\lambda, \alpha, L, P)$ is an example of Wright's generalized Bessel function; however it can not be expressed in terms of the more common Bessel function. To evaluate it the value of j is determined for which the function W_j reaches the maximum [15].

3. Parameter Estimation

We approximate the function $W(\lambda, \alpha, L, P) = \sum_{j=1}^{\infty} (\lambda^j (\alpha L)^{jP} e^{-\lambda} / j! \Gamma(jP)) = \sum_{j=1}^{\infty} W_j$ following the procedure by [15] where the value of j is determined for which W_j reaches maximum. We treat j as continuous so that W_j is differentiated with respect to j and set the derivative to zero. So for $L > 0$ we have the following.

Lemma 7 (see [15]). *The log maximum approximation of W_j is given by*

$$\log W_{\max} = \frac{L^{2-p}}{(2-p)\Theta} \left[\log \frac{L^P (p-1)^P}{\Theta^{(1-P)}(2-p)} + (1+P) \right.$$

$$\left. - P \log P - (1-P) \log \frac{L^{2-p}}{(2-p)\Theta} \right] - \log(2\pi) - \frac{1}{2} \quad (31)$$

$$\cdot \log P - \log \frac{L^{2-p}}{(2-p)\Theta},$$

where $j_{\max} = L^{2-p}/(2-p)\Theta$.

Proof.

$$W(\lambda, \alpha, L, P) = \sum_{j=1}^{\infty} \frac{\lambda^j (\alpha L)^{jP-1} e^{-\lambda}}{j! \Gamma(jP)}$$

$$= \sum_{j=1}^{\infty} \frac{\lambda^j L^{jP-1} e^{-L/\tau} e^{-\lambda}}{j! \tau^{Pj} \Gamma(Pj)} \quad \text{where } \tau = \frac{1}{\alpha}. \quad (32)$$

Substituting the values of λ, α in the above equation we have

$$W(\lambda, \alpha, L, P)$$

$$= \sum_{j=1}^{\infty} \frac{\left(\mu^{2-p}/\Theta(2-p)\right)^j L^{jP-1} \left[\Theta(1-p)\mu^{p-1}\right]^{jP} e^{-L/\tau} e^{-\lambda}}{j! \Gamma(Pj)}$$

$$= e^{-L/\tau-\lambda} L^{-1} \sum_{j=1}^{\infty} \frac{\mu^{(2-p)j} \left(\Theta(p-1)\mu^{p-1}\right)^{jP} L^{jP}}{\Theta^j (2-p)^j j! \Gamma(jP)} \quad (33)$$

$$= e^{-L/\tau-\lambda} L^{-1} \sum_{j=1}^{\infty} \frac{L^{jP}(p-1)^{jP} \mu^{(2-p)j+(p-1)jP}}{\Theta^{j(1-P)}(2-p)^j j! \Gamma(jP)}.$$

The term $\mu^{(2-p)j+(p-1)jP}$ depends on the L, p, P, Θ values so we maximize the summation

$$W(L, \Theta, P) = \sum_{j=1}^{\infty} \frac{L^{jP}(p-1)^{jP}}{\Theta^{j(1-P)}(2-p)^j j! \Gamma(jP)}$$

$$= \sum_{j=1}^{\infty} \frac{z^j}{j! \Gamma(jP)} \quad (34)$$

$$\text{where } z = \frac{L^P (p-1)^P}{\Theta^{(1-P)}(2-p)}$$

$$= W_j.$$

Considering W_j we have

$$\log W_j = j \log z - \log j! - \log(Pj)$$

$$= j \log z - \log \Gamma(j+1) - \log(Pj). \quad (35)$$

Using Stirling's approximation of Gamma functions we have

$$\log \Gamma(1+j) \approx (1+j) \log(1+j) - (1+j)$$

$$+ \frac{1}{2} \log\left(\frac{2\pi}{1+j}\right), \quad (36)$$

$$\log \Gamma(Pj) \approx Pj \log(Pj) - Pj + \frac{1}{2} \log\left(\frac{2\pi}{Pj}\right).$$

And hence we have

$$W_j \approx j \left[\log z + (1+P) - P \log P - (1-P) \log j\right]$$

$$- \log(2\pi) - \frac{1}{2} \log P - \log j. \quad (37)$$

For $1 < p < 2$ we have $P = (2-p)/(p-1) > 0$; hence the logarithms have positive arguments. Differentiating with respect to j we have

$$\frac{\partial \log W_j}{\partial j} \approx \log z - \frac{1}{j} - \log j - P \log(Pj)$$

$$\approx \log z - \log j - P \log(Pj), \quad (38)$$

where $1/j$ is ignored for large j. Solving for $(\partial \log Wj)/\partial j = 0$ we have

$$j_{\max} = \frac{L^{2-p}}{(2-p)\,\Theta}. \tag{39}$$

Substituting $j_{\max}$ in $\log W_j$ to find the maximum approximation of W_j we have

$$\log W_{\max} = \frac{L^{2-p}}{(2-p)\,\Theta} \left[\log \frac{L^P (p-1)^P}{\Theta^{(1-P)} (2-p)} + (1+P) \right.$$

$$\left. - P \log P - (1-P) \log \frac{L^{2-p}}{(2-p)\,\Theta} \right] - \log(2\pi) - \frac{1}{2} \tag{40}$$

$$\cdot \log P - \log \frac{L^{2-p}}{(2-p)\,\Theta}.$$

Hence the result follows. $\qquad \square$

It can be observed that $\partial W_j/\partial j$ is monotonically decreasing; hence $\log W_j$ is strictly convex as a function of j. Therefore W_j decays faster than geometrically on either side of $j_{\max}$ [15]. Therefore if we are to estimate $W(L, \Theta, P)$ by $\widehat{W}(L, \Theta, P) = \sum_{j=j_d}^{j_u} W_j$ the approximation error is bounded by geometric sum

$$W(L, \Theta, P) - \widehat{W}(L, \Theta, P)$$

$$< W_{j_d-1} \frac{1 - r_l^{j_d-1}}{1 - r_l} + W_{j_u+1} \frac{1}{1 - r_u},$$

$$r_l = \exp\left(\frac{\partial W_j}{\partial j}\right)\bigg|_j = j_d - 1, \tag{41}$$

$$r_u = \exp\left(\frac{\partial W_j}{\partial j}\right)\bigg|_j = j_u + 1.$$

For quick and accurate evaluation of $W(\lambda, \alpha, L, P)$, the series is summed for only those terms in the series which contribute significantly to the sum.

Generalized linear models extend the standard linear regression models to incorporate nonnormal response distributions and possibly nonlinear functions of the mean. The advantage of GLMs is that the fitting process maximizes the likelihood for the choice of the distribution for a random variable y and the choice is not restricted to normality unlike linear regression [16].

The exponential dispersion models are the response distributions for the generalized linear models. Tweedie distributions are members of the exponential dispersion models upon which the generalized linear models are based. Consequently fitting a Tweedie distribution follows the framework of fitting a generalized linear model.

Lemma 8. *In case of a canonical link function, the sufficient statistics for $\{\beta_j\}$ are $\{\sum_{i=1}^{n} y_i x_{ij}\}$.*

Proof. For n independent observations y_i of the exponential dispersion model (16) the log-likelihood function is

$$L(\beta) = \sum_{i=1}^{n} L_i = \sum_i \log f(y_i, \theta_i, \Theta)$$

$$= \sum_{i=1}^{n} \frac{y_i \theta_i - k(\theta_i)}{\Theta} + \sum_i \log a(y_i, \Theta). \tag{42}$$

But $\theta_i = \sum_j^p \beta_j x_{ij}$; hence

$$\sum_i^n y_i \theta_i = \sum_{i=1}^{n} y_i \sum_j^p \beta_j x_{ij} = \sum_j^p \beta_j \sum_{i=1}^{n} y_i x_{ij}. \tag{43}$$

$\qquad \square$

Proposition 9. *Given that y_i is distributed as (16) then its distribution depends only on its first two moments, namely, μ_i and $\mathrm{Var}(y_i)$.*

Proof. Let $g(\mu_i)$ be the link function of the GLM such that $\eta_i = \sum_{j=1}^{p} \beta_j x_{ij} = g(\mu_i)$. The likelihood equations are

$$\frac{\partial L(\beta)}{\partial \beta} = \sum_{i=1}^{n} \frac{\partial L_i}{\partial \beta_j} \quad \forall j. \tag{44}$$

Using chain rule we have

$$\frac{\partial L_i}{\partial \beta_j} = \frac{\partial L_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \frac{y_i - \mu_i}{\mathrm{Var}(y_i)} x_{ij} \frac{\partial \mu_i}{\partial \eta_i}. \tag{45}$$

Hence

$$\frac{\partial L(\beta)}{\partial \beta} = \frac{y_i - \mu_i}{\mathrm{Var}(y_i)} x_{ij} \frac{\partial \mu_i}{\partial \eta_i} = \frac{y_i - \mu_i}{\Theta \mu_i^p} x_{ij} \frac{\partial \mu_i}{\partial \eta_i}. \tag{46}$$

Since $\mathrm{Var}(y_i) = V(\mu_i)$, the relationship between the mean and variance characterizes the distribution. $\qquad \square$

Clearly a GLM only requires the first two moments of the response y_i; hence despite the difficulty of full likelihood analysis of Tweedie distribution as it can not be expressed in closed form for $1 < p < 2$ we can still fit a Tweedie distribution family. The likelihood is only required to estimate p and Θ as well as diagnostic check of the model.

Proposition 10. *Under the standard regularity conditions, for large n, the maximum likelihood estimator $\widehat{\beta}$ of β for generalized linear model is efficient and has an approximate normal distribution.*

Proof. From the log-likelihood, the covariance matrix of the distribution is the inverse of the information matrix $\mathbf{J} = \mathbf{E}(-\partial^2 L(\beta)/\partial \beta_h \partial \beta_j)$.

So

$$\mathbf{J} = \mathbf{E}\left(-\frac{\partial^2 L(\beta)}{\partial \beta_h \partial \beta_j}\right) = \mathbf{E}\left[\left(\frac{\partial^2 L_i}{\partial \beta_h}\right)\left(\frac{\partial^2 L_i}{\partial \beta_j}\right)\right]$$

$$= \left[\left(\frac{y_i - \mu_i}{\mathrm{Var}(y_i)} x_{ih}\frac{\partial \mu_i}{\partial \eta_i}\right)\left(\frac{y_i - \mu_i}{\mathrm{Var}(y_i)} x_{ij}\frac{\partial \mu_i}{\partial \eta_i}\right)\right] \quad (47)$$

$$= \frac{x_{ih} x_{ij}}{\mathrm{Var}(y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2.$$

Hence

$$\mathbf{E}\left(-\frac{\partial^2 L(\beta)}{\partial \beta_h \partial \beta_j}\right) = \sum_i^n \frac{x_{ih} x_{ij}}{\mathrm{Var}(y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 = \left(X^T W X\right), \quad (48)$$

where $W = \mathrm{diag}[(1/\mathrm{Var}(y_i))(\partial \mu_i/\partial \eta_i)^2]$.

Therefore $\widehat{\beta}$ has an approximate $N[\beta, (X^T W X)^{-1}]$ with $\mathrm{Var}(\widehat{\beta}) = (X^T \widehat{W} X)^{-1}$, where $\widehat{W}$ is evaluated at $\widehat{\beta}$. □

To compute $\widehat{\beta}$ we use the iteratively reweighted least square algorithm proposed by Dobson and Barnett [17] where the iterations use the working weights w_i:

$$\frac{w_i}{V(\mu_i)\,\dot{g}(\mu_i)^2}, \quad (49)$$

where $V(\mu_i) = \mu_i^p$.

However estimating p is more difficult than estimating β and Θ such that most researchers working with Tweedie densities have p a priori. In this study we use the procedure in [15] where the maximum likelihood estimator of p is obtained by directly maximizing the profile likelihood function. For any given value of p we find the maximum likelihood estimate of β, Θ and compute the log-likelihood function. This is repeated several times until we have a value of p which maximizes the log-likelihood function.

Given the estimated values of p and β, then the unbiased estimator of Θ is given by

$$\widehat{\Theta} = \sum_{i=1}^n \frac{\left[L_i - \mu_i\left(\widehat{\beta}\right)\right]^2}{\mu_i\left(\widehat{\beta}\right)^{\widehat{p}}}. \quad (50)$$

Since for $1 < p < 2$ the Tweedie density can not be expressed in closed form, it is recommended that the maximum likelihood estimate of Θ must be computed iteratively from full data [15].

4. Data and Model Fitting

4.1. Data Analysis. Daily rainfall data of Balaka district in Malawi covering the period 1995–2015 is used. The data was obtained from Meteorological Surveys of Malawi. Figure 1 shows a plot of the data.

In summary the minimum value is 0 mm which indicates that there were no rainfall on particular days, whereas the maximum amount is 123.7 mm. The mean rainfall for the whole period is 3.167 mm.
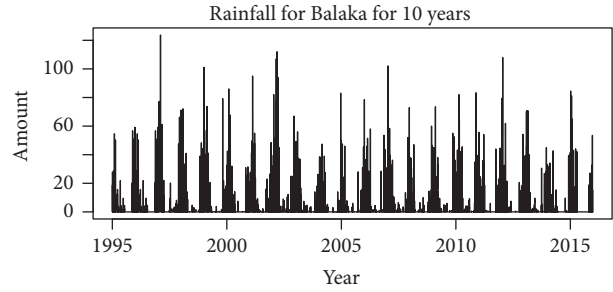
Figure 1: Daily rainfall amount for Balaka district.
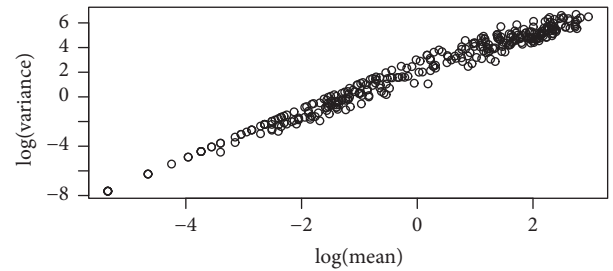
Figure 2: Variance mean relationship.

We investigated the relationship between the variance and the mean of the data by plotting the log(variance) against log(mean) as shown in Figure 2. From the figure we can observe a linear relationship between the variance and the mean which can be expressed as

$$\log(\mathrm{Variance}) = \alpha + \beta \log(\mathrm{mean}) \quad (51)$$

$$\mathrm{Variace} = A * \mathrm{mean}^\beta, \quad A \in \mathbb{R}. \quad (52)$$

Hence the variance can be expressed as some power $\beta \in \mathbb{R}$ of the mean agreeing with the Tweedie variance function requirement.

4.2. Fitted Model. To model the daily rainfall data we use sin and cos as predictors due to the cyclic nature and seasonality of rainfall. We have assumed that February ends on 28th for all the years to be uniform in our modeling.

The canonical link function is given by

$$\log \mu_i = a_0 + a_1 \sin\left(\frac{2\pi i}{365}\right) + a_2 \cos\left(\frac{2\pi i}{365}\right), \quad (53)$$

where $i = 1, 2, \ldots, 365$ corresponds to days of the year and a_0, a_1, a_2 are the coefficients of regression.

In the first place we estimate $\widehat{p}$ by maximizing the profile log-likelihood function. Figure 3 shows the graph of the profile log-likelihood function. As can be observed the value of p that maximizes the function is 1.5306.

From the results obtained after fitting the model, both the cyclic cosine and sine terms are important characteristics for daily rainfall Table 1. The covariates were determined to take into account the seasonal variations in the stochastic model.

TABLE 1: Estimated parameter values.

Parameter	Estimate	Std. error	t value	$\Pr(> |t|)$
$\hat{a}_0$	0.1653	0.0473	3.4930	0.0005^{***}
$\hat{a}_1$	0.9049	0.0572	15.81100	$<2e-16^{***}$
$\hat{a}_2$	2.0326	0.0622	32.6720	$<2e-16^{***}$
$\widehat{\Theta}$	14.8057	-	-	-

With *signif* code: $0 * * *$.
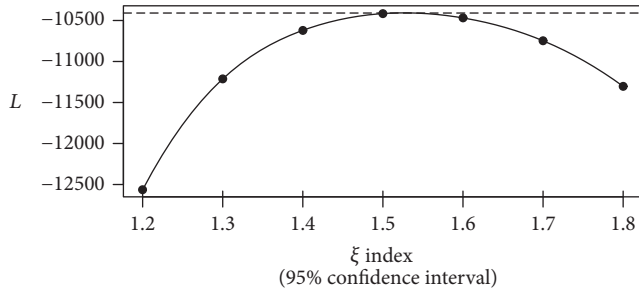
FIGURE 3: Profile likelihood.
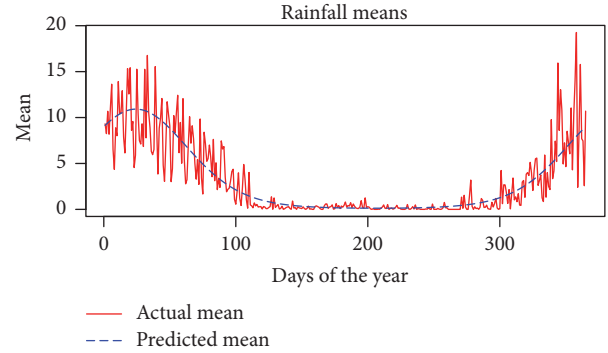
FIGURE 4: Actual versus predicted mean.

The predicted $\hat{\mu}_i, \hat{p}, \widehat{\Theta}$ for each day only depends on the day's conditions so that for each day i we have

$$\hat{\mu}_i = \exp\left[0.1653 + 0.9049 \sin\left(\frac{2\pi i}{365}\right)\right.$$

$$\left. + 2.0326 \cos\left(\frac{2\pi i}{365}\right)\right], \tag{54}$$

$$\hat{p} = 1.5306,$$

$$\widehat{\Theta} = 14.8057.$$

From these estimated values we can calculate the parameter $(\hat{\lambda}_i, \hat{\alpha}_i, \hat{P})$ from the corresponding formulas above as

$$\hat{\lambda}_i = \frac{1}{6.5716}\left(\exp\left[0.1653 + 0.9049 \sin\left(\frac{2\pi i}{365}\right)\right.\right.$$

$$\left.\left. + 2.03263 \cos\left(\frac{2\pi i}{365}\right)\right]\right)^{0.4694},$$

$$\hat{\alpha} = 7.4284\left(\exp\left[0.1653 + 0.9049 \sin\left(\frac{2\pi i}{365}\right)\right.\right. \tag{55}$$

$$\left.\left. + 2.0326 \cos\left(\frac{2\pi i}{365}\right)\right]\right)^{0.5306},$$

$$\hat{P} = 0.8847.$$

Comparing the actual means and the predicted means for 2 July we have $\hat{\mu} = 0.3820$, whereas $\mu = 0.4333$; similarly for 31 December we have $\hat{\mu} = 9.0065$ and $\mu = 10.6952$, respectively. Figure 4 shows the estimated mean and actual mean where the model behaves well generally.

4.3. Goodness of Fit of the Model. Let the maximum likelihood estimate of θ_i be $\hat{\theta}_i$ for all i and $\hat{\mu}$ as the model's mean

estimate. Let $\tilde{\theta}_i$ denote the estimate of θ_i for the saturated model with corresponding $\tilde{\mu} = y_i$.

The goodness of fit is determined by deviance which is defined as

$$-2\left[\frac{\text{maximum likelihood of the fitted model}}{\text{Maximum likelihood of the saturated model}}\right]$$

$$= -2\left[L\left(\hat{\mu}; y\right) - L\left(y, y\right)\right]$$

$$= 2\sum_{i=1}^{n} \frac{y_i\tilde{\theta}_i - k\left(\tilde{\theta}_i\right)}{\Theta} - 2\sum_{i=1}^{n} \frac{y_i\hat{\theta}_i - k\left(\hat{\theta}_i\right)}{\Theta} \tag{56}$$

$$= 2\sum_{i=1}^{n} \frac{y_i\left(\tilde{\theta}_i - \hat{\theta}_i\right) - k\left(\tilde{\theta}_i\right) + k\left(\hat{\theta}_i\right)}{\Theta} = \frac{\text{Dev}\left(y, \hat{\mu}\right)}{\Theta}.$$

$\text{Dev}(y, \hat{\mu})$ is called the deviance of the model and the greater the deviance, the poorer the fitted model as maximizing the likelihood corresponds to minimizing the deviance.

In terms of Tweedie distributions with $1 < p < 2$, the deviance is

$$\text{Dev}_p$$

$$= 2\sum_{i=1}^{n}\left(\frac{y_i^{2-p} - (2-p)y_i\mu_i^{1-p} + (1-p)\mu_i^{2-p}}{(1-p)(2-p)}\right). \tag{57}$$

Based on results from fitting the model, the residual deviance is 43144 less than the null deviance 62955 which implies that the fitted model explains the data better than a null model.

4.4. Diagnostic Check. The model diagnostic is considered as a way of residual analysis. The fitted model faces challenges
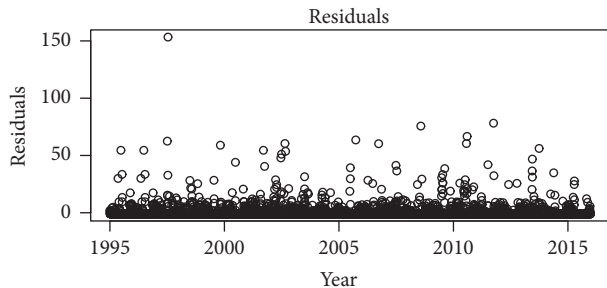
FIGURE 5: Residuals of the model.

FIGURE 7: Simulated rainfall and observed rainfall.

FIGURE 6: Q-Q plot of the quantile residuals.
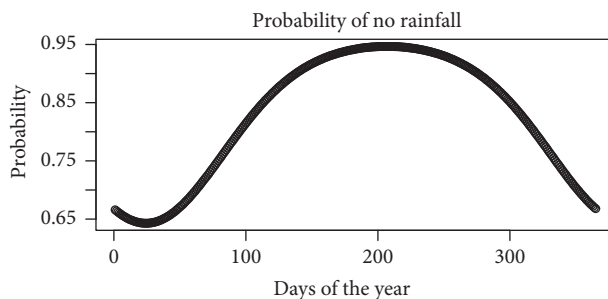
FIGURE 8: Probability of rainfall occurrence.

to be assessed especially for days with no rainfall at all as they produce spurious results and distracting patterns similarly as observed by [15]. Since this is a nonnormal regression, residuals are far from being normally distributed and having equal variances unlike in a normal linear regression. Here the residuals lie parallel to distinct values; hence it is difficult to make any meaningful decision about the fitted model (Figure 5).

So we assess the model based on quantile residuals which remove the pattern in discrete data by adding the smallest amount of randomization necessary on the cumulative probability scale.

The quantile residuals are obtained by inverting the distribution function for each response and finding the equivalent standard normal quantile.

Mathematically, let $a_i = \lim_{y \uparrow y_i} F(y; \widehat{\mu}_i, \widehat{\Theta})$ and $b_i = F(y_i; \widehat{\mu}_i, \widehat{\Theta})$, where F is the cumulative function of the probability density function $f(y; \mu, \Theta)$; then the randomized quantile residuals for y_i are

$$r_{q,i} = \Phi^{-1}(u_i) \tag{58}$$

with u_i being the uniform random variable on $(a_i, b_i]$. The randomized quantile residuals are distributed normally barring the variability in $\widehat{\mu}$ and $\widehat{\Theta}$.

Figure 6 shows the normalized Q-Q plot and as can be observed there are no large deviations from the straight line, only small deviations at the tail. The linearity observed indicates an acceptable fitted model.

5. Simulation

The model is simulated to test whether it produces data with similar characteristics to the actual observed rainfall. The simulation is done for a period of two years where one was the last year of the data (2015) and the other year (2016) was a future prediction. Then comparison was done with a graph for 2015 data as shown in Figure 7.

The different statistics of the simulated data and actual data are shown in Table 2 for comparisons.

The main objective of simulation is to demonstrate that the Poisson-Gamma can be used to predict and forecast rainfall occurrence and intensity simultaneously. Based on the results above (Figure 8), the model has shown that it works well in predicting the rainfall intensity and hence can be used in agriculture, actuarial science, hydrology, and so on.

However the model performed poorly in predicting probability of rainfall occurrence as it underestimated the probability of rainfall occurrence. It is suggested here that probably the use of truncated Fourier series can improve this estimation as compared to the sinusoidal.

But it performed better in predicting probability of no rainfall on days where there was little or no rainfall as indicated in Figure 8.

It can also be observed that the model produces synthetic precipitation that agrees with the four characteristics of a stochastic precipitation model as suggested by [4] as follows. The probability of rainfall occurrence obeys a seasonal pattern (Figure 8); in addition we can also tell that a probability of a rain in a day is higher if the previous day was wet which is the basis of precipitation models that involve the

TABLE 2: Data statistics.

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Predicted data	0.00	0.00	0.00	3.314	0.00	116.5
Actual data [10 yrs]	0.00	0.00	0.00	3.183	0.300	123.7
Actual data [2015]	0.00	0.00	0.00	3.328	0.00	84.5

Markov process. From Figure 7 we can also observe variation of rainfall intensity based on time of the season.

In addition the model allows modeling of exact zeros in the data and is able to predict a probability of no rainfall event simultaneously.

6. Conclusion

A daily stochastic rainfall model was developed based on a compound Poisson process where rainfall events follow a Poisson distribution and the intensity is independent of events following a Gamma distribution. Unlike several researches that have been carried out into precipitation modeling whereby two models are developed for occurrence and intensity, the model proposed here is able to model both processes simultaneously. The proposed model is also able to model the exact zeros, the event of no rainfall, which is not the case with the other models. This precipitation model is an important tool to study the impact of weather on a variety of systems including ecosystem, risk assessment, drought predictions, and weather derivatives as we can be able to simulate synthetic rainfall data. The model provides mechanisms for understanding the fine scale structure like number and mean of rainfall events, mean daily rainfall, and probability of rainfall occurrence. This is applicable in agriculture activities, disaster preparedness, and water cycle systems.

The model developed can easily be used for forecasting future events and, in terms of weather derivatives, the weather index can be derived from simulating a sample path by summing up daily precipitation in the relevant accumulation period. Rather than developing a weather index which is not flexible enough to forecast future events, we can use this model in pricing weather derivatives.

Rainfall data is generally zero inflated in that the amount of rainfall received on a day can be zero with a positive probability but continuously distributed otherwise. This makes it difficult to transform the data to normality by power transforms or to model it directly using continuous distribution. The Poisson-Gamma distribution has a complicated probability density function whose parameters are difficult to estimate. Hence expressing it in terms of a Tweedie distribution makes estimating the parameters easy. In addition, Tweedie distributions belong to the exponential family of distributions upon which generalized linear models are based; hence there is an already existing framework in place for fitting and diagnostic testing of the model.

The model developed allows the information in both zero and positive observations to contribute to the estimation of all parts of the model unlike the other model [3, 4, 9] which conditions rainfall intensity based on probability of occurrence. In addition the introduction of the dispersion parameter in the model helps in reducing underestimation of overdispersion of the data which is also common in the aforementioned models.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] A. Hussain, "Stochastic modeling of rainfall processes: a markov chain-mixed exponential model for rainfalls in different climatic conditions".

[2] M. Cao, A. Li, and J. Z. Wei, "Precipitation modeling and contract valuation: A frontier in weather derivatives," *The Journal of Alternative Investments*, vol. 7, no. 2, pp. 93–99, 2004.

[3] G. Leobacher and P. Ngare, "On modelling and pricing rainfall derivatives with seasonality," *Applied Mathematical Finance*, vol. 18, no. 1, pp. 71–91, 2011.

[4] M. Odening, O. Musshoff, and W. Xu, "Analysis of rainfall derivatives using daily precipitation models: Opportunities and pitfalls," *Agricultural Finance Review*, vol. 67, no. 1, pp. 135–156, 2007.

[5] D. S. Wilks, "Multisite generalization of a daily stochastic precipitation generation model," *Journal of Hydrology*, vol. 210, no. 1–4, pp. 178–191, 1998.

[6] C. Onof, R. E. Chandler, A. Kakou, P. Northrop, H. S. Wheater, and V. Isham, "Rainfall modelling using poisson-cluster processes: a review of developments," *Stochastic Environmental Research and Risk Assessment*, vol. 14, no. 6, pp. 384–411, 2000.

[7] R. Carmona and P. Diko, "Pricing precipitation based derivatives," *International Journal of Theoretical and Applied Finance*, vol. 8, no. 7, pp. 959–988, 2005.

[8] F. E. Benth and J. S. Benth, *Modeling and pricing in financial markets for weather derivatives*, vol. 17, World Scientific, 2012.

[9] B. López Cabrera, M. Odening, and M. Ritter, "Pricing rainfall futures at the CME," Technical report, Humboldt University, Collaborative Research Center.

[10] T. I. Harrold, A. Sharma, and S. J. Sheather, "A nonparametric model for stochastic generation of daily rainfall occurrence," *Water Resources Research*, vol. 39, no. 10, 2003.

[11] D. Lord, "Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed

dispersion parameter," *Accident Analysis & Prevention*, vol. 38, no. 4, pp. 751–766, 2006.

[12] J.-P. Wang, "Estimating species richness by a Poisson-compound gamma model," *Biometrika*, vol. 97, no. 3, pp. 727–740, 2010.

[13] C. S. Withers and S. Nadarajah, "On the compound Poisson-gamma distribution," *Kybernetika*, vol. 47, no. 1, pp. 15–37, 2011.

[14] E. W. Frees, R. Derrig, and G. Meyers, *Regression modeling with actuarial and financial applications*, vol. 1, Cambridge University Press, Cambridge, UK, 2014.

[15] P. K. Dunn and G. K. Smyth, "Evaluation of Tweedie exponential dispersion model densities by Fourier inversion," *Statistics and Computing*, vol. 18, no. 1, pp. 73–86, 2008.

[16] A. Agresti, *Foundations of linear and generalized linear models*, John Wiley & Sons, 2015.

[17] A. J. Dobson and A. G. Barnett, *An Introduction to Generalized Linear Models*, Texts in Statistical Science Series, CRC Press, Boca Raton, Fla, USA, 3rd edition, 2008.

U-Statistic for Multivariate Stable Distributions

Mahdi Teimouri, Saeid Rezakhah, and Adel Mohammadpour

Department of Statistics, Faculty of Mathematics and Computer Science, Amirkabir University of Technology (Tehran Polytechnic), 424 Hafez Ave., Tehran 15914, Iran

Correspondence should be addressed to Adel Mohammadpour; adel@aut.ac.ir

A *U*-statistic for the tail index of a multivariate stable random vector is given as an extension of the univariate case introduced by Fan (2006). Asymptotic normality and consistency of the proposed *U*-statistic for the tail index are proved theoretically. The proposed estimator is used to estimate the spectral measure. The performance of both introduced tail index and spectral measure estimators is compared with the known estimators by comprehensive simulations and real datasets.

1. Introduction

In recent years, stable distributions have received extensive use in a vast number of fields including physics, economics, finance, insurance, and telecommunications. Different sorts of data found in applications arise from heavy tailed or asymmetric distribution, where normal models are clearly inappropriate. In fact, stable distributions have theoretical underpinnings to accurately model a wide variety of processes. Stable distribution has originated with the work of Lévy [1]. There are a variety of ways to introduce a stable random vector. In the following, two definitions are proposed for a stable random vector; see Samorodnitsky and Taqqu [2].

Definition 1. A random vector $\mathbf{X} = (X_1, \ldots, X_d)^T$ is said to be stable in $\mathbb{R}^d$ if for any positive numbers A and B there are a positive number C and a vector $\mathbf{D} \in \mathbb{R}^d$ such that

$$AX_1 + BX_2 \overset{d}{=} CX + \mathbf{D}, \tag{1}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent and identical copies of $\mathbf{X}$ and $C = (A^\alpha + B^\alpha)^{1/\alpha}$.

Definition 2. Let $0 < \alpha < 2$. Then $\mathbf{X}$ is a non-Gaussian α-stable random vector in $\mathbb{R}^d$ if there exist a finite measure Γ on the unit sphere $\mathbb{S}^d = \{\mathbf{x} = (x_1, \ldots, x_d)^T \in \mathbb{R}^d \mid \langle \mathbf{x}, \mathbf{x} \rangle = 1\}$ and a vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^T \in \mathbb{R}^d$ such that

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \log E\left(\exp\left(i\left\langle \mathbf{t}, \mathbf{X}\right\rangle\right)\right) = \begin{cases} -\int_{\mathbb{S}^d} |\langle \mathbf{t}, \mathbf{s}\rangle|^\alpha \left[1 - i \operatorname{sgn}\langle \mathbf{t}, \mathbf{s}\rangle \tan\left(\dfrac{\pi\alpha}{2}\right)\right]\Gamma(d\mathbf{s}) + i\langle \mathbf{t}, \boldsymbol{\mu}\rangle, & \alpha \neq 1, \\[4mm] -\int_{\mathbb{S}^d} |\langle \mathbf{t}, \mathbf{s}\rangle| \left[1 + i \operatorname{sgn}\langle \mathbf{t}, \mathbf{s}\rangle \dfrac{2}{\pi}\log|\langle \mathbf{t}, \mathbf{s}\rangle|\right]\Gamma(d\mathbf{s}) + i\langle \mathbf{t}, \boldsymbol{\mu}\rangle, & \alpha = 1, \end{cases} \tag{2}$$

where $\langle \mathbf{t}, \mathbf{s}\rangle = \sum_{i=1}^{d} t_i s_i$ for $\mathbf{t} = (t_1, \ldots, t_d)^T$, $\mathbf{s} = (s_1, \ldots, s_d)^T$, $i^2 = -1$, and $\operatorname{sgn}(\cdot)$ denotes the sign function. The pair $(\Gamma, \boldsymbol{\mu})$ is unique.

The parameter α, in Definitions 1 and 2, is called tail index. A random vector $\mathbf{X}$ is said to be a strictly α-stable random

vector in $\mathbb{R}^d$ if $\boldsymbol{\mu} = \mathbf{0}$ for $\alpha \neq 1$; see Samorodnitsky and Taqqu [2]. We note that $\mathbf{X}$ is strictly α-stable, in the sense of Definition 1, if $\mathbf{D} = \mathbf{0}$. Throughout we assume that $\mathbf{X}$ is strictly α-stable and $\alpha \neq 1$. The probability density function of a stable distribution has no closed-form expression and moments with orders greater than or equal to α are not

finite for the members of this class. The two aforementioned difficulties make statistical inference about the parameters of a stable distribution hard. However, a series of contributions has permitted inference about the parameters of univariate and multivariate stable distributions. For example, in the univariate case, maximum likelihood (ML) estimation was studied first by DuMouchel (1971) and then by Nolan [3]. Although the ML approach leads to an efficient estimate for samples of large size, it involves numerical complexities. A program, called STABLE uses a cubic spline interpolation of stable densities for this purpose; see Nolan [4]. STABLE estimates all four parameters of a stable distribution for $\alpha \geq$ 0.4. Sample quantile (SQ) technique is another approach proposed by McCulloch [5]. The results are simple and consistent estimators of all four parameters based on five sample quantiles. The empirical characteristic function (CF) is suggested by Kogon and Williams [6]. The CF and SQ methods work well but are not as efficient as the ML method. As the last approach considered here, U-statistics for the tail index and scale parameters of a univariate strictly stable distribution are introduced by Fan [7]. In multivariate case, the focus of interest is the spectral measure estimation. Among them, we refer to Nolan et al. [8], Pivato and Seco [9], Ogata [10], and Mohammadi et al. [11].

The structure of the paper is as follows. In Section 2, new estimators for the tail index and spectral measure of a strictly stable distribution are presented which is an extension of the U-statistic proposed by Fan [7] for the univariate case. A comprehensive simulation study is performed in Section 3 to compare the performance of the introduced estimators and the known estimators. Two real data sets are analyzed in this section to illustrate the performance of the proposed method.

2. New Estimators

This section consists of two subsections. Firstly, we propose an estimator for the tail index. Secondly, an estimator for the spectral measure is given.

2.1. Estimation of Tail Index. The main result of this section is given in Theorem 4, which gives U-statistic for the inverse of tail index of a strictly stable distribution. We present the main result in the light of Lemma 3 given as follows. The proofs are given in the Appendix.

Lemma 3. *Let* $\mathbf{X} = (X_1, \ldots, X_d)^T$ *be a d-dimensional strictly stable random vector. Then,* $\operatorname{Var} \log \|\mathbf{X}\|$ *is finite, where* $\| \cdot \|$ *denotes the Euclidean norm.*

Theorem 4. *Let* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *be a sequence of n observations from a d-dimensional strictly stable random vector. Then*

$$U_n = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} H\left(\mathbf{x}_i, \mathbf{x}_j\right), \qquad (3)$$

where

$$H\left(\mathbf{x}_i, \mathbf{x}_j\right) = \frac{\log \|\mathbf{x}_i + \mathbf{x}_j\|}{\log 2} - \frac{\log \|\mathbf{x}_i\| + \log \|\mathbf{x}_j\|}{2 \log 2} \qquad (4)$$

is the U-statistic for $1/\alpha$.

As it is seen, from Theorem 4, the introduced U-statistic is an unbiased estimator for $1/\alpha$. Hereafter, we write $\widehat{\alpha}_{\mathrm{MU}} = 1/U_n$ as introduced estimator for α. Here, subscript MU indicates that $\widehat{\alpha}_{\mathrm{MU}}$ is constructed based on multivariate U-statistic defined in Theorem 4. It should be noted that when the true value of α is near two, the kernel given in (4) could be less than 0.5. So, $\widehat{\alpha}_{\mathrm{MU}}$ is greater than two. In this case, we set $\widehat{\alpha}_{\mathrm{MU}} = 2$.

2.2. Spectral Measure Estimation. We use $\widehat{\alpha}_{\mathrm{MU}}$ to estimate an m-point discrete approximation to the exact spectral measure of the form

$$\Gamma(\cdot) = \sum_{j=1}^{m} \gamma_j I_{\mathbf{s}_j}(\cdot), \qquad (5)$$

where γ_j is a mass at point $\mathbf{s}_j$ in the unit sphere $\mathbb{S}^d$ and $I_{\mathbf{s}_j}(\cdot)$ is an indicator function at point $\mathbf{s}_j$; for $j = 1, \ldots, m$, see Byczkowski et al. [12]. To estimate $\Gamma(\cdot)$, we replace Definition 2 for a strictly d-dimensional stable random vector with

$$\varphi_{\mathbf{X}}(\mathbf{t}) = -\sum_{j=1}^{m} \left|\langle \mathbf{t}, \mathbf{s}_j \rangle\right|^{\alpha} \left[1 - i \operatorname{sgn}\langle \mathbf{t}, \mathbf{s}_j \rangle G\left(\alpha, \mathbf{t}, \mathbf{s}_j\right)\right] \gamma_j,$$

$$(6)$$

$$= -\sum_{j=1}^{m} \psi\left(\mathbf{t}, \mathbf{s}_j, \alpha\right) \gamma_j,$$

where $G(\alpha, \mathbf{t}, \mathbf{s}_j) = \tan(\pi\alpha/2)$ for $\alpha \neq 1$ and $-2/\pi \log |\langle \mathbf{t}, \mathbf{s}_j \rangle|$ for $\alpha = 1$. Define

$$\mathbf{\Lambda}$$

$$= \begin{pmatrix} \psi\left(\mathbf{t}_1, \mathbf{s}_1, \alpha\right) & \psi\left(\mathbf{t}_1, \mathbf{s}_2, \alpha\right) & \cdots & \psi\left(\mathbf{t}_1, \mathbf{s}_m, \alpha\right) \\ \psi\left(\mathbf{t}_2, \mathbf{s}_1, \alpha\right) & \psi\left(\mathbf{t}_2, \mathbf{s}_2, \alpha\right) & \cdots & \psi\left(\mathbf{t}_2, \mathbf{s}_m, \alpha\right) \\ \vdots & \vdots & \ddots & \vdots \\ \psi\left(\mathbf{t}_m, \mathbf{s}_1, \alpha\right) & \psi\left(\mathbf{t}_m, \mathbf{s}_2, \alpha\right) & \cdots & \psi\left(\mathbf{t}_m, \mathbf{s}_m, \alpha\right) \end{pmatrix}, \quad (7)$$

$$\mathbf{V} = \left(-\log \varphi_{\mathbf{X}}\left(\mathbf{t}_1\right), \ldots, -\log \varphi_{\mathbf{X}}\left(\mathbf{t}_m\right)\right)^T, \qquad (8)$$

where $\mathbf{t}_j = (t_{j1}, \ldots, t_{jd})^T \in \mathbb{S}^d$, for $j = 1, \ldots, m$. Using (7) and (8), both sides of (6) are connected together through the following linear system:

$$\mathbf{V} = \mathbf{\Lambda}\boldsymbol{\gamma}, \qquad (9)$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^T$. Assuming that $\mathbf{\Lambda}$ in (9) is nonsingular, then $\boldsymbol{\gamma} = \mathbf{\Lambda}^{-1}\mathbf{V}$. Hence, we estimate the vector of the masses as

$$\widehat{\boldsymbol{\gamma}} = \widehat{\mathbf{\Lambda}}^{-1}\widehat{\mathbf{V}}, \qquad (10)$$

where

$$\widehat{\mathbf{\Lambda}}$$

$$= \begin{pmatrix} \psi\left(\mathbf{t}_1, \mathbf{s}_1, \widehat{\alpha}_{\mathrm{MU}}\right) & \psi\left(\mathbf{t}_1, \mathbf{s}_2, \widehat{\alpha}_{\mathrm{MU}}\right) & \cdots & \psi\left(\mathbf{t}_1, \mathbf{s}_m, \widehat{\alpha}_{\mathrm{MU}}\right) \\ \psi\left(\mathbf{t}_2, \mathbf{s}_1, \widehat{\alpha}_{\mathrm{MU}}\right) & \psi\left(\mathbf{t}_2, \mathbf{s}_2, \widehat{\alpha}_{\mathrm{MU}}\right) & \cdots & \psi\left(\mathbf{t}_2, \mathbf{s}_m, \widehat{\alpha}_{\mathrm{MU}}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \psi\left(\mathbf{t}_m, \mathbf{s}_1, \widehat{\alpha}_{\mathrm{MU}}\right) & \psi\left(\mathbf{t}_m, \mathbf{s}_2, \widehat{\alpha}_{\mathrm{MU}}\right) & \cdots & \psi\left(\mathbf{t}_m, \mathbf{s}_m, \widehat{\alpha}_{\mathrm{MU}}\right) \end{pmatrix},$$

$$\widehat{\mathbf{V}}$$

$$= \left(-\log \frac{1}{n}\sum_{i=1}^{n}\exp\left\{i\left\langle \mathbf{t}_1, \mathbf{x}_i\right\rangle\right\},\ldots,-\log \frac{1}{n}\sum_{i=1}^{n}\exp\left\{i\left\langle \mathbf{t}_m, \mathbf{x}_i\right\rangle\right\}\right)^{T},$$

(11)

in which $\mathbf{x}_i$ is i-th vector observation in random sample of size n.

Due to the standard error of $\widehat{\mathbf{V}}$, we have two problems with direct use of (10). Firstly, $\widehat{\gamma}$ may be complex, and secondly, its real part may be quite negative. Since $\widehat{\Lambda}$ and $\widehat{\mathbf{V}}$ are complex while gamma is constrained to be real (and nonnegative), the Euclidean norm used by McCulloch [13] and Nolan et al. [8] must be replaced with the complex modulus to solve both problems in a novel way. For this, we use the nnls($\cdot$) library in the R package. In the next section, the estimated spectral measure $\widehat{\gamma}$, based on $\widehat{\alpha}_{\mathrm{MU}}$, is shown by $\widehat{\gamma}_{\mathrm{MU}}$. We note that another estimator of γ can be constructed by separating both of the real and imaginary parts in the structure of $\widehat{\mathbf{V}}$. But simulation results show that constructed estimator gives the same performance.

3. Simulation Study

This section is in three parts. Firstly, we study the performance of the proposed estimator with the known ones for estimating the tail index. Secondly, we compare the performance of the spectral measure estimator developed through the introduced tail index estimator with the known approaches. In the last subsection, we give a real data example to illustrate the efficiency of the proposed estimators.

3.1. Performance Analysis of the Tail Index Estimators. Here, we perform a simulation study to compare the performance of $\widehat{\alpha}_{\mathrm{MU}}$ and four other estimators for α, including (1) $\widehat{\alpha}_{\mathrm{ML}}$, (2) $\widehat{\alpha}_{\mathrm{SQ}}$, (3) $\widehat{\alpha}_{\mathrm{CF}}$, and (4) $\widehat{\alpha}_{\mathrm{MM}}$. The first three competitors are ML, SQ, and CF estimations for the tail index, respectively. Each of three competitors is obtained as $\widehat{\alpha}_{\mathrm{PROJ}} = 1/m \sum_{j=1}^{m} \widehat{\alpha}(\mathbf{u}_j)$ after projecting the d-dimensional stable random vector using $\langle \mathbf{u}_j, \mathbf{X}\rangle$. Here, m is the number of masses, $\mathbf{u}_j = (u_{j1},\ldots,u_{jd})^{T}$ is an arbitrary unit vector, and $\mathbf{X}$ is the d-dimensional stable random vector. It is worth noting that the first three competitors are computed by the help of STABLE software after projecting. The fourth estimator, that is, $\widehat{\alpha}_{\mathrm{MM}}$, is the second estimator for tail index proposed by Mohammadi et al. [11]. We compare both the bias and root mean-squared error (RMSE) of estimators for 500 replications of samples of size $n = 500$ and 5000 of a bivariate stable random vector generated by the method given in Modarres and Nolan [14]. We use two settings for discrete spectral measure with $m = 8$ masses, including $\gamma_1 = (0.1, 0.2, 0.3, 0.4, 0.1, 0.2, 0.3, 0.4)^{T}$ and $\gamma_2 = (0, 0.1, 0.7, 0.3, 0.7, 0.3, 0.7, 0.1)^{T}$. In both cases, masses are concentrated on points $\mathbf{s}_j = (\cos(2\pi(j-1)/m), \sin(2\pi(j-1)/m))^{T}$ for $j = 1,\ldots,8$. In the first case that data are coming from a stable distribution with γ_1, we generate $\mathbf{t}_j$ from a uniform distribution on the unit sphere $\mathbb{S}^{d}$. For the second case, we set $\mathbf{t}_j = \mathbf{s}_j$. Biases and RMSEs

for $\alpha = (0.1 : 0.1 : 0.9, 0.95, 1.05, 1.1 : 0.1 : 1.9, 1.95, 2)$ are shown in Figures 1 and 2. As Figure 1 shows, when $\gamma_1 = (0.1, 0.2, 0.3, 0.4, 0.1, 0.2, 0.3, 0.4)^{T}$, we observe that $\widehat{\alpha}_{\mathrm{MU}}$ is more efficient than $\widehat{\alpha}_{\mathrm{SQ}}$ for $n = 5000$. Also, it works better than $\widehat{\alpha}_{\mathrm{MM}}$ in terms of RMSE (for $\alpha \leq 1.8$). Based on Figure 2, when $\gamma_2 = (0, 0.1, 0.7, 0.3, 0.7, 0.3, 0.7, 0.1)^{T}$, we observe that $\widehat{\alpha}_{\mathrm{MU}}$ is more efficient than other methods when $\alpha < 1.4$ and $n = 5000$ in the sense of RMSE. Also, when $\alpha < 1.7$ and $n = 500$, $\widehat{\alpha}_{\mathrm{MU}}$ is more efficient than $\widehat{\alpha}_{\mathrm{SQ}}$, $\widehat{\alpha}_{\mathrm{CF}}$, and $\widehat{\alpha}_{\mathrm{MM}}$ with respect to RMSE.

3.2. Performance Analysis of the Spectral Measure Estimators. Here, we compare the performance of the estimator for masses of spectral measure $\gamma = (\gamma_1,\ldots,\gamma_m)^{T}$ constructed based on U-statistic, $\widehat{\gamma}_{\mathrm{MU}}$ with the other four known estimators for the spectral measure. The competitors are three types of estimators for γ based on empirical characteristic function method: (1) $\widehat{\gamma}_{\mathrm{MLE\text{-}cf}}$; (2) $\widehat{\gamma}_{\mathrm{SQ\text{-}cf}}$; (3) $\widehat{\gamma}_{\mathrm{CF\text{-}cf}}$; and (4) Mohammadi et al. [11] estimator for γ, $\widehat{\gamma}_{\mathrm{MM}}$. For computing $\widehat{\gamma}_{\mathrm{MLE\text{-}cf}}$, $\widehat{\gamma}_{\mathrm{SQ\text{-}cf}}$, and $\widehat{\gamma}_{\mathrm{CF\text{-}cf}}$, we use command mvstable.fit($\mathbf{x}$, nspectral, method1d, method2d, param) in the STABLE program, where $\mathbf{x}$ is data vector, nspectral is number of spectral measure masses, method1d is the method to use for estimating parameters of univariate stable distribution, that is, MLE, SQ, and CF (corresponding codes in STABLE are 1, 2, and 3, respectively), method2d is the method to use for estimating parameters of bivariate stable distribution (we set method2d = 2 which corresponds to empirical characteristic function approach, cf), and param refers to kind of parameterization. Here, we set param = 1 since we are using the characteristic function in (2). More information about the first three competitors is given in Robust Analysis Inc. [15]. The estimators $\widehat{\gamma}_{\mathrm{MLE\text{-}cf}}$, $\widehat{\gamma}_{\mathrm{SQ\text{-}cf}}$, $\widehat{\gamma}_{\mathrm{CF\text{-}cf}}$, and $\widehat{\gamma}_{\mathrm{MM}}$ are obtained by substituting $\widehat{\alpha}_{\mathrm{ML}}$, $\widehat{\alpha}_{\mathrm{SQ}}$, $\widehat{\alpha}_{\mathrm{CF}}$, and $\widehat{\alpha}_{\mathrm{MM}}$ into (7) and then solving linear system (10), respectively. Comparisons are based on the RMSE of $\widehat{\gamma}_j$, for $j = 1,\ldots,m$, which is defined as $\sqrt{1/N \sum_{i=1}^{N}(\widehat{\gamma}_{ij} - \gamma_{ij})^2}$, where N is the number of iterations and $\widehat{\gamma}_{ij}$ is the estimation of jth component of $\widehat{\gamma}$ at ith iteration. We consider five scenarios for the structure of discrete spectral measure as follows.

(1) Independent case: $\gamma = (1/4, 0, 1/4, 0, 1/4, 0, 1/4, 0)^{T}$.

(2) Symmetric case: $\gamma = (0.1, 0.2, 0.3, 0.4, 0.1, 0.2, 0.3, 0.4)^{T}$.

(3) Uniform case: $\gamma = (1/8, 1/8, 1/8, 1/8, 1/8, 1/8, 1/8, 1/8)^{T}$.

(4) Triangle case: $\gamma = (0, 0.1, 0.7, 0.3, 0.7, 0.3, 0.7, 0.1)^{T}$.

(5) Exchangeable case: $\gamma = (0.1, 0.2, 0.1, 0.4, 0.3, 0.2, 0.3, 0.4)^{T}$.

We note that the first and the third scenarios above are similar to Examples 2 and 1 of Nolan et al. [8], respectively. The fourth scenario is called Triangle since corresponding density contour plot is similar to a triangle. For each of the above five scenarios, we arrange the settings of simulation as $m = 8$, $\alpha = 1.25; 1.75$; $n = 2000; 5000$ (n is sample size),
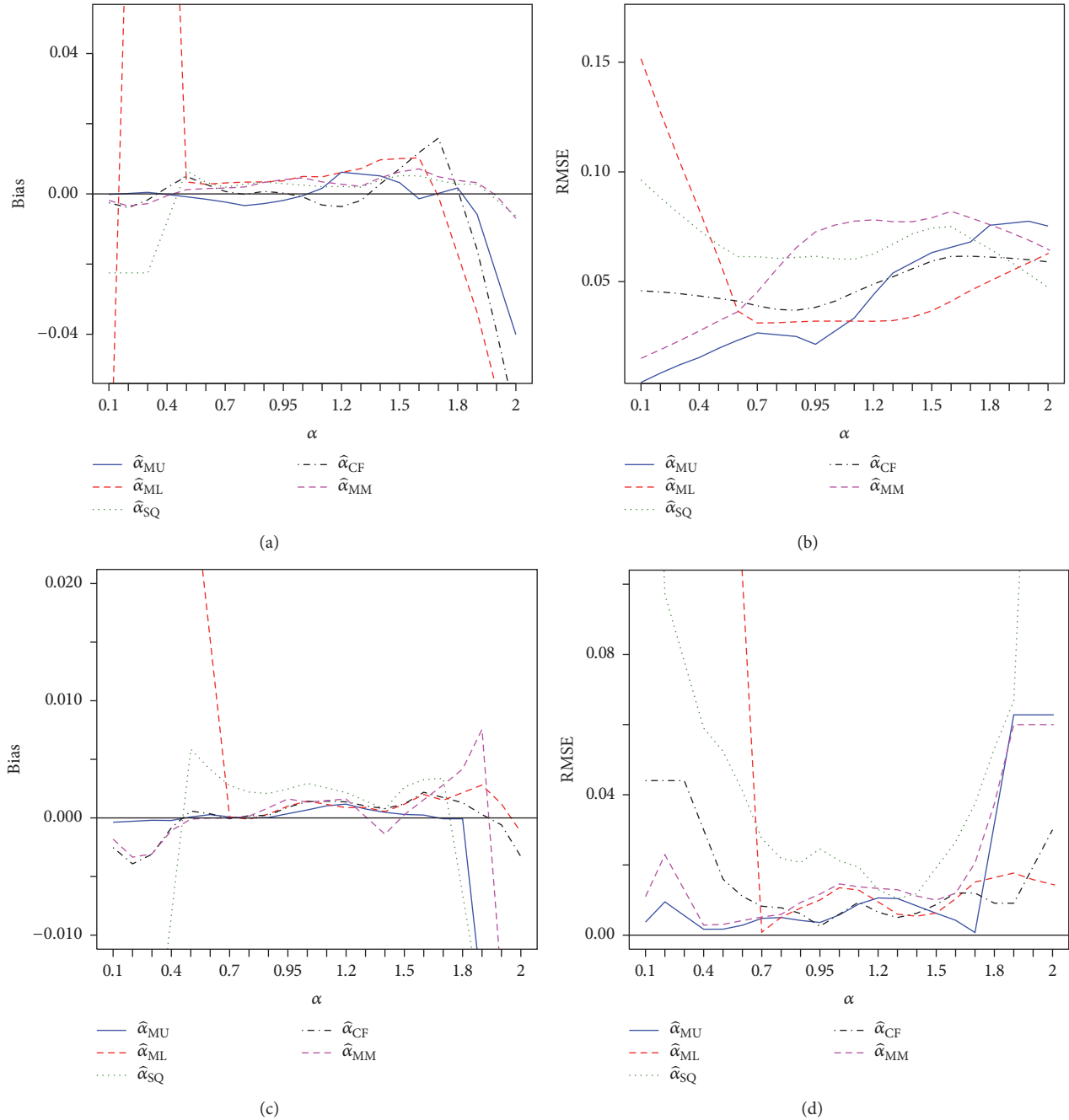
FIGURE 1: Biases and RMSEs of estimators when data are generated from a strictly stable distribution with discrete spectral measure $\gamma_1 = (0.1, 0.2, 0.3, 0.4, 0.1, 0.2, 0.3, 0.4)^T$. (a) Bias when $n = 500$, (b) RMSE when $n = 500$, (c) bias when $n = 5000$, and (d) RMSE when $n = 5000$.

and $N = 500$. It should be noted that masses are located at $\mathbf{s}_j = (\cos(2\pi(j-1)/m), \sin(2\pi(j-1)/m))^T$, for $j = 1, \ldots, m$, and components of $\mathbf{t}_j = (t_{j1}, \ldots, t_{jd})^T$ are generated from a uniform distribution on the unit sphere $\mathbb{S}^d$. The results of simulations are given in Figures 3–6. As it is seen, $\widehat{\gamma}_{MU}$ shows better performance than $\widehat{\gamma}_{MM}$.

3.3. Real Data Analysis. Here, we give two examples. In the first example, adjusted daily log-return (in percent) for the 30

stocks at the Dow Jones index is collected between January 3, 2000, and December 31, 2004. The log-return percent of 1247 closing prices has been computed for AXP (American Express Company) and MRK (Merck & Co. Inc.) stocks after multiplying the daily log-return by 100; see Nolan [16]. The scatter plot of AXP and MRK stocks log-return percent values, $\mathbf{X} = (AXP, MRK)^T$, is shown in Figure 7. We use a bivariate α-stable distribution with $m = 12$ points of masses for spectral measure addressed by $\mathbf{s}_j = (\cos(2\pi(j-1)/m), \sin(2\pi(j-1)/m))^T$, for $j = 1, \ldots, 12$. We estimate the

(a)

(b)

(c)

(d)
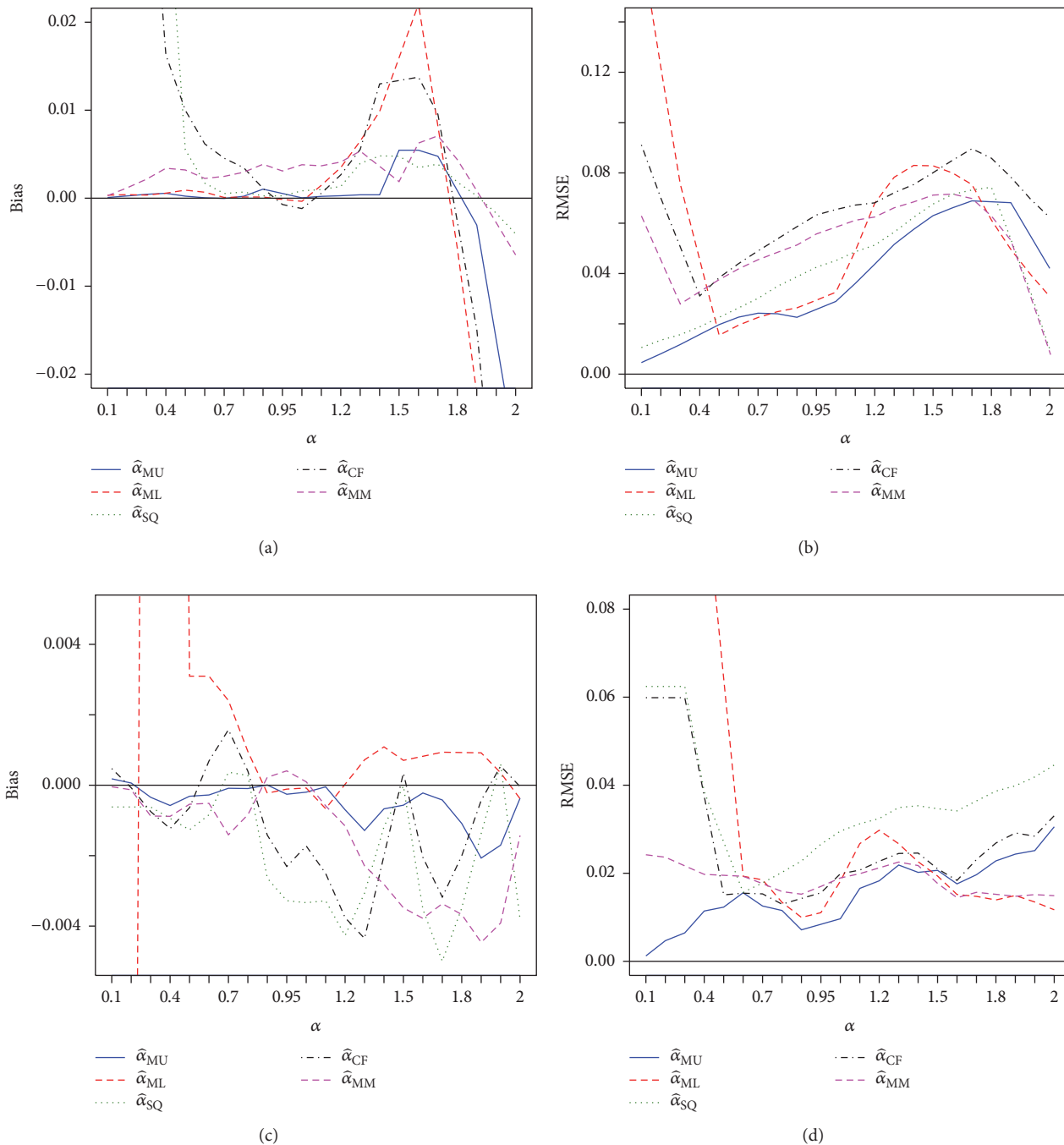
FIGURE 2: Biases and RMSEs of estimators when data are generated from a strictly stable distribution with discrete spectral measure $\gamma_2 = (0, 0.1, 0.7, 0.3, 0.7, 0.3, 0.7, 0.1)^T$. (a) Bias when $n = 500$, (b) RMSE when $n = 500$, (c) bias when $n = 5000$, and (d) RMSE when $n = 5000$.

location parameter as $\widehat{\boldsymbol{\mu}}_{\text{ML-cf}} = (-3.438E - 07, -2.402E - 07)^T$. So, a strictly α-stable distribution is fitted to the $\mathbf{Y} = (\mathbf{X} - \widehat{\boldsymbol{\mu}}_{\text{ML-cf}})^T$. For this, we set $\mathbf{t}_j = \mathbf{s}_j$, for $j = 1, \ldots, 12$. Table 1 shows the results for modelling data through five methods. We note that estimated tail indices are $\widehat{\alpha}_{\text{MU}} = 1.581$, $\widehat{\alpha}_{\text{MM}} = 1.734$, $\widehat{\alpha}_{\text{ML-cf}} = 1.618$, $\widehat{\alpha}_{\text{SQ-cf}} = 1.493$, and $\widehat{\alpha}_{\text{CF-cf}} = 1.723$. As it is seen, estimated tail indices through estimators $\widehat{\alpha}_{\text{MU}}$ and $\widehat{\alpha}_{\text{ML-cf}}$ are closer together than

the other estimators. In the second example, we focus on the cubic-root of the monthly average of river discharge. We choose discharge of the Odra and Wisla rivers in Poland during 1901 to 1986 (raw data are in m^3/s. They are available at https://nelson.wisc.edu/sage/data-and-models/riverdata/). The scatter plot for cubic-root of Odra river discharge versus cubic-root of Wisla river discharge is shown in Figure 8. Setting $m = 8$, $\mathbf{s}_j = (\cos(2\pi(j-1)/m), \sin(2\pi(j-1)/m))^T$, and
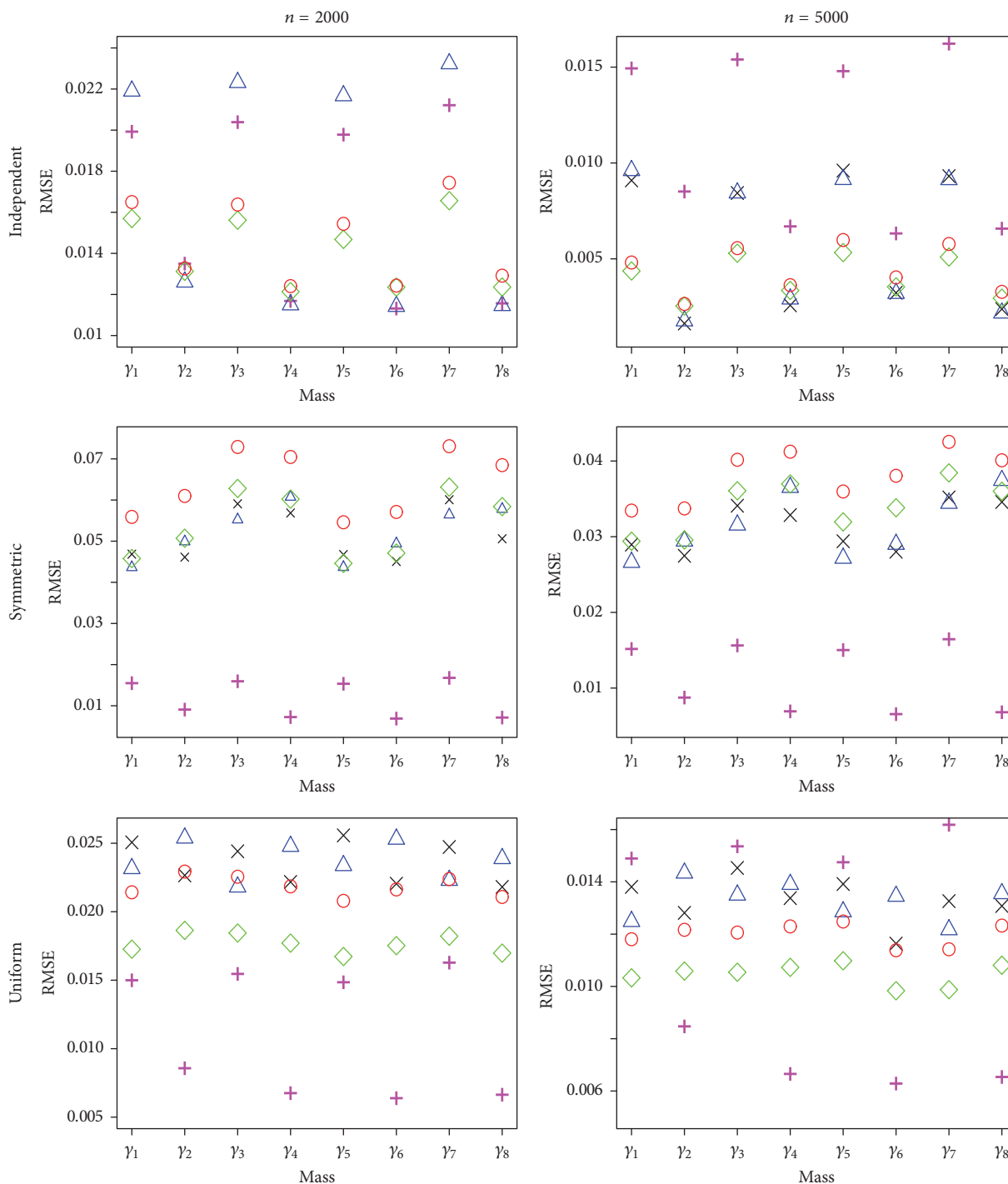
FIGURE 3: RMSEs of $\widehat{\gamma}$ under different scenarios when $\alpha = 1.25$. We use the following symbol scheme: $\diamond$ for $\widehat{\gamma}_{\text{UM}}$, $\bigcirc$ for $\widehat{\gamma}_{\text{MM}}$, $+$ for $\widehat{\gamma}_{\text{ML-cf}}$, $\times$ for $\widehat{\gamma}_{\text{SQ-cf}}$, and $\triangle$ for $\widehat{\gamma}_{\text{CF-cf}}$.

$\mathbf{t}_j = \mathbf{s}_j$, for $j = 1, \ldots, 8$, we obtain $\widehat{\boldsymbol{\mu}}_{\text{ML-cf}} = (7.9837, 9.9947)^T$. After fitting a strictly α-stable distribution to the shifted data, results for estimating spectral measure are given in Table 2. Estimated tail indices are $\widehat{\alpha}_{\text{MU}} = 1.860$, $\widehat{\alpha}_{\text{MM}} = 1.312$, $\widehat{\alpha}_{\text{ML-cf}} = 1.813$, $\widehat{\alpha}_{\text{SQ-cf}} = 1.936$, and $\widehat{\alpha}_{\text{CF-cf}} = 1.962$. Based on

results given in Table 2, estimated masses through estimators $\widehat{\boldsymbol{\gamma}}_{\text{MU}}$, $\widehat{\boldsymbol{\gamma}}_{\text{ML-cf}}$, and $\widehat{\boldsymbol{\gamma}}_{\text{SQ-cf}}$ are closer together than the other estimators. We compare here $\widehat{\boldsymbol{\gamma}}_{\text{MU}}$ with $\widehat{\boldsymbol{\gamma}}_{\text{ML-cf}}$ and $\widehat{\boldsymbol{\gamma}}_{\text{SQ-cf}}$ since the latter estimators are among the best estimators for the masses as shown in the previous subsection.
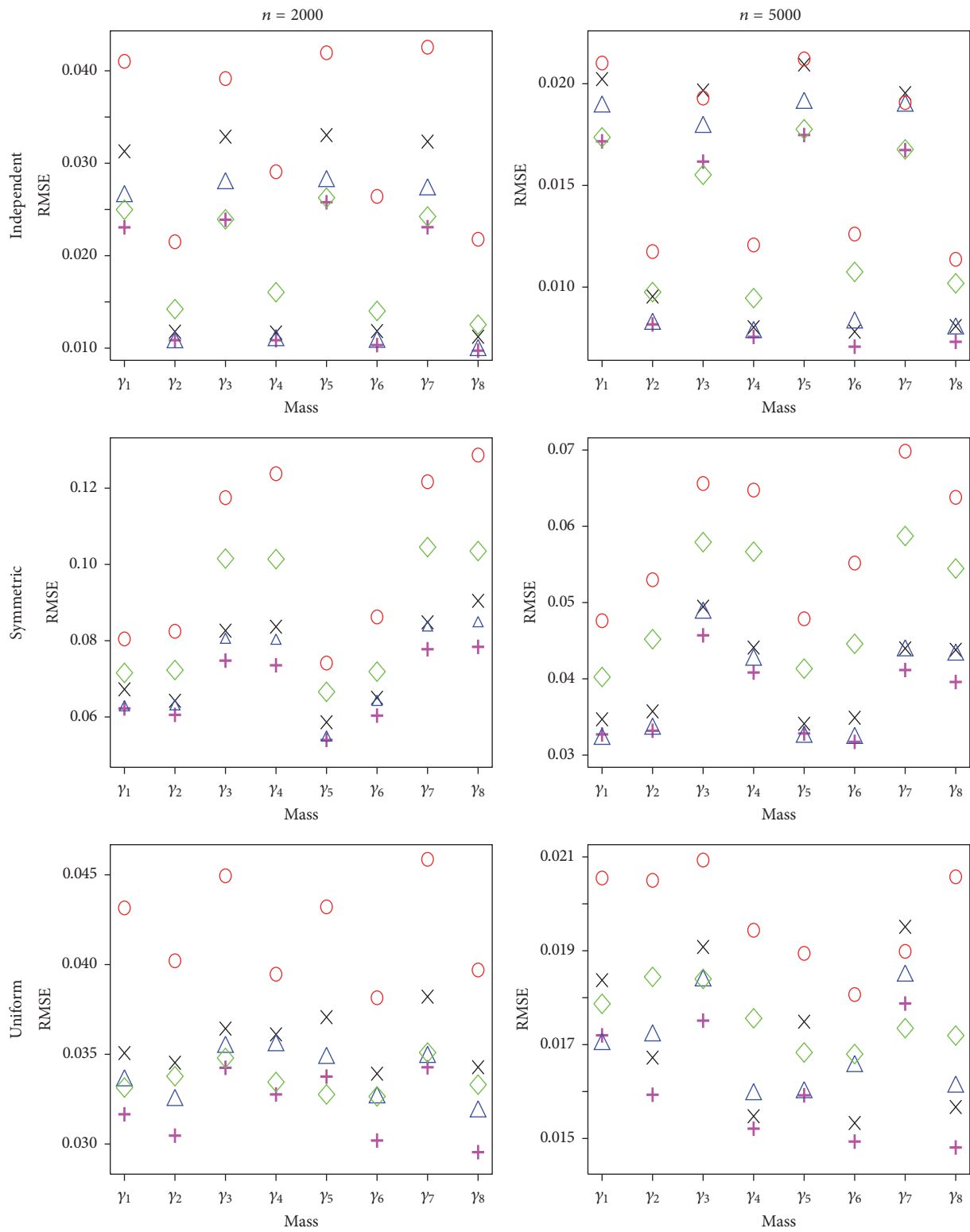
FIGURE 4: RMSEs of $\widehat{\gamma}$ under different scenarios when $\alpha = 1.75$. We use the following symbol scheme: $\diamond$ for $\widehat{\gamma}_{\text{UM}}$, $\bigcirc$ for $\widehat{\gamma}_{\text{MM}}$, $+$ for $\widehat{\gamma}_{\text{ML-cf}}$, $\times$ for $\widehat{\gamma}_{\text{SQ-cf}}$, and $\triangle$ for $\widehat{\gamma}_{\text{CF-cf}}$.
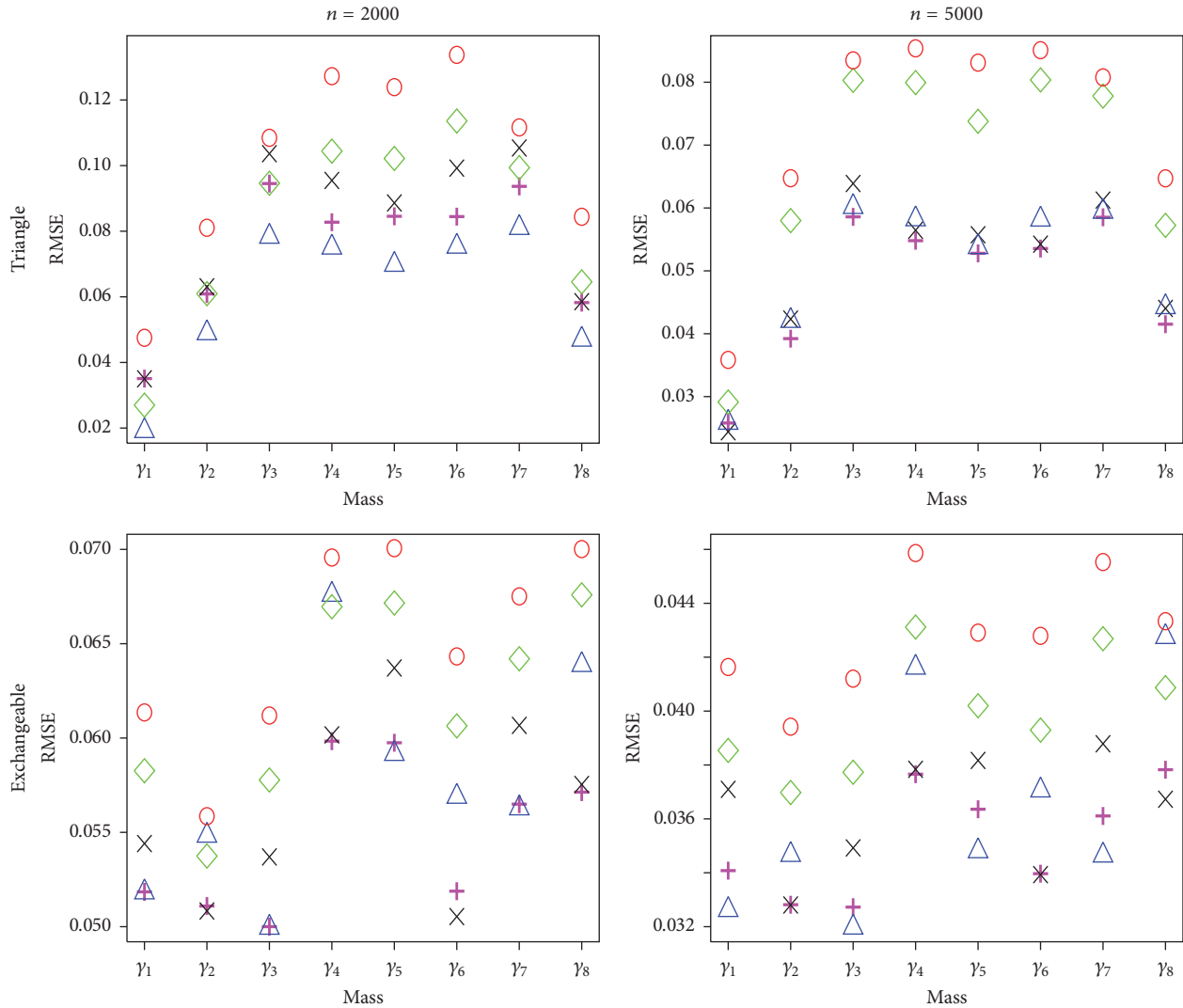
FIGURE 5: RMSEs of $\widehat{\gamma}$ under different scenarios when $\alpha = 1.25$. We use the following symbol scheme: $\diamond$ for $\widehat{\gamma}_{\text{UM}}$, $\circ$ for $\widehat{\gamma}_{\text{MM}}$, $+$ for $\widehat{\gamma}_{\text{ML-cf}}$, $\times$ for $\widehat{\gamma}_{\text{SQ-cf}}$, and $\triangle$ for $\widehat{\gamma}_{\text{CF-cf}}$.

4. Conclusion

We compare the performance of the introduced U-statistic for the tail index with the well-known methods, including maximum likelihood, empirical characteristic function, sample quantile, and that introduced in Mohammadi et al. [11] through a simulation study. In the sense of root mean-squared error, it is proved that proposed tail index estimator always outperforms Mohammadi et al. [11] and SQ methods when $\alpha \leq 1.4$. This is while ML and CF methods show better performance than the proposed estimator for large α, say $\alpha > 1.4$ in terms of root mean-squared error. Simulation studies for estimating the discrete spectral measure γ under five scenarios prove that estimator of γ based on introduced U-statistic shows, in terms of root mean-squared error, better performance than Mohammadi et al. [11] estimator. Analysis of two sets of real data reveals that estimator of the tail index and γ based on U-statistic shows expedient performance. As some possible future works, firstly, we aim to introduce

a U-statistic for the case of a nonzero location parameter. Secondly, we look for methodology possibly based on a U-statistic, to estimate tail, masses, and location parameters simultaneously. Finally, recalling that the approach employed in this work is based on characteristic function, the discrete spectral measure using $\widehat{\alpha}_{\text{MU}}$ can be estimated through projection approach.

Appendix

Proof of Lemma 3. We show that $E(\log^2 \|\mathbf{X}\|) < \infty$. Suppose $d = 2$ and $p^- = P(\|\mathbf{X}\| \leq 1)$, $p^+ = 1 - p^-$, $p^{--} = P(X_1 < 0, X_2 < 0)$, $p^{-+} = P(X_1 < 0, X_2 > 0)$, $p^{+-} = P(X_1 > 0, X_2 < 0)$, and $p^{++} = P(X_1 > 0, X_2 > 0)$. So

$$E\left(\log^2 \|\mathbf{X}\|\right) = E\left(\log^2 \|\mathbf{X}\| \mid \|\mathbf{X}\| < 1\right) p^-$$

$$+ E\left(\log^2 \|\mathbf{X}\| \mid \|\mathbf{X}\| \geq 1\right) p^+$$
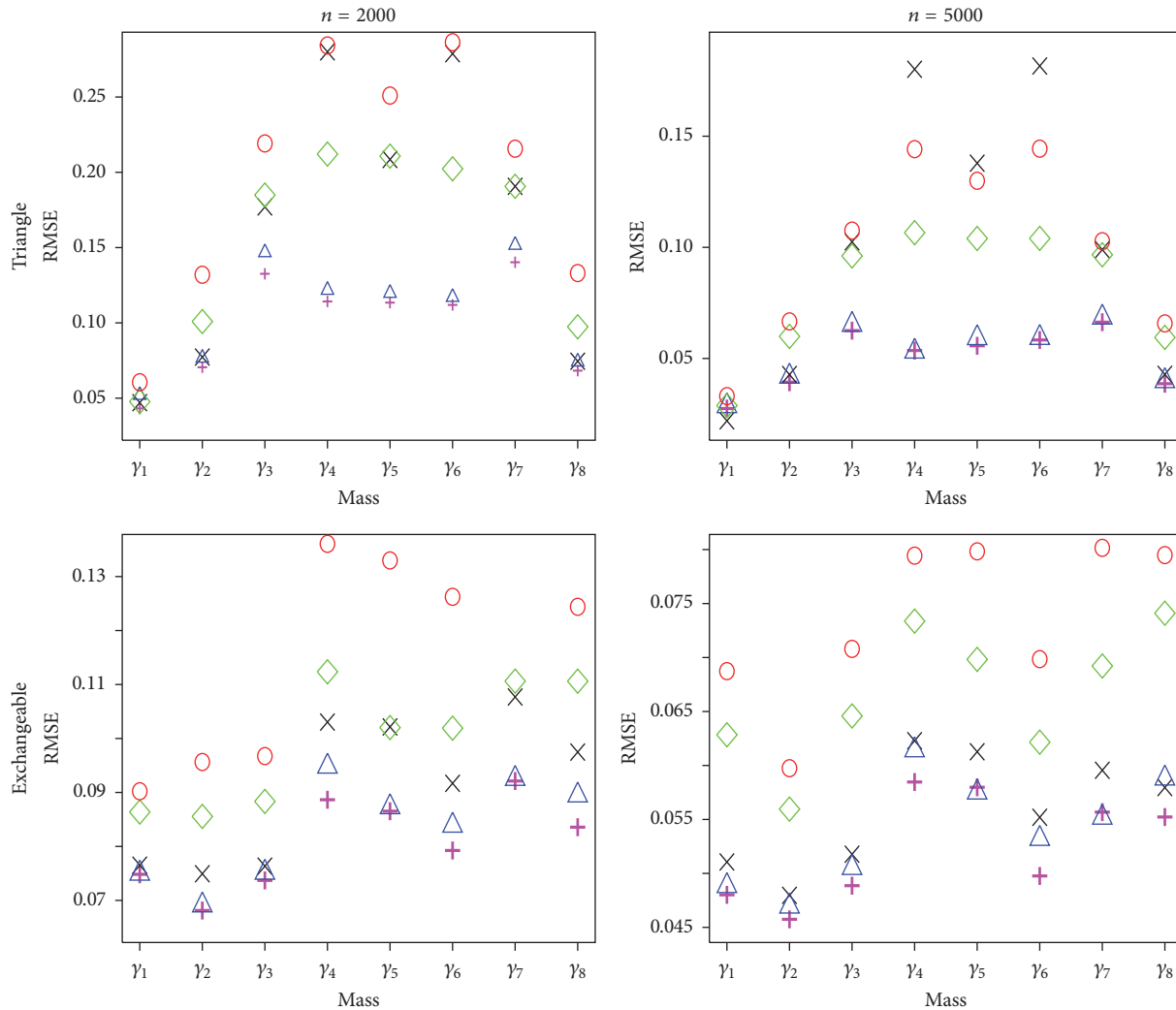
FIGURE 6: RMSEs of $\widehat{\gamma}$ under different scenarios when $\alpha = 1.75$. We use the following symbol scheme: $\diamond$ for $\widehat{\gamma}_{\text{UM}}$, $\circ$ for $\widehat{\gamma}_{\text{MM}}$, $+$ for $\widehat{\gamma}_{\text{ML-cf}}$, $\times$ for $\widehat{\gamma}_{\text{SQ-cf}}$, and $\triangle$ for $\widehat{\gamma}_{\text{CF-cf}}$.

TABLE 1: Estimation results after fitting a strictly bivariate α-stable distribution to AXP and MRK stocks data.

Estimator	$\widehat{\gamma}_1$	$\widehat{\gamma}_2$	$\widehat{\gamma}_3$	$\widehat{\gamma}_4$	$\widehat{\gamma}_5$	$\widehat{\gamma}_6$	$\widehat{\gamma}_7$	$\widehat{\gamma}_8$	$\widehat{\gamma}_9$	$\widehat{\gamma}_{10}$	$\widehat{\gamma}_{11}$	$\widehat{\gamma}_{12}$
$\widehat{\gamma}_{\text{MU}}$	0.396	0.070	0.077	0.433	0.087	0	0.231	0.378	0	0.439	0.078	0
$\widehat{\gamma}_{\text{MM}}$	0.373	0	0	0.463	0.021	0	0.333	0.479	0	0.563	0.100	0
$\widehat{\gamma}_{\text{ML-cf}}$	0.338	0.156	0.162	0.412	0	0	0.490	0.206	0.055	0.425	0.122	0
$\widehat{\gamma}_{\text{SQ-cf}}$	0.232	0.216	0.089	0.404	0.012	0	0.480	0.115	0.177	0.254	0.156	0
$\widehat{\gamma}_{\text{CF-cf}}$	0.421	0.129	0.237	0.434	0	0	0.537	0.209	0	0.527	0.089	0

TABLE 2: Estimation results after fitting a strictly bivariate α-stable distribution to Odra and Wisla discharge data.

Estimator	$\widehat{\gamma}_1$	$\widehat{\gamma}_2$	$\widehat{\gamma}_3$	$\widehat{\gamma}_4$	$\widehat{\gamma}_5$	$\widehat{\gamma}_6$	$\widehat{\gamma}_7$	$\widehat{\gamma}_8$
$\widehat{\gamma}_{\text{MU}}$	0.006	1.459	0.498	0	0	0	0	0
$\widehat{\gamma}_{\text{MM}}$	0	0.800	0.257	0	0.026	0.483	0.165	0
$\widehat{\gamma}_{\text{ML-cf}}$	0	1.380	0.508	0	0	0	0	0
$\widehat{\gamma}_{\text{SQ-cf}}$	0	1.490	0.532	0	0	0	0	0
$\widehat{\gamma}_{\text{CF-cf}}$	0	0.196	0.199	0	0.009	1.283	0.339	0
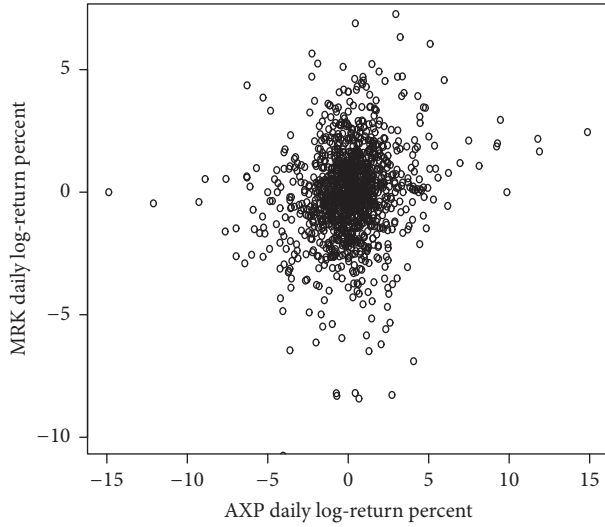
FIGURE 7: Scatter plot for AXP versus MRK daily log-return percent, $\mathbf{X} = (\text{AXP}, \text{MRK})^T$.
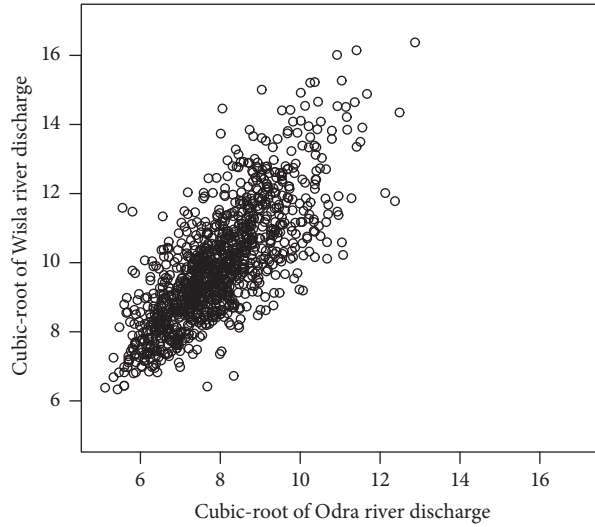
FIGURE 8: Scatter plot for cubic-root of Odra and Wisla rivers discharge.

$$\leq E\left(\log^2 |X| \mid \|\mathbf{X}\| < 1\right) p^-$$
$$+ E\left(\log^2 \|\mathbf{X}\| \mid \|\mathbf{X}\| \geq 1\right) p^+$$
$$\leq E\left(\log^2 |X|\right)$$
$$+ E\left(\log^2 \|\mathbf{X}\| \mid \|\mathbf{X}\| \geq 1\right),$$

(A.1)

where X, in the above, denotes one of the components of vector $\mathbf{X}$. It should be noted that inequality $E(\log^2 \|\mathbf{X}\| \mid \|\mathbf{X}\| < 1) \leq E(\log^2 |X| \mid \|\mathbf{X}\| < 1)$ holds irrespective of d. On the other hand,

$$E\left(\log^2 \|\mathbf{X}\| \mid \|\mathbf{X}\| \geq 1\right)$$
$$= E\left(\log^2 \|\mathbf{X}\| \mid \|\mathbf{X}\| \geq 1, X_1 < 0, X_2 < 0\right) p^{--}$$

$$+ E\left(\log^2 \|\mathbf{X}\| \mid \|\mathbf{X}\| \geq 1, X_1 < 0, X_2 > 0\right) p^{-+}$$
$$+ E\left(\log^2 \|\mathbf{X}\| \mid \|\mathbf{X}\| \geq 1, X_1 > 0, X_2 < 0\right) p^{+-}$$
$$+ E\left(\log^2 \|\mathbf{X}\| \mid \|\mathbf{X}\| \geq 1, X_1 > 0, X_2 > 0\right) p^{++}$$
$$\leq E\left(\log^2 |X_1 + X_2| \mid \|\mathbf{X}\| \geq 1, X_1 < 0, X_2 < 0\right)$$
$$\cdot p^{--}$$
$$+ E\left(\log^2 |-X_1 + X_2| \mid \|\mathbf{X}\| \geq 1, X_1 < 0, X_2 > 0\right)$$
$$\cdot p^{-+}$$
$$+ E\left(\log^2 |X_1 - X_2| \mid \|\mathbf{X}\| \geq 1, X_1 > 0, X_2 < 0\right)$$
$$\cdot p^{+-}$$
$$+ E\left(\log^2 |X_1 + X_2| \mid \|\mathbf{X}\| \geq 1, X_1 > 0, X_2 > 0\right)$$
$$\cdot p^{++} \leq 2E\left(\log^2 |X_1 + X_2|\right)$$
$$+ 2E\left(\log^2 |X_1 + X_2|\right).$$

(A.2)

Thus,

$$E\left(\log^2 \|\mathbf{X}\|\right) \leq E\left(\log^2 |X|\right) + 2E\left(\log^2 |X_1 + X_2|\right)$$
$$+ 2E\left(\log^2 |X_1 - X_2|\right).$$

(A.3)

Generally, for $d \geq 2$, one can write

$$E\left(\log^2 \|\mathbf{X}\|\right)$$
$$\leq E\left(\log^2 |X|\right)$$
$$+ \sum_{i=0}^{d} \binom{d}{i} E\left(\log^2 \left| -\sum_{j=1}^{i} X_j + X_{i+1} + \cdots + X_d \right| \right),$$

(A.4)

where we adopt this convention that $\sum_{j=1}^{0} X_j = 0$. Let $S(\alpha, \beta, \sigma, \mu = 0)$ stands for a univariate strictly stable random variable with tail index α, scale parameter σ, and skewness parameter β. It is well known that if $\mathbf{X} = (X_1, \ldots, X_d)^T$ is an α-stable random vector, then any linear combination of its components such as $\langle \mathbf{b_i}, \mathbf{X} \rangle = \sum_{j=1}^{d} b_{ij} X_j$, for $\mathbf{b}_i = (\overbrace{-1, \ldots, -1}^{i}, \overbrace{1, \ldots, 1}^{d-i})^T$, follows a stable distribution with tail index α,

$$\sigma_i = \left(\int_{\mathbb{S}^d} |\langle \mathbf{b_i}, \mathbf{s} \rangle|^\alpha \, \Gamma(d\mathbf{s}) \right)^{1/\alpha},$$

$$\beta_i = \frac{1}{\sigma_i^\alpha} \left(\int_{\mathbb{S}^d} |\langle \mathbf{b_i}, \mathbf{s} \rangle|^{[\alpha]} \, \Gamma(d\mathbf{s}) \right)^{1/\alpha},$$

(A.5)

where $\Gamma(\cdot)$ is spectral measure and $|x|^{[r]} = \operatorname{sgn}(x)|x|^r$; see Samorodnitsky and Taqqu [2]. It follows, from Kuruoglu [17], that if $X \sim S(\alpha, \beta, \sigma, 0)$, then

$$e(\alpha, \beta, \sigma) = E\left(\log^2 |X|\right)$$

$$= \frac{7\pi^2 - 6\theta^2 + 6\left[\log\left(\sigma/\cos\left(\theta\right)\right) + \gamma\left(1 - \alpha\right)\right]^2 - \pi^2\alpha^2}{6\alpha^2}, \quad \text{(A.6)}$$

where $\theta = \arctan(\beta \tan(\pi\alpha/2))$. Also,

$$e(\alpha, \beta_i, \sigma_i) = E\left(\log^2 |Z|\right)$$

$$= \frac{7\pi^2 - 6\theta_i^2 + 6\left[\log\left(\sigma_i/\cos\left(\theta_i\right)\right) + \gamma\left(1 - \alpha\right)\right]^2 - \pi^2\alpha^2}{6\alpha^2}, \quad \text{(A.7)}$$

where $Z = -\sum_{j=1}^{i} X_j + X_{i+1} + \cdots + X_d$, for $i = 0, \ldots, d$ and $\theta_i = \arctan(\beta_i \tan(\pi\alpha/2))$. Parameters σ_i and β_i are defined in (A.5). Finally,

$$E\left(\log^2 \|\mathbf{X}\|\right) \leq e(\alpha, \beta, \sigma) + \sum_{i=0}^{d} \binom{d}{i} e(\alpha, \beta_i, \sigma_i). \quad \text{(A.8)}$$

The proof is complete since all terms on the right-hand side of (A.8) are finite. $\square$

Proof of Theorem 4. We rewrite Definition 1 as

$$A\mathbf{X}_1 + B\mathbf{X}_2 \overset{d}{=} C\mathbf{X}_1 + \mathbf{D}. \quad \text{(A.9)}$$

Setting $A = 1$, $B = 1$, and $\mathbf{D} = \mathbf{0}$ in (A.9), it yields

$$\mathbf{X}_1 + \mathbf{X}_2 \overset{d}{=} 2^{1/\alpha}\mathbf{X}_1. \quad \text{(A.10)}$$

By applying log-transformation, after taking the Euclidean norm, to both sides of (A.10), we have

$$\frac{1}{\alpha} = \frac{\log \|\mathbf{X}_1 + \mathbf{X}_2\| - \log \|\mathbf{X}_1\|}{\log 2}. \quad \text{(A.11)}$$

The right-hand side of (A.11) can be used to define a symmetric kernel of the form

$$H(\mathbf{X}_1, \mathbf{X}_2) = \frac{\log \|\mathbf{X}_1 + \mathbf{X}_2\|}{\log 2}$$

$$- \frac{\log \|\mathbf{X}_1\| + \log \|\mathbf{X}_2\|}{2 \log 2}. \quad \text{(A.12)}$$

To guarantee the asymptotic normality of the introduced U-statistics for $1/\alpha$ with kernel (A.12), we need to check that $E(H(\mathbf{X}_1, \mathbf{X}_2))^2 < \infty$. It suffices to show that $\operatorname{Var} H(\mathbf{X}_1, \mathbf{X}_2) < \infty$. For this, the result of Lemma 3 shows that $\operatorname{Var} \log \|\mathbf{X}_1\|$ is finite. On the other hand, from (A.10) it turns out that

$$\operatorname{Var} \log \|\mathbf{X}_1 + \mathbf{X}_2\| = \operatorname{Var} \log \|\mathbf{X}_1\| = \operatorname{Var} \log \|\mathbf{X}_2\|. \quad \text{(A.13)}$$

We use property (A.13) to calculate variance of the right-hand side of (A.12) as

$$\operatorname{Var} H(\mathbf{X}_1, \mathbf{X}_2) = \frac{\operatorname{Var} \log \|\mathbf{X}_1\|}{\log^2 2} + \frac{\operatorname{Var} \log \|\mathbf{X}_1\|}{2 \log^2 2}$$

$$- \frac{K}{\log^2 2}, \quad \text{(A.14)}$$

where

$$K = \operatorname{Cov}\left(\log \|\mathbf{X}_1 + \mathbf{X}_2\|, \log \|\mathbf{X}_1\| + \log \|\mathbf{X}_2\|\right)$$

$$\leq 2\sqrt{\operatorname{Var} \log \|\mathbf{X}_1 + \mathbf{X}_2\|}\sqrt{\operatorname{Var} \log \|\mathbf{X}_1\|}. \quad \text{(A.15)}$$

Applying property (A.13) again on the right-hand side of (A.15), we have

$$\operatorname{Var} H(\mathbf{X}_1, \mathbf{X}_2) \leq \frac{7 \operatorname{Var} \log \|\mathbf{X}_1\|}{2 \log^2 2}, \quad \text{(A.16)}$$

where we used the result of Lemma 3 to get the right-hand side of (A.16). This means that

$$\lambda = \operatorname{Var}\left(E\left(H(\mathbf{X}_1, \mathbf{X}_2) \mid \mathbf{X}_1\right)\right) \leq \operatorname{Var} H(\mathbf{X}_1, \mathbf{X}_2)$$

$$< \infty. \quad \text{(A.17)}$$

Therefore, $E(H(\mathbf{X}_1, \mathbf{X}_2))^2 < \infty$. Now, we define U-statistic for $1/\alpha$ with kernel given in (A.12) as

$$U_n = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} H(\mathbf{x}_i, \mathbf{x}_j). \quad \text{(A.18)}$$

By definition, given U-statistic in (A.18) is unbiased estimator for $1/\alpha$. $\square$

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] P. Lévy, "Théorie des erreurs. La loi de Gauss et les lois exceptionnelles," *Bulletin De La Société Mathématique De France*, vol. 52, pp. 49–85, 1924.

[2] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models and Infinite Variance*, Stochastic Modeling, Chapman & Hall, New York, NY, USA, 1994.

[3] J. P. Nolan, "Maximum likelihood estimation of stable parameters," in *Lévy Processes: Theory and Applications*, O. E. Barndorff-Nielsen, T. Mikosch, and I. Resnick, Eds., pp. 379–400, Birkhöuser, Boston, Mass, USA, 2001.

[4] J. P. Nolan, "Numerical calculation of stable densities and distribution functions," *Communications in Statistics. Stochastic Models*, vol. 13, no. 4, pp. 759–774, 1997.

[5] J. H. McCulloch, "Simple consistent estimators of stable distribution parameters," *Communications in Statistics—Simulation and Computation*, vol. 15, no. 4, pp. 1109–1136, 1986.

[6] S. M. Kogon and D. B. Williams, "Characteristic function based estimation of stable parameters," in *A Practical Guide to Heavy Tailed Data*, R. Adler, R. Feldman, and M. Taqqu, Eds., pp. 311–338, Birkhäuser, Boston, Mass, USA, 1998.

[7] Z. Fan, "Parameter estimation of stable distributions," *Communications in Statistics. Theory and Methods*, vol. 35, no. 1-3, pp. 245–255, 2006.

[8] J. P. Nolan, A. K. Panorska, and J. H. McCulloch, "Estimation of stable spectral measures," *Mathematical and Computer Modelling*, vol. 34, no. 9–11, pp. 1113–1122, 2001.

 [9] M. Pivato and L. Seco, "Estimating the spectral measure of a multivariate stable distribution via spherical harmonic analysis," *Journal of Multivariate Analysis*, vol. 87, no. 2, pp. 219–240, 2003.

[10] H. Ogata, "Estimation for multivariate stable distributions with generalized empirical likelihood," *Journal of Econometrics*, vol. 172, no. 2, pp. 248–254, 2013.

[11] M. Mohammadi, A. Mohammadpour, and H. Ogata, "On estimating the tail index and the spectral measure of multivariate $\alpha\alpha$-stable distributions," *Metrika*, vol. 78, no. 5, pp. 549–561, 2015.

[12] T. Byczkowski, J. P. Nolan, and B. Rajput, "Approximation of multidimensional stable densities," *Journal of Multivariate Analysis*, vol. 46, no. 1, pp. 13–31, 1993.

[13] J. H. McCulloch, "Estimation of the bivariate stable spectral representation by the projection method," *Computational Economics*, vol. 16, no. 1-2, pp. 47–62, 2000.

[14] R. Modarres and J. P. Nolan, "A method for simulating stable random vectors," *Computational Statistics*, vol. 9, no. 1, pp. 11–19, 1994.

[15] Robust Analysis Inc, *User Manual for STABLE 5.0*, Software and User Manual, 2010, http://www.robustanalysis.com.

[16] J. P. Nolan, "Multivariate elliptically contoured stable distributions: theory and estimation," *Computational Statistics*, vol. 28, no. 5, pp. 2067–2089, 2013.

[17] E. E. Kuruoglu, "Density parameter estimation of skewed /spl alpha/-stable distributions," *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2192–2201, 2001.

On a Power Transformation of Half-Logistic Distribution

S. D. Krishnarani

Department of Statistics, Farook College, Kozhikode, Kerala 673632, India

Correspondence should be addressed to S. D. Krishnarani; krishnaranisd@gmail.com

Academic Editor: Aera Thavaneswaran

A new continuous distribution on the positive real line is constructed from half-logistic distribution, using a transformation and its analytical characteristics are studied. Some characterization results are derived. Classical procedures for the estimation of parameters of the new distribution are discussed and a comparative study is done through numerical examples. Further, different families of continuous distributions on the positive real line are generated using this distribution. Application is discussed with the help of real-life data sets.

1. Introduction

Distributions defined on the positive real line are widely used in modeling of survival data. The Weibull, Pareto, and Exponential distributions have major roles in fitting data sets in the fields of computer science, engineering, biology, and so forth. Half-logistic distribution (HLD) is another life distribution used in reliability analysis by many researchers. The half-logistic random variable studied by Balakrishnan [1, 2] has survival function

$$\overline{F}(x) = \frac{2}{1 + e^x}, \quad x > 0. \tag{1}$$

By imparting the location and scale parameters, its probability density function (pdf) is

$$f(x; \mu, \sigma) = \frac{2e^{((x-\mu)/\sigma)}}{\sigma \cdot \left(1 + e^{((x-\mu)/\sigma)}\right)^2}, \quad x \geq \mu, \ \sigma > 0. \tag{2}$$

In the past few years many researchers have paid much attention to this distribution and several generalizations have been introduced. Srinivasa Rao et al. [3], Olapade [4–6], Cordeiro et al. [7], Kantam et al. [8], and so forth are some of the recent works in this area.

It is well known that through power transformation, the Weibull is an extension of exponential while the power function distribution is that of uniform. Hence, it is of interest to know what would be the distribution of similar power transformation of half-logistic distributions. Motivated by this, in the present paper, we introduce a new continuous distribution on the positive real line using a transformation of half-logistic random variable. We study the properties and applications of this so-called generalization of half-logistic distribution.

The remaining part of the paper is organized as follows. In Section 2, Power Half-Logistic Distribution is introduced and its properties are studied. In Section 3, some characterization results are derived. Estimation of the parameters is done in Section 4 and numerical illustrations are given therein. Section 5 deals with extensions of this new transformed distribution. Application to real data sets is considered in Section 6 followed by a concluding section at the end.

2. Power Half-Logistic Distribution

Let Y be a random variable following half-logistic distribution with survival function

$$\overline{F}_Y(y) = \frac{2}{1 + e^{\beta y}}, \quad y > 0, \ \beta > 0. \tag{3}$$

Consider the transformation, $X = Y^{1/\alpha}$. Then the survival function of X is

$$\overline{F}_X(x) = \frac{2}{1 + e^{\beta x^\alpha}}, \quad x > 0, \ \alpha, \beta > 0 \tag{4}$$
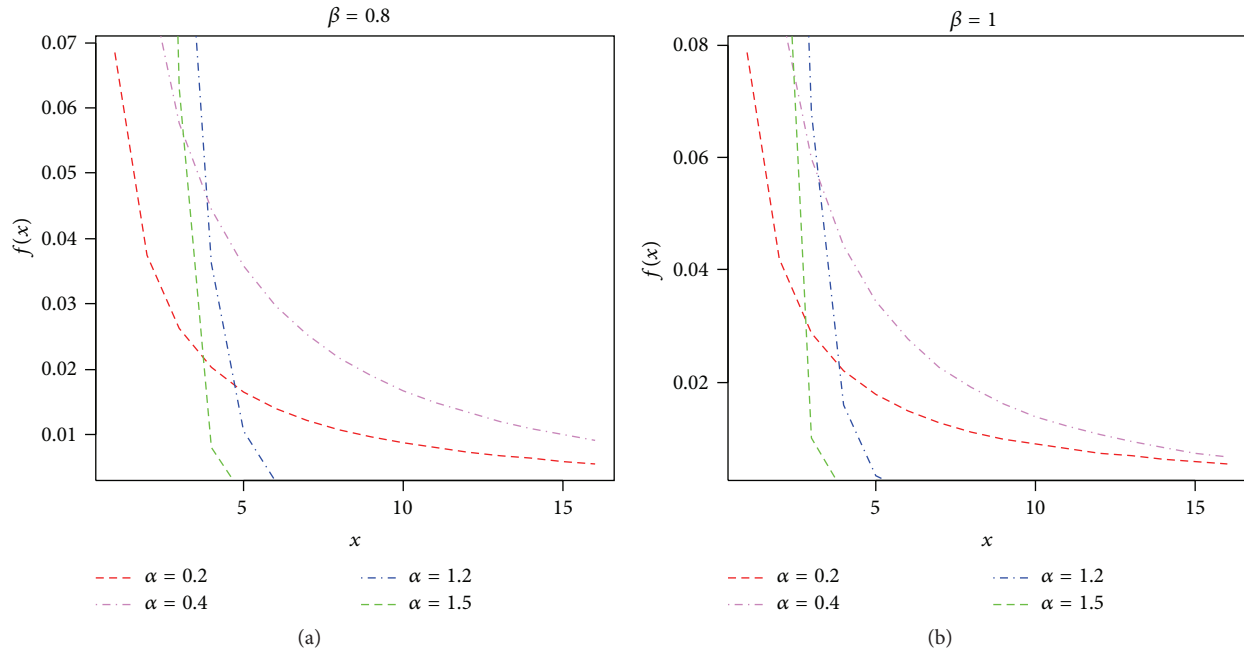
FIGURE 1: pdfs of PHLD for various values of the parameters.

and its pdf is

$$f(x) = \frac{2\alpha\beta e^{\beta x^\alpha} x^{\alpha-1}}{\left(1 + e^{\beta x^\alpha}\right)^2}, \quad x > 0, \ \alpha > 0, \ \beta > 0. \tag{5}$$

Hereinafter we call the random variable X with pdf (5) as the Power Half-Logistic Distribution (PHLD).

The graphical form of the pdf for various values of α and β is given in Figure 1 ($\beta = 0.8$ and $\beta = 1$). When β is fixed and α is increasing in $(0, 1)$, the pdf becomes more convex. But when α is >1 it becomes concave. Another characteristic is that for $0 < \alpha < 1$, the distribution is heavy tailed but for $\alpha > 1$, the distribution is light tailed as compared to HLD (see Figure 2). So it can be used to model data sets having tail probability less or greater than HLD. The plots of survival function and hazard rates for different values of α when $\beta = 0.5$ and 1.5 are shown in Figures 3 and 4, respectively, for an alternate view of the behavior of the distribution.

Next we explore the analytical properties of the PHLD, deriving its moments, median, quantiles, hazard function, and log odds function, and summarize them below.

Properties. (1) The sth moment $E(X^s) = (2s/(\beta^s \cdot \alpha))\Gamma(s/\alpha) \sum_{j=0}^\infty ((-1)^j/(j+1)^{s/\alpha})$.

(2) Median = $((\log 3)/\beta)^{1/\alpha}$.

(3) The pth quantile is $((1/\beta) \log((1+p)/(1-p)))^{1/\alpha}$.

(4) Hazard rate $r(x) = \alpha\beta e^{\beta x^\alpha} x^{\alpha-1}/(1 + e^{\beta x^\alpha})$. It can be seen in Figure 4 that for $\alpha > 1$, the distribution has increasing failure rate (IFR), but for $0 < \alpha < 1$ the distribution has decreasing failure rate (DFR).

(5) The log odds function is $\log[(e^{\beta x^\alpha} - 1)/2]$.

It may be noted that in a recent study on the exponentiated half-logistic family by Cordeiro et al. [7] a special case of it called half-logistic Weibull distribution has been just

mentioned without any elaborate study. They proposed a new exponentiated half-logistic (EHL) family as a competitive alternative for lifetime data analysis. For any parent continuous distribution G they defined the corresponding EHL-G distribution with distribution function $[(1 - \overline{G}(x)^\lambda)/(1 + \overline{G}(x)^\lambda)]^\alpha$. This new family extends several common distributions such as Frechet, normal, log-normal, Gumbel, and log-logistic distributions. It is interesting to observe that PHLD is the same as the distribution pointed out there when $G(x) = 1 - \exp(-x^\beta)$.

Next we derive some characterization results of PHLD.

3. Characterizations

In the first characterization we establish a relationship between the PHLD and Weibull distribution.

Result 1. Suppose $\overline{F}(x)$ and $\overline{G}(x)$ are survival functions with respective pdfs f and g. Then in the equation

$$\frac{d}{dx}\left(\frac{1}{\overline{G}(x)}\right) = \frac{1}{2}\frac{d}{dx}\left(\frac{1}{\overline{F}(x)}\right) \tag{6}$$

or

$$\frac{g(x)}{\left(\overline{G}(x)\right)^2} = \frac{1}{2}\frac{f(x)}{\left(\overline{F}(x)\right)^2}, \tag{7}$$

$g(x)$ has PHLD if, and only if, $f(x)$ is Weibull.

Proof. Suppose the pdf $g(x)$ is PHLD with the form as in (5) and then substituting in (7) and further on integrating, we get

$$\overline{F}(x) = e^{-\beta x^\alpha}, \tag{8}$$

which is the survival function of Weibull random variable.
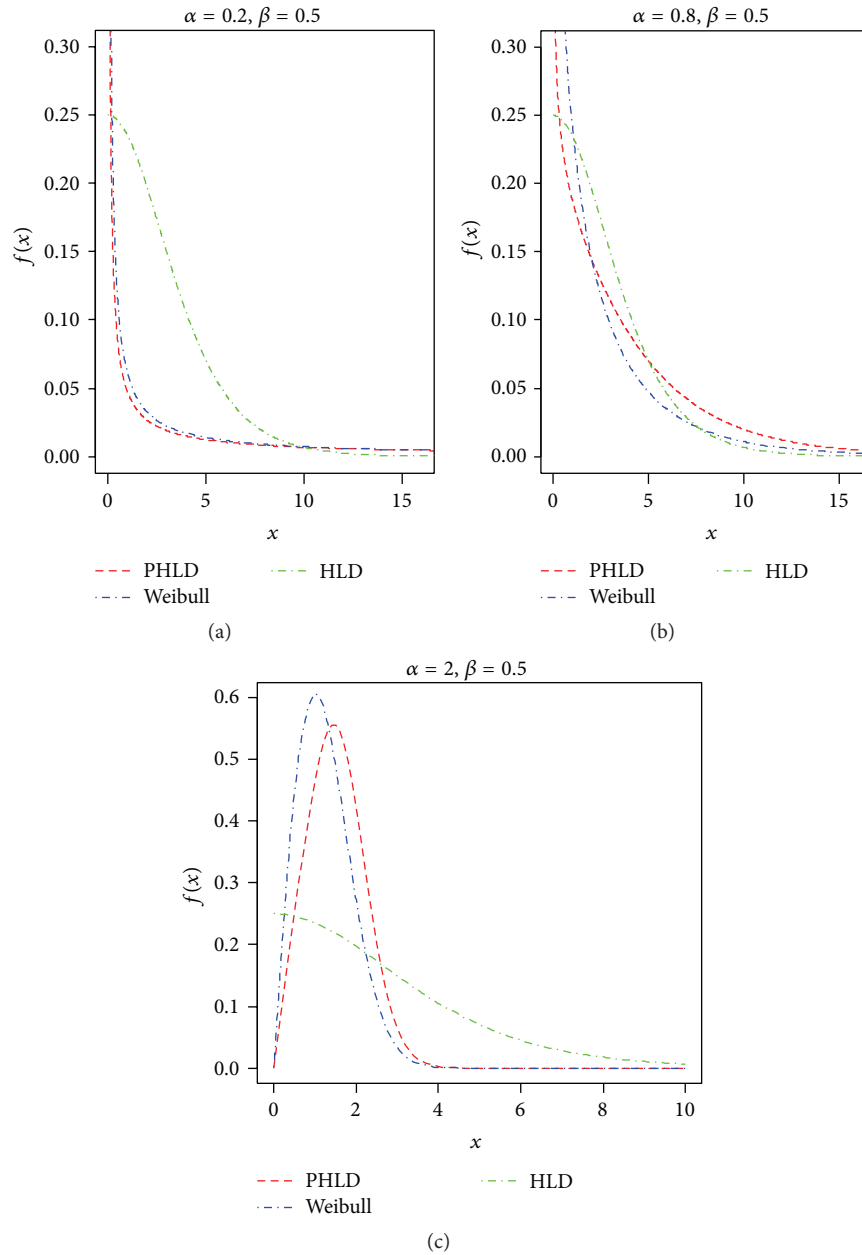
FIGURE 2: Comparison of PHLD, HLD, and Weibull densities.

Conversely assuming $f(x)$ as Weibull with pdf

$$f(x) = e^{-\beta x^\alpha} \alpha \beta x^{\alpha-1} \tag{9}$$

and substituting in (7), we get

$$\overline{G}(x) = \frac{2}{1 + e^{\beta x^\alpha}}, \tag{10}$$

which is the survival function of PHLD. □

Result 2. The function $g(x)$ in

$$x^{1/\alpha} \frac{g\left(x^{1/\alpha}\right)}{x} = \alpha f(x) \tag{11}$$

is the pdf of PHLD if, and only if, $f(x)$ is the pdf of HLD.

Proof. The proof easily follows. □

Result 3. For a survival function $\overline{G}(x)$, the functional equation

$$f\left(\left(x^\alpha + y^\alpha\right)^{1/\alpha}\right) = f(x) f(y) \tag{12}$$

(a variant of Cauchy's equation) is satisfied by $f(x) = 2/\overline{G}(x) - 1$ if, and only if, $\overline{G}(x) = 2/(1 + e^{cx^\alpha})$.

Proof. Suppose $f(x) = 2/\overline{G}(x) - 1$ satisfies the given functional equation.

Then there exists a constant c such that $f(x) = e^{cx^\alpha}$ (see [9]). Hence, $\overline{G}(x) = 2/(1 + e^{cx^\alpha})$.

(a)   (b)
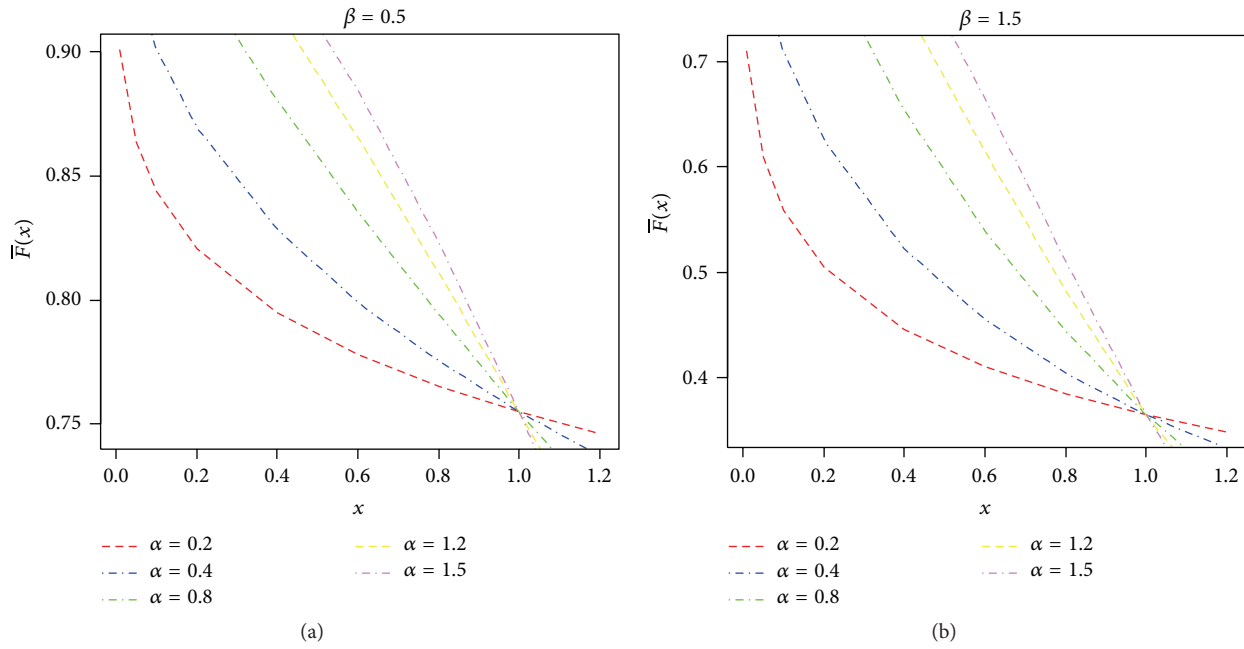
FIGURE 3: Survival function of PHLD ($\beta = 0.5$ and 1.5).

(a)   (b)
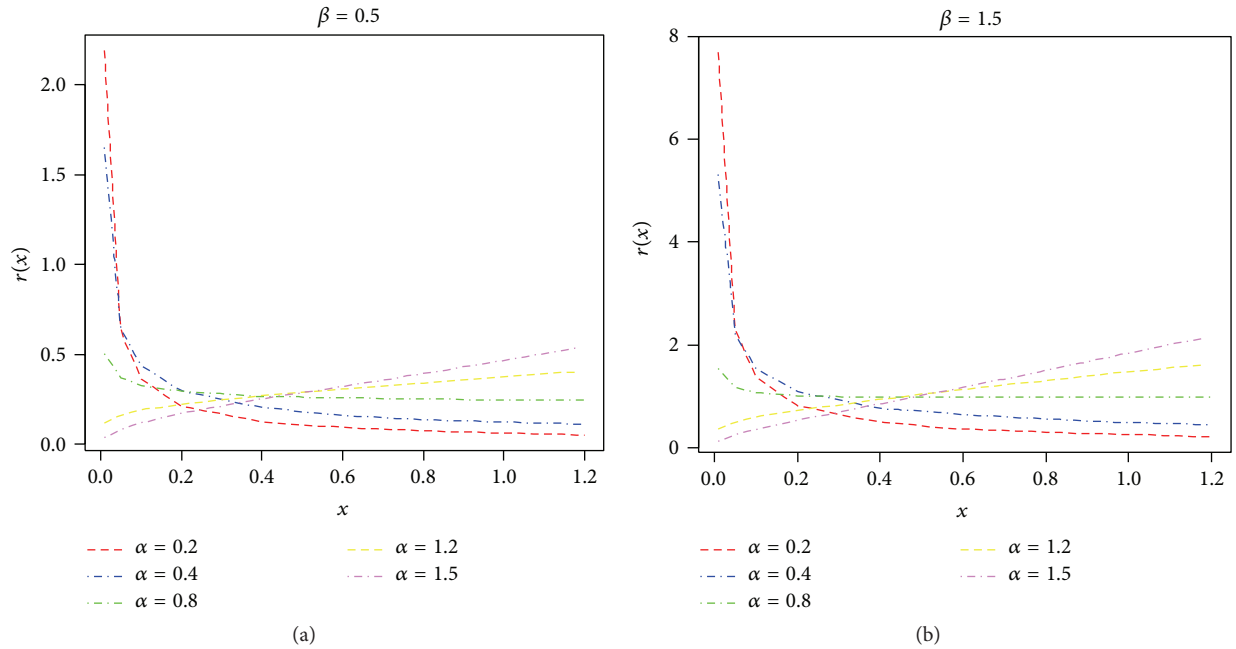
FIGURE 4: Hazard rate of PHLD ($\beta = 0.5$ and 1.5).

Converse part easily follows by assuming $\overline{G}(x) = 2/(1 + e^{cx^\alpha})$. □

One may derive further characterizations of PHLD by taking the $(1/\alpha)$th power of half-logistic variables in the results of Olapade [4], which described some characterizations of half-logistic distribution.

4. Estimation of the Parameters

We use the following three common methods for estimation purpose. Numerical illustrations are also done subsequently.

4.1. Maximum Likelihood Estimation. Suppose a sample of size n is taken from PHLD with density function (5). By taking logarithms and finding the derivative with respect to α

and β we have two nonlinear equations which can be solved simultaneously and numerically.

We have

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \beta x_i^\alpha \log x_i + \sum_{i=1}^{n} \log x_i$$

$$- 2\sum_{i=1}^{n} \frac{1}{1 + e^{\beta x_i^\alpha}} \log x_i x_i^\alpha \beta e^{\beta x_i^\alpha}, \qquad (13)$$

$$\frac{\partial \log L}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^{n} x_i^\alpha - \frac{2}{1 + e^{\beta x_i^\alpha} x_i^\alpha}. \qquad$$

4.2. Method of Moments. Method of moment estimation is another common method used for estimation of parameters. Equating the first and second raw moments to corresponding central moments, the following are the equations obtained:

$$\frac{\sum_{i=1}^{n} x_i}{n} = \frac{2}{\alpha\beta} \Gamma\left(\frac{1}{\alpha}\right) \sum_{j=0}^{\infty} \frac{(-1)^j}{(j+1)^{1/\alpha}},$$

$$\frac{\sum_{i=1}^{n} x_i^2}{n} = \frac{4}{\alpha\beta^2} \Gamma\left(\frac{2}{\alpha}\right) \sum_{j=0}^{\infty} \frac{(-1)^j}{(j+1)^{2/\alpha}}. \qquad (14)$$

4.3. Least Square Method. Least square estimation method involves the least squares regression to estimate the two parameters based on the linearized PHLD distribution function. For details of this procedure see Krishnaiah [10]. The basis of this method is the transformation of PHLD survival function

$$\overline{F}(x) = \frac{2}{1 + e^{\beta x^\alpha}}, \qquad (15)$$

in the form

$$\log x = \frac{-\log \beta}{\alpha} + \frac{1}{\alpha} \log\left[\log\left(\frac{2}{1 - F(x)} - 1\right)\right]. \qquad (16)$$

On putting $Y = \log x$ and $X = \log[\log((2/(1 - F(x))) - 1)]$, it becomes a linear function of X and Y in the form

$$Y = \frac{1}{\alpha}(-\log \beta) + \frac{1}{\alpha} X. \qquad (17)$$

Note that $1/\alpha$ is the slope of this equation and $(1/\alpha) \cdot (-\log \beta)$ is the intercept.

Let $x_1 < x_2 < \cdots < x_n$ be the times of failure arranged in ascending order and n is the sample size. Then $F(x_i)$ is estimated as in Zaka and Akhter [11], using Bernards' median rank method given by

$$\widehat{F}(x_i) = \frac{i - 0.3}{n + 0.4}. \qquad (18)$$

Now the least square estimates of α and β are

$$\widehat{\alpha} = \frac{n \sum X_i^2 - (\sum X_i)^2}{n \sum (X_i Y_i) - \sum X_i \sum Y_i}, \qquad (19)$$

$$\widehat{\beta} = e^{(-1/n)(\widehat{\alpha} \cdot \sum Y_i - \sum X_i)},$$

where X_i and Y_i are the values corresponding to the ordered failure times x_i.

TABLE 1: Parameter estimates.

Sample size	Parameters α, β	Method	Estimates $\widehat{\alpha}, \widehat{\beta}$	K-S distance	p value
$n = 100$	0.5, 0.8	MLE	0.5052, 0.8049	0.03	1.0
		LSE	0.4973, 0.8069	0.04	1.0
$n = 50$	0.5, 0.8	MLE	0.5140, 0.8039	0.06	1.0
		LSE	0.5008, 0.8122	0.06	1.0
$n = 20$	0.5, 0.8	MLE	0.5441, 0.7908	0.15	0.9831
		LSE	0.5107, 0.8200	0.15	0.9831
$n = 100$	1.2, 1.0	MLE	1.1948, 1.0082	0.03	1.0
		LSE	1.2218, 1.0012	0.03	1.0
$n = 50$	1.2, 1.0	MLE	1.2363, 0.9993	0.04	1.0
		LSE	1.1968, 1.0147	0.04	1.0
$n = 20$	1.2, 1.0	MLE	1.2968, 1.0093	0.1	1.0
		LSE	1.2270, 1.0323	0.1	1.0

4.4. Numerical Examples. Samples of sizes 100, 50, and 20 are generated from PHLD for different values of parameters. The standard method of generation in R-programming is used for the generation of samples. We repeat this process 1000 times and compute simulated average, standard errors, confidence intervals, and coverage probabilities in each case. Comparison of the maximum likelihood estimation (MLE) and least square estimation (LSE) methods mentioned above is done. The computations are performed using R-programme and results are shown in Table 1. Kolmogorov-Smirnov (K-S) statistic and corresponding p values are used for comparing the estimation methods. The 95% confidence intervals for the parameters using maximum likelihood estimates are also constructed in Table 2. Value of the K-S statistic is the same for both methods in most of the cases. Also p value is the same in both cases. As there is no clear supremacy of a method over the other, we suggest MLE method since it is more prevalent and the estimates have better appealing properties. The distribution functions are considered for the generated sequence for given parameter values and also using estimated parameters and those functions are plotted in Figure 5. Similarly histograms and superimposed density curves for estimated values of the parameters are shown in Figure 6. From this we conclude that the above two methods of estimation are in agreement. Coverage probabilities for the parameters for different sample sizes are given in Table 3 and it is clear that they are higher in the case of LSE method.

TABLE 2: Confidence interval for the parameters.

Sample size	Parameters α, β	Estimates $\widehat{\alpha}, \widehat{\beta}$	Confidence intervals	MSE
n = 100	0.5	0.5052	(0.4219, 0.5885)	0.0018
	0.8	0.8049	(0.60095, 1.0089)	0.0108
n = 50	0.5	0.5140	(0.3879, 0.6400)	0.0041
	0.8	0.8039	(0.5302, 1.0776)	0.0194
n = 20	0.5	0.5441	(0.3176, 0.7707)	0.0133
	0.8	0.7908	(0.3631, 1.2186)	0.0476
n = 100	1.2	1.1948	(1.0142, 1.4293)	0.0112
	1.0	1.0082	(0.7976, 1.2049)	0.0107
n = 50	1.2	1.2363	(0.9437, 1.5288)	0.0228
	1.0	0.9993	(0.7065, 1.2920)	0.0223
n = 20	1.2	1.2968	(0.7858, 1.8088)	0.0679
	1.0	1.0093	(0.5248, 1.4938)	0.0611

TABLE 3: Coverage probabilities for the parameters based on estimation methods.

Method	Sample size	Parameters $\alpha = 0.5, \beta = 0.8$	Parameters $\alpha = 1.2, \beta = 1$
MLE	n = 100	0.875, 0.864	0.834, 0.832
LSE		0.942, 0.955	0.944, 0.963
MLE	n = 50	0.844, 0.812	0.734, 0.724
LSE		0.946, 0.95	0.939, 0.951
MLE	n = 20	0.836, 0.529	0.575, 0.593
LSE		0.94, 0.956	0.942, 0.955

5. Extensions of PHLD

Several types of extensions are possible using this distribution. Some of them are very much related with already existing distributions in the literature. These are discussed in this section.

5.1. Log Power Half-Logistic Distributions.
Consider the distribution of some log transformations and we call the distributions obtained as log Power Half-Logistic Distributions. When $Y = e^X$ and X has PHLD, then

$$f(y) = \frac{2\alpha\beta e^{\beta(\log y)^\alpha} (\log y)^{\alpha-1}}{y (1 + e^{\beta(\log y)^\alpha})^2}, \quad y > 1. \tag{20}$$

This distribution is called as log positive Power Half-Logistic Distribution.

In a similar way we define the distribution of $Y = e^{-x}$ as log negative Power Half-Logistic Distribution with support $0 < y < 1$.

If $Y = e^{x^\alpha}$, where $X \sim$ PHLD, the pdf of Y is given by

$$g(y) = \frac{2}{(1 + y)^2}, \quad y > 1. \tag{21}$$

Similarly the pdf of $Z = e^{-x^\alpha}$ is

$$g(z) = \frac{2}{(1 + z)^2}, \quad 0 < z < 1. \tag{22}$$

So we get two distributions with the same structure but defined at two disjoint intervals $[1, \infty)$ and $[0, 1]$ which is the characteristic of a nonnegative random variable with respect to log transformations. These are having the same form as we transform half-logistic distribution by the transformations $Y = e^x$ and $Z = e^{-x}$.

Immediately, we have the following result, which may be exploited for generating random variables from PHLD.

Result 4. If $X \sim U(0, 1)$, then the random variable $Z = [(1/\beta) \log(a - X)/X]^{1/\alpha}$ has truncated PHLD (TPHLD).

Proof. If $X \sim U(0, 1)$, then

$$P(Z \geq z) = P\left(\left[\left(\frac{1}{\beta}\right) \log\left(\frac{a - X}{X}\right) \right]^{1/\alpha} \geq z \right)$$
$$= \frac{a}{1 + e^{\beta z^\alpha}}, \tag{23}$$

which is a new distribution with density function

$$f(x) = \frac{\alpha a \beta e^{\beta x^\alpha} x^{\alpha-1}}{(1 + e^{\beta x^\alpha})^2};$$
$$x > \left[\frac{1}{\beta} \log(a - 1) \right]^{1/\alpha}, \quad \alpha > 0, \ \beta > 0, \ a \geq 2. \tag{24}$$

□

We call this as Truncated Power Half-Logistic Distribution (TPHLD). Note that when $\alpha = 1$, we have the Truncated Half-Logistic Distribution; when $a = 2$, $\alpha = 1$, it gives HLD; and when $a = 2$, $\alpha > 0$, it gives PHLD.

5.2. Families of Distributions Generated from PHLD.
We have, in the literature, quite a few families of distributions generated from Beta and Gamma distributions (see [7, 12, 13]). This type of distributions is generalizations of many existing families. Here we generate families of distributions from the PHLD. A detailed study of this type of distributions, its properties, applications, and so forth is not attempted in this paper for brevity, but would be carried out in future.
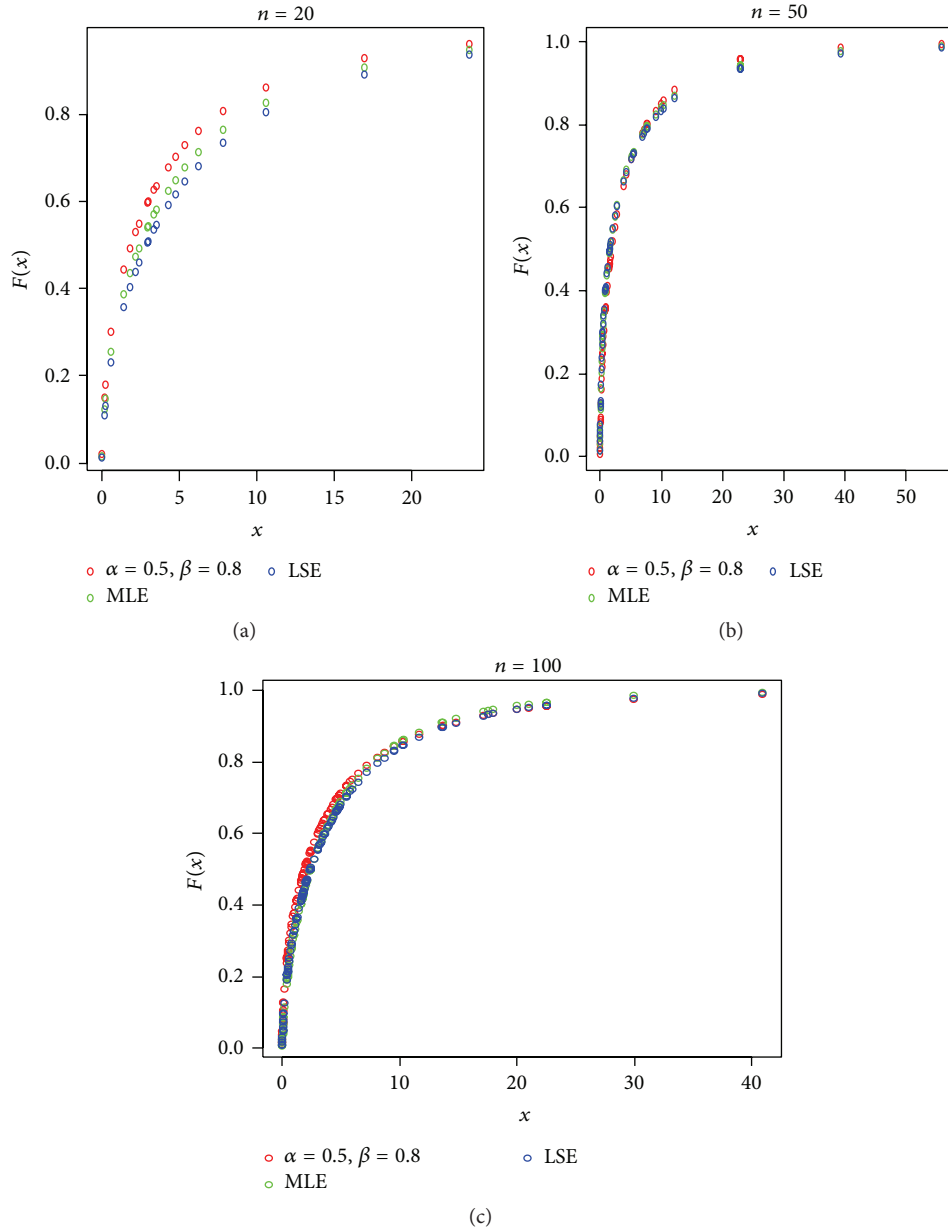
Figure 5: Distribution functions when $n = 20, 50, 100$.

Define a new transformation

$$G_1(x) = \int_0^{\phi(x)} d(H(x)), \qquad (25)$$

where $H(x)$ is the distribution function of PHLD or its generalizations and $\phi(x)$ takes different forms of $\overline{G}(x)$, the survival function of a random variable. This transformation gives us very interesting results as summarized in Table 4. Note that when $\beta = 1$, $\overline{G}(x)$ in Result number 1 in Table 4 has Marshall and Olkin [14] form with parameter 2. So a general structure is needed for constructing Marshall-Olkin form with parameter α. This is explained in the following remark.

Remark 1. Consider a new distribution, called General Power Half-Logistic Distribution (GPHLD), by adding a skewness parameter γ,

$$h(x) = (1 + \gamma) \frac{\gamma\beta\alpha e^{\beta x^\alpha} x^{\alpha-1}}{(1 + \gamma e^{\beta x^\alpha})^2}; \quad x > 0, \ \alpha > 0. \qquad (26)$$

The survival function is

$$\overline{H}(x) = \frac{1 + \gamma}{1 + \gamma e^{\beta x^\alpha}}. \qquad (27)$$
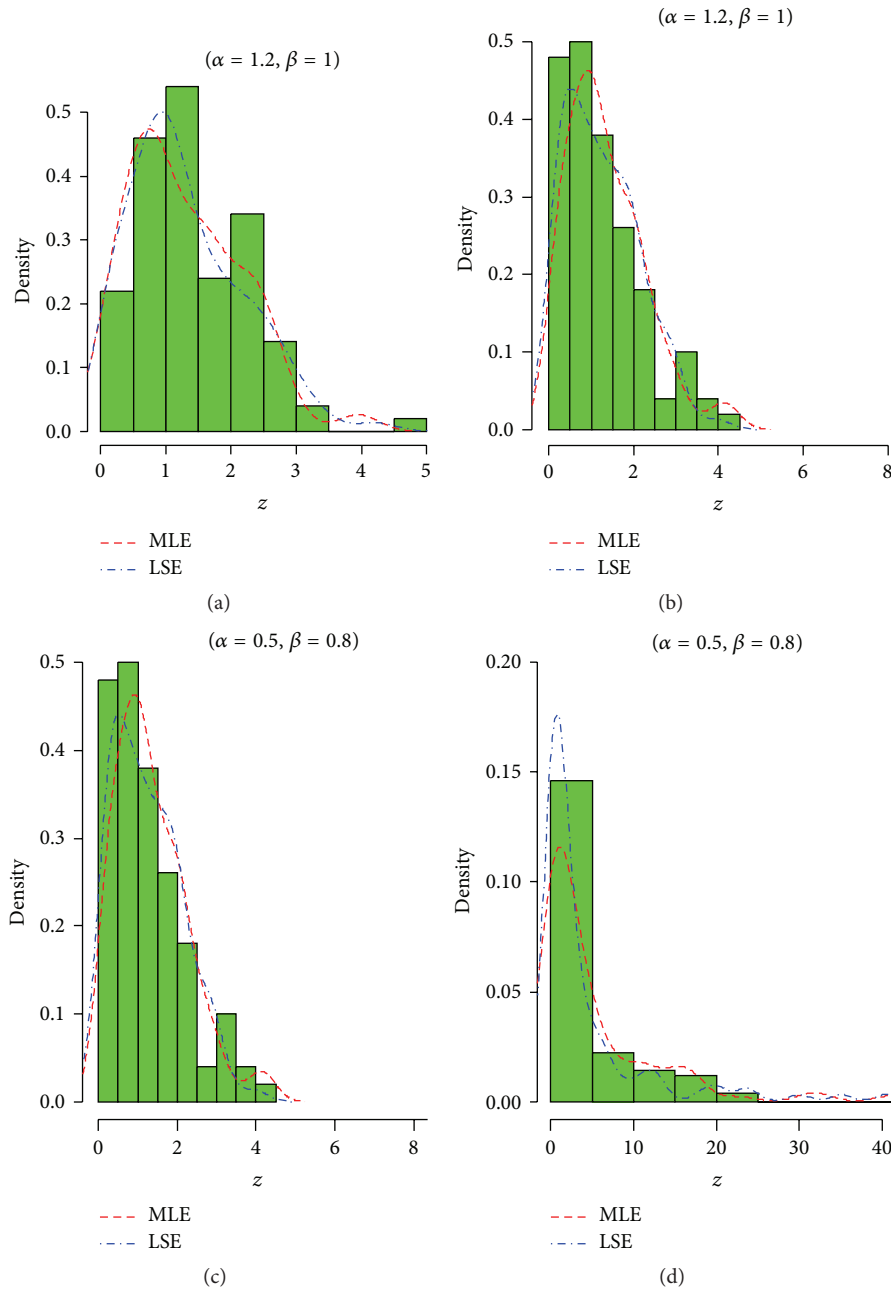
FIGURE 6: Fitted density curves based on estimation.

Using this in (25) we get a new family of life distributions, with survival function

$$\overline{H}(x) = \frac{\overline{G}(x)^{\beta}(1+\gamma)}{\gamma + \overline{G}(x)^{\beta}} \qquad (28)$$

which can be considered as a generalization of Marshall and Olkin [14] (M-O) form with survival function $\alpha\overline{F}(x)/(1-\overline{\alpha}\overline{F}(x))$, $\alpha > 0$. Let $\beta = 1$ in (28); then corresponding density function is given by $h(x) = (1+\gamma)\gamma g(x)/(\gamma + \overline{G}(x))^2$ and

hazard rate is $(\gamma/(\gamma + \overline{G}(x)))r(x)$ where $r(x)$ is the hazard rate function of $g(x)$. We can see that the parameters γ in (28) and α in M-O are related as $\gamma = 1/(\alpha - 1)$.

Interestingly we have noted that (27) is the Weibull-geometric distribution introduced by Barreto-Souza et al. [15] with parameter θ in $[-1, \infty)$.

Remark 2. Result number 3 of Table 4 is obtained by taking the γ th power of distribution function of PHLD (called as Type I PHLD) which is the same as in Cordeiro et al. [7].

TABLE 4: Families of distributions generated.

Result number	$H(x), x > 0,$ $\alpha > 0, \beta > 0$	$\phi(x)$	$\overline{G}_1(x)$
1	$1 - \dfrac{2}{1 + e^{\beta x^{\alpha}}}$	$\left(-\log \overline{G}(x)\right)^{1/\alpha}$	$\dfrac{2\overline{G}^{\beta}}{1 + \overline{G}^{\beta}}$
2	$1 - \dfrac{2}{1 + e^{\beta x^{\alpha}}}$	$\dfrac{G}{\overline{G}}$	$\dfrac{2}{1 + e^{\beta(G/\overline{G})^{\alpha}}}$
3	$\left[1 - \dfrac{2}{1 + e^{\beta x^{\alpha}}}\right]^{\gamma}$	$\left(-\log \overline{G}(x)\right)^{1/\alpha}$	$1 - \left[\dfrac{1 - \overline{G}(x)^{\beta}}{1 + \overline{G}(x)^{\beta}}\right]^{\gamma}$
4	$1 - \left(\dfrac{2}{1 + e^{\beta x^{\alpha}}}\right)^{\gamma}$	$\left(-\log \overline{G}(x)\right)^{1/\alpha}$	$\left[\dfrac{2\overline{G}(x)^{\beta}}{1 + \overline{G}(x)^{\beta}}\right]^{\gamma}$

TABLE 5: Fitting based on data set 1.

Distribution	Parameter estimates	Log likelihood	K-S distance	p value
Weibull	$\widehat{\alpha} = 2.08, \widehat{\beta} = 0.0001$	−113.6902	0.1304	0.9897
PHLD	$\widehat{\alpha} = 1.77, \widehat{\beta} = 0.0006$	−114.0720	0.2609	0.4143
Type I HLD	$\widehat{\alpha} = 3.41, \widehat{\beta} = 0.0350$	−113.0407	0.2174	0.6487
GPHLD	$\widehat{\alpha} = 2.01, \widehat{\beta} = 0.0002, \widehat{\gamma} = 4.96$	−113.7000	0.1304	0.9897

TABLE 6: Fitting based on data set 2.

Distribution	Parameter estimates	Log likelihood	K-S distance	p value
Weibull	$\widehat{\alpha} = 5.78, \widehat{\beta} = 0.06$	−15.2070	0.2063	0.1367
PHLD	$\widehat{\alpha} = 5.05, \widehat{\beta} = 0.13$	−13.8273	0.1587	0.4055
Type I HLD	$\widehat{\alpha} = 5.04, \widehat{\beta} = 0.13$	−13.8300	0.1587	0.4055
GPHLD	$\widehat{\alpha} = 3.202, \widehat{\beta} = 0.695, \widehat{\gamma} = 0.064$	−12.0336	0.0952	0.9375

Remark 3. $\overline{G}_1(x)$ in Result number 4 of Table 4 is a member of the Lehmann family of distributions and this is Type II PHLD.

6. Applications

In this section we use three sets of real-life data to fit the distributions. The analysis is done using R-programming software. The first set is discussed by Gupta and Kundu [16] in the fitting of exponentiated exponential distribution.

Data Set 1. The first data set is taken from Lawless [17, page 98]. The data are the number of million revolutions before failure for each of the 23 ball bearings in the life test and they are 17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.80, 51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04, and 173.40. We consider the Weibull, Type I HLD (Kantam et al. [8]), PHLD, and GPHLD for this particular data set.

The likelihood value as noted in Table 5 is greatest for Type 1 HLD and also based on p values of K-S statistic, we conclude that Type 1 HLD is a good fit for the data.

Data Set 2. This data set is from Smith and Naylor [18] representing strengths of 1.5 cm glass fibres. The data set is 0.55, 0.93, 1.25, 1.36, 1.49, 1.52, 1.58, 1.61, 1.64, 1.68, 1.73, 1.81, 2,

0.74, 1.04, 1.27, 1.39, 1.49, 1.53, 1.59, 1.61, 1.66, 1.68, 1.76, 1.82, 2.01, 0.77, 1.11, 1.28, 1.42, 1.5, 1.54, 1.6, 1.62, 1.66, 1.69, 1.76, 1.84, 2.24, 0.81, 1.13, 1.29, 1.48, 1.5, 1.55, 1.61, 1.62, 1.66, 1.7, 1.77, 1.84, 0.84, 1.24, 1.3, 1.48, 1.51, 1.55, 1.61, 1.63, 1.67, 1.7, 1.78, and 1.89 (see [19]). They have fitted different distributions to this data set. When we use this data set for the four distributions, Weibull, Type I HLD, PHLD, and GPHLD, the results are as follows.

Morais and Barreto-Souza [19] have shown that the Weibull-geometric distribution is better fit to this data set. The log-likelihood values, K-S distance, and p values in Table 6 reveal that GPHLD is better than the other three models (see Remark 1 in this context).

Data Set 3. This data set is of camber of 497 lead wires taken from Leone et al. [20]. Cooray et al. [21] considered this data and fitted folded logistic distribution. They got log-likelihood = −1698.24, K-S distance = 0.06, and p value = 0.32. Results are illustrated in Table 7 and it is clear that PHLD is the most suitable for this data set.

7. Conclusions

A new distribution on the positive real line is constructed using power transformation on half-logistic distribution.

TABLE 7: Fitting based on data set 3.

Distribution	Parameter estimates	Log likelihood	K-S distance	p value
Weibull	$\widehat{\alpha} = 1.860, \widehat{\beta} = 0.0059$	−1692.670	0.0795	0.10
PHLD	$\widehat{\alpha} = 1.595, \widehat{\beta} = 0.0180$	−1692.180	0.0785	0.10
Type I HLD	$\widehat{\alpha} = 2.095, \widehat{\beta} = 0.1450$	−1698.953	0.0845	0.06
GPHLD	$\widehat{\alpha} = 1.824, \widehat{\beta} = 0.0069, \widehat{\gamma} = 9.934$	−1692.484	0.0986	0.02

Analytical properties, some characterizations, and estimation of the parameters are done. New families of distributions are generated from this new distribution which generalizes many existing families of distributions. Applications are discussed with the help of three data sets. The properties, characteristics, and applications of the newly generated families of distributions are further topics for future research work. Some of these families are generated using the odds function.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] N. Balakrishnan, "Order statistics from the half logistic distribution," *Journal of Statistical Computation and Simulation*, vol. 20, no. 4, pp. 287–309, 1985.

[2] N. Balakrishnan, *Handbook of the Logistic Distribution*, vol. 123 of *Statistics: A Series of Textbooks and Monographs*, Marcel Dekker, New York, NY, USA, 1992.

[3] B. Srinivasa Rao, S. Nagendram, and K. Rosaiah, "Exponential Halflogistic additive failure rate model," *International Journal of Scientific and Research Publications*, vol. 3, no. 5, pp. 1–10, 2013.

[4] A. K. Olapade, "On Characterizations of the Half Logistic Distribution," Interstat, February Issue, Number 2, 2003, http://interstat.statjournals.net/.

[5] A. K. Olapade, "On type III generalized half logistic distribution," *Journal of the Iranian Statistical Society*, vol. 13, no. 1, 2008.

[6] A. K. Olapade, "The type I generalized half logistic distribution," *Journal of Iranian Statistical Society*, vol. 13, no. 1, pp. 69–88, 2014.

[7] G. M. Cordeiro, M. Alizadeh, and E. M. M. Ortega, "The exponentiated half-logistic family of distributions: properties and applications," *Journal of Probability and Statistics*, vol. 2014, Article ID 864396, 21 pages, 2014.

[8] R. R. L. Kantam, V. Ramakrishna, and M. S. Ravikumar, "Estimation and testing in Type I generalized half logistic distribution," *Journal of Modern Applied Statistical Methods*, vol. 12, no. 1, pp. 198–206, 2013.

[9] A. W. Marshall and I. Olkin, *Life Distributions: Structure of Nonparametric, Semiparametric, and Parametric Families*, Springer Series in Statistics, Springer, New York, NY, USA, 2007.

[10] K. Krishnaiah, *Applied Statistical Quality Control and Improvement*, Eastern Economy Edition, PHL Learning Private Limited, 2014.

[11] A. Zaka and A. S. Akhter, "Methods for estimating the parameters of the power function distribution," *Pakistan Journal of Statistics and Operation Research*, vol. 9, no. 2, pp. 213–224, 2013.

[12] N. Eugene, C. Lee, and F. Famoye, "Beta-normal distribution and its applications," *Communications in Statistics—Theory and Methods*, vol. 31, no. 4, pp. 497–512, 2002.

[13] K. Zografos and N. Balakrishnan, "On families of beta- and generalized gamma-generated distributions and associated inference," *Statistical Methodology*, vol. 6, no. 4, pp. 344–362, 2009.

[14] A. W. Marshall and I. Olkin, "A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families," *Biometrika*, vol. 84, no. 3, pp. 641–652, 1997.

[15] W. Barreto-Souza, A. L. De Morais, and G. M. Cordeiro, "The Weibull-geometric distribution," *Journal of Statistical Computation and Simulation*, vol. 81, no. 5, pp. 645–657, 2011.

[16] R. D. Gupta and D. Kundu, "Exponentiated exponential family: an alternative to gamma and Weibull distributions," *Biometrical Journal*, vol. 43, no. 1, pp. 117–130, 2001.

[17] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2003.

[18] R. L. Smith and J. C. Naylor, "A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution," *Journal of the Royal Statistical Society. Series C. Applied Statistics*, vol. 36, no. 3, pp. 358–369, 1987.

[19] A. L. Morais and W. Barreto-Souza, "A compound class of Weibull and power series distributions," *Computational Statistics and Data Analysis*, vol. 55, no. 3, pp. 1410–1425, 2011.

[20] F. C. Leone, L. S. Nelson, and R. B. Nottingham, "The folded normal distribution," *Technometrics*, vol. 3, no. 4, pp. 543–550, 1961.

[21] K. Cooray, S. Gunasekera, and M. M. A. Ananda, "The folded logistic distribution," *Communications in Statistics—Theory and Methods*, vol. 35, no. 1–3, pp. 385–393, 2006.

Properties of Matrix Variate Confluent Hypergeometric Function Distribution

Arjun K. Gupta,[1] Daya K. Nagar,[2] and Luz Estela Sánchez[2]

[1]*Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403-0221, USA*
[2]*Instituto de Matemáticas, Universidad de Antioquia, Calle 67, No. 53–108, Medellín, Colombia*

Correspondence should be addressed to Daya K. Nagar; dayaknagar@yahoo.com

We study matrix variate confluent hypergeometric function kind 1 distribution which is a generalization of the matrix variate gamma distribution. We give several properties of this distribution. We also derive density functions of $X_2^{-1/2} X_1 X_2^{-1/2}$, $(X_1 + X_2)^{-1/2} X_1 (X_1 + X_2)^{-1/2}$, and $X_1 + X_2$, where $m \times m$ independent random matrices X_1 and X_2 follow confluent hypergeometric function kind 1 and gamma distributions, respectively.

1. Introduction

The matrix variate gamma distribution has many applications in multivariate statistical analysis. The Wishart distribution, which is the distribution of the sample variance covariance matrix when sampling from a multivariate normal distribution, is a special case of the matrix variate gamma distribution.

The purpose of this paper is to give a generalization of the matrix variate gamma distribution and study its properties.

