ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Volume 18, No.4
October 2020

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E., 2003- . - 30 cm
Trimestral. - Continuação de : Revista de Estatística = ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726 ; e-ISSN 2183-0371

INDEX

Abstracted/indexed in: *Current Index to Statistics, Google Scholar, Mathematical Reviews, Science Citation Index Expanded®, Zentralblatt für Mathematic, Scimago Journal & Country Rank* and *SCOPUS.*

REVSTAT – Statistical Journal
Volume 18, Number 4, October 2020, 397–414

BLOCK BOOTSTRAP PREDICTION INTERVALS FOR GARCH PROCESSES

Authors: Beste Hamiye Beyaztas
– Department of Statistics, Istanbul Medeniyet University, Istanbul, Turkey
beste.sertdemir@medeniyet.edu.tr

Ufuk Beyaztas
– Department of Economics and Finance, Piri Reis University, Istanbul, Turkey
ubeyaztas@pirireis.edu.tr

Received: December 2017 Revised: April 2018 Accepted: April 2018

Abstract:

- In this paper, we propose a new resampling algorithm based on block bootstrap to obtain prediction intervals for future returns and volatilities of GARCH processes. The finite sample properties of the proposed methods are illustrated by an extensive simulation study and they are applied to Japan Yen (JPY) / U.S. dollar (USD) daily exchange rate data. Our results indicate that:
  (**i**) the proposed algorithm is a good competitor or even better and
  (**ii**) computationally more efficient than traditional method(s).

Keywords:

- *financial time series; prediction; resampling methods; exchange rate.*

AMS Subject Classification:

- 62F40, 92B84, 62M20.

1. INTRODUCTION

Many macroeconomic and financial time series vary over wide range around mean, and very large or small prediction errors may occur in practice. Since financial markets are sensitive to political events, speculations, changes in monetary policy etc., this variability in the error terms may occur. This implies that the variance of the errors may not be constant and it changes over time so the errors can be serially correlated in financial data. Additionally, one of the uncertain and decisive factors in financial time series analysis is the volatility as a measure of dispersion and an indicator of magnitude of fluctuations of the asset price series. Hence, measuring volatility as well as construction of valid predictions for future returns and volatilities have an important role in assessing risk and uncertainty in the financial market. Since volatility is the unobservable component of financial time series, it should be modeled correctly to obtain efficient parameter estimation and improve the accuracy of prediction intervals for assessing uncertainty in risk management. In this context, the generalized autoregressive conditionally heteroscedastic (GARCH) model proposed by [7] is one of the most commonly used technique for modeling volatility and obtaining dynamic prediction intervals for returns as well as volatilities. See [4], [22], [13] and [28] for recent studies on GARCH model in modelling volatility. Also see [1], [2], [3], and [15] for detailed information about construction of prediction intervals for future returns in financial time series analysis. However, those works only consider point forecast of volatility even though prediction intervals provide better inference taking into account uncertainty of unobservable sequence of volatilities. On the other hand, construction of prediction intervals requires some distributional assumptions which are generally unknown in practice. Moreover, they can be affected due to any departure from the assumptions and may lead us to unreliable results. One remedy to construct prediction intervals without considering distributional assumptions is to apply the well known resampling methods, such as the bootstrap.

For the serially correlated data, the method of block bootstrap is one of the most general tool to approximate the properties of estimators. In this technique the underlying idea is to construct a resample of the data of size n by dividing the data into several blocks with a sufficiently large block length ℓ and choosing among them till the bootstrap sample is obtained. Then, the dependence structure of the original data is attempted to be captured by these ℓ consecutive observations in each block drawn independently. The commonly used block bootstrap procedures called "non-overlapping" and "overlapping" are first proposed by [16] in the context of spatial data. Then [10] and [20], respectively, adapted the non-overlapping block bootstrap (NBB) and moving block bootstrap (MBB) approaches to the univariate time series context. In addition to these methods, [26] introduced the circular block bootstrap (CBB) method by wrapping the data around a circle before blocking them. Also, the stationary bootstrap (SB) method which deals with random block lengths is proposed by [25]. Moreover, Ordered non-overlapping block bootstrap (ONBB), which orders the bootstrapped blocks according to given labels to each original block, was suggested by [6] to improve the performance of the block bootstrap technique by taking into account the correlations between the blocks.

Bootstrap-based prediction intervals of autoregressive conditionally heteroscedastic (ARCH) model for future returns and volatilities are proposed by [23] and [27]. [24] further extends the previous works to GARCH$(1,1)$ model. Later, [11] suggests computationally

efficient bootstrap prediction intervals for ARCH and GARCH processes in the context for financial time series. All of these methods are based on resampling the residuals. The block bootstrap methods are not suitable for construction of prediction intervals in conditionally heteroskedastic time series models because of their poor finite sample performances. On the other hand, it is possible to construct valid block bootstrap based prediction intervals for GARCH processes by using the autoregressive-moving average (ARMA) representation of the GARCH models. For instance, [5] proposed to use the ONBB method to obtain prediction intervals for GARCH process and they obtained better prediction intervals for returns and volatilities compared to the existing residual based bootstrap method(s). Also, [19] introduced a stationary bootstrap prediction interval for GARCH models. In this paper, following the idea of [19], we propose a new bootstrap algorithm to obtain prediction intervals for future returns and volatilities under GARCH processes. In summary, our extension works as follows: First, we use the squares of the GARCH process, which have the ARMA representation, to make the parameter estimation process linear. The ordinary least squares estimators of the ARMA model are calculated by a high order autoregressive model of order m, and the residuals are computed. Then the block bootstrap methods are applied to the data to obtain the bootstrap sample of the returns which are used to calculate the bootstrap estimators of the ARMA coefficients and the bootstrap sample of the volatilities. Finally, the future values of the returns and volatilities of the GARCH process are obtained by means of bootstrap replicates and quantiles of the Monte Carlo estimates of the generated bootstrap distribution.

The rest of the paper is organized as follows. We describe our proposed methods in Section 2. An extensive Monte Carlo simulation is conducted to examine the finite sample performance of the proposed methods and the results are presented in Section 3. In Section 4, the JPY/USD daily exchange rate data is analyzed using the new methods and the results are presented. Section 5 concludes the paper.

2. METHODOLOGY

We use ARMA parameterization of a GARCH model and its least squares (LS) estimators in order to employ block bootstrap methods for constructing prediction intervals.

The GARCH(p, q) process considered in this study has the following representation:

$$\begin{aligned} y_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \omega + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \quad t = 1, ..., T, \end{aligned} \tag{2.1}$$

where $\{\epsilon_t\}$ is a sequence of white noise random variables and $E(\epsilon^4) < \infty$, ω, α_i and β_j are unknown parameters satisfying $\omega > 0$, $\alpha_i \geq 0$ and $\beta_j \geq 0$ for $i = 1, ..., p$ and $j = 1, ..., q$. The stochastic process σ_t is assumed to be independent of ϵ_t. Throughout this paper, we assume that the process $\{y_t\}$ is strictly stationary, i.e., $\sum_{i=1}^{r}(\alpha_i + \beta_i) < 1$, where $r = max(p, q)$, $\alpha_i = 0$ for $i > p$ and $\beta_i = 0$ for $i > q$; see [8] and [9]. A GARCH(p, q) process $\{y_t\}$ is represented in

the form of ARMA as follows:

$$y_t^2 = \omega + \sum_{i=1}^{r}(\alpha_i + \beta_i)y_{t-i}^2 + \nu_t - \sum_{j=1}^{q}\beta_j \nu_{t-j}, \tag{2.2}$$

where the innovation $\nu_t = y_t^2 - \sigma_t^2$ is a white noise (not i.i.d. in general) and identically distributed under the strict stationary assumption of y_t. Using the unconditional mean of the ARMA model given in (2.2), we have

$$E(y_t^2) = \frac{\omega}{1 - \sum_{i=1}^{r}(\alpha_i + \beta_i)}. \tag{2.3}$$

According to [18], the LS estimators of an ARMA model are obtained as follows:

(**a**) First, a high order autoregressive model of order m, AR(m), with $m > max(p, q)$, is fitted to the data by Yule-Walker method to obtain $\widehat{\nu}_t$, where m is determined from the data by using Akaike information criteria or Bayesian information criteria.

(**b**) Then a linear regression of y_t^2 onto $y_{t-1}^2, ..., y_{t-r}^2, \widehat{\nu}_{t-1}, ..., \widehat{\nu}_{t-q}$ is fitted to estimate the parameter vector $\phi = ((\alpha_1 + \beta_1), ..., (\alpha_r + \beta_r), -\beta_1, ..., -\beta_q)'$.

In matrix notations, let $\mathbf{Z}_T$ and $\mathbf{X}$ be as follows:

$$\mathbf{Z}_T = \begin{bmatrix} y_{m+1}^2 \\ \vdots \\ y_T^2 \end{bmatrix}$$

and

$$\mathbf{X} = \begin{bmatrix} y_m^2 & y_{m-1}^2 & \cdots & y_{m-p+1}^2 & \hat{\nu}_m & \hat{\nu}_{m-1} & \cdots & \hat{\nu}_{m-q+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{T-1}^2 & y_{T-2}^2 & \cdots & y_{T-p}^2 & \hat{\nu}_{T-1} & \hat{\nu}_{T-2} & \cdots & \hat{\nu}_{T-q} \end{bmatrix}.$$

Then, the LS estimator $\widehat{\phi} = (\widehat{(\alpha_1 + \beta_1)}, ..., \widehat{(\alpha_r + \beta_r)}, -\widehat{\beta}_1, ..., -\widehat{\beta}_q)'$ is obtained as

$$\widehat{\phi} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_T, \tag{2.4}$$

given $\boldsymbol{X}'\boldsymbol{X}$ is non-singular. The corresponding $\widehat{\alpha}_i$'s are calculated as $\widehat{\alpha}_i = \widehat{(\alpha_i + \beta_i)} - \widehat{\beta}_i$, for $i = 1, ..., p$.

For clarity, we next describe the complete algorithm of the proposed block bootstrap prediction intervals for future returns and volatilities.

Step 1. For a realization of GARCH(p, q) process, $\{y_{1-r}, ..., y_0, y_1, ..., y_T\}$, calculate the LS estimates of ARMA coefficients as in (2.4), and the corresponding $\widehat{\omega}$ is calculated by using (2.3) such that $\widehat{\omega} = E(\hat{y}_t^2)\left[1 - \sum_{i=1}^{r}\widehat{(\alpha_i + \beta_i)}\right]$, where $E(\hat{y}_t^2) = T^{-1}\sum_{t=1}^{T} y_t^2$.

Step 2. For $t = r, ..., T$, calculate the residuals $\widehat{\epsilon}_t = y_t/\widehat{\sigma}_t$ where $\widehat{\sigma}_t^2 = \widehat{\omega} + \sum_{i=1}^{p}\widehat{\alpha}_i y_{t-i}^2 + \sum_{j=1}^{q}\widehat{\beta}_j\widehat{\sigma}_{t-j}^2$ and $\widehat{\sigma}_0^2 = \widehat{\omega}/(1 - \sum_{i=1}^{r}(\widehat{\alpha}_i + \widehat{\beta}_i))$. Let $\widehat{F}_\epsilon$ be the empirical distribution function of the centered and rescaled residuals.

Step 3. Compute the error term as $\hat{\xi} = \mathbf{Z}_T - \mathbf{X}\hat{\phi}$ and construct the design matrix $\mathbf{Y} = (\mathbf{X}, \xi)$.

$$\mathbf{Y} = \begin{bmatrix} y_{t-1}^2 & y_{t-2}^2 & \cdots & y_{t-r}^2 & \hat{\nu}_{t-1} & \hat{\nu}_{t-2} & \cdots & \hat{\nu}_{t-q} & \hat{\xi}_t \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ y_{T-1}^2 & y_{T-2}^2 & \cdots & y_{T-r}^2 & \hat{\nu}_{T-1} & \hat{\nu}_{T-2} & \cdots & \hat{\nu}_{T-q} & \hat{\xi}_T \end{bmatrix}.$$

Let $\mathbf{Y}_t = (y_{t-1}^2, y_{t-2}^2, ..., y_{t-r}^2, \hat{\nu}_{t-1}, \hat{\nu}_{t-2}, ..., \hat{\nu}_{t-q}, \hat{\xi}_t)$, $t = 1, ..., T$, denotes the tth row of the design matrix $\mathbf{Y}$. Let also $\mathbf{B}^{(k)}$, for $k = 1, 2, 3$, respectively, represents the block vectors of NBB, MBB and CBB methods obtained from $\mathbf{Y}$ such that $\mathbf{B}_j^{(1)} = \{Y_{(j-1)\ell+1}, ..., Y_{j\ell}\}$ where $b = \lfloor T/\ell \rfloor$ and $j = 1, ..., b$, $\mathbf{B}_j^{(2)} = \{Y_j, ..., Y_{j+l-1}\}$ where $1 \le j \le N$ and $N = T - \ell + 1$ and $\mathbf{B}_j^{(3)} = \{Y_j, ..., Y_{j+l-1}\}$ where $1 \le j \le T$. Then obtain the block bootstrap observations $\{Y_1^*, ..., Y_T^*\}$, where $\mathbf{Y}_t^* = (y_{t-1}^{2*}, y_{t-2}^{2*}, ..., y_{t-r}^{2*}, \hat{\nu}_{t-1}^*, \hat{\nu}_{t-2}^*, ..., \hat{\nu}_{t-q}^*, \hat{\xi}_t^*)$, by sampling with replacement from $\mathbf{B}^{(k)}$. The ONBB and SB observations are obtained as follows:

- ONBB observations are obtained as ordering the bootstrapped non-overlapping blocks according to given labels to each original block. Suppose the data is divided into the four independent non-overlapping blocks. Then, the labels are determined as $B_1 = 1, B_2 = 2, B_3 = 3$ and $B_4 = 4$, and let the bootstrapped blocks are $B_1^* = B_4, B_2^* = B_2, B_3^* = B_3$ and $B_4^* = B_3$. As a consequence, the ONBB data is obtained as $\{B_2 \vdots B_3 \vdots B_3 \vdots B_4\}$.
- Let $B(i\ell) = (Y_i, ..., Y_{i+\ell-1})$, for $i \ge 1$, be the blocks of ℓ consecutive observations starting from Y_i. The observed time series data is wrapped around a circle in order to ensure that all starting points have equal probability of selection. Let $I_1, I_2, ...$ be the independently and identically distributed discrete uniform random variables on $\{1, ..., T\}$ so that $P(I_1 = i) = 1/T$, for $i = 1, ..., T$. Let also $L_1, L_2, ...$ be the i.i.d. geometric random variables with parameter ρ such that $0 < \rho < 1$ and the probability mass function $P(L_1 = \ell) = \rho(1-\rho)^{\ell-1}$, for $\ell = 1, 2, ...$ We assume that two sets $\{I_1, I_2, ...\}$ and $\{L_1, L_2, ...\}$ are independent and $\rho \to 0$ as $T\rho \to \infty$. Then, the SB data $\{Y_1^*, ..., Y_T^*\}$ are generated by sampling from $\{B_{I_1 L_1}, B_{I_2 L_2}, ...\}$ where $B_{I_r L_r} = \{Y_{I_r}, ..., Y_{I_r + L_r - 1}\}$ for $r \ge 1$.

Step 4. Let $\mathbf{X}^*$ be the bootstrap analogue of X such that

$$\mathbf{X}^* = \begin{bmatrix} y_m^{*2} & y_{m-1}^{*2} & \cdots & y_{m-p+1}^{*2} & \hat{\nu}_m^* & \hat{\nu}_{m-1}^* & \cdots & \hat{\nu}_{m-q+1}^* \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{T-1}^{*2} & y_{T-2}^{*2} & \cdots & y_{T-p}^{*2} & \hat{\nu}_{T-1}^* & \hat{\nu}_{T-2}^* & \cdots & \hat{\nu}_{T-q}^* \end{bmatrix}.$$

Then calculate the block bootstrap estimators of ARMA coefficients as

$$\widehat{\phi}^* = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{Z}_T^* = ((\widehat{\alpha_1 + \beta_1})^*, ..., (\widehat{\alpha_r + \beta_r})^*, -\widehat{\beta}_1^*, ..., -\widehat{\beta}_q^*)',$$

where $\mathbf{Z}_T^* = \mathbf{X}^*\hat{\phi} + \hat{\xi}$. Also, calculate the corresponding $\widehat{\alpha}_i^*$'s as $\widehat{\alpha}_i^* = (\widehat{\alpha_i + \beta_i})^* - \widehat{\beta}_i^*$, for $i = 1, ..., p$, and $\widehat{\omega}^*$'s as in Step 1 but using bootstrap observations.

