ISSN 1645-6726

REVSTAT

Statistical Journal

vol. 23 - n. 3 - July 2025

Instituto Nacional de Estatística
Statistics Portugal
1935-2025

REVSTAT — Statistical Journal, vol.23, n. 3 (July 2025)

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

Continues: Revista de Estatística = ISSN 0873-4275.

ISSN 1645-6726 ; e-ISSN 2183-0371

Editorial Board (2024-2025)

Publisher – Statistics Portugal

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

Digital edition

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

INDEX

Abstracted/indexed in: *Journal Citation Reports™ – JCR (Clarivate); DOAJ – Directory of Open Access Journals; Current Index to Statistics; Google Scholar; Mathematical Reviews® (MathSciNet®); Zentralblatt für Mathematic; SCImago Journal & Country Rank; Scopus®.*

REVSTAT – Statistical Journal
Volume 23, Number 3, July 2025, 301–323
https://doi.org/10.57805/revstat.v23i3.568

A Bivariate Weibull Distribution and Its Applications in Reliability

Authors: Katherine Davies ✉
– Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada
daviek5@mcmaster.ca

Debanjan Mitra
– Operations Management, Quantitative Methods and Information Systems, Indian Institute of Management Udaipur, Rajasthan, India
debanjan.mitra@iimu.ac.in

William Volterman
– Economical Insurance, Kitchener, ON, Canada
william.volterman@gmail.com

Received: April 2022 Revised: June 2023 Accepted: June 2023

Abstract:

- In this article, the bivariate exponential distribution proposed by Downton ("Bivariate exponential distributions in reliability theory", *Journal of the Royal Statistical Society Series B*, 1970) is extended to a bivariate Weibull distribution, and it is called the Downton's bivariate Weibull (DBW) distribution. Statistical properties of the DBW distribution are explored and likelihood inference developed based on complete as well as right-censored bivariate data are discussed. Through extensive Monte Carlo simulations, performance of the point and interval estimates are evaluated. Two real datasets are analyzed for illustrative purposes. It is concluded that the DBW distribution is very useful to model bivariate data.

Keywords:

- *bivariate lifetime data; bivariate Weibull distribution; maximum likelihood estimation; asymptotic confidence intervals; right censoring; moment-based estimates.*

AMS Subject Classification:

- 49A05, 78B26.

✉ Corresponding author

1. INTRODUCTION

Lifetime data can often be two dimensional. For example, Jung and Bai (2007) analysed field data under two-dimensional warranty, using bivariate data on age and mileage of cars. Bivariate life data pertaining to trucks working in the mines were analysed by Fuqing *et al.* (2017). In such applications, one of the main goals of analyses is to capture the dependence between components of the two-dimensional lifetimes. In some cases, the dependence is captured by using an approach based on copulas; see, for example Fuqing *et al.* (2017). In some other cases, particular bivariate distributions are used; for example, see Yuan (2016), among others.

In this article, we present and discuss a bivariate Weibull distribution that can be conveniently used for modelling two-dimensional lifetime data. The distribution presented here is flexible, computationally convenient, and has desirable properties that can be exploited for modelling purposes.

A bivariate exponential distribution, known as the Downton's bivariate exponential (DBE) distribution (Downton, 1970), is as follows. If a random vector (Y_1, Y_2) follows the DBE distribution, the corresponding joint probability density function (PDF) is given by

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{\theta_1\theta_2(1-\rho)} \exp\left\{-\frac{1}{1-\rho}\left(\frac{y_1}{\theta_1}+\frac{y_2}{\theta_2}\right)\right\} I_0\left(\frac{2(\rho y_1 y_2)^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right), \quad y_1, y_2 > 0, \tag{1.1}$$

with $\theta_1, \theta_2 > 0$ and $0 \le \rho < 1$. We express this as $(Y_1, Y_2) \sim \text{DBE}(\theta_1, \theta_2, \rho)$. Here, $I_0(\cdot)$ is the modified Bessel function of the first kind of order zero. Downton (1970) explained this model from the perspective of a failure model in the reliability context; see also Kotz *et al.* (2000) for a comprehensive description.

The DBE distribution has some excellent mathematical properties. For example, the DBE distribution originates from a shock model where two components of a system are subject to independent streams of shocks with exponentially distributed interarrival times. Also, the number of such shocks to cause failure is a random variable distributed according to the geometric distribution. For the DBE distribution, the marginal distributions of Y_1 and Y_2 are exponential, with scale parameters θ_1 and θ_2, respectively. The dependence between Y_1 and Y_2 is captured through the correlation parameter ρ; when $\rho = 0$, Y_1 and Y_2 are independent. The model is quite convenient to be used in reliability modelling due to its tractable mathematical nature. Especially, as the corresponding marginal distributions of Y_1 and Y_2 are univariate exponential distributions, the DBE distribution is particularly useful in situations where the hazard rates of the marginal distributions can be assumed (or verified) to be constants.

However, for bivariate lifetime data with non-constant marginal hazard rates, the DBE distribution is clearly not suitable. In such scenarios, it is more appropriate to model the bivariate life data by using a bivariate distribution that has univariate Weibull marginals, which can accommodate the non-constant marginal hazard rates. The motivation of this work is to develop a model that can accommodate non-constant hazard rates of the marginal lifetimes corresponding to the two components of a system, within the same context of arrival of independent streams of shocks to the two components. To achieve this, we have consider a non-linear transformation of the marginal lifetimes as described below. The resulting model is more appropriate for real datasets having non-constant marginal hazard rates.

Therefore, in this article, we discuss a generalization of the DBE distribution to a bivariate Weibull distribution which we call the Downton's bivariate Weibull (DBW) distribution. The DBW distribution partially retains the mathematical advantages of the DBE distribution with respect to its analytical tractability. In fact, the DBW distribution loses some of the interpretations of the DBE distribution, such as exponential interarrival times of the independent shock. However, as the DBW distribution has Weibull marginals, it is very flexible for modelling purposes; for example, it can accommodate bivariate lifetime data with non-constant marginal hazard rates.

In this article, we discuss the DBW distribution, its statistical properties, and its applications in detail. First, after constructing the DBW distribution, we explore some of its important properties which will be helpful in applying the distribution to real data. Then, we discuss likelihood inference for the DBW distribution, based on complete as well as right-censored lifetime data. Moment-based estimators for the distribution based on complete data are also discussed. The inferential methods are examined through extensive Monte Carlo simulation studies. Then, as an application, a two-dimensional warranty model is discussed. Analyses of two real datasets for illustrative purposes are presented. Based on our exploration, we conclude that the DBW distribution has desirable statistical properties, is computationally convenient, and is quite flexible for modelling purposes. We recommend its extensive use in modelling bivariate data, especially arising from lifetime experiments.

The paper is organized as follows. In Section 2, the DBW distribution with its statistical properties, and data generation algorithm are presented. Likelihood inference based on complete data from the DBW distribution, and a Monte Carlo simulation study to evaluate the point and interval estimates are presented in Section 3. This section also contains discussion of a procedure for obtaining moment-based estimators. In Section 4, inferences for the DBW distribution based on bivariate right-censored data are discussed, and a simulation study examining the performance of the estimates are presented. An application in the form of a two-dimensional warranty model is presented in Section 5. Two case studies based on real datasets are provided in Section 6 for illustrative purposes. Finally, the paper is concluded with some remarks in Section 7.

2. DOWNTON'S BIVARIATE WEIBULL DISTRIBUTION

Suppose $(Y_1, Y_2) \sim \text{DBE}(\theta_1, \theta_2, \rho)$. Consider the transformations $X_1 = Y_1^{\alpha_1}$ and $X_2 = Y_2^{\alpha_2}$, with $\alpha_1, \alpha_2 > 0$. The joint density of the transformed random variables X_1 and X_2 is then given by

(2.1)
$$f_{X_1,X_2}(x_1,x_2) = \frac{\eta_1\eta_2 x_1^{\eta_1-1}x_2^{\eta_2-1}}{\theta_1\theta_2(1-\rho)} \exp\left\{-\frac{1}{1-\rho}\left(\frac{x_1^{\eta_1}}{\theta_1}+\frac{x_2^{\eta_2}}{\theta_2}\right)\right\} I_0\left(\frac{2(\rho x_1^{\eta_1}x_2^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right), \quad x_1, x_2 > 0,$$

with $\theta_1, \theta_2, \eta_1, \eta_2 > 0$ and $0 \le \rho < 1$, where $\eta_1 = 1/\alpha_1$ and $\eta_2 = 1/\alpha_2$. Thus, through this power transformation of the variables Y_1 and Y_2, we obtain a bivariate distribution which we call the Downton's bivariate Weibull (DBW) distribution. We write this as $(X_1, X_2) \sim \text{DBW}(\eta_1, \theta_1, \eta_2, \theta_2, \rho)$. A similar distribution, called the Nakagami-m distribution, was discussed by Sagias and Karagiannidis (2005) in the context of modelling fading channels relating to digital communication systems. However, Sagias and Karagiannidis (2005) derived this distribution starting from a bivariate Rayleigh distribution.

Using an infinite series representation of the Bessel function provided in Gradshteyn and Ryzhik (2000) as

$$I_0(z) = \sum_{k=0}^{\infty} \frac{1}{(k!)^2} \left(\frac{z}{2}\right)^{2k},$$

we can obtain another representation of the PDF of the DBW distribution as

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= \eta_1\eta_2 \exp\left\{ -\frac{1}{1-\rho}\left(\frac{x_1^{\eta_1}}{\theta_1} + \frac{x_2^{\eta_2}}{\theta_2}\right)\right\} \\ &\quad \times \sum_{k=0}^{\infty} \frac{\rho^k}{(k!)^2(1-\rho)^{2k+1}} \frac{x_1^{-1+(k+1)\eta_1} x_2^{-1+(k+1)\eta_2}}{(\theta_1\theta_2)^{k+1}}, \quad x_1, x_2 > 0. \end{aligned} \tag{2.2}$$

The corresponding joint cumulative density function (CDF) of X_1 and X_2 can be obtained as

$$\begin{aligned} F_{X_1,X_2}(x_1,x_2) &= 1 - \exp\left(-\frac{x_1^{\eta_1}}{\theta_1}\right) Q_1\left(\sqrt{\frac{2}{1-\rho}}\frac{x_2^{\eta_2/2}}{\sqrt{\theta_2}}, \sqrt{\frac{2\rho}{1-\rho}}\frac{x_1^{\eta_1/2}}{\sqrt{\theta_1}}\right) \\ &\quad - \exp\left(-\frac{x_2^{\eta_2}}{\theta_2}\right)\left[1 - Q_1\left(\sqrt{\frac{2\rho}{1-\rho}}\frac{x_2^{\eta_2/2}}{\sqrt{\theta_2}}, \sqrt{\frac{2}{1-\rho}}\frac{x_1^{\eta_1/2}}{\sqrt{\theta_1}}\right)\right], \end{aligned} \tag{2.3}$$

where $Q_1(\cdot,\cdot)$ is the first order Marcum's Q-function (Marcum, 1950). The bivariate survival function of X_1 and X_2 is given by

$$\begin{aligned} S_{X_1,X_2}(x_1,x_2) &= 1 - F_{X_1}(x_1) - F_{X_2}(x_2) + F_{X_1,X_2}(x_1,x_2) \\ &= \exp\left(-\frac{x_1^{\eta_1}}{\theta_1}\right)\left[1 - Q_1\left(\sqrt{\frac{2}{1-\rho}}\frac{x_2^{\eta_2/2}}{\sqrt{\theta_2}}, \sqrt{\frac{2\rho}{1-\rho}}\frac{x_1^{\eta_1/2}}{\sqrt{\theta_1}}\right)\right] \\ &\quad + \exp\left(-\frac{x_2^{\eta_2}}{\theta_2}\right) Q_1\left(\sqrt{\frac{2\rho}{1-\rho}}\frac{x_2^{\eta_2/2}}{\sqrt{\theta_2}}, \sqrt{\frac{2}{1-\rho}}\frac{x_1^{\eta_1/2}}{\sqrt{\theta_1}}\right), \end{aligned} \tag{2.4}$$

where F_{X_1} and F_{X_2} are the marginal CDFs of X_1 and X_2, respectively. In Figures 1 and 2, the joint density function is plotted for different values of the parameters.

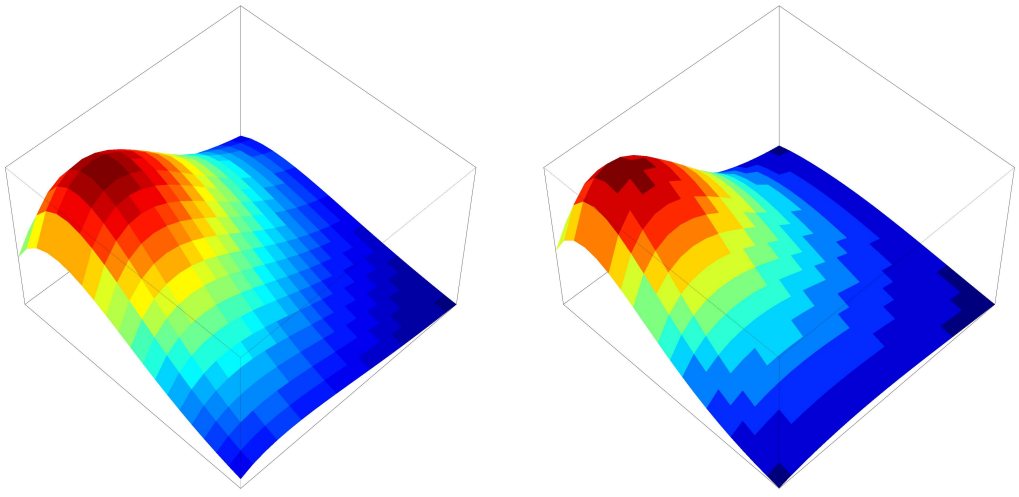

Figure 1: Plot of joint density: left: $\eta_1 = \eta_2 = 1.5$, $\theta_1 = \theta_2 = 1$, $\rho = 0$; right: $\eta_1 = \eta_2 = 1.5$, $\theta_1 = \theta_2 = 1$, $\rho = 0.25$.

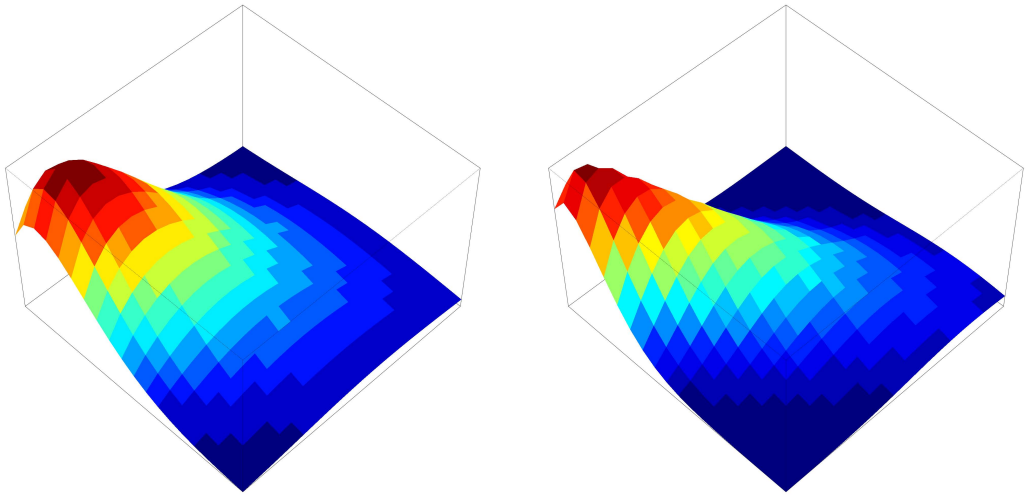

Figure 2: Plot of joint density: left: $\eta_1 = \eta_2 = 1.5$, $\theta_1 = \theta_2 = 1$, $\rho = 0.5$; right: $\eta_1 = \eta_2 = 1.5$, $\theta_1 = \theta_2 = 1$, $\rho = 0.75$.

2.1. Some properties

If $(X_1, X_2) \sim \text{DBW}(\eta_1, \theta_1, \eta_2, \theta_2, \rho)$, the marginal distributions of X_1 and X_2 are univariate Weibull with parameters (η_1, θ_1) and (η_2, θ_2), respectively, with PDF

$$f_{X_i}(x_i) = \frac{\eta_i}{\theta_i} x_i^{\eta_i - 1} \exp\left\{ -\frac{x_i^{\eta_i}}{\theta_i} \right\}, \quad x_i > 0, \tag{2.5}$$

and CDF

$$F_{X_i}(x_i) = 1 - \exp\left\{ -\frac{x_i^{\eta_i}}{\theta_i} \right\}, \quad x_i > 0, \tag{2.6}$$

for $i = 1, 2$.

The conditional distributions can be obtained by using the joint density and the marginal densities. If $(X_1, X_2) \sim \text{DBW}(\eta_1, \theta_1, \eta_2, \theta_2, \rho)$, then the PDFs of the conditional distributions of X_1 given $X_2 = x_2$, and X_2 given $X_1 = x_1$ are given by

$$f_{X_1|X_2=x_2}(x_1) = \frac{\eta_1 x_1^{\eta_1 - 1}}{\theta_1(1-\rho)} \exp\{-(A_1 + A_2)\} \sum_{k=0}^{\infty} \frac{1}{(k!)^2} \frac{\rho^k}{(1-\rho)^{2k}} \left(\frac{x_1^{\eta_1} x_2^{\eta_2}}{\theta_1 \theta_2} \right)^k, \quad x_1 > 0, \tag{2.7}$$

and

$$f_{X_2|X_1=x_1}(x_2) = \frac{\eta_2 x_2^{\eta_2 - 1}}{\theta_2(1-\rho)} \exp\{-(B_1 + B_2)\} \sum_{k=0}^{\infty} \frac{1}{(k!)^2} \frac{\rho^k}{(1-\rho)^{2k}} \left(\frac{x_1^{\eta_1} x_2^{\eta_2}}{\theta_1 \theta_2} \right)^k, \quad x_2 > 0, \tag{2.8}$$

respectively. The CDFs of the conditional distributions of X_1 given $X_2 = x_2$, and X_2 given $X_1 = x_1$ are given by

$$F_{X_1|X_2=x_2}(x_1) = \exp\{-A_2\} \times \sum_{k=0}^{\infty} \frac{1}{(k!)^2} A_2^k \gamma(k+1, A_1), \quad x_1 > 0, \tag{2.9}$$

and

$$F_{X_2|X_1=x_1}(x_2) = \exp\{-B_2\} \times \sum_{k=0}^{\infty} \frac{1}{(k!)^2} B_2^k \gamma(k+1, B_1), \quad x_2 > 0, \tag{2.10}$$

respectively, where $\gamma(s,t) = \int_0^t u^{s-1}e^{-u}du$ is the lower incomplete gamma function, and

$$A_1 = \frac{1}{1-\rho}\frac{x_1^{\eta_1}}{\theta_1},\ A_2 = \frac{\rho}{1-\rho}\frac{x_2^{\eta_2}}{\theta_2},\ B_1 = \frac{1}{1-\rho}\frac{x_2^{\eta_2}}{\theta_2},\ B_2 = \frac{\rho}{1-\rho}\frac{x_1^{\eta_1}}{\theta_1}.$$

Alternative expressions involving the modified Bessel function for the conditional PDFs and CDFs can be obtained. If $(X_1, X_2) \sim \text{DBW}(\eta_1, \theta_1, \eta_2, \theta_2, \rho)$, then alternative expressions for the PDFs of the conditional distributions of X_1 given $X_2 = x_2$, and X_2 given $X_1 = x_1$, respectively, are given by

$$f_{X_1|X_2=x_2}(x_1) = \frac{\eta_1 x_1^{\eta_1-1}}{\theta_1(1-\rho)} \exp\big\{-(A_1+A_2)\big\}\, I_0\left(\frac{2(\rho x_1^{\eta_1}x_2^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right), \quad x_1 > 0, \tag{2.11}$$

and

$$f_{X_2|X_1=x_1}(x_2) = \frac{\eta_2 x_2^{\eta_2-1}}{\theta_2(1-\rho)} \exp\big\{-(B_1+B_2)\big\}\, I_0\left(\frac{2(\rho x_1^{\eta_1}x_2^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right), \quad x_2 > 0. \tag{2.12}$$

Similarly, alternative expressions for the CDFs of the conditional distributions of X_1 given $X_2 = x_2$, and X_2 given $X_1 = x_1$ are given by

$$\begin{aligned} F_{X_1|X_2=x_2}(x_1) &= A_1 \exp(-A_2) \int_0^1 e^{-A_1 t} I_0(2\sqrt{A_1A_2 t})dt, \\ F_{X_2|X_1=x_1}(x_2) &= B_1 \exp(-B_2) \int_0^1 e^{-B_1 t} I_0(2\sqrt{B_1B_2 t})dt, \end{aligned}$$

respectively, which can be shown as follows. Note that we can write

$$\begin{aligned} F_{X_1|X_2=x_2}(x_1) &= \exp\{-A_2\} \sum_{k=0}^{\infty} \frac{1}{(k!)^2} A_2^k \gamma(k+1, A_1) \\ &= \exp\{-A_2\} \sum_{k=0}^{\infty} \frac{1}{(k!)^2} A_2^k \left[k! \exp\{-A_1\} A_1^{k+1} \sum_{l=0}^{\infty} \frac{A_1^l}{(l+k+1)!}\right] \\ &= A_1 \exp\{-(A_1+A_2)\} \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \frac{(1)_l}{(2)_{l+k}} \frac{(A_1A_2)^k}{k!} \frac{A_1^l}{l!} \\ &= A_1 \exp\{-(A_1+A_2)\} \Phi_3(b=1, c=2, w=A_1, z=A_1A_2) \\ &= A_1 \exp\{-(A_1+A_2)\} \left[\exp\{A_1\} \int_0^1 e^{-A_1 t} I_0(2\sqrt{A_1A_2 t})dt\right] \\ &= A_1 \exp\{-A_2\} \int_0^1 e^{-A_1 t} I_0(2\sqrt{A_1A_2 t})dt, \end{aligned}$$

where

$$\Phi_3(b, c, w, z) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \frac{(b)_k}{(c)_{k+l}} \frac{w^k z^l}{k!l!}$$

is the Humbert function or the confluent Appell function (Brychkov, 2017).

The correlation parameter ρ captures the dependence between X_1 and X_2. If $(X_1, X_2) \sim \text{DBW}(\eta_1, \theta_1, \eta_2, \theta_2, \rho)$, X_1 and X_2 are independent if and only if $\rho = 0$. When $0 < \rho < 1$, X_1 and X_2 are correlated. However, note that ρ is not the Pearson's correlation coefficient between X_1 and X_2 for the DBW distribution.

Following Sagias and Karagiannidis (2005), an expression for the product moments of the DBW distribution, in terms of the generalized hypergeometric function, can be obtained to be

$$E(X_1^{m_1} X_2^{m_2}) = (1-\rho)^{1+\frac{m_1}{\eta_1}+\frac{m_2}{\eta_2}} \theta_1^{\frac{m_1}{\eta_1}} \theta_2^{\frac{m_2}{\eta_2}} \Gamma\Big(1+\frac{m_1}{\eta_1}\Big) \Gamma\Big(1+\frac{m_2}{\eta_2}\Big) \times {}_2F_1\Big(1+\frac{m_1}{\eta_1}, 1+\frac{m_2}{\eta_2}; 1; \rho\Big),$$

where ${}_2F_1(\cdot;\cdot;\cdot)$ is the generalized hypergeometric function. Therefore, the product moment correlation coefficient of X_1 and X_2 is given by

$$\rho_{\text{PMC}} = \frac{\Gamma\big(1+\frac{1}{\eta_1}\big)\Gamma\big(1+\frac{1}{\eta_2}\big)\Big[(1-\rho)^{1+\frac{1}{\eta_1}+\frac{1}{\eta_2}} \times {}_2F_1\Big(1+\frac{1}{\eta_1}, 1+\frac{1}{\eta_2}; 1; \rho\Big) - 1\Big]}{\sqrt{\Gamma\big(1+\frac{2}{\eta_1}\big) - \Big\{\Gamma\big(1+\frac{1}{\eta_1}\big)\Big\}^2}\sqrt{\Gamma\big(1+\frac{2}{\eta_2}\big) - \Big\{\Gamma\big(1+\frac{1}{\eta_2}\big)\Big\}^2}}.$$

The expression for the product moment correlation ρ_{PMC} clearly shows that ρ is not the correlation coefficient between X_1 and X_2 for the DBW distribution and in fact only takes values in [0,1). Note that ρ_{PMC} depends only on η_1, η_2, and ρ; the scale parameters θ_1 and θ_2 have no impact on the correlation coefficient. Figure 3 gives a plot of different values of ρ and ρ_{PMC}, for different values of η_1 and η_2.

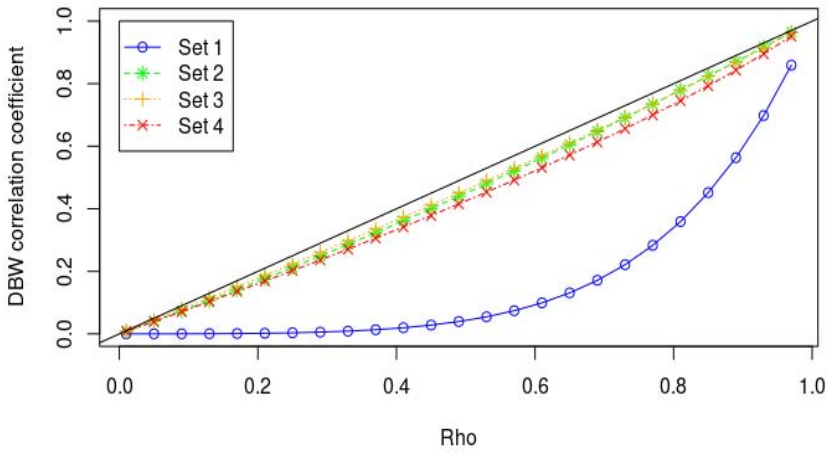

Figure 3: Plot of ρ and ρ_{PMC}; the values of the shape parameters for different sets are the following — Set 1: $\eta_1 = \eta_2 = 0.1$; Set 2: $\eta_1 = \eta_2 = 0.5$; Set 3: $\eta_1 = \eta_2 = 2.4$; Set 4: $\eta_1 = \eta_2 = 4.5$. The $y = x$ line is in black.

2.2. Data generation algorithm

Realisations from the DBW distribution can be generated using the marginal and conditional distributions. To generate an observation (x_1, x_2) from the DBW$(\eta_1, \theta_1, \eta_2, \theta_2, \rho)$ distribution, the simplest route would be to generate x_1 from the marginal Weibull distribution, and then for the given value x_1, to generate x_2 from its conditional distribution. The conditional distribution in equation (2.10) is computationally efficient in this regard. The process of generating random variables from the DBW distribution is as follows:

Algorithm 1:

Step 1: Generate x_1 from Weibull(η_1, θ_1).

Step 2: Transform x_1 into z_1, where $z_1 = \frac{\rho}{1-\rho}\frac{x_1^{\eta_1}}{\theta_1}$.

Step 3: Generate z_2 from the distribution with CDF

$$F_{Z_2|Z_1=z_1}(z_2) = z_2 \exp(-z_1) \int_0^1 e^{-z_2 t} I_0(2\sqrt{z_1 z_2 t})dt.$$

Step 4: Obtain x_2, using the transformation $x_2 = (\theta_2(1-\rho)z_2)^{1/\eta_2}$.

Note, however, that the above algorithm should only be used when $0 < \rho < 1$. For $\rho = 0$, x_1 and x_2 can be generated directly from their respective marginal distributions.

3. INFERENCE BASED ON COMPLETE BIVARIATE DATA

3.1. Likelihood inference

Based on observed bivariate data, maximum likelihood estimates (MLEs) for the parameters of the DBW distribution may be obtained. Let $(x_{1i}, x_{2i}), i = 1, ..., n$, denote the observed bivariate data. The likelihood function is given by

$$L(\boldsymbol{\omega}) = \prod_{i=1}^{n} f_{X_1,X_2}(x_{1i}, x_{2i}),$$

where $\boldsymbol{\omega} = (\eta_1, \theta_1, \eta_2, \theta_2, \rho)$ is the vector of model parameters. Using the joint PDF of X_1 and X_2 given in (2.1), the log-likelihood function is

$$\begin{aligned} \log L(\boldsymbol{\omega}) \;=\; & n\Big(\log\eta_1 + \log\eta_2 - \log\theta_1 - \log\theta_2 - \log(1-\rho)\Big) \\ & + (\eta_1 - 1)\sum_{i=1}^{n}\log x_{1i} + (\eta_2 - 1)\sum_{i=1}^{n}\log x_{2i} - \frac{1}{1-\rho}\sum_{i=1}^{n}\left(\frac{x_{1i}^{\eta_1}}{\theta_1} + \frac{x_{2i}^{\eta_2}}{\theta_2}\right) \\ & + \sum_{i=1}^{n}\log I_0\left(\frac{2(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right). \end{aligned} \tag{3.1}$$

The log-likelihood equations with respect to η_1, θ_1, η_2, θ_2, and ρ are as follows:

(3.2)
$$\frac{\partial \log L}{\partial \eta_1} = \frac{n}{\eta_1} + \sum_{i=1}^{n} \log x_{1i} - \frac{1}{1-\rho} \sum_{i=1}^{n} \frac{x_{1i} \log x_{1i}}{\theta_1} + \sum_{i=1}^{n} \frac{I_0'\left(\frac{2(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right)\left(\frac{(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right) \log x_{1i}}{I_0\left(\frac{2(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right)},$$

(3.3)
$$\frac{\partial \log L}{\partial \theta_1} = -\frac{n}{\theta_1} + \frac{1}{1-\rho} \sum_{i=1}^{n} \frac{x_{1i}^{\eta_1}}{\theta_1^2} - \sum_{i=1}^{n} \frac{I_0'\left(\frac{2(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right)\left(\frac{(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1^3\theta_2}}\right)}{I_0\left(\frac{2(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right)},$$

(3.4)
$$\frac{\partial \log L}{\partial \eta_2} = \frac{n}{\eta_2} + \sum_{i=1}^{n} \log x_{2i} - \frac{1}{1-\rho} \sum_{i=1}^{n} \frac{x_{2i} \log x_{2i}}{\theta_2} + \sum_{i=1}^{n} \frac{I_0'\left(\frac{2(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right)\left(\frac{(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right) \log x_{2i}}{I_0\left(\frac{2(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right)},$$

(3.5)
$$\frac{\partial \log L}{\partial \theta_2} = -\frac{n}{\theta_2} + \frac{1}{1-\rho} \sum_{i=1}^{n} \frac{x_{2i}^{\eta_2}}{\theta_2^2} - \sum_{i=1}^{n} \frac{I_0'\left(\frac{2(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right)\left(\frac{(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2^3}}\right)}{I_0\left(\frac{2(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right)},$$

(3.6)
$$\frac{\partial \log L}{\partial \rho} = \frac{n}{1-\rho} - \frac{1}{(1-\rho)^2} \sum_{i=1}^{n} \left(\frac{x_{1i}^{\eta_1}}{\theta_1} + \frac{x_{2i}^{\eta_2}}{\theta_2}\right) - \sum_{i=1}^{n} \frac{I_0'\left(\frac{2(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right)\left(\frac{(x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{\sqrt{\rho}\sqrt{\theta_1\theta_2}}\right)\frac{1+\rho}{(1-\rho)^2}}{I_0\left(\frac{2(\rho x_{1i}^{\eta_1} x_{2i}^{\eta_2})^{1/2}}{(1-\rho)\sqrt{\theta_1\theta_2}}\right)},$$

where $I_0'(z) = \frac{\partial}{\partial z} I_0(z)$. Using the relation

$$\frac{\partial}{\partial z} I_\nu(z) = \frac{\nu}{z} I_\nu(z) + I_{\nu+1}(z),$$

we readily obtain $I_0'(z) = I_1(z)$, where $I_1(z)$, the modified Bessel function of the first kind of order one, is given by the power series expansion

$$I_1(z) = \sum_{k=0}^{\infty} \frac{(z/2)^{1+2k}}{k!\Gamma(k+2)}.$$

It may be observed that it is not possible to obtain explicit solutions to these log-likelihood equations, and one has to rely on numerical methods for obtaining the MLE $\widehat{\boldsymbol{\omega}} = (\widehat{\eta_1}, \widehat{\theta_1}, \widehat{\eta_2}, \widehat{\theta_2}, \widehat{\rho})$. Numerical methods such as the Newton–Raphson technique, or its modified versions can be employed to obtain $\widehat{\boldsymbol{\omega}}$.

Asymptotic confidence intervals for the model parameters may be obtained by using the observed Fisher information matrix, and asymptotic normality of the MLEs. The observed Fisher information matrix, denoted by $\boldsymbol{I}(\boldsymbol{\omega})$, is defined as the negative of the Hessian of log-likelihood function in (3.1); that is,

$$\boldsymbol{I}(\boldsymbol{\omega}) = -\nabla^2(\log L(\boldsymbol{\omega})).$$

Now, under general regularity conditions, we know that the asymptotic distribution of the MLE $\widehat{\boldsymbol{\omega}}$ is a multivariate normal distribution, i.e.,

$$\sqrt{n}(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega}) \;\rightarrow\; \boldsymbol{N}_5(\boldsymbol{0}, \boldsymbol{I}^{-1}(\boldsymbol{\omega})|_{\boldsymbol{\omega}=\widehat{\boldsymbol{\omega}}}).$$

Using the above, asymptotic confidence intervals for the model parameters can be constructed. For example, for η_1, an asymptotic $100(1-\beta)\%$ confidence interval is given by

$$\left(\widehat{\eta_1} - \tau_{\beta/2}\sqrt{\boldsymbol{I}^{-1}_{1,1}}\;,\; \widehat{\eta_1} + \tau_{\beta/2}\sqrt{\boldsymbol{I}^{-1}_{1,1}}\right)$$

where $\boldsymbol{I}^{-1}_{i,j}$ is the (i,j)-th element of $\boldsymbol{I}^{-1}(\boldsymbol{\omega})|_{\boldsymbol{\omega}=\widehat{\boldsymbol{\omega}}}$, and τ_δ is the upper δ-percentile point of the standard normal distribution. Asymptotic confidence intervals for the parameters θ_1, η_2, θ_2, and ρ can be constructed in a similar fashion.

3.1.1. Choice of initial values

The numerical method for obtaining MLEs requires initial choices for the values of the parameters. Here, we develop a method for choosing initial parameter values. Note that the marginal distribution of X_1 (or X_2) does not depend on η_2 and θ_2 (or η_1 and θ_1). Based on this observation, we propose a method for computing closed form initial values.

When X follows a Weibull distribution with the PDF as given in (2.5), then $Z = \log X$ has an extreme-value distribution with PDF

$$f_Z(z) = \frac{1}{\sigma}\exp\left[\left(\frac{z-\mu}{\sigma}\right) - \exp\left(\frac{z-\mu}{\sigma}\right)\right], \qquad -\infty < z < \infty,\;\; -\infty < \mu < \infty,\;\; \sigma > 0,$$

where μ and σ are location and scale parameters, respectively, with

$$\mu = \frac{1}{\eta}\log\theta, \quad \sigma = \frac{1}{\eta}.$$

It can be shown that

$$E(Z) = \mu - \gamma\sigma, \quad \mathrm{Var}(Z) = \frac{\pi^2}{6}\sigma^2,$$

where $\gamma = 0.5722$ (approximately) is the Euler's constant. Using these results and relations, by equating the population moments with corresponding sample quantities and then by transforming back, we can easily work out approximate estimates for θ_1, θ_2, η_1, and η_2, as follows:

$$\theta_i = \exp\left(\bar{z}_i\left(\frac{\sqrt{6}}{\pi}s_i\right)^{-1} + \gamma\right), \quad \eta_i = \left(\frac{\sqrt{6}}{\pi}s_i\right)^{-1}, \quad i = 1, 2,$$

where $z_i = \log(x_i)$, $\bar{z}_i = \frac{1}{n}\sum_{i=1}^n z_i$, and $s_i = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (z_i - \bar{z})^2}$.

For the dependence parameter ρ, we choose the initial value by equating it to the sample Pearson's correlation coefficient estimated from the data, though it should be recalled here that ρ is not the Pearson's correlation coefficient for the DBW distribution. We have observed in our Monte Carlo simulation studies that the initial choices of parameter values obtained in this manner work quite efficiently for the numerical optimization technique to calculate the MLEs.

3.2. Estimators based on moments

Moment-based estimates of the parameters of the DBW distribution can be obtained by using results from the marginal and joint distributions. Note that, from the marginal distributions of X_1 and X_2, we have

$$E(X_i) = \theta_i^{1/\eta_i}\Gamma\left(1+\frac{1}{\eta_i}\right), \quad E(X_i^2) = \theta_i^{2/\eta_i}\Gamma\left(1+\frac{2}{\eta_i}\right), \quad i=1,2. \tag{3.7}$$

Suppose we have observed data as n pairs (x_{1j}, x_{2j}), $j=1,...,n$. Using (3.7), based on the observed data, we form the following equations, for $i=1,2$:

$$\frac{1}{n}\sum_{j=1}^{n} x_{ij} \;=\; \theta_i^{1/\eta_i}\Gamma\left(1+\frac{1}{\eta_i}\right), \tag{3.8}$$

$$\frac{1}{n}\sum_{j=1}^{n} x_{ij}^2 \;=\; \theta_i^{2/\eta_i}\Gamma\left(1+\frac{2}{\eta_i}\right), \tag{3.9}$$

which readily give

$$\frac{\Gamma\left(1+\frac{2}{\eta_i}\right)}{\left\{\Gamma\left(1+\frac{1}{\eta_i}\right)\right\}^2} = \frac{\frac{1}{n}\sum_{j=1}^{n} x_{ij}^2}{\overline{x}_i^2}, \tag{3.10}$$

where $\overline{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$. A moment-based estimator of η_i, say $\widetilde{\eta}_i$, can be obtained by numerically solving (3.10). Then, from (3.8), a moment-based estimator of θ_i, say $\widetilde{\theta}_i$, can be obtained as

$$\widetilde{\theta}_i = \left[\frac{\overline{x}_i}{\Gamma(1+\frac{1}{\widetilde{\eta}_i})}\right]^{\widetilde{\eta}_i}, \quad i=1,2. \tag{3.11}$$

Now, note that when $(X_1, X_2) \sim \mathrm{DBW}(\eta_1, \theta_1, \eta_2, \theta_2, \rho)$, we have $(Y_1, Y_2) \sim \mathrm{DBE}(\theta_1, \theta_2, \rho)$, where $Y_i = X_i^{\eta_i}$, for $i=1,2$. A closed form expression for the moment $\mu'_{r_1,r_2} = E(X_1^{r_1} X_2^{r_2})$ of the (r_1, r_2)-th order of the $\mathrm{DBE}(\theta_1, \theta_2, \rho)$ distribution can be obtained from Al-Saadi *et al.* (1979) as

$$\mu'_{r_1,r_2} \;=\; r_1!\, r_2!\, \theta_1^{r_1}\, \theta_2^{r_2} \sum_{j=0}^{r_2} \binom{r_2}{j} \binom{r_1+r_2-j}{r_1} \rho^{r_2-j}(1-\rho)^j,$$

from which it follows that

$$\mu'_{1,1} = (1+\rho)\,\theta_1\,\theta_2. \tag{3.12}$$

Al-Saadi and Young (1980) proposed moment-estimators for the parameters of the Downton's bivariate exponential (DBE) distribution based on this result. Balakrishnan and Ng (2001) proposed methods for improving estimates of ρ for the DBE distribution through resampling schemes. For the $\mathrm{DBW}(\eta_1, \theta_1, \eta_2, \theta_2, \rho)$ distribution, using the data (x_{1j}, x_{2j}), $j=1,...,n$, and using (3.12), we define

$$R = \frac{\sum_{j=1}^{n} y_{1j}\, y_{2j}}{n\,\widetilde{\theta}_1\,\widetilde{\theta}_2} - 1, \tag{3.13}$$

where $Y_i = X_i^{\eta_i}$, $i = 1, 2$, and $\widetilde{\theta}_1$ and $\widetilde{\theta}_2$ are as given in (3.11). Finally, a moment-based estimator for ρ can be proposed as

$$\widetilde{\rho} = \begin{cases} 0, & \text{if } R < 0, \\ R, & \text{if } 0 \le R < 1, \\ 1, & \text{if } R \ge 1. \end{cases} \tag{3.14}$$

The process described above can be organized in the form of an algorithm to compute the moment-based estimators $\widetilde{\theta}_1$, $\widetilde{\theta}_2$, $\widetilde{\eta}_1$, $\widetilde{\eta}_2$, and $\widetilde{\rho}$, using the data (x_{1j}, x_{2j}), $j = 1, ..., n$ as follows:

Algorithm 2:

Step 1: Obtain $\widetilde{\eta}_i$, $i = 1, 2$, by numerically solving (3.10).

Step 2: Obtain $\widetilde{\theta}_i$, $i = 1, 2$, from (3.11).

Step 3: Transform the data into (y_{1j}, y_{2j}), where $y_{ij} = x_{ij}^{\widetilde{\eta}_i}$, $i = 1, 2$, $j = 1, ..., n$.

Step 4: Based on the transformed data (y_{1j}, y_{2j}), obtain $\widetilde{\rho}$ from (3.14).

Note that the moment-based estimates of parameters may be used as initial values for numerical computation of the MLEs.

3.3. Numerical experiments: Monte Carlo simulations

To examine performance of the MLEs $\widehat{\eta}_1, \widehat{\theta}_1, \widehat{\eta}_2, \widehat{\theta}_2$, and $\widehat{\rho}$, and the asymptotic confidence intervals, extensive Monte Carlo simulations are carried out by using the R software (2023). For generating observations from the DBW distribution, the Algorithm 1 in Section 2.2 is followed.

The scale parameters θ_1 and θ_2 are fixed at unity, without any loss of generality. Different values of the shape parameters η_1 and η_2, and the dependence parameter ρ are used. The values of η_1 and η_2 are set at 0.50, 0.75, 1.00, and 1.25; ρ is taken as 0.25 (small), 0.5 (moderate), and 0.75 (high). Finally, all these settings are repeated for three different sample sizes: 20, 40, and 100. The detailed results of this detailed Monte Carlo simulation study are presented in the online supplementary materials of this paper in which we report bias and mean squared error (MSE) of the MLEs. Coverage probability of the asymptotic 95% confidence intervals are also reported.

It may be noted that the bias and MSE of the MLEs of all the parameters are quite reasonable. Clearly, as one would expect, the bias and MSE decrease with increasing sample size. The coverage probability of the confidence intervals for all parameters are reasonably close to the nominal level of 95%, and as expected, they improve with increasing sample size. The coverage probability of the confidence interval for ρ when $\rho = 0.25$ in smaller samples $(n = 20, 40)$ is somewhat less than the nominal level, but for larger samples $(n = 75, 100)$ it improves, and gets closer to the nominal level. Overall, the simulation results reflect that the MLEs are reasonably close to the true parameter values for a very wide range of simulation settings.

For the accuracy of the asymptotic confidence intervals, an important factor of the asymptotic normality of the MLEs. Here, we present histograms of the MLEs corresponding to one of the simulation settings in Figure 4. As pointed out by one of the reviewers, the asymptotic normality of the MLE of the parameter ρ can be questionable, especially when the true value of ρ is close to 0 and 1. To address this issue, suitable transformations of the MLEs may be considered. We optimized the likelihood function on a different scale transforming the parameters, to avoid any such problem. Natural logarithm transformation was considered for θ_1, θ_2, η_1, and η_2, and logit transformation was considered for ρ. The histograms of the transformed estimates are given in Figure 5. Similar results are obtained for all simulation settings considered.

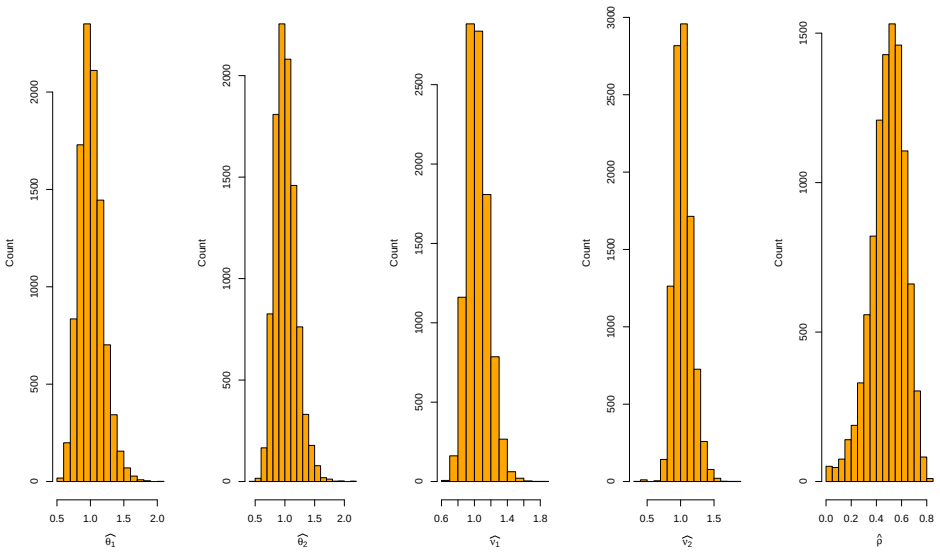

Figure 4: Histogram of the MLEs based on complete data when $\eta_1 = \eta_2 = 1$, $\theta_1 = \theta_2 = 1$, $\rho = 0.5$.

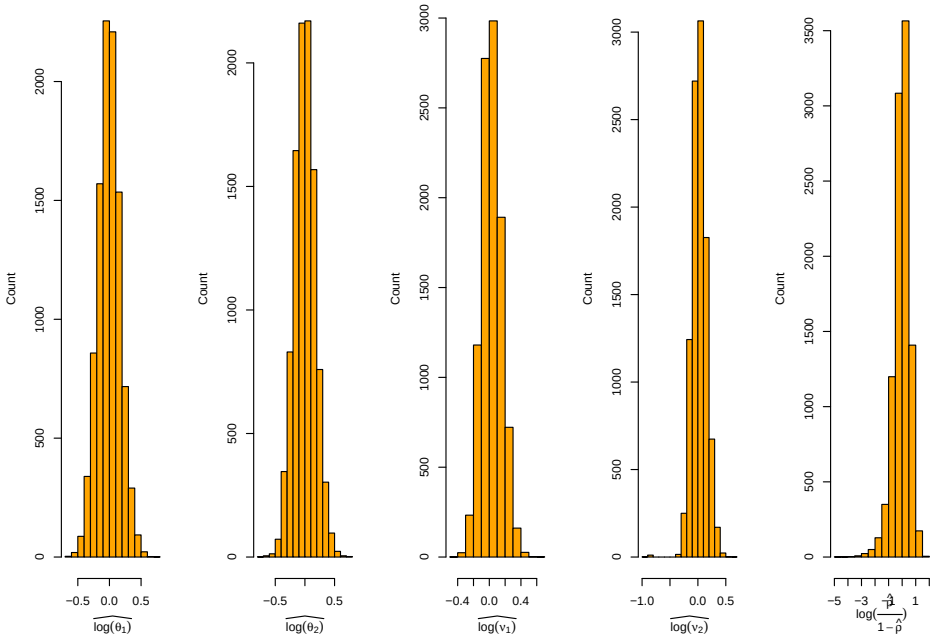

Figure 5: Histogram of the transformed MLEs based on complete data when $\eta_1 = \eta_2 = 1$, $\theta_1 = \theta_2 = 1$, $\rho = 0.5$.

4. INFERENCE BASED ON RIGHT-CENSORED DATA

4.1. Maximum likelihood estimation

In reliability experiments, right-censoring, and by that we mean Type-I right censoring, is perhaps the most common among different censoring mechanisms. For the univariate case, it may be described as follows. In the right-censoring scheme, units under study are followed up to a pre-fixed time-point, often called the right-censoring point. The units which are still in working condition at this point, are called the right-censored units. The data on lifetimes obtained from such a study are then a mixture of observed failures and right censored lifetimes. Inferential procedures based on univariate data with right-censoring have been developed by several authors; see for example Meeker and Escobar (1998).

For bivariate lifetime data, naturally the right-censoring model needs to consider censoring on both components of the lifetime vector. We consider the following model for right-censoring in the bivariate case. First, assume that all the n units under study are followed starting from a common time point, say τ, and are followed until a pre-fixed time point, say T; the endpoint T is assumed to be common to both components of the lifetime vector (X_1, X_2) for each unit (though this assumption can be relaxed without complications). At the end of the study, there are four possibilities for each unit:

(**a**) failure with respect to both components X_1 and X_2;
(**b**) failure with respect to X_1, but X_2 is right censored at T;
(**c**) failure with respect to X_2, but X_1 is right censored at T;
(**d**) both components are right-censored at T.

We assume non-informative censoring, i.e., the censoring mechanism is assumed to be independent of the lifetimes.

We can now construct the likelihood function for right-censored data. Let Δ_1 and Δ_2 denote indicators of censoring of a unit with corresponding to X_1 and X_2, respectively, i.e., for $i = 1, 2$,

$$\Delta_i = \begin{cases} 0, & \text{if } X_i > T - \tau, \\ 1, & \text{otherwise.} \end{cases}$$

Let S_{obs}, S_1, and S_2, and S_{cen} denote index sets such that

$$S_{\text{obs}} = \{j\colon \Delta_{1j} = 1, \Delta_{2j} = 1\}, \quad S_1 = \{j\colon \Delta_{1j} = 1, \Delta_{2j} = 0\},$$

$$S_2 = \{i\colon \Delta_{1j} = 0, \Delta_{2j} = 1\}, \quad S_{\text{cen}} = \{i\colon \Delta_{1j} = 0, \Delta_{2j} = 0\}.$$

Then the observed likelihood function for bivariate right-censored lifetime data is given by

$$\begin{aligned} L(\boldsymbol{\omega}) \;=\; & \prod_{j \in S_{\text{obs}}} f_{X_1,X_2}(x_{1j}, x_{2j}) \;\times \prod_{j \in S_1} f_{X_1}(x_{1j}) \Big[1 - F_{X_2|X_1=x_{1j}}(T-\tau)\Big] \\ & \times \prod_{j \in S_2} f_{X_2}(x_{2j}) \Big[1 - F_{X_1|X_2=x_{2j}}(T-\tau)\Big] \;\times \prod_{j \in S_{\text{cen}}} S_{X_1,X_2}(T-\tau, T-\tau). \end{aligned} \tag{4.1}$$

To further generalize the right-censoring model, consider the case where different units may start from different time points. For the j-th unit that starts from τ_j, the right-censoring time C_j is defined as the length of its life up to the fixed right censoring point T, and in that case, its observed lifetime corresponding to the two components X_1 and X_2 may be defined as

$$Z_{1j} = \text{Min}(X_{1j}, C_j), \quad Z_{2j} = \text{Min}(X_{2j}, C_j).$$

The likelihood function can then be generalized as

$$\begin{aligned} L(\boldsymbol{\omega}) \;=\; & \prod_{j=1}^{n} \Big\{ f_{X_1,X_2}(z_{1j}, z_{2j}) \Big\}^{\Delta_{1j}\Delta_{2j}} \Big\{ f_{X_1}(z_{1j}) \left[1 - F_{X_2|X_1=z_{1j}}(C_j)\right] \Big\}^{\Delta_{1j}(1-\Delta_{2j})} \\ & \times \Big\{ f_{X_2}(z_{2j}) \left[1 - F_{X_1|X_2=z_{2j}}(C_j)\right] \Big\}^{(1-\Delta_{1j})\Delta_{2j}} \Big\{ S_{X_1,X_2}(C_j, C_j) \Big\}^{(1-\Delta_{1j})(1-\Delta_{2j})}. \end{aligned} \tag{4.2}$$

It is quite clear that explicit, closed-form MLE of the parameters $\widehat{\boldsymbol{\omega}} = (\widehat{\eta_1}, \widehat{\theta_1}, \widehat{\eta_2}, \widehat{\theta_2}, \widehat{\rho})$ cannot be obtained by optimizing (4.1) or (4.2); numerical techniques must be employed to compute the MLEs. Asymptotic confidence intervals for the parameters may be constructed by using asymptotic variance-covariance matrix which is possible to obtain through the observed Fisher information matrix, in a similar fashion as discussed in Section 3. It is of importance to mention here that when all the observations are right-censored, i.e., $\Delta_{1i} = 0$ and $\Delta_{2i} = 0$ for all i, the likelihood estimates do not converge. In statistical analysis of reliability data, this is a quite common issue that occurs due to insufficiency of the available information regarding the underlying lifetime distribution in case of no observed failures.

4.2. Numerical experiments: Monte Carlo simulations

Detailed Monte Carlo simulations by using the R software are performed to assess the performance of the MLEs of the DBW distribution based on right-censored data. The bias and mean squared error (MSE) of the point estimates, and coverage probability of the asymptotic confidence intervals are estimated through simulations.

Three different sample sizes are used: $n = 20$ (small), 40 (moderate), and 100 (large). The scale parameters θ_1 and θ_2 are taken to be unity, without loss of generality. Three different values, namely, 0.75, 1.00, and 1.25 are used for the shape parameters η_1 and η_2. The dependence parameter ρ is fixed at 0.25 (moderate) and 0.75 (high). These settings are used along with different right-censoring rates. The detailed simulation results are provided in the online supplementary materials of this paper.

It is observed that bias and MSE of the MLEs in case of right-censored data are quite reasonable throughout, demonstrating expected trend of reducing with increasing sample size. The coverage probability of the confidence intervals for all parameters η_1, θ_1, η_2, and θ_2 are very close to the nominal level of 95%. The coverage probability of the confidence intervals for ρ is occasionally (e.g., when true value of ρ is 0.25) away from the nominal level of 95% for smaller sample sizes ($n = 20$), but it gradually improves with increase in sample size, and gets very close to the nominal level of 95% for larger sample size ($n = 100$). In summary, it may be concluded that the MLEs of parameters of the DBW distribution based on right-censored data are quite efficient in estimating the true parameters for a wide range of simulation settings.

In this case of time-censored data also, like the case with complete data, we optimized the likelihood function on transformed scale of the parameters. We provide histograms of the MLEs based on time-censored data, on the original scale as well as transformed scale, for one of the simulation settings in Figures 6 and 7. This demonstrates asymptotic normality of the MLEs.

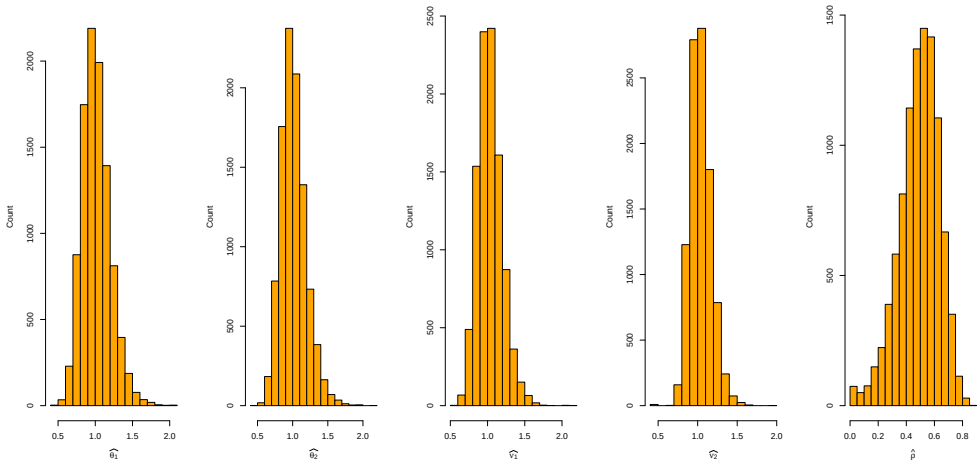

Figure 6: Histogram of the MLEs based on censored data when $\eta_1 = \eta_2 = 1$, $\theta_1 = \theta_2 = 1$, $\rho = 0.5$.

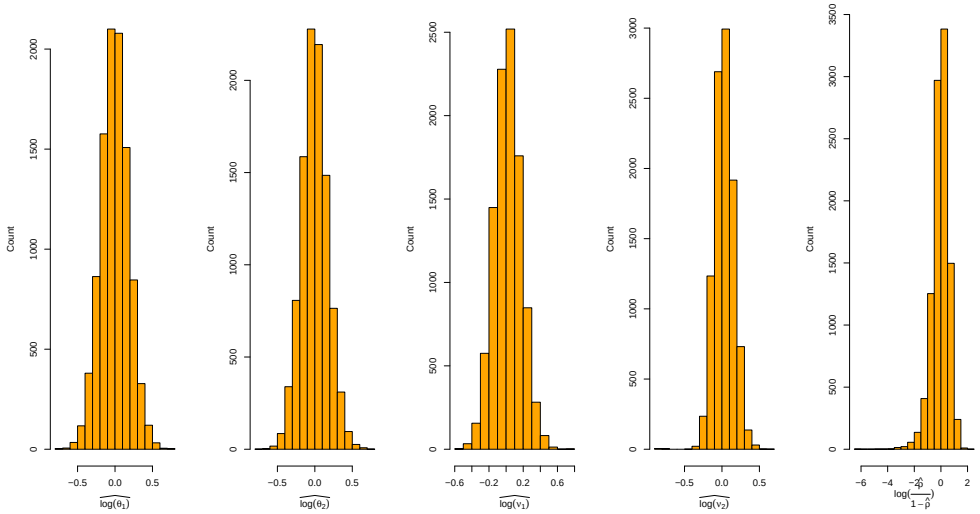

Figure 7: Histogram of the transformed MLEs based on censored data when $\eta_1 = \eta_2 = 1$, $\theta_1 = \theta_2 = 1$, $\rho = 0.5$.

5. APPLICATION TO A TWO-DIMENSIONAL WARRANTY MODEL

Several products such as automobiles, industry equipments, and heavy duty machines are sold under a two-dimensional warranty. For such products, their age as well as usage information are used to evaluate a warranty claim (Jung and Bai, 2007). For example, a manufacturer may sell a car under a warranty of 5 years or 50,000 kilometers.

Naturally, a bivariate distributional model will be suitable for analysing reliability data with a two-dimensional warranty. Jung and Bai (2007) proposed a model for analysing two-dimensional warranty data. Motivated by the work of Jung and Bai (2007), here we propose a methodology to model two-dimensional warranty data by using a bivariate distribution. Our methodology generalizes the method proposed by Jung and Bai (2007) by using more information while estimating the underlying bivariate distribution in terms of different status of failures of units sold under two-dimensional warranty.

Consider a product with two measurable quantities with respect to a two-dimensional warranty; let the random variables U and V denote these measurable quantities (for example, age and mileage of a car, respectively). Suppose the product is sold under a two-dimensional warranty (ψ, ξ); that is, a failure is considered to be within warranty when $(U, V) \in \psi \circledast \xi$, where

$$\psi \circledast \xi = \Big\{(u, v)\colon\ 0 \le u \le \psi,\ 0 \le v \le \xi\Big\}.$$

Suppose U and V follow a bivariate distribution with CDF $F_{U,V}(u, v; \boldsymbol{\omega})$ with corresponding PDF $f_{U,V}(u, v; \boldsymbol{\omega})$, where $\boldsymbol{\omega}$ is the vector of relevant model parameters. Our aim is to estimate the distribution $F_{U,V}(u, v; \boldsymbol{\omega})$ based on field-failure data of the product, so that relevant probabilities for the two-dimensional warranty can be estimated from $\widehat{F}_{U,V}(u, v; \widehat{\boldsymbol{\omega}})$.