We begin with a brief review of some definitions and notations. We adhere to standard notations (cf. Gupta and Nagar [1]). Let $A = (a_{ij})$ be an $m \times m$ matrix. Then, A' denotes the transpose of A; $\text{tr}(A) = a_{11} + \cdots + a_{mm}$; $\text{etr}(A) = \exp(\text{tr}(A))$; $\det(A) = $ determinant of A; norm of $A = \|A\| = $ maximum of absolute values of eigenvalues of the matrix A; $A > 0$ means that A is symmetric positive definite; and $A^{1/2}$ denotes the unique symmetric positive definite square root of $A > 0$. The multivariate gamma function $\Gamma_m(a)$ is defined by

$$\Gamma_m(a) = \int_{X>0} \text{etr}(-X) \det(X)^{a-(m+1)/2} \, dX$$

$$= \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma\left(a - \frac{i-1}{2}\right), \quad \text{Re}(a) > \frac{m-1}{2}. \tag{1}$$

The $m \times m$ symmetric positive definite random matrix X is said to have a matrix variate gamma distribution, denoted by $X \sim \text{Ga}(m, \nu, \theta, \Omega)$, if its probability density function (p.d.f.) is given by

$$\frac{\det(X)^{\nu-(m+1)/2} \text{etr}\left(-\Omega^{-1} X/\theta\right)}{\Gamma_m(\nu) \theta^{m\nu} \det(\Omega)^\nu}, \quad X > 0, \tag{2}$$

where Ω is a symmetric positive definite matrix of order m, $\theta > 0$, and $\nu > (m-1)/2$. For $\Omega = I_m$, the above density reduces to a standard matrix variate gamma density and in this case we write $X \sim \text{Ga}(m, \nu, \theta)$. Further, if $X_1 \sim \text{Ga}(m, \nu_1, \theta)$ and $X_2 \sim \text{Ga}(m, \nu_2, \theta)$ are independent gamma matrices, then the random matrix $(X_1 + X_2)^{-1/2} X_1 (X_1 + X_2)^{-1/2}$ follows a matrix variate beta type 1 distribution with parameters ν_1 and ν_2.

By replacing $\text{etr}(-\Omega^{-1} X/\theta)$ by the confluent hypergeometric function of matrix argument $_1F_1(\alpha; \beta; -\Omega^{-1} X/\theta)$, a generalization of the matrix variate gamma distribution can be defined by the p.d.f.:

$$C(\nu, \alpha, \beta, \theta, \Omega) \det(X)^{\nu-(m+1)/2} {}_1F_1\left(\alpha; \beta; -\frac{1}{\theta} \Omega^{-1} X\right), \tag{3}$$

where $X > 0$ and $C(\nu, \alpha, \beta, \theta, \Omega)$ is the normalizing constant. In Section 2, it has been shown that, for $\beta - \nu > (m-1)/2$,

$\alpha - \nu > (m-1)/2$, $\nu > (m-1)/2$, $\theta > 0$, and $\Omega > 0$, the normalizing constant can be evaluated as

$$C\left(\nu, \alpha, \beta, \theta, \Omega\right) = \frac{\Gamma_m\left(\alpha\right)\Gamma_m\left(\beta - \nu\right)\det\left(\theta\Omega\right)^{-\nu}}{\Gamma_m\left(\nu\right)\Gamma_m\left(\beta\right)\Gamma_m\left(\alpha - \nu\right)}. \quad (4)$$

Therefore, the p.d.f. in (3) can be written explicitly as

$$\frac{\Gamma_m\left(\alpha\right)\Gamma_m\left(\beta - \nu\right)\det\left(\theta\Omega\right)^{-\nu}}{\Gamma_m\left(\nu\right)\Gamma_m\left(\beta\right)\Gamma_m\left(\alpha - \nu\right)}\det\left(X\right)^{\nu-(m+1)/2}$$
$$\cdot {}_1F_1\left(\alpha; \beta; -\frac{1}{\theta}\Omega^{-1}X\right), \quad X > 0, \quad (5)$$

where $\beta - \nu > (m-1)/2$, $\alpha - \nu > (m-1)/2$, $\nu > (m-1)/2$, $\theta > 0$, $\Omega > 0$, and ${}_1F_1$ is the confluent hypergeometric function of the first kind of matrix argument (Gupta and Nagar [1]). Since the density given above involves the confluent hypergeometric function, we will call the corresponding distribution a *confluent hypergeometric function distribution*. We will write $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \Omega, \text{kind } 1)$ to say that the random matrix X has a confluent hypergeometric function distribution defined by the density (5). It has been shown by van der Merwe and Roux [2] that the above density can be obtained as a limiting case of a density involving the Gauss hypergeometric function of matrix argument. For $\alpha = \beta$, the density (5) reduces to a matrix variate gamma density and for $\Omega = I_m$ it slides to

$$\frac{\Gamma_m\left(\alpha\right)\Gamma_m\left(\beta - \nu\right)}{\theta^{m\nu}\Gamma_m\left(\nu\right)\Gamma_m\left(\beta\right)\Gamma_m\left(\alpha - \nu\right)}\det\left(X\right)^{\nu-(m+1)/2}$$
$$\cdot {}_1F_1\left(\alpha; \beta; -\frac{1}{\theta}X\right), \quad X > 0, \quad (6)$$

where $\beta - \nu > (m-1)/2$, $\alpha - \nu > (m-1)/2$, $\nu > (m-1)/2$, and $\theta > 0$. In this case we will write $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \text{kind } 1)$. The matrix variate confluent hypergeometric function kind 1 distribution occurs as the distribution of the matrix ratio of independent gamma and beta matrices. For $m = 1$, (6) reduces to a univariate confluent hypergeometric function kind 1 density given by (Orozco-Castañeda et al. [3])

$$\frac{\Gamma\left(\alpha\right)\Gamma\left(\beta - \nu\right)}{\theta^{\nu}\Gamma\left(\nu\right)\Gamma\left(\beta\right)\Gamma\left(\alpha - \nu\right)}x^{\nu-1}{}_1F_1\left(\alpha; \beta; -\frac{x}{\theta}\right), \quad x > 0, \quad (7)$$

where $\beta - \nu > 0$, $\alpha - \nu > 0$, $\nu > 0$, $\theta > 0$, and ${}_1F_1$ is the confluent hypergeometric function of the first kind (Luke [4]). The random variable x having the above density will be designated by $x \sim \mathrm{CH}(\nu, \alpha, \beta, \theta, \text{kind } 1)$. Since the matrix variate confluent hypergeometric function kind 1 distribution is a generalization of the matrix variate gamma distribution, it is reasonable to say that the matrix variate confluent hypergeometric function kind 1 distribution can be used as an alternative to the gamma distribution quite effectively.

Although ample information about matrix variate gamma distribution is available, little appears to have been done in the literature to study the matrix variate confluent hypergeometric function kind 1 distribution.

In this paper, we study several properties including stochastic representations of the matrix variate confluent hypergeometric function kind 1 distribution. We also derive the density function of the matrix quotient of two independent random matrices having confluent hypergeometric function kind 1 and gamma distributions. Further, densities of several other matrix quotients and matrix products involving confluent hypergeometric function kind 1, beta type 1, beta type 2, and gamma matrices are derived.

2. Some Definitions and Preliminary Results

In this section we give some definitions and preliminary results which are used in subsequent sections.

A more general integral representation of the multivariate gamma function can be obtained as

$$\Gamma_m\left(a\right) = \det\left(Y\right)^a \int_{R>0} \mathrm{etr}\left(-YR\right)\det\left(R\right)^{a-(m+1)/2}\,dR, \quad (8)$$

where $\mathrm{Re}(a) > (m-1)/2$ and $\mathrm{Re}(Y) > (m-1)/2$. The above result can be established for real $Y > 0$ by substituting $X = Y^{1/2}RY^{1/2}$ with the Jacobian $J(X \to R) = \det(Y)^{(m+1)/2}$ in (1) and it follows for complex Y by analytic continuation.

The multivariate generalization of the beta function is given by

$$B_m\left(a, b\right)$$
$$= \int_0^{I_m} \det\left(R\right)^{a-(m+1)/2}\det\left(I_m - R\right)^{b-(m+1)/2}\,dR \quad (9)$$
$$= \frac{\Gamma_m\left(a\right)\Gamma_m\left(b\right)}{\Gamma_m\left(a+b\right)} = B_m\left(b, a\right),$$

where $\mathrm{Re}(a) > (m-1)/2$ and $\mathrm{Re}(b) > (m-1)/2$.

The generalized hypergeometric function of one matrix, defined in Constantine [5], is given by

$$\begin{aligned}
&{}_pF_q\left(a_1, \ldots, a_p; b_1, \ldots, b_q; X\right)\\
&= \sum_{k=0}^{\infty}\sum_{\kappa \vdash k}\frac{\left(a_1\right)_\kappa \cdots \left(a_p\right)_\kappa}{\left(b_1\right)_\kappa \cdots \left(b_q\right)_\kappa}\frac{C_\kappa\left(X\right)}{k!},
\end{aligned} \quad (10)$$

where a_i, $i = 1, \ldots, p$, b_j, $j = 1, \ldots, q$ are arbitrary complex numbers, X is an $m \times m$ complex symmetric matrix, $C_\kappa(X)$ is the zonal polynomial of $m \times m$ complex symmetric matrix X corresponding to the ordered partition $\kappa = (k_1, \ldots, k_m)$, $k_1 \geq \cdots \geq k_m \geq 0$, $k_1 + \cdots + k_m = k$, and $\sum_{\kappa \vdash k}$ denotes the summation over all partitions κ. The generalized hypergeometric coefficient $(a)_\kappa$ used above is defined by

$$\left(a\right)_\kappa = \prod_{i=1}^{m}\left(a - \frac{i-1}{2}\right)_{k_i}, \quad (11)$$

where $(a)_r = a(a+1)\cdots(a+r-1)$, $r = 1, 2, \ldots$, with $(a)_0 = 1$. Conditions for convergence of the series in (10) are available in the literature. From (10), it follows that

$$_0F_0(X) = \sum_{k=0}^{\infty} \sum_{\kappa \vdash k} \frac{C_\kappa(X)}{k!} = \text{etr}(X),$$

$$_1F_0(a; X) = \sum_{k=0}^{\infty} \sum_{\kappa \vdash k} \frac{(a)_\kappa C_\kappa(X)}{k!} \tag{12}$$

$$= \det(I_m - X)^{-a}, \quad \|X\| < 1,$$

$$_1F_1(a; c; X) = \sum_{k=0}^{\infty} \sum_{\kappa \vdash k} \frac{(a)_\kappa C_\kappa(X)}{(c)_\kappa k!}, \tag{13}$$

$$_2F_1(a, b; c; X) = \sum_{k=0}^{\infty} \sum_{\kappa \vdash k} \frac{(a)_\kappa (b)_\kappa}{(c)_\kappa} \frac{C_\kappa(X)}{k!}, \quad \|X\| < 1. \tag{14}$$

By taking $a_p = b_q = c$ in (10), it can be observed that

$$_pF_q\left(a_1, \ldots, a_{p-1}, c; b_1, \ldots, b_{q-1}, c; X\right)$$
$$= {}_{p-1}F_{q-1}\left(a_1, \ldots, a_{p-1}; b_1, \ldots, b_{q-1}; X\right). \tag{15}$$

Substituting $p = 2$, $q = 1$ in (15) and using (12), the Gauss hypergeometric function $_2F_1(a, c; c; X)$ is reduced as

$$_2F_1(a, c; c; X) = {}_1F_0(a; X) = \det(I_m - X)^{-a},$$
$$\|X\| < 1. \tag{16}$$

The integral representations of the confluent hypergeometric function $_1F_1$ and the Gauss hypergeometric function $_2F_1$ are given by

$$_1F_1(a; c; X) = \frac{1}{B_m(a, c-a)} \int_0^{I_m} \det(R)^{a-(m+1)/2}$$
$$\cdot \det(I_m - R)^{c-a-(m+1)/2} \text{etr}(XR) \, dR, \tag{17}$$

$$_2F_1(a, b; c; X) = \frac{1}{B_m(a, c-a)}$$
$$\cdot \int_0^{I_m} \frac{\det(R)^{a-(m+1)/2} \det(I_m - R)^{c-a-(m+1)/2}}{\det(I_m - XR)^b} \, dR, \tag{18}$$

where $\text{Re}(a) > (m-1)/2$ and $\text{Re}(c-a) > (m-1)/2$.

Further generalizations of (8) and (9) in terms of zonal polynomials, due to Constantine [5], are given as

$$\int_{R>0} \text{etr}(-YR) \det(R)^{a-(m+1)/2} C_\kappa(XR) \, dR$$
$$= \Gamma_m(a)(a)_\kappa \det(Y)^{-a} C_\kappa\left(Y^{-1}X\right), \tag{19}$$

$$\int_0^{I_m} \det(R)^{a-(m+1)/2} \det(I_m - R)^{b-(m+1)/2} C_\kappa(XR) \, dR$$
$$= \frac{B_m(a, b)(a)_\kappa}{(a+b)_\kappa} C_\kappa(X), \tag{20}$$

respectively.

For $\text{Re}(\alpha) > (m-1)/2$ and $\text{Re}(\beta) > (m-1)/2$, we have

$$\int_{R>0} \det(R)^{\alpha-(m+1)/2} \text{etr}(-YR) \, _pF_q\left(a_1, \ldots, a_p; b_1, \ldots, b_q;\right.$$
$$\left. XR\right) dR = \Gamma_m(\alpha) \det(Y)^{-a} \, _{p+1}F_q\left(\alpha, a_1, \ldots, a_p;\right. \tag{21}$$
$$\left. b_1, \ldots, b_q; XY^{-1}\right),$$

$$\int_0^{I_m} \det(R)^{\alpha-(m+1)/2} \det(I_m$$
$$- R)^{\beta-(m+1)/2} \, _pF_q\left(a_1, \ldots, a_p; b_1, \ldots, b_q; XR\right) dR \tag{22}$$
$$= B_m(\alpha, \beta) \, _{p+1}F_{q+1}\left(\alpha, a_1, \ldots, a_p; \alpha+\beta, b_1, \ldots, b_q; X\right).$$

We can establish (21) and (22) by expanding $_pF_q$ in series form by using (10) and integrating term by term by applying (19) and (20) and finally summing the resulting series.

Note that the series expansions for $_1F_1$ and $_2F_1$ given in (13) and (14) can be obtained by expanding $\text{etr}(XR)$ and $\det(I_m - XR)^{-b}$, $\|XR\| < 1$, in (17) and (18) and integrating R using (20). Substituting $X = I_m$ in (18) and integrating, we obtain

$$_2F_1(a, b; c; I_m) = \frac{\Gamma_m(c) \Gamma_m(c-a-b)}{\Gamma_m(c-a) \Gamma_m(c-b)}, \tag{23}$$

where $\text{Re}(c-a-b) > (m-1)/2$, $\text{Re}(c-a) > (m-1)/2$, $\text{Re}(c-b) > (m-1)/2$, and $\text{Re}(c) > (m-1)/2$. The hypergeometric function $_1F_1(a; c; X)$ satisfies Kummer's relation

$$_1F_1(a; c; -X) = \text{etr}(-X) \, _1F_1(c-a; c; X). \tag{24}$$

For properties and further results on these functions the reader is referred to Constantine [5], James [6], Muirhead [7], and Gupta and Nagar [1]. The numerical computation of a hypergeometric function of matrix arguments is very difficult. However, some numerical methods are proposed in recent years; see, Hashiguchi et al. [8] and Koev and Edelman [9].

The generalized hypergeometric function with $m \times m$ complex symmetric matrices X and Y is defined by

$$_pF_q^{(m)}\left(a_1, \ldots, a_p; b_1, \ldots, b_q; X, Y\right)$$
$$= \sum_{k=0}^{\infty} \sum_{\kappa \vdash k} \frac{(a_1)_\kappa \cdots (a_p)_\kappa}{(b_1)_\kappa \cdots (b_q)_\kappa} \frac{C_\kappa(X) C_\kappa(Y)}{C_\kappa(I_m) k!}. \tag{25}$$

It is clear from the above definition that the order of X and Y is unimportant; that is,

$$_pF_q^{(m)}\left(a_1, \ldots, a_p; b_1, \ldots, b_q; X, Y\right)$$
$$= {}_pF_q^{(m)}\left(a_1, \ldots, a_p; b_1, \ldots, b_q; Y, X\right). \tag{26}$$

Also, if one of the argument matrices is the identity, this function reduces to the one argument function. Further, the two-matrix argument function $_pF_q^{(m)}$ can be obtained from the one-matrix function $_pF_q$ by averaging over the orthogonal

group $O(m)$ using a result given in James [6, Equation 23]; namely,

$$\int_{O(m)} C_\kappa \left(XHYH' \right) (\mathrm{d}H) = \frac{C_\kappa(X) C_\kappa(Y)}{C_\kappa(I_m)}, \qquad (27)$$

where $(\mathrm{d}H)$ denotes the normalized invariant measure on $O(m)$. That is,

$$\begin{aligned} {}_pF_q^{(m)} & \left(a_1, \ldots, a_p; b_1, \ldots, b_q; X, Y \right) \\ & = \int_{O(m)} {}_pF_q^{(m)} \left(a_1, \ldots, a_p; b_1, \ldots, b_q; XHYH' \right) (\mathrm{d}H), \end{aligned} \qquad (28)$$

given in James [6, Equation 30].

Finally, we define the inverted matrix variate gamma, matrix variate beta type 1, and matrix variate beta type 2 distributions. These definitions can be found in Gupta and Nagar [1] and Iranmanesh et al. [10].

Definition 1. An $m \times m$ random symmetric positive definite matrix X is said to have an inverted matrix variate gamma distribution with parameters μ, θ, and Ψ, denoted by $X \sim \mathrm{InvGa}(m, \mu, \theta, \Psi)$, if its p.d.f. is given by

$$\frac{\det(X)^{-\mu-(m+1)/2} \operatorname{etr}\left(-\Psi^{-1}X^{-1}/\theta \right)}{\det(\theta\Psi)^\mu \Gamma_m(\mu)}, \qquad X > 0, \qquad (29)$$

where $\mu > (m-1)/2$, $\theta > 0$, and Ψ is a symmetric positive definite matrix of order m.

Definition 2. An $m \times m$ random symmetric positive definite matrix U is said to have a matrix variate beta type 1 distribution with parameters $a(> (m-1)/2)$ and $b(> (m-1)/2)$, denoted as $U \sim \mathrm{B1}(m, a, b)$, if its p.d.f. is given by

$$\frac{\det(U)^{a-(m+1)/2} \det(I_m - U)^{b-(m+1)/2}}{B_m(a, b)}, \qquad 0 < U < I_m. \qquad (30)$$

Definition 3. An $m \times m$ random symmetric positive definite matrix V is said to have a matrix variate beta type 2 distribution with parameters $a(> (m-1)/2)$ and $b(> (m-1)/2)$, denoted as $V \sim \mathrm{B2}(m, a, b)$, if its p.d.f. is given by

$$\frac{\det(V)^{a-(m+1)/2} \det(I_m + V)^{-(a+b)}}{B_m(a, b)}, \qquad V > 0. \qquad (31)$$

Note that if $U \sim \mathrm{B1}(m, a, b)$, then $(I_m - U)^{-1}U \sim \mathrm{B2}(m, a, b)$. Further, if X_1 and X_2 are independent, $X_1 \sim \mathrm{Ga}(m, \nu_1, \theta)$ and $X_2 \sim \mathrm{Ga}(m, \nu_2, \theta)$, then $(X_1 + X_2)^{-1/2} X_1 (X_1 + X_2)^{-1/2} \sim \mathrm{B1}(m, \nu_1, \nu_2)$ and $X_2^{-1/2} X_1 X_2^{-1/2} \sim \mathrm{B2}(m, \nu_1, \nu_2)$.

We conclude this section by evaluating the normalizing constant $C(\nu, \alpha, \beta, \theta, \Omega)$ in (3). Since the density over its support set integrates to one, we have

$$[C(\nu, \alpha, \beta, \theta, \Omega)]^{-1}$$
$$= \int_{X>0} \det(X)^{\nu-(m+1)/2} {}_1F_1 \left(\alpha; \beta; -\frac{1}{\theta}\Omega^{-1}X \right) \mathrm{d}X. \qquad (32)$$

By rewriting ${}_1F_1$ using Kummer's relation (24) and integrating X by applying (21), we get

$$[C(\nu, \alpha, \beta, \theta, \Omega)]^{-1}$$
$$= \Gamma_m(\nu) \det(\theta\Omega)^\nu {}_2F_1(\nu, \beta - \alpha; \beta; I_m), \qquad (33)$$

where $\mathrm{Re}(\nu) > (m-1)/2$. Finally, writing ${}_2F_1(\nu, \beta - \alpha; \beta; I_m)$ in terms of multivariate gamma functions by using (23), we obtain

$$[C(\nu, \alpha, \beta, \theta, \Omega)]^{-1}$$
$$= \int_{X>0} \det(X)^{\nu-(m+1)/2} {}_1F_1 \left(\alpha; \beta; -\frac{1}{\theta}\Omega^{-1}X \right) \mathrm{d}X$$
$$= \frac{\det(\theta\Omega)^\nu \Gamma_m(\nu) \Gamma_m(\beta) \Gamma_m(\alpha - \nu)}{\Gamma_m(\alpha) \Gamma_m(\beta - \nu)}, \qquad (34)$$

where $\mathrm{Re}(\beta - \nu) > (m-1)/2$, $\mathrm{Re}(\alpha - \nu) > (m-1)/2$, $\theta > 0$, $\mathrm{Re}(\nu) > (m-1)/2$, and $\Omega > 0$.

3. Properties

In this section we study several properties of the confluent hypergeometric function kind 1 distribution defined in Section 1. For the sake of completeness, we first state the following results established in Gupta and Nagar [1].

(1) Let $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \Omega, \mathrm{kind}\ 1)$ and let A be an $m \times m$ constant nonsingular matrix. Then, $AXA' \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, A\Omega A', \mathrm{kind}\ 1)$.

(2) Let $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \mathrm{kind}\ 1)$ and let H be an $m \times m$ orthogonal matrix whose elements are either constants or random variables distributed independent of X. Then, the distribution of X is invariant under the transformation $X \to HXH'$ if H is a matrix of constants. Further, if H is a random matrix, then H and HXH' are independent.

(3) Let $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \Omega, \mathrm{kind}\ 1)$. Then, the cumulative distribution function (cdf) of X is derived as

$$\begin{aligned} P(X \le \Lambda) &= \frac{\Gamma_m(\alpha) \Gamma_m(\beta - \nu) \Gamma_m[(m+1)/2]}{\Gamma_m[\nu + (m+1)/2] \Gamma_m(\beta) \Gamma_m(\alpha - \nu)} \\ &\quad \cdot \det(\theta\Omega)^{-\nu} \det(\Lambda)^\nu \\ &\quad \cdot {}_2F_2 \left(\alpha, \nu; \beta, \nu + \frac{m+1}{2}; -\frac{1}{\theta}\Omega^{-1}\Lambda \right), \end{aligned} \qquad (35)$$

where $\Lambda > 0$.

(4) Let $X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}$, where X_{11} is a $q \times q$ matrix. Define $X_{11 \cdot 2} = X_{11} - X_{12}X_{22}^{-1}X_{21}$ and $X_{22 \cdot 1} = X_{22} - X_{21}X_{11}^{-1}X_{12}$. If $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \mathrm{kind}\ 1)$, then (i) X_{11} and $X_{22 \cdot 1}$ are independent, $X_{11} \sim \mathrm{CH}_q(\nu, \alpha - (m-q)/2, \beta - (m-q)/2, \theta, \mathrm{kind}\ 1)$ and $X_{22 \cdot 1} \sim \mathrm{CH}_{m-q}(\nu - q/2, \alpha - q/2, \beta - q/2, \theta, \mathrm{kind}\ 1)$, and (ii) X_{22} and $X_{11 \cdot 2}$ are independent, $X_{22} \sim \mathrm{CH}_{m-q}(\nu, \alpha - q/2, \beta - q/2, \theta, \mathrm{kind}\ 1)$ and $X_{11 \cdot 2} \sim \mathrm{CH}_q(\nu - (m-q)/2, \alpha - (m-q)/2, \beta - (m-q)/2, \theta, \mathrm{kind}\ 1)$.

(5) Let A be a $q \times m$ constant matrix of rank $q (\leq m)$. If $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \text{kind } 1)$, then $AXA' \sim \mathrm{CH}_q(\nu, \alpha-(m-q)/2, \beta-(m-q)/2, \theta, AA', \text{kind } 1)$ and $(AX^{-1}A')^{-1} \sim \mathrm{CH}_q(\nu-(m-q)/2, \alpha-(m-q)/2, \beta-(m-q)/2, \theta, (AA')^{-1}, \text{kind } 1)$.

(6) Let $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \text{kind } 1)$ and let $\mathbf{a}$ be a nonzero m-dimensional column vector of constants, then $(\mathbf{a}'\mathbf{a})^{-1}(\mathbf{a}'X\mathbf{a}) \sim \mathrm{CH}(\nu, \alpha-(m-1)/2, \beta-(m-1)/2, \theta, \text{kind } 1)$ and $\mathbf{a}'\mathbf{a}(\mathbf{a}'X^{-1}\mathbf{a})^{-1} \sim \mathrm{CH}(\nu-(m-1)/2, \alpha-(m-1)/2, \beta-(m-1)/2, \theta, \text{kind } 1)$. Further, if $\mathbf{y}$ is an m-dimensional random vector, independent of X, and $P(\mathbf{y} \neq \mathbf{0}) = 1$, then it follows that $(\mathbf{y}'\mathbf{y})^{-1}(\mathbf{y}'X\mathbf{y}) \sim \mathrm{CH}(\nu, \alpha-(m-1)/2, \beta-(m-1)/2, \theta, \text{kind } 1)$ and $\mathbf{y}'\mathbf{y}(\mathbf{y}'X^{-1}\mathbf{y})^{-1} \sim \mathrm{CH}(\nu-(m-1)/2, \alpha-(m-1)/2, \beta-(m-1)/2, \theta, \text{kind } 1)$.

It may also be mentioned here that properties (1)–(6) given above are modified forms of results given in Section 8.10 of Gupta and Nagar [1].

If the $m \times m$ random matrices X_1 and X_2 are independent, $X_1 \sim \mathrm{CH}_m(\nu, a_1 + \nu, 2\nu, \theta, \text{kind } 1)$ and $X_2 \sim \mathrm{CH}_m(\nu, a_2 + \nu, 2\nu, \theta, \text{kind } 1)$, $\nu = \gamma/2 + (m+1)/4$, then Roux and van der Merwe [11] have shown that $X_2^{-1/2}X_1 X_2^{-1/2}$ has matrix variate beta type 2 distributions with parameters a_2 and a_1.

The matrix variate confluent hypergeometric function kind 1 distribution can be derived as the distribution of the matrix ratio of independent gamma and beta matrices. It has been shown in Gupta and Nagar [1] that if $Y \sim \mathrm{Ga}(m, \nu, \theta)$ and $U \sim \mathrm{B1}(m, a, b)$, then $U^{-1/2}YU^{-1/2} \sim \mathrm{CH}_m(\nu, a + \nu, a + b + \nu, \theta, \text{kind } 1)$.

The expected values of X and X^{-1}, for $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \text{kind } 1)$, can easily be obtained from the above results. For any fixed $\mathbf{a} \in \mathbb{R}^m$, $\mathbf{a} \neq \mathbf{0}$,

$$E\left[\frac{\mathbf{a}'X\mathbf{a}}{\mathbf{a}'\mathbf{a}}\right] = E(\nu_1), \tag{36}$$

where $\nu_1 \sim \mathrm{CH}(\nu, \alpha-(m-1)/2, \beta-(m-1)/2, \theta, \text{kind } 1)$, and

$$E\left[\frac{\mathbf{a}'X^{-1}\mathbf{a}}{\mathbf{a}'\mathbf{a}}\right] = E\left(\frac{1}{\nu_2}\right), \tag{37}$$

where $\nu_2 \sim \mathrm{CH}(\nu-(m-1)/2, \alpha-(m-1)/2, \beta-(m-1)/2, \theta, \text{kind } 1)$. Hence, for all $\mathbf{a} \in \mathbb{R}^m$,

$$\mathbf{a}'E(X)\mathbf{a} = \mathbf{a}'\mathbf{a}E(\nu_1)$$
$$= \frac{\nu\theta\left[\beta - \nu - (m+1)/2\right]}{\left[\alpha - \nu - (m+1)/2\right]}\mathbf{a}'\mathbf{a},$$
$$\alpha - \nu > \frac{m+1}{2}, \quad \beta - \nu > \frac{m+1}{2}, \tag{38}$$

$$\mathbf{a}'E(X^{-1})\mathbf{a} = \mathbf{a}'\mathbf{a}E\left(\frac{1}{\nu_2}\right)$$
$$= \frac{\alpha - \nu}{\theta\left[\nu - (m+1)/2\right](\beta - \nu)}\mathbf{a}'\mathbf{a},$$
$$\nu > \frac{m+1}{2},$$

which implies that

$$E(X) = \frac{\nu\theta\left[\beta - \nu - (m+1)/2\right]}{\left[\alpha - \nu - (m+1)/2\right]}I_m,$$
$$\alpha - \nu > \frac{m+1}{2}, \quad \beta - \nu > \frac{m+1}{2}, \tag{39}$$

$$E(X^{-1}) = \frac{\alpha - \nu}{\theta\left[\nu - (m+1)/2\right](\beta - \nu)}I_m,$$
$$\nu > \frac{m+1}{2}.$$

The Laplace transform of the density of X, where $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \text{kind } 1)$, is given by

$$\frac{\theta^{-m\nu}\Gamma_m(\alpha)\Gamma_m(\beta - \nu)}{\Gamma_m(\nu)\Gamma_m(\beta)\Gamma_m(\alpha - \nu)}\int_{X>0} \mathrm{etr}\left(-ZX - \frac{X}{\theta}\right)$$
$$\cdot \det(X)^{\nu-(m+1)/2} \, {}_1F_1\left(\beta - \alpha; \beta; \frac{X}{\theta}\right)dX$$
$$= \frac{\Gamma_m(\alpha)\Gamma_m(\beta - \nu)}{\Gamma_m(\beta)\Gamma_m(\alpha - \nu)}\det(I_m + \theta Z)^{-\nu} \, {}_2F_1\left(\nu, \beta - \alpha; \beta; (I_m + \theta Z)^{-1}\right), \quad \mathrm{Re}(Z) > 0, \tag{40}$$

where we have used (24) and (21). From the above expression, the Laplace transform of the density of X, where $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \Omega, \text{kind } 1)$, is derived as

$$\frac{\Gamma_m(\alpha)\Gamma_m(\beta - \nu)}{\Gamma_m(\beta)\Gamma_m(\alpha - \nu)}\det(I_m + \theta\Omega Z)^{-\nu}$$
$$\cdot {}_2F_1\left(\nu, \beta - \alpha; \beta; (I_m + \theta\Omega Z)^{-1}\right), \quad \mathrm{Re}(Z) > 0. \tag{41}$$

Theorem 4. *Let $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \Omega, \text{kind } 1)$; then*

$$E\left[\det(X)^h\right]$$
$$= \frac{\det(\theta\Omega)^h \Gamma_m(\beta - \nu)\Gamma_m(\nu + h)\Gamma_m(\alpha - \nu - h)}{\Gamma_m(\nu)\Gamma_m(\alpha - \nu)\Gamma_m(\beta - \nu - h)}, \tag{42}$$

where $\mathrm{Re}(h + \nu) > (m-1)/2$, $\mathrm{Re}(h) < \alpha - \nu - (m-1)/2$, and $\mathrm{Re}(h) < \beta - \nu - (m-1)/2$.

Proof. From the density of X, we have

$$E\left[\det(X)^h\right] = \frac{\det(\theta\Omega)^{-\nu}\Gamma_m(\alpha)\Gamma_m(\beta - \nu)}{\Gamma_m(\nu)\Gamma_m(\beta)\Gamma_m(\alpha - \nu)}$$
$$\cdot \int_{X>0} \det(X)^{\nu+h-(m+1)/2} \, {}_1F_1\left(\alpha; \beta; -\frac{1}{\theta}\Omega^{-1}X\right)dX. \tag{43}$$

Now, evaluating the above integral by using (34), we get

$$E\left[\det(X)^h\right] = \frac{\det(\theta\Omega)^{-\nu}\Gamma_m(\alpha)\Gamma_m(\beta - \nu)}{\Gamma_m(\nu)\Gamma_m(\beta)\Gamma_m(\alpha - \nu)}$$
$$\cdot \frac{\det(\theta\Omega)^{\nu+h}\Gamma_m(\nu + h)\Gamma_m(\beta)\Gamma_m(\alpha - \nu - h)}{\Gamma_m(\alpha)\Gamma_m(\beta - \nu - h)}, \tag{44}$$

where $\text{Re}(h + v) > (m - 1)/2$, $\text{Re}(h) < \alpha - v - (m - 1)/2$, and $\text{Re}(h) < \beta - v - (m - 1)/2$. Finally, simplifying the above expression, we get the desired result. □

Corollary 5. *Let $x \sim \text{CH}(v, \alpha, \beta, \theta, \text{kind } 1)$; then*

$$E\left(x^h\right) = \frac{\theta^h \Gamma(\beta - v)\Gamma(v + h)\Gamma(\alpha - v - h)}{\Gamma(v)\Gamma(\alpha - v)\Gamma(\beta - v - h)}, \qquad (45)$$

where $\text{Re}(h + v) > 0$, $\text{Re}(h) < \alpha - v$, and $\text{Re}(h) < \beta - v$.

$$\text{Var}(\det(X)) = \theta^{2m}\det(\Omega)^2 \prod_{j=1}^m \left[\frac{[v - (j - 1)/2][\beta - v - (j + 1)/2]}{[\alpha - v - (j + 1)/2]} \left\{ \frac{[v - (j - 3)/2][\beta - v - (j + 3)/2]}{[\alpha - v - (j + 3)/2]} \right. \right.$$
$$\left. \left. - \frac{[v - (j - 1)/2][\beta - v - (j + 1)/2]}{[\alpha - v - (j + 1)/2]} \right\} \right], \qquad (47)$$

where $\beta > v + (m + 3)/2$ and $\alpha > v + (m + 3)/2$. For $m \times m$ symmetric matrix A, $E[C_\kappa(AX)]$ is derived as

$$E[C_\kappa(AX)] = \frac{\det(\theta\Omega)^{-v}\Gamma_m(\alpha)\Gamma_m(\beta - v)}{\Gamma_m(v)\Gamma_m(\beta)\Gamma_m(\alpha - v)}$$
$$\cdot \int_{X>0} C_\kappa(AX)\det(X)^{v-(m+1)/2} \qquad (48)$$
$$\cdot {}_1F_1\left(\alpha; \beta; -\frac{1}{\theta}\Omega^{-1}X\right)dX.$$

Replacing ${}_1F_1(\alpha; \beta; -\Omega^{-1}X/\theta)$ by its integral representation, namely,

$${}_1F_1\left(\alpha; \beta; -\frac{1}{\theta}\Omega^{-1}X\right) = \frac{\Gamma_m(\beta)}{\Gamma_m(\alpha)\Gamma_m(\beta - \alpha)}$$
$$\cdot \int_0^{I_m} \text{etr}\left(-\frac{1}{\theta}\Omega^{-1/2}X\Omega^{-1/2}Y\right)\det(Y)^{\alpha-(m+1)/2} \qquad (49)$$
$$\cdot \det(I_m - Y)^{\beta-\alpha-(m+1)/2} dY,$$

where $\text{Re}(\beta - \alpha) > (m - 1)/2$ and $\text{Re}(\alpha) > (m - 1)/2$, one obtains

$$E[C_\kappa(AX)] = \frac{\det(\theta\Omega)^{-v}\Gamma_m(\beta - v)}{\Gamma_m(v)\Gamma_m(\beta - \alpha)\Gamma_m(\alpha - v)}$$
$$\cdot \int_0^{I_m} \det(Y)^{\alpha-(m+1)/2}\det(I_m - Y)^{\beta-\alpha-(m+1)/2} \qquad (50)$$
$$\cdot \int_{X>0} C_\kappa(AX)\det(X)^{v-(m+1)/2}$$
$$\cdot \text{etr}\left(-\frac{1}{\theta}\Omega^{-1/2}X\Omega^{-1/2}Y\right)dX\,dY.$$

Using (42) the mean and the variance of $\det(X)$ are derived as

$$E[\det(X)] = \theta^m \det(\Omega)$$
$$\cdot \prod_{j=1}^m \frac{[v - (j - 1)/2][\beta - v - (j + 1)/2]}{[\alpha - v - (j + 1)/2]}, \qquad (46)$$

where $\beta > v + (m + 1)/2$, $\alpha > v + (m + 1)/2$, and

Now, evaluating the above integral by using (19), we obtain

$$E[C_\kappa(AX)] = \frac{\theta^k\Gamma_m(\beta - v)(v)_\kappa}{\Gamma_m(\beta - \alpha)\Gamma_m(\alpha - v)}$$
$$\cdot \int_0^{I_m}\det(Y)^{\alpha-v-(m+1)/2}\det(I_m - Y)^{\beta-\alpha-(m+1)/2} \qquad (51)$$
$$\cdot C_\kappa\left(\Omega^{1/2}A\Omega^{1/2}Y^{-1}\right)dY.$$

Finally, evaluating the integral involving Y by using (Khatrzi [12])

$$\int_0^{I_m}\det(R)^{a-(m+1)/2}\det(I_m - R)^{b-(m+1)/2}$$
$$\cdot C_\kappa\left(XR^{-1}\right)dR$$
$$= \frac{(-a - b + (m + 1)/2)_\kappa\Gamma_m(a)\Gamma_m(b)}{(-a + (m + 1)/2)_\kappa\Gamma_m(a + b)}C_\kappa(X), \qquad (52)$$
$$\text{Re}(a) > \frac{m - 1}{2} + k_1, \quad \text{Re}(b) > \frac{m - 1}{2},$$

we get

$$E[C_\kappa(AX)]$$
$$= \frac{\theta^k(v - \beta + (m + 1)/2)_\kappa(v)_\kappa}{(v - \alpha + (m + 1)/2)_\kappa}C_\kappa(\Omega A), \qquad (53)$$

where $\text{Re}(\alpha - v) > (m - 1)/2 + k_1$ and $\text{Re}(\beta - v) > (m - 1)/2 + k_1$. Proceeding similarly and using the result (Khatrzi [12])

$$\int_{R>0}\text{etr}(-XR)\det(R)^{a-(m+1)/2}C_\kappa\left(R^{-1}S\right)dR$$
$$= \frac{(-1)^k\Gamma_m(a)}{(-a + (m + 1)/2)_\kappa}\det(X)^{-a}C_\kappa(XS), \qquad (54)$$
$$\text{Re}(a) > \frac{m - 1}{2} + k_1,$$

the expected value of $C_\kappa(AX^{-1})$ is derived as

$$E\left[C_\kappa\left(AX^{-1}\right)\right] = \frac{\det(\theta\Omega)^{-\nu}\,\Gamma_m\,(\beta-\nu)}{\Gamma_m\,(\nu)\,\Gamma_m\,(\beta-\alpha)\,\Gamma_m\,(\alpha-\nu)}$$

$$\cdot \int_0^{I_m} \det(Y)^{\alpha-(m+1)/2}\det\left(I_m-Y\right)^{\beta-\alpha-(m+1)/2}$$

$$\cdot \int_{X>0} C_\kappa\left(AX^{-1}\right)\det(X)^{\nu-(m+1)/2}$$

$$\cdot \operatorname{etr}\left(-\frac{1}{\theta}\Omega^{-1/2}X\Omega^{-1/2}Y\right)\mathrm{d}X\,\mathrm{d}Y \tag{55}$$

$$= \frac{(-1)^k\,\theta^{-k}\Gamma_m\,(\beta-\nu)}{\Gamma_m\,(\beta-\alpha)\,\Gamma_m\,(\alpha-\nu)\,(-\nu+(m+1)/2)_\kappa}$$

$$\cdot \int_0^{I_m} \det(Y)^{\alpha-\nu-(m+1)/2}\det\left(I_m-Y\right)^{\beta-\alpha-(m+1)/2}$$

$$\cdot C_\kappa\left(\Omega^{-1/2}A\Omega^{-1/2}Y\right)\mathrm{d}Y,$$

where $\operatorname{Re}(\nu) > (m-1)/2 + k_1$. Finally, evaluating the above integral using (20) we obtain

$$E\left[C_\kappa\left(AX^{-1}\right)\right]$$

$$= \frac{(-1)^k\,\theta^{-k}\,(\alpha-\nu)_\kappa}{(\beta-\nu)_\kappa\,(-\nu+(m+1)/2)_\kappa}C_\kappa\left(\Omega^{-1}A\right), \tag{56}$$

$$\operatorname{Re}(\nu) > \frac{m-1}{2} + k_1.$$

In the next theorem, we derive the confluent hypergeometric function kind 1 distribution using independent beta and gamma matrices.

Theorem 6. *Let X_1 and X_2 be independent, $X_1 \sim \mathrm{Ga}(m,\nu,\theta)$ and $X_2 \sim \mathrm{B1}(m,a,b)$. Then, $X_2^{-1/2}X_1X_2^{-1/2} \sim \mathrm{CH}_m(\nu,a+\nu,a+b+\nu,\theta,\mathrm{kind}\ 1)$.*

Proof. See Gupta and Nagar [1]. $\qquad\square$

Theorem 7. *Let X_1 and X_2 be independent, $X_1 \sim \mathrm{Ga}(m,\nu,\theta)$ and $X_2 \sim \mathrm{B1}(m,a,b)$. Then, $X_1^{1/2}X_2^{-1}X_1^{1/2} \sim \mathrm{CH}_m(\nu,a+\nu,a+b+\nu,\theta,\mathrm{kind}\ 1)$.*

Proof. The result follows from Theorem 6 and the fact that $X_2^{-1/2}X_1X_2^{-1/2}$ and $X_1^{1/2}X_2^{-1}X_1^{1/2}$ have same eigenvalues, and the matrix variate confluent hypergeometric function kind 1 distribution is orthogonally invariant. $\qquad\square$

Theorem 8. *Let X_1 and X_2 be independent, $X_1 \sim \mathrm{Ga}(m,\nu,\theta)$ and $X_2 \sim \mathrm{B1}(m,a,b)$. Then, $(I_m-X_2)^{-1/2}X_1(I_m-X_2)^{-1/2} \sim \mathrm{CH}_m(\nu,b+\nu,a+b+\nu,\theta,\mathrm{kind}\ 1)$.*

Proof. Noting that $I_m-X_2 \sim \mathrm{B1}(m,b,a)$ and using Theorem 6 we get the result. $\qquad\square$

Theorem 9. *Let X_1 and X_2 be independent, $X_1 \sim \mathrm{Ga}(m,\nu,\theta)$ and $X_2 \sim \mathrm{B2}(m,a,b)$. Then, $(I_m+X_2)^{1/2}X_1(I_m+X_2)^{1/2} \sim \mathrm{CH}_m(\nu,b+\nu,a+b+\nu,\theta,\mathrm{kind}\ 1)$.*

Proof. The desired result is obtained by observing that $(I_m+X_2)^{-1} \sim \mathrm{B1}(m,b,a)$ and using Theorem 6. $\qquad\square$

Theorem 10. *Let X_1 and X_2 be independent, $X_1 \sim \mathrm{Ga}(m,\nu,\theta)$ and $X_2 \sim \mathrm{B2}(m,a,b)$. Then, $(I_m+X_2^{-1})^{1/2}X_1(I_m+X_2^{-1})^{1/2} \sim \mathrm{CH}_m(\nu,a+\nu,a+b+\nu,\theta,\mathrm{kind}\ 1)$.*

Proof. Noting that $(I_m+X_2^{-1})^{-1} \sim \mathrm{B1}(m,a,b)$ and using Theorem 6 we get the result. $\qquad\square$

Theorem 11. *Let X_1 and X_2 be independent, $X_1 \sim \mathrm{Ga}(m,\nu_1,\theta)$ and $X_2 \sim \mathrm{Ga}(m,\nu_2,\theta)$. Then, $(X_1+X_2)X_1^{-1}(X_1+X_2) \sim \mathrm{CH}_m(\nu_1+\nu_2,2\nu_1+\nu_2,2\nu_1+2\nu_2,\theta,\mathrm{kind}\ 1)$.*

Proof. It is well known that X_1+X_2 and $(X_1+X_2)^{-1/2}X_1(X_1+X_2)^{-1/2}$ are independent, $X_1+X_2 \sim \mathrm{Ga}(m,\nu_1+\nu_2,\theta)$ and $(X_1+X_2)^{-1/2}X_1(X_1+X_2)^{-1/2} \sim \mathrm{B1}(m,\nu_1,\nu_2)$. Therefore, using Theorem 7, $(X_1+X_2)X_1^{-1}(X_1+X_2) \sim \mathrm{CH}_m(\nu_1+\nu_2,2\nu_1+\nu_2,2\nu_1+2\nu_2,\theta,\mathrm{kind}\ 1)$. $\qquad\square$

Theorem 12. *Let X_1 and X_2 be independent, $X_1 \sim \mathrm{Ga}(m,\nu_1,\theta)$ and $X_2 \sim \mathrm{Ga}(m,\nu_2,\theta)$. Then, $(X_1+X_2)X_2^{-1}(X_1+X_2) \sim \mathrm{CH}_m(\nu_1+\nu_2,\nu_1+2\nu_2,2\nu_1+2\nu_2,\theta,\mathrm{kind}\ 1)$.*

Proof. The proof is similar to the proof of Theorem 11. $\qquad\square$

4. Distributions of Sum and Quotients

In statistical distribution theory it is well known that if X_1 and X_2 are independent, $X_1 \sim \mathrm{Ga}(m,\nu_1,\theta)$ and $X_2 \sim \mathrm{Ga}(m,\nu_2,\theta)$, then $X_2^{-1/2}X_1X_2^{-1/2} \sim \mathrm{B2}(m,\nu_1,\nu_2)$, $(X_1+X_2)^{-1/2}X_1(X_1+X_2)^{-1/2} \sim \mathrm{B1}(m,\nu_1,\nu_2)$, and $X_1+X_2 \sim \mathrm{Ga}(m,\nu_1+\nu_2)$. In this section we derive similar results when X_1 and X_2 are independent confluent hypergeometric function kind 1 and gamma matrices, respectively.

Theorem 13. *Let X_1 and X_2 be independent, $X_1 \sim \mathrm{CH}_m(\nu_1,\alpha_1,\beta_1,\theta,\mathrm{kind}\ 1)$ and $X_2 \sim \mathrm{Ga}(m,\nu_2,\theta)$. Then, the p.d.f. of $Z = X_2^{-1/2}X_1X_2^{-1/2}$ is given by*

$$\frac{\Gamma_m\,(\alpha_1)\,\Gamma_m\,(\beta_1-\nu_1)\,\Gamma_m\,(\nu_1+\nu_2)}{\Gamma_m\,(\beta_1)\,\Gamma_m\,(\alpha_1-\nu_1)\,\Gamma_m\,(\nu_1)\,\Gamma_m\,(\nu_2)}\frac{\det(Z)^{\nu_1-(m+1)/2}}{\det(I_m+Z)^{\nu_1+\nu_2}}$$

$$\cdot \,_2F_1\left(\beta_1-\alpha_1,\nu_1+\nu_2;\beta_1;(I_m+Z)^{-1}Z\right), \tag{57}$$

$$Z > 0.$$

Proof. Using the independence, the joint p.d.f. of X_1 and X_2 is given by

$$K\det\left(X_1\right)^{\nu_1-(m+1)/2}\det\left(X_2\right)^{\nu_2-(m+1)/2}$$

$$\cdot \operatorname{etr}\left[-\frac{1}{\theta}\left(X_1+X_2\right)\right]\,_1F_1\left(\beta_1-\alpha_1;\beta_1;\frac{1}{\theta}X_1\right), \tag{58}$$

$$X_1 > 0,\ \ X_2 > 0,$$

where

$$K = \frac{\theta^{-m(\nu_1+\nu_2)}\Gamma_m(\alpha_1)\,\Gamma_m(\beta_1-\nu_1)}{\Gamma_m(\nu_1)\,\Gamma_m(\nu_2)\,\Gamma_m(\beta_1)\,\Gamma_m(\alpha_1-\nu_1)}. \tag{59}$$

Making the transformation $Z = X_2^{-1/2}X_1X_2^{-1/2}$, $X_2 = X_2$, with the Jacobian $J(X_1, X_2 \rightarrow Z, X_2) = \det(X_2)^{(m+1)/2}$ we obtain the joint p.d.f. of Z and X_2 as

$$K \det(Z)^{\nu_1-(m+1)/2} \det(X_2)^{\nu_1+\nu_2-(m+1)/2}$$

$$\cdot \exp\left[-\frac{1}{\theta}(I_m+Z)\,X_2\right] {}_1F_1\left(\beta_1-\alpha_1;\beta_1;\frac{1}{\theta}ZX_2\right), \tag{60}$$

$$Z>0, \ X_2>0.$$

Now, integrating X_2 in (60) by applying (21) and substituting for K, we obtain the desired result. $\qquad\square$

Corollary 14. *Let* X_1 *and* X_2 *be independent,* $X_1 \sim$ *CH*$_m(\nu_1, \alpha_1, \beta_1, \theta, \text{kind } 1)$ *and* $X_2 \sim$ *Ga*(m, ν_2, θ). *Then, the p.d.f. of* $Z_1 = X_2^{1/2}X_1^{-1}X_2^{1/2}$ *is given by*

$$\frac{\Gamma_m(\alpha_1)\,\Gamma_m(\beta_1-\nu_1)\,\Gamma_m(\nu_1+\nu_2)}{\Gamma_m(\beta_1)\,\Gamma_m(\alpha_1-\nu_1)\,\Gamma_m(\nu_1)\,\Gamma_m(\nu_2)}$$

$$\cdot \frac{\det(Z_1)^{\nu_2-(m+1)/2}}{\det(I_m+Z_1)^{\nu_1+\nu_2}} \tag{61}$$

$$\cdot {}_2F_1\left(\beta_1-\alpha_1,\nu_1+\nu_2;\beta_1;(I_m+Z_1)^{-1}\right), \quad Z_1>0.$$

Corollary 15. *Let* X_1 *and* X_2 *be independent,* $X_1 \sim$ *Ga*(m, ν_1, θ) *and* $X_2 \sim$ *CH*$_m(\nu_2, \alpha_2, \beta_2, \theta, \text{kind } 1)$. *Then, the p.d.f. of* $Z_3 = X_2^{-1/2}X_1X_2^{-1/2}$ *is given by*

$$\frac{\Gamma_m(\alpha_2)\,\Gamma_m(\beta_2-\nu_2)\,\Gamma_m(\nu_1+\nu_2)}{\Gamma_m(\beta_2)\,\Gamma_m(\alpha_2-\nu_2)\,\Gamma_m(\nu_1)\,\Gamma_m(\nu_2)}$$

$$\cdot \frac{\det(Z_3)^{\nu_1-(m+1)/2}}{\det(I_m+Z_3)^{\nu_1+\nu_2}} \tag{62}$$

$$\cdot {}_2F_1\left(\beta_2-\alpha_2,\nu_1+\nu_2;\beta_2;(I_m+Z_3)^{-1}\right), \quad Z_3>0.$$

Proof. Interchanging subscripts 1 and 2 in Corollary 14, the p.d.f. of $Z_2 = X_1^{1/2}X_2^{-1}X_1^{1/2}$ is given by

$$\frac{\Gamma_m(\alpha_2)\,\Gamma_m(\beta_2-\nu_2)\,\Gamma_m(\nu_1+\nu_2)}{\Gamma_m(\beta_2)\,\Gamma_m(\alpha_2-\nu_2)\,\Gamma_m(\nu_1)\,\Gamma_m(\nu_2)}$$

$$\cdot \frac{\det(Z_2)^{\nu_1-(m+1)/2}}{\det(I_m+Z_2)^{\nu_1+\nu_2}} \tag{63}$$

$$\cdot {}_2F_1\left(\beta_2-\alpha_2,\nu_1+\nu_2;\beta_2;(I_m+Z_2)^{-1}\right), \quad Z_2>0,$$

where $X_1 \sim$ Ga(m, ν_1, θ) and $X_2 \sim$ CH$_m(\nu_2, \alpha_2, \beta_2, \theta, \text{kind } 1)$. Now, the result follows from the fact that $Z_2 = X_1^{1/2}X_2^{-1}X_1^{1/2}$ and $Z_3 = X_2^{-1/2}X_1X_2^{-1/2}$ have the same eigenvalues, and the distribution Z_2 is orthogonally invariant. $\quad\square$

Corollary 16. *Let* X_1 *and* X_2 *be independent,* $X_1 \sim$ *CH*$_m(\nu_1, \alpha_1, \nu_1+\nu_2, \theta, \text{kind } 1)$ *and* $X_2 \sim$ *Ga*(m, ν_2, θ). *Then,* $X_2^{-1/2}X_1X_2^{-1/2} \sim$ *B2*$(m, \nu_1, \alpha_1-\nu_1)$.

Corollary 17. *Let the random matrices* X_1 *and* X_2 *be independent,* $X_1 \sim$ *Ga*(m, ν_1, θ) *and* $X_2 \sim$ *CH*$_m(\nu_2, \alpha_2, \nu_2+\nu_1, \theta, \text{kind } 1)$. *Then,* $X_2^{-1/2}X_1X_2^{-1/2} \sim$ *B2*$(m, \alpha_2-\nu_2, \nu_2)$.

Corollary 18. *Let* X_1 *and* X_2 *be independent,* $X_1 \sim$ *Ga*(m, ν_1, θ) *and* $X_2 \sim$ *Ga*(m, ν_2, θ). *Then,* $X_2^{-1/2}X_1X_2^{-1/2} \sim$ *B2*(m, ν_1, ν_2).

Theorem 19. *Let* X_1, X_2, *and* X_3 *be independent,* $X_1 \sim$ *Ga*(m, μ, θ), $X_2 \sim$ *B1*(m, a, b), *and* $X_3 \sim$ *Ga*(m, ν, θ). *Then, the p.d.f. of* $Z = X_3^{-1/2}X_2^{-1/2}X_1X_2^{-1/2}X_3^{-1/2}$ *is given by*

$$\frac{\Gamma_m(a+\mu)\,\Gamma_m(a+b)\,\Gamma_m(\mu+\nu)}{\Gamma_m(a+b+\mu)\,\Gamma_m(a)\,\Gamma_m(\mu)\,\Gamma_m(\nu)}\,\frac{\det(Z)^{\mu-(m+1)/2}}{\det(I_m+Z)^{\mu+\nu}} \tag{64}$$

$$\cdot {}_2F_1\left(b, \mu+\nu; a+b+\mu; (I_m+Z)^{-1}Z\right), \quad Z>0.$$

Proof. Using the independence of X_1 and X_2 and Theorem 6, $X_2^{-1/2}X_1X_2^{-1/2} \sim$ CH$_m(\mu, a+\mu, a+b+\mu, \text{kind } 1)$. Further, using independence of $X_2^{-1/2}X_1X_2^{-1/2}$ and X_3 and Theorem 13, we obtain the desired result. $\quad\square$

Corollary 20. *Let* X_1, X_2, *and* X_3 *be independent,* $X_1 \sim$ *Ga*(m, μ, θ), $X_2 \sim$ *B1*(m, a, b), *and* $X_3 \sim$ *Ga*$(m, a+b, \theta)$. *Then,* $X_3^{-1/2}X_2^{-1/2}X_1X_2^{-1/2}X_3^{-1/2} \sim$ *B2*(m, μ, a).

Proof. For $\nu = a+b$, the p.d.f. of $Z = X_3^{-1/2}X_2^{-1/2}X_1X_2^{-1/2} \cdot X_3^{-1/2}$ given in the above theorem slides to

$$\frac{\Gamma_m(a+\mu)}{\Gamma_m(a)\,\Gamma_m(\mu)}\,\frac{\det(Z)^{\mu-(m+1)/2}}{\det(I_m+Z)^{\mu+a+b}}$$

$$\cdot {}_2F_1\left(b, \mu+a+b; a+b+\mu; (I_m+Z)^{-1}Z\right), \tag{65}$$

$$Z>0.$$

Now, simplifying the Gauss hypergeometric function as

$${}_2F_1\left(b, \mu+a+b; a+b+\mu; (I_m+Z)^{-1}Z\right)$$

$$= {}_1F_0\left(b; (I_m+Z)^{-1}Z\right) \tag{66}$$

$$= \det\left(I_m - (I_m+Z)^{-1}Z\right)^{-b} = \det(I_m+Z)^{b},$$

where we have used (16), the desired result is obtained. $\quad\square$

Corollary 21. *Let* V *and* U *be independent,* $V \sim$ *B2*(m, μ, ν) *and* $U \sim$ *B1*(m, a, b). *Then, the p.d.f. of* $Z = U^{-1/2}VU^{-1/2}$ *is given by*

$$\frac{\Gamma_m(a+\mu)\,\Gamma_m(a+b)\,\Gamma_m(\mu+\nu)}{\Gamma_m(a+b+\mu)\,\Gamma_m(a)\,\Gamma_m(\mu)\,\Gamma_m(\nu)}\,\frac{\det(Z)^{\mu-(m+1)/2}}{\det(I_m+Z)^{\mu+\nu}} \tag{67}$$

$$\cdot {}_2F_1\left(b, \mu+\nu; a+b+\mu; (I_m+Z)^{-1}Z\right), \quad Z>0.$$

Further, if $\nu = a+b$, *then* $U^{-1/2}VU^{-1/2} \sim$ *B2*(m, μ, a).

Proof. Observe that $Z = X_3^{-1/2}X_2^{-1/2}X_1X_2^{-1/2}X_3^{-1/2}$ and $X_2^{-1/2}X_3^{-1/2}X_1X_3^{-1/2}X_2^{-1/2}$ have same eigenvalues and the distribution of Z is orthogonally invariant. Therefore, the random matrices $X_3^{-1/2}X_2^{-1/2}X_1X_2^{-1/2}X_3^{-1/2}$ and $X_2^{-1/2}X_3^{-1/2}X_1X_3^{-1/2}X_2^{-1/2}$ have identical distribution. Now, setting $V = X_3^{-1/2}X_1X_3^{-1/2}$ and $U = X_2$, where $X_3^{-1/2}X_1X_3^{-1/2} \sim \text{B2}(m, \mu, \nu)$ and $X_2 \sim \text{B1}(m, a, b)$, we observe that $X_3^{-1/2}X_2^{-1/2}X_1X_2^{-1/2}X_3^{-1/2}$ and $U^{-1/2}VU^{-1/2}$ have identical distribution. □

Theorem 22. *Let X_1 and X_2 be independent, $X_1 \sim \text{CH}_m(\nu_1, \alpha_1, \beta_1, \theta, \text{kind 1})$ and $X_2 \sim \text{Ga}(m, \nu_2, \theta)$. Then, the p.d.f. of $R = (X_1 + X_2)^{-1/2}X_1(X_1 + X_2)^{-1/2}$ is given by*

$$\frac{\Gamma_m(\alpha_1)\,\Gamma_m(\beta_1 - \nu_1)\,\Gamma_m(\nu_1 + \nu_2)}{\Gamma_m(\beta_1)\,\Gamma_m(\alpha_1 - \nu_1)\,\Gamma_m(\nu_1)\,\Gamma_m(\nu_2)} \det(R)^{\nu_1 - (m+1)/2}$$

$$\cdot \det(I_m - R)^{\nu_2 - (m+1)/2} \tag{68}$$

$$\cdot \,_2F_1(\beta_1 - \alpha_1, \nu_1 + \nu_2; \beta_1; R), \quad 0 < R < I_m$$

and the p.d.f. of $S = X_1 + X_2$ is given by

$$\frac{\theta^{-m(\nu_1 + \nu_2)}\Gamma_m(\alpha_1)\,\Gamma_m(\beta_1 - \nu_1)}{\Gamma_m(\beta_1)\,\Gamma_m(\alpha_1 - \nu_1)\,\Gamma_m(\nu_1 + \nu_2)} \det(S)^{\nu_1 + \nu_2 - (m+1)/2}$$

$$\cdot \text{etr}\left(-\frac{1}{\theta}S\right) \,_2F_2\left(\nu_1, \beta_1 - \alpha_1; \nu_1 + \nu_2, \beta_1; \frac{1}{\theta}S\right), \tag{69}$$

$$S > 0.$$

Proof. Substituting $R = (X_1 + X_2)^{-1/2}X_1(X_1 + X_2)^{-1/2}$ and $S = X_1 + X_2$ with the Jacobian $J(X_1, X_2 \to R, S) = \det(S)^{(m+1)/2}$ in (58) we obtain the joint p.d.f. of R and S as

$$K \det(R)^{\nu_1 - (m+1)/2} \det(I_m - R)^{\nu_2 - (m+1)/2}$$

$$\cdot \det(S)^{\nu_1 + \nu_2 - (m+1)/2} \,\text{etr}\left(-\frac{1}{\theta}S\right) \tag{70}$$

$$\cdot \,_1F_1\left(\beta_1 - \alpha_1; \beta_1; \frac{1}{\theta}RS\right), \quad 0 < R < I_m, \, S > 0.$$

Now, integration of S by using (21) yields the density of R. The marginal density of S is obtained by integrating R by using (22). □

It may be remarked here that the density of R given in the above theorem can also be obtained from the density of $Z = X_2^{-1/2}X_1X_2^{-1/2}$ derived in Theorem 13 by making the transformation $R = (I_m + Z)^{-1}Z$.

Corollary 23. *Let X_1 and X_2 be independent random matrices, $X_1 \sim \text{Ga}(m, \nu_1, \theta)$ and $X_2 \sim \text{CH}_m(\nu_2, \alpha_2, \beta_2, \theta, \text{kind 1})$.*

Then, the p.d.f. of $R_1 = (X_1 + X_2)^{-1/2}X_1(X_1 + X_2)^{-1/2}$ is given by

$$\frac{\Gamma_m(\alpha_2)\,\Gamma_m(\beta_2 - \nu_2)\,\Gamma_m(\nu_1 + \nu_2)}{\Gamma_m(\beta_2)\,\Gamma_m(\alpha_2 - \nu_2)\,\Gamma_m(\nu_1)\,\Gamma_m(\nu_2)} \det(R_1)^{\nu_1 - (m+1)/2}$$

$$\cdot \det(I_m - R_1)^{\nu_2 - (m+1)/2} \tag{71}$$

$$\cdot \,_2F_1(\beta_2 - \alpha_2, \nu_1 + \nu_2; \beta_2; I_m - R_1), \quad 0 < R_1 < I_m$$

and the p.d.f. of $S_1 = X_1 + X_2$ is given by

$$\frac{\theta^{-m(\nu_1 + \nu_2)}\Gamma_m(\alpha_2)\,\Gamma_m(\beta_2 - \nu_2)}{\Gamma_m(\beta_2)\,\Gamma_m(\alpha_2 - \nu_2)\,\Gamma_m(\nu_1 + \nu_2)} \det(S_1)^{\nu_1 + \nu_2 - (m+1)/2}$$

$$\cdot \text{etr}\left(-\frac{1}{\theta}S_1\right) \,_2F_2\left(\nu_2, \beta_2 - \alpha_2; \nu_1 + \nu_2, \beta_2; \frac{1}{\theta}S_1\right), \tag{72}$$

$$S_1 > 0.$$

Proof. Interchanging subscripts 1 and 2 in Theorem 22, the p.d.f. of $R_2 = (X_1 + X_2)^{-1/2}X_2(X_1 + X_2)^{-1/2}$ is given by

$$\frac{\Gamma_m(\alpha_2)\,\Gamma_m(\beta_2 - \nu_2)\,\Gamma_m(\nu_1 + \nu_2)}{\Gamma_m(\beta_2)\,\Gamma_m(\alpha_2 - \nu_2)\,\Gamma_m(\nu_1)\,\Gamma_m(\nu_2)} \det(R_2)^{\nu_2 - (m+1)/2}$$

$$\cdot \det(I_m - R_2)^{\nu_2 - (m+1)/2} \tag{73}$$

$$\cdot \,_2F_1(\beta_2 - \alpha_2, \nu_1 + \nu_2; \beta_2; R_2), \quad 0 < R_2 < I_m,$$

where now $X_1 \sim \text{Ga}(m, \nu_1, \theta)$ and $X_2 \sim \text{CH}_m(\nu_2, \alpha_2, \beta_2, \theta, \text{kind 1})$. The desired result is now obtained by observing that $R_2 = I_m - R_1$. Similarly, the p.d.f. of S_1 is obtained by interchanging subscripts 1 and 2 in the p.d.f. of S. □

Corollary 24. *Let the random matrices X_1 and X_2 be independent, $X_1 \sim \text{CH}_m(\nu_1, \alpha_1, \nu_1 + \nu_2, \theta, \text{kind 1})$ and $X_2 \sim \text{Ga}(m, \nu_2, \theta)$. Then, $(X_1 + X_2)^{-1/2}X_1(X_1 + X_2)^{-1/2} \sim \text{B1}(m, \nu_1, \alpha_1 - \nu_1)$.*

Proof. The desired result is obtained by substituting $\beta_1 = \nu_1 + \nu_2$ in the p.d.f. of R and simplifying the resulting expression by using (16). □

Corollary 25. *Let the random matrices X_1 and X_2 be independent, $X_1 \sim \text{Ga}(m, \nu_1, \theta)$ and $X_2 \sim \text{CH}(m, \nu_2, \alpha_2, \nu_1 + \nu_2, \theta, \text{kind 1})$. Then, $(X_1 + X_2)^{-1/2}X_1(X_1 + X_2)^{-1/2} \sim \text{B1}(m, \alpha_2 - \nu_2, \nu_2)$.*

Proof. The desired result is obtained by substituting $\beta_2 = \nu_1 + \nu_2$ in the p.d.f. of R_1 and simplifying the resulting expression by using (16). □

Corollary 26. *Let the random matrices X_1 and X_2 be independent, $X_1 \sim \text{CH}_m(\nu_1, \alpha_1, \alpha_1 + \nu_1 + \nu_2, \theta, \text{kind 1})$ and $X_2 \sim \text{Ga}(m, \nu_2, \theta)$. Then $X_1 + X_2 \sim \text{CH}_m(\nu_1 + \nu_2, \alpha_1 + \nu_2, \alpha_1 + \nu_1 + \nu_2, \theta, \text{kind 1})$.*

Proof. The result is obtained by substituting $\beta_1 = \alpha_1 + \nu_1 + \nu_2$ in the p.d.f. of S and simplifying the resulting expression by using (15). □

Corollary 27. *Let the random matrices X_1 and X_2 be independent, $X_1 \sim \mathrm{Ga}(m, \nu_1, \theta)$ and $X_2 \sim \mathrm{CH}_m(\nu_2, \alpha_2, \alpha_2 + \nu_1 + \nu_2, \theta, \mathrm{kind}\ 1)$. Then $X_1 + X_2 \sim \mathrm{CH}_m(\nu_1 + \nu_2, \alpha_2 + \nu_1, \alpha_2 + \nu_1 + \nu_2, \theta, \mathrm{kind}\ 1)$.*

Proof. The result is obtained by substituting $\beta_2 = \alpha_2 + \nu_1 + \nu_2$ in the p.d.f. of S and simplifying the resulting expression by using (15). □

Corollary 28. *Let the random matrices X_1 and X_2 be independent, $X_1 \sim \mathrm{Ga}(m, \nu_1, \theta)$ and $X_2 \sim \mathrm{Ga}(m, \nu_2, \theta)$. Then, $(X_1 + X_2)^{-1/2} X_1 (X_1 + X_2)^{-1/2} \sim \mathrm{B1}(m, \nu_1, \nu_2)$ and $X_1 + X_2 \sim \mathrm{Ga}(m, \nu_1 + \nu_2, \theta)$.*

Proof. Substitute $\beta_1 = \alpha_1$ in the p.d.f. of R or $\beta_2 = \alpha_2$ in the p.d.f. of R_1 and simplify the resulting expression to get the desired result. □

5. Related Distributions

This section gives distributional results for the determinant of the random matrix distributed as confluent hypergeometric function kind 1.

In an unpublished report, Coelho et al. [13] have shown that if z is a positive random variable and $E(z^h)$ is defined for h in some neighborhood of zero, then the moments $E(z^h)$ uniquely identify the distribution of z. In the next theorem, we will use this result to derive the distribution of the product of two independent confluent hypergeometric function kind 1 variables.

Theorem 29. *If $x_1 \sim \mathrm{CH}(\nu, \alpha, \beta, \theta, \mathrm{kind}\ 1)$ and $x_2 \sim \mathrm{CH}(\nu + 1/2, \alpha + 1, \beta + 1, \theta, \mathrm{kind}\ 1)$ are independent, then $2\sqrt{x_1 x_2} \sim \mathrm{CH}(2\nu, 2\alpha, 2\beta, \theta, \mathrm{kind}\ 1)$.*

Proof. The hth moment of $2\sqrt{x_1 x_2}$ is derived as

$$
E\left[\left(2\sqrt{x_1 x_2}\right)^h\right]
$$

$$
= \frac{(2\theta)^h \Gamma(\beta - \nu) \Gamma(\nu + h/2) \Gamma(\alpha - \nu - h/2)}{\Gamma(\nu) \Gamma(\alpha - \nu) \Gamma(\beta - \nu - h/2)} \tag{74}
$$

$$
\cdot \frac{\Gamma(\beta - \nu + 1/2) \Gamma(\nu + 1/2 + h/2) \Gamma(\alpha - \nu + 1/2 - h/2)}{\Gamma(\nu + 1/2) \Gamma(\alpha - \nu + 1/2) \Gamma(\beta - \nu + 1/2 - h/2)}.
$$

Now, using the duplication formula for gamma function, namely,

$$
\Gamma(2z) = \frac{\Gamma(z) \Gamma(z + 1/2)}{2^{1-2z} \sqrt{\pi}}, \tag{75}
$$

the hth moment of $2\sqrt{x_1 x_2}$ is rewritten as

$$
E\left[\left(2\sqrt{x_1 x_2}\right)^h\right]
$$

$$
= \frac{\theta^h \Gamma(2\beta - 2\nu) \Gamma(2\nu + h) \Gamma(2\alpha - 2\nu - h)}{\Gamma(2\nu) \Gamma(2\alpha - 2\nu) \Gamma(2\beta - 2\nu - h)}, \tag{76}
$$

where $\nu > 0$, $\alpha - \nu > 0$, $\beta - \nu > 0$, $2\nu > h$, $2(\alpha - \nu) > h$, and $2(\beta - \nu) > h$.

Finally, comparison of the above expression with the one given in (45) yields the desired result. □

Theorem 30. *If $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \Omega, \theta, \mathrm{kind}\ 1)$, then $\det(\Omega^{-1/2} X \Omega^{-1/2})$ is distributed as $\prod_{i=1}^m z_i$, where $z_1, \ldots, z_m$ are independent, $z_i \sim \mathrm{CH}(\nu - (i-1)/2, \alpha - i + 1, \beta - i + 1, \theta, \mathrm{kind}\ 1)$, $i = 1, \ldots, m$.*

Proof. Writing multivariate gamma functions in terms of ordinary gamma function, (42) is rewritten as

$$
E\left[\det\left(\Omega^{-1/2} X \Omega^{-1/2}\right)^h\right]
$$

$$
= \prod_{i=1}^m \left[\frac{\theta^h \Gamma[\beta - \nu - (i-1)/2]}{\Gamma[\nu - (i-1)/2] \Gamma[\alpha - \nu - (i-1)/2]} \right. \tag{77}
$$

$$
\left. \cdot \frac{\Gamma[\nu - (i-1)/2 + h] \Gamma[\alpha - \nu - (i-1)/2 - h]}{\Gamma[\beta - \nu - (i-1)/2 - h]} \right].
$$

Now, comparing the above expression with (45), we get $E[\det(\Omega^{-1/2} X \Omega^{-1/2})^h] = \prod_{i=1}^m E(z_i^h)$. □

Corollary 31. *If $X \sim \mathrm{CH}_2(\nu, \alpha, \beta, \theta, \mathrm{kind}\ 1)$, then*

$$
2\det(X)^{1/2} \sim \mathrm{CH}\left(2\nu - 1, 2\alpha - 2, 2\beta - 2, \theta, \mathrm{kind}\ 1\right). \tag{78}
$$

Proof. For $m = 2$, $\det(X)$ is distributed as $z_1 z_2$, where z_1 and z_2 are independent, $z_1 \sim \mathrm{CH}(\nu, \alpha, \beta, \theta, \mathrm{kind}\ 1)$ and $z_2 \sim \mathrm{CH}(\nu - 1/2, \alpha - 1, \beta - 1, \theta, \mathrm{kind}\ 1)$. From Theorem 29, we have $2\sqrt{z_1 z_2} \sim \mathrm{CH}(2\nu - 1, 2\alpha - 2, 2\beta - 2, \theta, \mathrm{kind}\ 1)$. □

Corollary 32. *If $X \sim \mathrm{Ga}(m, \nu, \theta, \Omega)$, then $\det(\Omega^{-1/2} X \Omega^{-1/2})$ is distributed as $\prod_{i=1}^m z_i$, where $z_1, \ldots, z_m$ are independent, $z_i \sim \mathrm{Ga}(\nu - (i-1)/2, \theta)$, $i = 1, \ldots, m$.*

6. Distribution of Eigenvalues

In this section, we derive density of eigenvalues of random matrix distributed as confluent hypergeometric function kind 1.

Theorem 33. *Let A be a positive definite random matrix of order m with the p.d.f. $f(A)$. Then, the joint p.d.f. of the eigenvalues $l_1, l_2, \ldots, l_m$ of A is given by*

$$
\frac{\pi^{m^2/2}}{\Gamma_m(m/2)} \prod_{i<j}^m (l_i - l_j) \int_{O(m)} f\left(HLH'\right) (dH), \tag{79}
$$

where $l_1 > l_2 > \cdots > l_m > 0$, $L = \mathrm{diag}(l_1, l_2, \ldots, l_m)$, and (dH) is the unit invariant Haar measure on the group of orthogonal matrices $O(m)$.

Proof of Theorem 33 and several other results can be found in Muirhead [7].

Theorem 34. *If* $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \Omega, \mathrm{kind}\ 1)$, *then the joint p.d.f. of the eigenvalues* $x_1, x_2, \ldots, x_m$ *of* X *is given by*

$$
\frac{\pi^{m^2/2} \Gamma_m(\alpha) \Gamma_m(\beta - \nu) \det(\theta\Omega)^{-\nu}}{\Gamma_m(m/2) \Gamma_m(\nu) \Gamma_m(\beta) \Gamma_m(\alpha - \nu)} \left[\prod_{i<j}^m (x_i - x_j) \right]
$$

$$
\cdot \prod_{i=1}^m \left[x_i^{\nu - (m+1)/2} \right] {}_1F_1^{(m)}\left(\alpha; \beta; -\frac{1}{\theta} \Omega^{-1}, L \right), \tag{80}
$$

where $0 < x_m < \cdots < x_1 < \infty$, $L = \mathrm{diag}(x_1, \ldots, x_m)$, *and* ${}_1F_1^{(m)}$ *is the two-matrix argument confluent hypergeometric function.*

Proof. The p.d.f. of X is given by (5). Applying Theorem 33, we obtain the joint p.d.f. of the eigenvalues $x_1, \ldots, x_m$ of X as

$$
\frac{\pi^{m^2/2} \Gamma_m(\alpha) \Gamma_m(\beta - \nu) \det(\theta\Omega)^{-\nu}}{\Gamma_m(m/2) \Gamma_m(\nu) \Gamma_m(\beta) \Gamma_m(\alpha - \nu)} \left[\prod_{i<j}^m (x_i - x_j) \right]
$$

$$
\cdot \prod_{i=1}^m \left[x_i^{\nu - (m+1)/2} \right] \tag{81}
$$

$$
\cdot \int_{O(m)} {}_1F_1^{(m)}\left(\alpha; \beta; -\frac{1}{\theta} \Omega^{-1} H L H' \right) (\mathrm{d}H).
$$

Now, using (28), we obtain the desired result. $\qquad\square$

7. A Generalized Form

In this section, we give a more general form of the matrix variate confluent hypergeometric function kind 1 distribution by introducing an additional factor $\mathrm{etr}(-\Psi^{-1}X/\theta)$ in the p.d.f. (5). The p.d.f. of X, in this case, is given by

$$
\frac{\theta^{-m\nu} \det\left(\Psi^{-1} + \Omega^{-1}\right)^{\nu}}{\Gamma_m(\nu)\ {}_2F_1\left(\nu, \beta - \alpha; \beta; (\Psi^{-1} + \Omega^{-1})^{-1} \Omega^{-1}\right)} \mathrm{etr}\left(-\frac{1}{\theta}\right)
$$

$$
\cdot \det(X)^{\nu - (m+1)/2}\ {}_1F_1\left(\alpha; \beta; -\frac{1}{\theta} \Omega^{-1} X \right), \tag{82}
$$

where $X > 0$. We will write $X \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \Omega, \Psi, \mathrm{kind}\ 1)$ if the density of X is given by (82). For $\Psi = I_m$ and $\theta = 1$, the above p.d.f. reduces to

$$
\frac{\det\left(I_m + \Omega^{-1}\right)^{\nu}}{\Gamma_m(\nu)\ {}_2F_1\left(\nu, \beta - \alpha; \beta; (I_m + \Omega)^{-1}\right)} \mathrm{etr}(-X)
$$

$$
\cdot \det(X)^{\nu - (m+1)/2}\ {}_1F_1\left(\alpha; \beta; -\Omega^{-1} X \right), \quad X > 0, \tag{83}
$$

which is a special case of the generalized hypergeometric function density defined by Roux [14].

Theorem 35. *Let* $Z \mid \Sigma \sim \mathrm{InvGa}(m, \mu, \theta, \Sigma^{-1})$. *Further, let the prior distribution of* Σ *be a generalized matrix variate confluent hypergeometric kind 1 distribution with parameters* $\nu, \alpha, \beta, \theta, \Omega, \text{and } \Psi$, $\Sigma \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \Omega, \Psi, \mathrm{kind}\ 1)$. *Then,*

the marginal distribution of Z *is a generalized inverted matrix variate beta with the density*

$$
m(Z)
$$

$$
= \frac{\Gamma_m(\mu + \nu) \det\left(\Psi^{-1} + \Omega^{-1}\right)^{\nu}}{\Gamma_m(\nu) \Gamma_m(\mu)\ {}_2F_1\left(\nu, \beta - \alpha; \beta; (\Psi^{-1} + \Omega^{-1})^{-1} \Omega^{-1}\right)}
$$

$$
\cdot \frac{\det(Z)^{-\mu - (m+1)/2}}{\det\left(\Psi^{-1} + \Omega^{-1} + Z^{-1}\right)^{\mu + \nu}} \tag{84}
$$

$$
\cdot {}_2F_1\left(\mu + \nu, \beta - \alpha; \beta; \left(\Psi^{-1} + \Omega^{-1} + Z^{-1}\right)^{-1} \Omega^{-1} \right),
$$

where $Z > 0$.

Proof. By definition, the marginal density of Z, denoted by $m(Z)$, is obtained as

$$
m(Z) = \int_{\Sigma > 0} f(Z \mid \Sigma)\, \pi(\Sigma)\, \mathrm{d}\Sigma. \tag{85}
$$

Now, substituting for $f(Z \mid \Sigma)$ and $\pi(\Sigma)$, we get

$$
m(Z)
$$

$$
= \frac{\theta^{-m(\nu + \mu)} \det\left(\Psi^{-1} + \Omega^{-1}\right)^{\nu}}{\Gamma_m(\nu) \Gamma_m(\mu)\ {}_2F_1\left(\nu, \beta - \alpha; \beta; (\Psi^{-1} + \Omega^{-1})^{-1} \Omega^{-1}\right)}
$$

$$
\cdot \det(Z)^{-\mu - (m+1)/2} \int_{\Sigma > 0} \det(\Sigma)^{\mu + \nu - (m+1)/2} \tag{86}
$$

$$
\cdot \mathrm{etr}\left[-\frac{1}{\theta}\left(\Psi^{-1} + \Omega^{-1} + Z^{-1}\right)\Sigma \right]
$$

$$
\cdot {}_1F_1\left(\beta - \alpha; \beta; \frac{1}{\theta}\Omega^{-1}\Sigma \right) \mathrm{d}\Sigma.
$$

Finally, evaluating the above expression by using (21) and simplifying, we get the desired result. $\qquad\square$

Theorem 36. *Let* $Z \mid \Sigma \sim \mathrm{InvGa}(m, \mu, \Sigma^{-1})$. *Further, let the prior distribution of* Σ *be a generalized matrix variate confluent hypergeometric function kind 1 distribution parameters* $\nu, \alpha, \beta, \theta, \Omega, \text{and } \Psi$, $\Sigma \sim \mathrm{CH}_m(\nu, \alpha, \beta, \theta, \Omega, \Psi, \mathrm{kind}\ 1)$. *Then, the posterior distribution of* Σ *is a generalized matrix variate confluent hypergeometric kind 1 distribution with parameters* $\mu + \nu, \alpha, \beta, \theta, \Omega, \text{and } (\Psi^{-1} + Z^{-1})^{-1}$.

Proof. By definition and Theorem 35, we have

$$
\pi(\Sigma \mid Z) = \frac{f(Z \mid \Sigma)\, \pi(\Sigma)}{m(Z)}. \tag{87}
$$

Now, substituting appropriately, we get

$$
\pi(\Sigma \mid Z)
$$

$$
= \frac{\theta^{-m(\nu + \mu)} \det\left(\Psi^{-1} + \Omega^{-1} + Z^{-1}\right)^{\nu + \mu}}{\Gamma_m(\nu + \mu)\ {}_2F_1\left(\nu + \mu, \beta - \alpha; \beta; (\Psi^{-1} + \Omega^{-1} + Z^{-1})^{-1} \Omega^{-1}\right)}
$$

$$
\cdot \mathrm{etr}\left[-\frac{1}{\theta}\left(\Psi^{-1} + Z^{-1}\right)\Sigma \right] \det(\Sigma)^{\nu + \mu - (m+1)/2} \tag{88}
$$

$$
\cdot {}_1F_1\left(\alpha; \beta; -\frac{1}{\theta}\Omega^{-1}\Sigma \right), \quad \Sigma > 0,
$$

which is the desired result. $\qquad\square$

From the above results it is quite clear that the generalized matrix variate confluent hypergeometric function kind 1 distribution as the prior distribution is conjugate. Thus, this distribution may be used as an alternative to matrix variate gamma distribution.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, Chapman & Hall/CRC Press, Boca Raton, Fla, USA, 2000.

[2] G. J. van der Merwe and J. J. J. Roux, "On a generalized matrix-variate hypergeometric distribution," *South African Statistical Journal*, vol. 8, pp. 49–58, 1974.

[3] J. M. Orozco-Castañeda, D. K. Nagar, and A. K. Gupta, "Generalized bivariate beta distributions involving Appell's hypergeometric function of the second kind," *Computers and Mathematics with Applications*, vol. 64, no. 8, pp. 2507–2519, 2012.

[4] Y. L. Luke, *The Special Functions and Their Approximations*, vol. 1, Academic Press, New York, NY, USA, 1969.

[5] A. G. Constantine, "Some non-central distribution problems in multivariate analysis," *Annals of Mathematical Statistics*, vol. 34, pp. 1270–1285, 1963.

[6] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Annals of Mathematical Statistics*, vol. 35, pp. 475–501, 1964.

[7] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, NY, USA, 1982.

[8] H. Hashiguchi, Y. Numata, N. Takayama, and A. Takemura, "The holonomic gradient method for the distribution function of the largest root of a Wishart matrix," *Journal of Multivariate Analysis*, vol. 117, pp. 296–312, 2013.

[9] P. Koev and A. Edelman, "The efficient evaluation of the hypergeometric function of a matrix argument," *Mathematics of Computation*, vol. 75, no. 254, pp. 833–846, 2006.

[10] A. Iranmanesh, M. Arashi, D. K. Nagar, and S. M. Tabatabaey, "On inverted matrix variate gamma distribution," *Communications in Statistics—Theory and Methods*, vol. 42, no. 1, pp. 28–41, 2013.

[11] J. J. Roux and G. J. van der Merwe, "Families of multivariate distributions having properties usually associated with the Wishart distribution," *South African Statistical Journal*, vol. 8, pp. 111–117, 1974.

[12] C. G. Khatrzi, "On certain distribution problems based on positive definite quadratic functions in normal vectors," *Annals of Mathematical Statistics*, vol. 37, pp. 468–479, 1966.

[13] C. A. Coelho, R. P. Alberto, and L. M. Grilo, "When do the moments uniquely identify a distribution," CMA 13-2005, Centro de Matemática e Aplicações, Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2005.

[14] J. J. J. Roux, "On generalized multivariate distributions. South African Statist," *South African Statistical Journal*, vol. 5, pp. 91–100, 1971.

Extended Odd Fréchet-G Family of Distributions

Suleman Nasiru ⓘD

Department of Statistics, Faculty of Mathematical Sciences, University for Development Studies, Tamale, Ghana

Correspondence should be addressed to Suleman Nasiru; sulemanstat@gmail.com

The need to develop generalizations of existing statistical distributions to make them more flexible in modeling real data sets is vital in parametric statistical modeling and inference. Thus, this study develops a new class of distributions called the extended odd Fréchet family of distributions for modifying existing standard distributions. Two special models named the extended odd Fréchet Nadarajah-Haghighi and extended odd Fréchet Weibull distributions are proposed using the developed family. The densities and the hazard rate functions of the two special distributions exhibit different kinds of monotonic and nonmonotonic shapes. The maximum likelihood method is used to develop estimators for the parameters of the new class of distributions. The application of the special distributions is illustrated by means of a real data set. The results revealed that the special distributions developed from the new family can provide reasonable parametric fit to the given data set compared to other existing distributions.

1. Introduction

The fundamental reason for parametric statistical modeling is to identify the most appropriate model that adequately describes a data set obtained from experiment, observational studies, surveys, and so on. Most of these modeling techniques are based on finding the most suitable probability distribution that explains the underlying structure of the given data set. However, there is no single probability distribution that is suitable for different data sets. Thus, this has triggered the need to extend the existing classical distributions or develop new ones. Barrage of methods for defining new families of distributions have been proposed in literature for extending or generalizing the existing classical distributions in recent time. Some of these methods include Weibull-G [1], odd generalized exponential family [2], odd Lindley-G family [3], Topp-Leone odd log-logistic-G family [4], odd Burr-G family [5], odd Fréchet-G family [6], odd gamma-G family [7], transformed-transformer method [8], exponentiated transformed-transformer method [9], exponentiated generalized transformed-transformer method [10], alpha power transformed family [11], alpha logarithmic transformed family [12], Kumaraswamy-G family [13], beta-G family [14], Kumaraswamy transmuted-G family [15], transmuted geometric-G family [16], and beta extended Weibull family [17]. These methods are developed with the motivation

of defining new models with different kinds of failure rates (monotonic and nonmonotonic), constructing heavy-tailed distributions for modeling different kinds of data sets, developing distributions with symmetric, right skewed, left skewed, reversed J shape, and consistently providing a reasonable parametric fit to given data sets.

Recently, [6] developed the odd Fréchet family of distributions and defined its cumulative distribution function (CDF) as

$$H(x) = e^{-[(1-G(x;\psi))/G(x;\psi)]^\theta}, \quad x \in \mathbb{R}, \quad (1)$$

where $G(x;\psi)$ is the baseline CDF and ψ is a $p \times 1$ vector of associated parameters. Using the transformed-transformer method proposed by [8], an extension of the odd Fréchet family of distributions called the extended odd Fréchet-G (EOF-G) family of distributions is developed by integrating the Fréchet probability density function (PDF). Hence, the CDF of the EOF-G family is defined as

$$F(x) = \int_0^{G(x;\psi)^\alpha/(1-G(x;\psi)^\alpha)} \theta x^{-\theta-1} e^{-x^{-\theta}} dx$$

$$= e^{-[(1-G(x;\psi)^\alpha)/G(x;\psi)^\alpha]^\theta}, \quad \alpha > 0, \ \theta > 0, \ x \in \mathbb{R}, \quad (2)$$

where α and θ are extra shape parameters. The corresponding PDF of the new family is obtained by differentiating equation (2) and is given by

$f(x)$

$$= \frac{\alpha\theta g(x;\psi)\left(1 - G(x;\psi)^{\alpha}\right)^{\theta-1}}{G(x;\psi)^{\alpha\theta+1}} e^{-[(1-G(x;\psi)^{\alpha})/G(x;\psi)^{\alpha}]^{\theta}}, \quad (3)$$

$$\alpha > 0, \ \theta > 0, \ x \in \mathbb{R}.$$

The associated hazard rate function of the EOF-G family is defined as

$$h(x) = \frac{\alpha\theta g(x;\psi)\left(1 - G(x;\psi)^{\alpha}\right)^{\theta-1}}{G(x;\psi)^{\alpha\theta+1}\left(1 - e^{-[(1-G(x;\psi)^{\alpha})/G(x;\psi)^{\alpha}]^{\theta}}\right)}$$
$$\cdot e^{-[(1-G(x;\psi)^{\alpha})/G(x;\psi)^{\alpha}]^{\theta}}, \quad (4)$$

$$\alpha > 0, \ \theta > 0, \ x \in \mathbb{R}.$$

Hereafter, a random variable X following the EOF-G distribution is denoted by $X \sim \text{EOF–G}(x;\alpha,\theta,\psi)$ and for the purpose of simplicity, $G(x;\psi)$ can be written as $G(x)$. The CDF of the EOF-G family of distributions is tractable which makes it easy to generate random numbers provided that the CDF of the baseline distribution is also tractable. The uth quantile of the EOF-G family is given by

$$x_u = G^{-1}\left[\left(\frac{1}{1 + (-\log(u))^{1/\theta}}\right)^{1/\alpha}\right], \quad u \in [0,1], \quad (5)$$

where $G^{-1}(u)$ is the baseline quantile function. When $\alpha = 1$, the EOF-G family of distributions reduces to the odd Fréchet family of distributions. Adopting the interpretation of the CDF of the odd Weibull family as given in [18], the physical interpretation of the CDF of the EOF-G family is given as follows: Suppose Y is a lifetime random variable with continuous CDF, $G(x;\psi)^{\alpha}$. The odds ratio that an individual (component) having the lifetime Y will die (fail) at time x is $G(x;\psi)^{\alpha}/1 - G(x;\psi)^{\alpha}$. Given that the variability of these odds of death is denoted by the random variable X and that it follows the Fréchet distribution, then

$$\mathbb{P}(Y \le x) = \mathbb{P}\left(X \le \frac{G(x;\psi)^{\alpha}}{1 - G(x;\psi)^{\alpha}}\right) = F(x), \quad (6)$$

which is given in (2). The rest of the paper is organized as follows: In Section 2, special distributions of the EOF-G family are discussed. In Section 3, the mixture representation of the PDF and CDF of the EOF-G family is given. The statistical properties of the new family are derived in Section 4. In Section 5, the estimators for the parameters of the family are developed using the technique of maximum likelihood estimation. Monte Carlo simulations are performed in Section 6 to assess the performance of the estimators. In Section 7, the application of the special distributions is demonstrated using real data set. Finally, the concluding remarks of the study are given in Section 8.

2. Special Distributions of the EOF-G Family

In this section, two special distributions of the EOF-G family are discussed.

2.1. EOF-Nadarajah-Haghighi (EOFNH) Distribution.
Suppose the baseline CDF is that of the Nadarajah-Haghighi distribution; that is, $G(x;\beta,\lambda) = 1 - e^{(1-(1+\lambda x)^{\beta})}$ with corresponding PDF $g(x;\beta,\lambda) = \beta\lambda(1 + \lambda x)^{\beta-1}e^{(1-(1+\lambda x)^{\beta})}$ and positive parameters $\beta,\lambda > 0$. The PDF of the EOFNH distribution is given by

$f(x)$

$$= \frac{\alpha\beta\lambda\theta(1 + \lambda x)^{\beta-1}e^{(1-(1+\lambda x)^{\beta})}\left[1 - \left(1 - e^{(1-(1+\lambda x)^{\beta})}\right)^{\alpha}\right]^{\theta-1}}{\left(1 - e^{(1-(1+\lambda x)^{\beta})}\right)^{\alpha\theta+1}} \quad (7)$$
$$\cdot e^{-[(1-e^{(1-(1+\lambda x)^{\beta})})^{-\alpha}-1]^{\theta}},$$

where $\alpha,\beta,\theta > 0$ are shape parameters, $\lambda > 0$ is a scale parameter, and $x > 0$. Figure 1 shows the plots of the PDF of the EOFNH distribution for some selected parameter values. The density function exhibits different kinds of shapes.

The corresponding hazard rate function is given by

$h(x)$

$$= \frac{\alpha\beta\lambda\theta(1 + \lambda x)^{\beta-1}e^{(1-(1+\lambda x)^{\beta})}\left[1 - \left(1 - e^{(1-(1+\lambda x)^{\beta})}\right)^{\alpha}\right]^{\theta-1}}{\left(1 - e^{(1-(1+\lambda x)^{\beta})}\right)^{\alpha\theta+1}\left(1 - e^{-[(1-e^{(1-(1+\lambda x)^{\beta})})^{-\alpha}-1]^{\theta}}\right)} \quad (8)$$
$$\cdot e^{-[(1-e^{(1-(1+\lambda x)^{\beta})})^{-\alpha}-1]^{\theta}}, \quad x > 0.$$

The plots of the hazard rate function of the EOFNH distribution for some selected parameter values are shown in Figure 2. The hazard rate function can assume decreasing, bathtub, upside down bathtub, and other nonmonotonic failure rate forms.

The quantile function of the EOFNH distribution is given by

$$x_u = \frac{\left[1 - \log\left(1 - \left(1 + (-\log(u))^{1/\theta}\right)^{-1/\alpha}\right)\right]^{1/\beta} - 1}{\lambda}, \quad (9)$$

$$u \in [0,1].$$

Equation (9) can be used to generate random numbers from the EOFNH distribution. The first quartile, median, and upper quartile of the distribution are obtained by substituting $u = 0.25, 0.5$, and 0.75, respectively, into (9).

2.2. EOF-Weibull (EOFW) Distribution.
Consider the Weibull distribution with shape parameter $\beta > 0$ and scale parameter $\lambda > 0$, where the CDF and PDF for $x > 0$ are given by $G(x;\beta,\lambda) = 1 - e^{-\lambda x^{\beta}}$ and $g(x;\beta,\lambda) = \beta\lambda x^{\beta-1}e^{-\lambda x^{\beta}}$. Substituting the PDF and CDF of the Weibull distribution in (3), the PDF of the EOFW distribution is defined as

$$f(x) = \frac{\alpha\beta\lambda\theta x^{\beta-1}e^{-\lambda x^{\beta}}\left[1 - \left(1 - e^{-\lambda x^{\beta}}\right)^{\alpha}\right]^{\theta-1}}{\left(1 - e^{-\lambda x^{\beta}}\right)^{\alpha\theta+1}} \quad (10)$$
$$\cdot e^{-[(1-e^{-\lambda x^{\beta}})^{-\alpha}-1]^{\theta}},$$

where $\alpha,\beta,\theta > 0$ are shape parameters, $\lambda > 0$ is scale parameter, and $x > 0$. Figure 3 displays some of the possible shapes of
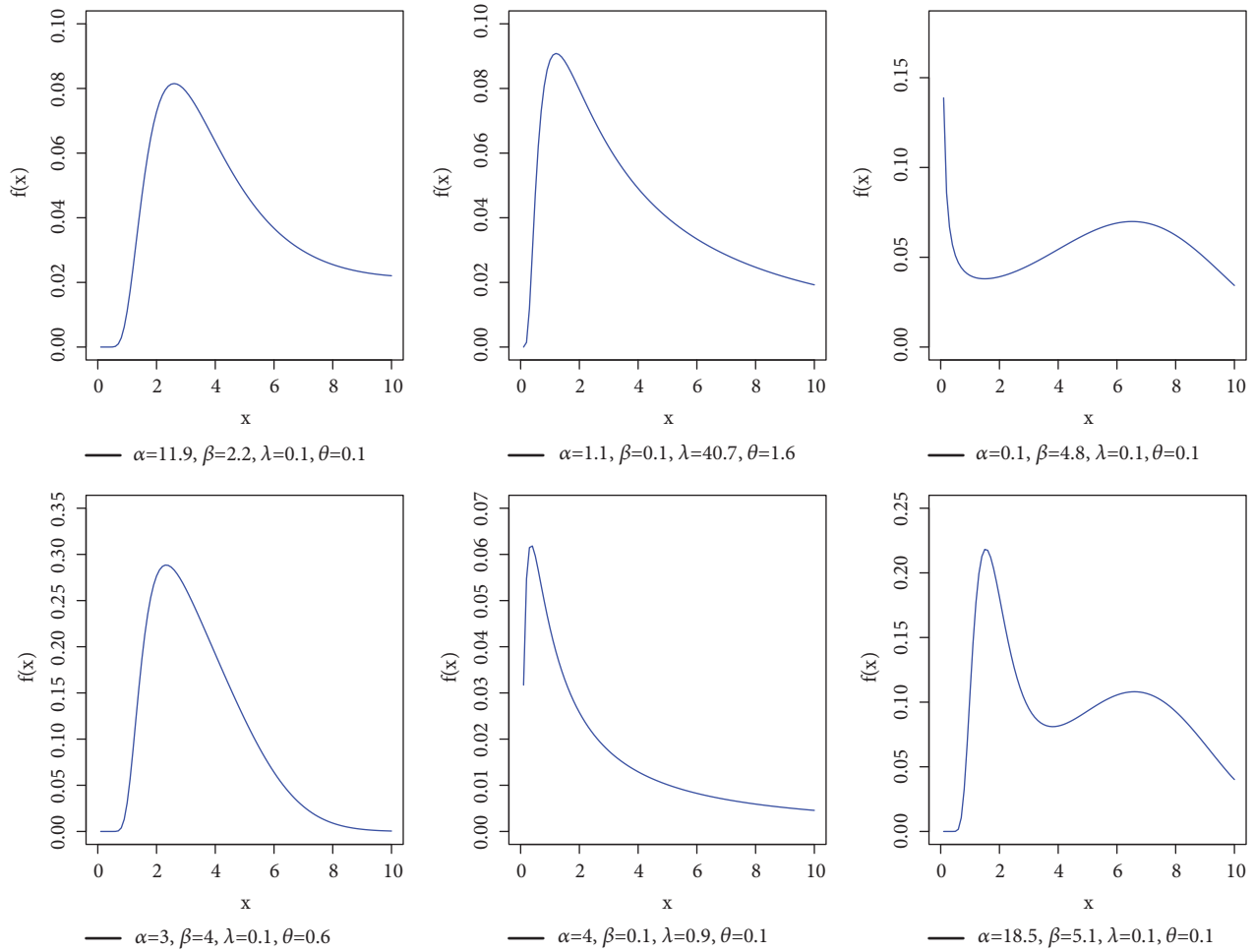
FIGURE 1: Plots of the EOFNH distribution density function.

the density function of the EOFW distribution. The density exhibits unimodal and reversed J-shape among others.

The hazard rate function of the EOFW distribution is given by

$$h(x) = \frac{\alpha\beta\lambda\theta x^{\beta-1} e^{-\lambda x^\beta} \left[1 - \left(1 - e^{-\lambda x^\beta} \right)^\alpha \right]^{\theta-1}}{\left(1 - e^{-\lambda x^\beta} \right)^{\alpha\theta+1} \left(1 - e^{-\left[(1 - e^{-\lambda x^\beta})^{-\alpha} - 1 \right]^\theta} \right)}$$

$$\cdot e^{-\left[(1 - e^{-\lambda x^\beta})^{-\alpha} - 1 \right]^\theta}, \quad x > 0.$$

The hazard rate function can assume decreasing, bathtub, and upside down bathtub forms for some selected parameter values as shown in Figure 4.

The quantile function of the EOFW distribution is defined as

$$x_u = \left\{ \frac{-\log\left[1 - \left(1 + (-\log(u))^{1/\theta} \right)^{-1/\alpha} \right]}{\lambda} \right\}^{1/\beta},$$

(12)

$$u \in [0, 1].$$

The generation of random numbers from the EOFW distribution can easily be done using (12).

3. Mixture Representation

In this section, the mixture representation of the PDF and CDF of the EOF-G family of distributions is discussed. The mixture representation is useful when deriving the statistical properties of this new family of distributions. Using the Taylor series expansion, the PDF can be written as

$$f(x) = \alpha\theta \sum_{i=0}^{\infty} \frac{(-1)^i g(x; \psi) \left[1 - G(x; \psi)^\alpha \right]^{\theta(i+1)-1}}{i! G(x; \psi)^{\alpha\theta(i+1)+1}}.$$

(13)

Equation (13) can be written as
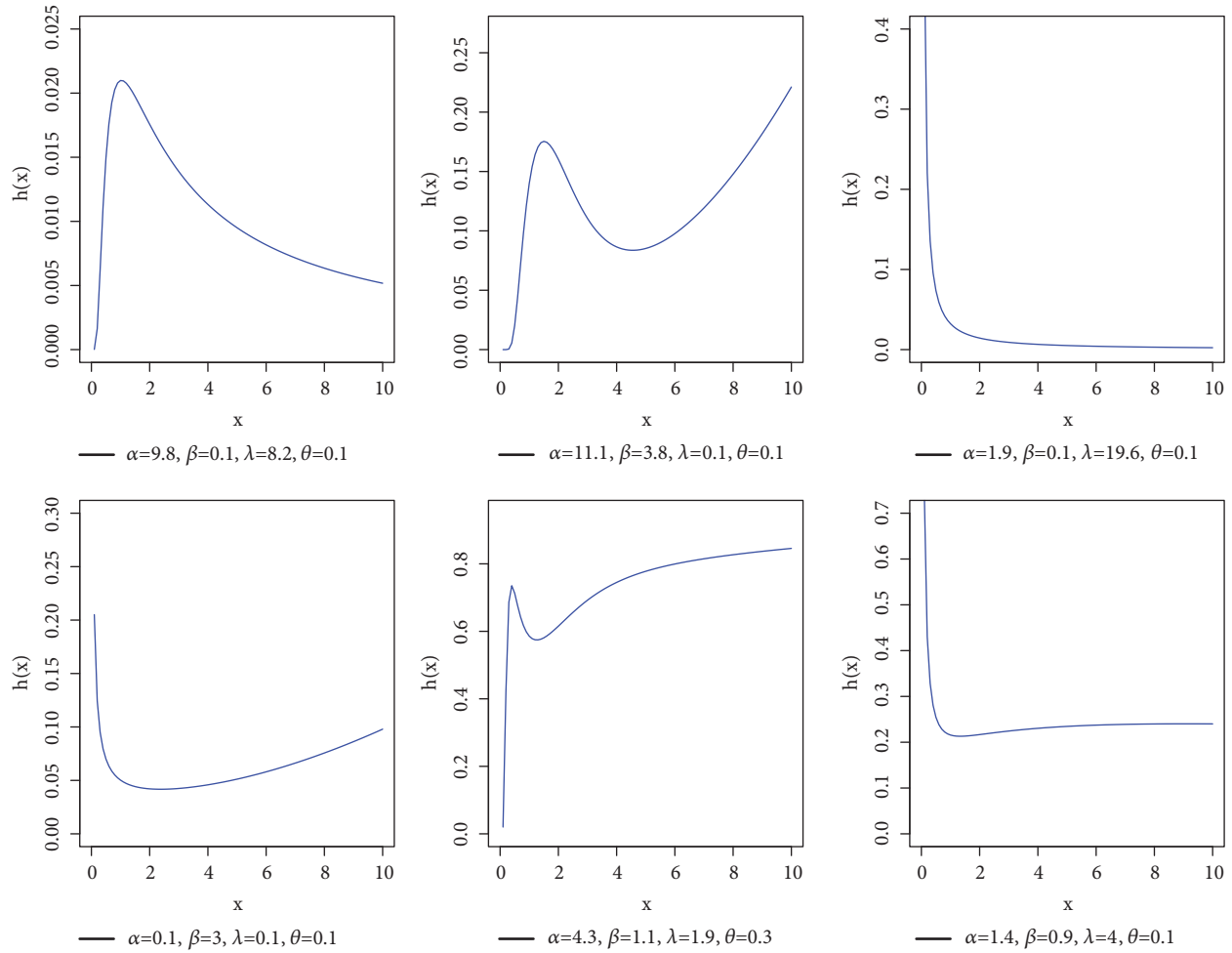
FIGURE 2: Plots of the EOFNH distribution hazard rate function.

$$f(x) = \alpha\theta \sum_{i=0}^{\infty} \frac{(-1)^i \, g(x;\psi) \left[1 - G(x;\psi)^{\alpha}\right]^{\theta(i+1)-1} \left[1 - (1 - G(x;\psi))\right]^{-[\alpha\theta(i+1)+1]}}{i!}. \tag{14}$$

Applying the generalized binomial series expansion yields

$$f(x) = \alpha\theta \sum_{i,j=0}^{\infty} \frac{(-1)^i}{i!} \binom{\alpha\theta(i+1)+j}{j} g(x;\psi) \cdot [1 - G(x;\psi)]^j \left[1 - G(x;\psi)^{\alpha}\right]^{\theta(i+1)-1}. \tag{15}$$

Now using the binomial series expansion, $(1 - z)^{\eta-1} = \sum_{j=0}^{\infty}(-1)^j \binom{\eta-1}{j} z^j$, $|z| < 1$, thrice yields

$$f(x) = \alpha\theta \sum_{i,j=0}^{\infty} \sum_{k,m=0}^{\infty} \sum_{q=0}^{m+j} \omega_{ijkmq} g(x;\psi) G(x;\psi)^q, \tag{16}$$

where

$$\omega_{ijkmq} = \frac{(-1)^{i+k+m+q}}{i!}$$
$$\cdot \binom{\alpha\theta(i+1)+j}{j} \binom{\theta(i+1)-1}{k} \binom{\alpha k}{m} \binom{m+j}{q}. \tag{17}$$

Alternatively (16) can be written in terms of the exponentiated-G (exp-G) density function as

$$f(x) = \alpha\theta \sum_{i,j=0}^{\infty} \sum_{k,m=0}^{\infty} \sum_{q=0}^{m+j} \omega_{ijkmq}^{*} \pi_{q+1}(x), \tag{18}$$

where $\omega_{ijkmq}^{*} = \omega_{ijkmq}/(q + 1)$ and $\pi_{q+1}(x) = (q + 1)g(x;\psi)G(x;\psi)^q$ is the exp-G density function with power
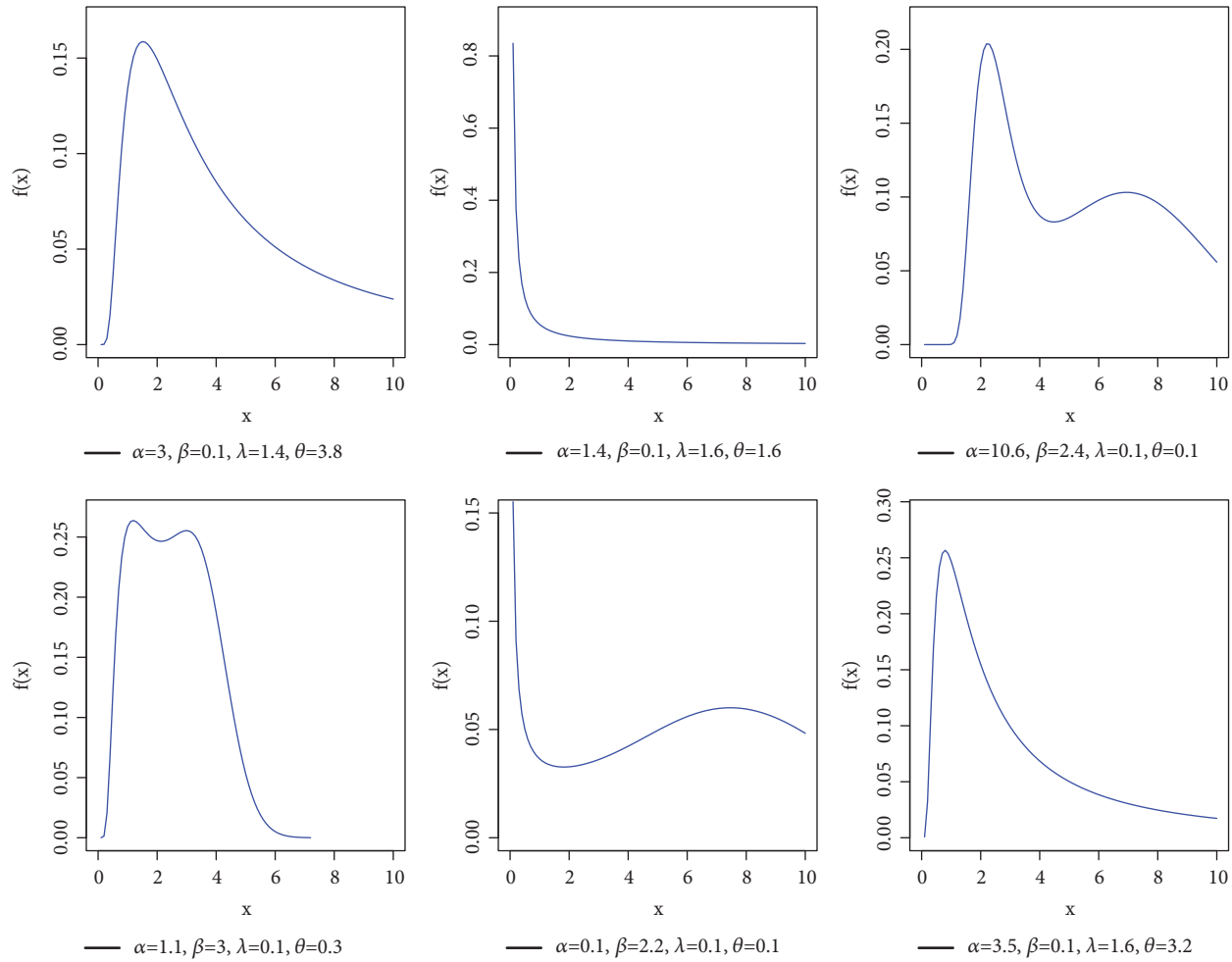
FIGURE 3: Plots of the EOFW distribution density function.

parameter $q + 1$. By integrating (18), the mixture representation of the CDF is given by

$$F(x) = \alpha\theta \sum_{i,j=0}^{\infty} \sum_{k,m=0}^{\infty} \sum_{q=0}^{m+j} \omega_{ijkmq}^{*} \Pi_{q+1}(x), \quad (19)$$

where $\Pi_{q+1}(x) = G(x; \boldsymbol{\psi})^{q+1}$ is the CDF of the exp-G family with power parameter $q + 1$.

4. Statistical Properties

In this section, the moments, incomplete moments, generating function, entropies, and order statistics of the EOF-G family are derived.

4.1. Moments. The rth noncentral moment of a random variable X is given by $E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx$. Hence, using this definition the rth noncentral moment of the EOF-G random variable is given by

$$E(X^r) = \alpha\theta \sum_{i,j=0}^{\infty} \sum_{k,m=0}^{\infty} \sum_{q=0}^{m+j} \omega_{ijkmq} \tau_{r,q}, \quad (20)$$

where $\tau_{r,q} = \int_{-\infty}^{\infty} x^r g(x; \boldsymbol{\psi}) G(x; \boldsymbol{\psi})^q dx$ is the probability weighted moment of the baseline distribution. The rth noncentral moment can also be expressed in terms of the quantile of the baseline distribution. Letting $G(x; (\boldsymbol{\psi})) = u$, the rth noncentral moment in terms of the quantile is given by

$$E(X^r) = \alpha\theta \sum_{i,j=0}^{\infty} \sum_{k,m=0}^{\infty} \sum_{q=0}^{m+j} \omega_{ijkmq} \int_{0}^{1} Q_G(u)^r u^q du, \quad (21)$$

where $Q_G(u)$ is the quantile function of the baseline distribution.

4.2. Incomplete Moments. The rth incomplete moment of a random variable X is defined as $m_r(y) = \int_{-\infty}^{y} x^r f(x) dx$. Thus, the rth incomplete moment of the EOF-G random variable is given by

$$m_r(y)$$

$$= \alpha\theta \sum_{i,j=0}^{\infty} \sum_{k,m=0}^{\infty} \sum_{q=0}^{m+j} \omega_{ijkmq} \int_{0}^{y} x^r g(x; \boldsymbol{\psi}) G(x; \boldsymbol{\psi})^q dx \quad (22)$$
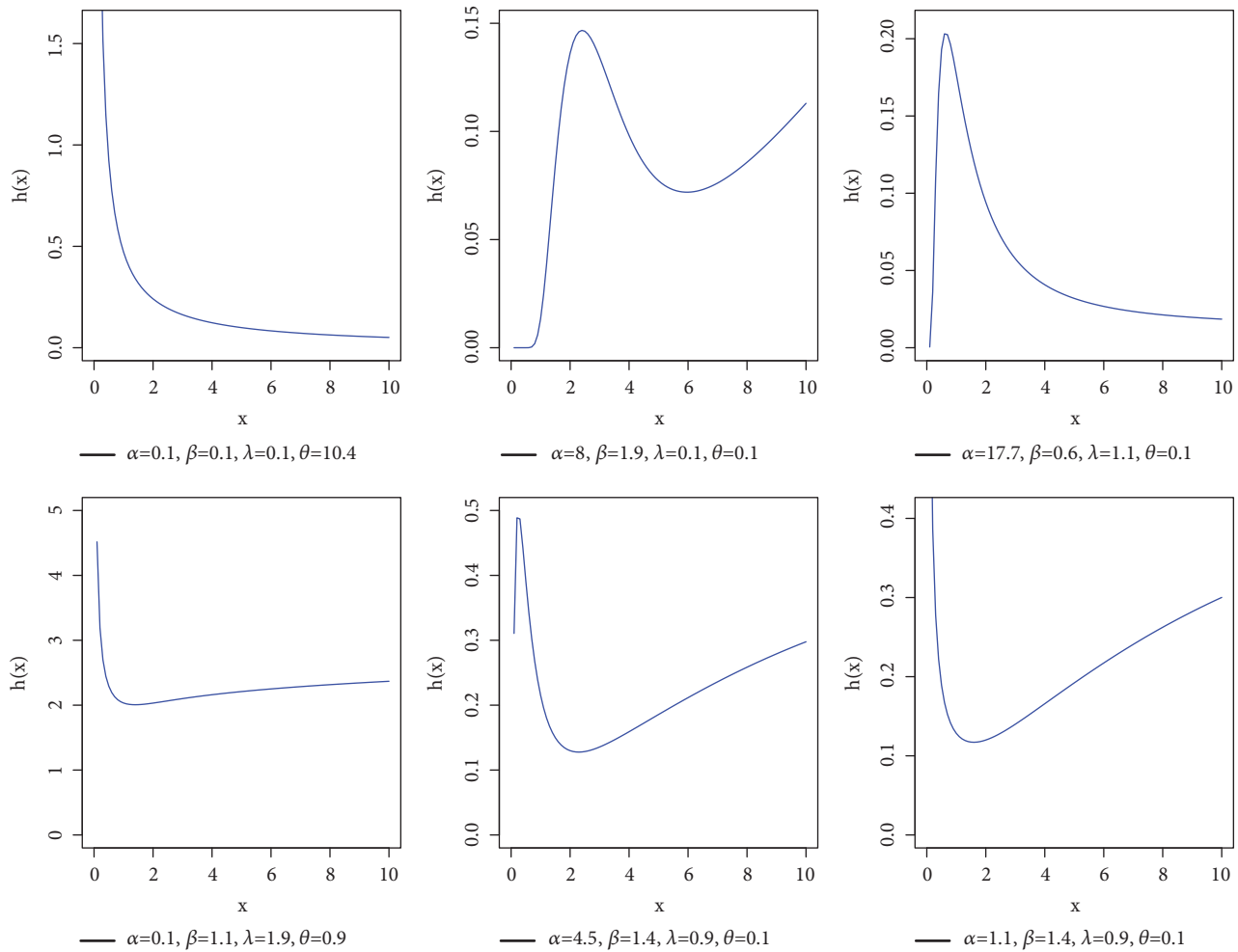
FIGURE 4: Plots of the EOFW distribution hazard rate function.

In terms of the quantile function of the baseline distribution, the rth incomplete moment is given by

$$m_r(y) = \alpha\theta \sum_{i,j=0}^{\infty} \sum_{k,m=0}^{\infty} \sum_{q=0}^{m+j} \omega_{ijkmq} \int_0^{G(y)} Q_G(u)^r u^q du. \quad (23)$$

Utilize the power series expansion of the quantile of the baseline; that is,

$$Q_G(u) = \sum_{h=0}^{\infty} e_h u^h, \quad (24)$$

where $e_h (h = 0, 1, \ldots)$ are suitably chosen real numbers that depend on the parameters of the $G(x; \psi)$ distribution. Furthermore, for positive integer r $(r \geq 1)$,

$$Q_G(u)^r = \left(\sum_{h=0}^{\infty} e_h u^h\right)^r = \sum_{h=0}^{\infty} e'_{r,h} u^h, \quad (25)$$

where $e'_{r,h} = (he_0)^{-1} \sum_{z=1}^{h} [z(r+1) - h] e_z e'_{r,h-z}$ and $e'_{r,0} = (e_0)^h$. For more details on quantile power series expansion, see [19]. Hence,

$$m_r(y) = \alpha\theta \sum_{i,j=0}^{\infty} \sum_{k,m=0}^{\infty} \sum_{q=0}^{m+j} \omega_{ijkmq} \int_0^{G(y)} \sum_{h=0}^{\infty} e'_{r,h} u^{h+q} du$$

$$= \alpha\theta \sum_{i,j=0}^{\infty} \sum_{k,m=0}^{\infty} \sum_{q=0}^{m+j} \omega_{ijkmq} e'_{r,h} \frac{G(y)^{h+q+1}}{h+q+1}. \quad (26)$$

The incomplete moments are used in the computation of other useful statistical measures such as the mean deviations about the mean $(\delta_1 = E(|X - \mu'_1|))$ and about the median $(\delta_2 = E(|X - M|))$. The mean deviation about the mean and about the median can further be expressed as

$$\delta_1 = 2\mu'_1 F(\mu'_1) - 2m_1(\mu'_1),$$

$$\delta_2 = \mu'_1 - 2m_1(M), \quad (27)$$

where $\mu'_1 = \mu$ is the mean obtained by putting $r = 1$ into (20), M is the median obtained by substituting $u = 0.5$ into (5), and $m_1(y) = \int_{-\infty}^{y} xf(x)dx$ is the first incomplete moment which can be obtained from (23) by substituting $r = 1$.

4.3. Generating Function.

In this subsection, two formulae for the computation of the moment generating function $M_X(t) = E(e^{tX})$ are given. Using the Taylor series expansion, $M_X(t) = E(e^{tX}) = \sum_{r=0}^{\infty} (t^r/r!)E(X^r)$. Thus, the moment generating function is given by

$$M_X(t) = \sum_{i,j=0}^{\infty} \sum_{k,m=0}^{\infty} \sum_{r=0}^{\infty} \sum_{q=0}^{m+j} \omega_{ijkmq} \tau_{r,q}. \tag{28}$$

Alternatively, the moment generating function can be expressed in terms of the quantile function of the baseline distribution as

$$M_X(t) = \sum_{i,j=0}^{\infty} \sum_{k,m=0}^{\infty} \sum_{q=0}^{m+j} \omega_{ijkmq} \int_0^1 e^{tQ_G(u)} u^q du. \tag{29}$$

4.4. Entropy Measures.

Entropies are measures of uncertainty or variation of a random variable. In this subsection, the Rényi, Shannon, and δ entropies are studied. The Rényi entropy [20] of a random variable X with PDF $f(x)$ is defined as

$$I_R(\delta) = \frac{1}{1-\delta} \log \left[\int_{-\infty}^{\infty} f(x)^\delta dx \right], \quad \delta > 0, \ \delta \neq 1. \tag{30}$$

Using similar concepts for expanding the PDF,

$$f(x)^\delta = (\alpha\theta)^\delta \sum_{i,j=0}^{\infty} \sum_{k,m=0}^{\infty} \sum_{q=0}^{m+j} \varpi_{ijkmq} g(x;\psi)^\delta G(x;\psi)^q, \tag{31}$$

where

$$\varpi_{ijkmq} = \frac{(-1)^{i+k+m+q} \delta^i}{i!} \cdot \binom{\alpha\theta(i+\delta)+\delta+j-1}{j} \binom{\theta(i+\delta)-\delta}{k} \binom{\alpha k}{m} \binom{m+j}{q}. \tag{32}$$

Hence,

$$I_R(\delta) = \frac{1}{1-\delta} \log \left[(\alpha\theta)^\delta \sum_{i,j=0}^{\infty} \sum_{k,m=0}^{\infty} \sum_{q=0}^{m+j} \varpi_{ijkmq} \right.$$
$$\left. \cdot \int_{-\infty}^{\infty} g(x;\psi)^\delta G(x;\psi)^q dx \right], \quad \delta > 0, \ \delta \neq 1. \tag{33}$$

The Shannon entropy [21] of a random variable X, say $\eta_X = E(-\log f(X))$. The Shannon entropy is a special case of the Rényi entropy when $\delta \uparrow 1$. The $\delta-$entropy is given by

$$H(\delta) = \frac{1}{\delta-1} \log \left[1 - \int_{-\infty}^{\infty} f(x)^\delta dx \right], \tag{34}$$
$$\delta > 0, \ \delta \neq 1.$$

Thus, the $\delta-$entropy is

$$H(\delta) = \frac{1}{\delta-1} \left[1 - (\alpha\theta)^\delta \sum_{i,j=0}^{\infty} \sum_{k,m=0}^{\infty} \sum_{q=0}^{m+j} \varpi_{ijkmq} \right.$$
$$\left. \cdot \int_{-\infty}^{\infty} g(x;\psi)^\delta G(x;\psi)^q dx \right], \quad \delta > 0, \ \delta \neq 1. \tag{35}$$

4.5. Order Statistics.

Let $X_1, X_2, \ldots, X_n$ represent a random sample from EOF-G family and $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the order statistics. Then the PDF, $f_{p:n}(x)$, of the pth order statistic $X_{p:n}$ is

$$f_{p:n}(x) = \frac{n!}{(p-1)!(n-p)!} \sum_{i=0}^{n-p} (-1)^i F(x)^{p+i-1} f(x). \tag{36}$$

Substituting the PDF and the CDF of the EOF-G random variable into the last equation yields

$$f_{p:n}(x) = \frac{n!\alpha\theta}{(p-1)!(n-p)!}$$
$$\cdot \sum_{j,k=0}^{\infty} \sum_{q,s=0}^{\infty} \sum_{w=0}^{k+s} \sum_{i=0}^{n-p} \varphi_{ijkqsw} g(x;\psi) G(x;\psi)^w, \tag{37}$$

after some algebraic manipulation, where

$$\varphi_{ijkqsw} = \frac{(-1)^{i+j+q+s+w} (p+i)^j}{j!}$$
$$\cdot \binom{n-p}{i} \binom{\alpha\theta(j+1)+k}{k} \binom{\theta(j+1)-1}{q} \binom{\alpha q}{s} \binom{k+s}{w}. \tag{38}$$

The PDF of the pth order statistic can be expressed in terms of the exp-G density function as

$$f_{n:p}(x)$$
$$= \frac{n!\alpha\theta}{(p-1)!(n-p)!} \sum_{j,k=0}^{\infty} \sum_{q,s=0}^{\infty} \sum_{w=0}^{k+s} \sum_{i=0}^{n-p} \varphi_{ijkqsw}^* \Delta_{w+1}(x), \tag{39}$$

where $\varphi_{ijkqsw}^* = \varphi_{ijkqsw}/(w+1)$ and $\Delta_{w+1}(x) = (w+1)g(x;\psi)G(x;\psi)^w$ is the exp-G density function with power parameter $w+1$.

5. Parameter Estimation

In this section, the maximum likelihood technique is employed to develop estimators for estimating the parameters of the EOF-G family of distributions. Suppose $x_1, x_2, \ldots, x_n$ are possible outcomes of a random sample obtained from

$X \sim \text{EOF} - \text{G}(x; \alpha, \theta, \boldsymbol{\psi})$ and $\boldsymbol{\vartheta} = (\alpha, \theta, \boldsymbol{\psi})^T$ is a parameter vector; then the total log-likelihood function is given by

$$
\begin{aligned}
\ell = {} & n \log(\alpha\theta) + \sum_{i=1}^{n} \log \text{g}(x_i; \boldsymbol{\psi}) \\
& + (\theta - 1) \sum_{i=1}^{n} \log\left[1 - G(x_i; \boldsymbol{\psi})^\alpha\right] \\
& - (\alpha\theta + 1) \sum_{i=1}^{n} \log G(x_i; \boldsymbol{\psi}) \\
& - \sum_{i=1}^{n} \left[\frac{1 - G(x_i; \boldsymbol{\psi})^\alpha}{G(x_i; \boldsymbol{\psi})^\alpha}\right]^\theta .
\end{aligned}
\tag{40}
$$

By finding the partial derivatives of (40), the components of the score vector $U(\boldsymbol{\vartheta}) = (\partial\ell/\partial\alpha, \partial\ell/\partial\theta, \partial\ell/\partial\boldsymbol{\psi})^T$ are

$$
\begin{aligned}
\frac{\partial\ell}{\partial\alpha} = {} & \frac{n}{\alpha} + (\theta - 1) \sum_{i=1}^{n} \frac{G(x_i; \boldsymbol{\psi})^\alpha \log G(x_i; \boldsymbol{\psi})}{1 - G(x_i; \boldsymbol{\psi})} \\
& - \theta \sum_{i=1}^{n} \log G(x_i; \boldsymbol{\psi}) \\
& + \theta \sum_{i=1}^{n} \frac{\left[1 - G(x_i; \boldsymbol{\psi})^\alpha\right]^{\theta-1} \log G(x_i; \boldsymbol{\psi})}{G(x_i; \boldsymbol{\psi})^{\alpha\theta}},
\end{aligned}
\tag{41}
$$

$$
\begin{aligned}
\frac{\partial\ell}{\partial\theta} = {} & \frac{n}{\theta} + \sum_{i=1}^{n} \log\left[1 - G(x_i; \boldsymbol{\psi})^\alpha\right] - \alpha \sum_{i=1}^{n} \log G(x_i; \boldsymbol{\psi}) \\
& - \sum_{i=1}^{n} \left[\frac{1 - G(x_i; \boldsymbol{\psi})^\alpha}{G(x_i; \boldsymbol{\psi})^\alpha}\right]^\theta \log\left[\frac{1 - G(x_i; \boldsymbol{\psi})^\alpha}{G(x_i; \boldsymbol{\psi})^\alpha}\right],
\end{aligned}
\tag{42}
$$

$$
\begin{aligned}
\frac{\partial\ell}{\partial\boldsymbol{\psi}} = {} & \sum_{i=1}^{n} \frac{\text{g}'(x_i; \boldsymbol{\psi})}{\text{g}(x_i; \boldsymbol{\psi})} \\
& + \alpha(\theta - 1) \sum_{i=1}^{n} \frac{G'(x_i; \boldsymbol{\psi}) G(x_i; \boldsymbol{\psi})^{\alpha-1}}{1 - G(x_i; \boldsymbol{\psi})} \\
& - (\alpha\theta + 1) \sum_{i=1}^{n} \frac{G'(x_i; \boldsymbol{\psi})}{G(x_i; \boldsymbol{\psi})} \\
& + \alpha\theta \sum_{i=1}^{n} \frac{G'(x_i; \boldsymbol{\psi}) \left[1 - G(x_i; \boldsymbol{\psi})^\alpha\right]^{\theta-1}}{G(x_i; \boldsymbol{\psi})^{\alpha\theta+1}},
\end{aligned}
\tag{43}
$$

where $\text{g}'(x_i; \boldsymbol{\psi}) = \partial\text{g}(x_i; \boldsymbol{\psi})/\partial\boldsymbol{\psi}$ and $G'(x_i; \boldsymbol{\psi}) = \partial G(x_i; \boldsymbol{\psi})/\partial\boldsymbol{\psi}$. In order to obtain the estimators for the parameters, we set (41), (42), and (43) to zero and solve the system numerically using methods such as the quasi-Newton algorithms since the equations do not have closed form. To obtain interval estimates of the parameters, a $p \times p$ observed information matrix can be estimated as $J(\boldsymbol{\vartheta}) = \partial^2 \ell / \partial q \partial r$ (for $q, r = \alpha, \theta, \boldsymbol{\psi}$), whose elements are evaluated numerically. To compute the approximate confidence intervals of the parameters, the multivariate normal distribution $N_p(\mathbf{0}, J(\widehat{\boldsymbol{\vartheta}})^{-1})$. Here, $J(\widehat{\boldsymbol{\vartheta}})$ is the observed information evaluated at $\widehat{\boldsymbol{\vartheta}}$. To investigate whether the EOF-G distributions are superior to the odd Fréchet family of distributions for given data sets, the likelihood ratio (LR) test can be performed using the following hypotheses: $H_0 : \alpha = 1$ versus $H_a : H_0$ is false. The LR test statistic is given by $LR = 2\{\ell(\widehat{\boldsymbol{\vartheta}}) - \ell(\overline{\boldsymbol{\vartheta}})\}$, where $\widehat{\boldsymbol{\vartheta}}$ is the vector of unrestricted estimates under H_a and $\overline{\boldsymbol{\vartheta}}$ is the vector of restricted maximum likelihood estimates under H_0. The LR test statistic is asymptotically distributed as Chi-square random variable with degrees of freedom equal to the difference between the numbers of parameters of the two models. As a decision rule, the null hypothesis is rejected when the LR test statistic exceeds the upper $100(1 - \eta)\%$ quantile of the Chi-square distribution.

6. Simulation Study

In this section, Monte Carlo simulations are performed to assess the accuracy and consistency of the maximum likelihood estimators. For the purpose of illustration, the simulations are performed using the estimators of the parameters of the EOFNH distribution. The quantile function given in (9) is used to generate random observations from the EOFNH distribution. The simulations are repeated $N = 1,000$ times each with sample size $n = 25, 75, 150, 300, 600, 800$ and parameter values I $: \alpha = 0.5, \beta = 0.5, \lambda = 0.5, \theta = 0.5$, II $: \alpha = 3.3, \beta = 0.8, \lambda = 0.2, \theta = 0.8$, and III $: \alpha = 0.9, \beta = 0.4, \lambda = 0.2, \theta = 0.6$. Table 1 presents the average bias (AB), the root mean square error (RMSE), and coverage probability (CP) of the 95% confidence intervals for the estimators of the parameters. The results indicated that the ABs and RMSEs decrease as the sample size increases. These results clearly show the accuracy and the consistency of the maximum likelihood estimators. Also, the CPs are quite close to the nominal value. Thus, the maximum likelihood technique works very well to estimate the parameters of the EOFNH distribution.

7. Application

In this section, the application of the EOFNH and EOFW distributions is illustrated using a real data set. The data consists of the Fatigue time of 101 6061-T6 aluminum coupons cut parallel to the direction of rolling and oscillated at 18 cycles per second. The data set given in Table 2 can be found in Birnbaum and Saunders [22]. The performance of the EOFNH and EOFW distributions is compared with that of the odd Fréchet Nadarajah-Haghighi (OFNH) and odd Fréchet Weibull (OFW) distributions using the Akaike information criterion (AIC) [23, 24] and Bayesian information criterion (BIC) [25]. The maximum likelihood estimates of the parameters of the fitted distributions are computed by maximizing the log-likelihood function via the subroutine *mle2* uisng the *bbmle* package in the R software [26].

TABLE 1: Monte Carlo simulation results: AB and RMSE and CP.

Parameter	n	I			II			III		
		AB	RMSE	CP	AB	RMSE	CP	AB	RMSE	CP
α	25	0.5406	1.9394	0.9780	24.9701	133.9291	0.9160	3.1057	17.0755	0.8780
	75	0.1710	0.5134	0.9470	12.5503	76.6534	0.9030	1.1174	3.9980	0.8850
	150	0.0919	0.3459	0.9120	5.6552	41.2149	0.9200	0.4664	1.3878	0.9020
	300	0.0242	0.2526	0.8670	2.1667	5.1107	0.9540	0.2038	0.6941	0.9280
	600	-0.0043	0.2076	0.8300	0.9715	2.8387	0.9640	0.1071	0.4109	0.9430
	800	-0.0010	0.1794	0.8670	0.7123	2.3503	0.9560	0.0623	0.3374	0.9430
β	25	0.0040	0.2147	0.9620	0.4450	1.3048	0.9810	0.0212	0.1894	0.9920
	75	0.0488	0.3523	0.9630	0.2138	0.5569	0.9880	0.0163	0.1195	0.9830
	150	0.0740	0.4896	0.9620	0.1105	0.3330	0.9890	0.0117	0.0705	0.9680
	300	0.0890	0.4095	0.975	0.0633	0.2147	0.9720	0.0068	0.0438	0.9620
	600	0.066	0.2745	0.9670	0.0324	0.1275	0.9630	0.0032	0.0296	0.9520
	800	0.0357	0.1378	0.9730	0.0264	0.1105	0.9620	0.0021	0.0255	0.9510
λ	25	12.5655	249.9064	0.9340	3.7073	32.1823	0.8930	4.9031	475403	0.8170
	75	0.6886	2.7878	0.8530	1.2894	15.0985	0.8850	1.4279	12.6625	0.8460
	150	0.3001	0.9320	0.8280	0.1493	2.7383	0.9200	0.2900	1.9911	0.8740
	300	0.1265	0.5992	0.8020	0.0298	0.1599	0.9490	0.0726	0.2735	0.8950
	600	0.0400	0.4286	0.7900	0.0075	0.0931	0.9800	0.0375	0.1450	0.9320
	800	0.0351	0.3629	0.8400	0.0053	0.0856	0.9670	0.0227	0.1175	0.9300
θ	25	0.0856	0.3163	0.9999	0.4125	1.7492	0.8970	0.1846	0.5637	0.9810
	75	0.0616	0.2515	0.9800	0.1192	0.8350	0.8370	0.0725	0.3927	0.9290
	150	0.0599	0.2400	0.9500	0.1242	0.7001	0.8280	0.0599	0.3320	0.9310
	300	0.0790	0.2422	0.9200	0.0602	0.4882	0.8550	0.0435	0.2675	0.9260
	600	0.0824	0.2313	0.8840	0.0506	0.3537	0.8940	0.1230	0.1618	0.9400
	800	0.0601	0.1957	0.9040	0.0396	0.3043	0.9130	0.0179	0.1502	0.9410

TABLE 2: Fatigue time of 101 6061-T6 aluminum coupons.

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 3: Maximum likelihood estimates and goodness-of-fit statistics.

Model	Parameter estimates	$-\ell$	AIC	BIC
EOFNH	$\widehat{\alpha} = 1.5838 \ (0.2023)$ $\widehat{\beta} = 1.1413 \ (0.4131)$ $\widehat{\lambda} = 0.0071 \ (0.0041)$ $\widehat{\theta} = 2.7505 \ (0.1088)$	-470.3600	948.7255	959.1860
OFNH	$\widehat{\alpha} = 0.4650 \ (0.4204)$ $\widehat{\lambda} = 0.0174 \ (0.0263)$ $\widehat{\theta} = 4.7456 \ (1.8584)$	-473.6000	953.1958	961.0412
EOFW	$\widehat{\alpha} = 2.2281 \ (1.0985)$ $\widehat{\beta} = 1.0205 \ (0.3152)$ $\widehat{\lambda} = 0.0099 \ (0.0166)$ $\widehat{\theta} = 2.3066 \ (0.6716)$	-471.2600	950.5259	960.9864
OFW	$\widehat{\beta} = 0.2785 \ (0.0181)$ $\widehat{\lambda} = 0.1823 \ (0.0156)$ $\widehat{\theta} = 13.2247 \ (0.0007)$	-473.8200	953.6325	961.4778

The PDFs of the OFNH and OFW distributions are, respectively, given by

$$f(x)$$

$$= \frac{\beta \lambda \theta (1 + \lambda x)^{\beta - 1} e^{(1 - (1 + \lambda x)^\beta)} \left[1 - \left(1 - e^{1 - (1 + \lambda x)^\beta} \right) \right]^{\theta - 1}}{\left(1 - e^{(1 - (1 + \lambda x)^\beta)} \right)^{\theta + 1}} \quad (44)$$

$$\cdot e^{-[(1 - e^{(1 - (1 + \lambda x)^\beta)})^{-1} - 1]^\theta}, \quad x > 0,$$

and

$$f(x) = \frac{\beta \lambda \theta x^{\beta - 1} e^{-\lambda x^\beta} \left[1 - \left(1 - e^{-\lambda x^\beta} \right) \right]^{\theta - 1}}{\left(1 - e^{-\lambda x} \right)^{\theta + 1}} \quad (45)$$

$$\cdot e^{-[(1 - e^{-\lambda x^\beta})^{-1} - 1]^\theta}, \quad x > 0.$$

Table 3 displays the maximum likelihood estimates of the parameters of the EOFNH, EOFW, OFNH, and OFW distributions with their corresponding standard errors in bracket and the model selection criteria. The results revealed that the EOFNH distribution provided the best fit for the data since it has the least values of AIC and the BIC. The EOFW distribution also performed better than the OFNH and OFW distributions. The OFNH distribution is a submodel

of the EOFNH distribution with $\alpha = 1$. Hence, testing $H_0 : \alpha = 1$ versus $H_a : \alpha \neq 1$ using the LR test gave a test statistic of 6.4703 with corresponding p-value of 0.01097. This implies that there is enough evidence to reject H_0 at the 5% significance level and conclude that the EOFNH distribution provides better fit to the data than the OFNH distribution. Similarly, the LR test was performed to compare the performances of the EOFW distribution and the OFW distribution. The analysis gave a test statistic of 5.1065 with a corresponding p-value of 0.0238. This implies that the EOFW distribution performs better than the OFW distribution at the 5% significance level.

Figure 5 displays the histogram of the data with the fitted densities and the empirical CDF with the fitted CDFs.

The P-P plots of the fitted distributions are displayed in Figure 6.

8. Conclusion

The development of new statistical distribution plays a critical role in parametric statistical inference. Because of this, researchers in the field of distribution theory attempt to develop generators for generalizing the existing distributions. In line with this, the study developed and studied a new class of distributions called the EOF-G family. The statistical properties including the moments, incomplete
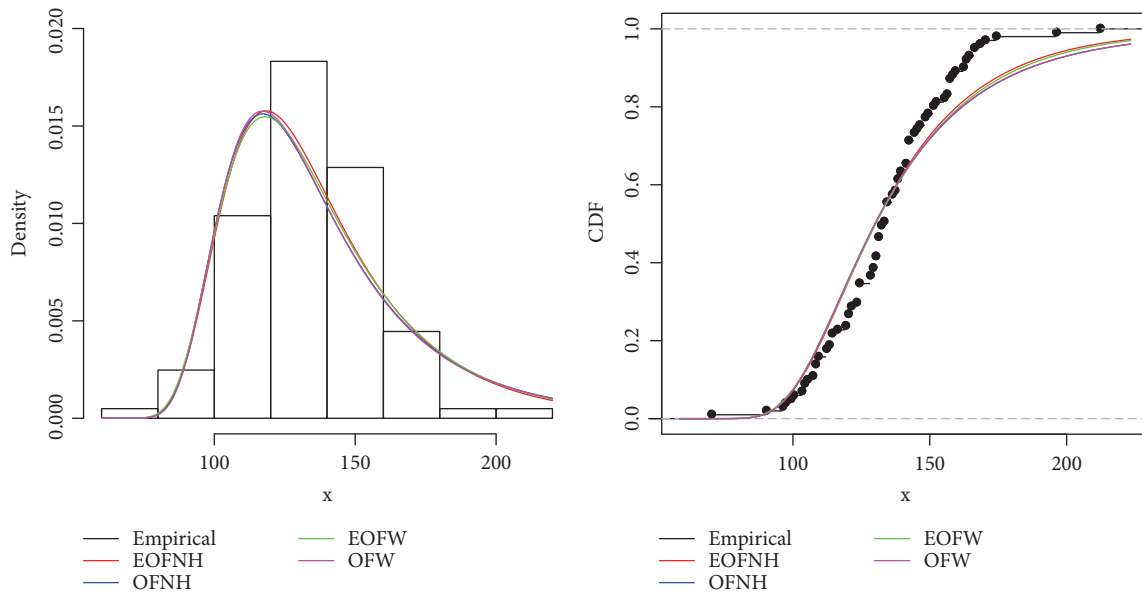
FIGURE 5: Plots of histogram of data and fitted densities; and empirical CDF and fitted CDF.
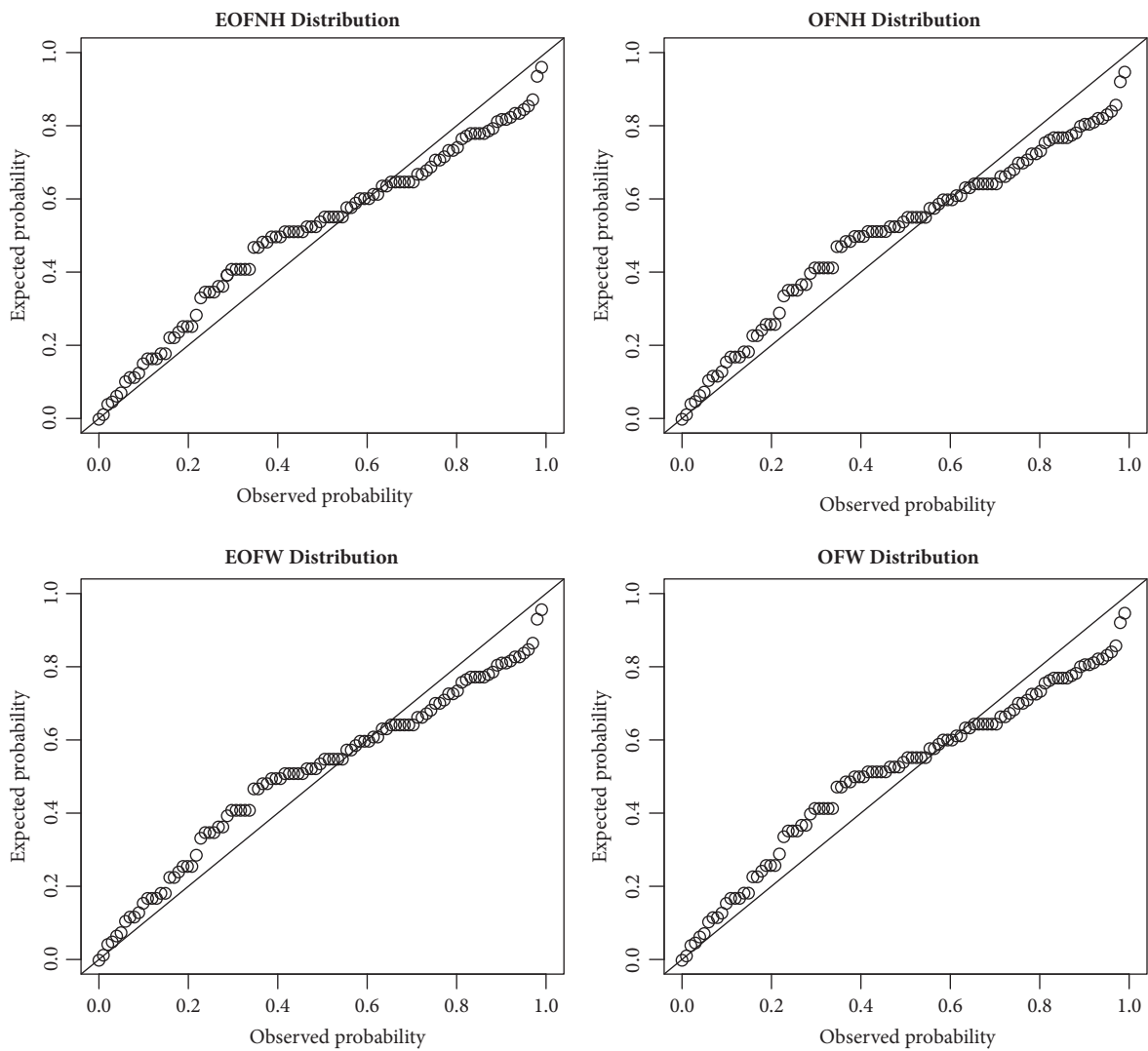
FIGURE 6: P-P plots of fitted distributions.

moments, generating function, entropies, and order statistics are derived. The maximum likelihood method is used to develop estimators for the parameters of the new family. The application of the special distributions developed using the EOF-G family is demonstrated using a real data set and the result compared with other existing distributions. From the application, it is evident that the special models developed from the EOF-G family can provide reasonable parametric fit to a given data set. Hence, it is hoped that the new class of distributions will attract wider applications in different fields of study.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this article.

References

[1] M. Bourguignon, R. B. Silva, and G. M. Cordeiro, "The weibull-G family of probability distributions," *Journal of Data Science*, vol. 12, pp. 53–68, 2014.

[2] M. H. Tahir, G. M. Cordeiro, M. Alizadeh, M. Mansoor, M. Zubair, and G. G. Hamedani, "The odd generalized exponential family of distributions with applications," *Journal of Statistical Distributions and Applications*, vol. 2, no. 1, pp. 1–28, 2015.

[3] F. Gomes-Silva, A. Percontini, E. de Brito, M. W. Ramos, R. Venâncio, and G. M. Cordeiro, "The odd Lindley-G family of distributions," *Austrian Journal of Statistics*, vol. 46, no. 1, pp. 65–87, 2017.

[4] E. Brito, G. M. Cordeiro, H. M. Yousof, M. Alizadeh, and G. O. Silva, "The Topp-Leone odd log-logistic family of distributions," *Journal of Statistical Computation and Simulation*, vol. 87, no. 15, pp. 3040–3058, 2017.

[5] M. A. Nasir, F. Jamal, G. O. Silva, and M. H. Tahir, "Odd Burr-G Poisson family of distributions," *Journal of Statistics Applications and Probability*, vol. 7, no. 1, pp. 9–28, 2018.

[6] M. A. Haq and M. Elgarhy, "The odd Fréchet-G family of probability distributions," *Journal of Statistics Applications & Probability*, vol. 7, no. 1, pp. 189–203, 2018.

[7] B. Hosseini, M. Afshari, and M. Alizadeh, "The generalized odd gamma-G family of distributions: properties and applications," *Austrian Journal of Statistics*, vol. 47, pp. 47–69, 2018.

[8] A. Alzaatreh, C. Lee, and F. Famoye, "A new method for generating families of continuous distributions," *METRON*, vol. 71, no. 1, pp. 63–79, 2013.

[9] A. Alzaghal, F. Famoye, and C. Lee, "Exponentiated T-X family of distributions with some applications," *International Journal of Statistics and Probability*, vol. 2, no. 3, pp. 31–49, 2013.

[10] S. Nasiru, P. N. Mwita, and O. Ngesa, "Exponentiated generalized Transformed-Transformer family of distributions," *Journal of Statistical and Econometric Methods* , vol. 6, no. 4, p. 17, 2017.

[11] A. Mahdavi and D. Kundu, "A new method for generating distributions with an application to exponential distribution,"

[12] V. Pappas, K. Adamidis, and S. Loukas, "A family of lifetime distributions," *International Journal of Quality, Statistics and Reliability*, vol. 2012, 6 pages, 2012.

[13] G. M. Cordeiro and M. de Castro, "A new family of generalized distributions," *Journal of Statistical Computation and Simulation*, vol. 81, no. 7, pp. 883–898, 2011.

[14] N. Eugene, C. Lee, and F. Famoye, "Beta-normal distribution and its applications," *Communications in Statistics—Theory and Methods*, vol. 31, no. 4, pp. 497–512, 2002.

[15] A. Z. Afify, G. M. Cordeiro, H. M. Yousof, A. Alzaatreh, and Z. M. Nofal, "The Kumaraswamy transmuted-G family of distributions: properties and applications," *Journal of Data Science*, pp. 245–270, 2016.

[16] A. Z. Afify, M. Alizadeh, H. M. Yousof, G. Aryal, and M. Ahmad, "The transmuted geometric-G family of distributions: theory and applications," *Pakistan Journal of Statistics*, vol. 32, no. 2, pp. 139–160, 2016.

[17] G. M. Cordeiro, G. O. Silva, and E. M. Ortega, "The beta extended Weibull family," *JPSS. Journal of Probability and Statistical Science*, vol. 10, no. 1, pp. 15–40, 2012.

[18] K. Cooray, "Generalization of the WEIbull distribution: the odd WEIbull family," *Statistical Modelling. An International Journal*, vol. 6, no. 3, pp. 265–277, 2006.

[19] I. S. Gradshteyn and I. M. Ryzhik, *Tables of integrals, series, and products*, Academic Press, NY, USA, 2007.

[20] A. Rényi, "On measures of entropy and information," in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 547–561, University of California Press, 1961.

[21] C. E. Shannon, "A mathematical theory of communication," *Bell Labs Technical Journal*, vol. 27, pp. 379–423, 1948.

[22] Z. W. Birnbaum and S. C. Saunders, "Estimation for a family of life distribution with applications to fatigue," *Journal of Applied Probability*, vol. 6, no. 2, pp. 328–347, 1969.

[23] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *International Symposium on Information Theory*, vol. 2nd, pp. 267–281, American SSR, Tsahkadsor, 1973.

[24] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, pp. 716–723, 1974.

[25] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[26] B. Bolker, "Tools for general maximum likelihood estimation," R development core team, 2014.

Communications in Statistics—Theory and Methods, vol. 46, no. 13, pp. 6543–6557, 2017.

Similarity Statistics for Clusterability Analysis with the Application of Cell Formation Problem

Yingyu Zhu and Simon Li ⓘ

Department of Mechanical and Manufacturing Engineering, University of Calgary, Alberta, Canada

Correspondence should be addressed to Simon Li; simoli@ucalgary.ca

This paper proposes the use of the statistics of similarity values to evaluate the clusterability or structuredness associated with a cell formation (CF) problem. Typically, the structuredness of a CF solution cannot be known until the CF problem is solved. In this context, this paper investigates the similarity statistics of machine pairs to estimate the potential structuredness of a given CF problem without solving it. One key observation is that a well-structured CF solution matrix has a relatively high percentage of high-similarity machine pairs. Then, histograms are used as a statistical tool to study the statistical distributions of similarity values. This study leads to the development of the U-shape criteria and the criterion based on the Kolmogorov-Smirnov test. Accordingly, a procedure is developed to classify whether an input CF problem can potentially lead to a well-structured or ill-structured CF matrix. In the numerical study, 20 matrices were initially used to determine the threshold values of the criteria, and 40 additional matrices were used to verify the results. Further, these matrix examples show that genetic algorithm cannot effectively improve the well-structured CF solutions (of high grouping efficacy values) that are obtained by hierarchical clustering (as one type of heuristics). This result supports the relevance of similarity statistics to preexamine an input CF problem instance and suggest a proper solution approach for problem solving.