Step 5. Obtain block bootstrap volatilities as $\widehat{\sigma}_t^{2*} = \widehat{\omega}^* + \sum_{i=1}^p \widehat{\alpha}_i^* y_{t-i}^{2*} + \sum_{j=1}^q \widehat{\beta}_j^* \widehat{\sigma}_{t-j}^{2*}$ with $\widehat{\sigma}_0^{2*} = \widehat{\omega}/(1 - \sum_{i=1}^r (\widehat{\alpha}_i + \widehat{\beta}_i))$.

Step 6. Calculate $h = 1, 2, ...$ steps ahead block bootstrap future returns and volatilities with the following recursions:

$$\begin{aligned}\widehat{\sigma}^{2*}_{T+h} &= \widehat{\omega}^* + \sum_{i=1}^{p} \widehat{\alpha}^*_i y^{2*}_{T+h-i} + \sum_{j=1}^{q} \widehat{\beta}^*_j \widehat{\sigma}^{2*}_{T+h-j}, \\ y^*_{T+h} &= \widehat{\sigma}^{2*}_{T+h} \widehat{\epsilon}^*_{T+h},\end{aligned}$$

where $y^*_{T+h} = y_{T+h}$ for $h \leq 0$ and $\widehat{\epsilon}^*_{T+h}$ is randomly drawn from $\widehat{F}_\epsilon$.

Step 7. Repeat Steps 3-6 B times to obtain bootstrap replicates of returns and volatilities $\{y^{*,1}_{T+h}, ..., y^{*,B}_{T+h}\}$ and $\{\widehat{\sigma}^{2*,1}_{T+h}, ..., \widehat{\sigma}^{2*,B}_{T+h}\}$ for each h. Note that B denotes the number of bootstrap replications.

As noted in [24], the one-step conditional variance is perfectly predictable if the model parameters are known, and the only uncertainty which is caused by the parameter estimation, is associated with the prediction of σ^2_{T+1}. On the other hand, there are further uncertainties about future errors when predicting two or more step ahead variances. Thus, it is more interesting issue to have prediction intervals for future volatilities. Now, let $G^*_y(k) = P(y^*_{T+h} \leq k)$ and $G^*_{\sigma^2}(k) = P(\widehat{\sigma}^{*2}_{T+h} \leq k)$ be the block bootstrap distribution functions of unknown distribution functions of y_{T+h} and σ^2_{T+h}, respectively. Also let $G^*_{y,B}(k) = \#(y^{*,b}_{T+h} \leq k)/B$ and $G^*_{\sigma^2,B}(k) = \#(\widehat{\sigma}^{2*,b}_{T+h} \leq k)/B$, for $b = 1, ..., B$, be the corresponding Monte Carlo (MC) estimates. Then, the $100(1-\gamma)\%$ bootstrap prediction intervals for y_{T+h} and σ^2_{T+h}, respectively, are given by

$$\begin{aligned}\left[LB^*_{y,B}, UB^*_{y,B}\right] &= \left[Q^*_{y,B}(\gamma/2), Q^*_{y,B}(1-\gamma/2)\right], \\ \left[LB^*_{\sigma^2,B}, UB^*_{\sigma^2,B}\right] &= \left[Q^*_{\sigma^2,B}(\gamma/2), Q^*_{\sigma^2,B}(1-\gamma/2)\right],\end{aligned}$$

where $Q^*_{y,B} = G^{*-1}_y$ and $Q^*_{\sigma^2,B} = G^{*-1}_{\sigma^2}$.

3. NUMERICAL RESULTS

We performed a simulation study to investigate the performances of the block bootstrap prediction intervals constructed through the GARCH(1, 1) model given in (3.1) below, and we compared our results with the method proposed by [24] (abbreviated as "PRR"). In brief, the PRR method uses quasi-maximum likelihood method to estimate the parameters and then, uses residual-based resampling to construct prediction intervals for future returns and volatilities. The comparison was made through the coverage probabilities and length of prediction intervals. It is worth the mention that we also checked the performances of the conventional block bootstrap methods. Roughly, we observed the coverage probabilities of other block bootstrap methods range in between 90%-94% for future returns while those range only in between 25%-60% for future volatilities. These results are not shown to save space, but are available from the authors upon request.

To discuss the numerical study we present here, let us start with the following GARCH(1, 1) model:

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= 0.05 + 0.1 y_{t-1}^2 + 0.85 \sigma_{t-1}^2, \end{aligned} \tag{3.1}$$

where ϵ_t follows a $N(0,1)$ distribution. The significance level γ is set to 0.05 to obtain 95% prediction intervals for future returns and volatilities. Since the block bootstrap methods are sensitive to the choice of the block length ℓ, we choose three different block lengths in our simulation study: $T^{1/3}, T^{1/4}, T^{1/5}$ as proposed by [17]. Let $h = 1, 2, ..., s$, $s \geq 1$, be defined as the lead time. We obtain the prediction intervals for next $s = 20$ observations. The experimental design is similar to those of [24] which is as follows:

Step 1. Simulate a GARCH(1, 1) series with the parameters given in equation (3.1), for $h = 1, ..., s$, generate $R = 1000$ future values y_{T+h} and σ_{T+h}^2 to calculate the average coverage probabilities and interval lengths (as well as their standard errors) for the prediction intervals.

Step 2. Calculate bootstrap future values $y_{T+h}^{*,b}$ and $\sigma_{T+h}^{2*,b}$ for $h = 1, ..., s$ and $b = 1, ..., B$. Then estimate the coverage probabilities (C^*) of bootstrap prediction intervals for y_{T+h}^* and σ_{T+h}^{2*} as

$$C_{y_{T+h}}^* = \frac{1}{R} \sum_{r=1}^{R} \mathbf{1}\{Q_{y_{T+h}}^*(\gamma/2) \leq y_{T+h}^{*,r} \leq Q_{y_{T+h}}^*(1-\gamma/2)\},$$

$$C_{\sigma_{T+h}^2}^* = \frac{1}{R} \sum_{r=1}^{R} \mathbf{1}\{Q_{\sigma_{T+h}^2}^*(\gamma/2) \leq \sigma_{T+h}^{2*,r} \leq Q_{\sigma_{T+h}^2}^*(1-\gamma/2)\},$$

where $\mathbf{1}$ represents the indicator function. The corresponding interval lengths (L^*) are calculated by

$$\begin{aligned} L_{y_{T+h}}^* &= Q_{y_{T+h}}^*(1-\gamma/2) - Q_{y_{T+h}}^*(\gamma/2), \\ L_{\sigma_{T+h}^2}^* &= Q_{\sigma_{T+h}^2}^*(1-\gamma/2) - Q_{\sigma_{T+h}^2}^*(\gamma/2). \end{aligned}$$

Step 3. Repeat Steps 1-2, $MC = 1000$ times to calculate the average values of $C_{y_{T+h}}^*$, $C_{\sigma_{T+h}^2}^*$, $L_{y_{T+h}}^*$ and $L_{\sigma_{T+h}^2}^*$.

Our results showed that the accuracy of the prediction intervals for volatilities are sensitive to the choice of block length parameter ℓ. The higher coverage probabilities are obtained for all the methods when $\ell = T^{1/5}$ is used, therefore to save space we present only the results obtained for the choices of block length parameter $\ell = T^{1/5}$. Table 1 summarizes the simulation results. More detailed results are presented in Figures 1–4. Our findings show that ONBB outperforms PRR and other block bootstrap methods in general. For coverage probabilities of future returns (see Figure 1), the performances of all the methods are almost the same. Also, all the proposed methods provide competitive interval lengths for returns (see Figure 3). For the prediction intervals of volatilities (please see Figure 4), the performance of ONBB is always better than PRR and other block bootstrap methods in small sample sizes especially for short-term forecasts, and it outperforms other methods also in large samples. PRR has better performances compared to non-ordered block bootstrap methods for short term forecasts, and all the methods have similar performances for long term forecasts. We note that the results obtained by MBB and CBB methods are quite similar, therefore to make the results more readable we present the results only for the CBB method.

Table 1: Prediction intervals for returns and volatilities of GARCH(1, 1) model.

Lead time	Sample size	Method	Average coverage for return (SE)	Average length for return (SE)	Average coverage for volatility (SE)	Average length for volatility (SE)
1	T	Empirical	0.95	3.814	0.95	—
	300	PRR	0.945 (0.021)	3.748 (0.874)	0.904 (0.295)	0.649 (0.520)
		ONBB	0.943 (0.022)	3.690 (0.704)	0.949 (0.220)	0.720 (0.592)
		NBB	0.941 (0.041)	3.739 (0.562)	0.847 (0.360)	0.986 (0.528)
		CBB	0.941 (0.042)	3.737 (0.558)	0.850 (0.357)	0.991 (0.536)
		SB	0.941 (0.042)	3.731 (0.564)	0.846 (0.361)	1.001 (0.544)
	3000	PRR	0.946 (0.011)	3.800 (0.863)	0.952 (0.214)	0.181 (0.194)
		ONBB	0.948 (0.015)	3.815 (0.793)	0.995 (0.070)	0.803 (0.740)
		NBB	0.948 (0.045)	3.889 (0.343)	0.892 (0.310)	1.224 (0.297)
		CBB	0.948 (0.046)	3.886 (0.340)	0.897 (0.304)	1.230 (0.297)
		SB	0.948 (0.045)	3.888 (0.347)	0.885 (0.319)	1.232 (0.300)
10	T	Empirical	0.95	3.946	0.95	1.389
	300	PRR	0.943 (0.026)	3.846 (0.712)	0.902 (0.117)	1.564 (1.387)
		ONBB	0.938 (0.025)	3.723 (0.530)	0.921 (0.113)	1.541 (1.181)
		NBB	0.937 (0.032)	3.738 (0.497)	0.898 (0.141)	1.547 (0.943)
		CBB	0.937 (0.032)	3.736 (0.503)	0.902 (0.136)	1.549 (0.944)
		SB	0.936 (0.032)	3.721 (0.499)	0.896 (0.141)	1.516 (0.923)
	3000	PRR	0.946 (0.012)	3.875 (0.604)	0.941 (0.036)	1.354 (0.653)
		ONBB	0.947 (0.014)	3.867 (0.584)	0.955 (0.059)	1.582 (0.967)
		NBB	0.947 (0.029)	3.901 (0.270)	0.939 (0.097)	1.670 (0.531)
		CBB	0.947 (0.029)	3.907 (0.275)	0.939 (0.098)	1.669 (0.533)
		SB	0.947 (0.029)	3.897 (0.278)	0.932 (0.103)	1.647 (0.541)
20	T	Empirical	0.95	3.948	0.95	1.661
	300	PRR	0.940 (0.026)	3.876 (0.647)	0.881 (0.122)	1.771 (1.515)
		ONBB	0.935 (0.026)	3.741 (0.507)	0.903 (0.119)	1.646 (0.990)
		NBB	0.934 (0.029)	3.746 (0.502)	0.895 (0.128)	1.635 (0.911)
		CBB	0.934 (0.029)	3.740 (0.498)	0.898 (0.125)	1.640 (0.900)
		SB	0.933 (0.029)	3.727 (0.499)	0.895 (0.126)	1.623 (0.919)
	3000	PRR	0.946 (0.012)	3.907 (0.444)	0.940 (0.033)	1.634 (0.627)
		ONBB	0.946 (0.014)	3.895 (0.460)	0.949 (0.063)	1.861 (0.972)
		NBB	0.946 (0.020)	3.910 (0.255)	0.948 (0.073)	1.876 (0.595)
		CBB	0.946 (0.020)	3.913 (0.255)	0.948 (0.071)	1.872 (0.583)
		SB	0.946 (0.020)	3.900 (0.259)	0.946 (0.073)	1.859 (0.598)

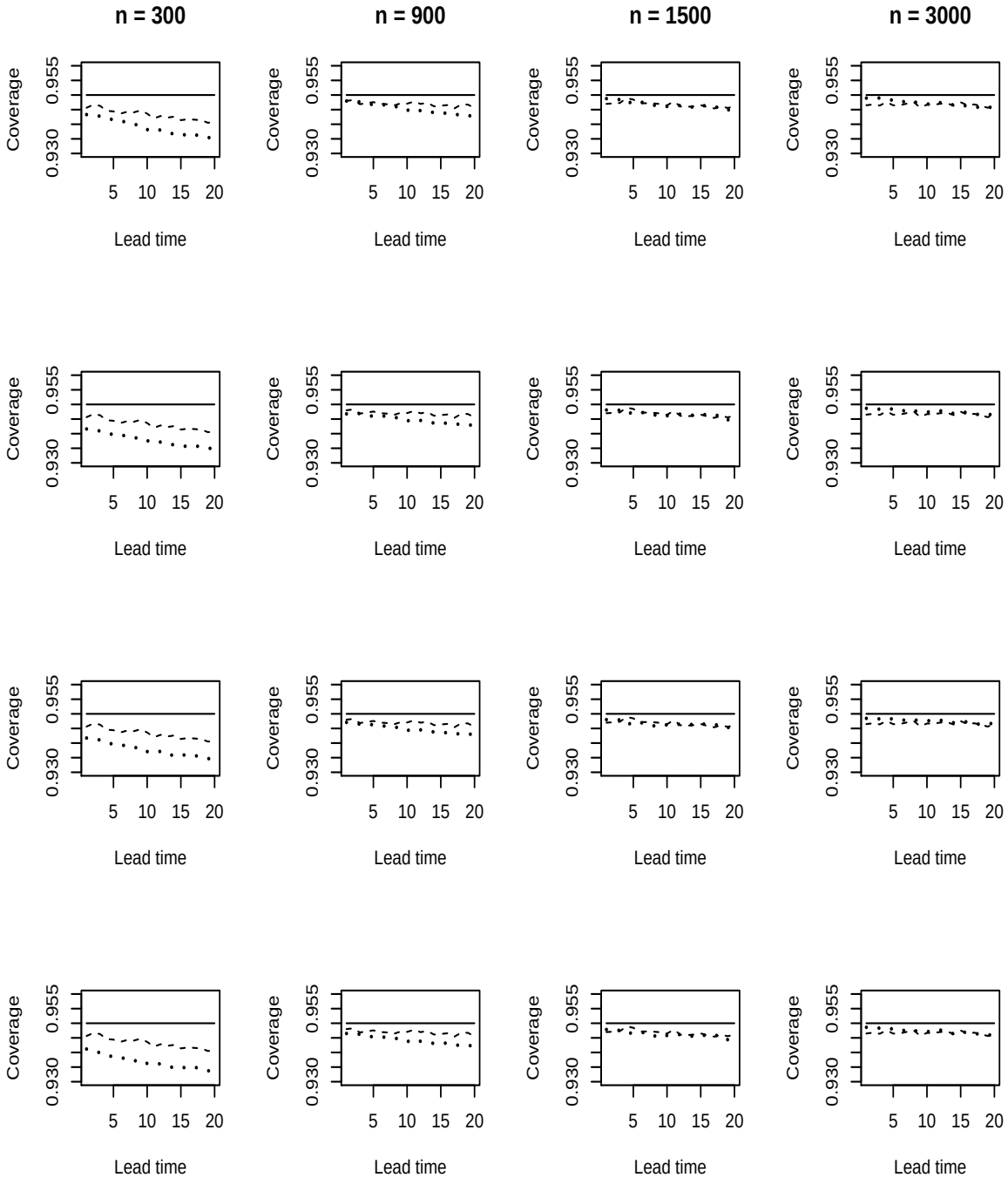

Figure 1: Estimated coverage probabilities of returns. First line: ONBB vs PRR, second line: NBB vs PRR, third line: CBB vs PRR, fourth line: SB vs PRR. Solid line represents the empirical coverage. Dashed line and dotted line represent the coverage probabilities obtained using PRR and proposed methods, respectively.