Consider n units of the concerned product that are sold at a common time $(U_{\text{start}}, V_{\text{start}}) = (0, 0)$, under a two-dimensional warranty (ψ, ξ). Further assume that the manufacturer observes the status of the units at $(U_{end}, V_{end}) = (\psi, \xi)$. For each unit, there are four possibilities at this stage:

- (**a**) failure within the warranty region, i.e., $(u, v) \in \psi \circledast \xi$ — in this case its contribution to the likelihood function would be $f_{U,V}(u, v)$;
- (**b**) warranty expired with respect to U, but not with respect to V, i.e., $U > \psi$ and $0 \le V \le \xi$ — in this case, its contribution to the likelihood function would be $f_U(u)\,[1 - F_{V|U=u}(\xi)]$;
- (**c**) warranty expired with respect to V, but not with respect to U, i.e., $0 \le U \le \psi$ and $V > \xi$ — in this case its contribution to the likelihood function would be $f_V(v)\,[1 - F_{U|V=v}(\psi)]$;
- (**d**) warranty expired with respect to both U and V, i.e., $U > \psi$ and $V > \xi$ — in this case its contribution to the likelihood function would be $S_{U,V}(\psi, \xi)$, where $S_{(U,V)}(\cdot, \cdot)$ is the joint survival function of (U, V).

Clearly, in cases (b), (c), and (d) above, at least one of the random variables is right-censored. Analysing product field-failure data under two-dimensional warranty is particularly challenging due to the right censored observations. Considering contributions of the n units under study according to their status with respect to (a), (b), (c), and (d), the likelihood function for field-failure data under two-dimensional warranty is given by

$$\begin{aligned} L(\boldsymbol{\omega}) \;=\; & \prod_{i=1}^{n} \Big\{f_{U,V}(u_i, v_i)\Big\}^{\Delta_{1i}\Delta_{2i}} \Big\{f_U(u_i)\,\big[1 - F_{V|U=u_i}(\xi)\big]\Big\}^{\Delta_{1i}(1-\Delta_{2i})} \\ & \times \Big\{f_V(v_i)\,\big[1 - F_{U|V=v_i}(\psi)\big]\Big\}^{(1-\Delta_{1i})\Delta_{2i}} \Big\{S_{U,V}(\psi, \xi)\Big\}^{(1-\Delta_{1i})(1-\Delta_{2i})}, \end{aligned} \tag{5.1}$$

where

$$\Delta_1 = \begin{cases} 0, & \text{if } U > \psi, \\ 1, & \text{otherwise,} \end{cases} \qquad \text{and} \qquad \Delta_2 = \begin{cases} 0, & \text{if } V > \xi, \\ 1, & \text{otherwise.} \end{cases}$$

It can be seen immediately that the likelihood in (5.1) follows from the general right-censored likelihood in (4.2). The likelihood in (5.1) can be maximized using numerical approach to obtain the MLE $\widehat{\boldsymbol{\omega}}$. Then, using $\widehat{\boldsymbol{\omega}}$, it is straightforward to estimate probabilities such as $\widehat{F}_{U,V}(\psi, \xi; \widehat{\boldsymbol{\omega}})$, i.e., the probability of a unit failing within the warranty region.

As the DBW distribution is a flexible model with Weibull marginals, it can accommodate a wide range of lifetime data, successfully capturing non-constant marginal hazard rates as well the dependence between the components. Therefore, we propose to model the joint distribution of U and V by the DBW distribution. When

$$(U, V) \sim \text{DBW}(\eta_1, \theta_1, \eta_2, \theta_2, \rho),$$

plugging-in the joint PDF, conditional CDFs, and joint survival function of the distribution in (5.1), we can estimate the parameters by the MLE $\widehat{\boldsymbol{\omega}} = (\widehat{\eta_1}, \widehat{\theta_1}, \widehat{\eta_2}, \widehat{\theta_2}, \widehat{\rho})$. Then, the probability of a unit failing within the warranty region is given by

$$\begin{aligned} \widehat{F}_{U,V}(\psi, \xi; \widehat{\boldsymbol{\omega}}) \;=\; & 1 - \exp\left(-\frac{\psi^{\widehat{\eta_1}}}{\widehat{\theta_1}}\right) Q_1\left(\sqrt{\frac{2}{1-\widehat{\rho}}}\,\frac{\xi^{\widehat{\eta_2}/2}}{\sqrt{\widehat{\theta_2}}}\,,\, \sqrt{\frac{2\widehat{\rho}}{1-\widehat{\rho}}}\,\frac{\psi^{\widehat{\eta_1}/2}}{\sqrt{\widehat{\theta_1}}}\right) \\ & - \exp\left(-\frac{\xi^{\widehat{\eta_2}}}{\widehat{\theta_2}}\right)\left[1 - Q_1\left(\sqrt{\frac{2\widehat{\rho}}{1-\widehat{\rho}}}\,\frac{\xi^{\widehat{\eta_2}/2}}{\sqrt{\widehat{\theta_2}}}\,,\, \sqrt{\frac{2}{1-\widehat{\rho}}}\,\frac{\psi^{\widehat{\eta_1}/2}}{\sqrt{\widehat{\theta_1}}}\right)\right]. \end{aligned} \tag{5.2}$$

An illustrative example of this warranty model is given in Section 6.1.

6. CASE STUDIES

In this section, we provide analyses of two real datasets. The first example is based on the car warranty data of Jung and Bai (2007). The second example is based on a general bivariate data on bone mineral density, originally from Johnson and Wichern (1999). The car warranty data demonstrates the use of the DBW distribution for reliability data, including its application in the two-dimensional warranty model discussed above. The the bone mineral density data is not a lifetime data as such, we present analysis of this data to demonstrate the versatility of the DBW distribution in modelling different types of data.

6.1. Real Data 1: Car warranty data

The car warranty dataset, analysed by Jung and Bai (2007), and Yuan (2016), contains data on age and mileage of 40 cars. Age and mileage of a car are two measurable quantities to assess its lifetime. Yuan (2016) analysed this data to obtain generalized moment estimates of a bivariate Weibull distribution, while Jung and Bai (2007) considered a two-dimensional warranty model.

Suppose U and V are the age and mileage of a car, respectively. In Figure 8, a scatterplot of U and V is given. It is clear from this scatterplot that age and mileage are highly correlated, which suggests that a bivariate distribution that captures the dependence between them will be suitable to model the car warranty data.

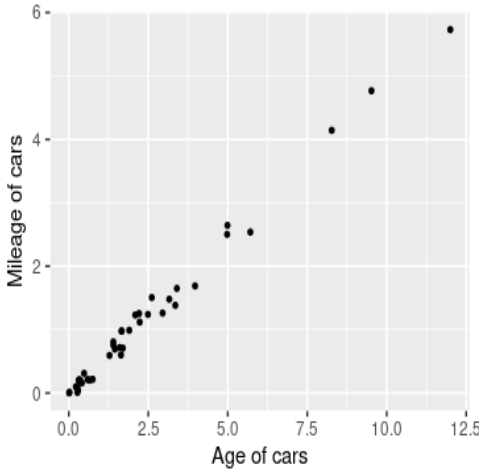

Figure 8: Scatterplot of Age (X_1) and Mileage (X_2) of cars for the car warranty data.

An option is to use the DBE distribution for this data. However, to observe that the marginal distributions of age and mileage have non-constant hazard rates, see Figure 9. From this plot, it is clear that the DBE distribution will not be suitable for the car warranty data, as the marginal distributions of U and V cannot be modelled by exponential distributions.

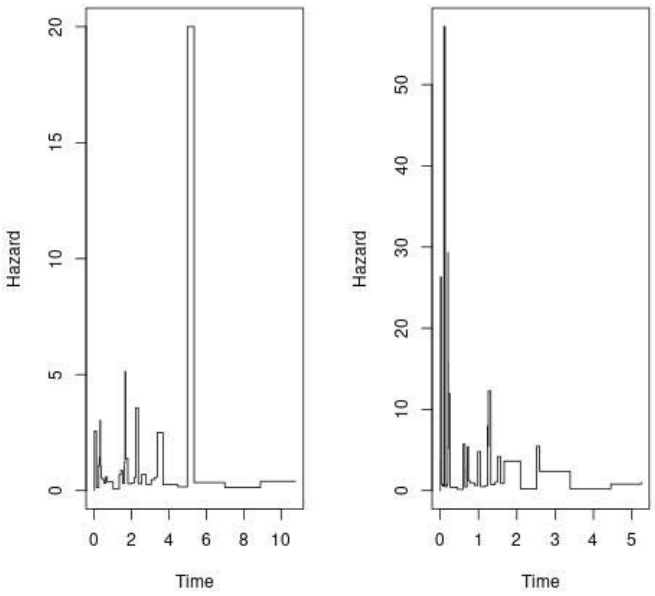

Figure 9: Plot of hazard functions for the marginal distributions for the car warranty data: car age (left), and car mileage (right).

Therefore, to accommodate the non-constant marginal hazards rates, we model the car warranty data by the DBW distribution, where the marginals are Weibull distributions. It may be noted here that Jung and Bai (2007) also provided justification for using the Weibull distribution as the marginal models for age and mileage in the car warranty data; though their bivariate model for age and mileage was not the DBW distribution.

The warranty data contains complete observations on U and V. However, for illustrative purposes, here we artificially impose right censoring to analyse the data; see the results presented in Table 1. Note that for greater amount of right-censored units, the confidence intervals of the parameters are wider, implying larger standard errors. Naturally, this implies that the estimates are relatively more reliable for smaller right censoring percentages, as expected. The MLEs can be used further to compute suitable probabilities from the estimated model. Table 1 also gives the estimated probabilities for units to fail within the warranty region, for different two-dimensional warranty schemes: $(\psi, \xi) = (0.5, 0.5)$, (1,1), (2,2). It is, of course, possible to use different values for ψ and ξ, for example, $(\psi, \xi) = (1,2)$ etc. for real life applications.

Table 1: MLEs and asymptotic confidence intervals of parameters of the DBW model based on Warranty Data for different right-censoring points.

Right-censoring Point	Parameter	MLE	Asymptotic 95% CI	$\widehat{F}_{U,V}(\psi,\xi)$
(0.5,0.5)	θ_1	3.909	(2.218, 6.891)	$\widehat{F}_{U,V}(0.5, 0.5) = 0.250$
	θ_2	2.679	(1.573, 4.562)	
	η_1	0.428	(0.258, 0.710)	
	η_2	0.379	(0.232, 0.620)	
	ρ	0.993	(0.983, 0.998)	
(1,1)	θ_1	2.537	(1.597, 4.032)	$\widehat{F}_{U,V}(1, 1) = 0.469$
	θ_2	1.579	(1.049, 2.376)	
	η_1	0.404	(0.254, 0.642)	
	η_2	0.414	(0.271, 0.632)	
	ρ	0.991	(0.977, 0.996)	
(2,2)	θ_1	1.778	(1.213, 2.609)	$\widehat{F}_{U,V}(2, 2; \widehat{\boldsymbol{\omega}}) = 0.763$
	θ_2	0.952	(0.695, 1.304)	
	η_1	0.719	(0.550, 0.941)	
	η_2	0.716	(0.546, 0.940)	
	ρ	0.985	(0.973, 0.992)	

6.2. Real Data 2: Bone mineral density data

The bone mineral density data from Johnson and Wichern (1999) consists of measurements on bone mineral density (BMD) in the unit g/cm^2 for 24 adults. Within each observation, the first value (X_1) represents the BMD of the bone dominant radius before the start of the study, and the second value (X_2) represents the BMD of the bone after the study. A scatterplot of the observed data indicates that the two variables are highly correlated.

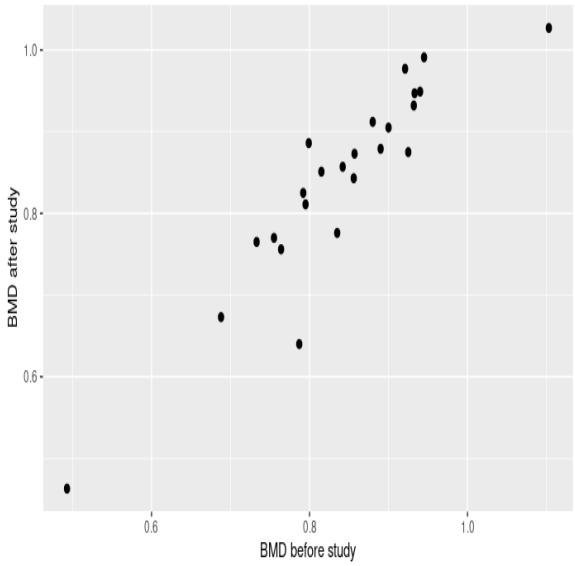

Figure 10: Plot of BMD before and after study.

Assuming DBW distribution for (X_1, X_2), we obtain the MLEs of the model parameters based on the given BMD data; Table 2 gives the results. We also calculate asymptotic 95% confidence intervals for the parameters.

Table 2: Estimates of parameters of the DBW model based on the BMD data.

Parameter	Estimate	95% CI
θ_1	0.407	(0.265, 0.623)
θ_2	0.372	(0.236, 0.586)
η_1	7.026	(5.427, 9.096)
η_2	8.100	(5.846, 11.222)
ρ	0.935	(0.860, 0.972)

To check if the model fit is reasonable, in Figure 11 we plot the estimated Weibull survival function $\widehat{S}(t) = \exp\left\{-\left(\frac{t^{\hat{\eta}}}{\hat{\theta}}\right)\right\}$ against the empirical survival function, evaluated at the ordered observed values, separately for X_1 and X_2. Noting that the plotted points roughly form a straight line which indicates that the estimated values of the survival function are in agreement, we conclude that the model fit is reasonable.

Finally, we can compare the suitability of the DBW model with other bivariate models available in the literature. For this purpose, here we consider the Marshall–Olkin bivariate Weibull (MOBW) distribution which was proposed by Marshall and Olkin as an extension of a bivariate exponential distribution that is now known as Marshall–Olkin bivariate exponential (MOBE) distribution (Marshall and Olkin, 1967). For the DBW and MOBW models, we calculate the Akaike's Information Criterion (AIC), and compare the values; the model with the lower AIC value would be more suitable for a given data. For the bone mineral density data, we obtain $\text{AIC}_{\text{DBW}} = -103.297$, and $\text{AIC}_{\text{MOBW}} = -55.457$. This suggests that the DBW distribution is a more suitable model for the bone mineral density data.

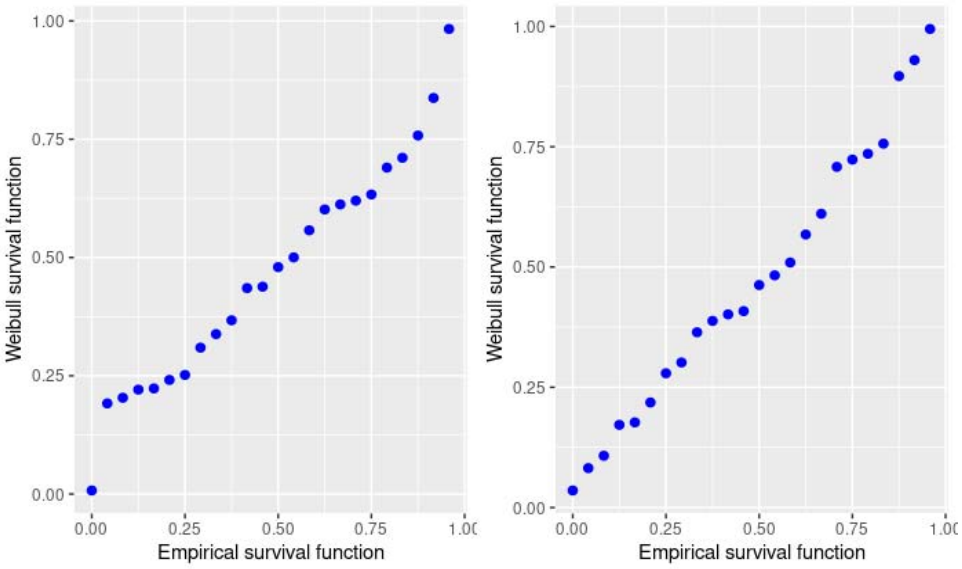

Figure 11: Plot of empirical survival function and Weibull survival function for the marginal distributions at the ordered observations with estimates parameters for X_1 (left) and X_2 (right).

7. CONCLUSION

In this article, a bivariate Weibull distribution, which we call the Downton's bivariate Weibull (DBW) distribution, is considered. Important statistical properties of the DBW distribution are studied. Then, inferences based on complete, and right-censored bivariate data are discussed for this distribution. Through extensive Monte Carlo simulation studies, it is observed that the point and interval estimates of the parameters of this distribution perform quite well — for both complete as well as right-censored data. A two-dimensional warranty model is discussed, and the application of the DBW distribution in the warranty model is considered. For illustrative purposes, two case studies based on real datasets are provided.

The DBW distribution has desirable statistical properties. It is quite flexible for modelling purposes, and it computationally convenient. It successfully captures the dependence between the components of a lifetime vector. In particular, the DBW distribution is very useful in modelling bivariate lifetime data when the marginal distributions indicate non-constant hazard rates. Based on our explorations presented in this paper, we strongly recommend its use in real life, particularly to model bivariate reliability data.

REFERENCES

Al-Saadi, S. D., Scrimshaw, D. F., and Young, D. H. (1979). Tests for independence of exponential variables. *Journal of Statistical Computation and Simulation*, 9:217–233.

Al-Saadi, S. D. and Young, D. H. (1980). Estimators for the correlation coefficient in a bivariate exponential distribution. *Journal of Statistical Computation and Simulation*, 11:13–20.

Balakrishnan, N. and Ng, H. K. T. (2001). Improved estimation of the correlation coefficient in a bivariate exponential distribution. *Journal of Statistical Computation and Simulation*, 68:173–184.

Brychkov, Y. A. (2017). Reduction formulas for the Appell and Humbert functions. *Integral Transforms and Special Functions*, 28:22–38.

Downton, F. (1970). Bivariate exponential distributions in reliability theory. *Journal of the Royal Statistical Society, Series B*, 32:408–417.

Fuqing, Y., Barabadi, A., and Jinmei, L. (2017). Reliability modelling on two-dimensional life data using bivariate Weibull distribution: with case study of truck in mines. *Maintenance and Reliability*, 19:650–659.

Gradshteyn, I. S. and Ryzhik, I. M. (2000). *Tables of Integrals, Series, and Products*, 6th edition. Academic Press.

Johnson, R. A. and Wichern, D. W. (1999). *Applied Multivariate Analysis*, 4th edition. Prentice-Hall, New Jersey, NJ.

Jung, M. and Bai, D. S. (2007). Analysis of field data under two-dimensional warranty. *Reliability Engineering and System Safety*, 92:135–143.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). *Continuous Multivariate Distributions*, Volume 1: *Models and Applications*, 2nd edition. John Wiley & Sons.

Marcum, J. (1950). Table of Q function. *Air Force Project RAND Research Memorandum M-339, ASTIA Document AD 1165451*. RAND Corporation, Santa Monica, California.

Marshall, A. W. and Olkin, I. (1967). A multivariate exponential distribution. *Journal of the American Statistical Association*, 62:30–44.

Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. John Wiley & Sons, New York.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Sagias, N. C. and Karagiannidis, G. K. (2005). Gaussian class multivariate Weibull distributions: theory and applications in fading channels. *IEEE Transactions on Information Theory*, 51:3608–3619.

Yuan, F. (2016). Parameter estimation for bivariate Weibull distribution using generalized moment method for reliability evaluation. *Quality and Reliability Engineering International*, 34:631–640.

REVSTAT – Statistical Journal
Volume 23, Number 3, July 2025, 325–340
https://doi.org/10.57805/revstat.v23i3.498

The Utilization of Partial Least Squares for Simultaneous Feature Selection and Extraction

Authors: Elisavet Beki
– Laboratory of Statistics and Data Analysis, University of the Aegean, Greece
elsampeki@hotmail.com

Kimon Ntotsis ✉
– Institute for Lung Health, NIHR Biomedical Research Centre – Respiratory & Infectious Diseases Theme, School of Medical Sciences, University of Leicester, United Kingdom
kn170@leicester.ac.uk

Received: July 2022 Revised: July 2023 Accepted: July 2023

Abstract:

- Several works concerning the utilization of Partial Least Squares as a supervised dimension reduction technique have been developed over the years in the field of chemometrics, among others, for regression purposes. However, Partial Least Squares can be a challenging procedure especially in the case of multivariate multiple regression due to data characteristics and complexity. Thus, in this work we propose the use of Partial Least Squares method as a variable selection technique in linear regression tasks that involve high dimensional spectral data sets. More precisely, we suggest the exploitation of the regression coefficients that Partial Least Squares estimates in order to identify and eject the insignificant predictor variables from the analysis. In such manner we are able to remove the uninformative variables and obtain in most cases better results than the classical Partial Least Squares regression but with simpler structure. We compare our proposed technique with the classical Partial Least Squares and Principal Component Analysis in both univariate and multivariate regression.

Keywords:

- *partial least squares; beta constrain; dimension reduction; variable selection.*

AMS Subject Classification:

- 62H25, 62J05, 62H12, 62H20.

✉ Corresponding author

1. INTRODUCTION

Recent advances in computer science and engineering enable scientists to collect and store datasets, typically characterized by high dimensionality. Such data sets, also referred to as Big Data, consist of numerous measured variables, the successful process of which aim to provide production of accurate predictions, informative visualizations and foremost, the in-depth understanding of the underlying patterns and relations between the variables of a dataset. However, the complexity of Big Data raises new challenges in their effective handling. The mitigation of noise accumulation and multicollinearity are two of the most common obstacles that the researcher is called upon to address and resolve. By noise, it is meant the part of data that brings insignificant information, while the term multicollinearity is used to describe the existence of high intercorrelations between variables. Further, in many cases, scientists have to deal with the existence of a small number of available observations compared with the larger number of measured variables, a situation referred as "small n, large p problem" (Ntotsis *et al.*, 2021).

Therefore, Dimensionality Reduction Techniques, hereafter DRT, have been developed and utilized to overcome these potential situations that can severely effect the modelling process when analyzing Big Data. DRT are defined as techniques that process high-dimensional datasets to produce representations of lower dimension, that retain as much as possible information from the original ones. Dimensionality reduction is used in statistical tasks that serve various purposes, such as classification, linear regression, etc, and it can be achieved either by feature selection or by feature extraction.

Any technique that selects and subsets a number of the initial variables that are considered statistically significant, and formulates a new dataset containing only these ones can be characterized as feature selection. By this process, variables' interpretation is preserved. Nevertheless, despite the advantage of interpretability, information captured in interactions and correlations between retained and removed variables is lost (Li and Zeng, 2009). Some of the most frequently used feature selection techniques, based either on wrapper methods (e.g. Forward Feature Selection, Backward Feature Elimination), or either on regularization methods (e.g. Least Absolute Shrinkage and Selection Operator). On the contrary, in feature extraction the produced data set is consisted of transformed variables and each variable is considered as a combination of all the initial ones, that project data points to a low dimensional space. These transformations are called components and they summarize information of initial variables by using them to form linear combinations. Although, in this case the interpretability of the initial variables is lost, feature extraction methods are very popular because they effectively replace initial variables by a few components that compress the relevant information contained in the data and make visualizations feasible. Among the most popular feature extraction techniques are Partial Least Squares Method (PLS), Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA) and Canonical Correlation Analysis (CCA).

Despite their different approach, both types of dimensionality reduction techniques can successfully mitigate noise accumulation and multicollinearity. Further, analysis of data of reduced complexity leads to models with improved accuracy/precision generated with less computational power and time, due to the parsimonious descriptions of the available observations.

As a result, their implementation in Big Data is substantial and the increased scientific interest about them has resulted in the proposal of various techniques.

In this article we attempt to evaluate the proficiency of a multilevel dimension reduction scheme: In the first place we estimate the regression coefficients of original variables utilizing PLS method. Combining testing thresholds and model assessment criteria, we are able to identify and remove uninformative variables for the modelling of response variable. Further, the dimension reduction is completed with the integration of Partial Least Squares Regression on the modified reduced data set. The efficiency of the proposed algorithm is investigated through the implementation of Principal Component Regression (PCR) on the reduced data set, independently from PLS, to compare the performance of the resulted models, in terms of AIC, Adjusted R^2 value and RMSECV information criteria.

In the past, many researchers have modified PLS and proposed refinements of the classical algorithm in order to achieve optimized results. Moving window partial least squares regression (MWPLS; Jiang *et al.*, 2002), group partial least squares (gPLS; Liquet *et al.*, 2016), sparse partial least squares (sPLS; Chung and Keles *et al.*, 2010, and Lê Cao *et al.*, 2008) and sparse group partial least squares (sgPLS; Liquet *et al.*, 2016) are among these attempts. In our proposal, the main idea is based on the fact that in linear regression those variables which are strongly related with the outcome and valuable for its prediction are associated with a large PLS regression (PLSR) coefficient (Wold *et al.*, 2001).

In Section 2 the basics of PCA is presented while Section 3 discuss the proposed feature selection PLS (FS-PLS) method. Some model assessment criteria are provided in Section 4. The application of FS-PLS and its comparison with PCA and PLS for the univariate and multivariate case is provided in Section 5. The paper ends with the concluding remarks and possible future expansions of this work.

2. PRINCIPAL COMPONENT ANALYSIS

Principal Components Analysis (PCA) is one of the most widespread dimensionality reduction techniques. It is a multivariate feature extraction method through which it is achieved the transformation of a data matrix X into a low dimension matrix. The newly generated matrix contains the principal components, i.e. the transformed variables of X matrix, that have been generated as linear combinations of the original variables. These components have the following desirable properties: they are uncorrelated and their number is significantly reduced compared to the corresponding number of the initial matrix. Thus, due to the construction algorithm, they summary the majority of the initial information discharged by noise.

Geometrically, through PCA data points are projected onto a low dimension space, the coordinate system of which is oriented in the directions of maximized variance of data points, the directions of Principal Components. The vectors containing the weights of the original variables in the linear combinations that define these directions are called loadings, while the coordinates of the available observations in the new space are called scores.

Making use of these quantities, an X matrix of size $n \times p$ can be written as:

$$X = T_m P_m^{\intercal} + E,$$

where T, P, E are correspondingly the matrices of scores, loadings and residual errors. The latter expresses the information that is lost through this analysis and it is the cost of the dimensionality reduction process. The index m indicates the dimension of the new space — it is the number of selected Principal Components and it is defined by the analyst.

Finally, the Principal Component Analysis of the original matrix is:

$$X = T_m P_m^{\intercal}.$$

This representation is an approximation of the initial matrix and can be used for modelling.

2.1. Construction algorithm

The main elements of PCA are scores and loadings matrices and they can be computed based on the Eigen-decomposition of either covariance matrix or correlation matrix of the initial variables (Ntotsis and Karagrigoriou, 2021).

Covariance matrix is used when all variables in X express the same measurement unit. In opposite case, correlation matrix is used, since correlation is independent from the scale of the variables. Regardless of the selected matrix, data standardization is highly recommended in the presence of extreme multicollinearity. It is achieved by replacing each x_{ij} element in X matrix by:

$$\frac{x_{ij} - \bar{x}_j}{s_j},$$

where s_j indicates the standard deviation of the j-th variable.

The selection and computation of the selected matrix is followed by the computation of its eigenvectors and its eigenvalues (λ_i). The latter are ordered and based on them, the determination of the number of the retained PCs (m) in the model is possible according to:

- Cumulative Percentage of total variation: The inclusion of PCs is interrupted when the first m of them achieve to absorb 80%–90% of total variation of X. The cumulative percentage of the PCs variation is computed by:

$$\sum_{j=1}^{m} q_j = 100 \frac{\sum_{j=1}^{m} \lambda_j}{\sum_{j=1}^{p} \lambda_j}.$$

- Kaiser's rule: According to this rule the dimension of the summarized matrix is equal to the number of eigenvalues that are larger than one.

- The scree graph: This graph illustrate the rank of eigenvalues against their value. Usually, the curve that connects the points forms an elbow-like shape, with the rank of the point located on its angle indicate the dimension of the new matrix.

The formation of the $p \times m$ loadings matrix P follows, which consists of the eigenvectors that correspond to the greatest m eigenvalues, set as columns. This matrix is useful for the computation of the matrix T, the matrix of Principal Components through the formula:

$$T = XP,$$

where they constitute its columns, and they can be used in linear regression tasks according to the following schema (Wehrens, 2011):

$$Y = XB + E = (TP^{\intercal})B + E' = T(P^{\intercal}B) + E' = TA + E',$$

where A matrix indicates the regression coefficients of components and they arise with the use of Ordinary Least Squares Method:

$$A = (T^{\intercal}T)^{-1}T^{\intercal}Y.$$

These coefficients can be transformed to coefficients of the original variables:

$$B = PA = P(T^{\intercal}T)^{-1}T^{\intercal}Y.$$

3. FEATURE SELECTION PARTIAL LEAST SQUARES (FS-PLS)

The FS-PLS procedure can be considered as an expansion of the original Partial Least Squares Method (PLS). Thus, its definition arises through PLS. The PLS is a dimension reduction technique that achieves feature extraction. It is quite popular in chemistry or chemometrics due to high correlations frequently encountered among variables in those fields. It shares a common approach with PCA: via its algorithm, few new uncorrelated variables that summarize the information in X matrix are produced. They are called latent variables and they arise as linear combinations of the initial ones. The difference in this approach is that these variables aim to summarize information that is directly related to the response variables in a regression problem. To accomplish this, both X and Y matrices are being analyzed through an iterative procedure that generates latent variables (LV) oriented in the direction that maximizes the covariance between the involved matrices. On the contrary, PCA focuses on the explanation of the variability in X matrix independently from Y, an approach that can result in loss of information, that could be proved valuable for prediction purposes. Furthermore, another difference between the two methods is that PLS disposes of a variation, that makes feasible the simultaneous modelling of multiple response variables, where the intercorrelations among them are considered. Instead, PCA runs regression steps multiple times to model multiple responses. PLS is often considered more appropriate than PCA in cases with small sample sizes, multicollinearity and missing data.

From a geometrical point of view, similarly to PCA, X but here also Y data are projected to low dimensional spaces defined by the latent variables. Making use of the scores and loadings of the observations in the new spaces, an X data matrix of size $n \times p$ and a Y data matrix of size $n \times k$ can be written as:

$$X = T_m P_m^{\intercal} + E,$$
$$Y = U_m Q_m^{\intercal} + F,$$

where T and U are score matrices, P and Q are loading matrices and E and F are matrices representing the information loss. The index m expresses the dimension of the new space and it is determined by the analyst. Eventually, the summarized approximation of the initial X matrix through PLS is:

$$X = T_m P_m^\intercal$$

and it can be used in linear regression.

3.1. Construction algorithm

The construction of a PLS model can be completed by a number of relative algorithms. In this article we use the Nonlinear Iterative Partial Least Squares algorithm (NIPALS), which was proposed by Wold *et al.* (2001).

According to the NIPALS algorithm (De Jong, 1993), a PLS model can be produced by a process in every iteration of which a latent variable emerges. More specifically, initially an X score vector (t) and a Y score vector (u) with maximized covariance are generated. Their directions are determined by weight vectors, w and c that correspond to X and Y matrices, respectively:

$$\begin{aligned} t &= Xw, \\ u &= Yc/(c^\intercal c). \end{aligned}$$

Next the computation of a loading vector p follows:

$$p = X^\intercal t/(t^\intercal t),$$

that is used in a deflation process of the X matrix and through which the information explained by the produced latent variable is subtracted:

$$X_{\text{new}} = X_{\text{old}} - tp^\intercal.$$

This deflated matrix is then used in the next iteration of the algorithm for obtaining a new latent variable.

The procedure of emerging new variables is terminated based on the indications of model selection criteria. The computation of their values follows the implementation of the regression step, demonstrating that in PLS dimension reduction and regression run simultaneously. The scheme in Partial Least Squares Regression (PLSR) is identical to Principal Component Regression (PCR):

$$Y = XB + E = (TP^\intercal)B + E' = T(P^\intercal B) + E' = TA + E',$$

where

$$A = (T^\intercal T)^{-1} T^\intercal Y.$$

Here, T and P matrices are formed by the apposition of the output scores vectors (t) and loading vectors (p) as columns, respectively. The i-th column in A matrix contains the regression coefficients for the i-th response variable in Y matrix. These coefficients refer to

the extracted latent variables, although the reference to the original variables is attainable by the formula:

$$B = RA = R(T^{\intercal}T)^{-1}T^{\intercal}Y.$$

The vectors in R matrix represent the weights of every original variable of X matrix at the extracted latent variables, unlike the weight vectors w_i that refer to the deflated matrices X_i and their apposition forms W matrix, which is related to R:

$$R = W(P^{\intercal}W)^{-1}.$$

As stated before, PLS is considered to be an effective dimension reduction technique when it comes to obtaining an optimal statistical model. However, like many similar feature extraction techniques, we end up with a model that involves all original variables, significant, or not. What if there was a way to take advantage of PLS algorithm in order to utilize it as a variable selection technique? The FS-PLS is a novel approach that allows the researcher to use the PLS procedure to remove non-significant variables from the original dataset and obtain a statistically significant model with minimum dimension when PLS is applied. FS-PLS provides a new dataset with simpler structure than the original one and still when its implementation is compared to PLS and PCA, the model arises from FS-PLS is more efficient than the corresponding models of PLS and PCA. This "superiority" is due to the fact that the constructed model of FS-PLS is easier to interpret since all irrelevant variables have been removed.

The beta coefficients (β) that emerge from the PLSR in conjunction with the number of selected latent variables can be seen as a general rule of thumb for disregarding variables from a dataset. We propose the following rule to determine if a variable is significant:

Let us assume a model with X_j, $j = 1, ..., m$, independent variables and let v be the number of latent variables that have been selected as optimal from the PLS regression of the aforementioned model. Let us also assume that β_j^v being the corresponding coefficient of X_j variable in the v-latent variable (each latent contains all original variables). Now, let us define equation (3.1) as follows:

$$|\beta_j^v| \leq c, \tag{3.1}$$

where $c \in \left[0.05, |\max\{\beta_j^v\}/2|\right)$ is a pre-determined non-negative value close to zero and $|\max\{\beta_j^v\}|$ is the maximum (absolute) value that exists in the coefficient matrix of the selected v latent variables. If (3.1) is satisfied for the j-th variable, i.e. $|\beta_j^1| \leq c$, and $|\beta_j^2| \leq c$, ... and $|\beta_j^v| \leq c$, then this variable can be labelled as non-significant. By integrating this β-based constraint in the PLS regression, it is feasible to discard the insignificant variables and still maintain a robust model. A fixed value c is expected to complement effectively all other aspects (purpose of the study, researcher's judgement, etc.) of the decision-making process. In that sense, it can be considered as a rule of thumb and is in the judgement of the researcher which value of c is the one that results the optimal PLS model without underfitting or overfitting the model under consideration. We recommend a step procedure of 0.05 units (i.e. 0.05, 0.10, 0.15, etc.) until model underfitting is observed based on the model selection criteria.

The FS-PLS algorithm consists of a two level implementation of the PLSR algorithm. Initially, PLS method is applied on the original dataset and the regression coefficients of the original variables are estimated with the use of models consisted of up to three latent variables.

Those variables with absolute values of regression coefficients lower than the testing threshold in all three models are considered insignificant for the prediction of response variable and they are removed from the dataset. This distinction between the variables is followed by the application of Partial Least Squares Regression to generate predictive models. Their competency is evaluated based on information criteria, such as AIC, Adjusted R^2 (R^2_{adj}), RMSECV and Adjusted Wold's R criterion, that lead to the final model selection.

The following algorithm displays the proposed procedure:

Algorithm 1 — Pseudocode for FS-PLS.

Input: A data set consisted by a $n \times p$ matrix X and a $n \times 1$ matrix Y, where each X_j and Y column represents a variable, and a constant threshold c.

Output: A data set consisted of the minimum variables that can result in the optimal PLS model.

Step 1: Application of PLSR on original data for the evaluation of regression coefficients.

Step 2: Usage of model selection criteria for number of optimal latent variables determination.

Step 3: Application of the constrain proposed in (3.1) for the location of the statistically insignificant variables.

Step 4: Removal from the input dataset the variables that **Step 3** indicate as insignificant.

Step 5: Repetition of **Step 1** on the minimized original data.

---

4. MODEL ASSESSMENT

In this section we briefly discuss classical model assessment criteria for PLS and/or PCA.

The selection of the optimal number of latent variables to retain in a PLS model is determined on the basis of the following criteria:

- Wold's R criterion: It is a criterion specially designed to evaluate PLSR models by comparing the contribution of a new extracted variable with the previous one, to the predictive ability of the model. For this purpose, cross validation technique is involved to compute Predicted Error Sum of Squares (PRESS) statistic and R ratio as follows (Li *et al.*, 2002):

$$R = \frac{\text{PRESS}(m+1)}{\text{PRESS}(m)},$$

where m denotes the number of retained latent variables in the model. The inclusion of the latent variable that makes R greater than one, terminates the construction algorithm and the produce of new latent variables. The first m of them are then included in the model.

- Adjusted Wold's R criterion: In this permutation of Wold's R criterion the ratio R is compared to the values 0.90 ($R^{0.90}_{\text{adj}}$) and 0.95 ($R^{0.95}_{\text{adj}}$) rather than 1, as in the original version. As it has been proven in Li *et al.* (2002), these variations give better results due to sample variability.

In many cases, when researchers deal with high dimensional datasets, variable selection leads up to the construction of a PLS and/or a PCA model, in order to remove insignificant variables in a preparatory level. As a result, the production of sets of models that differ in the number of predictors they arise from and also differ in terms of complexity, occurs. The selection of the optimal model can emerge from various model selection criteria (Faraway, 2002). The most frequently utilized criteria that one can use when in PCA, PLS and similar techniques have been documented below:

- R-squared (R^2) value: It expresses the percentage of the explained variability in the response variable and it is computed by:

$$R^2 = \frac{\sum_i^n (\hat{y}_i - \bar{y})^2}{\sum_i^n (y_i - \bar{y})^2}.$$

  It varies from 0 to 1 and higher values indicate more sufficient model performance.

- Adjusted R^2 value: It is a modification of R^2 criterion that penalizes models of higher complexity. It is computed by:

$$R^2_{\text{adj}} = 1 - \frac{(n-1)}{(n-p-1)}(1 - R^2),$$

  where n refers to the number of available observations and p to the number of retained components.

- Akaike's Information Criterion: This measure consider both the predictive accuracy and the parsimony of a model to evaluate it, according to the following formula:

$$\text{AIC}_m = -2\,(\textit{maximum log-likelihood}) + 2m,$$

  where m is the number of the retained predictors in the model. Among the comparable models, the most sufficient performance is presented by the model with the lowest AIC value.

- RMSECV: This measure involves cross validation to give an estimation of the variation/divergence of the predicted values from the true values of unseen observations, in lack of available data that could be used as a test set. The criterion uses the cross-validation approach and its value is computed as:

$$\text{RMSECV} = \sqrt{\frac{\sum_j \frac{\sum_i (y_{ij} - \hat{y}_{ij})^2}{N_j}}{k}},$$

  where $\hat{y}_{ij}$ is the estimation of y_{ij}, N_j is the number of observations in the j-th fold and k is the number of folds in cross validation procedure. Lower RMSECV values indicate better predictive capacity of the compared models.

By utilizing R^2_{adj}, we aimed to assess the models' ability to explain the variability in the dependent variable while considering the complexity of the regression models. This metric allows for a fair comparison between the PLS and PCA regression models, taking into account their respective number of latent variables or principal components and potential differences in the predictive performance. Moreover, adjusted R-squared aligns with the principle of parsimony, which encourages the use of simpler models that explain the data effectively. The inclusion of adjusted R-squared as one of our evaluation criteria supports the selection of models that strike a balance between explanatory power and model complexity.

We acknowledge that R^2_{adj} should not be the sole criterion for model evaluation, and we have also incorporated other well-established metrics such as AIC, Wold's R criterion, and RMSECV. These metrics provide complementary insights into model performance, including goodness of fit, model complexity, and prediction accuracy.

By utilizing a combination of these metrics, we aimed to provide a comprehensive evaluation of the PLS and PCA regression models, taking into consideration various aspects of model performance. This approach ensures a robust and thorough assessment of the models and allows for informed comparisons between them.

To address the issue of information control in our proposed methodology, we evaluate the same chemometrics datasets that have been extensively used in previous studies on PLS regression. These studies, have already addressed the concern of information preservation within PLS models. For instance, in first row of Table 1 (case where c is '—'), we present a typical PLS model used as a baseline in previous works. Our goal is to simplify this model by removing variables while maintaining its information content and robustness. To ensure information control, we employ the Wold's R criterion, specifically developed for evaluating PLS models. Additionally, we utilize other established metrics such as AIC, adjusted R-squared, and RMSECV to comprehensively assess the performance of our methodology. By incorporating these evaluation metrics and techniques, we ensure that the simplification process retains the essential information captured by the original PLS model. This allows us to strike a balance between model complexity and interpretability while preserving the predictive performance of the PLS regression. It is important to note that the information control aspect has already been addressed in the relevant literature on PLS regression, which serves as the foundation for our work.

5. NUMERICAL APPLICATIONS

In this section the application of the FS-PLS on near infrared (NIR) spectroscopy data is presented.

5.1. Univariate FS-PLS regression — FS-PLSR

In the first case, in the *gasoline* dataset, which is found in the `pls` package, X matrix includes 401 diffuse reflectance measurements and Y matrix is consisted of one response variable, that corresponds to the number of octanes of the total 60 observations (Kalivas, 1997).

Due to the multicollinearity and the rate of available observations to X-variables, dimensionality reduction is demanded in order to generate a linear regression model. Applying the FS-PLS optimization, we first computed the estimators of PLS-regression coefficients of all 401 variables in models built with up to three components. Their absolute values were then compared with predefined constant c of 0.10, 0.20, 0.25 and 0.30. The final X data matrices contextually included only the predictive variables with absolute values of PLS-coefficients higher than the testing threshold in one-, two- and three-component models (1LV, 2LV, and 3LV). In the next step we reapplied the PLSR method to the selected variables and the resulted models were evaluated based on AIC, R^2_{adj}, $R^{0.90}_{\text{adj}}$, and $R^{0.95}_{\text{adj}}$ and RMSECV. Table 1 and Table 2 summarize the results.

Table 1: Information criteria values of FS-PLSR models, where Attributes is the number of original variables.

c	Attributes	$R^{0.90}_{\text{adj}}$	$R^{0.95}_{\text{adj}}$	RMSECV	1 LV		2 LV		3 LV		4 LV		5 LV		6 LV	
					AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}
—	401	4	4	7	203	30%	52	94%	3	97%	−2	97%	−25	98%	−36	98%
0.10	374	4	4	7	203	30%	52	94%	4	97%	−3	97%	−25	98%	−36	98%
0.20	307	4	4	7	203	31%	52	94%	6	97%	−3	97%	−24	98%	−35	98%
0.25	245	4	6	6	202	31%	51	94%	9	97%	−4	98%	−23	98%	−35	98%
0.30	217	4	4	6	202	31%	51	94%	12	97%	−5	98%	−22	98%	−35	98%

In Table 1, the reduction in AIC values in all two-component models and the simultaneous increase of their R^2_{adj} values is noteworthy. These changes strongly indicate the outstanding enhancement of the corresponding models when the second component is retained in the model. Further, the most sufficient FS-PLSR model is proposed, the four-component model, which is based on the 0.30 testing threshold and it includes only 217 variables in X matrix, which consist of 46% of the initial observations. This choice is established in accordance with the adjusted Wold criterion, which is specialized to evaluate PLS models, complemented by the high R^2_{adj} value and the significant reduction in AIC value. It should be noted that AIC values tend to decrease as more components are added to the model. However, the rate of decrease is approximately fixed after the addition of the fourth component. Moreover, the criteria values of the models that resulted from the thresholds 0.25 and 0.30 are alike, though the latter constraint conveys to further dimensional reduction. At this point, it should be mentioned that more restrictive thresholds were tested; they were found to lead to over-fitted models and rejected.

The results of the PCA regression (PCR) models, generated with the datasets arising from the aforementioned thresholds, are displayed in Table 2. As the most adequate model is proposed the five-component model of the last threshold, since R^2_{adj} value is close to 1 and AIC value does not change sufficiently with the addition of more components in the model. In contradiction to the FS-PLSR models, the inclusion of the second component does not improve the model performance in any case, while the minimization of RMSECV values proposes much more complicated models than in FS-PLSR cases.

Table 2: Information criteria values of PCR models.

c	RMSECV	1 LV		2 LV		3 LV		4 LV		5 LV		6 LV	
		AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}
—	17	213	17%	215	17%	192	43%	6	97%	6	97%	8	97%
0.10	17	213	17%	215	17%	189	46%	6	97%	6	97%	7	97%
0.20	15	213	17%	214	17%	174	58%	6	97%	5	97%	7	97%
0.25	15	213	17%	214	18%	147	73%	7	97%	5	97%	5	97%
0.30	14	213	17%	214	18%	133	79%	7	97%	5	97%	5	97%

Artigue and Smith (2019) have examined several challenges related to PCR. However, our study specifically focuses on PLSR, which addresses a key limitation of PCR by considering both the independent and dependent variables, thereby offering a solution to the mentioned issue which is one of the main arguments of the article.

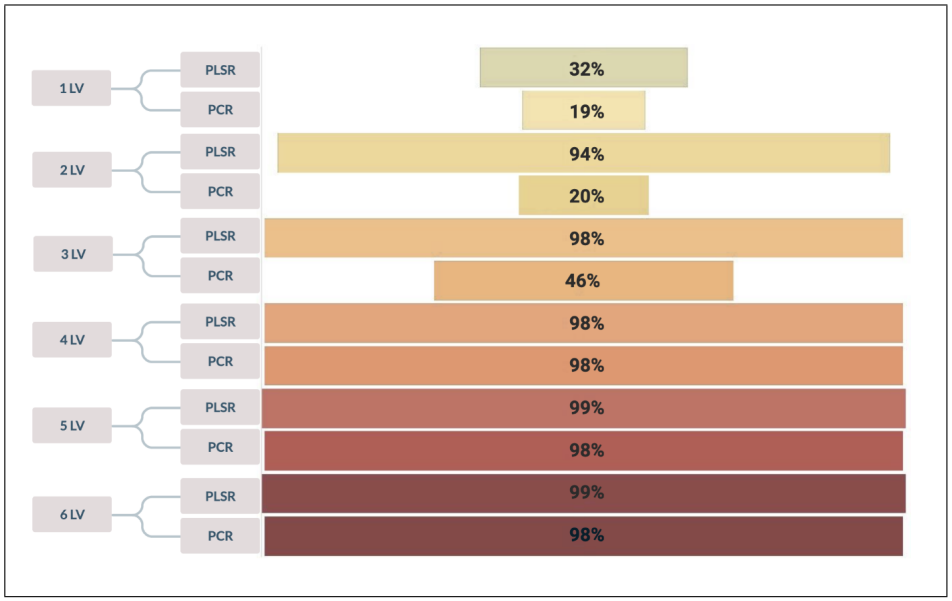

Figure 1: Percentage of explained variability of FS-PLSR and PCR models.

Additionally, taking into account the R^2_{adj} criterion and the percentages of explained variability in the models, as displayed in Figure 1, we conclude that in FS-PLSR the two-component and three-component models can lead to reliable results, preserving the advantage of visualization. Note that all c constrains resulted in similar explained variability and thus only one's results are being presented in Figure 1. These FS-PLSR models expose high R^2_{adj} values, while they leave unexplained a negligible percentage of the response variable. In PCR instead, the inclusion of the first four components fails to provide a model with sufficient performance. Finally, the comparison of these methods in terms of AIC values verifies the predominance of FS-PLSR against PCR: all AIC values in PCR models (with up to three components) are significantly smaller than the corresponding FS-PLSR model (Table 2).

5.2. Multivariate FS-PLS regression — FS-MPLSR

In the second case, the performance of the proposed FS-MPLSR optimization over a multivariate response is investigated (the multivariate case of FS-PLSR will be addressed by FS-MPLSR in the remaining article). In the corn data set, that we processed, the Y matrix consists of four variables, -moisture, oil, protein and starch, and X matrix includes 700 NIR spectroscopic attributes. Note that the FS-MPLSR algorithmic procedure is similar to the FS-PLSR with the only deference to be the number of response variables that form the latent variables. In the multivariate case, the modelling process aims to reveal and enable chemists to predict the moisture, oil, protein and starch content in different samples. In this situation, the implementation of Ordinary Least Squares as a linear regression method would be an inappropriate choice, since the X matrix is characterized by the existence of multicollinearity. Its mitigation is achieved through dimensional reduction, based on the absolute values of the FS-PLS-regression coefficients, in a similar way as in the univariate case. The FS-MPLSR algorithm was applied on the initial dataset to estimate these values. The computation of Adjusted Wold's R criterion $R_{\text{adj}}^{0.90}$ led to the conclusion that the sufficient modelling of the four Y-variables requires the inclusion of first 5, 21, 7 and 8 FS-PLS-components respectively. Based on this conclusion and the use of testing thresholds we defined the final reduced set of predictors as the intersection of the following four subsets:

- The first subset included the variables considered as statistically significant for $Y1$. The absolute values of regression coefficients of these variables are higher than the tested thresholds in one- to five-component models.
- The second subset included the variables considered as statistically significant for $Y2$. The absolute values of regression coefficients of these variables are higher than the tested thresholds in one- to twenty one-component models.
- The third subset included the variables considered as statistically significant for $Y3$. The absolute values of regression coefficients of these variables are higher than the tested thresholds in one- to seven-component models.
- The fourth subset included the variables considered as statistically significant for $Y4$. The absolute values of regression coefficients of these variables are higher than the tested thresholds in one- to eight-component models.

The thresholds that we tested were 2, 2.25, 2.50 and they resulted in the removal of 44, 69, and 99 variables from the original dataset, correspondingly. The new reduced data matrices were then processed via the FS-MPLSR and PCR methods. Based on the values of the aforementioned model selection criteria we inferred that the third threshold examined (2.50) generated the most efficient models. Table 3 and Table 4 summarize the values. The other options led to over-fitted or under-fitted models.

The optimum FS-MPLSR model retained twenty one components. The model complexity was determined in accordance with the theory (Wold *et al.*, 2001), which states that the FS-MPLSR model should include every component that is found to be significant for at least one variable of the set of responses. In this way information in Y matrix is significantly explained, as the percentages of the explained variability are 99.95% for $Y1$, 97.06% for $Y2$, 99.43% for $Y3$ and 99.55% for $Y4$, while the overall information of the new X matrix is utilized.

Table 3: Information criteria values of the FS-MPLSR model based on the remaining 601 attributes (700-99).

	1 LV		2 LV		5 LV		7 LV		8 LV		omitted LV	21 LV	
	AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}	⋯	AIC	R^2_{adj}
Y1	1	51%	2	52%	−50	80%	−142	96%	−166	97%	⋯	−366	99%
Y2	−62	27%	−62	27%	−68	37%	−87	55%	−154	86%	⋯	−214	95%
Y3	76	17%	65	32%	−35	88%	−56	92%	−60	92%	⋯	−181	99%
Y4	153	0%	145	13%	63	78%	28	88%	7	92%	⋯	−131	99%

Table 4: Information criteria values of the PCR model based on the remaining 601 attributes.

	1 LV		2 LV		5 LV		7 LV		8 LV		omitted LV	21 LV	
	AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}	⋯	AIC	R^2_{adj}
Y1	1	52%	1	53%	−115	93%	−153	96%	−151	96%	⋯	−371	99%
Y2	−62	27%	−61	26%	−69	38%	−116	72%	−124	76%	⋯	−158	88%
Y3	76	17%	74	22%	22	68%	−24	86%	−56	92%	⋯	−112	97%
Y4	153	0%	151	0%	120	44%	76	74%	52	83%	⋯	−47	97%

We can infer that the substantial dimensionality reduction that we achieved through the PLS-optimization, resulted in the generation of a unique model capable to predict the four responses at the same time, with the cost of an insignificant percentage of unexplained information. Nevertheless, a less strict consideration of the theoretical frame would yield an eight-component model with the profit of further dimensionality reduction and with the cost of a less accurate, but yet sufficient, prediction of $Y2$ response variable.

On the contrary, the PCR method generated four individual models, one for the prediction of each response variable, that needed twenty one components to capture 99.95%, 92.51%, 98.21% and 98.17% of the variability of the responses. These percentages, in combination with the results of the information criteria presented in Table 3, demonstrate that FS-MPLSR model is more adequate in all four responses.

6. CONCLUDING REMARKS

The aim of this study is to introduce PLS as a method for variable selection in a variety of fields, including time series analysis. Although this method is commonly used in a regression analysis, it can also be implemented in various other applications such as discriminant analysis, and hierarchical modelling. It can handle complex data sets and situations that cannot be solved by standard methods.

FS-PLS is considered optimal in evaluating more complex structures with a more realistic and holistic view. It has been proved to be a non-time consuming process and statistically efficient method with high prediction accuracy. As a recently found technique in the field, many aspects of its underlying mechanism have recently been revealed and yet, there is no strictly defined frame for its application. As a result, the method is considered to be very flexible and many modifications and experimentations can be tested. In this work the utilization of PLS approach was used as a variable selection criterion and by expansion as a dimension reduction technique. The FS-PLS procedure was able to remove up to 45% and 14% of the original variables in two frequently used datasets in chemometrics, one univariate set and one more complex multivariate one.

Although PLS is considered to be useful in small datasets, through the FS-PLS methodology it has been found to be useful in high-dimensional and/or big data analysis. Although, the applications is chosen from the field of chemometrics, the applicability was quite wide covering biology, physics, chemistry, business, and social sciences among others.

In the univariate case, the final selected model is based on only 217 predictors out of an initial set of 401. The three-component model, which is suggested as optimum, explains the major part of information captured in the data, while it is parsimonious, with high prediction ability and can easily be used for visualizations. The comparison with the corresponding PCR model, which was based on information criteria AIC, R^2_{adj}, and RMSECV, demonstrates that FS-PLSR model gave more sufficient results.

In the multivariate case, the problem appears to be more complicated. Initially FS-MPLSR was implemented on the data out of necessity, due to the fact that correlations were observed between the response variables. We estimated regression coefficients and we determined the significant components for each response variable. We compared the absolute values of the coefficients in significant components with thresholds and then, we defined four sets of predictors, which contained the important predictors for the individual responses, respectively. Their intersection consisted the final set of predictors for the multivariate regression model. This way, in the final selected model 99 less predictors than in the initial set were included. The simultaneous process of the response variables generated a single regression model with AIC values lower than the individual PCR models in all four response variables. The increased number of constructed models in the PCR method is associated with high complexity and computational cost of the whole analysis. This, in combination with the fact that less variability is explained in the second response variable with the PCR method, leads to the suggestion that a FS-PLSR model is optimum also in the multivariate case.

While our work primarily focuses on the predictive performance and feature selection aspects of the proposed FS-PLS methodology, we acknowledge the importance of addressing the issue of interpretability in regression analysis. It is worth noting that interpretability is a complex aspect in high-dimensional settings, and various techniques have been proposed in the literature to enhance it. In future research, we can explore the application of these techniques, such as sparse PLS (SPLS), non-negative matrix factorization (NMF), and independent component analysis (ICA) have been proposed in the literature to enhance interpretability in high-dimensional settings and improve techniques like PCA and PLS. By incorporating these approaches, we can potentially provide a more comprehensive analysis that combines predictive accuracy, feature selection, and enhanced interpretability. This avenue of research holds promise for advancing the field of PLS regression and its applicability in practical domains.

ACKNOWLEDGMENTS

We express our appreciation to the two referees and the Associate Editor for providing valuable suggestions and recommendations that greatly improved both the quality and the presentation of the manuscript. Furthermore, we gratefully acknowledge the invaluable guidance and supervision of Professor Alex Karagrigoriou from the Laboratory of Statistics and Data Analysis, which greatly contributed to the successful completion of this work.

REFERENCES

Ajana, S., Acar, N., Bretillon, L., Hejblum, B. P., Jacqmin-Gadda, H., and Delcourt, C. (2019). Benefits of dimension reduction in penalized regression methods for high-dimensional grouped data: a case study in low sample size. *Bioinformatics*, 35(19):3628–3634.

Artigue, H. and Smith, G. (2019). The principal problem with principal components regression. *Cogent Mathematics and Statistics*, 6: 1622190.

Chung, D. and Keles, S. (2010). Sparse Partial Least Squares classification for high dimensional data. *Statistical Applications in Genetics and Molecular Biology*, 9(1): Article 17.

Corn data set, available from: `http://software.eigenvector.com/data/index.html`

De Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3):251–263.

Faraway, J. J. (2002). *Practical Regression and Anova using R*. Retrieved November 1, 2020, from: `https://cran.r-project.org/doc/contrib/Faraway-PRA.pdf`

Höskuldsson, A. (1988). PLS regression methods. *Journal of Chemometrics*, 2(3):211–228.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*, 1st edition. Springer.

Jiang, J.-H., Berry, R. J., Siesler, H. W., and Ozaki, Y. (2002). Wavelength interval selection in multicomponent spectral analysis by moving window Partial Least-Squares regression with applications to mid-infrared and near-infrared spectroscopic data. *Analytical Chemistry*, 74(14):3555–3565.

Kalivas, J. H. (1997). Two data sets of near infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, 37(2):255–259.

Lê Cao, K. A., Rossouw, D., Robert-Granié, C., and Besse, P. (2008). A sparse PLS for variable selection when integrating omics data. *Statistical Applications in Genetics and Molecular Biology*, 7(1): Art. 35.

Li, B., Morris, J., and Martin, E. B. (2002). Model selection for partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 64(1):79–89.

Li, G.-Z. and Zeng, X. Q. (2009). Feature selection for Partial Least Square based dimension reduction. *Foundations of Computational Intelligence*, 5:3–37.

Liquet, B., de Micheaux, P. L., Hejblum, B. P., and Thiébaut, R. (2016). Group and sparse group partial least square approaches applied in genomics context. *Bioinformatics*, 32(1):35–42.

Ntotsis, K., Artemiou, A., and Karagrigoriou, A. (2021). Interdependency pattern recognition in econometrics: a penalized regularization antidote. *Econometrics*, 9:44.

Ntotsis, K. and Karagrigoriou, A. (2021). *The impact of multicollinearity on big data multivariate analysis modeling*. In "Applied Modeling Techniques and Data Analysis 1: Computational Data Analysis Methods and Tools", iSTE Wiley and Sons, 187–202.

Wehrens, H. R. M. J. (2011). *Chemometrics with R: Multivariate Data Analysis in the Natural Sciences and Life Sciences*. Springer, New York.

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130.