1. Introduction

The research of this paper is like a crossroad of manufacturing systems and computer science. Based on our disciplinary background, we initially study the cell formation (CF) problem that seeks for the clustering of similar machines and parts to support mass customization in [1]. In other words, a CF problem is a two-mode clustering problem [2]. Due to the NP-hard nature of the CF problem [3], many algorithms, including exact, metaheuristic, and heuristic approaches, have been proposed (to be discussed in Section 2.2.3). In the study of hierarchical clustering (abbreviated as HC, classified as a greedy-based heuristic approach), although HC is not the most powerful in searching for near-optimal solutions, it can yield satisfactory results comparable to some powerful metaheuristic approaches (e.g., genetic algorithms) for "well-structured" solutions. In this context, this research investigates the conditions based on the statistics of similarity values to estimate the potential structuredness of a given CF problem without solving it.

In the domain of computer science, the notion of structuredness somehow corresponds to the clusterability concept [4]. Intuitively, clusterability can be interpreted as a measure of an "intrinsic structure" of a dataset to be clustered [5]. Computer scientists have observed that a dataset of good clusterability can be clustered quite effectively (i.e., less impact from the NP-hard nature of the clustering problem). This observation has been summarized in a statement that "clustering is difficult only when it does not matter" (abbreviated as the CDNM thesis) [4, 6].

Notably, the measure of clusterability remains an open topic in computer science. Ackerman and Ben-David [4] have surveyed different definitions of clusterability and shown their incompatibility in pairwise comparisons. Nowakowska et al. [5] argued that a clusterability measure should be partition-independent so that it does not depend on the clustering algorithms and the resulting solutions. Ackerman et al. [7] proposed the use of the statistical distributions of pairwise distances between any two objects to evaluate clusterability.
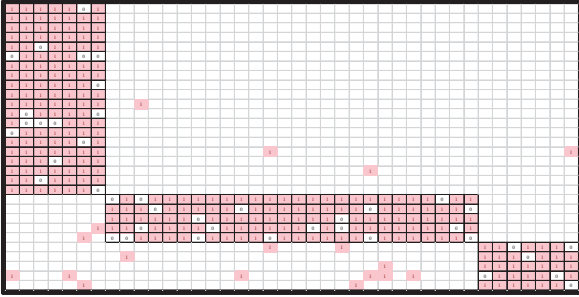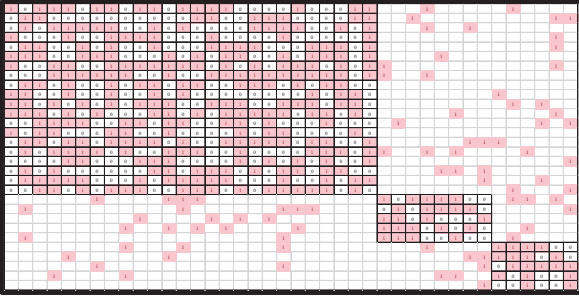
Example of a well-structured matrix	Example of an ill-structured matrix
Grouping efficacy = 0.8142	Grouping efficacy = 0.4780
81 (out of 435) machine pairs have similarity values higher than or equal to 0.80	4 (out of 435) machine pairs have similarity values higher than or equal to 0.50

FIGURE 1: Comparison of well-structured and ill-structured matrices.

Back to the context of the CF problem, in response to the CDNM thesis, we also observed that a heuristic approach (e.g., HC in our case) can yield satisfactory results. To further utilize this observation in practice, this research develops the criteria that assess the potential structuredness (corresponding to clusterability in computer science) of a given CF problem and suggest either using HC or genetic algorithm (GA) for problem solving. To verify the development, we have applied numerical examples to examine the results of the structuredness criteria and the quality of CF solutions via HC and GA.

Though developed independently, we want to acknowledge that our approach of evaluating the structuredness criteria is similar to the statistical approach by Ackerman et al. [7]. The difference lies in our application's focus on the CF problem, while Ackerman et al. [7] have focused on the relatively high-level development for clustering tasks. This difference explains our use of similarity measures (instead of distances) in statistical analysis since they are common for the CF problem and allow for some normalization in setting the structuredness criteria. Further, our work numerically checks the relations between structuredness criteria and the solution quality by two different clustering approaches (i.e., HC and GA).

Notably, this paper was extended from our conference paper [8] with the improvement of the techniques (e.g., the threshold setting and the normalization approach). Also, additional numerical examples have been used in the evaluation.

The rest of this paper is organized as follows. Section 2 will overview the CF problem and discuss the three properties of a well-structured CF solution in order to clarify the logical relation of similarity statistics. Section 3 will introduce the histogram analysis of similarity values and develop the U-shape criteria. Section 4 will introduce the Kolmogorov-Smirnov (K-S) test, which is used to develop another criterion to inform the matrix's structuredness. Section 5 will discuss the procedure that applies the developed criteria to classify

well-structured and ill-structured matrices. Section 6 will examine the structuredness criteria via numerical examples, which are also used to check the effectiveness of metaheuristics via a two-stage solution process. Section 7 will conclude this paper.

2. Background: Cell Formation Problem

2.1. Problem Introduction. In the design of a cellular manufacturing system, one early and important decision is the formation of machine groups and part families, and it is often referred to as the cell formation (CF) problem. A simple CF problem can be compactly captured by a machine-part incidence matrix. Let $M = \{m_i\}$ (for $i = 1$ to m) be the set of machines and $P = \{p_j\}$ (for $j = 1$ to n) be the set of parts. Then, an incidence matrix, denoted as $B = [b_{ij}]$, indicates whether machine m_i is required to produce part p_j (if so, $b_{ij} = 1$; otherwise, $b_{ij} = 0$). After solving the CF problem, the matrix's rows and columns can be reordered to reveal which subset of machines (i.e., a machine group) is highly related to which subset of parts (i.e., a part family).

By using the incidence matrices to represent CF solutions (i.e., block-diagonal matrices), they can be roughly classified into two types: well-structured and ill-structured matrix [2, 9]. As illustrated in Figure 1, a well-structured matrix has few nonzero matrix entries outside the blocks (defined as exceptional elements) and few zero matrix entries inside the blocks (defined as voids). Precisely, exceptional elements are the matrix entries of $b_{ij} = 1$ with m_i and p_j in different cells, and voids are the matrix entries of $b_{ij} = 0$ with m_i and p_j in the same call. The opposite conditions apply for an ill-structured matrix (i.e., a matrix solution with many exceptional elements and voids). A well-structured matrix implies that part families can be produced quite exclusively by some machine groups so that the changes of few part families will not be adversely impacting the production of other parts. This is one desirable feature of cellular manufacturing systems [1].

To quantify the structuredness of a CF matrix solution, we use the traditional grouping efficacy (denoted as μ), which is formulated as follows [10].

$$\mu = \frac{n_e - n_{out}}{n_e + n_{in}} \tag{1}$$

where n_e, n_{out}, n_{in} are the total number of nonzero matrix entries, exceptional elements, and voids, respectively. In a perfect CF solution where $n_{out} = n_{in} = 0$, the grouping efficiency is equal to its maximum value, i.e., one. When there are more exceptional elements (n_{out}) and voids (n_{in}), the grouping efficacy value will become smaller.

Yet, not all incidence matrices can be converted to a well-structured matrix due to the original complex interdependency of the production requirements among machines and parts. This situation cannot be resolved by advanced optimization techniques as the root cause stems from the original inputs of the CF problem. However, we cannot practically know whether a given CF problem is going to have a well-structured matrix or not until we actually solve this problem. In this context, the purpose of this paper is to assess the structuredness of a given CF problem by analyzing the similarity of machines without actually solving it. In the traditional CF notion, two machines can be said similar if they are required mainly to produce a subset of common parts. In this work, the Jaccard similarity coefficient is applied [11, 12]. Let s_{xy} be the similarity value between machines m_x and m_y. The formulation of the Jaccard similarity coefficient is provided below.

$$s_{xy} = \frac{a_{xy}}{a_{xy} + b_{xy} + c_{xy}} \tag{2}$$

where a_{xy} is the number of parts that need both machine m_x and m_y; b_{xy} is the number of parts that need machine m_x but not machine m_y; c_{xy} is the number of parts that need machine m_y but not machine m_x. Conceptually, the Jaccard similarity coefficient focuses on the number of common features (e.g., a_{xy}) that is normalized by the total number of relevant features (e.g., a_{xy}, b_{xy}, and c_{xy}). Notably, similarity is only evaluated for any two machines (i.e., a machine pair).

After specifying the notion of machine similarity, let us revisit the two examples in Figure 1. Each example has 30 machines, leading to 30×(30-1)/2 = 435 machine pairs. By examining the similarity of any two machines (or machine pairs), we find that the well-structured matrix has a higher number of machine pairs with high-similarity values. In the examples of Figure 1, we can get the following two statements concerning the statistics of the machine similarity values.

(i) Well-structured matrix: 81 (out of 435) machine pairs have similarity values higher than or equal to 0.80.

(ii) Ill-structured matrix: 4 (out of 435) machine pairs have similarity values higher than or equal to 0.50.

In this illustration, it is roughly identified that a well-structured matrix can have quite a different statistical distribution of machine similarity values as compared to an ill-structured matrix. This observation leads to an investigation question on the statistical conditions in which a well-structured matrix can be classified. This investigation is the focus of this paper. By knowing such statistical conditions, engineers in the design of cellular manufacturing systems can initially assess their production requirements via the statistics of machine similarity. If the statistical data shows unfavorable results (i.e., chance of getting a well-structured matrix is low), they can either modify the production requirements (e.g., buy more machines) or seek for other manufacturing systems. It can save the efforts to solve the CF problem with such initial assessment. Also, this paper will show that a well-structured matrix can be satisfactorily obtained by some less time-consuming heuristics (where complex optimization methods may not bring additional benefits).

2.2. Properties of a Well-Structured CF Solution. To investigate the statistical conditions of the structuredness of a CF solution, this section will discuss the three properties of a well-structured matrix. These three properties include (1) high grouping efficacy, (2) high percentage of high-similarity machine pairs, and (3) relative ease of obtaining satisfactory CF solutions. Afterward, a research plan will be discussed.

2.2.1. Property I: High Grouping Efficacy. The original formulation of the grouping efficacy (GE) in (1) can be found in Kumar and Chandrasekharan [10], and it is intended to replace a weighted sum function with a simple ratio to assess the goodness of a CF solution (in a block-diagonal form). Since then, the GE measure has become popular in the CF research (e.g., [9, 13]). Despite its popularity, some researchers have criticized its "built-in weights" [14], where a lower number of voids (i.e., n_{in}) tend to give a better GE measure (as compared to exceptional elements (i.e., n_{out})). Brusco [15] has commented that the nonlinearity of the GE measure has incurred a challenge for finding the exact solutions for the CF problems. As commented by Sarker and Mondal [16] in their survey paper, it is not easy to develop a standard measure that fits all CF problems. It is generally recognized that the GE measure is good to discern the structuredness of the matrix-based CF solutions [2]. Thus, we choose the GE measure in this study.

Based on its definition, a well-structured matrix should have few exceptional elements and voids, leading to a high value of GE. While GE is effective in indicating the structuredness of a CF solution (high value ⟶ well-structured matrix), this value cannot be known until the CF problem is solved. Thus, in this research, GE is used as a verification measure to examine how well machine similarity can be related to the structuredness of a CF solution.

2.2.2. Property II: High Percentage of High-Similarity Machine Pairs. Compared to the property of high grouping efficacy, it is less obvious to know that a well-structured matrix has a high percentage of high-similarity machine pairs. In view of the Jaccard similarity coefficient in (2), there are two types of factors used to assess the machine similarity. While a_{xy} (i.e., the number of common parts) is taken as a commonality factor, both b_{xy} and c_{xy} (i.e., the number of parts processed in one machine but not another one) serve as differentiating factors to normalize the similarity measure. In turn, if the similarity value of both machines is high, a_{xy} cannot be zero

and the values of b_{xy} and c_{xy} should be small, implying not only commonality but also exclusiveness of these two machines to process their common parts. This feature can potentially lead to smaller numbers of voids and exceptional numbers, leading to a well-structured matrix.

In literature, the notion of similarity has been applied for many years to address the CF problem, and the Jaccard similarity coefficient is one of the early applications [11]. Since then, many similarity coefficients have been proposed, and the comparison study of similarity coefficients can be found in Sarker [17], Mosier et al. [18], and Yin and Yasuda [19]. Notably, similarity is a context-dependent concept, and it depends on the application and relevant information to assess how similar between two objects. In our investigation, we choose the Jaccard similarity coefficient because its notion on the commonality and differentiating factors is straightforward to the simple CF application.

While similarity coefficients have been studied extensively for CF problems, the statistical distribution of similarity values of a CF problem has not been investigated reasonably in our understanding. Notably, these similarity values can be found without solving the CF problems. Then, if we know the relation between the statistical distribution of similarity values and the GE measure, we can use the statistical distribution of similarity values to assess the potential of yielding a well-structured matrix for a CF problem. This is the major aim of this paper.

2.2.3. Property III: Relative Ease of Obtaining Satisfactory CF Solutions.

At this point, we may wonder why it is important to know the potential of yielding a well-structured matrix before solving the CF problems. First of all, it has been recognized that a CF problem is a NP-hard problem [3] so that there will be less likely to find a practical algorithm that can guarantee an exact solution for a moderate-size problem. As a result, the effort required to solve a CF problem is not trivial. In literature, many metaheuristic algorithms have been proposed to solve the CF problems such as genetic algorithms [20, 21] and simulated annealing [22, 23]. Related comprehensive reviews can be found in Papaioannou and Wilson [24] and Renzi et al. [25]. While metaheuristic algorithms have capacities to yield high-quality solutions, they generally require users to have good mathematical skills to understand these algorithms [26] and good experiences to make some "implementation decisions" [15, p. 293] (e.g., terminating conditions in genetic algorithms).

In contrast to metaheuristic algorithms, heuristic algorithms are easier to implement but the quality of their solutions is often targeted [27, p. 159]; [24]. In a nutshell, a common feature of heuristic algorithms is their greedy or hill-climbing approaches that focus on best solutions at a stage without backtracking for other solution possibilities. This feature allows them to converge to some feasible solutions quickly with the trade-off of checking a smaller solution space (thus, potentially weaker solution quality). Hierarchical clustering (HC), which was one early approach for CF problems [11], is one example of heuristic algorithms since HC always groups the object pairs with the highest similarity values progressively without backtracking.

As its third property, it is observed that a well-structured matrix can be obtained relatively easily by a heuristic approach (referred to HC specifically in this paper), where the metaheuristic approach does not necessarily have an advantage for getting higher-quality solutions. Alternately, the advantage of the metaheuristic approach is observed more often in the case of ill-structured matrices. As discussed before, a well-structured matrix demonstrates sharp differences between similar and dissimilar machine pairs. This feature supports the "greedy" nature of the heuristic approach, which can easily distinguish high-similarity pairs in the progressive grouping process. In contrast, an ill-structured matrix has more machine pairs with middle-similarity values so that some borderline cases can potentially lead to solutions of lower quality. While this third property may not be obvious, more verifying examples will be reported later in Section 6.3 as part of the investigation effort of this paper.

Given this third property of a well-structured matrix, the statistical analysis of similarity values can then lead to another application, i.e., supporting the choice of the algorithmic approach for solving CF problems. If the statistical analysis shows a high potential to obtain a well-structured matrix, we can choose a heuristic approach to solve the CF problems. Alternately, if it indicates a high chance of getting an ill-structured matrix, we may consider revising the input incidence matrix (e.g., adding more machines or changing some part requirements). Also, we can prepare to use the metaheuristic approach to seek for high-quality solutions. In sum, the statistical analysis can preliminarily probe the structure of a given CF problem in order to determine the next problem solving step.

2.3. Research Plan.

In view of the three properties of a well-structured matrix discussed above, the research and development questions are set as follows.

(i) What are the criteria related to the statistics of similarity values to assess the potential of getting a well-structured matrix?

(ii) How do we decide on whether using a metaheuristic or heuristic approach for solving a CF problem?

To address the first question, this paper will utilize two statistical tools: histogram and the Kolmogorov-Smirnov (K-S) test. Histogram will be used to analyze the distribution of machine similarity values of a given CF problem, and twenty CF solutions will be set to investigate the threshold values for informing the potential structuredness of a matrix. The K-S test will be used to assess the normality of the distribution of machine similarity values. That is, if the set of similarity values roughly follow the normal distribution, it means that many machine pairs have the average similarity value, implying a low proportion of high-similarity values (i.e., an ill-structured matrix).

Based on the investigation using the histogram and the K-S test, we will develop a procedure to probe the structure of a given CF matrix and suggest whether using a metaheuristic or heuristic for problem solving (i.e., address the second question). In this paper, we have implemented
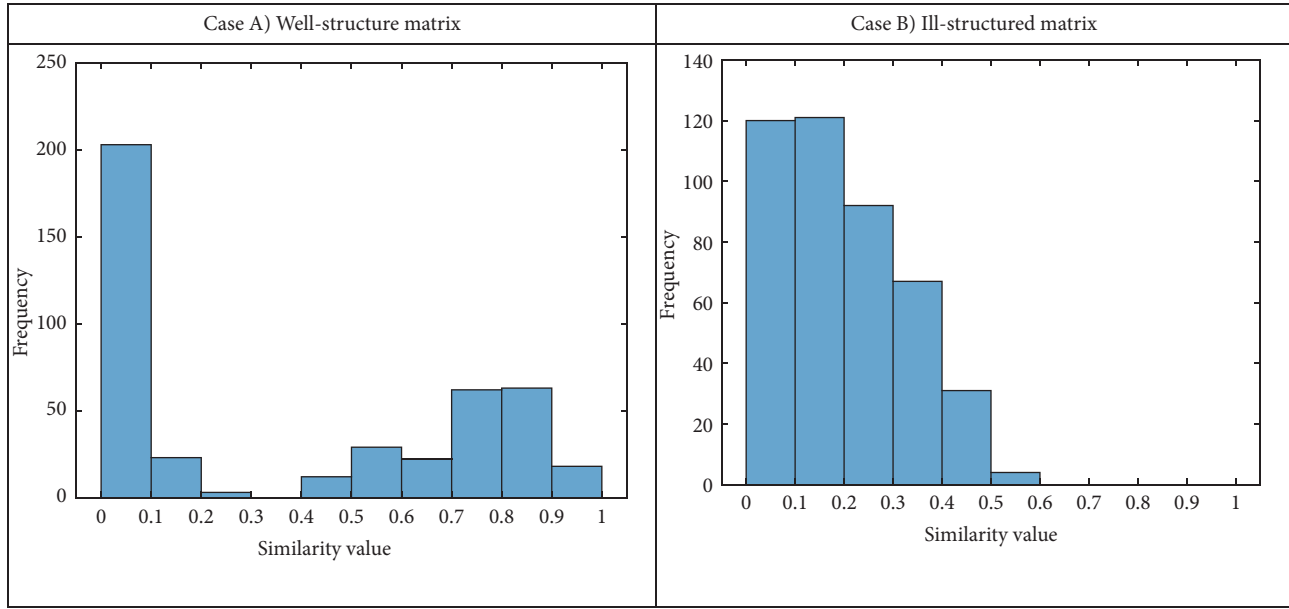
FIGURE 2: Histograms of well-structured and ill-structured matrices.

genetic algorithm (GA) and hierarchical clustering (HC) as the metaheuristic and heuristic approaches, respectively, for solving the CF problems. To verify the procedure, additional forty CF matrices will be set. These CF matrices will be solved by HC and then genetic algorithm to observe the relation between the matrix's structuredness and the utility of the metaheuristic approach for better CF solutions.

3. Histogram Analysis of Similarity Values

3.1. Histogram and the U-Shape. In this study, histograms are used to report the frequency distribution of machine similarity values with an increment of 0.1. Figure 2 shows two histograms for the well-structured and ill-structured matrices of Figure 1, respectively. In these histograms, the horizontal axis stands for the machine similarity values ranging from 0 to 1, and the vertical axis stands for the number of machine pairs within those ranges of similarity values. Notably, these histograms are independent of the orders of a matrix's rows and columns. That is, we can get these histograms of similarity values without solving the CF problem.

From these two histograms, it is observed that a well-structured matrix tends to yield an U-shape histogram, i.e., relatively high numbers of extreme similarity values. The right peak of the U-shape can be explained by the property of high percentage of high-similarity machine pairs discussed in Section 2.2.2. While the numbers of low-similarity machine pairs are high in both cases of well-structured and ill-structured matrices, a well-structured matrix has a low number of machine pairs of similarity values between 0.2 and 0.4. In contrast, an ill-structured matrix has a good number of those middle-similarity machine pairs, which cause a challenge of clear grouping in cell formation. Given this general U-shape observation, the next subsections will discuss the criteria that classify the structuredness of a matrix

(i.e., well-structured or ill-structured) based on the histogram data.

3.2. Setup of 20 Benchmark Matrices. Since the frequency distribution of a histogram will not be altered by the orders of a matrix's rows and columns, we can set the CF solution matrices with known structuredness and then observe their histograms to develop the structuredness criteria. In this investigation, twenty 30×40 solution matrices (i.e., 30 machines and 40 parts) with three cells (or blocks) are set. These matrices are varied by two factors: (1) block sizes and (2) numbers of exceptional elements and voids. Concerning the block sizes, five cases are set as follows, where each bracket indicates the size of a block as (number of machines × number of parts).

(i) Case A ⟶ even case: (10×13) (10×13) (10×14)

(ii) Case B ⟶ uneven case with a large block: (20×26) (5×7) (5×7)

(iii) Case C ⟶ uneven numbers of machines and parts in two blocks: (20×7) (5×26) (5×7)

(iv) Case D ⟶ uneven numbers of machines and parts in three blocks: (20×5) (5×17) (5×18)

(v) Case E ⟶ uneven case with two large blocks: (14×18) (14×19) (2×3)

Besides, four cases are set below to characterize the structuredness of matrices via the control of the numbers of exceptional elements and voids.

(i) Case I (well-structured): few exceptional elements and no voids

(ii) Case II (well-structured): no exceptional elements and few voids
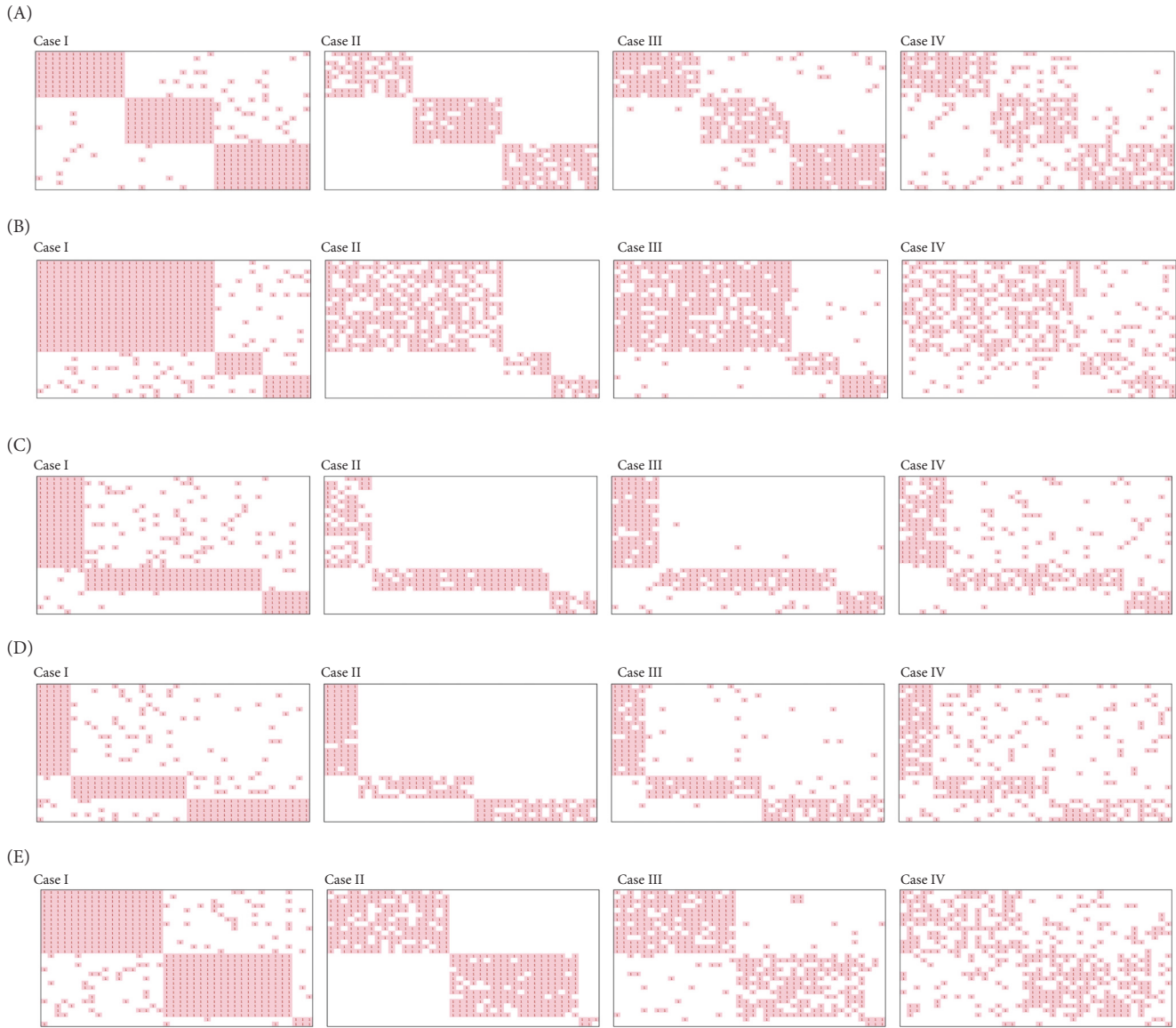
FIGURE 3: The resulting matrices of 20 benchmark cases.

(iii) Case III (well-structured): few exceptional elements and few voids

(iv) Case IV: (ill-structured): good numbers of exceptional elements and voids

The resulting 20 matrices are shown in Figure 3. As general inspections, the matrices in Cases I and II have clear boundaries of three cells. The matrices in Case III have more exceptional elements and voids but their structures are still quite discernible. In contrast, the structure of matrices in Case IV is messier with higher numbers of exceptional elements and voids. Based on these matrices, the next subsection will investigate their histograms and develop the U-shape criteria to classify the matrix's structuredness.

3.3. Histogram-Based U-Shape Criteria.
To inform the matrix's structuredness, two conditions as the U-shape

criteria are set toward the low and high-similarity values. Let $F_{left}(x)$ be the fraction of similarity values that are lower than x and $F_{right}(y)$ be the fraction of similarity values that are higher than y. Then, the general U-shape criteria can be expressed as follows.

$$F_{left}(x) \geq a \tag{3}$$

$$F_{right}(y) \geq b \tag{4}$$

where a and b are the thresholds of the minimum fractions of low and high-similarity values, respectively, to characterize the U-shape of a well-structured matrix. The setup of these parametric values (i.e., x, y, a, and b) will be based on the above 20 benchmark matrices.

Figure 4 shows the histograms of the 20 benchmark matrices. As the preliminary observations, the frequency
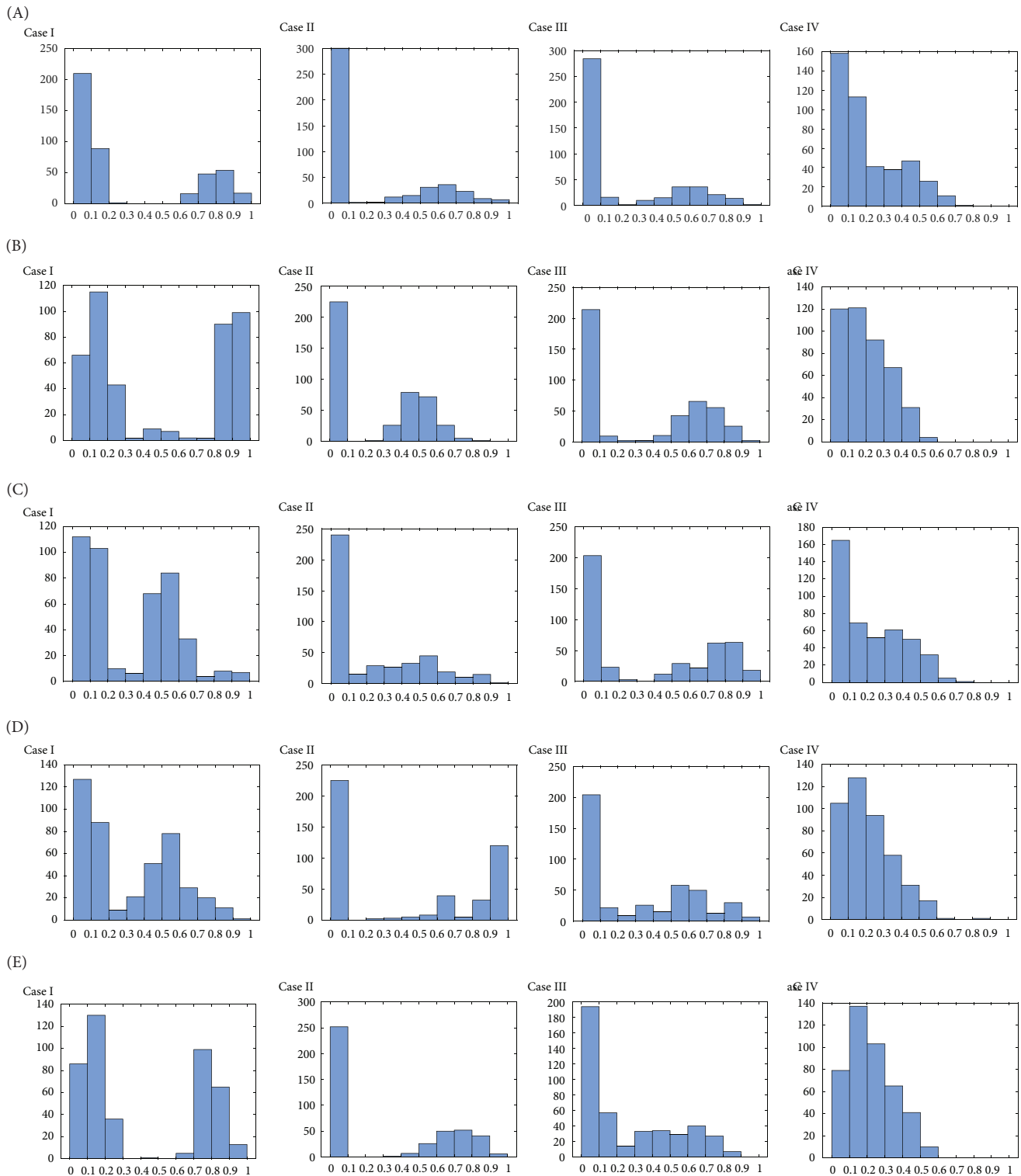
FIGURE 4: Histograms of 20 benchmark matrices.

distributions of these histograms are perceived quite different between the well-structured (i.e., Cases I, II, and III) and ill-structured matrices (i.e., Case IV). Yet, some U-shapes are not plainly obvious (e.g., Cases A-III and C-II), and the peaks of high-similarity values of the well-structured matrices are not

located at the rightmost region (e.g., Cases C-I and D-I). The U-shape criteria will then be set based on these observations.

Concerning the region of low-similarity values (i.e., the left side of the U-shape), it is found that both well-structured (i.e., Cases I, II, and III) and ill-structured (i.e., Case IV)

TABLE 1: Number of machine pairs with similarity values equal to zero.

	Case I	Case II	Case III	Case IV
A	34	300	110	32
B	7	225	83	23
C	23	240	121	57
D	22	225	116	15
E	11	252	73	14

TABLE 2: Number of machine pairs with similarity values greater than or equal to 0.5.

	Case I	Case II	Case III	Case IV
A	135	106	109	38
B	200	104	194	4
C	136	91	194	38
D	139	203	158	19
E	182	175	103	10

matrices have high proportions because many machines, as long as they are not in the same cell, have less common parts to work with in both cases. As a result, the proportions of low-similarity values from a well-structured matrix can become less discernible statistically. Thus, we choose to investigate the extreme value when the similarity values equal to zero, i.e., $F_{left}(x=0)$. Table 1 records the number of machine pairs with the similarity values equal to zero. As observed, while the matrices of Cases II and III have low right-side peaks, they have high proportions of such zero-similarity machine pairs. As the U-shape criteria will be used for the early screening, we set this criterion rather strictly as follows.

$$F_{left}(0) \geq 0.5 \qquad (5)$$

This criterion requires 50% of machine pairs to have zero-similarity values in order to qualify a well-structured matrix. By checking the benchmark matrices with 30 machines (i.e., 435 machine pairs), the threshold is 218 machine pairs, and the matrices in Case II pass this criterion.

Concerning the region of high-similarity values (i.e., the right side of the U-shape), as discussed earlier, not all well-structured matrices have high proportions of high-similarity values at the rightmost region. By inspecting the histograms in Figure 4, we identify a reasonable cut-off of high-similarity values should be 0.5, i.e., $F_{right}(y=0.5)$. Table 2 records the number of machine pairs with the similarity values greater than or equal to 0.5. As observed, the proportions of high-similarity values ($s_{xy} \geq 0.5$) in Case IV (i.e., ill-structured matrices) are relatively low. In contrast, Case C-II is the well-structured matrix with the lowest number of high-similarity values (i.e., 91), and the corresponding fraction is $91/435 \approx 0.21$. As a result, another U-shape criterion for the right-hand side is set as follows.

$$F_{right}(0.5) \geq 0.2 \qquad (6)$$

In sum, if an input incidence matrix satisfies one of the two U-shape criteria formulated in (5) and (6), this matrix has a good chance to yield a well-structured CF solution.

Notably, we treat the histogram-based U-shape criteria as a preliminary filter in this work. That is, if a matrix does not satisfy these criteria, it does not immediately imply that this matrix is ill-structured. In fact, other parameters of an input incidence matrix, such as the number of machines and the density of nonzero matrix entries, can impact the frequency distribution of a histogram. Thus, the next section will develop another criterion based on the K-S test.

4. Criterion Setting Based on the Kolmogorov-Smirnov (K-S) Test

4.1. Background. The Kolmogorov-Smirnov (K-S) test is one type of hypothesis testing in statistics (Corder and Foreman) [28]. As one of its applications, the K-S test is used in this paper to evaluate how well a dataset represents a normal distribution (i.e., the normality of the dataset). The use of the K-S test in this study is mainly motivated by the observation of the histograms in Figure 2 that a well-structure matrix will tend to give a U-shape. As the U-shape will generally exhibit two peaks in the histogram representation, the normality of the associated data (i.e., similarity values) will be weak in comparison to that of an ill-structured matrix.

Figure 5 illustrates the concept of the normality of similarity values with two cases: single-peak histogram and U-shape histogram. The K-S test essentially compares the curves of two cumulative distribution functions (CDFs) [29, 30]. While one CDF represents the empirical data points (i.e., empirical CDF, solid line), another CDF is based on the normal distribution curve fitted by the empirical data (i.e., hypothesized normal CDF, dashed line). As seen in Figures 5(c) and 5(d), the single-peak histogram has higher normality than the U-shape histogram since the single-peak histogram yields a closer match between the empirical and hypothesized normal CDFs. In contrast, the U-shape histogram yields its empirical CDF in Figure 5(d) with rapid increases at the beginning and the end, along with a relatively flat region in the middle, and this CDF curve significantly deviates from normality [31].

(a) A single-peak histogram

(b) A U-shape histogram

(c) CDFs of the single-peak histogram

(d) CDFs of the U-shape histogram
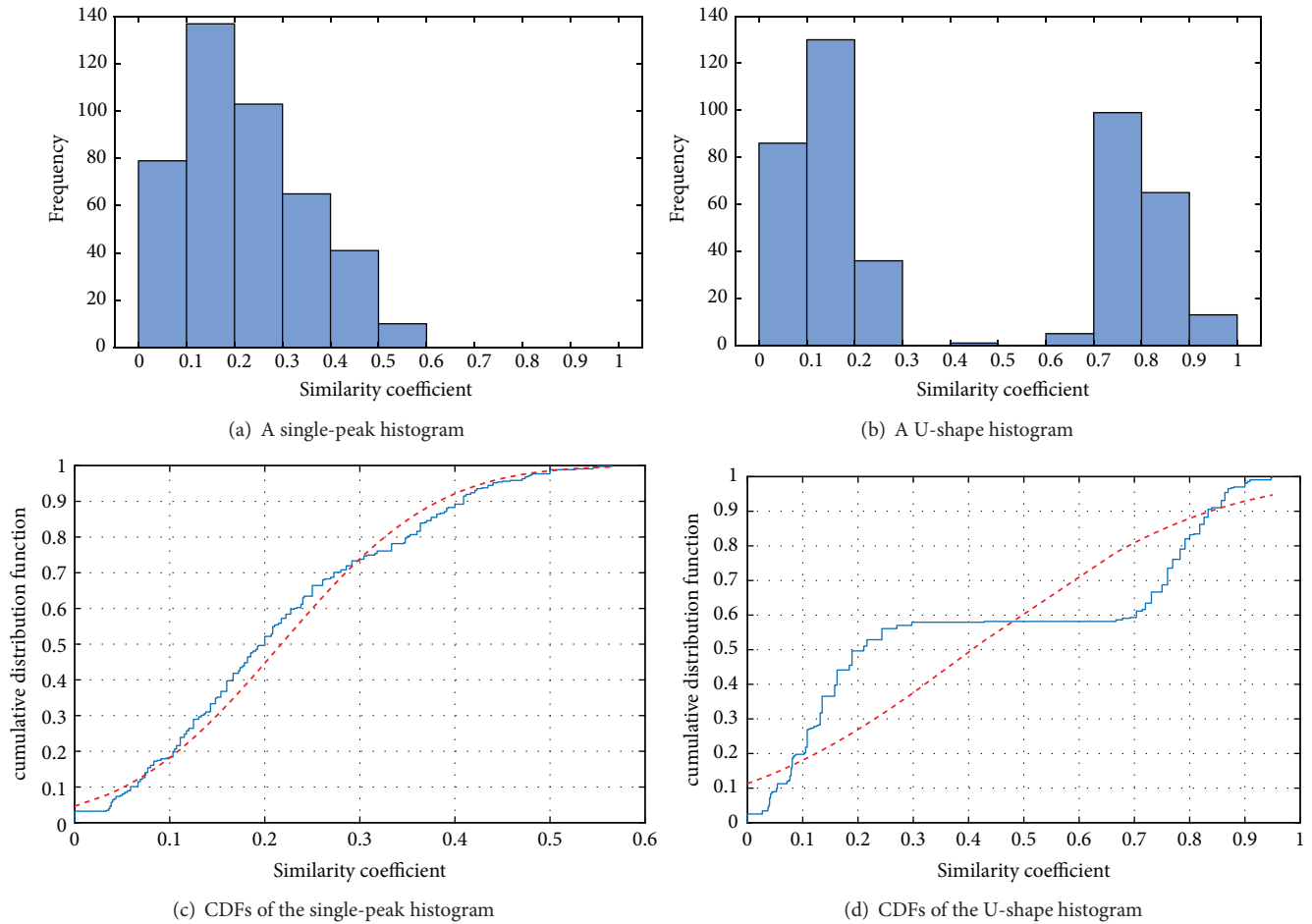
FIGURE 5: CDFs of similarity values (solid line: empirical CDF; dashed line: hypothesized normal CDF).

The P value is a common concept in hypothesis testing [32]. It can be interpreted as the smallest probability value associated with a given dataset to reject the null hypothesis (i.e., smaller P value $\longrightarrow$ more likely to reject the null hypothesis). In this work, we treat the P value of a K-S test as a proxy measure on the normality of a set of similarity values. That is, if the P value is smaller, the dataset tends to be *less*-normal [33]. Interpreted in our context, a *less*-normal condition implies a U-shape and thus a well-structured matrix. For example, the P value of the single-peak histogram in Figure 5(c) is 7.44×10^{-4}, and the P value of the U-shape histogram in Figure 5(d) is 9.27×10^{-22}.

Notably, the purpose of using the K-S test in this work is not about hypothesis testing, but only using its P value as a proxy measure to assess the normality of a set of similarity values and then inform the structuredness of a CF matrix. Yet, the P values in our applications tend to be very small. To conveniently handle this proxy measure, let P_{value} be the P value of a set of similarity values based on the K-S test, and an alternative proxy measure (denoted as L_p) is defined as follows:

$$L_p = -\log_{10} P_{value} \qquad (7)$$

As L_p is the negative logarithm of the P value, a higher value of L_p implies a higher tendency of having a U-shape of the dataset. For example, the values of L_p for the single-peak histogram (i.e., Figure 5(c)) and the U-shape histogram (i.e., Figure 5(d)) are 3.13 and 21.03, respectively. In other words, if a CF matrix yields a higher value of L_p, it has a better chance to be solved as a well-structured CF solution.

By knowing the property of the trend associated with L_p, it leads to the next investigation question on setting the threshold value of L_p to classify ill-structured and well-structured matrices. To do so, it is recognized that the values of L_p can be sensitive to the number of machines and the density of nonzero entries of a given matrix. Thus, the next subsection will investigate the upper bound of L_p of a given matrix to normalize the value of L_p. Then, we will apply the 20 benchmark matrices in Figure 3 to determine the threshold.

4.2. Estimate the Upper Bound of L_p for Normalization. The upper bound of L_p can be estimated by a perfect block-diagonal matrix, where the numbers of exceptional elements (n_{out}) and voids (n_{in}) are zero (i.e., the grouping efficacy $\mu = 1$). In this case, the machine pairs have similarity values equal to either one (when two machines belong to the same block)

TABLE 3: L_p, L_{bp} and their ratios of the benchmark matrices.

	Case I			Case II		
	L_p	L_{bp}	L_p/L_{bp}	L_p	L_{bp}	L_p/L_{bp}
Case A	43.44	62.39	0.70	69.58	85.40	0.81
Case B	22.03	31.35	0.70	44.35	72.93	0.61
Case C	14.80	70.55	0.21	42.57	97.73	0.44
Case D	13.13	75.90	**0.17**	43.58	96.69	0.45
Case E	23.06	39.67	0.58	53.56	65.65	0.82
	Case III			Case IV		
	L_p	L_{bp}	L_p/L_{bp}	L_p	L_{bp}	L_p/L_{bp}
Case A	39.67	76.05	0.52	10.77	73.53	**0.15**
Case B	26.84	54.38	0.49	3.77	72.19	0.05
Case C	23.06	88.52	0.26	7.69	86.15	0.09
Case D	18.77	91.49	0.21	3.13	83.92	0.04
Case E	17.56	67.29	0.26	1.97	67.14	0.03

or zero (when two machines are in different blocks). This kind of "bipolar" distribution can be viewed as a far extreme of the normal distribution, and the corresponding P value can be taken as the upper bound of L_p.

In the normalization process, we can first identify the size and the number of nonzero entries of a given matrix. Let m and n be the numbers of machines and parts, respectively, as the size of the matrix. The number of nonzero matrix entries has been denoted as n_e. Then, the density of nonzero entries of a matrix (denoted as D_s) can be determined as follows.

$$D_s = \frac{n_e}{m \times n} \qquad (8)$$

Given an incidence matrix, its upper bound of L_p can be considered in a case when its nonzero entries can be freely moved to form a nearly perfect block-diagonal matrix. By fixing the values of m, n, and D_s, there can be a corresponding theoretical upper bound of L_p. Let L_{bp} denote such an upper bound of L_p of a given matrix. Then, for any given matrix, we can determine its L_p and L_{bp}, where L_{bp} is treated as a normalizing factor. Since this paper focuses on machine similarity, we drop the consideration of n to simplify the investigation. Then, the next step is to determine the following function.

$$L_{bp} = f(m, D_s) \qquad (9)$$

To estimate the function of L_{bp}, our strategy is to systematically generate a good number of perfect block-diagonal matrices by varying the numbers of machines, parts, and even-size cells (note: the number of even-size cells will determine the number of nonzero entries). The ranges of these varying parameters in this work are listed as follows.

(i) Number of machines: from 10 to 50 machines

(ii) Number of parts: from 10 to 110 parts (with an increment of 10)

(iii) Number of even-size cells: from 2 to 14 cells (also restricted by the matrix's size to avoid extremely large and small cells)

Further details of the setup of these perfect matrices can be found in Zhu [34]. As a result, this work has generated 2519 perfect matrices. Then, the values of P value and L_p are determined for these matrices, giving 2519 points to approximate the function formulated in (9) via curve fitting techniques. The resulting regression equation is found as follows.

$$L_{bp}(m, D_s) = -29.08 + 2.164m + 132.3D_s$$
$$+ 0.1049m^2 - 10.35mD_s \qquad (10)$$

In practice, we can determine the values of L_p via (7) and L_{bp} via (10) for a given matrix. Then, we can check its ratio of L_p to L_{bp} and examine the U-shapeness and then the possible structuredness of the matrix. The next subsection will discuss the criterion based on the ratio of L_p to L_{bp}.

4.3. Ratio Criterion Based on L_p and L_{bp}. The setting of the ratio threshold for L_p and L_{bp} is based on the 20 benchmark matrices in Figure 3. The values of L_p, L_{bp} and their ratios are recorded in Table 3. As a recall, Cases I, II, and III are set to represent the well-structured matrices, and Case IV represents ill-structured matrices. As an initial assessment, the average of the ratios of Cases I, II, and III (i.e., well-structured matrices) is 0.48, while the ratio average of Case IV is 0.07. This observation indicates that the ratio L_p/L_{bp} can make distinctions between well-structured and ill-structured matrices quite effectively from a statistical standpoint.

Yet, when we examine the extreme situations, the lowest ratio of the well-structured cases is 0.17 (i.e., Case D-I, bold in Table 3), and the highest ratio of the ill-structured cases is 0.15 (i.e., Case A-IV, also bold in Table 3). As observed, the gap between the two is close, and we intend to impose a tight criterion to classify well-structured matrices. As a result, we set the threshold value at 0.2, formulated as follows.

$$\frac{L_p}{L_{bp}} \geq 0.2 \qquad (11)$$

At this point, Case D-I is the only well-structured matrix that does not satisfy this criterion. Yet, Case D-I satisfies one of
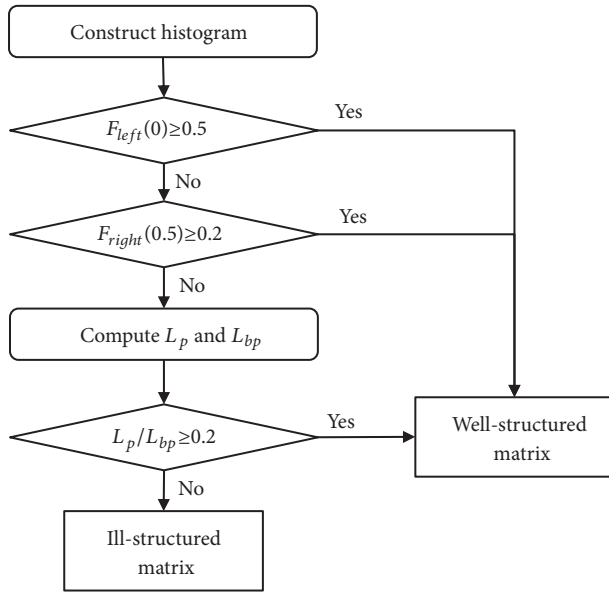
FIGURE 6: Procedure to assess the potential structuredness of an incidence matrix.

the earlier U-shape criteria. Thus, our next step is to combine the U-shape criteria and the ratio criterion in a procedure to examine the potential structuredness of an incidence matrix. That is, if a given matrix satisfies one of these criteria, it is indicated that this matrix has a high potential to yield a well-structured CF solution. The next section will discuss this procedure to apply these criteria to inform the potential structuredness of a given matrix.

5. Procedure

This section provides a four-step procedure below to assess the potential structuredness of an incidence matrix using the histogram-based U-shape criteria and the criterion based on the P value of the K-S test. Figure 6 illustrates the decision branches of this procedure.

Step 1 (construct histogram). By receiving an incidence matrix as an input, the similarity values of machine pairs are first determined based on (2). If there are m machines, there will be $m \times (m\text{-}1)/2$ machine pairs with their similarity values, forming the dataset of the statistical analysis. A histogram is then constructed to analyze these similarity values.

Step 2 (apply the histogram-based U-shape criteria). This represents the preliminary check based on the frequencies of having high and low-similarity values. If either one of the criteria $F_{left}(0) \geq 0.5$ or $F_{right}(0.5) \geq 0.2$ is satisfied, the incidence matrix is considered having a good potential to yield a well-structured CF solution. If none of these two criteria is satisfied, we will move on to the analysis based on the P value of the K-S test.

Step 3 (compute L_p and L_{bp}). The dataset of similarity values is treated as the input to determine the P value of the K-S

test in view of assessing the normality of the dataset. This calculation can be performed via some statistics software tools. In this work, we have used the statistics functions from Matlab to compute the P value. Then, the value of L_p can be evaluated using (7). With the incidence matrix, the value of L_{bp} can be evaluated using (10) by identifying the number of machines (i.e., m) and the density of nonzero entries (i.e., D_s).

Step 4 (apply the ratio criterion L_p / L_{bp}). With the values of L_p and L_{bp}, we can check the criterion if L_p / $L_{bp} \geq 0.2$. If this criterion is satisfied, the input matrix should have a good potential to yield a well-structured CF solution. If not, the input matrix would have a good chance to result in an ill-structured CF solution. The practitioners may consider modifying the input matrix by adding machines or revising the production requirements.

6. Application and Verification

To examine the statistical analysis of similarity values for CF problems in this paper, other 40 matrices (in addition to the earlier 20 benchmark matrices, making up a total of 60 matrices) will be generated and applied in this section. These 60 matrices will be used to examine the following two issues specifically.

(i) Given the three criteria for assessing the potential structuredness of a matrix, we are going to use these 60 matrices to examine their effectiveness to distinguish well-structured and ill-structured matrices.

(ii) While Property III (i.e., relative ease of obtaining satisfactory CF solutions) of a well-structured matrix has been discussed in Section 2.3, it will be verified via these 60 matrices by two stages of CF problem solving.

6.1. Setup of the 60 Incidence Matrices. The strategy to generate 60 matrices is based on the extension of getting the 20 benchmark matrices in Section 3.2. The additional varying factors include the following.

(i) In addition to the size of 30×40 matrix, another size of 40×100 matrix is set.

(ii) We add cases with more numbers of cells (from 3 to 6, 8, and 12 cells)

(iii) The evenness of cell sizes is also varied for each case.

Table 4 shows the setup of 60 matrices, where Cases A and E are repeated from Section 3.2 for comparison. Notably, the structuredness of matrices, which were classified as Cases I, II, III, and IV in Section 3.2, is also applied, leading to the study of 15×4 = 60 incidence matrices. As the intention of the setup, the matrices of Cases I and II have no voids and exceptional elements, respectively. Then, they should be classified as well-structured matrices. The matrices of Case III have only few exceptional elements and voids, and they should also be classified as well-structured matrices. In contrast, the matrices of Case IV have more exceptional elements and voids, and they should be classified as ill-structured matrices. The images and histograms of these 60

TABLE 4: Setup of incidence matrices.

	Matrix size	No. of cells	Cell sizes
Case A	30×40	3 cells	(10×13) (10×13) (10×14)
Case B	30×40	3 cells	(20×26) (5×7) (5×7)
Case C	30×40	3 cells	(20×7) (5×26) (5×7)
Case D	30×40	3 cells	(20×5) (5×17) (5×18)
Case E	30×40	3 cells	(14×18) (14×19) (2×3)
Case F	30×40	6 cells	(5×7) (5×7) (5×7) (5×7) (5×7) (5×5)
Case G	30×40	6 cells	(15×20) (3×4) (3×4) (3×4) (3×4) (3×4)
Case H	30×40	6 cells	(15×4) (3×20) (3×4) (3×4) (3×4) (3×4)
Case I	30×40	6 cells	(15×5) (3×7) (3×7) (3×7) (3×7) (3×7)
Case J	40×100	8 cells	(5×13) (5×13) (5×13) (5×13) (5×12) (5×12) (5×12) (5×12)
Case K	40×100	8 cells	(8×8) (8×8) (4×14) (4×14) (4×14) (4×14) (4×14) (4×14)
Case L	40×100	8 cells	(7×18) (7×18) (7×18) (7×17) (6×17) (2×4) (2×4) (2×4)
Case M	40×100	12 cells	(3×8) (3×8) (3×8) (3×8) (3×8) (3×8) (3×8) (4×9) (4×9) (4×9) (4×9)
Case N	40×100	12 cells	(5×5) (5×5) (4×5) (3×10) (3×10) (3×10) (3×10) (3×10) (2×9)
Case O	40×100	12 cells	(4×12) (4×12) (4×12) (4×11) (4×11) (2×2) (2×2) (2×2) (2×2)

matrices are provided as supplementary materials (available here).

6.2. Examination of the Criteria. To evaluate the effectiveness of the criteria to assess the structuredness of the matrices, we have evaluated the criteria values for the 60 matrices. The results are provided in Table 5, where the values satisfying the criteria of well-structured matrices are bold. As observed in these results, the structuredness criteria can discern the well-structured matrices of Cases I, II, and III, where each matrix there satisfies at least one criterion. In contrast, no matrices of Case IV satisfy any criteria of well-structured matrices.

In view of the effectiveness of individual criteria, it is observed that $F_{left}(0)$ is effective in filtering the matrices of Case II (i.e., few voids and no exceptional elements). Due to the absence of exceptional elements in this case, any two machines of different blocks will have similarity values equal to zero. This explains the high values of $F_{left}(0)$ observed in Case II. In contrast, $F_{right}(0.5)$ is less effectiveness when the matrices have more cells (e.g., Cases H and I) and large sizes (e.g., Cases J to O). Notably, the values of $F_{right}(0.5)$ for Case IV are quite low (ranging from 0.00 to 0.09). In this view, the criterion of $F_{right}(0.5)$ is quite tight.

By comparison, the ratio criterion (i.e., L_p/L_{bp}) seems effective in distinguishing well-structured matrices, where Case D-I is the only case not identified as a well-structured matrix by this criterion only. Notably, the discernible gap of well-structured matrices (lowest at 0.17 in Case D-I) and ill-structured matrices (highest 0.16 in Case L-IV) is small. It explains the need of having $F_{left}(0)$ and $F_{right}(0.5)$, along with the ratio criterion, in the assessment of the structuredness of the matrices.

6.3. Examination of Property III via Optimization. As a recall from Section 2.2.3, Property III states that a well-structured matrix can be fairly obtained via a heuristic approach, where more complex metaheuristics may not bring in additional benefits. To verify this property, the sixty matrices were tested with a two-stage solution process. First, each matrix will be solved by a hierarchical clustering (HC) method as one heuristic to yield a CF solution. Then, we examine if we can further optimize the obtained CF solution via the genetic algorithm (GA), representing a metaheuristic method. In this way, we can check the correlation between grouping efficiency and the percentage of improvement of solution quality by GA. The algorithmic details of the HC method and the implementation details of GA applied in this study can be found in Zhu [34].

Table 6 lists the grouping efficacy (μ) results for the 60 matrices after running hierarchical clustering (HC) and then genetic algorithm (HC+GA). Also, the percentages of improvement in view of grouping efficacy by GA are reported for comparison. As observed, the matrix solutions in Cases I and II cannot be further improved by GA, while three matrix solutions in Case III can be improved by GA with small percentages (between 0.20% and 0.25%). In contrast, the ill-structured matrix solutions in Case IV can be improved by GA in the percentages of improvement between 0.63% and 22.69%. Overall, we consider that the numerical results
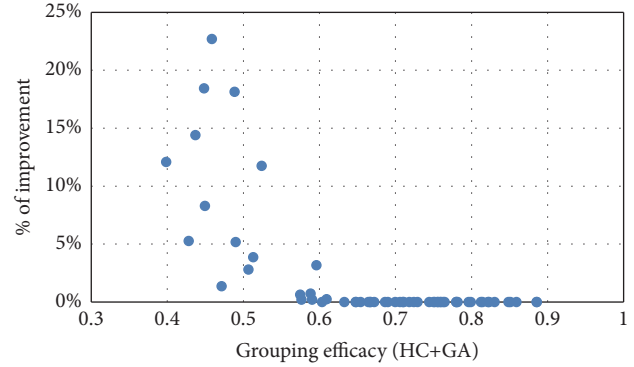
FIGURE 7: Percentage of solution improvement versus grouping efficacy.

generally follow Property III, given that the matrices in Case III are close to the boundary between well-structured and ill-structured matrices.

Figure 7 shows the plots of the percentages of solution improvement versus the values of grouping efficacy based on HC+GA. Based on the 60 matrices studied in this paper, GA did not improve the quality of matrix solutions that have 0.60 or higher grouping efficacy. For the data points of grouping efficacy values less than 0.60, we find that these data points are negatively correlated, where the correlation value [32, p. 173] is -0.62. In the statistical interpretation, we can state that a lower value of grouping efficacy tends to allow a larger room of improvement by GA but its linearity is not strong. Notably, the capabilities of HC and GA to yield high-quality solutions can depend on other factors (e.g., density of nonzero entries in a matrix). Thus, it is not easy to observe a linear correlation just between the percentage of improvement and the grouping efficacy. More control factors and samples should be required for an in-depth investigation.

7. Conclusions

This paper has explored the statistics of similarity values to investigate the structuredness of cell formation (CF) matrix solutions. Using grouping efficacy (μ) as one recognized index to inform the quality of a CF matrix, it is found that a well-structured matrix has a high percentage of high-similarity machine pairs (i.e., Property II). Accordingly, this paper sets up 20 benchmark matrices, with varying structuredness, to develop the U-shape criteria and the criterion based on the Kolmogorov-Smirnov test. Then, a procedure is developed to assess the potential structuredness of a CF matrix without solving the CF problem. The criteria for assessing structuredness of matrices are examined via additional 40 matrices, and agreeable results are observed. Genetic algorithm (GA) is used to see if it can improve the CF solutions obtained by hierarchical clustering (as one type of heuristics). The results show that the matrix solutions with high grouping efficacy values (i.e., well-structured matrices) cannot be effectively improved by GA.

While the worst-case computational complexity of clustering problems (e.g., NP hardness) is well recognized, the

TABLE 5: Results of the criteria values.

	Case I			Case II			Case III			Case IV		
	F_{left}	F_{right}	L_p/L_{bp}	F_{left}	F_{right}	L_p/L_{bp}	F_{left}	F_{right}	L_p/L_{bp}	F_{left}	F_{right}	L_p/L_{bp}
A	0.08	**0.31**	**0.70**	**0.69**	**0.24**	**0.81**	0.25	**0.25**	**0.52**	0.08	0.09	0.15
B	0.02	**0.46**	**0.70**	**0.52**	**0.24**	**0.61**	0.19	**0.45**	**0.49**	0.05	0.01	0.05
C	0.05	**0.31**	**0.21**	**0.55**	**0.21**	**0.44**	0.28	**0.45**	**0.26**	0.13	0.09	0.09
D	0.05	**0.32**	0.17	**0.52**	**0.47**	**0.45**	0.27	**0.36**	**0.21**	0.03	0.04	0.04
E	0.03	**0.42**	**0.58**	**0.58**	**0.40**	**0.82**	0.17	**0.24**	**0.26**	0.03	0.02	0.03
F	**0.51**	0.14	**0.39**	**0.86**	0.11	**0.92**	0.43	0.10	**0.36**	0.26	0.03	0.14
G	0.18	**0.26**	**0.41**	**0.72**	0.14	**0.80**	0.26	**0.26**	**0.39**	0.18	0.07	0.14
H	0.34	0.14	**0.27**	**0.72**	**0.23**	**0.66**	0.43	0.18	**0.27**	0.23	0.03	0.08
I	0.18	0.11	**0.23**	**0.72**	**0.22**	**0.68**	0.34	0.15	**0.23**	0.19	0.01	0.06
J	0.29	0.10	**0.43**	**0.90**	0.09	**0.95**	**0.53**	0.06	**0.47**	0.14	0.00	0.10
K	0.21	0.08	**0.34**	**0.88**	0.08	**0.91**	0.17	0.03	**0.25**	0.08	0.00	0.07
L	0.17	0.13	**0.45**	**0.87**	0.12	**0.96**	0.30	0.04	**0.28**	0.10	0.02	0.16
M	0.37	0.04	**0.31**	**0.94**	0.05	**0.94**	0.41	0.02	**0.27**	0.27	0.00	0.13
N	0.40	0.04	**0.28**	**0.93**	0.06	**0.94**	0.42	0.02	**0.26**	0.24	0.00	0.08
O	0.30	0.06	**0.30**	**0.93**	0.06	**0.97**	0.48	0.04	**0.34**	0.14	0.00	0.08

TABLE 6: Grouping efficacy results with the two-stage solution process.

	Case I			Case II		
	HC	HC+GA	% Improve	HC	HC+GA	% Improve
Case A	0.7817	0.7817	0	0.7550	0.7550	0
Case B	0.8859	0.8859	0	0.6542	0.6542	0
Case C	0.7587	0.7587	0	0.7180	0.7180	0
Case D	0.7514	0.7514	0	0.8218	0.8218	0
Case E	0.8590	0.8590	0	0.8302	0.8302	0
Case F	0.8230	0.8230	0	0.7800	0.7800	0
Case G	0.8511	0.8511	0	0.6861	0.6861	0
Case H	0.7059	0.7059	0	0.7500	0.7500	0
Case I	0.6475	0.6475	0	0.7444	0.7444	0
Case J	0.7645	0.7645	0	0.7600	0.7600	0
Case K	0.7106	0.7106	0	0.7241	0.7241	0
Case L	0.7561	0.7561	0	0.8122	0.8122	0
Case M	0.6720	0.6720	0	0.7798	0.7798	0
Case N	0.6327	0.6327	0	0.8484	0.8484	0
Case O	0.6644	0.6644	0	0.8854	0.8854	0

	Case III			Case IV		
	HC	HC+GA	% Improve	HC	HC+GA	% Improve
Case A	0.7627	0.7627	0	0.5776	0.5959	3.17%
Case B	0.7961	0.7961	0	0.4650	0.4713	1.35%
Case C	0.8142	0.8142	0	0.5842	0.5884	0.72%
Case D	0.7290	0.7290	0	0.4938	0.5129	3.87%
Case E	0.6898	0.6898	0	0.4927	0.5065	2.80%
Case F	0.7097	0.7097	0	0.4689	0.5240	11.75%
Case G	0.7990	0.7990	0	0.5711	0.5747	0.63%
Case H	0.6667	0.6667	0	0.4134	0.4884	18.14%
Case I	0.6481	0.6481	0	0.4067	0.4281	5.26%
Case J	0.6995	0.6995	0	0.3785	0.4483	18.44%
Case K	0.6035	0.6035	0	0.3819	0.4369	14.40%
Case L	0.6079	0.6094	0.25%	0.4657	0.4898	5.18%
Case M	0.5753	0.5765	0.21%	0.3738	0.4586	22.69%
Case N	0.5891	0.5903	0.20%	0.3556	0.3986	12.09%
Case O	0.6673	0.6673	0	0.4150	0.4494	8.29%

CDNM thesis (discussed in Section 1) has implied that not all clustering problems in practice are difficult to solve. This research corresponds to the "clustering pipeline" proposed by Ackerman et al. [7], where clusterability (or structuredness in our context) can be evaluated to inform the selection of effective clustering algorithms. In this view, one intended contribution of this work is to implement this idea in the context of the CF problem. In future work, we will explore more applications in manufacturing systems that require grouping and combinatorial decisions (e.g., product and systems modularity). Also, we can explore more statistical and machine learning techniques such as multimodality tests and random forest to replace the K-S test for better predication performance.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] N. L. Hyer and U. Wemmerlov, *Reorganizing the Factory: Competing through Cellular Manufacturing*, Productivity Press, Portland, OR, USA, 2002.

[2] M. J. Brusco, "An exact algorithm for maximizing grouping efficacy in part-machine clustering," *IIE Transactions*, vol. 47, no. 6, pp. 653–671, 2015.

[3] C. Liu, Y. Yin, K. Yasuda, and J. Lian, "A heuristic algorithm for cell formation problems with consideration of multiple production factors," *International Journal of Advanced Manufacturing Technology*, vol. 46, no. (9-12), pp. 1201–1213, 2010.

[4] M. Ackerman and S. Ben-David, "Clusterability: a theoretical study," in *Proceedings of the International Conference on Artificial Intelligence and Statistics, JMLR: Workshop and Conference Proceedings*, N. Lawrence, Ed., vol. 5, p. 18, 2009.

[5] E. Nowakowska, J. Koronacki, and S. Lipovetsky, "Clusterability assessment for Gaussian mixture models," *Applied Mathematics and Computation*, vol. 256, pp. 591–601, 2015.

[6] A. Daniely, N. Linial, and M. Saks, "Clustering is difficult only when it does not matter," *Computer Science (Machine Learning)*, 2012.

[7] M. Ackerman, A. Adolfsson, and N. Brownstein, "An effective and efficient approach for clusterability evaluation," *Computer Science (Machine Learning)*, 2016.

[8] Y. Zhu and S. Li, "Statistical Analysis of Similarity Measures for Solving Cell Formation Problems," *Procedia CIRP*, vol. 63, pp. 248–253, 2017.

[9] M. M. Paydar and M. Saidi-Mehrabad, "A hybrid genetic-variable neighborhood search algorithm for the cell formation problem based on grouping efficacy," *Computers & Operations Research*, vol. 40, no. 4, pp. 980–990, 2013.

[10] C. S. Kumar and M. P. Chandrasekharan, "Grouping efficacy: a quantitative criterion for goodness of block diagonal forms of binary matrices in group technology," *International Journal of Production Research*, vol. 28, no. 2, pp. 233–243, 1990.

[11] J. McAuley, "Machine grouping for efficient production," *Production Engineering Research and Development*, vol. 51, no. 2, pp. 53–57, 1972.

[12] P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy: the Principles and Practice of Numerical Classification*, Freeman, San Francisco, Calif, USA, 1973.

[13] T.-H. Wu, C.-C. Chang, and J.-Y. Yeh, "A hybrid heuristic algorithm adopting both Boltzmann function and mutation operator for manufacturing cell formation problems," *International Journal of Production Economics*, vol. 120, no. 2, pp. 669–688, 2009.

[14] G. J. K. Nair and T. T. Narendran, "Grouping index: a new quantitative criterion for goodness of block-diagonal forms in group technology," *International Journal of Production Research*, vol. 34, no. 10, pp. 2767–2782, 1996.

[15] M. J. Brusco, "An iterated local search heuristic for cell formation," *Computers & Industrial Engineering*, vol. 90, pp. 292–304, 2015.

[16] B. R. Sarker and S. Mondal, "Grouping efficiency measures in cellular manufacturing: A survey and critical review," *International Journal of Production Research*, vol. 37, no. 2, pp. 285–314, 1999.

[17] B. R. Sarker, "The resemblance coefficients in group technology: A survey and comparative study of relational metrics," *Computers & Industrial Engineering*, vol. 30, no. 1, pp. 103–116, 1996.

[18] C. T. Mosier, J. Yelle, and G. Walker, "Survey of similarity coefficient based methods as applied to the group technology configuration problem," *Omega*, vol. 25, no. 1, pp. 65–79, 1997.

[19] Y. Yin and K. Yasuda, "Similarity coefficient methods applied to the cell formation problem: A taxonomy and review," *International Journal of Production Economics*, vol. 101, no. 2, pp. 329–352, 2006.

[20] C. Zhao and Z. Wu, "A genetic algorithm for manufacturing cell formation with multiple routes and multiple objectives," *International Journal of Production Research*, vol. 38, no. 2, pp. 385–395, 2000.

[21] F. M. Defersha and M. Chen, "A linear programming embedded genetic algorithm for an integrated cell formation and lot sizing considering product quality," *European Journal of Operational Research*, vol. 187, no. 1, pp. 46–69, 2008.

[22] S. Lee and H. P. Wang, "Manufacturing cell formation: A dual-objective simulated annealing approach," *The International Journal of Advanced Manufacturing Technology*, vol. 7, no. 5, pp. 314–320, 1992.

[23] R. Tavakkoli-Moghaddam, N. Safaei, and F. Sassani, "A new solution for a dynamic cell formation problem with alternative routing and machine costs using simulated annealing," *Journal of the Operational Research Society*, vol. 59, no. 4, pp. 443–454, 2008.

[24] G. Papaioannou and J. M. Wilson, "The evolution of cell formation problem methodologies based on recent studies (1997–2008): Review and directions for future research," *European Journal of Operational Research*, vol. 206, no. 3, pp. 509–521, 2010.

[25] C. Renzi, F. Leali, M. Cavazzuti, and A. O. Andrisano, "A review on artificial intelligence applications to the optimal design of dedicated and reconfigurable manufacturing systems," *International Journal of Advanced Manufacturing Technology*, vol. 72, no. (1-4), pp. 403–418, 2014.

[26] A. Stawowy, "Evolutionary strategy for manufacturing cell design," *Omega* , vol. 34, no. 1, pp. 1–18, 2006.

[27] G. J. Nair and T. T. Narendran, "CASE: A clustering algorithm for cell formation with sequence data," *International Journal of Production Research*, vol. 36, no. 1, pp. 157–180, 1998.

[28] G. W. Corder and D. I. Foreman, *Nonparametric Statistics: A Step-By-Step Approach*, John Wiley & Sons, Hoboken, NY, USA, 2014.

[29] A. P. Bradley, "ROC curve equivalence using the Kolmogorov-Smirnov test," *Pattern Recognition Letters*, vol. 34, no. 5, pp. 470–475, 2013.

[30] T. Arnold and J. Emerson, "Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions," *The R Journal*, vol. 3, no. 2, pp. 34–39, 2011.

[31] A. Justel, D. Peña, and R. Zamar, "A multivariate Kolmogorov-Smirnov test of goodness of fit," *Statistics & Probability Letters*, vol. 35, no. 3, pp. 251–259, 1997.

[32] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*, Wiley, Hoboken, NJ, USA, 5th edition, 2011.

[33] Z. Drezner and O. Turel, "Normalizing variables with too-frequent values using a Kolmogorov–Smirnov test: A practical approach," *Computers & Industrial Engineering*, vol. 61, no. 4, pp. 1240–1244, 2011.

[34] Y. J. Zhu, *Hierarchical Clustering and Similarity Statistics for Solving and Investigating Cell Formation Problems*, Department of Mechanical and Manufacturing Engineering, University of Calgary, 2017.

A Mixture of Generalized Tukey's g Distributions

José Alfredo Jiménez and Viswanathan Arunachalam

Department of Statistics, Universidad Nacional de Colombia, Carrera 45 No. 26-85, Bogotá, Colombia

Correspondence should be addressed to Viswanathan Arunachalam; varunachalam@unal.edu.co

Mixtures of symmetric distributions, in particular normal mixtures as a tool in statistical modeling, have been widely studied. In recent years, mixtures of asymmetric distributions have emerged as a top contender for analyzing statistical data. Tukey's g family of generalized distributions depend on the parameters, namely, g, which controls the skewness. This paper presents the probability density function (pdf) associated with a mixture of Tukey's g family of generalized distributions. The mixture of this class of skewed distributions is a generalization of Tukey's g family of distributions. In this paper, we calculate a closed form expression for the density and distribution of the mixture of two Tukey's g families of generalized distributions, which allows us to easily compute probabilities, moments, and related measures. This class of distributions contains the mixture of Log-symmetric distributions as a special case.

1. Introduction

The main focus of interest in financial economics is the distribution of stock market returns. Mandelbrot [1] suggested the family of stable Paretian distributions for stock market returns. Fama [2] established that the normality assumption of the empirical data does not hold as the distribution is fat tailed. Kon [3] and Tse [4] used a mixture of normal distributions for stock return. Fielitz and Rozelle [5] proposed a mixture of nonnormal stable distributions for stock price. Consequently, greater emphasis has been placed on using distributions which have asymmetry and leptokurtic properties. Recently Jiménez et al. [6] proposed option pricing based mixture of log-skew-normal distributions. If extreme events tend to occur more frequently than normal events, then skewness and kurtosis of nonnormal distributions play an essential role for the volatility smile.

The most important and useful characteristic of Tukey's g-h family of distributions introduced by Tukey [7] is that it covers most of the Pearsonian family of distributions. It can also generate several known distributions, for example, lognormal, Cauchy, exponential, and Chi-squared (see Martínez and Iglewicz [8], page 363). From Tukey's g-h family of distribution, we obtain g distribution, which is closely related

to lognormal distribution and possesses similar properties of moments. Tukey's g-h family of distributions have been used to study financial markets. Badrinath and Chatterjee [9, 10] and Mills [11] used g-h to model the return on a stock index, as well as the return on shares in several markets. Dutta and Babbel [12] found that the skewness and leptokurtic behavior of LIBOR were modeled effectively using g-h distribution. Dutta and Babbel [13] used g-h to model interest rates and options on interest rates, while Tang and Wu [14] proposed a new method for the Decomposition of Portfolio VaR. Dutta and Perry [15] and recently Jiménez and Arunachalam [16] used g-h distribution to study the operational risk for heavy tailed severity models. Jiménez and Arunachalam [17] provided explicit expressions for VaR and CVaR calculations using the family of Tukey's g-h distributions. Currently Jiménez et al. [18] studied generalization of Tukey's g-h family of distributions, when the standard normal random variable is replaced by a continuous random variable U with mean 0 and variance 1.

The subfamily of g distributions exhibits skewness and has great importance in the study of asymmetric distributions for analyzing data. This kind of distribution allows us to obtain scaled Log-symmetric distributions. Vitiello and Poon [19] considered a simple mixture of two g distributions

for option pricing data. The purpose of this paper is to present a mixture of Tukey's g distributions and derive some statistical properties including the pdf and moment generating function and its properties.

The paper is organized as follows: Section 2 presents Tukey's g-h family of generalized distributions and its pdf, as well as the cumulative distribution function (cdf). In Section 3, some theoretical results of the mixture of two Tukey's g families of generalized distributions are presented and Section 4 explains the methodology of calculating estimation of parameters by the method of moments. Section 5 discusses the adjustment methodology of our proposed model to real data of Heating-Degree-Days (HDD) indices and finally, in Section 6, we conclude.

2. Tukey's g Family of Generalized Distributions

Tukey [7] introduced the family g-h distributions by means of two nonlinear transformations given by

$$Y = T_{g,h}(Z) = \frac{1}{g}\left(\exp\{gZ\} - 1\right)\exp\left\{\frac{hZ^2}{2}\right\} \quad (1)$$

with $g \neq 0$, $h \in \mathbb{R}$, where the distribution of Z is standard normal. When these transformations are applied to a continuous random variable U with mean 0 and variance 1 such that its pdf $f_U(\cdot)$ is symmetric about the origin and cdf $F_U(\cdot)$, the transformation $T_{g,h}(U)$ is obtained, which henceforth will be termed Tukey's g-h generalized distribution. If $h = 0$, Tukey's g-h generalized distribution reduces to

$$T_{g,0}(U) = \frac{1}{g}\left(\exp\{gU\} - 1\right) \quad (2)$$

which is known as Tukey's g generalized distribution.

In order to model an arbitrary random variable X using the transformation given in (2), Hoaglin and Peters [20] introduced two new parameters, A (location) and B (scale), and proposed the following linear transformation:

$$X = A + BY \quad \text{with } Y = T_{g,0}(U). \quad (3)$$

The following properties for pdf, cdf, and quantile functions of Tukey's g generalized distribution were established by Jiménez et al. [18] in terms of the pdf and cdf of X as follows:

$$\text{pdf: } f_X(x; \lambda, g, \theta) = \frac{1}{g(x-\theta)} f_U\left(\frac{1}{g}\ln\left(\frac{x-\theta}{B/g}\right)\right)$$
$$\text{if } g(x-\theta) > 0,$$

$$\text{cdf: } F_X(x; \lambda, g, \theta) = F_U\left(\frac{1}{g}\ln\left(\frac{x-\theta}{B/g}\right)\right)$$
$$\quad (4)$$
$$\text{if } g(x-\theta) > 0,$$

$$\text{qf: } F_U^{-1}(q) = u_q = \frac{1}{g}\ln\left(\frac{x_q-\theta}{B/g}\right)$$
$$\text{if } g(x_q-\theta) > 0,$$

where $\lambda = \ln(B/|g|)$ and $\theta = A - B/g$. We say that the random variable X has a Log-symmetric distribution (such distributions are all asymmetric; see for reference Johnson et al. [21] and Stuart and Ord [22]) with three parameters: threshold (θ), scale (λ), and shape (g), denoted by $X \sim LS(\lambda, g, \theta)$.

The first expression of (4) allows us to obtain the following pdf associated with Tukey's g distribution. Table 1 shows the parameters of the pdf of X that we obtain using a selected set of well known symmetrical distributions (from Jiménez et al. [18]).

The nth moment of the random variable $Y = T_{g,0}(U)$ is given by

$$\mu_n'(Y) = \frac{1}{g^n}\sum_{k=0}^{n}(-1)^k\binom{n}{k}M_U(\tilde{g}), \quad \text{if } g \neq 0, \quad (5)$$

where $\tilde{g} = (n-k)g$ and $M_U(t)$ is the moment generating function of the random variable U, which are even function; that is, $M_U(t) = M_U(-t)$. Table 2 shows parameters of the pdf and the moment generating function for a random variable U, using a selected set of well known symmetrical distributions.

Expression (5) allows us to obtain the moments of Tukey's g generalized distribution. The nth moment of the random variable X given by (3) can be obtained using the formula

$$\mathbb{E}\left[(X - \mathbb{E}[X])^n\right] = \mu_n(X)$$
$$= \left(\frac{B}{g}\right)^n\sum_{k=0}^{n}(-1)^k\binom{n}{k}M_U(\tilde{g})M_U^k(g), \quad (6)$$

where $B/g = \text{sgn}(g)e^\lambda$. Note that the above expression of the nth moment does not depend on the parameter θ. Formulas for calculating the standardized skewness, $\beta_1(X)$, and standardized excess kurtosis, $\beta_2(X)$, are given by

$$\beta_1(X) = \text{sgn}(g)$$
$$\cdot \frac{\left(M_U(3g) - M_U^3(g)\right) - 3\left(M_U(2g) - M_U^2(g)\right)M_U(g)}{\left[M_U(2g) - M_U^2(g)\right]^{3/2}}, \quad (7)$$
$$\beta_2(X) = \frac{M_U(4g) - 4M_U(3g)M_U(g) + 3M_U^2(2g)}{\left[M_U(2g) - M_U^2(g)\right]^2} - 3,$$

where $\text{sgn}(\cdot)$ denote the signum function. Note that these expressions only depend on the parameter g and its sign, respectively. Any LS distribution should satisfy the following test given in Stuart and Ord [22]:

$$\beta_2(X) - \beta_1^2(X) - 1 \geq 0. \quad (8)$$

3. The Mixture of Two g Distributions

We assume that Y follows a Log-Symmetric Mixture (LSMIX) distribution. Let us assume that $f_Y(y)$ is the weighted sum of m-component LS densities; that is,

$$f_Y(y; \Lambda) = \sum_{j=1}^{m}\omega_j f_U\left(y; \lambda_j, g_j, \theta_j\right). \quad (9)$$

TABLE 1: Parameters of the pdf of the random variable $Z = \ln(X)$.

Distribution of the r.v. U	Parameters			Distribution of the r.v. Z	Parameters	
	μ, a	σ, b	$g \neq 0$		μ, a	σ, b
Laplace	0	$\dfrac{\sqrt{2}}{2}$	$\|g\| < \dfrac{\sqrt{2}}{n}$	Log-Laplace	$\ln\left(\dfrac{B}{\|g\|}\right)$	$\dfrac{\sqrt{2}}{2}\|g\|$
Logistic	0	$\dfrac{\sqrt{3}}{\pi}$	$\|g\| < \dfrac{\pi}{\sqrt{3}n}$	Log-Logistic	$\ln\left(\dfrac{B}{\|g\|}\right)$	$\dfrac{\sqrt{3}}{\pi}\|g\|$
Normal	0	1	$g \in \mathbb{R}$	Lognormal	$\ln\left(\dfrac{B}{\|g\|}\right)$	$\|g\|$
HyperSec	0	$\dfrac{2}{\pi}$	$\|g\| < \dfrac{\pi}{2n}$	LoghyperSec	$\ln\left(\dfrac{B}{\|g\|}\right)$	$\dfrac{2}{\pi}\|g\|$
HyperCsc	0	$\dfrac{\sqrt{2}}{\pi}$	$\|g\| < \dfrac{\pi}{\sqrt{2}n}$	LoghyperCsc	$\ln\left(\dfrac{B}{\|g\|}\right)$	$\dfrac{\sqrt{2}}{\pi}\|g\|$

TABLE 2: Parameters of the pdf and moment generating functions of the random variable U.

Distribution of the r.v. U	Parameters			$M_U(g)$
	μ, a	σ, b	$g \neq 0$	
Laplace	0	$\dfrac{\sqrt{2}}{2}$	$\|g\| < \dfrac{\sqrt{2}}{n}$	$\dfrac{2}{2 - g^2}$
Logistic	0	$\dfrac{\sqrt{3}}{\pi}$	$\|g\| < \dfrac{\pi}{\sqrt{3}n}$	$\sqrt{3}g \csc\left(\sqrt{3}g\right)$
Normal	0	1	$g \in \mathbb{R}$	$\exp\left\{\dfrac{1}{2}g^2\right\}$
HyperSec	0	$\dfrac{2}{\pi}$	$\|g\| < \dfrac{\pi}{2n}$	$\sec(g)$
HyperCsc	0	$\dfrac{\sqrt{2}}{\pi}$	$\|g\| < \dfrac{\pi}{\sqrt{2}n}$	$\sec^2\left(\dfrac{g}{\sqrt{2}}\right)$

We use the notation $Y \sim \text{LSMIX}(\Lambda)$, where $\Lambda = (\xi_1, \ldots, \xi_m)$, and each element $\xi_j = (\omega_j, \lambda_j, g_j, \theta_j)$ is the parameter vector that defines the jth component and probability weights, ω_j, satisfying the conditions

$$\sum_{j=1}^{m} \omega_j = 1, \quad 0 < \omega_j < 1, \text{ for each } j. \tag{10}$$

$$F_X(x) = \begin{cases} 0, & \text{if } x \leq \theta_1, \\ \omega_1 F_U\left(\dfrac{\ln(x - \theta_1) - \lambda_1}{g_1}\right), & \text{if } \theta_1 < x \leq \theta_2, \\ \omega_1 F_U\left(\dfrac{\ln(x - \theta_1) - \lambda_1}{g_1}\right) + \omega_2 F_U\left(\dfrac{\ln(x - \theta_2) - \lambda_2}{g_2}\right), & \text{if } x > \theta_2, \end{cases} \tag{14}$$

where $\omega_2 = 1 - \omega_1$. Begin with the fact that the quartile function is the inverse of the cdf. Thus, replacing $x_q > \theta_2$ in (14), we obtain

$$F_X(x_q) = \omega_1 F_U\left(\dfrac{\ln(x_q - \theta_1) - \lambda_1}{g_1}\right) \\ + \omega_2 F_U\left(\dfrac{\ln(x_q - \theta_2) - \lambda_2}{g_2}\right). \tag{15}$$

According to Titterington et al. [23] the two-component mixture of known distributions is set by two weights. Let

$$X = A + BY \quad \text{with } Y \sim \text{LSMIX}(\Lambda). \tag{11}$$

Then we can assume that $f_X(x)$ is the weighted sum of two Tukey's g mixture densities such that $g_1 g_2 > 0$. Thus

$$f_X(x)$$
$$= \begin{cases} 0, & \text{if } x \leq \theta_1, \\ \dfrac{\omega_1}{g_1(x - \theta_1)} f_U(z_1), & \text{if } \theta_1 < x \leq \theta_2, \\ \dfrac{\omega_1}{g_1(x - \theta_1)} f_U(z_1) + \dfrac{1 - \omega_1}{g_2(x - \theta_2)} f_U(z_2), & \text{if } x > \theta_2, \end{cases} \tag{12}$$

where, without loss of generality, we let $\theta_1 < \theta_2$, $0 \leq \omega_1 \leq 1$ and for $j = 1, 2$

$$z_j = \dfrac{1}{g_j} \ln\left(\dfrac{x - \theta_j}{B/g_j}\right) \tag{13}$$

with $\theta_j = A - (B/g_j)$, $\lambda_j = (B/|g_j|)$. We use the notation $X \sim \text{GTMIX}(A, B, g_1, g_2, \omega_1)$. Vitiello and Poon [19] did not provide the piecewise nature of the mixture density function above in (12). In this case the cdf of X is given by

If we assume that $U \sim N(0, 1)$, (12) can be written as

$$f_X(x)$$
$$= \begin{cases} 0, & \text{if } x \leq \theta_1, \\ \dfrac{\omega_1}{g_1(x - \theta_1)} \varphi(z_1), & \text{if } \theta_1 < x \leq \theta_2, \\ \dfrac{\omega_1}{g_1(x - \theta_1)} \varphi(z_1) + \dfrac{1 - \omega_1}{g_2(x - \theta_2)} \varphi(z_2), & \text{if } x > \theta_2, \end{cases} \tag{16}$$

where $\varphi(z)$ is the standard normal pdf. Note that the expression above matches the pdf of a mixture of three-parameter lognormal distributions. Letting $\theta_1 = \theta_2 = 0$, the above pdf reduces to that of a mixture of two-parameter lognormal distributions.

Given that every normal pdf is a version of the standard normal pdf then if $U \sim N(\mu, \sigma^2)$ we have

$$f_U(u, \mu, \sigma) = \frac{1}{\sigma}\varphi\left(\frac{u-\mu}{\sigma}\right), \quad \text{with } \mu \in \mathbb{R}, \ \sigma > 0, \quad (17)$$

and (12) can be written as

$$f_X(x)$$

$$= \begin{cases} 0, & \text{if } x \le \theta_1^*, \\ \dfrac{\omega_1}{g_1^*(x-\theta_1^*)}\varphi(z_1), & \text{if } \theta_1^* < x \le \theta_2^*, \quad (18) \\ \dfrac{\omega_1}{g_1^*(x-\theta_1^*)}\varphi(z_1) + \dfrac{1-\omega_1}{g_2^*(x-\theta_2^*)}\varphi(z_2), & \text{if } x > \theta_2^*. \end{cases}$$

If the parameters g_j are scaled by σ, that is, $g_j^* = \sigma g_j$, then

$$z_j = \frac{\ln(x-\theta_j^*) - \lambda_j^*}{g_j^*} \quad (19)$$

with $\theta_j^* = A - (B\sigma/g_j^*)$, $\lambda_j^* = \ln(B\sigma/|g_j^*|) + (\mu/\sigma)g_j^*$. Note that the expression above matches the pdf of a mixture of three-parameter lognormal distributions, which is a generalization of the pdf given in (16), and we use the notation $X \sim \text{LSMIX}(\lambda_1^*, \lambda_2^*, g_1^*, g_2^*, \theta_1^*, \theta_2^*, \omega_1)$. Similarly, we can obtain pdf of a mixture of distributions for the random variables listed in Table 1.

4. Estimation of the Mixtures of Two Tukey's g Distributions

In this section, we explain the estimation of the mixture of two Tukey's g distributions. The expected value of X is given by

$$\mathbb{E}[X] = \mu_1' = \sum_{j=1}^{2}\omega_j\left(\theta_j + \frac{B}{g_j}M_U(g_j)\right). \quad (20)$$

The nth raw moment of the random variable X is given by

$$\mathbb{E}[X^n] = \sum_{j=1}^{2}\omega_j\sum_{k=0}^{n}\binom{n}{k}\left(\frac{B}{g_j}\right)^{n-k}\theta_j^k M_U(\widetilde{g}_j), \quad (21)$$

where $g_1 g_2 \ne 0$, $\widetilde{g}_j = (n-k)g_j$ and $M_U(t)$ is the moment generating function of the random variable U. The central moments μ_n of the random variable X are given by

$$\mathbb{E}\left[(X-\mu_1')^n\right] = \mu_n(X)$$

$$= \sum_{j=1}^{2}\omega_j\sum_{k=0}^{n}\binom{n}{k}\left(\frac{B}{g_j}\right)^{n-k}(\theta_j - \mu_1')^k M_U(\widetilde{g}_j). \quad (22)$$

The first five central moments are as follows:

$$\mu_1 = \omega_1\eta_1 + \omega_2\eta_2 = 0,$$

$$\mu_2 = \omega_1\left(\sigma_1^2 + \eta_1^2\right) + \omega_2\left(\sigma_2^2 + \eta_2^2\right),$$

$$\mu_3 = \omega_1\left(3\sigma_1^2 + \eta_1^2\right)\eta_1 + \omega_2\left(3\sigma_2^2 + \eta_2^2\right)\eta_2,$$

$$\mu_4 = \omega_1\left(3\sigma_1^4 + 6\eta_1^2\sigma_1^2 + \eta_1^4\right) \quad (23)$$

$$\qquad + \omega_2\left(3\sigma_2^4 + 6\eta_2^2\sigma_2^2 + \eta_2^4\right),$$

$$\mu_5 = \omega_1\left(15\sigma_1^4 + 10\eta_1^2\sigma_1^2 + \eta_1^4\right)\eta_1$$

$$\qquad + \omega_2\left(15\sigma_2^4 + 10\eta_2^2\sigma_2^2 + \eta_2^4\right)\eta_2,$$

where for $j = 1, 2$

$$\eta_j = \frac{B}{g_j}M_U(g_j) + \theta_j - \mu_1',$$

$$\sigma_j^2 = \left(\frac{B}{g_j}\right)^2\left[M_U(2g_j) - M_U^2(g_j)\right]. \quad (24)$$

Because $\theta_1 < \theta_2$, upon equating population moments to the corresponding sample moments, it follows from (23) that

$$\omega_1\left(\frac{B}{g_1}M_U(g_1) + \theta_1 - m_1\right)$$

$$\qquad + \omega_2\left(\frac{B}{g_2}M_U(g_2) + \theta_2 - m_1\right) = 0. \quad (25)$$

Left-hand side of system (23) is multiplied by $\omega_1 + \omega_2 = 1$; the equations take the following form:

$$\omega_1\eta_1 + \omega_2\eta_2 = 0,$$

$$\omega_1\left(\sigma_1^2 + \eta_1^2 - m_2\right) + \omega_2\left(\sigma_2^2 + \eta_2^2 - m_2\right) = 0,$$

$$\omega_1\left(3\eta_1\sigma_1^2 + \eta_1^3 - m_3\right) + \omega_2\left(3\eta_2\sigma_2^2 + \eta_2^3 - m_3\right) = 0,$$

$$\omega_1\left(3\sigma_1^4 + 6\eta_1^2\sigma_1^2 + \eta_1^4 - m_4\right) \quad (26)$$

$$\qquad + \omega_2\left(3\sigma_2^4 + 6\eta_2^2\sigma_2^2 + \eta_2^4 - m_4\right) = 0,$$

$$\omega_1\left(15\eta_1\sigma_1^4 + 10\eta_1^3\sigma_1^2 + \eta_1^5 - m_5\right)$$

$$\qquad + \omega_2\left(15\eta_2\sigma_2^4 + 10\eta_2^3\sigma_2^2 + \eta_2^5 - m_5\right) = 0,$$

where m_i $(i = 1, 2, \ldots)$ denote the ith central moment of the sample. Equations (26) accordingly constitute a system of five equations to be solved simultaneously for the estimates of the five parameters A, B, g_1, g_2, and ω_1.

Note that, from the first equation of system of (26), it follows that

$$\omega_1 = \frac{\eta_2}{\eta_2 - \eta_1}. \tag{27}$$

We eliminate ω_1 between the first and the subsequent equations of (26) in turn and thereby reduce the system to the following four equations in four unknowns A, B, g_1, and g_2:

$$\eta_1^{-1}\left(\sigma_1^2 + \eta_1^2 - m_2\right) = \eta_2^{-1}\left(\sigma_2^2 + \eta_2^2 - m_2\right),$$

$$\eta_1^{-1}\left(3\eta_1\sigma_1^2 + \eta_1^3 - m_3\right) = \eta_2^{-1}\left(3\eta_2\sigma_2^2 + \eta_2^3 - m_3\right),$$

$$\eta_1^{-1}\left(3\sigma_1^4 + 6\eta_1^2\sigma_1^2 + \eta_1^4 - m_4\right)$$
$$= \eta_2^{-1}\left(3\sigma_2^4 + 6\eta_2^2\sigma_2^2 + \eta_2^4 - m_4\right), \tag{28}$$

$$\eta_1^{-1}\left(15\eta_1\sigma_1^4 + 10\eta_1^3\sigma_1^2 + \eta_1^5 - m_5\right)$$
$$= \eta_2^{-1}\left(15\eta_2\sigma_2^4 + 10\eta_2^3\sigma_2^2 + \eta_2^5 - m_5\right).$$

These systems of equations are solved computationally by using scientific software package and we do not need to verify the unique solution of the system as the parameter estimates. We skip further details and numerical illustration owing to space constraint.

5. Illustration

In this section we discuss some examples and applications of the results derived in Section 3 with two examples. In the first example, we discuss the pricing of a call option using a mixture of two Tukey's g-generalized distributions as an example to illustrate the results of Section 3. In the second example, we examine the empirical real data of Heating-Degree-Day to demonstrate usefulness of our approach of mixture of LS distributions.

Jiménez et al. [24] derived the option price of an European option assuming that the terminal price distribution follows a g-generalized distribution. Instead if we use a mixture of two Tukey's classes of g-generalized distributions, then the price of the call option denoted by $C(t, \tau; K)$ with a strike price K and maturity date $T = t + \tau$ can be expressed as follows:

$$C(t, \tau; K) = \sum_{j=1}^{2} \omega_j e^{-r\tau}$$
$$\cdot \left[\int_{-\delta_j}^{\infty} e^{\lambda_j + g_j u} f_U(u)\, du - \left(K - \theta_j\right) F_U\left(\delta_j\right)\right], \tag{29}$$

where $K > \theta_2$ and

$$\delta_j = \frac{\lambda_j - \ln\left(K - \theta_j\right)}{g_j} \quad \text{for } j = 1, 2. \tag{30}$$

When $U \sim N(0, 1)$, (29) reduces to

$$C(t, \tau; K) = \sum_{j=1}^{2} \omega_j e^{-r\tau}$$
$$\cdot \left[e^{\lambda_j + (1/2)g_j^2}\Phi\left(\delta_j - g_j\right) - \left(K - \theta_j\right)\Phi\left(\delta_j\right)\right], \tag{31}$$

where $\Phi(\cdot)$ denotes the cdf of a standard univariate normal variable. If we assume that $\theta_1 = \theta_2 = 0$, then (31) reduces to

$$C(t, \tau; K)$$
$$= \sum_{j=1}^{2} \omega_j e^{-r\tau}\left[e^{\lambda_j + (1/2)g_j^2}\Phi\left(\delta_j - g_j\right) - K\Phi\left(\delta_j\right)\right]. \tag{32}$$

Note that when $g_j = \sigma_j\sqrt{\tau}$, these expressions coincide with the option pricing formula given in Bahra [25]. The authors also established closed form formula for the calculation of the sensitives measures of option pricing (Greek parameters of the option). Here we wish to observe that our mixture model uses less unknown parameters for calculating the option pricing, whereas Vitiello and Poon [19] used nine unknown parameters to obtain the same for the mixture of two g-distributions. It has been known that when we increase the number of parameters, we lose degrees of freedom and it is no longer acceptable for the best fit of data. This gives an advantage of our approach for the mixture of two g-generalized distributions.

We now present, as an example, the use of Heating-Degree-Days (HDD) in relation to winter temperature risk as a substitute for gas demand. HDD based contracts are listed on the Chicago Mercantile Exchange (CME). We consider an example that consists of monthly aggregate Heating-Degree-Day (HDD) data values at the Chicago O'Hare International Airport from December 1979 to December 2000 given in Wang [26] and explored also by Vitiello and Poon [19]. We describe first a LS distribution with three parameters based method to infer the implied risk-neutral probability density (RND). In Table 3, we present the estimated values of the three parameters of lognormal and Log-Logistic distributions; our interest is to compare with Vitiello and Poon [19] risk-neutral densities with our proposed mixture model.

The smaller value of the Kolmogorov-Smirnov (KS) test confirms that the data obeys the LS distributions with three parameters. We wish to observe that Anderson-Darling (AD) test is more sensitive to the tails of the LS distributions in comparison with KS test. In this case, we choose the Log-Logistic distribution as the best fit for the HDD data.

The implicit risk-neutral densities (RND) of LS distributions are shown in Figure 1 and compared with Figure 6 of Vitiello and Poon [19]. We have obtained a similar plot by our method with less unknown parameters than method given by Vitiello and Poon [19]. Furthermore, their KS test value of 13.6326% which is higher than the KS test values of Table 3 favors the best fit for the frequency of the LS distributions. Therefore, finite mixtures are attractive from the application viewpoint because of its flexibility and permit us to model various kinds of shaped distributions. In Table 4,

TABLE 3: Estimates for adjusting the LS(Λ).

Distribution of three parameters	Parameters			Test of adjusted	
	λ	g	θ	AD (%)	KS (%)
Lognormal	5.7407	0.5265	798.2540	37.53	12.3041
Log-Logistic	3.0225	283.8185	824.1814	32.68	11.1367

TABLE 4: Estimates for adjusting the mixture of LS(Λ).

Mixture of distributions	Parameters				KS (%) test
	$(\lambda_1; \lambda_2)$	$(g_1; g_2)$	$(\theta_1; \theta_2)$	ω	
Lognormal	6.5749	-0.1268	1797.9995	0.8188	8.0866
	11.0933	5.7574×10^{-4}	-64213.6284		
Log-Logistic	10.1338	-538.2433	1620.7711	0.8182	8.5928
	303.7429	6818.2894	-5309.4256		
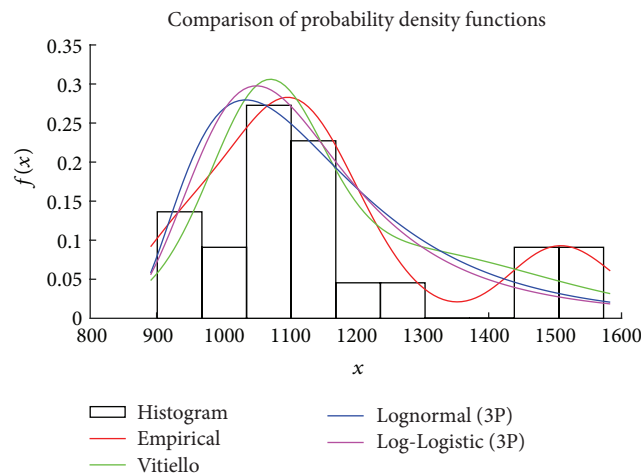
FIGURE 1: Empirical and LS(Λ) densities estimated from HDD.
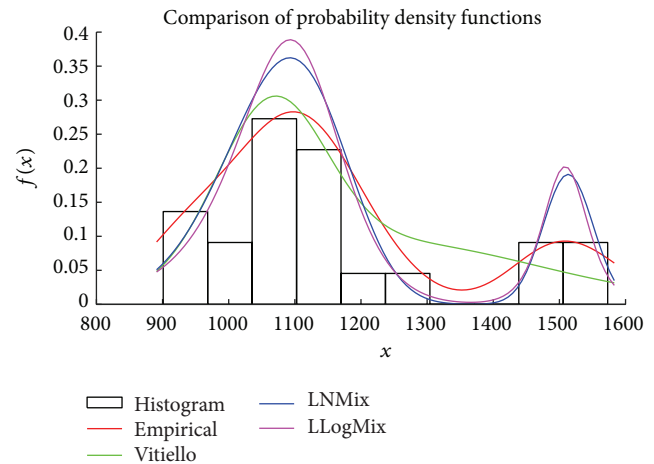
FIGURE 2: Empirical and two-g densities estimated from HDD.

we give the estimate values of the parameters of the mixture LS distributions. These parameters are estimated using (28). The estimated two g-densities and the implied risk-neutral densities (RND) are shown in Figure 2.

We observe that the bimodal LS mixture distribution has same fitting performance of the empirical distribution function (EDF) and lognormal mixture distribution gives best goodness of fit using the KS test.

6. Conclusions

This paper presents a mixture of Tukey's g-generalized distributions and its properties. The methodology of estimating the unknown parameters by the method of moments is also presented. The proposed model has the advantage that it provides flexibility, when skewness, kurtosis, or other moments of the underlying distribution do not follow a normal distribution.

Some special cases of well known distributions are obtained from the proposed model.

Competing Interests

The authors declare that they have no competing interests.

References

[1] B. Mandelbrot, "The variation of certain speculative prices," *The Journal of Business*, vol. 36, no. 4, pp. 394–419, 1963.

[2] E. F. Fama, "The behavior of stock-market prices," *The Journal of Business*, vol. 38, no. 1, pp. 34–105, 1965.

[3] S. J. Kon, "Models of stock returns—a comparison," *The Journal of Finance*, vol. 39, no. 1, pp. 147–165, 1984.

[4] Y. K. Tse, "Price and volume in the tokyo stock exchange: an exploratory study," Working Paper 1573, University of Illinois at Urbana-Champaign, 1989.

[5] B. D. Fielitz and J. P. Rozelle, "Stable distributions and the mixtures of distributions hypotheses for common stock returns," *Journal of the American Statistical Association*, vol. 78, no. 381, pp. 28–36, 1983.

[6] J. A. Jiménez, V. Arunachalam, and G. M. Serna, "Option pricing based on a log-skew-normal mixture," *International Journal of Theoretical and Applied Finance*, vol. 18, no. 8, Article ID 1550051, 22 pages, 2015.

[7] J. W. Tukey, "Modern techniques in data analysis," in *Proceedings of the NSF Sponsored Regional Research Conference*, Southeastern Massachusetts University, North Dartmouth, Mass, USA, 1977.

[8] J. Martínez and B. Iglewicz, "Some properties of the tukey g and h family of distributions," *Communications in Statistics—Theory and Methods*, vol. 13, no. 3, pp. 353–369, 1984.

[9] S. G. Badrinath and S. Chatterjee, "On measuring skewness and elongation in common stock return distributions: the case of the market index," *The Journal of Business*, vol. 61, no. 4, pp. 451–472, 1988.

[10] S. G. Badrinath and S. Chatterjee, "A data-analytic look at skewness and elongation in common-stock-return distributions," *Journal of Business & Economic Statistics*, vol. 9, no. 2, pp. 223–233, 1991.

[11] T. C. Mills, "Modelling skewness and kurtosis in the london stock exchange $ft - se$ index return distributions," *The Statistician*, vol. 44, no. 3, pp. 323–332, 1995.

[12] K. K. Dutta and D. F. Babbel, *On Measuring Skewness and Kurtosis in Short Rate Distributions: The Case of the US Dollar London Inter Bank Offer Rates*, The Warton School, University of Pennsylvania, Philadelphia, Pa, USA, 2004.

[13] K. K. Dutta and D. F. Babbel, "Extracting probabilistic information from the prices of interest rate options: tests of distributional assumptions," *Journal of Business*, vol. 78, no. 3, pp. 841–870, 2005.

[14] X. Tang and X. Wu, "A new method for the decomposition of portfolio VaR," *Journal of Systems Science and Information*, vol. 4, no. 4, pp. 721–727, 2006.

[15] K. K. Dutta and J. Perry, "A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital," Working Paper 06-13, Federal Reserve Bank of Boston, 2007.

[16] J. A. Jiménez and V. Arunachalam, "Evaluating operational risk by an inhomogeneous counting process based on Panjer recursion," *The Journal of Operational Risk*, vol. 11, no. 1, pp. 1–21, 2016.

[17] J. A. Jiménez and V. Arunachalam, "Using Tukey's g and h family of distributions to calculate value at risk and conditional value at risk," *Journal of Risk*, vol. 13, no. 4, pp. 95–116, 2011.

[18] J. A. Jiménez, V. Arunachalam, and G. M. Serna, "A generalization of Tukey's g-h family of distributions," *Journal of Statistical Theory and Applications*, vol. 14, no. 1, pp. 28–44, 2015.

[19] L. Vitiello and S.-H. Poon, "General equilibrium and risk neutral framework for option pricing with a mixture of distributions," *The Journal of Derivatives*, vol. 15, no. 4, pp. 48–60, 2008.

[20] D. C. Hoaglin and S. C. Peters, *Software for Exploring Distribution Shape*, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 1979.

[21] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 1, John Wiley & Sons, New York, NY, USA, 1994.

[22] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics: Distribution Theory*, vol. 1, Edward Arnold, London, UK; Halsted Press, New York, NY, USA, 6th edition, 1994.

[23] D. M. Titterington, A. F. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*, vol. 7, John Wiley & Sons, New York, NY, USA, 1985.

[24] J. A. Jiménez, V. Arunachalam, and G. M. Serna, "Option pricing based on the generalised Tukey distribution," *International Journal of Financial Markets and Derivatives*, vol. 3, no. 3, pp. 191–221, 2014.

[25] B. Bahra, "Implied risk-neutral probability density functions from option prices: a central bank perspective," in *Forecasting Volatility in the Financial Markets*, J. Knight and S. Satchell, Eds., pp. 201–226, Elsevier, 3rd edition, 1997.

[26] S. S. Wang, "A universal framework for pricing financial and insurance risks," *ASTIN Bulletin*, vol. 32, no. 2, pp. 213–234, 2002.

A Note on the Adaptive LASSO for Zero-Inflated Poisson Regression

**Prithish Banerjee,[1] Broti Garai,[2] Himel Mallick ⓘ[3,4]
Shrabanti Chowdhury,[5] and Saptarshi Chatterjee[6]**

[1]*JP Morgan Chase & Co., USA*
[2]*NBCUniversal, USA*
[3]*Department of Biostatistics, Harvard T.H. Chan School of Public Health, USA*
[4]*Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, USA*
[5]*Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, USA*
[6]*Eli Lilly and Company, USA*

Correspondence should be addressed to Himel Mallick; hmallick@hsph.harvard.edu

Prithish Banerjee, Broti Garai, and Himel Mallick contributed equally to this work.

We consider the problem of modelling count data with excess zeros using Zero-Inflated Poisson (ZIP) regression. Recently, various regularization methods have been developed for variable selection in ZIP models. Among these, EM LASSO is a popular method for simultaneous variable selection and parameter estimation. However, EM LASSO suffers from estimation inefficiency and selection inconsistency. To remedy these problems, we propose a set of EM adaptive LASSO methods using a variety of data-adaptive weights. We show theoretically that the new methods are able to identify the true model consistently, and the resulting estimators can be as efficient as oracle. The methods are further evaluated through extensive synthetic experiments and applied to a German health care demand dataset.

1. Introduction

Modern research studies routinely collect information on a broad array of outcomes including count measurements with excess amount of zeros. Modeling such zero-inflated count outcomes is challenging for several reasons. First, traditional count models such as Poisson and Negative Binomial are suboptimal in accounting for excess variability due to zero-inflation [1, 2]. Second, alternative zero-inflated models such as the **Z**ero-**I**nflated **P**oisson (ZIP) [2] and **Z**ero-**I**nflated **N**egative **B**inomial (ZINB) [1] models are computationally prohibitive in the presence of high-dimensional and collinear variables.

Regularization methods have been proposed as a powerful framework to mitigate these problems, which tend to exhibit significant advantages over traditional methods [3, 4]. Essentially all these methods enforce sparsity through a suitable penalty function and identify predictive features by means of a computationally efficient Expectation Maximization (EM) algorithm. Among these, EM LASSO is particularly attractive due to its capability to perform simultaneous model selection and stable effect estimation. However, recent research suggests that EM LASSO may not be fully efficient and its model selection result could be inconsistent [5, 6]. This led to a simple modification of the LASSO penalty, namely, the EM adaptive LASSO (EM AL). EM AL achieves "oracle selection consistency" by allowing different amounts of shrinkage for different regression coefficients.

Previous studies have not, however, investigated the EM AL at sufficient depth to evaluate its properties under diversified and realistic scenarios. It is not yet clear, for example, how reliable the resulting parameter estimates are in the presence of multicollinearity. In particular, the actual variable selection performance of EM AL depends on the proper

construction of the data-adaptive weight vector. When the features to be associated possess an inherent collinearity, EM AL is expected to produce suboptimal results, a phenomenon that is especially evident when the sample size is limited [7]. Several remedies have been suggested for linear and generalized linear models (GLMs) such as the standard error-adjusted adaptive LASSO (SEAL) [7, 8]. However, there is a lack of similar published methods for zero-inflated count regression models. In addition, complete software packages of these methods have not been made available to the community.

We address these issues by providing a set of flexible variable selection approaches to efficiently identify correlated features associated with zero-inflated count outcomes in a ZIP regression framework. We have implemented this method as AMAZonn (**A M**ulticollinearity-adjusted **A**daptive LASSO for **Z**ero-inflated C**o**u**n**t Regressio**n**). AMAZonn considers two data-adaptive weights: (i) the inverse of the maximum likelihood (ML) estimates (EM AL) and (ii) inverse of the ML estimates divided by their standard errors (EM SEAL). We show theoretically that AMAZonn is able to identify the true model consistently, and the resulting estimator is as efficient as oracle. Numerical studies confirmed our theoretical findings. The rest of the article is organized as follows. The AMAZonn method is proposed in the next section, and its theoretical properties are established in Section 3. Simulation results are reported in Section 4 and one real dataset is analyzed in Section 5. Then, the article concludes with a short discussion in Section 6. All technical details are presented in the Appendix.

2. Methods

2.1. Zero-Inflated Poisson (ZIP) Model. Zero-inflated count models assume that the observations originate either from a "susceptible" population that generates zero and positive counts according to a count distribution or from a "nonsusceptible" population, which produces additional zeros [1, 2]. Thus, while a subject with a positive count is considered to belong to the "susceptible" population, individuals with zero counts may belong to either of the two latent populations. We denote the observed values of the response variable as $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$. Following Lambert [2], a ZIP mixture distribution can be written as

$$P(y_i = k) = \begin{cases} p_i + (1 - p_i) e^{-\lambda_i} & \text{if } k = 0, \\ (1 - p_i) \dfrac{e^{-\lambda_i} \lambda_i^k}{k!} & \text{if } k = 1, 2, \ldots, \end{cases} \tag{1}$$

where p_i is the probability of belonging to the nonsusceptible population and λ_i is the Poisson mean corresponding to the susceptible population for the ith individual ($i = 1, \ldots, n$). It can be seen from (1) that ZIP reduces to the standard Poisson model when $p_i = 0$. Also, $P(y_i = 0) > e^{-\lambda_i}$, indicating zero-inflation. The probability of belonging to the "nonsusceptible" population, p_i, and the Poisson mean, λ_i, are linked to the explanatory variables through the logit and log links as

$$\text{logit}(p_i) = \mathbf{z}_i'\boldsymbol{\gamma} \text{ and} \tag{2}$$

TABLE 1: The AMAZonn data-adaptive weights. $\widehat{\beta}_{\text{ML}}$ and $\widehat{\gamma}_{\text{ML}}$ denote the ML estimates based on the unpenalized ZIP model, corresponding to count and zero submodels, respectively. SE denotes the standard errors of the corresponding ML estimates.

Weighting Scheme	Count	Zero
AMAZonn - EM AL	$\dfrac{1}{\left\|\widehat{\beta}_{j_{\text{ML}}}\right\|}$	$\dfrac{1}{\left\|\widehat{\gamma}_{j_{\text{ML}}}\right\|}$
AMAZonn - EM SEAL	$\dfrac{SE\left(\widehat{\beta}_{j_{\text{ML}}}\right)}{\left\|\widehat{\beta}_{j_{\text{ML}}}\right\|}$	$\dfrac{SE\left(\widehat{\gamma}_{j_{\text{ML}}}\right)}{\left\|\widehat{\gamma}_{j_{\text{ML}}}\right\|}$

$$\log(\lambda_i) = \mathbf{x}_i'\boldsymbol{\beta}, \tag{3}$$

where $\mathbf{x}_i$ and $\mathbf{z}_i$ are vectors of covariates for the ith subject ($i = 1, \ldots, n$) corresponding to the count and zero models, respectively, and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_q)'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$ are the corresponding regression coefficients including the intercepts.

For n independent observations, the ZIP log-likelihood function can be written as

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{y_i=0} \log\left\{ e^{z_i'\gamma} + e^{-e^{x_i'\beta}} \right\}$$

$$+ \sum_{y_i>0} \left\{ y_i x_i'\boldsymbol{\beta} + e^{-x_i'\beta} \right\} - \sum_{i=1}^{n} \log\left\{ 1 + e^{z_i'\gamma} \right\} \tag{4}$$

$$- \sum_{y_i>0} \log(y_i!).$$

2.2. The AMAZonn Method. AMAZonn considers two data-adaptive weights in the EM adaptive LASSO framework: (i) the inverse of the maximum likelihood (ML) estimates (EM AL) and (ii) inverse of the ML estimates divided by their standard errors (EM SEAL). As defined by Tang et al. [6], the EM adaptive LASSO formulation for ZIP regression is given by

$$\widehat{\boldsymbol{\theta}}^* = \arg\min\{-L(\boldsymbol{\theta})\} + \nu_1 \sum_{j=1}^{p} w_{1j} \left|\beta_j\right| + \nu_2 \sum_{j=1}^{p} w_{2j} \left|\gamma_j\right|, \tag{5}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\gamma}\}$ is the parameter vector of interest with known weights $w_1 = (w_{11}, \ldots, w_{1p})'$ and $w_2 = (w_{21}, \ldots, w_{2p})'$. As noted by Qian and Yang [7], the inverse of the maximum likelihood (ML) estimates as weights may not always be stable, especially when the multicollinearity of the design matrix is a concern. In order to adjust for this instability, AMAZonn additionally considers the inverse of the ML estimates divided by their standard errors as weights. We refer to these two methods as AMAZonn - EM AL and AMAZonn - EM SEAL, respectively (Table 1).

2.3. The EM Algorithm. In order to efficiently estimate the parameters in the above optimization problem (5), we resort to the EM algorithm. To this end, we define a set of latent variables z_i as follows:

$z_i = 1$ if y_i is from the zero state, and

$z_i = 0$ if y_i is from the count state, $\quad i = 1, \ldots, n.$

$$(6)$$

We consider the latent variables z_i's as the "missing data" and rewrite the complete-data log-likelihood function in (4) as follows:

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[z_i X_i \gamma - \log\left(1 + \exp\left(X_i \gamma\right)\right) \right.$$

$$(7)$$

$$\left. + \left(1 - z_i\right) \left\{ y_i X_i \beta - \left(y_i + 1\right) \log\left(1 + X_i \beta\right) \right\} \right].$$

With the above formulation, the objective function in (5) can be rewritten as

$$Q^*(\boldsymbol{\theta}) = -L(\boldsymbol{\theta}) + \nu_1 \sum_{j=1}^{p} w_{1j} \left| \beta_j \right| + \nu_2 \sum_{j=1}^{p} w_{2j} \left| \gamma_j \right|, \qquad (8)$$

which can be iteratively solved as follows:

(1) At iteration t, the **E step** computes the expectation of $Q^*(\boldsymbol{\theta})$ by substituting z_i with its conditional expectation given observed data and current parameter estimates

$$\widehat{z}_i^{(t)} = \begin{cases} \left(\left(1 + \left[\dfrac{\exp\left(-X_i \widehat{\gamma}^{(t)}\right)}{1 + \exp\left(-X_i \widehat{\beta}^{(t)}\right)} \right] \right) \right) & \text{if } y_i = 0, \\ 0 & \text{if } y_i > 0. \end{cases} \qquad (9)$$

(2) In the **M step**, the expected penalized complete-data log-likelihood (5) can be minimized the with respect to $\boldsymbol{\theta}$ as

$$Q^*\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}\right) = -2E(L\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}\right) + \nu_1 \sum_{j=1}^{p} w_{1j} \left| \beta_j \right|$$

$$(10)$$

$$+ \nu_2 \sum_{j=1}^{p} w_{2j} \left| \gamma_j \right|.$$

(3) Continue this process until convergence, $t = 1, 2, \ldots.$

It is to be noted that (10) can be further decomposed as

$$Q^*\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}\right) = Q_1^*\left(\boldsymbol{\beta} \mid \boldsymbol{\theta}^{(t)}\right) + Q_2^*\left(\boldsymbol{\gamma} \mid \boldsymbol{\theta}^{(t)}\right), \qquad (11)$$

where Q_1^* is the weighted penalized Poisson log-likelihood defined as

$$Q_1^*\left(\boldsymbol{\beta} \mid \boldsymbol{\theta}^{(t)}\right) = -2 \left[\sum_{i=1}^{n} \left(1 - \widehat{z}_i^{(t)}\right) \right.$$

$$\left. \cdot \left\{ y_i X_i \boldsymbol{\beta} - \left(y_i + 1\right) \log\left(1 + X_i \boldsymbol{\beta}\right) \right\} \right] \qquad (12)$$

$$+ \nu_1 \sum_{j=1}^{p} w_{1j} \left| \beta_j \right|,$$

and Q_2^* is the penalized logistic log-likelihood defined as

$$Q_2^*\left(\boldsymbol{\gamma} \mid \boldsymbol{\theta}^{(t)}\right) = -2 \left[\sum_{i=1}^{n} \widehat{z}_i^{(t)} X_i \boldsymbol{\gamma} - \log\left(1 + \exp\left(X_i \boldsymbol{\gamma}\right)\right) \right]$$

$$(13)$$

$$+ \nu_2 \sum_{j=1}^{p} w_{2j} \left| \gamma_j \right|,$$

both of which can be minimized separately using computationally efficient coordinate descent algorithms developed for GLMs [9].

2.4. Selection of Tuning Parameters. We select the tuning parameters based on the minimum BIC [10] criterion, which is known to provide better variable selection performance as compared to other information criteria [11]. This can be effortlessly incorporated in our formulation by using existing implementations for zero-inflated count models [3, 4, 6].

3. Oracle Properties

Recently, Tang et al. [6] showed that the EM adaptive LASSO (i.e., AMAZonn - EM AL) enjoys the so-called oracle properties, i.e., the estimator is able to identify the true model consistently, and the resulting estimator is as efficient as *oracle*. Here we extend these results to the AMAZonn - EM SEAL estimator and show that the AMAZonn - EM SEAL estimator also maintains the same theoretical properties. For the sake of completeness, we provide a combined general proof for both AMAZonn estimators.

Without being too rigorous mathematically, recall that the log-likelihood function for the ZIP regression model is given by

$$L(\boldsymbol{\theta}; \boldsymbol{v}_i) = \sum_{y_i = 0} \log\left[\psi_i + \left(1 - \psi_i\right) f\left(0; \lambda_i\right) \right]$$

$$(14)$$

$$+ \sum_{y_i > 0} \log\left[\left(1 - \psi_i\right) f\left(y_i; \lambda_i\right) \right],$$

where $\boldsymbol{v}_i$'s are the observed data (i.i.d observations from the ZIP distribution), $f(.; \lambda_i)$ is the probability mass function of Poisson distribution with parameter $\lambda_i = \exp(X_i \boldsymbol{\beta})$ and $\psi_i = \exp(X_i \boldsymbol{\gamma})/(1 + \exp(X_i \boldsymbol{\gamma})), i = 1, \ldots, n.$ The corresponding penalized log-likelihood is given by

$$Q(\boldsymbol{\theta}) = -L(\boldsymbol{\theta}; \boldsymbol{v}_i) + \nu_{1n} \sum_{j=1}^{p} w_{1j} \left| \beta_j \right| + \nu_{2n} \sum_{j=1}^{p} w_{2j} \left| \gamma_j \right|. \qquad (15)$$

Let us denote the true coefficient vector as $\boldsymbol{\theta_0} = (\boldsymbol{\beta_0}^T, \boldsymbol{\gamma_0}^T)^T.$ Decompose $\boldsymbol{\theta_0} = (\boldsymbol{\theta_{10}}^T, \boldsymbol{\theta_{20}}^T)^T$ and assume that $\boldsymbol{\theta_{20}}^T$ contains all zero coefficients. Let us denote the subset of true nonzero coefficients as $\mathscr{A} = \{j : \theta_{j0} \neq 0\}$ and the subset of selected nonzero coefficients as $\widehat{\mathscr{A}} = \{j : \widehat{\theta}_j \neq 0\}.$ With this formulation, the Fisher information matrix can be written as

$$I(\boldsymbol{\theta_0}) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}, \qquad (16)$$

where I_{11} is the Fisher information corresponding the true nonzero submodel. The oracle property of AMAZonn may be developed based on certain mild regularity conditions which are as follows:

(**A1**): The Fisher information matrix $I(\boldsymbol{\theta})$ is finite and positive definite for all values of $\boldsymbol{\theta}$.

(**A2**): There exists functions G_{jkl} such that

$$\frac{\partial^3 L(\boldsymbol{\theta}; \boldsymbol{v}_i)}{\partial \theta_j \partial \theta_k \partial \theta_l} \leq G_{jkl}(\boldsymbol{v}_i) \quad \forall \boldsymbol{\theta}, \tag{17}$$

where $g_{jkl} = E_{\boldsymbol{\theta}_0}(G_{jkl}(\boldsymbol{v}_i)) < \infty$ for all j, k, l.

Theorem 1. *Under (A1) and (A2), if $v_{1n} \longrightarrow \infty$, $v_{2n} \longrightarrow \infty$, $v_{1n}/\sqrt{n} \longrightarrow 0$, $v_{2n}/\sqrt{n} \longrightarrow 0$, then the AMAZonn estimators obey the following oracle properties:*

(1) *consistency in variable selection:* $\lim_n P(\widehat{\mathscr{A}} = \mathscr{A}) = 1$, *and*

(2) *asymptotic normality of the nonzero coefficients:* $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \longrightarrow_d \mathcal{N}(\mathbf{0}, I_{11}^{-1})$.

4. Simulation Studies

In this section, we conduct simulation studies to evaluate the finite sample performance of AMAZonn. For comparison purposes, the performance of both AMAZonn and EM LASSO is evaluated. For each simulated dataset, the associated tuning parameters are selected by the minimum BIC criterion for all the methods under consideration. All the examples reported in this section are obtained from published papers with slight modifications within the scope of the current study [11, 12].

Specially, three scenarios are considered: in the data generating models of Simulations 1 and 2, we consider all continuous predictors, whereas in Simulation 3, both continuous and categorical variables are included. For each experimental instance, we randomly partition the data into training and test sets: models are fitted on the training set and prediction errors based on mean absolute scaled error (MASE) are calculated on the held-out samples in the test set. For an exhaustive comparison, we considered three sets of sample sizes $\{n_T, n_P\} = \{200, 200\}, \{500, 500\}$, and $\{1000, 1000\}$, where n_T and n_P represent the size of the training and test data, respectively. The corresponding regression coefficients and intercepts are chosen so that a desired level of sparsity proportion (ϕ) is achieved. In order to remain as model-agnostic as possible, we consider the same set of predictors for both zero and count submodels (i.e., $\mathbf{X} = \mathbf{Z}$). Such models are common in many practical applications where no domain-specific prior information about the zero-inflation mechanism is available. Below we provide the detailed data generation steps for both simulation examples:

Simulation 1.

(1) Generate 40 predictors from the multivariate normal distribution with mean vector $\mathbf{0}$, variance vector $\mathbf{1}$,

and variance-covariance matrix V, where the elements of V are $\rho^{|j_1 - j_2|} \forall j_1 \neq j_2 = 1, \ldots, 40$. The values of pairwise correlation ρ varies from 0 (uncorrelated) to 0.4 (moderate collinearity) to 0.8 (high collinearity).

(2) The count and zero regression parameters are chosen as follows:

$$(\beta_1, \ldots, \beta_8)$$
$$= (-1, -0.5, -0.25, -0.1, 0.1, 0.25, 0.5, 0.75)',$$
$$(\beta_9, \ldots, \beta_{16}) = (0.2, \ldots, 0.2)',$$
$$(\beta_{17}, \ldots, \beta_{40}) = (0, \ldots, 0)',$$
$$(\gamma_1, \ldots, \gamma_8) \tag{18}$$
$$= (-0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.4)',$$
$$(\gamma_9, \ldots, \gamma_{16}) = (0.2, \ldots, 0.2)',$$
$$(\gamma_{17}, \ldots, \gamma_{40}) = (0, \ldots, 0)'.$$

(3) The zero-inflated count outcome y is simulated according to (1) with the above parameters and input data.

Simulation 2. It is similar to Simulation 1 except that the count and zero regression parameters are chosen as follows:

$$(\beta_1, \ldots, \beta_{10}) = (0.05, -0.25, 0.05, 0.25,$$
$$-0.15, 0.15, 0.25, -0.2, 0.25, -0.25)',$$
$$(\beta_{11}, \ldots, \beta_{30}) = (-0.2, 0.25, 0.15,$$
$$-0.25, 0.2, 0, \ldots, 0)',$$
$$(\beta_{31}, \ldots, \beta_{40}) = (0.27, -0.27, 0.14, 0.2,$$
$$-0.2, 0.2, 0, \ldots, 0)', \tag{19}$$
$$(\gamma_1, \ldots, \gamma_{10}) = (-0.5, -0.4, -0.3, -0.2,$$
$$-0.1, 0.1, 0.2, 0.3, 0.4, 0.5)',$$
$$(\gamma_{11}, \ldots, \gamma_{30}) = (-0.2, 0.25, 0.15, -0.25, 0.2, 0, \ldots, 0)',$$
$$(\gamma_{31}, \ldots, \gamma_{40}) = (0.27, -0.27, -0.14, -0.2,$$
$$-0.2, 0.2, 0, \ldots, 0)'.$$

Simulation 3.

(1) First simulate $X_1, \ldots, X_6$ independently from the standard normal distribution. Consider the following as the continuous predictors: $\{X_1\}, \{X_2\}$, $\{X_3, X_3^2, X_3^3\}, \{X_4\}, \{X_5\}$ and $\{X_6, X_6^2, X_6^3\}$.

(2) Simulate 5 continuous variables from the multivariate normal distribution with mean 0, variance 1, and AR(ρ) correlation structure for varying ρ in $\{0, 0.4,$

TABLE 2: Results of Simulations 1–3. Average (over 200 replications) of Mean Absolute Scale Errors (MASEs) of AMAZonn and EM LASSO is reported.

ρ	φ	n	Simulation 1			Simulation 2			Simulation 3		
			AMAZonn - EM SEAL	AMAZonn - EM AL	EM LASSO	AMAZonn - EM SEAL	AMAZonn - EM AL	EM LASSO	AMAZonn - EM SEAL	AMAZonn - EM AL	EM LASSO
	0.3	200	0.91	0.92	0.91	0.60	0.61	0.62	0.97	1.03	1.00
		500	0.90	0.90	0.91	0.60	0.60	0.61	0.97	0.99	1.00
		1000	0.91	0.91	0.92	0.58	0.58	0.60	0.97	0.98	0.98
	0.4	200	1.12	1.13	1.12	0.75	0.75	0.76	1.18	1.23	1.23
0.0		500	1.05	1.05	1.06	0.73	0.73	0.74	1.11	1.17	1.20
		1000	1.03	1.03	1.04	0.71	0.71	0.72	1.11	1.16	1.16
	0.5	200	1.28	1.28	1.27	0.87	0.87	0.87	1.40	1.46	1.46
		500	1.16	1.16	1.17	0.84	0.84	0.85	1.28	1.33	1.36
		1000	1.15	1.15	1.19	0.80	0.80	0.82	1.23	1.30	1.31
	0.3	200	1.05	1.06	1.09	0.63	0.63	0.63	0.96	1.01	0.99
		500	1.04	1.04	1.05	0.61	0.61	0.62	0.95	0.97	0.99
		1000	0.96	0.96	0.98	0.58	0.58	0.59	0.97	0.98	0.98
	0.4	200	1.21	1.22	1.22	0.75	0.75	0.76	1.19	1.22	1.23
0.4		500	1.18	1.18	1.21	0.71	0.71	0.72	1.14	1.19	1.22
		1000	1.13	1.14	1.18	0.68	0.68	0.70	1.13	1.18	1.17
	0.5	200	1.42	1.43	1.42	0.83	0.84	0.83	1.34	1.40	1.43
		500	1.26	1.26	1.32	0.80	0.81	0.82	1.27	1.32	1.35
		1000	1.23	1.23	1.30	0.75	0.75	0.77	1.27	1.34	1.33
	0.3	200	1.32	1.31	1.36	0.62	0.63	0.63	0.96	1.00	1.01
		500	1.13	1.13	1.23	0.59	0.59	0.61	0.97	0.99	1.01
		1000	1.13	1.13	1.21	0.56	0.56	0.58	0.95	0.96	0.96
	0.4	200	1.52	1.52	1.58	0.71	0.72	0.72	1.18	1.21	1.23
0.8		500	1.31	1.32	1.45	0.68	0.68	0.69	1.12	1.19	1.20
		1000	1.24	1.24	1.37	0.64	0.64	0.64	1.12	1.17	1.16
	0.5	200	1.56	1.58	1.61	0.78	0.78	0.78	1.37	1.42	1.44
		500	1.44	1.45	1.65	0.73	0.73	0.76	1.29	1.34	1.39
		1000	1.33	1.36	1.52	0.69	0.70	0.69	1.26	1.33	1.34

0.8} as before, and quantile-discretize each of them into 5 new variables based on their quantiles: $(-\infty, \Phi^{-1}(1/5)]$, $(\Phi^{-1}(1/5), \Phi^{-1}(2/5)]$, $(\Phi^{-1}(2/5), \Phi^{-1}(3/5)]$, $(\Phi^{-1}(3/5), \Phi^{-1}(4/5)]$, and $(\Phi^{-1}(4/5), \infty)$, leading to a total of 20 categorical variables.

(3) With the above input data and parameters, the zero-inflated count outcome y is simulated according to (1), where the two sets of regression parameters are chosen as follows:

$$(\beta_1, \ldots, \beta_{10}) = \left(0, 0, 0.1, 0.2, 0.1, 0, 0, \frac{2}{3}, -1, \frac{1}{3}\right),$$

$$(\beta_{11}, \ldots, \beta_{30}) = (-2, -1, 1, 2, 0, \ldots, 0),$$

$$(\gamma_1, \ldots, \gamma_{10}) = \left(0, 0, 0.1, 0.2, 0.1, 0, 0, \frac{2}{3}, -1, \frac{1}{3}\right),$$ (20)

$$(\gamma_{11}, \ldots, \gamma_{30}) = (-2, -1, 1, 2, 0, \ldots, 0).$$

The resulting performance measures iterated over 200 replications (Table 2) reveal that AMAZonn performs as well as or better than EM LASSO in most of the simulation scenarios. For highly collinear designs, AMAZonn - EM SEAL stands out to be the best estimator for almost every sample size and zero-inflation proportion, highlighting the benefit of incorporating data-adaptive weights based on both ML estimates and their standard errors. This phenomenon is also apparent in the analysis of German health care data in Section 5, where the parameter estimates from the AMAZonn - EM SEAL method appear to be more parsimonious than those from other methods.

5. Application to German Health Care Demand Data

Next, we apply our method to the German health care demand data [3], a subset of the German Socioeconomic Panel (GSOEP) dataset [13], which has also been used for

TABLE 3: Summary of predictors in German health care demand data.

Variables	Mean (sd) or Frequency	Description
health	6.84 (2.19)	health satisfaction: 0 (low) - 10 (high)
handicap	216 / 1596	1 : handicap, 0 : otherwise
hdegree	6.16 (18.49)	degree of handicap in percentage points
married	1257 / 555	1 : married, 0 : otherwise
schooling	11.83 (2.49)	years of schooling
hhincome	4.52 (2.13)	household income per month in German marks/1000
children	703 / 1109	1 : children under 16 in household, 0 : otherwise
self	153 / 1659	1 : self-employed, 0 : otherwise
civil	198 / 1614	1 : civil servant, 0 : otherwise
bluec	566 / 1246	1 : blue collar employee, 0 : otherwise
employed	1506 / 306	1 : employed, 0 : otherwise
public	1535 / 277	1 : public health insurance, 0 : otherwise
addon	33 / 1779	1 : addon insurance, 0 : otherwise
age30	1480 / 332	1 if age $\geq$ 30
age35	1176 / 636	1 if age $\geq$ 35
age40	919 / 893	1 if age $\geq$ 40
age45	716 / 1096	1 if age $\geq$ 45
age50	535 / 1227	1 if age $\geq$ 50
age55	351 / 1461	1 if age $\geq$ 55
age60	147 / 1665	1 if age $\geq$ 60

TABLE 4: Model selection performance of EM LASSO and AMA-Zonn on German health care data.

Methods	BIC	Time (in seconds)
EM LASSO	9062.744	50.252
AMAZonn - EM AL	9002.487	26.215
AMAZonn - EM SEAL	**8982.924**	26.528

illustration purposes in previous studies [3, 14]. The original data contains number of doctor office visits for $1,812$ West German men aged 25 to 65 years in the last three months of 1994 (response variable of interest), which is supplemented with complementary information on twelve annual waves from 1984 to 1995 including health care utilization, current employment status, and insurance arrangements under which subjects are protected [3]. The goal of the original study was to investigate how the employment characteristics of the German nationals are related to their health care demand. The distribution of the dependent variable (Figure 1) reveals that many doctor visits are zeros (41.2%), confirming that classical methods such as Poisson regression are inappropriate for modeling this outcome.

In the model fitting process, along with the original variables, the interactions between age groups and health condition are also considered, resulting in 28 candidate predictors (Table 3). The fitting results from the full models indicate that both EM adaptive LASSO methods provide competitive model selection performance (Table 4), often leading to sparser model selection than EM LASSO (Table 5). In addition, the AMAZonn - EM SEAL method appears to choose even fewer numbers of variables. Such feature of AMAZonn - EM SEAL can be appealing in many practical situations, where data collinearity between variables is a concern and a more aggressive feature selection is desired. While the computational overheads of both EM adaptive LASSO methods are similar, they are an order of magnitude faster than EM LASSO (Table 4), further confirming that AMAZonn offers a viable alternative to existing methods.

6. Discussion

In recent years, there has been a huge influx of zero-inflated count measurements spanning several disciplines including biology, public health, and medicine. This has motivated the widespread use of zero-inflated count models in many practical applications such as metagenomics, single-cell RNA sequencing, and health care research. In this article, we propose the AMAZonn method for adaptive variable selection in ZIP regression models. Both our simulation and real data experience suggest that AMAZonn can outperform EM LASSO under a variety of regression settings while maintaining the desired theoretical properties and computational convenience. Our preliminary results are rather encouraging, and for practical purposes, we provide a publicly available R package implementing this method: https://github.com/himelmallick/AMAZonn.

We envision a number of improvements that may further refine AMAZonn's performance. While AMAZonn relies on ML estimates to construct the weight vector, these estimates may not be available in ultrahigh dimensions [7]. Alternative initialization schemes could further improve on this such as the ridge estimates [15]. Extension to other zero-inflated models such as marginalized zero-inflated count regression [16, 17], two-part and hurdle models [18], and multiple-inflation models [19] can form a useful basis for further

TABLE 5: Estimated coefficients from the best-fitting ZIP models in German health care demand data analysis.

(a)

							Count Coefficients							
Methods	(Intercept)	hlth	handicap	ddegree	married	schooling	hhincome	children	self	civil	bluec	employed	public	addon
EM LASSO	2.322	-0.14	0.207	-0.002	-0.97	0.0	0.0	0.078	-0.178	-0.166	0.038	-0.106	0.089	0.205
AMAZonn - EM AL	2.305	-0.135	0.111	0.0	-0.947	0.0	0.0	0.079	-0.234	-0.245	0.0	-0.059	0.043	0.205
AMAZonn - EM SEAL	2.378	-0.142	0.098	0.0	-0.066	0.0	0.0	0.046	-0.189	-0.222	0.0	-0.055	0.0	0.14

							Count Coefficients							
Methods	ag30	ag35	ag40	ag45	ag50	ag55	ag60	ag30:hlth	ag35:hlth	ag40:hlth	ag45:hlth	ag50:hlth	ag55:hlth	ag60:hlth
EM LASSO	0.0	0.0	0.0	0.586	0.0	-0.27	0.081	0.0	0.0	-0.006	-0.076	0.006	0.082	-0.034
AMAZonn - EM AL	0.0	0.0	-0.047	0.769	0.0	-0.402	0.099	0.0	0.0	0.0	-0.101	0.0	0.106	-0.034
AMAZonn - EM SEAL	0.0	0.0	0.0	0.586	0.0	-0.25	0.0	0.0	0.0	0.0	-0.081	0.0	0.081	-0.017

(b)

							Zero Coefficients							
Methods	(Intercept)	hlth	handicap	ddegree	married	schooling	hhincome	children	self	civil	bluec	employed	public	addon
EM LASSO	-2.193	-0.262	-0.098	-0.003	-0.121	0.0	-0.012	0.253	0.112	0.134	0.0	0.0	-0.012	0.0
AMAZonn - EM AL	-2.226	-0.261	-0.162	0.0	0.0	0.0	0.0	0.163	0.0	0.0	0.0	0.0	0.0	0.0
AMAZonn - EM SEAL	-2.403	-0.283	0.0	0.0	-0.053	0.0	0.0	0.238	0.0	0.0	0.0	0.0	0.0	0.0

							Zero Coefficients							
Methods	ag30	ag35	ag40	ag45	ag50	ag55	ag60	ag30:hlth	ag35:hlth	ag40:hlth	ag45:hlth	ag50:hlth	ag55:hlth	ag60:hlth
EM LASSO	0.0	0.0	0.0	0.0	-0.459	0.0	-0.217	0.013	0.0	0.005	0.0	0.0	0.023	0.0
AMAZonn - EM AL	0.047	0.0	0.065	0.009	-0.527	0.0	-0.198	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AMAZonn - EM SEAL	0.0	0.0	0.0	0.0	-0.443	0.0	0.0	0.009	0.0	0.0	0.0	0.0	0.0	0.0

FIGURE 1: Number of doctor office visits in the German health care data.

investigations. Although we only focused on variable selection for fixed effects models, future work could include an extension to other regularization problems such as grouped variable selection [12, 20] as well as sparse mixed effects models [21].

Appendix

Proof. It is to be noted that both logistic and Poisson distributions belong to the exponential family. Since the objective function in (10) can be decomposed into weighted logistic and Poisson log-likelihoods (each belonging to the GLM family without the penalties), Theorem 1 is the direct application of Theorem 4 in Zou [22]. Therefore, if $\nu_{1n} \longrightarrow \infty$, $\nu_{2n} \longrightarrow \infty$, $\nu_{1n}/\sqrt{n} \longrightarrow 0$, and $\nu_{2n}/\sqrt{n} \longrightarrow 0$, then both the AMAZonn - EM AL and AMAZonn - EM SEAL estimators hold the oracle properties: with probability tending to 1, the estimate of zero coefficients is 0, and the estimate for nonzero coefficients has an asymptotic normal distribution with mean being the true value and variance which approximately equals the submatrix of the Fisher information matrix containing nonzero coefficients. Hence the proof is complete. $\square$

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Prithish Banerjee, Broti Garai, and Himel Mallick contributed equally to this work.

Acknowledgments

References

[1] W. H. Greene, *Accounting for excess zeros and sample selection in Poisson and negative binomial regression models*, New York University, New York, NY, 1994.

[2] D. Lambert, "Zero-inflated poisson regression, with an application to defects in manufacturing," *Technometrics*, vol. 34, no. 1, pp. 1–14, 1992.

[3] Z. Wang, S. Ma, and C.-Y. Wang, "Variable selection for zero-inflated and overdispersed data with application to health care demand in Germany," *Biometrical Journal*, vol. 57, no. 5, pp. 867–884, 2015.

[4] Z. Wang, S. Ma, C.-Y. Wang, M. Zappitelli, P. Devarajan, and C. Parikh, "EM for regularized zero-inflated regression models with applications to postoperative morbidity after cardiac surgery in children," *Statistics in Medicine*, vol. 33, no. 29, pp. 5192–5208, 2014.

[5] H. Mallick and H. K. Tiwari, "EM adaptive LASSO-a multilocus modeling strategy for detecting SNPs associated with zero-inflated count phenotypes," *Frontiers in Genetics*, vol. 7, 2016.

[6] Y. Tang, L. Xiang, and Z. Zhu, "Risk Factor Selection in Rate Making: EM Adaptive LASSO for Zero-Inflated Poisson Regression Models," *Risk Analysis*, vol. 34, no. 6, pp. 1112–1127, 2014.

[7] W. Qian and Y. Yang, "Model selection via standard error adjusted adaptive lasso," *Annals of the Institute of Statistical Mathematics*, vol. 65, no. 2, pp. 295–318, 2013.

[8] Z. Y. Algamal and M. H. Lee, "Adjusted Adaptive LASSO in High-dimensional Poisson Regression Model," *Modern Applied Science (MAS)*, vol. 9, no. 4, 2014.

[9] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software* , vol. 33, no. 1, pp. 1–22, 2010.

[10] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[11] J. Huang, S. Ma, H. Xie, and C.-H. Zhang, "A group bridge approach for variable selection," *Biometrika*, vol. 96, no. 2, pp. 339–355, 2009.

[12] S. Chatterjee, S. Chowdhury, H. Mallick, P. Banerjee, and B. Garai, "Group regularization for zero-inflated negative binomial regression models with an application to health care demand in Germany," *Statistics in Medicine*, vol. 37, no. 20, pp. 3012–3026, 2018.

[13] R. T. Riphahn, A. Wambach, and A. Million, "Incentive effects in the demand for health care: A bivariate panel count data estimation," *Journal of Applied Econometrics*, vol. 18, no. 4, pp. 387–405, 2003.

[14] M. Jochmann, "What belongs where? Variable selection for zero-inflated count models with an application to the demand for health care," *Computational Statistics*, vol. 28, no. 5, pp. 1947–1964, 2013.

[15] A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[16] D. L. Long, J. S. Preisser, A. H. Herring, and C. E. Golin, "A marginalized zero-inflated Poisson regression model with overall exposure effects," *Statistics in Medicine*, vol. 33, no. 29, pp. 5151–5165, 2014.

[17] V. A. Smith and J. S. Preisser, "Direct and flexible marginal inference for semicontinuous data," *Statistical Methods in Medical Research*, vol. 26, no. 6, pp. 2962–2965, 2016.

[18] V. A. Smith, B. Neelon, J. S. Preisser, and M. L. Maciejewski, "A marginalized two-part model for longitudinal semicontinuous data," *Statistical Methods in Medical Research*, vol. 26, no. 4, pp. 1949–1968, 2017.

[19] X. Su, J. Fan, R. A. Levine, X. Tan, and A. Tripathi, "Multiple-inflation Poisson model with L_1 regularization," *Statistica Sinica*, vol. 23, no. 3, pp. 1071–1090, 2013.

[20] S. Chowdhury, S. Chatterjee, H. Mallick, H. Banerjee, and B. Garai, "Group regularization for zero-inflated poisson regression models with an application to insurance ratemaking," *Journal of Applied Statistics*, 2018, In Press.

[21] A. Groll and G. Tutz, "Variable selection for generalized linear mixed models by L_1-penalized estimation," *Statistics and Computing*, vol. 24, no. 2, pp. 137–154, 2014.

[22] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418–1429, 2006.

Local Influence Analysis for Quasi-Likelihood Nonlinear Models with Random Effects

Tian Xia,[1] Jiancheng Jiang [ID],[2] and Xuejun Jiang [ID][3]

[1]*School of Mathematics and Statistics, Guizhou University of Finance and Economics, Guiyang 550025, China*
[2]*Department of Mathematics and Statistics, University of North Carolina at Charlotte, NC 28223, USA*
[3]*Department of Mathematics, Southern University of Science and Technology, Shenzhen 518055, China*

Correspondence should be addressed to Xuejun Jiang; jiangxj@sustc.edu.cn

We propose a quasi-likelihood nonlinear model with random effects, which is a hybrid extension of quasi-likelihood nonlinear models and generalized linear mixed models. It includes a wide class of existing models as examples. A novel penalized quasi-likelihood estimation method is introduced. Based on the Laplace approximation and a penalized quasi-likelihood displacement, local influence of minor perturbations on the data set is investigated for the proposed model. Four concrete perturbation schemes are considered in the local influence analysis. The effectiveness of the proposed methodology is illustrated by some numerical examinations on a pharmacokinetics dataset.

1. Introduction

In this paper, we propose a quasi-likelihood nonlinear model with random effects (QLNMWRE) and investigate local influence of the model. The QLNMWRE is a hybrid generalization of quasi-likelihood nonlinear models [1, 2] and generalized linear mixed models, and it combines the advantages of both models. Generalized linear mixed models (GLMMs) are extensions of the well-known generalized linear models [3] by adding random effects to the linear predictor. GLMMs are effective and flexible for modeling nonnormal responses, repeated measurements, and other forms of clustered data. Efficient inference for the GLMMs depends on the underlying distribution of the data. Nevertheless, the exact distribution is rarely known in practice. In contrast, the quasi-likelihood method [4] requires only the first and second moments assumptions about the distribution and has been widely applied in the theory and practice of statistics (see, e.g., [5–8]).

Detecting influential observations is important in data analysis. The local influence analysis has become a general tool for detecting a group of points with great influence on the fitted model through perturbation schemes [9]. This approach has been successfully applied in many models,

such as mixed models [10, 11], generalized linear models [12], generalized linear mixed models [13], exponential family nonlinear models [14], nonlinear reproductive dispersion mixed model [15], nonlinear mixed-effect models [16, 17], and multivariate threshold time series models [1]. However, in these references the local influence method severely depends on the likelihood displacement, which is rarely known in practice. Instead, quasi-likelihood methods do not require the exact likelihood function except the first two moments of the response variables. Hence, we conduct influence analysis of the QLNMWRE using a novel penalized quasi-likelihood estimation method. The proposed methodology is illustrated by analyzing the pharmacokinetics dataset.

The remainder of this paper is organized as follows. In Section 2, we introduce the QLNMWRE and the corresponding estimation method. A Fisher-scoring iteration algorithm is advanced to calculate the estimators. In Section 3, a penalized quasi-likelihood displacement (PQLD) is proposed, and assessment of local influence under four different perturbation schemes is investigated. In Section 4, the pharmacokinetics dataset is employed to illustrate the effectiveness of the proposed methodology. Finally, we make discussion in Section 5.

2. Models and Estimation Method

Let Y be a response vector of length n, and let X and Z be $n \times k$ and $n \times q$ matrices of explanatory variables associated with fixed and random effects, respectively. Conditional on the q-dimensional vector of random effects, $\mathbf{b}$, the observations, $\{y_i, i = 1, \ldots, n\}$, are independent and satisfy that

$$E(y_i \mid \mathbf{b}) = h(\mathbf{x}_i, \boldsymbol{\beta}) + \mathbf{z}_i^T \mathbf{b} \triangleq \mu_i$$

$$\text{var}(y_i \mid \mathbf{b}) = \sigma^2 v(\mu_i), \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$ $(p < n)$ is an unknown parameter vector defined in a compact subset $\mathscr{B} \subset R^p$, $\mathbf{x}_i$ and $\mathbf{z}_i$ are defined in a subset $\mathscr{X}$ of R^k and a subset $\mathscr{Z}$ of R^q, respectively, $v(\cdot)$ is a known variance function, σ^2 is a dispersion parameter that is known or can be estimated separately, $h(\cdot, \cdot)$ is a continuously differentiable function such that the derivative matrix $D = D_{\beta}(\boldsymbol{\beta}) = \partial h(\boldsymbol{\beta})/\partial \boldsymbol{\beta}^T$ has rank p for all $\boldsymbol{\beta}$, with $h(\boldsymbol{\beta}) = (h(\mathbf{x}_1, \boldsymbol{\beta}), \ldots, h(\mathbf{x}_n, \boldsymbol{\beta}))^T$, and the random effects $\mathbf{b}$ are assumed to be multivariate normally distributed:

$$\mathbf{b} \sim N(0, \sigma^2 \Sigma), \tag{2}$$

with Σ being a known nonnegative definite matrix. Following [2, 3, 18, 19], the conditional log quasi-likelihood on $\mathbf{b}$ is defined as

$$Q_1(\boldsymbol{\beta}; \mathbf{y}) = \sum_{i=1}^{n} \int_{y_i}^{\mu_i} \frac{y_i - t}{\sigma^2 v(t)} dt, \tag{3}$$

where $\mu_i = h(\mathbf{x}_i, \boldsymbol{\beta}) + \mathbf{z}_i^T \mathbf{b} \triangleq \mu_i(\boldsymbol{\beta})$. The model defined by (1)-(3) is the so-called QLNMWRE.

Clearly, this QLNMWRE encompasses some important special cases. If $\Sigma = 0$, then the above model is just the quasi-likelihood nonlinear model discussed by [2]; if $\mu_i(\boldsymbol{\beta}) = h(\mathbf{x}_i^T \boldsymbol{\beta}) + \mathbf{z}_i^T \mathbf{b}$, and y_i are independently drawn from a one-parameter exponential family of distributions with density

$$\exp\{\theta_i^T y_i - k(\theta_i)\} d\gamma(y_i), \quad i = 1, \ldots, n, \tag{4}$$

where $\gamma(\cdot)$ is a measure, then it reduces to generalized linear models with random effects (see [20, 21]). Hence, the QLNMWRE is a hybrid extension of the quasi-likelihood nonlinear models and the generalized linear models with random effects.

Let $p(\mathbf{b}|\sigma^2)$ be a probability density function of random effect $\mathbf{b}$. Then the joint log quasi-likelihood function of $\mathbf{y} = (y_1, \ldots, y_n)^T$ and $\mathbf{b}$ is

$$Q(\boldsymbol{\beta}, \sigma^2; \mathbf{y}, \mathbf{b}) = Q_1(\boldsymbol{\beta}; \mathbf{y}) + \log p(\mathbf{b} \mid \sigma^2)$$

$$= \sum_{i=1}^{n} \int_{y_i}^{\mu_i} \frac{y_i - t}{\sigma^2 v(t)} dt - \frac{q}{2} \log(2\pi\sigma^2) \tag{5}$$

$$- \frac{1}{2} \log |\Sigma| - \frac{1}{2\sigma^2} \mathbf{b}^T \Sigma^{-1} \mathbf{b}.$$

Similar to the relationship between the joint log-likelihood function and the marginal log-likelihood function, we have

$$Q(\boldsymbol{\beta}, \sigma^2; \mathbf{y}, \mathbf{b}) = Q(\boldsymbol{\beta}, \sigma^2; \mathbf{y}) + Q(\boldsymbol{\beta}, \sigma^2; \mathbf{b} \mid \mathbf{y}), \tag{6}$$

where $Q(\boldsymbol{\beta}, \sigma^2; \mathbf{y})$ is the marginal log quasi-likelihood function of $\mathbf{y}$ and $Q(\boldsymbol{\beta}, \sigma^2; \mathbf{b} \mid \mathbf{y})$ is the log quasi-likelihood function of $\mathbf{b}$ given $\mathbf{y}$, i.e.,

$$Q(\boldsymbol{\beta}, \sigma^2; \mathbf{y}) = \log \int \exp\{Q(\boldsymbol{\beta}, \sigma^2; \mathbf{y}, \mathbf{b})\} d\mathbf{b},$$

$$Q(\boldsymbol{\beta}, \sigma^2; \mathbf{b} \mid \mathbf{y}) = \log \left\{ \frac{\exp\{Q(\boldsymbol{\beta}, \sigma^2; \mathbf{y}, \mathbf{b})\}}{\int \exp\{Q(\boldsymbol{\beta}, \sigma^2; \mathbf{y}, \mathbf{b})\} d\mathbf{b}} \right\}. \tag{7}$$

Following the arguments in [20], the integrated log quasi-likelihood function used to estimate $\boldsymbol{\beta}$ is defined by

$$\exp\{Q(\boldsymbol{\beta}, \sigma^2; \mathbf{y})\} \propto |\Sigma|^{-1/2}$$

$$\cdot \int \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} d_i(y_i, \mu_i) - \frac{1}{2\sigma^2} \mathbf{b}^T \Sigma^{-1} \mathbf{b} \right\} d\mathbf{b}, \tag{8}$$

where $d_i(y; \mu) = -2 \int_{\mu}^{y} ((y - t)/v(t)) dt$ denotes the deviance measure of fit. If, conditional on $\mathbf{b}$, y_i is a member of the exponential family, then $-d_i(y_i; \mu_i)/(2\sigma^2)$ is the conditional log-likelihood of y_i given $\mathbf{b}$, and $E_{\mathbf{b}}[\sum_{i=1}^{n} d_i(y_i; \mu_i)/(2\sigma^2)]$ is the log-likelihood function.