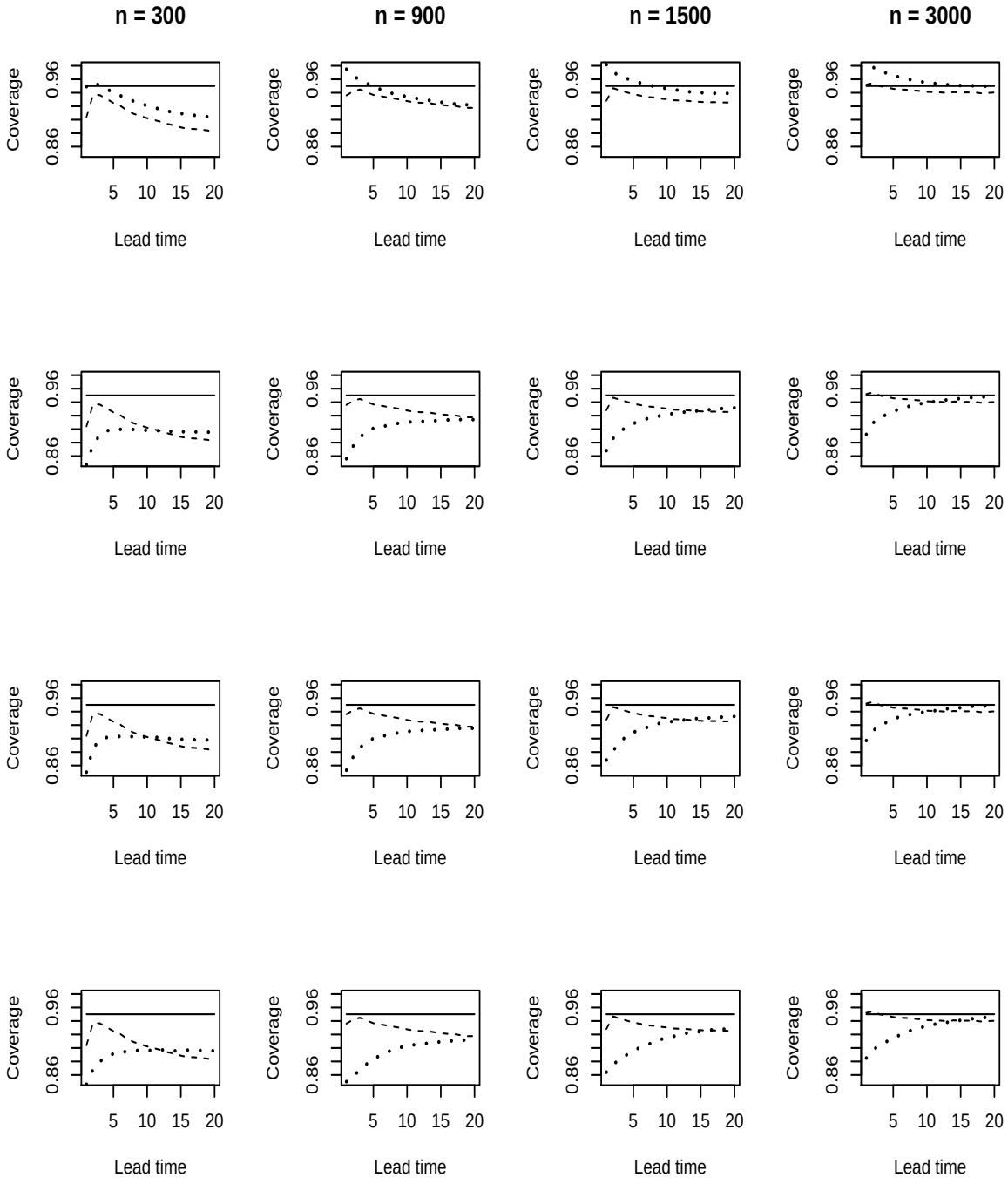

Figure 2: Estimated coverage probabilities of volatilities. First line: ONBB vs PRR, second line: NBB vs PRR, third line: CBB vs PRR, fourth line: SB vs PRR. Solid line represents the empirical coverage. Dashed line and dotted line represent the coverage probabilities obtained using PRR and proposed methods, respectively.

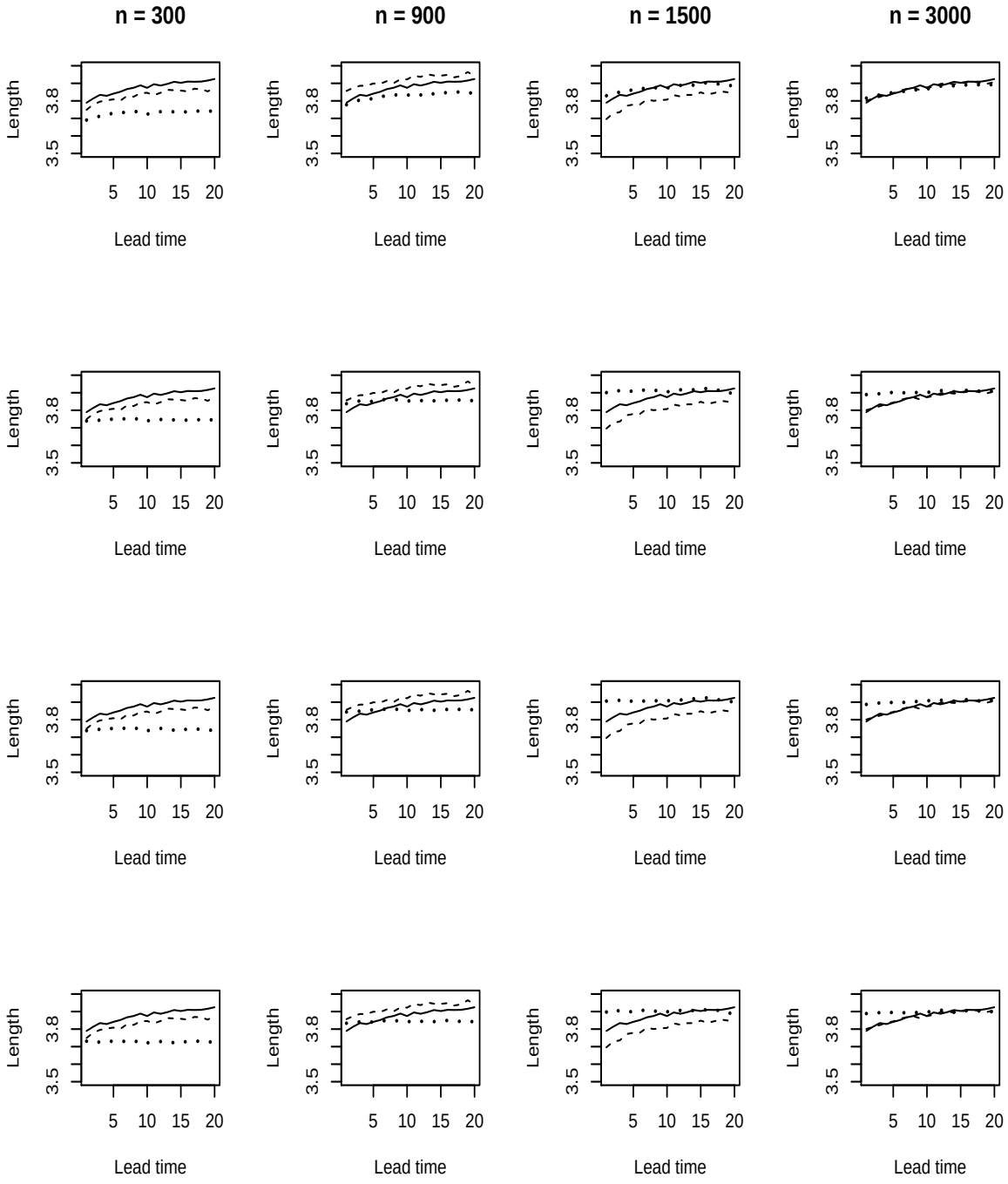

Figure 3: Estimated lengths of prediction intervals of returns. First line: ONBB vs PRR, second line: NBB vs PRR, third line: CBB vs PRR, fourth line: SB vs PRR. Solid line represents the empirical interval lengths. Dashed line and dotted line represent the interval lengths obtained using PRR and proposed methods, respectively.

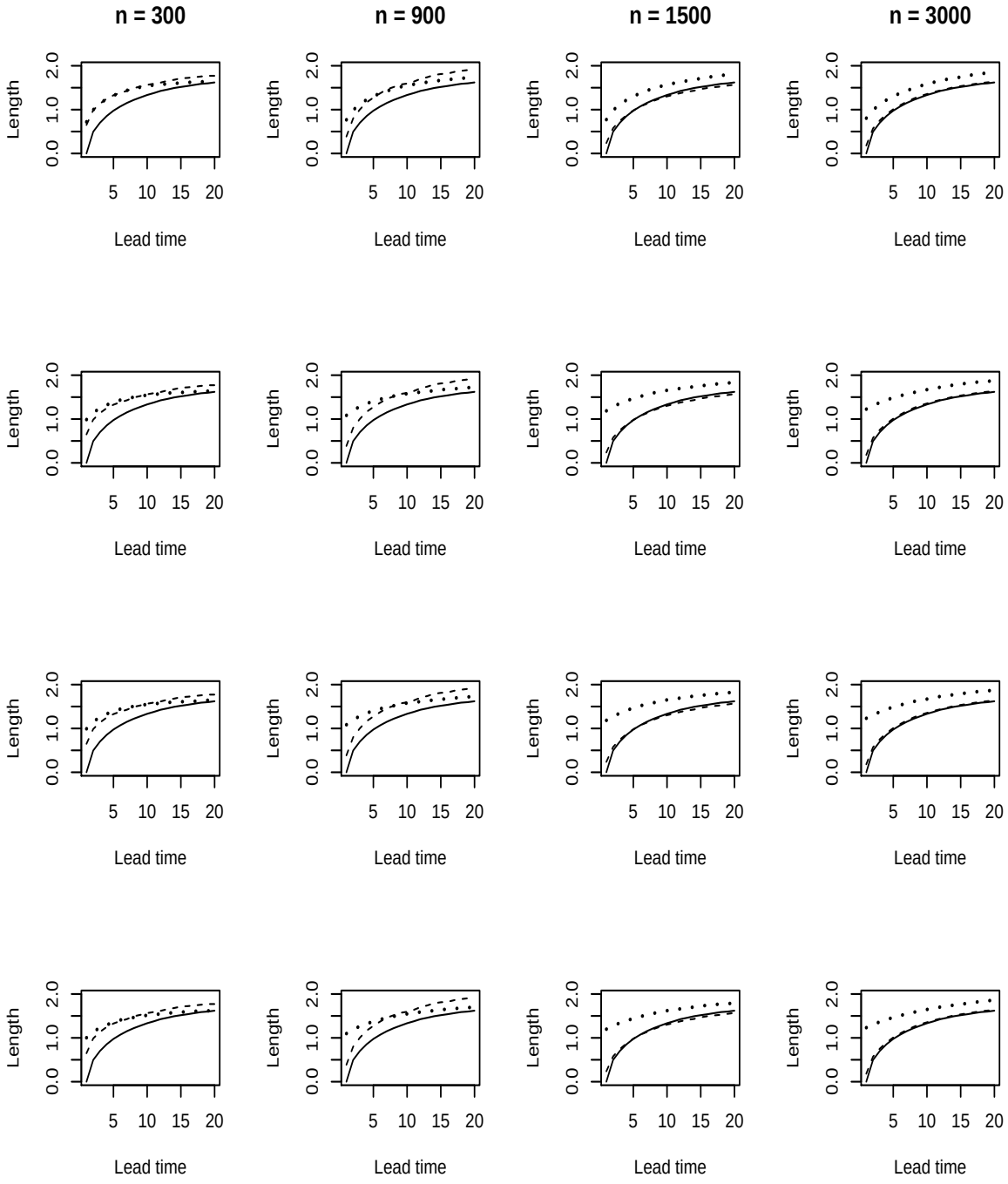

Figure 4: Estimated lengths of prediction intervals of volatilities. First line: ONBB vs PRR, second line: NBB vs PRR, third line: CBB vs PRR, fourth line: SB vs PRR. Solid line represents the empirical interval lengths. Dashed line and dotted line represent the interval lengths obtained using PRR and proposed methods, respectively.

We also compared our proposed algorithm with the PRR in terms of their computing times. Let c_1 and c_2 be the obtained computing times for PRR and proposed algorithm, respectively. Figure 5 represents the ratio of computing times, c_1/c_2, for various sample sizes based on $B = 1000$ bootstrap replications and only one Monte Carlo simulation. As presented in Figure 5, the proposed algorithm has considerably less computational time such that PRR requires about 36–12 times more computing time (in small and large samples, respectively) than the proposed algorithm.

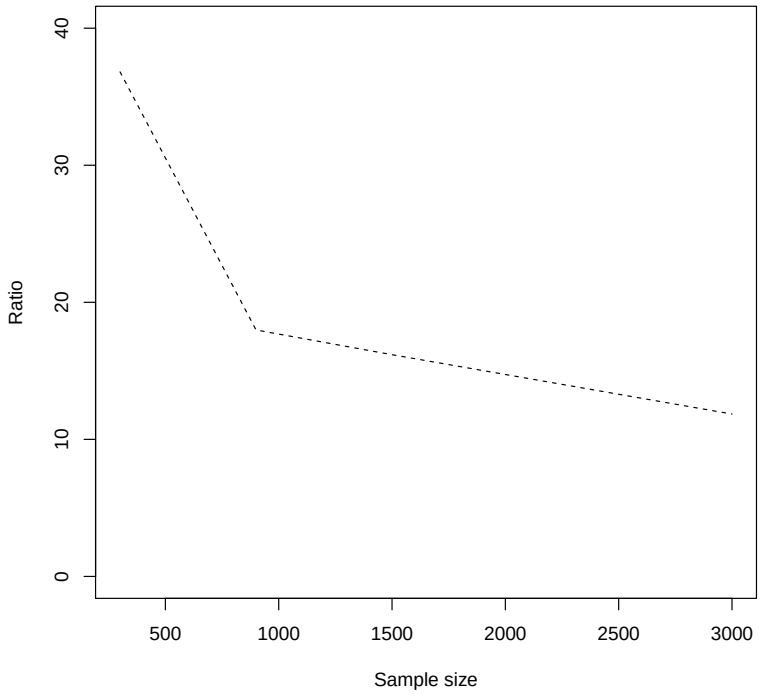

Figure 5: Ratio of the estimated computing times for PRR and proposed algorithm.

4. CASE STUDY

The JPY/USD daily exchange rate data were obtained starting from 3rd January, 2011 and ending on 30th April, 2015 (available at https://www.stlouisfed.org/). After excluding observations on weekends and inactive days, our final data consisted a total of 1071 observations. The daily logarithmic returns were obtained as $y_t = 100 * \log(P_t/P_{t-1})$, where P_t was the closing price on t-th day. The time series plots of the exchange rates and returns are presented in Figure 6. We checked the stationary status of the return series by applying the Ljung-Box and Augmented Dickey-Fuller t-statistic tests and small p-values reject the null hypothesis against stationary alternative and suggest that the return series is a mean-zero stationary process. Table 2 reports the sample statistics of y_t series, and it shows that the estimated kurtosis is higher than 3 which indicates that the distribution of the returns was leptokurtic. Next, we checked for the Gaussianity of the return series and the p-value $= 0.000$ of Jarque-Bera test indicated that y_t was not Gaussian. Further, we performed the Box-Pierce test to test for auto-correlations in the absolute and squared returns and smaller p-values indicated that the absolute and squared returns are highly auto-correlated.

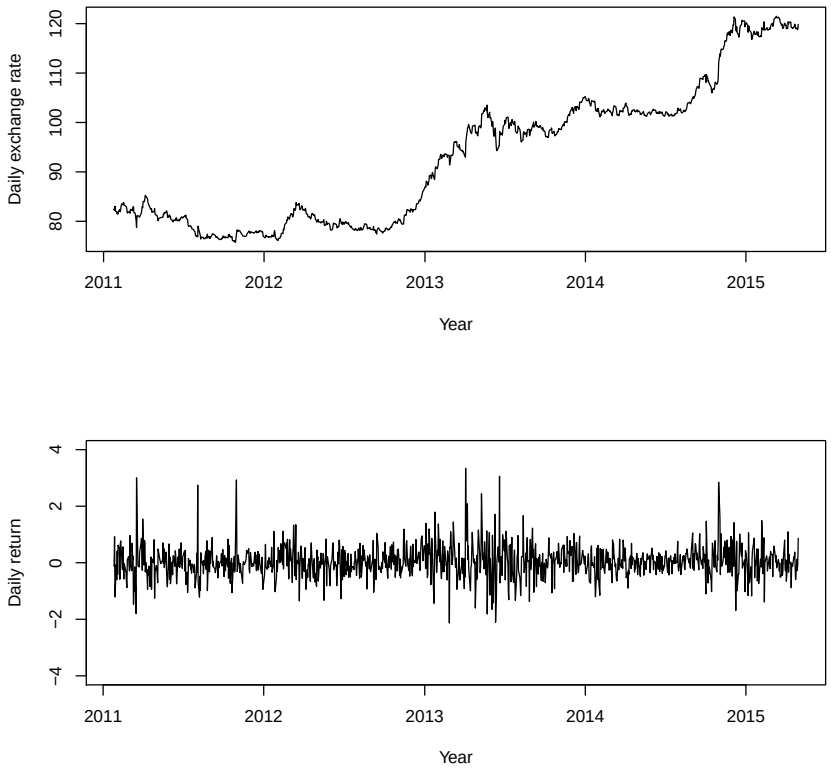

Figure 6: Time series plots of JPY/USD daily exchange rates and returns from 3rd January, 2011 to 30th April, 2015.

Table 2: Sample statistics for y_t.

T	Mean	Median	SD	Skewness	Kurtosis	Min.	Max.
1070	0.04	0.01	0.58	0.64	7.21	−2.13	3.34

The auto-correlations of returns, absolute and squared returns are presented in Table 3. All of our preliminary exploratory analyses suggested the presence of conditional heteroscedasticity in the series. To find the optimal lag for the GARCH model to model the return series we defined many possible subsets of the GARCH(p, q) models with different p and q values. To choose the best model we used Akaike information (AIC) criterion (since it is proposed to determine the best model for forecasting) and the results show that GARCH$(1, 1)$ model is optimal according to AIC.