REVSTAT – Statistical Journal
Volume 23, Number 3, July 2025, 341–358
https://doi.org/10.57805/revstat.v23i3.580

Longitudinal Data Regression Analysis Using Semiparametric Modelling

Authors: Abdulla Mamun
– Department of Mathematics, Gonzaga University, Spokane, WA, USA
mamun@gonzaga.edu

Sudhir Paul ✉
– Department of Mathematics and Statistics, University of Windsor, Windsor, ON, Canada
smjp@uwindsor.ca

Received: May 2022 Revised: July 2023 Accepted: July 2023

Abstract:

- Zhang, Leng and Tang (2015) propose joint parametric modelling of the means, variances, and the correlations by decomposing the correlation matrix via hyperspherical co-ordinates and show that this results unconstrained parameterization, fast computation, easy interpretation of the parameters, and model parsimony. With unconstrained structures, they also suggest future research on modelling the mean, the variance, and the correlations non-parametrically and semiparametrically. In this paper we explore semiparametric modelling via simulations and data analysis. Extensive simulations show that the semiparametric modelling produces similar bias and efficiency properties of the parameter estimates as those by the parametric modelling. However, model selection, using the AIC and the BIC, through the analysis of two real biomedical data sets show significant improvement in model parsimony.

Keywords:

- *B-spline; hyperspherical co-ordinates; joint mean-covariance models; longitudinal data; model parsimony; penalized spline; semiparametric models.*

AMS Subject Classification:

- 62F99, 62G99.

✉ Corresponding author

1. INTRODUCTION

In the statistical literature, the methods to understand the relationship of explanatory variables on each individual outcome variable are well developed and widely applied. However, in most health-related studies given the technological advancement and sophisticated methods of obtaining and storing data, a need to perform joint analysis of mean and covariance parameters simultaneously and accounting for the correlations is in high demand since a good covariance modelling approach improves statistical inference of the mean of interest (Liang and Zeger, 1986; Diggle *et al.*, 2002; He *et al.*, 2005). Furthermore, the covariance structure itself may be of scientific interest (Lin and Carroll, 2006).

Model parsimony in regression analysis, specially in the context of longitudinal data regression analysis, is important in biomedical fields. Zhang *et al.* (2015) propose joint parametric modelling of the means, the variances, and the correlations by decomposing the correlation matrix via hyperspherical co-ordinates and show that this results unconstrained parameterization, fast computation, easy interpretation of the parameters, and model parsimony. Zhang *et al.* (2015, pp. 237) comment that the decomposition of the correlation matrix opens many new avenues for future research and that with unconstrained structures, we can model the mean, the variance, and the correlation non-parametrically and semiparametrically. In this paper we deal with the semiparametric modelling. Our main aim, though, is to find whether semiparametric modelling improves model parsimony over the parametric modelling approach.

Suppose longitudinal measurements $\mathbf{y}_i = (y_{i1}, ..., y_{im_i})'$ and covariate vectors $(\mathbf{x}_{i1}, ..., \mathbf{x}_{im_i})'$ $(i = 1, ..., n)$, with $\mathbf{x}_{ij} = (x_{ij1}, ..., x_{ijp})'$ for $j = 1, ..., m_i$, collected from n subjects, are observed at times $\mathbf{t}_i = (t_{i1}, ..., t_{im_i})'$. In longitudinal data analysis it is important that statistical analysis takes into account that the repeated observations y_{ij}, $j = 1, ..., m_i$, are correlated (Liang and Zeger, 1986; Diggle *et al.*, 2002; He *et al.*, 2005). Accordingly we assume that $\mathbf{y}_i \sim N(\boldsymbol{\mu}_i, \Sigma_i)$, where $\boldsymbol{\mu}_i = (\mu_{i1}, ..., \mu_{im_i})'$, $\Sigma_i = D_i R_i D_i$, $D_i = \text{diag}(\sigma_{i1}, ..., \sigma_{im_i})$, and $R_i = (\rho_{ijk})_{j,k=1}^{m_i}$ is the correlation matrix of $\mathbf{y}_i$ with $\rho_{ijk} = \text{corr}(y_{ij}, y_{ik})$ being the correlation between the j^{th} and k^{th} measurements of the i^{th} subject. The main purpose in such longitudinal studies is to estimate the parameters involved in the means, the variances, and the correlation matrices. This can be done by maximizing the log-likelihood or by solving the maximum likelihood estimating equations. However, the constraints involved in the correlation parameters create a challenge. This can be overcome by decomposing the correlation matrix by Cholesky decomposition.

Zhang *et al.* (2015) proposed to parametrize the correlation matrix R_i for subject i (we suppress i) via hyperspherical co-ordinates by the Cholesky decomposition $R = TT'$, where $T = (T_{jk})$ is a lower triangular matrix given by

$$T = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ c_{21} & s_{21} & 0 & \cdots & 0 \\ c_{31} & c_{32}s_{31} & s_{32}s_{31} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ c_{m1} & c_{m2}s_{m1} & c_{m3}s_{m2}s_{m1} & \cdots & \prod_{l=1}^{m-1} s_{ml} \end{pmatrix},$$

with $c_{jk} = \cos(\phi_{jk})$ and $s_{jk} = \sin(\phi_{jk})$ are trigonometric functions of angles ϕ_{jk}.

For subject i, the total number of angles ϕ_{ijk} $(1 \leq k < j \leq m_i)$ is $m_i(m_i-1)/2$, which is the same as that of the free parameters in the correlation matrix. The decomposition of R has several advantages:

(**i**) diagonal elements of TT' are 1, and all other elements fall between -1 and 1;

(**ii**) TT' is always non-negative definite, satisfying the requirements of a correlation matrix;

(**iii**) the angles ϕ_{jk} of T as parameters are unconstrained in the range $[0, \pi)$.

It also establishes a hierarchical connection between the correlations and the angles (for further discussion on this see Zhang *et al.*, 2015). They then propose a joint regression model for the means, the variances, and the correlations as

$$g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta}, \quad \log \sigma^2_{ij} = \mathbf{z}'_{ij}\boldsymbol{\lambda}, \quad \text{and} \quad \phi_{ijk} = \mathbf{w}'_{ijk}\boldsymbol{\gamma}, \tag{1.1}$$

where $\mathbf{x}_{ij}$ are the usual known covariates as mentioned earlier, $\mathbf{z}_{ij}$ and $\mathbf{w}_{ijk}$ may contain baseline covariates, as well as polynomials in time (time related to longitudinal data) and their interactions. The unknown regression parameters $\boldsymbol{\beta}, \boldsymbol{\lambda}$, and $\boldsymbol{\gamma}$ are of dimensions $p \times 1$, $d \times 1$, and $q \times 1$, respectively. In practice we may choose $\mathbf{w}_{ijk}$ as a polynomial of time lag $(t_{ik} - t_{ij})$. Zhang *et al.* (2015) estimate the parameters $\boldsymbol{\beta}, \boldsymbol{\lambda}$, and $\boldsymbol{\gamma}$ of the model in equation (1) via quasi-Fisher scoring algorithm, a review of which is given in Section 2.

For investigating the properties of the estimates of the regression parameters through semiparametric modelling of the means and variances and to study the impact of this to model parsimony we consider three models:

Model 1: the parametric model (Zhang *et al.*, 2015) given above (equation (1.1));

Model 2: a model in which only the means are modelled semiparametrically, which is,

$$g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + f_1(t_{ij}), \quad \log(\sigma^2_{ij}) = \mathbf{z}'_{ij}\boldsymbol{\lambda}, \quad \text{and} \quad \phi_{ijk} = \mathbf{w}'_{ijk}\boldsymbol{\gamma};$$

Model 3: a model in which the means and the variances are modelled semiparametrically, which is,

$$g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + f_1(t_{ij}), \quad \log(\sigma^2_{ij}) = \mathbf{z}'_{ij}\boldsymbol{\lambda} + f_2(t_{ij}), \quad \text{and} \quad \phi_{ijk} = \mathbf{w}'_{ijk}\boldsymbol{\gamma}.$$

In Model 2 and Model 3, $f_1(\cdot)$ and $f_2(\cdot)$ are smooth functions parametrized by regression splines.

As in Zhang *et al.* (2015) we decompose the correlation matrix via hyperspherical coordinates and as in Leng, Zhang and Pan (2010) we use B-spline to estimate the unknown functions $f_1(\cdot)$ in Model 2 and $f_1(\cdot)$ and $f_2(\cdot)$ in Model 3. Five further investigations were conducted. First of these is to see the performance of the estimators of the regression parameters in terms of bias and efficiency and to see the effect of fixing the knots in spline smoothing. The second is to study the performance of the estimators of the regression parameters when some covariates are correlated, the third is a robustness study where the normality assumption of the error distribution is replaced by a mixture of normal distributions, the fourth is to see whether use of penalized spline results in improved estimation of the non-parametric functions in comparison to using B-spline, and the fifth is to see the effects of estimators by fitting the data generated from the semiparametric model to the parametric model and vice versa.

Section 2 deals with the method of estimation of the parameters of Model 3 in which the correlation matrix is decomposed via hyperspherical co-ordinates and the unknown functions $f_1(\cdot)$ and $f_2(\cdot)$ are estimated using B-spline basis functions. Estimation procedures in Model 2 and Model 1 are discussed as special cases of Model 3. An extensive simulation study is conducted and results are summarized in Section 3. Detailed analysis of two real data sets is given in Section 4 and a discussion follows in Section 5.

2. ESTIMATION IN JOINT SEMIPARAMETRIC MODELS

2.1. Estimation in Model 3 based on B-spline

We estimate the regression parameters of Model 3 (given above) where the means and the variances are modelled semiparametrically. As in Zhang *et al.* (2015), we parametrize R_i via hyperspherical co-ordinates. For simplicity, we assume that f_1 and f_2 have the same smoothness property. Without loss of generality, we assume that the domain of t_{ij} is in the interval $[0, 1]$ with partitions $0 = a_0 < a_1 < \cdots < a_{k_n} < a_{k_n+1} = 1$. Using the a_i's as knots, we have $K = k_n + l$ normalized B-spline basis functions of order l that form a basis for the linear spline space. The B-spline basis of order l, $B_i^{(l)}(t)$, is defined as

$$B_i^{(l)}(t) = \frac{t - a_i}{a_{i+l-1} - a_i} B_i^{(l-1)}(t) + \frac{a_{i+l} - t}{a_{i+l} - a_{i+1}} B_{i+1}^{(l-1)}(t),$$

and

$$B_i^{(1)}(t) = \begin{cases} 1, & a_i \le t < a_{i+1}, \\ 0, & \text{otherwise.} \end{cases}$$

Note that $B_i^{(l)}(t)$ is a polynomial function of degree $l - 1$. More details on the construction of B-spline basis can be found in Schumaker (1981). Thus $f_1(t)$ and $f_2(t)$ are approximated by $\boldsymbol{\pi}'(t)\boldsymbol{\alpha}$ and $\boldsymbol{\pi}'(t)\tilde{\boldsymbol{\alpha}}$, respectively, where $\boldsymbol{\pi}(t) = (B_1^{(l)}(t), ..., B_K^{(l)}(t))'$ is the vector of basis functions and $\boldsymbol{\alpha}, \tilde{\boldsymbol{\alpha}} \in \mathbb{R}^K$ are the spline coefficient vectors. Let $\boldsymbol{\pi}_{ij} = \boldsymbol{\pi}(t_{ij})$. With this notation, the nonlinear regression models can be linearized as in what follows:

$$g(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{\pi}'(t_{ij})\boldsymbol{\alpha} = \Pi'_{ij}\boldsymbol{\theta} \quad \text{and} \quad \log(\sigma_{ij}^2) = \mathbf{z}'_{ij}\boldsymbol{\lambda} + \boldsymbol{\pi}'(t_{ij})\tilde{\boldsymbol{\alpha}} = \Upsilon'_{ij}\boldsymbol{\rho},$$

where $\Pi'_{ij} = (\mathbf{x}'_{ij}, \boldsymbol{\pi}'_{ij})$, $\Upsilon'_{ij} = (\mathbf{z}'_{ij}, \boldsymbol{\pi}'_{ij})$, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$, and $\boldsymbol{\rho} = (\boldsymbol{\lambda}', \tilde{\boldsymbol{\alpha}}')'$. Suppose $\boldsymbol{\Pi}_i = (\Pi'_{i1}, \Pi'_{i2}, ..., \Pi'_{im_i})'$ and $\boldsymbol{\Upsilon}_i = \left(\Upsilon'_{i1}, \Upsilon'_{i2}, ..., \Upsilon'_{im_i}\right)'$.

Thus, now the parameters of interest are $\boldsymbol{\theta}$, $\boldsymbol{\rho}$, and $\boldsymbol{\gamma}$. Let $\mathbf{r}_i = \mathbf{y}_i - \boldsymbol{\mu}_i$. Then, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, ..., \epsilon_{im_i})' = T_i^{-1} D_i^{-1} \mathbf{r}_i \sim N(\mathbf{0}, I_{m_i})$. Denoting l to be the log-likelihood apart from a constant it can be shown that

$$-2l = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left(\log \sigma_{ij}^2 + \log T_{ijj}^2 + \epsilon_{ij}^2 \right).$$

From this, omitting details, by usual derivations, we can show that the maximum likelihood estimating equations for $\boldsymbol{\theta}, \boldsymbol{\rho}$, and $\boldsymbol{\gamma}$ are

$$\begin{aligned}
\boldsymbol{U}_1 &= \sum_{i=1}^{n} \boldsymbol{\Pi}_i' \Delta_i \Sigma_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\Pi}_i \boldsymbol{\theta})) = 0, \\
\boldsymbol{U}_2 &= \frac{1}{2} \sum_{i=1}^{n} \boldsymbol{\Upsilon}_i' (\mathbf{q}_i - \mathbf{1}_{m_i}) = 0, \\
\text{and} \quad \boldsymbol{U}_3 &= \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left[\frac{\partial \log T_{ijj}}{\partial \boldsymbol{\gamma}} (\epsilon_{ij}^2 - 1) + \epsilon_{ij} \sum_{k=1}^{j-1} b_{ijk} \epsilon_{ik} \right] = 0,
\end{aligned}$$

respectively, where $\Delta_i = \Delta_i(\boldsymbol{\Pi}_i \boldsymbol{\theta}) = \text{diag}\{\dot{g}^{-1}(\Pi_{i1}' \boldsymbol{\theta}), ..., \dot{g}^{-1}(\Pi_{im_i}' \boldsymbol{\theta})\}$, $\dot{g}^{-1}(\cdot)$ is the derivative of the inverse link function $g^{-1}(\cdot)$, $\mu(\cdot) = g^{-1}(\cdot)$, $\mathbf{q}_i = \text{diag}(R_i^{-1} D_i^{-1} \mathbf{r}_i \mathbf{r}_i' D_i^{-1})$, and $b_{ijk} = \sum_{l=k}^{j} \frac{\partial T_{ilk}}{\partial \gamma} a_{ijl}$ with a_{ijl} being the (j, l) element of T_i^{-1}. As in Zhang *et al.* (2015) these equations are solved by the quasi-Fisher scoring algorithm which is described in Appendix A.

2.2. Estimation in Model 2 based on B-spline

If we consider a semiparametric model with only the mean having semiparametric term as $g(\mu_{ij}) = \mathbf{x}_{ij}' \boldsymbol{\beta} + f_1(t_{ij})$, $\log(\sigma_{ij}^2) = \mathbf{z}_{ij}' \boldsymbol{\lambda}$, and $\phi_{ijk} = \mathbf{w}_{ijk}' \boldsymbol{\gamma}$, then we need to estimate $\boldsymbol{\theta}, \boldsymbol{\lambda}$, and $\boldsymbol{\gamma}$ which are obtained by solving

$$\begin{aligned}
\boldsymbol{V}_1 &= \sum_{i=1}^{n} \boldsymbol{\Pi}_i' \Delta_i \Sigma_i^{-1} \big(\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\Pi}_i \boldsymbol{\theta})\big) = 0, \\
\boldsymbol{V}_2 &= \frac{1}{2} \sum_{i=1}^{n} \mathbf{Z}_i' (\mathbf{q}_i - \mathbf{1}_{m_i}) = 0, \\
\text{and} \quad \boldsymbol{V}_3 &= \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left[\frac{\partial \log T_{ijj}}{\partial \boldsymbol{\gamma}} (\epsilon_{ij}^2 - 1) + \epsilon_{ij} \sum_{k=1}^{j-1} b_{ijk} \epsilon_{ik} \right] = 0.
\end{aligned}$$

These equations can also be solved using the algorithm in Appendix A. At convergence the variance-covariance of $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\lambda}}$, and $\hat{\boldsymbol{\gamma}}$ are obtained by inverting the Fisher information matrix given in Appendix B. Parameters of Model 1, are estimated, of course, by setting $f_1(t_{i,j}) = 0$ in the above three equations.

2.3. Estimation in Model 2 based on penalized Spline

The B-spline methodology, in some applications, produces over fitting of the data (Carroll *et al.*, 1999). In such cases penalized spline has been used to overcome this (Eilers and Marx, 1996). So, here, we further use the penalized spline in Model 2 instead of the B-spline to see whether it produces improvement in estimation of the non-parametric functions. As in Section 2.2, $f_1(\cdot)$ can be approximated by $\boldsymbol{\pi}'(t)\boldsymbol{\alpha}$. Now, we impose a penalization upon the parameters $\alpha_1, \alpha_2, ..., \alpha_K$, so that they are constrained such that $\sum_{i=1}^{K} \alpha_i^2 \leq C$.

With this constraint the log-likelihood apart from a constant can be written as

$$-2l = \sum_{i=1}^{n}\Big[\log|\Sigma_i| + \big(\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\Pi}_i\boldsymbol{\theta})\big)'\,\Sigma_i^1\big(\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\Pi}_i\boldsymbol{\theta})\big)\Big] - \tau\boldsymbol{\theta}'D\boldsymbol{\theta},$$

where $\tau > 0$ is a constant and $D = \begin{bmatrix} 0_{p\times p} & 0_{p\times K} \\ 0_{K\times p} & I_{K\times K} \end{bmatrix}$.

Using Lagrange multiplier the score equations for $\boldsymbol{\theta}, \boldsymbol{\lambda}$, and $\boldsymbol{\gamma}$ can be written as

$$\begin{aligned}
\boldsymbol{W}_1 &= -\sum_{i=1}^{n}\boldsymbol{\Pi}_i'\boldsymbol{\Delta}_i\Sigma_i^{-1}\boldsymbol{\Delta}_i(\mathbf{y}_i - \boldsymbol{\mu}_i) + \tau D\boldsymbol{\theta} = 0,\\
\boldsymbol{W}_2 &= \frac{1}{2}\sum_{i=1}^{n}\mathbf{Z}_i'(\mathbf{q}_i - \mathbf{1}_{m_i}) = 0,\\
\text{and}\quad \boldsymbol{W}_3 &= \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\frac{\partial \log T_{ijj}}{\partial\boldsymbol{\gamma}}(\epsilon_{ij}^2 - 1) + \epsilon_{ij}\sum_{k=1}^{j-1}b_{ijk}\epsilon_{ik}\right] = 0.
\end{aligned}$$

All these equations can then be solved using the same algorithm as what we use in Section 2.1. All of the block components of the Fisher information matrix remain the same as Model 2 except I_{11} which in this case is

$$I_{11} = -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\right] = \sum_{i=1}^{n}\boldsymbol{\Pi}_i'\boldsymbol{\Delta}_i\Sigma_i^{-1}\Delta_i\boldsymbol{\Pi}_i + \tau D.$$

2.4. Knot selection

The importance of knot selection in spline smoothing work has been well described in two pioneering papers by He, Fung and Zhu (2005) and Leng *et al.* (2010). These authors found that knot selection is less critical for the estimation of $\boldsymbol{\beta}$ than for the estimation of the nonparametric functions involved in Model 1 and Model 2 discussed in Section 1. However, in most situations, they found it appropriate to use the sample quantiles of $\{t_{ij}, i = 1, ..., n,\ j = 1, ..., m_i\}$ as knots. We follow their suggestion and note that the number of knots are not prespecified, rather, it depends on the total sample size $N = \sum_{i=1}^{n} m_i$. Through detailed asymptotic theoretical study Leng *et al.* (2010) show that the number of internal knots to be used is the integer part of $N^{1/5}$.

3. SIMULATION STUDY

As indicated in Section 1, an extensive simulation investigation is conducted in this section. Our purpose in this simulation, in addition to the study of the performance of the estimators of the regression parameters in terms of bias and efficiency, to study the effect of fixing the number of knots, the effect of correlated covariates, the effect of misspecifying the error distribution (robustness study), and compare performance of the estimation methods using B-spline and penalized spline. These simulations are performed in Sections 3.1, 3.2, 3.3, and 3.4. A further study, in Section 3.5, is conducted to compare the estimators by fitting the data generated from the semiparametric model to the parametric model and vice versa.

3.1. Study 1: Properties of the regression parameters

For this purpose we generate response data from each of the three models (Model 1, Model 2, and Model 3):

$$\begin{aligned} y_{ij} &= x_{ij1}\beta_1 + x_{ij2}\beta_2 + e_{ij}, \quad \log(\sigma_{ij}^2) = z_{ij1}\lambda_1 + z_{ij2}\lambda_2; \\ y_{ij} &= x_{ij1}\beta_1 + x_{ij2}\beta_2 + f_1(t_{ij}) + e_{ij}, \quad \log(\sigma_{ij}^2) = z_{ij1}\lambda_1 + z_{ij2}\lambda_2; \\ y_{ij} &= x_{ij1}\beta_1 + x_{ij2}\beta_2 + f_1(t_{ij}) + e_{ij}, \quad \log(\sigma_{ij}^2) = z_{ij1}\lambda_1 + z_{ij2}\lambda_2 + f_2(t_{ij}). \end{aligned}$$

For each model, values of parameters considered were $(\beta_1, \beta_2) = (1, 0.5)$, $(\gamma_1, \gamma_2) = (0.35, 0.5)$ and $(\lambda_1, \lambda_2) = (-0.5, 0.2)$. Following Leng *et al.* (2010) we generate the observation times as in what follows.

For each individual we consider a set of scheduled time points $\{0, 1, 2, ..., 12\}$. At each scheduled time, except time 0, each individual has a 20% probability of missing a fixed time point. To make it irregular and unequal time distances for different individuals a uniform $[0, 1]$ random variable is added to a non skipped scheduled time. This results in different observed time points t_{ij} per subject. However, while analysis t_{ij} is transformed onto $[0, 1]$.

For covariates, we take $x_{ij1} = t_{ij} + \delta_{ij}$, where δ_{ij} follows the standard normal distribution and x_{ij2} is generated from a Bernoulli(0.5) distribution. The nonparametric functions are taken as $f_1(t) = \cos(\pi t)$ and $f_2(t) = \sin(\pi t)$. The error $(e_{i1}, ..., e_{im_i})$ is generated from a multivariate normal distribution with mean 0 and covariance $\Sigma_i = D_i R_i D_i$, where $R_i = T_i T_i'$ with $\mathbf{w}_{ijk} = (1, t_{ij} - t_{ik})'$, and $\mathbf{z}_{ij} = \mathbf{x}_{ij}$. The expected sample size (for the calculation of the number of knots) is about 1040 ($=100 \times 13 \times 0.8$). The number of the knots is taken to be $4 \approx 1040^{1/5}$ (He *et al.*, 2005). Here, as can be seen, the number of knots is not prespecified. Parameters of the above models were estimated using the `R` package `jmcm` for the simulation studies.

Bias of the estimates of the parameters of all three models along with their standard errors, and MSE (S) of the non-parametric functions f_1 and f_2, based on 1000 replications, are given in Table 1.

Table 1: Bias and standard error of the estimated parameters based on 1000 replications; all the results are multiplied by a factor of 10^3.

Par.	Tr. val.	Zhang *et al.* (2015) (Model 1)		Semi. with mean (Model 2)		Semi. with mean and var (Model 3)		Semi. with mean and var when knot=10		Semi. with mean and var when knot=20	
		Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
β_1	1.0	0.00	0.02	−0.02	0.22	0.09	0.27	0.00	0.24	−0.01	0.25
β_2	0.5	0.00	0.05	0.01	0.11	−0.03	0.14	0.01	0.12	0.01	0.13
γ_1	0.35	−0.27	1.30	6.61	1.42	3.14	2.25	−21.51	2.42	−41.93	3.41
γ_2	0.5	−0.35	2.29	6.62	2.53	5.56	4.12	−3.62	4.08	−5.74	5.04
λ_1	−0.5	0.00	0.07	0.00	0.00	0.00	0.08	0.00	0.07	0.00	0.07
λ_2	0.2	0.00	0.13	−0.01	0.01	0.01	0.15	0.01	0.14	0.00	0.14
$S(\hat{f}_1)$					11.89		10.90		496.59		516.37
$S(\hat{f}_2)$							112.87		373.95		642.26

Table 1 shows that our semiparametric methods yield similar bias property of the estimates as compared to that for the parametric model. We redo the simulations by fixing the number of knots as $k_n = 10$ and $k_n = 20$. The results show that as the number of knots increase, the MSE of the estimated functions f_1 and f_2, bias, and standard error of the estimates of the parameters also increase.

3.2. Study 2: Properties of the regression parameters when some covariates are correlated

We have done a simulation study where all the covariates are correlated. The data sets are generated from the model

$$y_{ij} = x_{ij0}\beta_0 + x_{ij1}\beta_1 + x_{ij2}\beta_2 + f_1(t_{ij}) + e_{ij} \quad (i = 1, ..., n;\ j = 1, ..., m_i),$$

and

$$\phi_{ijk} = \gamma_0 + w_{ijk1}\gamma_1 + w_{ijk2}\gamma_2,$$

$$\log(\sigma_{ij}^2) = z_{ij0}\lambda_0 + z_{ij1}\lambda_1 + z_{ij2}\lambda_2 + f_2(t_{ij}).$$

where $(x_{ij0}, x_{ij1}, x_{ij2})'$ is generated from a multivariate normal distribution with mean 0, marginal variance 1, and correlation 0.5. We take $\boldsymbol{z}_{ij} = \boldsymbol{x}_{ij}$ and $\mathbf{w}_{ijk} = (1, t_{ij} - t_{ik}, (t_{ij} - t_{ik})^2)'$. t_{ij}, e_{ij}, $f_1(t)$, and $f_2(t)$ are same as study 3.1.

Table 2 shows that both parametric and semiparametric methods yield similar bias and standard error property of the estimates.

Table 2: Bias and standard error of the estimated parameters based on 1000 replications when the covariates are correlated; all the results are multiplied by a factor of 10^3.

Parameter	True value	Parametric		Semiparametric	
		Bias	SE	Bias	SE
β_0	1.0	0.00	0.02	0.05	0.12
β_1	−0.5	0.00	0.02	−0.05	0.11
β_2	0.5	0.00	0.01	0.03	0.02
γ_0	0.35	−0.04	1.43	3.42	2.34
γ_1	0.5	−0.08	3.41	5.71	4.81
γ_2	−0.3	0.05	1.93	−3.74	2.82
λ_0	−0.5	0.00	0.01	0.00	0.05
λ_1	0.5	0.00	0.02	0.00	0.04
λ_2	−0.3	0.00	0.01	0.00	0.05
MSE($\hat{f}_1$)					9.62
MSE($\hat{f}_2$)					98.84

3.3. Study 3: A robustness study. Properties of the regression parameters and the functions f_1 and f_2 when error follows mixture of normal distributions

Again, another study, similar to what was done in study 3.1, has been conducted, in which, for generating the response data of Model 2 and Model 3, we consider a mixture of two multivariate normal distributions with error distributions $N_{m_i}(0, \Sigma_i)$ and $N_{m_i}(0, 0.25^2\Sigma_i)$ with equal probability. The results for the simulation study are displayed in Table 3 which show that the bias and the standard errors of the estimates, and the MSE of f_1 remain almost the same as those in study 3.1. However, MSE of the fitted function f_2 in Model 3 increases significantly. Note that the function f_2 is associated with the variance and the function f_1 is associated with the mean.

Table 3: Simulation results for Study 3.3 in Model 2 and Model 3 over 1000 replications when error terms follow mixture of normal distribution; $n = 100$; all the results are multiplied by a factor of 10^3.

Par.	Tr. val.	Semi. (Model 2) (without mixed)		Semi. (Model 2) (with mixed)		Semi. (Model 3) (without mixed)		Semi. (Model 3) (with mixed)	
		Bias	SE	Bias	SE	Bias	SE	Bias	SE
β_1	1.0	−0.02	0.22	−0.02	0.21	0.09	0.27	0.11	0.22
β_2	0.5	0.01	0.11	0.01	0.13	−0.03	0.14	−0.04	0.11
γ_1	0.35	6.61	1.42	6.61	6.32	3.14	2.25	8.84	2.36
γ_2	0.5	6.62	2.53	−8.57	8.17	5.56	4.12	13.48	4.39
λ_1	−0.5	0.00	0.00	0.00	0.1	0.00	0.08	0.00	0.09
λ_2	0.2	−0.01	0.01	0.00	0.1	0.01	0.15	−0.02	0.18
MSE($\hat{f}_1$)			11.89		11.89		10.90		11.00
MSE($\hat{f}_2$)							112.87		693.49

Note that the mixed modeling affects only the MSE of the estimates of f_2 in the semiparametric model 3, but not f_1. Although we are not aware of a proof as to why mixed modelling affects only the MSE of the estimates of f_2, but not f_1, experience in this context and other similar contexts show that parameters or functions associated with the variance parameter are more difficult to estimate and show more bias and MSE than parameters or functions associated with the mean parameter. See, for example, Zhang *et al.* (2015, pp. 234), and Saha and Paul (2005, pp. 183). In estimating the mean parameter m and the dispersion parameter c, Saha and Paul (2005) find that estimates of the mean parameter m or regression parameters show much smaller bias and standard error than that of the dispersion parameter c.

3.4. Study 4: Comparison of using B-spline and penalized spline in Model 2

A further study is conducted to compare the B-spline and the penalized spline to estimate the nonparametric function f_1 in Model 2. The estimation procedure to estimate all parameters are discussed in Section 2.3. We have used generalized cross-validation approach

to find the penalty parameter for penalized splines. To generate response variable we consider the same mean and variance models as those in study 3.1 and the results are presented in Table 4. Results in Table 4 show no advantage of using the penalized spline over the B-spline.

Table 4: Bias and standard error of the estimated parameters in Model 2 based on 1000 replications using B-spline and penalized spline; all the results are multiplied by a factor of 10^3.

Parameter	True value	Semiparametric Model 2			
		B spline		Penalized spline	
		Bias	SE	Bias	SE
β_1	1.0	−0.02	0.22	−0.01	0.22
β_2	0.5	0.01	0.11	0.01	0.11
γ_1	0.35	6.61	1.42	6.60	1.38
γ_2	0.5	6.62	2.53	6.50	2.51
λ_1	−0.5	0.00	0.00	0.00	0.08
λ_2	0.2	−0.01	0.01	−0.02	0.15
MSE($\hat{f}_1$)			11.89		11.89

3.5. Study 5: Comparison of the estimators by fitting the data generated from the semiparametric model to the parametric model and vice versa

We have done a simulation study by fitting the data generated from the semiparametric model to the parametric model and vice versa. We generate the response variable from Model 1 similar to study 3.2 and estimate the regression parameters by fitting Model 3 and vice versa. The results for the simulation study are displayed in Table 5 which show that bias and standard errors of the estimates are reasonable by fitting the data generated from the parametric model and estimating the regression parameters by the semiparametric model.

Table 5: Bias and standard error of the estimated parameters based on 1000 replications by fitting the data generated from the semiparametric model to the parametric model and vice versa; all the results are multiplied by a factor of 10^3.

Parameter	True value	Parametric to Semiparametric		Semiparametric to Parametric	
		Bias	SE	Bias	SE
β_0	1.0	0.00	0.06	−46.48	7.63
β_1	−0.5	0.00	0.06	46.87	7.64
β_2	0.5	0.00	0.04	−27.46	7.84
γ_0	0.35	−1.74	2.53	473.47	15.79
γ_1	0.5	−2.83	5.47	−40.41	13.92
γ_2	−0.3	1.64	3.13	702.96	133.61
λ_0	−0.5	0.00	0.05	95.04	16.74
λ_1	0.5	0.00	0.05	−92.93	16.71
λ_2	−0.3	0.00	0.05	57.25	16.72

In contrast, bias and standard errors increase significantly if we generate response variable using the semiparametric model and estimate the regression parameters by the parametric model.

A note on the simulation results is that the choice of the values of the parameters, such as, β_1 and β_2 in Table 1, does not affect the results on the properties of the estimates. For example, the simulations for the results in Table 1 for Model 3 was re-run with $\beta_1 = 0.5$ and $\beta_2 = 2.5$ instead of $\beta_1 = 1$ and $\beta_2 = 0.5$ as in Table 1. The bias and SE of the estimates of all the parameters were virtually unchanged.

4. ANALYSIS OF TWO REAL DATA SETS ARISING IN BIOMEDICAL/ ENVIRONMENTAL STUDIES

The purpose of this section is to study the impact of semiparametric modelling on model parsimony through the analysis of two real life longitudinal data sets arising in biomedical/environmental studies. The first data set is regarding progesterone metabolite (pregnanediol-3-glucuronide, PdG) measures that was obtained by the Institute for Toxicology and Environmental Health at the University of California, Davis in collaboration with the Reproductive Epidemiology Section of the California Department of Health Services, Berkeley (Brumback and Rice, 1998). The second data set is regarding CD4 cell counts of 369 HIV-infected men obtained by Kaslow *et al.* (1987). The first data set involves polynomials in time and the second data set involves polynomials in time as well as real covariates. Full description and analysis of the these two data sets are given in Sections 4.1 and 4.2.

In order to select the most parsimonious model we need to identify the best integer triple representing, namely, the degrees of the three polynomial functions for the mean structure, the correlation structure and the variance structure. Similar to Pourahmadi (2000) and Pan and Mackenzie (2003), we use the Bayesian information criterion, BIC, to identify the best triple as follows:

$$\mathrm{BIC}(p, q, d) = -2\hat{l}_{\max}/n + (p + q + d + 3)\log(n)/n,$$

where $\hat{l}_{\max}$ is the maximum of the log-likelihood, p, q and d lie in the range of 0 to $(m-1)$ with $m = \max_{1 \le i \le n}\{m_i\}$.

Thus, (p', q', d') is the optimal triplet which minimizes $\mathrm{BIC}(p, q, d)$. Shao (1997) and Shai and Tsai (2002) demonstrated that BIC-criterion can identify the true model consistently. Although, in the literature BIC is preferred over the AIC (Akaike Information Criterion)

$$\mathrm{AIC}(p, q, d) = -2\hat{l}_{\max}/n + 2(p + q + d + 3)/n,$$

we include the later here for comparison in our model selection.

4.1. Progesterone metabolite data

The Brumback and Rice (1998) data consist of repeated progesterone metabolite (pregnanediol-3-glucuronide, PdG) measures from day −8 to day 15 in the menstrual cycle (day 0 denotes ovulation day) on a sample of 22 conceptive cycles from 22 women and 29 non-conceptive cycles from another 29 women to study of early pregnancy loss. Altogether 1130 observations were obtained from 51 women, with each woman contributing 9 to 24 observations over time. The data are reproduced in Table 1 of the Supplementary Material.

As in Brumback and Rice (1998), we take a log transformation of these data to make the normality assumption reasonable. Using our semiparametric modelling, the full model is

$$\begin{aligned} g(\mu_{ij}) &= \beta_0 + x_{ij}\beta_1 + \cdots + x_{ij}^p\beta_p + f_1(t_{ij}), \\ \text{and} \quad \phi_{ijk} &= \gamma_0 + w_{ijk}\gamma_1 + \cdots + w_{ijk}^q\gamma_q, \\ \log(\sigma_{ij}^2) &= \lambda_0 + z_{ij}\lambda_1 + \cdots + z_{ij}^d\lambda_d + f_2(t_{ij}), \end{aligned}$$

where $x_{ij} = z_{ij} = t_{ij}$, $w_{ijk} = t_{ij} - t_{ik}$ and p, q, and d lie in the range of 1 to 23. The parametric model, of course, will not have the non-parametric functions.

We now analyze the data using these parametric and semiparametric models. The results are given in Table 6. Note that here we give 9 most parsimonious models to capture the model having smallest BIC and/or smallest AIC. Results in Table 6 show that the most parsimonious model having the smallest BIC (BIC = 0.47) and the smallest AIC (AIC = −0.29) obtained by parametric modelling has $(p, q, d) = (7, 3, 7)$. Where as the corresponding most parsimonious model by semiparametric modelling has $(p, q, d) = (2, 2, 1)$ with BIC = −0.61 and AIC = −0.92. Thus the final models for these data are

$$\begin{aligned} g(\mu_{ij}) &= f_1(t_{ij}) + \beta_0 + x_{ij}\beta_1 + x_{ij}^2\beta_2, \\ \text{and} \quad \phi_{ijk} &= \gamma_0 + w_{ijk}\gamma_1 + w_{ijk}^2\gamma_2, \\ \log(\sigma_{ij}^2) &= f_2(t_{ij}) + \lambda_0 + z_{ij}\lambda_1, \end{aligned}$$

where $x_{ij} = z_{ij} = t_{ij}$ and $w_{ijk} = t_{ij} - t_{ik}$. Note that both the BIC and the AIC criteria choose the same model for these data.

Table 6: Progesterone hormone data: A comparison of various models using parametric (Zhang *et al.*, 2015) and semiparametric approaches.

(p, q, d)	No. of par.	Parametric			Semiparametric		
		$\hat{l}_{\max}$	BIC	AIC	$\hat{l}_{\max}$	BIC	AIC
(2,2,1)	8	−71.66	3.43	3.12	31.34	−0.61	−0.92
(2,2,2)	9	−66.59	3.31	2.96	31.34	−0.54	−0.88
(2,1,1)	7	−129.58	5.62	5.36	16.01	−0.09	−0.35
(2,1,4)	10	−66.98	3.40	4.53	16.38	0.12	−0.25
(3,1,4)	11	−28.97	1.98	3.09	16.38	0.21	−0.21
(3,1,5)	12	−28.20	2.03	3.00	19.58	0.16	−0.30
(2,1,6)	12	−65.84	3.51	4.42	23.73	−0.005	−0.46
(3,2,3)	11	−32.85	2.14	1.81	31.34	−0.38	−0.80
(7,3,7)	20	27.31	0.47	−0.29	42.06	−0.11	−0.87

4.2. Analysis of CD4 cell data

The data comprise CD4 cell counts of 369 HIV-infected men with six covariates including time since seroconversion (t_{ij}), age (relative to arbitrary origin, x_{ij1}), packs of cigarettes smoked per day (x_{ij2}), recreation drug use (x_{ij3}), number of sexual partners (x_{ij4}), and mental illness score (x_{ij5}). In total there are 2376 observations with multiple repeated measurements taken for each individual at different times, covering a period of approximately eight and a half years. The number of measurements for each individual varies from 1 to 12 taken at unequally spaced time points.

Now, for these data m_i varies from 1 to 12, so p, q, and d all lie in the range of 0 to 11. The values of p, q, and d for the most saturated model are 11. So, we have a total $11 \times 11 \times 11 = 1331$ models of which we present 6 most parsimonious models.

4.2.1. Analysis using polynomials in time

This data set has been analyzed by others in the past (for example, Zeger and Diggle, 1994, and Ye and Pan, 2006) using polynomials in time. Most recently Zhang *et al.* (2015), in order to jointly model the mean, correlation and variance structures, fit polynomial regressions of time

$$\begin{aligned} g(\mu_{ij}) &= \beta_0 + x_{ij}\beta_1 + \cdots + x_{ij}^p \beta_p, \\ \phi_{ijk} &= \gamma_0 + w_{ijk}\gamma_1 + \cdots + w_{ijk}^q \gamma_q, \\ \log(\sigma_{ij}^2) &= \lambda_0 + z_{ij}\lambda_1 + \cdots + z_{ij}^d \lambda_d, \end{aligned}$$

and

where $x_{ij} = z_{ij} = t_{ij}$ and $\mathbf{w}_{ijk} = t_{ij} - t_{ik}$.

We now analyze these data using the above parametric models and our method of semiparametric modelling Model 3 developed in Section 2.1. The results, in terms of $\hat{l}_{\max}$, BIC and AIC of the parametric and semiparametric models are given in Table 7.

Table 7: CD4 cell data: A comparison of various models using parametric (Zhang *et al.*, 2015) and semiparametric approaches with polynomials in time.

(p,q,d)	No. of par.	Parametric			Semiparametric		
		$\hat{l}_{\max}$	BIC	AIC	$\hat{l}_{\max}$	BIC	AIC
(4,1,1)	9	−4910.87	26.76	26.67	−4877.41	26.58	26.48
(3,1,1)	8	−4926.88	26.83	26.74	−4882.11	26.59	26.50
(8,1,1)	13	−4892.72	26.72	26.58	−4874.40	26.63	26.49
(8,3,1)	15	−4890.44	26.75	26.59	−4881.66	26.69	26.64
(3,3,3)	12	−4919.52	26.85	26.73	−4879.37	26.64	26.51
(8,3,3)	17	−4886.36	26.76	26.58	−4872.74	26.68	26.50

Note that here we provide all relevant information of 6 most parsimonious models. Of these, the most parsimonious model obtained from parametric modelling has $(p,q,d) = (8,1,1)$ with $\hat{l}_{\max} = -4892.72$, BIC $= 26.72$, and AIC $= 26.58$ (smallest BIC as well as smallest

AIC among all these 6 models). This finding is in agreement with what was found by Zhang *et al.* (2015). The semiparametric modelling, however, shows $(p, q, d) = (4, 1, 1)$ with $\hat{l}_{\max} = -4877.41$, BIC $= 26.58$, and $AIC = 26.48$ (smallest BIC as well as smallest AIC among all these 6 models). Including the intercept parameters the semiparametric model has 9 parameters as opposed to 13 parameters of the model obtained by the parametric model. Thus, the semiparametric modelling shows a significant improvement in model parsimony as compared to the parametric modelling and the most parsimonious model for these data is

$$\begin{aligned} g(\mu_{ij}) &= f(t_{ij}) + \beta_0 + x_{ij}\beta_1 + x_{ij}^2\beta_2 + x_{ij}^3\beta_3 + x_{ij}^4\beta_4, \\ \phi_{ijk} &= \gamma_0 + w_{ijk}\gamma_1, \\ \text{and} \qquad \log(\sigma_{ij}^2) &= g(t_{ij}) + \lambda_0 + z_{ij}\lambda_1, \end{aligned}$$

where $x_{ij} = z_{ij} = t_{ij}$ and $\mathbf{w}_{ijk} = t_{ij} - t_{ik}$.

Note that the same parsimonious model is obtained by both the BIC and the AIC procedures. To investigate the stability of AIC and BIC, a simulation study is conducted by choosing p, q, and d from $1, 2, ..., 11$ and obtain that the stability of AIC and BIC are very similar which is 62%. From the same simulation study, we further investigate the parametric and semiparametric models performances in terms of AIC and BIC and found that parametric models perform better than semiparametric models.

4.2.2. Analysis using polynomials in time and the covariates

We now analyze the CD4 data using the Model 3 regression models after including 5 covariates in the mean model. The results, in terms of $\hat{l}_{\max}$, BIC, and AIC of the parametric and the semiparametric models, are given in Table 8. Here also we provide all relevant information of 6 best (parsimonious) models. Of these, the most parsimonious model by the parametric modelling has $(c, p, q, d) = (3, 8, 1, 1)$ with 16 parameters by both the BIC and the AIC procedures, where c represents the number of covariates in the model. However, the most parsimonious model by the semiparametric modelling has $(c, p, q, d) = (3, 2, 1, 1)$ with 10 parameters using the BIC procedure (BIC $= 26.47$) and has $(c, p, q, d) = (3, 4, 1, 1)$ with 12 parameters using the AIC procedure (AIC $= 26.38$).

Thus, the most parsimonious models for the CD4 data involving time and the covariates is

$$\begin{aligned} g(\mu_{ij}) &= \beta_0 + x_{ij}\beta_1 + x_{ij}^2\beta_2 + x_{ij2}\beta_3 + x_{ij3}\beta_4 + x_{ij5}\beta_5 + f(t_{ij}), \\ \phi_{ijk} &= \gamma_0 + w_{ij}\gamma_1, \\ \text{and} \qquad \log(\sigma_{ij}^2) &= \lambda_0 + z_{ij}\lambda_1 + g(t_{ij}), \end{aligned}$$

where $x_{ij} = z_{ij} = t_{ij}$ and $\mathbf{w}_{ijk} = t_{ij} - t_{ik}$.

Our finding is that both the BIC and the AIC select same or similar models, although, BIC has some advantage over the AIC in the sense that, in some cases, the model selected by the BIC is slightly more parsimonious than the model selected by the AIC. Further, the semiparametric modelling obtains a model that is much more parsimonious than the model obtained by the parametric modelling approach.

Table 8: CD4 cell data: A comparison of various models using parametric (Zhang *et al.*, 2015) and semiparametric approaches with polynomials in time and covariates.

(c,p,q,d)	No. of par.	Parametric			Semiparametric		
		$\hat{l}_{\max}$	BIC	AIC	$\hat{l}_{\max}$	BIC	AIC
(3,2,1,1)	10	−4944.54	26.91	26.84	−4863.43	26.47	26.40
(5,2,1,1)	12	−4938.40	26.88	26.80	−4862.52	26.47	26.39
(3,4,1,1)	12	−4891.12	26.65	26.56	−4858.68	26.48	26.38
(5,4,1,1)	14	−4889.44	26.65	26.55	−4857.85	26.47	26.38
(5,3,1,1)	13	−4905.57	26.72	26.63	−4862.52	26.48	26.40
(3,8,1,1)	16	−4873.51	26.62	26.49	−4857.16	26.53	26.40

5. DISCUSSION

We develop joint estimation procedure for the mean (regression) and the variance parameters in longitudinal data using semiparametric modelling of the mean and the variance, regression spline, and by decomposing the correlation matrix via hyperspherical co-ordinates. Through an extensive simulation study we compare our method with the parametric method by Zhang *et al.* (2015). Further, the effect of the misspecification of the error distribution and of the number of knots used in the estimation of the nonparametric functions, and whether the penalized spline procedure improves the estimation of the nonparametric functions over the B-spline are investigated. Furthermore two real data sets arising from biomedical/environmental study are analyzed.

The main findings of the simulation study are:

(**a**) the parametric modelling and the semiparametric modelling produce similar bias and efficiency property of the regression parameters;

(**b**) good choice of number of knots can have significant impact of estimation and inference on non-linear trends (e.g., the nonparametric functions) using regression B-splines;

(**c**) use of the penalized spline does not improve the efficiency of the estimates of the nonparametric functions.

Through data analysis:

(**i**) our findings regarding the model selection procedures are that both the BIC and the AIC select same or similar models, although, BIC has some advantage over the AIC in the sense that, in some cases, the model selected by the BIC is slightly more parsimonious than the model selected by the AIC;

(**ii**) the main advantage of the semiparametric modelling over the parametric modelling is that the former produces a much more parsimonious model than the latter.

We also analyzed some other real bio-medical data sets and obtained similar conclusions, the details of which are not given here.

A. APPENDIX: Solution of the Estimating Equations of Model 3

We apply the quasi-Fisher scoring algorithm to solve estimating equations of Model 3 where the parameters $\boldsymbol{\theta}, \boldsymbol{\rho}$ and $\boldsymbol{\gamma}$ solve sequentially one by one with other parameter keep fixed in optimization:

Step 1: Choose initial values of the parameters as $\boldsymbol{\theta}^{(0)}, \boldsymbol{\rho}^{(0)}$ and $\boldsymbol{\gamma}^{(0)}$. Set $k = 0$.

Step 2: Calculate $\boldsymbol{\Sigma}_i$ by using $\boldsymbol{\rho}^{(k)}$ and $\boldsymbol{\gamma}^{(k)}$. Update $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + I_{11}^{-1}\boldsymbol{U}_1|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}}.$$

Step 3: Given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k+1)}$, update $\boldsymbol{\gamma}$ and $\boldsymbol{\rho}$ by using

$$\begin{pmatrix} \boldsymbol{\gamma}^{(k+1)} \\ \boldsymbol{\rho}^{(k+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\gamma}^{(k)} \\ \boldsymbol{\rho}^{(k} \end{pmatrix} + \left[\begin{pmatrix} I_{22} & I_{23} \\ I_{32} & I_{33} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{U}_3 \\ \boldsymbol{U}_2 \end{pmatrix} \right] |_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{(k)}, \boldsymbol{\rho}=\boldsymbol{\rho}^{(k)}}.$$

Step 4: Set $k \leftarrow k+1$ and repeat steps 2 and 3 until a desired convergence criteria is satisfied.

Note that block components of Fisher information matrix I are:

$$\begin{aligned}
I_{11} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] = \sum_{i=1}^{n} \boldsymbol{\Pi}_i' \boldsymbol{\Delta}_i \Sigma_i^{-1} \Delta_i \boldsymbol{\Pi}_i, \\
I_{12} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}'}\right] = -\sum_{i=1}^{n}\left[\boldsymbol{\Pi}_i' \Delta_i \frac{\partial \Sigma_i^{-1}}{\partial\boldsymbol{\gamma}'}(E(\mathbf{y}_i) - \boldsymbol{\mu}_i)\right] = 0, \\
I_{13} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\rho}'}\right] = 0, \\
I_{22} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right] = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[2\frac{\partial \log T_{ijj}}{\partial\boldsymbol{\gamma}}\frac{\partial \log T_{ijj}}{\partial\boldsymbol{\gamma}'} + \sum_{k=1}^{j-1} b_{ijk} b_{ijk}'\right], \\
I_{23} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\rho}'}\right] = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\frac{\partial \log T_{ijj}}{\partial\boldsymbol{\gamma}}\Upsilon_{ij}' + \frac{1}{2}\sum_{k=1}^{j-1} b_{ijk}\sum_{l=k}^{j} a_{ijl} T_{ilk}\Upsilon_{il}'\right], \\
I_{33} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\rho}\partial\boldsymbol{\rho}'}\right] = \frac{1}{4}\sum_{i=1}^{n} \boldsymbol{\Upsilon}_i'\left[I_{m_i} + R_i^{-1} \circ R_i\right]\boldsymbol{\Upsilon}_i,
\end{aligned}$$

where '$\circ$' represents the Hadamard product.

B. APPENDIX: Block Components of Fisher Information Matrix of the Estimating Equations of Model 2

$$
\begin{aligned}
I_{11} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] = \sum_{i=1}^{n} \boldsymbol{\Pi}_i' \boldsymbol{\Delta}_i \Sigma_i^{-1} \Delta_i \boldsymbol{\Pi}_i, \\
I_{12} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}'}\right] = -\sum_{i=1}^{n}\left[\boldsymbol{\Pi}_i' \Delta_i \frac{\partial \Sigma_i^{-1}}{\partial\boldsymbol{\gamma}'}(E(\mathbf{y}_i) - \boldsymbol{\mu}_i)\right] = 0, \\
I_{13} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\lambda}'}\right] = 0, \\
I_{22} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right] = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[2\frac{\partial \log T_{ijj}}{\partial\boldsymbol{\gamma}}\frac{\partial \log T_{ijj}}{\partial\boldsymbol{\gamma}'} + \sum_{k=1}^{j-1} b_{ijk} b_{ijk}'\right], \\
I_{23} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\lambda}'}\right] = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[\frac{\partial \log T_{ijj}}{\partial\boldsymbol{\gamma}} Z_{ij}' + \frac{1}{2}\sum_{k=1}^{j-1} b_{ijk}\sum_{l=k}^{j} a_{ijl} T_{ilk} Z_{il}'\right], \\
I_{33} &= -E\left[\frac{\partial^2 l}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}\right] = \frac{1}{4}\sum_{i=1}^{n} \mathbf{Z}_i'\left[I_{m_i} + R_i^{-1} \circ R_i\right]\mathbf{Z}_i,
\end{aligned}
$$

where '$\circ$' represents the Hadamard product.

ACKNOWLEDGMENTS

This research was partially supported by the Natural Sciences and Engineering Research Council of Canada. We also acknowledge the valuable suggestions from the referees.

REFERENCES

Zhang, W., Leng, C., and Tang, Y. C. (2015). A joint modelling approach for longitudinal studies. *Journal of the Royal Statistical Society* B, 77:219–238.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22.

Diggle, P. J., Heagerty, P., Liang, K. Y., and Zeger, S. L. (2002). *Analysis of Longitudinal Data*, 2nd edition. Oxford University Press.

He, X. M., Fung, W. K., and Zhu, Z. Y. (2005). Robust estimation in generalized partial linear models for clustered data. *Journal of the American Statistical Association*, 100:1176–1184.

Lin, X. and Carroll, R. J. (2006). Semiparametric estimation in general repeated measures problems. *Journal of the Royal Statistical Society* B, 68:69–88.

Leng, C., Zhang, W., and Pan, J. (2010). Semiparametric mean-covariance regression analysis for longitudinal data. *Journal of the American Statistical Association*, 105:181–193.

Schumaker, L. L. (1981). *Spline Functions*, 182, 190. Wiley, New York.

Carroll, R. H., Macam, J. D., and Ruppert, D. (1999). Nonparametric regression in the presence of measurement error. *Biometrika*, 86:541–554.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statist. Sci.*, 11:89–121.

Saha, K. K. and Paul, S. R. (2005). Bias-corrected maximum likelihood estimator of the intraclass correlation parameter for binary data. *Statistics in Medicine*, 24:3497–3512.

Brumback, B. A. and Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). *Journal of the American Statistical Association*, 93:961–994.

Kaslow, R. A., Ostrow, D. G., Detels, R. *et al.* (1987). The multicenter AIDS cohort study: Rationale, organization and selected characteristics of the participants. *American Journal of Epidemiology*, 126:310–318.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. *Biometrika*, 87:425–435.

Pan, J. and Mackenzie, G. (2003). Model selection for joint mean-covariance structures in longitudinal studies. *Biometrika*, 90:239–244.

Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica Sincia*, 221–264.

Shai, P. and Tsai, C. L. (2002). Regression model selection-a residual likelihood approach. *Journal of the Royal Statistical Society* B, 64:237–252.

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, 50:689–699.

Ye, H. and Pan, J. (2006). Modelling covariance structures in generalized estimating equations for longitudinal data. *Biometrika*, 93:927–941.

REVSTAT – Statistical Journal
Volume 23, Number 3, July 2025, 359–369
https://doi.org/10.57805/revstat.v23i3.625

A Closed-Form Expression for the Variance of Truncated Distribution and Its Uses

Authors: Roberto Vila ✉
– Department of Statistics, University of Brasília, Brazil
– Department of Mathematics and Statistics, McMaster University, Canada
rovig161@gmail.com

Narayanaswamy Balakrishnan
– Department of Mathematics and Statistics, McMaster University, Canada
bala@mcmaster.ca

Raul Matsushita
– Department of Statistics, University of Brasília, Brazil
raulmta@unb.br

Received: May 2023 Revised: September 2023 Accepted: September 2023

Abstract:

- This work sheds some light on the relationship between a distribution's standard deviation and its range, a topic that has been discussed extensively in the literature (see Shiffler and Harsha, 1980). While many previous studies have proposed inequalities or relationships that depend on the shape of the population distribution, the approach here is built on a family of bounded probability distributions based on skewing functions. We offer closed-form expressions for its moments and the asymptotic behavior as the support's semi-range tends to zero and ∞. We also establish an inequality in which the well-known Popoviciu's inequality is a special case. Finally, we provide an example using US dollar prices in four different currencies traded on foreign exchange markets to illustrate the results developed here.

Keywords:

- *truncated distribution; skewing function; Popoviciu's inequality.*

AMS Subject Classification:

- 60E05, 62Exx, 62Fxx.

✉ Corresponding author

1. INTRODUCTION

Relating the standard deviation (σ) to the range is a well-studied topic (see Tippett, 1925; Popoviciu, 1935; Shone, 1949; David *et al.*, 1954). However, most of the results found in the literature in this regard propose inequalities or relationships that depend on the shape of the population distribution. Matsushita *et al.* (2020, 2023) recently suggested a power law between σ and the semi-range ℓ without knowing the population distribution, but assuming symmetric truncated forms restricted to $\ell \ll 1$. They argued that truncation is a phenomenon naturally generated by the sampling process. In their approach, conditional distribution properties can link the usual unbounded distribution for describing unobserved data.

Here, we obtain a general closed-form expression for the variance of truncated distributions valid for all $\ell > 0$. We start with the general case (Section 2.1) by considering a truncated variable X over the interval $[a, b]$ based on a cumulative distribution function G with unbounded support as a skewing function. We present a general form for the centralized moment of order $p > 0$, $\mathbb{E}[(X - c)^p]$, where $c \in (a, b)$ denotes a centering parameter. We also derive its asymptotic behavior as the support's semi-range tends to zero and ∞, as well as some interesting inequalities.

Section 2.2 presents some properties and examples regarding the symmetrically truncated distribution as a particular case. Importantly, we deduce the form of the ratio σ/ℓ as a function of G and ℓ. We illustrate the established result with an example using actual financial data (Section 3). They consist of 16 million tick-by-tick returns of four currencies against the US dollar transacted on foreign exchange markets. Finally, Section 4 makes some brief concluding remarks.

2. MAIN RESULTS

2.1. General case

Suppose we have a random variable with cumulative distribution function (CDF) $G(x)$ and with infinite support. Based on G, given two real numbers a and b, such that $a < b$, we have

$$F_X(x) = \frac{G(x) - G(a)}{G(b) - G(a)}, \quad a < x < b, \tag{2.1}$$

to be the truncated CDF of a random variable X with support (a, b). Then, we have the following result concerning its moments.

Theorem 2.1. *Let c and p be real numbers such that $a < c < b$ and $p > 0$. Then, the p-th moment about c is given by*

$$\mathbb{E}[(X - c)^p] = \frac{1}{G(b) - G(a)} \left[(b - c)^p G(b) - (a - c)^p G(a) - p I_G(c; a - c, b - c, p)\right], \tag{2.2}$$

where $I_G(c; s, t, p)$ is defined as

$$I_G(c; s, t, p) = \int_s^t y^{p-1} G(y + c)\, \mathrm{d}y, \quad s < t.$$

Proof: From (2.1), we have

$$\begin{aligned}(2.3)\qquad \mathbb{E}[(X-c)^p] &= \frac{1}{G(b)-G(a)}\int_a^b (x-c)^p\,\mathrm{d}G(x)\\ &= \frac{p}{G(b)-G(a)}\left\{\int_a^c\left[\int_0^{x-c} y^{p-1}\,\mathrm{d}y\right]\mathrm{d}G(x)+\int_c^b\left[\int_0^{x-c} y^{p-1}\,\mathrm{d}y\right]\mathrm{d}G(x)\right\}.\end{aligned}$$

Upon changing the order of integration, the expression in (2.3) can be written as

$$\begin{aligned}&= \frac{p}{G(b)-G(a)}\left\{\int_0^{a-c} y^{p-1}\left[\int_a^{y+c}\mathrm{d}G(x)\right]\mathrm{d}y+\int_0^{b-c} y^{p-1}\left[\int_{y+c}^{b}\mathrm{d}G(x)\right]\mathrm{d}y\right\}\\ &= \frac{1}{G(b)-G(a)}[(b-c)^pG(b)-(a-c)^pG(a)]\\ &\quad-\frac{p}{G(b)-G(a)}\int_{a-c}^0 y^{p-1}G(y+c)\,\mathrm{d}y-\frac{p}{G(b)-G(a)}\int_0^{b-c} y^{p-1}G(y+c)\,\mathrm{d}y,\end{aligned}$$

which completes the proof of the theorem. □

Corollary 2.1. *Under the conditions of Theorem 2.1, we have*

$$\mathbb{E}[(X-c)^p]\leqslant\begin{cases}(b-c)^p[1-F_X(c)]+(a-c)^pF_X(c), & \text{if } p \text{ is even},\\ (b-c)^p[1-F_X(c)], & \text{if } p \text{ is odd},\end{cases}$$

where F_X is as given in (2.1).