In general, no analytical expressions are available for the integral in (8) and approximate techniques are needed. The simplest approach is the Laplace approximation [22, 23]. Obviously, the right-hand side of (8) is

$$c |\Sigma|^{-1/2} \int e^{-m(\mathbf{b})} d\mathbf{b}, \tag{9}$$

where $m(\mathbf{b}) = (1/2\sigma^2) \sum_{i=1}^{n} d_i(y_i, \mu_i) + (1/2\sigma^2) \mathbf{b}^T \Sigma^{-1} \mathbf{b}$. When the Laplace method is applied to approximate the integrated quasi-likelihood function (8), estimates of $\boldsymbol{\beta}$ for fixed σ^2 are obtained by maximizing the penalized quasi-likelihood (PQL) (8):

$$Q_p(\boldsymbol{\beta}) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} d_i(y_i, \widetilde{\mu}_i) - \frac{1}{2\sigma^2} \widetilde{\mathbf{b}}^T \Sigma^{-1} \widetilde{\mathbf{b}}$$

$$= \sum_{i=1}^{n} \int_{y_i}^{\widetilde{\mu}_i} \frac{y_i - t}{\sigma^2 v(t)} dt - \frac{1}{2\sigma^2} \widetilde{\mathbf{b}^T} \Sigma^{-1} \widetilde{\mathbf{b}} \tag{10}$$

$$= Q_1(\boldsymbol{\beta}; Y)|_{\mu_i = \widetilde{\mu}_i} - \frac{1}{2\sigma^2} \widetilde{\mathbf{b}}^T \Sigma^{-1} \widetilde{\mathbf{b}},$$

where $\widetilde{\mu}_i = f(\mathbf{x}_i, \boldsymbol{\beta}) + \mathbf{z}_i^T \widetilde{\mathbf{b}}$, and $\widetilde{\mathbf{b}} \triangleq \widetilde{\mathbf{b}}(\boldsymbol{\beta})$ is the root of $\partial m(\mathbf{b})/\partial \mathbf{b} = 0$ for fixed $\boldsymbol{\beta}$. We will use the penalized quasi-likelihood $Q_p(\boldsymbol{\beta})$ to estimate $\boldsymbol{\beta}$ and to conduct local influence analysis. To this end, we need the following assumptions.

Assumption A.

(i) $E(v(\mu_i))^{-1}(y_i - \mu_i)|_{\mu_i = \widetilde{\mu}_i} = 0, \forall i = 1, \ldots, n$;
(ii) there exists some constant $M > 0$ and some compact subset $\mathscr{B}_1 \subset \mathscr{B}$ such that

$$\sup_{i\geq 1, \beta\subseteq\mathcal{B}_1} E\left(\left(v\left(\mu_i\right)\right)^{-1}\left(y_i-\mu_i\right)\big|_{\mu_i=\tilde{\mu}_i}\right)^2 \leq M. \tag{11}$$

It is easily seen that Assumption A holds in generalized linear mixed models and exponential family nonlinear random effects models. Assumption A guarantees existence of the variance-covariance matrix of $\tilde{\mathbf{e}}$, where $\tilde{\mathbf{e}} = (\tilde{e}_1,\cdots,\tilde{e}_n)^T$ with $\tilde{e}_i = v(\mu_i)^{-1}(y_i-\mu_i)|_{\mu_i=\tilde{\mu}_i}$. Let $Z = (\mathbf{z}_1,\ldots,\mathbf{z}_n)^T$ and $K = \partial\tilde{\mathbf{e}}/\partial\boldsymbol{\mu}^T = \mathrm{diag}(k_1,\ldots,k_n)$, where

$$k_i = \frac{\partial}{\partial\mu_i}\left(\frac{y_i-\mu_i}{v\left(\mu_i\right)}\right)\Bigg|_{\mu_i=\tilde{\mu}_i}. \tag{12}$$

Put $\Omega_1 = K^{-1}-Z\Sigma Z^T, W = \partial^2 h(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T, \Omega = K_0^{-1}-Z\Sigma Z^T$, and $K_0 = E_y(K)$. Under Assumption A, we have the following result.

Theorem 1. *For the model defined by (1)-(3), conditional on $\tilde{\mathbf{b}}$, the quasi-score function, the quasi-observed information matrix, and the quasi-Fisher information matrix for $\boldsymbol{\beta}$ admit the following representations:*

$$S_n\left(\boldsymbol{\beta}\right) \triangleq \dot{Q}_p\left(\boldsymbol{\beta}\right) = \sigma^{-2}\left(\frac{\partial h\left(\boldsymbol{\beta}\right)}{\partial\boldsymbol{\beta}^T}\right)^T\tilde{\mathbf{e}} = \sigma^{-2}D^T\tilde{\mathbf{e}}, \tag{13}$$

$$H_n\left(\boldsymbol{\beta}\right) \triangleq -\ddot{Q}_p\left(\boldsymbol{\beta}\right) = -\frac{\partial^2 Q_p\left(\boldsymbol{\beta}\right)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}$$
$$= -\sigma^{-2}\left(\left[\tilde{\mathbf{e}}^T\right][W] + D^T\Omega_1 D\right), \tag{14}$$

$$F_n\left(\boldsymbol{\beta}\right) \triangleq E_y\left(-\ddot{Q}_p\left(\boldsymbol{\beta}\right)\right) = -\sigma^{-2}D^T\Omega^{-1}D, \tag{15}$$

where $[\cdot][\cdot]$ indicates the array multiplication.

Let $\hat{\boldsymbol{\beta}}_n$ denote the maximum quasi-likelihood estimator (MQLE) of $\boldsymbol{\beta}$, which is the solution of equation $\dot{Q}_p(\boldsymbol{\beta}) = 0$. Then the Fisher-scoring iteration method can be used for computing $\hat{\boldsymbol{\beta}}_n$ by iteratively solving the following equation (see [14, 24]):

$$\boldsymbol{\beta}^{i+1} = \boldsymbol{\beta}^i + F_n^{-1}\left(\boldsymbol{\beta}^i\right)S_n\left(\boldsymbol{\beta}^i\right)$$
$$= \boldsymbol{\beta}^i - \left(D^T\Omega^{-1}D\right)^{-1}D^T\tilde{\mathbf{e}} \quad (i = 0,1,2,\cdots) \tag{16}$$

where D, Ω, and $\tilde{\mathbf{e}}$ are all evaluated at $\boldsymbol{\beta}^i$ and $\mathbf{b}^i$.

On the other hand, it follows from (5) that the quasi-score function and the quasi-Fisher information matrix for $\mathbf{b}$ can be, respectively, expressed as

$$S\left(\mathbf{b}\right) \triangleq \frac{\partial Q\left(\boldsymbol{\beta},\sigma^2;\mathbf{y},\mathbf{b}\right)}{\partial\mathbf{b}} = \sigma^{-2}\left(Z^T\mathbf{e}-\Sigma^{-1}\mathbf{b}\right),$$

$$F\left(\mathbf{b}\right) \triangleq E_y\left(-\frac{\partial^2 Q\left(\boldsymbol{\beta},\sigma^2;\mathbf{y},\mathbf{b}\right)}{\partial\mathbf{b}\partial\mathbf{b}^T}\right) \tag{17}$$
$$= \sigma^{-2}\left(Z^T V^{-1}Z + \Sigma^{-1}\right),$$

where $\mathbf{e} = \mathbf{e}(\beta) = (e_1,\ldots,e_n)^T$ with $e_i = (y_i-\mu_i)/v(\mu_i)$, and

$$V^{-1} = \mathrm{diag}\left(v^{-1}\left(\mu_1\right),\ldots,v^{-1}\left(\mu_n\right)\right). \tag{18}$$

Hence, the Fisher-scoring iteration algorithm for computing the predictor of $\mathbf{b}^j$ under known $\boldsymbol{\beta}^j$ is given by

$$\mathbf{b}^{(j_i+1)} = \mathbf{b}^{(j_i)} + \left(Z^T V^{-1}Z + \Sigma^{-1}\right)^{-1}\left(Z^T\mathbf{e}-\Sigma^{-1}\mathbf{b}^{(j_i)}\right), \tag{19}$$
$$j_i = 0,1,2,\ldots,$$

where V and e are all evaluated at $\mathbf{b}^{(j_i)}$ and $\boldsymbol{\beta}^j$. As the iteration scheme (19) converges, $\mathbf{b}^{j_i}$ converges to $\mathbf{b}^j$.

In general, the choice of initial value $\boldsymbol{\beta}^0$ is important for the Fisher-scoring iteration algorithm. We use the algorithm in [2] for quasi-likelihood nonlinear models to find the starting values of parameter $\boldsymbol{\beta}$ for QLNMWRE with $b_0 = \cdots = b_q = 0$. Hence, the MQLE $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ can be obtained by solving (16) and (19) until convergence.

In order to investigate the statistical diagnostic measures for QLNMWRE, we rewrite (16)

$$\boldsymbol{\beta}^{i+1} = \left(D^T\Omega^{-1}D\right)^{-1}D^T\Omega^{-1}G\Big|_{\boldsymbol{\beta}^i}, \tag{20}$$

where $G = D\boldsymbol{\beta}^i - \Omega\tilde{\mathbf{e}}$. When $\boldsymbol{\beta}^i$ converges to $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}$ can be expressed as

$$\hat{\boldsymbol{\beta}} = \left(D^T\Omega^{-1}D\right)^{-1}D^{-1}\Omega^{-1}G\Big|_{\hat{\boldsymbol{\beta}}}, \tag{21}$$

where $G = D\hat{\boldsymbol{\beta}} - \Omega\hat{\mathbf{e}}$, D, Ω and $\hat{\mathbf{e}}$ are all evaluated at $\hat{\boldsymbol{\beta}}$.

3. Local Influence

The aim of local influence analysis is to investigate the behavior of some influence measure $T(\boldsymbol{\omega})$ when small perturbations are made in the model/data, where $\boldsymbol{\omega}$ is an m-dimensional vector of perturbations restricted to some open subset $\Theta \in R^m$. For simple statistical models, Cook constructed in [9] the likelihood displacement $LD(\boldsymbol{\omega})$ and used it to assess the local influence of a minor perturbation. Although this approach is very useful, serious difficulties are encountered when applying it to complicated models, because of the intractable likelihood function. For the sake of coping with those difficulties, some authors have considered alternatives to replace $LD(\boldsymbol{\omega})$. For instance, Zhu et al. proposed in [25] the Q-likelihood displacement and established an approach to assess local influence of statistical models with incomplete data, and Jung presented in [26] a quasi-likelihood displacement to obtain local influence analysis in generalized estimating equations. Inspired by [25, 26], we define in this work a new penalized quasi-likelihood displacement and then adapt the local influence approach introduced by [9] to the QLNMWRE.

Let $Q_p(\boldsymbol{\beta})$ and $Q_p(\boldsymbol{\beta}|\boldsymbol{\omega})$ be the penalized quasi-likelihood for the unperturbed and perturbed models, respectively. We assume that there is an $\boldsymbol{\omega}^0$ such as $Q_p(\boldsymbol{\beta}|\boldsymbol{\omega}^0) = Q_p(\boldsymbol{\beta})$. Let $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}(\boldsymbol{\omega})$ be the MQLE of $\boldsymbol{\beta}$ under the unperturbed

and perturbed models, respectively. Similar to the likelihood displacement [9], we define the penalized quasi-likelihood displacement (PQLD) as

$$PQLD(\boldsymbol{\omega}) = 2\left\{Q_p\left(\widehat{\boldsymbol{\beta}}\right) - Q_p\left(\widehat{\boldsymbol{\beta}}(\boldsymbol{\omega})\right)\right\}. \qquad (22)$$

The influence graph is defined as $\alpha(\boldsymbol{\omega}) = (\boldsymbol{\omega}^T, PQLD(\boldsymbol{\omega}))^T$. Following the approach developed in [9, 25, 26], the normal curvature C_l of $\alpha(\boldsymbol{\omega})$ at $\boldsymbol{\omega}^0$ in the direction of some unit vector $\mathbf{l}$ can be used to summarize the local behavior of the penalized quasi-likelihood displacement. As shown in [9], the normal curvature C_l in the unit direction $\mathbf{l}(\|\mathbf{l}\| = 1)$ at $\boldsymbol{\omega}^0$ is given by

$$C_l = 2\left|\mathbf{l}^T \ddot{F}\mathbf{l}\right|, \qquad (23)$$

where $\ddot{F} = -(\partial^2 Q_p(\boldsymbol{\beta} \mid \boldsymbol{\omega})/\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}^T)|_{\boldsymbol{\omega}=\boldsymbol{\omega}^0} = -\triangle^T \ddot{Q}_p^{-1}\triangle$, in which $\triangle = \partial^2 Q_p(\boldsymbol{\beta}|\boldsymbol{\omega})/\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^T$ is a $p \times m$ matrix evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}^0$, $\ddot{Q}_p = \partial^2 Q_p(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T$ is a $p \times p$ Hessian matrix evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$. The maximum curvature C_{max}, which is the largest absolute eigenvalue of $2\ddot{F}$, and the corresponding direction vector $\mathbf{l}_{max}$ are usually used for identifying locally influential observations. A large value of C_{max} is an indication of a serious local problem, and if the i-th element in $\mathbf{l}_{max}$ is relatively large special attention should be paid to the element being perturbed by ω_i. To apply the local influence method in [9] to the QLNMWRE, we consider the following four perturbation schemes.

3.1. Case-Weights Perturbation. Let $\boldsymbol{\omega}$ be an $n\times 1$ perturbation vector such that $\boldsymbol{\omega}^0 = (1,\cdots,1)^T$. The joint log quasi-likelihood function for the perturbed model is given by

$$Q\left(\boldsymbol{\beta},\sigma^2;\mathbf{y},\mathbf{b},\boldsymbol{\omega}\right) = C + \sum_{i=1}^{n}\omega_i \int_{y_i}^{\mu_i} \frac{y_i - t}{\sigma^2 v(t)}dt$$
$$- \frac{1}{2\sigma^2}\mathbf{b}^T\Sigma^{-1}\mathbf{b}, \qquad (24)$$

where $C = -(q/2)\log(2\pi\sigma^2) - (1/2)|\Sigma|$. Then the penalized quasi-likelihood function can be expressed as

$$Q_p\left(\boldsymbol{\beta} \mid \boldsymbol{\omega}\right) = \sum_{i=1}^{n}\omega_i \int_{y_i}^{\widetilde{\mu}_i} \frac{y_i - t}{\sigma^2 v(t)}dt - \frac{1}{2\sigma^2}\widetilde{\mathbf{b}}^T\Sigma^{-1}\widetilde{\mathbf{b}}, \qquad (25)$$

where $\widetilde{\mu}_i = h(\mathbf{x}_i, \boldsymbol{\beta}) + \mathbf{z}_i^T\widetilde{\mathbf{b}}$, $Z = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n)^T$, and $\widetilde{\mathbf{b}}$ satisfies

$$\sum_{i=1}^{n}\omega_i \frac{y_i - \widetilde{\mu}_i}{v(\widetilde{\mu}_i)}\mathbf{z}_i - \Sigma^{-1}\widetilde{\mathbf{b}} = 0. \qquad (26)$$

Hence, $\widetilde{\mathbf{b}} = \Sigma Z^T\overline{W}\widetilde{\mathbf{e}}$, where $\overline{W} = \text{diag}(\omega_1,\ldots,\omega_n)$. Then

$$\left.\frac{\partial^2 Q_p\left(\boldsymbol{\beta} \mid w\right)}{\partial\boldsymbol{\beta}\partial w_i}\right|_{\boldsymbol{\omega}^0,\widehat{\boldsymbol{\beta}}} = \sigma^{-2}D^T\delta_i\widehat{e}_i + \sigma^{-2}D^T K$$
$$\cdot K^{-1}\left(\Omega^{-1}K^{-1} - I\right)\delta_i\widehat{e}_i \qquad (27)$$
$$= \sigma^{-2}D^T\Omega_1^{-1}K^{-1}\delta_i\widehat{e}_i,$$

and thus

$$\triangle = \left.\frac{\partial^2 Q_p\left(\boldsymbol{\beta} \mid \boldsymbol{\omega}\right)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^T}\right|_{\boldsymbol{\omega}^0,\widehat{\boldsymbol{\beta}}} = \sigma^{-2}D^T\Omega_1^{-1}K^{-1}E, \qquad (28)$$

where $E = \text{diag}(\widehat{e}_1,\ldots,\widehat{e}_n)$.

3.2. Response Variable Perturbation. A perturbation of the response variables $(y_1,\cdots,y_n)^T$ is introduced by replacing y_i by $y_{i\omega} = y_i + \omega_i$, where $\boldsymbol{\omega} = (\omega_1,\cdots,\omega_n)^T$, and $\boldsymbol{\omega}^0 = (0,\ldots,0)^T$ represents the situation with no perturbation. In this case, the joint log quasi-likelihood function for the perturbed model is given by

$$Q\left(\boldsymbol{\beta},\sigma^2;\mathbf{y},\mathbf{b},\boldsymbol{\omega}\right) = C + \sum_{i=1}^{n}\int_{y_i+\omega_i}^{\mu_i} \frac{y_i + \omega_i - t}{\sigma^2 v(t)}dt$$
$$- \frac{1}{2\sigma^2}\mathbf{b}^T\Sigma^{-1}\mathbf{b}, \qquad (29)$$

where C is a constant. It follows from Section 2 that the penalized quasi-likelihood function is

$$Q_p\left(\boldsymbol{\beta} \mid \boldsymbol{\omega}\right) = \sum_{i=1}^{n}\int_{y_i+\omega_i}^{\widetilde{\mu}_i} \frac{y_i + \omega_i - t}{\sigma^2 v(t)}dt - \frac{1}{2\sigma^2}\widetilde{\mathbf{b}}^T\Sigma^{-1}\widetilde{\mathbf{b}}, \qquad (30)$$

where $\widetilde{\mu}_i = h(\mathbf{x}_i, \boldsymbol{\beta}) + \mathbf{z}_i^T\widetilde{\mathbf{b}}$, and $\widetilde{\mathbf{b}}$ satisfies

$$\sum_{i=1}^{n} \frac{y_i + \omega_i - \widetilde{\mu}_i}{v(\widetilde{\mu}_i)}\mathbf{z}_i - \Sigma^{-1}\widetilde{\mathbf{b}} = 0. \qquad (31)$$

It follows that $\widetilde{\mathbf{b}} = \Sigma Z^T(\widetilde{\mathbf{e}} + W_v)$, where $W_v = (w_{v1},\ldots,w_{vn})^T$ with $w_{vi} = (w_i/v(\mu_i))|_{\mu_i=\widetilde{\mu}_i}$. Then

$$\left.\frac{\partial^2 Q_p\left(\boldsymbol{\beta} \mid \boldsymbol{\omega}\right)}{\partial\boldsymbol{\beta}\partial\omega_i}\right|_{\boldsymbol{\omega}^0,\widehat{\boldsymbol{\beta}}} = \sigma^{-2}D^T\left.\left(\frac{\partial\widetilde{\mathbf{e}}}{\partial\omega_i} + \frac{\partial W_v}{\partial\omega_i}\right)\right|_{\boldsymbol{\omega}^0,\widehat{\boldsymbol{\beta}}}$$

$$= \sigma^{-2}D^T\left.\left(\left(\frac{\partial\widetilde{\mathbf{e}}}{\partial\widetilde{\boldsymbol{\mu}}^T}\right)^T\frac{\partial\widetilde{\boldsymbol{\mu}}}{\partial\omega_i} + \delta_i^T\frac{1}{v(\widetilde{\mu}_i)}\right)\right|_{\boldsymbol{\omega}^0,\widehat{\boldsymbol{\beta}}} \qquad (32)$$

$$= \sigma^{-2}D^T\left[\Omega_1^{-1}\left(K^{-1} - \Omega_1\right)\delta_i^T\frac{1}{v(\widetilde{\mu}_i)} + \delta_i^T\frac{1}{v(\widetilde{\mu}_i)}\right]$$

$$= \sigma^{-2}D^T\Omega_1^{-1}K^{-1}\delta_i^T\frac{1}{v(\widetilde{\mu}_i)},$$

and

$$\triangle = \left.\frac{\partial^2 Q_p\left(\boldsymbol{\beta} \mid w\right)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\omega}^T}\right|_{\boldsymbol{\omega}_0,\widehat{\boldsymbol{\beta}}} = \sigma^{-2}D^T\Omega_1^{-1}K^{-1}E^*, \qquad (33)$$

where $E^* = \text{diag}(e_1^*,\ldots,e_n^*)$ with $e_i^* = 1/v(\widetilde{\mu}_i)$ and $\widehat{\mu}_i = h(\mathbf{x}_i, \widehat{\boldsymbol{\beta}}) + \mathbf{z}_i^T\widehat{\mathbf{b}}$.

3.3. Explanatory Variables Perturbation. In this case, we focus on the perturbation of a specific explanatory variable. Under

this condition we have the perturbed explanatory matrix $X_{\boldsymbol{\omega}} = (\mathbf{x}_1, \cdots, \mathbf{x}_{t\omega}, \cdots, \mathbf{x}_n)^T$ with $\mathbf{x}_{t\omega} = \mathbf{x}_t + \boldsymbol{\omega}$, where $\mathbf{x}_t$ is a single explanatory variable of matrix $X_{\boldsymbol{\omega}}$ corresponding to y_t and $\boldsymbol{\omega}^0 = (0, \ldots, 0)^T$ denotes no perturbation. Then the joint log quasi-likelihood function for the perturbed model is

$$Q\left(\boldsymbol{\beta}, \sigma^2; \mathbf{y}, \mathbf{b}, \boldsymbol{\omega}\right) = C + \sum_{j=1}^{n} \int_{y_i}^{\mu_{j\omega}} \frac{y_i - t}{\sigma^2 v(t)} dt \\ - \frac{1}{2\sigma^2} \mathbf{b} \Sigma^{-1} \mathbf{b}, \tag{34}$$

where C is a constant, $\mu_{j\omega} = h(\mathbf{x}_j, \boldsymbol{\beta}) + \mathbf{z}_j^T \mathbf{b}$ $(j \neq i)$, and $\mu_{t\omega} = h(\mathbf{x}_{t\omega} \boldsymbol{\beta}) + \mathbf{z}_t^T \mathbf{b}$. It follows from Section 2 that

$$Q_p\left(\boldsymbol{\beta} \mid \boldsymbol{\omega}\right) = \sum_{j=1}^{n} \int_{y_i}^{\tilde{\mu}_{j\omega}} \frac{y_i - t}{\sigma^2 v(t)} dt - \frac{1}{2\sigma^2} \tilde{\mathbf{b}}^T \Sigma^{-1} \tilde{\mathbf{b}}, \tag{35}$$

where $\tilde{\mu}_{j\omega} = h(\mathbf{x}_j, \boldsymbol{\beta}) + \mathbf{z}_j^T \tilde{\mathbf{b}}$ $(j \neq t)$, $\tilde{\mu}_{t\omega} = h(\mathbf{x}_{t\omega}, \boldsymbol{\beta}) + \mathbf{z}_t^T \tilde{\mathbf{b}}$, and $\tilde{\mathbf{b}}$ satisfies

$$\sum_{j=1}^{n} \frac{y_i - \tilde{\mu}_{j\omega}}{v\left(\tilde{\mu}_{j\omega}\right)} \mathbf{z}_i - \Sigma^{-1} \tilde{\mathbf{b}} = 0. \tag{36}$$

Therefore, $\tilde{\mathbf{b}} = \Sigma Z^T \tilde{\mathbf{e}}_w$ and $\tilde{\mathbf{e}}_w = ((y_1 - \tilde{\mu}_1)/v(\tilde{\mu}_1), \cdots, (y_n - \tilde{\mu}_n)/v(\tilde{\mu}_n))^T$. Let $h_t^T = \partial h(\mathbf{x}_t, \beta)/\partial \mathbf{x}_t^T|_{\beta=\hat{\beta}}$ and $H_{bt} = \partial^2 h(\boldsymbol{\beta})/\partial \boldsymbol{\beta} \partial \mathbf{x}_t^T|_{\beta=\hat{\beta}}$. Then

$$\triangle = \left. \frac{\partial^2 Q_p\left(\boldsymbol{\beta} \mid \boldsymbol{\omega}\right)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^T} \right|_{\boldsymbol{\omega}^0, \hat{\boldsymbol{\beta}}}$$

$$= \sigma^{-2} \left[\hat{\mathbf{e}}_\omega^T \right] \left[\frac{\partial}{\partial \boldsymbol{\omega}^T} \left(\frac{\partial h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right) \right] \Big|_{\boldsymbol{\omega}^0, \hat{\boldsymbol{\beta}}} \tag{37}$$

$$+ \sigma^{-2} D^T \frac{\partial \tilde{\mathbf{e}}_w}{\partial \tilde{\boldsymbol{\mu}}} \frac{\partial \tilde{\boldsymbol{\mu}}}{\partial \boldsymbol{\omega}^T} \Big|_{\boldsymbol{\omega}^0, \hat{\boldsymbol{\beta}}}$$

$$= \sigma^{-2} \left[\hat{\mathbf{e}}^T \right] \left[H_{bt} \right] + \sigma^{-2} D^T \Omega_1^{-1} \delta_t h_t^T,$$

where $[\cdot][\cdot]$ indicates the array multiplication.

3.4. Perturbation of Covariates in Random Effects. Consider perturbing the data for the kth explanatory variable of Z, by modifying the data matrix Z as $Z_w = Z + \boldsymbol{\omega} \mathbf{d}_k^T$, where $\mathbf{d}_k$ is a q−vector with 1 at kth position and zeros elsewhere. Under this situation, the perturbed joint log quasi-likelihood can be expressed as

$$Q\left(\boldsymbol{\beta}, \sigma^2; \mathbf{y}, \mathbf{b} \mid \boldsymbol{\omega}\right) = C + \sum_{i=1}^{n} \int_{y_i}^{\mu_i} \frac{y_i - t}{\sigma^2 v(t)} dt \\ - \frac{1}{2\sigma^2} \mathbf{b}^T \Sigma^{-1} \mathbf{b}, \tag{38}$$

where C is a quantity that does not depend on $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$, and $\boldsymbol{\mu} = h(\boldsymbol{\beta}) + Z_w \mathbf{b}$. When $\boldsymbol{\omega}^0 = \mathbf{0}$, it indicates no perturbation. It follows from Section 2 that

$$Q_p\left(\boldsymbol{\beta} \mid \boldsymbol{\omega}\right) = \sum_{i=1}^{n} \int_{y_i}^{\tilde{\mu}_i} \frac{y_i - t}{\sigma^2 v(t)} dt - \frac{1}{2\sigma^2} \tilde{\mathbf{b}}^T \Sigma^{-1} \tilde{\mathbf{b}}, \tag{39}$$

where $\tilde{\boldsymbol{\mu}} = h(\boldsymbol{\beta}) + Z_w \tilde{\mathbf{b}}$, and $\tilde{\mathbf{b}}$ satisfies

$$Z_w^T \tilde{\mathbf{e}}_w - \Sigma^{-1} \tilde{\mathbf{b}} = 0, \tag{40}$$

and therefore, $\tilde{\mathbf{b}} = \Sigma Z_w^T \tilde{\mathbf{e}}_w$ with $\tilde{\mathbf{e}}_w = ((y_1 - \tilde{\mu}_1)/v(\tilde{\mu}_1), \cdots, (y_n - \tilde{\mu}_n)/v(\tilde{\mu}_n))^T$. Then

$$\left. \frac{\partial^2 Q_p\left(\boldsymbol{\beta} \mid \boldsymbol{\omega}\right)}{\partial \boldsymbol{\beta} \partial w_i} \right|_{\boldsymbol{\omega}^0, \hat{\boldsymbol{\beta}}} = \sigma^{-2} D^T \left(\frac{\partial \tilde{\mathbf{e}}_w}{\partial \tilde{\boldsymbol{\mu}}^T} \right)^T \frac{\partial \tilde{\boldsymbol{\mu}}}{\partial w_i} \Big|_{\boldsymbol{\omega}^0, \hat{\boldsymbol{\beta}}}$$

$$= \sigma^{-2} D^T K \left(I - Z\Sigma Z^T K \right)^{-1} \left[\delta_i \hat{b}_k + Z\Sigma d_k \hat{e}_i \right] \tag{41}$$

$$= \sigma^{-2} D^T \Omega_1^{-1} \left(\delta_i \hat{b}_k + Z\Sigma d_k \hat{e}_i \right)$$

Hence,

$$\triangle = \left. \frac{\partial^2 Q_p\left(\boldsymbol{\beta} \mid \boldsymbol{\omega}\right)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^T} \right|_{\boldsymbol{\omega}^0, \hat{\boldsymbol{\beta}}}$$

$$= \sigma^{-2} D^T \Omega_1^{-1} \left\{ \left[\delta_1 \hat{b}_k \quad \delta_2 \hat{b}_k \cdots \delta_n \hat{b}_k \right] \right. \tag{42}$$

$$+ Z\Sigma \left[d_k \hat{e}_1 \quad d_k \hat{e}_2 \cdots d_k \hat{e}_n \right] \right\} = \sigma^{-2} D^T \Omega_1^{-1} \left[I_n \hat{b}_k \right.$$

$$+ Z\Sigma d_k \hat{\mathbf{e}}^T \Big].$$

4. Numerical Results

To illustrate how to use the proposed methodology, we consider the data set reported by [27]. The data came from a study of the pharmacokinetics of indomethacin following bolus intravenous injection of the same dose in six human volunteers. For each subject, the plasma concentrations of indomethacin were measured at 11 time points from 15 min to 8 hours postinjection. Davidian et al. used nonlinear repeated model to analyze the dataset in [28]; we model it using the following QLNMWRE:

$$\mu_{ij} = f\left(x_{ij}, \boldsymbol{\beta}\right) + b_i \quad (i = 1, \cdots, 6; j = 1, \cdots, 11), \tag{43}$$

where response variables $y_{ij}|b_i$ belong to the Gumbel distribution (cf. [29]) with the density function

$$p\left(y_{ij} \mid b_i\right) = \exp\left\{ y_{ij} - \theta - \exp\left(y_{ij} - \theta\right) \right\}, \\ -\infty < y_{ij} < \infty, \quad -\infty < \theta < \infty, \tag{44}$$

$b_i \sim N(0, a)$, and $f(x, \boldsymbol{\beta}) = e^{\beta_1} \exp(-e^{\beta_2} x) + e^{\beta_3} \exp(-e^{\beta_4} x)$. By [29], we have $E(y) = \theta - \gamma$ and $\text{Var}(y) = \pi^2/6 = \sigma^2 v(\mu)$, where $\gamma = 0.5772$ is called the Euler constant, $\sigma^2 = \pi^2/6$ and $v(\cdot) = 1$. It is easily shown that Assumption
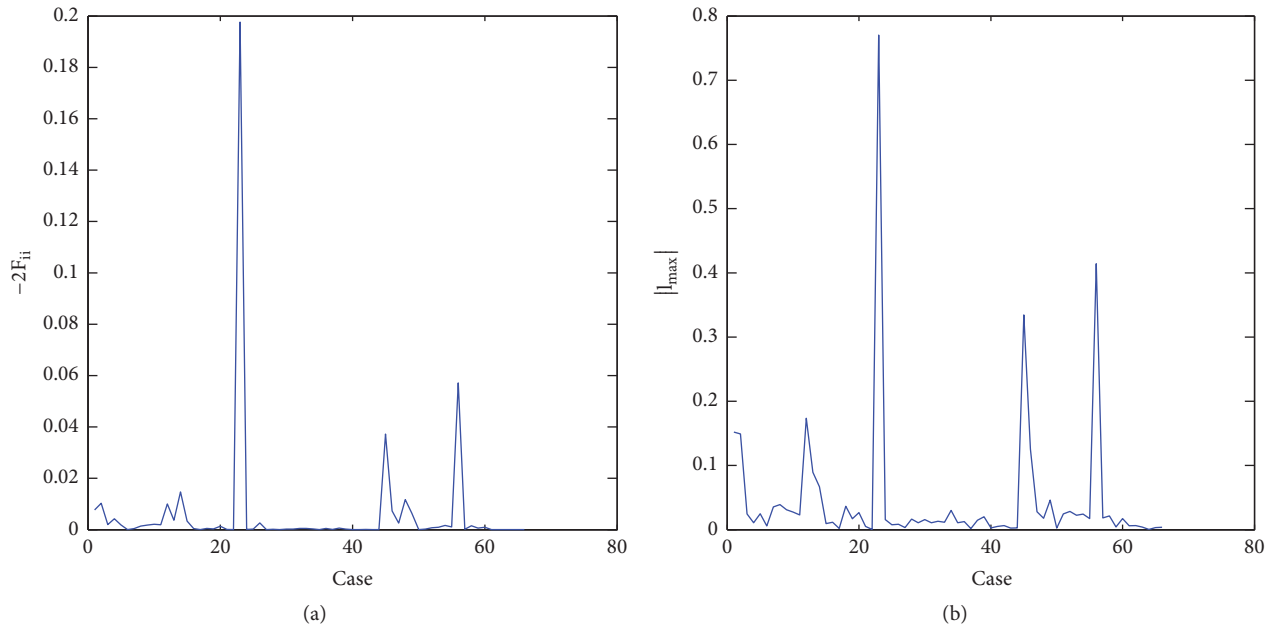
FIGURE 1: Index plots of $-F_{ii}$ and $|l_{max}|$ for case-weights perturbation.
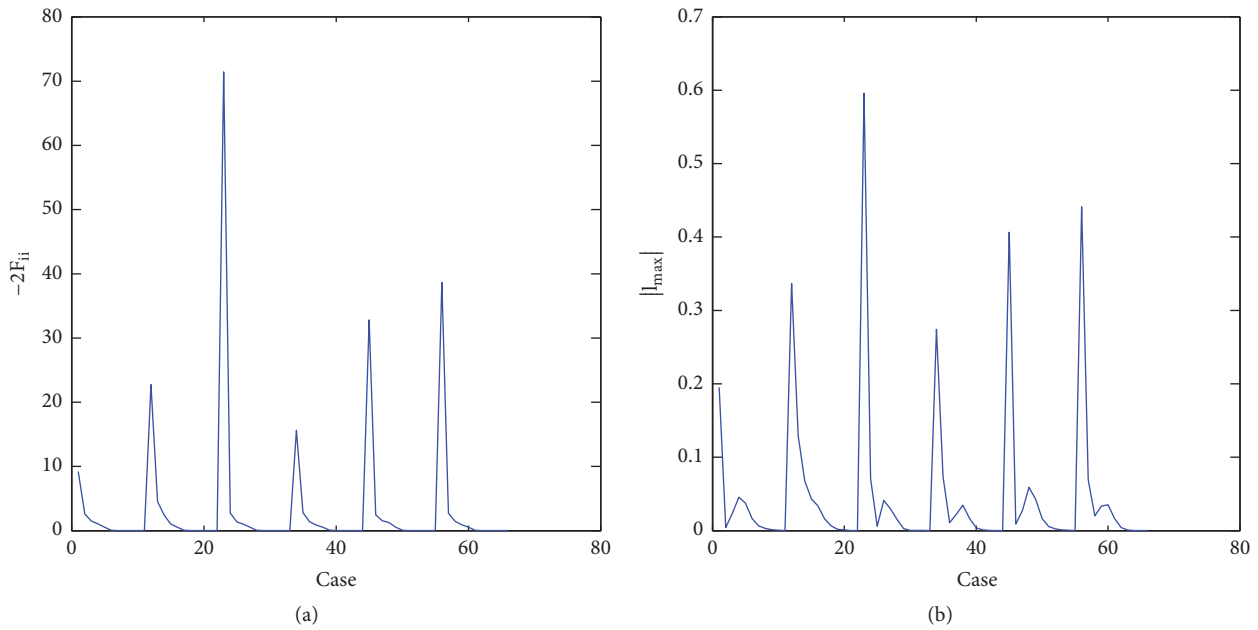
FIGURE 2: Index plots of perturbation of explanatory variables.

A holds for our proposed model. Therefore, we can apply our proposed methodology to estimate the parameters in model (43). Using the algorithm in Section 2, we obtain the MQLE of $\boldsymbol{\beta}$, the predictive values of $b_1, \cdots, b_6$ as follows: $\widehat{\boldsymbol{\beta}} = (0.8317, 0.0446, -13.2203, -1.2535)^T$ and

$$\widehat{\mathbf{b}} = (-0.1133, 0.0825, 0.1600, 0.0500,$$
$$- 0.0625, 0.1133)^T .$$

(45)

Now we present local influence analysis for the above fitting results. Under case-weight perturbation, cases 23, 45,

and 56 are most influential, as depicted as in Figure 1(a). Cases 1, 12, 23, 45, and 56 are identified as influential points, and case 23 is the most influential, as shown in Figure 1(b). The index plots of $-\ddot{F}$ and $|\mathbf{l}_{max}|$ for perturbation on explanatory variables are given in Figures 2(a) and 2(b), respectively. From Figure 2(a) we can see that cases 12, 23, 45, and 56 are identified as influential points. Figure 2(b) shows that cases 1, 12, 23, 34, 45, and 56 are influential. Figure 3 displays the index plots of $|\mathbf{l}_{max}|$ for the perturbation of random effects. For these types of perturbation, case 23 is identified as being the most influential. Note that case 23 exerts great
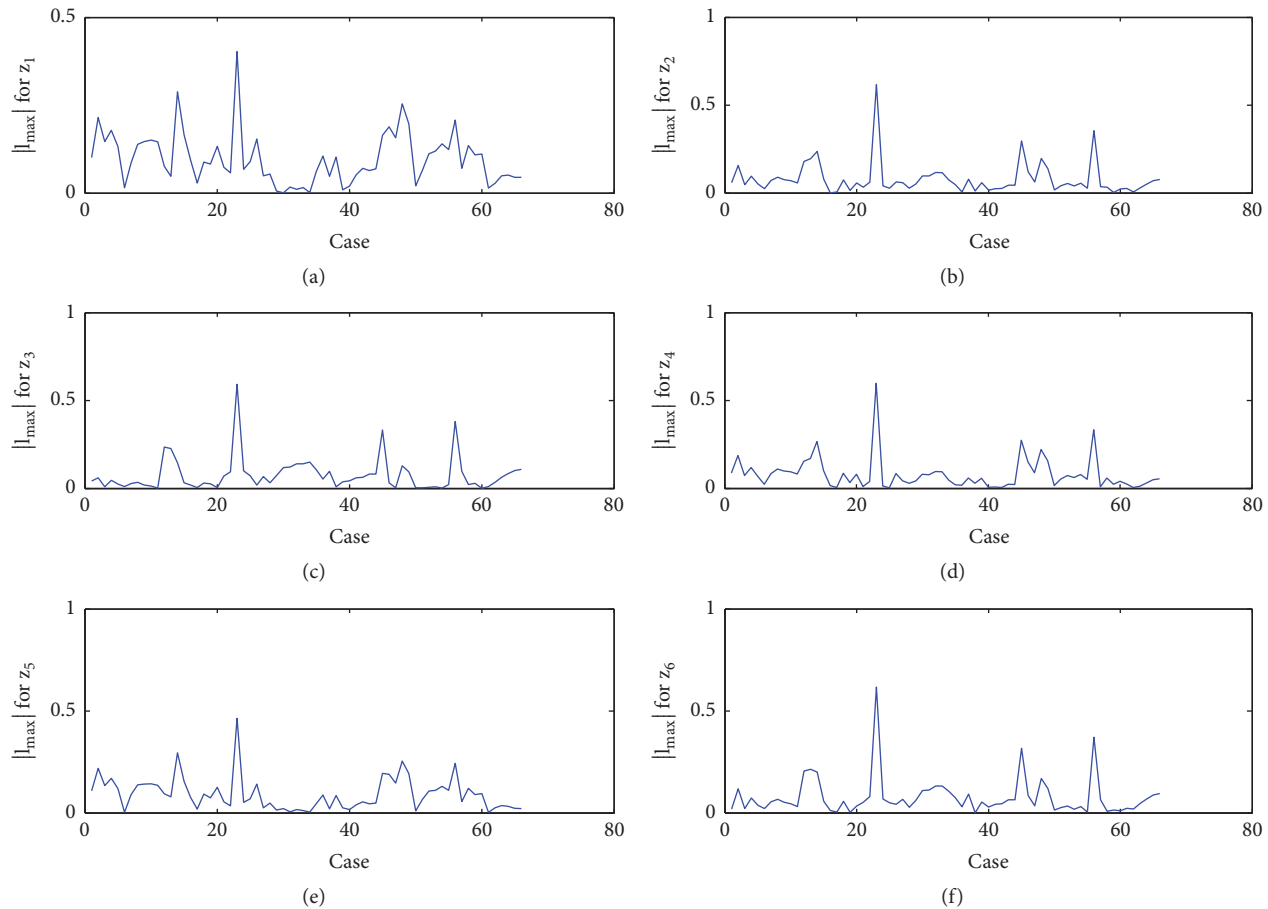
FIGURE 3: Index plots of $|l_{max}|$ for perturbation of random effects design matrix.

influence in each perturbation scheme, which indicates that the results obtained through different perturbation schemes are quite consistent. Special attention should be paid to those influential cases, which may be worthwhile to consider a more formal test to check whether they are outliers.

5. Conclusion

In this work, we have assessed the local influence of minor perturbations of our proposed models. The key idea of the previous approach is to study the behavior of the likelihood displacement obtained from a relevant perturbation. However, it is difficult to apply it directly to the proposed model due to the fact that the marginal quasi-likelihood function of the QLNMWRE involves the intractable integral. To solve this problem, we have employed Laplace's method to approximate the marginal quasi-likelihood function of the QLNMWRE, which results in the penalized quasi-likelihood (PQL). Based on the PQL and the penalized quasi-likelihood displacement, the estimates of unknown parameters have been proposed, and local influence analysis has been investigated. Our numerical example has demonstrated that our proposed local influence technique is rather useful in the detection of influential points. Although the focus of this article is on the assessment of influential points in the

QLNMWRE, the local influence approach can be extended to other complicated models.

Appendix

Proof of Theorem 1. Differentiating (10) with respect to $\boldsymbol{\beta}$ yields that

$$
\begin{aligned}
\dot{Q}_P(\boldsymbol{\beta}) &= \left.\frac{\partial Q_1(\mathbf{y};\boldsymbol{\mu})}{\partial \boldsymbol{\beta}}\right|_{\mu_i=\widetilde{\mu}_i} - \sigma^{-2}\left(\frac{\partial \widetilde{\mathbf{b}}}{\partial \boldsymbol{\beta}^T}\right)^T \Sigma^{-1}\widetilde{\mathbf{b}} \\
&= \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}^T}\right)^T \left.\frac{\partial Q_1(\mathbf{y};\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\widetilde{\boldsymbol{\mu}}} \\
&\quad - \sigma^{-2}\left(\frac{\partial \widetilde{\mathbf{b}}}{\partial \boldsymbol{\beta}^T}\right)^T \Sigma^{-1}\widetilde{\mathbf{b}} \qquad\qquad \text{(A.1)} \\
&= \left(D + Z\frac{\partial \widetilde{\mathbf{b}}}{\partial \boldsymbol{\beta}^T}\right)^T \left.\frac{\partial Q_1(\mathbf{y};\boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\widetilde{\boldsymbol{\mu}}} \\
&\quad - \sigma^{-2}\left(\frac{\partial \widetilde{\mathbf{b}}}{\partial \boldsymbol{\beta}^T}\right)^T \Sigma^{-1}\widetilde{\mathbf{b}}.
\end{aligned}
$$

It follows from the definition of $\widetilde{\mathbf{b}}$ and (10) that

$$\sum_{i=1}^{n} \frac{\partial}{\partial \mu_i} \left(\int_{y_i}^{\mu_i} \frac{y_i - t}{\sigma^2 \nu(t)} dt \right)\bigg|_{\mu_i = \widetilde{\mu}_i} \frac{\partial \mu_i}{\partial \mathbf{b}}\bigg|_{\mathbf{b}=\widetilde{\mathbf{b}}} - \sigma^{-2} \Sigma^{-1} \widetilde{\mathbf{b}}$$

$$= \sigma^{-2} \sum_{i=1}^{n} (y_i - \mu_i)(\nu(\mu_i))^{-1} \mathbf{z}_i\big|_{\mu_i = \widetilde{\mu}_i} - \sigma^{-2} \Sigma^{-1} \widetilde{\mathbf{b}} \quad \text{(A.2)}$$

$$= \sigma^{-2} Z^T \widetilde{\mathbf{e}} - \sigma^{-2} \Sigma^{-1} \widetilde{\mathbf{b}} = 0,$$

which implies

$$\widetilde{\mathbf{b}} = \Sigma Z^T \widetilde{\mathbf{e}}. \quad \text{(A.3)}$$

Substituting (A.3) into (A.1) yields (13). Differentiating (13) with respect to $\boldsymbol{\beta}$ leads to

$$\ddot{Q}_p(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}^T} \left\{ \left(\frac{\partial h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} \right)^T \frac{\partial Q_1(\mathbf{y}; \boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\bigg|_{\boldsymbol{\mu}=\widetilde{\boldsymbol{\mu}}} \right\}$$

$$= \left[\left(\frac{\partial Q_1(\mathbf{y}; \boldsymbol{\mu})}{\partial \boldsymbol{\mu}}\bigg|_{\boldsymbol{\mu}=\widetilde{\boldsymbol{\mu}}} \right)^T \right] \left[\frac{\partial^2 h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right]$$

$$+ D^T \frac{\partial^2 Q_1(\mathbf{y}; \boldsymbol{\mu})}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^T}\bigg|_{\boldsymbol{\mu}=\widetilde{\boldsymbol{\mu}}} \frac{\partial \widetilde{\boldsymbol{\mu}}}{\partial \boldsymbol{\beta}^T} \quad \text{(A.4)}$$

$$= \sigma^{-2} [\widetilde{e}^T][W] + \sigma^{-2} D^T K \frac{\partial \widetilde{\boldsymbol{\mu}}}{\partial \boldsymbol{\beta}^T}.$$

Differentiating (A.3) with respect to $\boldsymbol{\beta}$ yields

$$\frac{\partial \widetilde{\mathbf{b}}}{\partial \boldsymbol{\beta}} = \Sigma Z^T K \frac{\partial \widetilde{\boldsymbol{\mu}}}{\partial \boldsymbol{\beta}^T}. \quad \text{(A.5)}$$

Note that $\widetilde{\boldsymbol{\mu}} = h(X, \boldsymbol{\beta}) + Z\widetilde{\mathbf{b}}$; it follows that

$$\frac{\partial \widetilde{\boldsymbol{\mu}}}{\partial \boldsymbol{\beta}^T} = \frac{\partial h(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} + Z \frac{\partial \widetilde{\mathbf{b}}}{\partial \boldsymbol{\beta}^T}. \quad \text{(A.6)}$$

Combining (A.5) and (A.6) leads to

$$\frac{\partial \widetilde{\boldsymbol{\mu}}}{\partial \boldsymbol{\beta}^T} = D + Z \frac{\partial \widetilde{\mathbf{b}}}{\partial \boldsymbol{\beta}^T} = D + Z \Sigma Z^T K \frac{\partial \widetilde{\boldsymbol{\mu}}}{\partial \boldsymbol{\beta}^T}, \quad \text{(A.7)}$$

which implies

$$\frac{\partial \widetilde{\boldsymbol{\mu}}}{\partial \boldsymbol{\beta}^T} = \left(I - Z \Sigma Z^T K \right)^{-1} D$$

$$= K^{-1} \left(K^{-1} - Z \Sigma Z^T \right)^{-1} D = K^{-1} \Omega_1^{-1} D. \quad \text{(A.8)}$$

Substituting (A.8) into (A.4) yields (14). It follows from Assumption A and $K_0 = E_y(K)$ that (15) holds. Thus, the proof is completed. $\qquad\square$

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] X. Jiang, T. Xia, and X. Wang, "Asymptotic properties of maximum quasi-likelihood estimator in quasi-likelihood non linear models with stochastic regression," *Communications in Statistics—Theory and Methods*, vol. 46, no. 13, pp. 6229–6239, 2017.

[2] T. Xia, X.-R. Wang, and X.-J. Jiang, "Asymptotic properties of maximum quasi-likelihood estimator in quasi-likelihood non-linear models with misspecified variance function," *Statistics. A Journal of Theoretical and Applied Statistics*, vol. 48, no. 4, pp. 778–786, 2014.

[3] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman & Hall, London, UK, 2nd edition, 1989.

[4] R. W. Wedderburn, "Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method," *Biometrika*, vol. 61, pp. 439–447, 1974.

[5] J. Fan and I. Gijbels, *Local Polynomial Modeling and Its Application*, Chapman & Hall, London, UK, 1996.

[6] J. Jiang, X. Jiang, J. Li, Y. Liu, and W. Yan, "Spatial quantile estimation of multivariate threshold time series models," *Physica A: Statistical Mechanics and its Applications*, vol. 486, pp. 772–781, 2017.

[7] X. Jiang, J. Li, T. Xia, and W. Yan, "Robust and efficient estimation with weighted composite quantile regression," *Physica A: Statistical Mechanics and its Applications*, vol. 457, pp. 413–423, 2016.

[8] K. Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986.

[9] R. D. Cook, "Assessment of local influence," *ournal of the Royal Statistical Society. Series B (Methodological)*, vol. 48, no. 2, pp. 133–169, 1986.

[10] R. J. Beckman, C. J. Nachtsheim, and R. D. Cook, "Diagnostics for mixed-model analysis of variance," *Technometrics. A Journal of Statistics for the Physical, Chemical and Engineering Sciences*, vol. 29, no. 4, pp. 413–426, 1987.

[11] L. C. Montenegro, V. H. Lachos, and H. Bolfarine, "Local influence analysis for skew-normal linear mixed models," *Communications in Statistics—Theory and Methods*, vol. 38, no. 3-5, pp. 484–496, 2009.

[12] W. Thomas and R. D. Cook, "Assessing influence on predictions from generalized linear models," *Technometrics*, vol. 32, no. 1, pp. 59–65, 1990.

[13] L. Xiang, A. H. Lee, and S.-K. Tse, "Assessing local cluster influence in generalized linear mixed models," *Journal of Applied Statistics*, vol. 30, no. 4, pp. 349–359, 2003.

[14] B. C. Wei, *Exponential Family Nonlinear Models*, Springer, Singapore, 1998.

[15] N.-S. Tang, B.-C. Wei, and W.-Z. Zhang, "Influence diagnostics in nonlinear reproductive dispersion mixed models," *Statistics. A Journal of Theoretical and Applied Statistics*, vol. 40, no. 3, pp. 227–246, 2006.

[16] E. F. Vonesh and R. L. Carter, "Mixed-effects nonlinear regression for unbalanced repeated measures," *Biometrics - A Journal of the International Biometric Society*, vol. 48, no. 1, pp. 1–17, 1992.

[17] E. F. Vonesh, "A note on the use of Laplace's approximation for nonlinear mixed-effects models," *Biometrika*, vol. 83, no. 2, pp. 447–452, 1996.

[18] T. Xia, X. Jiang, and X. Wang, "Strong consistency of the maximum quasi-likelihood estimator in quasi-likelihood nonlinear models with stochastic regression," *Statistics & Probability Letters*, vol. 103, pp. 37–45, 2015.

[19] T. Xia, X. Jiang, and X. Wang, "Diagnostics for quasi-likelihood nonlinear models," *Communications in Statistics—Theory and Methods*, vol. 46, no. 18, pp. 8836–8851, 2017.

[20] N. E. Breslow and D. G. Clayton, "Approximate inference in generalized linear mixed models," *Journal of the American Statistical Association*, vol. 88, 9, p. 25, 1993.

[21] X. Lin, "Variance component testing in generalised linear models with random effects," *Biometrika*, vol. 84, no. 2, pp. 309–326, 1997.

[22] L. Tierney, R. E. Kass, and J. B. Kadane, "Approximate marginal densities of nonlinear functions," *Biometrika*, vol. 76, no. 3, pp. 425–433, 1989.

[23] R. Wolfinger, "Laplace's approximation for nonlinear mixed models," *Biometrika*, vol. 80, no. 4, pp. 791–795, 1993.

[24] Y. Lee and N. A. Nelder, "Hierarchical generalized linear models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 58, no. 4, pp. 619–678, 1996.

[25] H.-T. Zhu and S.-Y. Lee, "Local influence for incomplete-data models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 1, pp. 111–126, 2001.

[26] K.-M. Jung, "Local influence in generalized estimating equations," *Scandinavian Journal of Statistics*, vol. 35, no. 2, pp. 286–294, 2008.

[27] K. C. Kwan, G. O. Breault, E. R. Umbenhauer, F. G. McMahon, and D. E. Duggan, "Kinetics of indomethacin absorption, elimination, and enterohepatic circulation in man," *Journal of Pharmacokinetics and Biopharmaceutics*, vol. 4, no. 3, pp. 255–280, 1976.

[28] M. Davidian and D. M. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman and Hall, London, 1995.

[29] E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, 1958.

Robust Group Identification and Variable Selection in Regression

Ali Alkenani and Tahir R. Dikheel

Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Al Diwaniyah, Iraq

Correspondence should be addressed to Ali Alkenani; ali.alkenani@qu.edu.iq

The elimination of insignificant predictors and the combination of predictors with indistinguishable coefficients are the two issues raised in searching for the true model. Pairwise Absolute Clustering and Sparsity (PACS) achieves both goals. Unfortunately, PACS is sensitive to outliers due to its dependency on the least-squares loss function which is known to be very sensitive to unusual data. In this article, the sensitivity of PACS to outliers has been studied. Robust versions of PACS (RPACS) have been proposed by replacing the least squares and nonrobust weights in PACS with MM-estimation and robust weights depending on robust correlations instead of person correlation, respectively. A simulation study and two real data applications have been used to assess the effectiveness of the proposed methods.

1. Introduction

The latest developments in data aggregation have generated huge number of variables. The large amounts of data pose a challenge to most of the standard statistical methods. In many regression problems, the number of variables is huge. Moreover, many of these variables are irrelevant. Variable selection (VS) is the process of selecting significant variables for use in model construction. It is an important step in the statistical analysis. Statistical procedures for VS are characterized by improving the model's prediction, providing interpretable models while retaining computational efficiency. VS techniques, such as stepwise selection and best subset regression, may suffer from instability [1]. To tackle the instability problem, regularization methods have been used to carry out VS. They have become increasingly popular, as they supply a tool with which the VS is carried out during the process of estimating the coefficients in the model, for example, LASSO [2], SCAD [3], elastic-net [4], fused LASSO [5], adaptive LASSO [6], group LASSO [7], OSCAR [8], adaptive elastic-net [9], and MCP [10].

Searching for the correct model raises two matters: the exclusion of insignificant predictors and the combination of predictors with indistinguishable coefficients (IC) [11]. The above approaches can remove insignificant predictors but be unsuccessful to merge predictors with IC. Pairwise Absolute Clustering and Sparsity (PACS, [11]) achieves both goals. Moreover, PACS is an oracle method for simultaneous group identification and VS.

Unfortunately, PACS is sensitive to outliers due to its dependency on the least-squares loss function which is known as very sensitive to unusual data. In this article, the sensitivity of PACS to outliers has been studied. Robust versions of PACS (RPACS) have been proposed by replacing the least squares and nonrobust weights in PACS with MM-estimation and robust weights depending on robust correlations instead of person correlation, respectively. RPACS can completely estimate the parameters of regression and select the significant predictors simultaneously, while being robust to the existence of possible outliers.

The rest of this article proceeds as follows. In Section 2, PACS has been briefly reviewed. The robust extension of PACS is detailed in Section 3. Simulation studies under different settings are presented in Section 4. In Section 5, the proposed robust PACS has been applied to two real datasets. Finally, a discussion concludes in Section 6.

2. A Brief Review of PACS

Under the linear regression model setup with standardized predictors x_{ij} and centered response values y_i, $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, p$. Sharma et al. [11] proposed an oracle method PACS for simultaneous group identification and VS. PACS has less computational cost than OSCAR approach. In PACS, the equality of coefficients is attained by adding penalty to the pairwise differences and pairwise sums of coefficients. The PACS estimates are the minimizers of the following:

$$
\sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \left\{ \sum_{j=1}^{p} \omega_j \left| \beta_j \right| \right.
$$
$$
\left. + \sum_{1 \le j < k \le p} \omega_{jk(-)} \left| \beta_k - \beta_j \right| + \sum_{1 \le j < k \le p} \omega_{jk(+)} \left| \beta_k + \beta_j \right| \right\}, \tag{1}
$$

where $\lambda \ge 0$ is the regularization parameter and ω is the nonnegative weights.

The penalty in (1) consists of $\lambda\{\sum_{j=1}^{p} \omega_j|\beta_j|\}$ that encourages sparseness, $\lambda\{\sum_{1 \le j < k \le p} \omega_{jk(-)}|\beta_k - \beta_j|\}$, and $\lambda\{\sum_{1 \le j < k \le p} \omega_{jk(+)}|\beta_k + \beta_j|\}$ that encourages equality of coefficients. The second term of the penalty encourages the same sign coefficients to be set as equal, while the third term encourages opposite sign coefficients to be set as equal in magnitude.

Choosing of appropriate adaptive weights is very important for PACS to be an oracle procedure. Consequently, Sharma et al. [11] suggested adaptive PACS that incorporate correlations into the weights which are given as follows:

$$
\omega_j = \left| \widetilde{\beta}_j \right|^{-1},
$$
$$
\omega_{jk(-)} = \left(1 - r_{jk} \right)^{-1} \left| \widetilde{\beta}_k - \widetilde{\beta}_j \right|^{-1},
$$
$$
\omega_{jk(+)} = \left(1 + r_{jk} \right)^{-1} \left| \widetilde{\beta}_k + \widetilde{\beta}_j \right|^{-1} \tag{2}
$$
$$
\text{for } 1 \le j < k \le p,
$$

where $\widetilde{\beta}$ is $\sqrt{n}$ consistent estimator of β, such as the ordinary least squares (OLS) estimates or other shrinkage estimates like ridge regression estimates and r_{jk} is Pearson's correlation between the (j, k)th pair of predictors.

Sharma et al. [11] suggest using ridge estimates as initial estimates for β's to obtain weights perform well in studies with collinear predictors.

3. Robust PACS

3.1. Methodology of Robust PACS. The satisfactory performance of PACS under normal errors has been demonstrated in [11]. However, the high sensitivity to outliers is the main drawback of PACS where a single outlier can change the good performance of PACS estimate completely.

Note that, in (1), the least-squares criterion is used between the predictors and the response. Also, the weighted

penalty contains weights which depend on Pearson's correlation in their calculations. However, the least-squares criterion and Pearson's correlation are not robust to outliers. To achieve the robustness in estimation and select the informative predictors robustly, the authors propose replacing the least-squares criterion with MM-estimation [12] where the MM- estimators are efficient and have high breakdown points. Moreover, the nonrobust weights replaced with robust weights depend on robust correlations such as the fast consistent high breakdown (FCH) [13], reweighted multivariate normal (RMVN) [13], Spearman's correlation (SP), and Kendall's correlation (KN). The RPACS estimates minimizing the following:

$$
\sum_{i=1}^{N} \rho_1 \left(\frac{R_i(\beta)}{S_n} \right) + \lambda \left\{ \sum_{j=1}^{p} Ro\omega_j \left| \beta_j \right| \right.
$$
$$
+ \sum_{1 \le j < k \le p} Ro\omega_{jk(-)} \left| \beta_k - \beta_j \right| \tag{3}
$$
$$
\left. + \sum_{1 \le j < k \le p} Ro\omega_{jk(+)} \left| \beta_k + \beta_j \right| \right\},
$$

where $\lambda \ge 0$ is the regularization parameter and $Ro\omega$ is the robust version of the nonnegative weights which are describes in (2). $R_i(\beta) = y_i - \sum_{j=1}^{p} x_{ij}\beta_j$, S_n is M-estimate of scale of the residuals, and it is defined as a solution of

$$
\frac{1}{N} \sum_{i=1}^{N} \rho_0 \left(\frac{R_i}{S_n} \right) = K, \tag{4}
$$

where K is a constant and ρ_0 function satisfies the following conditions:

(1) ρ_0 is symmetric and continuously differentiable, and $\rho_0(0) = 0$.

(2) There exist $a > 0$ such that ρ_0 is strictly increasing on $[0, a]$ and constant on $[a, \infty)$.

(3) $K/\rho_0(a) = 1/2$.

The MM estimator in the first part of (3) is defined as an M-estimator of β using a redescending score function, $\psi(u) = \partial\rho_1(u)/\partial u$, and S_n obtained from (4). It is a solution to

$$
\sum_{i=1}^{N} x_{ij} \psi \left(\frac{R_i(\beta)}{S_n} \right) = 0 \quad j = 1, 2, \ldots, p, \tag{5}
$$

where ρ_1 is another bounded ρ function such that $\rho_1 \le \rho_0$.

3.2. Choosing the Robust Weights. The process of choosing the suitable weights is very important in order to obtain an oracle procedure [11]. The weights, which are described in (2), depend on Pearson's correlation in their calculations. From a practical point of view, it is well known that Pearson's correlation is not resistant to outliers and thus choosing weights in (2) based on this correlation will cause uncertain and deceptive results. Consequently, in order to get robust

weights, there is a need to estimate the correlation by using robust approaches. There are two types of robust versions for Pearson's correlation. The first type consists of those that are robust to the outliers, without interest in the general structure of the data, whereas the second type gives attention to the general structure of the data when dealing with outliers [14]. KN and MCD (minimum covariance determinant) are examples for the first and second types, respectively. Olive and Hawkins [13] proposed FCH and RMVN methods as practical consistent, outlier resistant estimators for multivariate location and dispersion. Alkenani and Yu [15] employed FCH and RMVN estimators instead of Pearson's correlation in the canonical correlation analysis (CCA) to obtain robust CCA. The authors showed that these estimators have good performance under different settings of outliers.

In this article, the FCH, RMVN, SP, and KN correlations have been employed instead of Pearson's correlation in order to obtain robust weights as follows:

$$Ro\omega_j = \left|\breve{\beta}_j\right|^{-1},$$

$$Ro\omega_{jk(-)} = \left(1 - Ror_{jk}\right)^{-1}\left|\breve{\beta}_k - \breve{\beta}_j\right|^{-1},$$

$$Ro\omega_{jk(+)} = \left(1 + Ror_{jk}\right)^{-1}\left|\breve{\beta}_k + \breve{\beta}_j\right|^{-1}$$

$$\text{for } 1 \le j < k \le p,$$

(6)

where Ror is a robust version of Pearson's correlation such as FCH, RMVN, SP, and KN correlations. $\breve{\beta}$ is a robust initial estimate for β and we suggest using robust ridge estimates as initial estimates for β's.

4. Simulation Study

In this section, five examples have been used to assess our proposed method RPACS by comparing it with PACS which is suggested in [11]. A regression model has been generated as follows:

$$y = X\beta + \epsilon \quad \epsilon \sim N\left(0, \sigma^2 I\right). \tag{7}$$

In all examples, predictors are standard normal. The distributions of the error term ϵ and the predictors are contaminated by two types of distributions, t distribution with 5 degrees of freedom ($t_{(5)}$) and Cauchy distribution with mean equal to 0 and variance equal to 1 (Cauchy $(0, 1)$). Also, different contamination ratios (5%, 10%, 15%, 20%, and 25%) were used. The performance of the methods is compared by using model error (ME) criterion for prediction accuracy which is defined by $(\widehat{\beta} - \beta)'V(\widehat{\beta} - \beta)$ where V represents the population covariance matrix of X. The sample sizes were 50 and 100 and the simulated model was replicated 1000 times.

Example 1. In this example, we choose the true parameters for the model of study as $\beta = (2, 2, 2, 0, 0, 0, 0, 0)^T$, $X \in \mathbb{R}^8$. The first three predictors are highly correlated with correlation equal to 0.7 and their coefficients are equal in magnitude, while the rest are uncorrelated.

Example 2. In this example, the true coefficients have been assumed as $\beta = (0.5, 1, 2, 0, 0, 0, 0, 0)^T$, $X \in \mathbb{R}^8$. The first three predictors are highly correlated with correlation equal to 0.7 and their coefficients differ in magnitude, while the rest are uncorrelated.

Example 3. In this example, the true parameters are $\beta = (1, 1, 1, 0.5, 1, 2, 0, 0, 0, 0)^T$, $X \in \mathbb{R}^{10}$. The first three predictors are highly correlated with correlation equal to 0.7 and their coefficients are equal in magnitude, while the second three predictors have lower correlation equal to 0.3 and different magnitudes. The rest of predictors are uncorrelated.

Example 4. In this example, true parameters are $\beta = (1, 1, 1, 0.5, 1, 2, 0, 0, 0, 0)^T$, $X \in \mathbb{R}^{10}$. The first three predictors are correlated with correlation equal to 0.3 and their coefficients are equal in magnitude, while the second three predictors have correlation equal to 0.7 and different magnitudes. The rest of predictors are uncorrelated.

Example 5. In this example, the true parameters are assumed as $\beta = (2, 2, 2, 1, 1, 0, 0, 0, 0, 0)^T$, $X \in \mathbb{R}^{10}$. The first three predictors are highly correlated with pairwise correlation equal to 0.7 and the second two predictors have pairwise correlation of 0.7, while the rest are uncorrelated. It can be observed that the groups of three and two highly correlated predictors have coefficients which are equal in magnitude.

To avoid repetition, the observations about the results in Tables 1–5 have been summarized as follows.

From Tables 1, 2, 3, 4, and 5, when there is no contamination data, PACS has good performance compared with our proposed methods. It is clear, when the contamination ratio of $t_{(5)}$ or Cauchy $(0, 1)$ goes up the performance of PACS goes down while RPACS with all the robust weights has a stable performance, and the preference is for RPACS.RMVN and RPACS.RFCH, respectively, for all the samples sizes. The variations in ME values for the RPACS estimates with all the robust weights are close under different setting of contamination and sample sizes, and they are less than the variations of PACS estimates.

5. Analysis of Real Data

In this section, the RPACS methods with all the robust weights and PACS method have been applied in real data. The NCAA sports data from Mangold et al. [16] and the pollution data from McDonald and Schwing [17] have been studied.

The response variable was centered and the predictors were standardized. To verify RPACS, the two data sets have been analyzed by including outliers in the response variable and the predictors. The two data sets have been contaminated with (5%, 10%, 15%, and 20%) data from multivariate t distribution with three degrees of freedom.

To evaluate the estimation accuracy of the RPACS methods, the correlation between the estimated parameters according to the different methods under consideration and the estimated parameters from PACS without outliers, denoted as $\text{Corr}(\widehat{\beta}, \widehat{\beta}_{\text{PACS},0})$, was presented. Also, the effective

TABLE 1: ME results of Example 1.

Dist.	n	Outliers%	PACS	RPACS.KN	RPACS.SP	RPACS.FCH	RPACS.RMVN
$t_{(5)}$	50	0	0.02304	0.02964	0.03083	0.02979	0.02902
		5	0.20135	0.08124	0.08135	0.05575	0.04655
		10	0.25043	0.14048	0.14543	0.06579	0.05664
		15	0.30788	0.17578	0.18152	0.07225	0.06153
		20	0.34708	0.19266	0.21286	0.08195	0.06939
		25	0.40692	0.21584	0.22533	0.10242	0.08238
	100	0	0.02004	0.02644	0.02863	0.02772	0.02700
		5	0.19100	0.07100	0.08030	0.05111	0.04025
		10	0.23012	0.13011	0.14002	0.06116	0.05013
		15	0.28715	0.15523	0.17137	0.06899	0.05902
		20	0.32520	0.18670	0.19234	0.07115	0.06005
		25	0.36692	0.20522	0.21404	0.09032	0.07784
Cauchy $(0, 1)$	50	5	0.18112	0.07004	0.07237	0.04390	0.03581
		10	0.23263	0.12001	0.12273	0.05472	0.04454
		15	0.28368	0.15274	0.16138	0.06237	0.05079
		20	0.33511	0.17162	0.18556	0.07381	0.05848
		25	0.38488	0.19330	0.20405	0.09342	0.07211
	100	5	0.17214	0.06111	0.07335	0.04277	0.03581
		10	0.22263	0.11001	0.11273	0.04672	0.03854
		15	0.27368	0.14274	0.15138	0.05237	0.04079
		20	0.31511	0.16162	0.17556	0.06381	0.04848
		25	0.35488	0.18330	0.19405	0.08342	0.06211

TABLE 2: ME results of Example 2.

Dist.	N	Outliers%	PACS	RPACS.KN	RPACS.SP	RPACS.FCH	RPACS.RMVN
$t_{(5)}$	50	0	0.11372	0.12032	0.12151	0.12047	0.11970
		5	0.29201	0.17191	0.17203	0.14644	0.13725
		10	0.34113	0.23117	0.23611	0.15646	0.14730
		15	0.39857	0.26647	0.27221	0.16294	0.15222
		20	0.43778	0.28336	0.30355	0.17263	0.16006
		25	0.49761	0.30653	0.31602	0.19312	0.17308
	100	0	0.10354	0.11022	0.11131	0.10407	0.10050
		5	0.28171	0.16170	0.17100	0.14180	0.13094
		10	0.32082	0.22080	0.23072	0.15185	0.14082
		15	0.37783	0.24591	0.26205	0.15967	0.14970
		20	0.41560	0.27700	0.28300	0.16185	0.15071
		25	0.45762	0.29592	0.30473	0.18101	0.16854
Cauchy $(0, 1)$	50	5	0.27182	0.16072	0.16306	0.13460	0.12650
		10	0.32333	0.21071	0.21342	0.14541	0.13523
		15	0.37434	0.24340	0.25204	0.15303	0.14145
		20	0.42581	0.26232	0.27626	0.16451	0.14918
		25	0.47558	0.284	0.29475	0.18412	0.16281
	100	5	0.26282	0.15181	0.16405	0.13345	0.12651
		10	0.31331	0.20071	0.20343	0.13742	0.12923
		15	0.36435	0.23341	0.24205	0.14304	0.13149
		20	0.40581	0.25232	0.26625	0.15451	0.13916
		25	0.44557	0.27400	0.28473	0.17412	0.15281

TABLE 3: ME results of Example 3.

Dist.	N	Outliers%	PACS	RPACS.KN	RPACS.SP	RPACS.FCH	RPACS.RMVN
$t_{(5)}$	50	0	0.14172	0.14831	0.14950	0.14844	0.14743
		5	0.32001	0.19991	0.20003	0.17441	0.16522
		10	0.36913	0.25915	0.26411	0.18444	0.17530
		15	0.42653	0.29443	0.30021	0.19094	0.18022
		20	0.46576	0.31135	0.33154	0.20063	0.18806
		25	0.52561	0.33453	0.34402	0.22112	0.20107
	100	0	0.13042	0.13501	0.13645	0.13344	0.13255
		5	0.30971	0.18971	0.19901	0.16982	0.15894
		10	0.34882	0.24883	0.25872	0.17985	0.16882
		15	0.40582	0.27391	0.29003	0.18765	0.17774
		20	0.44365	0.30501	0.31103	0.18983	0.17871
		25	0.48562	0.32392	0.33271	0.20901	0.19650
Cauchy (0, 1)	50	5	0.29982	0.18872	0.19106	0.1626	0.1545
		10	0.35133	0.23871	0.24142	0.17341	0.16323
		15	0.40234	0.2714	0.28004	0.18103	0.16945
		20	0.45381	0.29032	0.30426	0.19251	0.17718
		25	0.50358	0.312	0.32275	0.21212	0.19081
	100	5	0.32001	0.19991	0.20003	0.17445	0.16525
		10	0.36913	0.25917	0.26411	0.18441	0.17536
		15	0.42655	0.29444	0.30021	0.19093	0.18022
		20	0.46575	0.31134	0.33153	0.20063	0.18804
		25	0.525610	0.33453	0.34401	0.22112	0.20106

TABLE 4: ME results of Example 4.

Dist.	N	Outliers%	PACS	RPACS.KN	RPACS.SP	RPACS.FCH	RPACS.RMVN
$t_{(5)}$	50	0	0.15251	0.15910	0.16035	0.15921	0.15823
		5	0.33081	0.21070	0.21082	0.18520	0.17601
		10	0.37991	0.26993	0.27491	0.19523	0.18612
		15	0.43732	0.30521	0.31101	0.20175	0.19102
		20	0.47653	0.32216	0.34233	0.21143	0.19887
		25	0.53641	0.34531	0.35482	0.23192	0.21185
	100	0	0.13342	0.13901	0.14125	0.13814	0.13713
		5	0.32051	0.20051	0.20981	0.18062	0.16973
		10	0.35962	0.25965	0.26952	0.19067	0.17962
		15	0.41662	0.28471	0.30083	0.19847	0.18853
		20	0.45446	0.31581	0.32183	0.20066	0.18951
		25	0.49642	0.33472	0.34351	0.21981	0.20757
Cauchy (0, 1)	50	5	0.31062	0.19952	0.20188	0.1734	0.16538
		10	0.36216	0.24951	0.25222	0.18421	0.17404
		15	0.41316	0.2822	0.29087	0.19184	0.18025
		20	0.46461	0.30112	0.31507	0.20331	0.18798
		25	0.51438	0.32284	0.33357	0.22294	0.20161
	100	5	0.33083	0.21071	0.21083	0.18525	0.17606
		10	0.37993	0.26995	0.27491	0.19521	0.18613
		15	0.43733	0.30522	0.31101	0.20175	0.19102
		20	0.47653	0.32217	0.34233	0.21143	0.19886
		25	0.53641	0.34533	0.354814	0.23192	0.21188

TABLE 5: ME results of Example 5.

Dist.	N	Outliers%	PACS	RPACS.KN	RPACS.SP	RPACS.FCH	RPACS.RMVN
$t_{(5)}$	50	0	0.06031	0.06695	0.06815	0.06701	0.06602
		5	0.23861	0.11851	0.11862	0.09305	0.08381
		10	0.28771	0.177735	0.182712	0.10303	0.09392
		15	0.34512	0.21301	0.21881	0.10955	0.09886
		20	0.38433	0.22996	0.25015	0.11923	0.10667
		25	0.44424	0.25315	0.26262	0.13972	0.11965
	100	0	0.04125	0.04684	0.04908	0.04597	0.04496
		5	0.22837	0.108313	0.11765	0.08846	0.07755
		10	0.26744	0.16745	0.17733	0.09846	0.08743
		15	0.32445	0.19256	0.20865	0.10627	0.09636
		20	0.36228	0.22365	0.22966	0.10844	0.09733
		25	0.40425	0.24257	0.25131	0.12761	0.11537
Cauchy (0, 1)	50	0	0.06031	0.06695	0.06815	0.06701	0.06602
		5	0.21845	0.10737	0.10963	0.08125	0.07316
		10	0.26997	0.15734	0.16006	0.09206	0.08183
		15	0.32095	0.19007	0.19865	0.09963	0.08806
		20	0.37244	0.20896	0.22289	0.11115	0.09579
		25	0.42217	0.23067	0.24135	0.13073	0.10948
	100	0	0.04125	0.04684	0.04908	0.04597	0.04496
		5	0.23865	0.11854	0.11865	0.09308	0.08389
		10	0.28775	0.17779	0.18274	0.10303	0.09397
		15	0.34513	0.21304	0.21885	0.10958	0.09885
		20	0.38435	0.22998	0.25015	0.11926	0.10667
		25	0.44423	0.25314	0.26261	0.13977	0.11967

model size after accounting for equality of absolute coefficient estimates has been reported.

5.1. NCAA Sports Data. The NCAA sport data is taken from a study of the effects of sociodemographic indicators and the sports programs on graduation rates. The dataset is available from the website (http://www4.stat.ncsu.edu/~boos/var .select/ncaa.html). The data size is $n = 94$ and $p = 19$ predictors. The response variable is the average of 6 year graduation rate for 1996–1999. The predictors are students in top 10% HS (X1), ACT COMPOSITE 25TH (X2), on living campus (X3), first-time undergraduates (X4), Total Enrollment/1000 (X5), courses taught by TAs (X6), composite of basketball ranking (X7), in-state tuition/1000 (X8), room and board/1000 (X9), avg BB home attendance (X10), Full Professor Salary (X11), student to faculty ratio (X12), white (X13), assistant professor salary (X14), population of city where located (X15), faculty with PHD (X16), acceptance rate (X17), receiving loans (X18), and out of state (X19).

5.2. Pollution Data (PD). The PD is taken from a study of the effects of different air pollution indicators and sociodemographic factors on mortality. The dataset is available from the website (http://www4.stat.ncsu.edu/~boos/var .select/pollution.html). The data contains $n = 60$ observations and $p = 15$ predictors. The response is the total Age

Adjusted Mortality Rate (y). The predictors are Mean annual precipitation (X1), mean January temperature (X2), mean July temperature (X3), % population that is 65 years of age or over (X4), population per household (X5), median school years (X6), % of housing with facilities (X7), population per square mile (X8), % of population that is nonwhite (X9), % employment in white-collar occupations (X10), % of families with income under 3; 000 (X11), relative population potential (RPP) of hydrocarbons (X12), RPP of oxides of nitrogen (X13), RPP of sulfur dioxide (X14), and % relative humidity (X15).

From Tables 6 and 7, we have the following findings in terms of estimation accuracy and the effective model size:

(1) In case of no contamination, it can be observed that RPACS methods give comparable results as PACS. In addition, it can be seen that RPACS.RMVN and RPACS.FCH achieve better performance than RPACS.KN and RPACS.SP.

(2) In case of contamination, the performance of PACS is dramatically affected. Also, it is obvious that RPACS.RMVN and RPACS.FCH methods give very consistent results, even with the high contamination percentages. The performance of RPACS.KN and RPACS.SP is less efficient than RPACS.RMVN and RPACS.FCH especially for all the contamination percentages.

TABLE 6: The Corr($\widehat{\beta}, \widehat{\beta}_{\text{PACS},0}$) and the effective model size values for the methods under consideration based on the NCAA sport data.

Methods		Outliers%				
		0	5	10	15	20
Corr($\widehat{\beta}, \widehat{\beta}_{\text{PACS},0}$)	PACS	1	0.9033	0.8069	0.4112	0.1345
	RPACS.KN	0.9843	0.9839	0.9530	0.9019	0.8499
	RPACS.SP	0.9840	0.9837	0.9526	0.9006	0.8490
	RPACS.FCH	0.9850	0.9846	0.9843	0.9841	0.9839
	RPACS.RMVN	0.9856	0.9852	0.9850	0.9847	0.9845
The effective model size	PACS	5	6	7	9	10
	RPACS.KN	5	5	6	6	7
	RPACS.SP	5	5	6	6	7
	RPACS.FCH	5	5	5	5	5
	RPACS.RMVN	5	5	5	5	5

TABLE 7: The Corr($\widehat{\beta}, \widehat{\beta}_{\text{PACS},0}$) and the effective model size values for the methods under consideration based on the pollution data.

Methods		Outliers%				
		0	5	10	15	20
Corr($\widehat{\beta}, \widehat{\beta}_{\text{PACS},0}$)	PACS	1	0.9247	0.8259	0.7001	0.5925
	RPACS.KN	0.9882	0.9866	0.9552	0.9044	0.8518
	RPACS.SP	0.9877	0.9862	0.9545	0.9038	0.8511
	RPACS.FCH	0.9890	0.9887	0.9884	0.9882	0.9879
	RPACS.RMVN	0.9897	0.9895	0.9893	0.9890	0.9888
The effective model size	PACS	5	6	6	8	9
	RPACS.KN	5	5	6	7	7
	RPACS.SP	5	5	6	7	7
	RPACS.FCH	5	5	5	5	5
	RPACS.RMVN	5	5	5	5	5

6. Conclusions

In this paper, robust consistent group identification and VS procedures have been proposed (RPACS) which combine the strength of both robust and identifying relevant groups and VS procedure. The simulation studies and analysis of real data demonstrate that RPACS methods have better predictive accuracy and identifying relevant groups than PACS when outliers exist in the response variable and the predictors. In general, the preference is for RPACS.RMVN and RPACS.RFCH, respectively, for all the samples sizes.

Abbreviations

LASSO: Least absolute shrinkage and selection operator
PACS: Pairwise Absolute Clustering and Sparsity
RPACS: Robust Pairwise Absolute Clustering and Sparsity
VS: Variable selection
SCAD: Smoothly clipped absolute deviation
Fused LASSO: Fused least absolute shrinkage and selection operator
Adaptive LASSO: Adaptive least absolute shrinkage and selection operator
Group LASSO: Group least absolute shrinkage and selection operator
OSCAR: Octagonal shrinkage and clustering algorithm for regression
MCP: Minimax concave penalty
IC: Indistinguishable coefficients
FCH: Fast consistent high breakdown
RMVN: Reweighted multivariate normal
SP: Spearman's correlation
KN: Kendall's correlation
MCD: Minimum covariance determinant
CCA: Canonical correlation analysis
NCAA: National Collegiate Athletic Association
PD: Pollution data.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] L. Breiman, "Heuristics of instability and stabilization in model selection," *The Annals of Statistics*, vol. 24, no. 6, pp. 2350–2383, 1996.

[2] R. Tibshirani, "Regression shrinkage and selection via the lasso: A retrospective," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 73, no. 3, pp. 273–282, 1996.

[3] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.

[4] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.

[5] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society B: Statistical Methodology*, vol. 67, no. 1, pp. 91–108, 2005.

[6] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418–1429, 2006.

[7] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[8] H. D. Bondell and B. J. Reich, "Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR," *Biometrics*, vol. 64, no. 1, pp. 115–123, 2008.

[9] H. Zou and H. H. Zhang, "On the adaptive elastic-net with a diverging number of parameters," *The Annals of Statistics*, vol. 37, no. 4, pp. 1733–1751, 2009.

[10] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *The Annals of Statistics*, vol. 38, no. 2, pp. 894–942, 2010.

[11] D. B. Sharma, H. D. Bondell, and H. H. Zhang, "Consistent group identification and variable selection in regression with correlated predictors," *Journal of Computational and Graphical Statistics*, vol. 22, no. 2, pp. 319–340, 2013.

[12] V. c. Yohai, "High breakdown-point and high efficiency robust estimates for regression," *The Annals of Statistics*, vol. 15, no. 2, pp. 642–656, 1987.

[13] D. J. Olive and D. M. Hawkins, "Robust multivariate location and dispersion," http://lagrange.math.siu.edu/Olive/pphbmld .pdf, 2010.

[14] R. Wilcox, *Introduction to robust estimation and hypothesis testing*, Statistical Modeling and Decision Science, Academic press, 2005.

[15] A. Alkenani and K. Yu, "A comparative study for robust canonical correlation methods," *Journal of Statistical Computation and Simulation*, vol. 83, no. 4, pp. 690–720, 2013.

[16] W. D. Mangold, L. Bean, and D. Adams, "The Impact of Intercollegiate Athletics on Graduation Rates among Major NCAA Division I Universities: Implications for College Persistence Theory and Practice," *Journal of Higher Education*, vol. 74, no. 5, pp. 540–563, 2003.

[17] G. C. McDonald and R. C. Schwing, "Instabilities of regression estimates relating air pollution to mortality," *Technometrics*, vol. 15, no. 3, pp. 463–481, 1973.

Exploratory Methods for the Study of Incomplete and Intersecting Shape Boundaries from Landmark Data

Fathi M. O. Hamed[1] and Robert G. Aykroyd[2]

[1]*University of Benghazi, Benghazi, Libya*
[2]*University of Leeds, Leeds, UK*

Correspondence should be addressed to Robert G. Aykroyd; r.g.aykroyd@leeds.ac.uk

Structured spatial point patterns appear in many applications within the natural sciences. The points often record the location of key features, called landmarks, on continuous object boundaries, such as anatomical features on a human face. In other situations, the points may simply be arbitrarily spaced marks along a smooth curve, such as on handwritten numbers. This paper proposes novel exploratory methods for the identification of structure within point datasets. In particular, points are linked together to form curves which estimate the original shape from which the points are the only recorded information. Nonparametric regression methods are applied to polar coordinate variables obtained from the point locations and periodic modelling allows closed curves to be fitted even when data are available on only part of the boundary. Further, the model allows discontinuities to be identified to describe rapid changes in the curves. These generalizations are particularly important when the points represent shapes which are occluded or are intersecting. A range of real-data examples is used to motivate the modelling and to illustrate the flexibility of the approach. The method successfully identifies the underlying structure and its output could also be used as the basis for further analysis.

1. Introduction

Many scientific investigations involve the recording of spatially located data. This data might summarize objects within an image as digitized versions of continuous curves. Once the data are collected often the original context is lost and the aim of the analysis is to identify which points are associated with each other and to link the points to reconstruct the original shape. These can then be seen as estimates of continuous curves and object outlines. If the original scene contains multiple structures, then the analysis must also divide the points into groups with separate curves used to describe the points in each group. It is important to note that this is likely to form only the first part of an analysis and hence can be seen as exploratory data analysis.

This paper looks at the use of smoothing splines to identify and describe geometric patterns in sets of points. It is assumed that the points lie on smooth curves but that a dataset may contain multiple intersecting curves. It is vital that this be done in a nonparametric way so that the widest possible range of patterns can be highlighted. In general, these are closed, or nearly closed, curves and so a transformation to polar coordinates is used to simplify the analysis. Intersecting curves are described by allowing discontinuities in the fitted curves. These procedures are illustrated using simulated data and varied real datasets describing human faces, gorilla skulls, handwritten number 3's, and an archaeological site. These provide a wide variety of point patterns and reinforce the general usefulness of the proposed methods. For mathematical detailed description and applications of shape-based analysis of points, refer to, for example, Batschelet [1], Bookstein [2], Dryden and Mardia [3], and Lele and Richtsmeier [4].

To allow for this wide variety of possible curves a nonparametric fitting approach, such as splines, can be used (see, e.g., [5, 6]). The flexibility is helpful in the exploratory statistical analysis of a dataset, and the results can be used to suggest parametric equations for later analysis. Nonparametric regression is the general name for a range of curve fitting techniques which make few a priori assumptions about the true shape. In nonparametric regression, several different

families of basis functions can be used to describe curves; one of the common kinds of basis for smooth curves is the spline. Splines are generally defined as piecewise polynomials in which curve, or line, segments are joined together to form a continuous function. The spline smoothing approach to nonparametric regression is discussed, for example, by Silverman [7] and extended to deal with branching curves by defining a roughness penalty by Silverman and Wood [8]. For an introduction to natural cubic spline see Green and Silverman [9]. For more review of spline methods in statistics see Wegman and Wright [10], Silverman [11], Silverman [7], Nychka [12], and Wahba [13].

It is important to note that there are many existing general frameworks for performing spline-based regression. For example, multivariate adaptive regression splines (MARS) [14] or its more robust generalizations, RMARS [15] and RCMARS [16], with a good overview and comparison in [17]. These follow the general approach of general additive modelling [18] and give a formal framework for fitting and model selection.

A brief introduction to splines, along with the extension to circular data, is given in Section 2. The main results of this paper are given in Section 3 by considering modelling for single curves with occlusions and multiple intersecting curves. Although simulated examples are used to illustrate, the main real-data examples are given in Section 4. General discussion appears in Section 5.

2. Nonparametric Curve Estimation and Periodic Splines

A smoothing spline is a nonparametric curve estimator that is defined as the solution to a minimization problem. It provides a flexible smooth function for situations in which a simple polynomial or nonlinear regression model is not suitable. For a set of n observations $\chi = \{(x_i, y_i), \ i = 1, 2, \ldots, n\}$ consider a regression problem where the observations are assumed to satisfy

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where the errors ϵ_i are uncorrelated with zero mean and constant variance, σ^2. Then the spline smoothing method uses the data to construct a curve f by minimizing the objective function

$$J(f; \chi, \lambda) = \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 \\ + \lambda \int_{\infty}^{-\infty} \left(f^{(m)}(x) \right)^2 dx, \tag{2}$$

where $f^{(m)}$ represents the mth derivative of f, with m being a positive integer, and λ is a smoothing parameter. For more details of smoothing splines see, for example, Eubank [19], Eubank [6], and Cantoni and Hastie [20]. An alternative definition of the level of smoothing is in terms of an *equivalent degrees of freedom*, Df, which describes the amount of information in the data needed to estimate the

residuals. The function *smooth.spline* [21] allows λ or Df to be specified, but the degrees of freedom have been used in what follows as this gives a more intuitive interpretation.

The above objective function consists of two parts: the first measures the agreement of the function and the data and the second is a roughness penalty reflecting the total curvature—this can also be interpreted in a Bayesian setting as the likelihood and prior. Hence, for given Df, the estimate of f is given by

$$\widehat{f}(x, \text{Df}) = \min_{f} J(f; \chi, \text{Df}), \quad x \in \mathbf{R}. \tag{3}$$

If Df is large then the function is rough but closely fits the data, whereas when Df is small then the function is smooth but may not fit the data well. Here the choice of Df is made automatically using standard leave-one-out cross-validation [22]; that is,

$$\widehat{\text{Df}} = \min_{\text{Df}} \sum_{i=1}^{n} \left(y_i - \widehat{f}^{-i}(x_i, \text{Df}) \right)^2, \tag{4}$$

where $\widehat{f}^{-i}(\cdot, \text{Df})$ is the fitted spline curve, for given parameter Df, and with the ith data point, (x_i, y_i), being removed. Then $\widehat{f}(\cdot, \widehat{\text{Df}})$ is the fitted curve using the cross-validation estimate of the degrees of freedom.

Figure 1 shows fitted curves using splines with different degrees of freedom, Df. The true curve is a sine function with noise level $\sigma = 1/4$ which corresponds to a signal to noise ratio (SNR $= \sigma_f / \sigma$) of about 2. In (a) Df is about half the value found using cross-validation which is used in (b), with (c) using double the cross-validation degrees of freedom. The small degrees-of-freedom value gives a smoother fitted curve that ignores many of the points in the data whereas a large value produces a rougher fit which more closely follows the data. The automatic choice was $\widehat{\text{Df}} = 5.5$ which gives a very good fit to the data reproducing the sin curve well.

For this dataset, the periodic nature of the sin function has, so far, been ignored, and it is clear that the extreme left and right do not match exactly. For such datasets, made up of angles or directions, ignoring the periodic nature of the measurements when smoothing may produce unacceptable edge effects. A simple approach for dealing with this issue will now be considered.

Suppose that the dataset is made up of paired angles and distances which will be denoted as $\vartheta = \{(\theta_i, r_i) : i = 1, \ldots, n\}$ for a sample of size n. A simple approach for periodic data measured in the interval $(0, 2\pi)$, say, is to repeat the data. That is, for each angle θ_i, the corresponding new angular values are $(\theta_i - p\pi, \ldots, \theta_i - \pi, \ \psi, \ \theta_i + \pi, \ldots, \theta_i + p\pi)$, where $p = 1, 2, \ldots$, and similarly repeat the corresponding radial distances r_i to be $(r_i, \ldots, r_i)$. This produces a dataset, $\vartheta^p = \{(\theta_i, r_i) : i = 1, \ldots, n'\}$, with $n' = (2p + 1) \times n$ data values, and even for small p (e.g., $p = 1$ or 2) this gives a very good approximation to the full periodic spline. Cogburn and Davis [23] present the theory of periodic smoothing spline with application to the estimation of periodic functions and the R function `periodicSpline` from the package `splines` might provide an alternative computational approach.
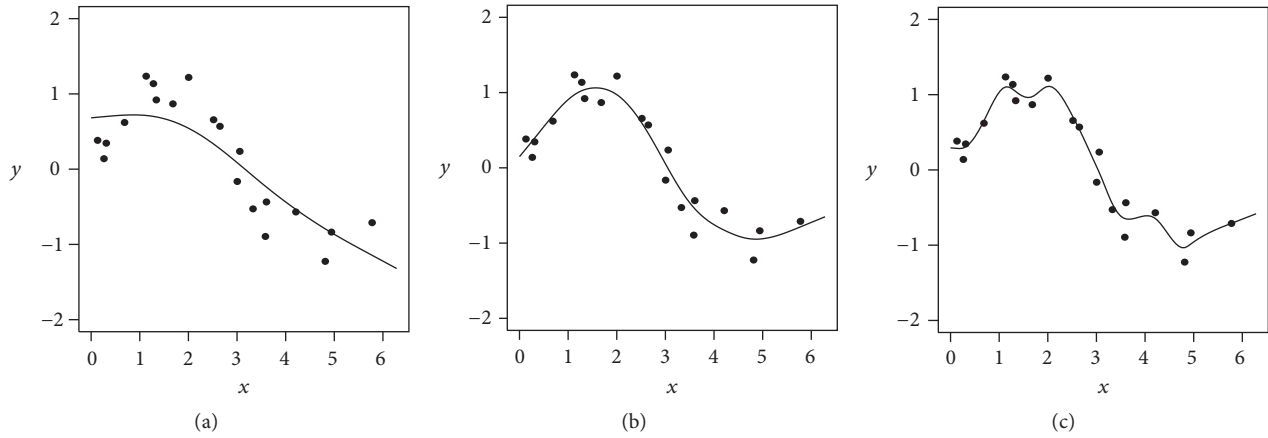
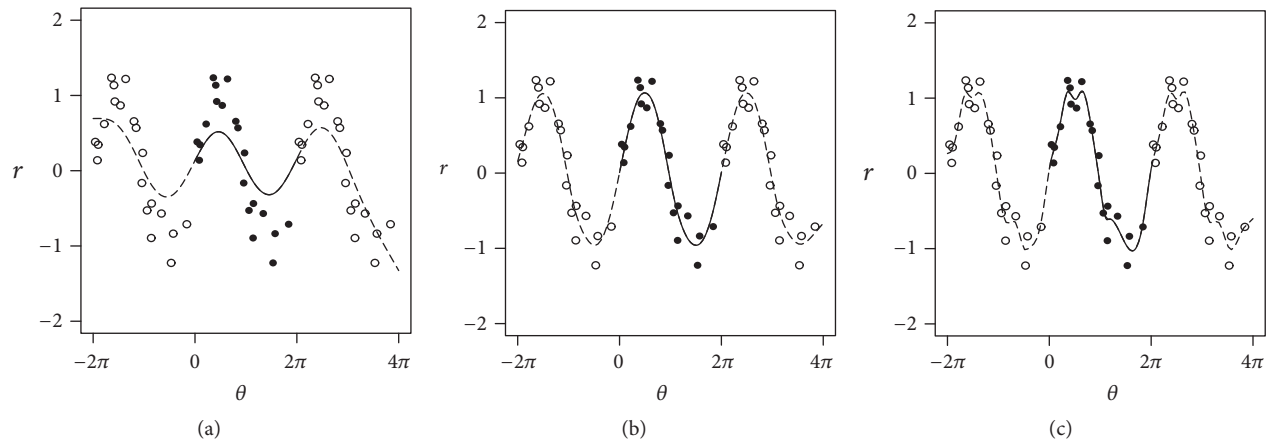FIGURE 1: Smoothing spline fits to sin data: (a) $\mathtt{Df} = 3$; (b) $\widehat{\mathtt{Df}} = 5.5$; (c) $\mathtt{Df} = 11$.

FIGURE 2: Periodic spline fits to sin data: (a) $\mathtt{Df} = 7$; (b) $\mathtt{Df}_{\mathrm{CV}} = 15$; (c) $\mathtt{Df} = 30$.

As illustration consider Figure 2 which shows fitted curves equivalent to those in Figure 1 but with $p = 1$. The solid circles are the original data with open circles representing the copied data points. Similarly, the solid line is the spline fitted curve over the original interval with the dashed line showing the fitted curve over the copied data points. In all cases the fit is better than in Figure 1, with the periodic nature, reproduced well, and as before the cross-validation choice of smoothing has produced an excellent reconstruction of the true sin curve.

Once fitted a residual sum of squares, RSS, calculated on the original data values, can be used as a measure of goodness-of-fit. Here this will be calculated using the radial distances with definition

$$\mathrm{RSS} = \sum_{i=1}^{n} \left(r_i - \widehat{r}\left(\theta_i, \widehat{\mathtt{Df}}\right)\right)^2 \tag{5}$$

but other versions could be used, for example, the Euclidean distance between fitted and observed points.

Of course, the approach could lead to a poor fit if the data is not periodic, but to prevent this it is possible to allow for a discontinuity in the relationship. Here the approach of Gu

[24], who considered discontinuities in cubic splines with a jump at a known location, will be extended to the periodic case and with an unknown discontinuity location.

Suppose that the points $\vartheta = \{(r_i, \theta_i) : i = 1, \ldots, n\}$ are partitioned into two groups with the first, ϑ_1, containing all the points with angles up to and including the change point and ϑ_2 those with angles above. Assuming that the points are ordered in increasing value of the angle, so that $\theta_1 \leq \cdots \leq \theta_n$, then let $\vartheta_1 = \{(\theta_i, r_i); i = 1, \ldots, k\}$ be the data before the change point and $\vartheta_2 = \{(\theta_i, r_i); i = k+1, \ldots, n\}$ the remaining data. For change point at θ_k two curves are fitted to the data such that

$$\widehat{r}\left(\theta, \widehat{\mathtt{Df}}\right) = \begin{cases} \min_r J\left(r; \vartheta_1, \widehat{\mathtt{Df}}_1\right), & \text{for } \theta \leq \theta_k \\ \min_r J\left(r; \vartheta_2, \widehat{\mathtt{Df}}_2\right), & \text{for } \theta > \theta_k, \end{cases} \tag{6}$$

where cross-validation is used separately on the two parts leading to two degrees of freedom, $\widehat{\mathtt{Df}} = (\widehat{\mathtt{Df}}_1, \widehat{\mathtt{Df}}_2)$. The significance of the change point could be assessed through a chi-squared test, but here a change point influence graph is considered based on the goodness-of-fit.

Consider the sin data shown in Figure 3(a) which has a change point of size about 1 introduced at $\theta = \theta_{[10]} \approx 2.3$.
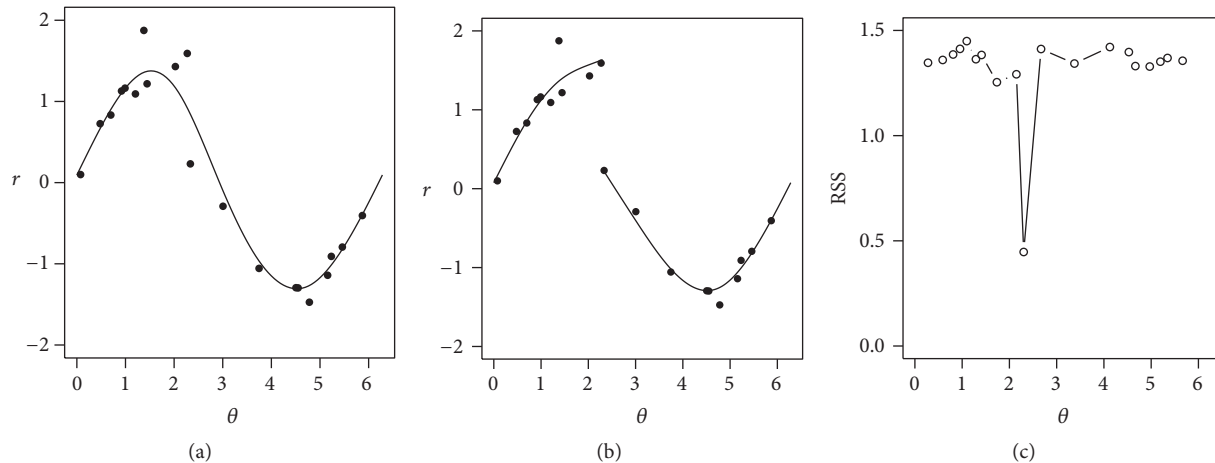
FIGURE 3: (a) Data from sin function with a discontinuity; (b) best two-part curve; (c) residual sum of square for two-part curves.

The curve in Figure 3(a) is fitted using a smoothing spline, but ignoring the change point, it is possible that the curves in Figure 3(b) are fitted using smoothing splines with change point at the estimated location. The automatically chosen value for the degrees of freedom, $\widehat{\text{Df}}$, for the single curve in (a) is 14.5 whereas for the two-part curve the overall degrees of freedom are $\widehat{\text{Df}}_1 + \widehat{\text{Df}}_2 = 11$. Figure 3(c) shows the residual sum of squares, RSS, for each possible change point location with a very clear minimum. The RSS for the curve in Figure 3(a) is 1.3 while, in (b), it has reduced to 0.45, which is substantially smaller and provides a much better description of the data. Hence this approach provides an intuitive approach to finding change points in data automatically.

3. A Model for Multiple Overlapping Curves

3.1. Motivation. To motivate the modelling, consider an unobserved true scene containing a few objects of various shape and sizes, with possible overlap. However, instead of the scene being recorded faithfully, only partial information is taken and, in particular, only points along the edges of the objects are recorded. These points might be chosen to identify features with special significance or they might simply be at equal or random locations along the edge. Further, due to overlaps, points from the full edge may not be in the dataset. Once collected, there is no record of which points are from which object, and no record is kept of possible object shapes nor even the number of objects. Hence, let the dataset consists of a collection of n points, $\chi = \{(x_i, y_y) : i = 1, \ldots, n\}$, recorded within some small region in 2D.

Figure 4 shows example datasets which will be analysed later. Panel (a) shows a human face profile with the forehead, eyes, nose, mouth, and chin clearly identifiable on the left—the points on the right locate the back of the neck and the hairline. Panel (b) shows points located along a handwritten number 3 at approximately equally spaced intervals.

3.2. Modelling a Single Curve with Occlusion. Before the periodic smoothing spline approach can be applied it is necessary

for the data to be first transformed to polar coordinates. First define a *centre*, (ξ, ζ), which can be estimated using the data centroid $(\widehat{\xi}, \widehat{\zeta}) = (\overline{x}, \overline{y})$ and then use the one-to-one transformation

$$r_i = \left((x_i - \overline{x})^2 + (y_i - \overline{y})^2 \right)^{1/2},$$

$$\theta_i = \tan^{-1}\left(\frac{(y_i - \overline{y})}{(x_i - \overline{x})} \right), \tag{7}$$

$$i = 1, \ldots, n.$$

This gives rise to an alternative data representation via the centre $(\overline{x}, \overline{y})$ and polar coordinates $\varphi = \{(r_i, \theta_i) : i = 1, \ldots, n\}$. Note that although this representation contains $n + 2$ pieces of information, by construction, the polar coordinate variables are not independent. Of course, other estimates of centre could be considered, such as the point which minimizes the variance of the radii. In particular, this measure should be more robust to presence of occlusions.

To illustrate the transformation and the subsequent spline smoothing consider the simulated data in Figure 5. Panel (a) shows the given points along with the sample centre marked with a "+"; the points in (b) are the corresponding polar coordinates relative to this centre. Also shown in (b) are the nonperiodic smoothing spline (continuous black line) and the period smoothing spline (dashed red line). These are all closely aligned except at the extreme angles. Once transformed back into Cartesian coordinates, as shown in panel (c), the slight discrepancies between the fitted splines are more clearly visible. At the far right of the plot, the periodic spline curve is closed and more naturally represents a possible object, whereas the nonperiodic spline is not closed making it difficult to interpret if this were part of the edge of a real object.

Figure 6 shows a second elliptical dataset but where part of the ellipse is missing. The Cartesian data are shown in (a) and (c), with the polar transformed data in (b). Panel (b) shows the nonperiodic spline and the period spline with dramatic differences which are even more obvious when the
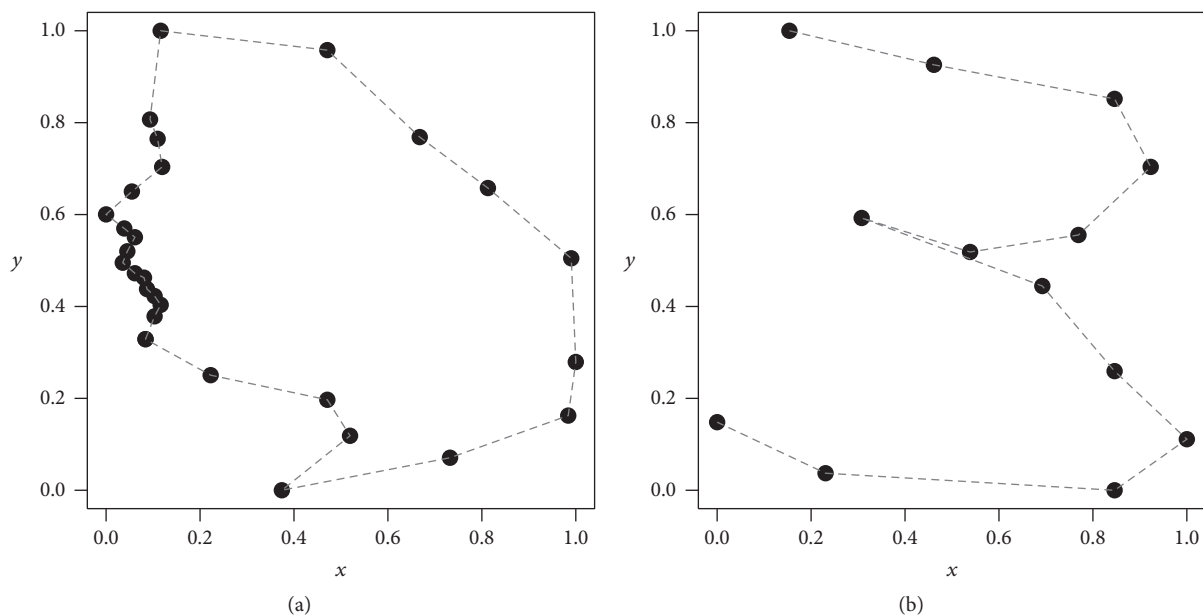
FIGURE 4: Real datasets: (a) human face and (b) handwritten number 3.

FIGURE 5: (a) Simulated data; (b) polar coordinate data with fitted spline curve; (c) data with back-transformed fitted curves. In (b) and (c) the solid curves use standard splines, whereas the dashed use the periodic spline.
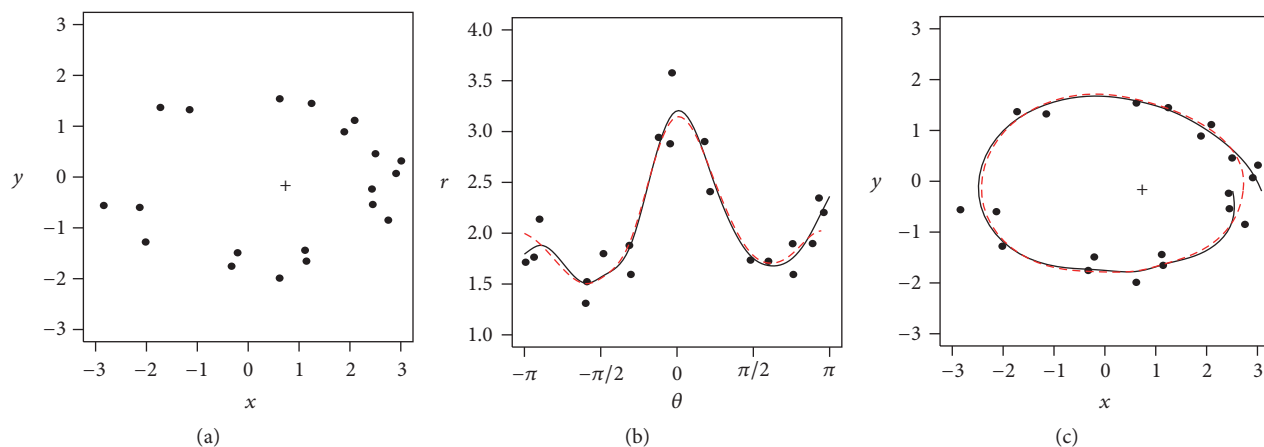

FIGURE 6: (a) Occluded data; (b) polar coordinate data with fitted spline curve; (c) data with back-transformed fitted curves. In (b) and (c) the solid curves use standard splines, whereas the dashed use the periodic spline.
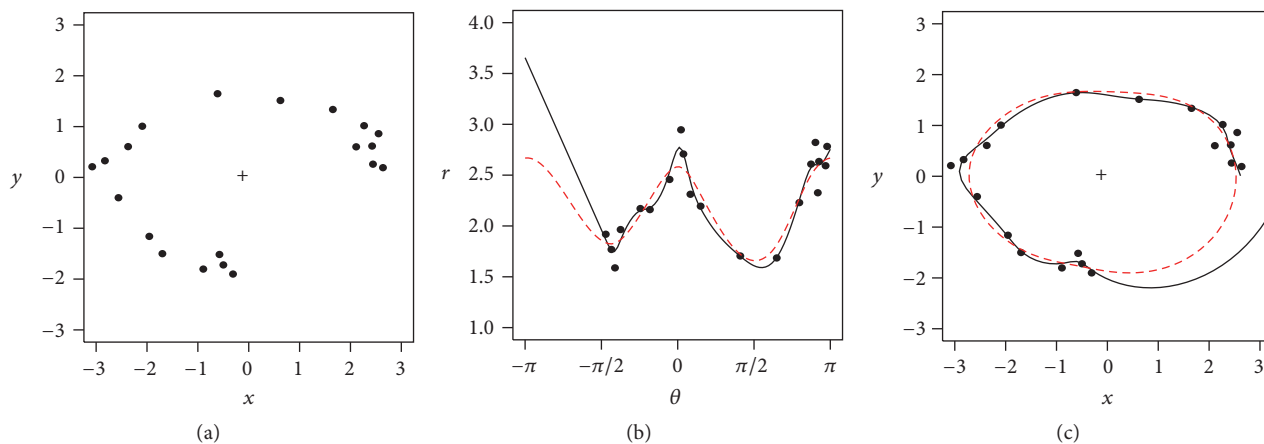
fitted curves are transformed back into Cartesian coordinates, as shown in panel (c). The periodic smoothing spline has done a very good job of interpolating the missing part of the curve and the results can easily be relied upon in further analysis. In particular, slight changes in the position of a few critical point will lead to very different shapes for the nonperiodic spline.

To summarize, application of smoothing splines to periodic point data has proved very successful. The modification of the duplicated data is a simple, yet effective way to create closed curves and to interpolate where data are missing. The approach has provided a robust and informative reconstruction of the unknown curve from the data.

3.3. Modelling Multiple Intersecting Curves.

To allow for intersecting and overlapping curves the points are partitioned into m groups, S_j, where $j = 1, 2, \ldots, m$. That is, $S_j \subseteq (1, \ldots, n)$ with $S_i \cap S_j = \emptyset$ when $i \neq j$ and $S_1 \cup \cdots \cup S_m = (1, \ldots, n)$. To record group membership a matrix $W_{n \times m} = (w_{ij})$ is defined, where $w_{ij} = 1$ if point i belongs to group j ($i \in S_j$) and $w_{ij} = 0$ otherwise. Then, $\sum_j w_{ij} = 1$ and $\sum_i w_{ij} = n_j$, where n_j is the number of points in the jth group; that is, $n_j = |S_j|$. For each group, working in polar coordinates, there is a centre, (ξ_j, ζ_j), and coordinates relative to the centre, $\vartheta_j = \{(r_{ij}, \theta_{ij}) : i = 1, \ldots, n_j\}$, with the full set of parameters denoted as $\vartheta = \{\vartheta_j : j = 1, \ldots, m\}$. The corresponding Cartesian coordinates can be written as $\Gamma_j = \{(\mu_{ij}, \nu_{ij}) : i = 1, \ldots, n_j\}$, with

$$\mu_{ij} = \xi_j + r_{ij} \cos\left(\theta_{ij}\right),$$

$$\nu_{ij} = \zeta_j + r_{ij} \sin\left(\theta_{ij}\right), \tag{8}$$

$$\text{for } i = 1, \ldots, n_j, \ \ j = 1, \ldots, m,$$

and the full collection of data as $\Gamma = \{\Gamma_j : j = 1, \ldots, m\}$. Further, it is assumed that the point locations are recorded with error giving observed measurements

$$x_{ij} = \mu_{ij} + \epsilon_{ij},$$

$$y_{ij} = \nu_{ij} + \varepsilon_{ij}, \tag{9}$$

$$\text{for } i = 1, \ldots, n_j, \ \ j = 1, \ldots, m,$$

where ϵ and ε are independent Gaussian random variables with zero mean and constant variance σ^2.

In what follows the full dataset will, without further explanation, be referred to using either $\chi = \{(x_{ij}, y_{ij}) : i = 1, \ldots, n_j, \ j = 1, \ldots, m\}$ and $\vartheta = \{(\theta_{ij}, r_{ij}) : i = 1, \ldots, n_j, \ j = 1, \ldots, m\}$ or equivalently, but without explicit reference to the group membership, $\chi = \{(x_i, y_i) : i = 1, \ldots, n\}$ and $\vartheta = \{(\theta_i, r_i) : i = 1, \ldots, n\}$ as is most convenient and intuitive.

3.4. Estimation with Multiple Intersecting Curves.

Now consider estimation of the model unknowns from observed data. Start by supposing that a dataset is available but that the group membership information is intact; then the group centres could be estimated as

$$\widehat{\xi}_j = \overline{x}_j = \frac{\sum_{i=1}^{n} w_{ij} x_i}{\sum_i w_{ij}},$$

$$\widehat{\zeta}_j = \overline{y}_j = \frac{\sum_{i=1}^{n} w_{ij} y_i}{\sum_i w_{ij}} \tag{10}$$

and, although some of these are unimportant, corresponding polar coordinate representation of point i relative to group centre j is

$$\widehat{r}_{ij} = \left(\left(x_i - \overline{x}_j\right)^2 + \left(y_i - \overline{y}_j\right)^2\right)^{1/2}, \tag{11}$$

$$\widehat{\theta}_{ij} = \tan^{-1}\left(\frac{y_i - \overline{y}_j}{x_i - \overline{x}_j}\right), \tag{12}$$

where $i = 1, \ldots, n$ and $j = 1, \ldots, m$. The overall residual sum of squares is then the sum of the separate components

$$\text{RSS} = \sum_{j=1}^{m} \text{RSS}_j = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \left(r_{ij} - \widehat{r}\left(\theta_{ij}, \widehat{\text{Df}}_j\right)\right)^2. \tag{13}$$

Now consider the case when the group membership is unknown and must be inferred from the data. The aim is to find linked points by fitting curves. Some datasets have more than one curve and some have intersecting curves. Then classifying the points into groups may help to fit the correct curves that represent the data.

In general, this can be thought of as a change point problem, as already discussed, to address the lack of stationarity in the values. A change point occurs at some point in the data if all of the values up to and including it share a common curve while all those after the change point share another. This is exactly the same situation as the discussion in Section 2 and hence the same method of solution is applied.

4. Application to Real Data

4.1. General.

The previous sections have illustrated the proposed exploratory data analysis tools on simulated example, whereas in this section the success of the approach is demonstrated on a varied range of real datasets. There is no wish to construct formal equations to define the shape but to stimulate further analyses.

4.2. Example 1: Face Data.

The first experiment is conducted on data extracted from the human face [25] in a study looking at changes in shape due to growth in children. Figure 7(a) shows the data with points joining the points; then (b) shows the points transformed to polar coordinates along with fitted spline curves. Figure 7(c) shows the data set with back-transformation fitted values, and the solid curve shows those from the standard spline while the dotted curve shows those from the periodic spline. It is clear from the fitted curves that there is not much difference between the periodic and
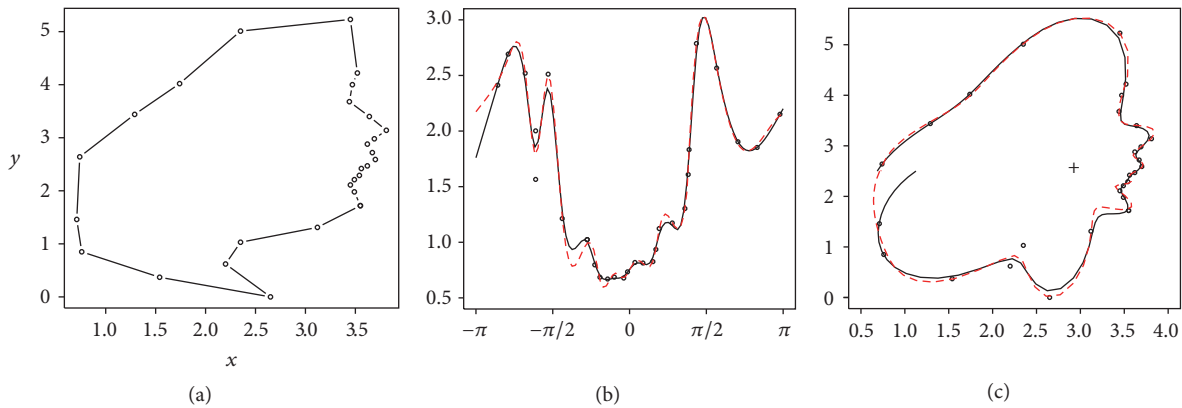
FIGURE 7: (a) Face data; (b) polar coordinate data with fitted spline curves; (c) back-transformed fitted curves. In (b) and (c) the solid curves use standard splines, whereas the dotted use periodic splines.
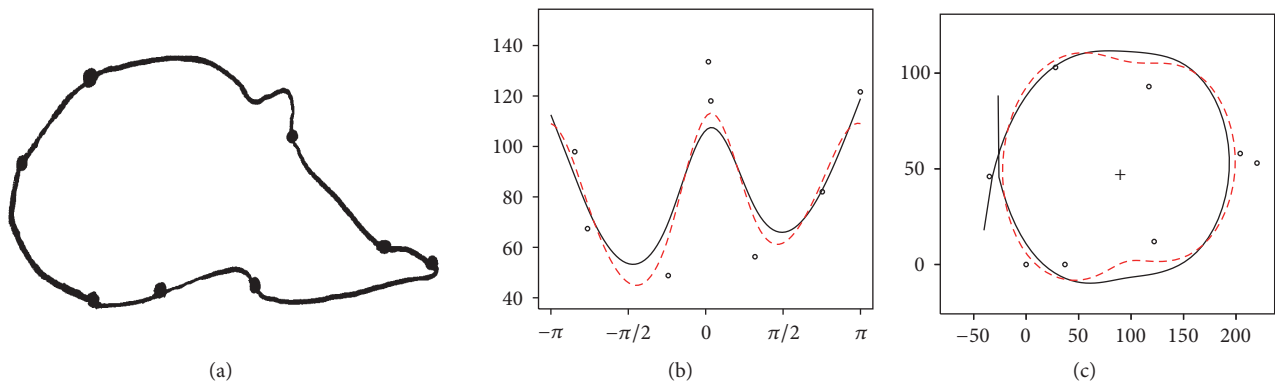
FIGURE 8: (a) Schematic diagram of a gorilla skull with anatomical landmarks for a male gorilla; (b) landmarks in polar coordinate and spline curves; (c) landmarks along with back-transformed fitted spline curves. In (b) and (c) the solid curves use standard splines, whereas the dotted use the periodic spline.

the standard smoothing splines. Both produce well fitted curves for the face. It is worth noting that the fitted curve can be evaluated at arbitrarily close locations, not only at the data points, and hence a smoothly interpolated curve can be drawn.

4.3. Example 2: Gorilla Skulls. This dataset, taken from Dryden and Mardia [3], is composed of 8 anatomical landmarks from the skulls of 29 male and 30 female gorillas. A landmark is defined as *a point of correspondence on each object that matches between and within populations* [3]. Figure 8(a) shows a schematic diagram of a typical skull with the landmarks indicated.

Figure 8(c) shows landmarks for one of the male gorillas and Figure 8(b) the corresponding points in polar coordinates along with spline fitting to the dataset and in Figure 8(c) after back-transforming. For both, the fits are good but at the expense of low smoothing in the spline. This fitting procedure was repeated for the other gorilla skulls and surprisingly the smoothed curves give good summaries allowing the skulls to be easily categorised into four main groups covering mainly male skulls which are rather elongated and two covering mainly female skulls which appear more rounded. The males

lead to generally larger values of the degrees of freedom ($6 < \mathtt{Df} < 8$) than the females ($\mathtt{Df} \approx 2$). In fact, the automatic choice of the degrees-of-freedom parameter can be used as a simple discrimination variable giving only 8 out of 59 incorrectly classified skulls. It is important to note that this was not a preconceived discriminator but was identified by the exploratory analysis. This has highlighted the usefulness of simple and flexible tools as a preliminary step in a more wide-ranging investigation.

4.4. Example 3: The Number 3. Another dataset, again taken from Dryden and Mardia [3], is made up of 13 landmarks from 30 handwritten number 3's; see Figure 9(a). Suppose the data are divided into two subsets with n_1 and n_2 observations, respectively. The best partition is made according to the minimum value of the overall residual sum of squares, RSS, which is displayed in panel (c). Each subset is transformed to polar coordinates using the different centres marked "+" in panel (a). Each subset is indicated by different marks along with their fitted spline curves as plotted in panel (b) with the back-transformed fitted curves in panel (a). Clearly, this has described the two-part curves very well. Again, this demonstrates the flexibility of the procedure.
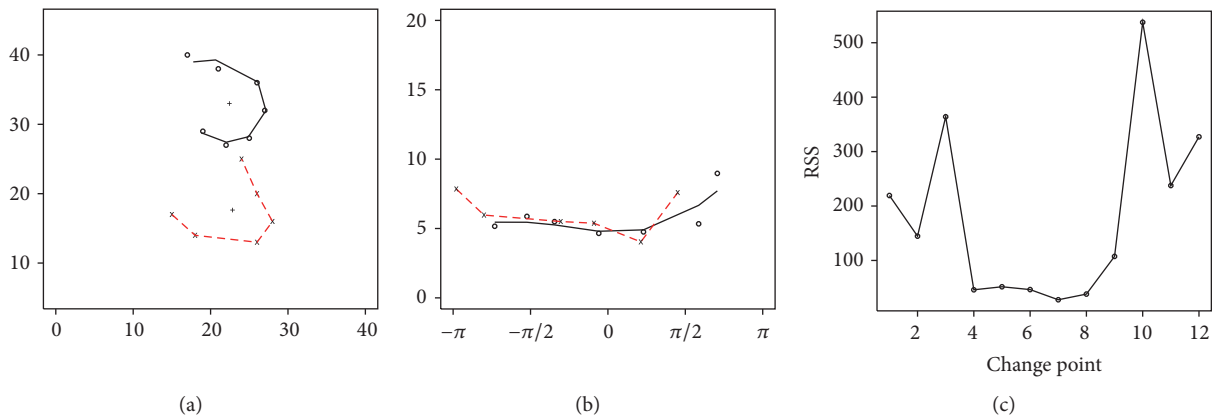
FIGURE 9: (a) A typical *number 3* dataset along with the back-transformed fitted two-part spline; (b) points in polar coordinates and two-part spline curve; (c) RSS plotted against change point position.
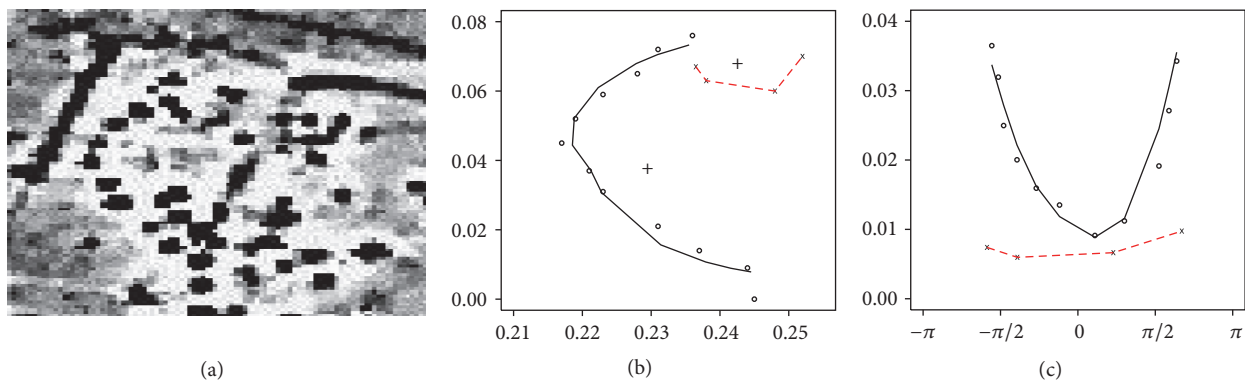
FIGURE 10: (a) Magnetic survey data for part of an iron-age archaeological site; (b) selected pits along with back-transformed fitted two-part spline curve; (c) polar coordinates and fitted two-part spline curve.

4.5. Example 4: Archaeological Site Data. The data in Figure 10(a) shows part of a typical image dataset (supplied by Alistair Marshall of the Guiting Power Amenity Trust; see Aykroyd et al. [26] for details) from a magnetic survey of an archaeological site. As well as linear features, which represent ditches, there are also several drifts of pits, but blurring and noise tend to camouflage the exact locations. Panel (b) shows the locations of some of these pits, appearing as small circles, and panel (c) shows the corresponding polar coordinates relative to the two data centres (marked "+" in (b)). According to the minimum value of the residual sum of squares, RSS, the observations can be classified automatically into two groups.

The data centres are calculated for each subset, the small circles are the data in the first subset, and "×" are the data in the second subset, with the fitted curves plotted in Figure 10(c). Then the fitted curves are back-transformed into Cartesian coordinate as shown in panel (b). The solid curve is for the first subset while the dotted curve is for the second subset. The aim of the analysis is to identify which points are associated with each other and to fit curves to the points, and this has been achieved well. The resulting linked points might then form part of further analysis or aid physical excavation.

5. Discussion

Making sense of clouds of points, apparently randomly placed across a 2D region, is a key task in many statistical investigations. When the points are recorded without additional information, the first task is to infer structure by linking points using a data-driven approach. This paper has proposed and investigated a simple, yet effective method based on change point identification and nonparametric spline smoothing. It provides an intuitive explanatory tool to identify patterns in the point locations. When it is assumed that the structures form lines and curves, the change points divide the data into subsets, with the splines providing a flexible method to infer the shape of the structures. The method has easily dealt with occlusions and intersections in scenes with multiple curves. Similar results might be achieved by applying more general modelling approaches, such as MARS, RARS, RCMARS; for details see, for example, [17], but we believe that a more straightforward and intuitive approach can have equal impact by bringing a range of easy-to-use tool to a wider audience. Further, for all users the methods considerations can be used to suggest further analyses based on more sophisticated approaches.

There is scope for extending the approach to include larger numbers of curves where it is not possible to divide the curves with a single change point. The nature of the problem is closely related to classification where the group membership is missing. This strongly suggests that a probabilistic approach might be considered based on statistical distribution models. This would then fit into the general framework where the EM algorithm has proven very useful. Also, there is a need to extend the approach to deal with unordered points and ones which are not star-shaped. These are areas of possible future work. Further, it is of interest to develop a similar procedure which would allow more formal modelling and model section, perhaps following the approach of general additive modelling [18].

The applications are various and varied with an illustrative example of the method when the data points are anatomical landmarks defined by geometrical features, equally spaced but blindly placed points along smooth curves and from extreme intensity points in grey-scale images. Further, the results of the analysis have provided new variables which could be the starting point for other analyses. Hence there is potential for this to be a valuable exploratory data analysis method in the tool-kit of applied statisticians and applied scientists.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] E. Batschelet, *Circular Statistics in Biology*, Academic Press, London, UK, 1981.

[2] F. L. Bookstein, *Morphometric Tools for Landmark Data: Geometry and Biology*, Cambridge Univesity Press, Cambridge, UK, 1991.

[3] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Chichester, UK, 1998.

[4] S. Lele and J. Richtsmeier, *An Invariant Approach to Statistical Analysis of Shapes*, Chapman & Hall/CRC, 2001.

[5] T. P. Ryan, *Modern Regression Methods*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, John Wiley & Sons, New York, NY, USA, 1997.

[6] R. L. Eubank, *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, New York, NY, USA, 1999.

[7] B. W. Silverman, "Some aspects of the spline smoothing approach to nonparametric regression curve fitting," *Journal of the Royal Statistical Society B*, vol. 47, no. 1, pp. 1–52, 1985.

[8] B. W. Silverman and J. T. Wood, "The nonparametric estimation of branching curves," *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 551–558, 1987.

[9] P. J. Green and B. W. Silverman, *Non Parametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman and Hall, 1994.

[10] E. J. Wegman and I. W. Wright, "Splines in statistics," *Journal of the American Statistical Association*, vol. 78, no. 382, pp. 351–365, 1983.

[11] B. W. Silverman, "A fast and efficient cross-validation method for smoothing parameter choice in spline regression," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 584–589, 1984.

[12] D. Nychka, "Splines as local smoothers," *The Annals of Statistics*, vol. 23, no. 4, pp. 1175–1197, 1995.

[13] G. Wahba, "Splines in nonparametric regression," in *Encyclopedia of Environmetrics*, John Wiley & Sons, New York, NY, USA, 2006.

[14] J. H. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, vol. 19, pp. 1–67, 1991.

[15] A. Özmen and G. W. Weber, "RMARS: robustification of multivariate adaptive regression spline under polyhedral uncertainty," *Journal of Computational and Applied Mathematics*, vol. 259, pp. 914–924, 2014.

[16] A. Özmen, G. W. Weber, I. Batmaz, and E. Kropat, "RCMARS: robustication of CMARS with different scenarios under polyhedral uncertainty set," *Communications in Nonlinear Science and Numerical Simulation*, vol. 16, pp. 1780–1787, 2011.

[17] A. Özmen, *Robust Optimization of Spline Models and Complex Regulatory Networks*, Contributions to Management Science, Springer, 2016.

[18] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*, Chapman & Hall/CRC, 1990.

[19] R. L. Eubank, "Diagnostics for smoothing splines," *Journal of the Royal Statistical Society. Series B. Methodological*, vol. 47, no. 2, pp. 332–341, 1985.

[20] E. Cantoni and T. Hastie, "Degrees-of-freedom tests for smoothing splines," *Biometrika*, vol. 89, no. 2, pp. 251–263, 2002.

[21] R Core Team, *R: A Language and Environment for Statistical Computing,*, R Foundation for Statistical Computing, Vienna, Austria, 2016, http://www.R-project.org/.

[22] P. Craven and G. Wahba, "Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation," *Numerical Mathematics*, vol. 31, pp. 377–403, 1979.

[23] R. Cogburn and H. T. Davis, "Periodic splines and spectral estimation," *The Annals of Statistics*, vol. 2, pp. 1108–1126, 1974.

[24] C. Gu, "Multivariate spline regression," in *Smoothing and Regression: Approaches, Computation and Application*, M. G. Schimek, Ed., pp. 329–354, John Wiley & Sons, New York, NY, USA, 2000.

[25] R. J. Morris, J. T. Kent, K. V. Mardia, R. G. Aykroyd, M. Fidrich, and A. Linney, *Exploratory Analysis of Facial Growth*, Leeds University Press, 1999.

[26] R. G. Aykroyd, J. G. Haigh, and G. T. Allum, "Bayesian methods applied to survey data from archeological magnetometry," *Journal of the American Statistical Association*, vol. 96, no. 453, pp. 64–76, 2001.

Numerical Reconstruction of the Covariance Matrix of a Spherically Truncated Multinormal Distribution

Filippo Palombi,[1,2] Simona Toti,[1] and Romina Filippini[1]

[1]*Istituto Nazionale di Statistica (ISTAT), Via Cesare Balbo 16, 00184 Rome, Italy*
[2]*Italian Agency for New Technologies, Energy and Sustainable Economic Development (ENEA),*
Via Enrico Fermi 45, 00044 Frascati, Italy

Correspondence should be addressed to Filippo Palombi; filippo.palombi@enea.it

We relate the matrix $\mathfrak{S}_{\mathscr{B}}$ of the second moments of a spherically truncated normal multivariate to its full covariance matrix Σ and present an algorithm to invert the relation and reconstruct Σ from $\mathfrak{S}_{\mathscr{B}}$. While the eigenvectors of Σ are left invariant by the truncation, its eigenvalues are nonuniformly damped. We show that the eigenvalues of Σ can be reconstructed from their truncated counterparts via a fixed point iteration, whose convergence we prove analytically. The procedure requires the computation of multidimensional Gaussian integrals over an Euclidean ball, for which we extend a numerical technique, originally proposed by Ruben in 1962, based on a series expansion in chi-square distributions. In order to study the feasibility of our approach, we examine the convergence rate of some iterative schemes on suitably chosen ensembles of Wishart matrices. We finally discuss the practical difficulties arising in sample space and outline a regularization of the problem based on perturbation theory.

1. Introduction

It has been more than forty years since Tallis [1] worked out the moment-generating function of a normal multivariate $X \equiv \{X_k\}_{k=1}^{\nu} \sim \mathscr{N}_\nu(0, \Sigma)$, subject to the conditional event

$$X \in \mathscr{E}_\nu\left(\rho; \Sigma\right), \quad \mathscr{E}_\nu\left(\rho; \Sigma\right) \equiv \left\{ x \in \mathbb{R}^\nu : x^{\mathrm{T}}\Sigma^{-1}x \le \rho \right\}. \quad (1)$$

The perfect match between the symmetries of the ellipsoid $\mathscr{E}_\nu(\rho; \Sigma)$ and those of $\mathscr{N}_\nu(0, \Sigma)$ allows for an exact analytic result, from which the complete set of multivariate truncated moments can be extracted upon differentiation. Consider, for instance, the matrix $\mathfrak{S}_{\mathscr{E}}(\rho; \Sigma)$ of the second truncated moments, expressing the covariances among the components of X within $\mathscr{E}_\nu(\rho; \Sigma)$. From Tallis' paper it turns out that

$$\mathfrak{S}_{\mathscr{E}}\left(\rho; \Sigma\right) = c_{\mathrm{T}}\left(\rho\right) \Sigma, \quad c_{\mathrm{T}}\left(\rho\right) \equiv \frac{F_{\nu+2}\left(\rho\right)}{F_\nu\left(\rho\right)}, \quad (2)$$

with F_ν denoting the cumulative distribution function of a χ^2-variable with ν degrees of freedom. Inverting (2)—so as to express Σ as a function of $\mathfrak{S}_{\mathscr{E}}$—is trivial, since $c_{\mathrm{T}}(\rho)$ is a scalar damping factor independent of Σ. In this paper, we shall refer to such inverse relation as the *reconstruction* of Σ from $\mathfrak{S}_{\mathscr{E}}$. Unfortunately, life is not always so easy. In general, the effects produced on the expectation of functions of X by cutting off the probability density outside a generic domain $\mathscr{D} \subset \mathbb{R}^\nu$ can be hardly calculated in closed form, especially if the boundary of $\mathscr{D}$ is shaped in a way that breaks the ellipsoidal symmetry of $\mathscr{N}_\nu(0, \Sigma)$. Thus, for instance, unlike (2), the matrix of the second truncated moments is expected in general to display a nonlinear/nontrivial dependence upon Σ.

In the present paper, we consider the case where $\mathscr{D}$ is a ν-dimensional Euclidean ball with center in the origin and square radius ρ. Specifically, we discuss the reconstruction of Σ from the matrix $\mathfrak{S}_{\mathscr{B}}$ of the spherically truncated second moments. To this aim, we need to mimic Tallis' calculation, with (1) being replaced by the conditional event

$$X \in \mathscr{B}_\nu\left(\rho\right), \quad \mathscr{B}_\nu\left(\rho\right) \equiv \left\{ x \in \mathbb{R}^\nu : x^{\mathrm{T}}x \le \rho \right\}. \quad (3)$$

This is precisely an example of the situation described in the previous paragraph: although $\mathscr{B}_\nu(\rho)$ has a higher degree of symmetry than $\mathscr{E}_\nu(\rho; \Sigma)$, still there is no possibility

of obtaining a closed-form relation between Σ and $\mathfrak{S}_{\mathscr{B}}$, since $\mathscr{B}_v(\rho)$ breaks the ellipsoidal symmetry of $\mathscr{N}_v(0, \Sigma)$: technically speaking, in this case we cannot perform any change of variable under the defining integral of the moment-generating function, which may reduce the dimensionality of the problem, as in Tallis' paper.

In spite of that, the residual symmetries characterizing the truncated distribution help simplify the problem in the following respects: (i) the reflection invariance of the whole setup still yields $\mathbb{E}[X_k \mid X \in \mathscr{B}_v(\rho)] = 0\ \forall k$ and (ii) the rotational invariance of $\mathscr{B}_v(\rho)$ preserves the possibility of defining the principal components of the distribution just like in the unconstrained case. In particular, the latter property means that $\mathfrak{S}_{\mathscr{B}}$ and Σ share the same orthonormal eigenvectors. In fact, the reconstruction of Σ from $\mathfrak{S}_{\mathscr{B}}$ amounts to solving a system of nonlinear integral equations, having the eigenvalues $\lambda \equiv \{\lambda_k\}_{k=1}^v$ of Σ as unknown variables and the eigenvalues $\mu \equiv \{\mu_k\}_{k=1}^v$ of $\mathfrak{S}_{\mathscr{B}}$ as input parameters. In a lack of analytic techniques to evaluate exactly the integrals involved, we resort to a numerical algorithm, of which we investigate feasibility, performance, and optimization.

The paper is organized as follows. In Section 2, we describe a few examples illustrating the occurrence of spherical truncations in practical situations. In Section 3, we show that the aforementioned integral equations have the analytic structure of a fixed point vector equation; that is to say, $\lambda = T(\lambda)$. This suggests achieving the reconstruction of λ numerically via suitably chosen iterative schemes. In Section 4, we prove the convergence of the simplest of them by inductive arguments, the validity of which relies upon the monotonicity properties of ratios of Gaussian integrals over $\mathscr{B}_v(\rho)$. In Section 5, we review some numerical techniques for the computation of Gaussian integrals over $\mathscr{B}_v(\rho)$ with controlled systematic error. These are based on and extend a classic work by Ruben [2] on the distribution of quadratic forms of normal variables. For the sake of readability, we defer proofs of statements made in this section to the appendix. In Section 6, we report on our numerical experiences: since the simplest iterative scheme, namely, the Gauss–Jacobi iteration, is too slow for practical purposes, we investigate the performance of its improved version based on overrelaxation; as expected, we find that the latter has a higher convergence rate; yet it still slows down polynomially in $1/\rho$ as $\rho \to 0$ and exponentially in v as $v \to \infty$; in order to reduce the slowing down, we propose an acceleration technique, which boosts the higher components of the eigenvalue spectrum. A series of Monte Carlo simulations enables us to quantify the speedup. In Section 7 we discuss the problems arising when μ is affected by statistical uncertainty and propose a regularization technique based on perturbation theory. To conclude, we summarize our findings in Section 8.

2. Motivating Examples

Spherical truncations of multinormal distributions may characterize different kinds of experimental devices and may occur in various problems of statistical and convex analysis. In this section, we discuss two motivating examples.

2.1. A Two-Dimensional Gedanken Experiment in Classical Particle Physics. Consider the following ideal situation. An accelerator physicist prepares an elliptical beam of classical particles with Gaussian transversal profile. The experimenter knows *a priori* the spatial distribution of the beam, that is, the covariance matrix Σ of the two-dimensional coordinates of the particles on a plane orthogonal to their flight direction. We can assume with no loss of generality that the transversal coordinate system has origin at the maximum of the beam intensity and axes along the principal components of the beam; thus it holds $\Sigma = \text{diag}(\lambda_1, \lambda_2)$. The beam travels straightforward until it enters a linear coaxial pipeline with circular profile, schematically depicted in Figure 1, where the beam is longitudinally accelerated. While the outer part of the beam is stopped by an absorbing wall, the inner part propagates within the pipeline. At the end of the beam flight the physicist wants to know if the transversal distribution of the particles is changed, due to unknown disturbance factors arisen within the pipeline. Accordingly, he measures again the spatial covariance matrix of the beam. Unfortunately, the absorbing wall has cut off the Gaussian tail, thus damping the covariance matrix and making it no more comparable to the original one. To perform such a comparison in the general case $\lambda_1 \neq \lambda_2$, the covariance matrix of the truncated circular beam has to go through the reconstruction procedure described in next sections.

2.2. A Multivariate Example: Connections to Compositional Data Analysis. Compositional Data Analysis (CoDA) has been the subject of a number of papers, pioneered by Aitchison [3] over the past forty years. As a methodology of statistical investigation, it finds application in all cases where the main object of interest is a multivariate with strictly positive continuous components to be regarded as portions of a total amount κ (the normalization $\kappa = 1$ is conventionally adopted in the mathematical literature). In other words, compositional variates belong to the κ-simplex:

$$\mathcal{S}_v = \left\{ z \in \mathbb{R}_+^v : |z|_1 = \kappa \right\}, \quad v \geq 2, \tag{4}$$

with $|z|_1 = \sum_{k=1}^v z_k$ the taxi-cab norm of z, while compositions with different norms can be always projected onto $\mathcal{S}_v$ via the closure operator $\mathscr{C} \cdot x \equiv \{\kappa x_1/|x|_1, \ldots, \kappa x_v/|x|_1\}$. There are countless types of compositional data, whose analysis raises problems of interest for statistics [4], for example, geochemical data, balance sheet data, and election data. The simplex constraint induces a kind of dependency among the parts of a composition that goes beyond the standard concept of covariance. This invalidates many ordinary techniques of statistical analysis.

In order to measure distances on $\mathcal{S}_v$, Aitchison introduced a positive symmetric function $d_A : \mathcal{S}_v \times \mathcal{S}_v \to \mathbb{R}_+$, explicitly defined by

$$d_A(x, y) = \sqrt{\frac{1}{2v} \sum_{i,k=1}^v \left[\log\left(\frac{x_i}{x_k}\right) - \log\left(\frac{y_i}{y_k}\right) \right]^2}. \tag{5}$$

The Aitchison distance is a key tool in CoDA. It is scale invariant in both its first and second arguments; that is, it
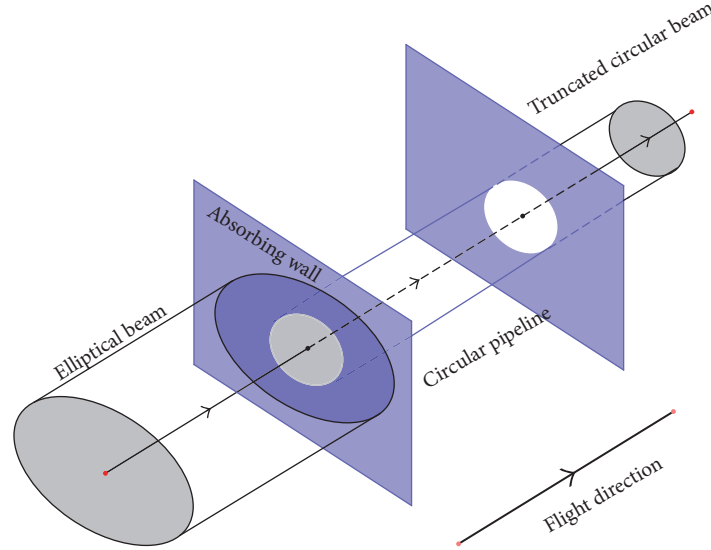
FIGURE 1: A classical particle beam with elliptical transversal profile is cut off upon entering a circular coaxial pipeline.

is left invariant by redefinitions $z \rightarrow \{\alpha z_1, \ldots, \alpha z_\nu\}$ with $\alpha \in \mathbb{R}_+$. Accordingly, its support can be extended to $\mathbb{R}_+^\nu \times \mathbb{R}_+^\nu$ by imposing

$$d_A(x, y) \equiv d_A(\mathscr{C} \cdot x, \mathscr{C} \cdot y), \quad x, y \in \mathbb{R}_+^\nu. \tag{6}$$

It was proved in [5] that d_A is a norm-induced metric on $\mathcal{S}_\nu$, provided the latter is given an appropriate normed vector space structure. Owing to the compositional constraint $|\cdot|_1 = \kappa$, it holds $\dim \mathcal{S}_\nu = \nu - 1$. Accordingly, the description of $\mathcal{S}_\nu$ in terms of ν components is redundant: an essential representation requires compositions to be properly mapped onto $(\nu-1)$-tuples. Among various possibilities, the Isometric Log-Ratio (ILR) transform introduced in [5] is the only known map of this kind leaving d_A invariant. More precisely, the ILR fulfills

$$d_A(x, y) = d_E(\mathrm{ilr}(x), \mathrm{ilr}(y)),$$
$$d_E(u, v) \equiv \sqrt{\sum_{k=1}^{\nu-1}(u_k - v_k)^2}. \tag{7}$$

It is known from [6] that if $X \sim \log \mathcal{N}_\nu(\mu, \Sigma)$ is a log-normal ν-variate, then $\mathscr{C} \cdot X \sim L_\nu(\mu', \Sigma')$ is a logistic-normal ν-variate (the reader should notice that in [6] the simplex is defined by $\mathcal{S}_\nu = \{z \in \mathbb{R}_+^\nu : |z|_1 < 1\}$; thus property 2.2 of [6] is here reformulated so as to take into account such difference), with a known relation between (μ, Σ) and (μ', Σ'). Analogously, it is not difficult to show that $\mathrm{ilr}(\mathscr{C} \cdot X) \sim \mathcal{N}_{\nu-1}(\mu'', \Sigma'')$ is a normal $(\nu-1)$-variate, with (μ'', Σ'') related to (μ', Σ') via the change of basis matrices derived in [5]. Just to sum up, it holds

$$X \sim \log \mathcal{N}_\nu(\mu, \Sigma) \Longrightarrow$$
$$\mathscr{C} \cdot X \sim L_\nu(\mu', \Sigma') \Longleftrightarrow \tag{8}$$
$$\mathrm{ilr}(\mathscr{C} \cdot X) \sim \mathcal{N}_{\nu-1}(\mu'', \Sigma'').$$

Now, suppose that (i) X fulfills (8) and has a natural interpretation as a composition, (ii) a representative set $\mathscr{D}_X$ of observations of X is given, and (iii) we wish to select from $\mathscr{D}_X$ those units which are compositionally closer to the center of the distribution, according to the Aitchison distance. To see that the problem is well posed, we first turn $\mathscr{D}_X$ into a set $\mathscr{D}_{\mathscr{C} \cdot X} \equiv \{y : y = \mathscr{C} \cdot x, x \in \mathscr{D}_X\}$ of compositional observations of $Y = \mathscr{C} \cdot X$. Then, we consider the special point $\mathrm{cen}[Y] = \mathscr{C} \cdot \exp\{\mathbb{E}[\ln Y]\}$, representing the center of the distribution of Y in a compositional sense: $\mathrm{cen}[Y]$ minimizes the expression $\mathbb{E}[d_A^2(Y, \mathrm{cen}[Y])]$ over $\mathcal{S}_\nu$ and fulfills $\mathrm{cen}[Y] = \mathrm{ilr}^{-1}(\mathbb{E}[\mathrm{ilr}(Y)])$, see [7]. By virtue of (8) this entails $\mathrm{ilr}(\mathrm{cen}[Y]) = \mathbb{E}[\mathrm{ilr}(Y)] = \mu''$. In order to select the observations which are closer to $\mathrm{cen}[Y]$, we set a threshold $\delta > 0$ and consider only those elements $y \in \mathscr{D}_{\mathscr{C} \cdot X}$ fulfilling $d_A^2(y, \mathrm{cen}[y]) < \delta$, with $\mathrm{cen}[y]$ being a sample estimator of $\mathrm{cen}[Y]$ on $\mathscr{D}_{\mathscr{C} \cdot X}$. Such selection rule operates a well-defined truncation of the distribution of Y. Moreover, in view of (7) and (8), we have

$$\mathbb{P}\left[d_A^2(Y, \mathrm{cen}[Y]) < \delta \mid Y \sim L_\nu(\mu', \Sigma')\right]$$
$$= \mathbb{P}\left[d_E^2(Z, \mu'') < \delta \mid Z \sim \mathcal{N}_{\nu-1}(\mu'', \Sigma'')\right], \tag{9}$$

with $Z = \mathrm{ilr}(\mathscr{C} \cdot X)$. As a consequence, we see that a compositional selection rule based on the Aitchison distance and (8) is equivalent to a spherical truncation of a multinormal distribution. Obviously, once Z has been spherically truncated, the covariance matrix of the remaining data is damped; thus an estimate of the full covariance matrix requires the reconstruction procedure described in next sections.

2.3. Compositional Outlier Detection via Forward Search Techniques.
Outlier detection in CoDA is a practical problem where the considerations made in Section 2.2 find concrete application. Most of the established methods for outlier

detection make use of the Mahalanobis metric. This is however unfit to describe the distance between compositions. Log-ratio transforms allow to get rid of the compositional constraint and make it possible to apply standard statistical methods [8]. Here we consider an innovative approach, namely, the Forward Search Algorithm (FSA), introduced in [9] and thoroughly discussed in [10]. The FSA admits an elegant extension to CoDA, of which the covariance reconstruction is a key step. In the next few lines we sketch the original algorithm and outline its extension to CoDA.

2.3.1. Construction of the Signal.
In its standard formulation the FSA applies to normal data. It assumes a dataset $\mathscr{D}_X = \{x^{(k)}\}_{k=1}^N$ with $x^{(k)} \in \mathbb{R}_+^{(v)}$ for $k = 1, \ldots, N$. The null hypothesis is that all the elements of $\mathscr{D}_X$ are independent realizations of a stochastic variable $X \sim \mathcal{N}(\mu_0, \Sigma_0)$. The FSA consists of a sequence of steps where data are recursively sorted. Along the recursion a signal is built and tested.