Table 3: Autocorrelations of y_t at lag k, $k = 1, 2, 5, 10, 16, ..., 20$.

Autocorrelations	r(1)	r(2)	r(5)	r(10)	r(16)	r(17)	r(18)	r(19)	r(20)
y_t	0.008	−0.006	−0.009	−0.027	−0.085	0.015	−0.016	0.011	0.097
$\lvert y_t \rvert$	0.117	0.107	0.111	0.125	0.119	0.070	0.084	0.132	0.096
y_t^2	0.094	0.091	0.066	0.085	0.070	0.027	0.021	0.097	0.083

To obtain out-of sample prediction intervals for the real data, we divide the full data into the following two parts: The model is constructed based on the observations from 3rd January, 2011 to 19th March, 2015 (1041 observations in total) to calculate 30 steps ahead predictions from 20th March to 30th April, 2015 and compare with the actual values. The fitted models for the PRR and proposed block bootstrap methods are obtained as in equations (4.1) and (4.2), respectively:

$$y_t^2 = 0.0054 + 0.0569y_{t-1}^2 + 0.9283\hat{\sigma}_{t-1}^2, \tag{4.1}$$

$$y_t^2 = 0.0150 + 0.9556y_{t-1}^2 + \nu_t - 0.8805\nu_{t-1}, \tag{4.2}$$

where $\widehat{\omega} = 0.0150$, $\widehat{\alpha}_1 = 0.0750$ and $\widehat{\beta}_1 = 0.8805$ for the model estimated by (4.2). The 30 steps ahead prediction intervals for returns y_{T+h} based on the models given in equations (4.1) and (4.2), together with the true returns are presented in Figure 7. The intervals obtained using all the methods are similar and they include all of the true values of returns (only PRR fails to cover the 13th point).

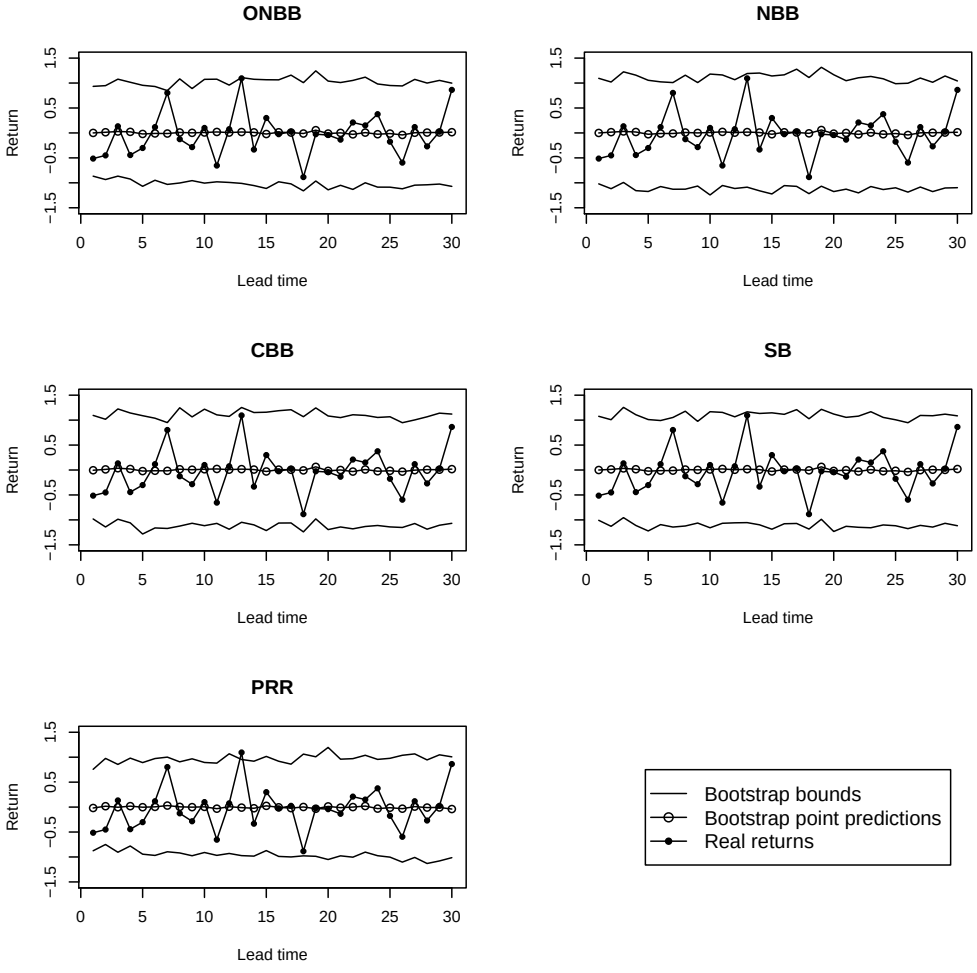

Figure 7: 95% prediction intervals of returns from 20th March, 2015 to 30th April, 2015.

Figure 8 shows the predicted intervals for 30 steps ahead volatilities σ^2_{T+h}. The true values of the volatilities can not be observed directly. We calculate the realized volatility by summing squared returns at day t, $\sigma^2_t = y^2_{t,1} + ... + y^2_{t,n}$, where n is the number of observations recorded during day t as proposed by [1]. Since our data is from 24 hour open trading market, the realized volatilities are computed by using one-minute returns based on tick-by-tick prices such that $n = 1440$ approximately. Figure 8 indicates that the PRR and ONBB methods produce narrower prediction intervals than the one obtained by other block bootstrap methods.

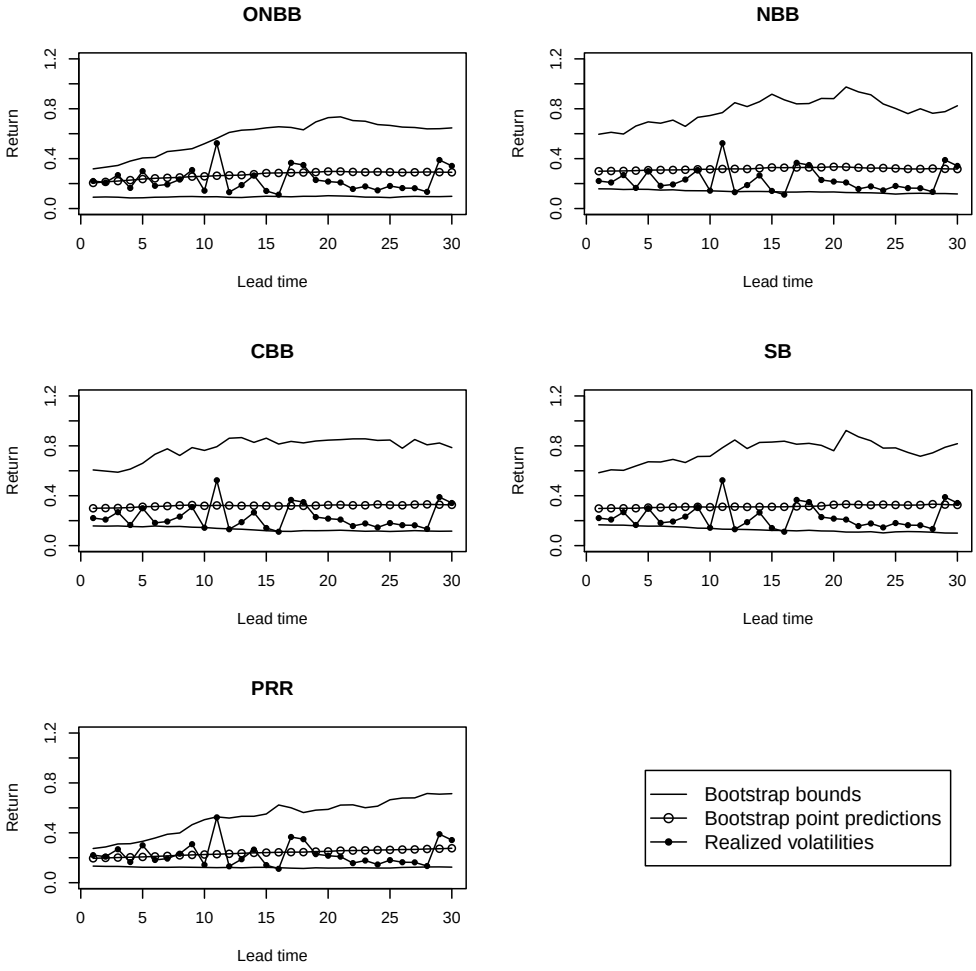

Figure 8: 95% prediction intervals of volatilities from 20th March, 2015 to 30th April, 2015.

5. CONCLUSION

In this paper, we propose a novel resampling algorithm to obtain prediction intervals for returns and volatilities under GARCH models, and we compare the performances of the methods by both simulations and a case study. Our idea is based on using the ARMA representation of the GARCH models. Under ARMA representation, estimation of parameters becomes linear, which allows us to have a valid prediction intervals for the block bootstrapping procedure.

Our findings show that our proposed ONBB method:

(**i**) is a good competitor or even better,

(**ii**) is computationally more efficient than traditional method(s).

Also, the proposed algorithm improves the performances of the non-ordered block bootstrap methods significantly compared to their conventional counterparts.

As a future research, the performances of the proposed methods can also be studied for forecasting time series with BOOT.EXPOS procedure as studied by [12] or they can also be used in other statistical inference problems for dependent data.

ACKNOWLEDGMENTS

We thank the anonymous referees for his/her careful reading of our manuscript and valuable suggestions and comments, which have helped us produce a significantly better paper. We are also grateful to the Editor and Associate Editor for offering the opportunity to publish our work.

REFERENCES

[1] Andersen, T.G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review*, **39**(4), 885–905.

[2] Andersen, T.G.; Bollerslev, T.; Diebold, F.X. and Labys, P. (2001). The distribution of realized exchange rate volatility, *Journal of the American Statistical Association*, **96**(453), 42–55.

[3] Baillie, R.T. and Bollerslev, T. (1992). Prediction in dynamic models with time-dependent conditional variances, *Journal of Econometrics*, **52**(1), 91–113.

[4] Bentes, S.R. (2015). A comparative analysis of the predictive power of implied volatility indices and GARCH forecasted volatility, *Physica A: Statistical Mechanics and its Applications*, **424**(1), 105–112.

[5] Beyaztas, B.H.; Beyaztas, U.; Bandyopadhyay, S. and Huang, W.M. (2017). New and fast block bootstrap-based prediction intervals for GARCH(1, 1) process with application to exchange rates, *Sankhya A*, **Doi: 10.1007/s13171-017-0098-2**(0), 1–27.

[6] Beyaztas, B.H.; Firuzan, E. and Beyaztas, U. (2017). New block bootstrap methods: Sufficient and/or ordered, *Communications in Statistics – Simulation and Computation*, **46**(5), 3942–3951.

[7] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**(3), 307–327.

[8] Bougerol, P. and Picard, N. (1992a). Strict stationarity of generalized autoregressive processes, *The Annals of Probability*, **20**(4), 1714–1730.

[9] Bougerol, P. and Picard, N. (1992b). Stationarity of GARCH processes and of some nonnegative time series, *Journal of Econometrics*, **52**(1), 115–127.

[10] Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence, *Journal of Econometrics*, **14**(3), 1171–1179.

[11] Chen, B.; Gel, Y.R.; Balakrishna, N. and Abraham, B. (2011). Computationally efficient bootstrap prediction intervals for returns and volatilities in ARCH and GARCH processes, *Journal of Forecasting*, **30**(1), 51–71.

[12] Cordeiro, C. and Neves, M.M. (2009). Forecasting time series with boot.expos procedure, *REVSTAT*, **7**(2), 135–149.

[13] Dyhrberg, A.H. (2016). Bitcoin, gold and the dollar? A GARCH volatility analysis, *Finance Research Letters*, **16**(1), 85–92.

[14] Efron, B. (1979). Bootstrap methods: another look at the jackknife, *The Annals of Statistics*, **7**(1), 1–26.

[15] Engle, R.F. and Patton, A.J. (2001). What good is a volatility model, *Quantitative Finance*, **1**(2), 237–245.

[16] Hall, P. (1985). Resampling a coverage pattern, *Stochastic Processes and their Applications*, **20**(2), 231–246.

[17] Hall, P.; Horowitz, J.L. and Jing, B. (1985). On blocking rules for the bootstrap with dependent data, *Biometrika*, **82**(3), 561–574.

[18] Hannan, E.J. and Rissanen, J. (1982). Recursive estimation of mixed autoregressive-moving average order, *Biometrika*, **69**(1), 81–94.

[19] Hwang, E. and Shin, D.W. (2013). Stationary bootstrap prediction intervals for GARCH(p, q), *Communications for Statistical Applications and Methods*, **20**(1), 41–52.

[20] Kunsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations, *The Annals of Statistics*, **17**(3), 1217–1241.

[21] Lahiri, S.N. (2003). *Resampling Methods for Dependent Data*, Springer, Verlag.

[22] Lama, A.; Jha, G.K.; Paul, R.K. and Gurung, B. (2015). Modelling and forecasting of price volatility: An application of GARCH and EGARCH models, *Agricultural Economics Research Review*, **28**(1), 365–382.

[23] Miguel, J.A. and Olave, P. (1999). Bootstrapping forecast intervals in ARCH models, *Test*, **8**(2), 345–364.

[24] Pascual, L.; Romo, J. and Ruiz, E. (2006). Bootstrap prediction for returns and volatilities in garch models, *Computational Statistics & Data Analysis*, **50**(9), 2293–2312.

[25] Politis, D.N. and Romano, J.P. (1994). The stationary bootstrap, *Journal of the American Statistical Association*, **89**(489), 1303–1313.

[26] Politis, D.N. and Romano, J.P. (2000). *A circular block-resampling procedure for stationary data*. In: "Exploring the Limits of Bootstrap" (R. LePage and L. Billard, Eds.), Wiley, New York, 263–270.

[27] Reeves, J.J. (2005). Bootstrap prediction intervals for ARCH models, *International Journal of Forecasting*, **21**(2), 237–248.

[28] Sotiriadis, M.S.; Tsotsos, R. and Kosmidou, K. (2016). Price and volatility interrelationships in the wholesale spot electricity markets of the Central-Western European and Nordic region: a multivariate GARCH approach, *Energy Systems*, **7**(1), 5–32.

REVSTAT – Statistical Journal
Volume 18, Number 4, October 2020, 415–436

A QUANTILE REGRESSION MODEL FOR BOUNDED RESPONSES BASED ON THE EXPONENTIAL-GEOMETRIC DISTRIBUTION

Authors: Pedro Jodrá
– Departamento de Métodos Estadísticos, Universidad de Zaragoza, Zaragoza, Spain
pjodra@unizar.es

María Dolores Jiménez-Gamero
– Departamento de Estadística e Investigación Operativa, Universidad de Sevilla, Sevilla, Spain
dolores@us.es

Received: March 2017 Revised: January 2018 Accepted: April 2018

Abstract:

- The paper first introduces a new two-parameter continuous probability distribution with bounded support from the extended exponential-geometric distribution. Closed-form expressions are given for the moments, moments of the order statistics and quantile function of the new law; it is also shown that the members of this family of distributions can be ordered in terms of the likelihood ratio order. The parameter estimation is carried out by the method of maximum likelihood and a closed-form expression is given for the Fisher information matrix, which is helpful for asymptotic inferences. Then, a new regression model is introduced by considering the proposed distribution, which is adequate for situations where the response variable is restricted to a bounded interval, as an alternative to the well-known beta regression model, among others. It relates the median response to a linear predictor through a link function. Extensions for other quantiles can be similarly performed. The suitability of this regression model is exemplified by means of a real data application.

Keywords:

- *exponential-geometric distribution; bounded support; regression model.*

AMS Subject Classification:

- 60E05, 62J02.

1. INTRODUCTION

The development of new parametric probability distributions attracts a great deal of attention with the aim of providing useful models in many different areas. Some recent contributions can be found in Bakoban and Abu-Zinadah [7], Gómez-Déniz et al. [18] and Jodrá et al. [24], among others. With respect to models with bounded support, considerable effort has been focussed on providing alternatives to the beta distribution. A prominent alternative is the two-parameter Kumaraswamy distribution introduced by Kumaraswamy [28] and thoroughly studied by Jones [25]. Other less known two-parameter models are the transformed Leipnik distribution (see Jorgensen [26, pp. 196–197]) and the recently introduced Log–Lindley law (see Gómez-Déniz et al. [17] and Jodrá and Jiménez-Gamero [23]). There are more proposals such as the four-parameter Kumaraswamy Weibull distribution (Cordeiro et al. [10]) and the five-parameter Kumaraswamy generalized gamma distribution (Pascoa et al. [35]), that present the drawback of having a high number of parameters and in these cases the parameter estimation often presents some difficulties.