Proof: It is evident that

$$(2.4)\qquad I_G(c;a-c,b-c,p)=I_G(c;a-c,0,p)+I_G(c;0,b-c,p).$$

Assuming p to be even, for $a-c<y<0$, we have $y^{p-1}G(c)<y^{p-1}G(y+c)<y^{p-1}G(a)$, and for $0<y<b-c$, we have $y^{p-1}G(c)<y^{p-1}G(y+c)<y^{p-1}G(b)$. Consequently, we get

$$(2.5)\qquad I_G(c;a-c,0,p)\geqslant-\frac{G(c)(c-a)^p}{p}\quad\text{and}\quad I_G(c;0,b-c,p)\geqslant\frac{G(c)(b-c)^p}{p}.$$

Now, upon using the inequalities in (2.5) into (2.4), we obtain

$$(2.6)\qquad I_G(c;a-c,b-c,p)\geqslant\frac{G(c)}{p}\left[(b-c)^p-(c-a)^p\right].$$

Using the inequality in (2.6) in the expression (2.2) of Theorem 2.1, we get

$$(2.7)\quad \mathbb{E}[(X-c)^p]\leqslant\frac{1}{G(b)-G(a)}\left\{(b-c)^pG(b)-(a-c)^pG(a)-G(c)\left[(b-c)^p-(c-a)^p\right]\right\}.$$

From (2.1), the right-hand side of the inequality (2.7) can be rewritten as

$$=(b-c)^p[1-F_X(c)]+(a-c)^pF_X(c).$$

This proves the inequality for the case when p is even.

The inequality for the case when p is odd can be established in an analogous manner. □

When p is even, a lower bound for $\mathbb{E}[(X-c)^p]$ can be established as below.

Proposition 2.1. *Under the conditions of Theorem 2.1, for p even, we have*

$$\mathbb{E}[(X-c)^p] \geqslant (t-c)^p[F_X(b)-F_X(t)], \quad \text{if } c<t<b.$$

Proof: Suppose p is even. Then, as $(X-c)^p \geqslant 0$ and $(X-c)^p \geqslant (t-c)^p$ for $X>t>c$, it is clear that

$$\mathbb{E}[(X-c)^p] \geqslant \mathbb{E}\big[(X-c)^p \mathbb{1}_{\{X>t\}}\big] \geqslant (t-c)^p\, \mathbb{E}\big[\mathbb{1}_{\{X>t\}}\big],$$

which yields the required result. □

Proposition 2.2. *Under the conditions of Theorem 2.1, we have*

$$\min_{a<c<b} \mathbb{E}[(X-c)^p] \leqslant \begin{cases} \left(\dfrac{b-a}{2}\right)^p, & \text{if } p \text{ is even,} \\ 0, & \text{if } p \geqslant 1 \text{ is odd.} \end{cases}$$

Proof: Suppose p is even. In this case, from Corollary 2.1, we have

$$\mathbb{E}[(X-c)^p] \leqslant (b-c)^p[1-F_X(c)] + (a-c)^p F_X(c) \leqslant S(F(c))T(c), \tag{2.8}$$

where $S(F(c)) = \max\{1-F_X(c), F_X(c)\}$ and $T(c) = (b-c)^p + (a-c)^p$. As $F_X(c) = 1/2$ is a minimum point of $M(F(c))$, we have

$$\min_{0<F(c)<1} S(F(c)) = \frac{1}{2}.$$

Taking the minimum over $0<F(c)<1$ in (2.8), we get

$$\mathbb{E}[(X-c)^p] \leqslant \frac{1}{2}T(c). \tag{2.9}$$

Now, taking the minimum over $a<c<b$ in (2.9) and using the fact that the function $T(c)$ reaches a minimum value at the point $c=(a+b)/2$, we get

$$\min_{a<c<b} \mathbb{E}[(X-c)^p] \leqslant \frac{1}{2}T\left(\frac{a+b}{2}\right) = \left(\frac{b-a}{2}\right)^p.$$

This proves the inequality for the case when p is even. Further, the inequality for the case when p is odd trivially follows from Corollary 2.1. □

Proposition 2.3. *The variance $\sigma^2 = \mathrm{Var}(X)$ can be expressed as*

$$\sigma^2 = \frac{1}{G(b)-G(a)}\left[(b-\mu)^2 G(b) - (a-\mu)^2 G(a) - 2I_G(\mu; a-\mu, b-\mu)\right],$$

where $\mu = \mathbb{E}(X)$ and

$$I_G(\mu; s, t) = \int_s^t yG(y+\mu)\,\mathrm{d}y, \quad s<t.$$

Proof: By taking $c=\mu$ and $p=2$ in Theorem 2.1, we readily obtain the required result. □

Proposition 2.4. *The Popoviciu inequality (Popoviciu, 1935) on variances given by*

$$\sigma^2 \leqslant \left(\frac{b-a}{2}\right)^2,$$

follows from Proposition 2.2.

Proof: The proof follows immediately by setting $p = 2$ in Proposition 2.2. □

A reverse form of Popoviciu's inequality can be obtained upon taking $p = 2$, $c = \mu = \mathbb{E}(X)$ and $t = (\mu + b)/2$ in Proposition 2.1.

Proposition 2.5. *We have*

$$\sigma^2 \geqslant \left(\frac{b-\mu}{2}\right)^2 \left[F_X(b) - F_X\left(\frac{\mu+b}{2}\right)\right].$$

Asymptotic behavior

In the following theorem, we establish the asymptotic behaviour of the p-th moment.

Theorem 2.2. *If X is distributed as in (2.1), then we have*

$$\lim_{\frac{b-a}{2} \longrightarrow 0^+} \mathbb{E}\left[\left(\frac{X-a}{b-a}\right)^p\right] = \frac{1}{p+1}, \quad p > -1,$$

and

$$\lim_{\frac{b-a}{2} \longrightarrow \infty} \mathbb{E}\left[\left(\frac{X-a}{b-a}\right)^p\right] = \frac{1}{2^p}.$$

Proof: From (2.1), we find (for $0 < z < 1$)

$$F_{\frac{X-a}{b-a}}(z) = \frac{G\big(z(b-a)+a\big) - G(a)}{G(b) - G(a)} = \frac{G\big(\frac{b-a}{2}(2z-1) + \frac{a+b}{2}\big) - G\big(\frac{a+b}{2} - \frac{b-a}{2}\big)}{G\big(\frac{b-a}{2} + \frac{a+b}{2}\big) - G\big(\frac{a+b}{2} - \frac{b-a}{2}\big)}.$$

Now, it is a simple task to verify that

$$\lim_{\frac{b-a}{2} \longrightarrow 0} F_{\frac{X-a}{b-a}}(z) = F_U(z), \quad \forall z \in \mathbb{R},$$

and

$$\lim_{\frac{b-a}{2} \longrightarrow \infty} F_{\frac{X-a}{b-a}}(z) = F_Y(z), \quad \forall z \neq \frac{1}{2},$$

where $U \sim U(0,1)$ and Y is a discrete random variable such that $\mathbb{P}(Y = 1/2) = 1$.

Moreover, we note that $[(X-a)/(b-a)]^p$ is uniformly integrable because $0 < (X-a)/(b-a) < 1$. As convergence in distribution along with uniform integrability imply convergence in mean (cf. Billingsley, 1968, Theorem 5.4), we have

$$\lim_{\frac{b-a}{2}\longrightarrow 0^+} \mathbb{E}\left[\left(\frac{X-a}{b-a}\right)^p\right] = \mathbb{E}(U^p), \quad p > -1,$$

and

$$\lim_{\frac{b-a}{2}\longrightarrow \infty} \mathbb{E}\left[\left(\frac{X-a}{b-a}\right)^p\right] = \mathbb{E}(Y^p),$$

which completes the proof of the theorem. □

Corollary 2.2. *We further have*

$$\lim_{\frac{b-a}{2}\longrightarrow 0^+} \frac{\sigma^2+(\mu-a)^2}{(b-a)^2} = \frac{1}{3} \quad \text{and} \quad \lim_{\frac{b-a}{2}\longrightarrow \infty} \frac{\sigma^2}{(b-a)^2} = 0,$$

where $\mu = \mathbb{E}(X)$ *and* $\sigma^2 = \mathrm{Var}(X)$.

Proof: By taking $p = 2$ in Theorem 2.2, the above results follow readily. □

2.2. Symmetric case

In this section, we assume that G is a skewing function, i.e., it is such that $G(x) \geqslant 0$ and $G(-x) = 1 - G(x)$, and X is as distributed in (2.1) with $a = -\ell$ and $b = \ell > 0$; that is, X has its CDF as

$$F_X(x) = \frac{G(x)+G(\ell)-1}{2G(\ell)-1}, \quad -\ell < x < \ell. \tag{2.10}$$

Proposition 2.6. *The variance* σ^2 *can be expressed as*

$$\sigma^2 = \ell^2 H(\ell),$$

where

$$H(\ell) = 1 - \frac{2C(\ell)-1}{2G(\ell)-1},$$

with $C(\ell) = C(\ell, G)$ *being defined as*

$$C(\ell) = \frac{2}{\ell^2}\int_0^\ell yG(y)\,\mathrm{d}y.$$

Moreover, $1/2 \leqslant C(\ell) \leqslant G(\ell)$.

Proof: As G is a skewing function, we have $\mu = \mathbb{E}(X) = 0$. Moreover, $I_G(\mu; a-\mu, b-\mu)$ in Proposition 2.3 satisfies the identity

$$I_G(\mu; a-\mu, b-\mu) = \ell^2 \left[C(\ell) - \frac{1}{2} \right]. \tag{2.11}$$

Upon substituting (2.11) in Proposition 2.3 and carrying out some simple algebraic steps, the required result follows. □

Remark 2.1. It is useful to observe that, knowing $C(\ell)$ (see Table 1 for some explicit examples of these constants), Proposition 2.6 gives a more informative result than Popoviciu's inequality and present in particular a method for the exact calculation of the variance of truncated distributions of the form in (2.10).

Table 1: Some examples of constants $C(\ell)$, for use in Proposition 2.6.

Distribution	$G(x)$	$C(\ell)$
Normal	$\Phi(x)$	$\frac{1}{2\ell^2}\left\{\ell\left[\ell + \exp\left(-\frac{\ell^2}{2}\right)\sqrt{\frac{2}{\pi}}\right] + (\ell^2-1)\ \mathrm{erf}\left(\frac{\ell}{\sqrt{2}}\right)\right\}$
Student-t $(\nu = 2)$	$\frac{1}{2}\left(1 + \sqrt{\frac{x^2}{x^2+1}}\right)$	$\frac{1}{2\ell^2}\left\{\ell^2 + \frac{1}{\ell}\sqrt{\frac{\ell^2}{2+\ell^2}}\left[2\ell + \ell^3 - 2\sqrt{2+\ell^2}\ \mathrm{arcsinh}\left(\frac{\ell}{\sqrt{2}}\right)\right]\right\}$
Cauchy	$\frac{1}{\pi}\ \arctan(x) + \frac{1}{2}$	$\frac{1}{2\pi\ell^2}\left[\ell(\ell\pi - 2) + 2(1+\ell^2)\ \arctan(\ell)\right]$
Laplace	$\frac{1}{2} + \frac{1}{2}\ \mathrm{sgn}(x)\left[1 - \exp(-\lvert x\rvert)\right]$	$\frac{1}{\ell^2}\ \exp(-\ell)\left[1 + \ell + \exp(\ell)(\ell^2 - 1)\right]$
Logistic	$\frac{1}{1+\exp(-x)}$	$\frac{2}{\ell^2}\left\{\frac{\pi^2}{12} + \ell \log[1+\exp(\ell)] + \mathrm{Li}_2[-\exp(\ell)]\right\}$

In Table 1, Φ is the CDF of a standard normal distribution, $\mathrm{Li}_2[z] = -\int_0^z \log(1-x)/x \mathrm{d}x$ is the polylogarithm function of order 2, and $\mathrm{erf}(z) = (2/\sqrt{\pi})\int_0^z \exp(-x^2)\mathrm{d}x$ is the error function.

Proposition 2.7. *We further have*

$$\lim_{\ell \longrightarrow 0^+} H(\ell) = \lim_{\ell \longrightarrow 0^+} \frac{\sigma^2}{\ell^2} = \frac{1}{3} \quad \text{and} \quad \lim_{\ell \longrightarrow \infty} H(\ell) = \lim_{\ell \longrightarrow \infty} \frac{\sigma^2}{\ell^2} = 0.$$

Proof: Upon taking $a = -\ell$, $b = \ell$ and $\mu = 0$ in Corollary 2.2, the required result follows. □

Next, we present two further examples in addition to these in Table 1.

Example 2.1. Let X have a truncated symmetric standard Cauchy distribution with density function $f(x) = (2\arctan\ell)^{-1} \cdot (1+x^2)^{-1}$ if $|x| < \ell$, with $\ell \leqslant 1$, and $f(x) = 0$ if $|y| > \ell$. As its variance is $\sigma^2 = \ell/\arctan\ell - 1$ (see Johnson *et al.*, 1994, p. 322), we obtain

$$H(\ell) = \frac{1}{\ell \arctan\ell} - \frac{1}{\ell^2}.$$

Example 2.2. Let X have a symmetrically truncated standard Gaussian distribution with density function $f(x) = \{\sqrt{2\pi}[1 - 2\Phi(-\ell)]\}^{-1} \exp(-x^2/2)$ if $|x| < \ell$, where $\ell \leqslant 1$ and Φ is the standard Gaussian cumulative distribution function, and $f(x) = 0$ if $|x| > \ell$. As its variance can be expressed as $\sigma^2 = 1 - 2\ell f(\ell)$ (see Johnson *et al.*, 1994, p. 158), we find

$$H(\ell) = \frac{1}{\ell^2} - \frac{2}{\ell^3 f(\ell)}.$$

In both these examples, we observe that $H(\ell) \to 1/3$ as $\ell \downarrow 0$ and $H(\ell) \uparrow 0$ as $\ell \uparrow \infty$, as stated in Proposition 2.7. However,

$$\ell H(\ell) = \frac{\sigma^2}{\ell}$$

behaves differently as $\ell \uparrow \infty$. In the first example, $\ell H(\ell) \to 2/\pi$, but in the second example, we find $\ell H(\ell) \to 0$.

3. ILLUSTRATION WITH FINANCIAL DATA

We illustrate the results developed here with intraday spot exchange rate data of four currencies against the US dollar transacted on the foreign exchange (Forex) market. There are 16 million tick-by-tick returns of bid prices provided by Tick Data, LLC (Table 2). Following the discussion regarding the truncated nature of the past (Matsushita *et al.*, 2020, 2023), we consider the symmetric case here.

Table 2: Intraday spot exchange rate data description.

Country	Currency	Code	Period	Number of days (d)	Data points (n)
Britain	British pound	GBP	31 Aug 08 – 12 Jun 15	2,116	2,754,615
Canada	Canadian dollar	CAD	12 Jun 00 – 12 Jun 15	4,419	3,931,202
Japan	Japanese yen	JPY	30 May 00 – 12 Jun 15	4,598	4,804,463
Switzerland	Swiss franc	CHF	30 May 00 – 12 Jun 15	4,587	4,838,100
			Total	15,720	16,328,380

Source: Tick Data, LLC.

For each currency, let $\{X_{ij}\}$ denote the j-th return observed on day $i = 1, ..., d$ (Table 2). Taking the daily sample standard deviation s_i and the maximum daily absolute return $\ell_i = \max_{1 \leq j \leq n_i}\{X_{ij}\}$, where n_i denotes the sample size on day i with $n = \sum_{i=1}^{d} n_i$, Figure 1 depicts the daily sample ratios $\{s_i/\ell_i\}$ in the form of dots.

Now, consider the general sequence ignoring days as $\{X_t\}$, where $t = 1, ..., n$. Letting $\ell^* = \max_{1 \leq t \leq n}\{X_t\}$, we generated a grid of 1,000 truncation points, $\{\ell : \ell = m\ell^*/1000$, where $m = 1, ..., 1000\}$. For each ℓ over this grid, we obtained the sample standard deviation of the conditional (truncated) data $\{X_t : |X_t| \leq \ell\}$. In this way, we empirically find the form of the ratio σ/ℓ for the returns of each currency (Figure 1). Then, Proposition 2.6 provides a feasible and practical way of describing the relationship between the variance and the cutoff ℓ.

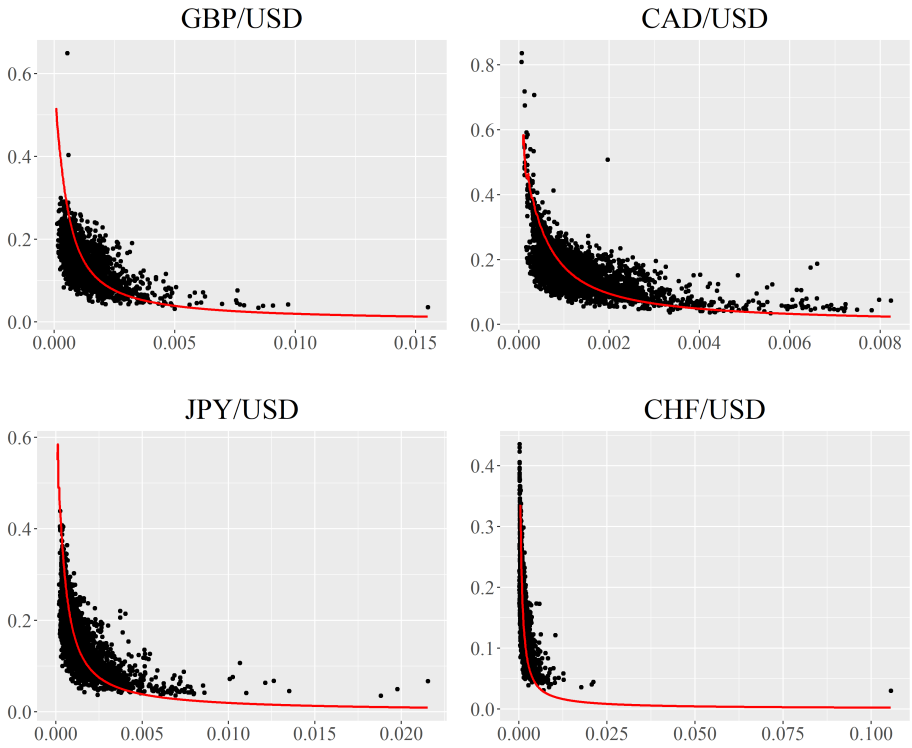

Figure 1: Display of ℓ *versus* the sample ratio σ/ℓ (lines) from data described in Table 2, where dots represent daily values.

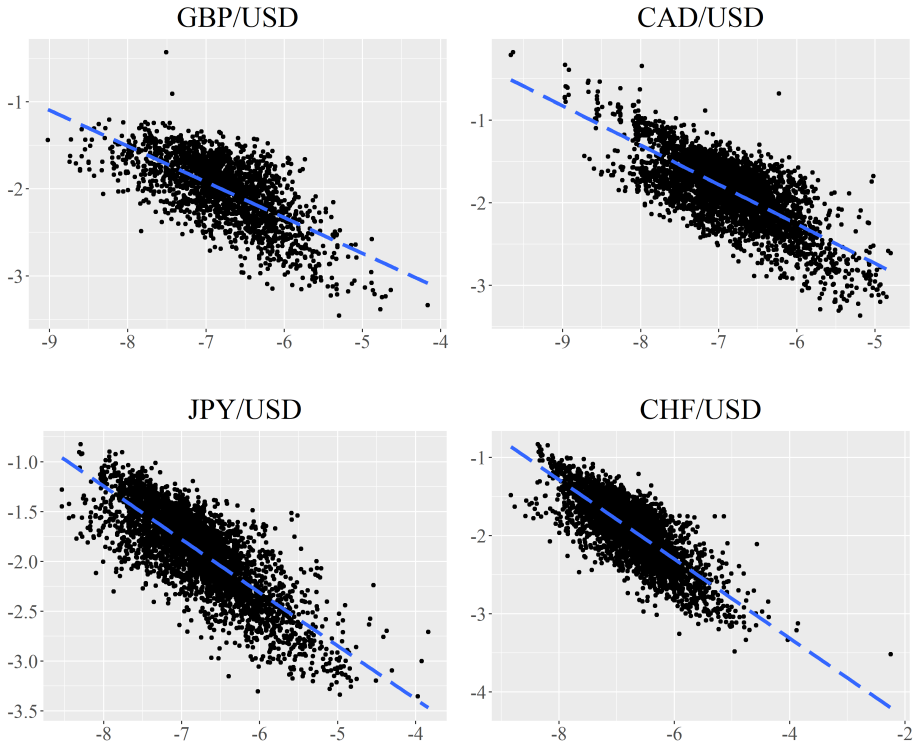

Figure 2: Display of $\ln \ell$ *versus* $\ln(\sigma/\ell)$ (dashed lines) from data described in Table 2, where dots represent daily values.

For small ℓ, Matsushita *et al.* (2020, 2023) proposed the power law $\sigma^{\beta}/\ell \approx \zeta$ from a second-order approximation, where $\beta > 0$ and ζ are real constants. So, we may approximate σ/ℓ as

$$\frac{\sigma}{\ell} \approx \zeta^{1/\beta} \ell^{-1+1/\beta}.$$

Figure 2 depicts the log-log plots of this approximate result, and shows the validity of such a power law approximation for σ/ℓ. Considering the symmetric case, Figure 1 illustrates how $H^{1/2}(\ell) = \sigma/\ell$ goes to zero as ℓ increases from the actual data. This decay approximately follows a power law of form $H^{1/2}(\ell) \propto \ell^{-1+1/\beta}$ (Figure 2), which is consistent with previous findings (Matsushita *et al.*, 2020). However, while we put forward a general form for any $\ell > 0$, the power law is valid only for ℓ close to zero (Matsushita *et al.*, 2023). Thus, our results can be applied to analyze other types of data besides financial data.

4. CONCLUDING REMARKS

In this work, we have presented a general approach for understanding the relationship between the variance and the range of a general family of truncated distributions based on skewing functions. We have established a closed-form expression for its moments and their asymptotic behavior as the support's semi-range tends to zero and ∞.

As discussed previously by Matsushita *et al.* (2020, 2023), if the truncated form arises naturally from the past, the function relating truncation length and standard deviation may assist in connecting the bounded past and unbounded future data. For this reason, we expect our results to be useful in many practical situations like the one demonstrated here with a real financial data.

ACKNOWLEDGMENTS

Roberto Vila gratefully acknowledges financial support from CNPq, CAPES, and FAP-DF, Brazil. Raul Matsushita acknowledges financial support from CNPq, CAPES, FAP-DF, and DPI/DPG/UnB, Brazil. The authors also express their sincere thanks to the Editor, Associate editor and the anonymous reviewers for their useful comments and suggestions on an earlier version of this manuscript which led to this improved version.

REFERENCES

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons, New York.

David, H. A., Hartley, H. O., and Pearson, E. S. (1954). The distribution of the ratio, in a single normal sample, of range to standard deviation. *Biometrika*, 41:482–493.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. 1, 2nd edition. John Wiley & Sons, New York.

Matsushita, R., Da Silva, S., Da Fonseca, R., and Nagata, M. (2020). Bypassing the truncation problem of truncated Lévy flights. *Physica A*, 559:125035.

Matsushita, R., Brom, P., Nagata, M., and Da Silva, S. (2023). Retrodicting with the truncated Lévy flight. *Communications in Nonlinear Science and Numerical Simulation*, 126:106900.

Popoviciu, T. (1935). Sur les équations algébriques ayant toutes leurs racines réelles. *Mathematica (Cluj)*, 9:129–145.

Shiffler, R. E. and Harsha, P. D. (1980). Upper and lower bounds for the sample standard deviation. *Teaching Statistics*, 2:84–86.

Shone, K. J. (1949). Relations between the standard deviation and the distribution of range in non-normal populations. *Journal of the Royal Statistical Society, Series B*, 11:85–88.

Tippett, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biometrika*, 17:364–387.

REVSTAT – Statistical Journal
Volume 23, Number 3, July 2025, 371–396
https://doi.org/10.57805/revstat.v23i3.599

A Novel Discrete Distribution Based on the Mixture of Poisson and Sum of Two Lindley Random Variables

Authors: Christophe Chesneau ✉
– Université de Caen Basse-Normandie, LMNO,
UFR de Sciences, F-14032, Caen, France
christophe.chesneau@gmail.com

Veena D'cruz
– Department of Statistics, Cochin University of Science and Technology,
Cochin 682 022, Kerala, India
veenadicruz@gmail.com

R. Maya
– Department of Statistics, Govt. College for Women,
Trivandrum 695 014, Kerala, India
publicationsofmaya@gmail.com

M. R. Irshad
– Department of Statistics, Cochin University of Science and Technology,
Cochin 682 022, Kerala, India
irshadm24@gmail.com

Received: December 2021 Revised: September 2023 Accepted: September 2023

Abstract:

- The distribution of the sum of two independent Lindley random variables having the same parameter defines the 2S-Lindley distribution recently introduced in the statistical literature. In this paper, we use it to create a new one-parameter discrete compound distribution called the Poisson 2S-Lindley distribution. More precisely, it is obtained by compounding the Poisson and 2S-Lindley distributions with the idea of combining their desirable functionalities. The mathematical and statistical properties of the proposed distribution are investigated and analyzed systematically. Subsequently, a statistical standpoint is adopted. Various estimation methods are considered, and an extensive simulation study is conducted to compare them. A count regression model as well as a first-order integer-valued autoregressive process based on the proposed model are constructed. In total, four real data sets in different fields are used to prove the empirical importance of the distribution.

Keywords:

- *2S-Lindley distribution; compounding; moments; stress strength reliability; simulation; count regression model; INAR(1) process.*

AMS Subject Classification:

- 60E05, 62E15, 62F10.

✉ Corresponding author

1. INTRODUCTION

Numerous applied as well as theoretical fields, such as medicine, sports, insurance, etc., deal with a large number of count data sets. Hence, count data modelling is of paramount importance. The Poisson (P) and negative binomial (NB) distributions are the most commonly used for this purpose. When an over-dispersion problem occurs, the P distribution is not considered ideal due to its equi-dispersion nature. The NB distribution is then more useful. However, when data sets have large skewness and kurtosis, the NB distribution is not well-adapted, and thus we look for more flexible distributions.

Mixed P distributions were then introduced for modelling non-homogeneous populations (see Karlis and Xekalaki, 2005, for more details). It consists of mixing the P distribution with well-known distributions. Mixed P distributions include the P Weibull distribution in Bereta *et al.* (2011), P transmuted Lindley distribution in Abd El-Monsef and Sohsah (2016), P transmuted exponential distribution in Bhati *et al.* (2017), P quasi Lindley distribution in Grine and Zeghdoudi (2017), new P weighted exponential distribution in Altun (2020), P Bilal distribution in Altun (2020), and P xgamma distribution in Altun *et al.* (2022), among others.

For other purposes, the Lindley distribution was originally proposed in Lindley (1958) in the context of Bayesian statistics. Various extensions and generalizations of the Lindley distribution have been proposed in the literature over the last decade. Chesneau *et al.* (2020) developed a continuous probability distribution with simple operations on two independent Lindley random variables. In particular, the distribution of the sum of two independent Lindley random variables with the same parameter is studied and named the 2S-Lindley (2S-L) distribution. They investigated the applications and structural properties of the corresponding model.

On the other hand, to model over-dispersed data, various count regression models based on many discrete distributions, including mixed P distributions with auxiliary variables, were introduced recently. They aim to predict the behavior of the count response variable similar to that of the P regression model. A count regression model based on the generalized Waring distribution was developed in Rodríguez-Avi *et al.* (2009). A study on crash data was done in Lord and Geedipally (2011) based on the NB-Lindley regression model. Sáez-Castillo and Conde-Sánchez (2013) introduced the hyper-P regression model. Cheng *et al.* (2013) analyzed the P Weibull regression model with real-life data sets. The P-weighted exponential distribution with its count regression model was developed in Zamani *et al.* (2014). Bhati *et al.* (2017) introduced a regression model for the over-dispersed dependent variable based on the P transmuted exponential distribution. Also, many authors compounded P with various generalized versions of the Lindley distribution and hence introduced count regression models based on it (see Mahmoudi and Zakerzadeh, 2010; Wongrin and Bodhisuwan, 2017; and Altun, 2021).

When the P distribution is used as the innovation distribution for the INAR(1) process, equi-dispersion (empirical mean equals empirical variance) turns out to be a disadvantage for modelling over-dispersed (empirical mean less than empirical variance) time series of counts, primarily in McKenzie (1985) and Al-Osh and Alzaid (1987), independently. Hence, to solve this, researchers came up with various alternatives. The INAR(1) process with P-Lindley

innovations (INAR(1)PL) in Lívio *et al.* (2018), INAR(1) process with new P weighted exponential innovations ((INAR(1)NPWE)) in Altun (2020), INAR(1) process with discrete three-parameter Lindley innovations in Eliwa *et al.* (2020), INAR(1) process with P quasi xgamma innovations (INAR(1)PQX) in Altun *et al.* (2021), INAR(1) process with Bell innovations (INAR(1)BL) in Huang and Zhu (2021), INAR(1) with P transmuted exponential innovations (INAR(1)PTE) in Altun and Khan (2022), INAR(1) with discrete pseudo Lindley innovations (INAR(1)DPsL) in Irshad *et al.* (2021) are some of the recently developed over-dispersed INAR(1) processes.

In this paper, we construct a new one-parameter compound distribution based on the P distribution and the probability density function (pdf) of the 2S-L distribution as described above. The newly proposed distribution, named the P2S-L distribution, is a one-parameter distribution having simple and explicit forms for its probability mass function (pmf), cumulative distribution function (cdf), moments, probability generating function (pgf), moment generating function (mgf), stress-strength reliability function, etc. Under the P2S-L distribution, a count regression model and an INAR(1) process are proposed to model over-dispersed data sets. Four data sets are used to illustrate its applicability.

The important feature of the P2S-L distribution is that it has a simpler explicit expression for moments as well as other statistical properties, and the P2S-L distribution is effective in modelling over-dispersed data sets. Furthermore, the 2S-L distribution was chosen over the Lindley and exponential distributions because the convolution operation defining its pdf makes it more flexible and suitable.

The rest of the paper is organized as follows: In Section 2, the proposed distribution is defined and some statistical properties are derived. Section 3 describes the estimation methods used to estimate the unknown parameter, and their finite sample performance through a simulation study is checked. Section 4 introduces a new count regression model, and Section 5 defines a new INAR(1)P2S-L process. Section 6 deals with four real-life data sets to prove the potential of the proposed distribution. Section 7 includes concluding observations.

2. THE POISSON 2S-LINDLEY DISTRIBUTION

Let us first define the 2S-L distribution proposed in Chesneau *et al.* (2020). Suppose that Y_1 and Y_2 are two independent random variables following the one-parameter Lindley distribution in Lindley (1958) with the same parameter denoted by θ. Then the random variable Y defined as the sum of Y_1 and Y_2, i.e., $Y = Y_1 + Y_2$, possesses the following pdf:

$$f(y;\theta) = \frac{\theta^4}{(1+\theta)^2} y \left(\frac{y^2}{6} + y + 1 \right) e^{-\theta y}, \quad y, \theta > 0.$$

It is understood that $f(y;\theta) = 0$ for $y \leq 0$. This pdf characterizes the 2S-L distribution. The corresponding cdf is given as

$$F(y;\theta) = 1 - \frac{1}{6(1+\theta)^2} \left[\theta^3 y \left(y^2 + 6y + 6 \right) + 3\theta^2 \left(y^2 + 4y + 2 \right) + 6\theta(y+2) + 6 \right] e^{-\theta y}, \quad y, \theta > 0.$$

The 2S-L distribution is the basis of our proposed one-parameter discrete compound distribution, which is mathematically formulated in the definition below.

Definition 2.1. Suppose that X is a random variable having the Poisson 2S-Lindley (P2S-L) distribution such that

$$X|\lambda \sim \mathrm{P}(\lambda) \quad \text{and} \quad \lambda|\theta \sim \text{2S-L}(\theta),$$

where $\lambda > 0, \theta > 0$, and $\sim \mathcal{D}$ denotes "follows the $\mathcal{D}$ distribution". Then, the unconditional pmf of X is

$$P(x;\theta) = \frac{\theta^4(1+x)}{6(1+\theta)^{6+x}}\big[x^2 + 6(\theta+2)^2 + x(11+6\theta)\big], \quad x = 0, 1, 2, ... \tag{2.1}$$

The distribution defined with the pmf (2.1) is referred to as the P2S-L distribution with parameter θ.

Proof: Using the procedure of compounding and integral developments, we have

$$\begin{aligned} P(x;\theta) &= \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!}\frac{\theta^4}{(1+\theta)^2}\lambda\left(\frac{\lambda^2}{6} + \lambda + 1\right)e^{-\theta\lambda}\mathrm{d}\lambda \\ &= \frac{\theta^4}{x!(\theta+1)^2}\left[\frac{1}{6}\int_0^\infty e^{-\lambda(\theta+1)}\lambda^{x+3}\mathrm{d}\lambda + \int_0^\infty e^{-\lambda(\theta+1)}\lambda^{x+2}\mathrm{d}\lambda + \int_0^\infty e^{-\lambda(\theta+1)}\lambda^{x+1}\mathrm{d}\lambda\right] \\ &= \frac{\theta^4(1+x)}{6(1+\theta)^{6+x}}\big[x^2 + 6(\theta+2)^2 + x(11+6\theta)\big]. \end{aligned}$$

The proof is completed. □

After some mathematical developments, the corresponding cdf of the P2S-L distribution is

$$\begin{aligned} F(x;\theta) = \frac{1}{6(1+\theta)^{6+x}}\Big\{&6\big(-1+(1+\theta)^x\big) + \theta\Big[-6(6+x) - 3(5+x)(6+x)\theta \\ &- (4+x)(5+x)(6+x)\theta^2 - 6\big(11+x(7+x)\big)\theta^3 - 6(2+x)\theta^4 \\ &+ 6(1+\theta)^x(2+\theta)(1+\theta+\theta^2)\big(3+\theta(3+\theta)\big)\Big]\Big\}. \end{aligned}$$

Proposition 1. The pmf given in (2.1) is log concave.

Proof: We need to prove that $\frac{P(x+1;\theta)}{P(x;\theta)}$ is a decreasing function in x. We have

$$\frac{P(x+1;\theta)}{P(x;\theta)} = \frac{(2+x)\big[x^2 + 6(2+\theta)(3+\theta) + x(13+6\theta)\big]}{(1+x)(1+\theta)\big[x^2 + 6(2+\theta)^2 + x(11+6\theta)\big]}.$$

Hence, by considering this ratio as a continuous function with respect to x and taking its derivative, we obtain

$$\frac{\partial}{\partial x}\frac{P(x+1;\theta)}{P(x;\theta)} = -\frac{3\left[\frac{1}{\theta+1} + \frac{(4\theta+4x+10)+2(x+1)\big(2\theta(x+1)+3x+4\big)}{\big(6(\theta+2)^2+x^2+(6\theta+11)x\big)^2}\right]}{(x+1)^2},$$

which is clearly negative, implying that $\frac{P(x+1;\theta)}{P(x;\theta)}$ is a decreasing function in x for every possible value of θ. This ends the proof. □

Figure 1 shows the possible pmf shapes of the P2S-L distribution for various values of θ.

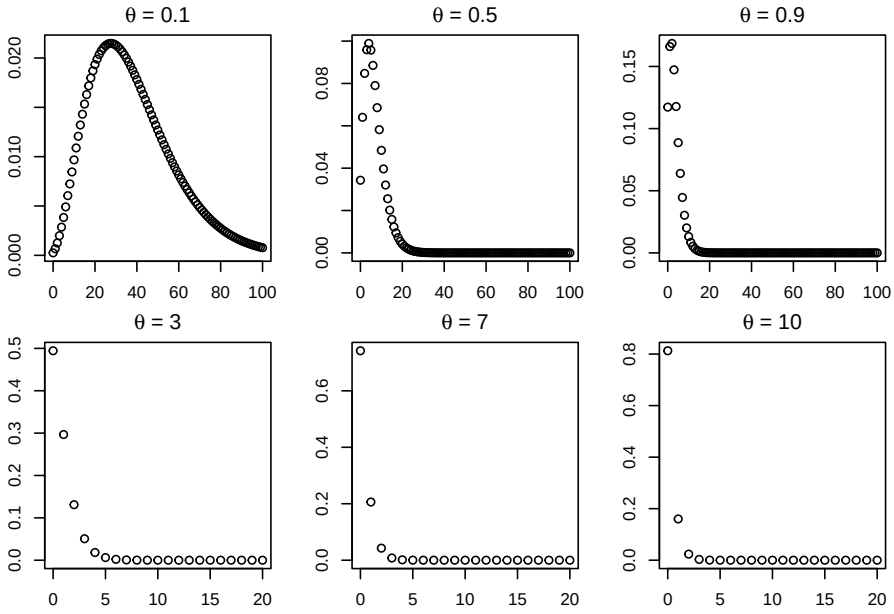

Figure 1: Some plots of the pmf of the P2S-L distribution for chosen values for θ.

The figure clearly indicates that the P2S-L distribution is right-skewed and, as θ goes larger, the mass will concentrate more on nearer to 0 values.

2.1. Moments, skewness and kurtosis

Some moment functions, skewness and kurtosis of the P2S-L distribution are now under investigation. If X represents a random variable having the P2S-L distribution, then the pgf of X is

$$G(s) = \mathrm{E}(s^X) = \sum_{x=0}^{\infty} s^x P(x;\theta) = \frac{\theta^4(2-s+\theta)^2}{(1+\theta)^2(1-s+\theta)^4}, \tag{2.2}$$

for $|s| < 1+\theta$. The mgf of X can be obtained by substituting s in (2.2) by e^t, and we obtain

$$M(t) = \mathrm{E}(e^{tX}) = \frac{\theta^4\big(2-e^t+\theta\big)^2}{(1+\theta)^2(1-e^t+\theta)^4},$$

for $t \le \ln(1+\theta)$. The r-th factorial moment of X is described in the next result.

Proposition 2. The r-th factorial moment of X is derived as

$$\mu_{[r]} = \mathrm{E}\big(X(X-1)\cdots(X-r+1)\big) = \frac{\big[r^2+6(1+\theta)^2+r(5+6\theta)\big](r+1)!}{6\theta^r(1+\theta)^2}.$$

Proof: The r-th factorial moment of X is obtained by integrating. That is,

$$\begin{aligned}\mu_{[r]} &= \int_0^\infty \frac{\theta^4}{(1+\theta)^2}\lambda^{r+1}\left(\frac{\lambda^2}{6}+\lambda+1\right)e^{-\theta\lambda}\mathrm{d}\lambda \\ &= \frac{\theta^4}{(\theta+1)^2}\left(\frac{1}{6}\int_0^\infty \lambda^{r+3}e^{-\theta\lambda}\mathrm{d}\lambda + \int_0^\infty \lambda^{r+2}e^{-\theta\lambda}\mathrm{d}\lambda + \int_0^\infty \lambda^{r+1}e^{-\theta\lambda}\mathrm{d}\lambda\right) \\ &= \frac{\left[r^2+6(1+\theta)^2+r(5+6\theta)\right](r+1)!}{6\theta^r(1+\theta)^2}.\end{aligned}$$

Hence the proof. □

Based on Proposition 2, the first four moments of X follow as

$$\begin{aligned}\mathrm{E}(X) &= \frac{2(2+\theta)}{\theta(1+\theta)}, \\ \mathrm{E}(X^2) &= \frac{2\left(\theta^3+6\theta^2+14\theta+10\right)}{\theta^2(\theta+1)^2}, \\ \mathrm{E}(X^3) &= \frac{2\left(\theta^4+12\theta^3+50\theta^2+90\theta+60\right)}{\theta^3(\theta+1)^2}, \\ \mathrm{E}(X^4) &= \frac{2\left(\theta^5+24\theta^4+158\theta^3+490\theta^2+720\theta+420\right)}{\theta^4(\theta+1)^2},\end{aligned}$$

and the variance of X is

$$\mathrm{Var}(X) = \frac{2\left(\theta^3+4\theta^2+6\theta+2\right)}{\theta^2(\theta+1)^2}.$$

It is worth noting that when θ tends to 0, the above moment measures tend to ∞, and when θ tends to ∞, they tend to 0. The skewness and kurtosis of the P2S-L distribution can be calculated using the following formulas:

$$\mathrm{Skewness}(X) = \frac{\mathrm{E}(X^3)-3\mathrm{E}(X^2)\mathrm{E}(X)+2[\mathrm{E}(X)]^3}{[\mathrm{Var}(X)]^{3/2}} \tag{2.3}$$

and

$$\mathrm{Kurtosis}(X) = \frac{\mathrm{E}(X^4)-4\mathrm{E}(X^3)\mathrm{E}(X)+6\mathrm{E}(X^2)[\mathrm{E}(X)]^2-3[\mathrm{E}(X)]^4}{[\mathrm{Var}(X)]^2}, \tag{2.4}$$

respectively. Both equations (2.3) and (2.4) will result in explicit forms.

2.2. Dispersion index and coefficient of variation

The expressions of the coefficient of variation (CV) and dispersion index (DI) of the P2S-L distribution are obtained in this section. The coefficient of variation is calculated as

$$\mathrm{CV}(X) = \frac{\sqrt{2(\theta^3+4\theta^2+6\theta+2)}}{2(\theta+2)}.$$

The dispersion index is given by

$$\mathrm{DI}(X) = 1+\frac{1}{\theta}+\frac{1}{\theta+1}-\frac{1}{\theta+2}.$$

Clearly, DI(X) is always greater than 1. As a result, the P2S-L distribution has over-dispersed properties.

Numerical values for the mean, variance, DI, skewness, and kurtosis for different sets of parameter values are reported in Table 1.

Table 1: Numerical values for some moment measures of the P2S-L distribution for various values of θ.

Measures	θ					
	0.1	0.5	1.9	2.0	5.0	8.0
Mean	38.1818	6.6667	1.4156	1.3333	0.4667	0.2778
Variance	436.5289	21.7778	2.2858	2.1111	0.5711	0.3156
DI	11.4329	3.2667	1.6147	1.5833	1.2238	1.1361
Skewness	1.0120	1.1297	1.4551	1.4731	1.9126	2.2638
Kurtosis	4.5261	4.8364	5.9078	5.9709	7.5983	9.0977

From Table 1, the P2S-L distribution would be appropriate for modelling over-dispersed, right-skewed and leptokurtic data sets.

2.3. Stress-strength analysis

Stress-strength reliability has wide applications in almost all fields of engineering and machine learning. Let X_{stress} and $X_{strength}$ be discrete random variables with positive integer values that model the stress and strength of a system, respectively. Then, the expected reliability can be calculated using the following formula:

$$R_{\text{Stress-Strength}} = \Pr(X_{\text{Stress}} < X_{\text{Strength}}) = \sum_{x=0}^{\infty} P_{X_{\text{Stress}}}(x)\left(1 - F_{X_{\text{Strength}}}(x)\right),$$

where $P_X(x)$ and $F_X(x)$ are the pmf and cdf of a random variable X, respectively. The expression for stress-strength reliability when $X_{\text{Stress}} \sim$ P2S-L with parameter θ_1 and $X_{\text{Strength}} \sim$ P2S-L with parameter θ_2, and X_{Stress} and X_{Strength} are independent, is obtained in explicit form as follows:

$$\begin{aligned}
R_{\text{Stress-Strength}} &= \frac{\theta_2^4}{(1+\theta_1)^2(1+\theta_2)^2\big(\theta_2+\theta_1(1+\theta_2)\big)^7} \\
&\times \theta_1^3\Big\{35+\theta_1(5+2\theta_1)\Big\{30+\theta_1\Big[39+\theta_1\big[30+\theta_1(11+2\theta_1)\big]\Big]\Big\}\Big\} \\
&+ \theta_1^2\Big\{21+\theta_1\Big\{217+\theta_1\Big[711+\theta_1\big[1111+\theta_1\big(953+\theta_1\big(471+16\theta_1(8+\theta_1)\big)\big)\big]\Big]\Big\}\Big\}\theta_2 \\
&+ \theta_1(1+\theta_1)(7+\theta_1)\Big\{91+\theta_1\Big\{434+\theta_1\Big[918+\theta_1\big[987+\theta_1\big(573+\theta_1(177+25\theta_1)\big)\big]\Big]\Big\}\Big\}\theta_2^2 \\
&+ (1+\theta_1)^2\Big\{1+\theta_1\Big\{21+\theta_1\Big[126+\theta_1\big[350+\theta_1\big(477+\theta_1\big(331+\theta_1(119+19\theta_1)\big)\big)\big]\Big]\Big\}\Big\}\theta_2^3 \\
&+ (1+\theta_1)^3\Big\{2+\theta_1\Big\{19+\theta_1\Big[65+\theta_1\big(110+\theta_1\big(92+\theta_1(39+7\theta_1)\big)\big)\Big]\Big\}\Big\}\theta_2^4 + (1+\theta_1)^9\theta_2^5.
\end{aligned}$$

Some numerical values for $R_{\text{Stress-Strength}}$ for different values of the parameters are given in Table 2.

Table 2: Numerical values for $R_{\text{Stress-Strength}}$ for different values of θ_1 and θ_2.

$\theta_1 \rightarrow$ $\theta_2 \downarrow$	0.1	0.2	0.5	0.9	1	1.5	2	2.3
0.1	0.507498	0.196622	0.031747	0.008971	0.007219	0.003305	0.002029	0.001639
0.2	0.817939	0.514575	0.154903	0.058235	0.048808	0.025438	0.016749	0.013911
0.5	0.974228	0.871305	0.534687	0.310976	0.27908	0.183003	0.13734	0.120468
0.9	0.993732	0.958349	0.765952	0.560421	0.52376	0.397247	0.326012	0.297247
1	0.995124	0.966264	0.796984	0.602639	0.56673	0.439911	0.366428	0.336299
1.5	0.998087	0.984944	0.886158	0.742172	0.712211	0.597391	0.523796	0.49193
2	0.998971	0.991334	0.925507	0.815831	0.79141	0.693082	0.626112	0.596133
2.3	0.999227	0.993309	0.939357	0.844368	0.822612	0.733147	0.670611	0.642212

From Table 2, it is clear that the stress-strength reliability increases as $\theta_2 \rightarrow \infty$ and it decreases as $\theta_1 \rightarrow \infty$.

2.4. Generating random values from the P2S-L distribution

Random values from the P2S-L distribution can be generated using the steps given below. For $i = 1, 2, 3, ..., n$, consider:

1. $\lambda_i \sim \text{Lindley}(\theta)$ and $\lambda_i^* \sim \text{Lindley}(\theta)$, independently;

2. an observation x_i of $X_i \sim \text{P}(\lambda_i + \lambda_i^*)$.

The obtained generated values are $x_1, x_2, ..., x_n$.

3. ESTIMATION METHODS

To accurately predict the behavior of a given distribution, the unknown parameter(s) must be precisely estimated. Here, four classical estimation methods, such as the method of moments (MM), maximum likelihood (ML), least square (LS) and weighted least square (WLS) estimation methods, are used for this purpose. We assume that n observations of a random variable X with the P2S-L distribution having the unknown parameter θ are $x_1, x_2, ..., x_n$.

3.1. Method of moments

In the first method of estimation, the parameter of the P2S-L distribution is estimated with the MM. The theoretical and empirical moments are equated and solved to obtain the same.

Based on the first sample moment given as $\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, the MM estimate (MME) of θ, say $\hat{\theta}_{\text{MME}}$, satisfies $\text{E}(X) = \overline{x}_n$, which implies that

$$\hat{\theta}_{\text{MME}} = \frac{2 - \overline{x}_n + \sqrt{4 + 12\overline{x}_n + \overline{x}_n^2}}{2\overline{x}_n},$$

provided that $\overline{x}_n > 0$. A biased property of the MME is provided in the next proposition.

Proposition 3. The estimate $\hat{\theta}_{\text{MME}}$ is positively biased.

Proof: We have $\hat{\theta}_{\text{MME}} = g(\overline{x}_n)$, where $g(t) = \frac{2-t+\sqrt{4+12t+t^2}}{2t}$, with $t > 0$. Now, since

$$g''(t) = \frac{8\Big[2+\sqrt{4+t(12+t)}\Big] + 2t\left\{12\Big[3+\sqrt{4+t(12+t)}\Big] + t\Big[30+3t+\sqrt{4+t(12+t)}\Big]\right\}}{t^3\big[4+t(12+t)\big]^{3/2}} > 0,$$

$g(t)$ is strictly convex. The Jensen inequality states that $\text{E}\big(g(\bar{X}_n)\big) > g\big(\text{E}(\bar{X}_n)\big)$ when considering the random version of $\overline{x}_n$, denoted by $\bar{X}_n$. This implies that the random version of $\hat{\theta}_{\text{MME}}$, say $\tilde{\theta}_{\text{MME}}$, satisfies the inequality $\text{E}(\tilde{\theta}_{\text{MME}}) > \theta$. Therefore, $\hat{\theta}_{\text{MME}}$ is positively biased. This ends the proof. □

3.2. Maximum likelihood estimation

The second method of estimation consists of finding the ML estimate (MLE) of θ. To this end, the likelihood function is

$$\text{L} = \prod_{i=1}^{n} \frac{\theta^4(1+x_i)}{6(1+\theta)^{6+x_i}}\big[x_i^2 + 6(\theta+2)^2 + x_i(11+6\theta)\big]$$

and the log of the likelihood function is

$$\begin{aligned}\log \text{L} \;=\; & 4n\log\theta - 6n\log(1+\theta) - n\log(6) + \sum_{i=1}^{n}\log(1+x_i) \\ & + \sum_{i=1}^{n}\log\big[x_i^2 + 6(\theta+2)^2 + x_i(11+6\theta)\big] - \log(\theta+1)\sum_{i=1}^{n} x_i.\end{aligned}$$

Now, the MLE of θ, say $\hat{\theta}_{\text{MLE}}$, satisfies

$$\hat{\theta}_{\text{MLE}} = \text{argmax}_{\theta>0}\,\text{L} \quad \text{or} \quad \hat{\theta}_{\text{MLE}} = \text{argmax}_{\theta>0}\log\text{L},$$

where $\text{argmax}_{\{c\}}\, G(c)$ are the points c of some function G at which the function values are maximized. It can be obtained by solving the following non-linear equation:

$$\frac{\partial \log \text{L}}{\partial\theta} = 0 \implies \frac{4n}{\theta} - \frac{6n}{\theta+1} - \frac{1}{1+\theta}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\frac{6(4+x_i+2\theta)}{x_i^2 + 6(2+\theta)^2 + x_i(11+6\theta)} = 0. \tag{3.1}$$

We obtain $\hat{\theta}_{\text{MLE}}$ by numerically solving (3.1) using a mathematical software like `MATHEMATICA`, `MATHCAD` and `R`. In our study, we employ the optim function of the `R` software to obtain $\hat{\theta}_{\text{MLE}}$.

The variance and standard error (SE) of $\hat{\theta}_{\text{MLE}}$ are calculated using the inverse of the observed scalar information matrix. The `fdHess` function of the `R` software is used for the same purpose. For $\alpha \in (0,1)$, the asymptotic $100(1-\alpha)\%$ confidence interval (CI) for the parameter θ is

$$\left[\hat{\theta}_{\text{MLE}} \pm z_{\alpha/2}\sqrt{\text{Var}(\hat{\theta}_{\text{MLE}})}\right],$$

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

3.3. Least square and weighted least square estimation

Let the ordered values of $x_1, x_2, ..., x_n$ be $x_{(1)}, x_{(2)}, ..., x_{(n)}$ and set

$$\text{LS} = \sum_{i=1}^{n}\left[F\big(x_{(i)};\theta\big) - \frac{i}{n+1}\right]^2$$

and

$$\text{WLS} = \sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[F\big(x_{(i)};\theta\big) - \frac{i}{n+1}\right]^2.$$

We obtain the LS estimate (LSE) and WLS estimate (WLSE) of θ by

$$\hat{\theta}_{LSE} = \text{argmin}_{\theta>0}\,\text{LS} \quad \text{and} \quad \hat{\theta}_{WLSE} = \text{argmin}_{\theta>0}\,\text{WLS},$$

respectively, where $\text{argmin}_{\{c\}}\,G(c)$ are the points c of some function G at which the function values are minimized. The LS and WLS estimates are optimized using the `optim` function of the `R` software.

3.4. Simulation study

An extensive simulation study is used in this section to compare the above-mentioned methods for estimating the unknown parameter θ. Some arbitrarily chosen values for θ, i.e., $\theta = 0.2, 0.5, 0.8, 1.3, 1.7, 2.2$, are considered for different sample sizes ($n =$ 25, 50, 100, 200, 400 and 800). Also, $N = 1000$ replications are used for the same. In addition, measures such as bias and mean square error (MSE) are calculated with the following formulas:

$$\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}|\hat{\theta}_i - \theta|, \quad \text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2,$$

where the index i refers to the i-th experiment. Simulation results, including the bias and MSEs of the MME, MLE, LSE, and WLSE, are reported in Table 3.

Table 3: Simulation results for the P2S-L distribution.

$\theta = 0.2$								
n	MME		MLE		LSE		WLSE	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
25	0.002781	0.000533	0.002770	0.000533	0.005383	0.000513	0.005375	0.000487
50	0.001035	0.000257	0.001018	0.000257	0.005271	0.000272	0.005362	0.000257
100	0.000993	0.000126	0.000991	0.000126	0.005270	0.000146	0.005299	0.000145
200	0.000832	0.000061	0.000833	0.000061	0.005268	0.000085	0.005206	0.000081
400	0.000315	0.000028	0.000314	0.000028	0.005161	0.000056	0.005172	0.000055
800	0.000109	0.000015	0.000111	0.000015	0.004723	0.000042	0.004943	0.000042

$\theta = 0.5$								
n	MME		MLE		LSE		WLSE	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
25	0.008855	0.003876	0.008701	0.003872	0.034054	0.004318	0.036604	0.004153
50	0.003064	0.001945	0.002984	0.001943	0.034036	0.002693	0.036014	0.002635
100	0.002575	0.001028	0.002517	0.001028	0.033953	0.001911	0.035450	0.001949
200	0.001375	0.000482	0.001366	0.000481	0.033330	0.001505	0.035239	0.001593
400	0.000750	0.000225	0.000731	0.000226	0.033178	0.001346	0.034293	0.001471
800	0.000674	0.000123	0.000667	0.000123	0.031546	0.001264	0.032465	0.001431

$\theta = 0.8$								
n	MME		MLE		LSE		WLSE	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
25	0.012386	0.012897	0.012017	0.012838	0.089365	0.014970	0.110187	0.015583
50	0.008235	0.006071	0.007991	0.006056	0.089177	0.011887	0.105940	0.012316
100	0.002922	0.003083	0.002862	0.003090	0.089012	0.009989	0.102034	0.012215
200	0.002850	0.001628	0.002788	0.001627	0.087833	0.008724	0.096044	0.011598
400	0.001502	0.000753	0.001477	0.000754	0.086976	0.008421	0.088295	0.011250
800	0.000509	0.000358	0.000511	0.000358	0.078654	0.008254	0.088295	0.010699

$\theta = 1.3$								
n	MME		MLE		LSE		WLSE	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
25	0.029639	0.048833	0.034871	0.052769	0.245340	0.072775	0.338869	0.115103
50	0.013136	0.022640	0.016509	0.022893	0.243812	0.067841	0.324150	0.105609
100	0.006243	0.011242	0.009289	0.010663	0.243389	0.064562	0.310919	0.097875
200	0.005737	0.005420	0.006728	0.005411	0.241816	0.060674	0.297188	0.091021
400	0.005690	0.002699	0.002866	0.002688	0.240467	0.060552	0.281971	0.085582
800	0.000949	0.001443	0.001221	0.001348	0.230707	0.059784	0.2664455	0.084874

$\theta = 1.7$								
n	MME		MLE		LSE		WLSE	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
25	0.067738	0.109771	0.053360	0.103309	0.417770	0.188135	0.590104	0.348501
50	0.024998	0.044997	0.034529	0.047393	0.417183	0.182087	0.569252	0.324724
100	0.007982	0.023980	0.017156	0.022596	0.416915	0.181277	0.546202	0.299772
200	0.006952	0.011087	0.007388	0.010343	0.414872	0.175299	0.524531	0.278230
400	0.002851	0.004889	0.001480	0.005598	0.411906	0.175281	0.498656	0.256103
800	0.000955	0.002690	0.000496	0.002784	0.399852	0.174821	0.465326	0.233700

$\theta = 2.2$								
n	MME		MLE		LS		WLS	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
25	0.116734	0.290218	0.115604	0.290072	0.688772	0.483882	0.954364	0.911138
50	0.037157	0.095885	0.036454	0.095643	0.686642	0.482239	0.925549	0.857396
100	0.012432	0.045527	0.012031	0.045469	0.686362	0.480097	0.894067	0.801097
200	0.009737	0.024009	0.009611	0.024004	0.686263	0.475233	0.863627	0.749363
400	0.005487	0.009871	0.005440	0.009866	0.684225	0.472904	0.826796	0.692542
800	0.004243	0.005719	0.004180	0.005725	0.666975	0.472544	0.780233	0.629468

According to Table 3, as n increases, the bias and MSE tend to 0. As a result, it is clear that MME and MLE perform almost identically when estimating the unknown parameter θ, and thus both methods are appropriate for small and large sample sizes.

4. P2S-L REGRESSION MODEL

In Section 2.1, it is clearly mentioned that the P2S-L model is capable of modelling over-dispersed data sets. Such data sets are of paramount importance since they are most often observed in real life. Hence, we specialize in over-dispersion. This section introduces a count regression model based on the P2S-L distribution for modelling over-dispersed data sets.