As a preliminary step, m_0 observations are randomly chosen from the bulk of $\mathscr{D}_X$. Let $S(m_0)$ be the set of these observations. $\mathscr{S}(m_0)$ is used to provide initial estimates $\mu(m_0)$, $\Sigma(m_0)$ of the true mean μ_0 and the true covariance matrix Σ_0, respectively. For $m = m_0 + 1, m_0 + 2, \ldots$, the $(m - m_0)$th step of the algorithm goes as follows:

(i) sort the elements of $\mathscr{D}_X$ according to the increasing values of the square Mahalanobis distance:

$$
\begin{aligned}
d_m(x)^2 \\
= \left[x - \mu(m-1)\right]^{\mathrm{T}} \Sigma(m-1)^{-1} \left[x - \mu(m-1)\right];
\end{aligned}
\tag{10}
$$

(ii) take the mth element $x^{(m)}$ of the newly sorted dataset and regard $s_m = d_m(x^{(m)})^2$ as the $(m - m_0)$th component of the signal;

(iii) let $\mathscr{S}(m)$ be the set of the first m observations of the newly sorted dataset;

(iv) use $\mathscr{S}(m)$ to provide new estimates $\mu(m)$, $\Sigma(m)$ of the true mean and the true covariance matrix, respectively.

Notice that $\mathscr{S}(m)$ is a truncated dataset. Therefore, $\Sigma(m)$ must include the Tallis' correction factor, (2) with $\rho = s_m$. While the recursion proceeds, the inliers of $\mathscr{D}_X$ populate progressively $\mathscr{S}(m)$. The recursion stops at the $\overline{m}$th step, when the first outlier $x^{(\overline{m})}$ of $\mathscr{D}_X$ produces a statistically relevant discontinuity in the signal.

2.3.2. Statistical Test of the Signal.
Statistical tests are needed to assess the relevance of discontinuities observed in the signal. At each step of the algorithm a new test is actually performed. Specifically, at the mth step, s_m is computed together with the lowest and highest values, respectively, $\delta s_{m,\alpha}$ and $\delta s_{m,1-\alpha}$, which are theoretically admissible for s_m under the null hypothesis at $(1 - \alpha)$ significance level for some α. These values are nothing but the α- and $(1 - \alpha)$-percentage points of s_m. Their envelopes for $m - m_0 > 0$ generate two curves that surround the signal when plotted

versus m. More precisely, one curve lies below it and the other lies above, provided the null hypothesis is not broken. The violation of one of the curves by the signal is interpreted as the result of the entrance of an outlier into $\mathscr{S}(m)$. Although the distribution of s_m cannot be calculated in closed form, its percentage points are obtained from a general formula, first derived in [11], yielding

$$
\delta s_{m,\alpha} = \left(\chi_v^2\right)^{-1} \left(\frac{m}{m + (N - m - 1) f_{2(N-m-1),2m;1-\alpha}}\right), \tag{11}
$$

with $f_{a,b;\alpha}$ the α-percentage point of the Fisher distribution with parameters (a, b). Equation (11) holds with decent approximation, as confirmed by numerical simulations.

2.3.3. Extension of the Forward Search to CoDA.
When $\mathscr{D}_X$ is a compositional dataset, it is unnatural to assume that its elements are realizations of a multivariate $X \sim \mathcal{N}(\mu_0, \Sigma_0)$. In this case the use of the FSE as outlined above does not make sense at all. Sometimes it is reasonable to assume $X \sim L_v(\mu_0, \Sigma_0)$, as first argued in [6]. In this case we can use the FSA to find outliers, provided that we first modify the algorithm in two respects:

(i) we replace the Mahalanobis distance by the Aitchison one;

(ii) we perform statistical tests consistently with the change of null hypothesis.

Specifically, at the mth step of the algorithm, we sort $\mathscr{D}_X$ according to the increasing values of the square Aitchison distance:

$$
d_A(x)^2 = \frac{1}{2v} \sum_{i,k=1}^{v} \left[\log\left(\frac{x_i}{x_k}\right) - \log\left(\frac{(c_{m-1})_i}{(c_{m-1})_k}\right)\right]^2, \tag{12}
$$

where $c_{m-1} = \mathrm{cen}[y \mid y \in \mathscr{S}(m-1)]$ is the center of $\mathscr{S}(m-1)$. Analogously, given the mth element $x^{(m)}$ of the newly sorted dataset, we regard $s_m = d_A(x^{(m)})^2$ as the mth component of the signal. The percentage points of s_m are obtained from $\mathrm{ilr}(\mathscr{S}(m))$ by using the probability correspondence established in (9). Since $\mathrm{ilr}(\mathscr{S}(m))$ is a spherically truncated dataset, the estimate of the covariance matrix $\Sigma(m)$ derived from it must undergo the reconstruction procedure described in next sections.

2.4. General Covariance Reconstruction Problem.
The examples discussed in the previous subsections are special cases of a more general inverse problem, namely, the reconstruction of the covariance matrix Σ of a normal multivariate X on the basis of the covariance matrix $\mathfrak{S}_\mathscr{D}$ of X truncated to some (convex) region $\mathscr{D}$. This is the simplest yet nontrivial inverse problem, which can be naturally associated with the normal distribution. The case $\mathscr{D} = \mathscr{B}_v(\rho)$ corresponds to a setup where theoretical and practical aspects of the problem can be investigated with relatively modest mathematical effort. It is certainly a well-defined framework where to study regularization techniques for nonlinear inverse problems in statistics, for which there is still much room for interesting work [12, 13].

3. Definitions and Setup

Let $X \in \mathbb{R}^\nu$ be a random vector with jointly normal distribution $\mathcal{N}_\nu(0, \Sigma)$ in $\nu \geq 1$ dimensions. The probability that X falls within $\mathcal{B}_\nu(\rho)$ is measured by the Gaussian integral

$$\alpha(\rho; \Sigma) \equiv \mathbb{P}\left[X \in \mathcal{B}_\nu(\rho)\right]$$
$$= \frac{1}{(2\pi)^{\nu/2} |\Sigma|^{1/2}} \int_{\mathcal{B}_\nu(\rho)} d^\nu x \exp\left\{-\frac{1}{2} x^{\mathrm{T}} \Sigma^{-1} x\right\}. \quad (13)$$

Since Σ is symmetric positive definite, it has orthonormal eigenvectors $\Sigma v^{(i)} = \lambda_i v^{(i)}$. Let us denote by $R \equiv \{v_i^{(j)}\}_{i,j=1}^\nu$ the orthogonal matrix having these vectors as columns and by $\Lambda \equiv \mathrm{diag}(\lambda) = R^{\mathrm{T}} \Sigma R$ the diagonal counterpart of Σ. From the invariance of $\mathcal{B}_\nu(\rho)$ under rotations, it follows that α depends upon Σ just by way of λ. Accordingly, we rename the Gaussian probability content of $\mathcal{B}_\nu(\rho)$ as

$$\alpha(\rho; \lambda) \equiv \int_{\mathcal{B}_\nu(\rho)} d^\nu x \prod_{m=1}^\nu \delta(x_m, \lambda_m),$$
$$\delta(y, \eta) = \frac{1}{\sqrt{2\pi\eta}} \exp\left\{-\frac{y^2}{2\eta}\right\}. \quad (14)$$

Note that (14) is not sufficient to fully characterize the random vector X under the spherical constraint, for which we need to calculate the distribution law $\mathbb{P}[X \in A \mid X \in \mathcal{B}_\nu(\rho)]$ as a function of $A \subset \mathbb{R}^\nu$. Alternatively, we can describe X in terms of the complete set of its truncated moments

$$m_{k_1 \cdots k_\nu}(\rho; \Sigma) \equiv \mathbb{E}\left[X_1^{k_1} X \cdots \in X_\nu^{k_\nu} \mathcal{B}_\nu(\rho)\right],$$
$$\{k_i\}_{i=1,\dots,\nu} \in \mathbb{N}^\nu. \quad (15)$$

As usual, these can be all obtained from the moment-generating function

$$\alpha m(t)$$
$$= \frac{1}{(2\pi)^{\nu/2} |\Sigma|^{1/2}} \int_{\mathcal{B}_\nu(\rho)} d^\nu x \exp\left\{t^{\mathrm{T}} x - \frac{1}{2} x^{\mathrm{T}} \Sigma^{-1} x\right\}, \quad (16)$$
$$t \in \mathbb{R}^\nu,$$

by differentiating the latter an arbitrary number of times with respect to the components of t; namely,

$$m_{k_1 \cdots k_\nu}(\rho; \Sigma) = \left.\frac{\partial^{k_1 + \cdots + k_\nu} m(t)}{(\partial t_1)^{k_1} \cdots (\partial t_\nu)^{k_\nu}}\right|_{t=0}. \quad (17)$$

It will be observed that $m(t)$ is in general not invariant under rotations of t. Therefore, unlike α, the moments $m_{k_1 \cdots k_\nu}$ depend effectively on both λ and R. For instance, for the matrix of the second moments $\mathfrak{S}_\mathcal{B} \equiv \{\partial^2 m/\partial t_i \partial t_j|_{t=0}\}_{i,j=1}^\nu$ such dependence amounts to

$$\alpha(\mathfrak{S}_\mathcal{B})_{ij} = \sum_{k,\ell=1}^\nu R_{ki} R_{\ell j} \int_{\mathcal{B}_\nu(\rho)} d^\nu x \, x_k x_\ell \prod_{m=1}^\nu \delta(x_m, \lambda_m). \quad (18)$$

By parity, the only nonvanishing terms in the above sum are those with $k = \ell$. Hence, it follows that Σ and $\mathfrak{S}_\mathcal{B}$ share R as a common diagonalizing matrix. In other words, if $M \equiv \mathrm{diag}(\mu)$ is the diagonal matrix of the eigenvalues of $\mathfrak{S}_\mathcal{B}$, then $M = R^{\mathrm{T}} \mathfrak{S}_\mathcal{B} R$. Moreover, μ_k is related to λ_k by

$$\mu_k = \lambda_k \frac{\alpha_k}{\alpha},$$
$$\alpha_k(\rho; \lambda) \equiv \int_{\mathcal{B}_\nu(\rho)} d^\nu x \frac{x_k^2}{\lambda_k} \prod_{m=1}^\nu \delta(x_m, \lambda_m), \quad (19)$$
$$k = 1, \dots, \nu.$$

The ratios α_k/α are naturally interpreted as adimensional correction factors to the eigenvalues of Σ, so they play the same role as $c_{\mathrm{T}}(\rho)$ in (2). However, α_k/α depends explicitly on the subscript k; thus each eigenvalue is damped differently from the others as a consequence of the condition $X \in \mathcal{B}_\nu(\rho)$.

Remark 1. In practical terms, (18)-(19) tell us that estimating the sample covariance matrix of $X \sim \mathcal{N}_\nu(0, \Sigma)$ from a spherically truncated population $\{x^{(m)}\}_{m=1}^M$, made of M independent units, via the classical estimator $Q_{ij} = (M - 1)^{-1} \sum_{m=1}^M (x_i^{(m)} - \tilde{x}_i^{(m)})(x_j^{(m)} - \tilde{x}_j^{(m)})$, being $\tilde{x} = M^{-1} \sum_{m=1}^M x^{(m)}$ the sample mean, yields a damped result. Nonetheless, the damping affects only the eigenvalues of the estimator, whereas its eigenvectors are left invariant.

3.1. Monotonicity Properties of Ratios of Gaussian Integrals. Equations (14) and (19) suggest introducing a general notation for the Gaussian integrals over $\mathcal{B}_\nu(\rho)$, under the assumption $\Sigma = \Lambda$. So, we define

$$\alpha_{k\ell m\cdots}(\rho; \lambda) \equiv \int_{\mathcal{B}_\nu(\rho)} d^\nu x \frac{x_k^2}{\lambda_k} \frac{x_\ell^2}{\lambda_\ell} \frac{x_m^2}{\lambda_m} \cdots \prod_{n=1}^\nu \delta(x_n, \lambda_n), \quad (20)$$

with each subscript q on the *left-hand side* addressing an additional factor of x_q^2/λ_q under the integral sign on the *right-hand side* (no subscripts means α). Several analytic properties of such integrals are discussed in [14]. In the following proposition, we lay emphasis on some issues concerning the monotonicity trends of the ratios α_k/α.

Proposition 2 (monotonicities). *Let* $\lambda_{(k)} \equiv \{\lambda_i\}_{i=1,\dots,\nu}^{i\neq k}$ *denote the set of the full eigenvalues without* λ_k. *The ratios* α_k/α *fulfill the following properties:*

(p_1) $\lambda_k(\alpha_k/\alpha)(\rho; \lambda)$ *is a monotonic increasing function of* λ_k *at fixed* ρ *and* $\lambda_{(k)}$;

(p_2) $(\alpha_k/\alpha)(\rho; \lambda)$ *is a monotonic decreasing function of* λ_k *at fixed* ρ *and* $\lambda_{(k)}$;

(p_3) $(\alpha_k/\alpha)(\rho; \lambda)$ *is a monotonic decreasing function of* λ_i $(i \neq k)$ *at fixed* ρ *and* $\lambda_{(i)}$,

where an innocuous abuse of notation has been made on writing $(\alpha_k/\alpha)(\rho; \lambda)$ *in place of* $\alpha_k(\rho; \lambda)/\alpha(\rho; \lambda)$.

Proof. Let the symbol $\partial_k \equiv \partial/\partial \lambda_k$ denote a derivative with respect to λ_k. In order to prove property (p_1), we apply the

chain rule of differentiation to $\lambda_k \alpha_k / \alpha$ and then we pass ∂_k under the integral sign in $\partial_k \alpha$ and $\partial_k \alpha_k$. In this way, we obtain

$$\partial_k \left(\lambda_k \frac{\alpha_k}{\alpha} \right) = \frac{1}{2} \left(\frac{\alpha_{kk}}{\alpha} - \frac{\alpha_k^2}{\alpha^2} \right)$$

$$= \frac{1}{2\lambda_k^2} \left\{ \mathbb{E} \left[X_k^4 \mid X \in \mathscr{B}_v(\rho) \right] \right.$$

$$\left. - \mathbb{E} \left[X_k^2 \mid X \in \mathscr{B}_v(\rho) \right]^2 \right\} = \frac{1}{2\lambda_k^2} \tag{21}$$

$$\cdot \operatorname{var} \left(X_k^2 \mid X \in \mathscr{B}_v(\rho) \right) \geq 0.$$

Moreover, since the truncated marginal density of X_k^2 is positive within a set of nonzero measure in $\mathbb{R}$, the monotonic trend of $\lambda_k \alpha_k / \alpha$ in λ_k is strict.

Properties (p_2) and (p_3) are less trivial than (p_1). Indeed, the same reasoning as above now yields on the one hand

$$\lambda_k \partial_k \left(\frac{\alpha_k}{\alpha} \right) = \partial_k \left(\lambda_k \frac{\alpha_k}{\alpha} \right) - \frac{\alpha_k}{\alpha}$$

$$= \frac{1}{2\lambda_k^2} \left\{ \operatorname{var} \left(X_k^2 \mid X \in \mathscr{B}_v(\rho) \right) \right. \tag{22}$$

$$\left. - 2\lambda_k \mathbb{E} \left[X_k^2 \mid X \in \mathscr{B}_v(\rho) \right] \right\} \leq 0,$$

and on the other

$$\lambda_i \partial_i \left(\frac{\alpha_k}{\alpha} \right) = \frac{1}{2} \left(\frac{\alpha_{ik}}{\alpha} - \frac{\alpha_i \alpha_k}{\alpha^2} \right)$$

$$= \frac{1}{2\lambda_i \lambda_k} \operatorname{cov} \left(X_i^2, X_k^2 \mid X \in \mathscr{B}_v(\rho) \right) \leq 0 \tag{23}$$

$$(i \neq k).$$

Despite being not *a priori* evident, the *right-hand side* of both (22) and (23) is negative (and vanishes in the limit $\rho \to \infty$). The inequalities $\operatorname{var}(X_k^2) \leq 2\lambda_k \mathbb{E}[X_k^2]$ within Euclidean balls have been first discussed in [14], while the inequalities $\operatorname{cov}(X_j^2, X_k^2)$ within generalized Orlicz balls have been discussed in [15, 16] for the case where the probability distribution of X is flat instead of being normal. More recently, a complete proof of both inequalities has been given in [17]. Despite the technical difficulties in proving them, their meaning should be intuitively clear. The variance inequality quantifies the squeezing affecting X_k^2 as a consequence of the truncation (in the unconstrained case it would be $\operatorname{var}(X_k^2) = 2\lambda_k^2$). The covariance inequality follows from the opposition arising among the square components in proximity of the boundary of $\mathscr{B}_v(\rho)$. Indeed, if $X_j^2 \nearrow \rho$, then $X_k^2 \searrow 0 \; \forall k \neq j$ in order for X to stay within $\mathscr{B}_v(\rho)$. $\quad \square$

3.2. Definition Domain of the Reconstruction Problem. A consequence of Proposition 2 is represented by the following.

Corollary 3. *Given v, ρ, and λ, μ_k is bounded by*

$$\frac{\rho}{r(v, \rho/2\lambda_k)} \leq \mu_k \leq \frac{\rho}{3},$$

$$r(v, z) \equiv (2v + 1) \frac{M(v, v + 1/2, z)}{M(v, v + 3/2, z)}, \tag{24}$$

with M denoting the Kummer function; namely,

$$M(a, b, z) = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{(a)_n}{(b)_n} z^n,$$

$$(x)_n \equiv \frac{\Gamma(x + n)}{\Gamma(x)}. \tag{25}$$

Proof. The upper bound of (24) corresponds to the value of μ_k in the v-tuple limit $\lambda_k \to \infty$, $\lambda_{(k)} \to \{0, \dots, 0\}$. This is indeed the maximum possible value allowed for μ_k according to properties (p_1) and (p_3) of Proposition 2. In order to perform this limit, we observe that

$$\lim_{\eta \to 0^+} \delta(y, \eta) = \delta(y), \tag{26}$$

with the δ symbol on the *right-hand side* representing the Dirac delta function (the reader who is not familiar with the theory of distributions may refer, for instance, to [18] for an introduction). Accordingly,

$$\lim_{\lambda_k \to \infty} \lim_{\lambda_{(k)} \to \{0, \dots, 0\}} \mu_k = \frac{\int_{-\sqrt{\rho}}^{+\sqrt{\rho}} dx_k \, x_k^2}{\int_{-\sqrt{\rho}}^{+\sqrt{\rho}} dx_k} = \frac{\rho}{3}. \tag{27}$$

The lower bound corresponds instead to the value taken by μ_k as $\lambda_{(k)} \to \{\infty, \dots, \infty\}$ and λ_k is kept fixed. In this limit, all the Gaussian factors in the probability density function except the kth one flatten to one. Hence,

$$\lim_{\lambda_{(k)} \to \{\infty, \dots, \infty\}} \mu_k$$

$$= \lim_{\lambda_{(k)} \to \{\infty, \dots, \infty\}} \frac{\int_{-\sqrt{\rho}}^{+\sqrt{\rho}} dx_k x_k^2 \delta(x_k, \lambda_k) \alpha^{(v-1)}(\rho - x_k^2; \lambda_{(k)})}{\int_{-\sqrt{\rho}}^{+\sqrt{\rho}} dx_k \delta(x_k, \lambda_k) \alpha^{(v-1)}(\rho - x_k^2; \lambda_{(k)})}$$

$$= \frac{\int_{-\sqrt{\rho}}^{+\sqrt{\rho}} dx_k x_k^2 e^{-x_k^2/2\lambda_k} (\rho - x_k^2)^{v-1}}{\int_{-\sqrt{\rho}}^{+\sqrt{\rho}} dx_k e^{-x_k^2/2\lambda_k} (\rho - x_k^2)^{v-1}} \tag{28}$$

$$= \rho \frac{\int_0^1 dx_k x_k^2 e^{-(\rho/2\lambda_k)x_k^2} (1 - x_k^2)^{v-1}}{\int_0^1 dx_k e^{-(\rho/2\lambda_k)x_k^2} (1 - x_k^2)^{v-1}}.$$

Numerator and denominator of the rightmost ratio are easily recognized to be integral representations of Kummer functions (see, e.g., [19, ch. 13]). $\quad \square$

The upper bound of (24) can be sharpened, as clarified by the following.

Proposition 4 (bounds on the truncated moments). *Let ν, ρ, and λ be given. If $\{i_1, \ldots, i_\nu\}$ is a permutation of $\{1, \ldots, \nu\}$ such that $\mu_{i_1} \le \mu_{i_2} \le \cdots \le \mu_{i_\nu}$, then the following upper bounds hold:*

(i) $\displaystyle\sum_{k=1}^{\nu} \mu_k \le \rho$;

(ii) $\displaystyle \mu_{i_k} \le \frac{\rho}{\nu - k + 1}$,

$$\qquad\qquad (29)$$

$$k = 1, \ldots, \nu.$$

Proof. The overall upper bound on the sum of truncated moments follows from

$$\sum_{k=1}^{\nu} \mu_k = \frac{1}{\alpha} \sum_{k=1}^{\nu} \lambda_k \alpha_k$$

$$= \frac{1}{\alpha} \int_{\mathcal{B}_\nu(\rho)} \mathrm{d}^\nu x \left(\sum_{k=1}^{\nu} x_k^2 \right) \prod_{m=1}^{\nu} \delta\left(x_m, \lambda_m\right) \le \rho. \qquad (30)$$

At the same time, the sum can be split and bounded from below by

$$\sum_{k=1}^{\nu} \mu_{i_k} = \sum_{k=1}^{n} \mu_{i_k} + \sum_{k=n+1}^{\nu} \mu_{i_k} \ge \sum_{k=1}^{n} \mu_{i_k} + (\nu - n)\, \mu_{i_{n+1}}, \qquad (31)$$

$$n = 0, 1, \ldots, \nu - 1.$$

The single upper bounds on the μ_k's are then obtained from (30)-(31). It will be noted that (29) (ii) is sharper than the upper bound of (24) only for $\nu > 3$ and $k < \nu - 2$. $\qquad\square$

From now on, we shall assume, with no loss of generality, that the eigenvalues of Σ are increasingly ordered, namely, $0 < \lambda_1 \le \cdots \le \lambda_\nu$ (we can always permute the labels of the coordinate axes, so as to let this be the case). An important aspect related to the eigenvalue ordering is provided by the following.

Proposition 5 (eigenvalue ordering). *Let ν, ρ, and λ be given. If $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_\nu$, then $\mu_1 \le \mu_2 \le \cdots \le \mu_\nu$ holds as well.*

Proof. In order to show that the spherical truncation does not violate the eigenvalue ordering, we make repeated use of the monotonicity properties of Proposition 2. Specifically, if $i < j$, then

$$\mu_i = \lambda_i \frac{\alpha_i}{\alpha}\left(\rho; \{\lambda_1, \ldots, \lambda_i, \ldots, \lambda_j, \ldots, \lambda_\nu\}\right)$$

$$\le \lambda_j \frac{\alpha_i}{\alpha}\left(\rho; \{\lambda_1, \ldots, \lambda_j, \ldots, \lambda_j, \ldots, \lambda_\nu\}\right)$$

$$\rightsquigarrow \text{ increasing monotonicity of } \lambda_i \frac{\alpha_i}{\alpha}$$

$$= \lambda_j \frac{\alpha_j}{\alpha}\left(\rho; \{\lambda_1, \ldots, \lambda_j, \ldots, \lambda_j, \ldots, \lambda_\nu\}\right)$$

$$\rightsquigarrow \text{ exchange symmetry } i \longleftrightarrow j$$

$$\le \lambda_j \frac{\alpha_j}{\alpha}\left(\rho; \{\lambda_1, \ldots, \lambda_i, \ldots, \lambda_j, \ldots, \lambda_\nu\}\right)$$

$$\rightsquigarrow \text{ decreasing monotonicity of } \frac{\alpha_j}{\alpha}$$

$$= \mu_j,$$

$$\qquad\qquad (32)$$

where the symbol "$\rightsquigarrow$" is used to explain where the inequality sign preceding it comes from and the "exchange symmetry" refers to the formal property of the one-index Gaussian integrals over $\mathcal{B}_\nu(\rho)$ to fulfill $\alpha_i(\rho; \{\lambda_1, \ldots, \lambda_i, \ldots, \lambda_j, \ldots, \lambda_\nu\}) = \alpha_j(\rho; \{\lambda_1, \ldots, \lambda_j, \ldots, \lambda_i, \ldots, \lambda_\nu\})$. $\qquad\square$

Let us now focus on (19). They have to be solved in order to reconstruct λ from μ. Formally, if we introduce a family of truncation operators $\tau_\rho : \mathbb{R}_+^\nu \rightarrow \mathbb{R}_+^\nu$ (parametrically depending on ρ), such that

$$\left(\tau_\rho \cdot \lambda\right)_k \equiv \lambda_k \frac{\alpha_k}{\alpha}(\rho; \lambda), \quad k = 1, \ldots, \nu, \qquad (33)$$

then the reconstruction of λ from μ amounts to calculating $\lambda = \tau_\rho^{-1} \cdot \mu$. One should be aware that τ_ρ is not a surjective operator in view of Corollary 3 and Proposition 4. Therefore, τ_ρ^{-1} is only defined within a bounded domain $\mathscr{D}(\tau_\rho^{-1})$. If we define

$$\mathscr{D}_0 = \left\{ \mu \in \mathbb{R}_+^\nu : \mu_1 \le \cdots \le \mu_\nu, \ \mu_k = \lambda_k \frac{\alpha_k}{\alpha} \text{ for } k \right.$$

$$\qquad\qquad (34)$$

$$\left. = 1, \ldots, \nu, \text{ for some } \lambda \in \mathbb{R}_+^\nu \right\},$$

then we have $\mathscr{D}(\tau_\rho^{-1}) = \{\mu : \mu = \sigma \cdot \mu_0 \text{ for some } \mu_0 \in \mathscr{D}_0, \ \sigma \in S_\nu\}$, where S_ν is the set of permutations of ν elements. From Proposition 4 we conclude that $\mathscr{D}_0 \subseteq H_\nu(\rho)$, being

$$H_\nu(\rho) \equiv \left\{ x \in \mathbb{R}_+^\nu : x_k \le \min\left\{ \frac{\rho}{3}, \frac{\rho}{\nu - k + 1} \right\}, \sum_{k=1}^{\nu} x_k \right.$$

$$\qquad\qquad (35)$$

$$\left. \le \rho, \ \forall k \right\}.$$

In fact, there are vectors $\mu \in \mathbb{R}_+^\nu$ fulfilling $\mu \in H_\nu(\rho)$ and $\mu \notin \mathscr{D}_0$; thus we conclude that $\mathscr{D}_0$ is a proper subset of $H_\nu(\rho)$. Numerical experiences based on the techniques discussed in the next sections show indeed that

$$\mathscr{D}\left(\tau_\rho^{-1}\right) = \bigcap_{k=1}^{\nu} \left\{ \mu \in \mathbb{R}_+^\nu : \sum_{\substack{j=1 \\ j \ne k}}^{1 \cdots \nu} \mu_j + 3\mu_k \le \rho \right\}. \qquad (36)$$
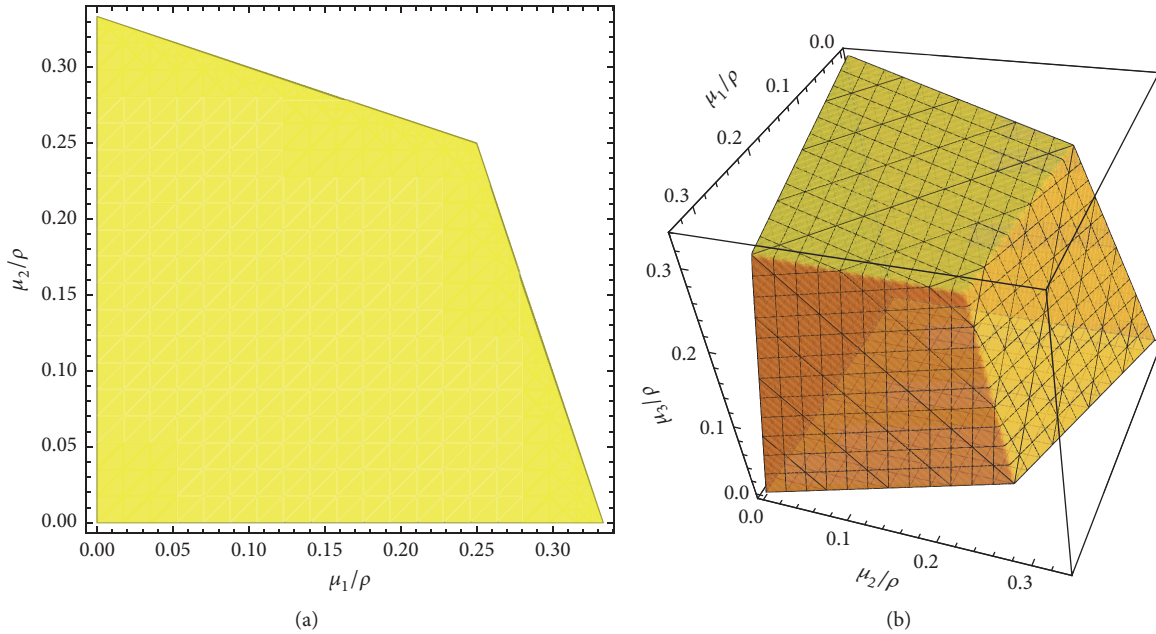
FIGURE 2: (a) Numerical reconstruction of $\mathscr{D}(\tau_\rho^{-1})$ in $\nu = 2$ dimensions. (b) Numerical reconstruction of $\mathscr{D}(\tau_\rho^{-1})$ in $\nu = 3$ dimensions.

A graphical representation of (36) in $\nu = 2$ and $\nu = 3$ dimensions is depicted in Figure 2. The reader should note that until Section 7 we shall always assume that μ comes from the application of τ_ρ to some λ; thus $\mu \in \mathscr{D}(\tau_\rho^{-1})$ by construction.

Now, we observe that (19) can be written in the equivalent form

$$\lambda = T(\lambda; \mu; \rho), \tag{37}$$

$$T : \mathbb{R}_+^\nu \times \mathbb{R}_+^\nu \times \mathbb{R}_+ \longrightarrow \mathbb{R}_+^\nu;$$
$$T_k(\lambda; \mu; \rho) = \mu_k \frac{\alpha}{\alpha_k}(\rho; \lambda), \quad k = 1, \ldots, \nu. \tag{38}$$

Since ρ and μ are (nonindependent) input parameters for the covariance reconstruction problem (and in order to keep the notation light), in the sequel we shall leave the dependence of T upon ρ and μ implicitly understood; that is, we shall write (37) as $\lambda = T(\lambda)$. Hence, we see that the full eigenvalue spectrum λ is a fixed point of the operator T. This suggests obtaining it as the limit of a sequence

$$\lambda^{(0)} = \mu,$$

$$\lambda^{(n+1)} = T\left(\lambda^{(n)}\right), \tag{39}$$

$$n = 0, 1, \ldots,$$

$$\lambda = \lim_{n \to \infty} \lambda^{(n)}, \tag{40}$$

provided that this can be shown to converge. Note that since $\alpha_k < \alpha$, it follows that $T_k(\lambda^{(n)}) > \mu_k \ \forall n$, so the sequence is bounded from below by μ. In particular, this holds for $n = 0$.

Therefore, the sequence moves to the right direction at least at the beginning. A formal proof of convergence, based on the monotonicity properties stated by Proposition 2, is given in the next section.

4. Convergence of the Fixed Point Equation

We split our argument into three propositions, describing different properties of the sequence $\lambda^{(n)}$. They assert, respectively, that (i) the sequence is componentwise monotically increasing; (ii) the sequence is componentwise bounded from above by any fixed point of T; and (iii) if T has a fixed point, this must be unique. Statements (i) and (ii) are sufficient to guarantee the convergence of the sequence to a finite limit (the unconstrained spectrum is a fixed point of T). In addition, the limit is easily recognized to be a fixed point of T. Hence, statement (iii) guarantees that the sequence converges to the unconstrained eigenvalue spectrum. We remark that all the monotonicities discussed in Proposition 2 are strict; that is, the ratios α_k/α have no stationary points at finite ρ and λ, which is crucial for the proof.

Proposition 6 (increasing monotonicity). *Given ν, ρ, and $\mu \in \mathscr{D}(\tau_\rho^{-1})$, the sequence $\lambda^{(0)} = \mu$, $\lambda^{(n+1)} = T(\lambda^{(n)})$, $n = 0, 1, \ldots$, is monotically increasing; namely, $\lambda_k^{(n+1)} > \lambda_k^{(n)} \ \forall k = 1, \ldots, \nu$.*

Proof. The proof is by induction. We first notice that

$$\lambda_k^{(1)} = T_k\left(\lambda^{(0)}\right) = T_k(\mu) = \mu_k \frac{\alpha}{\alpha_k}(\rho; \mu) > \mu_k = \lambda_k^{(0)},$$
$$\tag{41}$$
$$k = 1, \ldots, \nu;$$

the inequality follows from $\alpha_k(\rho;\mu) < \alpha(\rho;\mu)$. Suppose now that the property of increasing monotonicity has been checked off up to the nth element of the sequence. Then,

$$\lambda_k^{(n+1)} = \mu_k \frac{\alpha}{\alpha_k}\left(\rho;\lambda^{(n)}\right) > \mu_k \frac{\alpha}{\alpha_k}\left(\rho;\lambda^{(n-1)}\right) = \lambda_k^{(n)}; \quad (42)$$

the inequality follows this time from the inductive hypothesis and from properties (p_2) and (p_3) of Proposition 2. $\qquad\square$

Proposition 7 (boundedness). *Given v, ρ, and $\mu \in \mathcal{D}(\tau_\rho^{-1})$, the sequence $\lambda^{(0)} = \mu$, $\lambda^{(n+1)} = T(\lambda^{(n)})$, $n = 0, 1, \ldots$, is bounded from above; namely, $\lambda_k^{(n)} < \lambda_k^* \ \forall k = 1, \ldots, v$, λ^* being a fixed point of T.*

Proof. We proceed again by induction. We first notice that

$$\lambda_k^{(0)} = \mu_k < \mu_k \frac{\alpha}{\alpha_k}\left(\rho;\lambda^*\right) = \lambda_k^*, \quad k = 1, \ldots, v; \quad (43)$$

the inequality follows as previously from $\alpha_k(\rho;\lambda^*) < \alpha(\rho;\lambda^*)$. Suppose now that the property of boundedness has been checked off up to the nth element of the sequence. Then,

$$\lambda_k^{(n+1)} = \mu_k \frac{\alpha}{\alpha_k}\left(\rho;\lambda^{(n)}\right) = \lambda_k^* \frac{\alpha_k}{\alpha}\left(\rho;\lambda^*\right)\frac{\alpha}{\alpha_k}\left(\rho;\lambda^{(n)}\right)$$

$$< \lambda_k^* \frac{\alpha_k}{\alpha}\left(\rho;\lambda^{(n)}\right)\frac{\alpha}{\alpha_k}\left(\rho;\lambda^{(n)}\right) = \lambda_k^*; \quad (44)$$

the inequality follows for the last time from the inductive hypothesis and from properties (p_2) and (p_3) of Proposition 2. $\qquad\square$

According to Propositions 6 and 7, the sequence converges. Now, let $\tilde{\lambda} = \lim_{n\to\infty}\lambda^{(n)}$ be the limit of the sequence. Effortlessly, we prove that $\tilde{\lambda}$ is a fixed point of T. Indeed,

$$\tilde{\lambda}_k = \lim_{n\to\infty}\lambda_k^{(n)} = \lim_{n\to\infty}\mu_k\frac{\alpha}{\alpha_k}\left(\rho;\lambda^{(n-1)}\right)$$

$$= \mu_k\frac{\alpha}{\alpha_k}\left(\rho;\lim_{n\to\infty}\lambda^{(n-1)}\right) = T_k\left(\tilde{\lambda}\right). \quad (45)$$

Note that passing the limit over n under the integral sign is certainly allowed for Gaussian integrals.

Proposition 8 (uniqueness of the fixed point). *Let $\lambda' = T(\lambda')$ and $\lambda'' = T(\lambda'')$ be two fixed points of T, corresponding to the same choice of v, ρ, and $\mu \in \mathcal{D}(\tau_\rho^{-1})$. Then, it must be that $\lambda' = \lambda''$.*

Proof. According to the hypothesis, λ' and λ'' fulfill the equations

$$\lambda_k' = \mu_k\frac{\alpha}{\alpha_k}\left(\rho;\lambda'\right) \Longrightarrow$$

$$\mu_k = \lambda_k'\frac{\alpha_k}{\alpha}\left(\rho;\lambda'\right),$$

$$\lambda_k'' = \mu_k\frac{\alpha}{\alpha_k}\left(\rho;\lambda''\right) \Longrightarrow \quad (46)$$

$$\mu_k = \lambda_k''\frac{\alpha_k}{\alpha}\left(\rho;\lambda''\right).$$

Hence,

$$0 = \mu_k - \mu_k = \lambda_k'\frac{\alpha_k}{\alpha}\left(\rho;\lambda'\right) - \lambda_k''\frac{\alpha_k}{\alpha}\left(\rho;\lambda''\right)$$

$$= \sum_{\ell=1}^{v}\left[\int_0^1 dt\, J_{k\ell}\left(\rho;\lambda'' + t\left(\lambda' - \lambda''\right)\right)\right]\left(\lambda_\ell' - \lambda_\ell''\right), \quad (47)$$

where J denotes the Jacobian matrix of τ_ρ and is given by

$$J_{k\ell}\left(\rho;\lambda\right) = \partial_k\left(\lambda_\ell\frac{\alpha_\ell}{\alpha}\left(\rho;\lambda\right)\right) = \frac{1}{2}\frac{\lambda_\ell}{\lambda_k}\left(\frac{\alpha_{k\ell}}{\alpha} - \frac{\alpha_k\alpha_\ell}{\alpha^2}\right)$$

$$= \left[\Lambda^{-1}\Omega\left(\rho;\lambda\right)\Lambda\right]_{k\ell}, \quad (48)$$

having set $\Omega_{k\ell} \equiv (1/2)(\alpha_{k\ell}/\alpha - \alpha_k\alpha_\ell/\alpha^2)$. It will be noted that $\Omega = \{\Omega_{k\ell}\}_{k,\ell=1}^{v}$ is essentially the covariance matrix of the square components of X under spherical truncation (we have come across its matrix elements in (21)–(23)). As such, Ω is symmetric positive definite. Indeed,

$$\Omega_{k\ell} = \frac{1}{2\lambda_k\lambda_\ell}\,\text{cov}\left(X_k^2, X_\ell^2 \mid X \in \mathcal{B}_v\left(\rho\right)\right) = \frac{1}{2\lambda_k\lambda_\ell}$$

$$\cdot\,\mathbb{E}\left[\left(X_k^2 - \mathbb{E}\left[X_k^2 \mid X \in \mathcal{B}_v\left(\rho\right)\right]\right)\right. \quad (49)$$

$$\left.\cdot\left(X_\ell^2 - \mathbb{E}\left[X_\ell^2 \mid X \in \mathcal{B}_v\left(\rho\right)\right]\right) \mid X \in \mathcal{B}_v\left(\rho\right)\right].$$

On setting $Z_k = (X_k^2 - \mathbb{E}[X_k^2 \mid X \in \mathcal{B}_v(\rho)])/\sqrt{2}\lambda_k$, we can represent Ω as $\Omega = \mathbb{E}[ZZ^{\mathrm{T}} \mid X \in \mathcal{B}_v(\rho)]$. If $x \in \mathbb{R}^v$ is not the null vector, then $x^{\mathrm{T}}\Omega x = \mathbb{E}[x^{\mathrm{T}}ZZ^{\mathrm{T}}x \mid X \in \mathcal{B}_v(\rho)] = \mathbb{E}[(x^{\mathrm{T}}Z)^2 \mid X \in \mathcal{B}_v(\rho)] > 0$. Moreover, the eigenvalues of Ω fulfill the secular equation

$$0 = \det\left(\Omega - \phi\mathbb{I}_v\right) = \det\left[\Lambda^{-1}\left(\Omega - \phi\mathbb{I}_v\right)\Lambda\right]$$

$$= \det\left(\Lambda^{-1}\Omega\Lambda - \phi\mathbb{I}_v\right) = \det\left(J - \phi\mathbb{I}_v\right), \quad (50)$$

whence it follows that J is positive definite as well (though it is not symmetric). Since the sum of positive definite matrices is positive definite, we conclude that $\int_0^1 dt\, J(\rho;\lambda'' + t(\lambda'-\lambda''))$ is positive definite too. As such, it is nonsingular. Therefore, from (47), we conclude that $\lambda' = \lambda''$. $\qquad\square$

5. Numerical Computation of Gaussian Integrals over $\mathcal{B}_v(\rho)$

Let us now see how to compute $\alpha_{k\ell m\cdots}$ with controlled precision. Most of the relevant work has been originally done by Ruben in [2], where the case of α is discussed. We extend Ruben's technique to Gaussian integrals containing powers of the integration variable. Specifically, it is shown in [2] that $\alpha(\rho;\lambda)$ can be represented as a series of chi-square cumulative distribution functions:

$$\alpha\left(\rho;\lambda\right) = \sum_{m=0}^{\infty}c_m\left(s;\lambda\right)F_{v+2m}\left(\frac{\rho}{s}\right). \quad (51)$$

The scale factor s has the same physical dimension as ρ and λ. It is introduced in order to factorize the dependence of α upon ρ and λ at each order of the expansion. The series on the *right-hand side* of (51) converges uniformly on every finite interval of ρ. The coefficients c_m are given by

$$c_m(s;\lambda) = \frac{1}{m!} \frac{s^{v/2+m}}{|\Lambda|^{1/2}} \frac{\Gamma(v/2+m)}{\Gamma(v/2)} \mathbb{M}\left[(-Q)^m\right],$$

(52)

$$m = 0, 1, \ldots,$$

having defined $Q(x) \equiv x^{\mathrm{T}}[\Lambda^{-1} - s^{-1}\mathbb{1}_v]x$ for $x \in \mathbb{R}^v$ and $\mathbb{M}$ as the uniform average operator on the $(v-1)$-sphere $\partial\mathcal{B}_v(1) \equiv \{u \in \mathbb{R}^v : u^{\mathrm{T}}u = 1\}$; namely,

$$\mathbb{M}[\phi] \equiv \frac{\Gamma(v/2)}{2\pi^{v/2}} \int_{\partial\mathcal{B}_v(1)} \mathrm{d}u\, \phi(u),$$

(53)

$$\forall \phi \in \mathscr{C}^0\left(\partial\mathcal{B}_v(1)\right) \text{ a.e.}$$

Unfortunately, (52) is not particularly convenient for numerical computations, since $\mathbb{M}[(-Q)^m]$ is only given in integral form. However, it is also shown in [2] that the coefficients c_m can be extracted from the Taylor expansion (at $z_0 = 0$) of the generating function

$$\psi(z) = \prod_{k=1}^{v}\left(\frac{s}{\lambda_k}\right)^{1/2}\left[1 - \left(1 - \frac{s}{\lambda_k}\right)z\right]^{-1/2},$$

(54)

that is $\psi(z) = \sum_{m=0}^{\infty} c_m(s;\lambda)z^m$.

This series converges uniformly for $|z| < \min_i|1 - s/\lambda_i|^{-1}$. On evaluating the derivatives of $\psi(z)$, it is then shown that c_m's fulfill the recursion:

$$c_0 = \prod_{m=1}^{v}\sqrt{\frac{s}{\lambda_m}};$$

$$c_n = \frac{1}{2n}\sum_{r=0}^{n-1} g_{n-r}c_r,$$

(55)

$$n = 1, 2, \ldots;$$

$$g_n \equiv \sum_{m=1}^{v}\left(1 - \frac{s}{\lambda_m}\right)^n.$$

Finally, the systematic error produced on considering only the lowest k terms of the chi-square series of (51) is estimated by

$$\mathcal{R}_n(\rho;\lambda) \equiv \left|\sum_{m=n}^{\infty} c_m(s;\lambda)F_{v+2m}\left(\frac{\rho}{s}\right)\right| \le c_0(s;\lambda)$$

$$\cdot \frac{\Gamma\left(\frac{v}{2}+n\right)}{\Gamma\left(\frac{v}{2}\right)}\frac{\eta^n}{n!}(1-\eta)^{-\left(\frac{v}{2}+n\right)}F_{v+2n}\left[\frac{(1-\eta)\rho}{s}\right]$$

(56)

$$\equiv \mathfrak{R}_n,$$

with $\eta = \max_i|1 - s/\lambda_i|$.

Now, as mentioned, it is possible to extend the above expansion to all Gaussian integrals $\alpha_{k\ell m\cdots}$. Here, we are interested only in α_k and α_{jk}, since these are needed in order to implement the fixed point iteration and to compute the Jacobian matrix of τ_ρ. The extension is provided by the following.

Theorem 9 (Ruben's expansions). *The integrals α_k and α_{jk} admit the series representations:*

$$\alpha_k(\rho;\lambda) = \sum_{m=0}^{\infty} c_{k;m}(s;\lambda)F_{v+2(m+1)}\left(\frac{\rho}{s}\right),$$

(57)

$$\alpha_{jk}(\rho;\lambda) = \sum_{m=0}^{\infty} c_{jk;m}(s;\lambda)F_{v+2(m+2)}\left(\frac{\rho}{s}\right),$$

(58)

with s being an arbitrary positive constant. The series coefficients are given, respectively, by

$$c_{k;m}(s;\lambda) = \frac{s}{\lambda_k}\frac{v+2m}{m!}\frac{s^{v/2+m}}{|\Lambda|^{1/2}}\frac{\Gamma(v/2+m)}{\Gamma(v/2)}$$

$$\cdot \mathbb{M}\left[(-Q)^m u_k^2\right],$$

(59)

$$c_{jk;m}(s;\lambda) = (1 + 2\delta_{jk})\frac{s}{\lambda_j}\frac{s}{\lambda_k}\frac{(v+2m+2)(v+2m)}{m!}$$

$$\cdot \frac{s^{v/2+m}}{|\Lambda|^{1/2}}\frac{\Gamma(v/2+m)}{\Gamma(v/2)}\mathbb{M}\left[(-Q)^m u_j^2 u_k^2\right],$$

(60)

with δ_{jk} denoting the Kronecker symbol. The series on the right-hand side of (57)-(58) converge uniformly on every finite interval of ρ. The functions

$$\psi_k(z) = \left(\frac{s}{\lambda_k}\right)^{3/2}\left[1 - \left(1 - \frac{s}{\lambda_k}\right)z\right]^{-3/2}\prod_{i\neq k}\left(\frac{s}{\lambda_i}\right)^{1/2}$$

(61)

$$\cdot \left[1 - \left(1 - \frac{s}{\lambda_i}\right)z\right]^{-1/2},$$

$$\psi_{kk}(z) = 3\left(\frac{s}{\lambda_k}\right)^{5/2}\left[1-\left(1-\frac{s}{\lambda_k}\right)z\right]^{-5/2}$$

$$\cdot\prod_{i\neq k}\left(\frac{s}{\lambda_i}\right)^{1/2}\left[1-\left(1-\frac{s}{\lambda_i}\right)z\right]^{-1/2}, \tag{62}$$

$$\psi_{jk}(z) = \left(\frac{s}{\lambda_j}\frac{s}{\lambda_k}\right)^{3/2}$$

$$\cdot\left\{\left[1-\left(1-\frac{s}{\lambda_j}\right)z\right]\left[1-\left(1-\frac{s}{\lambda_k}\right)z\right]\right\}^{-3/2} \tag{63}$$

$$\times\prod_{i\neq j,k}\left(\frac{s}{\lambda_i}\right)^{1/2}\left[1-\left(1-\frac{s}{\lambda_i}\right)z\right]^{-1/2}\quad(j\neq k)$$

are generating functions, respectively, for the coefficients $c_{k;m}$, $c_{kk;m}$ and $c_{jk;m}$ $(j\neq k)$; that is, they fulfill

$$\psi_k(z) = \sum_{m=0}^{\infty} c_{k;m}(s;\lambda)\,z^m,$$

$$\psi_{jk}(z) = \sum_{m=0}^{\infty} c_{jk;m}(s;\lambda)\,z^m, \tag{64}$$

for $|z| < \min_i|1-s/\lambda_i|^{-1}$. Finally, the coefficients $c_{k;m}$, $c_{kk;m}$, and $c_{jk;m}$ $(j\neq k)$ can be obtained iteratively from the recursions

$$c_{k;0} = \left(\frac{s}{\lambda_k}\right)c_0;$$

$$c_{k;m} = \frac{1}{2m}\sum_{r=0}^{m-1} g_{k;m-r}c_{k;r};$$

$$g_{k;m} \equiv \sum_{i=1}^{\nu} e_{k;i}\left(1-\frac{s}{\lambda_i}\right)^m, \tag{65}$$

$$m\geq 1;$$

$$c_{jk;0} = \left(1+2\delta_{jk}\right)\left(\frac{s}{\lambda_j}\right)\left(\frac{s}{\lambda_k}\right)c_0;$$

$$c_{jk;m} = \frac{1}{2m}\sum_{r=0}^{m-1} g_{jk;m-r}c_{jk;r}; \tag{66}$$

$$g_{jk;m} \equiv \sum_{i=1}^{\nu} e_{jk;i}\left(1-\frac{s}{\lambda_i}\right)^m,$$

$$m\geq 1,$$

where the auxiliary coefficients $e_{k;i}$ and $e_{jk;i}$ are defined by

$$e_{k;i} = \begin{cases} 3 & \text{if } i = k \\ 1 & \text{otherwise,} \end{cases} \tag{67}$$

$$e_{kk;i} = \begin{cases} 5 & \text{if } i = k \\ 1 & \text{otherwise,} \end{cases}$$

$$\tag{68}$$

$$e_{jk;i} = \begin{cases} 3 & \text{if } i = j \text{ or } k \\ 1 & \text{otherwise} \end{cases} \quad(j\neq k).$$

It is not difficult to further generalize this theorem, so as to provide a chi-square expansion for any Gaussian integral $\alpha_{k\ell m\cdots}$. The proof follows closely the original one given by Ruben. We reproduce it in the appendix for α_k, just to highlight the differences arising when the Gaussian integral contains powers of the integration variable.

Analogously to (56), it is possible to estimate the systematic error produced when considering only the lowest k terms of the chi-square series of α_k and α_{jk}. Specifically, we find that

$$\mathscr{R}_{k;n} \equiv \left|\sum_{m=n}^{\infty} c_{k;m}(s;\lambda)\,F_{\nu+2(m+1)}\left(\frac{\rho}{s}\right)\right| \leq c_{k;0}$$

$$\cdot\frac{\eta^n}{n!}\left(1-\eta\right)^{-(\nu/2+n+1)}\frac{\Gamma(\nu/2+n+1)}{\Gamma(\nu/2)}$$

$$\cdot F_{\nu+2(n+1)}\left[\frac{(1-\eta)\rho}{s}\right] \equiv \mathfrak{R}_{k;n},$$

$$\tag{69}$$

$$\mathscr{R}_{jk;n} \equiv \left|\sum_{m=n}^{\infty} c_{jk;m}(s;\lambda)\,F_{\nu+2(m+2)}\left(\frac{\rho}{s}\right)\right| \leq c_{jk;0}$$

$$\cdot\frac{\eta^n}{n!}\left(1-\eta\right)^{-(\nu/2+n+2)}\frac{\Gamma(\nu/2+n+2)}{\Gamma(\nu/2)}$$

$$\cdot F_{\nu+2(n+2)}\left[\frac{(1-\eta)\rho}{s}\right] \equiv \mathfrak{R}_{jk;n}.$$

In order to evaluate all Ruben series with controlled uncertainty, we first set (see once more [2] for an exhaustive discussion on how to choose s) $s = 2\lambda_1\lambda_\nu/(\lambda_1+\lambda_\nu)$; then we choose a unique threshold ε representing the maximum tolerable systematic error; for example, $\varepsilon_{\text{dp}} = 1.0\times10^{-14}$ (roughly corresponding to double floating-point precision), for all α, α_k, and α_{jk}, and finally for each α_X we compute the integer

$$k_{\text{th}} \equiv \min_{n\geq 1}\{n : \mathfrak{R}_{X;n} < \varepsilon\}, \tag{70}$$

providing the minimum number of chi-square terms, for which the upper bound $\mathfrak{R}_{X;n}$ to the residual sum $\mathscr{R}_{X;n}$ lies below ε. Of course, this procedure overshoots the minimum number of terms really required for the $\mathscr{R}$'s to lie below ε, since we actually operate on the $\mathfrak{R}$'s instead of the $\mathscr{R}$'s. Nevertheless, the computational overhead is acceptable, as it will be shown in the next section. For the sake of

completeness, it must be said that typically the values of k_{th} for α, α_k, and α_{jk} with the same ϵ (and ρ, λ) are not much different from each other.

To conclude, we notice that k_{th} depends nontrivially upon λ. By contrast, since $F_\nu(x)$ is monotically increasing in x, we clearly see that k_{th} is monotically increasing in ρ. Now, should one evaluate α and the like for a given λ at several values of ρ, say $\rho_1 \leq \rho_2 \leq \cdots \leq \rho_{\max}$, it is advisable to save computing resources and work out Ruben coefficients just once, up to the order k_{th} corresponding to $\rho_{\max}$, since $k_{\text{th}}(\rho_1) \leq \cdots \leq k_{\text{th}}(\rho_{\max})$. We made use of this trick throughout our numerical experiences, as reported in the sequel.

6. Numerical Analysis of the Reconstruction Process

The fixed point (39) represents the simplest iterative scheme that can be used in order to reconstruct the solution $\lambda = \tau_\rho^{-1} \cdot \mu$. In the literature of numerical methods, this scheme is known as a nonlinear Gauss–Jacobi (GJ) iteration (see, e.g., [20]). Accordingly, we shall rewrite it as $\lambda_{\text{GJ},k}^{(n+1)} = T_k(\lambda_{\text{GJ}}^{(n)})$. As we have seen, the sequence $\lambda_{\text{GJ}}^{(n)}$ converges with no exception as $n \to \infty$, provided $\mu \in \mathscr{D}(\tau_\rho^{-1})$. Given $\epsilon_{\text{T}} > 0$, the number of steps n_{it} needed for an approximate convergence with relative precision ϵ_{T}, that is,

$$n_{\text{it}} \equiv \min_{n \geq 1} \left\{ n : \frac{\left\| \lambda_{\text{GJ}}^{(n)} - \lambda_{\text{GJ}}^{(n-1)} \right\|_\infty}{\left\| \lambda_{\text{GJ}}^{(n-1)} \right\|_\infty} < \epsilon_{\text{T}} \right\}, \qquad (71)$$

depends not only upon ϵ_{T}, but also on ρ and μ (note that the stopping rule is well conditioned, since $\|\lambda^{(n)}\|_\infty > 0$ $\forall n$ and also $\lim_{n \to \infty} \|\lambda^{(n)}\|_\infty > 0$). In order to characterize statistically the convergence rate of the reconstruction process, we must integrate out the fluctuations of n_{it} due to changes of μ; that is, we must average n_{it} by letting μ fluctuate across its own probability space. In this way, we obtain the quantity $\bar{n}_{\text{it}} \equiv \mathbb{E}_\mu[n_{\text{it}} \mid \epsilon_{\text{T}}, \rho]$, which better synthesizes the cost of the reconstruction for given ϵ_{T} and ρ. It should be evident that carrying out this idea analytically is hard, for on the one hand n_{it} depends upon μ nonlinearly and on the other hand μ has a complicated distribution, as we briefly explain below.

6.1. Choice of the Eigenvalue Ensemble. Since λ is the eigenvalue spectrum of a full covariance matrix, it is reasonable to assume its distribution to be a Wishart $\mathscr{W}_\nu(p, \Sigma_0)$ for some scale matrix Σ_0 and for some number of degrees of freedom $p \geq \nu$. In the sequel, we shall make the ideal assumption $\Sigma_0 = p^{-1} \cdot \mathbb{I}_\nu$, so that the probability measure of λ is (see, e.g., [21])

$$\mathrm{d}w_\nu(p; \lambda) = p^{(p+\nu^2-1)/2}$$

$$\cdot \frac{\pi^{\nu^2/2} \prod_{k=1}^\nu \lambda_k^{(p-\nu-1)/2} \exp\left((-p/2) \sum_{k=1}^\nu \lambda_k\right) \prod_{k<j} (\lambda_j - \lambda_k)}{2^{\nu p/2} \Gamma_\nu(p/2) \Gamma_\nu(\nu/2)} \quad (72)$$

$$\cdot \mathrm{d}^\nu \lambda.$$

Under this assumption, the probability measure of μ is obtained by performing the change of variable $\lambda = \tau_\rho^{-1} \cdot \mu$ in (72). Unfortunately, we have no analytic representation of τ_ρ^{-1}. Thus, we have neither an expression for the distribution of μ. However, μ can be extracted numerically as follows:

   (i) generate randomly $\Sigma \sim \mathscr{W}_\nu(p, p^{-1} \cdot \mathbb{I}_\nu)$ by means of the Bartlett decomposition [22];

   (ii) take the ordered eigenvalue spectrum λ of Σ;

   (iii) obtain μ by applying the truncation operator τ_ρ to λ.

Note that since $\mathscr{W}_\nu(p, p^{-1} \cdot \mathbb{I}_\nu)$ is only defined for $p \geq \nu$, we need to rescale p as ν increases. The simplest choice is to keep the ratio p/ν fixed. The larger this ratio, the closer Σ fluctuates around $\mathbb{I}_\nu$ (recall that if $\Sigma \sim \mathscr{W}_\nu(p, p^{-1} \cdot \mathbb{I}_\nu)$, then $\mathbb{E}[\Sigma_{ij}] = \delta_{ij}$ and $\text{var}(\Sigma_{ij}) = p^{-1}[1 + \delta_{ij}]$). In view of this, large values of p/ν are to be avoided, since they reduce the probability of testing the fixed point iteration on eigenvalue spectra characterized by large condition numbers $n_{\text{cond}} \equiv \lambda_\nu / \lambda_1$. For this reason, we have set $p = 2\nu$ in our numerical study.

Having specified an ensemble of matrices from which the eigenvalue spectra are extracted, we are now ready to perform numerical simulations. To begin with, we report in Figure 3 the marginal probability density function of the ordered eigenvalues $\{\lambda_k\}_{k=1}^\nu$ and their truncated counterparts $\{\mu_k\}_{k=1}^\nu$ for the Wishart ensemble $\mathscr{W}_{10}(20, 20^{-1} \cdot \mathbb{I}_{10})$ at $\rho = 1$, as obtained numerically from a rather large sample of matrices ($\approx 10^6$ units). It will be noted that (i) the effect of the truncation is severe on the largest eigenvalues, as a consequence of the analytic bounds of Corollary 2.1 and Proposition 2.2; (ii) while the skewness of the lowest truncated eigenvalues is negative, it becomes positive for the largest ones. This is due to a change of relative effectiveness of (29) (*i*) with respect to (29) (*ii*).

6.2. Choice of the Simulation Parameters. In order to explore the dependence of $\bar{n}_{\text{it}}$ upon ρ, we need to choose one or more simulation points for the latter. Ideally, it is possible to identify three different regimes in our problem: $\rho \lesssim \lambda_1$ (strong truncation regime), $\lambda_1 \lesssim \rho \lesssim \lambda_\nu$ (crossover), and $\rho \lesssim \lambda_\nu$ (weak truncation regime). We cover all of them with the following set of points:

$$\rho \in \{\text{Mo}(\lambda_1), \ldots, \text{Mo}(\lambda_\nu)\}$$
$$\cup \left\{ \frac{1}{2} \text{Mo}(\lambda_1), 2\text{Mo}(\lambda_\nu) \right\}, \qquad (73)$$

where $\text{Mo}(\cdot)$ stands for the mode. In principle, it is possible to determine $\text{Mo}(\lambda_k)$ with high accuracy by using analytic representations of the marginal probability densities of the ordered eigenvalues [24]. In practice, the latter become computationally demanding at increasingly large values of ν: for instance, the determination of the probability density of λ_2 requires $(\nu!)^2$ sums, which is unfeasible even at $\nu \sim 10$. Moreover, to our aims, it is sufficient to choose approximate values, provided that these lie not far from the exact ones. Accordingly, we have determined the eigenvalue modes
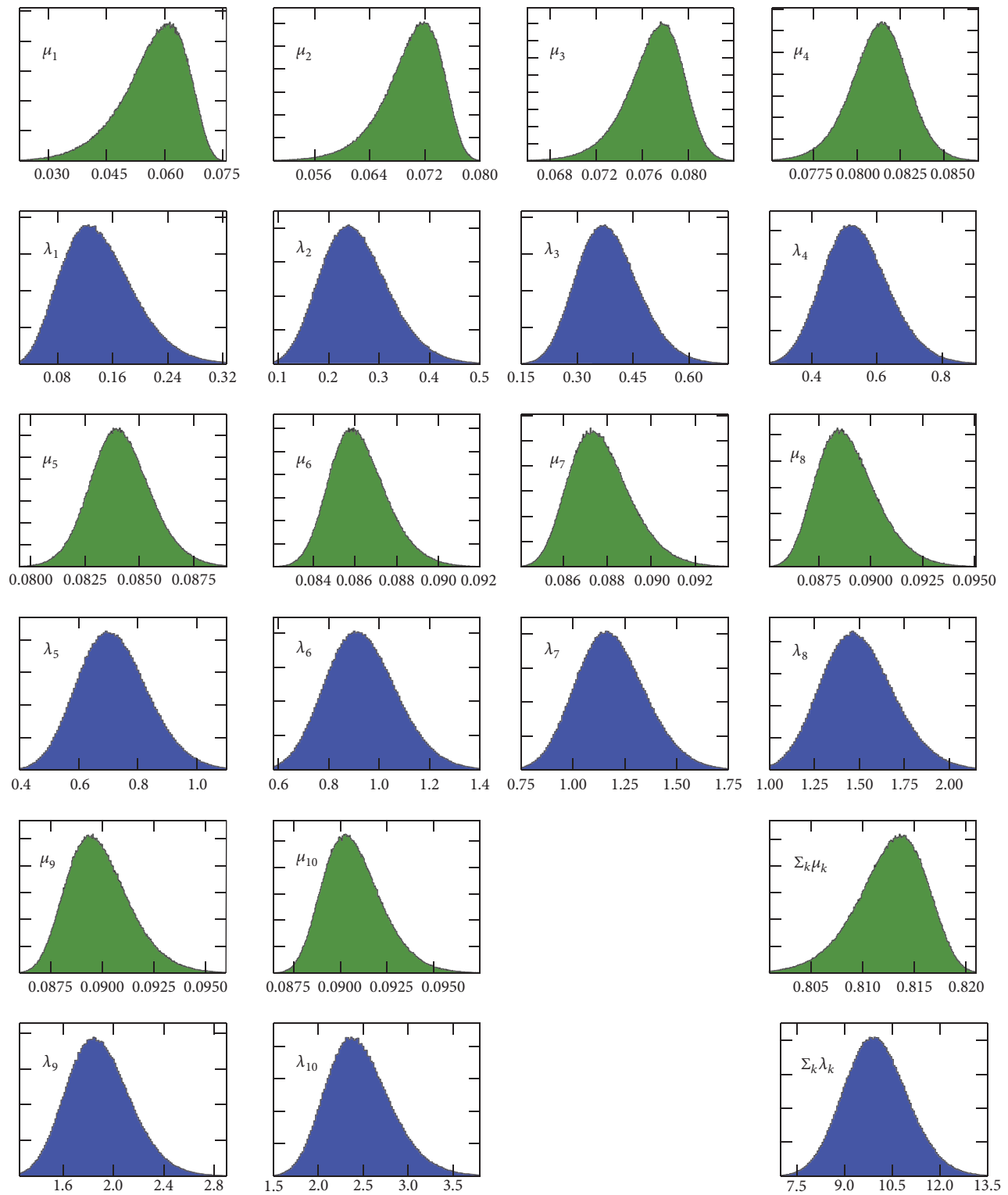
FIGURE 3: Monte Carlo simulation of the probability density function of the ordered eigenvalues λ_k (even rows) and their truncated counterparts μ_k at $\rho = 1$ (odd rows) for the Wishart ensemble $\mathscr{W}_{10}(20, 20^{-1} \cdot \mathbb{I}_{10})$. The last two plots (bottom right) display the distribution of the sum of eigenvalues.

TABLE 1: Numerical estimates of the mode of the ordered eigenvalues $\{\lambda_1, \ldots, \lambda_\nu\}$ of $\Sigma \sim \mathcal{W}_\nu(2\nu, (2\nu)^{-1} \cdot \mathbb{I}_\nu)$ with $\nu = 3, \ldots, 10$. The estimates have been obtained from Grenander's mode estimator [23].

	$\nu = 3$	$\nu = 4$	$\nu = 5$	$\nu = 6$	$\nu = 7$	$\nu = 8$	$\nu = 9$	$\nu = 10$
$\widehat{\mathrm{Mo}}(\lambda_1)$	0.1568	0.1487	0.1435	0.1383	0.1344	0.1310	0.1269	0.1258
$\widehat{\mathrm{Mo}}(\lambda_2)$	0.6724	0.4921	0.4017	0.3424	0.3039	0.2745	0.2554	0.2399
$\widehat{\mathrm{Mo}}(\lambda_3)$	1.6671	1.0112	0.7528	0.6071	0.5138	0.4543	0.4048	0.3693
$\widehat{\mathrm{Mo}}(\lambda_4)$	—	1.8507	1.2401	0.9621	0.7854	0.6684	0.5858	0.5288
$\widehat{\mathrm{Mo}}(\lambda_5)$	—	—	2.0150	1.4434	1.1269	0.9263	0.7956	0.7032
$\widehat{\mathrm{Mo}}(\lambda_6)$	—	—	—	2.1356	1.5789	1.2559	1.0527	0.9111
$\widehat{\mathrm{Mo}}(\lambda_7)$	—	—	—	—	2.2190	1.6764	1.3673	1.1603
$\widehat{\mathrm{Mo}}(\lambda_8)$	—	—	—	—	—	2.2763	1.7687	1.4624
$\widehat{\mathrm{Mo}}(\lambda_9)$	—	—	—	—	—	—	2.3210	1.8473
$\widehat{\mathrm{Mo}}(\lambda_{10})$	—	—	—	—	—	—	—	2.3775

numerically from samples of $N \simeq 10^6$ Wishart matrices. Our estimates are reported in Table 1 for $\nu = 3, \ldots, 10$. They have been obtained from Grenander's estimator [23]:

$$\mathrm{Mo}\,(\lambda_k)_{rs} = \frac{1}{2} \frac{\sum_{i=1}^{N-r} \left(\lambda_k^{(i)} + \lambda_k^{(i+r)}\right) \left(\lambda_k^{(i)} - \lambda_k^{(i+r)}\right)^{-s}}{\sum_{i=1}^{N-r} \left(\lambda_k^{(i)} - \lambda_k^{(i+r)}\right)^{-s}}, \quad (74)$$

with properly chosen parameters r, s.

We are now in the position to investigate numerically how many terms in Ruben's expansions must be considered as ε is set to $\varepsilon_{\mathrm{dp}} = 1.0 \times 10^{-14}$, for our choice of the eigenvalue ensemble $\lambda \sim \mathcal{W}_\nu(2\nu, (2\nu)^{-1} \cdot \mathbb{I}_\nu)$ and with ρ set as in Table 1. As an example, we report in Figure 4 the discrete distributions of k_{th} for the basic Gaussian integral α at $\nu = 10$, the largest dimension that we have simulated. As expected, we observe an increase of k_{th} with ρ. Nevertheless, we see that the number of Ruben's components to be taken into account for a double precision result keeps altogether modest even in the weak truncation regime, which proves the practical usefulness of the chi-square expansions.

6.3. Fixed Point Iteration at Work. The GJ iteration is too slow to be of practical interest. For instance, at $\nu = 10$, $\rho \simeq \mathrm{Mo}(\lambda_1)$ and $\epsilon_{\mathrm{T}} = 1.0 \times 10^{-7}$ (corresponding to a reconstruction of λ with single floating-point precision), it is rather easy to extract realizations of μ which require $n_{\mathrm{it}} \simeq 15,000$ to converge. An improvement of the GJ scheme is achieved via overrelaxation (GJOR); that is,

$$\lambda_{\mathrm{GJOR},k}^{(0)} = \mu_k,$$

$$\lambda_{\mathrm{GJOR},k}^{(n+1)} = \lambda_{\mathrm{GJOR},k}^{(n)} + \omega \left[T_k \left(\lambda_{\mathrm{GJOR}}^{(n)}\right) - \lambda_{\mathrm{GJOR},k}^{(n)} \right], \quad (75)$$

$$k = 1, \ldots, \nu.$$

Evidently, at $\omega = 1$, the GJOR scheme coincides with the standard GJ one. The optimal value ω_{opt} of the relaxation factor ω is not obvious even in the linear Jacobi scheme, where ω_{opt} depends upon the properties of the coefficient matrix of the system. For instance, if the latter is symmetric positive definite, it is demonstrated that the best choice is provided by $\omega_{\mathrm{opt}} \equiv 2(1 + \sqrt{1 - \sigma^2})^{-1}$, σ being the spectral radius of the

Jacobi iteration matrix [25]. In our numerical tests with the GJOR scheme, we found empirically that the optimal value of ω at $\rho \lesssim \lambda_\nu$ is close to the linear prediction, provided that σ is replaced by $\|J\|_\infty$, J being defined as in Section 3 (note that $\|J\|_\infty < 1$).

By contrast, the iteration diverges after few steps with increasing probability as $\rho/\lambda_\nu \to \infty$ if ω is kept fixed at $\omega = \omega_{\mathrm{opt}}$; in order to restore the convergence, ω must be lowered towards $\omega = 1$ as such limit is taken.

To give an idea of the convergence rate of the GJOR scheme, we show in Figure 5(a) a joint box-plot of the distributions of n_{it} at $\nu = 10$ and $\epsilon_{\mathrm{T}} = 1.0 \times 10^{-7}$. From the plot we observe that the distribution of n_{it} shifts rightwards as ρ decreases: clearly, the reconstruction is faster if ρ is in the weak truncation regime (where μ is closer to λ), whereas it takes more iterations in the strong truncation regime. The dependence of $\overline{n}_{\mathrm{it}}$ upon ρ, systematically displayed in Figure 6, is compatible with a scaling law

$$\log \overline{n}_{\mathrm{it}} \left(\rho, \nu, \epsilon_{\mathrm{T}}\right) = a\left(\nu, \epsilon_{\mathrm{T}}\right) - b\left(\nu, \epsilon_{\mathrm{T}}\right) \log \rho, \quad (76)$$

apart from small corrections occurring at large ρ. Equation (76) tells us that $\overline{n}_{\mathrm{it}}$ increases polynomially in $1/\rho$ at fixed ν. In order to estimate the parameters a and b in the strong truncation regime (where the algorithm becomes challenging), we performed jackknife fits to (76) of data points with $\rho \lesssim 1$. Results are collected in Figure 5(b), showing that b is roughly constant, while a increases almost linearly in ν. Thus, while the cost of the eigenvalue reconstruction is only polynomial in $1/\rho$ at fixed ν, it is exponential in ν at fixed ρ. The scaling law of the GJOR scheme is therefore better represented by $\overline{n}_{\mathrm{it}} = C e^{\kappa \nu}/\rho^b$, with C being a normalization constant independent of ρ and ν and κ representing approximately the slope of a as a function of ν. Although the GJOR scheme improves the GJ one, the iteration reveals to be still inefficient in a parameter subspace, which is critical for the applications.

6.4. Boosting the GJOR Scheme. A further improvement can be obtained by letting ω depend on the eigenvalue index in the GJOR scheme. Let us discuss how to work out such an adjustment. On commenting on Figure 3, we have already noticed that the largest eigenvalues are affected by the truncation to a larger extent than the smallest ones. Therefore,
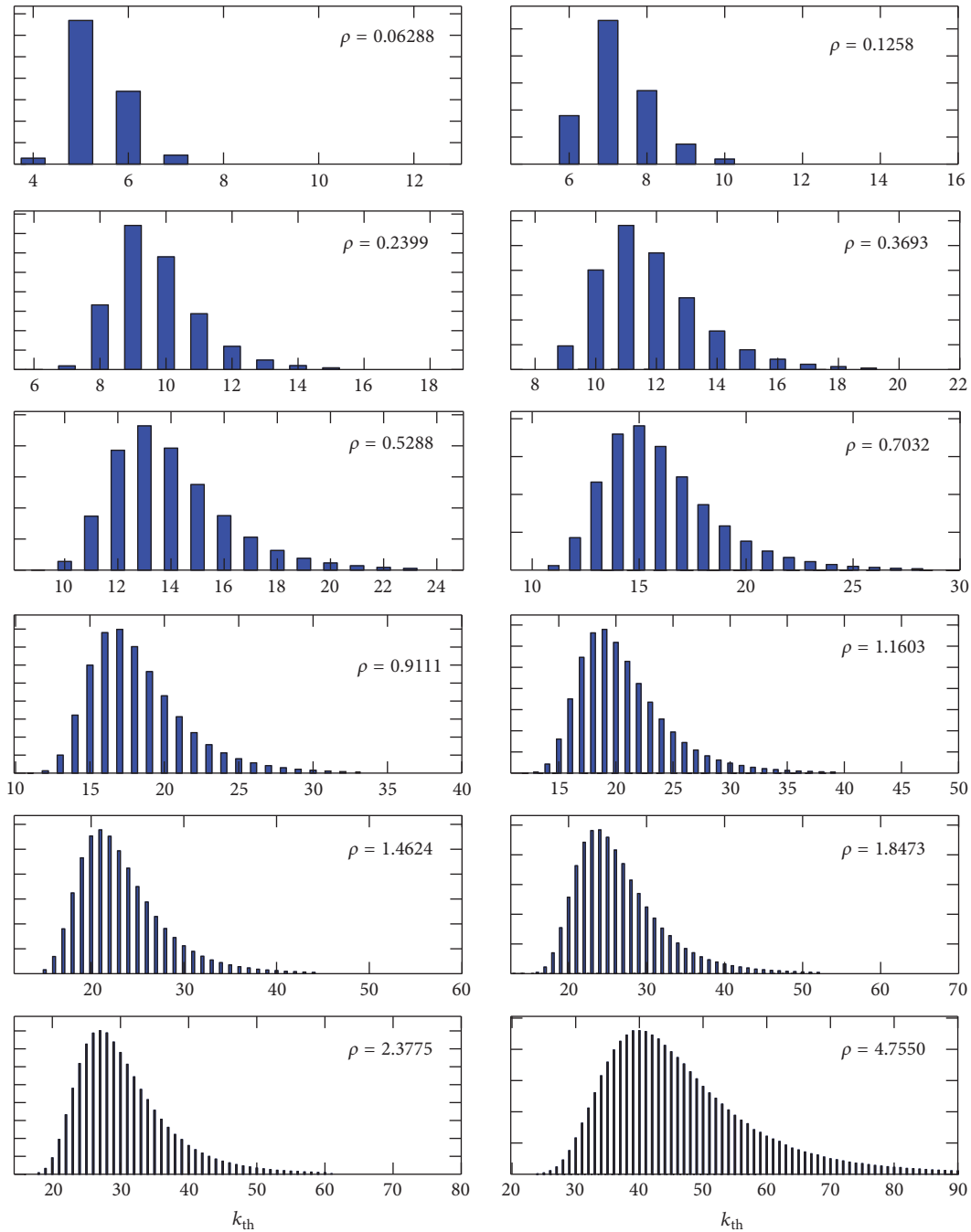
FIGURE 4: Monte Carlo simulation of the probability mass function of the parameter k_{th} for the Gaussian probability content α. The histograms refer to the eigenvalue ensemble λ of $\Sigma \sim \mathscr{W}_{10}(20, 20^{-1} \cdot \mathbb{I}_{10})$, with ρ chosen as in Table 1 and $\varepsilon = 1.0 \times 10^{-14}$.

they must perform a longer run through the fixed point iteration, in order to converge to the untruncated values. This is a possible qualitative explanation for the slowing down of the algorithm as $\rho \to 0$. In view of it, we expect to observe some acceleration of the convergence rate, if ω is replaced, for instance, by

$$\omega \longrightarrow \omega_k \equiv \left(1 + \beta \cdot k\right) \omega_{\mathrm{opt}}, \quad \beta \geq 0, \ k = 1, \ldots, \nu. \quad (77)$$

The choice $\beta = 0$ corresponds obviously to the standard GJOR scheme. Any other choice yields $\omega_k > \omega_{\mathrm{opt}}$. Therefore, the new scheme is also expected to display a higher rate of failures than the GJOR one at $\rho \gg \lambda_\nu$, for the reason explained in Section 5.3. The componentwise overrelaxation proposed in (77) is only meant to enhance the convergence speed in the strong truncation regime and in the crossover, where the improvement is actually needed.

v	$\widehat{a}$(jk. err.)	$\widehat{b}$(jk. err.)
3	4.93(3)	0.940(1)
4	5.25(2)	0.940(1)
5	5.44(2)	0.948(1)
6	5.65(1)	0.953(1)
7	5.85(1)	0.948(1)
8	6.02(1)	0.951(1)
9	6.18(1)	0.946(1)
10	6.31(1)	0.948(1)

(a)　　　　　　　　　　　　　　(b)
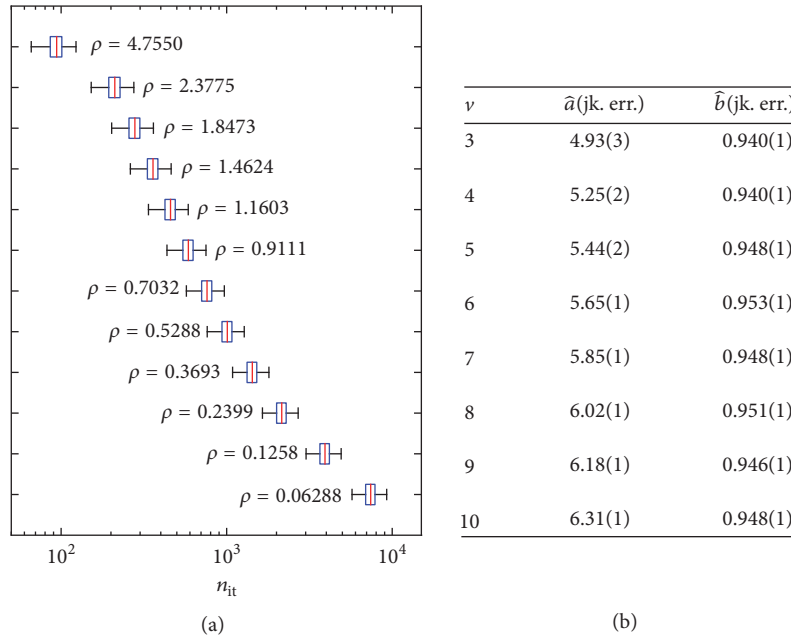
FIGURE 5: (a) Box-plot of n_{it} in the GJOR scheme at $v = 10$, with $\epsilon_T = 1.0 \times 10^{-7}$ and ρ chosen as in Table 1. The distributions have been reconstructed from a sample of $N \simeq 10^3$ eigenvalue spectra extracted from $\mathcal{W}_{10}(20, 20^{-1} \cdot \mathbb{1}_{10})$. The whiskers extend to the most extreme data point within $(3/2)(75\%-25\%)$ data range. (b) Numerical estimates of the scaling parameters a and b of the GJOR scheme, as obtained from jackknife fits to (76) of data points with $\rho \lesssim 1$ and $\epsilon_T = 1.0 \times 10^{-7}$. We quote in parentheses the jackknife error.

In order to confirm this picture, we have explored systematically the effect of β on $\overline{n}_{it}$ by simulating the reconstruction process at $v = 3, \ldots, 10$, with β varying from 0 to 2 in steps of $1/5$. First of all, we have observed that the rate of failures at large ρ is fairly reduced if the first $30 \div 50$ iterations are run with $\omega_k = \omega_{opt}$, and only afterwards β is switched on. Having minimized the failures, we have checked that for each value of β, the scaling law assumed in (76) is effectively fulfilled. Then, we have computed jackknife estimates of the scaling parameters a and b. These are plotted in Figure 7 as functions of v. Each trajectory (represented by a dashed curve) corresponds to a given value of β. Those with darker markers refer to smaller values of β and the other way round. From the plots we notice that

(i) all the trajectories with $\beta > 0$ lie below the one with $\beta = 0$;

(ii) the trajectories of a display a clear increasing trend with v; yet their slope lessens as β increases. By contrast, the trajectories of b develop a mild increasing trend with v as β increases, though this is not strictly monotonic;

(iii) the trajectories of both a and b seem to converge to a limit trajectory as β increases; we observe a saturation phenomenon, which thwarts the benefit of increasing β beyond a certain threshold close to $\beta_{max} \simeq 2$.

We add that pushing β beyond β_{max} is counterproductive, as the rate of failures becomes increasingly relevant in the crossover and eventually also in the strong truncation regime.

By contrast, if $\beta \lesssim \beta_{max}$ the rate of failures keeps very low for essentially all simulated values of ρ.

Our numerical results signal a strong reduction of the slowing down of the convergence rate. Indeed, (i) means qualitatively that C and b are reduced as β increases. (ii) means that κ is reduced as β increases (this is the most important effect, as κ is mainly responsible for the exponential slowing down with v). The appearance of a slope in the trajectories of b as β increases indicates that a mild exponential slowing down is also developed at denominator of the scaling law $\overline{n}_{it} = Ce^{\kappa v}/\rho^b$, but the value of b is anyway smaller than at $\beta = 0$. Finally, (iii) means that choosing $\beta > \beta_{max}$ has a minor impact on the performance of the algorithm. In Figure 8, we report a plot of the parameter κ (obtained from least-squares fits of data to a linear model $a = a_0 + \kappa \cdot v$) as a function of β. We see that $\kappa(\beta = 0)/\kappa(\beta = 2) \simeq 4$. This quantifies the maximum exponential speedup of the convergence rate, which can be achieved by our proposal. When β is close to β_{max}, $\overline{n}_{it}$ amounts to few hundreds at $v = 10$ and $\rho \simeq \lambda_1/2$.

7. On the Ill-Posedness of the Reconstruction in Sample Space

So far we have discussed the covariance reconstruction under the assumption that $\mu = \tau_\rho \cdot \lambda$ represents the exact truncated counterpart of some $\lambda \in \mathbb{R}^v$ and we have looked at the algorithmic properties of the iteration schemes which operatively define τ_ρ^{-1}. Such analysis is essential in order to characterize τ_ρ^{-1} mathematically; yet it is not sufficient in real situations, specifically when μ is perturbed by statistical noise.
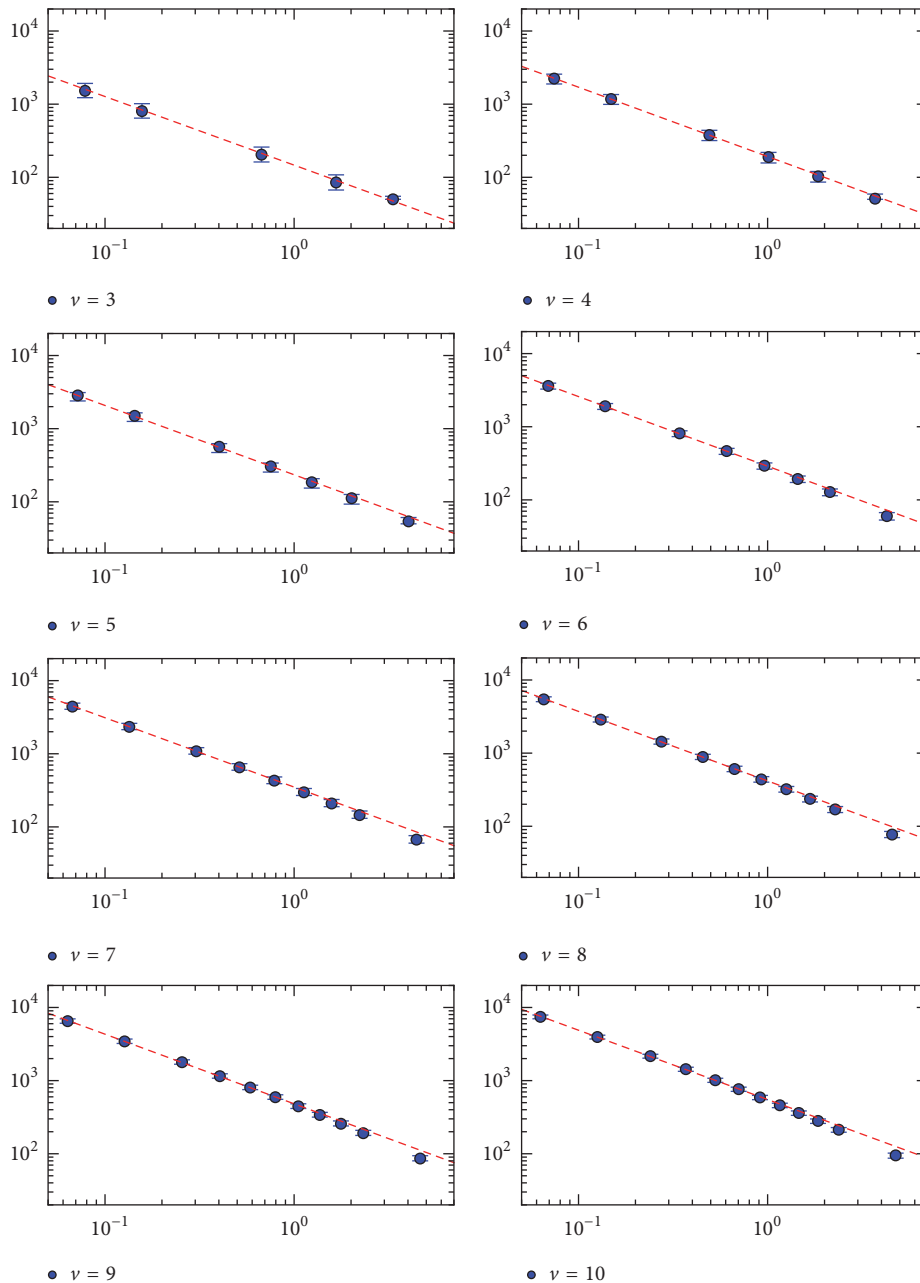
FIGURE 6: Log-log plots of $\overline{n}_{it}$ versus ρ in the GJOR scheme at $\epsilon_T = 1.0 \times 10^{-7}$. The parameter ρ has been chosen as in Table 1. The (red) dashed line in each plot represents our best jackknife linear fit to (76) of data points with $\rho \lesssim 1$.

In this section, we examine the difficulties arising when performing the covariance reconstruction in sample space. We first recall that, according to Hadamard [26], a mathematical problem is well posed provided that the following conditions are fulfilled:

(H_1) there exists always a solution to the problem;

(H_2) the solution is unique;

(H_3) the solution depends smoothly on the input data.

Inverse problems are often characterized by violation of one or more of them; see, for instance, [13]. In such cases,

the standard practice consists in regularizing the inverse operator, that is, in replacing it by a stable approximation. With regard to our problem, the reader will recognize that (H_1) is violated (and the problem becomes ill-posed) as soon as the space of the input data is allowed to be a superset of $\mathscr{D}(\tau_\rho^{-1})$: once clarifying how μ is concretely estimated in the applications (Sections 7.1 and 7.2), we propose a perturbative regularization of τ_ρ^{-1}, which improves effectively the fulfillment of (H_1) (Section 7.3). By contrast, Proposition 8 guarantees that whenever a solution exists, it is also unique; thus (H_2) is never of concern. Finally, the fulfillment of (H_3)
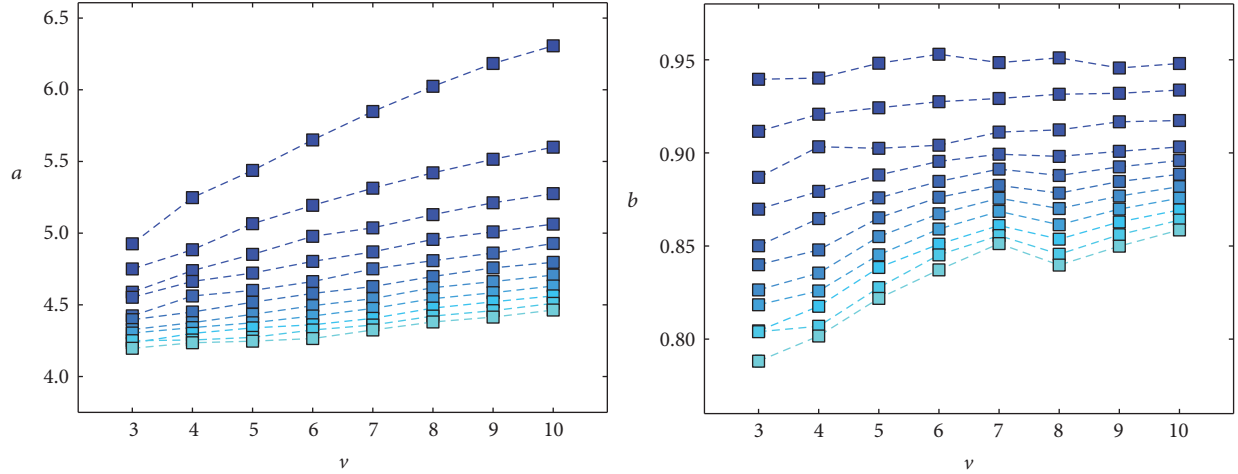
FIGURE 7: Scaling parameters a and b of the modified GJOR scheme as functions of v at $\epsilon_{\mathrm{T}} = 1.0 \times 10^{-7}$, with β varying in the range $0 \div 2$ in steps of $1/5$. Each trajectory (represented by a dashed curve) refers to a different value of β. Those with darker markers correspond to smaller values of β and the other way round.
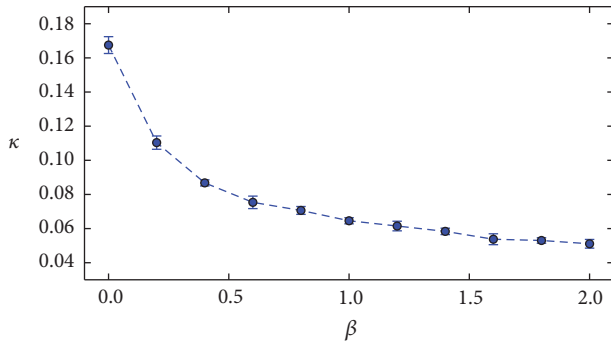
FIGURE 8: The parameter κ as a function of β. Estimates of κ are obtained from least-squares fits of data to a linear model $a = a_0 + \kappa \cdot v$.

depends on how the statistical noise on μ is nonlinearly inflated by the action of τ_ρ^{-1}. For the sake of conciseness, in the present paper, we just sketch the main ideas underlying the perturbative regularization of τ_ρ^{-1}, whereas a technical implementation of it and a discussion of (H_3) are deferred to a separate paper [27].

7.1. Definition of the Sample Truncated Covariance Matrix. The examples of Section 2 assume that (i) spherical truncations are operated on a representative sample $\mathscr{P}_N = \{x^{(k)}\}_{k=1}^N$ of $X \sim \mathscr{N}_v(0, \Sigma)$ with finite size N, (ii) ρ is known exactly, and (iii) the input budget for the covariance reconstruction is given by the subset

$$\mathcal{Q}_M = \left\{ x \in \mathscr{P}_N : \|x\|^2 < \rho \right\}, \quad \text{with } |\mathcal{Q}_M| = M \leq N. \quad (78)$$

As usual in the analysis of stochastic variables in sample space, we assume that the observations $x^{(k)}$ are realizations

of i.i.d. stochastic variables $X^{(k)} \sim \mathscr{N}_v(0, \Sigma)$, $k = 1, \ldots, N$. Thus, M is itself a stochastic variable in sample space, where it reads

$$M = \sum_{k=1}^N \mathbb{I}_{\mathscr{B}_v(\rho)} \left(X^{(k)} \right) \equiv \sum_{k=1}^N \mathbb{I}_k, \quad (79)$$

with $\mathbb{I}_{\mathscr{B}_v(\rho)}(\cdot)$ denoting the characteristic function of $\mathscr{B}_v(\rho)$ and $\mathbb{I}_k \equiv \mathbb{I}_{\mathscr{B}_v(\rho)}(X^{(k)})$ being just a shortcut for its extended counterpart. It is easily recognized that $M \sim B(N, \alpha)$ is a binomial variate. If we indeed denote by $\mathfrak{E}$ the sample expectation operator (*i.e.*, the integral with respect to the product measure of the joint variables $\{X^{(k)}\}_{k=1}^N$), then a standard calculation yields

$$\mathfrak{E}[M] = \mathfrak{E}\left[\sum_{k=1}^N \mathbb{I}_k \right] = \sum_{k=1}^N \mathfrak{E}[\mathbb{I}_k] = \sum_{k=1}^N \alpha = \alpha N,$$

$$\mathfrak{var}[M] = \mathfrak{E}[M^2] - \mathfrak{E}[M]^2$$

$$= \sum_{k=1}^N \mathfrak{E}\left[\mathbb{I}_k^2 \right] + \sum_{k,s:k \neq s}^{1 \cdots N} \mathfrak{E}[\mathbb{I}_k \mathbb{I}_s] - \alpha^2 N^2$$

$$= \sum_{k=1}^N \mathfrak{E}[\mathbb{I}_k] + \sum_{k,s:k \neq s}^{1 \cdots N} \mathfrak{E}[\mathbb{I}_k] \mathfrak{E}[\mathbb{I}_s] - \alpha^2 N^2$$

$$= \alpha N + \alpha^2 N(N-1) - \alpha^2 N^2$$

$$= \alpha(1 - \alpha) N.$$

$$(80)$$

Hence, we see that the relative dispersion of M is $O(N^{-1/2})$. Now, the simplest way to measure Σ and $\mathfrak{S}_{\mathscr{B}}$, respectively, from the sets $\mathscr{P}_N$ and $\mathcal{Q}_M$ is via the classical estimators

$$
\widehat{\Sigma}_{ij} = \frac{1}{N-1} \sum_{x \in \mathscr{P}_N} (x - \overline{x})_i \, (x - \overline{x})_j,
$$

$$
\overline{x}_i = \frac{1}{N} \sum_{x \in \mathscr{P}_N} x_i, \tag{81}
$$

$$
\left(\widehat{\mathfrak{S}}_{\mathscr{B}}\right)_{ij} = \frac{1}{M-1} \sum_{x \in \mathcal{Q}_M} (x - \widetilde{x})_i \, (x - \widetilde{x})_j,
$$

$$
\widetilde{x}_i = \frac{1}{M} \sum_{x \in \mathcal{Q}_M} x_i. \tag{82}
$$

We define the sample estimates $\widehat{\lambda}$ and $\widehat{\mu}$, respectively, of λ and μ as the eigenvalue spectra of $\widehat{\Sigma}$ and $\widehat{\mathfrak{S}}_{\mathscr{B}}$. By symmetry arguments we see that $\widetilde{x}_i$ is unbiased. Indeed, it holds

$$
\mathfrak{E}\left[\widetilde{x}_i\right] = \sum_{k=1}^{N} \mathfrak{E}\left[X_i^{(k)} \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s} \right]. \tag{83}
$$

The *right-hand side* of (83) makes only sense if we conventionally define the integrand to be zero in the integration subdomain $\{X^{(k)} \notin \mathscr{B}_\nu(\rho), \; \forall k\}$ or equivalently if we interpret $\mathfrak{E}[\widetilde{x}_i]$ as the conditional one $\mathfrak{E}[\widetilde{x}_i \mid M > 0]$ (the event $M > 0$ occurs a.s. only as $N \to \infty$). Since the sample measure is even under $X^{(k)} \to -X^{(k)}$ while the integrand is odd, we immediately conclude that $\mathfrak{bias}[\widetilde{x}_i] = 0$.

7.2. Bias of the Sample Truncated Covariance Matrix. The situation gets somewhat less trivial with $\widehat{\mathfrak{S}}_{\mathscr{B}}$: the normalization factor $(M-1)^{-1}$, which has been chosen in analogy with (81), is not sufficient to remove completely the bias of $\widehat{\mathfrak{S}}_{\mathscr{B}}$ at finite N, though we aim at showing here that the residual bias is exponentially small and asymptotically vanishing. In order to see this, we observe that

$$
\mathfrak{E}\left[\left(\widehat{\mathfrak{S}}_{\mathscr{B}}\right)_{ij}\right]
$$

$$
= \sum_{k=1}^{N} \mathfrak{E}\left[\left(X_i^{(k)} - \widetilde{X}_i\right)\left(X_j^{(k)} - \widetilde{X}_j\right) \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s - 1} \right]
$$

$$
= \sum_{\ell, r=1}^{\nu} R_{i\ell} R_{jr}
$$

$$
\cdot \sum_{k=1}^{N} \mathfrak{E}_{\text{diag}}\left[\left(X_\ell^{(k)} - \widetilde{X}_\ell\right)\left(X_r^{(k)} - \widetilde{X}_r\right) \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s - 1} \right], \tag{84}
$$

with $\mathfrak{E}_{\text{diag}}$ denoting the sample expectation corresponding to a multinormal measure with diagonal covariance matrix $\Lambda = \text{diag}(\lambda) = R^{\mathrm{T}} \Sigma R$, conditioned to $M > 1$. Having diagonalized the product measure, we observe that the integrand on the *right-hand side* is odd for $\ell \neq r$ and even for $\ell = r$ under the joint change of variables $X_\ell^{(k)} \to -X_\ell^{(k)}$ for $k = 1, \ldots, N$,

similarly to what we did in Section 2. As a consequence, it holds

$$
\mathfrak{E}\left[\left(\widehat{\mathfrak{S}}_{\mathscr{B}}\right)_{ij}\right]
$$

$$
= \sum_{\ell=1}^{\nu} R_{i\ell} R_{j\ell} \left\{ \sum_{k=1}^{N} \mathfrak{E}_{\text{diag}}\left[\left(X_\ell^{(k)} - \widetilde{X}_\ell\right)^2 \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s - 1} \right] \right\}, \tag{85}
$$

whence we infer that the matrix $\mathfrak{E}[\widehat{\mathfrak{S}}_{\mathscr{B}}]$ is diagonalized by the same matrix R as Σ. From (85) we also conclude that

$$
\mathfrak{bias}\left[\widehat{\mathfrak{S}}_{\mathscr{B}}\right] = R \, \text{diag}\,(w)\, R^{\mathrm{T}},
$$

$$
w_i \equiv \sum_{k=1}^{N} \mathfrak{E}_{\text{diag}}\left[\left(X_i^{(k)} - \widetilde{X}_i\right)^2 \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s - 1} \right] \tag{86}
$$

$$
- \mu_i,
$$

$i = 1, \ldots, \nu$. It should be observed that in general $w_i \neq \mathfrak{bias}[\widehat{\mu}_i]$ since the computation of $\widehat{\mu}_i$ requires the diagonalization of $\widehat{\mathfrak{S}}_{\mathscr{B}}$, which is in general performed by a diagonalizing matrix $\widehat{R} \neq R$. Nevertheless, if w vanishes then $\mathfrak{bias}[\widehat{\mathfrak{S}}_{\mathscr{B}}]$ vanishes too. Now, we observe that w_i splits into three contributions:

$$
w_{i1} = \sum_{k=1}^{N} \mathfrak{E}_{\text{diag}}\left[\left(X_i^{(k)}\right)^2 \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s - 1} \right] - \mu_i,
$$

$$
w_{i2} = -2 \sum_{k=1}^{N} \mathfrak{E}_{\text{diag}}\left[X_i^{(k)} \widetilde{X}_i \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s - 1} \right], \tag{87}
$$

$$
w_{i3} = \sum_{k=1}^{N} \mathfrak{E}_{\text{diag}}\left[\left(\widetilde{X}_i\right)^2 \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s - 1} \right],
$$

which can be exactly calculated and expressed in terms of μ_i, α, and N. For instance,

$$
w_{i1} = N \mathfrak{E}_{\text{diag}}\left[\frac{1}{M-1} \left(X_i^{(1)}\right)^2 \mathbb{I}_1 \mid M > 1 \right] - \mu_i
$$

$$
= N \sum_{m=2}^{N} \frac{1}{m-1} \mathfrak{E}_{\text{diag}}\left[\left(X_i^{(1)}\right)^2 \mathbb{I}_1 \mid M = m \right] - \mu_i
$$

$$
= N \sum_{m=2}^{N} \frac{1}{m-1} \alpha \mu_i \binom{N-1}{m-1} \alpha^{m-1} (1-\alpha)^{N-m} - \mu_i \tag{88}
$$

$$
= \mu_i \sum_{m=2}^{N} \frac{m}{m-1} \binom{N}{m} \alpha^m (1-\alpha)^{N-m} - \mu_i.
$$

Analogously, we have

$$
w_{i2} = -2\mu_i \sum_{m=2}^{N} \frac{1}{m-1} \binom{N}{m} \alpha^m (1-\alpha)^{N-m},
$$

$$
w_{i3} = \mu_i \sum_{m=2}^{N} \frac{1}{m-1} \binom{N}{m} \alpha^m (1-\alpha)^{N-m}. \tag{89}
$$

Hence, it follows that

$$w_i = -\mu_i \left[1 + \alpha \left(N - 1\right)\right] \left(1 - \alpha\right)^{N-1} . \tag{90}$$

Since $\alpha > 0$, we see that $\lim_{N \to \infty} w_i = 0$. Thus, we conclude that $\widehat{\mathfrak{S}}_{\mathcal{B}}$ is asymptotically unbiased.

A discussion of the variance of the sample truncated covariance matrix is beyond the scope of the present paper. We just observe that, apart from the above calculation, studying the sample properties of the truncated spectrum is made hard by the fact that eigenvalues and eigenvectors of a diagonalizable matrix are intimately related from their very definition; thus such study would require a careful consideration of the distribution of the sample diagonalizing matrix $\widehat{R} \neq R$ of $\widehat{\mathfrak{S}}_B$.

7.3. Perturbative Regularization of τ_ρ^{-1}. When μ is critically close to the internal boundary of $\mathscr{D}(\tau_\rho^{-1})$, a sample estimate $\widehat{\mu}$ may fall outside of it due to statistical fluctuations. In that case the iterative procedure described in the previous sections diverges. On the quantitative side, the ill-posedness of the reconstruction problem is measured by the failure probability

$$p_{\text{fail}} \left(\rho, \Sigma, N\right) = \mathbb{P}\left[\widehat{\mu} \notin \mathscr{D}\left(\tau_\rho^{-1}\right) \mid X^{(k)} \sim \mathcal{N}_\nu\left(0; \Sigma\right), \; k \right.$$
$$\left. = 1, \ldots, N\right], \tag{91}$$

which is a highly nontrivial function of ρ, Σ, and N. An illustrative example of it is reported in Figure 9(a), which refers to a specific case with $\nu = 4$ and $\Sigma = \text{diag}(0.1, 0.3, 0.8, 2.2)$. The plot suggests that the iterative procedure becomes severely ill-posed in the regime of strong truncation.

In order to regularize the problem, we propose to go back to (19) and consider it from a different perspective. Specifically, we move from the observation that a simplified framework occurs in the special circumstance when the eigenvalue spectra are fully degenerate, which is essentially equivalent to the setup of [1]. If $\mu_1 = \cdots = \mu_\nu \equiv \widetilde{\mu}$, by symmetry arguments it follows that $\lambda_1 = \cdots = \lambda_\nu \equiv \widetilde{\lambda}$ and the other way round. Equation (19) reduces in this limit to

$$\widetilde{\mu} = \widetilde{\lambda} \frac{F_{\nu+2}}{F_\nu}\left(\frac{\rho}{\widetilde{\lambda}}\right) \equiv \mathscr{T}_\rho\left(\widetilde{\lambda}\right). \tag{92}$$

It can be easily checked that the function $\mathscr{T}_\rho(\widetilde{\lambda})$ is monotonically increasing in $\widetilde{\lambda}$. In addition, we have

$$\begin{array}{ll} \text{(i)} & \lim_{\widetilde{\lambda} \to 0} \mathscr{T}_\rho\left(\widetilde{\lambda}\right) = 0, \\[2mm] \text{(ii)} & \lim_{\widetilde{\lambda} \to \infty} \mathscr{T}_\rho\left(\widetilde{\lambda}\right) = \frac{\rho}{\nu + 2}; \end{array} \tag{93}$$

thus (92) can be surely (numerically) inverted provided that $0 < \widetilde{\mu} < \rho/(\nu + 2)$. We can regard (92) as an approximation to the original problem (19). When μ is not degenerate, we must define $\widetilde{\mu}$ in terms of the components of μ. One possibility is to average them, that is, to choose

$$\widetilde{\mu} = \frac{1}{\nu} \sum_{i=1}^{\nu} \mu_i. \tag{94}$$

Subject to this, we expect $\widetilde{\lambda}$ to lie somewhere between λ_1 and λ_ν. Equation (92) can be thought of as the lowest order approximation of a perturbative expansion of (19) around the point $\lambda_T = \{\widetilde{\lambda}, \ldots, \widetilde{\lambda}\}$. If the condition number of Σ is not extremely large, such an expansion is expected to quickly converge, so that a few perturbative corrections to λ_T should be sufficient to guarantee a good level of approximation.

As mentioned above, a technical implementation of the perturbative approach and a thorough discussion of its properties are deferred to a separate paper [27]. Here, we limit ourselves to observing that the definition domain of perturbation theory is ultimately set by its lowest order approximation, since corrections to (92) are all algebraically built in terms of it, with no additional constraints. Following (94), the domain of $\mathscr{T}_\rho^{-1}$ comes to be defined as

$$\mathscr{D}\left(\mathscr{T}_\rho^{-1}\right) = \left\{\mu \in \mathbb{R}_+^\nu : \sum_{i=1}^{\nu} \mu_i \leq \frac{\rho\nu}{\nu + 2}\right\}, \tag{95}$$

and it is clear that $\mathscr{D}(\tau_\rho^{-1}) \subset \mathscr{D}(\mathscr{T}_\rho^{-1})$ (it is sufficient to sum term by term all the inequalities contributing to (36)). In Figure 10, we show the set difference $\mathscr{D}(\mathscr{T}_\rho^{-1}) \setminus \mathscr{D}(\tau_\rho^{-1})$ in $\nu = 2$ and $\nu = 3$ dimensions. When $\mu \in \mathscr{D}(\tau_\rho^{-1})$ but its estimate $\widehat{\mu} \notin \mathscr{D}(\tau_\rho^{-1})$, it may well occur $\widehat{\mu} \in \mathscr{D}(\mathscr{T}_\rho^{-1})$; that is, the set difference acts as an absorbing shield of the statistical noise. Therefore, if we define the failure probability of the perturbative reconstruction as

$$q_{\text{fail}}\left(\rho, \Sigma, N\right) = \mathbb{P}\left[\widehat{\mu} \notin \mathscr{D}\left(\mathscr{T}_\rho^{-1}\right) \mid X^{(k)} \right.$$
$$\left. \sim \mathcal{N}_\nu\left(0; \Sigma\right), \; k = 1, \ldots, N\right], \tag{96}$$

we expect the inequality $q_{\text{fail}}(\rho, \Sigma, N) \ll p_{\text{fail}}(\rho, \Sigma, N)$ to generously hold. An example is given in Figure 9(b): we see that q_{fail} becomes lower than p_{fail} by orders of magnitude as soon as ρ and N are not exceedingly small. In this sense, the operator $\mathscr{T}_\rho^{-1}$ can be regarded as the lowest order approximation of a regularizing operator for τ_ρ^{-1}.

8. Conclusions

In this paper we have studied how to reconstruct the covariance matrix Σ of a normal multivariate $X \sim \mathcal{N}_\nu(0, \Sigma)$ from the matrix $\mathfrak{S}_{\mathcal{B}}$ of the spherically truncated second moments, describing the covariances among the components of X when the probability density is cut off outside a centered Euclidean ball. We have shown that Σ and $\mathfrak{S}_{\mathcal{B}}$ share the same eigenvectors. Therefore, the problem amounts to relating the eigenvalues of Σ to those of $\mathfrak{S}_{\mathcal{B}}$. Such relation entails the inversion of a system of nonlinear integral equations, which admits unfortunately no closed-form solution. Having found a necessary condition for the invertibility of the system, we have shown that the eigenvalue reconstruction can be achieved numerically via a converging fixed point iteration. In order to prove the convergence, we rely ultimately upon some probability inequalities, known in the literature as *square correlation inequalities*, which have been recently proved in [17].
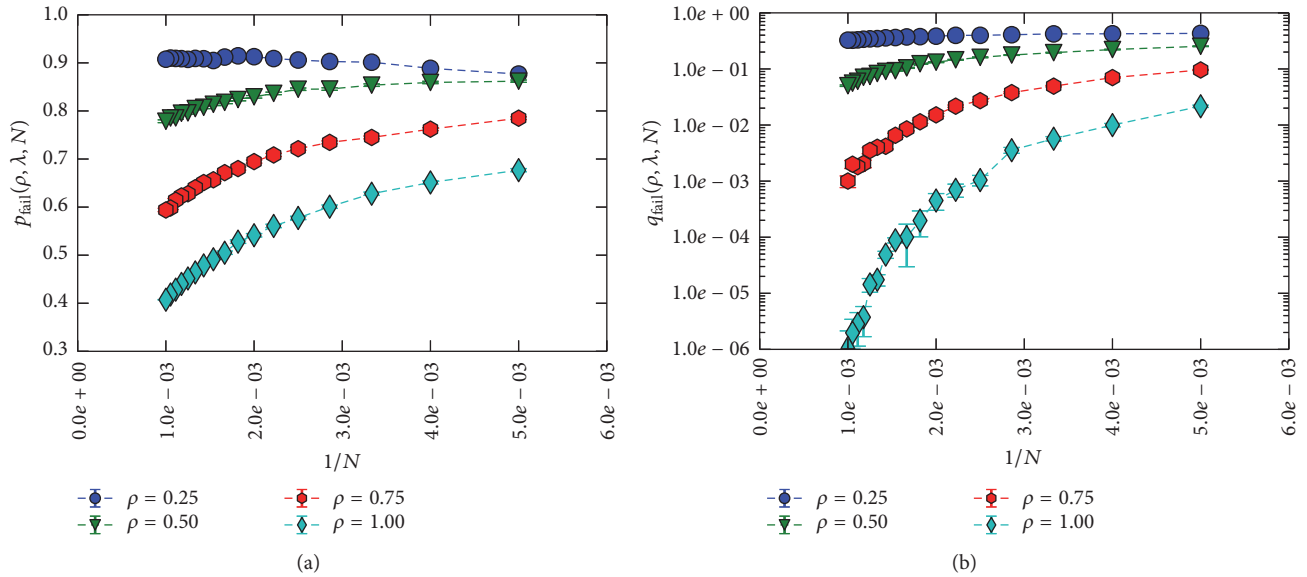
FIGURE 9: (a) Numerical reconstruction of the failure probability of the iterative procedure as $\nu = 4$ and $\Sigma = \mathrm{diag}(0.1, 0.3, 0.8, 2.2)$, for several values of ρ and for $N = 200, 250, \ldots, 1000$. (b) Failure probability of the perturbative regularization with the same parameters.
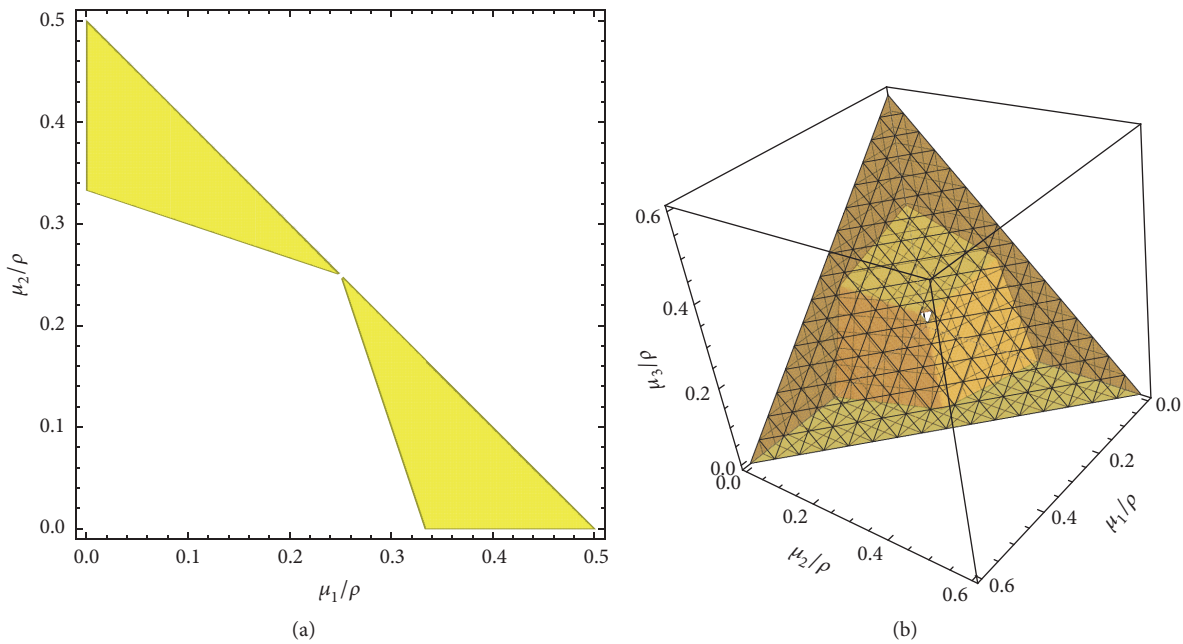
FIGURE 10: (a) Set difference $\mathscr{D}(\mathscr{T}_\rho^{-1}) \setminus \mathscr{D}(\tau_\rho^{-1})$ in $\nu = 2$ dimensions. (b) Set difference $\mathscr{D}(\mathscr{T}_\rho^{-1}) \setminus \mathscr{D}(\tau_\rho^{-1})$ in $\nu = 3$ dimensions.

In order to explore the convergence rate of the fixed point iteration, we have implemented some variations of the nonlinear Gauss–Jacobi scheme. Specifically, we have found that overrelaxing the basic iteration enhances the convergence rate by a moderate factor. However, the overrelaxed algorithm still slows down exponentially in the number of eigenvalues and polynomially in the truncation radius of the Euclidean ball. We have shown that a significant reduction of the slowing down can be achieved in the regime of strong truncation by adapting the relaxation parameter to the eigenvalue that is naturally associated with, so as to boost the higher components of the spectrum.

We have also discussed how the iterative procedure works when the eigenvalue reconstruction is performed on sample estimates of the truncated covariance spectrum. Specifically, we have shown that the statistical fluctuations make the

problem ill-posed. We have sketched a possible way out based on perturbation theory, which is thoroughly discussed in a separate paper [27].

A concrete implementation of the proposed approach requires the computation of a set of multivariate Gaussian integrals over the Euclidean ball. For this, we have extended to the case of interest a technique, originally proposed by Ruben for representing the probability content of quadratic forms of normal variables as a series of chi-square distributions. In the paper, we have shown the practical feasibility of the series expansion for the integrals involved in our computations.

Appendix

Proof of Theorem 9. As already mentioned in Section 4, the proof follows in the tracks of the original one of [2]. We detail the relevant steps for α_k, while for α_{jk} we only explain why it is necessary to distinguish between equal or different indices and the consequences for either case.

In order to prove (57), we first express α_k in spherical coordinates; that is, we perform the change of variable $x = ru$, being $r = \|x\|$ and $u \in \partial \mathcal{B}_v(1)$ (recall that $d^v x = r^{v-1} dr\, du$, with du embodying the angular part of the spherical Jacobian and the differentials of $v - 1$ angles); then we insert a factor of $1 = \exp(r^2/2s)\exp(-r^2/2s)$ under the integral sign. Hence, α_k reads

$$
\alpha_k(\rho;\lambda) = \frac{1}{(2\pi)^{v/2}|\Lambda|^{1/2}} \int_0^{\sqrt{\rho}} dr\, r^{v-1} \frac{r^2}{\lambda_k} \exp\left(-\frac{r^2}{2s}\right)
$$
$$
\cdot \int_{\partial \mathcal{B}_v(1)} du\, u_k^2 \exp\left(-\frac{Q(u)\,r^2}{2}\right).
$$
(A.1)

The next step consists in expanding the inner exponential in Taylor series (in his original proof, Ruben considers a more general setup, with the center of the Euclidean ball shifted by a vector $b \in \mathbb{R}^v$ from the center of the distribution. In that case, the Gaussian exponential looks different and must be expanded in series of Hermite polynomials. Here, we work in a simplified setup, where the Hermite expansion reduces to Taylor's); namely,

$$
\exp\left(-\frac{Q(u)\,r^2}{2}\right) = \sum_{m=0}^{\infty} \frac{1}{m!} \frac{r^{2m}}{2^m} (-Q)^m.
$$
(A.2)

This series converges uniformly in u. We review the estimate just for the sake of completeness:

$$
\left| \sum_{m=0}^{\infty} \frac{1}{m!} \frac{r^{2m}}{2^m} (-Q)^m \right| \le \sum_{m=0}^{\infty} \frac{1}{m!} \frac{r^{2m}}{2^m} q_0^m = \exp\left(\frac{r^2 q_0}{2}\right),
$$
(A.3)

where $q_0 = \max_i |1/s - 1/\lambda_i|$. It follows that we can integrate the series term by term. With the help of the uniform average operator introduced in (53), α_k is recast to

$$
\alpha_k(\rho;\lambda)
$$
$$
= \sum_{m=0}^{\infty} \frac{1}{m!} \frac{1}{\lambda_k} \frac{1}{|\Lambda|^{1/2}} \mathbb{M}\left[(-Q)^m u_k^2\right] \frac{1}{2^{v/2+m-1}\Gamma(v/2)}
$$
(A.4)
$$
\cdot \int_0^{\sqrt{\rho}} dr\, r^{v+2m+1} \exp\left(-\frac{r^2}{2s}\right).
$$

The presence of an additional factor of u_k^2 in the angular average is harmless, since $|u_k^2| < 1$. We finally notice that the radial integral can be expressed in terms of a cumulative chi-square distribution function on replacing $r \to \sqrt{rs}$; namely,

$$
\int_0^{\sqrt{\rho}} dr\, r^{v+2m+1} \exp\left(-\frac{r^2}{2s}\right)
$$
(A.5)
$$
= 2^{v/2+m} s^{v/2+m+1} \Gamma\left(\frac{v}{2} + m + 1\right) F_{v+2(m+1)}\left(\frac{\rho}{s}\right).
$$

Inserting (A.5) into (A.4) results in Ruben's representation of α_k. This completes the first part of the proof.

As a next step, we wish to demonstrate that the function ψ_k of (61) is the generating function of the coefficients $c_{k;m}$. To this aim, we first recall the identities

$$
a^{-1/2} = (2\pi)^{-1/2} \int_{-\infty}^{\infty} dx \exp\left(-\frac{a}{2}x^2\right),
$$
(A.6)
$$
a^{-3/2} = (2\pi)^{-1/2} \int_{-\infty}^{\infty} dx\, x^2 \exp\left(-\frac{a}{2}x^2\right),
$$

valid for $a > 0$. On setting $a_i = [1 - (1 - s/\lambda_i)z]$, we see that ψ_k can be represented in the integral form

$$
\psi_k(z) = \left(\frac{s}{\lambda_k}\right)^{3/2} (2\pi)^{-v/2} \int_{-\infty}^{\infty} dx_k x_k^2
$$
$$
\cdot \exp\left(-\frac{1}{2}\left[1 - \left(1 - \frac{s}{\lambda_k}\right)z\right]x_k^2\right) \times \prod_{i\neq k}\left(\frac{s}{\lambda_i}\right)^{1/2}
$$
$$
\cdot \int_{-\infty}^{\infty} dx_i \exp\left(-\frac{1}{2}\left[1 - \left(1 - \frac{s}{\lambda_k}\right)z\right]x_i^2\right) = \frac{s}{\lambda_k}
$$
(A.7)
$$
\cdot \frac{s^{v/2}}{(2\pi)^{v/2}|\Lambda|^{1/2}} \int_{\mathbb{R}^v} d^v x\, x_k^2
$$
$$
\cdot \exp\left(-\frac{1}{2}zsQ(x) - \frac{x^T \cdot x}{2}\right),
$$

provided $|z| < \min_i |1 - s/\lambda_i|^{-1}$. As previously done, we introduce spherical coordinates $x = ru$ and expand $\exp\{-(1/2)zsQ(x)\} = \exp\{-(1/2)zsr^2Q(u)\}$ in Taylor series.

By the same argument as above, the series converges uniformly in u (the factor of zs does not depend on u), thus allowing term-by-term integration. Accordingly, we have

$$\psi_k(z) = \frac{s}{\lambda_k} \frac{s^{\nu/2}}{(2\pi)^{\nu/2} |\Lambda|^{1/2}} \sum_{m=0}^{\infty} z^m \frac{s^m}{2^m m!}$$

$$\cdot \int_0^{\infty} dr\, r^{\nu+2(m+1)-1} e^{-r^2/2} \qquad (A.8)$$

$$\cdot \int_{\partial \mathcal{B}_\nu(1)} du\, [-Q(u)]^m u_k^2.$$

We see that the *right-hand side* of (A.8) looks similar to (A.4), the only relevant differences being the presence of the factor of z^m under the sum sign and the upper limit of the radial integral. With some algebra, we arrive at

$$\psi_k(z) = \sum_{m=0}^{\infty} z^m \left\{ \frac{2}{m!} \frac{s}{\lambda_k} \frac{s^{\nu/2+m}}{|\Lambda|^{1/2}} \frac{\Gamma(\nu/2+m+1)}{\Gamma(\nu/2)} \right.$$

$$\left. \cdot \mathbb{M}\left[(-Q)^m u_k^2\right] \right\}. \qquad (A.9)$$

The series coefficients are recognized to be precisely those of (59).

In the last part of the proof, we derive the recursion fulfilled by the coefficients $c_{k;m}$. To this aim, the mth derivative of ψ_k has to be evaluated at $z = 0$ and then identified with $m! c_{k;m}$. The key observation is that differentiating ψ_k reproduces ψ_k itself; that is to say,

$$\psi_k'(z) = \Psi_k(z) \psi_k(z), \qquad (A.10)$$

with

$$\Psi_k(z) = \frac{1}{2} \sum_{i=1}^{\nu} e_{k;i} \left(1 - \frac{s}{\lambda_i}\right) \left[1 - \left(1 - \frac{s}{\lambda_i}\right)z\right]^{-1}, \qquad (A.11)$$

and the auxiliary coefficient $e_{k;i}$ being defined as in (67). Equation (A.10) lies at the origin of the recursion. Indeed, from (A.10), it follows that that ψ_k'' is a function of ψ_k' and ψ_k; namely, $\psi_k'' = \Psi_k' \psi_k + \Psi_k \psi_k'$. Proceeding analogously yields the general formula

$$\psi_k^{(m)}(z) = \sum_{r=0}^{m-1} \binom{m-1}{r} \Psi_k^{(m-r-1)}(z) \psi_k^{(r)}(z). \qquad (A.12)$$

At $z = 0$, this reads

$$m! c_{k;m} = \sum_{r=0}^{m-1} \frac{(m-1)!}{(m-r-1)! r!} \Psi_k^{(m-r-1)}(0) r! c_{k;r}. \qquad (A.13)$$

The last step consists in proving that

$$\Psi_k^{(m)}(0) = \frac{1}{2} m! g_{k;m+1}, \qquad (A.14)$$

with $g_{k;m}$ defined as in (65). This can be done precisely as explained in [2].

Having reiterated Ruben's proof explicitly in a specific case, it is now easy to see how the theorem is extended to any other Gaussian integral. First of all, from (A.1), we infer that each additional subscript in $\alpha_{k\ell m\cdots}$ enhances the power of the radial coordinate under the integral sign by 2 units. This entails a shift in the number of degrees of freedom of the chi-square distributions in Ruben's expansion, amounting to twice the number of subscripts. For instance, since α_{jk} has two subscripts, its Ruben's expansion starts by $F_{\nu+4}$, independently of whether $j = k$ or $j \neq k$. In second place, we observe that in order to correctly identify the generating functions of Ruben's coefficients for a higher-order integral $\alpha_{k\ell m\cdots}$, we need to take into account the multiplicities of the indices $k, \ell, m, \ldots$. As an example, consider the case of ψ_{jk} ($j \neq k$) and ψ_{kk}. By going once more through the argument presented in (A.7), we see that (A.6) are sufficient to show that (63) is the generating function of α_{jk}. By contrast, in order to repeat the proof for the case of ψ_{kk}, we need an additional integral identity; namely,

$$a^{-5/2} = \frac{1}{3} (2\pi)^{-1/2} \int_{-\infty}^{+\infty} dx\, x^4 \exp\left(-\frac{a}{2} x^2\right), \qquad (A.15)$$

valid once more for $a > 0$. Hence, we infer that ψ_{kk} must depend upon λ_k via a factor of $[1 - (1 - s/\lambda_k)z]^{-5/2}$, whereas ψ_{jk} ($j \neq k$) must depend on λ_j and λ_k via factors of, respectively, $[1 - (1 - s/\lambda_j)z]^{-3/2}$ and $[1 - (1 - s/\lambda_k)z]^{-3/2}$. The different exponents are ultimately responsible for the specific values taken by the auxiliary coefficients $e_{kk;i}$ and $e_{jk;i}$ of (68).

To conclude, we observe that the estimates of the residuals $\mathcal{R}_{k;m}$ and $\mathcal{R}_{jk;m}$, presented in Section 4 without an explicit proof, do not require any further technical insight than already provided by [2] plus our considerations. We leave them to the reader, since they can be obtained once more in the tracks of the original derivation of $\mathcal{R}_m$. □

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] G. M. Tallis, "Elliptical and radial truncation in normal populations," *The Annals of Mathematical Statistics*, vol. 34, no. 3, pp. 940–944, 1963.

[2] H. Ruben, "Probability content of regions under spherical normal distributions. IV. The distribution of homogeneous and non-homogeneous quadratic functions of normal variables," *The Annals of Mathematical Statistics*, vol. 33, pp. 542–570, 1962.

[3] J. Aitchison, *The Statistical Analysis of Compositional Data*, Monographs on Statistics and Applied Probability, Chapman and Hall, London, UK, 1986.

[4] V. Pawlowsky-Glahn and A. Buccianti, *Compositional Data Analysis: Theory and Applications*, John Wiley & Sons, 2011.

[5] J. J. Egozcue, V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal, "Isometric logratio transformations for compositional data analysis," *Mathematical Geology*, vol. 35, no. 3, pp. 279–300, 2003.

[6] J. Aitchison and S. M. Shen, "Logistic-normal distributions: some properties and uses," *Biometrika*, vol. 67, no. 2, pp. 261–272, 1980.

[7] V. Pawlowsky-Glahn and J. J. Egozcue, "Geometric approach to statistical analysis on the simplex," *Stochastic Environmental Research and Risk Assessment*, vol. 15, no. 5, pp. 384–398, 2001.

[8] P. Filzmoser and K. Hron, "Outlier detection for compositional data using robust methods," *Mathematical Geosciences*, vol. 40, no. 3, pp. 233–248, 2008.

[9] A. C. Atkinson, "Fast very robust methods for the detection of multiple outliers," *Journal of the American Statistical Association*, vol. 89, no. 428, pp. 1329–1339, 1994.

[10] M. Riani, A. C. Atkinson, and A. Cerioli, "Finding an unknown number of multivariate outliers," *Journal of the Royal Statistical Society B*, vol. 71, part 2, pp. 447–466, 2009.

[11] W. C. Guenther, "An easy method for obtaining percentage points of order statistics," *Technometrics*, vol. 19, no. 3, pp. 319–321, 1977.

[12] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems. Mathematics and Its Applications*, vol. 375, Springer, Berlin, Germany, 1996.

[13] L. Cavalier, "Inverse problems in statistics," in *Inverse Problems and High-Dimensional Estimation: Stats in the Château Summer School, August 31–September 4, 2009*, P. Alquier, E. Gautier, and G. Stoltz, Eds., vol. 203 of *Lecture Notes in Statistics*, pp. 3–96, Springer, Berlin, Germany, 2011.

[14] F. Palombi and S. Toti, "A note on the variance of the square components of a normal multivariate within a Euclidean ball," *Journal of Multivariate Analysis*, vol. 122, pp. 355–376, 2013.

[15] M. Anttila, K. Ball, and I. Perissinaki, "The central limit problem for convex bodies," *Transactions of the American Mathematical Society*, vol. 355, no. 12, pp. 4723–4735, 2003.

[16] J. O. Wojtaszczyk, "The square negative correlation property for generalized Orlicz balls," in *Proceedings of the Geometric Aspects of Functional Analysis, Israel Seminar*, pp. 305–313, 2004-2005.

[17] R. Mukerjee and S. H. Ong, "Variance and covariance inequalities for truncated joint normal distribution via monotone likelihood ratio and log-concavity," *Journal of Multivariate Analysis*, vol. 139, pp. 1–6, 2015.

[18] F. G. Friedlander and M. S. Joshi, *Introduction to the Theory of Distributions*, Cambridge University Press, Cambridge, UK, 2nd edition, 1998.

[19] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications, New York, 1964., New York, NY, USA, 1964.

[20] K. L. Judd, *Numerical Methods in Economics*, The MIT Press, Cambridge, Mass, USA, 1998.

[21] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, New York, NY, USA, 2nd edition, 1984.

[22] M. S. Bartlett, "XX.—on the theory of statistical regression," *Proceedings of the Royal Society of Edinburgh*, vol. 53, pp. 260–283, 1934.

[23] U. Grenander, "Some direct estimates of the mode," *The Annals of Mathematical Statistics*, vol. 36, no. 1, pp. 131–138, 1965.

[24] A. Zanella, M. Chiani, and M. Z. Win, "On the marginal distribution of the eigenvalues of wishart matrices," *IEEE Transactions on Communications*, vol. 57, no. 4, pp. 1050–1060, 2009.

[25] D. M. Young, *Iterative Solution of Large Linear Systems*, Computer Science and Applied Mathematics, Academic Press, 1971.

[26] J. Hadamard, "Sur les problèmes aux dérivés partielles et leur signification physique," *Princeton University Bulletin*, vol. 13, pp. 49–52, 1902.

[27] F. Palombi and S. Toti, "A perturbative approach to the reconstruction of the eigenvalue spectrum of a normal covariance matrix from a spherically truncated counterpart," https://arxiv.org/abs/1207.1256.

[28] G. Ponti, F. Palombi, D. Abate et al., "The role of medium size facilities in the HPC ecosystem: the case of the new CRESCO4 cluster integrated in the ENEAGRID infrastructure," in *Proceedings of the International Conference on High Performance Computing and Simulation (HPCS '14)*, 6903807, pp. 1030–1033, July 2014.

A Generalized Class of Exponential Type Estimators for Population Mean under Systematic Sampling Using Two Auxiliary Variables

Mursala Khan

Department of Mathematics, COMSATS Institute of Information Technology, Abbottabad 22060, Pakistan

Correspondence should be addressed to Mursala Khan; mursala.khan@yahoo.com

We have proposed a generalized class of exponential type estimators for population mean under the framework of systematic sampling using the knowledge of two auxiliary variables. The expressions for the mean square error of the proposed class of estimators have been corrected up to first order of approximation. Comparisons of the efficiency of the proposed class of estimators under the optimal conditions with the other existing estimators have been presented through a real secondary data. The statistical study provides strong evidence that the proposed class of estimators in survey estimation procedure results in substantial efficiency improvements over the other existing estimation approaches.

1. Introduction

In the literature of survey sampling, it is well known that the efficiencies of the estimators of the population parameters of the variable of interest can be increased by the use of auxiliary information related to auxiliary variable x, which is highly correlated with the variable of interest y. Auxiliary information may be efficiently utilized either at planning stage or at design stage to arrive at an improved estimator compared to those estimators, not utilizing auxiliary information. A simple technique of utilizing the known knowledge of the population parameters of the auxiliary variables is through ratio, product, and regression method of estimations using different probability sampling designs such as simple random sampling, stratified random sampling, cluster sampling, systematic sampling, and double sampling.

In the present paper we will use knowledge of the auxiliary variables under the framework of systematic sampling. Due to its simplicity, systematic sampling is the most commonly used probability design in survey of finite populations; see W. G. Madow and L. H. Madow [1]. Apart from its simplicity, systematic sampling provides estimators which are more efficient than simple random sampling or stratified random sampling for certain types of population; see Cochran [2], Gautschi [3], and Hajeck [4]. Later on the problem of estimating the population mean using information on auxiliary variable has also been discussed by various authors including Quenouille [5], Hansen et al. [6], Swain [7], Singh [8], Shukla [9], Srivastava and Jhajj [10], Kushwaha and Singh [11], Bahl and Tuteja [12], Banarasi et al. [13], H. P. Singh and R. Singh [14], Kadilar and Cingi [15], Koyuncu and Kadilar [16], Singh et al. [17], Singh and Solanki [18], Singh and Jatwa [19], Tailor et al. [20], Khan and Singh [21], and Khan and Abdullah [22].

Let us consider a finite population P of size N of distinct and identifiable units, $P_1, P_2, P_3, \ldots, P_N$ and number it from 1 to N unitsin some order. A random sample of size n units is selected from the first k units and then every kth subsequent unit is selected; thus there will be k samples (clusters), each of size n and observe the study variable y and auxiliary variable x for each and every unit selected in the sample. Let (y_{ij}, x_{ij}) for $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n$: denote the value of jth unit in the ith sample. Then the systematic sample means are defined as $\overline{y}^* = (1/n) \sum_{j=1}^{n} y_{ij}$ and $\overline{x}^* = (1/n) \sum_{j=1}^{n} x_{ij}$

are the unbiased estimators of the population means $\overline{Y} = (1/N)\sum_{j=1}^N y_{ij}$ and $\overline{X} = (1/N)\sum_{j=1}^N x_{ij}$, respectively.

Further let

$$\rho_y^* = \{1 + (n-1)\rho_y\},$$

$$\rho_x^* = \{1 + (n-1)\rho_x\}, \tag{1}$$

$$\rho_z^* = \{1 + (n-1)\rho_z\},$$

where

$$\rho_y = \frac{\left(y_{ij} - \overline{Y}\right)\left(y_{ij'} - \overline{Y}\right)}{E\left(y_{ij} - \overline{Y}\right)^2},$$

$$\rho_x = \frac{\left(x_{ij} - \overline{X}\right)\left(x_{ij'} - \overline{X}\right)}{E\left(x_{ij} - \overline{X}\right)^2}, \tag{2}$$

$$\rho_z = \frac{\left(z_{ij} - \overline{Z}\right)\left(z_{ij'} - \overline{Z}\right)}{E\left(z_{ij} - \overline{Z}\right)^2}$$

are the corresponding intraclass correlation coefficients for the study variable y and the auxiliary variables x and z, respectively.

Similarly $\rho_{yx} = S_{yx}/S_y S_x$, $\rho_{yz} = S_{yz}/S_y S_z$, and $\rho_{xz} = S_{xz}/S_x S_z$ are the correlation coefficients of the study and the auxiliary variables, respectively, where S_y, S_x, and S_z are the population standard deviation of study variable y and auxiliary variables x and z, respectively. Also S_{yx}, S_{yz}, and S_{xz} are the population covariances between y and x, y and z, and z and x, respectively. Also let C_y and C_x and C_z be the population coefficients of variation of the study and the auxiliary variables, respectively.

The variance of the classical estimator unbiased estimator $\overline{y}_1$ is given by

$$V\left(\overline{y}_1\right) = \theta\overline{Y}^2 \rho_y^* C_y^2, \tag{3}$$

where $\theta = ((N-1)/nN)$.

Swain [7] proposed a ratio estimator in systematic sampling given by

$$\overline{y}_2 = \overline{y}^* \left(\frac{\overline{X}}{\overline{x}^*}\right). \tag{4}$$

The mean squared error of the above estimator is as follows:

$$\text{MSE}\left(\overline{y}_2\right) = \theta\overline{Y}^2 \left[\rho_y^* C_y^2 + \rho_x^* C_x^2\left(1 - 2k\sqrt{\rho^{**}}\right)\right], \tag{5}$$

where $\rho^{**} = \rho_y^*/\rho_x^*$ and $k = \rho_{yx}C_y/C_x$.

Shukla [9] suggested the following product estimator for population mean of the study variable; the suggested estimator and their mean squared error are given as follows:

$$\overline{y}_3 = \overline{y}^* \exp\left(\frac{\overline{z}^*}{\overline{Z}}\right), \tag{6}$$

$$\text{MSE}\left(\overline{y}_3\right) = \theta\overline{Y}^2 \left[\rho_y^* C_y^2 + \rho_z^* C_z^2\left(1 + 2k^*\sqrt{\rho_2^{**}}\right)\right], \tag{7}$$

where $\rho_2^{**} = \rho_y^*/\rho_z^*$ and $k^* = \rho_{yz}C_y/C_z$.

The usual regression estimator for population mean under systematic sampling is given as follows:

$$\overline{y}_4 = \overline{y}^* + b_{yx}\left(\overline{X} - \overline{x}^*\right), \tag{8}$$

where b_{yx} is the sample regression coefficient between x and y.

The variance of the estimator $\overline{y}_4$, up to first order of approximation, is as follows:

$$\text{MSE}\left(\overline{y}_4\right) = \theta\rho_y^* S_y^2 \left[1 - \rho_{yx}^2\right]. \tag{9}$$

Singh et al. [17] recommended ratio-product type exponential estimators and are given by

$$\overline{y}_5 = \overline{y}^* \exp\left(\frac{\overline{X} - \overline{x}^*}{\overline{X} + \overline{x}^*}\right),$$

$$\overline{y}_6 = \overline{y}^* \exp\left(\frac{\overline{x}^* - \overline{X}}{\overline{x}^* + \overline{X}}\right). \tag{10}$$

The mean square errors of the Singh et al. [17], using first order of approximation, are given as follows:

$$\text{MSE}\left(\overline{y}_5\right) = \theta\overline{Y}^2 \left[\rho_y^* C_y^2 + \frac{\rho_x^* C_x^2}{4}\left(1 - 4k\sqrt{\rho^{**}}\right)\right], \tag{11}$$

$$\text{MSE}\left(\overline{y}_6\right) = \theta\overline{Y}^2 \left[\rho_y^* C_y^2 + \frac{\rho_x^* C_x^2}{4}\left(1 + 4k\sqrt{\rho^{**}}\right)\right]. \tag{12}$$

Tailor et al. [20] suggested a ratio-cum-product estimator for finite population mean; the recommended estimator and their first order mean square error are shown as follows:

$$\overline{y}_7 = \overline{y}^* \left(\frac{\overline{X}}{\overline{x}^*}\right)\left(\frac{\overline{z}^*}{\overline{Z}}\right),$$

$$\text{MSE}\left(\overline{y}_7\right) = \theta\overline{Y}^2 \left[\rho_y^* C_y^2 + \rho_x^* C_x^2\left(1 - 2k\sqrt{\rho^{**}}\right)\right. \tag{13}$$

$$\left. + \rho_z^* C_z^2\left(1 - 2k^{**}\sqrt{\rho_1^{**}}\right) + 2k^* C_z^2\sqrt{\rho_y^* \rho_z^*}\right],$$

where $\rho_1^{**} = \rho_x^*/\rho_z^*$.

2. The Generalized Class of Exponential Estimators

In this section, we have proposed a generalized class of exponential type estimators for population mean of the study variable y, under the framework of systematic sampling as given by

$$\overline{y}_p$$

$$= \overline{y}^* \exp\left(\frac{f\left(\overline{X} - \overline{x}^*\right)}{\overline{X} + (g-1)\overline{x}^*}\right)\exp\left(\frac{h\left(\overline{Z} - \overline{z}^*\right)}{\overline{Z} + (\eta-1)\overline{z}^*}\right), \tag{14}$$

where $-\infty < f < \infty$, $-\infty < h < \infty$, $g > 0$, and $\eta > 0$.

TABLE 1: Some members of the proposed class of estimators.

Estimator	Values of constants			
	f	h	g	η
$\overline{y}_{p1} = \overline{y}^*$ [simple estimator]	0	0	0	0
$\overline{y}_{p2} = \overline{y}^* \exp\left(\dfrac{\overline{X} - \overline{x}^*}{\overline{X} + \overline{x}^*}\right)$ Singh et al. [17]	1	0	2	0
$\overline{y}_{p3} = \overline{y}^* \exp\left(\dfrac{\overline{Z} - \overline{z}^*}{\overline{Z} + \overline{z}^*}\right)$	0	1	0	2
$\overline{y}_{p4} = \overline{y}^* \exp\left(\dfrac{\overline{X} - \overline{x}^*}{\overline{X}}\right)$	1	0	1	0
$\overline{y}_{p5} = \overline{y}^* \exp\left(\dfrac{\overline{Z} - \overline{z}^*}{\overline{Z}}\right)$	0	1	0	1
$\overline{y}_{p6} = \overline{y}^* \exp\left(\dfrac{\overline{X} - \overline{x}^*}{\overline{X} + (g-1)\overline{x}^*}\right)$	1	0	g	η
$\overline{y}_{p7} = \overline{y}^* \exp\left(\dfrac{\overline{Z} - \overline{z}^*}{\overline{Z} + (\eta-1)\overline{z}^*}\right)$	0	1	g	η
$\overline{y}_{p8} = \overline{y}^* \exp\left(\dfrac{f\left(\overline{X} - \overline{x}^*\right)}{\overline{X} + \overline{x}^*}\right)$	f	0	2	η
$\overline{y}_{p9} = \overline{y}^* \exp\left(\dfrac{h\left(\overline{Z} - \overline{z}^*\right)}{\overline{Z} + \overline{z}^*}\right)$	0	h	g	2
$\overline{y}_{p10} = \overline{y}^* \exp\left(\dfrac{\overline{X} - \overline{x}^*}{\overline{X} + \overline{x}^*}\right)\exp\left(\dfrac{\overline{Z} - \overline{z}^*}{\overline{Z} + \overline{z}^*}\right)$	1	1	g	η
$\overline{y}_{p11} = \overline{y}^* \exp\left(\dfrac{\overline{X} - \overline{x}^*}{\overline{X}}\right)\exp\left(\dfrac{\overline{Z} - \overline{z}^*}{\overline{Z}}\right)$	1	1	1	1
$\overline{y}_{p12} = \overline{y}^* \exp\left(\dfrac{\overline{x}^* - \overline{X}}{\overline{x}^* + \overline{X}}\right)$ Singh et al. [17]	-1	0	2	η
$\overline{y}_{p13} = \overline{y}^* \exp\left(\dfrac{\overline{z}^* - \overline{Z}}{\overline{z}^* + \overline{Z}}\right)$	0	-1	g	2
$\overline{y}_{p14} = \overline{y}^* \exp\left(\dfrac{\overline{x}^* - \overline{X}}{\overline{x}^* + \overline{X}}\right)\exp\left(\dfrac{\overline{z}^* - \overline{Z}}{\overline{z}^* + \overline{Z}}\right)$	-1	-1	2	2
$\overline{y}_{p15} = \overline{y}^* \exp\left(\dfrac{\overline{X} - \overline{x}^*}{\overline{X} + \overline{x}^*}\right)\exp\left(\dfrac{\overline{z}^* - \overline{Z}}{\overline{z}^* + \overline{Z}}\right)$	1	-1	2	2
$\overline{y}_{p16} = \overline{y}^* \exp\left(\dfrac{\overline{x}^* - \overline{X}}{\overline{x}^* + \overline{X}}\right)\exp\left(\dfrac{\overline{Z} - \overline{z}^*}{\overline{Z} + \overline{z}^*}\right)$	-1	1	2	2
$\overline{y}_{p17} = \overline{y}^* \exp\left(\dfrac{\overline{X} - \overline{x}^*}{\overline{X}}\right)\exp\left(\dfrac{\overline{z}^* - \overline{Z}}{\overline{Z}}\right)$	1	-1	1	1
$\overline{y}_{p18} = \overline{y}^* \exp\left(\dfrac{\overline{x}^* - \overline{X}}{\overline{X}}\right)\exp\left(\dfrac{\overline{Z} - \overline{z}^*}{\overline{Z}}\right)$	-1	1	1	1

A set of some new and known members of the generalized class of exponential estimators generated from (14) for some suitable values of f, h, g, and η are listed in Table 1.

To obtain the properties of the proposed class of estimators up to first-order approximation, we use the following relative errors, symbols, and notations:

$$\psi_0 = \frac{\overline{y}^* - \overline{Y}}{\overline{Y}},$$

$$\psi_1 = \frac{\overline{x}^* - \overline{X}}{\overline{X}}, \qquad (15)$$

$$\psi_2 = \frac{\overline{z}^* - \overline{Z}}{\overline{Z}},$$

such that

$$E(\psi_0) = E(\psi_1) = E(\psi_2) = 0; \qquad (16)$$

also

$$E\left(\psi_0^2\right) = \theta \rho_y^* C_y^2,$$

$$E\left(\psi_1^2\right) = \theta \rho_x^* C_x^2,$$

$$E\left(\psi_2^2\right) = \theta \rho_z^* C_z^2,$$

$$E(\psi_0 \psi_1) = \theta k C_x^2 \sqrt{\rho_y^* \rho_x^*},$$

$$E\left(\psi_0\psi_2\right) = \theta k^* C_z^2 \sqrt{\rho_y^* \rho_z^*},$$

$$E\left(\psi_1\psi_2\right) = \theta k^{**} C_z^2 \sqrt{\rho_x^* \rho_z^*}.$$

$$(17)$$

Expanding (14) in terms of ψ's up to the first order of approximation, we have

$$\bar{y}_p = \bar{Y}\left(1 + \psi_0\right) \exp\left(\frac{-f\psi_1}{1 + (g-1)\left(1 + \psi_1\right)}\right)$$

$$\cdot \exp\left(\frac{-h\psi_2}{1 + (\eta - 1)\left(1 + \psi_2\right)}\right). \qquad (18)$$

Further simplify

$$\bar{y}_p - \bar{Y} = \bar{Y}\left[\psi_0 - \frac{f}{g}\psi_1 - \frac{h}{\eta}\psi_2 - \frac{f}{g}\psi_0\psi_1 - \frac{h}{\eta}\psi_0\psi_2 \right.$$

$$\left. + \delta_1\psi_1^2 + \delta_2\psi_2^2 + \delta_3\psi_1\psi_2\right], \qquad (19)$$

where $\delta_1 = f/g - f/g^2 + f^2/2g^2$, $\delta_2 = h/\eta - h/\eta^2 + h^2/2\eta^2$, and $\delta_3 = fh/g\eta$.

On squaring and taking expectation on both sides of (19), we get the mean square error of $\bar{y}_p$, up to the first degree of approximation, as

$$\text{MSE}\left(\bar{y}_p\right) = \theta\bar{Y}^2\left[\rho_y^* C_y^2 + \lambda_1^2 \rho_x^* C_x^2 + \lambda_2^2 \rho_z^* C_z^2\right.$$

$$- 2\lambda_1 k C_x^2 \sqrt{\rho_y^* \rho_x^*} - 2\lambda_2 k^* C_z^2 \sqrt{\rho_y^* \rho_z^*} \qquad (20)$$

$$\left. + 2\lambda_1\lambda_2 k^{**} C_z^2 \sqrt{\rho_x^* \rho_z^*}\right],$$

where $\lambda_1 = f/g$ and $\lambda_2 = h/\eta$.

By partially differentiating (20) with respect to λ_1 and λ_2, we get the optimum value of λ_1 and λ_2 as given by

$$\lambda_1 = \frac{\delta_1 \sqrt{\rho^{**}}}{\delta_2},$$

$$\lambda_2 = \frac{\delta_3 C_x^2 \sqrt{\rho_2^{**}}}{\delta_2}, \qquad (21)$$

where $\delta_1 = kC_x^2 - k^* k^{**} C_z^2$, $\delta_2 = C_x^2 - k^{2**} C_z^2$, and $\delta_3 = k^* - kk^{**}$.

Substituting the optimal values of λ_1 and λ_2 in (20) we obtain the minimum mean square error of the estimator $\bar{y}_m$ as follows:

$$\text{MSE}\left(\bar{y}_p\right)_{\min} = \theta\bar{Y}^2 \rho_y^* \left[C_y^2\right.$$

$$+ \frac{C_x^2}{\delta_2^2}\left\{\left(\delta_1^2 + \delta_3^2 C_x^2 C_z^2 + 2k^{**} C_z^2 \delta_1\delta_3\right)\right. \qquad (22)$$

$$\left.\left. - 2\delta_2\left(k\delta_1 + k^* C_z^2 \delta_3\right)\right\}\right].$$

3. Comparison of Efficiency

In this section, we have found some theoretical efficiencies conditions under which the proposed estimator performs better than the other relevant existing estimators by comparing the generalized class of exponential type estimators with other existing estimators.

(i) By (22) and (3), $\text{MSE}(\bar{y}_p) \leq \text{MSE}(\bar{y}_1)$ if

$$\left[2\delta_1\delta_2 k + 2\delta_2\delta_3 k^* C_z^2 - \delta_1^2 - \delta_3^2 C_x^2 C_z^2 - 2k^{**} C_z^2 \delta_1\delta_3\right]$$

$$\geq 0. \qquad (23)$$

(ii) By (22) and (5), $\text{MSE}(\bar{y}_p) \leq \text{MSE}(\bar{y}_2)$ if

$$\left[\delta_2^2 \rho_x^* \left(1 - 2k\sqrt{\rho^{**}}\right) - \rho_y^*\left(\delta_1^2 + \delta_3^2 C_x^2 C_z^2\right.\right.$$

$$\left.\left. + 2\delta_1\delta_3 k^{**} C_z^2 - 2\delta_1\delta_2 k - 2\delta_2\delta_3 k^* C_z^2\right)\right] \geq 0. \qquad (24)$$

(iii) By (22) and (7), $\text{MSE}(\bar{y}_p) \leq \text{MSE}(\bar{y}_3)$ if

$$\left[\rho_z^* C_z^2 \delta_2^2 \left(1 + 2k^* \sqrt{\rho_2^{**}}\right) - \rho_y^* C_x^2\left(\delta_1^2 + \delta_3^2 C_x^2 C_z^2\right.\right.$$

$$\left.\left. + 2k^{**} C_z^2 \delta_1\delta_3 - 2k\delta_1\delta_2 - 2\delta_2\delta_3 k^* C_z^2\right)\right] \geq 0. \qquad (25)$$

(iv) By (22) and (9), $\text{MSE}(\bar{y}_p) \leq \text{MSE}(\bar{y}_4)$ if

$$\left[C_x^2\left(2k\delta_1\delta_2 + 2\delta_2\delta_3 k^* C_z^2 - \delta_1^2 - \delta_3^2 C_x^2 C_z^2\right.\right.$$

$$\left.\left. - 2k^{**} C_z^2 \delta_1\delta_3\right) - \rho_{yx}^2 \delta_2^2 C_y^2\right] \geq 0. \qquad (26)$$

(v) By (22) and (11), $\text{MSE}(\bar{y}_p) \leq \text{MSE}(\bar{y}_5)$ if

$$\left[\frac{\rho_x^*}{4}\left(1 - 4k\sqrt{\rho^{**}}\right) - \frac{\rho_y^*}{\delta_2^2}\left(\delta_1^2 + \delta_3^2 C_x^2 C_z^2\right.\right.$$

$$\left.\left. + 2k^{**} C_z^2 \delta_1\delta_3 - 2k\delta_1\delta_2 - 2\delta_2\delta_3 k^* C_z^2\right)\right] \geq 0. \qquad (27)$$

(vi) By (22) and (12), $\text{MSE}(\bar{y}_p) \leq \text{MSE}(\bar{y}_6)$ if

$$\left[\frac{\rho_x^*}{4}\left(1 + 4k\sqrt{\rho^{**}}\right) - \frac{\rho_y^*}{\delta_2^2}\left(\delta_1^2 + \delta_3^2 C_x^2 C_z^2\right.\right.$$

$$\left.\left. + 2k^{**} C_z^2 \delta_1\delta_3 - 2k\delta_1\delta_2 - 2\delta_2\delta_3 k^* C_z^2\right)\right] \geq 0. \qquad (28)$$

(vii) By (22) and (13), $\text{MSE}(\bar{y}_p) \leq \text{MSE}(\bar{y}_7)$ if

TABLE 2: The mean square errors (MSEs) of the estimators and the percent relative efficiencies (PREs) with respect to $\overline{y}_1$.