This paper introduces a new two-parameter probability distribution with bounded support derived from the extended exponential-geometric (EEG) distribution. The EEG law is a continuous probability distribution studied by Adamidis et al. [2] to model lifetime data. More precisely, a random variable Y is said to have an EEG distribution if the probability density function (pdf) is given by

$$f_Y(y;\alpha,\beta) = \frac{\alpha(1+\beta)e^{-\alpha y}}{(1+\beta\, e^{-\alpha y})^2}, \qquad y>0, \quad \alpha>0, \ \beta>-1,$$

where α and β are the model parameters. In particular, the case $\alpha>0$ and $\beta\in(-1,0)$ corresponds to the exponential-geometric distribution proposed by Adamidis and Loukas [3]. A generalization of the EEG law is the three-parameter Weibull-geometric distribution introduced by Barreto-Souza et al. [8].

From the EEG distribution, we define a new random variable X with support in the standard unit interval $(0,1)$ by means of the transformation $X=\exp(-Y)$. It is easy to check that X has the following pdf and cumulative distribution function (cdf),

$$f(x;\alpha,\beta) = \frac{\alpha(1+\beta)x^{\alpha-1}}{(1+\beta\, x^{\alpha})^2}, \qquad 0<x<1, \quad \alpha>0, \ \beta>-1, \tag{1.1}$$

and

$$F(x;\alpha,\beta) = \frac{(1+\beta)x^{\alpha}}{1+\beta\, x^{\alpha}}, \qquad 0<x<1, \quad \alpha>0, \ \beta>-1,$$

respectively. In the sequel, the random variable defined by (1.1) will be referred to as the Log-extended exponential-geometric (LEEG) distribution. The LEEG distribution presents an advantage with respect to the beta distribution since it does not include special functions in its formulation. Figure 1 represents the pdf of X for several values of the parameters. It is interesting to note that the special case $\beta=0$ corresponds to the power function distribution, which includes the uniform distribution for $\alpha=1$.

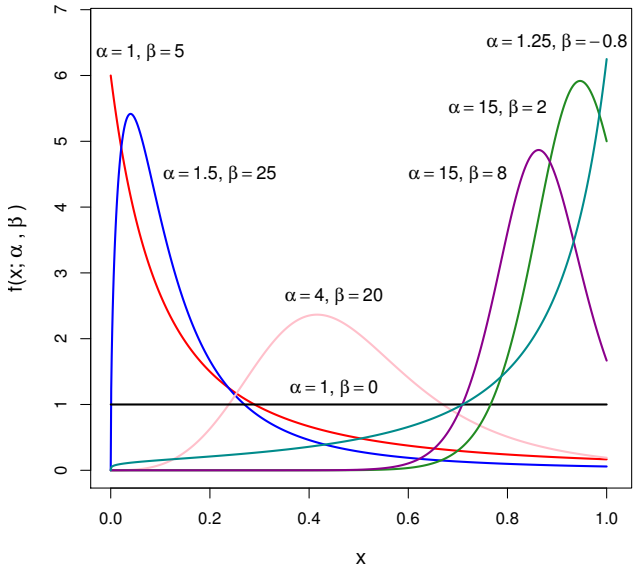

Figure 1: $f(x;\alpha,\beta)$ for different values of α and β.

Clearly, the LEEG distribution can be used to model real data taking values in the unit interval. Furthermore, as a linear transformation $(b-a)X+a$ moves a random variable X defined on $(0,1)$ to any other bounded support (a,b), with $a<b$, the LEEG law can be extended to any bounded domain in a straightforward manner, so there is no need to explain such an extension.

On the basis of the proposed distribution, we introduce a new regression model which assumes that the response variable takes values in the standard unit interval, as an alternative to the well-known beta regression model (see Ferrari and Cribari-Neto [15]). Other regression models for bounded responses can be found in [33, 34, 36]. Regression models usually express a location measure of a distribution as a function of covariates. The location measure is commonly taken the mean (which is the case of classical regression models) or some quantile (which is the case of quantile regression, see, for example, the book by Koenker [27]). With this aim, it is noted that the LEEG distribution can be easily reparametrized in terms of any of its quantiles. As the median is a robust central tendency measure, we choose to reparametrize the LEEG law with its median and construct the associated regression model, which relates the median response to a linear predictor through a link function. Nevertheless, it will become evident that any other quantile could be used.

The literature on parametric quantile regression is rather scarce. An example is the parametric regression quantile in Noufaily and Jones [32], designed for a positive response, while our proposal is for a bounded response. In addition to this evident distinctive feature, the main difference between our approach and that in [32] lies in the following: Noufaily and Jones [32] assume a distribution for the response (specifically, the generalized gamma with three parameters) and consider parametric forms for the dependence of the parameters (or some subset of them) on the covariate (they only assume a unique covariate, although their proposal can be extended to more covariates); then they replace the parameters in the expression of the quantile function of the assumed model by the fitted regression equations for the parameters. By contrast, we reparametrize the distribution in terms of the median

(although we could consider any other quantile) and assume a parametric form for the dependence of the median on the covariates (we do not limit the number of covariates). In our proposal, only one of the parameters is allowed to depend on the covariates, but it would be an obvious extension to express both of them as functions of the covariates. Note that our strategy is closer, in spirit, to Koenker [27], which assumes a regression model for a quantile; if the quantile is changed then the regression model also changes. In our scheme, if the distribution is parametrized in terms of another quantile (different from the median), the model parameters will change. On the contrary, in Noufaily and Jones [32] the model parameters are the same for each quantile since they do not fit a genuine quantile regression model, they just allow the distribution parameters to vary with the covariates and then replace them in the expression of the quantile function.

The remainder of this paper is organized as follows. In Section 2, some statistical properties of the LEEG distribution are studied. Precisely, it is shown that the LEEG law can be derived as the distribution of the minimum or maximum of a geometric random number of independent random variables with power function distribution, the moments, as well as the moments of the order statistics, can be expressed analytically in terms of the Lerch transcendent function, the quantile function can be given in closed form and the members of the new family of distributions can be ordered in terms of the likelihood ratio order. For the sake of clarity, the proofs of this section are deferred to Appendix B. Section 3 deals with the parameter estimation problem. Specifically, the method of maximum likelihood is theoretically and numerically studied. In addition, an explicit expression for the Fisher information matrix is obtained, which is useful for asymptotic inferences on the parameters. The proof of these results is deferred to Appendix C. Some numerical results studying the finite sample performance of the maximum likelihood estimators as well a real data set application are also displayed in this section. Section 4 shows how to construct a regression model for bounded responses on the basis of the LEEG distribution. A real data application demonstrates that such model may be more appropriate than others previously proposed. For the sake of completeness, Appendix A presents a known result concerning a logarithmic integral, which is used to provide unified proofs in Appendices B and C.

2. STATISTICAL PROPERTIES

This section studies some statistical properties of the LEEG distribution. Specifically, an stochastic representation is provided together with the shape of the pdf, the computation of moments, the computer-generation of pseudo-random data and the computation of moments of the order statistics. In all cases, closed-form expressions are given. Additionally, it is shown that the new family of distributions can be ordered in terms of the likelihood ratio order.

2.1. Stochastic representation

The LEEG distribution has been defined in (1.1) via an exponential transformation of the EEG distribution. It should be noted that the LEEG law can also be derived as follows.

Let N be a random variable having a geometric distribution with probability mass function (pmf) given by

$$P(N=n)=\left(1-\frac{1}{1+\beta}\right)^{n-1}\frac{1}{1+\beta}, \qquad n=1,2,...,$$

with $\beta>0$. Let M be a random variable having a geometric distribution with pmf given by

$$P(M=m)=(-\beta)^{m-1}(1+\beta), \qquad m=1,2,...,$$

with $\beta\in(-1,0)$. Let $T_1,T_2,...$ be independent identically distributed random variables having T_i a power function distribution with parameter $\alpha>0$, that is, its cdf is given by $F_{T_i}(t;\alpha)=t^\alpha$, $0<t<1$. Assume that N and M are independent of T_i, $i=1,2,....$

Proposition 2.1.

(**i**) *The random variable* $V=\min\{T_1,T_2,...,T_N\}$ *has a LEEG distribution with parameters* $\alpha>0$ *and* $\beta>0$.

(**ii**) *The random variable* $W=\max\{T_1,T_2,...,T_M\}$ *has a LEEG distribution with parameters* $\alpha>0$ *and* $\beta\in(-1,0)$.

2.2. Shape and mode

As it can be seen from Figure 1, the pdf of the LEEG distribution has a wide variety of shapes. The next result characterizes the shape of the pdf in terms of the parameter values.

Proposition 2.2. *Let* X *be a LEEG distribution with parameters* $\alpha>0$ *and* $\beta>-1$.

(**i**) *For any* $\alpha>1$, *if* $\beta>(\alpha-1)/(1+\alpha)$ *then* X *has a mode at* $x=\left(\frac{\alpha-1}{(1+\alpha)\beta}\right)^{1/\alpha}$ *and if* $\beta\in(-1,(\alpha-1)/(1+\alpha)]$ *then* (1.1) *is an increasing function.*

(**ii**) *For any* $0<\alpha<1$, *if* $\beta\in(-1,(\alpha-1)/(1+\alpha))$ *then* (1.1) *has a minimum at* $x=\left(\frac{\alpha-1}{(1+\alpha)\beta}\right)^{1/\alpha}$ *and if* $\beta\geq(\alpha-1)/(1+\alpha)$ *then* (1.1) *is a decreasing function.*

(**iii**) *If* $\alpha=1$ *and* $\beta=0$, *then* (1.1) *is the pdf of the uniform distribution on* $(0,1)$.

2.3. Moments

The moments of X can be expressed in closed form in terms of the Lerch transcendent function, Φ. Remind that Φ is defined as the analytic continuation of the series

$$\Phi(z,\lambda,v)=\sum_{i=0}^{\infty}\frac{z^i}{(i+v)^\lambda},$$

which converges for any real number $v>0$ if z and λ are any complex numbers with either $|z|<1$ or $|z|=1$ and $\mathrm{Re}(\lambda)>1$ (see Apostol [5] for further details).

Proposition 2.3. *Let X be a LEEG distribution with parameters $\alpha > 0$ and $\beta > -1$. The moments of X are given by*

$$E[X^k] = 1 - \frac{(1+\beta)k}{\alpha}\,\Phi\left(-\beta, 1, 1+\frac{k}{\alpha}\right), \qquad k = 1, 2, ... \tag{2.1}$$

It is interesting to note that the Lerch transcendent function is available in computer algebra systems such as Maple (function `LerchPhi(z,λ,v)`) and Mathematica (function `LerchPhi[z,λ,v]`). Accordingly, usual statistical measures involving $E[X^k]$ can be efficiently computed from equation (2.1).

2.4. Quantile function

An interesting advantage of the LEEG distribution with respect to the beta distribution is that the cdf of X is readily invertible.

Proposition 2.4. *The quantile function of the LEEG distribution with parameters $\alpha > 0$ and $\beta > -1$ is given by*

$$F^{-1}(u;\alpha,\beta) = \left(\frac{u}{1+\beta-\beta\,u}\right)^{1/\alpha}, \qquad 0 < u < 1.$$

From Proposition 2.4, the quartiles of the LEEG law are given by

$$Q_1 = \left(\frac{1}{4+3\beta}\right)^{1/\alpha}, \quad Q_2 = \left(\frac{1}{2+\beta}\right)^{1/\alpha}, \quad Q_3 = \left(\frac{3}{4+\beta}\right)^{1/\alpha}.$$

The explicit expression in Proposition 2.4 is helpful in simulation studies because pseudo-random data from the LEEG distribution can be generated by computer using the inverse transform method.

2.5. Order statistics

Next, analytical expressions to compute the moments of the order statistics are provided. To this end, it is shown that the moments of the largest order statistic of the LEEG law can be given in terms of a finite sum involving the Lerch transcendent function Φ and the generalized Stirling numbers of the first kind R_n^j (see Appendix A for the definition and calculation of these numbers).

Let $X_1, ..., X_n$ be a random sample of size n from the LEEG distribution with parameters $\alpha > 0$ and $\beta > -1$. Let $X_{1:n} \le X_{2:n} \le ... \le X_{n:n}$ be the order statistics obtained by arranging X_i, $i = 1, ..., n$, in non-decreasing order of magnitude. For any $n = 1, 2, ...$ and $k = 1, 2, ...$, denote by $E[X_{r:n}^k]$ the kth moment of $X_{r:n}$, $r = 1, ..., n$.

Proposition 2.5. *Let $X_1, ..., X_n$ be a random sample of size n from a LEEG distribution with parameters $\alpha > 0$ and $\beta > -1$. Let $X_{n:n}$ be the largest order statistic. Then*

$$E[X_{n:n}^k] = \frac{(1+\beta)^n}{\Gamma(n)} \sum_{j=0}^{n} R_n^j(k/\alpha, 1)\Phi\left(-\beta, 1-j, n+\frac{k}{\alpha}\right), \qquad k = 1, 2, ...$$

The result in Proposition 2.5 is useful to evaluate the moments of $X_{r:n}$, for $r = 1, ..., n-1$, thanks to the following well-known formula (see, for example, David and Nagaraja [13, p. 45])

$$E[X_{r:n}^k] = \sum_{j=r}^{n} (-1)^{(j-r)} \binom{j-1}{r-1} \binom{n}{j} E[X_{j:j}^k], \qquad r = 1, ..., n-1.$$

2.6. Stochastic orderings

To conclude Section 2, it is shown that the members of the new distribution can be ordered in terms of the likelihood ratio order, which is defined as follows (see, for example, Shaked and Shanthikumar [40, Chapter 1]).

Definition 2.1. Let X_1 and X_2 be two continuous random variables with pdfs f_1 and f_2, respectively, such that $f_2(x)/f_1(x)$ is non-decreasing over the union of the supports of X_1 and X_2. Then X_1 is said to be smaller than X_2 in the likelihood ratio order, denoted by $X_1 \leq_{LR} X_2$.

The likelihood ratio order is stronger than the hazard rate order and the usual stochastic order, which are defined as follows.

Definition 2.2. Let X_1 and X_2 be two random variables with cdfs F_1 and F_2 and hazard rates h_1 and h_2, respectively. Then

- **(i)** X_1 is said to be stochastically smaller than X_2, denoted by $X_1 \leq_{ST} X_2$, if $F_1(x) \geq F_2(x)$ for all x.
- **(ii)** X_1 is said to be smaller than X_2 in the hazard rate, denoted by $X_1 \leq_{HR} X_2$, if $h_1(x) \leq h_2(x)$ for all x.

The LEEG family can be ordered in the following way.

Proposition 2.6. *Let X_1 and X_2 be two random variables having a LEEG distribution with parameters (α, β_1) and (α, β_2), respectively, for some $\alpha > 0$, $\beta_1, \beta_2 > -1$. If $\beta_1 \geq \beta_2$ then $X_1 \leq_{LR} X_2$.*

As an immediate consequence of Proposition 2.6 and the well-known fact that

$$X_1 \leq_{LR} X_2 \Rightarrow X_1 \leq_{HR} X_2 \Rightarrow X_1 \leq_{ST} X_2,$$

the following corollary is stated.

Corollary 2.1. *Let X_1 and X_2 be two random variables having a LEEG distribution with parameters (α, β_1) and (α, β_2), respectively, for some $\alpha > 0$, $\beta_1, \beta_2 > -1$. If $\beta_1 \geq \beta_2$ then*

(**i**) $E(X_1^k) \leq E(X_2^k)$, $\forall k > 0$.

(**ii**) $h_1(x) \leq h_2(x)$, $\forall x \in (0,1)$.

As a special case of Corollary 2.1 (i) it follows that, for fixed $\alpha > 0$, the mean of the LEEG distribution decreases as β increases.

3. PARAMETER ESTIMATION

This section considers the estimation of the parameters of the LEEG distribution. Specifically, Subsection 3.1 describes the maximum likelihood (ML) method. A closed-form expression for the Fisher information matrix is provided in Subsection 3.2. The performance of the ML method is evaluated via a Monte Carlo simulation study in Subsection 3.3. Finally, a real data application is presented in Subsection 3.4.