4.1. Model construction

Let Y be the random variable which represents the counted number of occurrences of an event. We assume that it has the P2S-L distribution and satisfies $\mathrm{E}(Y) = \mu$, implying the following reparametrization:

$$\theta = \frac{2 - \mu + \sqrt{4 + 12\mu + \mu^2}}{2\mu}.$$

Assume we have n observations of Y, which is also the response variable, with the i-th observation being a realization of a random variable Y_i for $i = 1, 2, ..., n$, and the mean of Y_i is linked to the random covariate $X^\top = (X_1, X_2, ..., X_k)$ using the log-link function given by

$$\mu_i = \mathrm{E}(Y_i) = e^{x_i^\top \gamma}, \tag{4.1}$$

where $x_i^\top = (x_{i1}, x_{i2}, ..., x_{ik})$ is the covariate vector and $\gamma = (\gamma_0, \gamma_1, \gamma_2, ..., \gamma_k)$ is the unknown vector of regression coefficients. Based on (4.1), a linear form for the pmf of $Y_i | X_i^\top = x_i^\top$ which follows the P2S-L distribution with parameter μ_i is obtained as

$$\begin{aligned} P(y_i; e^{x_i^\top \gamma}) = & \frac{\left[\frac{2 - e^{x_i^\top \gamma} + \sqrt{4 + 12e^{x_i^\top \gamma} + \left(e^{x_i^\top \gamma}\right)^2}}{2e^{x_i^\top \gamma}}\right]^4 (1 + y_i)}{6\left\{1 + \left[\frac{2 - e^{x_i^\top \gamma} + \sqrt{4 + 12e^{x_i^\top \gamma} + \left(e^{x_i^\top \gamma}\right)^2}}{2e^{x_i^\top \gamma}}\right]\right\}^{6 + y_i}} \\ & \times \left\{ y_i^2 + 6\left[\left(\frac{2 - e^{x_i^\top \gamma} + \sqrt{4 + 12e^{x_i^\top \gamma} + \left(e^{x_i^\top \gamma}\right)^2}}{2e^{x_i^\top \gamma}}\right) + 2\right]^2 \right. \\ & \left. + \; y_i \left[11 + 6\left(\frac{2 - e^{x_i^\top \gamma} + \sqrt{4 + 12e^{x_i^\top \gamma} + \left(e^{x_i^\top \gamma}\right)^2}}{2e^{x_i^\top \gamma}}\right)\right] \right\}, \end{aligned}$$

where $y_i = 0, 1, 2, ...$

4.2. Estimation of the model

The ML method is used to estimate the regression coefficients in γ. The log of the likelihood function, say U, of the P2S-L count regression model is given by

$$\begin{aligned}\log \mathrm{U} \;=\; & 4\sum_{i=1}^{n}\left[\log\left(2-e^{x_i^\top\gamma}+\sqrt{4+12e^{x_i^\top\gamma}+\left(e^{x_i^\top\gamma}\right)^2}\right)-\log\left(2e^{x_i^\top\gamma}\right)\right]+\sum_{i=1}^{n}\log(1+y_i)\\ & -\; n\log 6\;-\;\sum_{i=1}^{n}(6+y_i)\log\left[1+\left(\frac{2-e^{x_i^\top\gamma}+\sqrt{4+12e^{x_i^\top\gamma}+\left(e^{x_i^\top\gamma}\right)^2}}{2e^{x_i^\top\gamma}}\right)\right]\\ & +\sum_{i=1}^{n}\log\left\{y_i^2+6\left[\left(\frac{2-e^{x_i^\top\gamma}+\sqrt{4+12e^{x_i^\top\gamma}+\left(e^{x_i^\top\gamma}\right)^2}}{2e^{x_i^\top\gamma}}\right)+2\right]^2\right.\\ & \left.+\;y_i\left[11+6\left(\frac{2-e^{x_i^\top\gamma}+\sqrt{4+12e^{x_i^\top\gamma}+\left(e^{x_i^\top\gamma}\right)^2}}{2e^{x_i^\top\gamma}}\right)\right]\right\}. \qquad (4.2)\end{aligned}$$

Now an estimation of the unknown vector γ is obtained by maximizing (4.2):

$$\hat{\gamma}=\mathrm{argmax}_{\gamma}\log \mathrm{U}.$$

In this regard, we use the `optim` function in the `R` software. Also, the SE of this estimate is calculated using the `fdHess` function in the `R` software.

4.3. Simulation of the P2S-L regression model

In this part, the ML method employed to estimate the unknown regression parameters is analyzed via a simulation study. By taking $k=2$, the parametric combinations ($\gamma_0=0.5$, $\gamma_1=0.1$, $\gamma_2=0.4$) and ($\gamma_0=0.2$, $\gamma_1=0.7$, $\gamma_2=0.5$) are used to generate $N=1000$ samples of sizes $n=50$, 250, 500 and 1000 from the model, $\log(\mu_i)=\gamma_0+\gamma_1x_{i1}+\gamma_2x_{i2}$.

Table 4: Simulation results for the P2S-L regression model.

n	Parameters	$\gamma_0=0.5,\ \gamma_1=0.1,\ \gamma_2=0.4$			$\gamma_0=0.2,\ \gamma_1=0.7,\ \gamma_2=0.5$		
		Estimates	Bias	MSE	Estimates	Bias	MSE
50	γ_0	0.459330	0.040670	0.069193	0.158021	0.041979	0.078119
	γ_1	0.115366	0.015366	0.077800	0.716583	0.016583	0.073192
	γ_2	0.403656	0.003656	0.079126	0.502873	0.002873	0.079101
250	γ_0	0.492112	0.007888	0.011192	0.190608	0.009392	0.013913
	γ_1	0.097662	0.002338	0.014621	0.702769	0.002769	0.014656
	γ_2	0.402847	0.002847	0.014029	0.500762	0.000762	0.013577
500	γ_0	0.494108	0.005892	0.006232	0.193702	0.006298	0.006675
	γ_1	0.098186	0.001814	0.007140	0.702319	0.002319	0.007309
	γ_2	0.401727	0.001727	0.007149	0.500547	0.000547	0.006969
1000	γ_0	0.496372	0.003628	0.002971	0.197913	0.002087	0.003236
	γ_1	0.099694	0.000307	0.003512	0.702143	0.002143	0.003694
	γ_2	0.400233	0.000233	0.003518	0.500348	0.000348	0.003456

We assume that x_{i1} and x_{i2} are generated from the U$(0,1)$ distribution. Here, measures such as estimates, bias, and MSEs are used to prove the asymptotic property of the MLEs. Table 4 reports the simulation results.

From Table 4, it is clear that as sample size increases, the bias and MSEs are nearer to 0, implying the consistency property of the MLEs for estimating the regression parameters.

5. INAR(1) PROCESS WITH P2S-L INNOVATIONS

In this section, we use the P2S-L distribution as an innovation distribution for the INAR(1) process for modelling over-dispersed count data sets. We have seen in the earlier sections that the P2S-L distribution is appropriate for over-dispersed count data sets. Also, in the empirical results section discussed later, we clearly prove that the P2S-L model provides better fits and results than some other recently developed overdispersed models. So the P2S-L distribution is applied here as an innovation distribution, and then the statistical properties of the process are investigated. A simulation study is included, and the parameters are estimated using the conditional maximum likelihood.

The process $\{X_t\}_{t\in\mathbb{Z}}$ defined by

$$X_t = p \circ X_{t-1} + \varepsilon_t, \tag{5.1}$$

with $p \in [0,1)$, is called the INAR(1) process if the innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are independent and identically distributed (iid). The symbol $\circ$ in (5.1) denotes the binomial thinning operator, which is described as

$$p \circ X_{t-1} = \sum_{j=1}^{X_{t-1}} C_j,$$

where $\{C_j\}_{j\in\mathbb{Z}}$ is a sequence of iid Bernoulli random variables with parameter p. Therefore, the corresponding INAR(1) process has the following one step transition probability:

$$\Pr\big(X_t = k \,|\, X_{t-1} = l\big) = \sum_{i=0}^{\min(k,l)} \Pr\big(B(l,p) = i\big) \Pr\big(\varepsilon_t = k - i\big), \qquad k, l = 0, 1, ...,$$

where $B(l,p) \sim$ Binomial(l,p) and $p \in [0,1)$. Historically, the P distribution was the first to be used with the INAR(1) process (for more information, see McKenzie, 1985 and Al-Osh and Alzaid, 1987).

If $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ has a P2S-L distribution with parameter θ, then the corresponding INAR(1)P2S-L process has the one step transition probability given as

$$\begin{aligned}\Pr\big(X_t = k \,|\, X_{t-1} = l\big) = {} & \sum_{i=0}^{\min(k,l)} \binom{l}{i} p^i (1-p)^{l-i} \\ & \times \frac{\theta^4[1+(k-i)]}{6(1+\theta)^{6+(k-i)}} \big[(k-i)^2 + 6(\theta+2)^2 + (k-i)(11+6\theta)\big].\end{aligned}$$

By using the expressions for the mean, variance, DI, conditional mean and conditional variance of $\{X_t\}_{t\in\mathbb{Z}}$ (see Wei, 2018 and Al-Osh and Alzaid, 1987), the mean, variance, DI,

conditional mean and conditional variance of the INAR(1)P2S-L process are derived as

$$\begin{aligned}
\mu_* &= \mathrm{E}(X_t) = \frac{2(2+\theta)}{(1-p)\theta(1+\theta)},\\
\sigma_*^2 &= \mathrm{Var}(X_t) = \frac{2\big(\theta^3(p+1)+\theta^2(3p+4)+2\theta(p+3)+2\big)}{\theta^2(\theta+1)^2(1-p^2)},\\
\mathrm{DI}_* &= \mathrm{DI}(X_t) = \frac{\theta^3(p+1)+\theta^2(3p+4)+2\theta(p+3)+2}{\theta(\theta+1)(\theta+2)(p+1)},
\end{aligned}$$

$$\mathrm{E}(X_t \,|\, X_{t-1}) = pX_{t-1} + \frac{2(2+\theta)}{\theta(1+\theta)} \tag{5.2}$$

and

$$\mathrm{Var}(X_t \,|\, X_{t-1}) = p(1-p)X_{t-1} + \frac{2\big(\theta^3+4\theta^2+6\theta+2\big)}{\theta^2(\theta+1)^2}. \tag{5.3}$$

5.1. Estimation of the process

Here, the estimation procedure using the conditional maximum likelihood (CML) method for the INAR(1)P2S-L process is examined. Let us consider T observations $x_1, x_2, ..., x_T$ of the INAR(1)P2S-L process from the random sample $X_1, X_2, ..., X_T$. Then the conditional log likelihood function is

$$\begin{aligned}
\ell(\Theta) &= \sum_{t=2}^{T} \log\Big[\Pr\big(X_t = x_t \,|\, X_{t-1} = x_{t-1}\big)\Big]\\
&= \sum_{t=2}^{T} \log\Bigg\{\sum_{i=0}^{\min(x_t, x_{t-1})} \binom{x_{t-1}}{i} p^i (1-p)^{x_{t-1}-i}\\
&\qquad \times \frac{\theta^4[1+(x_t-i)]}{6(1+\theta)^{6+(x_t-i)}}\Big[(x_t-i)^2 + 6(\theta+2)^2 + (x_t-i)(11+6\theta)\Big]\Bigg\},
\end{aligned}$$

where $\Theta = (\theta, p)^T$ is the unknown parametric vector to be estimated. The CML estimates (CMLEs) are obtained by

$$\hat{\Theta} = \mathrm{argmax}_{\Theta}\, \ell(\Theta).$$

The optim and `fdHess` functions of the `R` program are used to obtain the CMLEs, observed information matrix, and therefore the SEs of the parameter estimates.

5.2. Simulation of the INAR(1)P2S-L process

A simulation study is used to thoroughly investigate the CML method for estimating the parameters of the INAR(1)P2S-L process. Hence, we generated $N = 1000$ samples, each of sizes $n = 20, 50, 100, 500, 1000$, from the INAR(1)P2S-L process for the following four sets of parameter values: ($\theta = 0.2$, $p = 0.3$), ($\theta = 0.25$, $p = 0.5$), ($\theta = 1.2$, $p = 0.7$) and ($\theta = 1.6$, $p = 0.9$). For each n, the average bias and MSE for the parameters were calculated, and the simulation results are presented in Table 5.

Table 5: Simulation results for the INAR(1)P2S-L process.

n	Parameters	$\theta = 0.2,\ p = 0.3$			$\theta = 0.25,\ p = 0.5$		
		Estimates	Bias	MSE	Estimates	Bias	MSE
20	θ	0.219149	0.019149	0.002210	0.265957	0.015957	0.003318
	p	0.337932	0.037932	0.008395	0.510654	0.010654	0.006919
50	θ	0.210206	0.010206	0.000677	0.258585	0.008585	0.001116
	p	0.323072	0.023072	0.003635	0.507034	0.007034	0.002768
100	θ	0.205219	0.005219	0.000241	0.251768	0.001768	0.000489
	p	0.308560	0.008560	0.001226	0.498929	-0.001071	0.001371
500	θ	0.203217	0.003217	0.000062	0.250827	0.000827	0.000094
	p	0.311741	0.011741	0.000414	0.500856	0.000856	0.000272
1000	θ	0.204506	0.004506	0.000043	0.250165	0.000165	0.000045
	p	0.312342	0.012342	0.000288	0.500012	0.000012	0.000136

n	Parameters	$\theta = 1.2,\ p = 0.7$			$\theta = 1.6,\ p = 0.9$		
		Estimates	Bias	MSE	Estimates	Bias	MSE
20	θ	1.326347	0.126347	0.134854	1.751818	0.151818	0.402999
	p	0.724984	0.024984	0.006737	0.897135	-0.002865	0.001089
50	θ	1.289355	0.089355	0.055774	1.647628	0.047628	0.097429
	p	0.716579	0.016579	0.002509	0.897772	-0.002228	0.000349
100	θ	1.282011	0.082011	0.014919	1.625991	0.025991	0.048762
	p	0.711249	0.011249	0.000982	0.899145	-0.000855	0.000189
500	θ	1.243593	0.043593	0.008646	1.600812	0.000812	0.008673
	p	0.705575	0.005575	0.000652	0.899731	-0.000269	0.000034
1000	θ	1.220952	0.020952	0.008867	1.602889	0.002889	0.004258
	p	0.702504	0.002504	0.000261	0.900023	0.000023	0.000018

Table 5 makes it clear that the CML method is suitable for the estimation of the parameters since the bias and MSEs decrease to 0 quickly for small as well as large sample sizes.

6. EMPIRICAL STUDY

In this section, the proposed models under the P2S-L distribution are compared with some discrete models using four real data sets to prove their efficiency. The first two data sets illustrate the performance of the P2S-L distribution. The new P weighted exponential (NPWE) distribution (see Altun, 2020), P xgamma (PX) distribution (see Altun *et al.*, 2022), P Bilal (PBl) distribution (see Altun, 2020), P Lindley (PL) distribution (see Sankaran, 1970) and P distribution are considered for this purpose. The estimates of the involved parameters are computed along with their SEs, CIs (in the form of lower and upper bounds (LCI and UCI, respectively)), goodness of fit statistic (χ^2 statistic), degrees of freedom, p-values, and model adequacy measures such as the minus maximized log likelihood function (-Log L), Akaike information criterion (AIC) and Bayesian information criterion (BIC). The third data set is used to check the efficiency of the P2S-L regression model by comparing it with the new P generalized Lindley (NPGL) regression model (see Altun, 2021), the PX regression model, the PL regression model, and the P regression model by using the measures -Log L, AIC, and BIC. The last data set is used to check the capability of an integer-valued auto-regressive process with P2S-L innovations, the INAR(1)P2S-L process. The INAR(1)P2S-L process is compared with the INAR(1)PTE process (see Altun and Khan, 2022), INAR(1)PX process (see Altun *et al.*, 2022), INAR(1)PL process (see Lívio *et al.*, 2018), INAR(1)P process, and the INAR(1)G process by means of -Log L, AIC, BIC, and fitted measures such as mean, variance, and DI.

6.1. Bacterial clumps

This data set contains measures of the distribution of bacterial clumps per field in a milk film. A microscope slide was split into 400 regions of equal area, and the number of bacterial clumps on each was counted. Bliss and Fisher (1953) considered this data set and fitted it by the NB distribution rather than the P distribution. Table 6 contains some descriptive measures of the fitted P2S-L distribution using this data set. Hence, it is evident that the fitted P2S-L distribution is over-dispersed, right-skewed, and leptokurtic.

Table 6: Values of some descriptive statistics of the P2S-L distribution for the bacterial clumps data set.

Mean	Variance	DI	Skewness	Kurtosis
2.4625	5.7029	2.3159	2.1017	10.7154

The MLEs with their corresponding SEs, CIs under the form (LCI,UCI) for the parameter(s) and goodness-of-fit statistic for the bacterial clumps data set are given in Table 7.

Table 7: The MLE, LCI, UCI, -Log L, AIC, BIC, χ^2 and p-values for the distributions considered using the bacterial clumps data set.

X	Observed frequency	Expected frequency					
		P2S-L	NPWE	PX	PBl	PL	P
0	56	74.13	116.28	105.96	82.04	100.07	34.86
1	104	88.85	82.48	79.90	89.32	83.40	85.07
2	80	76.84	58.50	61.50	73.14	64.29	103.78
3	62	57.42	41.49	46.40	53.38	47.28	84.41
4	42	39.36	29.43	33.95	36.63	33.68	51.49
5	27	25.46	20.88	24.09	24.20	23.44	25.13
6	9	15.79	14.81	16.63	15.58	16.04	10.22
7	9	9.48	10.50	11.21	9.85	10.83	3.56
8	5	5.55	7.45	7.40	6.15	7.24	1.09
9	3	3.18	5.28	4.81	3.80	4.79	0.29
10	2	1.79	3.75	3.08	2.33	3.15	0.07
11	0	0.99	2.66	1.94	1.42	2.06	0.02
12	1	1.17	6.49	3.13	2.17	3.74	0.00
Total	400	400	400	400	400	400	400
θ	MLE	1.1915	0.3724	0.8540	0.3411	0.6523	2.4400
	SE	0.0023	0.2782	0.0018	0.0008	0.0015	0.0039
	LCI	1.1870	0	0.8505	0.3395	0.6493	2.4323
	UCI	1.1960	0.9177	0.8576	0.3427	0.6552	2.4477
α	MLE	—	0.1004	—	—	—	—
	SE	—	0.8219	—	—	—	—
	LCI	—	0	—	—	—	—
	UCI	—	1.7114	—	—	—	—
χ^2		10.9593	73.5180	52.3447	18.3150	43.9591	42.7049
degrees of freedom		8	7	8	8	8	8
p-value		0.2785	< 0.001	< 0.001	0.0317	< 0.001	< 0.001
-Log L		−795.5053	−829.4187	−817.7587	−798.9790	−813.0958	−831.6082
AIC		1593.0106	1660.8373	1637.5174	1599.9580	1628.1917	1665.2165
BIC		1597.0021	1664.8288	1641.5089	1603.9495	1632.1831	1669.2079

According to Table 7, the P2S-L distribution performs well because it has the lowest AIC, BIC, and χ^2 values with a higher p-value. Also, the estimated pmfs of the fitted distributions are diagrammatically represented in Figure 2.

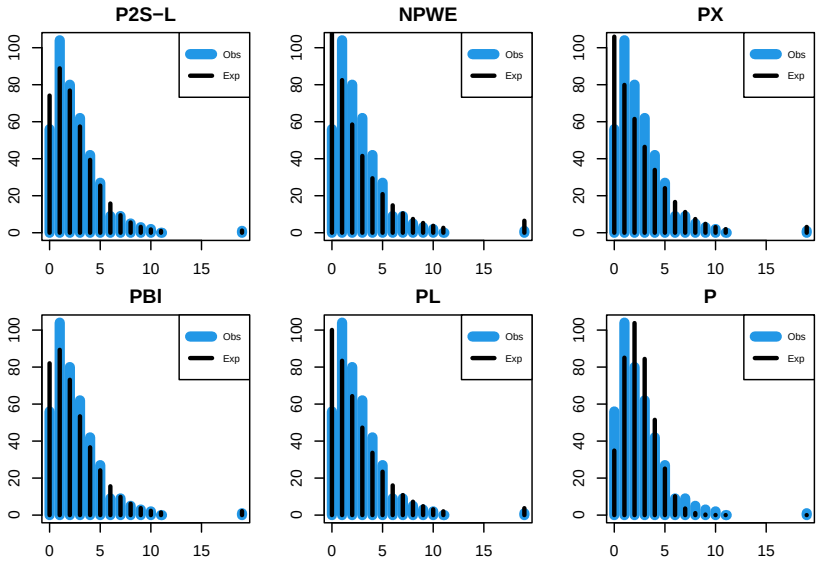

Figure 2: The estimated pmfs of the fitted distributions for the bacterial clumps data set.

From Figure 2 also, we can infer that the P2S-L distribution yields the best fit among other fitted distributions.

6.2. Yeast cells

The second data set contains measurements of yeast cell distribution per square in a hemocytometer (see Bliss and Fisher, 1953). The authors of this reference also used this data set to claim the fit of the NB distribution because it lacks randomness. Table 8 contains some descriptive measures of the fitted P2S-L distribution for this data set. Hence, here also, it is evident that the fitted P2S-L distribution is over-dispersed, right-skewed, and leptokurtic.

Table 8: Values of some descriptive statistics of the P2S-L distribution for the yeast cells data set.

Mean	Variance	DI	Skewness	Kurtosis
0.69	0.91118	1.3205	1.7254	6.7372

Table 9 shows the MLEs with their corresponding SEs, CIs in the form (LCI, UCI) for the parameter(s), and goodness-of-fit statistic for the yeast cells data set.

From Table 9, it is clear that the P2S-L distribution is the best among the considered competitive models since it has the lowest -Log L, AIC, BIC and χ^2 value with the highest p-value.

Table 9: The MLE, LCI, UCI, -Log L, AIC, BIC, χ^2 and p-values for distributions considered using the yeast cells data set.

X	Observed frequency	Expected frequency					
		P2S-L	NPWE	PX	PBl	PL	P
1	213	221.22	237.74	237.14	199.88	234.04	202.14
2	128	114.24	96.44	95.85	116.68	99.41	137.96
3	37	43.56	39.12	39.87	51.43	40.50	47.08
4	18	14.58	15.87	16.41	20.29	16.04	10.71
5	3	4.53	6.44	6.59	7.55	6.22	1.83
6	1	1.86	4.39	4.14	4.17	3.79	0.28
Total	400	400	400	400	400	400	400
θ	MLE	3.5683	1.1235	2.3330	0.9990	1.9502	0.6825
	SE	0.0109	2.2302	0.0072	0.0031	0.0064	0.0021
	LCI	3.5470	0	2.3189	0.9930	1.9377	0.6785
	UCI	3.5897	5.4947	2.3470	1.0050	1.9627	0.6865
α	MLE	-	0.3041	-	-	-	-
	SE	-	2.5888	-	-	-	-
	LCI	-	0	-	-	-	-
	UCI	-	5.3782	-	-	-	-
χ^2		3.0015	13.8457	14.4230	9.1358	11.0452	10.0362
degrees of freedom		2	1	2	2	2	2
p-value		0.3914	0.0031	0.0024	0.0275	0.0115	< 0.001
-Log L		-447.3560	-454.4330	-454.3928	-451.9304	-452.6185	-449.5038
AIC		896.7120	910.8660	910.7857	905.8609	907.2370	901.0076
BIC		900.7035	914.8574	914.7771	909.8524	911.2285	904.9990

From Figure 3, we can infer that the P2S-L distribution yields the best fit among other fitted distributions.

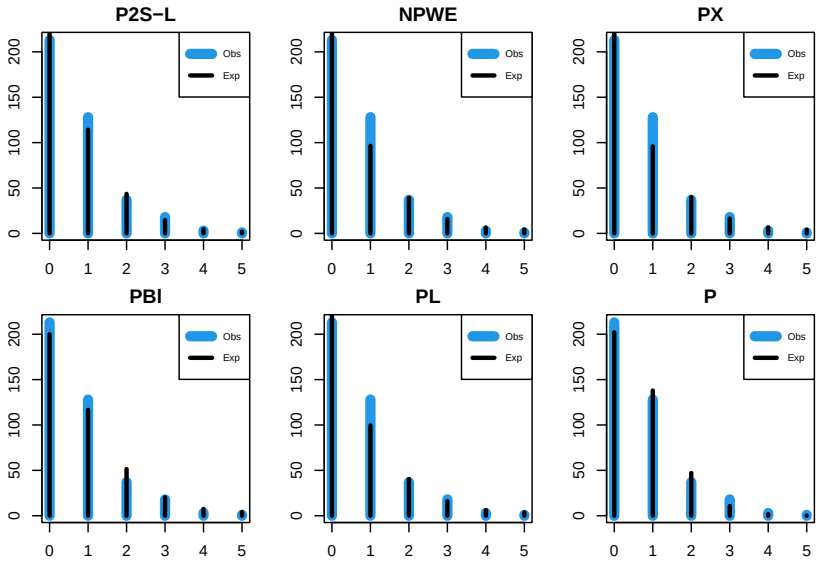

Figure 3: The estimated pmfs of the fitted distributions for the yeast cells data set.

6.3. Length of hospital stay

The third data set is utilized to examine the proficiency of the count regression model proposed under the P2S-L distribution. The data set consists of 3589 observations from the files of 1991 Arizona cardiovascular patients found in the COUNT package in R software. The P2S-L regression model is used to model the length of stay (y_i) by using the covariates: cardiovascular procedure (x_{1i}) (1 = CABG, 0 = PTCA), sex (x_{2i}) (1 = male, 0 = female), type of admission (x_{3i}) (1 = urgent, 0 = elective) and age (x_{4i}) (1 = age > 75, 0 = age ≤ 75). Given below is the regression structure that will be fitted by the P2S-L, NPGL, PX, PL, and P regression models:

$$\mu_i = e^{\gamma_0+\gamma_1 x_{1i}+\gamma_2 x_{2i}+\gamma_3 x_{3i}+\gamma_4 x_{4i}}.$$

The mean and variance of the dependent variable are calculated as 8.831 and 47.973, respectively, stating the clear over-dispersion. Table 10 gives the parameter estimates and results of the information criteria.

Table 10: The MLE, -Log L, AIC and BIC of the fitted regression models for the length of stay data set.

Covariates	P		PL		PX		NPGL		P2SL	
	Estimate SE	p-value	Estimate SE	p-value	Estimate SE	p- value	Estimate SE	p-value	Estimate SE	p-value
γ_0	1.4560 0.0158	< 0.001	1.4133 0.0372	< 0.001	1.3996 0.0349	< 0.001	1.4044 0.0353	< 0.001	1.4159 0.0298	< 0.001
γ_1	0.9606 0.0122	< 0.001	0.9843 0.0291	< 0.001	0.9725 0.0270	< 0.001	0.9761 0.0274	< 0.001	0.9835 0.0231	< 0.001
γ_2	−0.1240 0.0118	< 0.001	−0.1288 0.0304	< 0.001	−0.1269 0.0280	< 0.001	−0.1267 0.0285	< 0.001	−0.1262 0.0240	< 0.001
γ_3	0.3266 0.0121	< 0.001	0.3843 0.0302	< 0.001	0.3732 0.0280	< 0.001	0.3759 0.0284	< 0.001	0.3746 0.0240	< 0.001
γ_4	0.1224 0.0124	< 0.001	0.1193 0.0323	< 0.001	0.1202 0.0298	< 0.001	0.1198 0.0303	< 0.001	0.1197 0.0255	< 0.001
-Log L	−11189.8976		−10625.5957		−10569.8162		−10563.2551		−10164.8969	
AIC	22389.7952		21239.1913		21127.6324		21114.5102		20317.7938	
BIC	22420.7233		21202.0775		21090.5187		21077.3964		20280.6801	

Altun (2021) used this data set to prove the better fit of the NPGL regression model. Hence, the obtained -Log L, AIC and BIC were better than those of the NB regression model. From Table 10, it is clear that the P2S-L regression model is also better than the NB regression model. Now, since the P2S-L regression model has minimized values for its -Log L, AIC, and BIC, we conclude that it will be a more appropriate model than the other models for modelling this data set. As a result, we can say that the length of hospital stay increases when people have CABG cardiovascular surgery, are admitted urgently, and are over the age of 75. Additionally, female individuals have a longer hospital stay than male individuals.

6.4. Number of weakly sales

The fourth data set that we consider is the series of weekly sales (in integer units) of a particular soap product in a supermarket. The information comes from the Kilts Center for Marketing, Graduate School of Business, University of Chicago, and can be found at: `http://gbswww.uchicago.edu/kilts/research/db/dominicks`. (The product is 'Zest White Water 15 oz.', with code 3700031165). We consider it to prove the applicability of the P2S-L distribution as an innovation distribution for the INAR(1) process. The one step transition probabilities of the competitive INAR(1) models used here for comparison are given in Appendix A.

The mean, variance, and DI of the data set are 242, 5.44, and 15.40, respectively. The test proposed in Schweer and Wei (2014) for determining statistically significant overdispersion shows a p-value less than 0.001, indicating the data possess significant overdispersion. The autocorrelation function (ACF), partial ACF (PACF), histogram, and time series plots are displayed in Figure 4 and in the PACF plot, the only first-lag significance proves that this data set can be used for modelling the INAR(1) process.

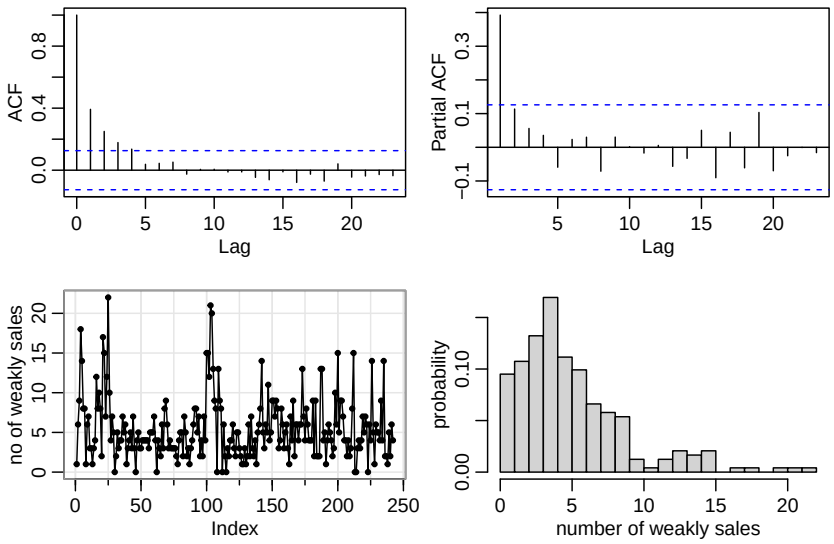

Figure 4: ACF, PACF, time series and histogram plots for the numbers of weekly sales data.

The fitted INAR(1) processes with the P2S-L innovation and other corresponding innovations used for comparison yield the parameter estimates along with SE, AIC, BIC, theoretical mean, variance, and DI as given in Table 11. The minimum values of AIC and BIC statistics for the INAR(1)P2S-L process prove that it provides a better fit than the INAR(1)PTE, INAR(1)PX, INAR(1)PL, INAR(1)P and INAR(1)G processes. Furthermore, the theoretical DI value of the INAR(1)P2S-L process is close to the empirical DI value. As a result, it is convincing that the INAR(1)P2S-L process effectively explains the data set's characteristics.

Table 11: The estimates and modelling adequacy statistics of the fitted distributions for the number of weekly sales data set.

Model	Parameters	Estimates	SE	AIC	BIC	μ_*	σ_*^2	DI_*
INAR(1)P2S-L	θ p	0.7569 0.2407	0.0485 0.0426	1244.4780	1251.4559	5.3926	12.6378	2.3436
INAR(1)PTE	α θ p	−0.9999 0.3839 0.2860	0.2946 0.0353 0.0516	1246.4137	1256.8805	1.6240	3.1983	1.9694
INAR(1)PX	θ p	0.6086 0.3241	0.0397 0.0369	1250.5878	1257.5656	3.1074	7.1419	2.2983
INAR(1)PL	θ p	0.4533 0.3202	0.0325 0.0369	1249.0689	1256.0467	5.4545	15.2780	2.8010
INAR(1)P	θ p	4.1853 0.2341	0.2110 0.0324	1363.8119	1370.7898	5.5041	5.4045	0.9819
INAR(1)G	θ p	0.2253 0.3713	0.0152 0.0324	1260.3368	1267.3147	0.3347	0.3315	0.9904
Empirical						5.4421	15.3376	2.8183

A residual analysis is done to check whether the fitted INAR(1)P2S-L process is statistically accurate. For that, the Pearson residuals $r_1, ..., r_T$ are calculated, where, for any $t = 1, 2, ..., n$,

$$r_t = \frac{x_t - \mathrm{E}(X_t \,|\, X_{t-1} = x_{t-1})}{\mathrm{Var}(X_t \,|\, X_{t-1} = x_{t-1})^{1/2}},$$

where $\mathrm{E}(X_t \,|\, X_{t-1})$ and $\mathrm{Var}(X_t \,|\, X_{t-1})$ are given in (5.2) and (5.3), respectively. The statistical validity of the fitted INAR(1) process is proved by acquiring a zero mean and unit variance for the uncorrelated Pearson residuals (see Harvey and Fernandes, 1989). The mean and variance of the Pearson residuals of the INAR(1)P2S-L process were 0.0419 and 1.0033, respectively, which are very close to the desired values. This proves our INAR(1)P2S-L process is statistically valid for the data set. According to the findings in Jazi *et al.* (2012), the INAR(1)P2S-L process for the data is such that

$$X_t = 0.2407 \circ X_{t-1} + \varepsilon_t,$$

where the innovation process satisfies $\varepsilon_t \sim$ P2S-L(0.7569).

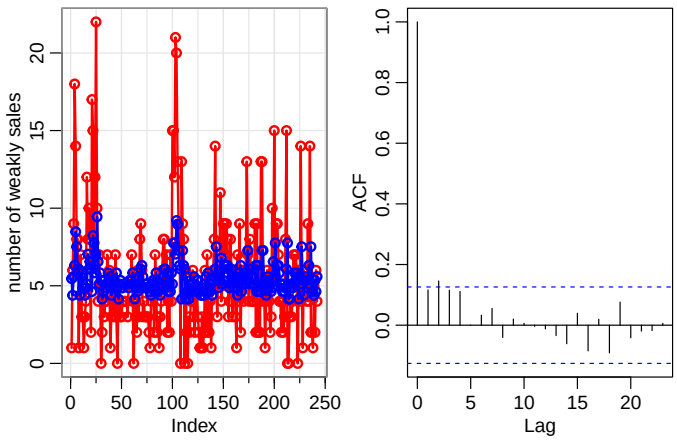

Figure 5: The predicted values of the number of weekly sales data set (left) and the ACF plot of the Pearson residuals (right).

Additionally, the ACF plot of the Pearson residuals in Figure 5 specifies that there is no presence of autocorrelation for the Pearson residuals. The plot for actual and predicted values of the monthly number of claims is displayed on the left side of Figure 5.

7. CONCLUDING REMARKS

A one-parameter discrete compound distribution, namely the Poisson 2S-Lindley (P2S-L) distribution, was proposed. It was shown that it can effectively model over-dispersed data sets. Flexibility is attained by having a closed form for almost every statistical and mathematical property. Methods such as the MM, ML, LS, and WLS methods were used for estimating the unknown parameter and were compared using a broad simulation study. More importantly, a new count regression model and an INAR(1) process based on the P2S-L distribution were introduced and described in detail. Four real-life data sets were evaluated to determine the applicability of the proposed model. Thus, the P2S-L distribution-based models will be effective in modelling count data, particularly over-dispersed count data sets.

A. APPENDIX

Transition probabilities for the processes used for comparison in Section 6.4.

1. Transition probability for the INAR(1)PTE process:

$$\Pr\big(X_t = k \,|\, X_{t-1} = l\big) \;=\; \sum_{i=0}^{\min(k,l)} \binom{l}{i} p^i (1-p)^{l-i} \theta \bigg(\frac{1-\alpha}{(1+\theta)^{1+k-i}} + \frac{2\alpha}{(1+2\theta)^{1+k-i}} \bigg),$$

with $|\alpha| \leq 1$ and $\theta > 0$.

2. Transition probability for the INAR(1)PX process:

$$\Pr\big(X_t = k \,|\, X_{t-1} = l\big) \;=\; \sum_{i=0}^{\min(k,l)} \binom{l}{i} p^i (1-p)^{l-i} \times \frac{\theta^2 \big[2(1+\theta)^2 + \theta(k-i+2)(k-i+1)\big]}{2(1+\theta)^{k-i+4}},$$

with $\theta > 0$.

3. Transition probability for the INAR(1)PL process:

$$\Pr\big(X_t = k \,|\, X_{t-1} = l\big) \;=\; \sum_{i=0}^{\min(k,l)} \binom{l}{i} p^i (1-p)^{l-i} \frac{\theta^2 (k-i+\theta+2)}{(\theta+1)^{k-i+3}},$$

with $\theta > 0$.

4. Transition probability for the INAR(1)P process:

$$\Pr\big(X_t = k \,|\, X_{t-1} = l\big) \;=\; \sum_{i=0}^{\min(k,l)} \binom{l}{i} p^i (1-p)^{l-i} \frac{e^{-\theta} \theta^{k-i}}{(k-i)!},$$

with $\theta > 0$.

5. Transition probability for the INAR(1)G process:

$$\Pr\big(X_t = k \,|\, X_{t-1} = l\big) \;=\; \sum_{i=0}^{\min(k,l)} \binom{l}{i} p^i (1-p)^{l-i} \theta (1-\theta)^{k-i},$$

with $\theta \in (0,1)$.

ACKNOWLEDGMENTS

We acknowledge the valuable suggestions from the learned referees.

REFERENCES

Abd El-Monsef, M. M. E. and Sohsah, N. M. (2016). Poisson–transmuted Lindley distribution. *Journal Of Advances In Mathematics*, 11(9):5631–5638.

Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, 8(3):261–275.

Altun, E. (2020). A new generalization of geometric distribution with properties and applications. *Communications in Statistics – Simulation and Computation*, 49(3):793–807.

Altun, E. (2020). A new one-parameter discrete distribution with associated regression and integer-valued autoregressive models. *Mathematica Slovaca*, 70(4):979–994.

Altun, E. (2021). A new two-parameter discrete Poisson-generalized Lindley distribution with properties and applications to healthcare data sets. *Computational Statistics*, 36:2841–2861.

Altun, E. and Khan, N. M. (2022). Modelling with the Novel INAR(1) PTE process. *Methodology and Computing in Applied Probability*, 24:1735–1751.

Altun, E., Bhati, D., and Khan, N. M. (2021). A new approach to model the counts of earthquakes: INARPQX(1) process. *SN Applied Sciences*, 3(2):1–17.

Altun, E., Cordeiro, G. M., and Ristić, M. M. (2022). An one-parameter compounding discrete distribution. *Journal of Applied Statistics*, 49(8):1935–1956

Bereta, E. M. P., Louzanda, F., and Franco, M. A. P. (2011). The Poisson–Weibull distribution. *Advances and Applications in Statistics*, 22(2):107–118.

Bhati, D., Kumawat, P., and Gómez-Déniz, E. (2017). A new count model generated from mixed Poisson transmuted exponential family with an application to health care data. *Communications in Statistics – Theory and Methods*, 46(22):11060–11076.

Bliss, C. I. and Fisher, R. A. (1953). Fitting the negative binomial distribution to biological data. *Biometrics*, 9(2):176–200.

Cheng, L., Geedipally, S. R., and Lord, D. (2013). The Poisson–Weibull generalized linear model for analyzing motor vehicle crash data. *Safety Science*, 54):38–42.

Chesneau, C., Tomy, L., and Gillariose, J. (2020). On a sum and difference of two Lindley distributions: theory and applications. *REVSTAT – Statistical Journal*, 18(5):673–696.

Eliwa, M. S., Altun, E., El-Dawoody, M., and El-Morshedy, M. (2020). A new three-parameter discrete distribution with associated INAR(1) process and applications. *IEEE Access*, 8:91150–91162.

Grine, R. and Zeghdoudi, H. (2017). On Poisson quasi-Lindley distribution and its applications. *Journal of Modern Applied Statistical Methods*, 16(2):21.

Harvey, A. C. and Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business & Economic Statistics*, 7(4):407–417.

Huang, J. and Zhu, F. (2021). A new first-order integer-valued autoregressive model with Bell innovations. *Entropy*, 23(6):713.

Irshad, M. R., Chesneau, C., D'cruz, V., and Maya, R. (2021). Discrete pseudo Lindley distribution: properties, estimation and application on INAR(1) process. *Mathematical and Computational Applications*, 26(4):76.

Jazi, M. A., Jones, G., and Lai, C. D. (2012). First-order integer valued AR processes with zero inflated Poisson innovations. *Journal of Time Series Analysis*, 33(6):954–963.

Karlis, D. and Xekalaki, E. (2005). Mixed Poisson distributions. *International Statistical Review/ Revue Internationale de Statistique*, 73(1):35–58.

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

Lívio, T., Khan, N. M., Bourguignon, M., and Bakouch, H. S. (2018). An INAR(1) model with Poisson–Lindley innovations. *Economics Bulletin*, 38(3):1505–1513.

Lord, D. and Geedipally, S. R. (2011). The negative binomial–Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros. *Accident Analysis & Prevention*, 43(5):1738–1742.

Mahmoudi, E. and Zakerzadeh, H. (2010). Generalized Poisson–Lindley distribution. *Communications in Statistics – Theory and Methods*, 39(10):1785–1798.

McKenzie, E. (1985). Some simple models for discrete variate time series 1. *JAWRA Journal of the American Water Resources Association*, 21(4):645–650.

Rodríguez-Avi, J., Conde-Sánchez, A., Sáez-Castillo, A. J., Olmo-Jiménez, M. J., and Martínez-Rodríguez, A. M. (2009). A generalized Waring regression model for count data. *Computational statistics & data analysis*, 53(10):3717–3725.

Sáez-Castillo, A. J. and Conde-Sánchez, A. (2013). A hyper-Poisson regression model for overdispersed and underdispersed count data. *Computational Statistics & Data Analysis*, 61:148–157.

Sankaran, M. (1970). The discrete Poisson–Lindley distribution. *Biometrics*, 26(1):145–149.

Schweer, S. and Wei, C. H. (2014). Compound Poisson INAR(1) processes: stochastic properties and testing for overdispersion. *Computational Statistics & Data Analysis*, 77:267–284.

Wei, C. H. (2018). *An Introduction to Discrete-Valued Time Series*. John Wiley & Sons.

Wongrin, W. and Bodhisuwan, W. (2017). Generalized Poisson–Lindley linear model for count data. *Journal of Applied Statistics*, 44(15):2659–2671.

Zamani, H., Faroughi, P., and Ismail, N. (2014). Bivariate Poisson-weighted exponential distribution with applications. *AIP Conference Proceedings*, 1602(1):964–968.

REVSTAT – Statistical Journal
Volume 23, Number 3, July 2025, 397–422
https://doi.org/10.57805/revstat.v23i3.634

Information Matrices for Composite Type Distributions

Authors: Jiahang Lyu
– Department of Mathematics, University of Manchester
UK
jiahang.lyu@manchester.ac.uk

Saralees Nadarajah ✉
– Department of Mathematics, University of Manchester
UK
mbbsssn2@manchester.ac.uk

Received: August 2023 Revised: September 2023 Accepted: September 2023

Abstract:

- Composite type distributions are being increasingly used to model insurance data. Yet no expressions seem to be available for observed information matrices. In this paper, we give expressions for the matrices for two-piece, three-piece and m-piece composite distributions in their most general forms. Expressions for a number of particular cases and a simulation study showing practical use are also given.

Keywords:

- *m-piece model; three-piece model; two-piece model.*

AMS Subject Classification:

- Primary: 62E15.

✉ Corresponding author

1. INTRODUCTION

Composite type distributions have attracted much attention for modeling insurance data lately. Nearly all of the proposed models are made up of two pieces. The most general two-piece composite distribution has its probability density and cumulative distribution functions specified by

$$f(x) = \begin{cases} \dfrac{1}{1+\Phi}\dfrac{f_1(x)}{F_1(\theta)}, & x \le \theta, \\ \dfrac{\Phi}{1+\Phi}\dfrac{f_2(x)}{1-F_2(\theta)}, & \theta < x, \end{cases} \tag{1.1}$$

and

$$F(x) = \begin{cases} \dfrac{1}{1+\Phi}\dfrac{f_1(x)}{F_1(\theta)}, & x < \theta, \\ \dfrac{\Phi}{1+\Phi}\left[1+\Phi\dfrac{F_2(x)-F_2(\theta)}{1-F_2(\theta)}\right], & \theta < x, \end{cases} \tag{1.2}$$

respectively, where f_1, f_2 are smooth and valid probability density functions, F_1, F_2 are the corresponding cumulative distribution functions and $\Phi > 0$.

Various particular cases of (1.1) and (1.2) have been studied in the literature: Cooray and Ananda (2005), Scollnik (2007), Teodorescu (2010), Lee (2011), Pigeon and Denuit (2011), Eliazar and Cohen (2012), Pak (2013), Bee (2015), Cooray and Cheng (2015), Luckstead and Devadoss (2017), Luckstead *et al.* (2017) and Mutali and Vernic (2020) considered lognormal and Pareto distributions for f_1 and f_2, respectively; Teodorescu and Vernic (2009), Aminzadeh and Deng (2018) and Gencturk *et al.* (2018) considered exponential and Pareto distributions for f_1 and f_2, respectively; Cooray (2009), Teodorescu and Panaitescu (2009), Scollnik and Sun (2012), Calderin-Ojeda (2018) and Deng and Aminzadeh (2019) considered Weibull and Pareto distributions for f_1 and f_2, respectively; Cooray *et al.* (2010) considered Weibull and inverse Weibull distributions for f_1 and f_2, respectively; Teodorescu and Vernic (2013) considered the Pareto distribution for f_2; Nadarajah and Bakar (2013, 2014) considered the lognormal distribution for f_1; Maghsoudi *et al.* (2014) considered Weibull and inverse transformed gamma distributions for f_1 and f_2, respectively; Calderin-Ojeda (2015) considered Weibull and Burr distributions for f_1 and f_2, respectively; Calderin-Ojeda and Kwok (2016) considered the Stoppa distribution for f_1; Kim *et al.* (2016) and Park and Baek (2018) considered lognormal and generalized Pareto distributions for f_1 and f_2, respectively; Maghsoudi and Abu Bakar (2017) considered transformed gamma and inverse transformed gamma distributions for f_1 and f_2, respectively; Aminzadeh and Deng (2019) considered inverse gamma and Pareto distributions for f_1 and f_2, respectively; Kim *et al.* (2019) considered gamma and generalized Pareto distributions for f_1 and f_2, respectively; Benatmane *et al.* (2020) considered Rayleigh and Pareto distributions for f_1 and f_2, respectively; and so on. Liu and Ananda (2022, 2023) studied an exponentiated version of (1.1) and (1.2). Grün and Miljkovic (2019) provided a comprehensive analysis on two-piece composite distributions. Liu and Ananda (2023) proposed an exponentiated version of inverse gamma-Pareto composite distribution.

Yet we are aware of no work giving the observed information matrix for the general form of composite distributions. Even the particular cases mentioned in the cited papers do not appear to have derived the observed information matrices. Observed information matrices are important for tests of hypothesis and interval estimation.

Observed information matrices are commonly computed numerically. This can be prone to errors especially for composite type distributions because they involve continuity and differentiability conditions at θ. The aim of this paper is to derive explicit expressions for the observed information matrices for the most general two-piece, three-piece and m-piece composite distributions. These expressions are given in Sections 2, 3 and 4. Some particular cases of these expressions when f_1 is taken to correspond to exponential and Weibull distributions are given in the supplementary file. A simulation study showing practical use of the expressions in Section 4 is described in Section 5.

m-piece composite distributions when $m > 2$ can be useful models for portfolio losses when the portfolio consists of more than two assets behaving differently. Examples could include portfolios consisting of losses relating to food, energy and metals.

All of the differentiation was performed manually. The correctness of differentiation was verified numerically. Mathematica code implementing the expressions in Sections 2 to 4 can be obtained from the corresponding author.

2. GENERAL TWO-PIECE COMPOSITE DISTRIBUTION

Suppose that the body and tail are described probability density functions f_1 and f_2, respectively. Then the probability density and cumulative distribution functions of the most general two-piece composite distribution are (1.1) and (1.2), respectively. Suppose f_1 and f_2 are parameterized by $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$, respectively. Conditions for continuity and differentiability at θ ensure that $\theta = \theta(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)$ and $\Phi = \Phi(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) = \frac{f_1(\theta)[1-F_2(\theta)]}{F_1(\theta)f_2(\theta)}$. If $x_1, x_2, ..., x_n$ is a random sample from (1.1)–(1.2) then the log-likelihood function of $(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)$ is

$$\begin{aligned}\log L(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2) \;=\; & \sum_{x_i \le \theta} \log f_1(x_i) + \sum_{\theta < x_i} \log f_2(x_i) - a[\log(1+\Phi) + \log F_1(\theta)] \\ & + (n-a)\{\log \Phi - \log(1+\Phi) - \log[1 - F_2(\theta)]\},\end{aligned}$$

where

$$a = \sum_{i=1}^{n} I\{x_i \le \theta\}.$$

Given the facts that

$$\frac{\partial}{\partial \theta} I\{x \ge \theta\} = \Delta(\theta)$$

and

$$\frac{\partial}{\partial \theta} I\{x < \theta\} = -\Delta(\theta)$$

(see https://en.wikipedia.org/wiki/Indicator_function), where $\Delta(\cdot)$ denotes the Dirac-delta function, we obtain the first order partial derivatives as

$$\begin{aligned}\frac{\partial \log L}{\partial \boldsymbol{\Omega}_1} &= \sum_{x_i \le \theta} \frac{\frac{\partial f_1(x_i)}{\partial \boldsymbol{\Omega}_1}}{f_1(x_i)} + \frac{\partial \theta}{\partial \boldsymbol{\Omega}_1} \sum_{i=1}^{n} \Delta(x_i - \theta)[\log f_1(x_i) - \log f_2(x_i)] \\ &\quad - \frac{\partial a}{\partial \boldsymbol{\Omega}_1}[\log(1+\Phi) + \log F_1(\theta)] \\ &\quad - \frac{\partial a}{\partial \boldsymbol{\Omega}_1}\{\log \Phi - \log(1+\Phi) - \log[1 - F_2(\theta)]\} \\ &\quad - a\left[\frac{\frac{\partial \Phi}{\partial \boldsymbol{\Omega}_1}}{1+\Phi} + \frac{f_1(\theta)}{F_1(\theta)}\frac{\partial \theta}{\partial \boldsymbol{\Omega}_1}\right] + (n-a)\left[\left(\frac{1}{\Phi} - \frac{1}{1+\Phi}\right)\frac{\partial \Phi}{\partial \boldsymbol{\Omega}_1} + \frac{f_2(\theta)}{1 - F_2(\theta)}\frac{\partial \theta}{\partial \boldsymbol{\Omega}_1}\right]\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \log L}{\partial \boldsymbol{\Omega}_2} &= \sum_{\theta < x_i} \frac{\frac{\partial f_2(x_i)}{\partial \boldsymbol{\Omega}_2}}{f_2(x_i)} + \frac{\partial \theta}{\partial \boldsymbol{\Omega}_2} \sum_{i=1}^{n} \Delta(x_i - \theta)[\log f_1(x_i) - \log f_2(x_i)] \\ &\quad - \frac{\partial a}{\partial \boldsymbol{\Omega}_2}[\log(1+\Phi) + \log F_1(\theta)] \\ &\quad - \frac{\partial a}{\partial \boldsymbol{\Omega}_2}\{\log \Phi - \log(1+\Phi) - \log[1 - F_2(\theta)]\} \\ &\quad - a\left[\frac{\frac{\partial \Phi}{\partial \boldsymbol{\Omega}_2}}{1+\Phi} + \frac{f_1(\theta)}{F_1(\theta)}\frac{\partial \theta}{\partial \boldsymbol{\Omega}_2}\right] + (n-a)\left[\left(\frac{1}{\Phi} - \frac{1}{1+\Phi}\right)\frac{\partial \Phi}{\partial \boldsymbol{\Omega}_2} + \frac{f_2(\theta)}{1 - F_2(\theta)}\frac{\partial \theta}{\partial \boldsymbol{\Omega}_2}\right].\end{aligned}$$

Hence, the observed information matrix is

$$\mathbf{J} = -\begin{pmatrix} \dfrac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1^2} & \dfrac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} \\ \dfrac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} & \dfrac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_2^2} \end{pmatrix}, \tag{2.1}$$

where

$$\begin{aligned}\frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1^2} &= \sum_{x_i \le \theta} \frac{\frac{\partial^2 f_1(x_i)}{\partial \boldsymbol{\Omega}_1^2} f_1(x_i) - \left[\frac{\partial f_1(x_i)}{\partial \boldsymbol{\Omega}_1}\right]^2}{[f_1(x_i)]^2} + 2\frac{\partial \theta}{\partial \boldsymbol{\Omega}_1} \sum_{i=1}^{n} \Delta(x_i - \theta)\frac{\frac{\partial f_1(x_i)}{\partial \boldsymbol{\Omega}_1}}{f_1(x_i)} \\ &\quad + \frac{\partial^2 \theta}{\partial \boldsymbol{\Omega}_1^2} \sum_{i=1}^{n} \Delta(x_i - \theta)[\log f_1(x_i) - \log f_2(x_i)] \\ &\quad - \frac{\partial^2 a}{\partial \boldsymbol{\Omega}_1^2}[\log(1+\Phi) + \log F_1(\theta)] - \frac{\partial^2 a}{\partial \boldsymbol{\Omega}_1^2}\{\log \Phi - \log(1+\Phi) - \log[1 - F_2(\theta)]\} \\ &\quad - 2\frac{\partial a}{\partial \boldsymbol{\Omega}_1}\left[\frac{\frac{\partial \Phi}{\partial \boldsymbol{\Omega}_1}}{1+\Phi} + \frac{f_1(\theta)}{F_1(\theta)}\frac{\partial \theta}{\partial \boldsymbol{\Omega}_1}\right] - 2\frac{\partial a}{\partial \boldsymbol{\Omega}_1}\left[\left(\frac{1}{\Phi} - \frac{1}{1+\Phi}\right)\frac{\partial \Phi}{\partial \boldsymbol{\Omega}_1} + \frac{f_2(\theta)}{1 - F_2(\theta)}\frac{\partial \theta}{\partial \boldsymbol{\Omega}_1}\right] \\ &\quad - a\frac{\frac{\partial^2 \Phi}{\partial \boldsymbol{\Omega}_1^2}(1+\Phi) - \left(\frac{\partial \Phi}{\partial \boldsymbol{\Omega}_1}\right)^2}{(1+\Phi)^2} - a\frac{\frac{\partial f_1(\theta)}{\partial \boldsymbol{\Omega}_1}F_1(\theta) - \frac{\partial \theta}{\partial \boldsymbol{\Omega}_1}[f_1(\theta)]^2}{[F_1(\theta)]^2}\frac{\partial \theta}{\partial \boldsymbol{\Omega}_1} - a\frac{f_1(\theta)}{F_1(\theta)}\frac{\partial^2 \theta}{\partial \boldsymbol{\Omega}_1^2} \\ &\quad + (n-a)\left[\frac{1}{(1+\Phi)^2} - \frac{1}{\Phi^2}\right]\left(\frac{\partial \Phi}{\partial \boldsymbol{\Omega}_1}\right)^2 + (n-a)\left(\frac{1}{\Phi} - \frac{1}{1+\Phi}\right)\frac{\partial^2 \Phi}{\partial \boldsymbol{\Omega}_1^2} \\ &\quad + (n-a)\frac{\frac{\partial f_2(\theta)}{\partial \boldsymbol{\Omega}_1}[1 - F_2(\theta)] + [f_2(\theta)]^2\frac{\partial \theta}{\partial \boldsymbol{\Omega}_1}}{[1 - F_2(\theta)]^2}\frac{\partial \theta}{\partial \boldsymbol{\Omega}_1} + (n-a)\frac{f_2(\theta)}{1 - F_2(\theta)}\frac{\partial^2 \theta}{\partial \boldsymbol{\Omega}_1^2},\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} &= \frac{\partial \theta}{\partial \boldsymbol{\Omega}_2} \sum_{i=1}^{n} \Delta(x_i - \theta) \frac{\frac{\partial f_1(x_i)}{\partial \boldsymbol{\Omega}_1}}{f_1(x_i)} + \frac{\partial^2 \theta}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} \sum_{i=1}^{n} \Delta(x_i - \theta)[\log f_1(x_i) - \log f_2(x_i)] \\
&\quad - \frac{\partial \theta}{\partial \boldsymbol{\Omega}_1} \sum_{i=1}^{n} \Delta(x_i - \theta) \frac{\frac{\partial f_2(x_i)}{\partial \boldsymbol{\Omega}_2}}{f_2(x_i)} - \frac{\partial^2 a}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} [\log(1 + \Phi) + \log F_1(\theta)] \\
&\quad - \frac{\partial^2 a}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} \{\log \Phi - \log(1 + \Phi) - \log[1 - F_2(\theta)]\} - \frac{\partial a}{\partial \boldsymbol{\Omega}_2} \left[\frac{\frac{\partial \Phi}{\partial \boldsymbol{\Omega}_1}}{1 + \Phi} + \frac{f_1(\theta)}{F_1(\theta)} \frac{\partial \theta}{\partial \boldsymbol{\Omega}_1} \right] \\
&\quad - \frac{\partial a}{\partial \boldsymbol{\Omega}_1} \left[\frac{\frac{\partial \Phi}{\partial \boldsymbol{\Omega}_2}}{1 + \Phi} + \frac{f_1(\theta)}{F_1(\theta)} \frac{\partial \theta}{\partial \boldsymbol{\Omega}_2} \right] - \frac{\partial a}{\partial \boldsymbol{\Omega}_1} \left[\left(\frac{1}{\Phi} - \frac{1}{1 + \Phi} \right) \frac{\partial \Phi}{\partial \boldsymbol{\Omega}_2} + \frac{f_2(\theta)}{1 - F_2(\theta)} \frac{\partial \theta}{\partial \boldsymbol{\Omega}_2} \right] \\
&\quad - \frac{\partial a}{\partial \boldsymbol{\Omega}_2} \left[\left(\frac{1}{\Phi} - \frac{1}{1 + \Phi} \right) \frac{\partial \Phi}{\partial \boldsymbol{\Omega}_1} + \frac{f_2(\theta)}{1 - F_2(\theta)} \frac{\partial \theta}{\partial \boldsymbol{\Omega}_1} \right] \\
&\quad - a \frac{\frac{\partial^2 \Phi}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2}(1 + \Phi) - \frac{\partial \Phi}{\partial \boldsymbol{\Omega}_1} \frac{\partial \Phi}{\partial \boldsymbol{\Omega}_2}}{(1 + \Phi)^2} - a \frac{\frac{\partial f_1(\theta)}{\partial \boldsymbol{\Omega}_2} F_1(\theta) - [f_1(\theta)]^2 \frac{\partial \theta}{\partial \boldsymbol{\Omega}_2}}{[F_1(\theta)]^2} \frac{\partial \theta}{\partial \boldsymbol{\Omega}_1} \\
&\quad - a \frac{f_1(\theta)}{F_1(\theta)} \frac{\partial^2 \theta}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} \\
&\quad + (n - a) \left[\frac{1}{(1 + \Phi)^2} - \frac{1}{\Phi^2} \right] \frac{\partial \Phi}{\partial \boldsymbol{\Omega}_1} \frac{\partial \Phi}{\partial \boldsymbol{\Omega}_2} + (n - a) \left(\frac{1}{\Phi} - \frac{1}{1 + \Phi} \right) \frac{\partial^2 \Phi}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} \\
&\quad + (n - a) \frac{\partial \theta}{\partial \boldsymbol{\Omega}_1} \frac{\frac{\partial f_2(\theta)}{\partial \boldsymbol{\Omega}_2}[1 - F_2(\theta)] + [f_2(\theta)]^2 \frac{\partial \theta}{\partial \boldsymbol{\Omega}_2}}{[1 - F_2(\theta)]^2} + \frac{f_2(\theta)}{1 - F_2(\theta)} \frac{\partial^2 \theta}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_2^2} &= \sum_{x_i > \theta} \frac{\frac{\partial^2 f_2(x_i)}{\partial \boldsymbol{\Omega}_2^2} f_2(x_i) - \left[\frac{\partial f_2(x_i)}{\partial \boldsymbol{\Omega}_2} \right]^2}{[f_2(x_i)]^2} + 2 \frac{\partial \theta}{\partial \boldsymbol{\Omega}_2} \sum_{i=1}^{n} \Delta(x_i - \theta) \frac{\frac{\partial f_2(x_i)}{\partial \boldsymbol{\Omega}_2}}{f_2(x_i)} \\
&\quad + \frac{\partial^2 \theta}{\partial \boldsymbol{\Omega}_2^2} \sum_{i=1}^{n} \Delta(x_i - \theta)[\log f_1(x_i) - \log f_2(x_i)] \\
&\quad - \frac{\partial^2 a}{\partial \boldsymbol{\Omega}_2^2} [\log(1 + \Phi) + \log F_1(\theta)] - \frac{\partial^2 a}{\partial \boldsymbol{\Omega}_2^2} \{\log \Phi - \log(1 + \Phi) - \log[1 - F_2(\theta)]\} \\
&\quad - 2 \frac{\partial a}{\partial \boldsymbol{\Omega}_2} \left[\frac{\frac{\partial \Phi}{\partial \boldsymbol{\Omega}_2}}{1 + \Phi} + \frac{f_1(\theta)}{F_1(\theta)} \frac{\partial \theta}{\partial \boldsymbol{\Omega}_2} \right] - 2 \frac{\partial a}{\partial \boldsymbol{\Omega}_2} \left[\left(\frac{1}{\Phi} - \frac{1}{1 + \Phi} \right) \frac{\partial \Phi}{\partial \boldsymbol{\Omega}_2} + \frac{f_2(\theta)}{1 - F_2(\theta)} \frac{\partial \theta}{\partial \boldsymbol{\Omega}_2} \right] \\
&\quad - a \frac{\frac{\partial^2 \Phi}{\partial \boldsymbol{\Omega}_2^2}(1 + \Phi) - \left(\frac{\partial \Phi}{\partial \boldsymbol{\Omega}_2} \right)^2}{(1 + \Phi)^2} - a \frac{\frac{\partial f_1(\theta)}{\partial \boldsymbol{\Omega}_2} F_1(\theta) - \frac{\partial \theta}{\partial \boldsymbol{\Omega}_2} [f_1(\theta)]^2}{[F_1(\theta)]^2} \frac{\partial \theta}{\partial \boldsymbol{\Omega}_2} - a \frac{f_1(\theta)}{F_1(\theta)} \frac{\partial^2 \theta}{\partial \boldsymbol{\Omega}_2^2} \\
&\quad + (n - a) \left[\frac{1}{(1 + \Phi)^2} - \frac{1}{\Phi^2} \right] \left(\frac{\partial \Phi}{\partial \boldsymbol{\Omega}_2} \right)^2 + (n - a) \left(\frac{1}{\Phi} - \frac{1}{1 + \Phi} \right) \frac{\partial^2 \Phi}{\partial \boldsymbol{\Omega}_2^2} \\
&\quad + (n - a) \frac{\frac{\partial f_2(\theta)}{\partial \boldsymbol{\Omega}_2}[1 - F_2(\theta)] + [f_2(\theta)]^2 \frac{\partial \theta}{\partial \boldsymbol{\Omega}_2}}{[1 - F_2(\theta)]^2} \frac{\partial \theta}{\partial \boldsymbol{\Omega}_2} + (n - a) \frac{f_2(\theta)}{1 - F_2(\theta)} \frac{\partial^2 \theta}{\partial \boldsymbol{\Omega}_2^2}.
\end{aligned}
$$