Estimator	Population MSE $(\overline{y}_*)$	PRE $(\overline{y}_*, \overline{y}_1)$
$\overline{y}_1$	1455.08	100.00
$\overline{y}_2$	373.32	389.62
$\overline{y}_3$	768.06	189.45
$\overline{y}_4$	43.74	3326.66
$\overline{y}_5$	820.09	177.43
$\overline{y}_6$	1044.42	139.32
$\overline{y}_7$	187.08	777.79
$\overline{y}_p$	23.63	6158.08

$$\rho_y^* \leq \frac{\delta_2^2 \left[\rho_x^* C_x^2 \left(1 - 2k\sqrt{\rho^{**}}\right) + \rho_z^* C_z^2 \left(1 - 2k^{**}\sqrt{\rho_1^{**}}\right) + 2k^* C_z^2 \sqrt{\rho_y^* \rho_z^*} \right]}{C_x^2 \left(\delta_1^2 + \delta_3^2 C_x^2 C_z^2 + 2k^{**}C_z^2 \delta_1 \delta_3 - 2k\delta_1\delta_2 - 2\delta_2\delta_3 k^* C_z^2\right)}. \tag{29}$$

4. Empirical Study

To examine the merits of the proposed estimator over the other existing estimators at optimum conditions, we have considered natural population data sets from the literature. The sources of population are given as follows.

Population (Source: Tailor et al. [20]). Consider

$$N = 15,$$

$$n = 3,$$

$$\overline{X} = 44.47,$$

$$\overline{Y} = 80,$$

$$\overline{Z} = 48.40,$$

$$C_y = 0.56,$$

$$C_x = 0.28,$$

$$C_z = 0.43,$$

$$S_y^2 = 2000,$$

$$S_x^2 = 149.55,$$

$$S_z^2 = 427.83,$$

$$S_{yx} = 538.57,$$

$$S_{yz} = -902.86,$$

$$S_{xz} = -241.06,$$

$$\rho_{yx} = 0.9848,$$

$$\rho_{yz} = -0.9760,$$

$$\rho_{xz} = -0.9530,$$

$$\rho_y = 0.6652,$$

$$\rho_x = 0.707,$$

$$\rho_z = 0.5487. \tag{30}$$

The percent relative efficiencies (PREs) of the stated estimators with respect to the usual unbiased estimator are obtained from the following mathematical formula:

$$\text{PRE}\left(\overline{y}_*, \overline{y}_1\right) = \frac{\text{MSE}\left(\overline{y}_1\right)}{\text{MSE}\left(\overline{y}_*\right)} \times 100, \tag{31}$$

where $* = 1, 2, 3, 4, 5, 6, 7,$ and p.

5. Conclusion

In this paper we proposed a generalized class of exponential type estimators for the population mean of study variable y, when information is available on two auxiliary variables under the framework of systematic sampling scheme. The properties of the proposed estimator are derived up to first order of approximation. The proposed estimator is compared with other present estimators, both as theoretical and empirical efficiency comparisons. We have also judged the performance of the proposed estimator for a known natural population dataset; see Tailor et al. [20]. Results are given in Table 2 which shows that performances of the proposed generalized class of exponential type estimator are more efficient than the other existing estimators by smaller mean square errors and the higher percent relative efficiencies of the estimators. Hence it is preferable to use the proposed estimator in practical surveys.

Competing Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

References

[1] W. G. Madow and L. H. Madow, "On the theory of systematic sampling. I," *Annals of Mathematical Statistics*, vol. 15, pp. 1–24, 1944.

[2] W. G. Cochran, "Relative accuracy of systematic and stratified random samples for a certain class of populations," *The Annals of Mathematical Statistics*, vol. 17, pp. 164–177, 1946.

[3] W. Gautschi, "Some remarks on systematic sampling," *Annals of Mathematical Statistics*, vol. 28, pp. 385–394, 1957.

[4] J. Hajeck, "Optimum strategy and other problems in probability sampling," *Casopis pro Pestovani Matematiky*, vol. 84, pp. 387–423, 1959.

[5] M. H. Quenouille, "Notes on bias in estimation," *Biometrika*, vol. 43, pp. 353–360, 1956.

[6] M. H. Hansen, W. N. Hurwitz, and M. Gurney, "Problems and methods of the sample survey of business," *Journal of the American Statistical Association*, vol. 41, no. 234, pp. 173–189, 1946.

[7] A. K. P. C. Swain, "The use of systematic sampling ratio estimate," *Journal of the Indian Statistical Association*, vol. 2, pp. 160–164, 1964.

[8] M. P. Singh, "Ratio-cum-product method of estimation," *Metrika*, vol. 12, pp. 34–42, 1967.

[9] N. D. Shukla, "Systematic sampling and product method of estimation," in *Proceeding of the all India Seminar on Demography and Statistics*, Varanasi, India, 1971.

[10] S. K. Srivastava and H. S. Jhajj, "A class of estimators of the population mean using multi-auxiliary information," *Calcutta Statistical Association Bulletin*, vol. 32, no. 125-126, pp. 47–56, 1983.

[11] K. S. Kushwaha and H. P. Singh, "Class of almost unbiased ratio and product estimators in systematic sampling," *Journal of the Indian Society of Agricultural Statistics*, vol. 41, no. 2, pp. 193–205, vi, 1989.

[12] S. Bahl and R. K. Tuteja, "Ratio and product type exponential estimators," *Journal of Information & Optimization Sciences*, vol. 12, no. 1, pp. 159–164, 1991.

[13] Banarasi, S. N. S. Kushwaha, and K. S. Kushwaha, "A class of ratio, product and difference (R.P.D.) estimators in systematic sampling," *Microelectronics Reliability*, vol. 33, no. 4, pp. 455–457, 1993.

[14] H. P. Singh and R. Singh, "Almost unbiased ratio and product type estimators in systematic sampling," *Questiio*, vol. 22, no. 3, pp. 403–416, 1998.

[15] C. Kadilar, and H. Cingi, "An improvement in estimating the population mean by using the correlation coefficient," *Hacettepe Journal of Mathematics and Statistics*, vol. 35, no. 1, pp. 103–109, 2006.

[16] N. Koyuncu and C. Kadilar, "Family of estimators of population mean using two auxiliary variables in stratified random sampling," *Communications in Statistics: Theory and Methods*, vol. 38, no. 13–15, pp. 2398–2417, 2009.

[17] H. P. Singh, R. Tailor, and N. K. Jatwa, "Modified ratio and product estimators for population mean in systematic sampling," *Journal of Modern Applied Statistical Methods*, vol. 10, no. 2, pp. 424–435, 2011.

[18] H. P. Singh and R. S. Solanki, "An efficient class of estimators for the population mean using auxiliary information in systematic sampling," *Journal of Statistical Theory and Practice*, vol. 6, no. 2, pp. 274–285, 2012.

[19] H. P. Singh and N. K. Jatwa, "A class of exponential type estimators in systematic sampling," *Economic Quality Control*, vol. 27, no. 2, pp. 195–208, 2012.

[20] T. Tailor, N. Jatwa, and H. P. Singh, "A ratio-cum-product estimator of finite population mean in systematic sampling," *Statistics in Transition*, vol. 14, no. 3, pp. 391–398, 2013.

[21] M. Khan and R. Singh, "Estimation of population mean in chain ratio-type estimator under systematic sampling," *Journal of Probability and Statistics*, vol. 2015, Article ID 248374, 2 pages, 2015.

[22] M. Khan and H. Abdullah, "A note on a differencetype estimator for population mean under twophase sampling design," *SpringerPlus*, vol. 5, article 723, 7 pages, 2016.

Stochastic Restricted Biased Estimators in Misspecified Regression Model with Incomplete Prior Information

Manickavasagar Kayanan (iD) **[1,2] and Pushpakanthie Wijekoon[3]**

[1]*Department of Physical Science, Vavuniya Campus, University of Jaffna, Vavuniya, Sri Lanka*
[2]*Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka*
[3]*Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka*

Correspondence should be addressed to Manickavasagar Kayanan; mgayanan@vau.jfn.ac.lk

The analysis of misspecification was extended to the recently introduced stochastic restricted biased estimators when multicollinearity exists among the explanatory variables. The Stochastic Restricted Ridge Estimator (SRRE), Stochastic Restricted Almost Unbiased Ridge Estimator (SRAURE), Stochastic Restricted Liu Estimator (SRLE), Stochastic Restricted Almost Unbiased Liu Estimator (SRAULE), Stochastic Restricted Principal Component Regression Estimator (SRPCRE), Stochastic Restricted r-k (SRrk) class estimator, and Stochastic Restricted r-d (SRrd) class estimator were examined in the misspecified regression model due to missing relevant explanatory variables when incomplete prior information of the regression coefficients is available. Further, the superiority conditions between estimators and their respective predictors were obtained in the mean square error matrix (MSEM) sense. Finally, a numerical example and a Monte Carlo simulation study were used to illustrate the theoretical findings.

1. Introduction

Misspecification due to left out relevant explanatory variables is very often when considering the linear regression model, which causes these variables to become a part of the error term. Consequently, the expected value of error term of the model will not be zero. Also, the omitted variables may be correlated with the variables in the model. Therefore, one or more assumptions of the linear regression model will be violated when the model is misspecified, and hence the estimators become biased and inconsistent. Further, it is well-known that the ordinary least squares estimator (OLSE) may not be very reliable if multicollinearity exists in the linear regression model. As a remedial measure to solve multicollinearity problem, biased estimators based on the sample model $y = X\beta + \varepsilon$ with prior information which can be exact or stochastic restrictions have received much attention in the statistical literature. The intention of this work is to examine the performance of the recently introduced stochastic restricted biased estimators in the misspecified regression model with incomplete prior knowledge about

regression coefficients when there exists multicollinearity among explanatory variables.

When we consider the biased estimation in misspecified regression model without any restrictions on regression parameters, Sarkar [1] discussed the consequences of exclusion of some important explanatory variables from a linear regression model when multicollinearity exists. Şiray [2] and Wu [3] examined the efficiency of the r-d class estimator and r-k class estimator, respectively, in the misspecified regression model. Chandra and Tyagi [4] studied the effect of misspecification due to the omission of relevant variables on the dominance of the r-(k, d) class estimator. Recently, Kayanan and Wijekoon [5] examined the performance of existing biased estimators and the respective predictors based on the sample information in a misspecified linear regression model without considering any prior information about regression coefficients.

It is recognized that the mixed regression estimator (MRE) introduced by Theil and Goldberger [6] outperforms ordinary least squares estimator (OLSE) when the regression

model is correctly specified. The biased estimation with stochastic linear restrictions in the misspecified regression model due to inclusion of an irrelevant variable with the incorrectly specified prior information was discussed by Teräsvirta [7]. Later Mittelhammer [8], Ohtani and Honda [9], Kadiyala [10], and Trenkler and Wijekoon [11] discussed the efficiency of MRE under misspecified regression model due to exclusion of a relevant variable with correctly specified prior information. Further, the superiority of MRE over the OLSE under the misspecified regression model with incorrectly specified sample and prior information was discussed by Wijekoon and Trenkler [12]. Hubert and Wijekoon [13] have considered the improvement of Liu estimator (LE) under a misspecified regression model with stochastic restrictions and introduced the Stochastic Restricted Liu Estimator (SRLE).

In this paper, the performance of the recently introduced stochastic restricted estimators, namely, the Stochastic Restricted Ridge Estimator (SRRE) proposed by Li and Yang [14], Stochastic Restricted Almost Unbiased Ridge Estimator (SRAURE), and Stochastic Restricted Almost Unbiased Liu Estimator (SRAULE) proposed by Wu and Yang [15], Stochastic Restricted Principal Component Regression Estimator (SRPCRE) proposed by He and Wu [16], Stochastic Restricted r-k (SRrk) class estimator, and Stochastic Restricted r-d (SRrd) class estimator proposed by Wu [17], was examined in the misspecified regression model when multicollinearity exists among explanatory variables. Further, a generalized form to represent these estimators is also proposed.

The rest of this article is organized as follows. The model specification and the estimators are written in Section 2. In Section 3, the mean square error matrix (MSEM) comparison between two estimators and respective predictors is considered. In Section 4, a numerical example and a Monte Carlo simulation study are given to illustrate the theoretical results in Scalar Mean Square Error (SMSE) criterion. Finally, some concluding remarks are mentioned in Section 5. The references and appendixes are given at the end of the paper.

2. Model Specification and the Estimators

Assume that the true regression model is given by

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon = X_1\beta_1 + \delta + \varepsilon, \tag{1}$$

where y is the $n \times 1$ vector of observations on the dependent variable, X_1 and X_2 are the $n \times l$ and $n \times p$ matrices of observations on the $m = l + p$ regressors, β_1 and β_2 are the $l \times 1$ and $p \times 1$ vectors of unknown coefficients, and ε is the $n \times 1$ vector of disturbances such that $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \Omega = \sigma^2 I$.

Let us say that the researcher misspecifies the regression model by excluding p regressors as

$$y = X_1\beta_1 + u. \tag{2}$$

Let us also assume that there exists prior information on β_1 in the form of

$$r = R\beta_1 + g + v, \tag{3}$$

where r is the $q \times 1$ vector, R is the given $q \times l$ matrix with rank q, g is the $q \times 1$ unknown fixed vector, v is the $q \times 1$ vector of disturbances such that $E(v) = 0$, $D(v) = E(vv') = \Psi = \sigma^2 W$, where W is positive definite, and $E(vu') = 0$

By combining sample model (2) and prior information (3), Theil and Goldberger [6] proposed the mixed regression estimator (MRE) as

$$\begin{aligned}\widehat{\beta}_{\text{MRE}} &= \left(X_1'\Omega^{-1}X_1 + R'\Psi^{-1}R\right)^{-1}\left(X_1'\Omega^{-1}y + R'\Psi^{-1}r\right) \\ &= \left(X_1'X_1 + R'W^{-1}R\right)^{-1}\left(X_1'y + R'W^{-1}r\right).\end{aligned} \tag{4}$$

To combat multicollinearity, several researchers introduce different types of stochastic restricted estimators in place of MRE. Seven such estimators are SRRE, SRAURE, SRLE, SRALUE, SRPCRE, SRrk class estimator, and SRrd class estimator defined below, respectively:

$$\widehat{\beta}_{\text{SRRE}} = \left(X_1'X_1 + kI\right)^{-1} X_1'X_1\widehat{\beta}_{\text{MRE}}$$

$$\widehat{\beta}_{\text{SRAURE}} = \left(I - k^2\left(X_1'X_1 + kI\right)^{-2}\right)\widehat{\beta}_{\text{MRE}}$$

$$\widehat{\beta}_{\text{SRLE}} = \left(X_1'X_1 + I\right)^{-1}\left(X_1'X_1 + dI\right)\widehat{\beta}_{\text{MRE}}$$

$$\widehat{\beta}_{\text{SRAULE}} = \left(I - (1-d)^2\left(X_1'X_1 + I\right)^{-2}\right)\widehat{\beta}_{\text{MRE}} \tag{5}$$

$$\widehat{\beta}_{\text{SRPCRE}} = T_hT_h'\widehat{\beta}_{\text{MRE}}$$

$$\widehat{\beta}_{\text{SRrk}} = T_hT_h'\left(X_1'X_1 + kI\right)^{-1}X_1'X_1\widehat{\beta}_{\text{MRE}}$$

$$\widehat{\beta}_{\text{SRrd}} = T_hT_h'\left(X_1'X_1 + I\right)^{-1}\left(X_1'X_1 + dI\right)\widehat{\beta}_{\text{MRE}},$$

where $k > 0, 0 < d < 1$, and $T_h = (t_1, t_2, \ldots, t_h)$ are the first h columns of $T = (t_1, t_2, \ldots, t_h, \ldots, t_l)$ which is an orthogonal matrix of the standardized eigenvectors of $X_1'X_1$.

According to Kadiyala [10], now we apply the simultaneous decomposition to the two symmetric matrices $X_1'X_1$ and $R'\Psi^{-1}R$, as

$$B'X_1'X_1B = I,$$

$$B'R'\Psi^{-1}RB = \Lambda, \tag{6}$$

where $X_1'X_1$ is a positive definite matrix and $R'\Psi^{-1}R$ is a positive semidefinite matrix, B is a $l \times l$ nonsingular matrix, and Λ is a $l \times l$ diagonal matrix with eigenvalues $\lambda_i > 0$ for $i = 1, 2, \ldots, q$ and $\lambda_i = 0$ for $i = q + 1, \ldots, l$.

Let $X_* = X_1B$, $R_* = RB$, $\gamma = B^{-1}\beta_1$, $X_*'X_* = I$, and $R_*'\Psi^{-1}R_* = \Lambda$; then the models (1), (2), and (3) can be written as

$$y = X_*\gamma + \delta + \varepsilon, \tag{7}$$

$$y = X_*\gamma + u, \tag{8}$$

$$r = R_*\gamma + g + v. \tag{9}$$

According to Wijekoon and Trenkler [12], the corresponding MRE is given by

$$\hat{\gamma}_{\text{MRE}} = \left(X'_* X_* + R'_* \Psi^{-1} R_*\right)^{-1} \left(X'_* y + R'_* W^{-1} r\right)$$
$$= \left(I + \sigma^2 \Lambda\right)^{-1} \left(X'_* y + R'_* W^{-1} r\right). \tag{10}$$

Hence, the respective expectation vector, bias vector, and dispersion matrix are given by

$$E\left(\hat{\gamma}_{\text{MRE}}\right) = \gamma + \left(I + \sigma^2 \Lambda\right)^{-1} \left(X'_* \delta + R'_* W^{-1} g\right),$$

$$\text{Bias}\left(\hat{\gamma}_{\text{MRE}}\right) = \left(I + \sigma^2 \Lambda\right)^{-1} \left(X'_* \delta + R'_* W^{-1} g\right), \tag{11}$$

$$D\left(\hat{\gamma}_{\text{MRE}}\right) = \sigma^2 \left(I + \sigma^2 \Lambda\right)^{-1}.$$

In the case of misspecification, now the SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, and SRrd for model (7) can be written as

$$\hat{\gamma}_{\text{SRRE}} = \left(X'_* X_* + kI\right)^{-1} X'_* X_* \hat{\gamma}_{\text{MRE}}$$
$$= (1 + k)^{-1} \hat{\gamma}_{\text{MRE}} = C_k \hat{\gamma}_{\text{MRE}}$$

$$\hat{\gamma}_{\text{SRAURE}} = \left(I - k^2 \left(X'_* X_* + kI\right)^{-2}\right) \hat{\gamma}_{\text{MRE}}$$
$$= (1 + k)^{-2} (1 + 2k) \hat{\gamma}_{\text{MRE}}$$
$$= (1 + 2k) (C_k)^2 \hat{\gamma}_{\text{MRE}} = C_k^* \hat{\gamma}_{\text{MRE}}$$

$$\hat{\gamma}_{\text{SRLE}} = \left(X'_* X_* + I\right)^{-1} \left(X'_* X_* + dI\right) \hat{\gamma}_{\text{MRE}}$$
$$= 2^{-1} (1 + d) \hat{\gamma}_{\text{MRE}} = C_d \hat{\gamma}_{\text{MRE}}$$

$$\hat{\gamma}_{\text{SRAULE}} = \left(I - (1 - d)^2 \left(X'_* X_* + I\right)^{-2}\right) \hat{\gamma}_{\text{MRE}} \tag{12}$$
$$= 2^{-2} (1 + d) (3 - d) \hat{\gamma}_{\text{MRE}}$$
$$= 2^{-1} (3 - d) C_d \hat{\gamma}_{\text{MRE}} = C_d^* \hat{\gamma}_{\text{MRE}}$$

$$\hat{\gamma}_{\text{SRPCRE}} = T_h T'_h \hat{\gamma}_{\text{MRE}} = C_h \hat{\gamma}_{\text{MRE}}$$

$$\hat{\gamma}_{\text{SRrk}} = (1 + k)^{-1} T_h T'_h \hat{\gamma}_{\text{MRE}} = C_k C_h \hat{\gamma}_{\text{MRE}}$$
$$= C_{hk} \hat{\gamma}_{\text{MRE}}$$

$$\hat{\gamma}_{\text{SRrd}} = 2^{-1} (1 + d) T_h T'_h \hat{\gamma}_{\text{MRE}} = C_d C_h \hat{\gamma}_{\text{MRE}}$$
$$= C_{hd} \hat{\gamma}_{\text{MRE}},$$

respectively, where $C_k = (1 + k)^{-1}$, $C_k^* = (1 + 2k)(C_k)^2$, $C_d = 2^{-1}(1 + d)$, $C_d^* = 2^{-1}(3 - d)C_d$, $C_h = T_r T'_r$, $C_{hk} = C_k C_h$, and $C_{hd} = C_d C_h$.

It is clear that C_k, C_k^*, C_d, and C_d^* are positive definite and C_h, C_{hk}, and C_{hd} are nonnegative definite.

Since all these estimators can be written by incorporating $\hat{\gamma}_{\text{MRE}}$, now we write a generalized form to represent SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, and SRrd as given below:

$$\hat{\gamma}_{(j)} = G_{(j)} \hat{\gamma}_{\text{MRE}}, \tag{13}$$

where $G_{(j)}$ is positive definite matrix if it stands for C_k, C_k^*, C_d, and C_d^*, and it is nonnegative definite matrix if it stands for C_h, C_{hk}, and C_{hd}.

Now the expectation vector, bias vector, the dispersion matrix, and the mean square error matrix can be written as

$$E\left(\hat{\gamma}_{(j)}\right) = G_{(j)} E\left(\hat{\gamma}_{\text{MRE}}\right)$$

$$= G_{(j)} \left(\gamma + \left(I + \sigma^2 \Lambda\right)^{-1} \left(X'_* \delta + R'_* W^{-1} g\right)\right)$$

$$= G_{(j)} (\gamma + \tau A)$$

$$\text{Bias}\left(\hat{\gamma}_{(j)}\right) = E\left(\hat{\gamma}_{(j)} - \gamma\right) = G_{(j)} (\gamma + \tau A) - \gamma$$

$$= \left(G_{(j)} - I\right) \gamma + G_{(j)} \tau A$$
$$\tag{14}$$

$$D\left(\hat{\gamma}_{(j)}\right) = G_{(j)} D\left(\hat{\gamma}_{\text{MRE}}\right) G'_{(j)} = \sigma^2 G_{(j)} \left(I + \sigma^2 \Lambda\right)^{-1}$$

$$\cdot G'_{(j)} = \sigma^2 G_{(j)} \tau G'_{(j)}$$

$$\text{MSEM}\left(\hat{\gamma}_{(j)}\right) = E\left(\hat{\gamma}_{(j)} - \gamma\right) \left(\hat{\gamma}_{(j)} - \gamma\right)' = D\left(\hat{\gamma}_{(j)}\right)$$

$$+ \text{Bias}\left(\hat{\gamma}_{(j)}\right) \text{Bias}\left(\hat{\gamma}_{(j)}\right)' = \sigma^2 G_{(j)} \tau G'_{(j)}$$

$$+ \left(\left(G_{(j)} - I\right) \gamma + G_{(j)} \tau A\right) \left(\left(G_{(j)} - I\right) \gamma + G_{(j)} \tau A\right)',$$

where $\tau = (I + \sigma^2 \Lambda)^{-1}$ and $A = (X'_* \delta + R'_* W^{-1} g)$.

Based on (14), the respective bias vector, dispersion matrix, and MSEM of the MRE, SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, and SRrd can easily be obtained and are given in Table B1 in Appendix B.

By using the approach of Kadiyala [10] and (3) and (4), the generalized prediction function can be defined as follows:

$$y_0 = X_* \gamma + \delta$$
$$\hat{y}_{(j)} = X_* \hat{\gamma}_{(j)}, \tag{15}$$

where y_0 is the actual value and $\hat{y}_{(j)}$ is the corresponding predictor.

The MSEM of the generalized predictor is given by

$$\text{MSEM}\left(\hat{y}_{(j)}\right) = E\left(\hat{y}_{(j)} - y_0\right) \left(\hat{y}_{(j)} - y_0\right)'$$

$$= X_* \left(\text{MSEM}\left(\hat{\gamma}_{(j)}\right)\right) X'_*$$

$$- X_* \left(\text{Bias}\left(\hat{\gamma}_{(j)}\right)\right) \delta' \tag{16}$$

$$- \delta \left(\text{Bias}\left(\hat{\gamma}_{(j)}\right)\right)' X'_* + \delta \delta'.$$

Note that the predictors based on the MRE, SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, and SRrd are denoted by $\hat{y}_{\text{MRE}}, \hat{y}_{\text{SRRE}}, \hat{y}_{\text{SRAURE}}, \hat{y}_{\text{SRLE}}, \hat{y}_{\text{SRAULE}}, \hat{y}_{\text{SRPCRE}}, \hat{y}_{\text{SRrk}}$, and $\hat{y}_{\text{SRrd}}$, respectively.

3. Mean Square Error Matrix (MSEM) Comparisons

If two generalized biased estimators $\hat{\gamma}_{(i)}$ and $\hat{\gamma}_{(j)}$ are given, the estimator $\hat{\gamma}_{(j)}$ is said to be superior to $\hat{\gamma}_{(i)}$ with respect to

MSEM sense if and only if $\text{MSEM}(\widehat{\gamma}_{(i)}) - \text{MSEM}(\widehat{\gamma}_{(j)}) \geq 0$. Also, if two generalized predictors $\widehat{y}_{(i)}$ and $\widehat{y}_{(j)}$ are given, the predictor $\widehat{y}_{(j)}$ is said to be superior to $\widehat{y}_{(i)}$ with respect to MSEM sense if and only if $\text{MSEM}(\widehat{y}_{(i)}) - \text{MSEM}(\widehat{y}_{(j)}) \geq 0$.

Now let $D_{(i,j)} = D(\widehat{\gamma}_{(i)}) - D(\widehat{\gamma}_{(j)})$, $b_{(i)} = \text{Bias}(\widehat{\gamma}_{(i)})$, $b_{(j)} = \text{Bias}(\widehat{\gamma}_{(j)})$, and $\Delta_{(i,j)} = \text{MSEM}(\widehat{\gamma}_{(i)}) - \text{MSEM}(\widehat{\gamma}_{(j)}) = D_{(i,j)} + b_{(i)}b'_{(i)} - b_{(j)}b'_{(j)}$.

By applying Lemma A1 (see Appendix A), the following theorem can be stated for the superiority of $\widehat{\gamma}_{(j)}$ over $\widehat{\gamma}_{(i)}$ with respect to the MSEM criterion.

Theorem 1. *If $D_{(i,j)}$ is positive definite, then $\widehat{\gamma}_{(j)}$ is superior to $\widehat{\gamma}_{(i)}$ in MSEM sense when the regression model is misspecified due to excluding relevant variables if and only if*

$$b'_{(j)} \left(D_{(i,j)} + b_{(i)}b'_{(i)} \right)^{-1} b_{(j)} \leq 1. \tag{17}$$

Proof. Let $D_{(i,j)}$ be a positive definite matrix. According to Lemma A1 (see Appendix A), $\Delta_{(i,j)}$ is nonnegative definite matrix if $b'_{(j)}(D_{(i,j)} + b_{(i)}b'_{(i)})^{-1}b_{(j)} \leq 1$. This completes the proof. $\square$

The following theorem can be stated for the superiority of $\widehat{y}_{(j)}$ over $\widehat{y}_{(i)}$ with respect to the MSEM criterion.

Theorem 2. *If $A \geq 0$, $\widehat{y}_{(j)}$ is superior to $\widehat{y}_{(i)}$ in MSEM sense when the regression model is misspecified due to excluding relevant variables if and only if $\theta \in \Re(A)$ and $\theta' A^{-1}\theta \leq 1$, where $A = X_*\Delta_{(i,j)}X'_* + X_*(b_{(i)} - b_{(j)})(b_{(i)} - b_{(j)})'X'_* + \delta\delta'$, $\theta = \delta + X_*(b_{(i)} - b_{(j)})$, and $\Re(A)$ stands for column space of A and A^{-1} is an independent choice of g-inverse of A.*

Proof. According to (16), we can write $\text{MSEM}(\widehat{y}_{(i)}) - \text{MSEM}(\widehat{y}_{(j)})$ as

$$\begin{aligned}
&\text{MSEM}\left(\widehat{y}_{(i)}\right) - \text{MSEM}\left(\widehat{y}_{(j)}\right) \\
&= X_*\left(\text{MSEM}\left(\widehat{\gamma}_{(i)}\right) - \text{MSEM}\left(\widehat{\gamma}_{(j)}\right)\right)X'_* \\
&\quad - X_*\left(\text{Bias}\left(\widehat{\gamma}_{(i)}\right) - \text{Bias}\left(\widehat{\gamma}_{(j)}\right)\right)\delta' \\
&\quad - \delta\left(\text{Bias}\left(\widehat{\gamma}_{(i)}\right) - \text{Bias}\left(\widehat{\gamma}_{(j)}\right)\right)'X'_* \\
&= X_*\Delta_{(i,j)}X'_* - X_*\left(b_{(i)} - b_{(j)}\right)\delta' \\
&\quad - \delta\left(b_{(i)} - b_{(j)}\right)'X'_*.
\end{aligned} \tag{18}$$

After some straight forward calculation, it can be written as

$$\text{MSEM}\left(\widehat{y}_{(i)}\right) - \text{MSEM}\left(\widehat{y}_{(j)}\right) = A - \theta\theta', \tag{19}$$

where $A = X_*(\Delta_{(i,j)} + (b_{(i)} - b_{(j)})(b_{(i)} - b_{(j)})')X'_* + \delta\delta'$ and $\theta = \delta + X_*(b_{(i)} - b_{(j)})$.

Due to Lemma A3 (see Appendix A), $\text{MSEM}(\widehat{y}_{(i)}) - \text{MSEM}(\widehat{y}_{(j)})$ is nonnegative definite matrix if and only if $A \geq 0$, $\theta \in \Re(A)$ and $\theta' A^{-1}\theta \leq 1$, where $\Re(A)$ stands for column space of A and A^{-1} is an independent choice of g-inverse of A. This completes the proof. $\square$

Based on Theorems 1 and 2, we can define Corollaries C1–C28, written in Appendix C, for the superiority conditions between two selected estimators and for the respective predictors by substituting the relevant expressions for $\text{Bias}(\widehat{\gamma}_{(i)})$, $\text{Bias}(\widehat{\gamma}_{(j)})$, $D(\widehat{\gamma}_{(i)})$, and $D(\widehat{\gamma}_{(j)})$ given in Table B1 in Appendix B.

4. Illustration of Theoretical Results

4.1. Numerical Example. To illustrate the theoretical results, we considered the dataset which gives the total National Research and Development Expenditures as a Percent of Gross National Product by Country from 1972 to 1986. The dependent variable Y of this dataset is the percentage spent by the United States, and the four other independent variables are X_1, X_2, X_3, and X_4. The variable X_1 represents the percent spent by the former Soviet Union, X_2 that spent by France, X_3 that spent by West Germany, and X_4 that spent by the Japan. The data has been analysed by Gruber [18], Akdeniz and Erol [19], and Li and Yang [14], among others. Now we assemble the data as follows:

$$X = \begin{pmatrix}
1.9 & 2.2 & 1.9 & 3.7 \\
1.8 & 2.2 & 2.0 & 3.8 \\
1.8 & 2.4 & 2.1 & 3.6 \\
1.8 & 2.4 & 2.2 & 3.8 \\
2.0 & 2.5 & 2.3 & 3.8 \\
2.1 & 2.6 & 2.4 & 3.7 \\
2.1 & 2.6 & 2.6 & 3.8 \\
2.2 & 2.6 & 2.6 & 4.0 \\
2.3 & 2.8 & 2.8 & 3.7 \\
2.3 & 2.7 & 2.8 & 3.8
\end{pmatrix}$$

$$y = \begin{pmatrix}
2.3 \\
2.2 \\
2.2 \\
2.3 \\
2.4 \\
2.5 \\
2.6 \\
2.6 \\
2.7 \\
2.7
\end{pmatrix}. \tag{20}$$

Note that the eigenvalues of $X'X$ are 302.96, 0.728, 0.044, and 0.035, the condition number is 93, and the Variance Inflation Factor (VIF) values are 6.91, 21.58, 29.75, and 1.79. This implies the existence of serious multicollinearity in the dataset.

The corresponding OLS estimator of β is $\widehat{\beta} = (X'X)^{-1}X'y = (0.645, 0.089, 0.143, 0.152)$ and the estimate of σ^2 is $\widehat{\sigma}^2 = 0.00153$. In this example, we consider

$R = (1, -2, -2, -2)$ and $g = c(1, -1, 2, 0)$. The SMSE values of the estimators are summarized in Tables B2-B3 in Appendix B.

Table B2 shows the estimated SMSE values of MRE, SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, and SRrd for the regression model when $(l, p) = (4, 0)$, $(l, p) = (3, 1)$, and $(l, p) = (2, 2)$ with respect to shrinkage parameters (k/d), where l denotes the number of variables in the model and p denotes the number of misspecified variables. Table B3 shows the estimated SMSE values of the predictor of MRE, SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, and SRrd for the regression model when $(l, p) = (4, 0)$, $(l, p) = (3, 1)$, and $(l, p) = (2, 2)$ for some selected shrinkage parameters (k/d).

Note that when $(l, p) = (4, 0)$ the model is correctly specified, when $(l, p) = (3, 1)$ one variable is omitted from the model, and when $(l, p) = (2, 2)$ two variables are omitted from the model. For simplicity, we choose shrinkage parameter values k and d in the range $(0, 1)$.

From Table B2, we can observe that the MRE is superior to the other estimators when $(l, p) = (4, 0)$ and SRAULE, SRRE, SRLE, and SRAURE outperform the other estimators for $(k/d) < 0.2$, $0.2 \leq (k/d) < 0.5$, $0.5 \leq (k/d) < 0.7$, and $(k/d) \geq 0.7$, respectively, when $(l, p) = (3, 1)$. Similarly, SRLE and SRRE are superior to the other estimators for $(k/d) < 0.5$ and $(k/d) \geq 0.5$, respectively, when $(l, p) = (2, 2)$.

From Table B3, we further observe that predictors based on SRLE and SRRE outperform the other predictors for $(k/d) < 0.5$ and $(k/d) \geq 0.5$, respectively, when $(l, p) = (4, 0)$ and $(l, p) = (3, 1)$, and predictors based on SRrd and SRrk are superior to the other predictors for $(k/d) < 0.5$ and $(k/d) \geq 0.5$, respectively, when $(l, p) = (2, 2)$.

4.2. Simulation. For further clarification, a Monte Carlo simulation study is done at different levels of misspecification using R 3.2.5. Following McDonald and Galarneau [20], we can generate the explanatory variables as follows:

$$x_{ij} = \left(1 - \rho^2\right)^{1/2} z_{ij} + \rho z_{i,m};$$

$$i = 1, 2, \ldots, n. \ j = 1, 2, \ldots, m,$$

(21)

where z_{ij} is an independent standard normal pseudorandom number and ρ is specified so that the theoretical correlation between any two explanatory variables is given by ρ^2. A dependent variable is generated by using the following equation:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \varepsilon_i;$$

$$i = 1, 2, \ldots, n,$$

(22)

where ε_i is a normal pseudorandom number with mean zero and variance one. Also, we select $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ as the normalized eigenvector corresponding to the largest eigenvalue of $X'X$ for which $\beta'\beta = 1$. Further we choose $R = (1, 1, 1, 1, 1)$ and $g = (1, -2, 0, 3, 1)$.

Then the following setup is considered to investigate the effects of different degrees of multicollinearity on the estimators:

(i) $\rho = 0.9$, condition number = 9.49, and VIF = $(5.99, 5.88, 5.94, 5.96, 20.47)$.

(ii) $\rho = 0.99$, condition number = 34.77, and VIF = $(57.66, 56.50, 57.26, 57.31, 225.06)$.

(iii) $\rho = 0.999$, condition number = 115.66, and VIF = $(574.3, 562.8, 570.7, 570.8, 2271.4)$.

Three different sets of observations are considered by selecting $(l, p) = (5, 0)$, $(l, p) = (4, 1)$, and $(l, p) = (3, 2)$ when $n = 50$, where l denotes the number of variables in the model and p denotes the number of misspecified variables. Note that when $(l, p) = (5, 0)$ the model is correctly specified, when $(l, p) = (4, 1)$ one variable is omitted from the model, and when $(l, p) = (3, 2)$ two variables are omitted from the model. For simplicity, we select values k and d in the range $(0, 1)$.

The simulation is repeated 2000 times by generating new pseudorandom numbers and the simulated SMSE values of the estimators and predictors are obtained using the following equations:

$$\text{SMSE}\left(\widehat{\gamma}_{(j)}\right) = \frac{1}{2000} \sum_{r=1}^{2000} \text{tr}\left(\text{MSEM}\left(\widehat{\gamma}_{(j)r}\right)\right),$$

$$\text{SMSE}\left(\widehat{y}_{(j)}\right)$$

$$= \frac{1}{2000} \sum_{r=1}^{2000} \text{tr}\left(\text{MSEM}\left(\widehat{y}_{(j)r}\right)\right) \text{ respectively.}$$

(23)

The simulation results are summarized in Tables B4–B9 in Appendix B.

Tables B4, B5, and B6 show the estimated SMSE values of the estimators for the regression model when $(l, p) = (5, 0)$, $(l, p) = (4, 1)$, and $(l, p) = (3, 2)$ and $\rho = 0.9$, $\rho = 0.99$, and $\rho = 0.999$ for the selected values of shrinkage parameters (k/d), respectively. Tables B7, B8, and B9 show the corresponding estimated SMSE values of the predictors for the above regression models, respectively.

From Table B4, we can observe that MRE and SRAULE outperform the other estimators for $(k/d) < 0.8$ and $(k/d) \geq 0.8$, respectively, when $(l, p) = (5, 0)$ and $(l, p) = (4, 1)$. Further, SRLE and SRRE are superior to the other estimators for $(k/d) < 0.5$ and $(k/d) \geq 0.5$, respectively, when $(l, p) = (3, 2)$ under $\rho = 0.9$.

From Table B5, we can observe that SRAULE, MRE, and SRAURE outperform the other estimators for $(k/d) < 0.3$, $0.3 \leq (k/d) < 0.7$, and $(k/d) \geq 0.7$, respectively, when $(l, p) = (5, 0)$. Similarly, SRAULE, SRRE, SRLE, and SRAURE are superior to the other estimators when $(k/d) < 0.2$, $0.2 \leq (k/d) < 0.5$, $0.5 \leq (k/d) < 0.7$, and $(k/d) \geq 0.7$, respectively, when $(l, p) = (4, 1)$, and both SRLE and SRRE outperform the other estimators for $(k/d) < 0.5$ and $(k/d) \geq 0.5$, respectively, when $(l, p) = (3, 2)$ and $\rho = 0.99$.

The results in Table B6 indicate that MRE is superior to the other estimators when $(l, p) = (5, 0)$, and SRAULE, SRRE, SRLE, and SRAURE outperform the other estimators for $(k/d) < 0.2$, $0.2 \leq (k/d) < 0.5$, $0.5 \leq (k/d) < 0.7$, and $(k/d) \geq 0.7$, respectively, when $(l, p) = (4, 1)$. Further, SRLE and SRRE outperform the other estimators for $(k/d) < 0.5$

FIGURE 1: SMSE values of the estimators in the misspecified regression model $((l, p) = (3, 2))$ when $n = 50$ and $\rho = 0.9$.

FIGURE 2: SMSE values of the estimators in the misspecified regression model when $n = 50$ and $\rho = 0.99$.

FIGURE 3: SMSE values of the estimators in the misspecified regression model when $n = 50$ and $\rho = 0.999$.

FIGURE 4: SMSE values of the predictors in the misspecified regression model when $n = 50$ and $\rho = 0.9$.

and $(k/d) \geq 0.5$, respectively, when $(l, p) = (3, 2)$ and $\rho = 0.999$.

From Tables B7–B9, we further observe that the predictors based on SRrd and SRrk always outperform the other predictors for $(k/d) < 0.5$ and $(k/d) \geq 0.5$, respectively, when $(l, p) = (5, 0)$, $(l, p) = (4, 1)$, and $(l, p) = (3, 2)$.

The SMSE values of the selected estimators are plotted with different ρ values to demonstrate the results graphically when $(l, p) = (3, 2)$. Figures 1–3 show the graphical illustration of the performance of estimators in the misspecified regression model $((l, p) = (3, 2))$ when $\rho = 0.9$, $\rho = 0.99$,

and $\rho = 0.999$, respectively. Similarly, Figures 4–6 present the graphical illustration of the performance of predictors in the misspecified regression model $((l, p) = (3, 2))$ when $\rho = 0.9$, $\rho = 0.99$, and $\rho = 0.999$, respectively.

5. Conclusion

Theorems 1 and 2 give the common form of superiority conditions to compare the estimators (MRE, SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, and SRrd) and their respective predictors in MSEM criterion in the misspecified linear
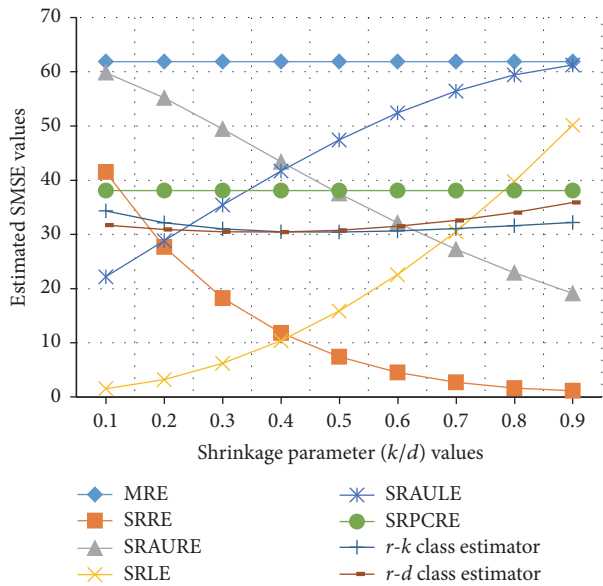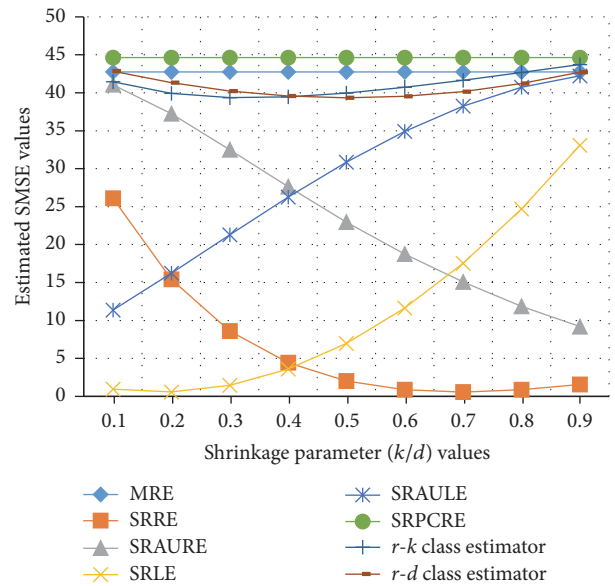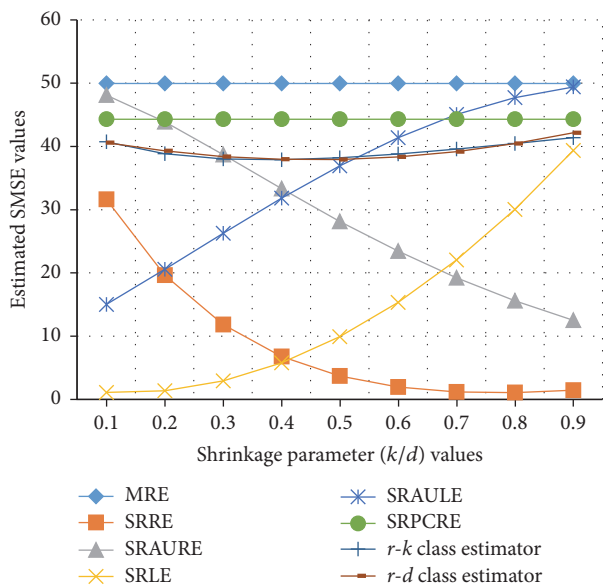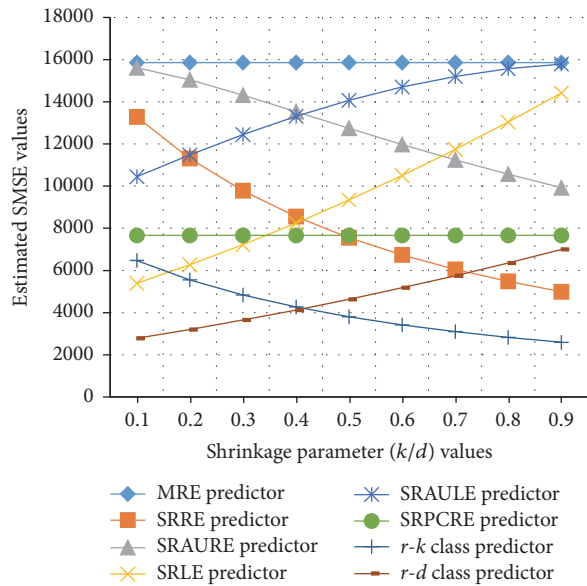
Figure 5: SMSE values of the predictors in the misspecified regression model when $n = 50$ and $\rho = 0.99$.

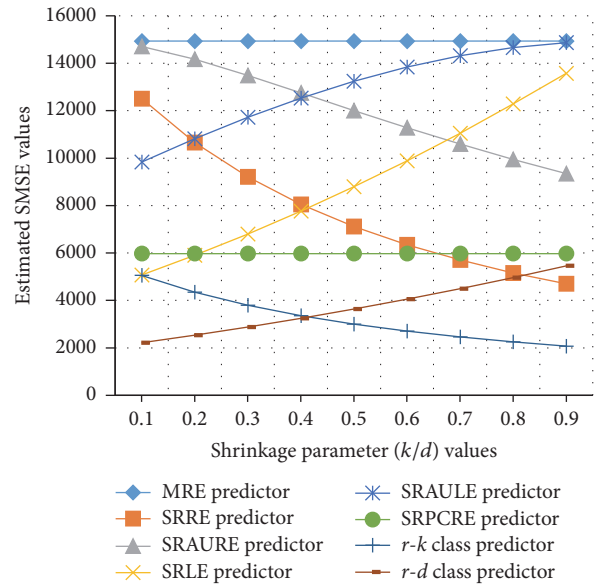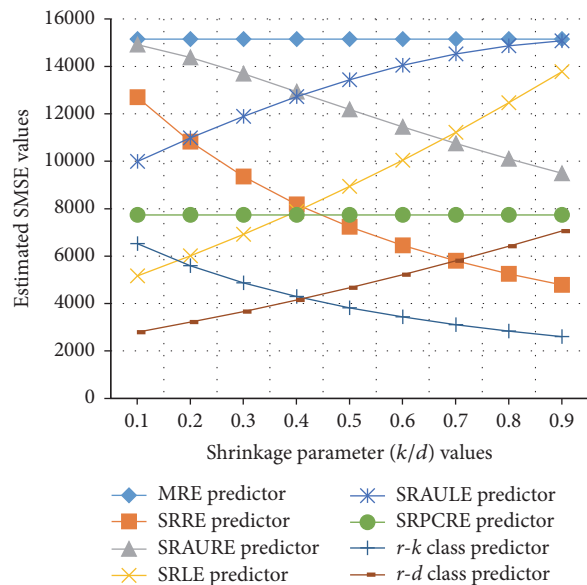Figure 6: SMSE values of the predictors in the misspecified regression model when $n = 50$ and $\rho = 0.999$.

regression model when the prior information of the regression coefficients is incomplete, and the multicollinearity exists among the explanatory variables.

From the simulation study, the superior estimators and predictors over the others when the conditions are different can be identified. The results obtained in this research will produce significant improvements in the parameter estimation in misspecified regression models with incomplete prior information, and the results are applicable to real-world applications.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] N. Sarkar, "Comparisons among some estimators in misspecified linear models with multicollinearity," *Annals of the Institute of Statistical Mathematics*, vol. 41, no. 4, pp. 717–724, 1989.

[2] G. Şiray, "r-d class estimator under misspecification," *Communications in Statistics—Theory and Methods*, vol. 44, no. 22, pp. 4742–4756, 2015.

[3] J. Wu, "Superiority of the r-k class estimator over some estimators in a misspecified linear model," *Communications in Statistics—Theory and Methods*, vol. 45, no. 5, pp. 1453–1458, 2016.

[4] S. Chandra and G. Tyagi, "On the performance of some biased estimators in a misspecified model with correlated regressors," in *STATISTICS IN TRANSITION new series*, pp. 27–52, 2017.

[5] M. Kayanan and P. Wijekoon, "Performance of Existing Biased Estimators and the respective Predictors in a Misspecified Linear Regression Model," *Open Journal of Statistics*, pp. 876–900, 2017.

[6] H. Theil and A. S. Goldberger, "On Pure and Mixed Statistical Estimation in Economics," *International Economic Review*, vol. 2, no. 1, pp. 65–78, 1961.

[7] T. Teräsvirta, *Linear restrtctions in misspecified linear models and polynomial distributed lag estimation*, vol. 5, Departntent of Statistics University of Helsinki, Helsinki, Finland, 1980.

[8] R. C. Mittelhammer, "On specification error in the general linear model and weak mean square error superiority of the mixed estimator," *Communications in Statistics—Theory and Methods*, vol. 10, no. 2, pp. 167–176, 1981.

[9] K. Ohtani and Y. Honda, "On small sample properties of the mixed regression predictor under misspecification," *Communications in Statistics - Theory and Methods*, pp. 2817–2825, 1984.

[10] K. Kadiyala, "Mixed regression estimator under misspecification," *Economics Letters*, vol. 21, no. 1, pp. 27–30, 1986.

[11] G. Trenkler and P. Wijekoon, "Mean square error matrix superiority of the mixed regression estimator under misspecification," *Statistica*, vol. 49, no. 1, pp. 65–71, 1989.

[12] P. Wijekoon and G. Trenkler, "Mean Square Error Matrix Superiority of Estimators under Linear Restrictions and Misspecification," *Economics Letters*, vol. 30, pp. 141–149, 1989.

[13] M. Hubert and P. Wijekoon, "Superiority of the stochastic restricted Liu estimator under misspecification," *Statistica*, vol. 64, no. 1, pp. 153–162, 2004.

[14] Y. Li and H. Yang, "A new stochastic mixed ridge estimator in linear regression model," *Statistical Papers*, pp. 315–323, 2010.

[15] J. Wu and H. Yang, "On the stochastic restricted almost unbiased estimators in linear regression model," *Communications in Statistics - Simulation and Computation*, vol. 43, no. 2, pp. 428–440, 2014.

[16] D. He and Y. Wu, "A Stochastic Restricted Principal Components Regression Estimator in the Linear Model," *The Scientific World Journal*, vol. 2014, Article ID 231506, 6 pages, 2014.

[17] J. Wu, "On the Stochastic Restricted r-k Class Estimator and Stochastic Restricted r-d Class Estimator in Linear Regression Model," *Journal of Applied Mathematics*, vol. 2014, Article ID 173836, 6 pages, 2014.

Forecasting Time Series Movement Direction with Hybrid Methodology

Salwa Waeto,[1,2] Khanchit Chuarkham,[3] and Arthit Intarasit[1,2]

[1]*Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Pattani 94000, Thailand*
[2]*Centre of Excellence in Mathematics, Commission on Higher Education, Ratchathewi, Bangkok 10400, Thailand*
[3]*Faculty of Commerce and Management, Prince of Songkla University, Trang Campus, Trang 92000, Thailand*

Correspondence should be addressed to Arthit Intarasit; a.intarasit@gmail.com

Forecasting the tendencies of time series is a challenging task which gives better understanding. The purpose of this paper is to present the hybrid model of support vector regression associated with Autoregressive Integrated Moving Average which is formulated by hybrid methodology. The proposed model is more convenient for practical usage. The tendencies modeling of time series for Thailand's south insurgency is of interest in this research article. The empirical results using the time series of monthly number of deaths, injuries, and incidents for Thailand's south insurgency indicate that the proposed hybrid model is an effective way to construct an estimated hybrid model which is better than the classical time series model or support vector regression. The best forecast accuracy is performed by using mean square error.

1. Introduction

Time series modeling and forecasting are a challenge for describing dynamic phenomena and pattern behavior of the time series. In recent years, the issue of accurate Thailand's south insurgency trends has been receiving more attention. There are many research papers that studied the unrest in southern Thailand. According to the database of Deep South Watch [1], Jitpiromsri and Mccargo [2] and Jitpiromsri [3] reported the trends of Thailand's south insurgency using diagram for comparing the monthly number of the unrest incidents. By applying a polynomial least-square regression, they provided the forecasting model for describing the unrest incidents in the south of Thailand. This polynomial is not indeed fitting the monthly number of the unrest incidents as well.

In this study, we would like to identify patterns and trends of Thailand's south insurgency and to evaluate the accuracy of model for modeling and forecasting. By doing this, we use the traditional regression models such as Autoregressive

(AR), Moving Average (MA), Autoregressive Moving Average (ARMA), and Autoregressive Integrated Moving Average (ARIMA). These models are also called the Box-Jenkins models.

In general, time series data of Thailand's south insurgency can be categorized as nonstationary time by using Box-Jenkins methodology. Then an estimated model of time series data of Thailand's south insurgency can be obtained by support vector regression (SVR). We aim to combine ARIMA and SVR for making an adequately estimated model in order to forecast time series of Thailand's south insurgency.

This paper is organized as follows. Section 2 provides some backgrounds of mathematical theories related to time series modeling and forecasting and SVR. The detail of proposed hybrid model is explained in Section 3. Section 4 gives experimental results obtaining the proposed hybrid model with the first difference in time series of Thailand's south insurgency. Finally, the main conclusions are summarized in Section 5.

2. Background and Mathematical Theory

2.1. Autoregressive Integrated Moving Average Modeling.
Three basic methods for forecasting time series are naïve model, exponential smoothing model, and ARIMA model. The first two models relate to a random walk as the formulation of the model. In this section, ARIMA model will be reviewed.

An autoregressive model of order p abbreviated as AR(p) model is

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + w_t = \sum_{i=1}^{p} \phi_i y_{t-i}, \quad (1)$$

where y_t is stationary, $\phi_1, \ldots, \phi_p$ are constants ($p \neq 0$), and w_t is a white noise series with zero mean and variance σ_w^2. AR(p) model of (1) predicts the current value y_t by the p past function $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ which explains y_t as a linear combination of $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$.

The Moving Average model of order q abbreviated as MA(q) model is

$$y_t = \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q} + w_t = \sum_{j=1}^{q} \theta_j w_{j-q}, \quad (2)$$

where y_t is stationary, $\phi_1, \ldots, \phi_q$ are constants ($q \neq 0$), and w_t is a Gaussian white noise series with mean zero. MA(q) model of (2) explains the current value y_t by a linear combination of the q white noise $w_{t-1}, w_{t-2}, \ldots, w_{t-q}$.

Autoregressive Moving Average model abbreviated as ARMA(p, q) model developed by Box and Jenkins [4] is defined by the combined autoregressive and the Moving Average model. It has the form

$$y_t = \delta + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j w_{t-j} + w_t. \quad (3)$$

According to the original Box-Jenkins methodology, an integrated process is the stationary process obtained by differenced a nonstationary process. The stationary ARMA(p,q) process after being differenced d times is denoted by ARIMA(p, d, q):

$$\Delta^d y_t = \delta + \phi_1 \Delta^d y_{t-1} + \cdots + \phi_p \Delta^d y_{t-p} + \theta_1 w_{t-1} + \cdots \\ + \theta_q w_{t-q} + w_t, \quad (4)$$

where Δ^d denoted dth difference time series [5]. These models are as foundation model for time series forecasting.

2.2. Box-Jenkins Methodology. Plots of autocorrelation function (acf) and partial autocorrelation function (pacf) are the main tools in order to identify parameters for AR, MA, ARMA, and ARIMA models. AR(p) is used to obtain an estimated model for time series when the acf exhibits tendency to die down quickly, either by an exponential decay or by a damped sine wave whereas the pacf exhibits tendency to show spike (significant autocorrelation) for lags up to p and then will die down immediately.

Opposite to AR(p), MA(q) is used to obtain an estimated model of time series when the acf exhibits tendency to die down quickly, either by an exponential decay or by a damped sine wave whereas the pacf exhibits tendency to show spike (significant autocorrelation) for lags up to p and then will die down immediately.

A mixed process ARMA(p, q) is suggested when either the acf or the pacf exits tend to show spike for lags up, respectively, to p and q and then die down quickly, either by an exponential decay or by a damped sine wave. Proceeding diagnostic checking to identify p and q for the mixed process ARMA(p, q) which is able to fit to times series is the best performance [6].

This identification as described in this section will be important to diagnose a model of our study.

2.3. Hybrid Models. In recent years, the forecasting model used in the literature can be classified into three categories: statistical models, artificial intelligence model (AI), and hybrid model.

Statistical models are known as time series models including naïve model, AR model, MA model, ARIMA model, exponential smoothing, and generalized autoregressive conditional heteroskedasticity (GARCH) volatility which aim to utilize time series analysis to identify the pattern of time series and provide the future value based on the obtained pattern.

ARIMA model is known as Box-Jenkins model [4] which includes AR and MA models identified by Box-Jenkins methodology. These models are based on the assumption that the time series under study are stationary and linear which means that the relationship between the input and output series is linear.

AI models are the second kinds of forecast time series, practically artificial neural networks (ANNs), genetic algorithm (GA), and supported vector machine (SVM). AI models can capture nonlinear pattern and improved forecast performance.

Many of the literatures introduce a hybrid model in order to capture the linear and nonlinear characteristics in time series. Wang et al. [7] reported that using a statistical model alone or using an AI model alone are not adequate in making forecasts for stock price time series.

2.4. Hybrid Methodology. A hybrid model is described by a combination of models with mixed methodology for formulation. Many literatures suggested that time series consists of linear L_t and nonlinear N_t as in the form

$$y_t = L_t + N_t. \quad (5)$$

An estimated model of (5) is formulated as follows: using linear statistic model to obtain an estimated model of linear component L_t denoted by $\widehat{L}_t$ and after that modeling the residual $y_t - \widehat{L}_t$ which contains only the nonlinear relationship to obtain an estimated model of nonlinear component N_t denoted by $\widehat{N}_t$.

Zhang [8] utilized the hybrid model by introducing the estimated model of (5) in the form $\widehat{y}_t = \widehat{L}_t + \widehat{N}_t$,

where $\widehat{L}_t$ is prescribed by ARIMA model and $\widehat{N}_t$ is prescribed by feedforward neural networks model. Modified Zhang's hybrid approach with estimated $\widehat{N}_t$ by support vector machine (SVM) model can be found in many literatures, for example, De Oliveira and Ludermir [9], while Aladag et al. [10] estimated $\widehat{N}_t$ by Elman's recurrent neural networks (ERNN) model and applied to Canadian Lynx data.

2.5. Supported Vector Regression. Let the dot product pace $\mathfrak{R}^d$ be our data universe with vectors $\mathbf{x} \in \mathfrak{R}^d$ as objects. Let S be a sample set such that $S \subset \mathfrak{R}^d$. Let $f : \mathfrak{R}^d \to \mathfrak{R}$ be the target function. Let $D = \{(\mathbf{x}, y) \mid \mathbf{x} \in S$ and $y = f(\mathbf{x})\}$ be the training set.

The regression problem is to find the best approximate model $\widetilde{f} : \mathfrak{R}^d \to y$ for the true underlying function f mapping input $\mathbf{x}$ to output y by using D such that $\widetilde{f}(\mathbf{x}) \cong f(\mathbf{x})$.

The regression problem is classified as linear or nonlinear type. For the linear regression model, the best approximate model $\widetilde{f}$ can be obtained from the set of possible functions with the following set of specifications:

$$\left\{\widetilde{f} \mid \widetilde{f}(\mathbf{x}) = \boldsymbol{\omega}^T \mathbf{x} + b, \ \boldsymbol{\omega} \in \mathfrak{R}^d, \ b \in \mathfrak{R}\right\}, \quad (6)$$

where $\boldsymbol{\omega}$ is a weight vector and b is a constant.

Generally, in order to describe nonlinear relationship between input and output, the SVR allied $\Phi : \mathfrak{R}^d \to \mathfrak{F}$ transform the nonlinear regression problem in the lower dimension input space $\mathfrak{R}^d$ into a linear regression problem in a high dimension feature space $\mathfrak{F}$. In the new space $\mathfrak{F}$, a linear model $\widetilde{f}$ is formulated, which represents a nonlinear model in the original space:

$$\widetilde{y} = \widetilde{f}(\mathbf{x}, \boldsymbol{\omega}) = \langle \boldsymbol{\omega}, \Phi(\mathbf{x}) \rangle + b, \quad (7)$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product in $\mathfrak{F}$. Linear SVR model $\widetilde{f}$ in (6) is obtained from (7) by using the identity function $\Phi(\mathbf{x}) \to \mathbf{x}$.

Performing SVR to fit linear regression $\widetilde{f}$ to the training data by estimate $\boldsymbol{\omega}$ and b in (7) as minimization of the following regularized function:

$$\underset{\mathbf{w},b}{\text{minimize}} \quad R(C) = \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{i=1}^{\ell} L_2^\varepsilon(\widetilde{f}(\mathbf{x}_i), y_i), \quad (8)$$

where both C and ε are user-given parameters and $L_2^\varepsilon(\widetilde{f}(\mathbf{x}), y)$ is quadratic ε-insensitive loss function defined by $L_2^\varepsilon(y, f(\mathbf{x})) = |y - f(\mathbf{x})|_\varepsilon^2$.

The following two propositions related to the formulation of an estimated model. These propositions are modified from [11, 12] for our study.

Proposition 1. *Given a regression training set $D = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_\ell, y_\ell)\} \subseteq \mathbb{R}^n \times \mathbb{R}$ the optimal support vector regression model is computed by $\widetilde{f}^*(\mathbf{x}) = \langle \boldsymbol{\omega}^*, \mathbf{x} \rangle - b^*$, where*

the parameters $\boldsymbol{\omega}^*$ and b^* solved the following optimization problem:

$$\underset{\mathbf{w},b}{\text{minimize}} \quad R(C) = \frac{1}{2}\|\boldsymbol{\omega}\|^2 + \frac{C}{2}\sum_{i=1}^{\ell}\left(\xi_i^2 + \widehat{\xi}_i^2\right),$$

$$\text{subject to} \quad y_i - \widetilde{f}(\mathbf{x}_i) \leq \varepsilon + \xi_i,$$

$$\widetilde{f}(\mathbf{x}_i) - y_i \leq \varepsilon + \widehat{\xi}_i, \quad (9)$$

$$\xi_i, \widehat{\xi}_i \geq 0,$$

$$\forall i = 1, \ldots, \ell$$

hold with $\widetilde{y} = \widetilde{f}(\mathbf{x}) = \langle \boldsymbol{\omega}, \mathbf{x} \rangle + b$.

The constant C is called the penalty constant which is trade-off between margin maximization and the minimization of the slack variables.

Proposition 2. *Given a regression training set $D = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_\ell, y_\ell)\} \subseteq \mathbb{R}^n \times \mathbb{R}$ the optimal support vector regression model is computed by $\widetilde{f}^*(\mathbf{x}) = \boldsymbol{\omega}^* \bullet \mathbf{x} - b^*$, where*

$$\boldsymbol{\omega}^* = \sum_{i=1}^{\ell}(\alpha_i^* - \widetilde{\alpha}_i^*)\mathbf{x}_i,$$

$$b^* = \frac{1}{\ell}\sum_{i=1}^{\ell}\boldsymbol{\omega}^* \bullet \mathbf{x}_i - \left(y_i + \varepsilon + \frac{\alpha_i^*}{C}\right) \quad (10)$$

and α^* are the parameters solved by the following dual quadratic optimization problem:

$$\underset{\boldsymbol{\alpha}, \widetilde{\boldsymbol{\alpha}}}{\text{minimize}} \quad \sum_{i=1}^{\ell} y_i(\alpha_i - \widetilde{\alpha}_i) - \varepsilon\sum_{i=1}^{\ell}(\alpha_i + \widetilde{\alpha}_i)$$

$$- \frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}\left((\alpha_i - \widetilde{\alpha}_i)(\alpha_j - \widetilde{\alpha}_j)\mathbf{x}_i \bullet \mathbf{x}_j + \frac{1}{C}\delta_{ij}\right), \quad (11)$$

$$\text{subject to} \quad \sum_{i=1}^{\ell}(\alpha_i - \widetilde{\alpha}_i) = 0, \quad \alpha_i \geq 0, \ \widetilde{\alpha}_i \geq 0, \ \forall i = 1, \ldots, \ell.$$

The parameter $\boldsymbol{\omega}^*$ is obtained by α_i^* and $\widetilde{\alpha}_i^*$ which satisfied optimization (11). The parameter b^* is solved as follows: obtain $\xi_i = \alpha_i^*/C$ from (11) and substitute $\widetilde{f}^*(\mathbf{x}) = \boldsymbol{\omega}^* \bullet \mathbf{x} - b^*$ and $\xi_i = \alpha_i^*/C$ in the constraint $\widetilde{f}(\mathbf{x}_i) = y_i + \varepsilon + \widehat{\xi}_i$; then solve for b_i^*. Define b^* as the average of b_i^*.

The optimal regression model is obtained by substituting $\boldsymbol{\omega}^*$ into b^* and into $\widetilde{f}^*(\mathbf{x}) = \boldsymbol{\omega}^* \bullet \mathbf{x} - b^*$ where we have the following lemma.

Lemma 3. *The optimal regression model is*

$$\widetilde{f}^*(\mathbf{x}) = \sum_{i=1}^{\ell}(\alpha_i^* - \widetilde{\alpha}_i^*)\mathbf{x}_i \bullet \mathbf{x} - \frac{1}{\ell}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell}(\alpha_i^* - \widetilde{\alpha}_i^*)\mathbf{x}_i$$

$$\bullet \mathbf{x}_j - \left(y_j + \varepsilon + \frac{\alpha_j^*}{C}\right), \quad (12)$$

where the coefficient $(\alpha_i^* - \tilde{\alpha}_i^*)$ is nonzero as support vector. The optimal regression model $\tilde{f}^*(\mathbf{x})$ depends only on the support vectors.

3. Formulation of the Proposed Model

In this section, we want to formulate the proposed model. We begin by using the hybrid models that combine several models in order to reduce the risk of using an inappropriate model, obtain the results that are more accurate than the previous one, and improve overall forecasting performance.

Assume that (y_t) is the under-study time series based on the assumption of linear and stationary time series. Then, we use the Box-Jenkins methodology to check behavior of (y_t). After this step, we can get a suitable model of AR(p) or MA(q) or ARIMA(p, d, q) in order to estimate $\tilde{y}_t^A$. By fitting under-study time series with (2) or (3) or (4), we can get $\tilde{y}_t^A$ in the form $g^\Delta(a_{t-1}, a_{t-2}, \ldots, a_{t-p}, w_{t-1}, w_{t-2}, \ldots, w_{t-q})$. According to Lemma 3, perform SVR for under-study time series in order to evaluate $\tilde{y}_t^S$ from (12). This model is a function of its past N values in the form $g^*(\beta_{t_1}, \beta_{t_2}, \ldots, \beta_{t_N}, b^*)$ with $t_1 \leq \cdots \leq t_{N-1} \leq t_N$.

Consider a time set $\{t_1, \ldots, t_k, t_{k+1}, \ldots, t_N\}$ with $y_{t_1} = a$ and $y_{t_N} = b \neq a$. There is only one single time point (necessarily from time t_1 to time t_N) precisely on time $\tilde{y} \simeq \tilde{y}_{t^*}^S$ satisfying $e_{t^*}^S < e_{t^*}^A$ for all $t^* \in \{t_1, \ldots, t_k\}$ and $\tilde{y} \simeq \tilde{y}_{t^\triangle}^A$ satisfying $e_{t^\triangle}^A < e_{t^\triangle}^S$ for all $t^\triangle \in \{t_{k+1}, \ldots, t_N\}$. The proposed hybrid model $\tilde{y}$ is the estimated model defined by setting $\tilde{y} \simeq \tilde{y}_{t^*}^S$ for all $t^* \in \{t_1, \ldots, t_k\}$ and vanishing otherwise and setting $\tilde{y} \simeq \tilde{y}_{t^\triangle}^A$ for all $t^\triangle \in \{t_{k+1}, \ldots, t_N\}$ and vanishing otherwise. The proposed hybrid model can be extended to include two or more time intervals.

The under-study time series (y_t) is initially modeled by the proposed hybrid model as follows:

$$y_t = \tilde{y}_t + e_t = \left(\tilde{y}_{t^*}^S + \tilde{y}_{t^\triangle}^A\right) + e_t, \tag{13}$$

where e_t is residuals of the time series model in the time t that is as obtained from (13),

$$e_t = y_t - \tilde{y}_t = \left(y_{t^*} - \tilde{y}_{t^*}^S\right) + \left(y_{t^\triangle} - \tilde{y}_{t^\triangle}^A\right) = e_{t^*}^S + e_{t^\triangle}^A, \tag{14}$$

where $e_{t^*}^S$ and $e_{t^\triangle}^A$ are residuals of the under-study time series model in the time t of the estimated model of $\tilde{y}_{t^*}^S$, respectively, to $e_{t^\triangle}^A$.

4. Application of the Proposed Hybrid Model to Thailand's South Insurgency Movement Direction Forecasting

4.1. Data Set. In this research, we are interested in studying the unrest in the four southern provinces of Thailand, particularly in Pattani, Yala, Narathiwat, and parts of Songkla. We consider the monthly number of deaths, injuries, and incidents in these provinces. At the time of working research, we can get the latest data from Deep South Watch (DSW) [1] and Deep South Coordination Center (DSCC) [13]. By
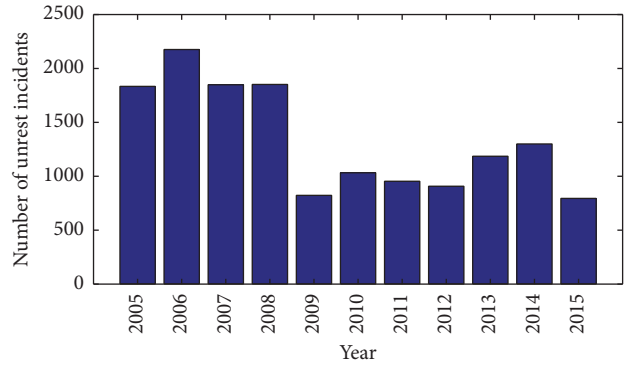
FIGURE 1: Number of unrest incidents in the four southern provinces of Thailand (Pattani, Yala, Narathiwat, and Songkla) from 2005 to 2015.
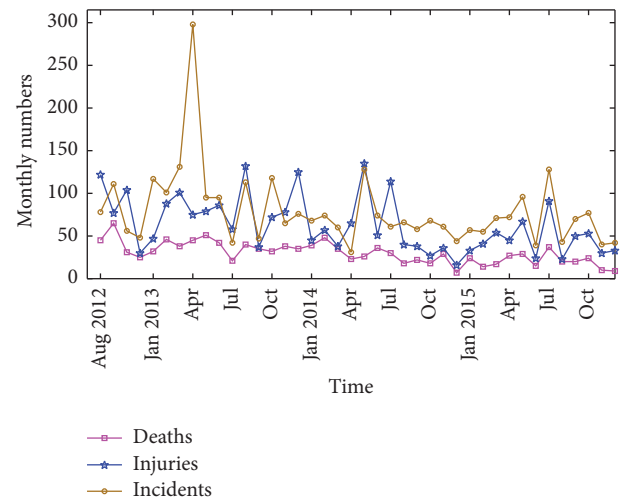
FIGURE 2: Monthly number of deaths, injuries, and incidents for unrest in the four southern provinces of Thailand.

using the proposed hybrid model, our aim is to formulate an estimate model for the trend of the number of deaths, injuries, and incidents in these regions.

Figure 1 illustrates a diagram of the number of unrest incidents in the four southern provinces from 2005 to 2015. This diagram presents a high frequency of the number of unrest incidents with a small fluctuation in the first period (2005 to 2008), a decreasing frequency of the number of unrest incidents in the middle period (2009 to 2012), and an increasing frequency of the number of unrest incidents from 2013 to 2014, the lowest frequency in 2015.

The data series of our study consists of 40 months of deaths, injuries, and incidents in the four southern provinces of Thailand from September 2012 to December 2015.

Figure 2 presents three graphs describing three data series of monthly number for deaths, injuries, and incidents. It shows that the graph of deaths is in the bottom for all periods of time, while the graph of injuries is in the middle between the graphs of deaths and incidents in almost all periods of time. Moreover, the graph of incidents is in the top in almost all periods of time.
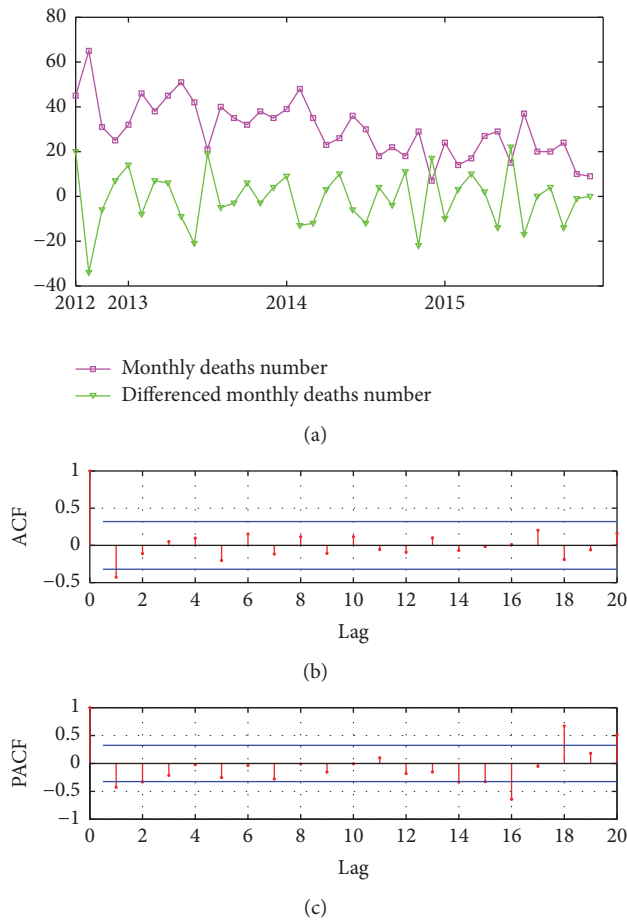
FIGURE 3: (a) Monthly number of deaths is plotted against its first differenced series, acf (b) and pacf (c) plots for the first difference in monthly number of deaths.

FIGURE 4: (a) Monthly number of injuries is plotted against its first differenced series, acf (b) and pacf (c) plots for the first difference in monthly number of injuries.

From Figure 2, we can see that the number of incidents is not necessary to be equal to the sum of numbers of deaths and injuries. Sometimes, there is an unrest incident; no deaths or no injures occurs. Or there are high numbers of deaths and injuries in some incidents.

Monthly numbers of injures and incidents are apparently stationary. A candidate model for monthly number of two data series can be determined by plotting of acf and pacf. However, the monthly number of deaths exhibits a linear trend in the mean since it has a clear downward slope.

Figure 3 shows comparing of monthly number of deaths plotted against its first differenced series for monthly number of deaths (a) and plotting of acf (b) and pacf (c). The data series of injures plotted against its first differenced series is shown in Figure 4 and the data series of incidents plotted against its first differenced series is shown in Figure 5.

Plotting of the first differenced series (Figures 3, 4, and 5) shows that it looks like a stationary process, although plotting acf and pacf of series of deaths, injuries, and incidents cannot clearly identify parameter for constructing an estimated model formulated by the ARIMA model.

The acf for the first difference in monthly number of deaths tends to die down quickly whereas the pacf tends to

show spike for lags up to 1 which ignores significant spikes in each plot when it is outside the limits. This suggests that the first difference in monthly number of deaths can be a model as an AR(1).

Similarly, the first differenced series of injures and incidents can be a model as an AR(1). After checking of residual in diagnosis stage, this indicates that ARMA(2, 3) is a candidate model for formulating an estimate model for the first difference in monthly number of deaths and injuries. MA(1) is also a candidate model for the first difference in monthly number of incidents. With notation of ARIMA(p, d, q), ARIMA(2, 1, 3) is an estimated model for monthly number of deaths and injuries and ARIMA(0, 1, 1) for monthly number of incidents.

Table 1 reports mean square error (mse) of three estimated models for monthly number of deaths, injuries, and incidents formulated by ARIMA, SVR, and hybrid. The mean square error of the formal model is calculated by choosing the best trajectory: 1×10^6 trajectories simulated by ARIMA for each series.

Plotting a convergent of mean square error is calculated from monthly number and an estimated model with 2,500,

TABLE 1: Some reports of mean square error for fitting and forecasting the series.

Time series	Model		
	ARIMA*	SVR	Hybrid
Deaths series	9.4383	7.0882	0.7922
Injuries series	28.0352	20.4161	0.9921
Incident series	41.8077	31.7669	1.469

Note. ARIMA(2,1,3) for monthly number of deaths and injuries and ARIMA(0,1,1) for monthly number of incidents.
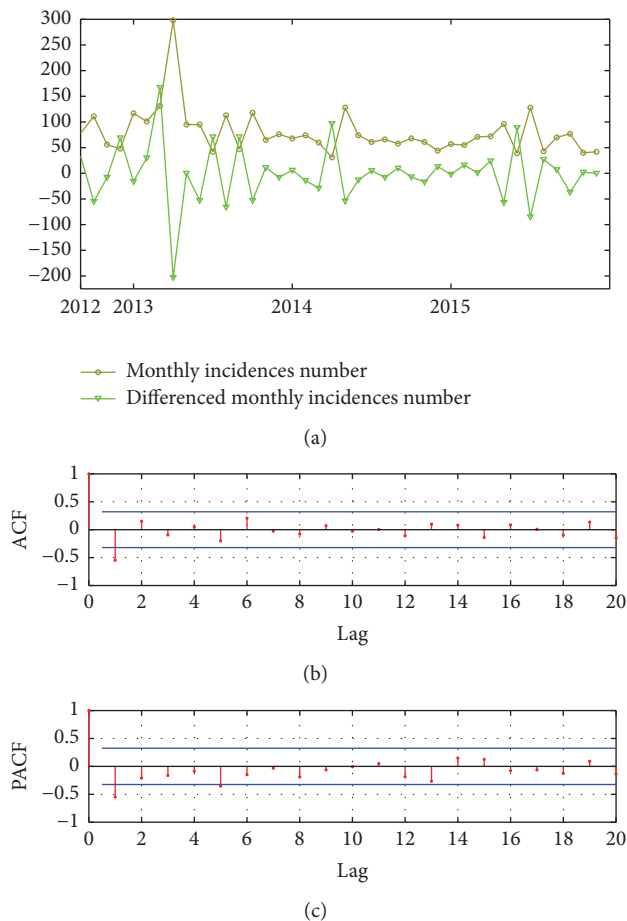
FIGURE 5: (a) Monthly number of incidents is plotted against its first differenced series, acf (b) and pacf (c) plots for its first difference in monthly number of incidents.

$5,000, \ldots, 1 \times 10^6$ trajectories for monthly number of deaths, injuries, and incidents illustrated in Figure 6.

Setting $\varepsilon = 0.0025$, $c = 150000$, $\gamma = 3.25$, and $b = 2.75$ for SVR model and using ARIMA(2, 1, 3) model in order to select from the best trajectory from 1×10^6 trial trajectories, then both models are combined in order to formulate an estimated model for monthly number of deaths: $\widetilde{y}_t = \widetilde{y}_t^S + \widetilde{y}_t^A$, where

$$
\begin{aligned}
\widetilde{y}_t^A = -1.079 &- 0.046\Delta y_{t-1} - 0.751\Delta y_{t-2} - 0.915w_{t-1} \\
&- 0.915w_{t-2} - w_{t-3} + w_t,
\end{aligned}
\tag{15}
$$
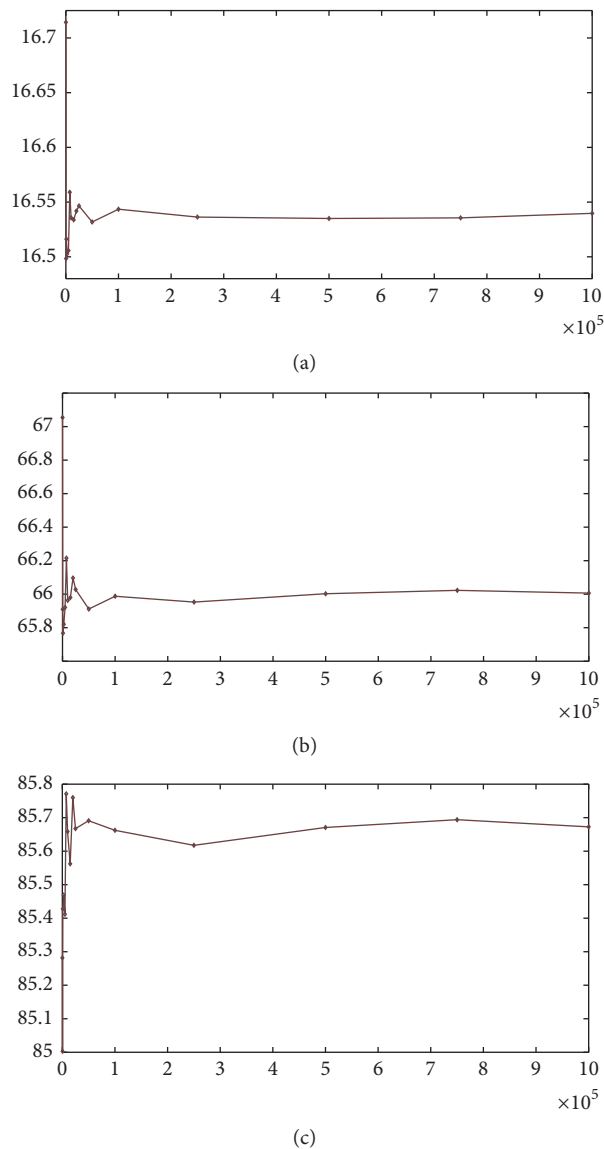
FIGURE 6: Fitting performance for monthly number of deaths (a) and injuries (b) with ARIMA(2, 1, 3) and for monthly number of incidents (c) with ARIMA(0, 1, 1).

$t \in \{35, \ldots, 40\}$, and $\widetilde{y}_t^S = \sum_{i=1}^{35} \beta_i^* K(\mathbf{x_i}, x) + b^*$ where $t \in \{1, \ldots, 36\}$, $b^* = -0.7115$, and β_i^*, $i = 1, \ldots, 35$, where it has specified a value.
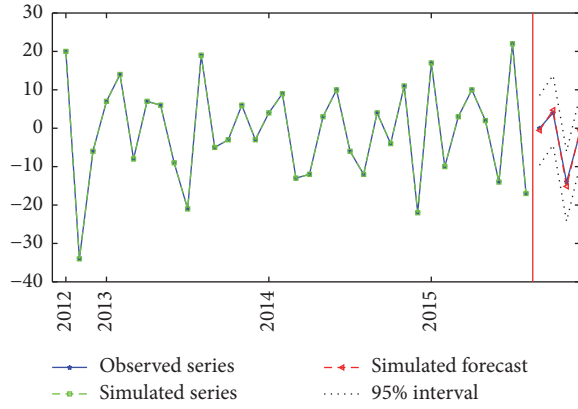
FIGURE 7: The actual, fitted, and forecasted series by hybrid model for series of deaths.

FIGURE 8: The actual, fitted, and forecasted series by hybrid model for series of injuries.

Predictive performance of SVR-ARIMA$(2, 1, 3)$ hybrid model for monthly number of deaths and injuries, respectively, is shown in Figure 7.

In the same way, for monthly number of injures, setting $\varepsilon = 0.025$, $c = 350000$, $\gamma = 2.755$, and $b = 0.00125$ for SVR model and using ARIMA$(2, 1, 3)$ model in order to select from the best trajectory from 1×10^6 trial trajectories, both models are combined in order to formulate an estimated model for monthly number of injuries: $\tilde{y}_t = \tilde{y}_t^S + \tilde{y}_t^A$, where

$$\tilde{y}_t^A = -5.851 - 1.560\Delta y_{t-1} - 0.773\Delta y_{t-2} - 0.618 w_{t-1}$$
$$- 0.618 w_{t-2} - w_{t-3} + w_t, \tag{16}$$

$t \in \{35, \ldots, 40\}$, and $\tilde{y}_t^S = \sum_{i=1}^{35} \beta_i^* K(\mathbf{x_i}, x) + b^*$, where $t \in \{1, \ldots, 36\}$, $b^* = -3.0611$, and β_i^*, $i = 1, \ldots, 35$, where it has specified a value.

Predictive performance of SVR-ARIMA$(2, 3)$ hybrid model for difference monthly number of injuries is shown in Figure 8.

For monthly number of incidents, set $\varepsilon = 0.025$, $c = 200000$, $\gamma = 1.555$, and $b = 0.725$ for SVR model and use ARIMA$(0, 1, 1)$ model to select from the best trajectory

FIGURE 9: The actual, fitted, and forecasted series by hybrid model for number of incidents.

from 1×10^6 trial trajectories. Then these two models are combined in order to formulate an estimated model for monthly number of incidents: $\tilde{y}_t = \tilde{y}_t^S + \tilde{y}_t^A$, where

$$\tilde{y}_t^A = -0.193 - w_{t-1} + w_t, \tag{17}$$

$t \in \{36, \ldots, 40\}$, and $\tilde{y}_t^S = \sum_{i=1}^{35} \beta_i^* K(\mathbf{x_i}, x) + b^*$, where $t \in \{1, \ldots, 35\}$, $b^* = -1.4121$, and β_i^*, $i = 1, \ldots, 35$, where it has specified a value.

Predictive performance of SVR-ARIMA$(0, 1, 1)$ hybrid model for monthly number of incidents is shown in Figure 9.

5. Conclusions

The hybrid SVR-ARIMA model has been investigated to formulate time series model of monthly number of Thailand's south insurgency in this study. In particular, we consider the first difference in monthly number of deaths, injuries, and incidents in Pattani, Yala, Narathiwat, and Songkla provinces in 40 months from September 2012 to December 2015. According to the hybrid methodology, the SVR-ARIMA(p, d, q) model is obtained by combining ARIMA(p, d, q) and SVR model. Plotting of autocorrelation and partial autocorrelation indicates that the first difference in monthly number of deaths, injuries, and incidents is linear and stationary.

The test results of the estimated model are obtained from the proposed hybrid model and compared with the estimated model of the AR(p), MA(q), ARIMA(p, d, q), and SVM models. This presents the fact that the proposed hybrid model performs better than the remaining models. For time series of Thailand's south insurgency, SVR-ARIMA$(2, 1, 3)$ is the estimated model for monthly number of deaths and injuries and SVR-ARIMA$(0, 1, 1)$ is the estimated model for monthly number of incidents. In particular, SVR-ARIMA$(2, 1, 3)$ consists of two components: the first component uses the SVR model in order to formulate the estimated model for historical data and the second component uses the ARIMA model in order to formulate the estimated model for the unseen value in the short future.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] Deep South Watch (DSW). Deep South Incident Database for the Thailand. http://www.deepsouthwatch.org/dsid.

[2] S. Jitpiromsri and D. Mccargo, "The Southern Thai Conflict Six Years On: Insurgency, Not Just Crime," *Contemporary Southeast Asia*, vol. 32, no. 2, pp. 156–183, 2010.

[3] S. Jitpiromsri, *An Inconvenient Truth about the Deep South Violent Conflict: A Decade of Chaotic, Constrained Realities and Uncertain Resolution*, Center for Conflict Studies and Cultural Diversity (CSCD), Deep South Watch (DSW), Prince of Songkla University, 2010.

[4] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, Calif, USA, 1976.

[5] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, Prentice Hall, Englewood Cliffs, NJ, USA, 3rd edition, 1994.

[6] D. Asteriou and S. G. Hall, *Applied Econometrics*, Palgrave Macmillan, China, 2nd edition.

[7] J.-J. Wang, J.-Z. Wang, Z.-G. Zhang, and S.-P. Guo, "Stock index forecasting based on a hybrid model," *Omega*, vol. 40, no. 6, pp. 758–766, 2012.

[8] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[9] J. F. L. De Oliveira and T. B. Ludermir, "A hybrid evolutionary system for parameter optimization and lag selection in time series forecasting," in *Proceedings of the 3rd Brazilian Conference on Intelligent Systems, BRACIS 2014*, pp. 73–78, Sao Paulo, Brazil, October 2014.

[10] C. H. Aladag, E. Egrioglu, and C. Kadilar, "Forecasting nonlinear time series with a hybrid methodology," *Applied Mathematics Letters*, vol. 22, no. 9, pp. 1467–1470, 2009.

[11] L. Hamal, Knowledge*Discovery with Support Vector Machines*, John Wiley & Sons, Hoboken, NJ, USA, 2009.

[12] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, UK, 2000.

[13] Deep South Coordination Center (DSCC). Database of the violent event and victims in southern Thailand. http://dsrd.pn.psu.ac.th/webnew/index.php/database.html.

Applications of Fuss-Catalan Numbers to Success Runs of Bernoulli Trials

S. J. Dilworth[1] and S. R. Mane[2]

[1]*Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA*
[2]*Convergent Computing Inc., P.O. Box 561, Shoreham, NY 11786, USA*

Correspondence should be addressed to S. R. Mane; srmane001@gmail.com

In a recent paper, the authors derived the exact solution for the probability mass function of the geometric distribution of order k, expressing the roots of the associated auxiliary equation in terms of generating functions for Fuss-Catalan numbers. This paper applies the above formalism for the Fuss-Catalan numbers to treat additional problems pertaining to occurrences of success runs. New exact analytical expressions for the probability mass function and probability generating function and so forth are derived. First, we treat sequences of Bernoulli trials with $r \geq 1$ occurrences of success runs of length k with ℓ-overlapping. The case $\ell < 0$, where there must be a gap of at least $|\ell|$ trials between success runs, is also studied. Next we treat the distribution of the waiting time for the rth nonoverlapping appearance of a pair of successes separated by at most $k - 2$ failures ($k \geq 2$).

1. Introduction

In a recent paper [1], the authors derived the exact analytical solution for the probability mass function of the geometric distribution of order k. The roots of the auxiliary equation of the associated recurrence relation were derived in terms of generating functions for Fuss-Catalan numbers. (See the text by Graham et al. [2] for details about Fuss-Catalan numbers.) In this paper, we employ our formalism for the Fuss-Catalan numbers to treat additional problems pertaining to occurrences of success runs in sequences of Bernoulli trials. Throughout our paper, we treat only sequences of independent identically distributed (i.i.d.) Bernoulli trials with constant success probability p (and failure probability $q = 1 - p$). The theory of success runs is discussed extensively in the texts by Balakrishnan and Koutras [3] and Johnson et al. [4]. Our formalism provides a new perspective to treat problems of success runs in sequences of Bernoulli trials and complements and extends results derived by previous authors (especially Feller [5, pp. 322–326]). Citations and comparisons to the works of others will be presented in Sections 3 and 4, after we have derived our results.

We treat two main problems in this paper. First, we consider sequences with multiple $r \geq 1$ occurrences of success runs of length k. The success runs are permitted to overlap, with a maximum of $\ell \geq 0$ overlaps between success runs. This is known as "ℓ-overlapping." The case $\ell < 0$ is perhaps surprising at first sight but is also of interest. In this case there must be a gap or "buffer" of at least $|\ell|$ trials (of arbitrary outcomes) between success runs. We call this "$|\ell|$-buffering." We also consider the scenario in which the length of the sequence n is held fixed and the number of success runs $r \geq 0$ is allowed to vary. This is the binomial distribution of order k with ℓ-overlapping success runs. An encyclopedia article on binomial distributions of order k has been published by Philippou and Antzoulakos [6]. Using Fuss-Catalan numbers, we present new concise expressions for the probability mass functions of these distributions.

In Section 4 we study a different problem. We analyze the distribution of the waiting time for the rth nonoverlapping appearance of a pair of successes separated by at most $k - 2$ failures ($k \geq 2$). Our main reference for this problem is the elegant analysis by Koutras [7], who also gives an excellent bibliography on the subject. For $r = 1$ and $k \geq 2$, the problem is a special case of the detection waiting game when a 2-out-of-k moving (or sliding) window detector is employed. See Koutras [7] for additional details and references. Note that

our material in Section 4 is self-contained and is a different problem from that mentioned above.

2. Notation and Definitions

We summarize the basic notation and definitions presented in our earlier paper [1]. For a sequence of independent identically distributed Bernoulli trials with success probability p (and failure probability $q = 1 - p$), let X_k be the waiting time for the first run of k consecutive successes. Then X_k is said to have the geometric distribution of order k. This distribution was studied by Feller in his classic text [5, pp. 322–326]. It is also known as the negative binomial distribution of order k with parameter $(1, p)$; see Philippou [8]. The probability mass function f_k of X_k satisfies the recurrence relation, for $n > k$,

$$f_k(n) = q f_k(n-1) + pq f_k(n-2) + p^2 q f_k(n-3)$$
$$+ \cdots + p^{k-1} q f_k(n-k). \tag{1}$$

The initial conditions are $f_k(n) = 0$ for $n = 1, \ldots, k-1$ and $f_k(k) = p^k$. We define the auxiliary polynomial

$$\mathcal{A}_{p,k}(z) = z^k - qz^{k-1} - qpz^{k-2} - qp^2 z^{k-3} - \cdots$$
$$- qp^{k-1}. \tag{2}$$

The auxiliary equation is $\mathcal{A}_{p,k}(z) = 0$. We will drop the subscripts p and k unless necessary. Feller [5] proved that the roots of the auxiliary equation are distinct and also that there is a unique positive real root, and it lies in $(0, 1)$, and the real positive root has a strictly larger magnitude than all the other roots. Additional properties of the roots were derived in [1]. We denote the roots by $\lambda_j(p, k)$, $j = 0, 1, \ldots, k-1$, where λ_0 is the positive real root. We call λ_0 the "principal root" and the other roots "secondary roots." Unless required, we will omit the arguments p and k. It is useful to multiply $\mathcal{A}(z)$ by $(z - p)$ to obtain the polynomial

$$\mathcal{B}(z) = (z - p)\mathcal{A}(z) = z^k(z-1) + p^k(1-p). \tag{3}$$

Remark 1 (Fuss-Catalan numbers and roots of auxiliary polynomial). Relevant definitions, formulas, and identities for the Fuss-Catalan numbers can be found in the text by Graham et al. [2]. The Fuss-Catalan numbers are given by

$$A_m(\nu, r) = \frac{r}{m!} \prod_{j=1}^{m-1} (m\nu + r - j)$$
$$= \frac{r}{\Gamma(m+1)} \frac{\Gamma(m\nu + r)}{\Gamma(m(\nu-1) + r + 1)}. \tag{4}$$

The first form (finite product) is valid in general, while the second form (Gamma functions) is well defined provided $m\nu + r \neq 0$. The generating function of the Fuss-Catalan numbers is $B_\nu(z)$ and [2, p. 363]

$$B_\nu(z) = \sum_{m=0}^{\infty} A_m(\nu, 1) z^m, \tag{5a}$$

$$B_\nu(z)^r = \sum_{m=0}^{\infty} A_m(\nu, r) z^m. \tag{5b}$$

We will also require the following formula:

$$B_\nu(z) - 1 = zB_\nu(z)^\nu. \tag{6}$$

It was proved in [1] that, for all $0 < p < 1$,

$$\lambda_j = 1 - \frac{1}{B_{1+1/k}\left(e^{2\pi ij/k} pq^{1/k}\right)} \quad (1 \le j < k). \tag{7}$$

For $k/(k+1) < p < 1$, the above expression also applies for λ_0, while, for $0 < p < k/(k+1)$,

$$\lambda_0 = \frac{1}{B_{k+1}(p^k q)}. \tag{8}$$

For ease of reference, we list several relevant properties of the roots in the following. The proofs of all the results were given in [1], or references cited therein, and are omitted in the following.

Remark 2. All the roots of the auxiliary equation are distinct.

Remark 3. For $p \in (0, 1)$, the auxiliary equation has a unique positive real root, which lies in $(0, 1)$. We denote the positive real root by λ_0, or $\lambda_0(p, k)$, as stated above. For any $p \in (0, 1)$, exactly one of the three following statements is true:

(i) $0 < p < k/(k+1) < \lambda_0 < 1$,

(ii) $0 < \lambda_0 < k/(k+1) < p < 1$,

(iii) $\lambda_0 = p = k/(k+1)$.

Remark 4. For $p \in (0, 1)$, the principal root λ_0 has a strictly greater magnitude than all the other roots of the auxiliary equation; that is, $0 < |z_r| < \lambda_0 < 1$, where $z_r \in \mathbb{C} \setminus \lambda_0$ is a root of $\mathcal{A}(z)$. We employ the term "secondary roots" for the set $\{\lambda_j, \ j = 1, \ldots, k-1\}$.

Remark 5. For any $p \in (0, 1)$, the secondary roots λ_j, $j = 1, \ldots, k-1$ satisfy the inequality

$$0 < |\lambda_j| < \min\{p, \lambda_0\} \le \frac{k}{k+1} \le \max\{p, \lambda_0\} < 1. \tag{9}$$

The inequalities involving $k/(k+1)$ are strict if $p \neq k/(k+1)$.

Remark 6. For $p \in (0, 1)$, let $R(p)$ denote the set of $k+1$ roots of the equation $z^k(1-z) = p^k(1-p)$. Let $p_1, p_2 \in (0, 1)$. Then $R(p_1) = R(p_2)$ if $p_2 = p_1$ or $p_2 = \lambda_0(p_1)$; otherwise $R(p_1) \cap R(p_2) = \varnothing$.

In addition to the above properties of the roots, we will also need the following two results, which were not proved in [1], as well as a lemma about sums of series.

Proposition 7 (distinctness of roots for different k). *Consider fixed $p \in (0, 1)$. Suppose z_r is a root of the auxiliary equation $\mathcal{A}(z) = 0$ for $k = k_1$. Then z_r is not a root for any other value of k.*

Proof. We are given that

$$z_r^{k_1}(1 - z_r) = p^{k_1}q. \tag{10}$$

Suppose that z_r is also a root for $k = k_2 \neq k_1$. Then by hypothesis

$$z_r^{k_2}\left(1 - z_r\right) = p^{k_2}q. \tag{11}$$

From (9), $z_r \neq 0$ and $z_r \neq 1$ for $0 < p < 1$; hence we can divide the two equations to deduce

$$z_r^{k_2-k_1} = p^{k_2-k_1}. \tag{12}$$

Hence $|z_r| = |p|$. Also from (9), $|z_r| < p$ for all the secondary roots. Thus the only possibility is that z_r is the positive real root. Hence $z_r = p$, but this is a root if and only if $p = k/(k+1)$ (see Remark 3). However, for arbitrary $p \in (0, 1)$, the constraint $p = k/(k+1)$ has either no solution for k or at most one solution for k. $\qquad\square$

Proposition 8 (comparison of principal roots for different k, for fixed p). *For fixed $p \in (0, 1)$, if $k_1 < k_2$ the principal roots $\lambda_0(p, k_1)$ and $\lambda_0(p, k_2)$, for $k = k_1, k_2$ respectively, satisfy the inequality*

$$\lambda_0\left(p, k_1\right) < \lambda_0\left(p, k_2\right). \tag{13}$$

Proof. To exhibit the dependence on k, we denote the auxiliary polynomial by $\mathscr{A}_{p,k}(z)$. Then from (2)

$$\mathscr{A}_{p,k+1}\left(z\right) = z\mathscr{A}_{p,k}\left(z\right) - qp^k. \tag{14}$$

Then set $z = \lambda_0(p, k)$ to obtain

$$\mathscr{A}_{p,k+1}\left(\lambda_0\left(p, k\right)\right) = \lambda_0\left(p, k\right)\mathscr{A}_{p,k}\left(\lambda_0\left(p, k\right)\right) - qp^k$$
$$= -qp^k < 0. \tag{15}$$

Now $\mathscr{A}_{p,k+1}(z)$ has exactly one real root in $z \in (0, 1)$, which is $\lambda_0(p, k+1)$. Also $\mathscr{A}_{p,k+1}(0) < 0$ and $\mathscr{A}_{p,k+1}(1) > 0$. It follows that

$$\lambda_0\left(p, k\right) < \lambda_0\left(p, k+1\right). \tag{16}$$

By extension, this establishes (13) for all $k_1 < k_2$. $\qquad\square$

Lemma 9. *For $n \geq m \geq 1$ and $\xi \neq 1$,*

$$\xi^{m-1}\sum_{i=m}^{n}\binom{i-1}{m-1}\xi^{i-m}$$
$$= \frac{1}{(m-1)!}\left[\frac{d^{m-1}}{dx^{m-1}}\frac{1 - (\xi x)^n}{1 - \xi x}\right]_{x=1}. \tag{17}$$

Next, for $j \geq n \geq 0$ and $\xi \neq 1$,

$$\left[\frac{d^n}{dx^n}\left(\frac{x^j}{1 - \xi x}\right)\right]_{x=1} = n!\sum_{i=0}^{n}\binom{j}{i}\frac{\xi^{n-i}}{(1 - \xi)^{n-i+1}}. \tag{18}$$

The expressions on the right hand sides of both equations are clearly well defined for all $\xi \neq 1$.

Proof. To derive (17) we define the sum

$$Z\left(x\right) = \frac{1 + \xi x + (\xi x)^2 + \cdots + (\xi x)^{n-1}}{(m-1)!}$$
$$= \frac{1 - (\xi x)^n}{(m-1)!\,(1 - \xi x)}. \tag{19}$$

Then differentiate the sum in (19) $m-1$ times to obtain

$$\frac{d^{m-1}Z\left(x\right)}{dx^{m-1}} = \xi^{m-1}\left[\binom{m-1}{m-1} + \binom{m}{m-1}\xi x\right.$$
$$\left. + \binom{m+1}{m-1}(\xi x)^2 + \cdots + \binom{n-1}{m-1}(\xi x)^{n-m}\right]. \tag{20}$$

Evaluation at $x = 1$ yields (17). The derivation of (18) is an application of Leibniz's rule:

$$\frac{d^n}{dx^n}\left(\frac{x^j}{1 - \xi x}\right) = \sum_{i=0}^{n}\binom{n}{i}\left(\frac{d^i x^j}{dx^i}\right)\left(\frac{d^{n-i}}{dx^{n-i}}\frac{1}{1 - \xi x}\right)$$
$$= \sum_{i=0}^{n}\frac{n!}{i!\,(n-i)!}\frac{j!}{(j-i)!}x^{j-i}\frac{(n-i)!\xi^{n-i}}{(1 - \xi x)^{n-i+1}} \tag{21}$$
$$= n!\sum_{i=0}^{n}\frac{j!}{i!\,(j-i)!}\frac{\xi^{n-i}x^{j-i}}{(1 - \xi x)^{n-i+1}}.$$

Evaluation at $x = 1$ yields (18). $\qquad\square$

3. Multiple $r \geq 1$ Success Runs in Sequences of Bernoulli Trials

3.1. Probability Generating Function, Mean, and Variance. We now turn to the first problem of interest in this paper, namely, the waiting time to obtain $r > 1$ success runs of length k. We begin by displaying the following expressions for the case $r = 1$. They were derived by Feller [5] and will be required in the following.

Remark 10 (probability generating function for $r = 1$). Let $s \in \mathbb{C}$. The probability generating function (p.g.f.) for the geometric distribution of order k is [5, eq. (7.6)]

$$\phi\left(s, k\right) = \frac{p^k s^k\left(1 - ps\right)}{1 - s + qp^k s^{k+1}}$$
$$= \frac{p^k s^k}{1 - qs\left(1 + ps + \cdots + p^{k-1}s^{k-1}\right)}. \tag{22}$$

The p.g.f. exists for $|s| < 1/\lambda_0(p, k)$ [5]. The mean and variance are given by [5, eq. (7.7)]

$$\mu\left(k\right) = \frac{1 - p^k}{qp^k},$$
$$\sigma^2\left(k\right) = \frac{1}{\left(qp^k\right)^2} - \frac{2k+1}{qp^k} - \frac{p}{q^2}. \tag{23}$$

The dependences on p and k will be omitted in the following unless necessary.

We now calculate the probability generating function, mean, and variance for multiple $r > 1$ overlapping runs. The success runs have length k and there can be at most $0 \le \ell < k$ overlaps between consecutive success runs. We denote the waiting time by $X_{r,k,\ell}$. The case $\ell = 0$ of nonoverlapping runs was extensively analyzed by Philippou [8], who named the distribution as the negative binomial distribution of order k with vector parameter (r, p). Ling (1989) [9] and Hirano et al. [10] derived results for the special case $\ell = k - 1$. The text by Balakrishnan and Koutras [3] lists the special cases $\ell = 0$ and $\ell = k - 1$ as, respectively, Type I and Type III negative binomial distributions of order k.

Proposition 11 (probability generating function for $r \ge 1$).
Let the probability generating function for $r \ge 1$ success runs of length k with at most ℓ overlaps be $\phi_{r,k,\ell}(s)$. We will omit the subscripts k and ℓ unless necessary. Then

$$\phi_{r,k,\ell}(s) = \frac{\phi(s,k)^r}{\phi(s,\ell)^{r-1}}. \tag{24}$$

Notice that, for $r = 1$, we obtain $\phi_{1,k,\ell}(s) = \phi(s,k)$ (see (22)), as required.

Proof. Define $X_j(= X_{j,k,\ell})$ as the waiting time to complete j success runs. So suppose we have completed $j - 1$ success runs. Hence by definition the last k trials are all successes. Then exactly one of the following $k - \ell + 1$ mutually exclusive events will occur:

  (i) The next $k - \ell$ trials are all successes. This yields the jth success run.

  (ii) The next κ trials are successes, followed by a failure, where $\kappa = 0, 1, \ldots, k - \ell - 1$. Then we restart the waiting time for the next success run from scratch (conditioned on an initial failure). We denote this additional waiting time by Y_j. Clearly, Y_j has the same distribution as X_1.

Since the events are mutually exclusive, we add the probabilities to obtain

$$\mathbb{E}\left(\exp\left(tX_j\right)\right) = p^{k-\ell}\mathbb{E}\left(\exp\left(t\left(X_{j-1} + k - \ell\right)\right)\right)$$

$$+ \sum_{\kappa=0}^{k-\ell-1} p^\kappa q \mathbb{E}\left(\exp\left(t\left(X_{j-1} + Y_j + \kappa + 1\right)\right)\right)$$

$$= p^{k-\ell}e^{t(k-\ell)}\mathbb{E}\left(\exp\left(tX_{j-1}\right)\right) + \sum_{\kappa=0}^{k-\ell-1} p^\kappa q e^{t(\kappa+1)}$$

$$\cdot \mathbb{E}\left(\exp\left(tX_{j-1}\right)\right)\mathbb{E}\left(\exp\left(tY_j\right)\right) \tag{25}$$

$$= \mathbb{E}\left(\exp\left(tX_{j-1}\right)\right)$$

$$\cdot \left[p^{k-\ell}e^{t(k-\ell)} + \phi\left(e^t\right)\sum_{\kappa=0}^{k-\ell-1} p^\kappa q e^{t(\kappa+1)}\right].$$

Now set $s = e^t$ and note that $\phi_j(s) = \mathbb{E}(\exp(tX_j))$. Hence we obtain the following recurrence relation and solution for $\phi_r(s)$:

$$\phi_r(s) = \phi_{r-1}(s)\left[(ps)^{k-\ell} + \phi(s)\sum_{\kappa=0}^{k-\ell-1} p^\kappa q s^{\kappa+1}\right], \tag{26a}$$

$$= \phi(s)\left[(ps)^{k-\ell} + \phi(s)qs\frac{1 - (ps)^{k-\ell}}{1 - ps}\right]^{r-1}. \tag{26b}$$

Define Z as the term in the brackets. After some tedious algebra we obtain

$$Z = (ps)^{k-\ell} + \frac{(ps)^k(1-ps)}{1 - s + qp^k s^{k+1}}\frac{qs\left(1 - (ps)^{k-\ell}\right)}{1 - ps}$$

$$= \frac{(ps)^k(1-ps)}{1 - s + qp^k s^{k+1}}\frac{1 - s + qp^\ell s^{\ell+1}}{(ps)^\ell(1 - ps)} = \frac{\phi(s,k)}{\phi(s,\ell)}. \tag{27}$$

In the last line it is necessary to exhibit the dependences on k and ℓ explicitly. Then (24) follows immediately. $\square$

Proposition 12 (domain of convergence). *The probability generating function for $r \ge 1$ success runs $\phi_{r,k,\ell}(s)$ converges for*

$$|s| < \frac{1}{\lambda_0(p,k)}. \tag{28}$$

Hence the domain of convergence of the probability generating function is the same for all $r \ge 1$.

Proof. Clearly, the function $\phi_{r,k,\ell}(s)$ is well defined if and only if the sums of the series for $\phi(s,k)$ and $\phi(s,\ell)$ both converge. It was proved by Feller [5] that $\phi(s,k)$ exists for $|s| < 1/\lambda_0(p,k)$. Hence $\phi_{r,k,\ell}(s)$ exists if

$$|s| < \min\left\{\frac{1}{\lambda_0(p,k)}, \frac{1}{\lambda_0(p,\ell)}\right\}. \tag{29}$$

However, because $\ell < k$, it follows from Proposition 8 that $\lambda_0(p,\ell) < \lambda_0(p,k)$. Hence $1/\lambda_0(p,k) < 1/\lambda_0(p,\ell)$. This proves (28). $\square$

Proposition 13 (mean and variance). *The mean $\mu_{r,k,\ell}$ and variance $\sigma^2_{r,k,\ell}$ for the waiting time for $r \ge 1$ success runs are given by*

$$\mu_{r,k,\ell} = r\mu(k) - (r-1)\mu(\ell), \tag{30a}$$

$$\sigma^2_{r,k,\ell} = r\sigma^2(k) - (r-1)\sigma^2(\ell). \tag{30b}$$

Proof. We put $s = e^t$ and differentiate with respect to t and evaluate at $t = 0$. We differentiate $\ln\phi_r(e^t)$ to obtain

$$\frac{\phi'_r(e^t)}{\phi_r(e^t)} = r\frac{\phi'(e^t,k)}{\phi(e^t,k)} - (r-1)\frac{\phi'(e^t,\ell)}{\phi(e^t,\ell)}. \tag{31}$$

Evaluating at $t = 0$ and noting that $\phi(1, k) = \phi(1, \ell) = 1$ yield

$$\mu_{r,k,\ell} = r\mu(k) - (r - 1)\mu(\ell). \tag{32}$$

This proves (30a). We differentiate again to obtain

$$\frac{\phi_r''(e^t)}{\phi_r(e^t)} - \frac{\phi_r'^2(e^t)}{\phi_r(e^t)^2}$$

$$= r\left(\frac{\phi''(e^t, k)}{\phi(e^t, k)} - \frac{\phi'^2(e^t, k)}{\phi(e^t, k)^2}\right) \tag{33}$$

$$- (r - 1)\left(\frac{\phi''(e^t, \ell)}{\phi(e^t, \ell)} - \frac{\phi'^2(e^t, \ell)}{\phi(e^t, \ell)^2}\right).$$

We again evaluate at $t = 0$ to obtain

$$\sigma_{r,k,\ell}^2 = r\sigma^2(k) - (r - 1)\sigma^2(\ell). \tag{34}$$

This proves (30b). □

We now show that various results derived by other authors are special cases of our results above. As stated above, the case $\ell = 0$ of nonoverlapping runs was solved by Philippou [8], while Ling (1989) [9] and Hirano et al. [10] treated the case $\ell = k - 1$ of overlapping runs.

Remark 14 (Philippou (1984)). Philippou [8, Lemma 2.2] stated "Let X be a *rv* distributed as $NB_k(x; r, p)$. Then its probability generating function, to be denoted by $\gamma_{k,r}(s)$, is given by"

$$\gamma_{k,r}(s) = \left[\frac{p^k s^k (1 - ps)}{(1 - s + qp^k s^{k+1})}\right]^r, \quad |s| \le 1. \tag{35}$$

The mean and variance are given by [8, Proposition 2.1]

$$\mathbb{E}(X) = \frac{r(1 - p^k)}{qp^k}, \tag{36a}$$

$$\sigma^2(X) = \frac{r\left[1 - (2k + 1)qp^k - p^{2k+1}\right]}{q^2 p^{2k}}. \tag{36b}$$

Proof. By definition, Philippou's notation $\gamma_{k,r}(s)$ is the same as our $\phi_{r,k,\ell}(s)$ with $\ell = 0$. From (22), it is easy to show that $\phi(s, \ell) = 1$ for $\ell = 0$, whence $\phi_{r,k,0}(s) = \phi(s, k)^r$ and (35) follows. Note that Philippou [8] stated the domain of convergence to be $|s| \le 1$, but we have shown that it is $|s| \le 1/\lambda_0(p, k)$, which is a larger domain. Next, it is also easy to show that $\mu(\ell) = \sigma^2(\ell) = 0$ for $\ell = 0$, whence $\mu_{r,k,0} = r\mu(k)$ and $\sigma_{r,k,0}^2 = r\sigma^2(k)$ which yield (36a) and (36b), respectively. □

Remark 15 (Ling (1989)). Ling (1989) [9, Theorem 4.1] stated that, for $|s| \le 1$,

$$\phi_r^{(k)}(s) = s\phi_{r-1}^{(k)}(s)\left[p + q\phi_1^{(k)}(s)\right], \tag{37a}$$

$$\phi_r^{(k)}(s) = \phi_1^{(k)}(s)\left[s\left(p + q\phi_1^{(k)}(s)\right)\right]^{r-1}. \tag{37b}$$

Proof. Ling wrote $\phi_r^{(k)}(s)$ and $\phi_1^{(k)}(s)$ where we have written $\phi_{r,k,\ell}(s)$ and $\phi(s)$, but the connection between the notations is clear. Setting $\ell = k - 1$ in (26a) yields

$$\phi_r(s) = \phi_{r-1}(s)\left[ps + qs\phi(s)\right]$$
$$= s\phi_{r-1}(s)\left[p + q\phi(s)\right]. \tag{38}$$

This yields (37a). Next (37b) follows immediately by solving the recurrence relation

$$\phi_r(s) = \phi(s)\left[s\left(p + q\phi(s)\right)\right]^{r-1}. \tag{39}$$

Similar to Philippou [8], Ling (1989) [9] also stated the domain of convergence to be $|s| \le 1$, but we have shown that it is larger, given by $|s| \le 1/\lambda_0(p, k)$. □

Remark 16 (Hirano et al. (1991)). Hirano et al. [10, Theorem 4.1] stated that "Let $\phi_r^{NB}(t)$ be the p.g.f. of $NB_k^{III}(r, p)$." They wrote

$$\phi_r^{NB}(t) = \left\{pt + qt\phi_1^{NB}(t)\right\}^{r-1}\phi_1^{NB}(t). \tag{40}$$

They also wrote

$$\phi_1^{NB}(t) = \frac{(pt)^k}{1 - q\sum_{i=0}^{k-1} p^i t^{i+1}}. \tag{41}$$

Then they derived the solution

$$\phi_r^{NB}(t) = \frac{(pt - 1)(pt)^{k+r-1}\left(-qp^{k-1}t^k + t - 1\right)^{r-1}}{\left(-qp^k t^{k+1} + t - 1\right)^r}. \tag{42}$$

They also gave expressions for the mean and variance. The mean is

$$\mathbb{E}\left(N_r^{(k)}\right) = \frac{qr + p - p^k}{p^k q}. \tag{43}$$

Proof. The connection between the notations is that they write $\phi_r^{NB}(t)$ and $\phi_1^{NB}(t)$ where we write $\phi_r(s)$ and $\phi(s)$, respectively. They employ t as the independent variable, where we use s. It is simple to derive that their expression for $\phi_1^{NB}(t)$ in (42) equals that for $\phi(s, k)$ in (22). Next, (40) is simply (39) with the changes of notation listed above. Next, setting $\ell = k - 1$ and changing the independent variable from s to t,

$$\phi_{r,k,k-1}(t) = \frac{\phi(t, k)^r}{\phi(t, k - 1)^{r-1}}$$

$$= \frac{(pt)^{kr}(1 - pt)^r}{(1 - t + qp^k t^{k+1})^r}\frac{\left(1 - t + qp^{k-1}t^k\right)^{r-1}}{(pt)^{(k-1)(r-1)}(1 - pt)^{r-1}} \tag{44}$$

$$= \frac{(pt - 1)(pt)^{k+r-1}\left(-qp^{k-1}t^k + t - 1\right)^{r-1}}{\left(-qp^k t^{k+1} + t - 1\right)^r}.$$

This is exactly (42). From (30a), the mean for $\ell = k - 1$ is

$$\mu_{r,k,k-1} = r\mu(k) - (r-1)\mu(k-1)$$

$$= r\frac{1-p^k}{p^k q} - (r-1)\frac{1-p^{k-1}}{p^{k-1}q} = \frac{qr + p - p^k}{p^k q}. \quad (45)$$

This is exactly (43). Hirano et al. [10] also displayed an expression for the variance. The proof of equivalence with our expression involves merely tedious algebra and is omitted.   □

3.2. Probability Mass Function. We derive an expression for $f_{r,k,\ell}(n)$, the probability mass function (p.m.f.) that the rth success run of length k with ℓ-overlapping occurs at the nth Bernoulli trial, where $r \geq 1$ and $n \geq 1$. Clearly $f_{r,k,\ell}(n) = 0$ for $n < \ell + r(k - \ell) = rk - (r-1)\ell$ and $f_{r,k,\ell}(n) = p^{rk-(r-1)\ell}$ for $n = rk - (r-1)\ell$. An expression for the p.m.f. for the case $r = 1$ was derived in [1]. By definition, the probability generating function is related to the probability mass function via

$$\phi_{r,k,\ell}(s) = \sum_{n=0}^{\infty} s^n f_{r,k,\ell}(n). \quad (46)$$

We derived an expression for $\phi_{r,k,\ell}(s)$ above and we will use it to derive an expression for $f_{r,k,\ell}(n)$ in the following. From the second form for $\phi(s,k)$ in (22), with $z = 1/s$,

$$\psi(z) = \phi\left(\frac{1}{z}\right) = \frac{p^k}{\mathscr{A}_{p,k}(z)}. \quad (47)$$

Hence for $r \geq 1$ success runs,

$$\psi_{r,k,\ell}(z) \equiv \phi_{r,k,\ell}\left(\frac{1}{z}\right) = \frac{\psi(z,k)^r}{\psi(z,\ell)^{r-1}}$$

$$= p^{rk-(r-1)\ell}\frac{\mathscr{A}_{p,\ell}(z)^{r-1}}{\mathscr{A}_{p,k}(z)^r}. \quad (48)$$

The right hand side is a rational function of two polynomials. From Proposition 7, the auxiliary polynomials $\mathscr{A}_{p,\ell}(z)$ and $\mathscr{A}_{p,k}(z)$ have no roots in common. Furthermore, because $\ell < k$, the numerator polynomial is of a lower degree than the denominator polynomial. We also know that all the roots of the auxiliary polynomials are distinct. Hence we can expand $\psi_{r,k,\ell}(z)$ as a sum of partial fractions with repeated roots (of the denominator polynomial)

$$\psi_{r,k,\ell}(z) = \sum_{j=0}^{k-1} \sum_{m=1}^{r} \frac{a_{jm}}{\left(z - \lambda_j(p,k)\right)^m}. \quad (49)$$

Here the coefficients a_{jm} are parameters which depend on r, k, and ℓ but not on z. For brevity, we drop the subscripts k and ℓ on $\psi_{r,k,\ell}$ and also write the roots as λ_j in the following. The coefficients a_{jm} can be evaluated explicitly in terms of the roots $\{\lambda_i(p,k), \ i = 0, \ldots, k-1\}$ via the standard residues formula

$$a_{jm} = \frac{1}{(r-m)!}\left[\frac{d^{r-m}}{dz^{r-m}}\left(\left(z - \lambda_j\right)^r \psi_r(z)\right)\right]_{z=\lambda_j}. \quad (50)$$

Returning to the use of $s = 1/z$, we see that

$$\phi_r(s) = \psi_r\left(\frac{1}{s}\right) = \sum_{j=0}^{k-1} \sum_{m=1}^{r} \frac{a_{jm}s^m}{\left(1 - \lambda_j s\right)^m}. \quad (51)$$

We expand the right hand side using the negative binomial theorem and equate $f_{r,k,\ell}(n)$ to the coefficient of s^n.

Proposition 17. *The probability mass function $f_{r,k,\ell}(n)$ for the rth success run of length k with ℓ-overlapping is given by*

$$f_{r,k,\ell}(n) = \sum_{j=0}^{k-1} \sum_{m=1}^{r} \binom{n-1}{m-1} a_{jm}\lambda_j(p,k)^{n-m}. \quad (52)$$

Hence $f_{r,k,\ell}(n)$ is given by a sum of exactly rk terms, independently of n. Recall from above that $f_r(n) = 0$ for $n < rk - (r-1)\ell$ so the above formula is only required for $n \geq rk - (r-1)\ell$; hence the binomial coefficients are well defined.

The derivation of the above expression has already been given above, where all notation has been defined and explained.

3.3. Binomial Distribution of Order k with ℓ-Overlapping. Consider a sequence of Bernoulli trials of fixed length $n > 0$ and let $N_{n,k,\ell}$ denote the number of success runs of length k with a maximum of ℓ overlaps between success runs. This is the binomial distribution of order k with ℓ-overlapping success runs and has been reviewed in the encyclopedia article by Philippou and Antzoulakos [6]. Good overviews have also been given by Makri and Philippou [11] and Makri et al. [12]; see the bibliographies in both references. Ling (1988) [13] introduced the case of $\ell = k - 1$ and called it the "Type II binomial distribution of order k."

The case $r = 0$ is the probability that the longest success run in the first n trials has length less than k. It is also known as the probability that the waiting time to attain the first success run of length k exceeds n trials. This scenario has been solved by many authors. For example, Feller [5] presented an asymptotic solution in terms of the principal root. In our paper [1], we extended Feller's solution to include all the roots. Solutions have also been derived by Burr and Cane [14], Godbole [15], Philippou and Makri [16], and Muselli [17], all of whom expressed their results using (possibly nested) binomial or multinomial sums.

Let $g_{n,k,\ell}(r) = P(N_{n,k,\ell} = r)$, where $r = 0, 1, \ldots$, be the probability mass function for $N_{n,k,\ell}$. We derive an expression for $g_{n,k,\ell}(r)$ in the following. Note that, to obtain a nontrivial distribution, we must have $n \geq \ell + r(k - \ell)$ so for fixed n we must have $0 \leq r \leq r_* \equiv \lfloor (n - \ell)/(k - \ell) \rfloor$.

Proposition 18. *The probability mass function $g_{n,k,\ell}(r) = P(N_{n,k,\ell} = r)$ for the binomial distribution of order k with ℓ-overlapping is given by*

$$g_{n,k,\ell}(r) = \frac{f_{r+1,k,\ell}(n+k+1)}{p^k q} - \frac{f_{r,k,\ell}(n+\ell+1)}{p^\ell q}. \quad (53)$$

Here $r = 0, 1, \ldots, \lfloor (n - \ell)/(k - \ell) \rfloor$. Note that the last term vanishes for $r = 0$; hence the above expression agrees with our result in [1].

Proof. We solve the problem as follows. Suppose the rth success run is completed on the nth trial. Then by definition the outcomes of the last k trials are all successes. There are now two mutually exclusive and exhaustive possibilities, according as the $(r-1)$th success run is contiguous with (and possibly overlaps) the rth success run, or there is at least one failure between the runs.

  (i) In the former case, the $(r-1)$th success run terminates at the trial $n - k + \ell$. (This event is null if $r = 1$.) The outcome of the $(n - k)$th trial is a success.

  (ii) In the latter case, the outcome of the $(n - k)$th trial is a failure. The first $n - k - 1$ trials contain exactly $r - 1$ success runs.

Since the events are mutually exclusive and exhaustive, we add the probabilities to obtain

$$f_{r,k,\ell}(n) = p^{k-\ell} f_{r-1,k,\ell}(n - k + \ell) \\ + p^k q g_{n-k-1,k,\ell}(r-1). \tag{54}$$

Rearranging terms and replacing n by $n + k + 1$ and r by $r + 1$ yield (53). Our expression for $g_{n,k,\ell}(r)$ is given by a sum of exactly $k(r+1)$ terms, independently of n. Note, however, that $f_{r+1,k,\ell}(n + k + 1)$ and $f_{r,k,\ell}(n + \ell + 1)$ must be calculated for each r. In practice, this means we must calculate $f_{r,k,\ell}(n)$ in (52) for $r = 1, \ldots, r_*$. This requires a total of $r_*(r_* + 1)/2$ sums, to obtain the full probability mass distribution. $\square$

We now summarize results for the p.m.f. and p.g.f. derived by other authors. Aki and Hirano (2000) [18, Proposition 2.2] derived an expression for the p.g.f. as a nested sum of multinomial terms. Makri and Philippou [11, Theorem 2.1] derived the p.m.f. $P(N_{n,k,\ell} = x)$, $x = 0, 1, \ldots, \lfloor (n - \ell)/(k - \ell) \rfloor$ as a sum of multinomial terms. They also derived an alternative expression for the p.m.f. [11, Theorem 2.2] in terms of $C(n, m, r)$, which is the number of possible ways of distributing n identical balls into m urns such that the maximum allowed number of balls in any one urn is r [11, Lemma 2.1]. They also calculated the mean [11, Proposition 2.1]

$$\mathbb{E}\left(N_{n,k,\ell}\right)$$

$$= p^\ell \sum_{r=1}^{r_*} \left\{ 1 + (1 - p)\left[n - \ell - r(k - \ell)\right] \right\} p^{r(k-\ell)}. \tag{55}$$

The special case $\ell = 0$ is Proposition 2.4 of Aki and Hirano (1988) [19]; see also Antzoulakos and Chadjiconstantinidis [20]. The special case $\ell = k - 1$ is equivalent to Theorem 4.1(i) of Ling (1988) [13]

$$\mathbb{E}\left(N_{n,k,k-1}\right) = (n - k + 1) p^k. \tag{56}$$

Ling (1988) [13, Theorem 3.1] gave a recursive relation for the p.m.f. and also an explicit expression for the p.m.f.

[13, Theorem 3.2], in terms of nested multinomial sums. Ling derived the mean [13, Theorem 4.1(i)] and the variance [13, Theorem 4.1(ii)] and a recurrence relation for the m.g.f. [13, Theorem 4.1(iii)]. Inoue and Aki [21, Proposition 3] derived an explicit expression for the p.g.f. $\phi_n^{(+)}(z)$ in terms of restricted multiple sums and multinomials. They stated that their expression for the special case $\ell = k - 1$ was derived by Inoue and Aki [21, Proposition 4]. Hirano et al. [10] studied the case $\ell = k - 1$ in some detail. They give an explicit expression for the p.g.f. $\phi_n(t)$ in terms of restricted multiple sums and multinomials [10, Theorem 2.2]. Hirano et al. [10, Theorem 2.3] give an explicit expression for the p.m.f. in terms of nested multinomial sums, but different from Ling (1988) [13, Theorem 3.1]. Han and Aki [22, Theorem 2.1] presented a recurrence formula to calculate the p.m.f.

3.4. Success Runs with $\ell < 0$. The case $\ell < 0$ is not without interest. In this scenario, there must be a gap or buffer of at least $|\ell|$ trials (of arbitrary outcomes) between success runs. We call this scenario "$|\ell|$-buffering." First, we derive the probability mass function, probability generating function, mean, and variance of the negative binomial distribution of order k for $r \geq 1$ success runs of length k with ℓ-overlapping. Next, we treat sequences with a fixed total length n and study the binomial distribution of order k with buffer $|\ell|$. The value of r of the number of success runs spans the interval $0 \leq r \leq \lfloor (n - \ell)/(k - \ell) \rfloor$. This is the same formula as for $\ell \geq 0$. We derive an expression for the probability mass function for the above distribution.

Most of the published literature for the case $\ell < 0$ has treated sequences of fixed length n. Inoue and Aki [21, Section 4.2] published results for sequences of Markov trials. They derived an expression for the p.g.f. as a nested sum of multinomial terms [21, Proposition 4]. Han and Aki [22] treated sequences of i.i.d. Bernoulli trials. They derived a recurrence relation for the p.g.f. [22, Theorem 4.1].

The results for the negative binomial case (fixed $r \geq 1$, variable $n \geq 1$) are straightforward to derive for $\ell < 0$. The following results are stated without proof.

Proposition 19. *For $\ell < 0$, the probability mass function $f_{r,k,\ell}(n)$ for r success runs of length k with $|\ell|$-buffering satisfies the obvious identity*

$$f_{r,k,\ell}(n) = f_{r,k,0}(n + (r - 1)\ell). \tag{57}$$

The probability generating function is then given by

$$\phi_{r,k,\ell}(s) = \phi_{r,k,0}(s) s^{(r-1)\ell} = \phi(s, k)^r s^{(r-1)\ell}. \tag{58}$$

The domain of convergence of the p.g.f. is clearly the same as in the case $\ell \geq 0$ (see (28)) and is $|s| < 1/\lambda_0(p, k)$. It follows easily from (58) that the mean and variance are given by

$$\mu_{r,k,\ell} = r\mu(k) + (r - 1)\ell, \tag{59a}$$

$$\sigma_{r,k,\ell}^2 = r\sigma^2(k). \tag{59b}$$

For sequences of fixed length n, the analysis of the binomial distribution of order k is nontrivial for $\ell < 0$. We first state the following obvious result for all ℓ.

Remark 20. For any $\ell \in \mathbb{Z}$, let $h_{r,k,\ell}(n)$ be the probability of attaining, after n trials, r or fewer success runs of length k with ℓ-overlapping for $\ell \geq 0$ or $|\ell|$-buffering for $\ell < 0$. Then clearly

$$h_{r,k,\ell}(n) = \sum_{j=1}^{\infty} f_{r,k,\ell}(n+j). \tag{60}$$

Hence, for fixed n, the probability mass function $g_{n,k,\ell}(r)$ for $r = 0, 1, \ldots, \lfloor (n-\ell)/(k-\ell) \rfloor$ is given by

$$\begin{aligned} g_{n,k,\ell}(r) &= h_{r,k,\ell}(n) - h_{r-1,k,\ell}(n) \\ &= \sum_{j=1}^{\infty} \left[f_{r,k,\ell}(n+j) - f_{r-1,k,\ell}(n+j) \right]. \end{aligned} \tag{61}$$

The above expression is valid for all $\ell \in \mathbb{Z}$ but requires the summation of an infinite series. For $\ell \geq 0$, (53) offers a more concise expression for $g_{n,k,\ell}(r)$. For $\ell < 0$, we can also derive a more concise expression for $g_{n,k,\ell}(r)$ as follows.

Proposition 21. *For fixed n and fixed $\ell < 0$, the probability mass function $g_{n,k,\ell}(r) = P(N_{n,k,\ell} = r)$ for the binomial distribution of order k with $|\ell|$-buffering is given by*

$$\begin{aligned} g_{n,k,\ell}(r) &= \frac{f_{r+1,k,\ell}(n+k-\ell) - p^{k+1} f_{r,k,\ell}(n)}{p^k q} \\ &\quad + \sum_{j=1}^{|\ell|-1} f_{r+1,k,\ell}(n+j). \end{aligned} \tag{62}$$

Here $r = 0, 1, \ldots, \lfloor (n-\ell)/(k-\ell) \rfloor$. Note by definition that $f_{r,k,\ell}(n) = 0$ for $r = 0$.

Proof. We omit the indices k and ℓ in the following. Consider $f_r(n)$, where r success runs have taken place, ending at trial n. Hence the last k outcomes are all successes. We then have the following mutually exclusive and exhaustive possibilities.

(i) The outcome of trial $n - k$ is a success. Then the $(r - 1)$th success run must end at trial $n - k + \ell$. The $|\ell|$ trials in the sequence from $n - k + \ell + 1$ through $n - k$ constitute the buffer between the two success runs. The probability of this event is $p^{k+1} f_{r-1}(n-k+\ell)$. (This event is null if $r = 1$. Note that $f_{0,k,\ell}(n-k+\ell) = 0$.)

(ii) The outcome of trial $n - k$ is a failure. Then we must attain $r - 1$ success runs by trial $n - k + \ell$. However, we must subtract the possibility that the rth success run ends at one of the trials $n - k + \ell + 1$ through $n - k - 1$. (Note that if $\ell = -1$, this set is empty.) The probability of this event is $p^k q(g_{n-k+\ell}(r-1) - \sum_{j=1}^{|\ell|-1} f_r(n-k+\ell+j))$.

The events are mutually exclusive and exhaustive (noting that there can be at most one success run completed from trials

$n - k + \ell + 1$ through $n - k - 1$ because of the buffering requirement); hence we add the probabilities to obtain

$$\begin{aligned} f_r(n) &= p^{k+1} f_{r-1}(n-k+\ell) \\ &\quad + p^k q \left[g_{n-k+\ell}(r-1) - \sum_{j=1}^{|\ell|-1} f_r(n-k+\ell+j) \right]. \end{aligned} \tag{63}$$

Rearranging terms and replacing n by $n + k - \ell$ and r by $r + 1$ yield (62). For $\ell = -1$, the last sum is absent and the above expression is the same as (53). $\qquad\square$

4. Pairs of Successes Separated by At Most $k-2$ Failures

In this section we study a different problem. We treat the distribution of waiting time for the rth nonoverlapping appearance of a pair of successes separated by at most $k - 2$ failures ($k \geq 2$). Our main reference is the elegant analysis by Koutras [7], who also gives an excellent bibliography on the subject. To avoid cluttering the notation in this paper with too many symbols, we will reuse some of the symbols such as $f_{r,k}(n)$ for the probability mass function, and so forth. It should be understood that we are treating a new problem, and the following notation is self-contained. We begin with $r = 1$. Koutras [7] gave a recurrence relation for the probability mass function $f_{r,k}(n)$. We will suppress the indices r and k unless required. We derive the exact solutions for the roots of the auxiliary polynomial associated with the recurrence relation, in terms of Fuss-Catalan numbers. We also derive various pertinent properties of the roots. We then solve a Vandermonde matrix system of equations to derive an expression for the p.m.f. as a sum over powers of the roots. We also derive an expression for the probability of the waiting time to exceed n trials.

Let us denote the waiting time by $T_{r,k}$. Note that Koutras [7] writes $T_{k,r}$, but we write $T_{r,k}$ to maintain consistency with the notation in the earlier parts of our paper. We begin with the case $r = 1$ and drop the subscripts.

Remark 22 (Koutras [7], Theorem 3.1). The probability mass function $f(n) = P(T = n)$ satisfies the recurrence relation [7, eq. 3.1]

$$f(n) = qf(n-1) + pq^{k-1} f(n-k) \quad (n > k). \tag{64}$$

The initial conditions are [7, eq. 3.2]

$$f(0) = f(1) = 0, \tag{65a}$$

$$f(n) = (n-1) p^2 q^{n-2}, \quad 1 < n \leq k. \tag{65b}$$

The auxiliary polynomial associated with the above recurrence relation is

$$\mathscr{A}(z) = z^k - qz^{k-1} - pq^{k-1}. \tag{66}$$

The auxiliary equation is $\mathscr{A}(z) = 0$.

Proposition 23 (properties of roots). *For fixed $0 < p < 1$, the roots of the auxiliary polynomial have the following properties:*

(a) *There are no repeated roots.*

(b) *There is a unique positive real root.*

(c) *The positive real root lies in $(q, 1)$.*

(d) *If k is odd, there are no other real roots. If k is even, there is exactly one negative real root.*

(e) *The magnitude of the positive real root exceeds that of all the other roots.*

Proof. Both $\mathscr{A}(z)$ and $\mathscr{A}'(z)$ must vanish simultaneously at a repeated root. Next

$$\mathscr{A}'(z) = kz^{k-1} - q(k-1)z^{k-2}$$
$$= kz^{k-2}\left[z - \frac{q(k-1)}{k}\right]. \tag{67}$$

Hence $\mathscr{A}'(z)$ vanishes at $z = 0$ (not a root of $\mathscr{A}(z)$) or $z = q(k-1)/k$. Note that $0 < q(k-1)/k < q$. Now for $0 < z < q$,

$$z^k - qz^{k-1} = (z-q)z^{k-1} < 0. \tag{68}$$

Hence for $0 < z < q$,

$$\mathscr{A}(z) = (z-q)z^{k-1} - pq^{k-1} < -pq^{k-1} < 0. \tag{69}$$

Hence $\mathscr{A}(z) \neq 0$ for $z = q(k-1)/k$. Hence $\mathscr{A}(z)$ has no repeated roots. Next note that $\mathscr{A}(0) = -pq^{k-1}$, $\mathscr{A}(q) = -pq^{k-1}$, and $\mathscr{A}(1) = p(1 - q^{k-1})$. Hence $\mathscr{A}(z)$ has an odd number of positive real roots for $z \in (q, 1)$. Now from (67), $\mathscr{A}'(z) > 0$ for $z > q(1 - 1/k)$, so $\mathscr{A}'(z) > 0$ for $z > q$. It follows that $\mathscr{A}(z)$ has exactly one positive real root, and it lies in the interval $z \in (q, 1)$. Also if k is odd then $\mathscr{A}(z) < 0$ for $z < 0$ and there are no negative real roots. If k is even then $\mathscr{A}(z)$ increases as z decreases through negative values; hence for even k, $\mathscr{A}(z)$ has exactly one negative real root. Next, if z_r is a root, by the triangle inequality,

$$\left|z_r\right|^k = \left|qz_r^{k-1} + pq^{k-1}\right| \le q\left|z_r\right|^{k-1} + pq^{k-1}. \tag{70}$$

Hence

$$\left|z_r\right|^k - q\left|z_r\right|^{k-1} - pq^{k-1} \le 0. \tag{71}$$

The inequality is strict unless z_r is real and positive (so that both z_r^k and z_r^{k-1} are real and positive) and we have shown that there is only one real positive root. Hence the real positive root has a larger magnitude than all the other roots. □

We will call the positive real root the "principal root" and refer to all the other roots as "secondary roots." We will denote the roots by $\mu_j(p, k)$, $j = 0, \ldots, k - 1$, where the principal root is μ_0. Although our the following analysis is for $0 < p < 1$, it is helpful to note the following limiting cases for $p = 0$ and $p = 1$.

Proposition 24 (limiting cases for roots). *If $p = 0$, the principal root is $\mu_0 = 1$. If $p = 1$, the principal root is $\mu_0 = 0$. All the secondary roots vanish for both $p = 0$ and $p = 1$. None of the roots vanish if $0 < p < 1$.*

Proof. We have already seen that $\mathscr{A}(0) = -pq^{k-1}$; hence obviously $z = 0$ is not a root if $0 < p < 1$. If $p = 1$, the auxiliary equation is $z^k = 0$; hence all the roots vanish. If $p = 0$, the auxiliary equation is $z^{k-1}(z - 1) = 0$, so one root is $z = 1$ and the others are all $z = 0$. Hence $\mu_0 = 1$ for $p = 0$ and $\mu_0 = 0$ for $p = 1$, and all the secondary roots vanish for both $p = 0$ and $p = 1$. □

Proposition 25 (principal root decreases monotonically with increasing p). *For fixed $k \ge 2$, let $0 < p_1 < p_2 < 1$ and denote the respective principal roots by $\mu(p_1)$ and $\mu(p_2)$. Then $\mu(p_2) < \mu(p_1)$.*

Proof. Note that the auxiliary polynomial can be expressed in the following alternative form:

$$\mathscr{A}_p(z) = z^k - qz^{k-1} - pq^{k-1}$$
$$= z^k - (1-p)z^{k-1} - pq^{k-1} \tag{72}$$
$$= z^{k-1}(z-1) + p\left(z^{k-1} - q^{k-1}\right).$$

For brevity write $\mu_* = \mu_0(p_1)$. Then by definition

$$0 = \mathscr{A}_{p_1}(\mu_*) = \mu_*^{k-1}(\mu_* - 1) + p_1\left(\mu_*^{k-1} - q_1^{k-1}\right). \tag{73}$$

Then because $q_2 < q_1$, it follows that $\mu_*^{k-1} - q_2^{k-1} > \mu_*^{k-1} - q_1^{k-1}$ and so (because $p_2 > p_1$) it also follows that $p_2(\mu_*^{k-1} - q_2^{k-1}) > p_1(\mu_*^{k-1} - q_1^{k-1})$. Hence

$$\mathscr{A}_{p_2}(\mu_*) = \mu_*^{k-1}(\mu_* - 1) + p_2\left(\mu_*^{k-1} - q_2^{k-1}\right)$$
$$= \underbrace{\mathscr{A}_{p_1}(\mu_*)}_{=0}$$
$$+ \underbrace{p_2\left(\mu_*^{k-1} - q_2^{k-1}\right) - p_1\left(\mu_*^{k-1} - q_1^{k-1}\right)}_{>0} \tag{74}$$

$$> 0.$$

Hence $\mathscr{A}_{p_2}(\mu_0(p_1)) > 0$. Since we know $\mathscr{A}_{p_2}(q_2) < 0$ and $\mathscr{A}_{p_2}(1) > 0$ and also $q_2 < q_1 < \mu_0(p_1) < 1$ and the real positive root $\mu_0(p_2)$ is unique, it follows that $\mu_0(p_2) < \mu_0(p_1)$. □

As already stated earlier in this paper, details about the Fuss-Catalan numbers can be found in the text by Graham et al. [2]. See the expressions above, in (4), (5a), and (5b), which will be essential in the following. We require the following lemma for the domain of convergence of the generating functions of the Fuss-Catalan numbers.

Lemma 26 (domain of convergence). *The generating function of the Fuss-Catalan numbers $B_\nu(z)$ converges for*

$$|z| \le \rho_* \equiv \frac{|\nu - 1|^{\nu-1}}{\nu^\nu}. \tag{75}$$

The proof was derived in our earlier paper [1, Corollary 22 and Proposition 23] and is omitted. The proof was actually derived for $B_{k+1}(z)$ but note (this is important) that k does not have to be an integer; hence we write ν above. It was also proved in [1] that the domain of convergence includes the circle of convergence.

Proposition 27 (solutions for roots using Fuss-Catalan numbers). *Define $p_* = \rho_*/(1 + \rho_*)$. For $0 < p \le p_*$, the principal and secondary roots are given by*

$$\mu_0(p,k) = \frac{q}{B_k(-p/q)}, \tag{76a}$$

$$\mu_j(p,k) = q - \frac{q}{B_{k/(k-1)}\left(e^{\pi i(2j-1)/(k-1)}(p/q)^{1/(k-1)}\right)} \tag{76b}$$

$$1 \le j \le k-1.$$

For $p_* \le p \le 1$, the principal and secondary roots are given by

$$\mu_j(p,k) = \frac{q}{1 - B_{1/k}\left(-e^{-2\pi ij/k}(q/p)^{1/k}\right)} \tag{77}$$

$$0 \le j \le k-1.$$

For $p = p_*$, either set of solutions may be employed.

Proof. It is simpler to set $z = q\zeta$ and solve for ζ. Then the auxiliary equation is

$$\zeta^k - \zeta^{k-1} - \frac{p}{q} = 0. \tag{78}$$

We also require $\omega_k = e^{2\pi i/k}$ and $\omega_{k-1} = e^{2\pi i/(k-1)}$. It is simpler to begin with the case $p_* \le p < 1$, because all the roots are given by a unified formula. We rewrite (78) in the form $\zeta^k(1 - 1/\zeta) = p/q$. Now take the kth root to obtain $\zeta(1 - 1/\zeta)^{1/k} = \omega_k^j(p/q)^{1/k}$. Next put $(1 - 1/\zeta)^{1/k} = C^{1/k}$, so $\zeta = 1/(1-C)$ and then

$$C(\zeta) - 1 = -\omega_k^{-j}\left(\frac{q}{p}\right)^{1/k} C(\zeta)^{1/k}. \tag{79}$$

We now employ (6) with $\nu = 1/k$ and $\zeta = -\omega_k^{-j}(q/p)^{1/k}$. Then

$$C(\zeta) = B_{1/k}\left(-\omega_k^{-j}\left(\frac{q}{p}\right)^{1/k}\right). \tag{80}$$

This yields (77). We must establish the set of values of p where the above result is valid. Using (75), the above solution is valid for

$$\left|\frac{q}{p}\right| \le \left(\frac{|1/k - 1|^{1/k-1}}{(1/k)^{1/k}}\right)^k = \frac{k^k}{|k-1|^{k-1}} = \frac{1}{\rho_*}. \tag{81}$$

Hence $(1-p)/p \le 1/\rho_*$ or $p_* \le p < 1$. This establishes (77).

Next we derive (76a). We rewrite (78) in the form $\zeta^{k-1}(\zeta - 1) = p/q$. Then put $\zeta = 1/\widetilde{D}$ and we obtain

$$\widetilde{D}(\zeta) - 1 = -\left(\frac{p}{q}\right)\widetilde{D}(\zeta)^k. \tag{82}$$

We now employ (6) with $\nu = k$ and $\zeta = -(p/q)$. Then

$$\widetilde{D}(\zeta) = B_k\left(-\frac{p}{q}\right). \tag{83}$$

The domain of convergence is clearly $|p/(1 - p)| \le \rho_*$ or $0 < p \le p_*$. This establishes (76a). We now solve for the remaining (secondary) roots. We again rewrite (78) in the form $\zeta^{k-1}(\zeta - 1) = p/q$. We now take the $(k - 1)$th root to obtain

$$\zeta(\zeta - 1)^{1/(k-1)} = \omega_{k-1}^j\left(\frac{p}{q}\right)^{1/(k-1)}. \tag{84}$$

Now put $(\zeta - 1)^{1/(k-1)} = (-1/D)^{1/(k-1)}$ and then $\zeta = 1 - 1/D$. Hence

$$\left(1 - D^{-1}\right)\left(\frac{1}{D}\right)^{1/(k-1)}$$
$$= (-1)^{-1/(k-1)}\,\omega_{k-1}^j\left(\frac{p}{q}\right)^{1/(k-1)}. \tag{85}$$

After some algebra this yields

$$D(\zeta) - 1 = e^{\pi i(2j-1)/(k-1)}\left(\frac{p}{q}\right)^{1/(k-1)} D(\zeta)^{k/(k-1)}. \tag{86}$$

We now employ (6) with $\nu = k/(k - 1)$ and $\zeta = e^{\pi i(2j-1)/(k-1)}(p/q)^{1/(k-1)}$. Then

$$D = B_{k/(k-1)}\left(e^{\pi i(2j-1)/(k-1)}\left(\frac{p}{q}\right)^{1/(k-1)}\right). \tag{87}$$

This yields (76b). Note that this solution yields only $k - 1$ distinct roots. If k is odd all the roots are complex. If k is even, then for $j = k/2$ we obtain $e^{\pi i(2j-1)/(k-1)} = e^{i\pi} = -1$; hence $\zeta = -(p/q)^{1/(k-1)}$, which yields the negative real root. For all k, we can index the roots by $1 \le j \le k - 1$. Using (75), the above solution is valid for

$$\left|\frac{p}{q}\right| \le \left(\frac{|k/(k-1) - 1|^{k/(k-1)-1}}{(k/(k-1))^{k/(k-1)}}\right)^{k-1} = \frac{|k-1|^{k-1}}{k^k} \tag{88}$$

$$= \rho_*.$$

Hence $p/(1 - p) \le \rho_*$ or $0 < p \le p_*$. This establishes (76b). □

We now derive an expression for the probability mass function in terms of a sum over the roots.