3.1. Maximum likelihood method

Let $X_1, ..., X_n$ be a random sample of size n from a LEEG distribution with unknown parameters $\alpha > 0$ and $\beta > -1$ and denote by $x_1, ..., x_n$ the observed values. From the likelihood function, $L(\alpha, \beta) = \prod_{i=1}^n f(x_i; \alpha, \beta)$, the log-likelihood function is given by

$$\log L(\alpha, \beta) = n \log \alpha + n \log (1 + \beta) + (\alpha - 1) \sum_{i=1}^{n} \log x_i - 2 \sum_{i=1}^{n} \log (1 + \beta\, x_i^{\alpha}). \tag{3.1}$$

The ML estimates of α and β are the values $\hat{\alpha}$ and $\hat{\beta}$ that maximize $\log L(\alpha, \beta)$. The partial derivatives of $\log L(\alpha, \beta)$ with respect to each parameter are the following:

$$\frac{\partial}{\partial \alpha} \log L(\alpha, \beta) = \frac{n}{\alpha} + \sum_{i=1}^{n} \log x_i - 2\beta \sum_{i=1}^{n} \frac{x_i^{\alpha} \log x_i}{1 + \beta\, x_i^{\alpha}}, \tag{3.2}$$

$$\frac{\partial}{\partial \beta} \log L(\alpha, \beta) = \frac{n}{1 + \beta} - 2 \sum_{i=1}^{n} \frac{x_i^{\alpha}}{1 + \beta\, x_i^{\alpha}}. \tag{3.3}$$

The ML estimates of the parameters satisfy the system that results from equating to 0 the equations (3.2) and (3.3). Nevertheless, since such system does not have an explicit solution, in order to obtain the ML estimates it is preferable to maximize the function (3.1). Subsection 3.3 will deal with this practical issue.

Another practical point is the possible presence of extreme values in the data. Although we are assuming that the data are continuous, which implies that the probability of observing the values zero and one is null, in applications, due to rounding errors, these extreme cases may appear in the observations. By looking at the expression of the log-likelihood (3.1), the

presence of ones involves no problem; on the other hand, the presence of zeroes implies that the log-likelihood cannot be calculated. In such a case, we recommend replacing all zeroes by a positive small quantity.

3.2. Fisher information matrix

Below, an analytical expression for the Fisher information matrix is given, which let us explicitly calculate the asymptotic covariance matrix of the ML estimators. To this end, the polylogarithm function, which is a particular case of the Lerch transcendent function (see Appendix A), plays an important role.

Proposition 3.1. *Let $X_1, ..., X_n$ be a random sample of size n from a LEEG distribution with parameters $\alpha > 0$ and $\beta > -1$. For $\alpha > 0$ and $\beta \in (-1, 0) \cup (0, \infty)$ the Fisher information matrix is given by*

$$I(\alpha, \beta) = \begin{bmatrix} \dfrac{n}{\alpha^2} - \dfrac{2n}{3\alpha^2\beta}\left\{(1+\beta)\mathrm{Li}_2(-\beta) + \beta\right\} & \dfrac{n(1+\beta)}{3\alpha\beta}\left(\dfrac{1}{(1+\beta)^2} - \dfrac{\log(1+\beta)}{\beta}\right) \\ \dfrac{n(1+\beta)}{3\alpha\beta}\left(\dfrac{1}{(1+\beta)^2} - \dfrac{\log(1+\beta)}{\beta}\right) & \dfrac{n}{3(1+\beta)^2} \end{bmatrix},$$

where Li_2 *denotes the polylogarithm function of order two. For $\alpha > 0$ and $\beta = 0$,*

$$I(\alpha, 0) = \begin{bmatrix} \dfrac{n}{\alpha^2} & -\dfrac{n}{2\alpha} \\ -\dfrac{n}{2\alpha} & \dfrac{n}{3} \end{bmatrix}.$$

As it is well-known, it is useful to have an explicit expression for $I(\alpha, \beta)$ since by inverting this matrix we get the asymptotic covariance matrix of the ML estimators and it can be used to approximate their standard errors. Denote by N_2 a bivariate normal distribution and by $\xrightarrow{d}$ the convergence in distribution.

Proposition 3.2. *Let $X_1, ..., X_n$ be a random sample of size n from a LEEG distribution with parameters $\alpha > 0$ and $\beta > -1$. Let $\widehat{\boldsymbol{\theta}}$ denote the ML estimator of $\boldsymbol{\theta} = (\alpha, \beta)$. Then,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathrm{N}_2(\mathbf{0}, \Sigma),$$

where $\Sigma = \Sigma(\alpha, \beta)$ is such that for $\beta \neq 0$

$$\Sigma(\alpha, \beta) = \begin{bmatrix} -\dfrac{3\alpha^2\beta^4}{(1+\beta)c(\beta)} & -\dfrac{3\alpha\beta^2[(1+\beta)^2\log(1+\beta) - \beta]}{c(\beta)} \\ -\dfrac{3\alpha\beta^2[(1+\beta)^2\log(1+\beta) - \beta]}{c(\beta)} & \dfrac{3\beta^3(1+\beta)[2(1+\beta)\mathrm{Li}_2(-\beta) - \beta]}{c(\beta)} \end{bmatrix},$$

with

$$c(\beta) = (1+\beta)^3\log^2(1+\beta) - 2\beta(1+\beta)\log(1+\beta) + \beta^3[2\mathrm{Li}_2(-\beta) - 1] + \beta^2$$

and Li_2 *stands for the polylogarithm function of order two, and for $\beta = 0$*

$$\Sigma(\alpha, 0) = \begin{bmatrix} 4\alpha^2 & 6\alpha \\ 6\alpha & 12 \end{bmatrix}.$$

3.3. Simulation study

As discussed in Subsection 3.1, in order to obtain the ML estimates of the parameters the following optimization problem is solved

$$\begin{aligned} &\max \ \log L(\alpha, \beta) \\ &\text{s.t.} \quad \alpha > 0 \\ &\qquad \beta > -1, \end{aligned} \tag{3.4}$$

where $\log L(\alpha, \beta)$ is given in equation (3.1). In our simulations, problem (3.4) was solved by using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, available in the function `constrOptim` of the R programming language [37]. We chose the BFGS algorithm because (3.4) is an optimization problem with linear inequality constraints. The BFGS algorithm requires a starting point, which must be in the interior of the feasible region, together with the gradient function of $\log L(\alpha, \beta)$. As starting point we tried several options with little or no effect on the final solution. All numerical results in this paper were obtained by using as starting point the pair $(1, 1)$.

The performance of the ML estimators was assessed via a Monte Carlo simulation study. The following notation was used. The number of random samples generated is denoted by N and the size of each random sample is denoted by n. The following quantities were computed for the simulated estimates $\hat{\alpha}_j$, $j = 1, ..., N$:

- **(i)** The mean: $\bar{\alpha} = (1/N) \sum_{j=1}^{N} \hat{\alpha}_j$.
- **(ii)** The bias: $\text{Bias}(\hat{\alpha}) = \bar{\alpha} - \alpha$.
- **(iii)** The mean-square error: $\text{MSE}(\hat{\alpha}) = (1/N) \sum_{j=1}^{N} (\hat{\alpha}_j - \alpha)^2$.

The quantities $\bar{\beta}$, $\text{Bias}(\hat{\beta})$ and $\text{MSE}(\hat{\beta})$ are analogously defined and were also computed. In particular, we generated $N = 10,000$ random samples of different sizes n for several values of α and β. Some simulation results are shown in Table 1, where it is included the mean, bias and MSE of the simulated estimates together with the asymptotic variance of the estimators calculated directly from the diagonal elements of $(1/n)\Sigma(\alpha, \beta)$, with $\Sigma(\alpha, \beta)$ given by Proposition 3.2, and denoted by $\text{Var}[\hat{\alpha}]$ and $\text{Var}[\hat{\beta}]$. From the obtained results, it can be concluded that the ML method provides acceptable estimates of the parameters, although it should be noted that the ML method tended to slightly overestimate the value of both parameters in the cases considered in the present study.

Table 1: ML estimates of α and β.

	$\alpha = 0.25$				$\beta = -0.25$				$\alpha = 1.25$				$\beta = -0.80$			
	$\bar{\alpha}$	Bias($\hat{\alpha}$)	MSE($\hat{\alpha}$)	Var[$\hat{\alpha}$]	$\bar{\beta}$	Bias($\hat{\beta}$)	MSE($\hat{\beta}$)	Var[$\hat{\beta}$]	$\bar{\alpha}$	Bias($\hat{\alpha}$)	MSE($\hat{\alpha}$)	Var[$\hat{\alpha}$]	$\bar{\beta}$	Bias($\hat{\beta}$)	MSE($\hat{\beta}$)	Var[$\hat{\beta}$]
$n = 50$	0.2759	0.0259	0.0078	0.0060	−0.0663	0.1836	0.2755	0.1429	1.5817	0.3317	0.7677	0.4335	−0.7176	0.0823	0.0436	0.0170
$n = 75$	0.2664	0.0164	0.0047	0.0040	−0.1350	0.1150	0.1507	0.0952	1.4641	0.2141	0.4293	0.2890	−0.7479	0.0520	0.0214	0.0113
$n = 100$	0.2614	0.0114	0.0034	0.0030	−0.1702	0.0798	0.0994	0.0714	1.4136	0.1636	0.2892	0.2167	−0.7605	0.0394	0.0138	0.0085
$n = 200$	0.2562	0.0062	0.0016	0.0015	−0.2087	0.0411	0.0429	0.0357	1.3281	0.0781	0.1271	0.1083	−0.7812	0.0187	0.0055	0.0042
$n = 500$	0.2525	0.0025	0.0006	0.0006	−0.2341	0.0158	0.0153	0.0142	1.2798	0.0298	0.0468	0.0433	−0.7931	0.0068	0.0019	0.0017

	$\alpha = 1.0$				$\beta = 5.0$				$\alpha = 1.5$				$\beta = 10.0$			
	$\bar{\alpha}$	Bias($\hat{\alpha}$)	MSE($\hat{\alpha}$)	Var[$\hat{\alpha}$]	$\bar{\beta}$	Bias($\hat{\beta}$)	MSE($\hat{\beta}$)	Var[$\hat{\beta}$]	$\bar{\alpha}$	Bias($\hat{\alpha}$)	MSE($\hat{\alpha}$)	Var[$\hat{\alpha}$]	$\bar{\beta}$	Bias($\hat{\beta}$)	MSE($\hat{\beta}$)	Var[$\hat{\beta}$]
$n = 50$	1.0333	0.0333	0.0335	0.0310	6.2783	1.2783	18.0034	8.4895	1.5515	0.0515	0.0643	0.0565	12.8102	2.8102	77.0472	31.1192
$n = 75$	1.0251	0.0251	0.0224	0.0206	5.8739	0.8739	9.9312	5.6596	1.5341	0.0341	0.0412	0.0376	11.8078	1.8078	37.2463	20.7461
$n = 100$	1.0173	0.0173	0.0162	0.0155	5.6031	0.6031	6.0647	4.2447	1.5230	0.0230	0.0295	0.0282	11.1966	1.1966	23.6258	15.5596
$n = 200$	1.0081	0.0081	0.0079	0.0077	5.2955	0.2955	2.6045	2.1223	1.5126	0.0126	0.0143	0.0141	10.6480	0.6480	9.7978	7.7798
$n = 500$	1.0044	0.0044	0.0031	0.0031	5.1294	0.1294	0.9255	0.8489	1.5044	0.0044	0.0057	0.0056	10.2248	0.2248	3.4507	3.1119

	$\alpha = 15.0$				$\beta = 2.0$				$\alpha = 15.0$				$\beta = 10.0$			
	$\bar{\alpha}$	Bias($\hat{\alpha}$)	MSE($\hat{\alpha}$)	Var[$\hat{\alpha}$]	$\bar{\beta}$	Bias($\hat{\beta}$)	MSE($\hat{\beta}$)	Var[$\hat{\beta}$]	$\bar{\alpha}$	Bias($\hat{\alpha}$)	MSE($\hat{\alpha}$)	Var[$\hat{\alpha}$]	$\bar{\beta}$	Bias($\hat{\beta}$)	MSE($\hat{\beta}$)	Var[$\hat{\beta}$]
$n = 50$	15.6584	0.6584	10.9093	9.4940	2.6075	0.6075	4.0518	2.0166	15.4493	0.4493	6.2381	5.6530	12.6339	2.6339	71.0531	31.1192
$n = 75$	15.4822	0.4822	7.0518	6.3293	2.4081	0.4081	2.1360	1.3444	15.2776	0.2776	4.0164	3.7687	11.6095	1.6095	35.2727	20.7461
$n = 100$	15.3295	0.3295	5.1931	4.7470	2.2870	0.2870	1.4440	1.0083	15.2294	0.2294	3.0211	2.8265	11.2339	1.2339	24.5033	15.5596
$n = 200$	15.1849	0.1849	2.4365	2.3735	2.1432	0.1432	0.5975	0.5041	15.1187	0.1187	1.4387	1.4132	10.5759	0.5759	9.6097	7.7798
$n = 500$	15.0682	0.0682	0.9674	0.9494	2.0565	0.0565	0.2204	0.2016	15.0389	0.0389	0.5740	0.5653	10.2267	0.2267	3.3672	3.1119

3.4. A real data application

In this subsection, a real data set illustrates the practical usefulness of the LEEG distribution by showing that it may be a more appropriate model than other distributions with support in the standard unit interval.

The data set is available from the personal website of Professor E.W. Frees[1] and consists of 73 observations on 7 variables. The data were collected from a questionnaire carried out with the purpose of relating cost effectiveness to management philosophy of controlling the company's exposure to various property and casualty losses, after adjusting for company effects such as size and industry type. These data have been previously analyzed by Schmit and Roth [38], Frees [16, Chapter 6], Gómez-Déniz et al. [17] and Jodrá and Jiménez-Gamero [23].

In this section, interest is centered on the variable FIRMCOST (divided by 100), which is a measure of the cost effectiveness of the risk management practices of the firm. Based on Subsection 3.1, the LEEG law was fitted to the variable FIRMCOST/100. The ML estimates obtained were $\hat{\alpha} = 1.4322$ and $\hat{\beta} = 52.1069$. It can also be checked that the correlation coefficient between the theoretical and the empirical cumulative probabilities is 0.9956.

Additionally, we applied the following goodness-of-fit tests based on the empirical cdf: the Cramér von Mises statistic W^2, the Watson statistic U^2, the Anderson–Darling statistic A^2 and the Kolmogorov–Smirnov statistic D. A detailed definition together with simple formulae for computing these statistics can be found in D'Agostino and Stephens[12, Chapter 4]. To get the p-values we applied a parametric bootstrap generating 10,000 bootstrap samples (see Stute et al. [41] and Babu and Rao [6] for full details). We also applied two test based on the empirical characteristic function [19, 20] by using the integral transformation, as proposed in Meintanis et al. [30], taking as weight functions: the standard normal law, FC_1, and the pdf $w(t) = \{1 - \cos(t)\}/\pi t^2$, which is the choice recommended in Epps and Pulley [14] (see also Section 4 in [20]), FC_2. The results are shown in Table 2 and suggest that the LEEG law provides a satisfactory fit.

Table 2: Goodness-of-fit tests.

	W^2	U^2	A^2	D	FC_1	FC_2
Statistic value:	0.0571	0.0571	0.5133	0.0626	0.0011	0.1142
p-value:	0.2610	0.2610	0.1363	0.5320	0.1164	0.2663

The LEEG fitting was compared to the ones provided by other two-parameter distributions used to model data in the unit interval. Specifically, we considered the beta, Kumaraswamy, Log–Lindley and transformed Leipnik distributions. In order to compare these models, we calculated the Akaike information criterion AIC (see Akaike [4]), the consistent Akaike information criterion CAIC (see Bozdogan [9]) and the Bayesian informa-

[1] http://instruction.bus.wisc.edu/jfrees/jfreesbooks/Regression%20Modeling/BookWebDec2010/data.html, filename: RiskSurvey.

tion criterion BIC (see Schwarz [39]), which are defined as follows, $\text{AIC} = 2m - 2\log L$, $\text{CAIC} = m(1+\log n) - 2\log L$ and $\text{BIC} = m[\log n - \log(2\pi)] - 2\log L$, respectively, where m is the number of parameters, n is the sample size and L denotes the maximized value of the likelihood function. As it is well-known, the model with lowest values of AIC, CAIC and BIC is preferred. For each fitted distribution, Table 3 shows the ML estimated parameters together with the log-likelihood, AIC, CAIC and BIC values. Looking at Table 3, the LEEG distribution provides the best fit. Moreover, the Vuong test [42] was applied to compare the LEEG model to the beta, Kumaraswamy, Log–Lindley and transformed Leipnik distributions. In the four cases the Vuong statistic was very close to 0, so suggesting that all these distributions can be considered equally close to the data. In this regard, we consider the LEEG distribution an attractive alternative to the aforesaid models.

Table 3: Fitted distribution, ML estimates, log-likelihood, AIC, CAIC and BIC.