3. GENERAL THREE-PIECE COMPOSITE DISTRIBUTION

Suppose that the lower tail, middle part and the upper tail are described probability density functions f_1, f_2 and f_3, respectively. Then the probability density and cumulative distribution functions of the most general three-piece composite distribution are

$$f(x) = \begin{cases} \dfrac{1}{1+\Phi_2+\Phi_3}\dfrac{f_1(x)}{F_1(\theta_1)}, & x \le \theta_1, \\ \dfrac{\Phi_2}{1+\Phi_2+\Phi_3}\dfrac{f_2(x)}{F_2(\theta_2)-F_2(\theta_1)}, & \theta_1 < x \le \theta_2, \\ \dfrac{\Phi_3}{1+\Phi_2+\Phi_3}\dfrac{f_3(x)}{1-F_3(\theta_2)}, & \theta_2 < x, \end{cases} \tag{3.1}$$

and

$$F(x) = \begin{cases} \dfrac{1}{1+\Phi_2+\Phi_3}\dfrac{F_1(x)}{F_1(\theta_1)}, & x \le \theta_1, \\ \dfrac{1}{1+\Phi_2+\Phi_3}\left[1+\Phi_2\dfrac{F_2(x)-F_2(\theta_1)}{F_2(\theta_2)-F_2(\theta_1)}\right], & \theta_1 < x \le \theta_2, \\ \dfrac{1}{1+\Phi_2+\Phi_3}\left[1+\Phi_2+\Phi_3\dfrac{F_3(x)-F_3(\theta_2)}{1-F_3(\theta_2)}\right], & \theta_2 < x, \end{cases} \tag{3.2}$$

respectively. Suppose f_1, f_2 and f_3 are parameterized by $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$ and $\boldsymbol{\Omega}_3$, respectively. Conditions for continuity and differentiability at θ_1, θ_2 ensure that $\theta_1 = \theta_1(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{\Omega}_3)$, $\theta_2 = \theta_2(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{\Omega}_3)$, $\Phi_2 = \Phi_2(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{\Omega}_3) = \frac{f_1(\theta_1)[F_2(\theta_2)-F_2(\theta_1)]}{F_1(\theta_1)f_2(\theta_1)}$ and $\Phi_3 = \Phi_3(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{\Omega}_3) = \frac{f_1(\theta_1)f_2(\theta_2)[1-F_3(\theta_2)]}{F_1(\theta_1)f_2(\theta_1)f_3(\theta_2)}$. If $x_1, x_2, ..., x_n$ is a random sample from (3.1)–(3.2) then the log-likelihood function of $(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{\Omega}_3)$ is

$$\begin{aligned} \log L(\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{\Omega}_3) = & \sum_{x_i \le \theta_1} \log f_1(x_i) - a \log F_1(\theta_1) - n \log(1+\Phi_2+\Phi_3) \\ & + b\{\log \Phi_2 - \log[F_2(\theta_2) - F_2(\theta_1)]\} \\ & + \sum_{\theta_1 \le x_i \le \theta_2} \log f_2(x_i) + c\{\log \Phi_3 - \log[1 - F_3(\theta_2)]\} + \sum_{\theta_2 < x_i} \log f_3(x_i), \end{aligned}$$

where

$$a = \sum_{i=1}^{n} I\{x_i \le \theta_1\},$$

$$b = \sum_{i=1}^{n} I\{\theta_1 < x_i \le \theta_2\}$$

and

$$c = \sum_{i=1}^{n} I\{\theta_2 < x_i\}$$

with $a+b+c=n$. The first order partial derivatives are

$$\begin{aligned}\frac{\partial \log L}{\partial \boldsymbol{\Omega}_1} &= \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}\sum_{i=1}^{n}\Delta(x_i-\theta_1)\log f_1(x_i)+\sum_{x_i\le\theta_1}\frac{\frac{\partial f_1(x_i)}{\partial \boldsymbol{\Omega}_1}}{f_1(x_i)}-\frac{\partial a}{\partial \boldsymbol{\Omega}_1}\log F_1(\theta_1)-\frac{a f_1(\theta_1)\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}}{F_1(\theta_1)}\\ &-\frac{n\left(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_1}+\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_1}\right)}{1+\Phi_2+\Phi_3}+\frac{\partial b}{\partial \boldsymbol{\Omega}_1}\{\log\Phi_2-\log[F_2(\theta_2)-F_2(\theta_1)]\}\\ &+b\left[\frac{\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_1}}{\Phi_2}-\frac{\frac{f_2(\theta_2)\partial\theta_2}{\partial \boldsymbol{\Omega}_1}-\frac{f_2(\theta_1)\partial\theta_1}{\partial \boldsymbol{\Omega}_1}}{F_2(\theta_2)-F_2(\theta_1)}\right]-\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}\sum_{x_i\le\theta_2}\Delta(x_i-\theta_1)\log f_2(x_i)\\ &+\frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1}\sum_{\theta_1\le x}\Delta(x_i-\theta_2)\log f_2(x_i)\\ &+\frac{\partial c}{\partial \boldsymbol{\Omega}_1}\{\log\Phi_3-\log[1-F_3(\theta_2)]\}+c\left[\frac{\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_1}}{\Phi_3}+\frac{\frac{f_3(\theta_2)\partial\theta_2}{\partial \boldsymbol{\Omega}_1}}{1-F_3(\theta_2)}\right]-\frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1}\sum_{i=1}^{n}\Delta(x_i-\theta_2)\log f_3(x_i),\end{aligned}$$

$$\begin{aligned}\frac{\partial \log L}{\partial \boldsymbol{\Omega}_2} &= \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2}\sum_{i=1}^{n}\Delta(x_i-\theta_1)\log f_1(x_i)-\frac{\partial a}{\partial \boldsymbol{\Omega}_2}\log F_1(\theta_1)-\frac{a f_1(\theta_1)\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2}}{F_1(\theta_1)}\\ &-\frac{n\left(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_2}+\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_2}\right)}{1+\Phi_2+\Phi_3}+\frac{\partial b}{\partial \boldsymbol{\Omega}_2}\{\log\Phi_2-\log[F_2(\theta_2)-F_2(\theta_1)]\}\\ &+b\left[\frac{\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_2}}{\Phi_2}-\frac{\frac{f_2(\theta_2)\partial\theta_2}{\partial \boldsymbol{\Omega}_2}-\frac{f_2(\theta_1)\partial\theta_1}{\partial \boldsymbol{\Omega}_2}}{F_2(\theta_2)-F_2(\theta_1)}\right]-\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2}\sum_{x_i\le\theta_2}\Delta(x_i-\theta_1)\log f_2(x_i)\\ &+\frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2}\sum_{\theta_1\le x_i}\Delta(x_i-\theta_2)\log f_2(x_i)\\ &+\sum_{\theta_1\le x_i\le\theta_2}\frac{\frac{\partial f_2(x_i)}{\partial \boldsymbol{\Omega}_2}}{f_2(x_i)}+\frac{\partial c}{\partial \boldsymbol{\Omega}_2}\{\log\Phi_3-\log[1-F_3(\theta_2)]\}+c\left[\frac{\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_2}}{\Phi_3}+\frac{\frac{f_3(\theta_2)\partial\theta_2}{\partial \boldsymbol{\Omega}_2}}{1-F_3(\theta_2)}\right]\\ &-\frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2}\sum_{i=1}^{n}\Delta(x_i-\theta_2)\log f_3(x_i)\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \log L}{\partial \boldsymbol{\Omega}_3} &= \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3}\sum_{i=1}^{n}\Delta(x_i-\theta_1)\log f_1(x_i)-\frac{\partial a}{\partial \boldsymbol{\Omega}_3}\log F_1(\theta_1)-\frac{a f_1(\theta_1)\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3}}{F_1(\theta_1)}\\ &-\frac{n\left(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_3}+\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_3}\right)}{1+\Phi_2+\Phi_3}+\frac{\partial b}{\partial \boldsymbol{\Omega}_3}\{\log\Phi_2-\log[F_2(\theta_2)-F_2(\theta_1)]\}\\ &+b\left[\frac{\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_3}}{\Phi_2}-\frac{\frac{f_2(\theta_2)\partial\theta_2}{\partial \boldsymbol{\Omega}_3}-\frac{f_2(\theta_1)\partial\theta_1}{\partial \boldsymbol{\Omega}_3}}{F_2(\theta_2)-F_2(\theta_1)}\right]-\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3}\sum_{x_i\le\theta_2}\Delta(x_i-\theta_1)\log f_2(x_i)\\ &+\frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3}\sum_{\theta_1\le x_i}\Delta(x_i-\theta_2)\log f_2(x_i)+\frac{\partial c}{\partial \boldsymbol{\Omega}_3}\{\log\Phi_3-\log[1-F_3(\theta_2)]\}\\ &+c\left[\frac{\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_3}}{\Phi_3}+\frac{\frac{f_3(\theta_2)\partial\theta_2}{\partial \boldsymbol{\Omega}_3}}{1-F_3(\theta_2)}\right]-\frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3}\sum_{i=1}^{n}\Delta(x_i-\theta_2)\log f_3(x_i)+\sum_{\theta_2<x_i}\frac{\frac{\partial f_3(x_i)}{\partial \boldsymbol{\Omega}_3}}{f_3(x_i)}.\end{aligned}$$

Hence, the observed information matrix is

$$\mathbf{J} = -\begin{pmatrix} \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1^2} & \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} & \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3} \\ \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} & \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_2^2} & \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3} \\ \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3} & \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3} & \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_3^2} \end{pmatrix},$$

where

$$\begin{aligned}
\frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1^2} \;=\; & \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1^2} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \log f_1(x_i) + \frac{2 \partial \theta_1}{\partial \boldsymbol{\Omega}_1} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \frac{\frac{\partial f_1(x_i)}{\partial \boldsymbol{\Omega}_1}}{f_1(x_i)} \\
& + \sum_{x_i \le \theta_1} \frac{\frac{\partial^2 f_1(x_i)}{\partial \boldsymbol{\Omega}_1^2} f_1(x_i) - \left[\frac{\partial f_1(x_i)}{\partial \boldsymbol{\Omega}_1}\right]^2}{[f_1(x_i)]^2} - \frac{\partial^2 a}{\partial \boldsymbol{\Omega}_1^2} \log F_1(\theta_1) \\
& - \frac{\left[\frac{\partial a}{\partial \boldsymbol{\Omega}_1} f_1(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} + a \frac{\partial f_1(\theta_1)}{\partial \boldsymbol{\Omega}_1} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} + a f_1(\theta_1) \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1^2}\right] F_1(\theta_1) - a[f_1(\theta_1)]^2 \left(\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}\right)^2}{[F_1(\theta_1)]^2} \\
& - \frac{\partial a}{\partial \boldsymbol{\Omega}_1} \frac{f_1(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}}{F_1(\theta_1)} - \frac{n\left[\left(\frac{\partial^2 \Phi_2}{\partial \boldsymbol{\Omega}_1^2} + \frac{\partial^2 \Phi_3}{\partial \boldsymbol{\Omega}_1^2}\right)(1 + \Phi_2 + \Phi_3) - \left(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_1} + \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_1}\right)^2\right]}{(1 + \Phi_2 + \Phi_3)^2} \\
& + \frac{\partial^2 b}{\partial \boldsymbol{\Omega}_1^2} \{\log \Phi_2 - \log[F_2(\theta_2) - F_2(\theta_1)]\} + 2 \frac{\partial b}{\partial \boldsymbol{\Omega}_1} \left[\frac{\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_1}}{\Phi_2} - \frac{f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}}{F_2(\theta_2) - F_2(\theta_1)}\right] \\
& + \frac{b \frac{\partial^2 \Phi_2}{\partial \boldsymbol{\Omega}_1^2} \Phi_2 - b\left(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_1}\right)^2}{\Phi_2^2} \\
& - \frac{b\left[\frac{\partial f_2(\theta_2)}{\partial \boldsymbol{\Omega}_1} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} + f_2(\theta_2) \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_1^2} - \frac{\partial f_2(\theta_1)}{\partial \boldsymbol{\Omega}_1} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} - f_2(\theta_1) \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1^2}\right]}{F_2(\theta_2) - F_2(\theta_1)} \\
& + \frac{b\left[f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}\right]^2}{[F_2(\theta_2) - F_2(\theta_1)]^2} \\
& - 2 \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \Delta(x_i - \theta_2) \log f_2(x_i) \\
& - \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1^2} \sum_{x_i \le \theta_2} \Delta(x_i - \theta_1) \log f_2(x_i) \;+\; \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_1^2} \sum_{\theta_1 \le x_i} \Delta(x_i - \theta_2) \log f_2(x_i) \\
& + \frac{\partial^2 c}{\partial \boldsymbol{\Omega}_1^2} \{\log \Phi_3 - \log[1 - F_3(\theta_2)]\} \;+\; 2 \frac{\partial c}{\partial \boldsymbol{\Omega}_1} \left[\frac{\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_1}}{\Phi_3} + \frac{\frac{f_3(\theta_2) \partial \theta_2}{\partial \boldsymbol{\Omega}_1}}{1 - F_3(\theta_2)}\right] \\
& + c \frac{\Phi_3 \frac{\partial^2 \Phi_3}{\partial \boldsymbol{\Omega}_1^2} - \left(\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_1}\right)^2}{\Phi_3^2} \;-\; \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_1^2} \sum_{i=1}^{n} \Delta(x_i - \theta_2) \log f_3(x_i) \\
& + \frac{c\left[\frac{\partial f_3(\theta_2)}{\partial \boldsymbol{\Omega}_1} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} + f_3(\theta_2) \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_1^2}\right][1 - F_3(\theta_2)] + c\left[f_3(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1}\right]^2}{[1 - F_3(\theta_2)]^2},
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} \;=\; & \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \log f_1(x_i) \;+\; \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \frac{\frac{\partial f_1(x_i)}{\partial \boldsymbol{\Omega}_1}}{f_1(x_i)} \\
& - \frac{\partial^2 a}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} \log F_1(\theta_1) \;-\; \frac{\partial a}{\partial \boldsymbol{\Omega}_1} \frac{f_1(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2}}{F_1(\theta_1)} \\
& - \frac{n\Big(\frac{\partial^2 \Phi_2}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} + \frac{\partial^2 \Phi_3}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2}\Big)(1 + \Phi_2 + \Phi_3) - n\Big(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_1} + \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_1}\Big)\Big(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_2} + \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_2}\Big)}{(1 + \Phi_2 + \Phi_3)^2} \\
& - \frac{\frac{\partial a}{\partial \boldsymbol{\Omega}_2} f_1(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} + a \frac{\partial f_1(\theta_1)}{\partial \boldsymbol{\Omega}_2} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} + a f_1(\theta_1) \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2}}{F_1(\theta_1)} \;+\; \frac{a[f_1(\theta_1)]^2 \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2}}{[F_1(\theta_1)]^2} \\
& + \frac{\partial^2 b}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} \{\log \Phi_2 - \log[F_2(\theta_2) - F_2(\theta_1)]\} \\
& + \frac{\partial b}{\partial \boldsymbol{\Omega}_1} \left[\frac{\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_2}}{\Phi_2} - \frac{f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2}}{F_2(\theta_2) - F_2(\theta_1)} \right] \\
& + \frac{\partial b}{\partial \boldsymbol{\Omega}_2} \left[\frac{\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_1}}{\Phi_2} - \frac{f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}}{F_2(\theta_2) - F_2(\theta_1)} \right] + \frac{b \frac{\partial^2 \Phi_2}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} \Phi_2 - b \frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_1} \frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_2}}{\Phi_2^2} \\
& - \frac{b\Big[\frac{\partial f_2(\theta_2)}{\partial \boldsymbol{\Omega}_2} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} + f_2(\theta_2) \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} - \frac{\partial f_2(\theta_2)}{\partial \boldsymbol{\Omega}_2} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} - f_2(\theta_1) \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2}\Big]}{F_2(\theta_2) - F_2(\theta_1)} \\
& + \frac{b\Big[f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}\Big]\Big[f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} - f_2(\theta_1) \frac{\partial \theta_1)}{\partial \boldsymbol{\Omega}_2}\Big]}{[F_2(\theta_2) - F_2(\theta_1)]^2} \\
& - \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} \sum_{x_i \le \theta_2} \Delta(x_i - \theta_1) \log f_2(x_i) \\
& - \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \Delta(x_i - \theta_2) \log f_2(x_i) \\
& - \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} \sum_{x_i \le \theta_2} \Delta(x_i - \theta_1) \frac{\frac{\partial f_2(x_i)}{\partial \boldsymbol{\Omega}_2}}{f_2(x_i)} \\
& + \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} \sum_{\theta_1 \le x_i} \Delta(x_i - \theta_2) \log f_2(x_i) \\
& - \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \Delta(x_i - \theta_2) \log f_2(x_i) \\
& + \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} \sum_{\theta_1 \le x_i} \Delta(x_i - \theta_2) \frac{\frac{\partial f_2(x_i)}{\partial \boldsymbol{\Omega}_2}}{f_2(x_i)} + \frac{\partial^2 c}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} \{\log \Phi_3 - \log[1 - F_3(\theta_2)]\} \\
& + \left[\frac{\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_2}}{\Phi_3} + \frac{f_3(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2}}{1 - F_3(\theta_2)} \right] \frac{\partial c}{\partial \boldsymbol{\Omega}_1} + \frac{\partial c}{\partial \boldsymbol{\Omega}_2} \left[\frac{\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_1}}{\Phi_3} + \frac{f_3(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1}}{1 - F_3(\theta_2)} \right] \\
& + \frac{c \frac{\partial^2 \Phi_3}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} \Phi_3 - c \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_1} \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_2}}{\Phi_3^2} \\
& + \frac{c\Big[\frac{\partial f_3(\theta_2)}{\partial \boldsymbol{\Omega}_2} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} + f_3(\theta_2) \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2}\Big][1 - F_3(\theta_2)] + c[f_3(\theta_2)]^2 \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2}}{[1 - F_3(\theta_2)]^2} \\
& - \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_1 \boldsymbol{\Omega}_2} \sum_{i=1}^{n} \Delta(x_i - \theta_2) \log f_3(x_i),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3} &= \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \log f_1(x_i) + \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \frac{\frac{\partial f_1(x_i)}{\partial \boldsymbol{\Omega}_1}}{f_1(x_i)} \\
&- \frac{\partial^2 a}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3} \log F_1(\theta_1) \\
&- \frac{\frac{\partial a}{\partial \boldsymbol{\Omega}_3} f_1(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} + a \frac{\partial f_1(\theta_1)}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} + a f_1(\theta_1) \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3}}{F_1(\theta_1)} \\
&+ \frac{a[f_1(\theta_1)]^2 \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3}}{[F_1(\theta_1)]^2} - \frac{\partial a}{\partial \boldsymbol{\Omega}_1} \frac{f_1(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3}}{F_1(\theta_1)} \\
&- \frac{n\Big(\frac{\partial^2 \Phi_2}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3} + \frac{\partial^2 \Phi_3}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3}\Big)(1 + \Phi_2 + \Phi_3) - n\Big(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_1} + \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_1}\Big)\Big(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_3} + \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_3}\Big)}{(1 + \Phi_2 + \Phi_3)^2} \\
&+ \frac{\partial^2 b}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3} \{\log \Phi_2 - \log[F_2(\theta_2) - F_2(\theta_1)]\} \\
&+ \frac{\partial b}{\partial \boldsymbol{\Omega}_1} \left[\frac{\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_3}}{\Phi_2} - \frac{f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3}}{F_2(\theta_2) - F_2(\theta_1)} \right] \\
&+ \frac{\partial b}{\partial \boldsymbol{\Omega}_3} \left[\frac{\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_1}}{\Phi_2} - \frac{f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}}{F_2(\theta_2) - F_2(\theta_1)} \right] + \frac{b \frac{\partial^2 \Phi_2}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3} \Phi_2 - b \frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_1} \frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_3}}{\Phi_2^2} \\
&- \frac{b\Big[\frac{\partial f_2(\theta_2)}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} + f_2(\theta_2) \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3} - \frac{\partial f_2(\theta_2)}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} - f_2(\theta_2) \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3}\Big]}{F_2(\theta_2) - F_2(\theta_1)} \\
&+ \frac{b\Big[f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} - f_2(\theta_1) \frac{\partial \theta_1)}{\partial \boldsymbol{\Omega}_1}\Big]\Big[f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3}\Big]}{[F_2(\theta_2) - F_2(\theta_1)]^2} \\
&- \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3} \sum_{x_i \le \theta_2} \Delta(x_i - \theta_1) \log f_2(x_i) \\
&- \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \Delta(x_i - \theta_2) \log f_2(x_i) \\
&+ \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3} \sum_{\theta_1 \le x_i} \Delta(x_i - \theta_2) \log f_2(x_i) \\
&- \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \Delta(x_i - \theta_2) \log f_2(x_i) \\
&+ \frac{\partial^2 c}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3} \{\log \Phi_3 - \log[1 - F_3(\theta_2)]\} \\
&+ \left[\frac{\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_3}}{\Phi_3} + \frac{f_3(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3}}{1 - F_3(\theta_2)} \right] \frac{\partial c}{\partial \boldsymbol{\Omega}_1} + \frac{\partial c}{\partial \boldsymbol{\Omega}_3} \left[\frac{\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_1}}{\Phi_3} + \frac{f_3(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1}}{1 - F_3(\theta_2)} \right] \\
&+ \frac{c \frac{\partial^2 \Phi_3}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3} \Phi_3 - c \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_1} \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_3}}{\Phi_3^2} \\
&+ \frac{c\Big[\frac{\partial f_3(\theta_2)}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} + f_3(\theta_2) \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_3}\Big][1 - F_3(\theta_2)] + c[f_3(\theta_2)]^2 \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3}}{[1 - F_3(\theta_2)]^2} \\
&- \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_1 \boldsymbol{\Omega}_3} \sum_{i=1}^{n} \Delta(x_i - \theta_2) \log f_3(x_i) - \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_1} \sum_{i=1}^{n} \Delta(x_i - \theta_2) \frac{\frac{\partial f_3(x_i)}{\partial \boldsymbol{\Omega}_3}}{f_3(x_i)},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_2^2} \;=\; & \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_2^2} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \log f_1(x_i) \\
& - \frac{\partial^2 a}{\partial \boldsymbol{\Omega}_2^2} \log F_1(\theta_1) - \frac{\partial a}{\partial \boldsymbol{\Omega}_2} \frac{f_1(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2}}{F_1(\theta_1)} \\
& - \frac{\left[\frac{\partial a}{\partial \boldsymbol{\Omega}_1} f_1(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2} + a \frac{\partial f_1(\theta_1)}{\partial \boldsymbol{\Omega}_2} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2} + a f_1(\theta_1) \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_2^2}\right] F_1(\theta_1) - a[f_1(\theta_1)]^2 \left(\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2}\right)^2}{[F_1(\theta_1)]^2} \\
& - \frac{n\left(\frac{\partial^2 \Phi_2}{\partial \boldsymbol{\Omega}_2^2} + \frac{\partial^2 \Phi_3}{\partial \boldsymbol{\Omega}_2^2}\right)(1 + \Phi_2 + \Phi_3) - n\left(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_2} + \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_2}\right)^2}{(1 + \Phi_2 + \Phi_3)^2} \\
& + \frac{\partial^2 b}{\partial \boldsymbol{\Omega}_2^2} \{\log \Phi_2 - \log[F_2(\theta_2) - F_2(\theta_1)]\} \\
& + \frac{2 \partial b}{\partial \boldsymbol{\Omega}_2} \left[\frac{\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_2}}{\Phi_2} - \frac{f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2}}{F_2(\theta_2) - F_2(\theta_1)}\right] + \frac{b \frac{\partial^2 \Phi_2}{\partial \boldsymbol{\Omega}_2^2} \Phi_2 - b\left(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_2}\right)^2}{\Phi_2^2} \\
& - \frac{b\left[\frac{\partial f_2(\theta_2)}{\partial \boldsymbol{\Omega}_2} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} + f_2(\theta_2) \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_2^2} - \frac{\partial f_2(\theta_1)}{\partial \boldsymbol{\Omega}_2} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2} - f_2(\theta_1) \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_2^2}\right]}{F_2(\theta_2) - F_2(\theta_1)} \\
& + \frac{b\left[f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2}\right]^2}{[F_2(\theta_2) - F_2(\theta_1)]^2} \\
& - \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_2^2} \sum_{x_i \le \theta_2} \Delta(x_i - \theta_1) \log f_2(x_i) \;-\; \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2} \sum_{x_i \le \theta_2} \Delta(x_i - \theta_1) \frac{\frac{\partial f_2(x_i)}{\partial \boldsymbol{\Omega}_2}}{f_2(x_i)} \\
& - 2 \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \Delta(x_i - \theta_2) \log f_2(x_i) \\
& + \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} \sum_{\theta_1 \le x_i} \Delta(x_i - \theta_2) \frac{\frac{\partial f_2(x_i)}{\partial \boldsymbol{\Omega}_2}}{f_2(x_i)} \\
& + \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_2^2} \sum_{\theta_1 \le x_i} \Delta(x_i - \theta_2) \log f_2(x_i) \;+\; \frac{\partial^2 c}{\partial \boldsymbol{\Omega}_2^2} \{\log \Phi_3 - \log[1 - F_3(\theta_2)]\} \\
& + 2 \frac{\partial c}{\partial \boldsymbol{\Omega}_2} \left[\frac{\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_2}}{\Phi_3} + \frac{f_3(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2}}{1 - F_3(\theta_2)}\right] + c \frac{\Phi_3 \frac{\partial^2 \Phi_3}{\partial \boldsymbol{\Omega}_2^2} - \left(\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_2}\right)^2}{\Phi_3^2} \\
& - \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_2^2} \sum_{i=1}^{n} \Delta(x_i - \theta_2) \log f_3(x_i) \\
& + \frac{c\left[\frac{\partial f_3(\theta_2)}{\partial \boldsymbol{\Omega}_2} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} + f_3(\theta_2) \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_2^2}\right][1 - F_3(\theta_2)] + c\left[f_3(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2}\right]^2}{[1 - F_3(\theta_2)]^2} \\
& - \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2} \sum_{x_i \le \theta_2} \Delta(x_i - \theta_1) \frac{\frac{\partial f_2(x_i)}{\partial \boldsymbol{\Omega}_2}}{f_2(x_i)} \;+\; \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} \sum_{\theta_1 \le x_i} \Delta(x_i - \theta_2) \frac{\frac{\partial f_2(x_i)}{\partial \boldsymbol{\Omega}_2}}{f_2(x_i)} \\
& + \sum_{\theta_1 \le x_i \le \theta_2} \frac{\frac{\partial^2 f_2(x_i)}{\partial \boldsymbol{\Omega}_2^2} f_2(x_i) - \left[\frac{\partial f_2(x_i)}{\partial \boldsymbol{\Omega}_2}\right]^2}{[f_2(x_i)]^2},
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3} &= \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \log f_1(x_i) - \frac{\partial^2 a}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3} \log F_1(\theta_1) \\
&- \frac{\frac{\partial a}{\partial \boldsymbol{\Omega}_3} f_1(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2} + a \frac{\partial f_1(\theta_1)}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2} + a f_1(\theta_1) \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3}}{F_1(\theta_1)} \\
&+ \frac{a[f_1(\theta_1)]^2 \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3}}{[F_1(\theta_1)]^2} - \frac{\partial a}{\partial \boldsymbol{\Omega}_2} \frac{f_1(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3}}{F_1(\theta_1)} \\
&- \frac{n\Big(\frac{\partial^2 \Phi_2}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3} + \frac{\partial^2 \Phi_3}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3}\Big)(1 + \Phi_2 + \Phi_3) - n\Big(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_2} + \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_2}\Big)\Big(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_3} + \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_3}\Big)}{(1 + \Phi_2 + \Phi_3)^2} \\
&+ \frac{\partial^2 b}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3} \{\log \Phi_2 - \log[F_2(\theta_2) - F_2(\theta_1)]\} \\
&+ \frac{\partial b}{\partial \boldsymbol{\Omega}_2} \left[\frac{\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_3}}{\Phi_2} - \frac{f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3}}{F_2(\theta_2) - F_2(\theta_1)} \right] \\
&+ \frac{\partial b}{\partial \boldsymbol{\Omega}_3} \left[\frac{\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_2}}{\Phi_2} - f_2(\theta_2) \frac{\frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2}}{F_2(\theta_2) - F_2(\theta_1)} \right] + \frac{b \frac{\partial^2 \Phi_2}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3} \Phi_2 - b \frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_2} \frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_3}}{\Phi_2^2} \\
&- \frac{b\Big[\frac{\partial f_2(\theta_2)}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} + f_2(\theta_2) \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3} - \frac{\partial f_2(\theta_2)}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2} - f_2(\theta_1) \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3}\Big]}{F_2(\theta_2) - F_2(\theta_1)} \\
&+ \frac{b\Big[f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2}\Big]\Big[f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3}\Big]}{[F_2(\theta_2) - F_2(\theta_1)]^2} \\
&- \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3} \sum_{x_i \le \theta_2} \Delta(x_i - \theta_1) \log f_2(x_i) \\
&- \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_2} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \Delta(x_i - \theta_2) \log f_2(x_i) \\
&- \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3} \sum_{x_i \le \theta_2} \Delta(x_i - \theta_1) \frac{\frac{\partial f_2(x_i)}{\partial \boldsymbol{\Omega}_2}}{f_2(x_i)} + \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3} \sum_{\theta_1 \le x_i} \Delta(x_i - \theta_2) \frac{\frac{\partial f_2(x_i)}{\partial \boldsymbol{\Omega}_2}}{f_2(x_i)} \\
&+ \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3} \sum_{\theta_1 \le x_i} \Delta(x_i - \theta_2) \log f_2(x_i) \\
&- \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \Delta(x_i - \theta_2) \log f_2(x_i) \\
&+ \frac{\partial^2 c}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3} \{\log \Phi_3 - \log[1 - F_3(\theta_2)]\} \\
&+ \left[\frac{\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_3}}{\Phi_3} + \frac{f_3(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3}}{1 - F_3(\theta_2)} \right] \frac{\partial c}{\partial \boldsymbol{\Omega}_2} + \frac{\partial c}{\partial \boldsymbol{\Omega}_3} \left[\frac{\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_2}}{\Phi_3} + \frac{f_3(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2}}{1 - F_3(\theta_2)} \right] \\
&+ \frac{c \frac{\partial^2 \Phi_3}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3} \Phi_3 - c \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_2} \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_3}}{\Phi_3^2} \\
&+ \frac{c\Big[\frac{\partial f_3(\theta_2)}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} + f_3(\theta_2) \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_3}\Big][1 - F_3(\theta_2)] + c[f_3(\theta_2)]^2 \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3}}{[1 - F_3(\theta_2)]^2} \\
&- \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_2 \boldsymbol{\Omega}_3} \sum_{i=1}^{n} \Delta(x_i - \theta_2) \log f_3(x_i) - \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_2} \sum_{i=1}^{n} \Delta(x_i - \theta_2) \frac{\frac{\partial f_3(x_i)}{\partial \boldsymbol{\Omega}_3}}{f_3(x_i)}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_3^2} \;=\; & \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_3^2} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \log f_1(x_i) - \frac{\partial^2 a}{\partial \boldsymbol{\Omega}_3^2} \log F_1(\theta_1) - \frac{\partial a}{\partial \boldsymbol{\Omega}_3} \frac{f_1(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3}}{F_1(\theta_1)} \\
& - \frac{\left[\frac{\partial a}{\partial \boldsymbol{\Omega}_3} f_1(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3} + a \frac{\partial f_1(\theta_1)}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3} + a f_1(\theta_1) \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_3^2} \right] F_1(\theta_1) - a [f_1(\theta_1)]^2 \left(\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3} \right)^2}{[F_1(\theta_1)]^2} \\
& - \frac{n \left[\left(\frac{\partial^2 \Phi_2}{\partial \boldsymbol{\Omega}_3^2} + \frac{\partial^2 \Phi_3}{\partial \boldsymbol{\Omega}_3^2} \right) (1 + \Phi_2 + \Phi_3) - \left(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_3} + \frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_3} \right)^2 \right]}{(1 + \Phi_2 + \Phi_3)^2} \\
& + \frac{\partial^2 b}{\partial \boldsymbol{\Omega}_3^2} \{ \log \Phi_2 - \log [F_2(\theta_2) - F_2(\theta_1)] \} \\
& + 2 \frac{\partial b}{\partial \boldsymbol{\Omega}_3} \left[\frac{\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_3}}{\Phi_2} - f_2(\theta_2) \frac{\frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3}}{F_2(\theta_2) - F_2(\theta_1)} \right] \\
& + \frac{b \frac{\partial^2 \Phi_2}{\partial \boldsymbol{\Omega}_3^2} \Phi_2 - b \left(\frac{\partial \Phi_2}{\partial \boldsymbol{\Omega}_3} \right)^2}{\Phi_2^2} \\
& - \frac{b \left[\frac{\partial f_2(\theta_2)}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3} + f_2(\theta_2) \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_3^2} - \frac{\partial f_2(\theta_1)}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3} - f_2(\theta_1) \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_3^2} \right]}{F_2(\theta_2) - F_2(\theta_1)} \\
& + \frac{b \left[f_2(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3} - f_2(\theta_1) \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3} \right]^2}{[F_2(\theta_2) - F_2(\theta_1)]^2} \\
& - 2 \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \Delta(x_i - \theta_2) \log f_2(x_i) \\
& - \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_3^2} \sum_{x_i \le \theta_2} \Delta(x_i - \theta_1) \log f_2(x_i) + \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_3^2} \sum_{\theta_1 \le x_i} \Delta(x_i - \theta_2) \log f_2(x_i) \\
& + \frac{\partial^2 c}{\partial \boldsymbol{\Omega}_3^2} \{ \log \Phi_3 - \log [1 - F_3(\theta_2)] \} \\
& + 2 \frac{\partial c}{\partial \boldsymbol{\Omega}_3} \left[\frac{\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_3}}{\Phi_3} + \frac{f_3(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3}}{1 - F_3(\theta_2)} \right] + c \frac{\Phi_3 \frac{\partial^2 \Phi_3}{\partial \boldsymbol{\Omega}_3^2} - \left(\frac{\partial \Phi_3}{\partial \boldsymbol{\Omega}_3} \right)^2}{\Phi_3^2} \\
& - \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_3^2} \sum_{i=1}^{n} \Delta(x_i - \theta_2) \log f_3(x_i) \\
& + \frac{c \left[\frac{\partial f_3(\theta_2)}{\partial \boldsymbol{\Omega}_3} \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3} + f_3(\theta_2) \frac{\partial^2 \theta_2}{\partial \boldsymbol{\Omega}_3^2} \right] [1 - F_3(\theta_2)] + c \left[f_3(\theta_2) \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3} \right]^2}{[1 - F_3(\theta_2)]^2} \\
& - 2 \frac{\partial \theta_2}{\partial \boldsymbol{\Omega}_3} \sum_{i=1}^{n} \Delta(x_i - \theta_2) \frac{\frac{\partial f_3(x_i)}{\partial \boldsymbol{\Omega}_3}}{f_3(x_i)} + \sum_{\theta_2 < x_i} \frac{\frac{\partial^2 f_3(x_i)}{\partial \boldsymbol{\Omega}_3^2} f_3(x_i) - \left[\frac{\partial f_3(x_i)}{\partial \boldsymbol{\Omega}_3} \right]^2}{[f_3(x_i)]^2}.
\end{aligned}
$$

4. GENERAL m-PIECE COMPOSITE DISTRIBUTION

Suppose every piece of the distribution is described by f_i, F_i for $i = 1, 2, ..., m$. Then the probability density and cumulative distribution functions of the most general m-piece composite distribution are

$$f(x) = \begin{cases} \zeta \dfrac{f_1(x)}{F_1(\theta_1)}, & x \leq \theta_1, \\ \zeta \phi_i \dfrac{f_i(x)}{F_i(\theta_i) - F_i(\theta_{i-1})}, & \theta_{i-1} < x < \theta_i, \\ \zeta \phi_m \dfrac{f_m(x)}{1 - F_m(\theta_{m-1})}, & \theta_{m-1} < x, \end{cases} \tag{4.1}$$

and

$$F(x) = \begin{cases} \zeta \dfrac{F_1(x)}{F_1(\theta_1)}, & x \leq \theta_1, \\ \zeta \left[\displaystyle\sum_{j=1}^{i} \phi_j + \phi_i \frac{F_i(x) - F_i(\theta_{i-1})}{F_i(\theta_i) - F_i(\theta_{i-1})} \right], & \theta_{i-1} < x < \theta_i, \\ \zeta \phi_m \left[\displaystyle\sum_{j=1}^{m-1} \phi_j + \phi_m \frac{F_m(x) - F_m(\theta_{m-1})}{1 - F_m(\theta_{m-1})} \right], & \theta_{m-1} < x, \end{cases} \tag{4.2}$$

respectively, where θ_i, $i = 2, ..., m-1$, mark the section boundaries and $\zeta = 1/(1+\phi_2+\cdots+\phi_m)$. We take $\phi_1 = 1$. Suppose f_i are parameterized by $\boldsymbol{\Omega}_i$. Conditions for continuity and differentiability at θ_i ensure that $\theta_i = \theta_i(\boldsymbol{\Omega}_1, ..., \boldsymbol{\Omega}_m)$ and

$$\phi_i = \phi_i(\boldsymbol{\Omega}_1, ..., \boldsymbol{\Omega}_m) = \frac{\displaystyle\prod_{j=1}^{i-1} f_j(\theta_j)[F_i(\theta_i) - F_i(\theta_{i-1})]}{F_1(\theta_1) \displaystyle\prod_{j=2}^{i} f_j(\theta_{j-1})}.$$

Let

$$a_1 = \sum_{i=1}^{n} I\{x_i \leq \theta_1\},$$

$$a_m = \sum_{i=1}^{n} I\{\theta_{m-1} < x\}$$

and

$$a_j = \sum_{i=1}^{n} I\{\theta_{j-1} \leq x_i \leq \theta_j\}$$

for $j = 2, ..., m-1$. Note that $a_1 + a_2 + \cdots + a_m = n$. If $x_1, x_2, ..., x_n$ is a random sample from (4.1)–(4.2) then the log-likelihood function of $(\boldsymbol{\Omega}_1, ..., \boldsymbol{\Omega}_m)$ is

$$\begin{aligned}\log L(\boldsymbol{\Omega}_1, ..., \boldsymbol{\Omega}_m) \;=\; & n \log \zeta - a_1 \log[F_1(\theta_1)] - \sum_{i=2}^{m-1} a_i \log[F_i(\theta_i) - F_i(\theta_{i-1})] \\ & - a_m \log[1 - F_m(\theta_{m-1})] + \sum_{x_i \le \theta_1} \log f_1(x_i) \\ & + \sum_{i=2}^{m-1} \left[a_i \log \phi_i + \sum_{\theta_{i-1} \le x_i \le \theta_i} \log f_i(x_i) \right] + a_m \log \phi_m + \sum_{\theta_{m-1} < x_i} \log f_m(x_i).\end{aligned}$$

The first order partial derivatives are

$$\begin{aligned}\frac{\partial \log L}{\partial \boldsymbol{\Omega}_1} \;=\; & \frac{n}{\zeta}\frac{\partial \zeta}{\partial \boldsymbol{\Omega}_1} - \frac{\partial a_1}{\partial \boldsymbol{\Omega}_1} \log[F_1(\theta_1)] - \frac{\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} a_1 f_1(\theta_1)}{F_1(\theta_1)} - \sum_{i=2}^{m-1} \frac{\partial a_i}{\partial \boldsymbol{\Omega}_1} \log[F_i(\theta_i) - F_i(\theta_{i-1})] \\ & - \sum_{i=2}^{m-1} a_i \frac{\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_1} f_i(\theta_i) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_1} f_i(\theta_{i-1})}{F_i(\theta_i) - F_i(\theta_{i-1})} - \frac{\partial a_m}{\partial \boldsymbol{\Omega}_1} \log[1 - F_m(\theta_{m-1})] + a_m \frac{\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_1} f_m(\theta_{m-1})}{1 - F_m(\theta_{m-1})} \\ & + \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \log f_1(x_i) + \sum_{x_i \le \theta_1} \frac{\frac{\partial f_1(x_i)}{\partial \boldsymbol{\Omega}_1}}{f_1(x_i)} \\ & + \sum_{i=1}^{m-1} \left[\frac{\partial a_i}{\partial \boldsymbol{\Omega}_1} \log \phi_i + a_i \frac{\frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_1}}{\phi_i} - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_1} \sum_{x_i \le \theta_i} \Delta(x_i - \theta_{i-1}) \log f_i(x_i) \right] \\ & + \sum_{i=1}^{m-1} \frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_1} \sum_{\theta_{i-1} \le x_i} \Delta(x_i - \theta_i) \log f_i(x_i) \\ & + \frac{\partial a_m}{\partial \boldsymbol{\Omega}_1} \log \phi_m + a_m \frac{\frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_1}}{\phi_m} - \frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_1} \sum_{i=1}^{n} \Delta(x_i - \theta_{m-1}) \log f_m(x_i),\end{aligned}$$

$$\begin{aligned}\frac{\partial \log L}{\partial \boldsymbol{\Omega}_i} \;=\; & \frac{n}{\zeta}\frac{\partial \zeta}{\partial \boldsymbol{\Omega}_i} - \frac{\partial a_1}{\partial \boldsymbol{\Omega}_i} \log[F_1(\theta_1)] - \frac{\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_i} a_1 f_1(\theta_1)}{F_1(\theta_1)} - \sum_{i=2}^{m-1} \frac{\partial a_i}{\partial \boldsymbol{\Omega}_i} \log[F_i(\theta_i) - F_i(\theta_{i-1})] \\ & - \sum_{i=2}^{m-1} a_i \frac{\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_i} f_i(\theta_i) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_i} f_i(\theta_{i-1})}{F_i(\theta_i) - F_i(\theta_{i-1})} - \frac{\partial a_m}{\partial \boldsymbol{\Omega}_i} \log[1 - F_m(\theta_{m-1})] + a_m \frac{\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_i} f_m(\theta_{m-1})}{1 - F_m(\theta_{m-1})} \\ & + \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_i} \sum_{i=1}^{n} \Delta(x_i - \theta_1) \log f_1(x_i) + \sum_{i=2}^{m-1} \sum_{\theta_{i-1} \le x_i \le \theta_i} \frac{\frac{\partial f_i(x_i)}{\partial \boldsymbol{\Omega}_i}}{f_i(x_i)} \\ & + \sum_{i=1}^{m-1} \left[\frac{\partial a_i}{\partial \boldsymbol{\Omega}_i} \log \phi_i + a_i \frac{\frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_i}}{\phi_i} - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_i} \sum_{x_i \le \theta_i} \Delta(x_i - \theta_{i-1}) \log f_i(x_i) \right] \\ & + \sum_{i=1}^{m-1} \frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_i} \sum_{\theta_{i-1} \le x_i} \Delta(x_i - \theta_i) \log f_i(x_i) \\ & + \frac{\partial a_m}{\partial \boldsymbol{\Omega}_i} \log \phi_m + a_m \frac{\frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_i}}{\phi_m} - \frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_i} \sum_{i=1}^{n} \Delta(x_i - \theta_{m-1}) \log f_m(x_i)\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \log L}{\partial \boldsymbol{\Omega}_m} &= \frac{n}{\zeta}\frac{\partial \zeta}{\partial \boldsymbol{\Omega}_m} - \frac{\partial a_1}{\partial \boldsymbol{\Omega}_m}\log[F_1(\theta_1)] - \frac{\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_m}a_1 f_1(\theta_1)}{F_1(\theta_1)} - \sum_{i=2}^{m-1}\frac{\partial a_i}{\partial \boldsymbol{\Omega}_m}\log[F_i(\theta_i)-F_i(\theta_{i-1})] \\ &- \sum_{i=2}^{m-1} a_i \frac{\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_m}f_i(\theta_i) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_m}f_i(\theta_{i-1})}{F_i(\theta_i)-F_i(\theta_{i-1})} - \frac{\partial a_m}{\partial \boldsymbol{\Omega}_m}\log[1-F_m(\theta_{m-1})] + a_m\frac{\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_m}f_m(\theta_{m-1})}{1-F_m(\theta_{m-1})} \\ &+ \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_m}\sum_{i=1}^{n}\Delta(x_i-\theta_1)\log f_1(x_i) \\ &+ \sum_{i=1}^{m-1}\left[\frac{\partial a_i}{\partial \boldsymbol{\Omega}_m}\log\phi_i + a_i\frac{\frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_m}}{\phi_i} - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_m}\sum_{x_i\le\theta_i}\Delta(x_i-\theta_{i-1})\log f_i(x_i)\right] \\ &+ \sum_{i=1}^{m-1}\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_m}\sum_{\theta_{i-1}\le x_i}\Delta(x_i-\theta_i)\log f_i(x_i) \\ &+ \frac{\partial a_m}{\partial \boldsymbol{\Omega}_m}\log\phi_m + a_m\frac{\frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_m}}{\phi_m} - \frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_m}\sum_{i=1}^{n}\Delta(x_i-\theta_{m-1})\log f_m(x_i) + \sum_{\theta_{m-1}<x_i}\frac{\frac{\partial f_m(x_i)}{\partial \boldsymbol{\Omega}_m}}{f_m(x_i)}\end{aligned}$$

for $i = 2, ..., m-1$. Hence, the observed information matrix is

$$\mathbf{J} = -\begin{pmatrix} \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1^2} & \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} & \cdots & \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_m} \\ \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_2} & \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_2^2} & \cdots & \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_m} & \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_2 \partial \boldsymbol{\Omega}_m} & \cdots & \frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_m^2} \end{pmatrix},$$

where

$$\begin{aligned}\frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1^2} &= \frac{n}{\zeta}\frac{\partial^2 \zeta}{\partial \boldsymbol{\Omega}_1^2} - \frac{n}{\zeta^2}\left(\frac{\partial \zeta}{\partial \boldsymbol{\Omega}_1}\right)^2 - \frac{\partial^2 a_1}{\partial \boldsymbol{\Omega}_1^2}\log F_1(\theta_1) - \frac{\partial a_1}{\partial \boldsymbol{\Omega}_1}\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}\frac{f_1(\theta_1)}{F_1(\theta_1)} + \frac{\left(\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}\right)^2 a_1[f_1(\theta_1)]^2}{[F_1(\theta_1)]^2} \\ &- \frac{\frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1^2}a_1 f_1(\theta_1) + \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}\frac{\partial a_1}{\partial \boldsymbol{\Omega}_1}f_1(\theta_1) + \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}a_1\frac{\partial f_1(\theta_1)}{\partial \boldsymbol{\Omega}_1}}{F_1(\theta_1)} \\ &- \sum_{i=2}^{m-1}\frac{\partial^2 a_i}{\partial \boldsymbol{\Omega}_1^2}\log[F_i(\theta_i)-F_i(\theta_{i-1})] - 2\sum_{i=2}^{m-1}\frac{\partial a_i}{\partial \boldsymbol{\Omega}_1}\frac{f_i(\theta_i)\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_1} - f_i(\theta_{i-1})\frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_1}}{F_i(\theta_i)-F_i(\theta_{i-1})} \\ &- \sum_{i=2}^{m-1} a_i \frac{\left[\frac{\partial^2 \theta_i}{\partial \boldsymbol{\Omega}_1^2}f_i(\theta_i) + \frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_1}\frac{\partial f_i(\theta_i)}{\partial \boldsymbol{\Omega}_1} - \frac{\partial^2 \theta_{i-1}}{\partial \boldsymbol{\Omega}_1^2}f_i(\theta_{i-1}) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_1}\frac{\partial f_i(\theta_{i-1})}{\partial \boldsymbol{\Omega}_1}\right]}{F_i(\theta_i)-F_i(\theta_{i-1})} \\ &+ \sum_{i=2}^{m-1} a_i \frac{\left[\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_1}f_i(\theta_i) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_1}f_i(\theta_{i-1})\right]^2}{[F_i(\theta_i)-F_i(\theta_{i-1})]^2} - \frac{\partial^2 a_m}{\partial \boldsymbol{\Omega}_1^2}\log[1-F_m(\theta_{m-1})]\end{aligned}$$

$$
\begin{aligned}
&+ 2\frac{\partial a_m}{\partial \boldsymbol{\Omega}_1}\frac{\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_1} f_m(\theta_{m-1})}{1-F_m(\theta_{m-1})} + a_m \frac{\frac{\partial^2 \theta_{m-1}}{\partial \boldsymbol{\Omega}_1^2} f_m(\theta_{m-1}) + \frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_1}\frac{\partial f_m(\theta_{m-1})}{\partial \boldsymbol{\Omega}_1}}{1-F_m(\theta_{m-1})} \\
&+ a_m \frac{\left[\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_1} f_m(\theta_{m-1})\right]^2}{\left[1-F_m(\theta_{m-1})\right]^2} + \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1^2}\sum_{i=1}^{n}\Delta(x_i-\theta_1)\log f_1(x_i) \\
&+ 2\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}\sum_{i=1}^{n}\Delta(x_i-\theta_1)\frac{\frac{\partial f_1(x_i)}{\partial \boldsymbol{\Omega}_1}}{f_1(x_i)} + \sum_{x_i\le\theta_1}\frac{\frac{\partial^2 f_1(x_i)}{\partial \boldsymbol{\Omega}_1^2} f_1(x_i) - \left[\frac{\partial f_1(x_i)}{\partial \boldsymbol{\Omega}_1}\right]^2}{\left[f_1(x_i)\right]^2} \\
&+ \sum_{i=2}^{m-1}\left[\frac{\partial^2 a_i}{\partial \boldsymbol{\Omega}_1^2}\log\phi_i + 2\frac{\partial a_i}{\partial \boldsymbol{\Omega}_1}\frac{\frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_1}}{\phi_i} + a_i\frac{\frac{\partial^2 \phi_i}{\partial \boldsymbol{\Omega}_1^2}\phi_i - \left(\frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_1}\right)^2}{\phi_i^2} - \frac{\partial^2 \theta_{i-1}}{\partial \boldsymbol{\Omega}_1^2}\sum_{x_i\le\theta_i}\Delta(x_i-\theta_{i-1})\log f_i(x_i)\right] \\
&+ \sum_{i=2}^{m-1}\left[\frac{\partial^2 \theta_i}{\partial \boldsymbol{\Omega}_1^2}\sum_{\theta_{i-1}<x_i}\Delta(x_i-\theta_i)\log f_i(x_i) - 2\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_1}\frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_1}\sum_{i=1}^{n}\Delta(x_i-\theta_{i-1})\Delta(x_i-\theta_i)\log f_i(x_i)\right] \\
&+ \frac{\partial^2 a_m}{\partial \boldsymbol{\Omega}_1^2}\log\phi_m + 2\frac{\partial a_m}{\partial \boldsymbol{\Omega}_1}\frac{\frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_1}}{\phi_m} + a_m\frac{\frac{\partial^2 \phi_m}{\partial \boldsymbol{\Omega}_1^2}\phi_m - \left(\frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_1}\right)^2}{\phi_m^2} - \frac{\partial^2 \theta_{m-1}}{\partial \boldsymbol{\Omega}_1^2}\sum_{i=1}^{n}\Delta(x_i-\theta_{m-1})\log f_m(x_i),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_1 \boldsymbol{\Omega}_i} &= \frac{n}{\zeta}\frac{\partial^2 \zeta}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_i} - \frac{n}{\zeta^2}\frac{\partial \zeta}{\partial \boldsymbol{\Omega}_1}\frac{\partial \zeta}{\partial \boldsymbol{\Omega}_i} - \frac{\partial^2 a_1}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_i}\log[F_1(\theta_1)] - \frac{\partial a_1}{\partial \boldsymbol{\Omega}_1}\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_i}\frac{f_1(\theta_1)}{F_1(\theta_1)} \\
&+ \frac{\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_i} a_1 [f_1(\theta_1)]^2}{[F_1(\theta_1)]^2} - \frac{\frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_i} a_1 f_1(\theta_1) + \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1}\frac{\partial a_1}{\partial \boldsymbol{\Omega}_i} f_1(\theta_1) + \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_1} a_1 \frac{\partial f_1(\theta_1)}{\partial \boldsymbol{\Omega}_i}}{F_1(\theta_1)} \\
&- \sum_{i=2}^{m-1}\frac{\partial^2 a_i}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_i}\log[F_i(\theta_i) - F_i(\theta_{i-1})] \\
&- \sum_{i=2}^{m-1}\frac{\partial a_i}{\partial \boldsymbol{\Omega}_1}\frac{f_i(\theta_i)\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_i} - f_i(\theta_{i-1})\frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_i}}{F_i(\theta_i) - F_i(\theta_{i-1})} - \sum_{i=2}^{m-1}\frac{\partial a_i}{\partial \boldsymbol{\Omega}_i}\frac{f_i(\theta_i)\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_1} - f_i(\theta_{i-1})\frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_1}}{F_i(\theta_i) - F_i(\theta_{i-1})} \\
&- \sum_{i=2}^{m-1} a_i \frac{\left[\frac{\partial^2 \theta_i}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_i} f_i(\theta_i) + \frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_1}\frac{\partial f_i(\theta_i)}{\partial \boldsymbol{\Omega}_i} - \frac{\partial^2 \theta_{i-1}}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_i} f_i(\theta_{i-1}) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_1}\frac{\partial f_i(\theta_{i-1})}{\partial \boldsymbol{\Omega}_i}\right]}{F_i(\theta_i) - F_i(\theta_{i-1})} \\
&+ \sum_{i=2}^{m-1} a_i \frac{\left[\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_1} f_i(\theta_i) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_1} f_i(\theta_{i-1})\right]\left[\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_i} f_i(\theta_i) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_i} f_i(\theta_{i-1})\right]}{[F_i(\theta_i) - F_i(\theta_{i-1})]^2} \\
&- \frac{\partial^2 a_m}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_i}\log[1 - F_m(\theta_{m-1})] + \frac{\partial a_m}{\partial \boldsymbol{\Omega}_1}\frac{\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_i} f_m(\theta_{m-1})}{1 - F_m(\theta_{m-1})} + \frac{\partial a_m}{\partial \boldsymbol{\Omega}_i}\frac{\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_1} f_m(\theta_{m-1})}{1 - F_m(\theta_{m-1})} \\
&+ a_m \frac{\frac{\partial^2 \theta_{m-1}}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_i} f_m(\theta_{m-1}) + \frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_1}\frac{\partial f_m(\theta_{m-1})}{\partial \boldsymbol{\Omega}_i}}{1 - F_m(\theta_{m-1})} \\
&+ a_m \frac{\left[\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_1} f_m(\theta_{m-1})\right]\left[\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_i} f_m(\theta_{m-1})\right]}{[1 - F_m(\theta_{m-1})]^2} \\
&+ \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_1 \partial \boldsymbol{\Omega}_i}\sum_{i=1}^{n}\Delta(x_i-\theta_1)\log f_1(x_i) + \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_i}\sum_{i=1}^{n}\Delta(x_i-\theta_1)\frac{\frac{\partial f_1(x_i)}{\partial \boldsymbol{\Omega}_1}}{f_1(x_i)}
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{i=2}^{m-1}\left[\frac{\partial^2 a_i}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_i}\log\phi_i+\frac{\partial a_i}{\partial\boldsymbol{\Omega}_1}\frac{\frac{\partial\phi_i}{\partial\boldsymbol{\Omega}_i}}{\phi_i}+\frac{\partial a_i}{\partial\boldsymbol{\Omega}_i}\frac{\frac{\partial\phi_i}{\partial\boldsymbol{\Omega}_1}}{\phi_i}+a_i\frac{\frac{\partial^2\phi_i}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_i}\phi_i-\frac{\partial\phi_i}{\partial\boldsymbol{\Omega}_1}\frac{\partial\phi_i}{\partial\boldsymbol{\Omega}_i}}{\phi_i^2}\right]\\
&+\sum_{i=2}^{m-1}\left[\frac{\partial^2\theta_i}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_i}\sum_{\theta_{i-1}<x_i}\Delta(x_i-\theta_i)\log f_i(x_i)-\frac{\partial^2\theta_{i-1}}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_i}\sum_{x_i\le\theta_i}\Delta(x_i-\theta_{i-1})\log f_i(x_i)\right]\\
&-\sum_{i=2}^{m-1}\left[\left(\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_1}\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_i}+\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_i}\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_1}\right)\sum_{i=1}^{n}\Delta(x_i-\theta_{i-1})\Delta(x_i-\theta_i)\log f_i(x_i)\right]\\
&+\sum_{i=2}^{m-1}\left[\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_1}\sum_{\theta_{i-1}\le x_i}\Delta(x_i-\theta_i)\frac{\frac{\partial f_i(x_i)}{\partial\boldsymbol{\Omega}_i}}{f_i(x_i)}-\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_1}\sum_{x_i\le\theta_i}\Delta(x_i-\theta_{i-1})\frac{\frac{\partial f_i(x_i)}{\partial\boldsymbol{\Omega}_i}}{f_i(x_i)}\right]\\
&+\frac{\partial^2 a_m}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_i}\log\phi_m+\frac{\partial a_m}{\partial\boldsymbol{\Omega}_1}\frac{\frac{\partial\phi_m}{\partial\boldsymbol{\Omega}_i}}{\phi_m}+\frac{\partial a_m}{\partial\boldsymbol{\Omega}_i}\frac{\frac{\partial\phi_m}{\partial\boldsymbol{\Omega}_1}}{\phi_m}+a_m\frac{\frac{\partial^2\phi_m}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_i}\phi_m-\frac{\partial\phi_m}{\partial\boldsymbol{\Omega}_1}\frac{\partial\phi_m}{\partial\boldsymbol{\Omega}_i}}{\phi_m^2}\\
&-\frac{\partial^2\theta_{m-1}}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_i}\sum_{i=1}^{n}\Delta(x_i-\theta_{m-1})\log f_m(x_i)
\end{aligned}
$$