Proposition 28. *The probability mass function $f(n)$ is given by*

$$f(n) = p^2 q^{k-1} \sum_{j=0}^{k-1} \frac{(1 - \mu_j)\mu_j^{n-k+1}}{(\mu_j - q)^2\left(k\left(\mu_j - q\right) + q\right)}. \tag{89}$$

Proof. It is known from the theory of recurrence relations that we may express the p.m.f. in the form

$$f(n) = \sum_{i=0}^{k-1} c_i \mu_i^{n-1}. \tag{90}$$

We must solve for the coefficients c_i. Because there are no repeated roots, c_i do not depend on n. We can derive the solution using a Vandermonde matrix

$$
\begin{pmatrix}
1 & 1 & \cdots & 1 \\
\mu_0 & \mu_1 & \cdots & \mu_{k-1} \\
\mu_0^2 & \mu_1^2 & \cdots & \mu_{k-1}^2 \\
\vdots & & & \vdots \\
\mu_0^{k-1} & \mu_1^{k-1} & \cdots & \mu_{k-1}^{k-1}
\end{pmatrix}
\begin{pmatrix}
c_0 \\
c_1 \\
c_2 \\
\vdots \\
c_{k-1}
\end{pmatrix}
\tag{91}
$$

$$
= p^2
\begin{pmatrix}
0 \\
1 \\
2q \\
\vdots \\
(k-1)q^{k-2}
\end{pmatrix}.
$$

To solve this we need the Lagrange basis polynomials $L_i(z)$. The Lagrange basis polynomials have degree $k-1$ and have the property $L_i(\mu_j) = \delta_{ij}$ for $i, j = 0, \dots, k-1$. They can be written as

$$L_i(z) = \frac{\prod_{j \neq i} (z - \mu_j)}{\prod_{j \neq i} (\mu_i - \mu_j)}. \tag{92}$$

Let us write $L_i(z) = \sum_{j=0}^{k-1} L_{ij} z^j$. Then

$$c_i = \sum_{j=0}^{k-1} L_{ij} f(j+1) = p^2 \sum_{j=0}^{k-1} L_{ij} j q^{j-1} = p^2 L_i'(q). \tag{93}$$

Next express the ratio in (92) as $L_i(z) = N_i(z)/D_i$. Note that $D_i = N_i(\mu_i)$. Then

$$D_i = \mathscr{A}'(\mu_i) = (k(\mu_i - q) + q) \mu_i^{k-2}. \tag{94}$$

Next $N_i(z) = \mathscr{A}(z)/(z - \mu_i)$ so

$$N_i(z) = \frac{z^k - qz^{k-1} - pq^{k-1}}{z - \mu_i}. \tag{95}$$

Differentiate with respect to z to obtain

$$N_i'(z) = \frac{(kz - (k-1)q) z^{k-2}}{z - \mu_i} - \frac{z^k - qz^{k-1} - pq^{k-1}}{(z - \mu_i)^2}. \tag{96}$$

Then after some algebra we obtain

$$N_i'(q) = \frac{(kq - (k-1)q) q^{k-2}}{q - \mu_i} + \frac{pq^{k-1}}{(q - \mu_i)^2} \tag{97}$$

$$= \frac{(1 - \mu_i) q^{k-1}}{(q - \mu_i)^2}.$$

Hence the coefficient c_i is given by

$$c_i = p^2 L_i'(q) = \frac{(1 - \mu_i)}{(\mu_i - q)^2 (k(\mu_i - q) + q)} \frac{p^2 q^{k-1}}{\mu_i^{k-2}}. \tag{98}$$

Then

$$f(n) = p^2 q^{k-1} \sum_{j=0}^{k-1} \frac{(1 - \mu_j) \mu_j^{n-k+1}}{(\mu_j - q)^2 (k(\mu_j - q) + q)}. \tag{99}$$

This establishes (89). $\qquad\square$

The probability generating function $\phi(s)$ was derived by Koutras [7, Theorem 3.2]:

$$\phi(s) = \frac{(ps)^2}{1 - qs - pq^{k-1}s^k} \sum_{i=0}^{k-2} (qs)^i$$

$$= \frac{(ps)^2}{1 - qs - pq^{k-1}s^k} \frac{1 - (qs)^{k-1}}{1 - qs}, \quad |s| \leq 1. \tag{100}$$

Koutras stated that the p.g.f. exists for $|s| \leq 1$. Using our solutions for the roots above, we can state the full domain of convergence. Note that

$$\phi(s) = \sum_{n=0}^{\infty} s^n f(n) = \sum_{j=0}^{k-1} \frac{c_j}{1 - \mu_j s}. \tag{101}$$

The series in the above sum converge for $|s| < \min\{1/\mu_j\}$, $j = 0, \dots, k-1$. Since the principal root μ_0 has the largest magnitude of all the roots, the p.g.f. in (100) exists in the domain

$$|s| < \frac{1}{\mu_0(p, k)}. \tag{102}$$

This is a larger domain than $|s| \leq 1$ and is clearly the maximal domain where the p.g.f. exists. We can also see this using (100). The denominator of $\phi(s)$ is $\mathscr{A}(1/s) = s^{-k} \prod_{j=0}^{k-1}(1 - \mu_j s)$. Then $\phi(s)$ exists in an open neighborhood of $s = 0$ until one of the factors in the product vanishes, which also yields $|s| < \min\{1/\mu_j\}$, $j = 0, \dots, k-1$, thence (102). We have given an explicit expression for the principal root $\mu_0(p, k)$ in Proposition 27.

The probability for the waiting time $P(T > n)$ is clearly given by the sum $P(T > n) = \sum_{i=1}^{\infty} f(n + i)$, so

$$P(T > n) = p^2 q^{k-1} \sum_{j=0}^{k-1} \frac{\mu_j^{n-k+2}}{(\mu_j - q)^2 (k(\mu_j - q) + q)}. \tag{103}$$

Asymptotically, the sum is dominated by the principal root; hence for sufficiently large n we may retain only the principal root

$$P(T > n) \asymp \frac{p^2 q^{k-1} \mu_0^{n-k+2}}{(\mu_0 - q)^2 (k(\mu_0 - q) + q)}. \tag{104}$$

Next we treat the case of $r > 1$ nonoverlapping appearances of a pair of successes separated by at most $k - 2$ failures. Following Koutras [7], the probability mass function and probability generating function are defined via (also following Koutras we drop the subscript k and write $T_r = T_{r,k}$, etc.)

$$f_r(n) = P(T_r = n), \tag{105a}$$

$$\phi_r(s) = \sum_{n=0}^{\infty} s^n f_r(n). \tag{105b}$$

Following the procedure in the earlier part of our paper, we will derive an expression for the p.m.f. $f_r(n)$ via the p.g.f. $\phi_r(s)$ and a partial fraction decomposition. Since the enumeration is nonoverlapping and the trials are i.i.d. random variables, the p.g.f. is given by [7, Theorem 4.1] $\phi_r(s) = \phi(s)^r$. Then we set $z = 1/s$ and obtain

$$\psi_r(z) = \phi_r\left(\frac{1}{z}\right) = p^{2r} \left[\frac{\sum_{i=0}^{k-2} q^i z^{k-2-i}}{z^k - qz^{k-1} - pq^{k-1}} \right]^r. \tag{106}$$

This is a rational function of two polynomials in z. The numerator of the term inside the brackets has degree $k - 2$ while the denominator has degree k. Note also that the numerator equals $(z^{k-1} - q^{k-1})/(z - q)$ and vanishes at $z_* = qe^{2\pi ij/(k-1)}$, $j = 1, \ldots, k - 1$. However, at such values the denominator equals $(z_* - 1)q^{k-1}$, which is nonzero. Hence the numerator and denominator have no roots in common. Furthermore, the denominator polynomial has no repeated roots. Hence we may employ exactly the same reasoning as was used to derive (49), to obtain

$$\psi_r(z) = \sum_{j=0}^{k-1} \sum_{m=1}^{r} \frac{\tilde{a}_{jm}}{(z - \mu_j)^m}. \tag{107}$$

Here we employ the notation $\tilde{a}_{jm}$ to avoid confusion with a_{jm} in (49). The coefficients $\tilde{a}_{jm}$ depend on r and k but not on z. As with a_{jm}, the coefficients $\tilde{a}_{jm}$ are given via the standard residues formula

$$\tilde{a}_{jm} = \frac{1}{(r-m)!} \left[\frac{d^{r-m}}{dz^{r-m}} \left((z - \mu_j)^r \psi_r(z) \right) \right]_{z=\mu_j}. \tag{108}$$

Then (see (51))

$$\phi_r(s) = \psi_r\left(\frac{1}{s}\right) = \sum_{j=0}^{k-1} \sum_{m=1}^{r} \frac{\tilde{a}_{jm} s^m}{(1 - \mu_j s)^m}. \tag{109}$$

We expand the right hand side using the negative binomial theorem and equate $f_r(n)$ to the coefficient of s^n

$$f_r(n) = \sum_{j=0}^{k-1} \sum_{m=1}^{r} \binom{n-1}{m-1} \tilde{a}_{jm} \mu_j^{n-m}. \tag{110}$$

Hence $f_r(n)$ is given by a sum of exactly rk terms, independently of n. Note that $f_r(n) = 0$ for $n < 2r$, because of the nonoverlapping enumeration. An alternative method to

derive the probability mass function $f_r(n)$ for $r > 1$ was given by Koutras, where the p.m.f. is obtained via a recurrence relation (unnumbered equations after Theorem 4.1 in [7]). Our expression in (110) is essentially the solution of that recurrence.

Next consider a fixed number of trials n. Following Koutras [7], let $N_{n,k}$ denote the number of occurrences of a strand of k (at most) trials containing two successes in the first n outcomes. Then [7, eq. 2.1]

$$P(N_{n,k} \geq r) = P(T_{r,k} \leq n). \tag{111}$$

Note that trivially $P(N_{n,k} \geq r) = 1$ for $r = 0$. For $r \geq 1$, the right hand side is easily evaluated using (110). Clearly $P(N_{n,k} \geq r) = 0$ and $P(T_{r,k} \leq n) = 0$ for $n < 2r$, because of the nonoverlapping enumeration. Then for $r \geq 1$ and $n \geq 2r$, using (110),

$$P(N_{n,k} \geq r) = \sum_{i=2r}^{n} f_{r,k}(i)$$

$$= \sum_{i=2r}^{n} \sum_{j=0}^{k-1} \sum_{m=1}^{r} \binom{i-1}{m-1} \tilde{a}_{jm} \mu_j^{i-m}. \tag{112}$$

We now use (17) and (18) to evaluate the sum over i (note that $m \leq r$ so all the binomial coefficients are nonzero, also $\mu_j \neq 0$):

$$S_{jm} \equiv \sum_{i=2r}^{n} \binom{i-1}{m-1} \mu_j^{i-m} = \frac{1}{(m-1)! \mu_j^{m-1}} \left[\frac{d^{m-1}}{dx^{m-1}} \right.$$

$$\cdot \left. \frac{1 - (\mu_j x)^n}{1 - \mu_j x} \right]_{x=1} - \left[\frac{d^{m-1}}{dx^{m-1}} \frac{1 - (\mu_j x)^{2r-1}}{1 - \mu_j x} \right]_{x=1}$$

$$= \frac{1}{(m-1)! \mu_j^{m-1}} \left[\frac{d^{m-1}}{dx^{m-1}} \right.$$

$$\cdot \left. \frac{(\mu_j x)^{2r-1} - (\mu_j x)^n}{1 - \mu_j x} \right]_{x=1} = \frac{1}{\mu_j^{m-1}}$$

$$\cdot \sum_{i=0}^{m-1} \left[\mu_j^{2r-1} \binom{2r-1}{i} - \mu_j^n \binom{n}{i} \right] \frac{\mu_j^{m-1-i}}{(1 - \mu_j)^{m-i}}$$

$$= \sum_{i=0}^{m-1} \left[\mu_j^{2r-1} \binom{2r-1}{i} - \mu_j^n \binom{n}{i} \right] \frac{1}{\mu_j^i (1 - \mu_j)^{m-i}}. \tag{113}$$

This is a sum of exactly m terms, independently of n. Note also that $\mu_j \neq 1$ for $0 \leq j \leq k - 1$; hence the right hand side is well defined for all the roots. Substituting in (112) yields

$$P(N_{n,k} \geq r) = \sum_{j=0}^{k-1} \sum_{m=1}^{r} \binom{i-1}{m-1} \tilde{a}_{jm} S_{jm}. \tag{114}$$

Because each S_{jm} is itself a sum of m terms, the number of summands on the right hand side is $O(kr^2)$. However, the overall computational complexity is independent of n.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this work.

Acknowledgment

References

[1] S. J. Dilworth and S. R. Mane, "Success run waiting times and Fuss-Catalan numbers," *Journal of Probability and Statistics*, vol. 2015, Article ID 482462, 11 pages, 2015.

[2] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, New York, NY, USA, 2nd edition, 1994.

[3] N. Balakrishnan and M. V. Koutras, *Runs and Scans with Applications*, Wiley, New York, NY, USA, 2002.

[4] N. L. Johnson, A. W. Kemp, and S. Kotz, *Univariate Discrete Distributions*, Wiley-Interscience, Hoboken, NJ, USA, 3rd edition, 2005.

[5] W. Feller, *An Introduction to Probability Theory and Its Applications*, Wiley, New York, NY, USA, 3rd edition, 1968.

[6] A. N. Philippou and D. L. Antzoulakos, "Distributions of order k," in *International Encyclopedia of Statistical Science*, M. Lovric, Ed., pp. 400–402, Springer, Berlin, Germany, 2011.

[7] M. V. Koutras, "On a waiting time distribution in a sequence of Bernoulli trials," *Annals of the Institute of Statistical Mathematics*, vol. 48, no. 4, pp. 789–806, 1996.

[8] A. N. Philippou, "The negative binomial distribution of order k and some of its properties," *Biometrical Journal*, vol. 26, no. 7, pp. 789–794, 1984.

[9] K. D. Ling, "A new class of negative binomial distributions of order k," *Statistics & Probability Letters*, vol. 7, no. 5, pp. 371–376, 1989.

[10] K. Hirano, S. Aki, N. Kashiwagi, and H. Kuboki, "On Ling's binomial and negative binomial distributions of order k," *Statistics and Probability Letters*, vol. 11, no. 6, pp. 503–509, 1991.

[11] F. S. Makri and A. N. Philippou, "On binomial and circular binomial distributions of order k for l-overlapping success runs of length k," *Statistical Papers*, vol. 46, no. 3, pp. 411–432, 2005.

[12] F. S. Makri, A. N. Philippou, and Z. M. Psillakis, "On l-overlapping success runs distributions of order k," in *Proceedings of the 20th Panhellenic Statistics Conference*, pp. 479–487, Nicosia, Cyprus, April 2007.

[13] K. D. Ling, "On binomial distributions of order k," *Statistics & Probability Letters*, vol. 6, no. 4, pp. 247–250, 1988.

[14] E. J. Burr and G. Cane, "Longest run of consecutive observations having a specified attribute," *Biometrika*, vol. 48, pp. 461–465, 1961.

[15] A. P. Godbole, "Specific formulae for some success run distributions," *Statistics and Probability Letters*, vol. 10, no. 2, pp. 119–124, 1990.

[16] A. N. Philippou and F. S. Makri, "Longest success runs and fibonacci-type polynomials," *The Fibonacci Quarterly*, vol. 23, no. 4, pp. 338–346, 1985.

[17] M. Muselli, "Simple expressions for success run distributions in Bernoulli trials," *Statistics & Probability Letters*, vol. 31, no. 2, pp. 121–128, 1996.

[18] S. Aki and K. Hirano, "Numbers of success-runs of specified length until certain stopping time rules and generalized binomial distributions of order k," *Annals of the Institute of Statistical Mathematics*, vol. 52, no. 4, pp. 767–777, 2000.

[19] S. Aki and K. Hirano, "Some characteristics of the binomial distribution of order k and related distributions," in *Statistical Theory and Data Analysis II*, pp. 211–222, North-Holland Publishing, Amsterdam, The Netherlands, 1988.

[20] D. L. Antzoulakos and S. Chadjiconstantinidis, "Distributions of numbers of success runs of fixed length in Markov dependent trials," *Annals of the Institute of Statistical Mathematics*, vol. 53, no. 3, pp. 599–619, 2001.

[21] K. Inoue and S. Aki, "Generalized binomial and negative binomial distributions of order k by the l-overlapping enumeration scheme," *Annals of the Institute of Statistical Mathematics*, vol. 55, no. 1, pp. 153–167, 2003.

[22] S. Han and S. Aki, "A unified approach to binomial-type distributions of order k," *Communications in Statistics—Theory and Methods*, vol. 29, no. 8, pp. 1929–1943, 2000.

Classical and Bayesian Approach in Estimation of Scale Parameter of Nakagami Distribution

Kaisar Ahmad,[1] S. P. Ahmad,[1] and A. Ahmed[2]

[1]*Department of Statistics, University of Kashmir, Srinagar, Jammu and Kashmir 190006, India*
[2]*Department of Statistics and O.R., Aligarh Muslim University, Aligarh, India*

Correspondence should be addressed to Kaisar Ahmad; ahmadkaisar31@gmail.com

Nakagami distribution is considered. The classical maximum likelihood estimator has been obtained. Bayesian method of estimation is employed in order to estimate the scale parameter of Nakagami distribution by using Jeffreys', Extension of Jeffreys', and Quasi priors under three different loss functions. Also the simulation study is conducted in R software.

1. Introduction

Nakagami distribution can be considered as a flexible lifetime distribution. It has been used to model attenuation of wireless signals traversing multiple paths (for details see Hoffman [1]), fading of radio signals, data regarding communicational engineering, and so forth. The distribution may also be employed to model failure times of a variety of products (and electrical components) such as ball bearing, vacuum tubes, and electrical insulation. It is also widely considered in biomedical fields, such as to model the time to the occurrence of tumors and appearance of lung cancer. It has the applications in medical imaging studies to model the ultrasounds especially in Echo (heart efficiency test). Shanker et al. [2] and Tsui et al. [3] use the Nakagami distribution to model ultrasound data in medical imaging studies. This distribution is extensively used in reliability theory and reliability engineering and to model the constant hazard rate portion because of its memory less property. Yang and Lin [4] investigated and derived the statistical model of spatial-chromatic distribution of images. Through extensive evaluation of large image databases, they discovered that a two-parameter Nakagami distribution well suits the purpose. Kim and Latchman [5] used the Nakagami distribution in their analysis of multimedia.

The probability density function (pdf) of the Nakagami distribution is given as mentioned in Figure 1:

$$f(y; \theta, k) = \frac{2k^k}{\Gamma(k)\theta^k} y^{2k-1} \exp\left(\frac{-ky^2}{\theta}\right) \tag{1}$$

$$\text{for } y > 0, \ k, \theta > 0,$$

where θ and k are the scale and the shape parameters, respectively.

2. Materials and Methods

There are two main philosophical approaches to statistics. The first is called the classical approach which was founded by Professor R. A. Fisher in a series of fundamental papers round about 1930. In classical approach we use the same method as obtained by Ahmad et al. [6].

The alternative approach is the Bayesian approach which was first discovered by Reverend Thomas Bayes. In this approach, parameters are treated as random variables and data is treated as fixed. Recently Bayesian estimation approach has received great attention by most researchers among them are Al-Aboud [7] who studied Bayesian estimation for the extreme value distribution using progressive censored data and asymmetric loss. Ahmed et al. [8] considered
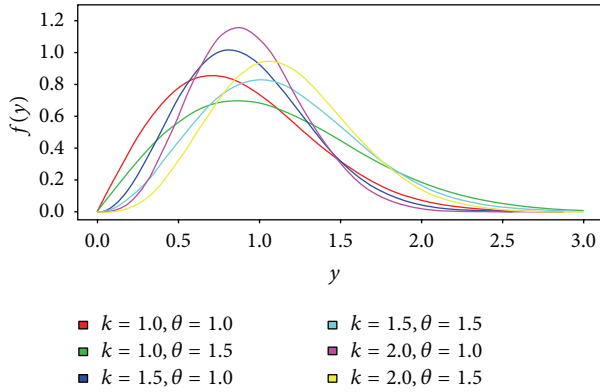
FIGURE 1: The pdf's of Nakagami distribution under various values of k and theta.

Bayesian Survival Estimator for Weibull distribution with censored data. An important prerequisite in this approach is the appropriate choice of prior(s) for the parameters. Very often, priors are chosen according to one's subjective knowledge and beliefs. The other integral part of Bayesian inference is the choice of loss function. A number of symmetric and asymmetric loss functions have been shown to be functional; see Pandey et al. [9], Al-Athari [10], S. P. Ahmad and K. Ahmad [11], Ahmad et al. [12, 13], and so forth.

Theorem 1. *Let $(y_1, y_2, \ldots, y_n)$ be a random sample of size n having pdf (1); then the maximum likelihood estimator of scale parameter θ, when the shape parameter k is known, is given by*

$$\widehat{\theta} = \frac{\sum_{i=1}^{n} y_i^2}{n}. \tag{2}$$

Proof. The likelihood function of the pdf (1) is given by

$$L(y; \theta, k) = \frac{(2k^k)^n}{(\Gamma k)^n \theta^{nk}} \prod_{i=1}^{n} y_i^{2k-1} \exp\left(\frac{-k}{\theta} \sum_{i=1}^{n} y_i^2\right). \tag{3}$$

The log likelihood function is given by

$$\log L(y; \theta, k) = n \log(2k^k) - n \log \Gamma k - nk \log \theta$$
$$+ (2k-1) \sum_{i=1}^{n} \log y_i - \frac{k}{\theta} \sum_{i=1}^{n} y_i^2. \tag{4}$$

Differentiating (4) with respect to θ and equating to zero, we get

$$\widehat{\theta} = \frac{\sum_{i=1}^{n} y_i^2}{n}. \tag{5}$$

□

2.1. Loss Functions Used in This Paper. (i) The quadratic loss function which is given by

$$L_{qd}(\widehat{\theta}, \theta) = \left(\frac{(\widehat{\theta} - \theta)}{\theta}\right)^2; \quad \theta > 0, \tag{6}$$

which is a symmetric loss function; θ and $\widehat{\theta}$ represent the true and estimated values of the parameter.

(ii) The Al-Bayyati new loss function is of the form

$$L_{nl}(\widehat{\theta}, \theta) = \theta^{c_1}(\widehat{\theta} - \theta)^2; \quad c_1 \varepsilon R, \tag{7}$$

which is an asymmetric loss function; θ and $\widehat{\theta}$ represent the true and estimated values of the parameter.

(iii) The entropy loss function is given by

$$L_{ef}(\widehat{\theta}, \theta) = \left(\frac{\widehat{\theta}}{\theta} - \log\left(\frac{\widehat{\theta}}{\theta}\right) - 1\right); \quad \theta > 0, \tag{8}$$

where θ and $\widehat{\theta}$ represent the true and estimated values of the parameter.

3. Bayesian Method of Estimation

In this section Bayesian estimation of the scale parameter of Nakagami distribution is obtained by using various priors under different symmetric and asymmetric loss functions.

3.1. Posterior Density under Jeffreys' Prior. Let $(y_1, y_2, \ldots, y_n)$ be a random sample of size n having the probability density function (1) and the likelihood function (2).

Jeffreys' prior for θ is given by

$$g(\theta) = \frac{1}{\theta}; \quad \theta > 0. \tag{9}$$

By using the Bayes theorem, we have

$$\pi_1(\theta \mid y) \propto L(y \mid \theta) g(\theta). \tag{10}$$

Using (2) and (9) in (10),

$$\pi_1(\theta \mid y)$$
$$\propto \frac{(2k^k)^n}{\Gamma(k)^n \theta^{nk+1}} \prod_{i=1}^{n} y_i^{2k-1} \exp\left(\frac{-k}{\theta} \sum_{i=1}^{n} y_i^2\right), \tag{11}$$

$$\pi_1(\theta \mid y) = \rho \frac{1}{\theta^{nk+1}} \exp\left(\frac{-k}{\theta} \sum_{i=1}^{n} y_i^2\right),$$

where ρ is independent of θ and

$$\rho = \frac{\left(k \sum_{i=1}^{n} y_i^2\right)^{nk}}{\Gamma nk}. \tag{12}$$

Using the value of ρ in (11),

$$\pi_1(\theta \mid y)$$
$$= \left(\frac{\left(k \sum_{i=1}^{n} y_i^2\right)^{nk}}{\Gamma nk} \frac{1}{\theta^{nk+1}} \exp\left(\frac{-k}{\theta} \sum_{i=1}^{n} y_i^2\right)\right). \tag{13}$$

3.2. Posterior Density under Extension of Jeffreys' Prior. Let $(y_1, y_2, \ldots, y_n)$ be a random sample of size n having the probability density function (1) and the likelihood function (2).

The extension of Jeffreys' for θ is given by

$$g_1(\theta) = \frac{1}{\theta^{2c}}; \quad \theta > 0. \tag{14}$$

By using the Bayes theorem, we have

$$\pi_2(\theta \mid \underline{y}) \propto L(y \mid \theta) g_1(\theta). \tag{15}$$

Using (2) and (14) in (15),

$$\pi_2(\theta \mid \underline{y})$$
$$\propto \frac{(2k^k)^n}{\Gamma(k)^n \theta^{nk+2c}} \prod_{i=1}^{n} y_i^{2k-1} \exp\left(\frac{-k}{\theta} \sum_{i=1}^{n} y_i^2\right). \tag{16}$$

Thus

$$\pi_2(\theta \mid \underline{y}) = \rho \frac{1}{\theta^{nk+2c}} \exp\left(\frac{-k}{\theta} \sum_{i=1}^{n} y_i^2\right), \tag{17}$$

$$\rho = \frac{\left(k \sum_{i=1}^{n} y_i\right)^{nk+2c-1}}{\Gamma(nk + 2c - 1)}. \tag{18}$$

By using the value of ρ in (17), we have

$$\pi_2(\theta \mid \underline{y}) = \left(\frac{\left((-k/\theta) \sum_{i=1}^{n} y_i^2\right)^{nk+2c-1}}{\Gamma(nk + 2c - 1)} \frac{1}{\theta^{nk+2c}}\right.$$
$$\left. \cdot \exp\left(\frac{-k}{\theta} \sum_{i=1}^{n} y_i^2\right)\right). \tag{19}$$

3.3. Posterior Density under Quasi Prior. Let $(y_1, y_2, \ldots, y_n)$ be a random sample of size n having the probability density function (1) and the likelihood function (2).

Quasi prior for θ is given by

$$g_2(\theta) = \frac{1}{\theta^d}; \quad \theta > 0, \ d > 0. \tag{20}$$

By using the Bayes theorem, we have

$$\pi_3(\theta \mid \underline{y}) \propto L(y \mid \theta) g_2(\theta). \tag{21}$$

Using (2) and (20) in (21),

$$\pi_3(\theta \mid \underline{y}) \propto \frac{(2k^k)^n}{\Gamma(k)^n \theta^{nk+d}} \prod_{i=1}^{n} y_i^{2k-1} \exp\left(\frac{-k}{\theta} \sum_{i=1}^{n} y_i^2\right) \tag{22}$$

$$\pi_3(\theta \mid \underline{y}) = \rho \frac{1}{\theta^{nk+d}} \exp\left(\frac{-k}{\theta} \sum_{i=1}^{n} y_i^2\right),$$

where ρ is independent of θ and

$$\rho = \frac{\left(k \sum_{i=1}^{n} y_i^2\right)^{nk+d-1}}{\Gamma(nk + d - 1)}. \tag{23}$$

Using the value of ρ in (22),

$$\pi_3(\theta \mid \underline{y})$$
$$= \left(\frac{\left(k \sum_{i=1}^{n} y_i^2\right)^{nk+d-1}}{\Gamma(nk + d - 1)} \frac{1}{\theta^{nk+d}} \exp\left(\frac{-k}{\theta} \sum_{i=1}^{n} y_i^2\right)\right). \tag{24}$$

4. Bayesian Estimation by Using Jeffreys' Prior under Different Loss Functions

Theorem 2. *Assuming the loss function $L_{qd}(\widehat{\theta}, \theta)$, the Bayes estimate of the scale parameter θ, if the shape parameter k is known, is of the form*

$$\widehat{\theta}_{qd} = \frac{\left(k \sum_{i=1}^{n} y_i^2\right)}{(nk + 1)}. \tag{25}$$

Proof. The risk function of the estimator θ under the quadratic loss function $L_{qd}(\widehat{\theta}, \theta)$ is given by the formula

$$R(\widehat{\theta}) = \int_0^{\infty} \left(\frac{(\widehat{\theta} - \theta)}{\theta}\right)^2 \pi_1(\theta \mid y) d\theta. \tag{26}$$

Using (13) in (26), we get

$$R(\widehat{\theta}) = \int_0^{\infty} \left(\frac{(\widehat{\theta} - \theta)}{\theta}\right)^2$$
$$\cdot \frac{\left(k \sum_{i=1}^{n} y_i^2\right)^{nk}}{\Gamma nk} \frac{1}{\theta^{nk+1}} \exp\left(\frac{-k}{\theta} \sum_{i=1}^{n} y_i^2\right) d\theta. \tag{27}$$

On solving (27), we get

$$R(\widehat{\theta}) = \frac{\widehat{\theta}^2 \Gamma(nk + 2)}{\Gamma nk \left(k \sum_{i=1}^{n} y_i^2\right)^2} - \frac{2\widehat{\theta} \Gamma(nk + 1)}{\Gamma nk \left(k \sum_{i=1}^{n} y_i^2\right)} + 1. \tag{28}$$

Minimization of the risk with respect to $\widehat{\theta}$ gives us the optimal estimator:

$$\widehat{\theta}_{qd} = \frac{\left(k \sum_{i=1}^{n} y_i^2\right)}{(nk + 1)}. \tag{29}$$

$\square$

Theorem 3. *Assuming the loss function $L_{nl}(\widehat{\theta}, \theta)$, the Bayes estimate of the scale parameter θ, if the shape parameter k is known, is of the form*

$$\widehat{\theta}_{nl} = \frac{\left(k \sum_{i=1}^{n} y_i^2\right)}{(nk - c_1 - 1)}. \tag{30}$$

Proof. The risk function of the estimator θ under the Al-Bayyati loss function $L_{nl}(\widehat{\theta}, \theta)$ is given by the formula

$$R\left(\widehat{\theta}\right) = \int_0^\infty \theta^{c_1} \left(\widehat{\theta} - \theta\right)^2 \pi_1\left(\theta \mid \underline{y}\right) d\theta. \qquad (31)$$

On substituting (13) in (31), we have

$$R\left(\widehat{\theta}\right) = \int_0^\infty \theta^{c_1} \left(\widehat{\theta} - \theta\right)^2 \left(\frac{\left(k \sum_{i=1}^n y_i^2\right)^{nk}}{\Gamma nk} \frac{1}{\theta^{nk+1}}\right. \qquad (32)$$

$$\left. \cdot \exp\left(\frac{-k}{\theta} \sum_{i=1}^n y_i^2\right)\right) d\theta.$$

Solving (32), we get

$$R\left(\widehat{\theta}\right) = \left[\frac{\widehat{\theta}^2 \left(k \sum_{i=1}^n y_i^2\right)^{c_1} \Gamma\left(nk - c_1\right)}{\Gamma nk}\right.$$

$$+ \frac{\left(k \sum_{i=1}^n y_i^2\right)^{c_1+2} \Gamma\left(nk - c_1 - 2\right)}{\Gamma nk} \qquad (33)$$

$$\left. - \frac{2\widehat{\theta}\left(k \sum_{i=1}^n y_i^2\right)^{c_1+1} \Gamma\left(nk - c_1 - 1\right)}{\Gamma nk}\right].$$

Minimization of the risk with respect to $\widehat{\theta}$ gives us the optimal estimator:

$$\widehat{\theta}_{nl} = \frac{\left(k \sum_{i=1}^n y_i^2\right)}{\left(nk - c_1 - 1\right)}. \qquad (34)$$

Theorem 4. *Assuming the loss function $L_{ef}(\widehat{\theta}, \theta)$, the Bayes estimate of the scale parameter θ, if the shape parameter k is known, is of the form*

$$\widehat{\theta}_{ef} = \frac{\left(k \sum_{i=1}^n y_i^2\right)}{nk}. \qquad (35)$$

Proof. The risk function of the estimator θ under entropy loss function $L_{ef}(\widehat{\theta}, \theta)$ is given by the formula

$$R\left(\widehat{\theta}\right) = \int_0^\infty \left(\frac{\widehat{\theta}}{\theta} - \log\left(\frac{\widehat{\theta}}{\theta}\right) - 1\right) \pi_1\left(\theta \mid \underline{y}\right) d\theta. \qquad (36)$$

Using (13) in (36), we get

$$R\left(\widehat{\theta}\right) = \int_0^\infty \left(\frac{\widehat{\theta}}{\theta} - \log\left(\frac{\widehat{\theta}}{\theta}\right) - 1\right)$$

$$\cdot \frac{\left(k \sum_{i=1}^n y_i^2\right)^{nk}}{\Gamma nk} \frac{1}{\theta^{nk+1}} \exp\left(\frac{-k}{\theta} \sum_{i=1}^n y_i^2\right) d\theta. \qquad (37)$$

On solving (37), we get

$$R\left(\widehat{\theta}\right) = \widehat{\theta}\frac{\Gamma\left(nk + 1\right)}{\Gamma\left(nk\right)\left(k \sum_{i=1}^n y_i^2\right)} - \log\left(\widehat{\theta}\right) + h\left(\theta\right) - 1. \qquad (38)$$

Minimization of the risk with respect to $\widehat{\theta}$ gives us the optimal estimator:

$$\widehat{\theta}_{ef} = \frac{\left(k \sum_{i=1}^n y_i^2\right)}{nk}. \qquad (39)$$

5. Bayesian Estimation by Using Extension Jeffreys' Prior under Different Loss Functions

Theorem 5. *Assuming the loss function $L_{qd}(\widehat{\theta}, \theta)$, the Bayes estimate of the scale parameter θ, if the shape parameter k is known, is of the form*

$$\widehat{\theta}_{qd} = \frac{\left(k \sum_{i=1}^n y_i^2\right)}{\left(nk + 2c\right)}. \qquad (40)$$

Proof. The risk function of the estimator θ under the quadratic loss function $L_{qd}(\widehat{\theta}, \theta)$ is given by the formula

$$R\left(\widehat{\theta}\right) = \int_0^\infty \left(\frac{\left(\widehat{\theta} - \theta\right)}{\theta}\right)^2 \pi_2\left(\theta \mid \underline{y}\right) d\theta. \qquad (41)$$

Using (19) in (41), we get

$$R\left(\widehat{\theta}\right) = \int_0^\infty \left(\frac{\left(\widehat{\theta} - \theta\right)}{\theta}\right)^2 \left(\frac{\left(k \sum_{i=1}^n y_i^2\right)^{nk+2c-1}}{\Gamma\left(nk + 2c - 1\right)} \frac{1}{\theta^{nk+2c}}\right. \qquad (42)$$

$$\left. \cdot \exp\left(\frac{-k}{\theta} \sum_{i=1}^n y_i^2\right)\right) d\theta.$$

On solving (42), we get

$$R\left(\widehat{\theta}\right) = \frac{\widehat{\theta}^2 \Gamma\left(nk + 2c + 1\right)}{\Gamma nk \left(k \sum_{i=1}^n y_i^2\right)^2} - \frac{2\widehat{\theta}\Gamma\left(nk + 2c\right)}{\Gamma nk \left(k \sum_{i=1}^n y_i^2\right)} + 1. \qquad (43)$$

Minimization of the risk with respect to $\widehat{\theta}$ gives us the optimal estimator:

$$\widehat{\theta}_{qd} = \frac{\left(k \sum_{i=1}^n y_i^2\right)}{\left(nk + 2c\right)}. \qquad (44)$$

Remark 6. By replacing $c = 1/2$ in (44), the same Bayes estimate is obtained as in (29).

Theorem 7. *Assuming the loss function $L_{nl}(\widehat{\theta}, \theta)$, the Bayes estimate of the scale parameter θ, if the shape parameter k is known, is of the form*

$$\widehat{\theta}_{nl} = \frac{\left(k \sum_{i=1}^n y_i^2\right)}{\left(nk + 2c - c_1 - 2\right)}. \qquad (45)$$

Proof. The risk function of the estimator θ under the Al-Bayyati loss function $L_{nl}(\widehat{\theta}, \theta)$ is given by the formula

$$R\left(\widehat{\theta}\right) = \int_0^\infty \theta^{c_1} \left(\widehat{\theta} - \theta\right)^2 \pi_2\left(\theta \mid \underline{y}\right) d\theta. \qquad (46)$$

On substituting (19) in (46), we have

$$R\left(\widehat{\theta}\right) = \int_0^\infty \theta^{c_1} \left(\widehat{\theta} - \theta\right)^2 \left(\frac{\left(k \sum_{i=1}^n y_i^2\right)^{nk}}{\Gamma nk} \frac{1}{\theta^{nk+2c}} \right.$$

(47)

$$\left. \cdot \exp\left(\frac{-k}{\theta} \sum_{i=1}^n y_i^2 \right) \right) d\theta.$$

Solving (47), we get

$$R\left(\widehat{\theta}\right) = \left[\frac{\widehat{\theta}^2 \left(k \sum_{i=1}^n y_i^2\right)^{c_1} \Gamma\left(nk + 2c - c_1 - 1\right)}{\Gamma\left(nk + 2c - 1\right)} \right.$$

$$+ \frac{\left(k \sum_{i=1}^n y_i^2\right)^{c_1+2} \Gamma\left(nk + 2c - c_1 - 3\right)}{\Gamma\left(nk + 2c - 1\right)} \quad (48)$$

$$\left. - \frac{2\widehat{\theta} \left(k \sum_{i=1}^n y_i^2\right)^{c_1+1} \Gamma\left(nk + 2c - c_1 - 2\right)}{\Gamma\left(nk + 2c - 1\right)} \right].$$

Minimization of the risk with respect to $\widehat{\theta}$ gives us the optimal estimator:

$$\widehat{\theta}_{\mathrm{nl}} = \frac{\left(k \sum_{i=1}^n y_i^2\right)}{\left(nk + 2c - c_1 - 2\right)}. \quad (49)$$

□

Remark 8. By replacing $c = 1/2$ in (49), the same Bayes estimate is obtained as in (34).

Theorem 9. *Assuming the loss function $L_{ef}(\widehat{\theta}, \theta)$, the Bayes estimate of the scale parameter θ, if the shape parameter k is known, is of the form*

$$\widehat{\theta}_{ef} = \frac{\left(k \sum_{i=1}^n y_i^2\right)}{\left(nk + 2c - 1\right)}. \quad (50)$$

Proof. The risk function of the estimator θ under entropy loss function $L_{ef}(\widehat{\theta}, \theta)$ is given by the formula

$$R\left(\widehat{\theta}\right) = \int_0^\infty \left(\frac{\widehat{\theta}}{\theta} - \log\left(\frac{\widehat{\theta}}{\theta}\right) - 1 \right) \pi_2\left(\theta \mid \underline{y}\right) d\theta. \quad (51)$$

Using (19) in (51), we get

$$R\left(\widehat{\theta}\right) = \int_0^\infty \left(\frac{\widehat{\theta}}{\theta} - \log\left(\frac{\widehat{\theta}}{\theta}\right) - 1 \right)$$

(52)

$$\cdot \frac{\left(k \sum_{i=1}^n y_i^2\right)^{nk+2c-1}}{\Gamma\left(nk + 2c - 1\right)} \frac{1}{\theta^{nk+2c}} \exp\left(\frac{-k}{\theta} \sum_{i=1}^n y_i^2 \right) d\theta.$$

On solving (52), we get

$$R\left(\widehat{\theta}\right) = \widehat{\theta} \frac{\Gamma\left(nk + 2c\right)}{\Gamma\left(nk + 2c - 1\right) \left(k \sum_{i=1}^n y_i^2\right)} - \log\left(\widehat{\theta}\right)$$

(53)

$$+ h\left(\theta\right) - 1.$$

Minimization of the risk with respect to $\widehat{\theta}$ gives us the optimal estimator:

$$\widehat{\theta}_{\mathrm{ef}} = \frac{\left(k \sum_{i=1}^n y_i^2\right)}{\left(nk + 2c - 1\right)}. \quad (54)$$

□

Remark 10. By replacing $c = 1/2$ in (54), the same Bayes estimate is obtained as in (39).

6. Bayesian Estimation by Using Quasi Prior under Different Loss Functions

Theorem 11. *Assuming the loss function $L_{qd}(\widehat{\theta}, \theta)$, the Bayes estimate of the scale parameter θ, if the shape parameter k is known, is of the form*

$$\widehat{\theta}_{qd} = \frac{\left(k \sum_{i=1}^n y_i^2\right)}{\left(nk + d\right)}. \quad (55)$$

Proof. The risk function of the estimator θ under the quadratic loss function $L_{qd}(\widehat{\theta}, \theta)$ is given by the formula

$$R\left(\widehat{\theta}\right) = \int_0^\infty \left(\frac{\left(\widehat{\theta} - \theta\right)}{\theta} \right)^2 \pi_3\left(\theta \mid \underline{y}\right) d\theta. \quad (56)$$

Using (24) in (56), we get

$$R\left(\widehat{\theta}\right) = \int_0^\infty \left(\frac{\left(\widehat{\theta} - \theta\right)}{\theta} \right)^2 \left(\frac{\left(k \sum_{i=1}^n y_i^2\right)^{nk+d}}{\Gamma\left(nk + d - 1\right)} \frac{1}{\theta^{nk+d}} \right.$$

(57)

$$\left. \cdot \exp\left(\frac{-k}{\theta} \sum_{i=1}^n y_i^2 \right) \right) d\theta.$$

On solving (57), we get

$$R\left(\widehat{\theta}\right) = \frac{\widehat{\theta}^2 \Gamma\left(nk + d + 1\right)}{\Gamma nk \left(k \sum_{i=1}^n y_i^2\right)^2} - \frac{2\widehat{\theta} \Gamma\left(nk + d\right)}{\Gamma nk \left(k \sum_{i=1}^n y_i^2\right)} + 1. \quad (58)$$

Minimization of the risk with respect to $\widehat{\theta}$ gives us the optimal estimator:

$$\widehat{\theta}_{\mathrm{qd}} = \frac{\left(k \sum_{i=1}^n y_i^2\right)}{\left(nk + d\right)}. \quad (59)$$

□

Remark 12. By replacing $d = 1$ in (59), the same Bayes estimate is obtained as in (29).

Theorem 13. *Assuming the loss function $L_{nl}(\widehat{\theta}, \theta)$, the Bayes estimate of the scale parameter θ, if the shape parameter k is known, is of the form*

$$\widehat{\theta}_{nl} = \frac{\left(\sum_{i=1}^n y_i\right)}{\left(nk + d - c_1 - 2\right)}. \quad (60)$$

Proof. The risk function of the estimator θ under the Al-Bayyati loss function $L_{nl}(\widehat{\theta}, \theta)$ is given by the formula

$$R\left(\widehat{\theta}\right) = \int_0^\infty \theta^{c_1} \left(\widehat{\theta} - \theta\right)^2 \pi_3\left(\theta \mid \underline{y}\right) d\theta. \tag{61}$$

On substituting (24) in (61), we have

$$R\left(\widehat{\theta}\right) = \int_0^\infty \theta^{c_1} \left(\widehat{\theta} - \theta\right)^2 \left(\frac{\left(k \sum_{i=1}^n y_i^2\right)^{nk+d-1}}{\Gamma\left(nk+d-1\right)} \frac{1}{\theta^{nk+d}}\right)$$
$$\cdot \exp\left(\frac{-k}{\theta} \sum_{i=1}^n y_i^2\right)\right) d\theta. \tag{62}$$

Solving (62), we get

$$R\left(\widehat{\theta}\right) = \left[\frac{\widehat{\theta}^2 \left(k \sum_{i=1}^n y_i^2\right)^{c_1} \Gamma\left(nk+d-c_1-1\right)}{\Gamma\left(nk+d-1\right)} \right.$$
$$+ \frac{\left(k \sum_{i=1}^n y_i^2\right)^{c_1+2} \Gamma\left(nk+d-c_1-3\right)}{\Gamma\left(nk+d-1\right)} \tag{63}$$
$$\left. - \frac{2\widehat{\theta}\left(k \sum_{i=1}^n y_i^2\right)^{c_1+1} \Gamma\left(nk+d-c_1-2\right)}{\Gamma\left(nk+d-1\right)} \right].$$

Minimization of the risk with respect to $\widehat{\theta}$ gives us the optimal estimator:

$$\widehat{\theta}_{nl} = \frac{\left(k \sum_{i=1}^n y_i^2\right)}{\left(nk+d-c_1-2\right)}. \tag{64}$$

$\square$

Remark 14. By replacing $d = 1$ in (64), the same Bayes estimate is obtained as in (34).

Theorem 15. *Assuming the loss function $L_{ef}(\widehat{\theta}, \theta)$, the Bayes estimate of the scale parameter θ, if the shape parameter k is known, is of the form*

$$\widehat{\theta}_{ef} = \frac{\left(k \sum_{i=1}^n y_i^2\right)}{\left(nk+d-1\right)}. \tag{65}$$

Proof. The risk function of the estimator θ under the entropy loss function $L_{ef}(\widehat{\theta}, \theta)$ is given by the formula

$$R\left(\widehat{\theta}\right) = \int_0^\infty \left(\frac{\widehat{\theta}}{\theta} - \log\left(\frac{\widehat{\theta}}{\theta}\right) - 1\right) \pi_3\left(\theta \mid \underline{y}\right) d\theta. \tag{66}$$

Using (24) in (66), we get

$$R\left(\widehat{\theta}\right) = \int_0^\infty \left(\frac{\widehat{\theta}}{\theta} - \log\left(\frac{\widehat{\theta}}{\theta}\right) - 1\right)$$
$$\cdot \left(\frac{\left(k \sum_{i=1}^n y_i^2\right)^{nk+d-1}}{\Gamma\left(nk+d-1\right)} \frac{1}{\theta^{nk+d}}\right. \tag{67}$$
$$\left. \cdot \exp\left(\frac{-k}{\theta} \sum_{i=1}^n y_i^2\right)\right) d\theta.$$

On solving (67), we get

$$R\left(\widehat{\theta}\right) = \widehat{\theta}\frac{\Gamma\left(nk+d\right)}{\Gamma\left(nk+d-1\right)\left(k \sum_{i=1}^n y_i^2\right)} - \log\left(\widehat{\theta}\right)$$
$$+ h\left(\theta\right) - 1. \tag{68}$$

Minimization of the risk with respect to $\widehat{\theta}$ gives us the optimal estimator:

$$\widehat{\theta}_{ef} = \frac{\left(k \sum_{i=1}^n y_i^2\right)}{\left(nk+d-1\right)}. \tag{69}$$

$\square$

Remark 16. By replacing $d = 1$ in (69), the same Bayes estimate is obtained as in (39).

7. Results and Discussion

We primarily studied the classical maximum likelihood estimation and Bayesian estimation for Nakagami distribution using Jeffreys', extension of Jeffreys', and Quasi priors under three different symmetric and asymmetric loss functions. Here our main focus was to find out the estimate of scale parameter for Nakagami distribution. The mathematical derivations were checked by using the different data sets and the estimate was obtained.

For descriptive manner, we generate different random samples of size 25, 50, and 100 to represent small, medium, and large data set for the Nakagami distribution in R Software; a simulation study was carried out 3,000 times for each pairs of (θ, k) where $(k = 0.5, 1.0)$ and $(\theta = 1.0, 1.5)$. The values of extension were $(C = 0.5, 1.0)$ and $(d = 1.0, 1.5)$. The value for the loss parameter was $(C_1 = -1$ and $1)$. This was iterated 2000 times and the estimates of scale parameter for each method were calculated. The results are presented in (Tables 1, 2, and 3), respectively.

8. Conclusion

In this paper we have generated three types of data sets with different sample sizes for Nakagami distribution. These data sets were simulated with the help of programs and the behavior of the data was checked in case of parameter estimation for Nakagami distribution in R Software. With these data sets we have obtained the estimate of scale parameter for Nakagami

TABLE 1: Estimates by using Jeffreys' prior under three different loss functions.

n	k	θ	θ_{ML}	θ_{qd}	θ_{ef}	θ_{nl}	
						$C_1 = -1$	$C_1 = 1$
25	0.5	1.0	221.9361	205.4964	221.9361	221.9361	264.2096
	1.0	1.5	20.05983	19.2883	20.05983	20.05983	21.80416
50	0.5	1.0	354.8246	341.1775	354.8246	354.8246	385.6789
	1.0	1.5	49.986	49.00588	49.986	49.986	52.06875
100	0.5	1.0	863.8767	846.938	863.8767	863.8767	899.8716
	1.0	1.5	122.1739	120.9643	122.1739	122.1739	124.6672

ML: maximum likelihood, qd: quadratic loss function, ef: entropy loss function, and nl: Al-Bayyati's new loss function.

TABLE 2: Estimates by using Extension Jeffreys' prior under three different loss functions.

n	k	θ	C	θ_{ML}	θ_{qd}	θ_{ef}	θ_{nl}	
							$C_1 = -1.0$	$C_1 = 1.0$
25	0.5	1.0	0.5	221.9361	205.4964	221.9361	221.931	264.2096
			1.0	221.9361	191.3242	205.4964	205.494	241.2349
	1.0	1.5	0.5	20.05983	19.2883	20.05983	20.05983	21.80416
			1.0	20.05983	18.57392	19.2883	19.2883	20.89565
50	0.5	1.0	0.5	354.8246	341.1775	354.8246	354.8246	385.6789
			1.0	354.8246	328.5413	341.1775	341.1775	369.6089
	1.0	1.5	0.5	49.986	49.00588	49.986	49.986	52.06875
			1.0	49.986	48.06346	49.00588	49.00588	51.00612
100	0.5	1.0	0.5	863.8767	846.938	863.8767	863.8767	899.8716
			1.0	863.8767	830.6507	846.938	846.938	881.5069
	1.0	1.5	0.5	122.1739	120.9643	122.1739	122.1739	124.6672
			1.0	122.1739	119.7783	120.9643	120.9643	123.408

ML: maximum likelihood, qd: quadratic loss function, ef: entropy loss function, and nl: Al-Bayyati's new loss function.

TABLE 3: Estimates by using Quasi prior under three different loss functions.

n	k	θ	d	θ_{ML}	θ_{qd}	θ_{ef}	θ_{nl}	
							$C_1 = -1$	$C_1 = 1.0$
25	0.5	1.0	1.0	221.9361	205.4964	221.9361	221.9361	264.2096
			1.5	221.9361	198.1572	213.4001	213.4001	252.2001
	1.0	1.5	1.0	20.05983	19.2883	20.05983	20.05983	21.80416
			1.5	20.05983	18.92437	19.6665	19.6665	21.34024
50	0.5	1.0	1.0	354.8246	341.1775	354.8246	354.8246	385.6789
			1.5	354.8246	334.7401	347.8672	347.8672	377.4729
	1.0	1.5	1.0	49.986	49.00588	49.986	49.986	52.06875
			1.5	49.986	48.5301	49.49109	49.49109	51.53196
100	0.5	1.0	1.0	863.8767	846.938	863.8767	863.8767	899.8716
			1.5	863.8767	838.7153	855.3235	855.3235	890.5946
	1.0	1.5	1.0	122.1739	120.9643	122.1739	122.1739	124.6672
			1.5	122.1739	120.3684	121.5661	121.5661	124.0344

ML: maximum likelihood, qd: quadratic loss function, ef: entropy loss function, and nl: Al-Bayyati's new loss function.

distribution under three different symmetric and asymmetric loss functions by using three different priors. With the help of these results we can also do comparison between loss functions and the priors.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] M. Nakagami, "The *m*-distribution—a general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation: Proceedings of a Symposium Held at the University of California, Los Angeles, June 18–20, 1958*, W. C. Hoffman, Ed., pp. 3–36, Pergamon Press, Oxford, UK, 1960.

[2] A. K. Shanker, C. Cervantes, H. Loza-Tavera, and S. Avudainayagam, "Chromium toxicity in plants," *Environment International*, vol. 31, no. 5, pp. 739–753, 2005.

[3] P.-H. Tsui, C.-C. Huang, and S.-H. Wang, "Use of Nakagami distribution and logarithmic compression in ultrasonic tissue characterization," *Journal of Medical and Biological Engineering*, vol. 26, no. 2, pp. 69–73, 2006.

[4] D. T. Yang and J. Y. Lin, "Food availability, entitlement and the Chinese famine of 1959–61," *Economic Journal*, vol. 110, no. 460, pp. 136–158, 2000.

[5] K. Kim and H. A. Latchman, "Statistical traffic modeling of MPEG frame size: experiments and analysis," *Journal of Systemics, Cybernetics and Informatics*, vol. 7, no. 6, pp. 54–59, 2009.

[6] K. Ahmad, S. P. Ahmad, and A. Ahmed, "Some important characterizing properties, information measures and estimations of weibull distribution," *International Journal of Modern Mathematical Sciences*, vol. 12, no. 2, pp. 88–97, 2014.

[7] F. M. Al-Aboud, "Bayesian estimations for the extreme value distribution using progressive censored data and asymmetric loss," *International Mathematical Forum*, vol. 4, no. 33, pp. 1603–1622, 2009.

[8] A. O. M. Ahmed, N. A. Ibrahim, J. Arasan, and M. B. Adam, "Extension of Jeffreys' prior estimate for weibull censored data using Lindley's approximation," *Australian Journal of Basic and Applied Sciences*, vol. 5, no. 12, pp. 884–889, 2011.

[9] B. N. Pandey, N. Dwividi, and B. Pulastya, "Comparison between Bayesian and maximum likelihood estimation of the scale parameter in Weibull distribution with known shape under linex loss function," *Journal of Scientific Research*, vol. 55, pp. 163–172, 2011.

[10] F. M. Al-Athari, "Parameter estimation for the double-pareto distribution," *Journal of Mathematics and Statistics*, vol. 7, no. 4, pp. 289–294, 2011.

[11] S. P. Ahmad and K. Ahmad, "Bayesian analysis of weibull distribution using R software," *Australian Journal of Basic and Applied Sciences*, vol. 7, no. 9, pp. 156–164, 2013.

[12] K. Ahmad, S. P. Ahmad, and A. Ahmed, "On parameter estimation of erlang distribution using bayesian method under different loss functions," in *Proceedings of International Conference on Advances in Computers, Communication, and Electronic Engineering*, pp. 200–206, University of Kashmir, 2015.

[13] K. Ahmad, S. P. Ahmad, and A. Ahmed, "Bayesian analysis of generalized gamma distribution using R software," *Journal of Statistics Applications & Probability*, vol. 4, no. 2, pp. 323–335, 2015.

The Half-Logistic Generalized Weibull Distribution

Masood Anwar [ID] **and Amna Bibi**

Department of Mathematics, COMSATS Institute of Information Technology, Park Road, Chak Shahzad, Islamabad, Pakistan

Correspondence should be addressed to Masood Anwar; masoodanwar@comsats.edu.pk

A new three-parameter generalized distribution, namely, half-logistic generalized Weibull (HLGW) distribution, is proposed. The proposed distribution exhibits increasing, decreasing, bathtub-shaped, unimodal, and decreasing-increasing-decreasing hazard rates. The distribution is a compound distribution of type I half-logistic-G and Dimitrakopoulou distribution. The new model includes half-logistic Weibull distribution, half-logistic exponential distribution, and half-logistic Nadarajah-Haghighi distribution as submodels. Some distributional properties of the new model are investigated which include the density function shapes and the failure rate function, raw moments, moment generating function, order statistics, L-moments, and quantile function. The parameters involved in the model are estimated using the method of maximum likelihood estimation. The asymptotic distribution of the estimators is also investigated via Fisher's information matrix. The likelihood ratio (LR) test is used to compare the HLGW distribution with its submodels. Some applications of the proposed distribution using real data sets are included to examine the usefulness of the distribution.

1. Introduction

Statistical distributions are the basic aspects of all parametric statistical techniques including inference, modeling, survival analysis, and reliability. For the analysis of lifetime data, it is an important task to fit the data by a statistical model. A number of lifetime distributions have been developed in literature for this purpose. The widely used lifetime models usually have a limited range of behaviors. Such type of distributions cannot give a better fit to model all the practical situations. Recently, several authors have developed a number of families of statistical models by applying different techniques. Various techniques have been introduced in the literature to derive new flexible models as discussed by Lai [1].

Marshall and Olkin [2] introduced an effective technique to extend a family of distributions by addition of another parameter. By applying this technique, they generalized the exponential and the Weibull distributions. Al-Zahrani and Sagor [3] proposed the Poisson Lomax model by compounding the Poisson Lomax distributions. Bidram and Nadarajah [4] introduced the exponentiated EG2 distribution by using the method of resilience parameter. Kus [5] considered compounding of Poisson and exponential distribution.

Gurvich et al. [6] generalized the Weibull distribution offering a new distribution function elucidating a wide range of functional forms of the effect of size on the strength distribution, using a simple method of evaluation of the basic statistical parameters. Nadarajah and Kotz [7], Lai et al. [8], Lai et al. [9], and Xie et al. [10] further discussed some modifications of the Weibull model.

In this paper, another extension of the extended Weibull distribution is introduced using half-logistic-G generator. So we propose the half-logistic generalized Weibull (HLGW) distribution without adding any extra parameter to the baseline model. The new model is the compound distribution of two previously known distributions, one of which follows the class proposed by Gurvich et al. [6] and the other is type I half-logistic-G model. The proposed model is able to depict more complex hazard rates and provides a good alternate to the Weibull distribution that does not exhibit upside-down bathtub-shaped or unimodal failure rates.

Dimitrakopoulou et al. [11] established a three-parameter lifetime model with PDF

$$
\begin{aligned}
g\left(x; \omega, \eta, \gamma\right) \\
= \omega \eta \gamma x^{\eta-1}\left(1 + \gamma x^{\eta}\right)^{\omega-1} \exp\left\{1 - \left(1 + \gamma x^{\eta}\right)^{\omega}\right\},
\end{aligned}
\tag{1}
$$

where $x > 0$ and $\omega, \eta > 0$ are the shape parameters and $\gamma > 0$ is a scale parameter. The CDF corresponding to (1) is

$$G(x; \omega, \eta, \gamma) = 1 - \exp\left\{1 - (1 + \gamma x^\eta)^\omega\right\}. \qquad (2)$$

In this paper, a three-parameter lifetime model is presented which is the compound model of the previously known models introduced by Dimitrakopoulou et al. [11] and half-logistic-G (HL-G) distribution called half-logistic generalized Weibull (HLGW) distribution. The half-logistic-G distribution is presented by Cordeiro et al. [12] with the CDF

$$G(x; \gamma, \theta) = \int_0^{-\ln(1-F(x;\theta))} \frac{2\gamma e^{-\gamma x}}{(1 + e^{-\gamma x})^2} dx$$
$$= \frac{1 - [1 - F(x; \theta)]^\gamma}{1 + [1 - F(x; \theta)]^\gamma}, \qquad (3)$$

where $F(x; \theta)$ is the CDF of the baseline distribution and $\gamma > 0$ is the shape parameter. As a special case, for $\gamma = 1$, the TIHL-G is the half-logistic-G (HL-G) model with cumulative distribution function

$$G(x; \theta) = \frac{F(x; \theta)}{1 + \overline{F}(x; \theta)}. \qquad (4)$$

The corresponding PDF to (4) is given by

$$g(x; \theta) = \frac{2f(x; \theta)}{\left[1 + \overline{F}(x; \theta)\right]^2}, \qquad (5)$$

where $f(x) = (d/dx)F(x)$ and $\overline{F}(x; \theta) = 1 - F(x; \theta)$.

The rest of the paper is unfolded as follows. Section 2 contains the introduction of the half-logistic generalized Weibull (HLGW) distribution and provides the plots of its density function. Section 3 explores the distributional properties of the HLGW model. In Section 4, the method of maximum likelihood estimation is used to obtain the estimators of unknown parameters. The asymptotic distribution of the estimators is also investigated in this section via Fisher's information matrix. A simulation study is discussed in Section 5 to check out the accuracy of point and interval estimates of the HLGW parameters. Section 6 involves some applications of the HLGW distribution using lifetime data sets to examine the fitness of the proposed model. Section 7 provides concluding remarks about the paper.

2. The Half-Logistic Generalized Weibull Distribution

Substitution of (1) and (2) in (5) results the following PDF of the HLGW distribution:

$$g(x; \omega, \eta, \gamma)$$
$$= \frac{2\omega\eta\gamma x^{\eta-1}(1 + \gamma x^\eta)^{\omega-1}\exp\left(1 - (1 + \gamma x^\eta)^\omega\right)}{\left[1 + \exp\left(1 - (1 + \gamma x^\eta)^\omega\right)\right]^2}, \qquad (6)$$

for $x > 0$.

The CDF associated with (6) is as follows:

$$G(x) = \frac{1 - \exp\left(1 - (1 + \gamma x^\eta)^\omega\right)}{1 + \exp\left(1 - (1 + \gamma x^\eta)^\omega\right)}, \qquad (7)$$

The parameters $\omega, \eta > 0$ are the shape parameters and $\gamma > 0$ is a scale parameter. From now on, a random variable X with PDF (6) will be written as $X \sim \text{HLGW}(\omega, \eta, \gamma)$.

3. Distributional Properties

This section deals with the investigation of the distributional properties of HLGW distribution. The statistical properties include the plots of the density function, the failure rate function, raw moments, moment generating function, order statistics, L-moments, and quantile function.

3.1. Special Cases. The HLGW distribution includes the following distributions as special cases:

(i) For $\omega = 1$, the HLGW model reduces to the half-logistic Weibull (HLW) model with the PDF

$$g(x; \eta, \gamma) = \frac{2\eta\gamma x^{\eta-1}\exp\left(-\gamma x^\eta\right)}{\left[1 + \exp\left(-\gamma x^\eta\right)\right]^2}, \qquad (8)$$

where η is the shape parameter and γ is the scale parameter. For $-\gamma = \alpha$, the half-logistic Weibull model is also called the power half-logistic distribution (PHLD) proposed and studied by Krishnarani [13].

(ii) For $\omega = \eta = 1$, the HLGW model generates a new model, the half-logistic exponential (HLE) model, with scale parameter γ and the PDF

$$g(x; \gamma) = \frac{2\gamma\exp\left(-\gamma x\right)}{\left[1 + \exp\left(-\gamma x\right)\right]^2}. \qquad (9)$$

(iii) For $\eta = 1$, the HLGW model gives another new model, the half-logistic Nadarajah-Haghighi (HLNH) model, with the PDF

$$g(x; \omega, \gamma) = \frac{2\omega\gamma(1 + \gamma x)^{\omega-1}\exp\left(1 - (1 + \gamma x)^\omega\right)}{\left[1 + \exp\left(1 - (1 + \gamma x)^\omega\right)\right]^2}, \qquad (10)$$

where ω and γ are the shape and scale parameters, respectively.

3.2. The Shapes of HLGW Distribution. The following are the conditions under which the PDF of the HLGW distribution (6) shows different behaviors:

(i) For $\eta < 1$, the PDF is monotone decreasing with $\lim_{x \to 0^+} g(x; \theta) = \infty$, $\lim_{x \to \infty} g(x; \theta) = 0$.

(ii) For $\eta = 1$, the same shape is exhibited with $\lim_{x \to 0^+} g(x; \theta) = \omega\gamma/2$ and $\lim_{x \to \infty} g(x; \theta) = 0$.
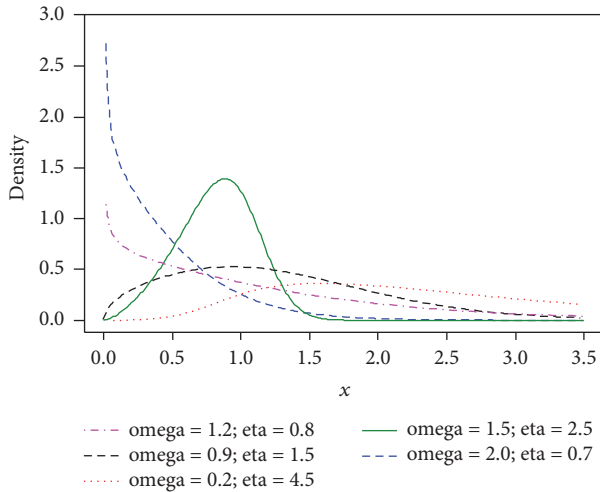
FIGURE 1: The shapes of the PDF of the HLGW distribution.

(iii) For $\eta > 1$, the PDF has the value zero at the origin; then it increases to a higher value and then decreases, falling towards the value of zero at infinity.

Various behaviors of the PDF are shown in Figure 1, for some parameter values; $\gamma = 1$. The mode of density (6) can be obtained from $(d/dx)[\log(g(x))]|_{x=x_0} = 0$.

3.3. Hazard Rate Function (HRF).

For X is a random variable, the HRF is given as $h(x) = g(x)/\overline{G}(x)$, where $\overline{G} = 1 - G$ represents the survival function given by

$$\overline{G}(x) = \frac{2\exp\left(1 - \left(1 + \gamma x^\eta\right)^\omega\right)}{1 + \exp\left(1 - \left(1 + \gamma x^\eta\right)^\omega\right)}. \tag{11}$$

The HRF of $X \sim \text{HLGW}(\omega, \eta, \gamma)$ can be written as

$$h(x; \omega, \eta, \gamma) = \frac{\omega\eta\gamma x^{\eta-1}\left(1 + \gamma x^\eta\right)^{\omega-1}}{\left[1 + \exp\left(1 - \left(1 + \gamma x^\eta\right)^\omega\right)\right]}, \tag{12}$$

for $x > 0$, demonstrating different shapes for different parameter values. By differentiating (12), it can be easily checked that

(a) for $\omega = 1$ and $\eta = 1$, the value of HRF $h(x)$ is zero at the origin; then it increases to its maximum; after that it is constant,

(b) for $\omega \geq 1$ and $\eta > 1$ or $\omega < 1$ and $\eta \leq 1$, $h(x)$ has increasing or decreasing behavior,

(c) for $\omega \geq 1$ and $\eta < 1$,

　　(i) if $\omega\eta \leq 1$, $h(x)$ is decreasing or decreasing-increasing-decreasing (DID),

　　(ii) if $\omega\eta > 1$, $h(x)$ is increasing or bathtub-shaped,

(d) for $\omega \leq 1$ and $\eta > 1$,

　　(i) if $\omega\eta < 1$, $h(x)$ has upside-down bathtub shape (unimodal),

　　(ii) if $\omega\eta \geq 1$, $h(x)$ is monotone increasing.

The different hazard shapes are shown in Figure 2 for different parameter values. By restricting $\omega(n + 1) - 1 \in N$, the failure rate function (12) can also be depicted as

$$
\begin{aligned}
&h(x; \omega, \eta, \gamma) \\
&= \sum_{k=1}^{\infty}\sum_{n=0}^{\infty}\sum_{r=0}^{\omega(n+1)-1} \frac{(-1)^{n+k-1}\,e^{k-1}\,(k-1)^n}{(n+1)!}\gamma_r\eta_r x^{\eta_r-1},
\end{aligned} \tag{13}
$$

for $x > 0$, where $\gamma_r = \binom{\omega(n+1)}{r+1}\gamma^{r+1}$ and $\eta_r = \eta(r+1)$. Thus the HRF can be written as the sum of ω terms and hence (13) is the failure rate function of a series system of ω components.

3.4. Moments.

In this section, the rth moment $\mu_r' = E[X^r]$ of the HLGW model is presented as an infinite sum representation. The first four moments for $r = 1, 2, 3, 4$ have been calculated accordingly.

Theorem 1. Let $X \sim \text{HLGW}(\omega, \eta, \gamma)$ be a random variable, where $\omega, \eta, \gamma > 0$, and the rth moment of X about the origin is as follows:

$$
\begin{aligned}
\mu_r' &= E(X^r) \\
&= \frac{2\omega}{\gamma^{r/\eta}}\sum_{j=0}^{r/\eta}\sum_{k=1}^{\infty}\sum_{n=0}^{\infty}\binom{\frac{r}{\eta}}{j}\frac{e^k k^{n+1}(-1)^{j+k+n+r/\eta}}{\left[j + \omega(n+1)\right]n!},
\end{aligned} \tag{14}
$$

where $r = 1, 2, 3, 4$ and $r/\eta \in \mathbb{N}$.

Corollary 2. Let $X \sim \text{HLGW}(\omega, \eta, \gamma)$ be a random variable, where $\omega, \eta, \gamma > 0$ and $r/\eta \in \mathbb{N}$. The first four moments of the random variable are

$$
\begin{aligned}
\mu' &= E[X] \\
&= \frac{2\omega}{\gamma^{1/\eta}}\sum_{j=0}^{1/\eta}\sum_{k=1}^{\infty}\sum_{n=0}^{\infty}\binom{\frac{1}{\eta}}{j}\frac{e^k k^{n+1}(-1)^{j+k+n+1/\eta}}{\left[j + \omega(n+1)\right]n!} \\[6pt]
\mu_2' &= E[X^2] \\
&= \frac{2\omega}{\gamma^{2/\eta}}\sum_{j=0}^{2/\eta}\sum_{k=1}^{\infty}\sum_{n=0}^{\infty}\binom{\frac{2}{\eta}}{j}\frac{e^k k^{n+1}(-1)^{j+k+n+2/\eta}}{\left[j + \omega(n+1)\right]n!} \\[6pt]
\mu_3' &= E[X^3] \\
&= \frac{2\omega}{\gamma^{3/\eta}}\sum_{j=0}^{3/\eta}\sum_{k=1}^{\infty}\sum_{n=0}^{\infty}\binom{\frac{3}{\eta}}{j}\frac{e^k k^{n+1}(-1)^{j+k+n+3/\eta}}{\left[j + \omega(n+1)\right]n!} \\[6pt]
\mu_4' &= E[X^4] \\
&= \frac{2\omega}{\gamma^{4/\eta}}\sum_{j=0}^{4/\eta}\sum_{k=1}^{\infty}\sum_{n=0}^{\infty}\binom{\frac{4}{\eta}}{j}\frac{e^k k^{n+1}(-1)^{j+k+n+4/\eta}}{\left[j + \omega(n+1)\right]n!}.
\end{aligned} \tag{15}
$$

3.5. Moment Generating Function (MGF).

The MGF of X is retrieved from $M_X(t) = E[e^{tX}] = \int e^{tx}g(x)dx$.
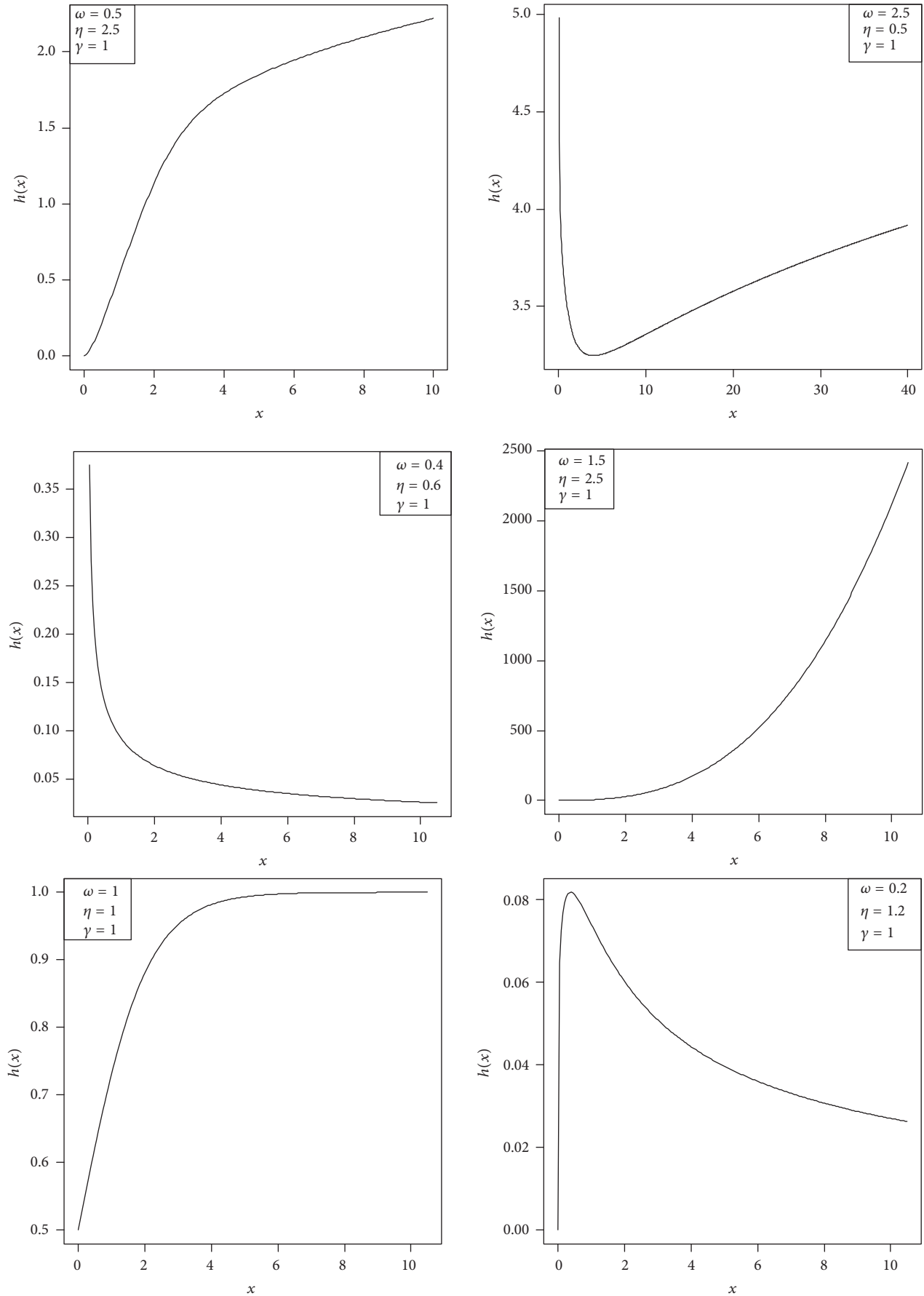
FIGURE 2: Continued.
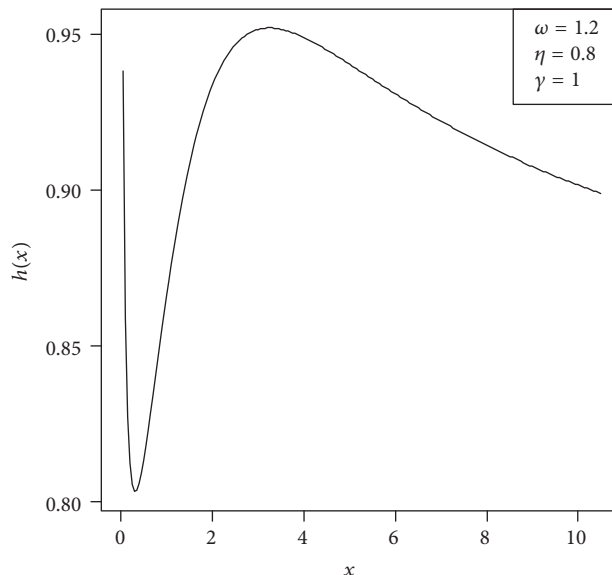
FIGURE 2: The HRF of HLGW distribution for different parameter values.

Theorem 3. *let $X \sim HLGW(\omega, \eta, \gamma)$ be a random variable, and the moment generating function (MGF) of X is given by*

$$M_X(t) = \frac{2\omega}{\gamma^{j/\eta}}$$

$$\cdot \sum_{k=1}^{\infty} \sum_{n=0}^{\infty} \sum_{j=0}^{n} \sum_{i=0}^{j/\eta} \binom{n}{j} \binom{j}{\eta} \binom{\frac{j}{\eta}}{i} \frac{e^k k^{n-j+1} t^j (-1)^{i+j+k+n+j/\eta}}{[i + \omega(n-j+1)] n!}, \quad (16)$$

where $j/\eta \in \mathbb{N}$.

3.6. Order Statistics. Let $X_1, X_2, \ldots, X_m$ be a random sample of size m from a distribution with PDF $g(x)$ and CDF $G(x)$ and $X_{1:m}, X_{2:m}, \ldots, X_{m:m}$ are the analogous order statistics. The PDF and CDF of $X_{r:m}$, $1 \leqslant r \leqslant m$, are

$$g_{r:m}(x) = \frac{1}{B(r, m-r+1)} g(x) [G(x)]^{r-1}$$

$$\cdot [1 - G(x)]^{m-r} = \frac{1}{B(r, m-r+1)} g(x)$$

$$\cdot \sum_{i=0}^{m-r} \binom{m}{j} \binom{m-r}{i} (-1)^i [G(x)]^{i+r-1}, \quad (17)$$

$$G_{r:m} = \sum_{k=r}^{m} \binom{m}{k} [G(x)]^k [1 - G(x)]^{m-k},$$

where $B(r, m - r + 1)$ is the beta function.

Theorem 4. *Let $g(x)$ and $G(x)$ be the PDF and CDF of random variable $X \sim HLGW(\omega, \eta, \gamma)$; then the PDF of $X_{r:m}$ is*

$$g_{r:m}(x) = 2\omega \eta \gamma C_{r:m} \sum_{i=0}^{m-r} \binom{m-r}{i} (-1)^i$$

$$\cdot \frac{x^{\eta-1} (1 + \gamma x^\eta)^{\omega-1} \exp\left[1 - (1 + \gamma x^\eta)^\omega\right]}{\left[1 + \exp\left(1 - (1 + \gamma x^\eta)^\omega\right)\right]^2}$$

$$\times \left[\frac{1 - \exp\left(1 - (1 + \gamma x^\eta)^\omega\right)}{1 + \exp\left(1 - (1 + \gamma x^\eta)^\omega\right)}\right]^{i+r-1}, \quad (18)$$

where $C_{r:m} = B[(r, m - r + 1)]^{-1}$.

The CDF corresponding to (18) is

$$G_{r:m}(x) = \sum_{k=r}^{m} \binom{m}{k} \left[\frac{1 - \exp\left(1 - (1 + \gamma x^\eta)^\omega\right)}{1 + \exp\left(1 - (1 + \gamma x^\eta)^\omega\right)}\right]^k$$

$$\cdot \left[\frac{2 \exp\left(1 - (1 + \gamma x^\eta)^\omega\right)}{1 + \exp\left(1 - (1 + \gamma x^\eta)^\omega\right)}\right]^{m-k}. \quad (19)$$

3.7. L-Moments. Suppose that we have a random sample $X_1, X_2, \ldots, X_n$ collected from $X \sim HLGW(\omega, \eta, \gamma)$. The rth population L-moments are as follows:

$$E[X_{r:n}] = \int_0^\infty x g(X_{r:n}) dx = \int_0^\infty x \sum_{i=0}^{n-r} \binom{n-r}{i} (-1)^i$$

$$\cdot \frac{2\omega \eta \gamma C_{r:n} x^{\eta-1} (1 + \gamma x^\eta)^{\omega-1} e^{1-(1+\gamma x^\eta)^\omega}}{\left(1 + e^{1-(1+\gamma x^\eta)^\omega}\right)^2} \quad (20)$$

$$\times \left(\frac{1 - e^{1-(1+\gamma x^\eta)^\omega}}{1 + e^{1-(1+\gamma x^\eta)^\omega}}\right)^{i+r-1} dx.$$

We use the substitution $y = 1 + \gamma x^\eta$, so $x = ((y-1)/\gamma)^{1/\eta}$ and $dx = (1/\gamma\eta)((y-1)/\gamma)^{1/\eta-1} dy$, where $1/\eta \in \mathbb{N}$. Therefore, we get

$$E\left[X_{r:n}\right] = 2\omega C_{r:n}\sum_{i=0}^{n-r}\sum_{k=0}^{i+r-1}\sum_{j=0}^{\infty}\sum_{l=0}^{1/\eta}\sum_{m=0}^{\infty}\binom{n-r}{i}\binom{i+r-1}{k}\binom{-(i+r+1)}{j}\binom{\frac{1}{\eta}}{l}\times\left(\frac{-1}{\gamma}\right)^{1/\eta}$$

$$\cdot\frac{(-1)^{j+k+l+m}\,e^{j+k+1}\,(j+k+1)^m}{m!}\int_1^{\infty}y^{l+m\omega+\omega-1}dy.\tag{21}$$

By working out the integration, we arrive at the following formula:

$$E\left[X_{r:n}\right]$$

$$= 2\omega\sum_{i=0}^{n-r}\sum_{k=0}^{i+r-1}\sum_{j=0}^{\infty}\sum_{l=0}^{1/\eta}\sum_{m=0}^{\infty}\left(\frac{-1}{\gamma}\right)^{1/\eta}\frac{(j+k+1)^m\,e^{j+k+1}}{l+\omega(m+1)\,m!}A_{ijkl},\tag{22}$$

where

$$A_{ijkl} = C_{r:n}\,(-1)^{j+k+l+m+1}$$

$$\cdot\binom{n-r}{i}\binom{i+r-1}{k}\binom{-(i+r+1)}{j}\binom{\frac{1}{\eta}}{l}.\tag{23}$$

The relation (22) can be used to find out the first L-moments of $X_{r:n}$; that is, for $r = n = 1$ we get $\lambda_1 = E[X_{1:1}]$.

$$\lambda_1 = 2\omega\sum_{j=0}^{\infty}\sum_{m=0}^{\infty}\sum_{l=0}^{1/\eta}\binom{-2}{j}\binom{\frac{1}{\eta}}{l}\left(\frac{-1}{\gamma}\right)^{1/\eta}$$

$$\cdot\frac{(-1)^{j+l+m+1}\,e^{j+1}\,(j+1)^m}{(l+\omega(m+1))\,m!}.\tag{24}$$

The other two moments, λ_2 and λ_3, are, respectively, given by

$$\lambda_2 = 4\omega\sum_{j=0}^{\infty}\sum_{l=0}^{1/\eta}\sum_{m=0}^{\infty}\binom{\frac{1}{\eta}}{l}\frac{(-1/\gamma)^{1/\eta}}{(l+\omega(m+1))\,m!}$$

$$\times\left[\sum_{i=0}^{1}\sum_{k=0}^{i}\binom{1}{i}\binom{i}{k}\binom{-2-i}{j}(-1)^{j+k+l+m+1}\right.$$

$$\cdot e^{j+k+l}\,(j+k+l)^m + \sum_{k=0}^{1}\binom{1}{k}\binom{-3}{j}(-1)^{j+k+l+m+1}$$

$$\left.\cdot e^{j+k+1}\,(j+k+1)^m\right].$$

$$\lambda_3 = 2\omega\sum_{j=0}^{\infty}\sum_{l=0}^{1/\eta}\sum_{m=0}^{\infty}\binom{\frac{1}{\eta}}{l}$$

$$\cdot\frac{(-1/\gamma)^{1/\eta}\,(-1)^{j+k+l+m+1}\,e^{j+k+1}\,(j+k+1)^m}{(l+\omega(m+1))\,m!}$$

$$\times\left[\sum_{i=0}^{2}\sum_{k=0}^{i}\binom{2}{i}\binom{i}{k}\binom{-2-i}{j}\right.$$

$$+ 3!\sum_{i=0}^{1}\sum_{k=0}^{i+1}\binom{1}{i}\binom{i+1}{k}\binom{-3-i}{j}$$

$$\left.+ 3\sum_{k=0}^{2}\binom{2}{k}\binom{-4-i}{j}\right]\tag{25}$$

3.8. Quantile Function.
For X to be a random variable with the PDF (6), the quantile function $Q(u)$ is

$$Q\left(u\right) = \inf\left\{x \in R : G\left(x\right) \geqslant u\right\},\quad\text{where }0 < u < 1.\tag{26}$$

The above relation is used to find the quantile function of HLGW distribution. Therefore, we have

$$Q\left(u\right) = \left[\frac{1}{\gamma}\left\{\left(1 - \ln\left(\frac{1-u}{1+u}\right)\right)^{1/\omega} - 1\right\}\right]^{1/\eta}.\tag{27}$$

Hence, the generator for X can be given by the following algorithm:

(1) Generate $U\sim$uniform$(0, 1)$.

(2) Use (27) and obtain an outcome of X by $X = Q(U)$.

By using the quantile function (27), we can examine the Bowley skewness [14] and Moors kurtosis [15] for HLGW as follows:

$$sk = \frac{Q\left(3/4\right) + Q\left(1/4\right) - 2Q\left(2/4\right)}{Q\left(3/4\right) - Q\left(1/4\right)}$$

$$\kappa = \frac{Q\left(3/8\right) - Q\left(1/8\right) + Q\left(7/8\right) - Q\left(5/8\right)}{Q\left(3/4\right) - Q\left(1/4\right)}.\tag{28}$$

Table 1 illustrates the values of skewness and kurtosis for the HLGW model for some values of ω, η, and γ. It can be noted that the skewness and kurtosis are free of parameter γ and they are decreasing functions of the parameters ω and η.

TABLE 1: Skewness and kurtosis of HLGW for different values of ω, η, and γ.

γ	ω	$\eta = 0.5$		$\eta = 1.0$		$\eta = 2.0$	
		sk	ku	sk	ku	sk	ku
	0.5	0.6530	2.4642	0.3366	1.4540	0.1303	1.2492
	1.0	0.4634	1.6286	0.1808	1.2395	0.0197	1.2074
0.5	1.5	0.3913	1.4463	0.1264	1.1952	−0.0196	1.2045
	2.0	0.3541	1.3700	0.0989	1.1779	−0.0396	1.2054
	2.5	0.3314	1.3286	0.0824	1.1691	−0.0517	1.2067
	0.5	0.6530	2.4642	0.3366	1.4540	0.1303	1.2492
	1.0	0.4634	1.6286	0.1808	1.2395	0.0197	1.2074
1.5	1.5	0.3913	1.4463	0.1264	1.1952	−0.0196	1.2045
	2.0	0.3541	1.3700	0.0989	1.1779	−0.0396	1.2054
	2.5	0.3314	1.3286	0.0824	1.1691	−0.0517	1.2067
	0.5	0.6530	2.4642	0.3366	1.4540	0.1303	1.2492
	1.0	0.4634	1.6286	0.1808	1.2395	0.0197	1.2074
2.0	1.5	0.3913	1.446	0.1264	1.1952	−0.0196	1.2045
	2.0	0.3541	1.3700	0.0989	1.1779	−0.0396	1.2054
	2.5	0.3314	1.3286	0.0824	1.1691	−0.0517	1.2067
	0.5	0.6530	2.4642	0.3366	1.4540	0.1303	1.2492
	1.0	0.4634	1.6286	0.1808	1.2395	0.0197	1.2074
3.0	1.5	0.3913	1.4463	0.1264	1.1956	−0.0196	1.2045
	2.0	0.3541	1.3700	0.0989	1.1779	−0.0396	1.2054
	2.5	0.3314	1.3286	0.0824	1.1691	−0.0517	1.2067

4. Estimation

The log-likelihood function is expressed as

$$
\begin{aligned}
\ell\left(x ; \omega, \eta, \gamma\right) \\
= n \ln (2) + n \ln (\omega) + n \ln (\eta) + n \ln (\gamma) \\
+ (\eta - 1) \sum_{i=1}^{n} \ln x_i + (\omega - 1) \sum_{i=1}^{n} \ln \left(1 + \gamma x_i^{\eta}\right) \\
+ \sum_{i=1}^{n} \left(1 - \left(1 + \gamma x_i^{\eta}\right)^{\omega}\right) \\
- 2 \sum_{i=1}^{n} \ln \left[1 + \exp \left(1 - \left(1 + \gamma x_i^{\eta}\right)^{\omega}\right)\right].
\end{aligned}
\tag{29}
$$

Taking the first partial derivatives of $\ell(x ; \omega, \eta, \gamma)$ with respect to ω, η, and γ and letting them equal zero, we obtain a nonlinear system of equations.

$$
\begin{aligned}
\frac{\partial \ell}{\partial \omega} = \frac{n}{\omega} + \sum_{i=1}^{n} \ln \left(1 + \gamma x_i^{\eta}\right) \\
- \sum_{i=1}^{n} \left(1 + \gamma x_i^{\eta}\right)^{\omega} \ln \left(1 + \gamma x_i^{\eta}\right) \\
+ 2 \sum_{i=1}^{n} \frac{\left(1 + \gamma x_i^{\eta}\right)^{\omega} \ln \left(1 + \gamma x_i^{\eta}\right) e^{1-\left(1+\gamma x_i^{\eta}\right)^{\omega}}}{1 + e^{1-\left(1+\gamma x_i^{\eta}\right)^{\omega}}} = 0.
\end{aligned}
$$

$$
\frac{\partial \ell}{\partial \eta} = \frac{n}{\eta} + \sum_{i=1}^{n} \ln x_i + (\omega - 1) \sum_{i=1}^{n} \frac{\gamma x_i^{\eta} \ln x_i}{1 + \gamma x_i^{\eta}}
$$

$$
- \omega \gamma \sum_{i=1}^{n} x_i^{\eta} \ln x_i \left(1 + \gamma x_i^{\eta}\right)^{\omega-1}
$$

$$
+ 2\omega \sum_{i=1}^{n} \frac{\gamma x_i^{\eta} \ln x_i \left(1 + \gamma x_i^{\eta}\right)^{\omega-1} e^{1-\left(1+\gamma x_i^{\eta}\right)^{\omega}}}{1 + e^{1-\left(1+\gamma x_i^{\eta}\right)^{\omega}}} = 0
$$

$$
\frac{\partial \ell}{\partial \gamma} = \frac{n}{\gamma} + (\omega - 1) \sum_{i=1}^{n} \frac{x_i^{\eta}}{1 + \gamma x_i^{\eta}} - \omega \sum_{i=1}^{n} \left(1 + \gamma x_i^{\eta}\right)^{\omega-1} x_i^{\eta}
$$

$$
+ 2\omega \sum_{i=1}^{n} \frac{x_i^{\eta} \left(1 + \gamma x_i^{\eta}\right)^{\omega-1} e^{1-\left(1+\gamma x_i^{\eta}\right)^{\omega}}}{1 + e^{1-\left(1+\gamma x_i^{\eta}\right)^{\omega}}} = 0
\tag{30}
$$

The above equations cannot be solved analytically, and statistical software can be used to solve them numerically via iterative methods and get the maximum likelihood estimate (MLE) of ω, η, and γ.

4.1. Asymptotic Distribution. In order to have approximate confidence intervals (CIs) of the involved parameters, we require the estimated values of the elements of variance-covariance matrix V of the MLEs. The variance-covariance matrix V is estimated by the observed information matrix $\widehat{V}$, where

$$
\widehat{V} = - \begin{bmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{bmatrix},
\tag{31}
$$

Table 2: Bias and MSE for the HLGW parameters.

ω	η	γ	n	Bias (ω)	MSE (ω)	Bias (η)	MSE (η)	Bias (γ)	MSE (γ)
			20	0.2103	0.0392	1.1935	3.2415	0.3454	0.9104
			40	0.1732	0.0314	1.1431	2.9381	0.2931	0.5830
			60	0.1526	0.0283	1.1063	2.4897	0.2480	0.3148
1.5	5	0.5	80	0.1473	0.0231	0.9733	2.1738	0.1812	0.1908
			100	0.1332	0.0207	0.8910	1.9318	0.1317	0.0813
			250	0.1171	0.0113	0.5337	1.3877	0.0811	0.0031
			500	0.1010	0.0061	0.2701	1.0814	0.0213	0.0008
			20	0.3118	0.0433	2.6754	15.9312	0.6125	10.2311
			40	0.2918	0.0395	2.3487	12.3471	0.2451	7.6401
			60	0.2554	0.0365	2.1174	9.2877	0.1615	4.3881
1.5	10	1	80	0.2375	0.0335	2.0968	6.2514	0.0941	1.9722
			100	0.2114	0.0317	2.0532	4.9934	0.0532	1.5531
			250	0.1783	0.0307	0.0065	1.9951	0.0113	0.4899
			500	0.1265	0.0299	0.0013	0.2164	0.0095	0.0989

where I_{ij}, $i, j = 1, 2, 3$, are the second partial derivatives of (29) with respect to ω, η, and γ. They are the entries of Fisher's information matrix analogous to ω, η, and γ, respectively, which are given in Appendix. The diagonal of matrix in (31) gives the variances of the MLEs of ω, η, and γ, respectively.

Approximation by a standard normal (SN) distribution of the distribution of $Z_{\hat{\theta}_k} = (\hat{\theta}_k - \theta_k)/\sqrt{\widehat{\text{Var}}(\hat{\theta}_k)}$, where $\hat{\theta} = (\hat{\omega}, \hat{\eta}, \hat{\gamma})$, results in an approximate $100(1 - \vartheta)\%$ confidence interval for θ_k as

$$\hat{\theta}_k \pm z_{\vartheta/2} \sqrt{\widehat{\text{Var}}\left(\hat{\theta}_k\right)}, \quad j = 1, 2, 3, \qquad (32)$$

where $z_{\vartheta/2}$ is the upper $(\vartheta/2)100$th percentile of SN distribution. We can use the likelihood ratio (LR) test to compare the fit of the HLGW distribution with its submodels for a given data set. For example, to test $\gamma = 0$, the LR statistic is $\mathbf{w} = 2[\ln(L(\hat{\omega}, \hat{\eta}, \hat{\gamma})) - \ln(L(\tilde{\omega}, \tilde{\eta}, 0))]$, where $\hat{\omega}, \hat{\eta}$, and $\hat{\gamma}$ are the unrestricted estimates and $\tilde{\omega}, \tilde{\eta}$ are the restricted estimates. The LR test rejects the null hypothesis if $\mathbf{w} > \chi_\epsilon^2$, where χ_ϵ^2 denotes the upper $100 \in \%$ point of the χ^2-distribution with 1 degree of freedom.

5. Simulation Study

The MLEs can be checked out by a simulation study. The following steps can be followed:

(1) By using (6), 5,000 samples of size n are achieved. The variates of the HLGW distribution are developed using

$$X = \left[\frac{1}{\gamma}\left\{\left(1 - \ln\left(\frac{1-u}{1+u}\right)\right)^{1/\omega} - 1\right\}\right]^{1/\eta}, \qquad (33)$$

for $u \sim U(0, 1)$.

(2) The MLEs are computed for the samples, say $\widehat{\Theta}_j = (\hat{\omega}_j, \hat{\eta}_j, \hat{\gamma}_j)$ for $j = 1, 2, \ldots, 5,000$.

(3) The mean square errors (MSEs) are calculated for every parameter.

The above steps were repeated for $n = 20, 40, 60, 80, 100, 250$, and 500 with $\omega = 1.5$, $\eta = 5$, $\gamma = 0.5$ and $\omega = 1.5, \eta = 10, \gamma = 1$. Table 2 shows the bias and MSEs of ω, η, and γ. It can be deduced through the table that MSEs for individual parameters fall to zero when sample size increases.

6. Applications

This section deals with the applications of the HLGW model to two lifetime data sets, that is, the data of 213 observed values of intervals between failures of air conditioning system of Boeing 720 jet airplanes and the data of waiting times (min) of 100 bank customers. Estimates of the parameters of HLGW distribution (standard errors in parentheses) and Cramer-von Mises W^*, Anderson Darling A^*, and K-S statistics are presented for the data sets. In general, the smaller the values of W^*, A^*, and K-S statistics, the better the fit. We compare the proposed model with other models for the same data sets to check the potentiality and flexibility of new model.

6.1. Air Conditioning Systems Failure Data. The first application concerns 213 observed values of intervals between failures of air conditioning system of Boeing 720 jet airplanes firstly analysed by Proschan [16]. We have compared the performance of the HLGW distribution with its submodels as well as with some other well-known models given below.

The Weibull Poisson (WP) distribution by Lu and Shi [17] with the PDF:

$$g(t; \alpha, \beta, \lambda) = \frac{\alpha\beta\lambda t^{\alpha-1}}{1 - e^{-\lambda}} e^{-\lambda - \beta t^\alpha + \lambda \exp(-\beta t^\alpha)}, \quad t > 0, \qquad (34)$$

for α, β, and $\lambda > 0$.

TABLE 3: Estimates of models for the air conditioning systems failure data.

Distributions	Estimates			$-2\ell(\widehat{\theta})$	Statistics			
	$\widehat{\omega}$	$\widehat{\eta}$	$\widehat{\gamma}$		W^*	A^*	K-S	p value
HLGW	0.3470	1.3814	0.0296	2349.674	0.0339	0.2587	0.0394	0.8957
	(0.0885)	(0.2462)	(0.0127)					
HLW		0.7804	0.0442	2360.630	0.2099	1.2851	0.0577	0.4770
		(0.0426)	(0.0098)					
HLE			0.0144	2383.325	0.3446	2.0606	0.1407	0.0004
			(0.0009)					
HLNH	0.5596		0.0482	2352.930	0.0850	0.5534	0.0445	0.7926
	(0.0515)		(0.0108)					
Weibull	0.0158	0.9226		2355.171	0.1373	0.8552	0.0514	0.6270
	(0.0038)	(0.0459)						
CWG	0.2890	0.0079	4.1050	2362.329	0.1519	1.0253	0.0502	0.6563
	(0.0238)	(0.0032)	(2.8847)					
WP	0.4015	0.4967	−8.0978	2351.850	0.0440	0.3305	0.0412	0.8624
	(0.0584)	(0.1838)	(2.5766)					

The complementary Weibull geometric (CWG) distribution by Tojeiro et al. [18] with the PDF:

$$g(t) = \frac{\gamma \lambda^\gamma \theta t^{\gamma-1} e^{-(\lambda t)^\gamma}}{[\theta + (1-\theta) e^{-(\lambda t)^\gamma}]^2}, \qquad (35)$$

for t, λ, γ, and $\theta > 0$.

The Weibull (W) distribution with PDF:

$$g(y; \beta, \gamma) = \beta \gamma y^{\gamma-1} \exp\left(-\beta y^\gamma\right), \qquad (36)$$

for $y > 0, \beta > 0$, and $\gamma > 0$.

The estimated values of the parameters with standard errors (SE) are found using the method of maximum likelihood estimation. Table 3 gives the estimated values of the parameters along with their standard errors and the test statistics have the smallest values of W^*, A^*, and K-S statistic for the data set under HLGW distribution as compared to the other models. Based on these values, it is concluded that HLGW distribution is the best model as compared to the other models to fit this data set. This conclusion can also be made by the CDF plots for empirical and fitted HLGW distributions using data of 213 values of intervals between failures of air conditioning system in Figure 3(a) for the data. In Figure 3(b), the TTT plot is shown for the data set. The failure rate shape for this data set is decreasing, as its TTT plot is convex.

The LR test statistics of hypotheses H_0: HLW versus H_a: HLGW, H_0: HLE versus H_a: HLGW, and H_0: HLNH versus H_a: HLGW are 10.956 (p value = 0.00093), 33.651 (p value = 0.00001), and 3.256 (p value = 0.071163), respectively. We conclude that there is a significant difference between HLW and HLGW distributions, HLE and HLGW, and also between HLNH and HLGW distributions at the 10% level.

6.2. Waiting Time Data. The data encountered in the second application involves data of waiting times (min) of 100 bank customers used by Bidram and Nadarajah [4]. We

have compared the performance of the HLGW distribution with its submodels as well as with some other well-known distributions, such as the Weibull and the complementary Weibull geometric distribution. We also use the LR test to compare the HLGW distribution and its submodels.

The estimated values of the parameters, the standard errors, and the goodness-of-fit test statistics W^*, A^*, and K-S statistic for the data of waiting time are given in Table 4. These tables illustrate that HLGW model shows a good fit for this data set as compared to the other distributions. The CDF plot for empirical and HLGW distributions using the data set of waiting time in Figure 4(a) also confirms the fitness of HLGW distribution. Figure 4(b) expresses the TTT plot for this data set. Since the TTT plot of the data set is concave, the data set has increasing hazard rate shape. The LR test statistics of the hypotheses H_0: HLW versus H_a: HLGW, H_0: HLE versus H_a: HLGW, and H_0: HLNH versus H_a: HLGW are 5.501 (p value = 0.019006), 11.85 (p value = 0.000577), and 9.271 (p value = 0.002328). The HLGW distribution is significantly better than HLW, HLE, and HLNH distributions. There is no difference between HLGW and Weibull distribution based on the LR test; however the goodness-of-fit statistics W^*, A^*, and K-S statistic clearly show that HLGW distribution is better than Weibull distribution for the data.

7. Conclusion

We have proposed a three-parameter lifetime generalized distribution, referred to as the half-logistic generalized Weibull (HLGW) distribution. The HLGW distribution has three other distributions like the half-logistic exponential, the half-logistic Weibull, and the half-logistic Nadarajah-Haghighi distributions, as its submodels. The new model exhibits a variety of shapes of the failure rate function, that is, increasing, increasing and then constant, decreasing, bathtub, unimodal, and decreasing-increasing-decreasing (DID) shapes. Various statistical properties of the HLGW distribution are derived and studied in detail. We estimated the parameters involved
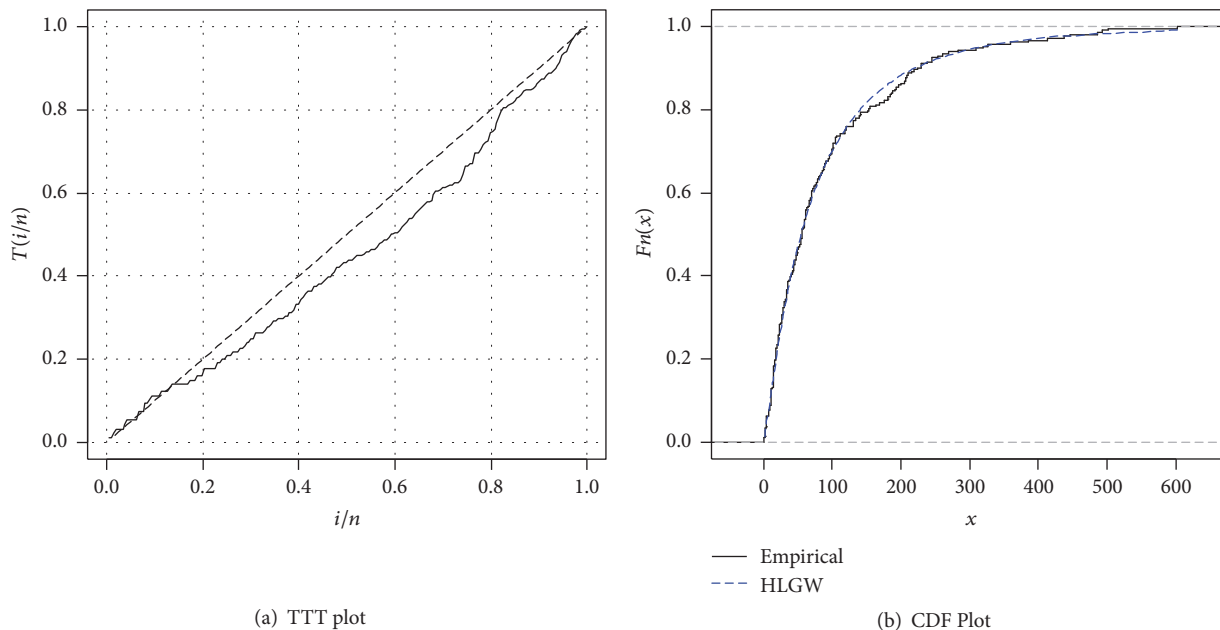
(a) TTT plot

(b) CDF Plot

FIGURE 3: TTT plot and CDF plot for the air conditioning systems failure data.

TABLE 4: Estimates of models for the waiting time data.

Distributions	Estimates			$-2\ell(\widehat{\theta})$	Statistics			
	$\widehat{\omega}$	$\widehat{\eta}$	$\widehat{\gamma}$		W^*	A^*	K-S	p value
HLGW	0.3610	2.1425	0.0773	634.226	0.0176	0.1277	0.0368	0.9992
	(0.1268)	(0.5257)	(0.0340)					
HLW		1.2336	0.0772	639.730	0.0934	0.5740	0.0627	0.8267
		(0.0978)	(0.0216)					
HLE			0.1446	646.076	0.0583	0.3624	0.1213	0.1053
			(0.0119)					
HLNH	1.4044		0.0846	643.497	0.1032	0.6355	0.0877	0.4257
	(0.3411)		(0.0316)					
Weibull	0.0305	1.4582		637.461	0.0629	0.3961	0.0577	0.8938
	(0.0095)	(0.1089)						
CWG	0.5066	0.0126	2.2254	645.742	0.0745	0.4653	0.0485	0.9726
	(0.0779)	(0.0097)	(1.7142)					

in the model by using the method of MLEs. Two real data sets are used to illustrate the flexibility, potentiality, and usefulness of HLGW distribution. It is concluded that HLGW model delivers better fitting than the other lifetime models and we hope that HLGW distribution may attract wider range of practical applications and this research may serve as a reference and benefit future research in the subject field of study.

Appendix

Fisher's Information Matrix

The elements of Fisher's information matrix analogous to ω, η, and γ:

$$\frac{\partial^2 \ell}{\partial \omega^2} = -\frac{n}{\omega^2} - \sum_{i=1}^{n} A_i^\omega \left(\ln A_i\right)^2$$

$$+ 2\sum_{i=1}^{n} \frac{\left(\ln A_i\right)^2 A_i^\omega e^{1-A_i^\omega}}{\left(1 + e^{1-A_i^\omega}\right)^2} \left(1 - A_i^\omega + e^{1-A_i^\omega}\right)$$

$$\frac{\partial^2 \ell}{\partial \omega \partial \eta} = \frac{\partial^2}{\partial \eta \partial \omega} = \sum_{i=1}^{n} \frac{\gamma x_i^\eta \ln x_i}{A_i} \left[1 - A_i^\omega \left(1 + \omega \ln A_i\right)\right.$$

$$+ \frac{2 A_i^\omega e^{1-A_i^\omega}}{1 + e^{1-A_i^\omega}}$$

$$\left. + \frac{2\omega A_i^\omega \ln A_i e^{1-A_i^\omega}}{\left(1 + e^{1-A_i^\omega}\right)^2} \left(1 - A_i^\omega + e^{1-A_i^\omega}\right)\right]$$

(a) TTT plot

—— Empirical
--- HLGW
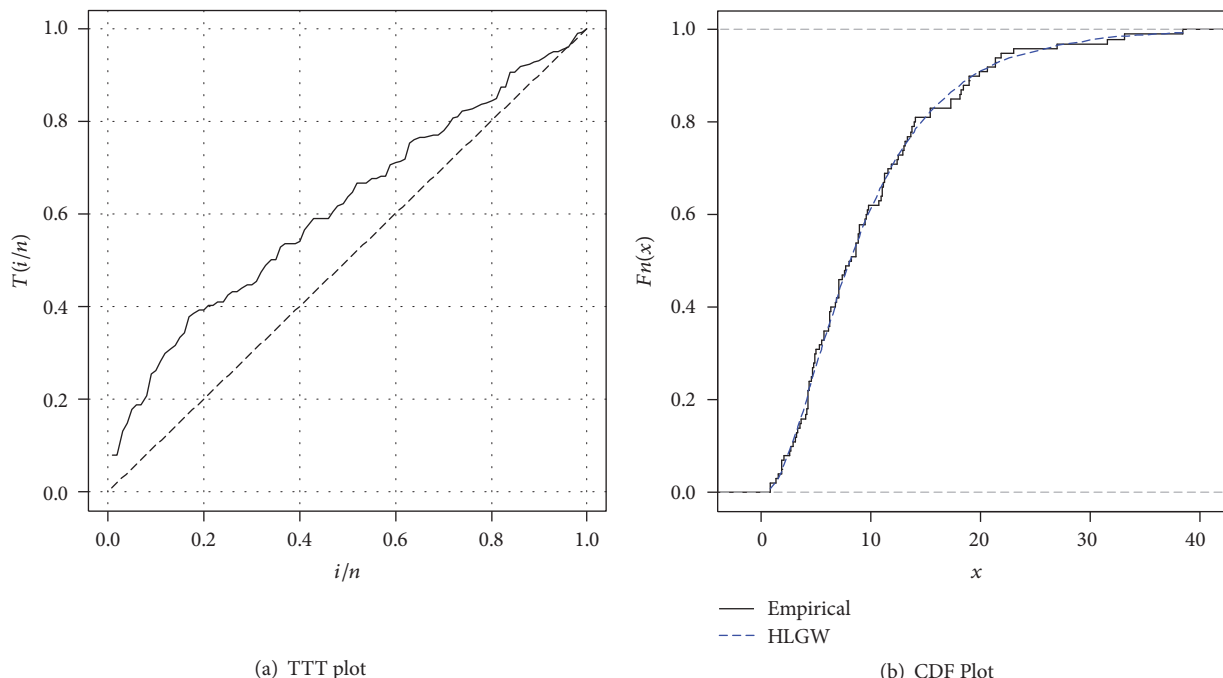
(b) CDF Plot

FIGURE 4: TTT plot and CDF plot for the waiting time data.

$$\frac{\partial^2 \ell}{\partial \omega \partial \gamma} = \frac{\partial^2 \ell}{\partial \gamma \partial \omega} = \sum_{i=1}^{n} \frac{x_i^{\eta}}{A_i} \left[1 - A_i^{\omega} \left(1 + \omega \ln A_i \right) \right.$$

$$+ \frac{2A_i^{\omega} e^{1-A_i^{\omega}}}{1 + e^{1-A_i^{\omega}}}$$

$$\left. + \frac{2\omega A_i^{\omega} \ln A_i e^{1-A_i^{\omega}}}{\left(1 + e^{1-A_i^{\omega}}\right)^2} \left(1 - A_i^{\omega} + e^{1-A_i^{\omega}} \right) \right]$$

$$\frac{\partial^2 \ell}{\partial \eta^2} = -\frac{n}{\eta^2} + \sum_{i=1}^{n} \frac{\gamma x_i^{\eta} \left(\ln x_i \right)^2}{A_i} \left[-\omega A_i^{\omega} + (\omega - 1) \right.$$

$$\cdot \left(1 - \frac{\gamma x_i^{\eta}}{A_i} - \frac{\omega A_i^{\omega} \gamma x_i^{\eta}}{A_i} + \frac{2\omega A_i^{\omega} \gamma x_i^{\eta} e^{1-A_i^{\omega}}}{A_i \left(1 + e^{1-A_i^{\omega}} \right)} \right)$$

$$\left. + \frac{2\omega A_i^{\omega} e^{1-A_i^{\omega}}}{1 + e^{1-A_i^{\omega}}} \left(1 - \frac{\omega A_i^{\omega} \gamma x_i^{\eta}}{A_i \left(1 + e^{1-A_i^{\omega}} \right)} \right) \right]$$

$$\frac{\partial^2 \ell}{\partial \eta \partial \gamma} = \frac{\partial^2 \ell}{\partial \gamma \partial \eta} = \sum_{i=1}^{n} \frac{x_i^{\eta} \ln x_i}{A_i} \left[-\omega A_i^{\omega} + (\omega - 1) \right.$$

$$\cdot \left(1 - \frac{\gamma x_i^{\eta}}{A_i} - \frac{\omega A_i^{\omega} \gamma x_i^{\eta}}{A_i} + \frac{2\omega A_i^{\omega} \gamma x_i^{\eta} e^{1-A_i^{\omega}}}{A_i \left(1 + e^{1-A_i^{\omega}} \right)} \right)$$

$$\left. + \frac{2\omega A_i^{\omega} e^{1-A_i^{\omega}}}{1 + e^{1-A_i^{\omega}}} \left(1 - \frac{\omega A_i^{\omega} \gamma x_i^{\eta}}{A_i \left(1 + e^{1-A_i^{\omega}} \right)} \right) \right]$$

$$\frac{\partial^2 \ell}{\partial \gamma^2} = -\frac{n}{\gamma^2} - \sum_{i=1}^{n} \frac{x_i^{2\eta}}{A_i^2} \left[(\omega - 1) \right.$$

$$\cdot \left(1 + \omega A_i^{\omega} - \frac{2\omega A_i^{\omega} e^{1-A_i^{\omega}}}{1 + e^{1-A_i^{\omega}}} \right) + \frac{2\omega^2 A_i^{2\omega} e^{1-A_i^{\omega}}}{\left(1 + e^{1-A_i^{\omega}} \right)^2} \right].$$

$$\text{(A.1)}$$

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] C. D. Lai, "Constructions and applications of lifetime distributions," *Applied Stochastic Models in Business and Industry*, vol. 29, no. 2, pp. 127–140, 2013.

[2] A. W. Marshall and I. Olkin, "A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families," *Biometrika*, vol. 84, no. 3, pp. 641–652, 1997.

[3] B. Al-Zahrani and H. Sagor, "The Poisson-Lomax distribution," *Revista Colombiana de Estadistica*, vol. 37, no. 1, pp. 225–245, 2014.

[4] H. Bidram and S. Nadarajah, "A new lifetime model with decreasing, increasing, bathtub-shaped, and upside-down bathtub-shaped hazard rate function," *Statistics: A Journal of Theoretical and Applied Statistics*, 2015.

[5] C. Kus, "A new lifetime distribution," *Computational Statistics & Data Analysis*, vol. 51, no. 9, pp. 4497–4509, 2007.

[6] M. R. Gurvich, A. T. DiBenedetto, and S. V. Ranade, "A new statistical distribution for characterizing the random strength of brittle materials," *Journal of Materials Science*, vol. 32, no. 10, pp. 2559–2564, 1997.

[7] S. Nadarajah and S. Kotz, "On some recent modifications of Weibull distribution," *IEEE Transactions on Reliability*, vol. 54, no. 4, pp. 561-562, 2005.

[8] C. D. Lai, M. Xie, and D. N. P. Murthy, "Reply to "On some recent modifications of Weibull distribution"," *IEEE Transactions on Reliability*, vol. 54, no. 4, p. 563, 2005.

[9] C. D. Lai, M. Xie, and D. N. P. Murthy, "A modified Weibull distribution," *IEEE Transactions on Reliability*, vol. 52, no. 1, pp. 33–37, 2003.

[10] M. Xie, Y. Tang, and T. N. Goh, "A modified Weibull extension with bathtub-shaped failure rate function," *Reliability Engineering & System Safety*, vol. 76, no. 3, pp. 279–285, 2002.

[11] T. Dimitrakopoulou, K. Adamidis, and S. Loukas, "A lifetime distribution with an upside-down bathtub-shaped hazard function," *IEEE Transactions on Reliability*, vol. 56, no. 2, pp. 308–311, 2007.

[12] G. M. Cordeiro, M. Alizadeh, and P. . Marinho, "The type I half-logistic family of distributions," *Journal of Statistical Computation and Simulation*, vol. 86, no. 4, pp. 707–728, 2016.

[13] S. D. Krishnarani, "On a power transformation of half-logistic distribution," *Journal of Probability and Statistics*, vol. 2016, Article ID 2084236, 2016.

[14] J. Kenney and E. S. Keeping, *Mathematics of Statistics*, Van Nostrand, Princeton, NJ, USA, 1962, Van Nostrand., Mathematics of Statistics.

[15] J. J. Moors, "A Quantile Alternative for Kurtosis," *The American Statistician*, vol. 37, no. 1, p. 25, 1988.

[16] F. Proschan, "Theoretical Explanation of Observed Decreasing Failure Rate," *Technometrics*, vol. 5, no. 3, pp. 375–383, 1963.

[17] W. Lu and D. Shi, "A new compounding life distribution: the Weibull-Poisson distribution," *Journal of Applied Statistics*, vol. 39, no. 1, pp. 21–38, 2012.

[18] C. Tojeiro, F. Louzada, M. Roman, and P. Borges, "The complementary Weibull geometric distribution," *Journal of Statistical Computation and Simulation*, vol. 84, no. 6, pp. 1345–1362, 2014.

A Novel Entropy-Based Decoding Algorithm for a Generalized High-Order Discrete Hidden Markov Model

Jason Chin-Tiong Chan ⓘ[1] **and Hong Choon Ong** ⓘ[2]

[1]*Ted Rogers School of Management, Ryerson University, 350 Victoria St., Toronto, ON, Canada M5B 2K3*
[2]*School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Gelugor, Penang, Malaysia*

Correspondence should be addressed to Jason Chin-Tiong Chan; chintiongjason.chan@ryerson.ca

The optimal state sequence of a generalized High-Order Hidden Markov Model (HHMM) is tracked from a given observational sequence using the classical Viterbi algorithm. This classical algorithm is based on maximum likelihood criterion. We introduce an entropy-based Viterbi algorithm for tracking the optimal state sequence of a HHMM. The entropy of a state sequence is a useful quantity, providing a measure of the uncertainty of a HHMM. There will be no uncertainty if there is only one possible optimal state sequence for HHMM. This entropy-based decoding algorithm can be formulated in an extended or a reduction approach. We extend the entropy-based algorithm for computing the optimal state sequence that was developed from a first-order to a generalized HHMM with a single observational sequence. This extended algorithm performs the computation exponentially with respect to the order of HMM. The computational complexity of this extended algorithm is due to the growth of the model parameters. We introduce an efficient entropy-based decoding algorithm that used reduction approach, namely, entropy-based order-transformation forward algorithm (EOTFA) to compute the optimal state sequence of any generalized HHMM. This EOTFA algorithm involves a transformation of a generalized high-order HMM into an equivalent first-order HMM and an entropy-based decoding algorithm is developed based on the equivalent first-order HMM. This algorithm performs the computation based on the observational sequence and it requires $O(T\widetilde{N}^2)$ calculations, where $\widetilde{N}$ is the number of states in an equivalent first-order model and T is the length of observational sequence.

1. Introduction

State sequence for the Hidden Markov Model (HMM) is invisible but we can track the most likelihood state sequence based on the model parameter and a given observational sequence. The restored state has many applications especially when the hidden state sequence has meaningful interpretations for making predictions. For example, Ciriza et al. [1] have determined the optimal printing rate based on the HMM model parameter and an optimal time-out based on the restored states. The classical Viterbi algorithm is the most common technique for tracking state sequence from a given observational sequence [2]. However, it does not measure the uncertainty present in the solution. Proakis and Salehi [3] proposed a method for measuring the error of a single state but this method is unable to measure the error of the entire state sequence. Hernando et al. [4] proposed a method of using entropy for measuring the uncertainty of the state sequence of a first-order HMM tracked from a single observational sequence with a length of T. The method is based on the forward recursion algorithm integrated with entropy for computing the optimal state sequence. Mann and McCallum [5] developed an algorithm for computing the subsequent constrained entropy of HMM which is similar to the probabilistic model conditional random fields (CRF). Ilic [6] developed an algorithm based on forward-backward recursion over the entropy semiring, namely, the Entropy Semiring Forward-Backward (ESRFB) algorithm for a first-order HMM with a single observational sequence. ESRFB has lower memory requirement as compared with Mann and McCallum's algorithm for subsequent constrained entropy computation.

This paper is organized as follows. In Section 2, we define the generalized HHMM and present the extended entropy-based algorithm for computing the optimal state sequence developed by Hernando et al. [4] from a first-order to a generalized HHMM. In Section 3, we first review the high-order transformation algorithm proposed by Hadar and Messer [7] and then we introduce EOTFA, an entropy-based order-transformation forward algorithm for computing the optimal state sequence for any generalized HHMM. We discuss future research in Section 4 on entropy associated with state sequence of a generalized high-order HMM.

2. Entropy-Based Decoding Algorithm with an Extended Approach

The uncertainty appearing in a HHMM can be quantified by entropy. This concept is applied for quantifying the uncertainty of the state sequence tracked from a single observational sequence and model parameters. The entropy of the state sequence equals 0 if there is only one possible state sequence that could have generated the observation sequence as there is no uncertainty in the solution. The higher this entropy the higher the uncertainty involved in tracking the hidden state sequence. We extend the entropy-based Viterbi algorithm developed by Hernando et al. [4] for computing the optimal state sequence from a first-order HMM to a high-order HMM, that is, kth-order, where $k \geq 2$. The state entropy in HHHM is computed recursively for the reason of reducing the computational complexity from $O(N^{kT})$ which used direct evaluation method to $O(TN^{k+1})$ in a HHMM where N is the number of states, T is the length of observational sequence, and k is the order of the Hidden Markov Model. In terms of memory space, the entropy-based Viterbi algorithm is more efficient which requires $O(N^{k+1})$ as compared to the classical Viterbi algorithm which requires $O(TN^{k+1})$. The memory space for the classical Viterbi algorithm is dependent on the length of the observational sequence due to the involvement of the process of "back tracking" in computing the optimal state sequence.

Before introducing the extended entropy-based Viterbi algorithm, we define a generalized high-order HMM, that is, kth-order HMM, where $k \geq 2$. These are followed by the definition of forward and backward probability variables for a generalized high-order HMM. These variables are required for computing the optimal state sequence in our decoding algorithm.

2.1. Elements of HHMM. HHMM involves two stochastic processes, namely, hidden state process and observation process. The hidden state process cannot be directly observed. However, it can be observed through the observation process. The observational sequence is generated by the observation process incorporated with the hidden state process. For a discrete HHMM, it must satisfy the following conditions.

The hidden state process $\{q_t\}_{t=2-k}^{T}$ is the kth-order Markov chain that satisfies

$$P\left(q_t \mid \{q_l\}_{l<t}\right) = P\left(q_t \mid \{q_l\}_{l=t-k}^{t-1}\right), \qquad (1)$$

where q_t denotes the hidden state at time t and $q_t \in S$, where S is the finite set of hidden states.

The observation process $\{o_t\}_{t=1}^{T}$ is incorporated with the hidden state process according to the state probability distribution that satisfies

$$P\left(o_t \mid \{o_l\}_{l<t}, \{q_l\}_{l\leq t}\right) = P\left(o_t \mid \{q_l\}_{l=t-k+1}^{t}\right), \qquad (2)$$

where o_t denotes the observation at time t and $o_t \in V$, where V is the finite set of observation symbols.

The elements for the kth-order discrete HMM are as follows:

(i) Number of distinct hidden states, N

(ii) Number of distinct observed symbols, M

(iii) Length of observational sequence, T

(iv) Observational sequence, $O = \{o_t, \ t = 1, 2, \ldots, T\}$

(v) Hidden state sequence, $Q = \{q_t, \ t = 2 - k, \ldots, T\}$

(vi) Possible values for each state, $S = \{s_i, \ i = 1, 2, \ldots, N\}$

(vii) Possible symbols per observation, $V = \{v_w, \ w = 1, 2, \ldots, M\}$

(viii) Initial hidden state probability vector, $\pi_{i_1}, \pi_{i_1 i_2}, \ldots, \pi_{i_1 \cdots i_k}$

where π_{i_1} is the probability that model will transit from state s_{i_1},

$$\pi_{i_1} = P\left(q_1 = s_{i_1}\right),$$

$$\sum_{i_1=1}^{N} \pi_{i_1} = 1, \qquad (3)$$

$$\pi_{i_1} \geq 0, \ 1 \leq i_1 \leq N$$

$\pi_{i_1 i_2}$ is the probability that model will transit from state s_{i_1} and state s_{i_2},

$$\pi_{i_1 i_2} = P\left(q_0 = s_{i_1}, q_1 = s_{i_2}\right),$$

$$\sum_{i_2=1}^{N} \pi_{i_1 i_2} = 1, \qquad (4)$$

$$\pi_{i_1 i_2} \geq 0, \ 1 \leq i_1, i_2 \leq N,$$

$$\vdots$$

$\pi_{i_1 \cdots i_k}$ is the probability that model will transit from state s_{i_1}, state $s_{i_2}, \ldots,$ and state s_{i_k},

$$\pi_{i_1 \cdots i_k} = P\left(q_{2-k} = s_{i_1}, q_{3-k} = s_{i_2}, \ldots, q_1 = s_{i_k}\right),$$

$$\sum_{i_k=1}^{N} \pi_{i_1 \cdots i_k} = 1, \qquad (5)$$

$$\pi_{i_1 \cdots i_k} \geq 0, \ 1 \leq i_1, i_2, \ldots, i_k \leq N$$

(ix) State transition probability matrix, $A_1 = \{a_{i_1 i_2}\}$, $A_2 = \{a_{i_1 i_2 i_3}\}, \ldots, A_k = \{a_{i_1 i_2 \cdots i_{k+1}}\}$,

where A_{j-1} is the j-dimensional state transition probability matrix and $a_{i_1 i_2 \cdots i_j}$, is the probability of a transition to state s_{i_j} given that it has had a transition from state s_{i_1} to state s_{i_2} to $\cdots$ and to state $s_{i_{j-1}}$ where $j = 2, \ldots, k+1$,

$$a_{i_1 \cdots i_j} = P\left(q_t = s_{i_j} \mid q_{t-1} = s_{i_{j-1}}, q_{t-2} = s_{i_{j-2}}, \ldots, q_{t-j+1}\right)$$

$$= s_{i_1}\Big),$$

$$\sum_{i_j=1}^{N} a_{i_1 i_2 \cdots i_j} = 1, \tag{6}$$

$$a_{i_1 i_2 \cdots i_j} \geq 0$$

(x) Emission probability matrix, $B_1 = \{b_{i_1}(v_m)\}$, $B_2 = \{b_{i_1 i_2}(v_m)\}, \ldots, B_k = \{b_{i_1 \cdots i_k}(v_m)\}$,

where B_1 is the two-dimensional emission probability matrix and $b_{i_1}(v_m)$ is a probability of observing v_m in state s_{i_1},

$$b_{i_1}(v_m) = P\left(o_t = v_m \mid q_t = s_{i_1}\right),$$

$$\sum_{m=1}^{M} b_{i_1}(v_m) = 1, \tag{7}$$

$$b_{i_1}(v_m) \geq 0, \ 1 \leq i_1 \leq N,$$

where B_j is the $j+1$-dimensional emission probability matrix and $b_{i_1 \cdots i_j}(v_m)$ is a probability of observing v_m in state s_{i_1} at time $t-j+1$, s_{i_2} at time $t-j+2, \ldots,$ and s_{i_j} at time t where $j = 2, \ldots, k,$

$$b_{i_1 \cdots i_j}(v_m)$$

$$= P\left(o_t = v_m \mid q_t = s_{i_j}, q_{t-1} = s_{i_{j-1}}, \ldots, q_{t-j+1} = s_{i_1}\right),$$

$$\sum_{m=1}^{M} b_{i_1 \cdots i_j}(v_m) = 1, \tag{8}$$

$$b_{i_1 \cdots i_j}(v_m) \geq 0, \ 1 \leq i_1, i_2, \ldots, i_j \leq N$$

For the kth-order discrete HMM, we summarize the parameters by using the components of $\lambda = (\pi_{i_1}, \pi_{i_1 i_2}, \ldots, \pi_{i_1 i_2 \cdots i_k}, A_1, A_2, \ldots, A_k, B_1, B_2, \ldots, B_k).$

Note that throughout this paper, we will use the following notations.

(i) $q_{1:t}$ denotes $q_1, q_2, \ldots, q_t$

(ii) $o_{1:t}$ denotes $o_1, o_2, \ldots, o_t$

2.2. Forward and Backward Probability.

The entropy-based algorithm proposed by Hernando et al. [4] for computing the optimal state sequence of a first-order HMM is incorporated with forward recursion process. Recently, high-order HMM are widely used in a variety of applications such as speech recognition [8, 9] and longitudinal data analysis [10, 11]. For the HHMM, the Markov assumption has been weakened since the next state not only depends on the current state but also depends on other historical states. The state dependency is subjected to the order of HMM. Hence we have to modify the classical forward and backward probability variables for the HHMM, that is, the kth-order HMM where $k \geq 2$ are shown as follows.

Definition 1. The forward variable $\alpha_t(i_2, i_3, \ldots, i_{k+1})$ in the kth-order HMM is a joint probability of the partial observation sequence $o_1, o_2, \ldots, o_t$ and the hidden state of s_{i_2} at time $t-k+1$, s_{i_3} at time $t-k+2, \ldots,$ $s_{i_{k+1}}$ at time t where $1 \leq t \leq T$. It can be denoted as

$$\alpha_t(i_2, i_3, \ldots, i_{k+1}) = P\left(o_1, o_2, \ldots, o_t, q_{t-k+1}\right)$$
$$= s_{i_2}, q_{t-k+2} = s_{i_3}, \ldots, q_t = s_{i_{k+1}} \mid \lambda\right). \tag{9}$$

From (9), $t = 1$ and $1 \leq i_2, i_3, \ldots, i_{k+1} \leq N$, we obtain the initial forward variable as

$$\alpha_1(i_2, i_3, \ldots, i_{k+1})$$

$$= P\left(o_1, q_{2-k} = s_{i_2}, q_{3-k} = s_{i_3}, \ldots, q_1 = s_{i_{k+1}} \mid \lambda\right)$$

$$= P\left(q_{2-k} = s_{i_2}, q_{3-k} = s_{i_3}, \ldots, q_1 = s_{i_{k+1}}\right) \tag{10}$$

$$\cdot P\left(o_1 \mid q_{2-k} = s_{i_2}, q_{3-k} = s_{i_3}, \ldots, q_1 = s_{i_{k+1}}\right)$$

$$= \pi_{i_2 i_3 \cdots i_{k+1}} b_{i_2 i_3 \cdots i_{k+1}}(o_1).$$

From (9), (10), and $1 \leq i_1, i_2, \ldots, i_k, i_{k+1} \leq N$, we obtain the recursive forward variable for $t = 2, \ldots, T,$

$$\alpha_t(i_2, i_3, \ldots, i_{k+1}) = P\left(o_1, o_2, \ldots, o_t, q_{t-k+1} = s_{i_2}, \right.$$

$$q_{t-k+2} = s_{i_3}, \ldots, q_t = s_{i_{k+1}} \mid \lambda\right)$$

$$= \sum_{i_1=1}^{N} P\left(o_1, o_2, \ldots, o_t, q_{t-k} = s_{i_1}, q_{t-k+1} = s_{i_2}, q_{t-k+2}\right.$$

$$= s_{i_3}, \ldots, q_t = s_{i_{k+1}} \mid \lambda\right)$$

$$= \sum_{i_1=1}^{N} P\left(o_1, o_2, \ldots, o_{t-1}, q_{t-k} = s_{i_1}, q_{t-k+1} \right. \tag{11}$$

$$= s_{i_2}, \ldots, q_{t-1} = s_{i_k} \mid \lambda\right) P\left(q_t = s_{i_{k+1}} \mid q_{t-k}\right.$$

$$= s_{i_1}, q_{t-k+1} = s_{i_2}, \ldots, q_{t-1} = s_{i_k}\right) \times P\left(o_t \mid q_{t-k+1}\right.$$

$$= s_{i_2}, q_{t-k+2} = s_{i_3}, \ldots, q_t = s_{i_{k+1}}\right)$$

$$= \left[\sum_{i_1=1}^{N} \alpha_{t-1}(i_1, i_2, \ldots, i_k) a_{i_1 i_2 \cdots i_{k+1}}\right] b_{i_2 i_3 \cdots i_{k+1}}(o_t).$$

Definition 2. The backward probability variable $\beta_t(i_1, i_2, \ldots, i_k)$ in the kth-order HMM is a conditional probability of the partial observation sequence $o_{t+1}, o_{t+2}, \ldots, o_T$ given the hidden state of s_{i_1} at time $t - k + 1$, s_{i_2} at time $t - k + 2, \ldots$, and s_{i_k} at time t. It can be denoted as

$$\beta_t (i_1, i_2, \ldots, i_k) = P\left(o_{t+1}, o_{t+2}, \ldots, o_T \mid q_{t-k+1}\right.$$

$$= s_{i_1}, q_{t-k+2} = s_{i_2}, \ldots, q_t = s_{i_k}, \lambda\big), \tag{12}$$

where $1 \leq t \leq T$, $1 \leq i_1, i_2, \ldots, i_k \leq N$.

We obtain the initial backward probability variable as

$$\beta_T (i_1, i_2, \ldots, i_k) = 1. \tag{13}$$

From (12) and (13), we obtain the recursive backward probability variable for $t = 1, 2, \ldots, T - 1$,

$$\beta_t (i_1, i_2, \ldots, i_k) = P\left(o_{t+1}, o_{t+2}, \ldots, o_T \mid q_{t-k+1} = s_{i_1},\right.$$

$$q_{t-k+2} = s_{i_2}, \ldots, q_t = s_{i_k}, \lambda\big)$$

$$= \sum_{i_{k+1}=1}^{N} P\left(o_{t+1}, o_{t+2}, \ldots, o_T, q_{t+1} = s_{i_{k+1}} \mid q_{t-k+1}\right.$$

$$= s_{i_1}, q_{t-k+2} = s_{i_2}, \ldots, q_t = s_{i_k}, \lambda\big)$$

$$= \sum_{i_{k+1}=1}^{N} P\left(o_{t+2}, \ldots, o_T \mid q_{t-k+2} = s_{i_2}, \ldots, q_{t+1}\right.$$

$$= s_{i_{k+1}}, \lambda\big) P\left(q_{t+1} = s_{i_{k+1}} \mid q_{t-k+1} = s_{i_1}, \ldots, q_t\right.$$

$$= s_{i_k}, \lambda\big) \times P\left(o_{t+1} \mid q_{t-k+2} = s_{i_2}, \ldots, q_{t+1} = s_{i_{k+1}}\right)$$

$$= \sum_{i_{k+1}=1}^{N} \beta_{t+1} (i_2, i_3, \ldots, i_{k+1}) \, a_{i_1 i_2 \cdots i_{k+1}} b_{i_2 i_3 \cdots i_{k+1}} (o_{t+1}). \tag{14}$$

The probability of the observational sequence given the model parameter for the first-order HMM can be represented by using the classical forward probability and backward probability variables [2]. We extend it to HHMM by using our modified forward probability and backward probability variables. The proof is due to Rabiner [2].

Definition 3. Let $\alpha_t(i_1, i_2, \ldots, i_k)$ and $\beta_t(i_1, i_2, \ldots, i_k)$ be the forward probability variable and backward probability variable, respectively; $P(O \mid \lambda)$ is presented using the forward and backward probability variables as

$$P(O \mid \lambda) = P(o_1, \ldots, o_T \mid \lambda)$$

$$= \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_k=1}^{N} \alpha_t (i_1, i_2, \ldots, i_k) \, \beta_t (i_1, i_2, \ldots, i_k). \tag{15}$$

Proof.

$$P(O \mid \lambda) = P(o_1, o_2, \ldots, o_T \mid \lambda) = \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_k=1}^{N} P\left(o_1, o_2, \ldots, o_T, q_{t-k+1} = s_{i_1}, q_{t-k+2} = s_{i_2}, \ldots, q_t = s_{i_k} \mid \lambda\right)$$

$$= \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_k=1}^{N} P\left(o_1, o_2, \ldots, o_t, q_{t-k+1} = s_{i_1}, q_{t-k+2} = s_{i_2}, \ldots, q_t = s_{i_k} \mid \lambda\right)$$

$$\times P\left(o_{t+1}, o_{t+2}, \ldots, o_T \mid q_{t-k+1} = s_{i_1}, q_{t-k+2} = s_{i_2}, \ldots, q_t = s_{i_k}, \lambda\right)$$

$$= \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_k=1}^{N} \alpha_t (i_1, i_2, \ldots, i_k) \, \beta_t (i_1, i_2, \ldots, i_k). \tag{16}$$

We now normalize both of the forward and backward probability variables. These normalized variables are required as the intermediate variables for the algorithm of state entropy computation. $\square$

Definition 4. The normalized forward probability variable $\widehat{\alpha}_t(i_2, i_3, \ldots, i_{k+1})$ in the kth-order HMM is defined as the probability of the hidden state of s_{i_2} at time $t - k + 1$, s_{i_3} at time $t - k + 2, \ldots$, $s_{i_{k+1}}$ at time t given the partial observation sequence $o_1, o_2, \ldots, o_t$ where $1 \leq t \leq T$.

$$\widehat{\alpha}_t (i_2, i_3, \ldots, i_{k+1}) = P\left(q_{t-k+1} = s_{i_2}, q_{t-k+2} = s_{i_3}, \ldots, q_t\right.$$

$$= s_{i_{k+1}} \mid o_1, o_2, \ldots, o_t\big). \tag{17}$$

From (10), (17), $t = 1$, and $1 \leq i_1, i_2, \ldots, i_k \leq N$, we obtain the initial normalized forward probability variable as

$$\widehat{\alpha}_1 (i_2, i_3, \ldots, i_{k+1})$$

$$= P\left(q_{2-k} = s_{i_2}, q_{3-k} = s_{i_3}, \ldots, q_1 = s_{i_{k+1}} \mid o_1\right)$$

$$= \frac{P\left(q_{2-k} = s_{i_2}, q_{3-k} = s_{i_3}, \dots, q_1 = s_{i_{k+1}}, o_1\right)}{P(o_1)}$$

$$= \frac{\pi_{i_2 i_3 \cdots i_{k+1}} b_{i_2 i_3 \cdots i_{k+1}}(o_1)}{r_0},$$

(18)

where

$$r_0 = \sum_{j_k=1}^{N} \cdots \sum_{j_1=1}^{N} \pi_{j_1 j_2 \cdots j_k} b_{j_1 j_2 \cdots j_k}(o_1). \tag{19}$$

From (11), (17), (18), and $t = 2, \dots, T$, $1 \le i_1, i_2, \dots, i_k, i_{k+1} \le N$, we obtain the recursive normalized forward probability variable as

$$\widehat{\alpha}_t\left(i_2, i_3, \dots, i_{k+1}\right)$$

$$= P\left(q_{t-k+1} = s_{i_2}, q_{t-k+2} = s_{i_3}, \dots, q_t = s_{i_{k+1}} \mid o_1, o_2, \dots, o_t\right)$$

$$= \frac{P\left(q_{t-k+1} = s_{i_2}, q_{t-k+2} = s_{i_3}, \dots, q_t = s_{i_{k+1}}, o_1, o_2, \dots, o_t\right)}{P(o_t \mid o_1, o_2, \dots, o_{t-1})} \quad (20)$$

$$= \frac{\left[\sum_{i_1=1}^{N} \widehat{\alpha}_{t-1}\left(i_1, i_2, \dots, i_k\right) a_{i_1 i_2 \cdots i_{k+1}}\right] b_{i_2 i_3 \cdots i_{k+1}}(o_t)}{r_t},$$

where

$$r_t = \sum_{j_k=1}^{N} \cdots \sum_{j_1=1}^{N} \sum_{i_1=1}^{N} \alpha_{t-1}\left(i_1, j_1, \dots, j_{k-1}\right)$$

$$\cdot a_{i_1 j_1 \cdots j_k} b_{j_1 j_2 \cdots j_k}(o_t).$$

(21)

Note that the normalization factor r_t ensures that the probabilities sum to one and it also represents the conditional observational probability [2].

Definition 5. The normalized backward probability variable $\widehat{\beta}_t(i_1, i_2, \dots, i_k)$ in the kth-order HMM is defined as the quotient of a conditional probability of the partial observation sequence $o_{t+1}, o_{t+2}, \dots, o_T$ given the hidden state of s_{i_1} at time $t - k + 1$, s_{i_2} at time $t - k + 2, \dots$, s_{i_k} at time t, and a conditional probability of the partial observation sequence $o_{t+1}, o_{t+2}, \dots, o_T$ given the entire observation sequence $o_1, o_2, \dots, o_T$. It can be denoted as

$$\widehat{\beta}_t\left(i_1, i_2, \dots, i_k\right)$$

$$= \frac{P\left(o_{t+1}, o_{t+2}, \dots, o_T \mid q_{t-k+1} = s_{i_1}, q_{t-k+2} = s_{i_2}, \dots, q_t = s_{i_k}\right)}{P(o_{t+1}, o_{t+2}, \dots, o_T \mid o_1, o_2, \dots, o_T)}, \quad (22)$$

where $1 \le t \le T$, $1 \le i_1, i_2, \dots, i_k \le N$

From (14) and (22), we obtain the recursive normalized backward probability variable as

$$\widehat{\beta}_t\left(i_1, i_2, \dots, i_k\right) = \frac{P\left(o_{t+1}, o_{t+2}, \dots, o_T \mid q_{t-k+1} = s_{i_1}, q_{t-k+2} = s_{i_2}, \dots, q_t = s_{i_k}\right)}{P\left(o_{t+1}, o_{t+2}, \dots, o_T \mid o_1, o_2, \dots, o_T\right)}$$

$$= \frac{\sum_{i_{k+1}=1}^{N} P\left(o_{t+1}, o_{t+2}, \dots, o_T \mid q_{t-k+1} = s_{i_1}, q_{t-k+2} = s_{i_2}, \dots, q_t = s_{i_k}, q_{t+1} = s_{i_{k+1}}\right)}{P\left(o_{t+1}, \dots, o_T \mid o_1, o_2, \dots, o_T\right)} \quad (23)$$

$$= \frac{\sum_{i_{k+1}=1}^{N} \widehat{\beta}_{t+1}\left(i_2, i_3, \dots, i_{k+1}\right) a_{i_1 i_2 \cdots i_{k+1}} b_{i_2 i_3 \cdots i_{k+1}}(o_{t+1})}{r_{t+1}},$$

where

$$r_{t+1} = \sum_{j_k=1}^{N} \cdots \sum_{j_1=1}^{N} \sum_{i_1=1}^{N} \alpha_t\left(i_1, j_1, \dots, j_{k-1}\right)$$

$$\cdot a_{i_1 j_1 \cdots j_k} b_{j_1 j_2 \cdots j_k}(o_{t+1}).$$

(24)

Our extended algorithm includes the normalized forward recursion given by (18) and (20). The extended algorithm for the kth-order HMM requires $O(TN^{k+1})$ calculations if we include either normalized forward recursion given by (18) and (20) or the normalized backward recursion given by (13) and (23). The direct evaluation method, in comparison, requires $O(N^{T+k-1})$ calculations where N is the number of states, T is the length of observational sequence, and k is the order of the Hidden Markov Model.

2.3. The Algorithm by Hernando et al. Hernando et al. [4] are pioneers for using entropy to compute the optimal state sequence of a first-order HMM with a single observational sequence. This algorithm is based on a first-order HMM normalized forward probability,

$$\widehat{\alpha}_t(j) = P\left(q_t = s_j \mid o_1, o_2, \dots, o_t\right), \tag{25}$$

auxiliary probability,

$$P\left(q_{t-1} = s_i \mid q_t = s_j, o_{1:t}\right), \tag{26}$$

and intermediate entropy,

$$H_t\left(s_j\right) = H\left(q_{1:t-1} \mid q_t = s_j, o_{1:t}\right). \tag{27}$$

The entropy-based algorithm for computing the optimal state sequence of a first-order HMM is as follows [4].

(1) Initialization. For $t = 1$ and $1 \leq j \leq N$,

$$H_1\left(s_j\right) = 0,$$

$$\widehat{\alpha}_1\left(j\right) = \frac{\pi_j b_j\left(o_1\right)}{\sum_{i=1}^{N} \pi_i b_i\left(o_1\right)}. \tag{28}$$

(2) Recursion. For $t = 2, \ldots, T - 1$, and $1 \leq j \leq N$,

$$\widehat{\alpha}_t\left(j\right) = \frac{\sum_{i=1}^{N} \widehat{\alpha}_{t-1}\left(i\right) a_{ij} b_j\left(o_t\right)}{\sum_{k=1}^{N} \sum_{i=1}^{N} \widehat{\alpha}_{t-1}\left(i\right) a_{ik} b_k\left(o_t\right)},$$

$$P\left(q_{t-1} = s_i \mid q_t = s_j, o_{1:t}\right) = \frac{a_{ij}\widehat{\alpha}_{t-1}\left(i\right)}{\sum_{k=1}^{N} \sum_{i=1}^{N} a_{ik}\widehat{\alpha}_{t-1}\left(i\right)},$$

$$H_t\left(s_j\right) = \sum_{i=1}^{N} \left[P\left(q_{t-1} = s_i \mid q_t = s_j, o_{1:t}\right) H_{t-1}\left(s_i\right)\right] \tag{29}$$

$$- \sum_{i=1}^{N} \left[P\left(q_{t-1} = s_i \mid q_t = s_j, o_{1:t}\right)\right.$$

$$\left. \cdot \log_2 P\left(q_{t-1} = s_i \mid q_t = s_j\right)\right].$$

(3) Termination

$$H_T\left(q_{1:T} \mid o_{1:T}\right) = \sum_{i=1}^{N} H_T\left(s_i\right) \widehat{\alpha}_T\left(i\right)$$

$$- \sum_{i=1}^{N} \widehat{\alpha}_T\left(i\right) \log_2 \widehat{\alpha}_T\left(i\right). \tag{30}$$

This algorithm performs the computation linearly with respect to the length of the observation sequence with computational complexity $O(TN^2)$. It requires the memory space of $O(N^2)$ which indicates that the memory space is independent of the observational sequence.

2.4. The Computation of the Optimal State Sequence for a HHMM. The extended classical Viterbi algorithm is commonly used for computing the optimal state sequence for HHMM. This algorithm provides the solution along with its likelihood. This likelihood probability can be determined as follows.

$$P\left(q_1, q_2, \ldots, q_T \mid o_1, o_2, \ldots, o_T\right)$$

$$= \frac{P\left(q_1, q_2, \ldots, q_T, o_1, o_2, \ldots, o_T\right)}{P\left(o_1, o_2, \ldots, o_T\right)}. \tag{31}$$

This probability can be used as a measure of quality of the solution. The higher the probability of our "solution," the better our "solution." Entropy can also be used for measuring the quality of the state sequence of the kth-order HMM. Hence, state entropy is proposed to be used for obtaining the optimal state sequence of a HHMM.

We define entropy of a discrete random variable as follows [12].

Definition 6. The entropy $H(X)$ of a discrete random variable X with a probability mass function $P(X = x)$ is defined as

$$H\left(X\right) = - \sum_{x \in X} P\left(x\right) \log_2 P\left(x\right). \tag{32}$$

When the log has a base of 2, the unit of the entropy is bits. Note that $0 \log 0 = 0$.

From (32), the entropy of the distribution for all possible state sequences is as follows:

$$H\left(q_1, q_2, \ldots, q_T \mid o_1, o_2, \ldots, o_T\right) = - \sum_{Q} \left[P\left(q_1\right.\right.$$

$$= s_{i_1}, q_2 = s_{i_2}, \ldots, q_T = s_{i_T} \mid o_1, o_2, \ldots, o_T\right)$$

$$\cdot \log_2 P\left(q_1 = s_{i_1}, q_2 = s_{i_2}, \ldots, q_T \tag{33}\right.$$

$$\left.\left. = s_{i_T} \mid o_1, o_2, \ldots, o_T\right)\right].$$

For the first-order HMM, if all N^T possible state sequences are equally likely to generate a single observational sequence with a length of T, then the entropy equals $T \log_2 N$. The entropy is $kT \log_2 N$ in the kth-order HMM if all N^{kT} possible state sequences are equally likely to produce the observational sequence.

For this extended algorithm, we require an intermediate state entropy variable, $H_t(s_{i_2}, s_{i_3}, \ldots, s_{i_{k+1}})$ that can be computed recursively using the previous variable, $H_{t-1}(s_{i_1}, s_{i_2}, \ldots, s_{i_k})$.

We define the state entropy variable for the kth-order HMM as follows.

Definition 7. The state entropy variable, $H_t(s_{i_2}, s_{i_3}, \ldots, s_{i_{k+1}})$, in the kth-order HMM is the entropy of all the state sequences that lead to state of s_{i_2} at time $t - k + 1$, s_{i_3} at time $t - k + 2, \ldots$, and $s_{i_{k+1}}$ at time t, given the observation sequence $o_1, o_2, \ldots, o_t$. It can be denoted as

$$H_t\left(s_{i_2}, s_{i_3}, \ldots, s_{i_{k+1}}\right) = H\left(q_{2-k:t-1} \mid q_{t-k+1}\right.$$

$$\left. = s_{i_2}, q_{t-k+2} = s_{i_3}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\right). \tag{34}$$

We analyse the state entropy for the kth-order HMM in detail, shown as follows.

From (34) and $t = 1$, we obtain the initial state entropy variable as

$$H_1\left(s_{i_2}, s_{i_3}, \ldots, s_{i_{k+1}}\right) = 0. \tag{35}$$

From (34) and (35) we obtain the recursion on the entropy for $t = 2, \ldots, T$, and $1 \leq i_1, i_2, \ldots, i_{k+1} \leq N$,

$$H_t\left(s_{i_2}, s_{i_3}, \ldots, i_{k+1}\right) = H\left(q_{2-k:t-1} \mid q_{t-k+1}\right.$$

$$\left. = s_{i_2}, q_{t-k+2} = s_{i_3}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\right)$$

$$= H\left(q_{t-k:t-1} \mid q_{t-k+1} = s_{i_2}, q_{t-k+2} = s_{i_3}, \ldots, q_t\right.$$

$$= s_{i_{k+1}}, o_{1:t}\right) + H\left(q_{2-k:t-2} \mid q_{t-k}, q_{t-k+1} = s_{i_2}, q_{t-k+2}\right.$$

$$\left. = s_{i_3}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\right),$$

(36)

where

$$H\left(q_{t-k:t-1} \mid q_{t-k+1} = s_{i_2}, q_{t-k+2} = s_{i_3}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\right)$$

$$= -\sum_{i_1=1}^{N} \left[P\left(q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k} \mid q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\right) \right.$$

$$\cdot \log_2 \left(P\left(q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k} \mid q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\right)\right)\right],$$

$$H\left(q_{2-k:t-2} \mid q_{t-k}, q_{t-k+1} = s_{i_2}, q_{t-k+2} = s_{i_3}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\right)$$

(37)

$$= \sum_{i_1=1}^{N} \left[P\left(q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k} \mid q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\right) \right.$$

$$\left. \cdot H\left(q_{2-k:t-2} \mid q_{t-k} = s_{i_1}, q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\right)\right]$$

$$= \sum_{i_1=1}^{N} \left[P\left(q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k} \mid q_{t-k+1} = s_{i_2}, q_{t-k+2} = s_{i_3}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\right) H_{t-1}\left(s_{i_1}, s_{i_2}, \ldots, s_{i_k}\right)\right].$$

The auxiliary probability $P(q_{t-k} = s_{i_1}, q_{t-k+1} = s_{i_2}, \ldots, q_{t-1} = s_{i_k} \mid q_{t-k+1} = s_{i_2}, \ldots, q_{t-1} = s_{i_k}, q_t = s_{i_{k+1}}, o_{1:t})$ is required for

our extended entropy-based algorithm. It can be computed as follows:

$$P\left(q_{t-k} = s_{i_1}, q_{t-k+1} = s_{i_2}, \ldots, q_{t-1} = s_{i_k} \mid q_{t-k+1} = s_{i_2}, \ldots, q_{t-1} = s_{i_k}, q_t = s_{i_{k+1}}, o_{1:t}\right)$$

$$= \frac{P\left(q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}}, o_{t-k+1}, \ldots, o_t \mid q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k}, o_{1:t-1}\right) P\left(q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k} \mid o_{1:t-1}\right)}{P\left(q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}}, o_{t-k+1}, \ldots, o_t \mid o_{1:t-1}\right)}$$

$$= \frac{P\left(o_{t-k+1}, \ldots, o_t \mid q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}}\right) P\left(q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}} \mid q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k}\right) P\left(q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k} \mid o_{1:t-1}\right)}{P\left(o_{t-k+1}, \ldots, o_t \mid q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}}\right) P\left(q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}} \mid o_{1:t-1}\right)}$$

(38)

$$= \frac{P\left(q_{t-k} = s_{i_2}, \ldots, q_t = s_{i_{k+1}} \mid q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k}\right) P\left(q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k} \mid o_{1:t-1}\right)}{\sum_{j_k=1}^{N} \sum_{j_{k-1}=1}^{N} \cdots \sum_{j_1=1}^{N} P\left(q_{t-k+1} = s_{j_2}, \ldots, q_t = s_{j_{k+1}} \mid q_{t-k} = s_{j_1}, \ldots, q_{t-1} = s_{j_k}\right) P\left(q_{t-k} = s_{j_1}, \ldots, q_{t-1} = s_{j_k} \mid o_{1:t-1}\right)}$$

$$= \frac{a_{i_1 i_2 \cdots i_k i_{k+1}} \widehat{\alpha}_{t-1}\left(i_1, i_2, \ldots, i_k\right)}{\sum_{j_k=1}^{N} \sum_{j_{k-1}=1}^{N} \cdots \sum_{j_1=1}^{N} a_{j_1 j_2 \cdots j_k i_{k+1}} \widehat{\alpha}_{t-1}\left(j_1, j_2, \ldots, j_k\right)}.$$

For the final process of our extended algorithm, we are required to compute the conditional entropy $H(q_{1:T} \mid o_{1:T})$ which can be expanded as follows:

$$H\left(q_{1:T} \mid o_{1:T}\right) = H\left(q_{1:T-k} \mid q_{T-k+1} = s_{i_1}, q_{T-k+2} = s_{i_2},\right.$$

$$\left. q_{T-k+3} = s_{i_3}, \ldots, q_T = s_{i_k}, o_{1:T}\right) + H\left(q_{T-k+1:T} \mid o_{1:T}\right)$$

$$= \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_k=1}^{N} H_T\left(s_{i_1}, s_{i_2}, \ldots, s_{i_k}\right)$$

$$\cdot \widehat{\alpha}_T\left(i_1, i_2, \ldots, i_k\right)$$

$$- \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \cdots \sum_{i_k=1}^{N} \widehat{\alpha}_T\left(i_1, i_2, \ldots, i_k\right)$$

$$\cdot \log_2 \left(\widehat{\alpha}_T\left(i_1, i_2, \ldots, i_k\right)\right).$$

(39)

The following basic properties of HMM and entropy are used for proving Lemma 8.