Distribution	ML estimates	$\log L$	AIC	CAIC	BIC
LEEG(α,β) $f(x;\alpha,\beta) = \dfrac{\alpha(1+\beta)x^{\alpha-1}}{(1+\beta\, x^{\alpha})^2}$	$\hat{\alpha} = 1.4322$ $\hat{\beta} = 52.1069$	93.63	−183.26	−176.68	−182.35
Beta(a,b) $f(x;a,b) = \dfrac{1}{\mathrm{B}\,(a,b)}\, x^{a-1}(1-x)^{b-1}$	$\hat{a} = 0.6125$ $\hat{b} = 3.7978$	76.11	−148.23	−141.65	−147.32
Kumaraswamy(a,b) $f(x;a,b) = abx^{a-1}(1-x^a)^{b-1}$	$\hat{a} = 0.6648$ $\hat{b} = 3.4407$	78.65	−153.30	−146.72	−152.40
Log–Lindley(a,b) $f(x;a,b) = a[b + a(b-1)\log x]x^{a-1}$	$\hat{a} = 0.6906$ $\hat{b} = 0.0231$	76.60	−149.20	−142.62	−148.30
Transformed Leipnik(μ,λ) $f(x;\mu,\lambda) = \dfrac{[x(1-x)]^{-\frac{1}{2}}}{\mathrm{B}\left(\frac{\lambda+1}{2},\frac{1}{2}\right)}\left(1+\dfrac{(x-\mu)^2}{x(1-x)}\right)^{-\frac{\lambda}{2}}$	$\hat{\mu} = 0.0261$ $\hat{\lambda} = 6.4061$	80.51	−157.02	−150.43	−156.11

4. A REGRESSION MODEL FOR BOUNDED RESPONSES

Regression models are commonly used to model the mean of a response variable as a function of a set of covariates (also called independent variables or regressors). As shown in Proposition 2.3, the moments of the LEEG distribution can be expressed in terms of the Lerch transcendent function, which implies that the mean does not possess a simple expression. This fact makes difficult to build a regression model which relates the mean response with covariates. By contrast, the expression of the quantiles of the LEEG distribution is quite tractable, so our proposal is to use them to construct a regression model. In principle, we could choose any quantile, but since the median is a robust measure of location and, in this regard, it is considered as a competitor of the mean, we will choose the median.

As a first step towards the construction of the regression model, the LEEG distribution is reparametrized in terms of the median Q_2 by equating Q_2 to a new parameter θ and solving

the resultant equation for β. The resulting pdf is

$$f(x;\alpha,\theta)=\frac{\alpha\theta^{\alpha}(1-\theta^{\alpha})x^{\alpha-1}}{[\theta^{\alpha}+(1-2\theta^{\alpha})x^{\alpha}]^2}\,, \qquad 0<x<1,\ \alpha>0,\ 0<\theta<1. \tag{4.1}$$

It should be noted that all properties studied for the parametrization (1.1) carry over for the above one with $\beta=(1-2\theta^{\alpha})/\theta^{\alpha}$.

Let $X_1,...,X_n$ be n independent random variables and denote by x_1, ..., x_n the observed values. Assume that each X_i has pdf $f(x;\alpha,\theta_i)$ given by (4.1). Suppose that the median of X_i satisfies $\theta_i=g(z_i^t\gamma)$, $i=1,...,n$, where $z_i=(z_{i1},...,z_{ik})^t$ is the vector of covariates associated to the response x_i, $\gamma=(\gamma_1,...,\gamma_k)$ is an unknown vector of regression coefficients and g is the link function. It is assumed that the link function g is a strictly monotonic and twice differentiable function. There are several possible choices for g satisfying the required conditions, such as the logit, probit, log-log, Cauchy, etc.

From equation (4.1), the log-likelihood function of the model with covariates is given by

$$\begin{aligned}\ell(\alpha,\gamma) \;=\; & n\log\alpha+(\alpha-1)\sum_{i=1}^{n}\log x_i+\alpha\sum_{i=1}^{n}\log\theta_i+\sum_{i=1}^{n}\log(1-\theta_i^{\alpha})\\ & -2\sum_{i=1}^{n}\log(\theta_i^{\alpha}+x_i^{\alpha}-2\theta_i^{\alpha}x_i^{\alpha}).\end{aligned}$$

The derivatives of $\ell(\alpha,\gamma)$ with respect to each parameter, which are required to compute the ML estimates of the parameters, are given by

$$\begin{aligned}\frac{\partial}{\partial\alpha}\ell(\alpha,\gamma) =\; & \frac{n}{\alpha}+\sum_{i=1}^{n}\log x_i+\sum_{i=1}^{n}\log\theta_i-\sum_{i=1}^{n}\frac{\theta_i^{\alpha}\log\theta_i}{1-\theta_i^{\alpha}}\\ & -2\sum_{i=1}^{n}\frac{\theta_i^{\alpha}\log\theta_i+x_i^{\alpha}\log x_i-2x_i^{\alpha}\theta_i^{\alpha}(\log\theta_i+\log x_i)}{\theta_i^{\alpha}+(1-2\theta_i^{\alpha})x_i^{\alpha}},\\ \frac{\partial}{\partial\gamma_r}\ell(\alpha,\gamma) =\; & \alpha\sum_{i=1}^{n}\frac{1}{\theta_i}\frac{\partial}{\partial\gamma_r}\theta_i-\alpha\sum_{i=1}^{n}\frac{\theta_i^{\alpha-1}}{1-\theta_i^{\alpha}}\frac{\partial}{\partial\gamma_r}\theta_i-2\alpha\sum_{i=1}^{n}\frac{(1-2x_i^{\alpha})\theta_i^{\alpha-1}}{\theta_i^{\alpha}+(1-2\theta_i^{\alpha})x_i^{\alpha}}\frac{\partial}{\partial\gamma_r}\theta_i,\end{aligned}$$

for $r=1,...,k$. The derivative $\frac{\partial}{\partial\gamma_r}\theta_i$ will depend on the chosen link function. For example, if it is considered the logit link, which is given by

$$\theta_i=\frac{\exp(z_i^t\gamma)}{1+\exp(z_i^t\gamma)},$$

then

$$\frac{\partial}{\partial\gamma_r}\theta_i=\theta_i(1-\theta_i)z_{ir}, \qquad i=1,...,n,\ r=1,...,k.$$

As in most regression models, for the proposed model it is possible to evaluate the marginal effects that each covariate has on the conditional median, given the covariates, by calculating (see, for example, [36, § 2.2.3])

$$\delta_{ij}=\frac{\partial\theta_i}{\partial z_{ij}}=\theta_i(1-\theta_i)\gamma_j, \qquad i=1,...,n,\ j=1,...,k. \tag{4.2}$$

This marginal effect indicates that a small change in the jth covariate, say ν, increases or decreases the conditional median θ_i by a quantity $\delta_{ij}\nu + o(\nu)$. As a summary measure of all these $k \times n$ effects, one can calculate the average marginal effects that each covariate has on the conditional median by evaluating the above derivative at $\bar{\theta} = \theta(\bar{z})$, obtaining

$$\bar{\delta_j} = \frac{\partial \bar{\theta}}{\partial z_{ij}} = \bar{\theta}(1-\bar{\theta})\gamma_j, \qquad j = 1, ..., k.$$

For the practical use of these quantities, all parameters must be replaced by estimators.

As an application, we analyze the data set considered in Subsection 3.4. The full data set consists of 73 observations on 7 variables: FIRMCOST, previously studied; ASSUME, the per occurrence retention amount as a percentage of total assets; CAP, which indicates that the firm owns a captive insurance company; SIZELOG, the logarithm of total assets; INDCOST, a measure of the firm industry risk; CENTRAL, a measure of the importance of the local managers in choosing the amount of risk to be retained; and SOPH, a measure of the degree of importance in using analytical tools.

As response variable we took x =FIRMCOST/100 and the other variables were considered as covariates. An intercept was also included in the regression model. The data were analyzed using the beta regression model and the LEEG regression model presented in this paper. Following [17], the logit link was considered in all cases. This data set was also analyzed in [17] by using the Log–Lindley regression model. Nevertheless, due to the problems observed in [23], we will not consider such model in our study. The response variables x and $1 - x$ were both studied. For the analysis of the beta regression model we used the package `betareg` (see [11]) of the R programming language [37]; to obtain the ML estimates of the parameters in the LEEG regression model we used the function `optim` of the R language. Table 4 reports the value of the log-likelihood function for the models under consideration.

Table 4: Values the of the log-likelihood with covariates for the responses x and $1 - x$.

	x	$1 - x$
Beta	87.72	87.72
LEEG	122.48	103.33

As expected, the values of the log-likelihood function for x and $1 - x$ for the beta fitting are identical, since if a random variable X has a beta law with parameters a and b, then $1 - X$ has a beta law with parameters b and a. On the other hand, the values of the log-likelihood for x and $1 - x$ for the LEEG fittings differ, since these laws do not possess the aforementioned property of the beta distribution. Hence, if the value of the log-likelihood function is used as a criterion for comparison, we see that the best fit is obtained for the LEEG regression model for the response variable x.

In addition, we applied the Vuong test [42] for testing the null hypothesis that both models are equally close to the actual model, against the alternative that one model is closer than the other. The test rejected the null hypothesis in favor of the hypothesis that the LEEG regression model is closer than the beta regression model (the p-value is 0.0012).

We also compared the Pearson residuals of both models. Figure 2 displays them.

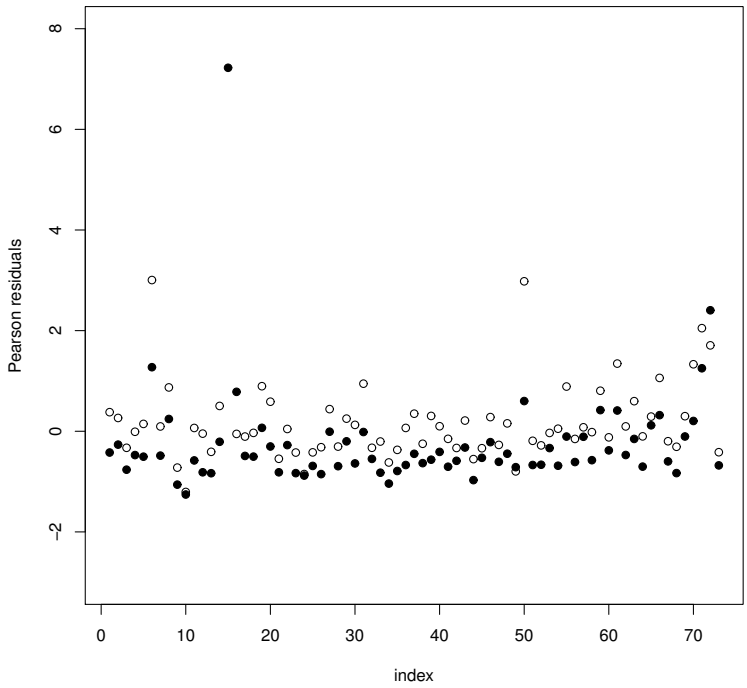

Figure 2: Pearson residuals for the beta regression model (black) and the LEEG regression model (white).

Table 5 displays the estimation results for the LEEG regression model with response variable x. The standard errors of the parameter estimates were approximated by means of the square root of the diagonal elements of the negative of the observed information matrix, that is, the matrix whose entries are the second order derivatives of the log-likelihood (its expression is omitted for the sake of brevity). The p-values of the Wald test for testing the nullity of each parameter were calculated by using the normal approximation.

Table 5: Parameter estimates for the LEEG regression model with response x and average marginal effects (a.m.e.).

Parameter	Estimate	S.E.	t-Wald	p-value	a.m.e.
α	2.20257	0.22661	9.71975	0.0000	
Intercept	3.98741	1.21128	3.29191	0.0010	
ASSUME	−0.01234	0.01216	−1.01482	0.3102	−0.00080
CAP	−0.05257	0.22327	−0.23545	0.8139	−0.00340
SIZELOG	−0.90907	0.12466	−7.29242	0.0000	−0.05884
INDCOST	2.34318	0.62296	3.76138	0.0002	0.15166
CENTRAL	−0.13648	0.08385	−1.62766	0.1036	−0.00883
SOPH	0.00932	0.01965	0.47398	0.6355	0.00060

From these results, it can be inferred that the covariates SIZELOG and INDCOST have a significant non-null effect on the response variable. These two covariates have the largest average marginal effects, negative for SIZELOG, indicating that an increase in SIZELOG

diminishes the median of the response variable, and positive for INDCOST, indicating that an increase in INDCOST increases the median of the response variable.

Before ending this section we would like to remark that the lack of a simple expression for the quantiles of the classic beta distribution hampers the development of a quantile regression based on it.

A. APPENDIX

This appendix is devoted to present a known result concerning a logarithmic integral. Such result will be used to solve in a unified manner the integrals arising in Appendices B and C.

For any real numbers $a \geq 0$, $s \geq 1$ and $z > -1$, denote by

$$\Gamma_n(z, s, a) = \int_0^1 \frac{u^a \log^{s-1}(1/u)}{(1+zu)^{n+1}}\, du, \qquad n = 1, 2, ... \tag{A.1}$$

Jodrá and Jiménez-Gamero [22] showed that $\Gamma_n(z, s, a)$ can be expressed as a finite sum involving the Lerch transcendent function together with the generalized Stirling numbers of the first kind. To be more precise, Mitrinović [31] defined the generalized Stirling numbers of the first kind, $R_n^j(\rho, \tau)$, by means of the following generating function

$$\prod_{j=0}^{n-1}(w - \rho - \tau j) = \sum_{j=0}^{n} R_n^j(\rho, \tau) w^j,$$

where n is a non-negative integer and ρ, τ are complex numbers with $\tau \neq 0$. Mitrinović [31] expressed these numbers in terms of the best-known signed Stirling numbers of the first kind $R_n^j(0, 1)$ (see Abramowitz and Stegun [1, p. 824])

$$R_n^j(\rho, \tau) = \sum_{k=0}^{n-j} \binom{j+k}{k} (-1)^k \rho^k \tau^{n-j-k} R_n^{j+k}(0, 1), \qquad \rho \neq 0, \tag{A.2}$$

which is important from a computational point of view since the numbers $R_n^j(0, 1)$ are available in most computer algebra systems. Jodrá and Jiménez-Gamero [22, Theorem 2.1] established that for any $a \geq 0$, $s \geq 1$ and $z > -1$,

$$\Gamma_n(z, s, a) = \frac{\Gamma(s)}{\Gamma(n+1)} \sum_{j=0}^{n} R_n^j(a - n + 1, 1)\Phi(-z, s - j, a + 1), \qquad n = 1, 2, ..., \tag{A.3}$$

which in the special case $z = 0$ becomes $\Gamma_n(0, s, a) = \Gamma(s)/(a+1)^s$. Additionally, (A.3) can be expressed in terms of the polylogarithm function if $a = 0, 1, ..., n-1$ (see [22, Corollary 2.6] and also [21]), specifically,

$$\Gamma_n(z, s, a) = \frac{\Gamma(s)}{(-z)^{a+1}\Gamma(n+1)} \sum_{j=1}^{n} R_n^j(a - n + 1, 1)\mathrm{Li}_{s-j}(-z). \tag{A.4}$$

It is interesting to note that the Lerch transcendent function includes as a particular case the polylogarithm function, more precisely, $\mathrm{Li}_\lambda(z) = z\Phi(z, \lambda, 1)$ (see Apostol [5]). In particular, the case $\lambda = 1$ corresponds to the natural logarithm, $\mathrm{Li}_1(z) = -\log(1 - z)$, and the case $\lambda = 2$ is known as dilogarithm or polylogarithm function of order two.

B. APPENDIX

Here, we give the proofs of the results stated in Section 2.