for $2\le i\le m-1$,

$$
\begin{aligned}
\frac{\partial^2\log L}{\partial\boldsymbol{\Omega}_1\boldsymbol{\Omega}_m}=\;&\frac{n}{\zeta}\frac{\partial^2\zeta}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}-\frac{n}{\zeta^2}\frac{\partial\zeta}{\partial\boldsymbol{\Omega}_1}\frac{\partial\zeta}{\partial\boldsymbol{\Omega}_m}-\frac{\partial^2 a_1}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}\log[F_1(\theta_1)]-\frac{\partial a_1}{\partial\boldsymbol{\Omega}_1}\frac{\partial\theta_1}{\partial\boldsymbol{\Omega}_m}\frac{f_1(\theta_1)}{F_1(\theta_1)}\\
&+\frac{\frac{\partial\theta_1}{\partial\boldsymbol{\Omega}_1}\frac{\partial\theta_1}{\partial\boldsymbol{\Omega}_m}a_1[f_1(\theta_1)]^2}{[F_1(\theta_1)]^2}-\frac{\frac{\partial^2\theta_1}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}a_1f_1(\theta_1)+\frac{\partial\theta_1}{\partial\boldsymbol{\Omega}_1}\frac{\partial a_1}{\partial\boldsymbol{\Omega}_m}f_1(\theta_1)+\frac{\partial\theta_1}{\partial\boldsymbol{\Omega}_1}a_1\frac{\partial f_1(\theta_1)}{\partial\boldsymbol{\Omega}_m}}{F_1(\theta_1)}\\
&-\sum_{i=2}^{m-1}\frac{\partial^2 a_i}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}\log[F_i(\theta_i)-F_i(\theta_{i-1})]\\
&-\sum_{i=2}^{m-1}\frac{\partial a_i}{\partial\boldsymbol{\Omega}_1}\frac{f_i(\theta_i)\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_m}-f_i(\theta_{i-1})\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_m}}{F_i(\theta_i)-F_i(\theta_{i-1})}-\sum_{i=2}^{m-1}\frac{\partial a_i}{\partial\boldsymbol{\Omega}_m}\frac{f_i(\theta_i)\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_1}-f_i(\theta_{i-1})\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_1}}{F_i(\theta_i)-F_i(\theta_{i-1})}\\
&-\sum_{i=2}^{m-1}a_i\frac{\left[\frac{\partial^2\theta_i}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}f_i(\theta_i)+\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_1}\frac{\partial f_i(\theta_i)}{\partial\boldsymbol{\Omega}_m}-\frac{\partial^2\theta_{i-1}}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}f_i(\theta_{i-1})-\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_1}\frac{\partial f_i(\theta_{i-1})}{\partial\boldsymbol{\Omega}_m}\right]}{F_i(\theta_i)-F_i(\theta_{i-1})}\\
&+\sum_{i=2}^{m-1}a_i\frac{\left[\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_1}f_i(\theta_i)-\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_1}f_i(\theta_{i-1})\right]\left[\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_m}f_i(\theta_i)-\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_m}f_i(\theta_{i-1})\right]}{[F_i(\theta_i)-F_i(\theta_{i-1})]^2}\\
&-\frac{\partial^2 a_m}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}\log[1-F_m(\theta_{m-1})]+\frac{\partial a_m}{\partial\boldsymbol{\Omega}_1}\frac{\frac{\partial\theta_{m-1}}{\partial\boldsymbol{\Omega}_m}f_m(\theta_{m-1})}{1-F_m(\theta_{m-1})}+\frac{\partial a_m}{\partial\boldsymbol{\Omega}_m}\frac{\frac{\partial\theta_{m-1}}{\partial\boldsymbol{\Omega}_1}f_m(\theta_{m-1})}{1-F_m(\theta_{m-1})}\\
&+a_m\frac{\frac{\partial^2\theta_{m-1}}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}f_m(\theta_{m-1})+\frac{\partial\theta_{m-1}}{\partial\boldsymbol{\Omega}_1}\frac{\partial f_m(\theta_{m-1})}{\partial\boldsymbol{\Omega}_m}}{1-F_m(\theta_{m-1})}\\
&+a_m\frac{\left[\frac{\partial\theta_{m-1}}{\partial\boldsymbol{\Omega}_1}f_m(\theta_{m-1})\right]\left[\frac{\partial\theta_{m-1}}{\partial\boldsymbol{\Omega}_m}f_m(\theta_{m-1})\right]}{[1-F_m(\theta_{m-1})]^2}\\
&+\frac{\partial^2\theta_1}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}\sum_{i=1}^{n}\Delta(x_i-\theta_1)\log f_1(x_i)+\frac{\partial\theta_1}{\partial\boldsymbol{\Omega}_m}\sum_{i=1}^{n}\Delta(x_i-\theta_1)\frac{\frac{\partial f_1(x_i)}{\partial\boldsymbol{\Omega}_1}}{f_1(x_i)}\\
&+\sum_{i=2}^{m-1}\left[\frac{\partial^2 a_i}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}\log\phi_i+\frac{\partial a_i}{\partial\boldsymbol{\Omega}_1}\frac{\frac{\partial\phi_i}{\partial\boldsymbol{\Omega}_m}}{\phi_i}+\frac{\partial a_i}{\partial\boldsymbol{\Omega}_m}\frac{\frac{\partial\phi_i}{\partial\boldsymbol{\Omega}_1}}{\phi_i}+a_i\frac{\frac{\partial^2\phi_i}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}\phi_i-\frac{\partial\phi_i}{\partial\boldsymbol{\Omega}_1}\frac{\partial\phi_i}{\partial\boldsymbol{\Omega}_m}}{\phi_i^2}\right]
\end{aligned}
$$

$$+\sum_{i=2}^{m-1}\left[\frac{\partial^2\theta_i}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}\sum_{\theta_{i-1}<x_i}\Delta(x_i-\theta_i)\log f_i(x_i)-\frac{\partial^2\theta_{i-1}}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}\sum_{x_i\le\theta_i}\Delta(x_i-\theta_{i-1})\log f_i(x_i)\right]$$
$$-\sum_{i=2}^{m-1}\left[\left(\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_1}\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_m}+\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_m}\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_1}\right)\sum_{i=1}^{n}\Delta(x_i-\theta_{i-1})\Delta(x_i-\theta_i)\log f_i(x_i)\right]$$
$$+\frac{\partial^2 a_m}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}\log\phi_m+\frac{\partial a_m}{\partial\boldsymbol{\Omega}_1}\frac{\frac{\partial\phi_m}{\partial\boldsymbol{\Omega}_m}}{\phi_m}+\frac{\partial a_m}{\partial\boldsymbol{\Omega}_m}\frac{\frac{\partial\phi_m}{\partial\boldsymbol{\Omega}_1}}{\phi_m}+a_m\frac{\frac{\partial^2\phi_m}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}\phi_m-\frac{\partial\phi_m}{\partial\boldsymbol{\Omega}_1}\frac{\partial\phi_m}{\partial\boldsymbol{\Omega}_m}}{\phi_m^2}$$
$$-\frac{\partial^2\theta_{m-1}}{\partial\boldsymbol{\Omega}_1\partial\boldsymbol{\Omega}_m}\sum_{i=1}^{n}\Delta(x_i-\theta_{m-1})\log f_m(x_i)-\frac{\partial\theta_{m-1}}{\partial\boldsymbol{\Omega}_1}\sum_{i=1}^{n}\Delta(x_i-\theta_{m-1})\frac{\frac{\partial f_m(x_i)}{\partial\boldsymbol{\Omega}_m}}{f_m(x_i)}$$

for $2\le i\le m-1$,

$$\frac{\partial^2\log L}{\partial\boldsymbol{\Omega}_i^2}=\frac{n}{\zeta}\frac{\partial^2\zeta}{\partial\boldsymbol{\Omega}_i^2}-\frac{n}{\zeta^2}\left(\frac{\partial\zeta}{\partial\boldsymbol{\Omega}_i}\right)^2-\frac{\partial^2 a_i}{\partial\boldsymbol{\Omega}_i^2}\log F_1(\theta_1)-\frac{\partial a_1}{\partial\boldsymbol{\Omega}_i}\frac{\partial\theta_1}{\partial\boldsymbol{\Omega}_i}\frac{f_1(\theta_1)}{F_1(\theta_1)}+\frac{\left(\frac{\partial\theta_1}{\partial\boldsymbol{\Omega}_i}\right)^2 a_1[f_1(\theta_1)]^2}{[F_1(\theta_1)]^2}$$
$$-\frac{\frac{\partial^2\theta_1}{\partial\boldsymbol{\Omega}_i^2}a_1f_1(\theta_1)+\frac{\partial\theta_1}{\partial\boldsymbol{\Omega}_i}\frac{\partial a_1}{\partial\boldsymbol{\Omega}_i}f_1(\theta_1)+\frac{\partial\theta_1}{\partial\boldsymbol{\Omega}_i}a_1\frac{\partial f_1(\theta_1)}{\partial\boldsymbol{\Omega}_i}}{F_1(\theta_1)}-\sum_{i=2}^{m-1}\frac{\partial^2 a_i}{\partial\boldsymbol{\Omega}_i^2}\log[F_i(\theta_i)-F_i(\theta_{i-1})]$$
$$-2\sum_{i=2}^{m-1}\frac{\partial a_i}{\partial\boldsymbol{\Omega}_i}\frac{f_i(\theta_i)\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_i}-f_i(\theta_{i-1})\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_i}}{F_i(\theta_i)-F_i(\theta_{i-1})}$$
$$-\sum_{i=2}^{m-1}a_i\frac{\left[\frac{\partial^2\theta_i}{\partial\boldsymbol{\Omega}_i^2}f_i(\theta_i)+\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_i}\frac{\partial f_i(\theta_i)}{\partial\boldsymbol{\Omega}_i}-\frac{\partial^2\theta_{i-1}}{\partial\boldsymbol{\Omega}_i^2}f_i(\theta_{i-1})-\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_i}\frac{\partial f_i(\theta_{i-1})}{\partial\boldsymbol{\Omega}_i}\right]}{F_i(\theta_i)-F_i(\theta_{i-1})}$$
$$+\sum_{i=2}^{m-1}a_i\frac{\left[\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_i}f_i(\theta_i)-\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_i}f_i(\theta_{i-1})\right]^2}{[F_i(\theta_i)-F_i(\theta_{i-1})]^2}-\frac{\partial^2 a_m}{\partial\boldsymbol{\Omega}_i^2}\log[1-F_m(\theta_{m-1})]$$
$$+2\frac{\partial a_m}{\partial\boldsymbol{\Omega}_i}\frac{\frac{\partial\theta_{m-1}}{\partial\boldsymbol{\Omega}_i}f_m(\theta_{m-1})}{1-F_m(\theta_{m-1})}+a_m\frac{\frac{\partial^2\theta_{m-1}}{\partial\boldsymbol{\Omega}_i^2}f_m(\theta_{m-1})+\frac{\partial\theta_{m-1}}{\partial\boldsymbol{\Omega}_i}\frac{\partial f_m(\theta_{m-1})}{\partial\boldsymbol{\Omega}_i}}{1-F_m(\theta_{m-1})}$$
$$+a_m\frac{\left[\frac{\partial\theta_{m-1}}{\partial\boldsymbol{\Omega}_i}f_m(\theta_{m-1})\right]^2}{[1-F_m(\theta_{m-1})]^2}+\frac{\partial^2\theta_1}{\partial\boldsymbol{\Omega}_i^2}\sum_{i=1}^{n}\Delta(x_i-\theta_1)\log f_1(x_i)$$
$$-2\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_i}\sum_{i=2}^{m-1}\sum_{x_i\le\theta_i}\Delta(x_i-\theta_{i-1})\frac{\frac{\partial f_i(x_i)}{\partial\boldsymbol{\Omega}_i}}{f_i(x_i)}+2\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_i}\sum_{i=2}^{m-1}\sum_{\theta_{i-1}\le x_i}\Delta(x_i-\theta_i)\frac{\frac{\partial f_i(x_i)}{\partial\boldsymbol{\Omega}_i}}{f_i(x_i)}$$
$$+\sum_{i=2}^{m-1}\sum_{\theta_{i-1}\le x_i\le\theta_i}\frac{\frac{\partial^2 f_i(x_i)}{\partial\boldsymbol{\Omega}_i^2}f_i(x_i)-\left[\frac{\partial f_i(x_i)}{\partial\boldsymbol{\Omega}_i}\right]^2}{[f_i(x_i)]^2}$$
$$+\sum_{i=2}^{m-1}\left[\frac{\partial^2 a_i}{\partial\boldsymbol{\Omega}_i^2}\log\phi_i+2\frac{\partial a_i}{\partial\boldsymbol{\Omega}_i}\frac{\frac{\partial\phi_i}{\partial\boldsymbol{\Omega}_i}}{\phi_i}+a_i\frac{\frac{\partial^2\phi_i}{\partial\boldsymbol{\Omega}_i^2}\phi_i-\left(\frac{\partial\phi_i}{\partial\boldsymbol{\Omega}_i}\right)^2}{\phi_i^2}-\frac{\partial^2\theta_{i-1}}{\partial\boldsymbol{\Omega}_i^2}\sum_{x_i\le\theta_i}\Delta(x_i-\theta_{i-1})\log f_i(x_i)\right]$$
$$+\sum_{i=2}^{m-1}\left[\frac{\partial^2\theta_i}{\partial\boldsymbol{\Omega}_i^2}\sum_{\theta_{i-1}<x_i}\Delta(x_i-\theta_i)\log f_i(x_i)-2\frac{\partial\theta_i}{\partial\boldsymbol{\Omega}_i}\frac{\partial\theta_{i-1}}{\partial\boldsymbol{\Omega}_i}\sum_{i=1}^{n}\Delta(x_i-\theta_{i-1})\Delta(x_i-\theta_i)\log f_i(x_i)\right]$$
$$+\frac{\partial^2 a_m}{\partial\boldsymbol{\Omega}_i^2}\log\phi_m+2\frac{\partial a_m}{\partial\boldsymbol{\Omega}_i}\frac{\frac{\partial\phi_m}{\partial\boldsymbol{\Omega}_i}}{\phi_m}+a_m\frac{\frac{\partial^2\phi_m}{\partial\boldsymbol{\Omega}_i^2}\phi_m-\left(\frac{\partial\phi_m}{\partial\boldsymbol{\Omega}_i}\right)^2}{\phi_m^2}-\frac{\partial^2\theta_{m-1}}{\partial\boldsymbol{\Omega}_i^2}\sum_{i=1}^{n}\Delta(x_i-\theta_{m-1})\log f_m(x_i)$$

for $2\le i\le m-1$,

$$
\begin{aligned}
\frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_i \boldsymbol{\Omega}_j} = {} & \frac{n}{\zeta}\frac{\partial^2 \zeta}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j} - \frac{n}{\zeta^2}\frac{\partial \zeta}{\partial \boldsymbol{\Omega}_i}\frac{\partial \zeta}{\partial \boldsymbol{\Omega}_j} - \frac{\partial^2 a_1}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j}\log[F_1(\theta_1)] - \frac{\partial a_1}{\partial \boldsymbol{\Omega}_i}\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_j}\frac{f_1(\theta_1)}{F_1(\theta_1)} \\
& + \frac{\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_i}\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_j} a_1 [f_1(\theta_1)]^2}{[F_1(\theta_1)]^2} - \frac{\frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j} a_1 f_1(\theta_1) + \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_i}\frac{\partial a_1}{\partial \boldsymbol{\Omega}_j} f_1(\theta_1) + \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_i} a_1 \frac{\partial f_1(\theta_1)}{\partial \boldsymbol{\Omega}_j}}{F_1(\theta_1)} \\
& - \sum_{i=2}^{m-1} \frac{\partial^2 a_i}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j}\log[F_i(\theta_i) - F_i(\theta_{i-1})] \\
& - \sum_{i=2}^{m-1} \frac{\partial a_i}{\partial \boldsymbol{\Omega}_j}\frac{f_i(\theta_i)\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_i} - f_i(\theta_{i-1})\frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_i}}{F_i(\theta_i) - F_i(\theta_{i-1})} - \sum_{i=2}^{m-1} \frac{\partial a_i}{\partial \boldsymbol{\Omega}_i}\frac{f_i(\theta_i)\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_j} - f_i(\theta_{i-1})\frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_j}}{F_i(\theta_i) - F_i(\theta_{i-1})} \\
& - \sum_{i=2}^{m-1} a_i \frac{\left[\frac{\partial^2 \theta_i}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j} f_i(\theta_i) + \frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_i}\frac{\partial f_i(\theta_i)}{\partial \boldsymbol{\Omega}_j} - \frac{\partial^2 \theta_{i-1}}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j} f_i(\theta_{i-1}) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_i}\frac{\partial f_i(\theta_{i-1})}{\partial \boldsymbol{\Omega}_j}\right]}{F_i(\theta_i) - F_i(\theta_{i-1})} \\
& + \sum_{i=2}^{m-1} a_i \frac{\left[\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_i} f_i(\theta_i) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_i} f_i(\theta_{i-1})\right]\left[\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_j} f_i(\theta_i) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_j} f_i(\theta_{i-1})\right]}{[F_i(\theta_i) - F_i(\theta_{i-1})]^2} \\
& - \frac{\partial^2 a_m}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j}\log[1 - F_m(\theta_{m-1})] + \frac{\partial a_m}{\partial \boldsymbol{\Omega}_i}\frac{\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_j} f_m(\theta_{m-1})}{1 - F_m(\theta_{m-1})} + \frac{\partial a_m}{\partial \boldsymbol{\Omega}_j}\frac{\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_i} f_m(\theta_{m-1})}{1 - F_m(\theta_{m-1})} \\
& + a_m \frac{\frac{\partial^2 \theta_{m-1}}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j} f_m(\theta_{m-1}) + \frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_i}\frac{\partial f_m(\theta_{m-1})}{\partial \boldsymbol{\Omega}_j}}{1 - F_m(\theta_{m-1})} \\
& + a_m \frac{\left[\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_i} f_m(\theta_{m-1})\right]\left[\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_j} f_m(\theta_{m-1})\right]}{[1 - F_m(\theta_{m-1})]^2} + \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j}\sum_{i=1}^{n} \Delta(x_i - \theta_1)\log f_1(x_i) \\
& + \sum_{i=2}^{m-1}\left[\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_j}\sum_{\theta_{i-1} \le x_i} \Delta(x_i - \theta_i)\frac{\frac{\partial f_i(x_i)}{\partial \boldsymbol{\Omega}_i}}{f_i(x_i)} - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_j}\sum_{x_i \le \theta_i} \Delta(x_i - \theta_{i-1})\frac{\frac{\partial f_i(x_i)}{\partial \boldsymbol{\Omega}_i}}{f_i(x_i)}\right] \\
& + \sum_{i=2}^{m-1}\sum_{\theta_{i-1} \le x_i \le \theta_i} \frac{\frac{\partial^2 f_i(x_i)}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j} f_i(x_i) - \frac{\partial f_i(x_i)}{\partial \boldsymbol{\Omega}_i}\frac{\partial f_i(x_i)}{\partial \boldsymbol{\Omega}_j}}{[f_i(x_i)]^2} \\
& + \sum_{i=2}^{m-1}\left[\frac{\partial^2 a_i}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j}\log \phi_i + \frac{\partial a_i}{\partial \boldsymbol{\Omega}_i}\frac{\frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_j}}{\phi_i} + \frac{\partial a_i}{\partial \boldsymbol{\Omega}_j}\frac{\frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_i}}{\phi_i} + a_i \frac{\frac{\partial^2 \phi_i}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j}\phi_i - \frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_i}\frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_j}}{\phi_i^2}\right] \\
& + \sum_{i=2}^{m-1}\left[\frac{\partial^2 \theta_i}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j}\sum_{\theta_{i-1} \le x_i} \Delta(x_i - \theta_i)\log f_i(x_i) - \frac{\partial^2 \theta_{i-1}}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j}\sum_{x_i \le \theta_i} \Delta(x_i - \theta_{i-1})\log f_i(x_i)\right] \\
& - \sum_{i=2}^{m-1}\left[\left(\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_i}\frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_j} + \frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_j}\frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_i}\right)\sum_{i=1}^{n} \Delta(x_i - \theta_{i-1})\Delta(x_i - \theta_i)\log f_i(x_i)\right] \\
& + \sum_{i=2}^{m-1}\left[\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_i}\sum_{\theta_{i-1} \le x_i} \Delta(x_i - \theta_i)\frac{\frac{\partial f_i(x_i)}{\partial \boldsymbol{\Omega}_j}}{f_i(x_i)} - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_i}\sum_{x_i \le \theta_i} \Delta(x_i - \theta_{i-1})\frac{\frac{\partial f_i(x_i)}{\partial \boldsymbol{\Omega}_j}}{f_i(x_i)}\right] \\
& + \frac{\partial^2 a_m}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j}\log \phi_m + \frac{\partial a_m}{\partial \boldsymbol{\Omega}_i}\frac{\frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_j}}{\phi_m} + \frac{\partial a_m}{\partial \boldsymbol{\Omega}_j}\frac{\frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_i}}{\phi_m} + a_m \frac{\frac{\partial^2 \phi_m}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j}\phi_m - \frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_i}\frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_j}}{\phi_m^2} \\
& - \frac{\partial^2 \theta_{m-1}}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_j}\sum_{i=1}^{n} \Delta(x_i - \theta_{m-1})\log f_m(x_i)
\end{aligned}
$$

for $2 \le i < j \le m-1$,

$$
\begin{aligned}
\frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_i \boldsymbol{\Omega}_m} &= \frac{n}{\zeta}\frac{\partial^2 \zeta}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m} - \frac{n}{\zeta^2}\frac{\partial \zeta}{\partial \boldsymbol{\Omega}_i}\frac{\partial \zeta}{\partial \boldsymbol{\Omega}_m} - \frac{\partial^2 a_1}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m}\log[F_1(\theta_1)] - \frac{\partial a_1}{\partial \boldsymbol{\Omega}_i}\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_m}\frac{f_1(\theta_1)}{F_1(\theta_1)} \\
&+ \frac{\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_i}\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_m} a_1 [f_1(\theta_1)]^2}{[F_1(\theta_1)]^2} - \frac{\frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m} a_1 f_1(\theta_1) + \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_i}\frac{\partial a_1}{\partial \boldsymbol{\Omega}_m} f_1(\theta_1) + \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_i} a_1 \frac{\partial f_1(\theta_1)}{\partial \boldsymbol{\Omega}_m}}{F_1(\theta_1)} \\
&- \sum_{i=2}^{m-1} \frac{\partial^2 a_i}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m}\log[F_i(\theta_i) - F_i(\theta_{i-1})] \\
&- \sum_{i=2}^{m-1} \frac{\partial a_i}{\partial \boldsymbol{\Omega}_m}\frac{f_i(\theta_i)\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_i} - f_i(\theta_{i-1})\frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_i}}{F_i(\theta_i) - F_i(\theta_{i-1})} - \sum_{i=2}^{m-1} \frac{\partial a_i}{\partial \boldsymbol{\Omega}_i}\frac{f_i(\theta_i)\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_m} - f_i(\theta_{i-1})\frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_m}}{F_i(\theta_i) - F_i(\theta_{i-1})} \\
&- \sum_{i=2}^{m-1} a_i \frac{\left[\frac{\partial^2 \theta_i}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m} f_i(\theta_i) + \frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_i}\frac{\partial f_i(\theta_i)}{\partial \boldsymbol{\Omega}_m} - \frac{\partial^2 \theta_{i-1}}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m} f_i(\theta_{i-1}) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_i}\frac{\partial f_i(\theta_{i-1})}{\partial \boldsymbol{\Omega}_m}\right]}{F_i(\theta_i) - F_i(\theta_{i-1})} \\
&+ \sum_{i=2}^{m-1} a_i \frac{\left[\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_i} f_i(\theta_i) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_i} f_i(\theta_{i-1})\right]\left[\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_m} f_i(\theta_i) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_m} f_i(\theta_{i-1})\right]}{[F_i(\theta_i) - F_i(\theta_{i-1})]^2} \\
&- \frac{\partial^2 a_m}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m}\log[1 - F_m(\theta_{m-1})] + \frac{\partial a_m}{\partial \boldsymbol{\Omega}_i}\frac{\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_m} f_m(\theta_{m-1})}{1 - F_m(\theta_{m-1})} + \frac{\partial a_m}{\partial \boldsymbol{\Omega}_m}\frac{\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_i} f_m(\theta_{m-1})}{1 - F_m(\theta_{m-1})} \\
&+ a_m \frac{\frac{\partial^2 \theta_{m-1}}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m} f_m(\theta_{m-1}) + \frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_i}\frac{\partial f_m(\theta_{m-1})}{\partial \boldsymbol{\Omega}_m}}{1 - F_m(\theta_{m-1})} \\
&+ a_m \frac{\left[\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_i} f_m(\theta_{m-1})\right]\left[\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_m} f_m(\theta_{m-1})\right]}{[1 - F_m(\theta_{m-1})]^2} \\
&+ \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m}\sum_{i=1}^{n} \Delta(x_i - \theta_1)\log f_1(x_i) \\
&+ \sum_{i=2}^{m-1}\left[\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_m}\sum_{\theta_{i-1} \le x_i} \Delta(x_i - \theta_i)\frac{\frac{\partial f_i(x_i)}{\partial \boldsymbol{\Omega}_i}}{f_i(x_i)} - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_m}\sum_{x_i \le \theta_i} \Delta(x_i - \theta_{i-1})\frac{\frac{\partial f_i(x_i)}{\partial \boldsymbol{\Omega}_i}}{f_i(x_i)}\right] \\
&+ \sum_{i=2}^{m-1}\left(\frac{\partial^2 a_i}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m}\log \phi_i + \frac{\partial a_i}{\partial \boldsymbol{\Omega}_i}\frac{\frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_m}}{\phi_i} + \frac{\partial a_i}{\partial \boldsymbol{\Omega}_m}\frac{\frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_i}}{\phi_i} + a_i \frac{\frac{\partial^2 \phi_i}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m}\phi_i - \frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_i}\frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_m}}{\phi_i^2}\right) \\
&+ \sum_{i=2}^{m-1}\left[\frac{\partial^2 \theta_i}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m}\sum_{\theta_{i-1} \le x_i} \Delta(x_i - \theta_i)\log f_i(x_i) - \frac{\partial^2 \theta_{i-1}}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m}\sum_{x_i \le \theta_i} \Delta(x_i - \theta_{i-1})\log f_i(x_i)\right] \\
&- \sum_{i=2}^{m-1}\left[\left(\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_i}\frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_m} + \frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_m}\frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_i}\right)\sum_{i=1}^{n} \Delta(x_i - \theta_{i-1})\Delta(x_i - \theta_i)\log f_i(x_i)\right] \\
&+ \frac{\partial^2 a_m}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m}\log \phi_m + \frac{\partial a_m}{\partial \boldsymbol{\Omega}_i}\frac{\frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_m}}{\phi_m} + \frac{\partial a_m}{\partial \boldsymbol{\Omega}_m}\frac{\frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_i}}{\phi_m} + a_m \frac{\frac{\partial^2 \phi_m}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m}\phi_m - \frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_i}\frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_m}}{\phi_m^2} \\
&- \frac{\partial^2 \theta_{m-1}}{\partial \boldsymbol{\Omega}_i \partial \boldsymbol{\Omega}_m}\sum_{i=1}^{n} \Delta(x_i - \theta_{m-1})\log f_m(x_i) - \frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_i}\sum_{i=1}^{n} \Delta(x_i - \theta_{m-1})\frac{\frac{\partial f_m(x_i)}{\partial \boldsymbol{\Omega}_m}}{f_m(x_i)}
\end{aligned}
$$

for $2 \le i \le m-1$, and

$$
\begin{aligned}
\frac{\partial^2 \log L}{\partial \boldsymbol{\Omega}_m^2} = {} & \frac{n}{\zeta}\frac{\partial^2 \zeta}{\partial \boldsymbol{\Omega}_m^2} - \frac{n}{\zeta^2}\left(\frac{\partial \zeta}{\partial \boldsymbol{\Omega}_m}\right)^2 - \frac{\partial^2 a_1}{\partial \boldsymbol{\Omega}_m^2}\log F_1(\theta_1) - \frac{\partial a_1}{\partial \boldsymbol{\Omega}_m}\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_m}\frac{f_1(\theta_1)}{F_1(\theta_1)} \\
& + \frac{\left(\frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_m}\right)^2 a_1 [f_1(\theta_1)]^2}{[F_1(\theta_1)]^2} - \frac{\frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_m^2} a_1 f_1(\theta_1) + \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_m}\frac{\partial a_1}{\partial \boldsymbol{\Omega}_m} f_1(\theta_1) + \frac{\partial \theta_1}{\partial \boldsymbol{\Omega}_m} a_1 \frac{\partial f_1(\theta_1)}{\partial \boldsymbol{\Omega}_m}}{F_1(\theta_1)} \\
& - \sum_{i=2}^{m-1} \frac{\partial^2 a_i}{\partial \boldsymbol{\Omega}_m^2}\log[F_i(\theta_i) - F_i(\theta_{i-1})] - 2\sum_{i=2}^{m-1}\frac{\partial a_i}{\partial \boldsymbol{\Omega}_m}\frac{f_i(\theta_i)\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_m} - f_i(\theta_{i-1})\frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_m}}{F_i(\theta_i) - F_i(\theta_{i-1})} \\
& - \sum_{i=2}^{m-1} a_i \frac{\left[\frac{\partial^2 \theta_i}{\partial \boldsymbol{\Omega}_m^2} f_i(\theta_i) + \frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_m}\frac{\partial f_i(\theta_i)}{\partial \boldsymbol{\Omega}_m} - \frac{\partial^2 \theta_{i-1}}{\partial \boldsymbol{\Omega}_m^2} f_i(\theta_{i-1}) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_m}\frac{\partial f_i(\theta_{i-1})}{\partial \boldsymbol{\Omega}_m}\right]}{F_i(\theta_i) - F_i(\theta_{i-1})} \\
& + \sum_{i=2}^{m-1} a_i \frac{\left[\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_m} f_i(\theta_i) - \frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_m} f_i(\theta_{i-1})\right]^2}{[F_i(\theta_i) - F_i(\theta_{i-1})]^2} \\
& - \frac{\partial^2 a_m}{\partial \boldsymbol{\Omega}_m^2}\log[1 - F_m(\theta_{m-1})] + 2\frac{\partial a_m}{\partial \boldsymbol{\Omega}_m}\frac{\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_m} f_m(\theta_{m-1})}{1 - F_m(\theta_{m-1})} \\
& + a_m \frac{\frac{\partial^2 \theta_{m-1}}{\partial \boldsymbol{\Omega}_m^2} f_m(\theta_{m-1}) + \frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_m}\frac{\partial f_m(\theta_{m-1})}{\partial \boldsymbol{\Omega}_m}}{1 - F_m(\theta_{m-1})} + a_m \frac{\left[\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_m} f_m(\theta_{m-1})\right]^2}{[1 - F_m(\theta_{m-1})]^2} \\
& + \frac{\partial^2 \theta_1}{\partial \boldsymbol{\Omega}_m^2}\sum_{i=1}^{n}\Delta(x_i - \theta_1)\log f_1(x_i) \\
& + \sum_{i=2}^{m-1}\left[\frac{\partial^2 a_i}{\partial \boldsymbol{\Omega}_m^2}\log\phi_i + 2\frac{\partial a_i}{\partial \boldsymbol{\Omega}_m}\frac{\frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_m}}{\phi_i} + a_i\frac{\frac{\partial^2 \phi_i}{\partial \boldsymbol{\Omega}_m^2}\phi_i - \left(\frac{\partial \phi_i}{\partial \boldsymbol{\Omega}_m}\right)^2}{\phi_i^2} - \frac{\partial^2 \theta_{i-1}}{\partial \boldsymbol{\Omega}_m^2}\sum_{x_i \le \theta_i}\Delta(x_i - \theta_{i-1})\log f_i(x_i)\right] \\
& + \sum_{i=2}^{m-1}\left[\frac{\partial^2 \theta_i}{\partial \boldsymbol{\Omega}_m^2}\sum_{\theta_{i-1} < x_i}\Delta(x_i - \theta_i)\log f_i(x_i) - 2\frac{\partial \theta_i}{\partial \boldsymbol{\Omega}_m}\frac{\partial \theta_{i-1}}{\partial \boldsymbol{\Omega}_m}\sum_{i=1}^{n}\Delta(x_i - \theta_{i-1})\Delta(x_i - \theta_i)\log f_i(x_i)\right] \\
& + \frac{\partial^2 a_m}{\partial \boldsymbol{\Omega}_m^2}\log\phi_m + 2\frac{\partial a_m}{\partial \boldsymbol{\Omega}_m}\frac{\frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_m}}{\phi_m} + a_m\frac{\frac{\partial^2 \phi_m}{\partial \boldsymbol{\Omega}_m^2}\phi_m - \left(\frac{\partial \phi_m}{\partial \boldsymbol{\Omega}_m}\right)^2}{\phi_m^2} - \frac{\partial^2 \theta_{m-1}}{\partial \boldsymbol{\Omega}_m^2}\sum_{i=1}^{n}\Delta(x_i - \theta_{m-1})\log f_m(x_i) \\
& - 2\frac{\partial \theta_{m-1}}{\partial \boldsymbol{\Omega}_m}\sum_{i=1}^{n}\Delta(x_i - \theta_{m-1})\frac{\frac{\partial f_m(x_i)}{\partial \boldsymbol{\Omega}_m}}{f_m(x_i)} + \sum_{\theta_{m-1} < x_i}\frac{\frac{\partial^2 f_m(x_i)}{\partial \boldsymbol{\Omega}_m^2} f_m(x_i) - \left[\frac{\partial f_m(x_i)}{\partial \boldsymbol{\Omega}_m}\right]^2}{[f_m(x_i)]^2}.
\end{aligned}
$$

5. A SIMULATION STUDY

In this section, we perform a simulation study to show the practical value of the expressions in the paper. We used the following scheme:

- (**i**) simulate a sample of size 1000 from the m-piece composite distribution in Section 4 with each piece described by an exponential distribution and $\theta_j = j$, $j = 1, 2, ..., m-1$;
- (**ii**) estimate the maximum likelihood estimates for the distribution;
- (**iii**) compute the information matrix given in Section 4;

(**iv**) compute the information matrix numerically using the package `numDeriv` in R (R Development Core Team, 2023);

(**v**) take the squared sum of elementwise differences of the two matrices in steps (iii) and (iv);

(**vi**) repeat steps (i) to (v) 100 times;

(**vii**) draw a boxplot of the 100 squared sums of elementwise differences of the two matrices;

(**viii**) repeat steps (i) to (vii) for $m = 2, 5, 10, 50$.

Figure 1 shows the boxplots. Figure 2 shows the boxplots when the lognormal distribution was used to model each piece in step (i). Figure 3 shows the boxplots when the Pareto distribution was used to model each piece in step (i).

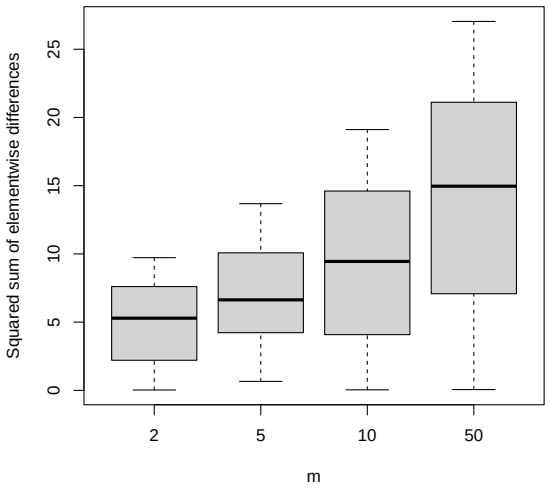

Figure 1: The squared sum of elementwise differences versus m when each piece was modeled by the exponential distribution.

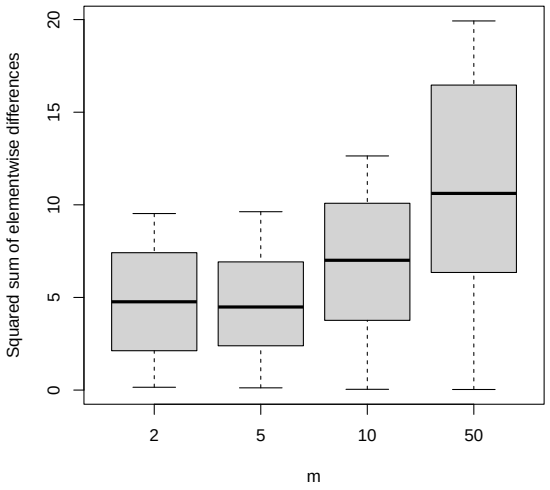

Figure 2: The squared sum of elementwise differences versus m when each piece was modeled by the lognormal distribution.

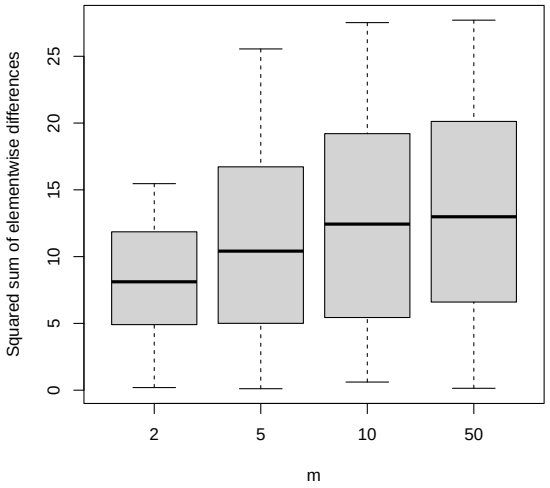

Figure 3: The squared sum of elementwise differences versus m when each piece was modeled by the Pareto distribution.

We can see that the squared sum of elementwise differences is significant whatever m is. The magnitude of the squared sum increases with increasing values of m. This is expected because of:

i) numerical errors;

ii) the increasing number of boundaries in the composite distribution as m increases;

iii) the increasing number of conditions on continuity;

iv) the increasing number of conditions on differentiability.

The differences appear as large as 25 when $m = 50$. Hence, the expressions in the paper should be trusted (over numerical derivatives) for accurate computation of the information matrices and hence for accurate modeling of insurance data especially when m is large.

For simplicity, we have chosen the exponential, lognormal and Pareto distributions to describe each piece of the composite distribution. But the results were similar when other distributions were considered. In particular, the squared sum of elementwise differences were significant for any m and always increased with increasing values of m.

ACKNOWLEDGMENTS

The authors would like to thank the Editor and the referee for careful and comments which improved this paper.

REFERENCES

Aminzadeh, M. S. and Deng, M. (2018). Bayesian predictive modeling for exponential-Pareto composite distribution. *Variance*, 12:59–68.

Aminzadeh, M. S. and Deng, M. (2019). Bayesian predictive modeling for inverse gamma-Pareto composite distribution. *Communications in Statistics – Theory and Methods*, 48:1938–1954.

Bee, M. (2015). Estimation of the lognormal-Pareto distribution using probability weighted moments and maximum likelihood. *Communications in Statistics – Simulation and Computation*, 44:2040–2060.

Benatmane, C., Zeghdoudi, H., Shanker, R. and Lazri, N. (2020). Composite Rayleigh–Pareto distribution: application to real fire insurance losses data set. *Journal of Statistics and Management Systems*. doi: 10.1080/09720510.2020.1759253

Calderin-Ojeda, E. (2015). On the composite Weibull–Burr model to describe claim data. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 1:59–69.

Calderin-Ojeda, E. (2018). A note on parameter estimation in the composite Weibull–Pareto distribution. *Risks*, 6. doi: 10.3390/risks6010011

Calderin-Ojeda, E. and Kwok, C. F. (2016). Modeling claims data with composite Stoppa models. *Scandinavian Actuarial Journal*, 9:817–836.

Cooray, K. (2009). The Weibull–Pareto composite family with applications to the analysis of unimodal failure rate data. *Communications in Statistics – Theory and Methods*, 38:1901–1915.

Cooray, K. and Ananda, M. M. A. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal*, 2005:321–334.

Cooray, K. and Cheng, C.-I. (2015). Bayesian estimators of the lognormal-Pareto composite distribution. *Scandinavian Actuarial Journal*, 2015:500–515.

Cooray, K., Gunasekera, S., and Ananda, M. (2010). Weibull and inverse Weibull composite distribution for modeling reliability data. *Model Assisted Statistics and Applications*, 5:109–115.

Deng, M. and Aminzadeh, M. S. (2019). Bayesian predictive analysis for Weibull–Pareto composite model with an application to insurance data. *Communications in Statistics – Simulation and Computation*. doi: 10.1080/03610918.2019.1699572

Eliazar, I. I. and Cohen, M. H. (2012). A Langevin approach to the log-Gauss–Pareto composite statistical structure. *Physica A: Statistical Mechanics and Its Applications*, 391:5598–5610.

Gencturk, Y., Yigiter, A., and Hamurkaroglu, C. (2018). A new model for data with structural change at threshold: Composite exponential-lognormal model. *Advances and Applications in Statistics*, 53:441–456.

Grün, B. and Miljkovic, T. (2019). Extending composite loss models using a general framework of advanced computational tools. *Scandinavian Actuarial Journal*, 2019:642–660.

Kim, B., Noh, J., and Baek, C. (2016). Threshold estimation for the composite lognormal-GPD models. *Korean Journal of Applied Statistics*, 29:807–822.

Kim, Y., Kim, H., Lee, G., and Min, K.-H. (2019). A modified hybrid gamma and generalized Pareto distribution for precipitation data. *Asia-Pacific Journal of Atmospheric Sciences*, 55:609–616.

Lee, K.-H. (2011). Estimation on composite lognormal-Pareto distribution based on doubly censored samples. *Journal of the Korean Data and Information Science Society*, 22:171–177.

Liu, B. and Ananda, M. M. A. (2022). A generalized family of exponentiated composite distributions. *Mathematics*, 10: article number 1895.

Liu, B. and Ananda, M. M. A. (2023). A new insight into reliability data modeling with an exponentiated composite exponential-Pareto model. *Applied Sciences*, 13: article number 645.

Liu, B. and Ananda, M. M. A. (2023). Analyzing insurance data with an exponentiated composite inverse gamma-Pareto model. *Communications in Statistics – Theory and Methods*, 52:7618–7631.

Luckstead, J. and Devadoss, S. (2017). Pareto tails and lognormal body of US cities size distribution. *Physica A: Statistical Mechanics and Its Applications*, 465:573–578.

Luckstead, J., Devadoss, S., and Danforth, D. (2017). The size distributions of all Indian cities. *Physica A: Statistical Mechanics and Its Applications*, 474:237–249.

Maghsoudi, M. and Abu Bakar, S. A. (2017). In search of best fitted composite model to the ALAE data set with transformed gamma and inversed transformed gamma families. *AIP Conference Proceedings*, 1842. doi: 10.1063/1.4982837

Maghsoudi, M., Abu Bakar, S. A., and Hamzah, N. A. (2014). Composite Weibull-inverse transformed gamma distribution and its actuarial application. *AIP Conference Proceedings*, 1605. doi: 10.1063/1.4887728

Mutali, S. and Vernic, R. (2020). On the composite lognormal-Pareto distribution with uncertain threshold. *Communications in Statistics – Simulation and Computation*. doi: 10.1080/03610918.2020.1743860

Nadarajah, S. and Bakar, S. A. A. (2013). CompLognormal: an R package for composite lognormal distributions. *R Journal*, 5:97–103.

Nadarajah, S. and Bakar, S. A. A. (2014). New composite models for the Danish fire insurance data. *Scandinavian Actuarial Journal*, 180–187.

Pak, R. J. (2013). A robust estimation for the composite lognormal-Pareto model. *Communications for Statistical Applications and Methods*, 20:311–319.

Park, S. and Baek, C. (2018). Time-varying modeling of the composite LN-GPD. *Korean Journal of Applied Statistics*, 31:109–122.

Pigeon, M. and Denuit, M. (2011). Composite lognormal-Pareto model with random threshold. *Scandinavian Actuarial Journal*, 177–192.

R Development Core Team (2023). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*. Vienna, Austria.

Scollnik, D. P. M. (2007). On composite lognormal-Pareto models. *Scandinavian Actuarial Journal*, 20–33.

Scollnik, D. P. M. and Sun, C. (2012). Modeling with Weibull–Pareto models. *North American Actuarial Journal*, 16:260–272.

Teodorescu, S. (2010). On the truncated composite lognormal-Pareto model. *Mathematical Reports*, 12:71–84.

Teodorescu, S. and Panaitescu, E. (2009). On the truncated composite Weibull–Pareto model. *Mathematical Reports*, 11:259–273.

Teodorescu, S. and Vernic, R. (2009). Some composite exponential Pareto models for actuarial prediction. *Romanian Journal of Economic Forecasting*, 82–100.

Teodorescu, S. and Vernic, R. (2013). On composite Pareto models. *Mathematical Reports*, 15:11–29.

REVSTAT – Statistical Journal
Volume 23, Number 3, July 2025, 423–440
https://doi.org/10.57805/revstat.v23i3.623

The Wald-Type Confidence Interval on the Mean Response Function of the Poisson Inverse Gaussian Ridge Regression

Authors: Muhammad Qasim
– Department of Economics, Finance and Statistics, Jönköping International Business School, Jönköping University, Sweden
muhammad.qasim@ju.se

Y. Murat Bulut ✉
– Department of Statistics, Science Faculty, Eskişehir Osmangazi University, Türkiye
ymbulut@ogu.edu.tr

Kristofer Månsson
– Department of Economics, Finance and Statistics, Jönköping International Business School, Jönköping University, Sweden
kristofer.mansson@ju.se

Received: June 2022 Revised: October 2023 Accepted: October 2023

Abstract:

- The negative binomial (NB) regression model is commonly used to model overdispersed count data. However, the NB regression model is not suitable for highly overdispersed data, for which the Poisson-inverse Gaussian (PIG) regression model is often used instead. The maximum likelihood (ML) estimator is typically used to estimate the coefficients of the PIG regression model. However, when multicollinearity exists among the explanatory variables, the ML estimator's variance can become inflated. To address this issue, we propose PIG ridge regression (PIGRR) and quantile-based ridge regression estimators for the PIG regression model. We also suggest using a Wald-type method to calculate the confidence interval on the mean response function of the PIGRR. To evaluate the performance of these proposed methods, we conducted a Monte Carlo simulation study, considering mean squared error and average confidence lengths as performance criteria. Additionally, we analyzed the traffic fatalities dataset to demonstrate the benefits of the proposed estimators for practitioners dealing with multicollinearity issues in real datasets.

Keywords:

- *multicollinearity; Poisson-inverse Gaussian distribution; ridge regression; confidence interval; traffic fatalities.*

AMS Subject Classification:

- 62J07, 62J12.

✉ Corresponding author

1. INTRODUCTION

Count data is commonly analyzed in many real-world applications. This includes examples such as the number of days lost due to work accidents, the number of occurrences of thunderstorms in a calendar year, or the number of insurance claims. The Poisson distribution is the primary choice for modeling such data due to its simplicity, but it assumes that the mean and variance of the distribution are equal, known as the equidispersion assumption. This makes the Poisson distribution unsuitable for modeling overdispersed data, where the variance is larger than the mean. Various factors can cause overdispersion, such as an insufficient number of interaction terms in the regression model or the omission of important explanatory variables.

Various statistical distributions have been proposed to model overdispersed data, with the negative binomial (NB) distribution being the most popular. The NB distribution allows for the variance to exceed the mean of the data and can be obtained as a mixture of Poisson and gamma distributions. In regression models, when the dependent variable, y_i, takes the form of non-negative integers or counts, the NB regression model, which is a particular case of generalized linear models (GLMs), can be applied. The NB regression model is widely used in applied areas such as social, economic, and health sciences (Månsson, 2012). The coefficients of the NB regression model can be estimated using the maximum likelihood (ML) estimation method, which involves solving a non-linear equation through iterative algorithms such as the iterative weighted least squares (IWLS) algorithm.

The PIG distribution is another distribution used to model overdispersed data that is discussed in the literature. It is a special case of the Sichel distribution with two parameters (Stein *et al.*, 1987). Like the NB distribution, the PIG distribution can also be expressed as a mixture of distributions (Willmot, 1987). However, the PIG distribution uses a mixture of the Poisson and inverse Gaussian distributions, while the NB distribution uses a gamma distribution as the mixing distribution. The PIG distribution is particularly useful for count data with longer tails and larger kurtosis, making it an alternative to the NB distribution (Dean *et al.*, 1989). The PIG distribution is better suited for data with a high initial peak that may be skewed to the right, giving it a modeling capacity superior to that of the NB distribution (Hilbe, 2014). Furthermore, the PIG distribution has an easily obtainable likelihood function that has a closed-form representation, making parameter estimation relatively simple (Zha *et al.*, 2016). These properties make the PIG distribution more effective than the NB distribution in dealing with highly overdispersed data. The PIG distribution has been applied in a variety of contexts, including modeling vehicle crash data in Zha *et al.* (2016) and comparing NB and PIG regression models for horseshoe crabs data in Putri *et al.* (2020).

In regression models, another issue that can arise is the problem of multicollinearity. This occurs when the explanatory variables are highly correlated with each other, as defined by Frisch (1934). Multicollinearity has a negative impact on the ML estimator, resulting in estimated coefficients with larger variances and unreliable inference. Confidence intervals for unknown parameters may also have a wider range than usual, and it becomes difficult to estimate the unique effects of the explanatory variables. To address this issue, Hoerl and Kennard introduced the ridge estimator in the linear regression model (LRM). Since then, many different biased estimators have been proposed to combat multicollinearity, including the Liu estimator (Liu, 1993), Liu-type estimator (Liu, 2003), and two-parameter estimator (Özkale and Kaçıranlar, 2007 and Qasim *et al.*, 2022), among others.

In GLMs, the problem of multicollinearity was first addressed by Schaefer *et al.* (1984), who proposed the ridge estimator for logistic regression, a particular case of GLMs (Schaefer *et al.*, 1984). Since then, the ridge estimator has been adapted for other special cases of GLMs. For instance, Månsson and Shukur (2011) introduced the ridge estimator in the Poisson regression model and Locking *et al.* (2013) defined it for the Probit regression model. Algamal (2019) proposed the ridge estimator for gamma and inverse Gaussian regression models, while Månsson (2012) developed it for the negative binomial model. Khalaf *et al.* (2014) introduced the ridge estimator in the Tobit regression model, and Qasim *et al.* (2021) proposed it for the Beta regression model. Recently, Awwad *et al.* (2022) proposed two-parameter estimators for the logistic regression model. Other biased estimators have also been proposed by different authors to address multicollinearity in GLMs (see, e.g., Abbasi and Özkale, 2021; Abonazel *et al.*, 2023; Ahmed *et al.*, 2021; Akram *et al.*, 2022; Algamal *et al.*, 2022; Alheety *et al.*, 2021; Amin *et al.*, 2019; Amin *et al.*, 2020; Asar and Genç, 2018; Bulut, 2021; Kibria, 2022; Lukman *et al.*, 2022; Lukman *et al.*, 2022; Mahmoudi *et al.*, 2020; Nooi Asl *et al.*, 2021; Omer *et al.*, 2021; Özkale, 2019; Qasim *et al.*, 2020; Qasim *et al.*, 2022; Toker *et al.*, 2019; Toker *et al.*, 2021). To the best of our knowledge, no ridge regression or quantile-based ridge regression estimators have been defined for the PIG regression model. Moreover, we suggest using a Wald-type method to calculate the confidence interval for the mean response function of the PIG ridge regression (PIGRR). Gómez-Deniz and Calderín-Ojeda (2018) investigated the mixture of Poisson-reciprocal inverse Gaussian distributions, which, under specific conditions, outperforms the PIG model. This paper introduces the PIGRR and a confidence interval that can be extended to a mixed Poisson regression model utilizing the reciprocal inverse Gaussian as the mixing distribution . The implementation of these enhancements holds substantial potential for advancing future research.

This paper is organized as follows. In Section 2, we introduce the PIG regression model, and an estimation of the parameters is given. In Section 3, we define the ridge estimator for the PIG regression model and propose some estimators of the ridge parameter, also, the Wald confidence interval is given. In Section 4, the Monte Carlo simulation study is conducted. In Section 5, we give a real data example. Finally, the paper ended with conclusions.

2. METHODOLOGY

In this section, we present the PIG regression model and introduce the ridge regression estimator, which was defined by Hoerl and Kennard (1970a,b) in the PIG regression model. Furthermore, we introduce the properties of matrix mean squared error (MMSE) and scalar mean squared error (MSE). We will also propose certain biasing parameters to improve the performance of the proposed estimator.

2.1. The Poisson-inverse Gaussian regression model

The probability mass function (pmf) of the PIG distribution is given as

$$P\left(y_i;\mu_i,\tau\right)=\left(\frac{2\alpha_i}{\pi}\right)^{\frac{1}{2}}\frac{\mu_i^{y_i}\exp\left(\frac{1}{\tau}\right)K_{y_i-\frac{1}{2}}(\alpha_i)}{(\alpha_i\tau)^{y_i}\,y_i!},\qquad y_i=0,1,2,..., \tag{2.1}$$

where $\alpha_i = \frac{\sqrt{1+2\mu_i\tau}}{\tau}$ and $K_s(\cdot)$ is the modified Bessel function of the second kind (Stein *et al.*, 1987). The mean and variance are given as respectively

$$\mathrm{E}(Y_i) = \mu_i, \tag{2.2}$$

$$\mathrm{Var}(Y_i) = \mu_i(1+\tau\mu_i). \tag{2.3}$$

In the regression framework, generally, the mean of the response variable is modeled. Because $\mathrm{Var}(Y_i) > \mathrm{E}(Y_i)$, the PIG model is suitable to model overdispersed data.

Let $Y_1, Y_2, ..., Y_n$ be n independent random variables from the PIG distribution with the parameters μ and τ ($Y_i \sim \mathrm{PIG}(\mu, \tau)$). Then, we assume that the mean of Y_i satisfies the following functional relation:

$$g(\mu_i) = \eta_i = x_i^\top \beta, \qquad i = 1, 2, ..., n, \tag{2.4}$$

where η_i is the linear predictor, $x_i^\top = (x_{i1}, x_{i2}, ..., x_{in})$ denotes the vector of covariates, and $\beta = (\beta_1, \beta_2, ..., \beta_p)$ is the regression coefficient's vector with size p. It is assumed that the mean link function $g\colon (0,\infty) \to \mathbb{R}$ is strictly monotonic and twice differentiable. In literature, there are different link functions such as identity link function ($g(\mu) = \mu$), logarithmic link function ($g(\mu) = \log(\mu)$), and square root link function ($g(\mu) = \sqrt{\mu}$) for more details about the link function, please see McCullagh and Nelder (1989). In this study, we use the log link function.