(i) According to the generalized high-order HMM, state $q_{t-k-j+1}$, $j \geq 2$ and q_t are statistically independent given $q_{t-k}, q_{t-k+1}, q_{t-k+2}, \ldots, q_{t-1}$. The same applies to $q_{t-k-j+1}$,

$j \geq 2$ and o_t are statistically independent given $q_{t-k}, q_{t-k+1}, q_{t-k+2}, \ldots, q_{t-1}$.

(ii) According to the basic property of entropy [12],

$$H(X \mid Y = y) = H(X) \tag{40}$$

if X and Y are independent.

We now introduce the following lemma for the kth-order HMM. The following proof is due to Hernando et al. [4].

Lemma 8. *For the kth-order HMM, the entropy of the state sequence up to time $t - k - 1$, given the states from time $t - k$ to time $t - 1$ and the observations up to time $t - 1$, is conditionally independent of the state and observation at time t*

$$H_{t-1}\left(s_{i_1}, s_{i_2}, \ldots, s_{i_k}\right) = H\left(q_{1:t-2} \mid q_{t-k} = s_{i_1}, q_{t-k+1} \right. \tag{41}$$

$$= s_{i_2}, q_{t-k+2} = s_{i_3}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\Big).$$

Proof.

$$H\left(q_{1:t-2} \mid q_{t-k} = s_{i_1}, q_{t-k+1} = s_{i_2}, q_{t-k+2} = s_{i_3}, \ldots, q_t \right.$$

$$= s_{i_{k+1}}, o_{1:t}\Big) = H\left(q_{1:t-2} \mid q_{t-k} = s_{i_1}, q_{t-k+1}\right.$$

$$= s_{i_2}, q_{t-k+2} = s_{i_3}, \ldots, q_{t-1} = s_{i_k}, o_{1:t-1}, q_t = s_{i_{k+1}}, o_t\Big)$$

$$= H\left(q_{1:t-2} \mid q_{t-k} = s_{i_1}, q_{t-k+1} = s_{i_2}, q_{t-k+2}\right.$$

$$= s_{i_3}, \ldots, q_{t-1} = s_{i_k}, o_{1:t-1}\Big) = H_{t-1}\left(s_{i_1}, s_{i_2}, \ldots, s_{i_k}\right). \tag{42}$$

Our extended entropy-based algorithm for computing the optimal state sequence is based on normalized forward recursion variable, state entropy recursion variable, and auxiliary probability. From (18), (20), (35), (36), (38), and (39), we construct the extended entropy-based decoding algorithm for the kth-order HMM as follows:

(1) Initialization. For $t = 1$ and $1 \leq i_2, i_3, \ldots, i_{k+1} \leq N$,

$$H_1\left(s_{i_2}, s_{i_3}, \ldots, s_{i_{k+1}}\right) = 0,$$

$$\widehat{\alpha}_1\left(i_2, i_3, \ldots, i_{k+1}\right) \tag{43}$$

$$= \frac{\pi_{i_2 i_3 \cdots i_{k+1}} b_{i_2 i_3 \cdots i_{k+1}}(o_1)}{\sum_{j_k=1}^{N} \sum_{j_{k-1}=1}^{N} \cdots \sum_{j_1=1}^{N} \pi_{j_1 j_2 \cdots j_k} b_{j_1 j_2 \cdots j_k}(o_1)}.$$

(2) Recursion. For $t = 2, \ldots, T-1$, and $1 \leq i_1, i_2, \ldots, i_{k+1} \leq N$,

$$\widehat{\alpha}_t\left(i_2, i_3, \ldots, i_{k+1}\right) = \frac{\sum_{i_1=1}^{N} \widehat{\alpha}_{t-1}\left(i_1, i_2, \ldots, i_k\right) a_{i_1 i_2 \cdots i_k i_{k+1}} b_{i_2 i_3 \cdots i_k i_{k+1}}(o_t)}{\sum_{j_k=1}^{N} \cdots \sum_{j_1=1}^{N} \sum_{i_1=1}^{N} \widehat{\alpha}_{t-1}\left(i_1, j_1, \ldots, j_{k-1}\right) a_{i_1 j_1 \cdots j_k} b_{j_2 j_3 \cdots j_{k+1}}(o_t)},$$

$$P\left(q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k} \mid q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\right) = \frac{a_{i_1 i_2 \cdots i_k i_{k+1}} \widehat{\alpha}_{t-1}\left(i_1, i_2, \ldots, i_k\right)}{\sum_{j_k=1}^{N} \sum_{j_{k-1}=1}^{N} \cdots \sum_{j_1=1}^{N} a_{j_1 j_2 \cdots j_k i_{k+1}} \widehat{\alpha}_{t-1}\left(j_1, j_2, \ldots, j_k\right)},$$

$$H_t\left(s_{i_2}, s_{i_3}, \ldots, s_{i_{k+1}}\right) = \sum_{i_1=1}^{N} \left[P\left(q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k} \mid q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\right) H_{t-1}\left(s_{i_1}, s_{i_2}, \ldots, s_{i_k}\right)\right] \tag{44}$$

$$- \sum_{i_1=1}^{N} \left[P\left(q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k} \mid q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\right)\right.$$

$$\cdot \log_2\left(P\left(q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k} \mid q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}}, o_{1:t}\right)\right)\Big].$$

(3) Termination

$$H\left(q_{1:T} \mid o_{1:T}\right) = \sum_{i_1=1}^{N} \cdots \sum_{i_k=1}^{N} H_T\left(s_{i_1}, s_{i_2}, \ldots, s_{i_k}\right)$$

$$\cdot \widehat{\alpha}_T\left(i_1, i_2, \ldots, i_k\right) \tag{45}$$

$$- \sum_{i_1=1}^{N} \cdots \sum_{i_k=1}^{N} \widehat{\alpha}_T\left(i_1, i_2, \ldots, i_k\right) \log_2 \widehat{\alpha}_T\left(i_1, i_2, \ldots, i_k\right).$$

This extended algorithm performs the computation of the optimal state sequence linearly with respect to the length of observational sequence which requires $O(TN^{k+1})$ calculation and it has memory space that is independent of the length

of observational sequence, $O(N^{k+1})$, since $\widehat{\alpha}_t(i_2, i_3, \ldots, i_{k+1})$, $H_t(s_{i_2}, s_{i_3}, \ldots, s_{i_{k+1}})$, $P(q_{t-k} = s_{i_1}, \ldots, q_{t-1} = s_{i_k} \mid q_{t-k+1} = s_{i_2}, \ldots, q_t = s_{i_{k+1}}, o_{1:t})$ should be computed only once in tth iteration and, having been used for the computation of $(t + 1)$th, they can be deleted from the space storage. □

2.5. Numerical Illustration for the Second-Order HMM. We consider a second-order HMM for illustrating our extended entropy-based algorithm in computing the optimal state sequence. Let us assume that this second-order HMM has the state space S, which is $S = \{s_1, s_2\}$ and the possible symbols per observation which is $O = \{v_1, v_2, v_3\}$.

The graphical representation of the first-order HMM that is used for the numerical example in this section is given in
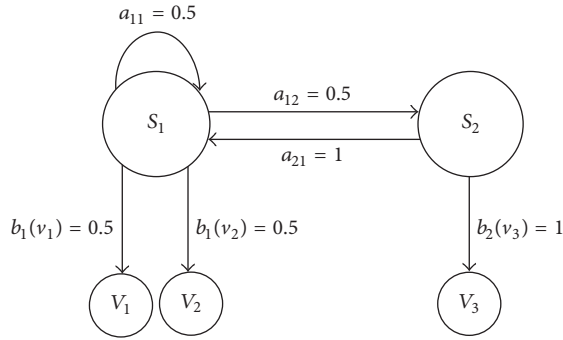
FIGURE 1: The graphical diagram shows a first-order HMM with 2 states and 3 observational symbols.
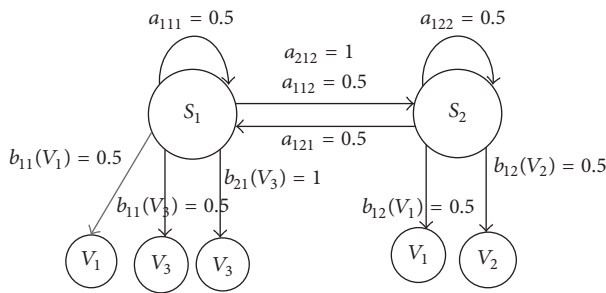

FIGURE 2: The graphical diagram shows a second-order HMM with 2 states and 3 observational symbols.

Figure 1. The second-order HMM in Figure 2 is developed based on the first-order HMM in Figure 1 which has two states and three observational symbols. A HMM of any order has the parameters of $\lambda = (\pi, A, B)$ where π is the initial state probability vector, A is the state transition probability matrix, and B is the emission probability matrix. Note that the matrices of A and B whose components are indicated as $a_{i_1 i_2}$, $a_{i_1 i_2 i_3}$, $b_{i_2}(o_t = v_m)$ and $b_{i_2 i_3}(o_t = v_m)$ where $1 \leq i_1, i_2, i_3 \leq 2$ and $1 \leq m \leq 3$ can be obtained from Figures 1 and 2. However, the initial state probability vector is not shown in the above graphical diagrams.

The initial state probability vectors for the first-order and second-order HMM are shown as follows:

$$\pi_1 = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix},$$

$$\pi_2 = \begin{bmatrix} 0.5 & 0 \end{bmatrix}, \qquad (46)$$

$$\pi_3 = \begin{bmatrix} 0.5 & 0 \end{bmatrix}.$$

$\pi_1 = \{\dot{\pi}_{i_2}\}$ is the initial state probability vector for the first-order HMM and $\pi_2 = \{\dot{\pi}_{i_2 1}\}$ and $\pi_3 = \{\dot{\pi}_{i_2 2}\}$ are the initial state probability vectors for the second-order HMM where $\dot{\pi}_{i_2} = P(q_1 = s_{i_2})$, $\dot{\pi}_{i_2 1} = P(q_1 = s_1, q_0 = s_{i_2})$, $\dot{\pi}_{i_2 2} = P(q_1 = s_2, q_0 = s_{i_2})$, and $1 \leq i_2 \leq 2$.

The state transition probability matrices for the first-order and second-order HMMs are shown as follows:

$$A_1 = \begin{bmatrix} 0.5 & 0.5 \\ 1 & 0 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} 0.5 & 0.5 \\ 0 & 0 \end{bmatrix}, \qquad (47)$$

$$A_3 = \begin{bmatrix} 0.5 & 0.5 \\ 1 & 0 \end{bmatrix}.$$

$A_1 = \{a_{i_1 i_2}\}$ is the state transition probability matrix for the first-order HMM and $A_2 = \{a_{i_1 i_2 1}\}$ and $A_3 = \{a_{i_1 i_2 2}\}$ are the state transition probability matrices for the second-order HMM where $a_{i_1 i_2} = P(q_t = s_{i_2} \mid q_{t-1} = s_{i_1})$, $a_{i_1 i_2 1} = P(q_t = s_1 \mid q_{t-1} = s_{i_2}, q_{t-2} = s_{i_1})$, $a_{i_1 i_2 2} = P(q_t = s_2 \mid q_{t-1} = s_{i_2}, q_{t-2} = s_{i_1})$, and $1 \leq i_1, i_2 \leq 2$

The emission probability matrices for the first-order and second-order HMMs are shown as follows:

$$B_1 = \begin{bmatrix} 0.5 & 0 \\ 0.5 & 0 \\ 0 & 1 \end{bmatrix},$$

$$B_2 = \begin{bmatrix} 0.5 & 0.5 \\ 0 & 0 \end{bmatrix},$$

$$B_3 = \begin{bmatrix} 0 & 0.5 \\ 0 & 0 \end{bmatrix}, \qquad (48)$$

$$B_4 = \begin{bmatrix} 0.5 & 0 \\ 1 & 0 \end{bmatrix}.$$

$B_1 = \{b_{i_2}(o_t = v_m)\}$ is the emission probability matrix for the first-order HMM and $B_2 = \{b_{i_2 i_3}(o_t = v_1)\}$, $B_3 = \{b_{i_2 i_3}(o_t = v_2)\}$, and $B_4 = \{b_{i_2 i_3}(o_t = v_3)\}$ are the emission probability matrices for the second-order HMM where $b_{i_2}(o_t = v_m) = P(o_t = v_m \mid q_t = s_{i_2})$, $b_{i_2 i_3}(o_t = v_1) = P(o_t = v_1 \mid q_t = s_{i_3}, q_{t-1} = s_{i_2})$, $b_{i_2 i_3}(o_t = v_2) = P(o_t = v_2 \mid q_t = s_{i_3}, q_{t-1} = s_{i_2})$, and $b_{i_2 i_3}(o_t = v_3) = P(o_t = v_3 \mid q_t = s_{i_3}, q_{t-1} = s_{i_2})$.

The following is the observational sequence that we used for illustrating our extended algorithm:

$$o_{1:6} = (o_1 = v_1, o_2 = v_1, o_3 = v_3, o_4 = v_2, o_5 = v_3, o_6 \qquad (49)$$
$$= v_1).$$

We applied our extended algorithm for computing the optimal state sequence based on state entropy. The computed value of the state entropy is shown in Figure 3.

The total entropy after each time step is displayed at the bottom of Figure 3. For example, after receiving the second observation, that is, $o_{1:2} = (o_1 = v_1, o_2 = v_1)$, it has produced two state sequences which are $q_{1:2} = (q_1 = s_1, q_2 = s_1)$ and $q_{1:2} = (q_1 = s_1, q_2 = s_2)$ as shown by the bold arrows. Each possible state sequence has a probability of 0.5; that is, $\hat{\alpha}_2(1, 1) = \hat{\alpha}_2(1, 2) = 0.5$, and hence the total entropy is 1 bit.

Obs / State	$o_1 = v_1$	$o_2 = v_1$	$o_3 = v_3$	$o_4 = v_2$	$o_5 = v_3$	$o_6 = v_1$
s_1	$H_1(1,1)=0$ $\hat{\alpha}_1(1,1)=0.5$ $H_1(2,1)=0$ $\hat{\alpha}_1(2,1)=0$	$H_2(1,1)=0$ $\hat{\alpha}_2(1,1)=0.5$ $H_2(2,1)=0$ $\hat{\alpha}_2(2,1)=0$	$H_3(1,1)=0.5$ $\hat{\alpha}_3(1,1)=0.33$ $H_3(2,1)=0.5$ $\hat{\alpha}_3(2,1)=0.67$	$H_4(1,1)=0$ $\hat{\alpha}_4(1,1)=0$ $H_4(2,1)=0$ $\hat{\alpha}_4(2,1)=0$	$H_5(1,1)=0$ $\hat{\alpha}_5(1,1)=0$ $H_5(2,1)=0$ $\hat{\alpha}_5(2,1)=1$	$H_6(1,1)=0$ $\hat{\alpha}_6(1,1)=0$ $H_6(2,1)=0$ $\hat{\alpha}_6(2,1)=0$
s_2	$H_1(1,2)=0$ $\hat{\alpha}_1(1,2)=0.5$ $H_1(2,2)=0$ $\hat{\alpha}_1(2,2)=0$	$H_2(1,2)=0$ $\hat{\alpha}_2(1,2)=0.5$ $H_2(2,2)=0$ $\hat{\alpha}_2(2,2)=0$	$H_3(1,2)=0$ $\hat{\alpha}_3(1,2)=0$ $H_3(2,2)=0$ $\hat{\alpha}_3(2,2)=0$	$H_4(1,2)=0$ $\hat{\alpha}_4(1,2)=1$ $H_4(2,2)=0$ $\hat{\alpha}_4(2,2)=0$	$H_3(1,2)=0$ $\hat{\alpha}_3(1,2)=0$ $H_3(2,2)=0$ $\hat{\alpha}_3(2,2)=0$	$H_6(1,2)=0$ $\hat{\alpha}_6(1,2)=1$ $H_6(2,2)=0$ $\hat{\alpha}_6(2,2)=0$
Total entropy	1	1	1.41	0	0	0

FIGURE 3: The evolution of the trellis structure of the second-order HMM with the observation sequence $o_{1:6} = (o_1 = v_1, o_2 = v_1, o_3 = v_3, o_4 = v_2, o_5 = v_3, o_6 = v_1)$.

However, after receiving the fourth observation, that is, $o_{1:4} = (o_1 = v_1, o_2 = v_1, o_3 = v_3, o_4 = v_2)$, it has produced one state sequence which is $q_{1:4} = (q_1 = s_1, q_2 = s_2, q_3 = s_1, q_4 = s_2)$ as shown by the dashed arrow. This possible state sequence has a probability of 1, that is, $\hat{\alpha}_4(1,2) = 1$, and hence the total entropy is 0 bit. After receiving the sixth observation, this second-order HMM has produced only one possible optimal state sequence; that is, $q_{1:6} = (q_1 = s_1, q_2 = s_2, q_3 = s_1, q_4 = s_2, q_5 = s_1, q_6 = s_2)$ with the total entropy of 0 which indicates that there is no uncertainty.

3. Entropy-Based Decoding Algorithm with a Reduction Approach

The extended entropy-based Viterbi algorithm in Section 2 has addressed only the issue related to memory space but this algorithm is not able to overcome the computational complexity. In this section, we introduce an efficient entropy-based algorithm that used reduction approach, namely, entropy-based order-transformation forward algorithm (EOTFA) to compute the optimal state sequence based on entropy of any generalized HHMM. This algorithm has addressed issues related to memory space and computational complexity.

3.1. Transforming a High-Order HMM with a Single Observational Sequence. This EOTFA algorithm involves a transformation of a generalized high-order HMM into an equivalent first-order HMM and an algorithm is developed based on the equivalent first-order HMM. This algorithm performs the computation based on the observational sequence and it requires $O(T\widetilde{N}^2)$ calculations, where $\widetilde{N}$ is the number of states in an equivalent first-order model and T is the length of observational sequence.

The transformation of a generalized high-order HMM into an equivalent first-order HMM is based on Hadar and Messer's method [7].

Suppose $\widetilde{Q}_t = (q_t, q_{t-1}, \ldots, q_{t-k+1})$ for $1 \le t \le T$; then the hidden state process $\{\widetilde{Q}_t\}_{t=1}^T$ of the kth-order Markov chain satisfies

$$P\left(\widetilde{Q}_t \mid \{\widetilde{Q}_l\}_{l<t}\right)$$
$$= P\left(q_t, q_{t-1}, \ldots, q_{t-k+1} \mid q_{t-1}, q_{t-2}, \ldots, q_{2-k}\right)$$
$$= P\left(q_t \mid q_{t-1}, q_{t-2}, \ldots, q_{t-k}\right) \tag{50}$$
$$= P\left(q_t, q_{t-1}, \ldots, q_{t-k+1} \mid q_{t-1}, q_{t-2}, \ldots, q_{t-k}\right)$$
$$= P\left(\widetilde{Q}_t \mid \widetilde{Q}_{t-1}\right),$$

where $\widetilde{Q}_t$ takes the value from the set of hidden states $\widetilde{S} = \{s_i, i = 1, 2, \ldots, N\}^k$. Hence, the hidden state process $\{\widetilde{Q}_t\}_{t=1}^T$ forms the first-order HMM Markov process.

The observation process $\{o_t\}_{t=1}^T$ satisfies

$$P\left(o_t \mid \{o_l\}_{l<t}, \{\widetilde{Q}_l\}_{l\le t}\right) = P\left(o_t \mid \{o_l\}_{l\le t-1}, \{q_l\}_{l\le t}\right)$$
$$= P\left(o_t \mid \{q_l\}_{l\le t}\right) = P\left(o_t \mid \{q_l\}_{l=t-k}^t\right) \tag{51}$$
$$= P\left(o_t \mid \widetilde{Q}_t\right).$$

Hence, the hidden state process $\{\widetilde{Q}_t\}_{t=1}^T$ and the observation process $\{o_t\}_{t=1}^T$ form the first-order HMM.

Remarks 9. (i)

$$P\left(\widetilde{Q}_t \mid \widetilde{Q}_{t-1}\right) = P\left(\widetilde{Q}_t\right.$$

$$= \left[q_{t-k+1} = s_{i_2}, q_{t-k+2} = s_{i_3}, \dots, q_t = s_{i_{k+1}}\right] \mid \widetilde{Q}_{t-1}$$

$$= \left[q_{t-k} = s_{i_1}, q_{t-k+1} = s_{i_2}, \dots, q_{t-1} = s_{i_k}\right]\right) = P\left(\widetilde{Q}_t \quad (52)\right.$$

$$= \left[s_{i_2}, s_{i_3}, \dots, s_{i_{k+1}}\right] \mid \widetilde{Q}_{t-1} = \left[s_{i_1}, s_{i_2}, \dots, s_{i_k}\right]\right)$$

$$= P\left(\widetilde{Q}_t = s_{i_2 i_3 \cdots i_{k+1}} \mid \widetilde{Q}_{t-1} = s_{i_1 i_2 \cdots i_k}\right),$$

where $[s_{i_1}, s_{i_2}, \dots, s_{i_k}]$ and $[s_{i_2}, s_{i_3}, \dots, s_{i_{k+1}}] \in \widetilde{S}$.

(ii)

$$P\left(o_t \mid \widetilde{Q}_t\right) = P\left(o_t \mid \widetilde{Q}_t\right.$$

$$= \left[q_{t-k+1} = s_{i_2}, q_{t-k+2} = s_{i_3}, \dots, q_t = s_{i_{k+1}}\right]\right)$$

$$= P\left(o_t \mid \widetilde{Q}_t = \left[s_{i_2}, \dots, s_{i_{k+1}}\right]\right) = P\left(o_t \mid \widetilde{Q}_t\right. \quad (53)$$

$$= s_{i_2 i_3 \cdots i_{k+1}}\right),$$

where $[s_{i_2}, s_{i_3}, \dots, s_{i_{k+1}}] \in \widetilde{S}$.

Note that we assume $s_{i_1}, s_{i_2}, \dots, s_{i_k} = s_{i_1 i_2 \cdots i_k}$ and $s_{i_2}, s_{i_3}, \dots, s_{i_{k+1}} = s_{i_2 i_3 \cdots i_{k+1}}$.

The elements for the transformation of a high-order into an equivalent first-order discrete HMM are as follows:

(i) Number of distinct hidden states, $\widetilde{N}$

(ii) Number of distinct observed symbols, M

(iii) Length of observational sequence, T

(iv) Observational sequence, $O = \{o_t, \ t = 1, 2, \dots, T\}$

(v) Hidden state sequence, $\widetilde{Q} = \{\widetilde{Q}_t, \ t = 1, 2, \dots, T\}$

(vi) Possible values for each state, $\widetilde{S} = \{s_i, \ i = 1, 2, \dots, N\}^k$

(vii) Possible symbols per observation, $\widetilde{V} = \{v_w, \ w = 1, 2, \dots, M\}$

(viii) Initial hidden state probability vector, $\widetilde{\pi} = \{\widetilde{\pi}_i\}$, and $\widetilde{\pi}_i$ is the probability that model will transit from state $\widetilde{s}_i = [s_{i_1}, s_{i_2}, \dots, s_{i_k}] = s_{i_1 i_2 \cdots i_k}$, where

$$\widetilde{\pi}_i = P\left(\widetilde{Q}_1 = \widetilde{s}_i\right),$$

$$\sum_{i=1}^{\widetilde{N}} \widetilde{\pi}_i = 1, \quad (54)$$

$$\widetilde{\pi}_i \geq 0$$

(ix) State transition probability matrix, $\widetilde{A} = \{\widetilde{a}_{ij}\}$ and a_{ij} is the probability of a transition from state

$\widetilde{s}_i = [s_{i_1}, s_{i_2}, \dots, s_{i_k}]$ at time $t-1$ to state $\widetilde{s}_j = [s_{i_2}, s_{i_3}, \dots, s_{i_{k+1}}]$ at time t where

$$\widetilde{a}_{ij} = P\left(\widetilde{Q}_t = \widetilde{s}_j \mid \widetilde{Q}_{t-1} = \widetilde{s}_i\right),$$

$$\sum_{j=1}^{\widetilde{N}} \widetilde{a}_{ij} = 1, \quad (55)$$

$$\widetilde{a}_{ij} \geq 0,$$

where the first $k-1$ entries of $\widetilde{s}_i$ are equal to the last $k-1$ entries of $\widetilde{s}_j$

(x) Emission probability matrix, $\widetilde{B} = \{\widetilde{b}_i(v_m)\}$, and $\widetilde{b}_i(v_m)$ is a probability of observing v_m in state $\widetilde{s}_i = [s_{i_1}, s_{i_2}, \dots, s_{i_k}]$ at time t:

$$\widetilde{b}_i(v_m) = P\left(o_t = v_m \mid \widetilde{Q}_t = \widetilde{s}_i\right),$$

$$\sum_{m=1}^{M} \widetilde{b}_i(v_m) = 1, \quad (56)$$

$$\widetilde{b}_i(v_m) \geq 0.$$

3.2. The Forward and Backward Probabilities Variables for the Transformed Model. In this subsection, we omit the derivations for forward and backward probability variables since the derivations are similar to the derivations in Section 2.2.

The forward recursion variable for the transformed model at time t is as follows:

$$\widetilde{\alpha}_t(j) = P\left(o_1, o_2, \dots, o_t, \widetilde{Q}_t = \widetilde{s}_j \mid \lambda\right)$$

$$= P\left(o_1, o_2, \dots, o_t, \widetilde{Q}_t = s_{i_2 i_3 \cdots i_k} \mid \lambda\right) \quad (57)$$

$$= \sum_{i=1}^{\widetilde{N}} \widetilde{\alpha}_{t-1}(i)\, \widetilde{a}_{ij} \widetilde{b}_j(o_t).$$

The backward recursion variable for the transformed model at time t is as follows:

$$\widetilde{\beta}_t(i) = P\left(o_{t+1}, o_{t+2}, \dots, o_T \mid \widetilde{Q}_t = \widetilde{s}_i, \lambda\right)$$

$$= P\left(o_{t+1}, o_{t+2}, \dots, o_T \mid \widetilde{Q}_t = s_{i_1 i_2 \cdots i_k}\right) \quad (58)$$

$$= \left[\sum_{j=1}^{\widetilde{N}} \widetilde{\beta}_{t+1}(j)\, \widetilde{a}_{ij}\right] \widetilde{b}_j(o_{t+1}).$$

The normalized forward variable at time t is as follows:

$$\widetilde{\alpha}_t^\bullet(j) = P\left(\widetilde{Q}_t \mid o_{1:t}\right) = \frac{\sum_{i=1}^{\widetilde{N}} \widetilde{\alpha}_{t-1}^\bullet(i)\, \widetilde{a}_{ij} \widetilde{b}_j(o_t)}{r_t^\bullet}, \quad (59)$$

where $r_t^\bullet = \sum_{j=1}^{\widetilde{N}} \sum_{i=1}^{\widetilde{N}} \widetilde{\alpha}_{t-1}(i) \widetilde{a}_{ij} \widetilde{b}_j(o_t)$.

The normalized backward variables at time t is as follows:

$$\widetilde{\beta}_t^\bullet(i) = \frac{P\left(o_{t+1:T} \mid \widetilde{Q}_t\right)}{P\left(o_{t+1:T} \mid o_{o:t}\right)} = \frac{\sum_{j=1}^{\widetilde{N}} \widetilde{\beta}_{t+1}^\bullet(j)\, \widetilde{a}_{ij} \widetilde{b}_j(o_{t+1})}{r_{t+1}^\bullet}, \quad (60)$$

where $r_{t+1}^\bullet = \sum_{j=1}^{\widetilde{N}} \sum_{i=1}^{\widetilde{N}} \widetilde{\alpha}_t(i) \widetilde{a}_{ij} \widetilde{b}_j(o_{t+1})$.

3.3. The Computation of the Optimal State Sequence for a HHMM.

For EOFTA algorithm, we require state entropy variable, $H_t(\widetilde{s}_j)$, that can be computed recursively using the previous variable, $H_{t-1}(\widetilde{s}_i)$.

We define the state entropy variable as follows.

Definition 10. The state entropy variable, $H_t(\widetilde{s}_j)$, in an order-transformation HMM, is the entropy of all the paths that lead to state of $\widetilde{s}_j$ at time t, given the observations $o_1, o_2, \ldots, o_t$. It can be denoted as

$$H_t\left(\widetilde{s}_j\right) = H\left(\widetilde{Q}_{1:t-1} \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right). \tag{61}$$

From (61) at $t = 1$, we obtain the initial state entropy variable as

$$H_1\left(\widetilde{s}_j\right) = 0. \tag{62}$$

From (61) and (62), we obtain the recursion on the entropy for $t = 2, \ldots, T - 1$, and $1 \le i, j \le \widetilde{N}$

$$
\begin{aligned}
H_t\left(\widetilde{s}_j\right) &= H\left(\widetilde{Q}_{1:t-1} \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right) \\
&= H\left(\widetilde{Q}_{1:t-2}, \widetilde{Q}_{t-1} \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right) \\
&= H\left(\widetilde{Q}_{t-1} \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right) \\
&\quad + H\left(\widetilde{Q}_{1:t-2} \mid \widetilde{Q}_{t-1}, \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right),
\end{aligned}
\tag{63}
$$

where

$$
\begin{aligned}
&H\left(\widetilde{Q}_{t-1} \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right) \\
&= -\sum_{i=1}^{\widetilde{N}} \left[P\left(\widetilde{Q}_{t-1} = \widetilde{s}_i \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right) \right. \\
&\quad \left. \cdot \log_2\left(P\left(\widetilde{Q}_{t-1} = \widetilde{s}_i \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right)\right) \right], \\
&H\left(\widetilde{Q}_{1:t-2} \mid \widetilde{Q}_{t-1}, \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right) \\
&= \sum_{i=1}^{\widetilde{N}} \left[P\left(\widetilde{Q}_{t-1} = \widetilde{s}_i \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right) \right. \\
&\quad \left. \cdot H\left(\widetilde{Q}_{1:t-2} \mid \widetilde{Q}_{t-1} = \widetilde{s}_i, \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right) \right] \\
&= \sum_{i=1}^{\widetilde{N}} \left[P\left(\widetilde{Q}_{t-1} = \widetilde{s}_i \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right) H_{t-1}\left(\widetilde{s}_i\right) \right].
\end{aligned}
\tag{64}
$$

The auxiliary probability $P(\widetilde{Q}_{t-1} = \widetilde{s}_i \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t})$ is required for our EOTFA algorithm. It can be computed as follows:

$$
\begin{aligned}
&P\left(\widetilde{Q}_{t-1} = \widetilde{s}_i \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right) = P\left(\widetilde{Q}_{t-1} = \widetilde{s}_i \mid \widetilde{Q}_t = \widetilde{s}_j, o_t, o_{1:t-1}\right) \\
&= \frac{P\left(\widetilde{Q}_t = \widetilde{s}_j, o_t \mid \widetilde{Q}_{t-1} = \widetilde{s}_i, o_{1:t-1}\right) P\left(\widetilde{Q}_{t-1} = \widetilde{s}_i \mid o_{1:t-1}\right)}{P\left(\widetilde{Q}_t = \widetilde{s}_j, o_t \mid o_{1:t-1}\right)} \\
&= \frac{P\left(o_t \mid \widetilde{Q}_t = \widetilde{s}_j\right) P\left(\widetilde{Q}_t = \widetilde{s}_j \mid \widetilde{Q}_{t-1} = \widetilde{s}_i\right) P\left(\widetilde{Q}_{t-1} = \widetilde{s}_i \mid o_{1:t-1}\right)}{P\left(o_t \mid \widetilde{Q}_t = \widetilde{s}_j\right) P\left(\widetilde{Q}_t = \widetilde{s}_j \mid o_{1:t-1}\right)}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{P\left(\widetilde{Q}_t = \widetilde{s}_j \mid q_{t-1} = \widetilde{s}_i\right) P\left(\widetilde{Q}_{t-1} = \widetilde{s}_i \mid o_{1:t-1}\right)}{\sum_{k=1}^{\widetilde{N}} P\left(\widetilde{Q}_t = \widetilde{s}_j \mid Q_{t-1} = \widetilde{s}_k\right) P\left(\widetilde{Q}_{t-1} = \widetilde{s}_k \mid o_{1:t-1}\right)} \\
&= \frac{\widetilde{a}_{ij} \overset{\bullet}{\widetilde{\alpha}}_{t-1}(i)}{\sum_{k=1}^{\widetilde{N}} \widetilde{a}_{kj} \overset{\bullet}{\widetilde{\alpha}}_{t-1}(k)}.
\end{aligned}
\tag{65}
$$

For the final process, we compute $H(q_{1:T} \mid o_{1:T})$ which can be expanded as follows:

$$
\begin{aligned}
H\left(\widetilde{Q}_{1:T} \mid o_{1:T}\right) &= H\left(\widetilde{Q}_{1:T-1} \mid \widetilde{Q}_T = \widetilde{s}_j, o_{1:T}\right) \\
&\quad + H\left(\widetilde{Q}_T \mid o_{1:T}\right) \\
&= \sum_{i=1}^{\widetilde{N}} H_T\left(\widetilde{s}_i\right) \overset{\bullet}{\widetilde{\alpha}}_T(i) \\
&\quad - \sum_{i=1}^{\widetilde{N}} \overset{\bullet}{\widetilde{\alpha}}_T(i) \log_2\left(\overset{\bullet}{\widetilde{\alpha}}_T(i)\right).
\end{aligned}
\tag{66}
$$

The basic entropy concept in (40) and the following basic properties of HMM are used for proving our Lemma 11. According to the transformation of a high-order into an equivalent first-order HMM, state $\widetilde{Q}_{t-r}$, $r \ge 2$, and $\widetilde{Q}_t$ are statistically independent given $\widetilde{Q}_{t-1}$. The same applies to $\widetilde{Q}_{t-r}$, $r \ge 2$ and o_t are statistically independent given $\widetilde{Q}_{t-1}$.

The following proof is due to Hernando et al. [4].

Lemma 11. *For the transformation of a high-order into an equivalent first-order HMM, the entropy of the state sequence up to time $t-2$, given the states at time $t-1$ and the observations up to time $t - 1$, is conditionally independent on the state and observation at time t*

$$H_{t-1}\left(\widetilde{s}_i\right) = H\left(\widetilde{Q}_{1:t-2} \mid \widetilde{Q}_{t-1} = \widetilde{s}_i, \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right). \tag{67}$$

Proof.

$$
\begin{aligned}
&H\left(\widetilde{Q}_{1:t-2} \mid \widetilde{Q}_{t-1} = \widetilde{s}_i, \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right) \\
&= H\left(\widetilde{Q}_{1:t-2} \mid \widetilde{Q}_{t-1} = \widetilde{s}_i, o_{1:t-1}, \widetilde{Q}_t = \widetilde{s}_j, o_t\right) \\
&= H\left(\widetilde{Q}_{1:t-2} \mid \widetilde{Q}_{t-1} = \widetilde{s}_i, o_{1:t-1}\right) = H_{t-1}\left(\widetilde{s}_i\right).
\end{aligned}
\tag{68}
$$

Our EOTFA algorithm for computing the optimal state sequence is based on the normalized forward recursion variable, state entropy recursion variable, and auxiliary probability. From (59), (60), (61), (62), (63), and (66), we construct our EOTFA algorithm as follows.

(1) Initialization. For $t = 1$ and $1 \le j \le \widetilde{N}$,

$$H_1\left(\widetilde{s}_j\right) = 0,$$

$$\overset{\bullet}{\widetilde{\alpha}}_1(j) = \frac{\widetilde{\pi}(j) \widetilde{b}_j(o_1)}{\sum_{i=1}^{\widetilde{N}} \widetilde{\pi}(i) \widetilde{b}_i(o_1)}. \tag{69}$$

(2) Recursion. For $t = 2, \ldots, T$ and $1 \leq j \leq \widetilde{N}$,

$$\overset{\bullet}{\widetilde{\alpha}}_t(j) = \frac{\sum_{i=1}^{\widetilde{N}} \overset{\bullet}{\widetilde{\alpha}}_{t-1}(i)\, \widetilde{a}_{ij} \widetilde{b}_j(o_t)}{\sum_{k=1}^{\widetilde{N}} \sum_{i=1}^{\widetilde{N}} \overset{\bullet}{\widetilde{\alpha}}_{t-1}(i)\, \widetilde{a}_{ik} \widetilde{b}_k(o_t)},$$

$$P\left(\widetilde{Q}_{t-1} = \widetilde{s}_i \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right) = \frac{\widetilde{a}_{ij} \overset{\bullet}{\widetilde{\alpha}}_{t-1}(i)}{\sum_{k=1}^{\widetilde{N}} \widetilde{a}_{kj} \overset{\bullet}{\widetilde{\alpha}}_{t-1}(k)},$$

$$H_t\left(\widetilde{s}_j\right) = \sum_{i=1}^{\widetilde{N}} H_{t-1}\left(\widetilde{s}_i\right) P\left(\widetilde{Q}_{t-1} = \widetilde{s}_i \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right)$$

$$\tag{70}$$

$$- \sum_{i=1}^{\widetilde{N}} P\left(\widetilde{Q}_{t-1} = \widetilde{s}_i \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right)$$

$$\cdot \log_2\left(P\left(\widetilde{Q}_{t-1} = \widetilde{s}_i \mid \widetilde{Q}_t = \widetilde{s}_j, o_{1:t}\right)\right).$$

(3) Termination

$$H\left(\widetilde{Q}_{1:T} \mid o_{1:T}\right) = \sum_{i=1}^{\widetilde{N}} H_T\left(\widetilde{s}_i\right) \overset{\bullet}{\widetilde{\alpha}}_T(i)$$

$$\tag{71}$$

$$- \sum_{i=1}^{\widetilde{N}} \overset{\bullet}{\widetilde{\alpha}}_T(i) \log_2\left(\overset{\bullet}{\widetilde{\alpha}}_T(i)\right).$$

The direct evaluation algorithm, Hernando et al.'s algorithm, and our algorithm perform the computation of state entropy exponentially with respect to the order of HMM. Our algorithm proposes the transformation of a generalized high-order into an equivalent first-order HMM and then compute the state entropy based on the equivalent first-order model; hence our algorithm is the most efficient in which it requires $O(T\widetilde{N}^2)$ calculations as compared to the direct evaluation method which requires $O(N^{T+k-1})$ calculations and the extended algorithm which requires $O(TN^{k+1})$ calculations where N is the number of states in a model, $\widetilde{N}$ is the number of states in an equivalent first-order model, T is the length of observational sequence, and k is the order of HMM. □

3.4. Numerical Illustration for an Equivalent First-Order HMM. We consider the second-order HMM in Section 2.5 for illustrating our EOTFA algorithm in computing the optimal state sequence. According to our proposed novel algorithm, we first transformed the second-order HMM in Section 2.5 into the equivalent first-order HMM by using Hadar and Messer method [7]. The equivalent first-order HMM has the following model parameters $\widetilde{\lambda} = (\widetilde{\pi}, \widetilde{A}, \widetilde{B})$, where $\widetilde{\pi}$ is the initial state probability vector, $\widetilde{A}$ is the state transition probability matrix, and $\widetilde{B}$ is the emission probability matrix.

$$\widetilde{\pi} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \end{bmatrix},$$

$$\widetilde{A} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\widetilde{B} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0.5 & 0 & 1 & 0 \end{bmatrix}.$$

$$\tag{72}$$

Note that the above state transition probability and the emission probability matrices whose components are indicated as $\widetilde{a}_{i_1 i_2}$ and $\widetilde{b}_{i_2}(o_t = v_m)$ where $1 \leq i_1, i_2 \leq 4$ and $1 \leq m \leq 3$ can be obtained from the graphical diagram in Figure 4.

The state space for the equivalent first-order HMM is $\widetilde{S} = \{\widetilde{s}_1, \widetilde{s}_2, \widetilde{s}_3, \widetilde{s}_4\}$, where $\widetilde{s}_1 = [s_1, s_1]$, $\widetilde{s}_2 = [s_1, s_2]$, $\widetilde{s}_3 = [s_2, s_1]$, and $\widetilde{s}_4 = [s_2, s_2]$, and the possible symbols per observation are $O = \{v_1, v_2, v_3\}$. Note that $\widetilde{\pi}_1 = \{\widetilde{\pi}_{i_2}\}$, where $\widetilde{\pi}_{i_2} = P(\widetilde{Q}_t = \widetilde{s}_{i_2})$, $\widetilde{A} = \{\widetilde{a}_{i_1 i_2}\}$, where $\widetilde{a}_{i_1 i_2} = P(\widetilde{Q}_t = \widetilde{s}_{i_2} \mid \widetilde{Q}_{t-1} = \widetilde{s}_{i_1})$, and $\widetilde{B} = \{\widetilde{b}_{i_2}(o_t = v_m)\}$, where $\widetilde{b}_{i_2}(o_t = v_m) = P(o_t = v_m \mid \widetilde{Q}_t = \widetilde{s}_{i_2})$.

The equivalent first-order HMM was developed based on Hadar and Messer's method [7] is shown in Figure 4.

Secondly, the optimal state sequence is computed based on the equivalent first-order HMM by using our proposed algorithm. Finally, the optimal state sequence of the second-order HMM is inferred from the optimal state sequence from the equivalent first-order HMM.

The following is the observational sequence used for illustrating our algorithm:

$$o_{1:6} = (o_1 = v_1, o_2 = v_1, o_3 = v_3, o_4 = v_2, o_5 = v_3, o_6$$

$$= v_1).$$

$$\tag{73}$$

We applied our EOFTA algorithm for computing the optimal state sequence based on the state entropy. The computed value of state entropy is shown in Figure 5.

The total entropy after each time step for the transformed model, that is, the second-order transformed into the equivalent first-order HMM is displayed at the bottom of Figure 5. For example, this model has produced only one possible state sequence; that is, $\widetilde{Q}_{1:5} = (\widetilde{Q}_1 = \widetilde{s}_1, \widetilde{Q}_2 = \widetilde{s}_2, \widetilde{Q}_3 = \widetilde{s}_3, \widetilde{Q}_4 = \widetilde{s}_2, \widetilde{Q}_5 = \widetilde{s}_3)$, as shown by the bold arrow with a probability of 1 after receiving the fifth observation. The total entropy is 0 at $t = 5$ which indicates that there is no uncertainty. After receiving the sixth observation, that is, $o_{1:6} = (o_1 = v_1, o_2 = v_1, o_3 = v_3, o_4 = v_2, o_5 = v_3, o_6 = v_1)$, this equivalent first-order HMM has produced one possible optimal state sequence $\widetilde{Q}_{1:6} = (\widetilde{Q}_1 = \widetilde{s}_1, \widetilde{Q}_2 = \widetilde{s}_2, \widetilde{Q}_3 = \widetilde{s}_3, \widetilde{Q}_4 = \widetilde{s}_2, \widetilde{Q}_5 = \widetilde{s}_3, \widetilde{Q}_6 = \widetilde{s}_2)$ which is similar to $q_{1:6} = (q_1 = s_1, q_2 = s_2, q_3 = s_1, q_4 = s_2, q_5 = s_1, q_6 = s_2)$ that is produced by the second-order HMM in Section 2.5 with a total entropy of 0 which indicates that there is no uncertainty. As a result, the optimal state sequence of the high-order HMM is inferred from the optimal state sequence of the equivalent first-order HMM.
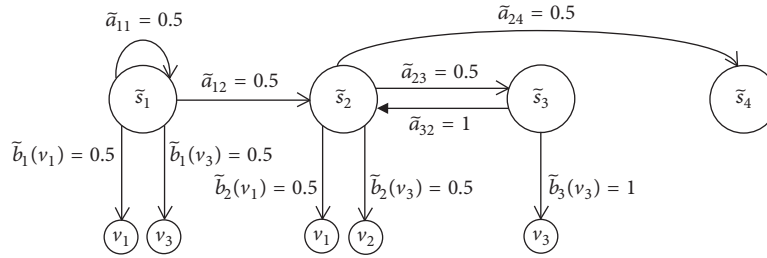
FIGURE 4: The graphical diagram shows an equivalent first-order HMM.

state \ obs	$o_1 = v_1$		$o_2 = v_1$		$o_3 = v_3$		$o_4 = v_2$		$o_5 = v_3$		$o_6 = v_1$
$\widetilde{s}_1$	$H_1(1) = 0$ $\widehat{\alpha}_1^{\bullet}(1) = 0.5$		$H_2(1) = 0$ $\widehat{\alpha}_2^{\bullet}(1) = 0.5$		$H_3(1) = 0.5$ $\widehat{\alpha}_3^{\bullet}(1) = 0.33$		$H_4(1) = 0$ $\widehat{\alpha}_4^{\bullet}(1) = 0$		$H_5(1) = 0$ $\widehat{\alpha}_5^{\bullet}(1) = 0$		$H_6(1) = 0$ $\widehat{\alpha}_6^{\bullet}(1) = 0$
$\widetilde{s}_2$	$H_1(2) = 0$ $\widehat{\alpha}_1^{\bullet}(2) = 0.5$		$H_2(2) = 0$ $\widehat{\alpha}_2^{\bullet}(2) = 0.5$		$H_3(2) = 0$ $\widehat{\alpha}_3^{\bullet}(2) = 0$		$H_4(2) = 0$ $\widehat{\alpha}_4^{\bullet}(2) = 1$		$H_5(2) = 0$ $\widehat{\alpha}_5^{\bullet}(2) = 0$		$H_6(2) = 0$ $\widehat{\alpha}_6^{\bullet}(2) = 1$
$\widetilde{s}_3$	$H_1(3) = 0$ $\widehat{\alpha}_1^{\bullet}(3) = 0$		$H_2(3) = 0$ $\widehat{\alpha}_2^{\bullet}(3) = 0$		$H_3(3) = 0.5$ $\widehat{\alpha}_3^{\bullet}(3) = 0.67$		$H_4(3) = 0$ $\widehat{\alpha}_4^{\bullet}(3) = 0$		$H_5(3) = 0$ $\widehat{\alpha}_5^{\bullet}(3) = 1$		$H_6(3) = 0$ $\widehat{\alpha}_6^{\bullet}(3) = 0$
$\widetilde{s}_4$	$H_1(4) = 0$ $\widehat{\alpha}_1^{\bullet}(4) = 0$		$H_2(4) = 0$ $\widehat{\alpha}_2^{\bullet}(4) = 0$		$H_3(4) = 0$ $\widehat{\alpha}_3^{\bullet}(4) = 0$		$H_4(4) = 0$ $\widehat{\alpha}_4^{\bullet}(4) = 0$		$H_5(4) = 0$ $\widehat{\alpha}_5^{\bullet}(4) = 0$		$H_6(4) = 0$ $\widehat{\alpha}_6^{\bullet}(4) = 0$
Total entropy	1		1		1.41		0		0		0

FIGURE 5: The evolution of the trellis structure for a transformation of a second-order into an equivalent first-order HMM with the observation sequence $o_{1:6} = (o_1 = v_1, o_2 = v_1, o_3 = v_3, o_4 = v_2, o_5 = v_3, o_6 = v_1)$.

Our proposed algorithm is based on the equivalent first-order HMM and only requires $O(T\widetilde{N}^2)$ computation and hence we can conclude that our EOTFA algorithm is more efficient.

4. Conclusion and Future Work

We have introduced a novel algorithm for computing the optimal state sequence for HHMM that requires $O(T\widetilde{N}^2)$ calculations and $O(\widetilde{N}^2)$ memory space where $\widetilde{N}$ is the number of states in an equivalent first-order HMM and T is the length of observational sequence. This algorithm is to be running with Viterbi algorithm in tracking the optimal state sequence as well as the entropy of the distribution of the state sequence. We have developed this algorithm for the case of a generalized discrete high-order HMM. This research can be also extended for continuous high-order HMMs and these models are widely used in speech recognition.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] V. Ciriza, L. Donini, J. Durand, and S. Girard, "Optimal time-outs for power management under renewal or hidden Markov processes for requests," Tech. Rep., 2011, http://hal.inria.fr/hal-00412509/en.

[2] L. R. Rabiner, "Tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[3] J. G. Proakis and M. Salehi, *Communications System Engineering*, Prentice-Hall, Upper Saddler River, NJ, USA, 2002.

[4] D. Hernando, V. Crespi, and G. Cybenko, "Efficient computation of the hidden Markov model entropy for a given observation sequence," *Institute of Electrical and Electronics*

Engineers Transactions on Information Theory, vol. 51, no. 7, pp. 2681–2685, 2005.

[5] G. S. Mann and A. McCallum, "Efficient computation of entropy gradient for semi-supervised conditional random fields," in *Proceedings of the Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics (NAACL '07)*, pp. 109–112, Association for Computational Linguistics, Morristown, NJ, USA, April 2007.

[6] V. M. Ilic, "Entropy Semiring Forward-backward Algorithm for HMM Entropy Computation," 2011, https://arxiv.org/abs/1108.0347.

[7] U. Hadar and H. Messer, "High-order hidden Markov models—estimation and implementation," in *Proceedings of the 15th IEEE/SP Workshop on Statistical Signal Processing (SSP '09)*, pp. 249–252, Wales, UK, September 2009.

[8] J. A. du Preez, "Efficient training of high-order hidden Markov models using first-order representations," *Computer Speech and Language*, vol. 12, no. 1, pp. 23–39, 1998.

[9] L. M. Lee and J. C. Lee, "A study on high-order hidden Markov Models and applications to speech recognition," in *Proceedings of the 19th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 682–690, 2006.

[10] R. M. Altman, "Mixed hidden MARkov models: an extension of the hidden MARkov model to the longitudinal data setting," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 201–210, 2007.

[11] A. Spagnoli, R. Henderson, R. J. Boys, and J. J. Houwing-Duistermaat, "A hidden Markov model for informative dropout in longitudinal response data with crisis states," *Statistics & Probability Letters*, vol. 81, no. 7, pp. 730–738, 2011.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications and Signal Processing, John Wiley & Sons, New York, NY, USA, 2006.

A Simple Empirical Likelihood Ratio Test for Normality Based on the Moment Constraints of a Half-Normal Distribution

C. S. Marange [iD] **and Y. Qin**

Department of Statistics, Faculty of Science and Agriculture, Fort Hare University, East London Campus, 5200, South Africa

Correspondence should be addressed to C. S. Marange; cmarange@ufh.ac.za

A simple and efficient empirical likelihood ratio (ELR) test for normality based on moment constraints of the half-normal distribution was developed. The proposed test can also be easily modified to test for departures from half-normality and is relatively simple to implement in various statistical packages with no ordering of observations required. Using Monte Carlo simulations, our test proved to be superior to other well-known existing goodness-of-fit (GoF) tests considered under symmetric alternative distributions for small to moderate sample sizes. A real data example revealed the robustness and applicability of the proposed test as well as its superiority in power over other common existing tests studied.

1. Introduction

Testing for distributional assumptions for normality is of paramount importance in applied statistical modelling. Several well-known numerical tests for normality are widely used by investigators to supplement the graphical techniques in assessing departures from normality. Amongst others, these tests include the Kolmogorov-Smirnov (KS) test [1], the Lilliefors (LL) test [2], the Anderson-Darling (AD) test [3, 4], the Shapiro-Wilks (SW) test [5], the Jarque-Bera (JB) test [6], and the DAgostino and Pearson (DP) test [7]. These tests differ on certain characteristics of the normal distribution on which they focus. That is, some focus on the empirical distribution function (EDF), some are moment based, and some are based on regression as well as correlation. Of these tests, some use normalized sample data whilst some use observed values. However, though these tests are commonly used in practice they do have major drawbacks. For example, some of these tests require complete specification of the null distribution, some require computation of critical values to be done for each specified null distribution, and some require ordering of the sample data when computing the test statistic. Generally, most of these tests are not supported when certain combinations of parameters of a specified distribution are estimated.

Of these, the most well-known goodness-of-fit (GoF) test is the SW test but it was originally restricted to small sample sizes (i.e., $n \leq 50$). Several modifications have been proposed by several researchers. These include Royston [8] who suggested a normalized transformation for the test in order to resolve the limitations on the sample size, Shapiro and Francia [9] who also modified the test so that it can be ideal for large sample sizes, Chen and Shapiro [10] who proposed normalized spacings for an alternative test of the SW test, and Rahman and Govindarajulu [11] who defined new weights for the SW test statistic. However, the major drawback of the SW test is computation time in dealing with large samples when computing the covariance matrix that corresponds to order statistics of the vector of weights and the standard normal distribution.

However, we also have GoF tests that are based on moment constraints such as the skewness and kurtosis coefficients and these are well known to be efficient tools for evaluating normality. These moment based tests include the skewness test, the kurtosis test, the DP test, and the JB test. These tests combine moment constraints to check for deviations from normality. They are often referred to as omnibus tests because of their ability to detect departures from normality whilst not depending upon the parameters of the normal distribution. The adoption of the use of moment

based tests coupled with the empirical likelihood methodology has recently attracted the attention of researchers in developing GoF tests for normality [12, 13]. Dong and Giles [12] proposed an empirical likelihood ratio (ELR) test utilizing the empirical likelihood (EL) methodology of Owen [14]. They monitored the first four moment conditions of the normal distribution and their test outperformed alternate common existing tests studied against several alternative distributions. Our study followed from the works of Shan et al. [13] who proposed a simple ELR test for normality based on moment constraints using a standardized normal variable. Their test proved to be more powerful than other well-known GoF tests on small to moderate sample sizes for several alternative distributions. In this study we adopted their approach and focused on the construction of a simple ELR test for normality using the moment constraints of the half-normal distribution. The next section will outline the development of our proposed test followed by Monte Carlo simulations. A real data example will be presented. Discussions and conclusion of the findings as well as potential areas of future research will be highlighted.

2. ELR Test Development

Let us assume we have independent and identically distributed (*i.i.d*) nonordered random variables $X_1, X_2, \ldots, X_n$. The intention being to assess whether the observed data is normally distributed. Thus we intend testing the following null hypothesis:

$$H_0 : X \sim N\left(\mu, \sigma^2\right), \tag{1}$$

where μ and σ^2 are considered to be unknown parameters. We proposed using the standardized random variables of the normal distribution by using the following transformations:

$$Z_i^\star = \frac{X_i - \mu}{SD}, \quad i = 1, 2, \ldots, n, \tag{2}$$

where $\mu = \overline{X} = (1/n)\sum_{i=1}^n X_i$ and SD is the standard deviation to be estimated by an unbiased quantity $s^2 = S/(n-1)$. One can also decide to use the maximum likelihood estimate (MLE) $\widehat{\sigma}^2 = S/n$, where $S = \sum_{i=1}^n (X_i - \overline{X})^2$ and $\overline{X} = (1/n)\sum_{i=1}^n X_i$. Both quantities s^2 and $\widehat{\sigma}^2$ are known to converge to σ^2 as n approaches ∞. We also used an alternative transformation following Lin and Mudholkar's [17] work which also eliminates the dependency that exists between μ and σ on the data distribution. Thus we also transformed our observations using

$$Z_i^* = \frac{\sqrt{n/(n-1)}\left(X_i - \overline{X}\right)}{SD_{-i}}, \quad i = 1, 2, \ldots, n, \tag{3}$$

where $\overline{X} = (1/n)\sum_{j=1}^n X_j$, $SD_{-i}^2 = (1/(n-2))\sum_{j=1, j\neq i}^n (X_j - \overline{X}_{-i})^2$, and $\overline{X}_{-i} = (1/(n-1))\sum_{j=1, j\neq i}^n X_j$. As n gets large the standardized data points $Z_1, Z_2, \ldots, Z_n$ become asymptotically independent. If $X \sim N(0, \sigma^2)$, then the absolute value $|X| \sim HN(\mu, \sigma^2)$. It also follows that if $X \sim N(\mu, \sigma^2)$, then the

modulus of the standardized normal random variables, $Z^\star$ and Z^*, follows a standardized half-normal random variable with mean $= \sqrt{2/\pi}$ and variance $= 1$. The standardized form of the half-normal distribution is also known as the χ^2-distribution with $\nu = 1$. The standardized half-normal random variable has a PDF that is given by

$$f_Z(z) = \begin{cases} \dfrac{2}{\sqrt{2\pi}} e^{-(1/2)z^2} & \text{for } z \geq 0, \\ 0 & \text{for } z < 0. \end{cases} \tag{4}$$

and we denote it as $Z \sim HN(\mu, \sigma^2)$. Following Prudnikov et al. [18], the kth moment of the standardized half-normal variable for some integer $k > 0$ is as outlined in the proposition below.

Proposition 1. *Let* $Z \sim HN(\sqrt{2/\pi}, 1)$, *for* $k = 1, 2, \ldots, n$, *and then the* k^{th} *moments are given by*

$$E\left(Z^k\right) = \mu_k = \frac{1}{\sqrt{\pi}} 2^{k/2} \Gamma\left(\frac{k+1}{2}\right), \tag{5}$$

where $\Gamma(\cdot)$ *denotes the gamma function.*

We then derived the first four moments using the function given in (5). These moments are easily obtained as follows.

Corollary 2. *Let* $Z \sim HN(\mu, \sigma^2)$. *The first two moments of Z, that is* μ *and* σ *are given by*

$$E(Z) = \mu = \sqrt{\frac{2}{\pi}} \Gamma(1) = \sqrt{\frac{2}{\pi}} \approx 0.7979, \tag{6}$$

$$\text{var}(Z) = \frac{2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) = 1. \tag{7}$$

Corollary 3. *Let* $Z \sim HN(\mu, \sigma^2)$. *The skewness and kurtosis coefficients of Z are given by*

$$\gamma(Z_1) = E\left(Z^3\right) = \mu_3 = \sqrt{\frac{2^3}{\pi}} \Gamma(2) = 2\sqrt{\frac{2}{\pi}} \approx 1.5958, \tag{8}$$

$$\gamma(Z_2) = E\left(Z^4\right) = \mu_4 = \frac{4}{\sqrt{\pi}} \Gamma\left(\frac{5}{2}\right) = 3. \tag{9}$$

In this study we used the first four moment constraints of the standardized half-normal distribution.

2.1. The ELR Based Test Statistic. We used an empirical likelihood ratio test (ELR) to construct our test statistic. Our aim was to compare the GoF test under H_0 against the alternative (H_a). In order to achieve this, we constructed our test statistic as follows. Let us consider n nonordered observations $X_1, X_2, \ldots, X_n$ that are independent and identically distributed and assumed to have unknown μ and σ. The intention is to perform a GoF test for the distributional assumption that $X_1, X_2, \ldots, X_n$ are consistent with a normal distribution. Now consider that the random variables $Z_1, Z_2, \ldots, Z_n$ are absolute standardized normal variables from the random

variables $X_1, X_2, \ldots, X_n$. Thus the transformed/standardized observations have a moment function given in Proposition 1 above. Following the EL methodology we assigned p_i, which is a probability parameter to each transformed observation Z_i, and then formulated the EL function that is given by

$$L(F) = \prod_{i=1}^{n} p_i, \qquad (10)$$

where p_i's satisfy the fundamental properties of probability; that is $0 \leq p_i \leq 1$ and $\sum_{i=1}^{n} p_i = 1$. Probability parameters, p_i's, will then be chosen subject to unbiased moment conditions and the EL method will utilize these p_i's in order to maximize the EL function. Following this EL technique, $E(Z^k)$ has sample moments $\sum_{i=1}^{n} p_i Z_i^k$ and the probability parameters (p_i's) are elements of the EL function. Under H_0, the four unbiased empirical moment equations have the form

$$\sum_{i=1}^{n} p_i Z_i^k - \mu_k = 0, \quad k = 1, 2, \ldots, n. \qquad (11)$$

The composite hypotheses for the ELR test are given by

$$H_0 : z_i's \sim HN\left(\mu, \sigma^2\right)$$
$$\text{vs } H_a : z_i's \nsim HN\left(\mu, \sigma^2\right). \qquad (12)$$

Alternatively considering the above unbiased empirical moment equations, the hypotheses for the ELR test can be written as

$$H_0 : E\left(Z^k\right) = \mu_k$$
$$\text{vs } H_a : E\left(Z^k\right) \neq \mu_k, \qquad (13)$$

The nonparametric empirical likelihood function corresponding to the given hypotheses has the form:

$$L(F) = L\left(Z_1, Z_2, \ldots, Z_n \mid \mu_k\right) = \prod_{i=1}^{n} p_i, \qquad (14)$$

where the unknown probability parameters and p_i's are attained under H_0 and H_a. Under H_0 the EL function is maximized with respect to the p_i's subject to two constraints

$$\sum_{i=1}^{n} p_i = 1,$$
$$\sum_{i=1}^{n} p_i Z_i^k = \mu_k. \qquad (15)$$

Following this, the weights of p_i's are identified as

$$p_1, p_2, \ldots, p_n = \sup \prod_{i=1}^{n} a_i \mid \quad \sum_{i=1}^{n} a_i = 1,$$
$$\sum_{i=1}^{n} a_i Z_i^k = \mu_k, \qquad (16)$$

where $0 \leq a_j \leq 1$, for $j = \{1, 2, \ldots, n\}$. If we then use the Lagrangian multipliers technique, it can be shown that the maximum EL function under H_0 can be expressed by the given form:

$$L\left(F_{H_0}\right) = L\left(Z_1, Z_2, \ldots, Z_n \mid \mu_k\right)$$
$$= \prod_{i=1}^{n} \frac{1}{n\left(1 + \lambda_k\left(Z_i^k - \mu_k\right)\right)}, \qquad (17)$$

where λ_k is a root of

$$\sum_{i=1}^{n} \frac{\left(Z_i^k - \mu_k\right)}{1 + \lambda_k\left(Z_i^k - \mu_k\right)} = 0. \qquad (18)$$

Under the alternative hypothesis, $\sum_{i=1}^{n} p_i Z_i^k = \mu_k$ is not required to identify the weights, p_i, in order to maximize the EL function but only $\sum_{i=1}^{n} p_i = 1$. Thus under H_a the nonparametric EL function is given by

$$L\left(F_{H_a}\right) = L\left(Z_1, Z_2, \ldots, Z_n\right) = \prod_{i=1}^{n}\left(\frac{1}{n}\right) = \left(\frac{1}{n}\right)^n. \qquad (19)$$

Now let us consider $(-2LLR)_k$ to be -2 log likelihood test statistic for the hypotheses $H_0 : E(Z^k) = \mu_k$ vs $H_a : E(Z^k) \neq \mu_k$. It should be noted that, under H_0, minus two times the log ELR has an asmymptotic χ^2 limiting distribution [19]. Thus considering the null and alternative hypotheses, the above test statistic will simply be transformed to

$$(-2LLR)_k = -2 \log \frac{L\left(F_{H_0}\right)}{L\left(F_{H_a}\right)}$$
$$= -2 \log \frac{L\left(Z_1, Z_2, \ldots, Z_n \mid \mu_k\right)}{L\left(Z_1, Z_2, \ldots, Z_n\right)}. \qquad (20)$$

With simple substitution the above can be simplified to

$$(-2LLR)_k = -2 \log \frac{\prod_{i=1}^{n}\left(1/n\left(1 + \lambda_k\left(Z_i^k - \mu_k\right)\right)\right)}{\prod_{i=1}^{n}\left(1/n\right)}$$
$$= 2 \sum_{i=1}^{n} \log\left[1 + \lambda_k\left(Z_i^k - \mu_k\right)\right]. \qquad (21)$$

We used the likelihood ratio to compare to size adjusted critical values in order to decide whether or not to reject H_0. We then proposed to reject the null hypothesis if

$$\text{ELR}_Z = \max_{k \in G} (-2LLR)_k > C_\alpha, \qquad (22)$$

where C_α is the test threshold and is $100(1-\alpha)\%$ percentile of the $\chi^2(1)$ distribution whilst G are integer values representing the set of moment constraints that maximizes the test statistic. As recommended by Dong and Giles [12], we used the first four moment constraints; that is, we set $G = \{1, 2, 3, 4\}$. In this study we used the abbreviation **ELR$_{Z1}$** to refer to the

first test where we transformed data using (2) and we used the abbreviation **ELR**$_{Z2}$ to refer to the second alternative test where we transformed data using (3). Our test statistic ELR$_Z$ = max$_{k \in G}(-2LLR)_k$ is a CUSUM-type statistic as classified by Vexler and Wu [20]. In their article, Vexler and Wu [20] stated that based on the change point literature, another common alternative is to utilize the Shiryaev-Roberts (SR) statistic in replacement of the CUSUM-type statistic (see, for example, [21, 22]). In our case the classical SR statistic was of the form $\sum_{k \in G} \exp(-2LLR)_k$. Vexler, Liu, and Pollak [23] showed that the classical SR statistic and the simple CUSUM-type statistic have almost equivalent optimal statistical properties due to their common null-martingale basis. Moreover, the classical SR statistic is adapted from the CUSUM-type statistic.

Shan et al. [13] used Monte Carlo experiments to compare the CUSUM-type statistic for their ELR test for normality with an equivalent classical SR statistic and based on the relative simplicity of the CUSUM-type statistic, as well as its power properties, the authors opted to use the CUSUM-type statistic for their study. We conducted a numerical experiment to compare power for the CUSUM-type and SR statistic for our proposed test statistics with increased moment constraints and, based on the same reasons given by Shan et al. [13], we decided to use the CUSUM-type statistic for our Monte Carlo comparisons. Also, from the results, ELR$_{Z2}$ outperformed ELR$_{Z1}$, hence ELR$_{Z2}$ was our preferred test. For all further comparisons, ELR$_{Z1}$ was excluded in this study. Findings for this Monte Carlo experiment are presented in Table 4. However, it should be noted from these findings that ELR$_{Z1}$ has the potential to be superior to ELR$_{Z2}$ under certain alternatives. Further investigations to uncover the alternatives in which ELR$_{Z1}$ is superior to ELR$_{Z2}$ are a potential area of future research which will not be further addressed in this study. The next section will outline the Monte Carlo simulation procedures using the R statistical package.

3. Monte Carlo Simulation Study

We used the R statistical package to implement our Monte Carlo simulation procedures in power comparisons as well as assessment of our preferred proposed test (ELR$_{Z2}$). It should be noted that other standard statistical packages can easily be used to implement our proposed tests. In order for us to conduct any assessments and evaluations of the proposed test, firstly we had to determine the size adjusted critical values.

3.1. Size Adjusted Critical Values. Since the proposed ELR test is an asymptotic test, we therefore computed the unknown actual sizes for finite samples using Monte Carlo simulations with 50,000 replications. Motivated by practical applications, we considered critical values for relatively small sample sizes, i.e., $10 \leq n \leq 200$ because most applied statistical sciences datasets fall within this range. The actual rejection rate for a given sample size (n) is considered to be the total number of the rejections divided by the total number of replications. Data was simulated from a standard normal distribution. The stored ordered test statistics were then used to determine

the percentiles of the empirical distribution. This makes it possible to obtain the 30%, 25%, 20%, . . . , 1%, size adjusted critical values.

3.2. ELR Test Assessment. The power of the proposed test (ELR$_{Z2}$) was compared to that of common existing GoF tests that include the Anderson-Darling (AD) test [3, 4] test, the modified Kolmogorov-Smirnov (KS) test [2] the Cramer-von Mises (CVM) test [24–26], the Jarque-Bera (JB) test [6], the Shapiro-Wilk (SW) test [5], the density based empirical likelihood ratio based (DB) test [16], and the simple and exact empirical likelihood test based on moment relations (SEELR) [13] at the 5% significance level. Power simulations were done using 5,000 replications for all tests with varying sample sizes ($n = 20, 30, 50$ and 80) against different alternative distributions. We adopted alternative distributions used by Shan et al. [13] which covers a wide range of both symmetric and asymmetric applied distributions. To assess robustness and applicability of our proposed test (ELR$_{Z2}$), we conducted a bootstrap study using some real data.

4. Results of the Monte Carlo Simulations

This section presents the findings of the power comparisons for the different categories of the alternative distributions considered. The results of the power comparisons are presented in Tables 5–8. Under symmetric cases defined on $(-\infty, \infty)$ our new test ELR$_{Z2}$ outperformed all other studied tests against the considered alternative distributions but slightly inferior to the JB test. For symmetric distributions defined on $(0, 1)$ our proposed test (ELR$_{Z2}$) was comparable to the DB test and significantly outperformed other alternate tests studied. However, when the alternative is Beta (0.5, 0.5), the ELR$_{Z2}$ test is comparable to the SW and SEELR tests whilst only outperforming the KS test, the CVM test and the JB test.

As for asymmetric distributions defined on $(0, \infty)$, the SW and SEELR are the most powerful tests and should be the preferred tests under these cases. The AD and DB tests are comparable and they performed better than the proposed test as well as the KS and CVM tests. Lastly, in the category of asymmetric alternative distributions defined on $(-\infty, \infty)$ the ELR$_{Z2}$ test was comparable to the SEELR test at low sample sizes (i.e., $n = 20, 30$) for the non-central t-distributions. The SW test outperformed all the tests considered in this study under these asymmetric alternative distributions. For the ELR based tests only the SEELR test was comparable to the common existing tests studied, that is, the AD test, the KS$_M$ test, the CVM test, and the JB test.

Overall, when considering all the normality tests with respect to all of the alternative distributions considered, it can be seen that, the JB, the ELR$_{Z2}$ and the SW tests are generally the most powerful tests given symmetric alternatives defined on $(-\infty, \infty)$, whilst the DB and the ELR$_{Z2}$ tests are the most powerful tests for symmetric alternatives defined on $(0, 1)$. On the other hand, the SEELR and the SW tests are the most powerful tests for asymmetric alternatives defined on $(0, \infty)$, whereas, the JB and SW tests are the most powerful tests for asymmetric alternatives defined on $(-\infty, \infty)$.

TABLE 1: Comparisons of computational times (in seconds) for the studied tests.

Test	Replications	Elapsed	Relative	User.self	Sys.self
AD	5,000	1.14	2.000	1.14	0.00
CVM	5,000	0.82	1.439	0.81	0.00
DB	5,000	17.00	29.825	16.94	0.04
ELR_{Z2}	5,000	44.83	78.649	44.78	0.01
JB	5,000	252.64	443.228	252.60	0.00
KS_M	5,000	0.89	1.561	0.90	0.00
SEELR	5,000	45.42	79.684	45.41	0.00
SW	5,000	0.57	1.000	0.58	0.00

TABLE 2: The baby boom data.

Times between births (in min)															
59	14	37	62	68	2	15	9	157	27	37	2	55	86	14
4	40	36	47	9	61	1	26	13	28	77	26	45	25	18
29	15	38	2	2	19	27	14	13	19	54	70	28		

Note. Data appeared in the newspaper the Sunday Mail on December 21, 1997 [15].

It was of paramount importance for us to determine the computational cost of the new algorithms by focusing on the computation time of the proposed test as compared to that of the considered existing tests. To assess this, we used the R benchmark tool on a notebook installed with 64 Bit Windows 10 Home addition. Equipped with a 4th generation Intel Core i5-4210U processor which has a speed of 1.7 GHz cache and memory (RAM) of 4 GB PC3 DDR3L SDRAM, we set our simulations to 5,000 for each test with sample size set at $n = 80$. The results (see Table 1) show only a clear advantage of our proposed approach to that of the widely known JB test. Also from the results, our proposed methods are comparable to the SEELR test but inferior to the DB test. The SW, CVM, KS and AD tests are computationally more efficient in terms of time than the rest of the studied tests.

5. A Real Data Example

In this example we used baby boom data from an observational study with records of forty-four (44) babies born at a 24-hour hospital in Brisbane, Australia. We opted for this dataset because it can be used to demonstrate applicability of various statistical procedures to some common applied distributions which include the normal (by modelling the birth weights), the binomial (inferences in the number of boys/girls born), the geometric (by considering the number of births until a boy/girl is born), the Poisson (births per hour for each hour), and the exponential (inference on times between births). Recently, Miecznikowski et al. [27] used the baby boom dataset in a resampling study on the application of their ELR based goodness-of-fit test. For more information regarding this dataset one can refer to Dunn [28]. For our application we opted to make use of the exponential distribution; thus we were interested in inference on the times between births. Table 2 shows the times between births which were computed by taking the differences between successive times of birth after midnight of birth times.
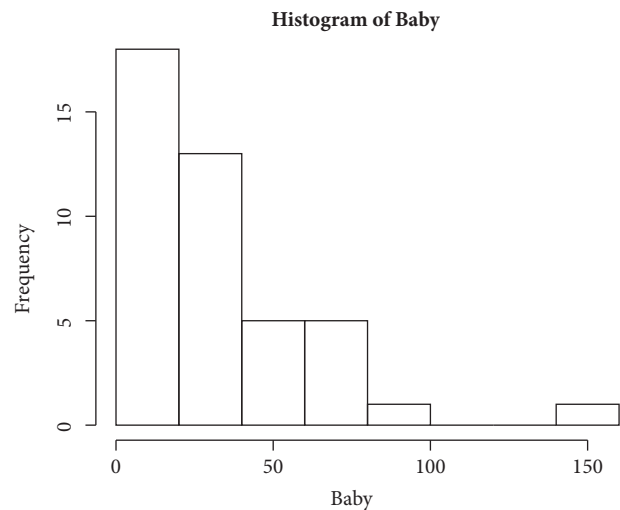

FIGURE 1: Histogram for times between births for baby boom data.

The goal of this example was to carry out a bootstrap study in assessing the robustness and applicability of our proposed test (ELR_{Z2}) on uniformly distributed data. However, the times between births are known to be consistent with the exponential distribution (see Figure 1). By assessing the histogram one can easily see that the data resembles the exponential distribution revealing that the times between births are exponentially consistent. We used the inverse exponential distribution to transform the times between births so that they can be uniformly distributed. We then used the density based empirical likelihood ratio based test (dbEmpLikeGOF) to check if the transformed baby boom data are uniformly distributed. The dbEmpLikeGOF test returned a p value of 0.6950 suggesting that the transformed data are consistent with the uniform distribution.

TABLE 3: Bootstrapping using the inverse exponential transformed baby boom data.

	Bootstrap power comparisons: H_0 : data is normally distributed							
Observations removed	AD	KS_M	CVM	JB	SW	DB	SEELR	ELR_{Z2}
3	0.0000	0.0000	0.0000	0.0000	0.1012	0.6040	0.0000	0.6132
8	0.0164	0.0006	0.0012	0.0000	0.1182	0.4486	0.0146	0.4054
13	0.0281	0.0073	0.0066	0.0011	0.1023	0.3488	0.0568	0.2714

TABLE 4: A numerical assessment on power using the Shiryaev-Roberts (S-R) and CUSUM-type (C-t) statistics for the proposed tests (ELR_{Z1} and ELR_{Z2}) with increased moment constraints at $\alpha = 0.05$.

	ELR_{Z1}				ELR_{Z2}			
n	$k = \{1, 2, 3, 4\}$		$k = \{1, 2, 3, 4, 5\}$		$k = \{1, 2, 3, 4\}$		$k = \{1, 2, 3, 4, 5\}$	
	S-R	C-t	S-R	C-t	S-R	C-t	S-R	C-t
				$t(2)$				
30	0.0416	0.0330	0.0020	0.0010	0.6980	**0.6998**	0.6166	0.5912
50	0.5142	0.4112	0.1666	0.1356	0.8766	**0.8774**	0.8320	0.8030
80	0.8732	0.8336	0.7476	0.7184	**0.9718**	0.9684	0.9544	0.9488
				Cauchy(0,1)				
30	0.3262	0.3438	0.0000	0.0000	**0.9560**	0.9556	0.9248	0.9192
50	0.9538	0.9344	0.7246	0.6754	0.9970	**0.9974**	0.9928	0.9900
80	0.9996	0.9996	0.9964	0.9940	**1.0000**	**1.0000**	0.9998	0.9996
				Uniform(0,1)				
30	**0.7230**	0.1958	0.7208	0.7206	0.5772	0.5986	0.6996	0.7004
50	0.9458	0.5222	**0.9532**	0.9474	0.9032	0.9122	0.9434	0.9398
80	0.9966	0.8462	0.9978	0.9980	0.9940	0.9956	**0.9986**	0.9976
				Exp(1)				
30	0.0094	0.0304	0.0070	0.0068	0.4638	**0.4818**	0.3874	0.3772
50	0.0836	**0.8096**	0.0022	0.0042	0.6274	0.6306	0.5628	0.5380
80	0.3764	**0.9972**	0.2504	0.2346	0.7942	0.8070	0.7558	0.7506
				$t(\delta = 1, \nu = 2)$				
30	0.0476	0.0136	0.0012	0.0028	0.7168	**0.7230**	0.6476	0.6280
50	0.5204	0.4172	0.1932	0.1676	0.8904	**0.8908**	0.8450	0.8294
80	0.8752	0.8610	0.7736	0.7700	0.9714	**0.9766**	0.9636	0.9558
				SN(0,1,5)				
30	0.0514	0.0520	0.0486	0.0442	**0.1394**	0.1242	0.1048	0.0944
50	0.0404	0.0362	0.0350	0.0352	0.1408	**0.1432**	0.1114	0.0904
80	0.0358	0.0338	0.0272	0.0204	0.1592	**0.1646**	0.1158	0.1226

Note. Our proposed tests are maximized on $k \in G$, where G can take any integer to represent the moment constraints used to maximise the test statistics for specified sample sizes at 5% level of significance using 5,000 simulations. n is the sample size. **Bold** represents the powerful test statistic for the given simulation scenarios.

For the resampling study we performed a power simulation study by randomly removing 3, 8, and 13 observations from the transformed baby boom data at 5% significance level using 20,000 replications for each simulation. For comparison's sake we considered the AD test, the modified KS test, the CVM test, the JB test, the SW test, the DB test, the SEELR test, and our proposed test (ELR_{Z2}). The Monte Carlo bootstrap simulation results are presented in Table 3. It is undeniably clear that our test outperformed all the common existing tests and therefore suggests its robustness and applicability on real data. It should be noted that we opted for uniformly distributed data for our application since our proposed test (ELR_{Z2}) proved to be more powerful for symmetric alternative distributions which are defined on $(0, 1)$.

6. Conclusion

An empirical likelihood ratio test for normality based on moment constraints of the half-normal distribution has been developed. Overall, the proposed ELR test has good power properties and significantly outperformed the considered well-known common existing tests against the studied alternative symmetric distributions. In our case, the attractive power properties of the proposed ELR test resulted from the EL method being able to integrate most of the available information by utilizing the first four moment constraints and also through the utilization of the EL function which leads to additional power benefits. We advocate for our proposed test (ELR_{Z2}) to be the preferred choice when one

TABLE 5: Results of the Monte Carlo power comparisons based on samples with sizes (n) from **symmetric alternative distributions** defined on $(-\infty, \infty)$ at $\alpha = 0.05$.

Distribution	n	AD	KS_M	CVM	JB	SW	DB	SEELR	ELR_{Z2}
t(2)	20	0.5068	0.4482	0.5138	0.5632	0.5282	0.2806	0.3774	0.5268
	30	0.6834	0.5832	0.6552	0.7016	0.6908	0.3946	0.4228	0.7004
	50	0.8538	0.7782	0.8370	0.8812	0.8572	0.5640	0.4800	0.8726
	80	0.9602	0.9200	0.9554	0.9646	0.9566	0.8010	0.5420	0.9658
t(4)	20	0.2270	0.1768	0.2114	0.2898	0.2410	0.0922	0.1698	0.2450
	30	0.3002	0.2182	0.2764	0.3788	0.3338	0.1084	0.2164	0.3398
	50	0.4150	0.3176	0.3794	0.5400	0.4520	0.1388	0.2468	0.4784
	80	0.5558	0.3994	0.5210	0.7064	0.6282	0.2094	0.2784	0.6760
t(7)	20	0.1162	0.0952	0.1006	0.1670	0.1398	0.0492	0.1066	0.1346
	30	0.1404	0.1008	0.1306	0.2222	0.1806	0.0552	0.1188	0.1664
	50	0.1806	0.1272	0.1578	0.2954	0.2362	0.0502	0.1422	0.2276
	80	0.2380	0.1618	0.2086	0.4010	0.3122	0.0650	0.1590	0.3324
Cauchy(0,1)	20	0.8780	0.8386	0.8898	0.8622	0.8674	0.7012	0.6368	0.8450
	30	0.9672	0.9410	0.9622	0.9574	0.9610	0.8606	0.6910	0.9542
	50	0.9976	0.9950	0.9964	0.9954	0.9958	0.9712	0.7424	0.9976
	80	1.0000	1.0000	0.9998	0.9998	0.9998	0.9992	0.8882	1.0000
Cauchy(0,5)	20	0.8778	0.8374	0.8796	0.8650	0.8704	0.6902	0.6454	0.8550
	30	0.9628	0.9414	0.9648	0.9512	0.9590	0.8664	0.6950	0.9542
	50	0.9968	0.9948	0.9976	0.9968	0.9966	0.9730	0.7468	0.9962
	80	1.0000	1.0000	1.0000	0.9998	1.0000	0.9996	0.8872	1.0000
Logistic	20	0.1090	0.0872	0.0982	0.1460	0.1138	0.0436	0.0944	0.1158
	30	0.1176	0.0908	0.1220	0.1982	0.1474	0.0452	0.1044	0.1482
	50	0.1562	0.1184	0.1456	0.2620	0.1986	0.0414	0.1216	0.1900
	80	0.2098	0.1406	0.1870	0.3474	0.2662	0.0468	0.1266	0.2908

Anderson-Darling (**AD**) test, Modified Kolmogorov-Smirnov (**KS$_M$**) test [2], Cramer-von Mises test (**CVM**) test, Jarque-Bera (**JB**) test, Shapiro-Wilk (**SW**) test, density based empirical likelihood ratio based (**DB**) test [16], simple and exact empirical likelihood ratio based (**SEELR**) test [13], and the proposed test **ELR$_{Z2}$**.

TABLE 6: Results of the Monte Carlo power comparisons based on samples with sizes (n) from **symmetric alternative distributions** defined on $(0, 1)$ at $\alpha = 0.05$.

Distribution	n	AD	KS_M	CVM	JB	SW	DB	SEELR	ELR_{Z2}
Beta(2,2)	20	0.0564	0.0544	0.0594	0.0052	0.0516	0.1310	0.0696	0.0970
	30	0.0786	0.0520	0.0812	0.0012	0.0768	0.2004	0.0550	0.1962
	50	0.1222	0.0852	0.1172	0.0010	0.1528	0.3468	0.0628	0.4252
	80	0.2340	0.1256	0.1834	0.0128	0.3170	0.5978	0.1128	0.7204
Beta(3,3)	20	0.0404	0.0474	0.0408	0.0076	0.0372	0.0780	0.0518	0.0620
	30	0.0786	0.0520	0.0812	0.0046	0.0768	0.1112	0.0392	0.1030
	50	0.0736	0.0524	0.0650	0.0014	0.0682	0.1654	0.0326	0.1906
	80	0.1076	0.0762	0.0826	0.0022	0.1128	0.2772	0.0298	0.3458
Beta(0.5,0.5)	20	0.6160	0.3098	0.5058	0.0066	0.7190	0.9094	0.7092	0.7015
	30	0.8576	0.4998	0.7332	0.0052	0.9392	0.9914	0.8830	0.8960
	50	0.9902	0.7976	0.9568	0.3822	0.9992	1.0000	0.9916	0.9956
	80	1.0000	0.9724	0.9990	0.9872	1.0000	1.0000	1.0000	1.0000
Uniform(0,1)	20	0.1640	0.1014	0.1396	0.0040	0.1886	0.4064	0.2598	0.3332
	30	0.3004	0.1422	0.2262	0.0020	0.3894	0.6622	0.3202	0.6002
	50	0.5780	0.2532	0.4282	0.0118	0.7546	0.9358	0.5624	0.9120
	80	0.8636	0.4578	0.7092	0.3706	0.9688	0.9990	0.8730	0.9944
Logit-norm(0,1)	20	0.0648	0.0442	0.0562	0.0056	0.0578	0.1294	0.0700	0.1010
	30	0.0858	0.0574	0.0748	0.0024	0.0796	0.1974	0.0658	0.1990
	50	0.1394	0.0812	0.1220	0.0010	0.1612	0.3420	0.0676	0.4156
	80	0.2630	0.1368	0.2114	0.0126	0.3408	0.5830	0.1094	0.7108
Logit-norm(0,2)	20	0.3758	0.1844	0.2934	0.0046	0.4366	0.7034	0.4806	0.5348
	30	0.6092	0.2884	0.4822	0.0030	0.7342	0.9150	0.6512	0.8258
	50	0.9016	0.5412	0.7814	0.1174	0.9742	0.9976	0.9006	0.9818
	80	0.9942	0.8170	0.9644	0.8594	1.0000	1.0000	0.9958	0.9996

Anderson-Darling (**AD**) test, Modified Kolmogorov-Smirnov (**KS$_M$**) test [2], Cramer-von Mises test (**CVM**) test, Jarque-Bera (**JB**) test, Shapiro-Wilk (**SW**) test, density based empirical likelihood ratio based (**DB**) test [16], simple and exact empirical likelihood ratio based (**SEELR**) test [13], and the proposed test **ELR$_{Z2}$**.

TABLE 7: Results of the Monte Carlo power comparisons based on samples with sizes (n) from **asymmetric alternative distributions** defined on $(0, \infty)$ at $\alpha = 0.05$.

		Asymmetric alternative distributions defined on $(0, \infty)$ at $\alpha = 0.05$							
Distribution	n	AD	KS_M	CVM	JB	SW	DB	SEELR	ELR_{Z2}
Exp(1)	20	0.7850	0.5722	0.7222	0.6230	0.8334	0.8384	0.8522	0.3642
	30	0.9296	0.7780	0.8922	0.8286	0.9646	0.9754	0.9996	0.4752
	50	0.9972	0.9594	0.9878	0.9756	0.9998	0.9992	1.0000	0.6400
	80	1.0000	0.9990	0.9998	0.9998	1.0000	1.0000	1.0000	0.8114
Gamma(2,1)	20	0.4590	0.3066	0.4136	0.4080	0.5380	0.4420	0.5684	0.2264
	30	0.6662	0.4776	0.6072	0.5852	0.7502	0.6876	0.8094	0.2844
	50	0.8960	0.6926	0.8436	0.8242	0.9500	0.9180	0.9668	0.3822
	80	0.9840	0.8962	0.9682	0.9782	0.9976	0.9914	0.9984	0.5210
Lognorm(0,1)	20	0.9080	0.7760	0.8846	0.8172	0.9350	0.9210	0.9418	0.6036
	30	0.9838	0.9304	0.9730	0.9466	0.9888	0.9906	1.0000	0.7418
	50	1.0000	0.9942	0.9998	0.9976	1.0000	1.0000	1.0000	0.9068
	80	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9838
Lognorm(0,2)	20	0.9986	0.9904	0.9970	0.9840	0.9990	0.9998	0.9999	0.8894
	30	0.9998	0.9998	1.0000	0.9994	1.0000	1.0000	1.0000	0.9684
	50	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9988
	80	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Weibull(2,1)	20	0.1348	0.0980	0.1142	0.1258	0.1582	0.1264	0.1626	0.0932
	30	0.1828	0.1306	0.1654	0.1704	0.2274	0.1958	0.2718	0.0892
	50	0.3050	0.2000	0.2530	0.2738	0.4086	0.3446	0.5202	0.1120
	80	0.4954	0.3186	0.4200	0.4346	0.6644	0.5634	0.7812	0.1080
Weibull(0.5,1)	20	0.9962	0.9810	0.9954	0.9562	0.9990	0.9996	0.9986	0.8014
	30	1.0000	0.9990	1.0000	0.9972	1.0000	1.0000	1.0000	0.9168
	50	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9866
	80	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Anderson-Darling (**AD**) test, Modified Kolmogorov-Smirnov (**KS_M**) test [2], Cramer-von Mises test (**CVM**) test, Jarque-Bera (**JB**) test, Shapiro-Wilk (**SW**) test, density based empirical likelihood ratio based (**DB**) test [16], simple and exact empirical likelihood ratio based (**SEELR**) test [13], and the proposed test **ELR_{Z2}**.

TABLE 8: Results of the Monte Carlo power comparisons based on samples with sizes (n) from **asymmetric alternative distributions** defined on $(-\infty, \infty)$ at $\alpha = 0.05$.

		Asymmetric alternative distributions defined on $(-\infty, \infty)$ at $\alpha = 0.05$							
Distribution	n	AD	KS_M	CVM	JB	SW	DB	SEELR	ELR_{Z2}
$t(\delta = 1, \nu = 2)$	20	0.6446	0.5692	0.6440	0.6556	0.6498	0.4612	0.5688	0.5542
	30	0.8060	0.7178	0.7934	0.8080	0.8072	0.6210	0.6678	0.7242
	50	0.9492	0.8900	0.9394	0.9414	0.9410	0.7820	0.7782	0.8872
	80	0.9928	0.9762	0.9892	0.9924	0.9924	0.9294	0.8410	0.9726
$t(\delta = 1, \nu = 4)$	20	0.3180	0.2368	0.2848	0.3606	0.3142	0.1638	0.2790	0.2744
	30	0.4086	0.3246	0.3884	0.4810	0.4518	0.2262	0.3584	0.3718
	50	0.5912	0.4618	0.5538	0.6592	0.6360	0.3202	0.4830	0.5344
	80	0.7642	0.6370	0.7296	0.8290	0.8108	0.4626	0.5826	0.7084
$t(\delta = 1, \nu = 7)$	20	0.1490	0.1138	0.1404	0.1934	0.1692	0.0766	0.1420	0.1492
	30	0.1958	0.1424	0.1736	0.2722	0.2318	0.0940	0.1876	0.1936
	50	0.2846	0.1968	0.2522	0.3834	0.3372	0.1194	0.2556	0.2738
	80	0.3848	0.2756	0.3542	0.5102	0.4556	0.1620	0.3300	0.3798
SN(0,1,2)	20	0.0896	0.0756	0.0882	0.1054	0.1068	0.0636	0.0978	0.0710
	30	0.1194	0.0912	0.1062	0.1214	0.1422	0.0784	0.1336	0.0792
	50	0.1666	0.1260	0.1488	0.1810	0.1968	0.1164	0.2080	0.0800
	80	0.2434	0.1918	0.2258	0.2712	0.2940	0.1550	0.3074	0.0890
SN(0,1,5)	20	0.2406	0.1744	0.2212	0.2060	0.2660	0.2092	0.2810	0.1144
	30	0.3586	0.2604	0.3152	0.3056	0.4230	0.3346	0.4742	0.1278
	50	0.5796	0.4098	0.5430	0.4768	0.6672	0.5250	0.7110	0.1372
	80	0.8080	0.6084	0.7554	0.7378	0.8888	0.7394	0.9020	0.1722
SC(0,2,5)	20	0.9660	0.9360	0.9658	0.9410	0.9736	0.9436	0.9462	0.8482
	30	0.9978	0.9884	0.9954	0.9910	0.9970	0.9882	0.9774	0.9524
	50	1.0000	0.9998	1.0000	0.9998	0.9998	0.9992	0.9844	0.9940
	80	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9976	0.9998

Anderson-Darling (**AD**) test, Modified Kolmogorov-Smirnov (**KS_M**) test [2], Cramer-von Mises test (**CVM**) test, Jarque-Bera (**JB**) test, Shapiro-Wilk (**SW**) test, density based empirical likelihood ratio based (**DB**) test [16], simple and exact empirical likelihood ratio based (**SEELR**) test [13], and the proposed test **ELR_{Z2}**.

is testing for departures from normality against symmetric alternative distributions for small to moderate sample sizes. However, our test has low power in the considered asymmetric alternatives and further modifications in improving the power of the test under these alternatives would be much appreciated.

In this study we used the moment constraints of the standardized variables of the half-normal distribution. It will be of interest for one to use the raw moments (nonstandardized data points) of the half-normal distribution. However, according to Dong and Giles [12], the power of the ELR test using standardized observations is within the same range as it is when using nonstandardized data points. Also of interest are the findings by Mittelhammer et al. [29] where they suggested that the power of ELR based tests increases as the moment constraints increase. From our numerical experiment we did not extensively explore this conjecture and this is a potential area of future research and it might be interesting to carry out a more detailed investigation for the proposed tests. We focused on tests for normality, which is a common distribution to test in applied statistical modelling and we believe that our proposed test will assist investigators to use empirical likelihood approaches using moment constraints for goodness-of-fit tests of other applied distributions in practice. By simply ignoring the absolute values of the transformed observations and utilizing standardized half-normal data points our proposed test will simply transform to a GoF test for assessing departures from half-normality.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] A. N. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione," *Giornalle dell'Instituto Italiano degli Attuari*, vol. 4, pp. 83–91, 1933.

[2] H. W. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *Journal of the American Statistical Association*, vol. 62, no. 318, pp. 399–402, 1967.

[3] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain goodness of fit criteria based on stochastic processes," *Annals of Mathematical Statistics*, vol. 23, pp. 193–212, 1952.

[4] T. W. Anderson and D. A. Darling, "A test of goodness of fit," *Journal of the American Statistical Association*, vol. 49, pp. 765–769, 1954.

[5] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality: Complete samples," *Biometrika*, vol. 52, pp. 591–611, 1965.

[6] C. M. Jarque and A. K. Bera, "A test for normality of observations and regression residuals," *International Statistical Review*, vol. 55, no. 2, pp. 163–172, 1987.

[7] R. DAgostino and E. S. Pearson, "Tests for departure from normality. Empirical results for the distributions of b2 and b1," *Biometrika*, vol. 60, no. 3, pp. 613–622, 1973.

[8] P. Royston, "Approximating the Shapiro-Wilk W-test for non-normality," *Statistics and Computing*, vol. 2, no. 3, pp. 117–119, 1992.

[9] S. S. Shapiro and R. S. Francia, "An approximate analysis of variance test for normality," *Journal of the American Statistical Association*, vol. 67, no. 337, pp. 215-216, 1972.

[10] L. Chen and S. S. Shapiro, "An alernative test for normality based on normalized spacings," *Journal of Statistical Computation and Simulation*, vol. 53, no. 3-4, pp. 269–287, 1995.

[11] M. M. Rahman and Z. Govindarajulu, "A modification of the test of Shapiro and Wilk for normality," *Journal of Applied Statistics*, vol. 24, no. 2, pp. 219–235, 1997.

[12] L. B. Dong and D. E. Giles, "An empirical likelihood ratio test for normality," *Communications in Statistics—Simulation and Computation*, vol. 36, no. 1–3, pp. 197–215, 2007.

[13] G. Shan, A. Vexler, G. E. Wilding, and A. D. Hutson, "Simple and exact empirical likelihood ratio tests for normality based on moment relations," *Communications in Statistics—Simulation and Computation*, vol. 40, no. 1, pp. 129–146, 2010.

[14] A. B. Owen, *Empirical Likelihood*, Chapman and Hall, New York, NY, USA, 2001.

[15] S. Steele, *Babies by the Dozen for Christmas: 24-Hour Baby Boom*, The Sunday Mail (Brisbane), 1997.

[16] A. Vexler and G. Gurevich, "Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy," *Computational Statistics & Data Analysis*, vol. 54, no. 2, pp. 531–545, 2010.

[17] C. C. Lin and G. S. Mudholkar, "A simple test for normality against asymmetric alternatives," *Biometrika*, vol. 67, no. 2, pp. 455–461, 1980.

[18] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integrals and Series*, vol. 1, Gordon and Breach Science Publishers, 1986.

[19] A. B. Owen, "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika*, vol. 75, no. 2, pp. 237–249, 1988.

[20] A. Vexler and C. Wu, "An optimal retrospective change point detection policy," *Scandinavian Journal of Statistics*, vol. 36, no. 3, pp. 542–558, 2009.

[21] G. Lorden and M. Pollak, "Nonanticipating estimation applied to sequential analysis and changepoint detection," *The Annals of Statistics*, vol. 33, no. 3, pp. 1422–1454, 2005.

[22] A. Vexler, "Guaranteed testing for epidemic changes of a linear regression model," *Journal of Statistical Planning and Inference*, vol. 136, no. 9, pp. 3101–3120, 2006.

[23] A. Vexler, A. Liu, and M. Pollak, "Transformation of change-point detection methods into a Shiryayev-Roberts form," Tech. Rep., Department of Biostatistics, The New York State University at Buffalo, 2006.

[24] H. Cramér, "On the composition of elementary errors: first paper: mathematical deductions," *Scandinavian Actuarial Journal*, vol. 11, pp. 13–74, 1928.

[25] R. Von Mises, "Wahrscheinlichkeitsrechnung und Ihre Anwen-
    dung in der Statistik und Theoretischen Physik," F. Deuticke,
    Leipzig, Vol. 6.1, 1931.

[26] N. V. Smirnov, "Sui la distribution de w2 (Criterium de M.R.v.
    Mises)," *Comptes Rendus Mathematique Academie des Sciences,
    Paris*, vol. 202, pp. 449–452, 1936.

[27] J. C. Miecznikowski, A. Vexler, and L. Shepherd, "DbEmp-
    LikeGOF: An R package for nonparametric likelihood ratio
    tests for goodness-of-fit and two-sample comparisons based on
    sample entropy," *Journal of Statistical Software*, vol. 54, no. 3, pp.
    1–19, 2013.

[28] P. K. Dunn, "A simple data set for demonstrating common
    distributions," *Journal of Statistics Education*, vol. 7, no. 3, 1999.

[29] R. C. Mittelhammer, G. G. Judge, and D. Miller, *Econometric
    Foundations*, Cambridge University Press, 2000.

Exact Interval Inference for the Two-Parameter Rayleigh Distribution Based on the Upper Record Values

Jung-In Seo,[1] Jae-Woo Jeon,[2] and Suk-Bok Kang[2]

[1]*Department of Statistics, Daejeon University, No. 62, Daehak-ro, Dong-gu, Republic of Korea*
[2]*Department of Statistics, Yeungnam University, No. 280, Daehak-ro, Gyeongsan, Republic of Korea*

Correspondence should be addressed to Suk-Bok Kang; sbkang@yu.ac.kr

The maximum likelihood method is the most widely used estimation method. On the other hand, it can produce substantial bias, and an approximate confidence interval based on the maximum likelihood estimator cannot be valid when the sample size is small. Because the sizes of the record values are considerably smaller than the original sequence observed in the majority of cases, a method appropriate for this situation is required for precise inference. This paper provides the exact confidence intervals for unknown parameters and exact predictive intervals for the future upper record values by providing some pivotal quantities in the two-parameter Rayleigh distribution based on the upper record values. Finally, the validity of the proposed inference methods was examined from Monte Carlo simulations and real data.

1. Introduction

The cumulative distribution function (cdf) and probability density function (pdf) of the random variable (RV), X, with the Rayleigh distribution are given, respectively, by

$$F(x) = 1 - \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \tag{1}$$

$$f(x) = \frac{x-\mu}{\sigma^2} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad x > \mu, \ \sigma > 0, \tag{2}$$

where μ is the location parameter and σ is the scale parameter. The Rayleigh distribution is used because the life of the model theory reliability plays an important role in modeling the life of the random phenomenon. Moreover, it is used in many applications, such as reliability, life tests, and survival analysis because its failure rate is a linear function of time. Therefore, this distribution has been studied by many authors in the case, where samples are censored due to a range of reasons. Dyer and Whisenand [1] examined the properties of the k-optimum best linear unbiased estimators (BLUEs) of the scale parameter in the Rayleigh distribution and provided an approximate k-optimum BLUE based on k order

statistics. Sinha and Howlader [2] derived the highest posterior density (HPD) credible interval for the scale parameter and the reliability function in a Rayleigh distribution. Ali Mousa and Al-Sagheer [3] obtained the maximum likelihood estimators (MLEs) and Bayes estimators for μ, σ, and the reliability function of the Rayleigh distribution obtained based on the progressively Type-II censored data. Raqab and Madi [4] discussed the Bayesian predictive methods for the total time on test using doubly censored data with a Rayleigh distribution and the scale parameter and applied the methods to a real data set that represented the deep-groove ball bearing failure times. For the same real data, Kim and Han [5] applied a Bayesian inference method based on the conjugate prior of the scale parameter of the Rayleigh distribution under general progressive censoring and S. Dey and T. Dey [6] applied this by providing point and interval estimation methods for the scale parameter of the Rayleigh distribution under progressive Type-II censoring with binomial removal. This paper considered a two-parameter Rayleigh distribution based on the upper record values that are used extensively to build statistical modeling arising in many real-life situations involving weather, sports, economics, and life tests. The record values are described as follows.

Let $\{X_i, i = 1, \ldots, n\}$ be a sequence of independent and identically distributed (iid) RVs from a continuous probability distribution. If $X_i > X_j$ for all $i < j$, then X_j is an upper record value. The indices at which the upper record values occur are given by the record times $\{U(k), k \geq 1\}$, where $U(k) = \min\{j \mid j > U(k-1), X_j > X_{U(k-1)}\}$, $k \geq 1$ with $U(1) = 1$. Chandler [7] first studied the record values and their basic properties. Ahsanullah [8] provided detailed descriptions of the general theory and applications for the well-known probability distributions based on the records. Seo and Kim [9] provided inference methods to estimate unknown parameters and predicted future upper record values from the extreme value distribution using both frequentist and Bayesian approaches. Note that the sizes of the record values are actually considerably smaller than the observed original sequence in the majority of case; a method appropriate for this situation is required for precise inference. The maximum likelihood method is the most extensively used estimation method. On the other hand, the approximate confidence intervals (CIs) based on the asymptotic normality of the MLE can yield inappropriate results because μ and σ are supported by $(-\infty, x)$ and $(0, \infty)$, respectively. Moreover, the asymptotic normality of the MLE requires the suitable regularity conditions and it is difficult to prove that the regularity conditions are satisfied when the record values are observed from the two-parameter Rayleigh distribution. This paper constructs exact CIs for unknown parameters (μ, σ) of the Rayleigh distribution based on the upper record values by providing some pivotal quantities, which are much more efficient than the maximum likelihood method in terms of computation cost. Another aim of this paper is to construct exact predictive intervals (PIs) for the future upper record values based on the past upper record values from the Rayleigh distribution because it is very important to correctly predict in many fields such as earthquakes, flood, and rainfall.

The remainder of the paper is structured as follows. Section 2 provides some pivotal quantities and derives the exact CIs for unknown parameters and PIs for the future upper record values in the Rayleigh distribution based on the upper record values. Section 3 assesses the validity of the proposed method through Monte Carlo simulations and real data. Section 4 concludes the paper.

2. Inference Based on Pivotal Quantity

The likelihood function for θ is given by Arnold et al., [10] as

$$L(\theta) = \prod_{i=1}^{k} f(x_i \theta). \tag{3}$$

Let $X_{U(1)}, \ldots, X_{U(k)}$ be the first k upper record values from the two-parameter Rayleigh distribution. The likelihood function based on record values is given by

$$L(\mu, \sigma) = \exp\left[-\frac{(x_{U(k)} - \mu)^2}{2\sigma^2}\right] \prod_{i=1}^{k} \frac{x_{U(i)} - \mu}{\sigma^2}. \tag{4}$$

The MLEs $\hat{\mu}$ and $\hat{\sigma}$ can be found by solving the following likelihood equations for μ and σ simultaneously:

$$\frac{\partial}{\partial \mu} \log L(\mu, \sigma) = \frac{x_{U(k)} - \mu}{\sigma^2} - \sum_{i=1}^{k} \frac{1}{x_{U(i)} - \mu} = 0, \tag{5}$$

$$\frac{\partial}{\partial \sigma} \log L(\mu, \sigma) = \frac{(x_{U(k)} - \mu)^2}{\sigma^3} - \frac{2k}{\sigma} = 0.$$

On the other hand, the MLEs cannot be expressed in closed form and their exact distributions are difficult to derive. Alternatively, by the asymptotic normality of the MLE, the approximate $100(1-\alpha)\%$ CIs for μ and σ can be obtained as

$$\hat{\mu} \pm Z_{\alpha/2} \sqrt{\mathrm{Var}(\hat{\mu})},$$

$$\hat{\sigma} \pm Z_{\alpha/2} \sqrt{\mathrm{Var}(\hat{\sigma})}, \tag{6}$$

where $Z_{\alpha/2}$ denotes the upper $\alpha/2$ point of the standard normal distribution and the variances $\mathrm{Var}(\hat{\mu})$ and $\mathrm{Var}(\hat{\sigma})$ are the diagonal elements of the asymptotic variance-covariance matrix obtained by inverting the Fisher information matrix for unknown parameters (μ, σ):

$$I(\mu, \sigma)$$

$$= \begin{bmatrix} E\left(-\dfrac{\partial^2}{\partial \mu^2} \log L(\mu, \sigma)\right) & E\left(-\dfrac{\partial^2}{\partial \mu \partial \sigma} \log L(\mu, \sigma)\right) \\ E\left(-\dfrac{\partial^2}{\partial \sigma \partial \mu} \log L(\mu, \sigma)\right) & E\left(-\dfrac{\partial^2}{\partial \sigma^2} \log L(\mu, \sigma)\right) \end{bmatrix}, \tag{7}$$

under certain regularity conditions. Nevertheless, it can provide inappropriate results because the supports of μ and σ do not coincide with that of the normal distribution and the record values are rarely observed, as mentioned before.

2.1. Confidence Interval. This subsection develops inference methods based on the pivotal quantities to construct exact CIs for unknown parameters (μ, σ) and PIs for the future upper record values. Note that the proposed method is much easier to calculate than the maximum likelihood method. The following provides some pivotal quantities.

Let

$$Z_i = -\log\left[1 - F(x_{U(i)})\right] = \frac{(X_{U(i)} - \mu)^2}{2\sigma^2}, \tag{8}$$

$$i = 1, \ldots, k.$$

$Z_1 < \cdots < Z_k$ are the upper record values with a standard exponential distribution. From this result, the following spacing can be obtained:

$$S_i = Z_i - Z_{i-1} = \frac{1}{2\sigma^2}\left[\left(X_{U(i)} - \mu\right)^2 - \left(X_{U(i-1)} - \mu\right)^2\right],$$

$$i = 1, \ldots, k \ \ (Z_0 \equiv 0), \tag{9}$$

which are the iid RVs from the standard exponential distribution (see Arnold et al., [10]). Based on the spacing, a pivotal quantity $T = 2S_1$ having a χ^2 distribution with 2 degrees of freedom and a pivotal quantity can be derived as

$$V = 2\sum_{i=2}^{k} S_i = \frac{\left(X_{U(k)} - \mu\right)^2}{\sigma^2} - \frac{\left(X_{U(1)} - \mu\right)^2}{\sigma^2}, \tag{10}$$

having the χ^2 distribution with $2(k-1)$ degrees of freedom. Because they have independent RVs, the following pivotal quantity is obtained:

$$W(\mu) = \frac{V/2(k-1)}{T/2} = \frac{1}{k-1}\left[\left(\frac{X_{U(k)} - \mu}{X_{U(1)} - \mu}\right)^2 - 1\right], \tag{11}$$

which has a F distribution with $2(k-1)$ and 2 degrees of freedom. An exact $100(1 - \alpha)\%$ CI for μ based on the pivotal quantity $W(\mu)$ can be constructed as

$$\left(\frac{X_{U(k)} - X_{U(1)}\sqrt{(k-1)F_{\alpha/2,(2(k-1),2)} + 1}}{1 - \sqrt{(k-1)F_{\alpha/2,(2(k-1),2)} + 1}} < \mu\right.$$

$$\left. < \frac{X_{U(k)} - X_{U(1)}\sqrt{(k-1)F_{1-\alpha/2,(2(k-1),2)} + 1}}{1 - \sqrt{(k-1)F_{1-\alpha/2,(2(k-1),2)} + 1}}\right), \tag{12}$$

for any $0 < \alpha < 1$, where $F_{\alpha,(\nu_1,\nu_2)}$ is the upper α percentile of the F distribution with ν_1 and ν_2 degrees of freedom.

Moreover, because $Q(\sigma) = T + V$ has the χ^2 distribution with $2k$ degrees of freedom, an exact $100(1 - \alpha)\%$ CI for σ based on the pivotal quantity Q can be constructed as

$$\left(\sqrt{\frac{\left(X_{U(k)} - \mu\right)^2}{\chi^2_{2k,\alpha}}} < \sigma < \sqrt{\frac{\left(X_{U(k)} - \mu\right)^2}{\chi^2_{2k,1-\alpha}}}\right) \tag{13}$$

$$\text{for any } 0 < \alpha < 1,$$

where $\chi^2_{\alpha,k}$ is the upper α percentile of the χ^2 distribution with k degrees of freedom. Note that because the precise CI (13) depends on the nuisance parameter μ, this paper shows how

to address the nuisance parameter μ based on a generalized pivotal quantity, and an exact CI for σ is proposed based on the generalized pivotal quantity.

Let μ^* be the unique solution of $W(\mu) = W$, where W has a F distribution with $2(k - 1)$ and 2 degrees of freedom. The unique solution can then be given by

$$\mu^* = \frac{X_{U(k)} - X_{U(1)}\sqrt{(k-1)W + 1}}{1 - \sqrt{(k-1)W + 1}}. \tag{14}$$

Moreover, let Q be the RV from the χ^2 distribution with $2k$ degrees of freedom. The generalized pivotal quantity from the pivotal quantity $Q(\sigma)$ is given by

$$W(\mu^*) = \sqrt{\frac{\left(X_{U(k)} - \mu^*\right)^2}{Q}}. \tag{15}$$

Here, the samples $W(\mu^*)_{(1)}, \ldots, W(\mu^*)_{(N)}$ can be obtained by generating $N(\geq 10{,}000)$ the RVs W and Q. $W(\mu^*)'_i$s are ordered as $W(\mu^*)_{(1)}, \ldots, W(\mu^*)_{(N)}$. Therefore, an exact $100(1 - \alpha)\%$ CI for σ based on the generalized pivotal quantity $W(\mu^*)$ can be constructed:

$$\left(W\left(\mu^*\right)_{[(N/100)\times\alpha/2]}, W\left(\mu^*\right)_{[(N/100)\times(1-\alpha/2)]}\right), \tag{16}$$

where $[z]$ denotes the largest integer less than or equal to z. In Section 3, the proposed CIs are examined in terms of the coverage probability (CPs) to determine if they are valid CIs.

2.2. Predictive Interval. This subsection develops a method for predicting the future upper record values based on the observed upper record values $x_{U(i)}, \ldots, x_{U(k)}$ by providing a pivotal quantity. Let $X_{U(s)}$ $(s > k)$ be a future upper record value. The conditional density function of $X_{U(s)}$, given $x_{U(k)}$, defined by Ahsanullah [11], is given by

$$f_{X_{U(s)}|x_{U(k)}}\left(x_{U(s)} \mid \mu, \sigma\right)$$

$$= \frac{1}{\Gamma(s-k)}\left[\log\left(1 - F\left(x_{U(k)}\right)\right)\right.$$

$$\left. - \log\left(1 - F\left(x_{U(s)}\right)\right)\right]^{s-k-1} \frac{f\left(x_{U(s)}\right)}{1 - F\left(x_{U(k)}\right)}, \tag{17}$$

$$\mu < x_{U(k)} < x_{U(s)},$$

from the Markov property of the record values. Assuming that the observed upper record values, $x_{U(i)}, \ldots, x_{U(k)}$, arise

from the Rayleigh distribution with the pdf (2), the conditional density function (17) is written as

$$f_{X_{U(s)}|x_{U(k)}}\left(x_{U(s)} \mid \mu, \sigma\right)$$

$$= \frac{1}{\Gamma(s-k)} \left[\frac{(x_{U(s)} - \mu)^2}{2\sigma^2} - \frac{(x_{U(k)} - \mu)^2}{2\sigma^2}\right]^{s-k-1} \tag{18}$$

$$\cdot \frac{x_{U(s)} - \mu}{\sigma^2} \exp\left[\frac{(x_{U(k)} - \mu)^2}{2\sigma^2} - \frac{(x_{U(s)} - \mu)^2}{2\sigma^2}\right].$$

Let

$$Y = \frac{1}{\sigma^2}\left[(X_{U(s)} - \mu)^2 - (x_{U(k)} - \mu)^2\right]. \tag{19}$$

Because the Jacobian of transformation is

$$\frac{d}{dy}X_{U(s)} = \frac{\sigma^2}{2\sqrt{\sigma^2 y + (x_{U(k)} - \mu)^2}}, \tag{20}$$

the density function of Y is given by

$$f(y) = \frac{1}{2^{s-k}\Gamma(s-k)} y^{s-k-1} e^{-y/2}, \quad 0 < y < \infty, \tag{21}$$

which is the pdf of the χ^2 distribution with $2(s-k)$ degrees of freedom. Suppose that μ and σ are known. An exact $100(1-\alpha)\%$ PI based on the pivotal quantity Y for the future upper record value $X_{U(s)}$ is obtained as

$$\left(\mu + \sqrt{\sigma^2 \chi^2_{1-\alpha/2,2(s-k)} + (x_{U(k)} - \mu)^2}, \mu\right.$$

$$\left. + \sqrt{\sigma^2 \chi^2_{\alpha/2,2(s-k)} + (x_{U(k)} - \mu)^2}\right). \tag{22}$$

When μ and σ are unknown, they can be substituted with by μ^* and $W(\mu^*)$ in PI (22) based on the fact that $W(\mu^*)$ is the generalized pivotal quantity for constructing the exact CI for σ. In the same way, the generalized pivotal quantity is given by

$$Y(\mu^*)$$

$$= \left(\mu^* + \sqrt{W(\mu^*)^2 \chi^2_{1-\alpha/2,2(s-k)} + (x_{U(k)} - \mu^*)^2}\right), \tag{23}$$

and an exact $100(1-\alpha)\%$ PI for $X_{U(s)}$ based on the generalized pivotal quantity $Y(\mu^*)$ can be constructed as follows:

$$\left(Y(\mu^*)_{[(N/100)\times\alpha/2]}, Y(\mu^*)_{[(N/100)\times(1-\alpha/2)]}\right). \tag{24}$$

3. Application

This section assesses the proposed methods through a Monte Carlo simulation and presents a real data set.

3.1. Simulation Study. The proposed exact CIs (12) and (16) are assessed in terms of their CPs and average lengths (ALs). The upper record values were first generated from the standard Rayleigh distribution with $\mu = 0$ and $\sigma = 1$ for different k, and the CIs (12) and (16) were calculated based on the generated samples by using the provided methods in Section 2.1. The CPs and ALs of the exact CIs were obtained over 10,000 simulations. These values are reported in Table 1.

Table 1 shows that the CPs matched their corresponding nominal levels even in a small sample size and that all ALs decrease with increasing sample size.

3.2. Real Data. To illustrate the proposed inference procedure, the survival times in (days) of a group of lung cancer patients (from Lawless [12, p. 319]) were considered as follows:

$$6.96, 9.30, 6.96, 7.24, 9.30, 4.90, 8.42, 6.05, 10.18, 6.82, 8.58, 7.77, 11.94, 11.25, 12.94, \text{ and } 12.94. \tag{25}$$

From the data, the observed upper record values were 6.96, 9.30, 10.18, 11.94, and 12.94. Soliman and Al-Aboud [13] showed that the Rayleigh distribution fits the observed record values well. These record values are employed to obtain the proposed CIs (12) and (16). Moreover, the exact PIs for the future upper record values $X_{U(s)}$ ($s = 6, 7$) were computed, as listed in Table 2.

4. Concluding Remarks

This paper proposes methods for inferencing the exact CIs for unknown parameters (μ, σ) in the Rayleigh distribution based on the upper record values and exact PIs for the future upper record values by providing some pivotal quantities.

Because the proposed exact CI (13) and PI (22) depend on the nuisance parameters, this study proposed generalized pivotal quantities $W(\mu^*)$ and $Y(\mu^*)$ to solve the drawback. The proposed methods were more computationally convenient than the maximum likelihood method. Moreover, the proposed exact CIs provide very good performance even in small sample sizes. If the location parameter of the Rayleigh distribution is of interest, the exact CI (12) should be used because it does not have any nuisance parameter.

Competing Interests

The authors declare that they have no competing interests.

TABLE 1: CPs (ALs) of exact 95% CIs for μ and σ.

k	μ	σ
5	0.949 (6.135)	0.952 (2.468)
7	0.953 (4.926)	0.954 (1.723)
9	0.951 (4.353)	0.950 (1.329)
11	0.947 (4.044)	0.953 (1.127)
13	0.949 (3.845)	0.954 (0.995)
15	0.950 (3.688)	0.951 (0.901)

TABLE 2: Results for real data.

μ	σ	$X_{U(6)}$	$X_{U(7)}$
(−13.771, 6.444)	(1.768, 9.840)	(12.963, 18.818)	(13.148, 21.961)

References

[1] D. D. Dyer and C. W. Whisenand, "Best linear unbiased estimator of the Rayleigh distribution," *IEEE Transactions on Reliability*, vol. 22, pp. 27–34, 1973.

[2] S. K. Sinha and H. A. Howlader, "Credible and hpd intervals of the parameter and reliability of rayleigh distribution," *IEEE Transactions on Reliability*, vol. R-32, no. 2, pp. 217–220, 1983.

[3] M. A. Ali Mousa and S. A. Al-Sagheer, "Statistical inference for the rayleigh model based on progressively type-II censored data," *Statistics. A Journal of Theoretical and Applied Statistics*, vol. 40, no. 2, pp. 149–157, 2006.

[4] M. Z. Raqab and M. T. Madi, "Bayesian prediction of the total time on test using doubly censored Rayleigh data," *Journal of Statistical Computation and Simulation*, vol. 72, no. 10, pp. 781–789, 2002.

[5] C. S. Kim and K. H. Han, "Estimation of the scale parameter of the Rayleigh distribution under general progressive censoring," *Journal of the Korean Statistical Society*, vol. 38, no. 3, pp. 239–246, 2009.

[6] S. Dey and T. Dey, "Statistical inference for the Rayleigh distribution under progressively Type-II censoring with binomial removal," *Applied Mathematical Modelling*, vol. 38, no. 3, pp. 974–982, 2014.

[7] K. N. Chandler, "The distribution and frequency of record values," *Journal of the Royal Statistical Society. Series B. Methodological*, vol. 14, pp. 220–228, 1952.

[8] M. Ahsanullah, *Record Values-Theory and Applications*, University Press of America, New York, NY, USA, 2004.

[9] J. I. Seo and Y. Kim, "Bayesian inference on extreme value distribution using upper record values," *Communications in Statistics-Theory and Methods*, 2016.

[10] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *Records*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, USA, 1998.

[11] M. Ahsanullah, *Record Statistics*, Nova Science, New York, NY, USA, 1995.

[12] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York, NY, USA, 1982.

[13] A. A. Soliman and F. M. Al-Aboud, "Bayesian inference using record values from Rayleigh model with application," *European Journal of Operational Research*, vol. 185, no. 2, pp. 659–672, 2008.

Performance of Synthetic Double Sampling Chart with Estimated Parameters Based on Expected Average Run Length

Huay Woon You

Pusat PERMATApintar Negara, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Correspondence should be addressed to Huay Woon You; hwyou@ukm.edu.my

A synthetic double sampling (SDS) chart is commonly evaluated based on the assumption that process parameters (namely, mean and standard deviation) are known. However, the process parameters are usually unknown and must be estimated from an in-control Phase-I dataset. This will lead to deterioration in the performance of a control chart. The average run length (ARL) has been implemented as the common performance measure in process monitoring of the SDS chart. Computation of ARL requires practitioners to determine shift size in advance. However, this requirement is too restricted as practitioners may not have the experience to specify the shift size in advance. Thus, the expected average run length (EARL) is introduced to assess the performance of the SDS chart when the shift size is random. In this paper, the SDS chart, with known and estimated process parameters, was evaluated based on EARL and compared with the performance measure, ARL.

1. Introduction

The quality of products and services is an essential factor in the world of business [1]. In order to enhance the quality of products and services, statistical process control (SPC) is used to monitor and attain the process of manufacturing and services. Among the SPC techniques, control charts are one of the most prominent techniques for detecting shifts in a process. The first control chart was proposed by Dr. Walter A. Shewhart, and it was named the Shewhart chart. The Shewhart chart is frequently used to detect large process mean shifts due to its simplicity [2]. However, the main limitation of the Shewhart chart is that it is insensitive in detecting moderate and small process mean shifts.

In recent years, Khoo et al. [3] suggested synthetic double sampling (SDS) chart, which combines double sampling (DS) subchart and conforming run length (CRL) subchart. From the findings, the SDS chart is efficient in detecting moderate and small process mean shifts compared to the synthetic chart and double sampling chart. The implementation of the SDS chart is based on the assumption that the process parameters are known. Nevertheless, the process parameters are generally unknown in practice. This requires for an estimation of the process parameters from the in-control Phase-I samples.

Saleh and Mahmoud [4] claimed that when the estimated process parameters are used in place of the known process parameters, the performance of the control chart is affected due to the existence of variability in the estimation. Woodall and Montgomery [5] recognised the importance of examining the effects of parameter estimation on the performance of a particular control chart. Therefore, the effects of parameter estimation need to be considered when designing a control chart. Among others, [6–8] have examined the impacts of Phase-I parameter estimation on the performance of a control chart.

The performance of a control chart is crucial in determining the appropriate control chart to be implemented in a process. A common performance measure in process monitoring is average run length (ARL). ARL is the number of samples (on average) plotted on a control chart before it signals an out-of-control [9]. By using ARL as a performance measure, the chart's user needs to determine the process shift size.

You et al. [10] investigated the ARL performance of the SDS chart when process parameters were estimated and this

has motivated the current research work. In most practical situations, it is usual that the next shift size is unknown in advance [11, 12]. To circumvent this problem, the expected average run length (EARL) is proposed to examine the performance of the SDS chart when the process shift size is unknown and random. Hence, the performance of the SDS chart, with known and estimated process parameters, is investigated using EARL. In addition, the proposed alternative performance measure, i.e., EARL of the SDS chart, will be compared with the corresponding chart using ARL.

The rest of this paper is structured as follows: Section 2 presents the operation and steps to implement the SDS chart. Moreover, the run length properties of the SDS chart with known and estimated process parameters are also given in Section 2. Section 3 illustrates the performance comparison of the SDS chart, based on EARL and ARL, for known process parameters and that of the corresponding chart with estimated process parameters. Finally, concluding remarks are drawn in the last section.

2. Materials and Methods

Khoo et al. [3] developed the SDS chart, which comprises the DS subchart and a CRL subchart. The CRL subchart is an attribute chart, with one lower limit, i.e., L_3. Figure 1 illustrates the operation of the SDS chart with known process parameters.

Step 1. Set the charting parameters n_1, n_2, L, L_1, L_2, and L_3.

Step 2. At sampling time $i = 1, 2, \ldots$, take the first sample of size n_1 and the sample mean, $\overline{Y}_{1,i} = \sum_{j=1}^{n_1} Y_{1j}/n_1$, is calculated.

Step 3. Compute the standardised statistic $Z_{1,i} = [(\overline{Y}_{1,i} - \mu_0)\sqrt{n_1}]/\sigma_0$ corresponding to the first sample.

Step 4. If $Z_{1,i}$ is in I_1, the ith sampling time is conforming and the control flow returns to Step 2.

Step 5. If $Z_{1,i}$ is in I_3, the ith sampling time is nonconforming and the control flow goes to Step 9.

Step 6. If $Z_{1,i}$ is in I_2, take a second sample with n_2 and compute the sample mean, $\overline{Y}_{2,i} = \sum_{j=1}^{n_2} Y_{2j}/n_2$.

Step 7. Calculate the sample mean of the combined samples $\overline{Y}_i = (n_1\overline{Y}_{1,i} + n_2\overline{Y}_{2,i})/(n_1 + n_2)$ and the standardised statistic $Z_i = [(\overline{Y}_i - \mu_0)\sqrt{n_1 + n_2}]/\sigma_0$ corresponding to the combined samples.

Step 8. If Z_i is in I_4, and the ith sampling time is conforming, then return to Step 2. Otherwise, the sampling time is regarded as nonconforming and the control flow goes to Step 9.

Step 9. Count the number of inspected sampling times between the present and last nonconforming sampling times inclusive of the present nonconforming sampling time, and denote it as the CRL value.
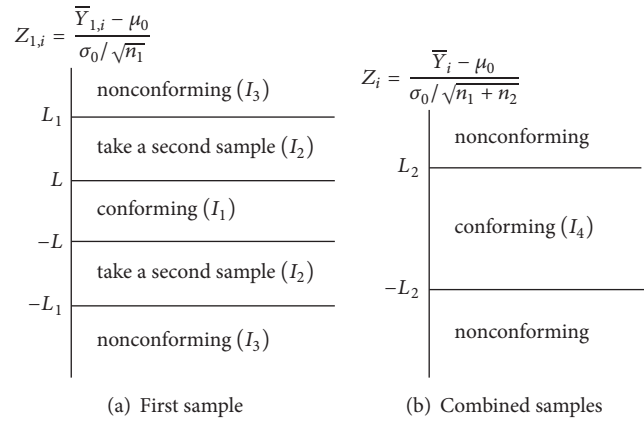

FIGURE 1: DS subchart.

Step 10. If CRL $> L_3$, the process is in-control and the control flow returns to Step 2. Otherwise, the process is out-of-control and immediate actions are required to eliminate the assignable cause(s). Then, return to Step 2.

Without loss of generality, the in-control mean, μ_0, and in-control standard deviation, σ_0, are assumed as known. Let $P = 1 - P_a - P_b$ be the probability of deciding that a sampling time is nonconforming in the DS subchart. Note that P_a and P_b can be expressed as follows [3]:

$$P_a = \Pr(Z_{1,i} \in I_1) = \Phi(L + \delta\sqrt{n_1}) - \Phi(-L + \delta\sqrt{n_1}),$$

$$P_b = \Pr(Z_i \in I_4, \ Z_{1,i} \in I_2)$$

$$= \int_{z \in I_2^*} \left[\Phi\left(cL_2 + rc\delta - z\sqrt{\frac{n_1}{n_2}} \right) \right.$$

$$\left. - \Phi\left(-cL_2 + rc\delta - z\sqrt{\frac{n_1}{n_2}} \right) \right] \phi(z)\, dz, \tag{1}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution function (cdf) and probability density function (pdf) for a standard normal random variable, respectively. Here, $r = \sqrt{n_1 + n_2}$, $c = r/\sqrt{n_2}$, and $I_2^* = [-L_1 + \delta\sqrt{n_1}, -L + \delta\sqrt{n_1}) \cup (L + \delta\sqrt{n_1}, L_1 + \delta\sqrt{n_1}]$.

Finally, for the SDS chart when process parameters are known, the ARL is equal to

$$\text{ARL} = \frac{1}{P} \times \frac{1}{1 - (1 - P)^{L_3}} \tag{2}$$

Moreover, when the exact shift size is unknown, it is essential to consider the EARL for an overall range of shifts $(\delta_{\min}, \delta_{\max})$, where $\delta_{\min}$ and $\delta_{\max}$ indicate the lower and upper bounds of the mean shift, respectively. The EARL of the SDS chart with known process parameters is

$$\text{EARL} = \int_{\delta_{\min}}^{\delta_{\max}} f_\delta(\delta)\, \text{ARL}\, d\delta, \tag{3}$$

where ARL can be obtained from (2) and $f_\delta(\delta)$ is the pdf of the shift size δ. Since the actual shape of $f_\delta(\delta)$ is usually

unknown, it is assumed that the shifts in process mean happen with equal probability; i.e., δ is uniformly distributed with $U(\delta_{\min}, \delta_{\max})$. Therefore, (3) reduces to

$$\text{EARL} = \frac{1}{\delta_{\max} - \delta_{\min}} \int_{\delta_{\min}}^{\delta_{\max}} \text{ARL} \, d\delta. \tag{4}$$

In reality, μ_0 and σ_0 are unknown and need to be estimated from m Phase-I samples, each of size n, i.e., $\{X_{i,1}, X_{i,2}, ..., X_{i,n}\}$, for $i = 1, 2, \ldots, m$. The estimators of μ_0 and σ_0 are [10]

$$\widehat{\mu}_0 = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} X_{i,j}, \tag{5}$$

$$\widehat{\sigma}_0 = \sqrt{\frac{1}{m(n-1)} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(X_{i,j} - \overline{X}_i\right)^2}, \tag{6}$$

respectively. As the values of μ_0 and σ_0 are both unknown and need to be estimated using $\widehat{\mu}_0$ and $\widehat{\sigma}_0$, the standardised statistic for the first sample and combined samples at the sampling time i of the DS subchart with estimated process parameters becomes

$$\widehat{Z}_{1,i} = \frac{\overline{X}_{1,i} - \widehat{\mu}_0}{\widehat{\sigma}_0 / \sqrt{n_1'}}, \tag{7}$$

$$\widehat{Z}_i = \frac{\overline{X}_i - \widehat{\mu}_0}{\widehat{\sigma}_0 / \sqrt{n_1' + n_2'}}, \tag{8}$$

respectively.

Here, n_1' and n_2' represent the sample sizes of the SDS chart with the estimated process parameters. Similarly, $L', L_1', L_2',$ and L_3' correspond to the limits of the SDS chart when the process parameters are estimated.

Then, the probability for the DS subchart with the estimated process parameters to identify a nonconforming sampling time is given as $\widehat{P} = 1 - \widehat{P}_a - \widehat{P}_b$; i.e., [10],

$$\widehat{P} = 1 - \left[\Phi \left(W \sqrt{\frac{n_1'}{mn}} + RL' - \delta \sqrt{n_1'} \right) \right.$$
$$\left. - \Phi \left(W \sqrt{\frac{n_1'}{mn}} - RL' - \delta \sqrt{n_1'} \right) \right] \tag{9}$$
$$- \left(\int_{z \in I_2} \widehat{P}_4 \times f_{\widehat{Z}_{1,i}} \left(z \mid \widehat{\mu}_0, \widehat{\sigma}_0 \right) dz \right),$$

with

$$\widehat{P}_4 = \Phi \left[W \sqrt{\frac{n_2'}{mn}} + R \left(\frac{L_2' \sqrt{n_1' + n_2'} - z \sqrt{n_1'}}{\sqrt{n_2'}} \right) \right.$$
$$\left. - \delta \sqrt{n_2'} \right] - \Phi \left[W \sqrt{\frac{n_2'}{mn}} \right.$$
$$\left. - R \left(\frac{L_2' \sqrt{n_1' + n_2'} + z \sqrt{n_1'}}{\sqrt{n_2'}} \right) - \delta \sqrt{n_2'} \right], \tag{10}$$

$$f_{\widehat{Z}_{1,i}} \left(z \mid \widehat{\mu}_0, \widehat{\sigma}_0 \right) = R\phi \left(W \sqrt{\frac{n_1'}{mn}} + Rz - \delta \sqrt{n_1'} \right).$$

Here, W and R are random variables denoted as

$$W = \left(\widehat{\mu}_0 - \mu_0 \right) \frac{\sqrt{mn}}{\sigma_0}, \tag{11}$$

$$R = \frac{\widehat{\sigma}_0}{\sigma_0}, \tag{12}$$

respectively.

As $\widehat{\mu}_0 \sim N[\mu_0, \sigma_0^2/(mn)]$, it can be deduced that the pdf of W is

$$f_W(w) = \phi(w). \tag{13}$$

For the random variable R, it is known that $\widehat{\sigma}_0^2 / \sigma_0^2 \sim \gamma(m(n-1)/2, 2/(m(n-1)))$, i.e., the gamma distribution with parameters $[m(n-1)]/2$ and $2/[m(n-1)]$. Using this property, the pdf of R is as follows:

$$f_R(r \mid m, n) = 2r f_\gamma \left(r^2 \mid \frac{m(n-1)}{2}, \frac{2}{m(n-1)} \right), \tag{14}$$

where $f_\gamma(\cdot)$ is the pdf of the gamma distribution with parameters $[m(n-1)]/2$ and $2/[m(n-1)]$. Note that, for complete and detailed derivation, reader can refer to You et al. [10].

Thus, the ARL of the SDS chart with estimated process parameters is

$$\text{ARL}_m = \int_{-\infty}^{+\infty} \int_0^{+\infty} \frac{1}{\widehat{P}} \times \frac{1}{1 - \left(1 - \widehat{P}\right)^{L_3'}} f_W(w)$$
$$\cdot f_R(r \mid m, n) \, dr \, dw. \tag{15}$$

Consequently, when the process parameters are estimated from the in-control Phase-I samples, the computation of the EARL is

$$\text{EARL}_m$$
$$= \int_{-\infty}^{+\infty} \int_0^{+\infty} \text{EARL} f_W(w) f_R(r \mid m, n) \, dr \, dw, \tag{16}$$

where EARL can be obtained from (3) by replacing P and L_3 with $\widehat{P}$ and L_3', respectively.

TABLE 1: Optimal charting parameters $(n_1, n_2, L, L_1, L_2, L_3)$ and the corresponding $EARL_1$s for $n = \{3, 4, 5, 6\}$ with different combinations of $(m, \delta_{min}, \delta_{max})$ when $EARL_0 = 370.4$.

n	δ_{min}	δ_{max}	n_1	n_2	L	L_1	L_2	L_3	m					
									30	50	80	200	500	$+\infty$
3	0.2	1.0	2	6	1.3830	5.2804	2.0239	9	3.56	3.36	3.26	3.18	3.15	3.13
	1.0	2.0	2	2	0.6745	4.8576	2.0503	2	1.02	1.02	1.02	1.01	1.01	1.01
4	0.2	1.0	3	6	1.3830	5.2804	2.0539	7	2.60	2.51	2.46	2.42	2.40	2.40
	1.0	2.0	3	3	0.9674	4.9920	2.0086	2	1.00	1.00	1.00	1.00	1.00	1.00
5	0.2	1.0	3	8	1.1503	5.0443	2.0618	6	2.05	2.00	1.98	1.96	1.96	1.95
	1.0	2.0	4	3	0.9674	4.9920	2.0363	2	1.00	1.00	1.00	1.00	1.00	1.00
6	0.2	1.0	4	9	1.2206	5.1630	2.0210	5	1.73	1.71	1.69	1.68	1.68	1.67
	1.0	2.0	5	3	0.9674	4.9920	2.0527	2	1.00	1.00	1.00	1.00	1.00	1.00

3. Results and Discussion

In practice, the exact shift size of a process is unknown. In this situation, if the corresponding optimal charting parameters are employed based on a particular shift size, the performance of the control chart will be significantly different if different shift occurred in the process. Therefore, it is essential to evaluate the performance of the SDS chart using alternative performance measure, i.e., EARL. In this paper, the optimal charting parameters $(n_1, n_2, L, L_1, L_2, L_3)$ of the SDS chart were computed using a nonlinear minimisation problem, i.e., optimal statistical design that minimises the out-of-control EARL ($EARL_1$). The programmes are written in the ScicosLab software version 4.4.2 (http://www.scicoslab.org).

The optimal charting parameters and the corresponding $EARL_1$s with different combinations of sample size, $n = \{3, 4, 5, 6\}$, the number of Phase-I samples, $m = \{30, 50, 80, 200, 500, +\infty\}$ with $(\delta_{min}, \delta_{max}) = (0.2, 1.0)$ and $(\delta_{min}, \delta_{max}) = (1.0, 2.0)$ with the in-control EARL, i.e., $EARL_0 = 370.4$, are presented in Table 1. Here, $m = +\infty$ denotes the known process parameters case, while $m = \{30, 50, 80, 200, 500\}$ denotes the estimated process parameters case. The performance of the SDS chart for both the known and estimated process parameters cases was calculated using the optimal charting parameters in columns 4–9, which were obtained by minimising $EARL_1$ when the process parameters are known.

From Table 1, for the same n, δ_{min}, and δ_{max}, the value of $EARL_1$ decreases with the increases in m. This is due to the fact that as the more Phase-I samples are taken, the performance of the estimated process parameters SDS chart approaches to the corresponding chart with known process parameters; i.e., $EARL_1$ value decreases to indicate better performance. For instance, when $n = 3$, $\delta_{min} = 0.2$, and $\delta_{max} = 1.0$, the optimal charting parameters $(n_1, n_2, L, L_1, L_2, L_3) = (2, 6, 1.3830, 5.2804, 2.0239, 9)$ yield the lowest $EARL_1 = 3.13$, when the process parameters are known. With these optimal charting parameters, $EARL_1 = \{3.56, 3.36, 3.26, 3.18, 3.15\}$ for $m = \{30, 50, 80, 200, 500\}$, respectively. It is noticeable that the $EARL_1$ for the estimated process parameters case is deviated from the known process parameters. However, the $EARL_1$ value approaches to the $EARL_1$ value which corresponds to the $m = +\infty$ when

the number of Phase-I samples increased. These findings show that more than 80 Phase-I samples are required to reduce the effects of process parameters estimation when estimating the process parameters from the in-control Phase-I samples.

To illustrate the implementation of the proposed optimal charting parameters, Table 2 presents the optimal charting parameters for the known process parameters SDS chart based on minimising out-of-control ARL (ARL_1) and the corresponding ARL_1s for the same combinations of (m, n) in Table 1. Here, the in-control ARL, i.e., $ARL_0 = 370.4$ is intended. Note that $\delta = \{0.2, 0.5, 0.9, 1.2, 1.5, 1.9\}$ are considered here to accommodate the $(\delta_{min}, \delta_{max})$ that are considered in Table 1; i.e., $\delta \in \{0.2, 0.5, 0.9\}$ and $\delta \in \{1.2, 1.5, 1.9\}$ are included in $(\delta_{min}, \delta_{max}) = (0.2, 1.0)$ and $(\delta_{min}, \delta_{max}) = (1.0, 2.0)$, respectively. In Table 1, when $n = 3$, $\delta_{min} = 0.2$, and $\delta_{max} = 1.0$, the $EARL_1 = 3.13$ is obtained using the optimal charting parameters $(n_1, n_2, L, L_1, L_2, L_3) = (2, 6, 1.3830, 5.2804, 2.0239, 9)$. Here, by using the same optimal charting parameters for $\delta = 0.5$ (i.e. $\delta \in (\delta_{min}, \delta_{max})$), it yields $ARL_1 = \{17.76, 14.32, 12.89, 11.72, 11.32\}$ when $m = \{30, 50, 80, 200, 500\}$ using the ScicolsLab program. It is observed that the ARL_1 value is almost the same to those in Table 2 when $n = 3$ and $\delta = 0.5$, although the optimal charting parameters based on minimising ARL_1 are different, i.e., $(n_1, n_2, L, L_1, L_2, L_3) = (2, 6, 1.3830, 5.2804, 2.1867, 18)$ (see Table 2). This indicates that the optimal charting parameters obtained based on minimising $EARL_1$ can be employed as long as $\delta \in (\delta_{min}, \delta_{max})$, i.e., when the practitioners do not have knowledge to determine the exact process shift size in advance.

4. Conclusions

In the production and manufacturing industries, it is a typical situation where quality practitioners are undecided about the process shift size to be implemented. The findings showed that the performance criterion EARL is capable of tackling the random shift size situation. Furthermore, the results also revealed that the performance of the SDS chart was adversely affected by process parameters estimation. This was proven when more than 80 Phase-I samples were needed for the chart

Table 2: Optimal charting parameters $(n_1, n_2, L, L_1, L_2, L_3)$ and the corresponding ARL_1s for $n = \{3, 4, 5, 6\}$ with different combinations of (m, δ) when $\mathrm{ARL}_0 = 370.4$.

| n | δ | n_1 | n_2 | L | L_1 | L_2 | L_3 | m | | | | | |
								30	50	80	200	500	$+\infty$
	0.2	2	6	1.3830	5.2804	2.4572	68	247.22	168.88	136.75	110.56	101.56	96.01
	0.5	2	6	1.3830	5.2804	2.1867	18	16.68	13.35	12.03	10.99	10.63	10.41
3	0.9	2	6	1.3830	5.2804	1.9945	8	2.83	2.72	2.67	2.62	2.60	2.59
	1.2	2	4	1.1503	5.0443	2.0178	4	1.64	1.62	1.60	1.59	1.58	1.58
	1.5	2	3	0.9674	4.9920	2.0523	3	1.24	1.23	1.23	1.22	1.22	1.22
	1.9	2	3	0.9674	4.9920	2.0523	3	1.06	1.06	1.06	1.05	1.05	1.05
	0.2	3	8	1.5341	5.1956	2.3954	60	149.45	113.15	95.73	80.45	75.01	71.62
	0.5	3	8	1.5341	5.1956	2.0910	15	9.65	8.45	7.93	7.49	7.33	7.23
4	0.9	3	6	1.3830	5.2804	2.0153	6	2.12	2.07	2.04	2.02	2.01	2.00
	1.2	3	4	1.1503	5.0443	2.0185	3	1.34	1.33	1.32	1.32	1.32	1.31
	1.5	3	3	0.9674	4.9920	2.0086	2	1.10	1.10	1.10	1.09	1.09	1.09
	1.9	3	3	0.9674	4.9920	2.0086	2	1.01	1.01	1.01	1.01	1.01	1.01
	0.2	4	10	1.6449	5.1247	2.3394	55	108.31	84.83	73.06	62.63	58.92	56.60
	0.5	3	10	1.2816	5.1041	2.1216	12	6.64	6.05	5.78	5.54	5.46	5.40
5	0.9	3	8	1.1503	5.0443	2.0186	5	1.73	1.71	1.69	1.68	1.68	1.67
	1.2	4	4	1.1503	5.0443	2.0615	3	1.20	1.19	1.19	1.18	1.18	1.18
	1.5	4	4	1.1503	5.0443	1.9647	2	1.05	1.04	1.04	1.04	1.04	1.04
	1.9	4	3	0.9674	4.9920	2.0363	2	1.00	1.00	1.00	1.00	1.00	1.00
	0.2	4	12	1.3830	5.2804	2.3744	44	84.09	66.77	58.05	50.35	47.62	45.91
	0.5	4	12	1.3830	5.2804	2.0727	11	5.02	4.69	4.54	4.40	4.35	4.32
6	0.9	4	8	1.1503	5.0443	2.0178	4	1.50	1.48	1.47	1.46	1.46	1.46
	1.2	5	5	1.2816	5.1041	2.0235	3	1.12	1.11	1.11	1.11	1.11	1.11
	1.5	5	4	1.1503	5.0443	1.9954	2	1.02	1.02	1.02	1.02	1.02	1.02
	1.9	5	3	0.9674	4.9920	2.0527	2	1.00	1.00	1.00	1.00	1.00	1.00

with the estimated process parameters to behave similarly like the one with known process parameters. Therefore, future research works can consider the optimal charting parameters by minimising EARL_1 for the SDS chart when the process parameters are estimated.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] S. Shang, Q. Zhou, M. Liu, and Y. Shao, "Sample size calculation for controlling false discovery proportion," *Journal of Probability and Statistics*, vol. 2012, Article ID 817948, 13 pages, 2012.

[2] D. C. Montgomery, *A Modern Introduction*, John Wiley & Sons, New York, NY, USA, 7th edition, 2013.

[3] M. B. C. Khoo, H. C. Lee, Z. Wu, C.-H. Chen, and P. Castagliola, "A synthetic double sampling control chart for the process mean," *IIE Transactions*, vol. 43, no. 1, pp. 23–38, 2010.

[4] N. A. Saleh and M. A. Mahmoud, "Accounting for phase I sampling variability in the performance of the MEWMA control chart with estimated parameters," *Communications in Statistics—Simulation and Computation*, vol. 46, no. 6, pp. 4333–4347, 2017.

[5] W. H. Woodall and D. C. Montgomery, "Some current directions in the theory and application of statistical process monitoring," *Journal of Quality Technology*, vol. 46, no. 1, pp. 78–94, 2014.

[6] S. Psarakis, A. K. Vyniou, and P. Castagliola, "Some recent developments on the effects of parameter estimation on control charts," *Quality and Reliability Engineering International*, vol. 30, no. 8, pp. 1113–1129, 2014.

[7] P. Castagliola, P. E. Maravelakis, and F. O. Figueiredo, "The EWMA median chart with estimated parameters," *Institute of Industrial Engineers (IIE). IIE Transactions*, vol. 48, no. 1, pp. 66–74, 2016.

[8] S. Du, X. Yao, and D. Huang, "Engineering model-based Bayesian monitoring of ramp-up phase of multistage manufacturing process," *International Journal of Production Research*, vol. 53, no. 15, pp. 4594–4613, 2015.

[9] S. Chakraborti, "Run length distribution and percentiles: The shewhart X̄chart with unknown parameters," *Quality Engineering*, vol. 19, no. 2, pp. 119–127, 2007.

[10] H. W. You, M. B. C. Khoo, M. H. Lee, and P. Castagliola, "Synthetic double sampling X chart with estimated process parameters," *Quality Technology and Quantitative Management*, vol. 12, no. 4, pp. 579–604, 2016.

[11] Y. Ou, Z. Wu, K. M. Lee, and S. Chen, "An optimal design algorithm of the SPRT chart for minimizing weighted ATS," *International Journal of Production Economics*, vol. 139, no. 2, pp. 564–574, 2012.

[12] W. L. Teoh, J. K. Chong, M. B. C. Khoo, P. Castagliola, and W. C. Yeong, "Optimal Designs of the Variable Sample Size X̄Chart Based on Median Run Length and Expected Median Run Length," *Quality and Reliability Engineering International*, vol. 33, no. 1, pp. 121–134, 2017.

Permissions

The contributors of this book come from diverse backgrounds, making this book a truly international effort. This book will bring forth new frontiers with its revolutionizing research information and detailed analysis of the nascent developments around the world.

We would like to thank all the contributing authors for lending their expertise to make the book truly unique. They have played a crucial role in the development of this book. Without their invaluable contributions this book wouldn't have been possible. They have made vital efforts to compile up to date information on the varied aspects of this subject to make this book a valuable addition to the collection of many professionals and students.

This book was conceptualized with the vision of imparting up-to-date information and advanced data in this field. To ensure the same, a matchless editorial board was set up. Every individual on the board went through rigorous rounds of assessment to prove their worth. After which they invested a large part of their time researching and compiling the most relevant data for our readers.

The editorial board has been involved in producing this book since its inception. They have spent rigorous hours researching and exploring the diverse topics which have resulted in the successful publishing of this book. They have passed on their knowledge of decades through this book. To expedite this challenging task, the publisher supported the team at every step. A small team of assistant editors was also appointed to further simplify the editing procedure and attain best results for the readers.

Apart from the editorial board, the designing team has also invested a significant amount of their time in understanding the subject and creating the most relevant covers. They scrutinized every image to scout for the most suitable representation of the subject and create an appropriate cover for the book.

The publishing team has been an ardent support to the editorial, designing and production team. Their endless efforts to recruit the best for this project, has resulted in the accomplishment of this book. They are a veteran in the field of academics and their pool of knowledge is as vast as their experience in printing. Their expertise and guidance has proved useful at every step. Their uncompromising quality standards have made this book an exceptional effort. Their encouragement from time to time has been an inspiration for everyone.

The publisher and the editorial board hope that this book will prove to be a valuable piece of knowledge for researchers, students, practitioners and scholars across the globe.

List of Contributors

Nelson Christopher Dzupire
Pan African University Institute of Basic Sciences, Technology and Innovation, Juja, Kenya

Philip Ngare
Pan African University Institute of Basic Sciences, Technology and Innovation, Juja, Kenya University of Nairobi, Nairobi, Kenya

Leo Odongo
Pan African University Institute of Basic Sciences, Technology and Innovation, Juja, Kenya Kenyatta University, Nairobi, Kenya

Mahdi Teimouri, Saeid Rezakhah and Adel Mohammadpour
Department of Statistics, Faculty of Mathematics and Computer Science, Amirkabir University of Technology (Tehran Polytechnic), 424 Hafez Ave., Tehran 15914, Iran

S. D. Krishnarani
Department of Statistics, Farook College, Kozhikode, Kerala 673632, India

Arjun K. Gupta
Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403-0221, USA

Daya K. Nagar and Luz Estela Sánchez
Instituto de Matemáticas, Universidad de Antioquia, Calle 67, No. 53–108, Medellín, Colombia

Suleman Nasiru
Department of Statistics, Faculty of Mathematical Sciences, University for Development Studies, Tamale, Ghana

Yingyu Zhu and Simon Li
Department of Mechanical and Manufacturing Engineering, University of Calgary, Alberta, Canada

José Alfredo Jiménez and Viswanathan Arunachalam
Department of Statistics, Universidad Nacional de Colombia, Carrera 45 No. 26-85, Bogotá, Colombia

Ali Alkenani and Tahir R. Dikheel
Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Al Diwaniyah, Iraq

Fathi M. O. Hamed
University of Benghazi, Benghazi, Libya

Robert G. Aykroyd
University of Leeds, Leeds, UK

Simona Toti and Romina Filippini
Istituto Nazionale di Statistica (ISTAT), Via Cesare Balbo 16, 00184 Rome, Italy

Filippo Palombi
Istituto Nazionale di Statistica (ISTAT), Via Cesare Balbo 16, 00184 Rome, Italy
Italian Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Via Enrico Fermi 45, 00044 Frascati, Italy

Mursala Khan
Department of Mathematics, COMSATS Institute of Information Technology, Abbottabad 22060, Pakistan

Manickavasagar Kayanan
Department of Physical Science, Vavuniya Campus, University of Jaffna, Vavuniya, Sri Lanka
Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka

Pushpakanthie Wijekoon
Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka

Salwa Waeto and Arthit Intarasit
Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Pattani 94000,Thailand
Centre of Excellence in Mathematics, Commission on Higher Education, Ratchathewi, Bangkok 10400,Thailand

Khanchit Chuarkham
Faculty of Commerce and Management, Prince of Songkla University, Trang Campus, Trang 92000, Thailand

S. J. Dilworth
Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA

S. R. Mane
Convergent Computing Inc., Shoreham, NY 11786, USA

Kaisar Ahmad and S. P. Ahmad
Department of Statistics, University of Kashmir, Srinagar, Jammu and Kashmir 190006, India

A. Ahmed
Department of Statistics and O.R., Aligarh Muslim University, Aligarh, India

Masood Anwar and Amna Bibi
Department of Mathematics, COMSATS Institute of Information Technology, Park Road, Chak Shahzad, Islamabad, Pakistan

Jason Chin-Tiong Chan
Ted Rogers School of Management, Ryerson University, 350 Victoria St., Toronto, ON, Canada M5B 2K3

Hong Choon Ong
School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Gelugor, Penang, Malaysia

C. S. Marange and Y. Qin
Department of Statistics, Faculty of Science and Agriculture, Fort Hare University, East London Campus, 5200, South Africa

Jung-In Seo
Department of Statistics, Daejeon University, No. 62, Daehak-ro, Dong-gu, Republic of Korea

Jae-Woo Jeon and Suk-Bok Kang
Department of Statistics, Yeungnam University, No. 280, Daehak-ro, Gyeongsan, Republic of Korea

Huay Woon You
Pusat PERMATApintar Negara, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Index