Generally, the estimate of β is usually found using the ML estimation method. The likelihood function of the PIG is given as

$$\begin{aligned} L(\mu, \tau; y) &= \prod_{i=1}^{n} P(y_i; \mu, \tau) \\ &= \prod_{i=1}^{n} \left(\frac{2\alpha_i}{\pi}\right)^{\frac{1}{2}} \frac{\mu_{y_i} \exp\left(\frac{1}{\tau}\right) K_{y_i - \frac{1}{2}}(\alpha_i)}{(\alpha_i \tau)^{y_i}\, y_i!}. \end{aligned} \tag{2.5}$$

The log-likelihood function is obtained by taking the natural logarithm of both sides of the equation (2.5) as follows:

$$\begin{aligned} \ell(\mu, \tau; y) = \sum_{i=1}^{n} \Bigg\{ & \frac{1}{2}\log(2\alpha_i) - \frac{1}{2}\log(\pi) + y_i \log(\mu_i) + \frac{1}{\tau} + \log\left(K_{y_i - \frac{1}{2}}(\alpha_i)\right) \\ & - y_i \log(\alpha_i) - y_i \log(\tau) - \log(y_i!) \Bigg\}. \end{aligned} \tag{2.6}$$

When we take $\mu_i = \exp\left(x_i^\top \beta\right)$, the log-likelihood function is rewritten as follows:

$$\begin{aligned} \ell(\beta, \tau; y) = \sum_{i=1}^{n} \Bigg\{ & \frac{1}{2}\log\left(2\,\frac{\sqrt{1+2\exp\left(x_i^\top\beta\right)\tau}}{\tau}\right) - \frac{1}{2}\log(\pi) \\ & + y_i \log\left(\exp\left(x_i^\top\beta\right)\right) + \frac{1}{\tau} + \log\left(K_{y_i - \frac{1}{2}}\left(\frac{\sqrt{1+2\exp\left(x_i^\top\beta\right)\tau}}{\tau}\right)\right) \\ & - y_i \log\left(\frac{\sqrt{1+2\exp\left(x_i^\top\beta\right)}}{\tau}\right) - y_i\log(\tau) - \log(y_i!) \Bigg\}. \end{aligned} \tag{2.7}$$

The vector of coefficients using the ML estimation method is estimated by solving the following equation:

$$
\begin{aligned}
S(\beta) &= \frac{\partial \ell(\beta,\tau;y)}{\partial \beta} \\
&= \sum_{i=1}^{n}\left\{ x_i \left[y_i - \frac{R_{y_i-\frac{1}{2}}\left(\frac{\sqrt{1+2\exp\left(x_i^\top\beta\right)\tau}}{\tau}\right)}{\left(1+2\exp\left(x_i^\top\beta\right)\tau\right)^{\frac{1}{2}}}\ \exp\left(x_i^\top\beta\right)\right]\right\} = 0,
\end{aligned}
\tag{2.8}
$$

where $R_\nu(z) = \frac{K_{\nu+1}(z)}{K_\nu(z)}$ is calculated recursively as $R_{-\frac{1}{2}}(z) = 1$; $R_\nu(z) = \frac{2\nu}{z} + \frac{1}{R_{\nu-1}(z)}$.

Since (2.8) is a nonlinear, $S(\beta)$ can not be solved explicitly. Iterative methods, such as Newton-Raphson or Iteratively Reweighted Least Squares (IRLS), are necessary to solve this equation. When estimating unknown coefficients using the IRLS algorithm, the unknown parameters are obtained in each iteration as follows:

$$
\beta^{(m+1)} = \beta^{(m)} + I^{-1}\left(\beta^{(m)}\right) S\left(\beta^{(m)}\right),
\tag{2.9}
$$

where $S\left(\beta^{(m)}\right)$ is the score vector evaluated at $\beta^{(m)}$ and $I^{-1}\left(\beta^{(m)}\right)$ is the Fisher information matrix evaluated at $\beta^{(m)}$. When convergence holds, the vector of coefficients can be obtained as follows:

$$
\widehat{\beta}_{\mathrm{ML}} = (\mathcal{D})^{-1} X^\top \widehat{W}\, \widehat{u},
\tag{2.10}
$$

where $\widehat{u}_i = \log(\widehat{\mu}_i) + \frac{(y_i-\widehat{\mu}_i)}{\widehat{\mu}_i}$, $W = \operatorname{diag}\left(\frac{\widehat{\mu}_i}{1+\widehat{\tau}\widehat{\mu}_i}\right)$ and $\mathcal{D} = X^\top \widehat{W} X$. Also, to estimate dispersion parameter τ, partial derivative of the log-likelihood function is obtained as

$$
\begin{aligned}
S(\tau) &= \frac{\partial \ell(\beta,\tau;y)}{\partial \tau} \\
&= \sum_{i=1}^{n}\left\{ -\frac{1}{\tau^2} - \frac{y_i}{\tau} + R_{y_i-\frac{1}{2}}\left(\frac{\sqrt{1+2\exp\left(x_i^\top\beta\right)\tau}}{\tau}\right) \frac{1+\tau\exp\left(x_i^\top\beta\right)}{\tau^2\left(1+2\exp\left(x_i^\top\beta\right)\right)^{\frac{1}{2}}}\right\} = 0.
\end{aligned}
$$

The covariance matrix of the ML estimator is obtained as follows:

$$
\operatorname{Cov}\left(\widehat{\beta}_{\mathrm{ML}}\right) \cong (\mathcal{D})^{-1}.
\tag{2.11}
$$

The asymptotic mean squared error (MSE) of ML estimator can be defined as

$$
\operatorname{MSE}\left(\widehat{\beta}_{\mathrm{ML}}\right) \cong \operatorname{tr}\left[(\mathcal{D})^{-1}\right] = \sum_{j=1}^{p} \frac{1}{\lambda_j},
\tag{2.12}
$$

where λ_j is the j-th eigenvalue of the $X^\top \widehat{W} X$ matrix. When the explanatory variables in a PIG regression model are highly intercorrelated, it results in a multicollinearity problem. This often causes the matrix of weighted cross products to become ill-conditioned, with some eigenvalues being small. As a consequence, the estimated MSE of the ML estimator is inflated. To overcome this issue, we propose a biased estimator, the PIGRR estimator.

3. Poisson inverse Gaussian ridge regression

We propose PIGRR method as a solution to the problem caused by multicollinearity applied to count data on overdispersion case. Let $\widehat{\mathcal{B}}$ be any estimator of the parameter vector β. The difference between the log-likelihood functions assessed at ML estimate and $\widehat{\mathcal{B}}$ is

$$\operatorname{diff}(\ell) = \ell_{\max} - \ell\big(\widehat{\mathcal{B}}\big).$$

By following Segerstedt (1992), we define the PIGRR as the solution of the optimization problem

$$\min \widehat{\mathcal{B}}^{\top}\widehat{\mathcal{B}} \qquad \text{subject to} \qquad \operatorname{diff}(\ell) = c_0,$$

where c_0 is the fixed number and by developing the Lagrange function, we define

$$\begin{aligned}\mathcal{Q}\big(\widehat{\mathcal{B}},k\big) \;=\; \widehat{\mathcal{B}}^{\top}\widehat{\mathcal{B}} - \left(\frac{2}{k}\right)\Bigg[\sum_{i=1}^{n}\Bigg\{\frac{1}{2}\log\left(\frac{2\alpha_i}{\pi}\right) + y_i\log\Big(\exp\big(x_i^{\top}\beta\big)\Big) + \frac{1}{\tau} \\ +\; \log\Big(K_{y_i-\frac{1}{2}}(\alpha_i)\Big) - y_i\log\left(\frac{\alpha_i}{\tau}\right) - \log(y_i!)\Bigg\} + (c_0 - \ell_{\max})\Bigg],\end{aligned}$$

where $\alpha_i = \frac{\sqrt{1+2\exp\left(x_i^{\top}\beta\right)\tau}}{\tau}$, $\frac{2}{k}$ is the Lagrange multiplier and $\mathcal{Q}\big(\widehat{\mathcal{B}},k\big)$ is a $p\times 1$ vector with elements $\mathfrak{q}_j\big(\widehat{\mathcal{B}},k\big)$:

$$\begin{aligned}\mathfrak{q}_j\big(\widehat{\mathcal{B}},k\big) \;&=\; \frac{\partial\mathcal{Q}\big(\widehat{\mathcal{B}},k\big)}{\partial\widehat{\mathcal{B}}_j} \\ &=\; 2\widehat{\mathcal{B}}_j - \left(\frac{2}{k}\right)\frac{\partial\ell\big(\widehat{\mathcal{B}}\big)}{\partial\widehat{\mathcal{B}}_j} \\ &=\; 2\widehat{\mathcal{B}}_j - \left(\frac{2}{k}\right)\sum_{i=1}^{n}\left[\left(\frac{y_i - \exp\big(x_i^{\top}\beta\big)}{\exp\big(x_i^{\top}\beta\big)\Big[1+\tau\exp\big(x_i^{\top}\beta\big)\Big]}\right)\frac{\partial\exp\big(x_i^{\top}\beta\big)}{\partial\eta_i}\;x_{ij}\right], \quad j=1,2,...,p. \qquad (3.1)\end{aligned}$$

Now define $\mathcal{H}\big(\widehat{\mathcal{B}},k\big)$ as a $p\times p$ matrix with elements $\mathfrak{h}_{jq}\big(\widehat{\mathcal{B}},k\big)$ and taking the second order derivates of $\mathfrak{q}_j\big(\widehat{\mathcal{B}},k\big)$ as

$$\mathfrak{h}_{jq}\big(\widehat{\mathcal{B}},k\big) \;=\; \frac{\partial^2\mathcal{Q}\big(\widehat{\mathcal{B}},k\big)}{\partial\widehat{\mathcal{B}}_j\,\partial\widehat{\mathcal{B}}_q} \;=\; 2\delta_{jq} - \left(\frac{2}{k}\right)\frac{\partial^2\ell\big(\widehat{\mathcal{B}}\big)}{\partial\widehat{\mathcal{B}}_j\,\partial\widehat{\mathcal{B}}_q}.$$

Taking the expectation of both sides of (3.1) as

$$\begin{aligned}\mathrm{E}\left[\frac{\partial^2\mathcal{Q}\big(\widehat{\mathcal{B}},k\big)}{\partial\widehat{\mathcal{B}}_j\partial\widehat{\mathcal{B}}_q}\right] \;&=\; 2\left[\delta_{jq} - \left(\frac{1}{k}\right)\mathrm{E}\left\{-\frac{\partial\ell\big(\widehat{\mathcal{B}}\big)}{\partial\widehat{\mathcal{B}}_j}\,\frac{\partial\ell\big(\widehat{\mathcal{B}}\big)}{\partial\widehat{\mathcal{B}}_q}\right\}\right], \\ &=\; 2\left[\delta_{jq} + \left(\frac{1}{k}\right)\sum_{i=1}^{n}\frac{x_{ij}\,x_{iq}}{\exp\big(x_i^{\top}\beta\big)\Big[1+\tau\exp\big(x_i^{\top}\beta\big)\Big]}\left\{\frac{\partial\exp\big(x_i^{\top}\beta\big)}{\partial\eta_i}\right\}^2\right], \qquad (3.2)\end{aligned}$$

where $\delta_{jq} = \begin{cases} 1, & \text{if } j = q, \\ 0, & \text{otherwise.} \end{cases}$ By means of the Fisher scoring method in this case yields

$$\mathcal{H}\big(\widehat{\mathcal{B}}^{(m)}, k\big)\,\widehat{\beta}(k)^{(m+1)} = \mathcal{H}\big(\widehat{\mathcal{B}}^{(m)}, k\big)\,\widehat{\beta}(k)^{(m)} + \mathcal{Q}\big(\widehat{\mathcal{B}}^{(m)}, k\big), \tag{3.3}$$

where $\mathfrak{q}_j\big(\widehat{\mathcal{B}}, k\big)$ and $\mathfrak{h}_{jq}\big(\widehat{\mathcal{B}}, k\big)$ are the elements of the vector $\mathcal{H}\big(\widehat{\mathcal{B}}^{(m)}, k\big)$ and matrix $\mathcal{Q}\big(\widehat{\mathcal{B}}^{(m)}, k\big)$, respectively, and both can be assessed at the preliminary estimate $\widehat{\beta}(k)^{(m)}$. From equations (3.1)–(3.2), identifying that

$$\begin{aligned} \mathcal{Q}\big(\widehat{\mathcal{B}}^{(m)}, k\big) &= 2\,\widehat{\beta}(k)^{(m)} - \left(\frac{2}{k}\right)\Big[X^{\top}\widehat{W}\,\widehat{u} - \mathcal{D}\,\widehat{\beta}(k)^{(m)}\Big], \\ \mathcal{H}\Big(\widehat{\mathcal{B}}^{(m)}, k\Big) &= \left(-\frac{2}{k}\right)\big[\mathcal{D} + k\,I\big]. \end{aligned}$$

Now, (3.3) is equivalent to

$$\begin{aligned} (\mathcal{D} + k\,I)\,\widehat{\beta}(k)^{(m+1)} &= \left(2\,\widehat{\beta}(k)^{(m)} - \left(\frac{2}{k}\right)\Big[X^{\top}Wu - \mathcal{D}\,\widehat{\beta}(k)^{(m)}\Big] - \left(\frac{2}{k}\right)(\mathcal{D} + k\,I)\,\widehat{\beta}(k)^{(m)}\right) \\ &\quad \times \left(-\frac{k}{2}\right)(\mathcal{D} + k\,I)^{-1}, \\ (\mathcal{D} + k\,I)\,\widehat{\beta}(k)^{(m+1)} &= X^{\top}Wu. \end{aligned}$$

When the successive estimates $\widehat{\beta}(k)^{(m)}$ converges to $\widehat{\beta}(k)$ as $m \to \infty$, then we find the following PIGRR estimator:

$$\widehat{\beta}_{\text{PIGRR}} = \widehat{\beta}(k) = \big(\mathcal{D}(k)\big)^{-1}X^{\top}Wu, \tag{3.4}$$

where k is the ridge parameter and $\mathcal{D}(k) = \mathcal{D} + k\,I$.

The mean squared error (MSE) of $\widehat{\beta}_{\text{PIGRR}}$ equals

$$\begin{aligned} \text{MSE}\big(\widehat{\beta}_{\text{PIGRR}}\big) &\cong \text{E}\Big[\big(\widehat{\beta}_{\text{PIGRR}} - \beta\big)^{\top}\big(\widehat{\beta}_{\text{PIGRR}} - \beta\big)\Big] \\ &\cong \text{tr}\Big[\big(\mathcal{D}(k)\big)^{-1}\mathcal{D}\big(\mathcal{D}(k)\big)^{-1}\Big] + k^2\beta^{\top}\big(\mathcal{D}(k)\big)^{-2}\beta \\ &\cong \sum_{j=1}^{p}\frac{\lambda_j}{(\lambda_j + k)^2} + \sum_{j=1}^{p}\frac{k^2\alpha_j^2}{(\lambda_j + k)^2}, \end{aligned} \tag{3.5}$$

where $\lambda_j(\lambda_1, \lambda_2, ..., \lambda_p) > 0$ are the eigenvalues of the matrix $\mathcal{D}$, α_j^2 is the j-th element of $\mathcal{T}^{\top}\widehat{\beta}_{\text{ML}}$ and $\mathcal{T}$ is the matrix whose columns are the eigenvectors of the matrix $\mathcal{D}$ such that $\mathcal{D} = \xi\Lambda\xi^{\top}$, where $\Lambda = \text{diag}(\lambda_j)$. Månsson and Shukur (2011) investigated the MSE properties of the Poisson ridge regression, and these properties are same for the PIGRR. Månsson and Shukur (2011) showed that Poisson ridge regression is superior to the ML estimator in terms of scalar MSE. Similarly, the $\widehat{\beta}_{\text{PIGRR}}$ is also superior to the $\widehat{\beta}_{\text{ML}}$ when $k > 0$.

3.1. Proposed estimators of ridge parameter k

The performance of the PIGRR method depends on the choice of ridge parameter (k). Therefore, we suggest some new methods for estimating k. First, we derive the optimal value of k by taking the derivative of $\mathrm{MSE}(\widehat{\beta}_{\mathrm{PIGRR}})$ with respect to k:

$$\frac{\partial \mathrm{MSE}(\widehat{\beta}_{\mathrm{PIGRR}})}{\partial k} = -2\sum_{j=1}^{p} \frac{\lambda_j(1-k\alpha_j^2)}{(\lambda_j+k)^3}. \tag{3.6}$$

By equating (3.6) to zero, we find the value of k that minimizes $\mathrm{MSE}(\widehat{\beta}_{\mathrm{PIGRR}})$ as

$$k_j = \frac{1}{\alpha_j^2}. \tag{3.7}$$

Different estimators of k recommended by Månsson and Shukur (2011) for the Poisson regression and some of these estimators are generalized for the PIGRR:

$$\widehat{k}_1 = \frac{1}{\alpha_{\max}^2}, \qquad \widehat{k}_2 = \left(\prod_{j=1}^{p} \frac{1}{m_j}\right)^{\frac{1}{p}}, \qquad \widehat{k}_3 = \max\left(\frac{1}{m_j}\right),$$

where $\alpha_{\max}^2$ is the maximum value of α_j^2 and $m_j = (k)^{\frac{1}{2}}$. By following Qasim *et al.* (2021), we suggest $\widehat{k}_4$ estimator for the PIGRR:

$$\widehat{k}_4 = \frac{p}{\sum_{j=1}^{p} \alpha_j^2}.$$

Now, we propose a quantile-based ridge estimation. Let $\mathcal{X}$ be a continuous random variable with a cumulative distribution function (cdf) $\mathcal{F}(x)$. Then the q-th quantile of a population is denoted as $\mathbb{Q}(\mathsf{q})$ and characterized the functional inverse of the cdf taken at q, such as

$$\mathbb{Q}(\mathsf{q}) = \mathcal{F}^{-1}(\mathsf{q}) = \inf\left[\mathsf{x}\colon \mathcal{F}(\mathsf{x}) \geq \mathsf{q}\right] \tag{3.8}$$

for predetermined $0 < q < 1$. (3.8) can also be defined as $P[\mathsf{x} \leq \mathbb{Q}(\mathsf{q})] = \mathsf{q}$, which indicates that $100q\%$ of the observations less than or equal to the population quantile $\mathbb{Q}(\mathsf{q})$.

By means of (3.7), we begin by setting $\mathcal{S}_1 = \big(L, k_{(1)}\big)$, $\mathcal{S}_2 = \big[\,k_{(1)}, k_{(2)}\big)$, ..., $\mathcal{S}_{p-1} = \big[\,k_{(p)}, U\big)$, where $k_{(1)}, k_{(2)}, ..., k_{(p)}$ are order statistics of k_j $(j = 1, 2, ..., p)$, and L and U are lower and upper bounds of k (which can be > 0 and ∞, respectively). In addition, the q-th quantile of $\mathbb{Q}_{\mathsf{q}}$ lies in one of these intervals. As a result, the quantile-based ridge estimator can be computed as

$$P\big[k \leq \mathbb{Q}(\mathsf{q})\big] = \mathsf{q}, \qquad 0 < \mathsf{q} < 1.$$

The q-th quantile can also be expressed as $P\big[k > \mathbb{Q}(\mathsf{q})\big] = 1 - \mathsf{q}$. We set four levels ($\mathsf{q} = 0.25$, 0.50, 0.75, 0.95) of the quantiles and therefore, we propose four different quantile-based ridge estimators for the PIGRR, namely, $\mathbb{Q}_{0.25}$, $\mathbb{Q}_{0.50}$, $\mathbb{Q}_{0.75}$ and $\mathbb{Q}_{0.95}$. These ridge estimators are proposed in order to get the lowest MSE of the PIGRR.

3.2. Wald confidence interval

We first examine the coverage properties of Wald-type confidence intervals on the mean response function, $\mu_i = \exp(x_i^\top \beta)$. Recall that in the LRM, a confidence interval on $\mathrm{E}(y \,|\, x = x_0)$ with p parameter is

$$\widehat{y}(x_0) \pm t_{\frac{\alpha}{2}, n-p}\, \widehat{\sigma} \sqrt{x_0^\top \big(X^\top X\big)^{-1} x_0}.$$

Myers and Montgomery (1997) provide an analogous illustration for the GLMs. The asymptotic $100(1-\alpha)\%$ confidence interval on the mean response function at the point x_0 as follows:

$$\widehat{\mu}(x_0) \pm z_{\frac{\alpha}{2}} \sqrt{c_0^\top \big(\mathcal{C}^\top \widehat{W}^{-1} \mathcal{C}\big)^{-1} c_0}, \tag{3.9}$$

where $\widehat{\mu}(x_0)$ is the estimated mean response function at the point of x_0, $\mathcal{C}$ is the matrix of derivatives stimulated by the Taylor series expansion applied in the GLMs, $W = \mathrm{diag}\{\mathrm{Var}(y_i)\}$, where $\mathrm{Var}(y_i) = \ddot{b}(\theta)\, a(\phi)$ from the exponential family of distribution, and c_0 is the vector of these derivatives. By following McCullagh and Nelder (1989) and Myers *et al.* (2002), the development of (3.9) is relatively simple. The PIG regression is the member of the exponential family of distribution. For a link, say, $g(\mu) = x^\top \beta$, we have $\mu = g^{-1}(x^\top \beta)$ and its estimated mean function $\widehat{\mu} = g^{-1}\big(x^\top \widehat{\beta}\big)$. We know that μ is a nonlinear function of β. We make use of the delta method to approximate $\mathrm{Var}[\widehat{\mu}(x_0)]$ for the confidence interval on $\mu(x_0)$. The delta method allows for approximation of the variance of a quantity that is a nonlinear function. Since $\widehat{\mu}(x_0)$ is a nonlinear function of parameter in estimates $\widehat{\beta}$ and by means of the delta method

$$\mathrm{Var}\big(\widehat{\mu}(x_0)\big) = c_0^\top \Big\{ \mathrm{Cov}\big(\widehat{\beta}\big) \Big\}\, c_0,$$

where $c_0 = \frac{\partial \widehat{\mu}(x_0)}{\partial \widehat{\beta}_{\mathrm{ML}}}$ is a vector of derivatives and $\mathrm{Cov}\big(\widehat{\beta}\big)$ is the asymptotic variance-covariance matrix of the ML estimator for the GLMs, such as $I\big(\widehat{\beta}\big)^{-1} = \big(\mathcal{C}^\top W^{-1} \mathcal{C}\big)^{-1}$ and therefore, $\mathrm{Var}\big(\widehat{\mu}(x_0)\big) = c_0^\top$ is approximated by $c_0^\top \big(\mathcal{C}^\top W^{-1} \mathcal{C}\big)^{-1} c_0$. We define the pivotal quantity

$$z = \frac{\widehat{\mu}(x_0) - \mu(x_0)}{\Big[c_0^\top \big(\mathcal{C}^\top W^{-1} \mathcal{C}\big)^{-1} c_0 \Big]^{\frac{1}{2}}}$$

has a normal distribution with mean 0 and variance 1. Hence the useful general form of the approximate $100(1-\alpha)\%$ confidence interval on $\mu(x_0)$ for the GLMs is $\widehat{\mu}(x_0) \pm z_{\frac{\alpha}{2}} \sqrt{c_0^\top \big(\mathcal{C}^\top \widehat{W}^{-1} \mathcal{C}\big)^{-1} c_0}$. Then, we consider the special case of the GLMs is PIG regression. By using the canonical link function $\mathcal{C} = \Big(\frac{WX}{a(\phi)}\Big) \big(\mathcal{C}^\top W^{-1} \mathcal{C}\big)^{-1} = \big(X^\top W^{-1} X\big)^{-1} \{a(\phi)\}^2$, $c_0 = \frac{\{\mathrm{Var}(y_0)\}\, x_0}{a(\phi)}$, and finally the required confidence interval is defined as

$$\widehat{\mu}(x_0) \pm z_{\frac{\alpha}{2}} \widehat{\mathrm{Var}}(y_0) \sqrt{x_0^\top \big(X^\top \widehat{W}^{-1} X\big)^{-1} x_0}.$$

Now we define the Wald confidence interval for $\mu_i = \exp(x_i^\top \beta)$ by using the mean and variance functions of the PIG regression

$$\exp(x_i^\top \widehat{\beta}_{\mathrm{ML}}) \pm z_{\frac{\alpha}{2}} \left[\exp(x_i^\top \widehat{\beta}_{\mathrm{ML}}) \left\{ 1 + \widehat{\tau} \exp(x_i^\top \widehat{\beta}_{\mathrm{ML}}) \right\} \right] \sqrt{x_i^\top \mathcal{D}^{-1} x_i}, \tag{3.10}$$

where $\mathcal{D} = X^\top \widehat{W}^{-1} X$. By using the asymptotic variance-covariance matrix of the PIGRR, we define the Wald confidence interval for $\mu_i = \exp(x_i^\top \beta)$

$$\exp(x_i^\top \widehat{\beta}_{\mathrm{PIGRR}}) \pm z_{\frac{\alpha}{2}} \left[\exp(x_i^\top \widehat{\beta}_{\mathrm{PIGRR}}) \left\{ 1 + \widehat{\tau} \exp(x_i^\top \widehat{\beta}_{\mathrm{PIGRR}}) \right\} \right] \sqrt{x_i^\top \mathcal{D}(k)^{-1} \mathcal{D} \mathcal{D}(k)^{-1} x_i}, \tag{3.11}$$

where $\mathcal{D}(k) = X^\top \widehat{W}^{-1} X + k I$.

4. MONTE CARLO SIMULATION

In this section, we conduct a Monte Carlo simulation to demonstrate the performance of the proposed estimator compared to the ML estimator. In the following subsection, we present the simulation design.

4.1. Design of the simulation

The factors considered in the simulation to evaluate the performance of the proposed estimator include sample size (n), degree of correlation (ρ), and the number of explanatory variables (p). To investigate the impact of these factors, we set up a simulation design that comprises four different sample sizes: 50, 100, 150, and 200, with three distinct correlation coefficient values (ρ) of 0.90, 0.95, and 0.99. Moreover, we use three different values for the number of explanatory variables, namely 3, 5, and 7.

We generate the dependent variable of the PIG regression model as follows:

$$y_i \sim \mathrm{PIG}(\mu_i, \tau), \tag{4.1}$$

where $\mu_i = \exp(x_i^\top \beta)$, $\tau = 2.5$, $\beta = (\beta_1, \beta_2, ..., \beta_p)$ with $\sum_{j=1}^{p} \beta_j^2 = 1$. We use the following method, which was given by McDonald and Galarneau (1975), to generate explanatory variables:

$$x_{ij} = \left(1 - \rho^2\right)^{\frac{1}{2}} z_{ij} + \rho z_{ip}, \tag{4.2}$$

where z_{ij} are independent pseudo-random numbers following the standard normal distribution, ρ represents the correlation between the explanatory variables and $i = 1, 2, ..., n$, $j = 1, 2, ..., p$. To compare the performance of the estimators, we estimate the MSE, which is calculated based on 1000 simulation replications for various combinations of n, ρ, and p:

$$\mathrm{MSE}(\widehat{\beta}) = \frac{\sum_{r=1}^{1000} (\widehat{\beta}_r - \beta)^\top (\widehat{\beta}_r - \beta)}{1000}, \tag{4.3}$$

where $\widehat{\beta}_r$ is the estimated value of coefficients in the r-th replication. Also, we obtain the average confidence lengths on the means response functions of ML estimator and PIGRR methods.

4.2. Results of the simulation study

Tables 1–2 present the results of simulation study, which aimed to investigate the impact of several factors on the accuracy of estimation. Specifically, we provide MSE values for different combinations of p, n, and ρ values. These factors include the number of explanatory variables (p), sample sizes (n), and correlation degrees (ρ), all of which can significantly affect the reliability of statistical analysis. In addition to the MSE values, we also report the average confidence lengths for the means response functions of ML and PIGRR methods. These measures provide insight into the precision and stability of the estimated response functions under varying conditions.

Table 1: MSE values of ML and PIGRR estimators with different ridge parameters.

p	n	ρ	ML	PIGRR							
				k_1	k_2	k_3	k_4	$\mathbb{Q}_{0.25}$	$\mathbb{Q}_{0.50}$	$\mathbb{Q}_{0.75}$	$\mathbb{Q}_{0.95}$
3	50	0.90	3.9629	3.6578	3.7991	3.5100	3.4228	3.3892	3.1003	2.5715	**2.3345**
		0.95	5.3212	4.8090	4.6749	3.9837	4.3833	4.2103	3.7195	3.0916	**2.8166**
		0.99	16.4580	14.2634	8.6817	7.3257	12.0915	10.3336	7.9850	6.8134	**6.2833**
	100	0.90	3.2531	3.0851	3.2157	3.1278	3.0607	3.0814	2.9287	2.4258	**2.2284**
		0.95	4.0130	3.7358	3.8521	3.6021	3.5256	3.5081	3.2874	2.8365	**2.6232**
		0.99	9.8350	8.6379	6.4043	5.3192	7.5223	6.8228	5.6322	4.9925	**4.7043**
	150	0.90	3.1694	3.0541	3.1495	3.0964	2.9636	2.9847	2.8415	2.4136	**2.2074**
		0.95	3.6547	3.4477	3.5705	3.4123	3.2946	3.2951	3.1528	2.6992	**2.4918**
		0.99	7.3670	6.5738	5.5960	4.6865	5.8786	5.4267	4.7384	4.2610	**4.0352**
	200	0.90	3.1255	3.0402	3.1128	3.0654	2.9555	2.9785	2.8237	2.3700	**2.1659**
		0.95	3.5378	3.3804	3.4881	3.3811	3.2627	3.2783	3.1339	2.6926	**2.4826**
		0.99	6.3801	5.7209	5.1625	4.3956	5.1791	4.8949	4.4018	3.9482	**3.7147**
5	50	0.90	5.9567	5.4809	5.6316	4.5465	4.8643	4.7143	4.0507	3.2758	**2.3346**
		0.95	9.7346	8.7749	8.0036	5.4935	7.4877	7.0213	5.7144	4.5515	**3.0213**
		0.99	42.3171	37.2487	14.5175	10.6302	29.5380	23.9425	15.8654	11.3122	**8.6180**
	100	0.90	4.2455	4.0043	4.2025	3.9079	3.7171	3.7109	3.4010	2.8741	**2.1255**
		0.95	5.9535	5.4818	5.6709	4.6344	4.8770	4.7407	4.1433	3.5055	**2.5924**
		0.99	20.7169	18.2691	10.4827	7.1336	14.6964	12.5546	9.1581	7.2057	**5.8155**
	150	0.90	3.7633	3.6038	3.7489	3.5885	3.4036	3.4129	3.1439	2.7543	**2.0351**
		0.95	5.0467	4.7354	4.9369	4.3738	4.3428	4.2698	3.8872	3.3723	**2.5377**
		0.99	15.1806	13.5483	9.4951	6.5243	11.1794	9.8631	7.6467	6.3033	**5.1333**
	200	0.90	3.6360	3.5172	3.6278	3.5229	3.3640	3.3716	3.1369	2.6800	**1.9944**
		0.95	4.5471	4.3075	4.4927	4.1243	4.0071	3.9783	3.6564	3.1451	**2.3418**
		0.99	11.8073	10.5958	8.6331	5.8646	8.8742	7.9941	6.4428	5.3743	**4.2514**
7	50	0.90	9.1405	8.4541	8.5601	6.0465	7.2886	7.0655	5.8112	4.1206	**2.3914**
		0.95	15.6450	14.2603	12.2415	6.9601	11.6641	10.6449	7.9813	5.6598	**3.4269**
		0.99	77.3166	69.2965	20.3857	14.8155	53.4698	44.5839	28.7129	17.5815	**11.8956**
	100	0.90	5.5511	5.2208	5.5038	4.7570	4.6679	4.6057	4.0392	3.0931	**1.9922**
		0.95	8.6180	7.9574	8.1922	5.8951	6.8185	6.5105	5.4209	4.1414	**2.7296**
		0.99	35.9445	32.2017	15.3524	8.6824	24.9931	22.1400	14.8486	9.5913	**6.4953**
	150	0.90	4.6103	4.3901	4.5963	4.2505	4.0346	4.0119	3.6539	2.8875	**1.8452**
		0.95	6.5878	6.1419	6.4549	5.2634	5.4149	5.2693	4.5961	3.6723	**2.3812**
		0.99	22.3684	20.0214	13.8088	7.5938	15.7448	14.5071	10.2449	7.2399	**5.2942**
	200	0.90	4.4384	4.2867	4.4326	4.1912	4.0171	4.0027	3.6143	2.8580	**1.8242**
		0.95	6.1816	5.8494	6.1113	5.1619	5.2816	5.1938	4.5491	3.6021	**2.3417**
		0.99	18.6922	16.8889	12.5442	7.5197	13.6177	12.4779	9.3996	7.0053	**5.0791**

Table 1 presents several important findings from our simulation study. Firstly, the results show that PIGRR methods with all ridge parameters outperform the ML method in terms of MSE values, which confirms our initial expectation. Additionally, the MSE values of the PIGRR estimator with the ridge parameter $\mathbb{Q}_{0.95}$ are consistently smaller than the MSE values of other PIGRR estimators and ML across all conditions. Furthermore, our simulation results reveal that correlation degrees have a negative impact on the MSE values for fixed values of p and n. Specifically, as the degree of correlation increases, the MSE values of all estimators increase, indicating reduced accuracy and reliability of statistical inference. Conversely, increasing the sample size has a positive effect on the MSE values, as larger samples provide more reliable estimates and reduce sampling variability. It is also important to note that the MSE values of ML are approximately twice as large as those of the PIGRR method with $\mathbb{Q}_{0.95}$ across all simulation scenarios. This suggests that the PIGRR method is a more accurate and reliable estimator than the ML method, particularly in high-dimensional settings. Finally, we observe that increasing the number of explanatory variables also leads to an increase in MSE values when other factors are held constant. This highlights the importance of careful variable selection and regularization techniques in statistical modeling.

Table 2 presents the average confidence lengths for the ML and PIGRR methods with $\mathbb{Q}_{0.95}$. Similar to the MSE results, we observe several important findings related to the average confidence lengths. Firstly, the results show that the average confidence lengths of the PIGRR method with $\mathbb{Q}_{0.95}$ are generally shorter than those of the ML method, indicating a higher precision and more accurate inference. Moreover, as the degree of correlation increases, the average confidence lengths of both methods also increase, highlighting the impact of correlation on statistical inference. We also find that the number of explanatory variables has an adverse effect on the average confidence lengths. Specifically, as the number of explanatory variables increases from 3 to 7, the average confidence length of both methods also increases, suggesting that careful variable selection and regularization are important in minimizing the impact of overfitting. Additionally, the simulation results show that increasing the sample size generally leads to a decrease in the average confidence lengths for both methods, as larger samples provide more reliable estimates and reduce sampling variability. However, we note that there are some exceptions to this trend, as the average confidence lengths do not always decrease when the sample size changes from 150 to 200.

Table 2: Comparison of 95% average confidence lengths on the means response functions of ML and PIGRR estimators.

p	ρ	ML				PIGRR[$\mathbb{Q}_{0.95}$]			
		n				n			
		50	100	150	200	50	100	150	200
3	0.90	139.84	48.42	40.63	58.54	26.68	14.84	13.01	11.00
	0.95	239.95	65.23	46.57	63.21	40.92	21.44	14.54	12.19
	0.99	273.31	66.00	59.94	68.10	194.36	44.19	21.76	18.22
5	0.90	1613.70	227.33	309.02	269.58	423.15	27.24	30.05	26.33
	0.95	2677.17	468.77	317.27	330.86	955.23	67.50	31.49	29.73
	0.99	14011.06	529.45	605.12	482.90	963.27	151.18	105.99	37.39
7	0.90	19381.12	1554.10	626.95	965.84	453.31	41.55	42.09	18.07
	0.95	25894.30	1563.55	716.83	1143.08	718.47	122.27	37.51	50.53
	0.99	29457.73	3702.62	10852.13	2389.50	4633.72	600.87	175.96	258.32

Overall, these findings provide important insights into the impact of various factors on the precision and accuracy of statistical inference and have important implications for researchers and practitioners in selecting appropriate methods and interpreting their results.

5. APPLICATION: TRAFFIC FATALITIES

In this application, we model the traffic fatalities for 48 US states (excluding Alaska and Hawaii) during 1988. The data is taken from the `AER` package in `R` and it is denoted Fatalities. We fit PIG regression for the number of night-time vehicle fatalities of the age group 21–24 years old (y_i). We consider the following regressors: per capita personal income in 1987 dollars (X_1), spirits that measures the spirits consumption (X_2), population which is the population in the respective age group such as, population age group 21–24 years old (X_3), total population (X_4) and miles-total measured as the total vehicle miles in millions (X_5). We see the impact of per capita personal income, spirits consumption, population age group 21–24 years old, miles-total, and total population on the number of night-time vehicle fatalities. The square root transformation is used to find the positive predictive confidence interval values of the number of night-time vehicle fatalities. Our analysis aimed to understand how these variables are related to the number of night-time vehicle fatalities. The impact of per capita personal income, spirits consumption, population age group 21–24 years old, miles-total, and total population on the number of night-time vehicle fatalities was explored using PIG regression. Our computations are carried out by using `R` programing language. The problem of multicollinearity is tested by the condition index,which is calculated by taking the square root of the maximum eigenvalue divided by the minimum eigenvalue. $\text{CI} = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} = 95.12$, which indicate a strong multicollinearity in the dataset (for more details, please see Abbasi and Özkale, 2021; Amin *et al.*, 2020; Bulut and Işılar, 2021). The eigenvalues are 2605.7, 92.5, 40.3, 11.6, 0.57 and 0.29.

The MSE of the conventional ML estimator is inflated in the presence of multicollinearity. Therefore, we used biased estimation methods to overcome the problem of multicollinearity. The benefit of our proposed estimator is shown by using this dataset in the PIG regression model. Table 3 presents the estimates of the parameters and scalar MSE of the ML and PIGRR estimators. Application results reveal that a substantial decrease of the MSE when applying PIGRR. This is in line with the simulated results where PIGRR always have a better performance than the ML in almost all considered conditions. Previous empirical results according to OECD (2015) is mixed but in general robustness as shown by our new estimator is desirable. The estimated coefficient using ML and PIGRR of per capita personal income is positive which shows the higher number of night-time vehicle fatalities. Furthermore, spirits are negative which is counterintuitive sine a higher consumption of alcohol should lead to more accidents on average. However, when using proposed PIGRR with quantile based ridge estimation the value of the coefficients is shrunken towards zero. Finally, total population is positive in age group 21–24 which according to OECD (2015) is in line with previous research. The values using PIGRR is shrunken towards zero but not as much as for income and spirits variables. Miles-total is actually positive, and it increases the number of night-time vehicle fatalities.

Table 3: Estimated coefficient and MSE of estimators.

Estimators		Intercept	X_1	X_2	X_3	X_4	X_5	MSE
		β_0	β_1	β_2	β_3	β_4	β_5	
ML		1.644	0.117	−0.126	−0.628	0.777	0.130	5.349
PIGRR	k_1	1.641	0.097	−0.094	−0.220	0.350	0.154	2.111
	k_2	1.639	0.094	−0.092	−0.194	0.326	0.154	2.063
	k_3	1.617	0.066	−0.075	−0.058	0.201	0.155	2.151
	k_4	1.615	0.063	−0.074	−0.051	0.195	0.155	2.174
	$\mathbb{Q}_{0.25}$	1.615	0.063	−0.074	−0.051	0.195	0.155	2.174
	$\mathbb{Q}_{0.50}$	1.564	0.017	−0.058	0.015	0.148	0.160	2.437
	$\mathbb{Q}_{0.75}$	1.504	−0.025	−0.049	0.044	0.136	0.166	2.577
	$\mathbb{Q}_{0.95}$	1.145	−0.128	−0.048	0.109	0.148	0.193	2.864

Note: The explanatory variables are defined as: the per capita personal income in 1987 dollars (X_1), spirits that measures the spirits consumption (X_2), population which is the population in the respective age group such as, population age group 21–24 years old (X_3), total population (X_4) and miles-total measured as the total vehicle miles in millions (X_5).

The results presented in Table 4 (Appendix A) demonstrate the Wald-type 95% confidence interval on the mean response function for each observation of the number of night-time vehicle fatalities using the ML and PIGRR with a shrinkage parameter k_2. It can be seen that in the presence of multicollinearity, the confidence intervals for the ML estimator are wider compared to those of PIGRR. This indicates that the PIGRR method is better at shrinking the coefficients towards zero and producing narrower confidence intervals. Moreover, the average 95% confidence length on the mean response function using the ML and PIGRR are 6.56 and 5.44, respectively, which further confirms the superiority of the PIGRR method in terms of producing narrower confidence intervals and coefficient shrinkage. These findings provide additional evidence that the proposed PIGRR method is more effective in reducing the impact of multicollinearity on the estimation of the mean response function. Furthermore, it is worth noting that the estimated MSE of the proposed PIGRR estimator is smaller than that of the traditional ML estimator, which highlights the advantage of using the PIGRR method in this context.

6. CONCLUSION

The Poisson inverse Gaussian ridge (PIGRR) regression and Wald-type confidence interval for the predictive response function are proposed and is compared to alternative ML estimator. Also, quantile-based ridge estimation ($\mathbb{Q}_{0.25}$, $\mathbb{Q}_{0.50}$, $\mathbb{Q}_{0.75}$, $\mathbb{Q}_{0.95}$) and shrinkage estimators (k_1, k_2, k_3, k_4) are suggested to obtain the optimal value of the PIGRR for the GLM. Results show that the PIGRR method outperforms the traditional ML estimator, providing narrower confidence intervals for the predictive response function. Monte Carlo simulations and an application to a traffic fatalities dataset further demonstrated the effectiveness of the proposed PIGRR estimator. In conclusion, the results of this study indicate that the PIGRR method can be used as an alternative estimator for the PIG regression and GLMs, especially in the presence of multicollinearity.

A. APPENDIX

Table 4: Comparison of 95% Confidence Interval (CI) on the means response function using ML and PIGRR estimators.

Observation	MLE		PIGRR		Length of 95% CI	
	μ_i	95% CI	μ_i	95% CI	MLE	PIGRR
1	4.84	(2.86, 6.82)	4.84	(3.08, 66.60)	3.96	3.52
2	4.66	(3.20, 6.12)	4.67	(3.42, 5.92)	2.91	2.50
3	4.47	(2.57, 6.37)	4.36	(2.58, 6.15)	3.81	3.56
4	20.30	(14.10, 26.51)	20.45	(14.25, 26.64)	12.41	12.39
5	4.76	(2.81, 6.71)	4.73	(2.88, 6.57)	3.90	3.69
6	5.78	(0.00, 14.00)	5.54	(0.00, 13.13)	14.00	13.13
7	3.85	(1.12, 6.59)	3.91	(1.35, 6.47)	5.47	5.12
8	8.34	(4.12, 12.57)	7.77	(4.98, 10.55)	8.45	5.57
9	5.22	(2.82, 7.63)	5.47	(3.94, 7.00)	4.81	3.06
10	4.18	(2.63, 5.73)	4.07	(2.64, 5.51)	3.10	2.87
11	7.47	(0.32, 15.27)	7.34	(2.80, 11.89)	14.95	9.09
12	5.53	(3.68, 7.38)	5.51	(3.97, 7.04)	3.70	3.07
13	5.09	(2.68, 7.50)	4.87	(2.74, 7.00)	4.82	4.26
14	5.15	(2.24, 8.06)	4.91	(2.38, 7.44)	5.82	5.06
15	4.63	(3.00, 6.26)	4.65	(3.04, 6.25)	3.27	3.21
16	4.37	(1.94, 6.81)	4.51	(2.42, 6.60)	4.87	4.18
17	4.01	(2.63, 5.38)	4.02	(2.73, 5.31)	2.75	2.59
18	5.31	(0.83, 9.79)	5.35	(1.19, 9.52)	8.96	8.34
19	5.63	(0.50, 11.76)	5.72	(0.22, 11.23)	11.26	11.01
20	6.50	(4.28, 8.71)	6.62	(4.83, 8.40)	4.43	3.57
21	5.11	(3.05, 7.17)	5.05	(3.09, 7.01)	4.12	3.92
22	4.02	(1.68, 6.35)	4.08	(1.82, 6.34)	4.67	4.52
23	5.65	(3.59, 7.70)	5.47	(3.75, 7.19)	4.11	3.44
24	3.87	(2.60, 5.14)	3.85	(2.61, 5.09)	2.54	2.48
25	4.52	(2.91, 6.14)	4.40	(2.90, 5.90)	3.24	3.00
26	2.83	(0.00, 7.41)	3.12	(0.00, 7.47)	7.41	7.47
27	3.12	(0.00, 7.88)	3.39	(0.00, 7.90)	7.88	7.90
28	7.44	(0.00, 16.28)	7.05	(0.94, 15.04)	16.28	14.10
29	4.10	(2.60, 5.60)	4.08	(2.63, 5.52)	3.01	2.89
30	9.91	(0.00, 33.90)	9.75	(0.00, 22.63)	33.90	22.63
31	5.45	(3.85, 7.05)	5.56	(4.21, 6.91)	3.20	2.71
32	3.63	(2.10, 5.16)	3.67	(2.24, 5.10)	3.06	2.86
33	7.82	(3.51, 12.13)	7.48	(4.44, 10.51)	8.62	6.07
34	4.93	(2.67, 7.19)	4.80	(3.04, 6.57)	4.52	3.53
35	4.84	(3.20, 6.48)	4.67	(3.31, 6.04)	3.28	2.74
36	8.31	(0.26, 16.37)	7.81	(3.33, 12.29)	16.10	8.96
37	4.09	(1.77, 6.40)	4.07	(1.95, 6.19)	4.63	4.23
38	4.06	(1.94, 6.18)	4.25	(2.38, 6.11)	4.24	3.73
39	3.80	(2.49, 5.11)	3.80	(2.52, 5.07)	2.62	2.55
40	5.24	(3.58, 6.90)	5.19	(3.68, 6.70)	3.32	3.02
41	10.63	(2.53, 18.72)	10.77	(5.06, 16.47)	16.18	11.41
42	4.38	(2.37, 6.39)	4.28	(2.37, 6.20)	4.03	3.84
43	3.85	(2.38, 5.33)	3.87	(2.50, 5.25)	2.94	2.75
44	6.01	(1.60, 10.42)	6.11	(2.40, 9.82)	8.82	7.42
45	5.58	(3.05, 8.12)	5.39	(3.22, 7.56)	5.07	4.34
46	4.42	(2.15, 6.68)	4.28	(2.20, 6.37)	4.53	4.17
47	5.05	(3.65, 6.45)	5.07	(3.78, 6.36)	2.81	2.58
48	3.81	(2.70, 4.92)	3.81	(2.73, 4.89)	2.23	2.17

ACKNOWLEDGMENTS

The authors would like to thank the referees and the associate editor for their constructive comments that led to substantial improvement of the paper.

REFERENCES

Abbasi, A. and Özkale, M. R. (2021). The r-k class estimator in generalized linear models applicable with simulation and empirical study using a Poisson and Gamma responses. *Hacettepe Journal of Mathematics and Statistics*, 50(2):594–611.

Abonazel, M. R., Dawoud, I., Awwad, F. A., and Eldin, E. T. (2023). New estimators for the probit regression model with multicollinearity. *Scientific African*, e01565.

Ahmed, E., Belaghi, R. A., Asar, Y., and Hussein, A. (2021). Ridge-type shrinkage estimators in low and high dimensional beta regression model with application in econometrics and medicine. *arXiv*, preprint arXiv:2111.13887.

Akram, M. N., Amin, M., Kibria, G. B. M., Arashi, M., Lukman, A. F., and Afzal, N. (2022). A new improved Liu estimator for the QSAR model with inverse Gaussian response. *Communications in Statistics – Simulation and Computation*, 1–16.

Algamal, Z. Y. (2019). Performance of ridge estimator in inverse Gaussian regression model. *Communications in Statistics – Theory and Methods*, 48(15):3836–3849.

Algamal, Z. Y., Lukman, A. F., Abonazel, M. R., and Awwad, F. A. (2022). Performance of the ridge and Liu estimators in the zero-inflated Bell regression model. *Journal of Mathematics*.

Alheety, M. I., Qasim, M., Månsson, K., and Kibria, G. B. M. (2021). Modified almost unbiased two-parameter estimator for the Poisson regression model with an application to accident data. *SORT – Statistics and Operations Research Transactions*, 45(2):121–142.

Awwad, F. A., Odeniyi, K. A., Dawoud, I., Algamal, Z., Abonazel, M. R., Kibria, G. B. M., and Eldin, E. T. (2022). New two-parameter estimators for the logistic regression model with multicollinearity. *WSEAS Transactions on Mathematics*, 21:403–414.

Amin, M., Qasim, M., and Amanullah, M. (2019). Performance of Asar and Genç and Huang and Yang's two-parameter estimation methods for the gamma regression model. *Iranian Journal of Science and Technology, Transactions A: Science*, 43(6):2951–2963.

Amin, M., Qasim, M., Amanullah, M., and Afzal, S. (2020). Performance of some ridge estimators for the gamma regression model. *Statistical Papers*, 61(3):997–1026.

Asar, Y. and Genç, A. (2018). A new two-parameter estimator for the Poisson regression model. *Iranian Journal of Science and Technology, Transactions A: Science*, 42(2):793–803.

Bulut, Y. M. and Işılar, M. (2021). Two parameter ridge estimator in the inverse Gaussian regression model. *Hacettepe Journal of Mathematics and Statistics*, 50(3):895–910.

Bulut, Y. M. (2021). Inverse Gaussian Liu-type estimator. *Communications in Statistics – Simulation and Computation*, 1–16.

Dean, C., Lawless, J. F., and Willmot, G. E. (1989). A mixed Poisson-inverse-Gaussian regression model. *The Canadian Journal of Statistics*, 17(2):171–181.

Frisch, R. (1934). *Statistical Confluence Analysis by means of Complete Regression Systems*. University Institute of Economics, Oslo.

Gómez-Deniz, E. and Calderín-Ojeda, E. (2018). Properties and applications of the Poisson-reciprocal inverse Gaussian distribution. *Journal of Statistical Computation and Simulation*, 88(2):269–289.

Hilbe, J. M. (2014). *Modeling Count Data*. New York, Cambridge University Press.

Hoerl, A. E. and Kennard, R. W. (1970a). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Hoerl, A. E. and Kennard, R. W. (1970b). Ridge regression: applications to nonorthogonal problems. *Technometrics*, 12(1):69–82.

Kaçıranlar, S. and Dawoud, I. (2018). On the performance of the Poisson and the negative binomial ridge predictors. *Communications in Statistics – Simulation and Computation*, 47(6):1751–1770.

Khalaf, G., Månsson, K., Sjölander, P., and Shukur, G. (2014). A tobit ridge regression estimator. *Communications in Statistics – Theory and Methods*, 43:131–140.

Kibria, G. B. M. (2022). More than hundred (100) estimators for estimating the shrinkage parameter in a linear and generalized linear ridge regression models. *Journal of Econometrics and Statistics*, 2(2):233–252.

Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics – Theory and Methods*, 22(2):393–402.

Liu, K. (2003). Using Liu-type estimator to combat collinearity. *Communications in Statistics – Theory and Methods*, 32(5):1009–1020.

Locking, H., Månsson, K., and Shukur, G. (2013). Performance of some ridge parameters for probit regression: with application to Swedish job search data. *Communications in Statistics – Simulation and Computation*, 42:698–710.

Lukman, A. F., Aladeitan, B. Ayinde, K., and Abonazel, M. R. (2022). Modified ridge-type for the Poisson regression model: simulation and application. *Communications in Statistics – Simulation and Computation*, 49(8):2124–2136.

Lukman, A. F., Ayinde, K., Kibria, G. B. M., and Adewuyi, E. T. (2022). Modified ridge-type estimator for the gamma regression model. *Communications in Statistics – Simulation and Computation*, 51(9):5009–5023.

Mahmoudi, A., Arabi Belaghi, R., and Mandal, S. (2020). A comparison of preliminary test, Stein-type and penalty estimators in gamma regression model. *Journal of Statistical Computation and Simulation*, 90(17):3051–3079.

Månsson, K. and Shukur, G. (2011). A Poisson ridge regression estimator. *Economic Modelling*, 28(4):1475–1481.

Månsson, K. (2012). On ridge estimators for the negative binomial regression model. *Economic Modelling*, 29(2):178–184.

McCullagh, P. and Nelder, J. (1989). *Generalized Liner Models*. London, Chapman and Hall.

McDonald, G. C. and Galarneau, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, 70(350):407–416.

Myers, R. H. and Montgomery, D. C. (1997). A tutorial on generalized linear models. *Journal of Quality Technology*, 29(3):274–291.

Myers, R. H., Montgomery, D. C., and Vining, G. G. (2002). *Generalized Linear Models: With Applications in Engineering and the Sciences*. New York, J. Wiley.

Nooi Asl, M., Bevrani, H., Arabi Belaghi, R., and Månsson, K. (2021). Ridge-type shrinkage estimators in generalized linear models with an application to prostate cancer data. *Statistical Papers*, 62: 1043–1085.

Omer, T., Sjölander, P., Månsson, K., and Kibria, G. B. M. (2021). Improved estimators for the zero-inflated Poisson regression model in the presence of multicollinearity: simulation and application of maternal death data. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 7(3):394–412.

Özkale, M. R. (2019). The r-d class estimator in generalized linear models: applications on gamma, Poisson and binomial distributed responses. *Journal of Statistical Computation and Simulation*, 89(4):615–640.

Özkale, M. R. and Kaçıranlar, S. (2007). The restricted and unrestricted two-parameter estimators. *Communications in Statistics – Theory and Methods*, 36(15):2707–2725.

Putri, G. N., Nurrohmah, S., and Fithriani, I. (2020). Comparing Poisson-inverse Gaussian model and negative binomial model on case study: horseshoe crabs data. *Journal of Physics: Conference Series*, 1442:1–5.

Qasim, M., Månsson, K., Amin, M., Kibria, G. B. M., and Sjölander, P. (2020). Biased adjusted Poisson ridge estimators-method and application. *Iranian Journal of Science and Technology, Transactions A: Science*, 44(6):1775–1789.

Qasim, M., Månsson, K., and Kibria, G. B. M. (2021). On some beta ridge regression estimators: method, simulation and applications. *Journal of Statistical Computation and Simulation*, 91(9):1699–1712.

Qasim, M., Månsson, K., Sjölander, P., and Kibria, G. B. M. (2022). A new class of efficient and debiased two-step shrinkage estimators: method and application. *Journal of Applied Statistics*, 49(16):4181–4205.

Schaefer, R. L., Roi, L. D., and Wolfe, R. A. (1984). A ridge logistic estimator. *Communications in Statistics – Theory and Methods*, 13:99–113.

Segerstedt, B. (1992). On ordinary ridge regression in generalized linear models. *Communications in Statistics – Theory and Methods*, 21(8):2227–2246.

Stein, G. Z., Zucchini, W., and Juritz, J. M. (1987). Parameter estimation for the Sichel distribution and its multivariate extension. *Journal of the American Statistical Association*, 82(399):938–944.

Toker, S., Şiray, G. Ü., and Qasim, M. (2019). *Developing a first order two parameter estimator for generalized linear model*. In "11th International Statistics Congress".

Toker, S., Özbay, N., Şiray, G. Ü., and Yenilmez, I. (2021). Tobit Liu estimation of censored regression model: an application to Mroz data and a Monte Carlo simulation study. *Journal of Statistical Computation and Simulation*, 91(6):1061–1091.

Willmot, G. E. (1987). The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. *Scandinavian Actuarial Journal*, 3-4:113–127.

Zha, L., Lord, D., and Zou, Y. (2016). The Poisson inverse Gaussian (PIG) generalized linear regression model for analyzing motor vehicle crash data. *Journal of Transportation Safety and Security*, 8(1):18–35.

REVSTAT — Statistical journal

Aims and Scope

The aim of REVSTAT — Statistical Journal is to publish articles of high scientific content, developing Statistical Science focused on innovative theory, methods, and applications in different areas of knowledge. Important survey/review contributing to Probability and Statistics advancement is also welcome.

Background

Statistics Portugal started in 1996 the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, a quarterly publication whose goal was the publication of papers containing original research results, and application studies, namely in the economic, social, and demographic fields. Statistics Portugal was aware of how vital statistical culture is in understanding most phenomena in the present-day world, and of its responsibilities in disseminating statistical knowledge.

In 1998 it was decided to publish papers in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work. At the time, the editorial board was mainly composed by Portuguese university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal. In 2001, the *Revista de Estatística* published a three volumes special issue containing extended abstracts of the invited and contributed papers presented at the 23rd European Meeting of Statisticians (EMS). During the EMS 2001, its editor-in-chief invited several international participants to join the editorial staff.

In 2003 the name changed to REVSTAT — Statistical Journal, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

Editorial policy

REVSTAT — Statistical Journal is an open access peer-reviewed journal published quarterly, in English, by Statistics Portugal.

The editorial policy of REVSTAT is mainly placed on the originality and importance of the research. The whole submission and review processes for REVSTAT are conducted exclusively online on the journal's webpage revstat.ine.pt based in Open Journal System (OJS). The only working language allowed is English. Authors intending to submit any work must register, login and follow the guidelines.

There are no fees for publishing accepted manuscripts that will be made available in open access.

All articles consistent with REVSTAT aims and scope will undergo scientific evaluation by at least two reviewers, one from the Editorial Board and another external. Authors can suggest

an editor or reviewer who is expert on the paper subject providing her/his complete information, namely: name, affiliation, email and, if possible, personal URL or ORCID number.

 Also, in the context of archiving policy, REVSTAT is a *blue* journal welcoming authors to deposit their works in other scientific repositories regarding the use of the published edition and providing its source.

Abstract and Indexing services

REVSTAT — Statistical Journal is covered *by Journal Citation Reports - JCR (Clarivate); Current Index to Statistics; Google Scholar; Mathematical Reviews® (MathSciNet®); Zentralblatt für Mathematic; Scimago Journal & Country Rank; Scopus*

Author guidelines

The whole submission and review processes for REVSTAT are conducted exclusively online on the journal's webpage https://revstat.ine.pt/ based in Open Journal System (OJS). Authors intending to submit any work must ***register***, ***login*** and follow the indications choosing ***Submissions***.

REVSTAT — Statistical Journal adopts the COPE guidelines on publication ethics.

Work presentation

- the only working language is English;
- the first page should include the name, ORCID iD (optional), Institution, country, and mail-address of the author(s);
- a summary of fewer than one hundred words, followed by a maximum of six keywords and the MSC 2020 subject classification should be included also in the first page;
- manuscripts should be typed only in black, in double-spacing, with a left margin of at least 3 cm, with numbered lines, and a maximum of 25 pages;
- the title should be with no more than 120 characters (with spaces);
- figures must be a minimum of 300dpi and will be reproduced online as in the original work, however, authors should take into account that the printed version is always in black and grey tones;
- authors are encouraged to submit articles using LaTeX which macros are available at *REVSTAT style*;
- citations in text should be included in the text by name and year in parentheses, as in the following examples: § article title in lowercase (Author 1980); § This theorem was proved later by AuthorB and AuthorC (1990); § This subject has been widely addressed (AuthorA 1990; AuthorB et al. 1995; AuthorA and AuthorB 1998);
- references should be listed in alphabetical order of the author's scientific surname at the end of the article;
- acknowledgments of people, grants or funds should be placed in a short section before the References title page. Note that religious beliefs, ethnic background, citizenship and political orientations of the author(s) are not allowed in the text;

- authors are welcome to suggest one of the Editors or Associate Editors or yet other reviewer expert on the subject providing a complete information, namely: name, affiliation, email, and personal URL or ORCID number in the Comments for the Editor (submission form).

Accepted papers

After final revision and acceptance of an article for publication, authors are requested to provide the corresponding LaTeX file, as in REVSTAT style.

Supplementary files may be included and submitted separately in .tiff, .gif, .jpg, .png .eps, .ps or .pdf format. These supplementary files may be published online along with an article, containing data, programming code, extra figures, or extra proofs, etc; however, REVSTAT is not responsible for any supporting information supplied by the author(s).

Copyright Notice

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, Statistics Portugal, to ensure the widest possible dissemination of information.

According to REVSTAT's archiving policy, after assigning the copyright form, authors may cite and use limited excerpts (figures, tables, etc.) of their works accepted/published in REVSTAT in other publications and may deposit only the published edition in scientific repositories providing its source as REVSTAT while the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.