Proof of Proposition 2.1: The conditional cdf of the random variable $V|N=n$ is $F_{V|N=n}(v;\alpha)=1-(1-v^\alpha)^n$, with $0<v<1$, $\alpha>0$ and $n=1,2,....$ Then, it is clear the following

$$P(V\le v, N=n)=[1-(1-v^\alpha)^n]\left(1-\frac{1}{1+\beta}\right)^{n-1}\frac{1}{1+\beta},$$

where $\beta>0$. Hence, part (i) follows from the fact that the marginal cdf of V is

$$F_V(v;\alpha,\beta)=\sum_{n=1}^{\infty}P(V\le v,N=n)=\frac{(1+\beta)v^\alpha}{1+\beta\, v^\alpha},\quad 0<v<1,\ \alpha>0,\ \beta>0.$$

The proof of part (ii) follows a similar pattern. The conditional cdf of $W|M=m$ is $F_{W|M=m}(w;\alpha)=w^{\alpha\, m}$, with $0<w<1$, $\alpha>0$ and $m=1,2,....$ Therefore, $P(W\le w, M=m)=w^{\alpha\, m}(-\beta)^{m-1}(1+\beta)$, where $\beta\in(-1,0)$. Finally, considering that $F_W(w;\alpha,\beta)=\sum_{m=1}^{\infty}P(W\le w,M=m)$ the result is obtained. □

Proof of Proposition 2.2: The first derivative of (1.1) is given by

$$\frac{\partial}{\partial x}f(x;\alpha,\beta)=-\frac{\alpha(1+\beta)}{(1+\beta x^\alpha)^3}\,[\beta(1+\alpha)x^\alpha-(\alpha-1)]. \tag{B.1}$$

The solution of the equation $(\partial/\partial x)f(x;\alpha,\beta)=0$ is $x_0=\left(\frac{\alpha-1}{(1+\alpha)\beta}\right)^{1/\alpha}$. Moreover, after some calculations, it can be checked that

$$\frac{\partial^2}{\partial x^2}f(x;\alpha,\beta)\Big|_{x=x_0}=-\frac{(1+\beta)(1+\alpha)^2(\alpha-1)^2}{8\alpha\beta}.$$

On the one hand, if $\alpha>1$ and $\beta>(\alpha-1)/(1+\alpha)$ then $x_0\in(0,1)$ and $\frac{\partial^2}{\partial x^2}f(x;\alpha,\beta)\big|_{x=x_0}<0$ which implies that x_0 is the mode of X. In addition, from (B.1), it can be seen that (1.1) is an increasing function if $\alpha>1$ and $\beta\in(-1,(\alpha-1)/(1+\alpha)]$ since $(\partial/\partial x)f(x;\alpha,\beta)>0$. This proves part (i). On the other hand, if $0<\alpha<1$ and $\beta<(\alpha-1)/(1+\alpha)$ then $x_0\in(0,1)$ and $\frac{\partial^2}{\partial x^2}f(x;\alpha,\beta)\big|_{x=x_0}>0$ which implies that (1.1) achieves a minimum at x_0. It can also be checked that (1.1) is a decreasing function if $0<\alpha<1$ and $\beta\ge(\alpha-1)/(1+\alpha)$. This proves part (ii). Part (iii) is directly obtained from (1.1). □

Proof of Proposition 2.3: For any $k=1,2,...$, the k-th moment of X can be computed as follows

$$E[X^k]=\int_0^1 x^k f(x;\alpha,\beta)dx=\int_0^1 x^k\frac{\alpha(1+\beta)x^{\alpha-1}}{(1+\beta\, x^\alpha)^2}\,dx=(1+\beta)\int_0^1\frac{u^{k/\alpha}}{(1+\beta\, u)^2}\,du,$$

where in the last equality we have made the change of variable $x^\alpha=u$. Hence, the k-th moment of X can be rewritten as below

$$E[X^k]=(1+\beta)\int_0^1\frac{u^{k/\alpha}}{(1+\beta\, u)^2}\,du=(1+\beta)\Gamma_1(\beta,1,k/\alpha),$$

where Γ_1 is given by equation (A.1). Using equation (A.3), we have

$$\Gamma_1(\beta, 1, k/\alpha) = R_1^1(k/\alpha, 1)\Phi\left(-\beta, 0, 1 + \frac{k}{\alpha}\right) + R_1^0(k/\alpha, 1)\Phi\left(-\beta, 1, 1 + \frac{k}{\alpha}\right).$$

By virtue of (A.2), $R_1^1(k/\alpha, 1) = 1$ and $R_1^0(k/\alpha, 1) = -k/\alpha$ since $R_1^0(0, 1) = 0$ and $R_1^1(0, 1) = 1$. Moreover, $\Phi(-\beta, 0, 1 + k/\alpha) = 1/(1 + \beta)$. Hence, the result is obtained. □

Proof of Proposition 2.4: The result is obtained directly by solving the equation $F(x; \alpha, \beta) = u$, $0 < u < 1$, with respect to the variable x. □

Proof of Proposition 2.5: For any $n = 1, 2, ...$, the k-th moment of the largest order statistic $X_{n:n}$ is given by

$$E[X_{n:n}^k] = n \int_0^1 x^k \left[F(x; \alpha, \beta)\right]^{n-1} f(x; \alpha, \beta) dx = n(1 + \beta)^n \int_0^1 \frac{u^{k/\alpha + n - 1}}{(1 + \beta u)^{n+1}} du,$$

where in the second equality we have made the change of variable $u = x^\alpha$. Now, taking into account equation (A.1), $E[X_{n:n}^k]$ can be written as follows

$$E[X_{n:n}^k] = n(1 + \beta)^n \Gamma_n \left(\beta, 1, \frac{k}{\alpha} + n - 1\right).$$

Finally, the claimed result follows by applying equation (A.3) in the above equation. □

Proof of Proposition 2.6: Let us denote $v(x) = \frac{\partial}{\partial x} \log\left(\frac{f(x;\alpha,\beta_2)}{f(x;\alpha,\beta_1)}\right) = \frac{num}{den}$, where $den = x(1 + \beta_1 x^\alpha)(1 + \beta_2 x^\alpha)$ and $num = 2\alpha x^\alpha(\beta_1 - \beta_2)$. It can be checked that $den > 0$ for any $x \in (0, 1)$, $\alpha > 0$ and $\beta_1, \beta_2 > -1$ and also that $num \geq 0$ for any $x \in (0, 1)$ and $\alpha > 0$ if and only if $\beta_1 \geq \beta_2$. Since $v(x) \geq 0$ implies that $\frac{f(x;\alpha,\beta_2)}{f(x;\alpha,\beta_1)}$ is non-decreasing in x, the result follows. □

C. APPENDIX

Here, we give the proofs of the results presented in Subsection 3.2.

Proof of Proposition 3.1: The Hessian matrix of $\log L(\alpha, \beta)$ is defined by

$$H(\alpha, \beta) = \begin{bmatrix} \dfrac{\partial^2 \log L(\alpha, \beta)}{\partial \alpha^2} & \dfrac{\partial^2 \log L(\alpha, \beta)}{\partial \alpha \partial \beta} \\ \dfrac{\partial^2 \log L(\alpha, \beta)}{\partial \beta \partial \alpha} & \dfrac{\partial^2 \log L(\alpha, \beta)}{\partial \beta^2} \end{bmatrix},$$

with

$$\frac{\partial^2}{\partial \alpha^2} \log L(\alpha, \beta) = -\frac{n}{\alpha^2} - 2\beta \sum_{i=1}^n \frac{x_i^\alpha (\log x_i)^2}{(1 + \beta\, x_i^\alpha)^2}, \tag{C.1}$$

$$\frac{\partial^2}{\partial \alpha \partial \beta} \log L(\alpha, \beta) = -2 \sum_{i=1}^n \frac{x_i^\alpha \log x_i}{(1 + \beta\, x_i^\alpha)^2}, \tag{C.2}$$

$$\frac{\partial^2}{\partial \beta^2} \log L(\alpha, \beta) = -\frac{n}{(1 + \beta)^2} + 2 \sum_{i=1}^n \frac{x_i^{2\alpha}}{(1 + \beta\, x_i^\alpha)^2}. \tag{C.3}$$

From (C.1)–(C.3), the Fisher information matrix, $I(\alpha,\beta) = -E[H(\alpha,\beta)]$, is given by

$$I(\alpha,\beta) = \begin{bmatrix} \dfrac{n}{\alpha^2} + 2\beta n \displaystyle\int_0^1 \frac{x^\alpha (\log x)^2}{(1+\beta\, x^\alpha)^2}\, f(x)dx & 2n \displaystyle\int_0^1 \frac{x^\alpha \log x}{(1+\beta\, x^\alpha)^2}\, f(x)dx \\ 2n \displaystyle\int_0^1 \frac{x^\alpha \log x}{(1+\beta\, x^\alpha)^2}\, f(x)dx & \dfrac{n}{(1+\beta)^2} - 2n \displaystyle\int_0^1 \frac{x^{2\alpha}}{(1+\beta\, x^\alpha)^2}\, f(x)dx \end{bmatrix},$$

where we have used the notation $f(x)$ instead of $f(x;\alpha,\beta)$ for brevity. Below, we consider each integral expression in the elements of $I(\alpha,\beta)$. Let us first assume that $\beta \neq 0$. Making the change of variable $u = x^\alpha$ and taking into account (A.1), those integrals can be expressed as follows

$$\begin{aligned} \int_0^1 \frac{x^\alpha (\log x)^2}{(1+\beta\, x^\alpha)^2} f(x)dx &= \frac{1+\beta}{\alpha^2} \int_0^1 \frac{u(\log(1/u))^2}{(1+\beta\, u)^4}\, du = \frac{1+\beta}{\alpha^2}\, \Gamma_3(\beta,3,1), \\ \int_0^1 \frac{x^\alpha \log x}{(1+\beta\, x^\alpha)^2} f(x)dx &= -\frac{1+\beta}{\alpha} \int_0^1 \frac{u \log(1/u)}{(1+\beta\, u)^4}\, du = -\frac{1+\beta}{\alpha}\, \Gamma_3(\beta,2,1), \\ \int_0^1 \frac{x^{2\alpha}}{(1+\beta\, x^\alpha)^2} f(x)dx &= (1+\beta) \int_0^1 \frac{u^2}{(1+\beta\, u)^4}\, du = (1+\beta)\Gamma_3(\beta,1,2). \end{aligned}$$

Now, by virtue of (A.4) and after some calculations we get

$$\begin{aligned} \Gamma_3(\beta,3,1) &= -\frac{1}{3\beta}\left(\frac{\mathrm{Li}_2(-\beta)}{\beta} + \frac{1}{1+\beta}\right), \\ \Gamma_3(\beta,2,1) &= \frac{1}{6\beta}\left(\frac{\log(1+\beta)}{\beta} - \frac{1}{(1+\beta)^2}\right), \\ \Gamma_3(\beta,1,2) &= \frac{1}{3(1+\beta)^3}, \end{aligned}$$

where Li_2 denotes the polylogarithm function of order two. Now, the stated result is obtained by substituting in the elements of $I(\alpha,\beta)$ the value of the corresponding integrals.

The result for $\beta = 0$ is derived by means of routine calculations, so we omit the details. □

Proof of Proposition 3.2: The result follows by using standard large sample theory results for ML estimators (for example, by applying Lehmann and Casella [29, Theorem 5.1, p. 463]). In particular, the asymptotic covariance matrix of the ML estimators, Σ, is obtained by inverting the expected Fisher information matrix $(1/n)I(\alpha,\beta)$, with $I(\alpha,\beta)$ provided in Proposition 3.1. □

ACKNOWLEDGMENTS

The authors thank the anonymous referee for his/her constructive comments, which led to an improvement of the paper. Research of P. Jodrá has been partially funded by grant of Diputación General de Aragón –Grupo E24-17R– and ERDF funds. Research of M.D. Jiménez-Gamero has been partially funded by grant MTM2017-89422-P of the Spanish Ministry of Economy, Industry and Competitiveness, ERDF support included.

REFERENCES

[1] Abramowitz, M. and Stegun, I.A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York.

[2] Adamidis, K.; Dimitrakopoulou, T. and Loukas, S. (2005). On an extension of the exponential-geometric distribution, *Statistics & Probability Letters*, **73**(3), 259–269.

[3] Adamidis, K. and Loukas, S. (1998). A lifetime distribution with decreasing failure rate, *Statistics & Probability Letters*, **39**(1), 35–42.

[4] Akaike, H. (1974). A new look at statistical model identification, *IEEE Transactions on Automatic Control*, **19**(6), 716–723.

[5] Apostol, T.M. (2010). *Zeta and related functions*. In "NIST Handbook of Mathematical Functions" (F.W.F. Olver, D.W. Lozier, R.F. Boisvert and C.W. Clark, Eds.), National Institute of Standards and Technology, Washington, DC, and Cambridge University Press, Cambridge, 601–616.

[6] Babu, G.J. and Rao, C.R. (2004). Goodness-of-fit tests when parameters are estimated, *Sankhyā*, **66**(1), 63–74.

[7] Bakoban, R.A. and Abu-Zinadah, H.H. (2017). The beta generalized inverted exponential distribution with real data applications, *REVSTAT Statistical Journal*, **15**(1), 65–88.

[8] Barreto-Souza, W.; Lemos de Morais, A. and Cordeiro, G.M. (2011). The Weibull-geometric distribution, *Journal of Statistical Computation and Simulation*, **81**(5), 645–657.

[9] Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, **52**(3), 345–370.

[10] Cordeiro, G.M.; Ortega, E.M.M. and Nadarajah, S. (2010). The Kumaraswamy Weibull distribution with application to failure data, *Journal of the Franklin Institute. Engineering and Applied Mathematics*, **347**(8), 1399–1429.

[11] Cribari-Neto, F. and Zeileis, A. (2010). Beta Regression in R, *Journal of Statistical Software*, **34**(2), 1–24.

[12] D'Agostino, R.B. and Stephens, M.A. (Eds.) (1986). *Goodness-of-Fit-Techniques*, Marcel Dekker, New York.

[13] David, H.A. and Nagaraja, H.N. (2003). *Order Statistics*, 3rd edition, John Wiley & Sons, Hoboken, New Jersey.

[14] Epps, T.W. and Pulley, L.B. (1983). A test for normality based on the empirical characteristic function, *Biometrika*, **70**(3), 723–726.

[15] Ferrari, S.L.P. and Cribari-Neto F. (2004). Beta regression for modelling rates and proportions, *Journal of Applied Statistics*, **31**(7), 799–815.

[16] Frees, E.W. (2010). *Regression Modeling with Actuarial and Financial Applications*, International Series on Actuarial Science, Cambridge University Press, Cambridge.

[17] Gómez-Déniz, E.; Sordo, M.A. and Calderín-Ojeda, E. (2014). The Log-Lindley distribution as an alternative to the beta regression model with applications in insurance, *Insurance: Mathematics & Economics*, **54**, 49–57.

[18] Gómez-Déniz, E.; Vázquez-Polo, F.J. and García, V. (2017). The Modified Borel-Tanner (MBT) regression model, *REVSTAT Statistical Journal*, **15**(3), 425–442.

[19] Jiménez-Gamero, M.D.; Alba-Fernández, M.V.; Jodrá, P. and Barranco-Chamorro, I. (2015). An approximation to the null distribution of a class of Cramér-von Mises statistics, *Mathematics and Computers in Simulation*, **118**, 258–272.

[20] Jiménez-Gamero, M.D.; Alba-Fernández, M.V.; Muñoz-García, J. and Chalco-Cano, Y. (2009). Goodness-of-fit tests based on empirical characteristic functions, *Computational Statistics & Data Analysis*, **53**(12), 3957–3971.

[21] Jodrá, P. (2008). On a connection between the polylogarithm function and the Bass diffusion model, *Proceedings of The Royal Society of London. Series A. Mathematical, Physical and Engineering Sciences*, **464**(2099), 3081–3088.

[22] Jodrá, P. and Jiménez-Gamero, M.D. (2014). On a logarithmic integral and the moments of order statistics from the Weibull-geometric and half-logistic families of distributions, *Journal of Mathematical Analysis and Applications*, **410**(2), 882–890.

[23] Jodrá, P. and Jiménez-Gamero, M.D. (2016). A note on the Log-Lindley distribution, *Insurance: Mathematics & Economics*, **71**, 186–194.

[24] Jodrá, P.; Jiménez-Gamero, M.D. and Alba-Fernández, M.V. (2015). On the Muth Distribution, *Mathematical Modelling and Analysis*, **20**(3), 291–310.

[25] Jones, M.C. (2009). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages, *Statistical Methodology*, **6**(1), 70–81.

[26] Jorgensen, B. (1997). *The Theory of Dispersion Models*, Chapman & Hall, London.

[27] Koenker, R. (2005). *Quantile Regression*, Cambridge University Press, New York.

[28] Kumaraswamy, P. (1980). Generalized probability density-function for double-bounded random-processes, *Journal of Hydrology*, **46**,(1-2), 79–88.

[29] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd edition, Springer Texts in Statistics, Springer-Verlag, New York.

[30] Meintanis, S.G.; Jiménez-Gamero, M.D. and Alba-Fernández, M.V. (2014). A class of goodness-of-fit tests based on transformation, *Communications in Statistics. Theory and Methods*, **43**(8), 1708–1735.

[31] Mitrinović, D.S. (1961). Sur une classe de nombres reliés aux nombres de Stirling, *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences de Paris*, **252**, 2354–2356.

[32] Noufaily, A. and Jones, M.C. (2013). Parametric quantile regression based on the generalized gamma distribution, *Journal of the Royal Statistical Society. Series C. Applied Statistics*, **62**(5), 723–740.

[33] Papke, L.E. and Wooldridge, J.M. (1996). Econometric methods for fractional response variables with an application to 401 (K) plan participation rates, *Journal of Applied Econometrics*, **11**, 619–632.

[34] Papke, L.E. and Wooldridge, J.M. (2008). Panel data methods for fractional response variables with an application to test pass rates, *Journal of Econometrics*, **145**(1-2), 121–133.

[35] Pascoa, M.A.R.; Ortega, E.M.M. and Cordeiro, G.M. (2011). The Kumaraswamy generalized gamma distribution with application in survival analysis, *Statistical Methodology*, **8**(5), 411–433.

[36] Pérez-Rodríguez, J.V. and Gómez-Déniz, E. (2015). Spread component costs and stock trading characteristics in the Spanish Stock Exchange. Two flexible fractional response models, *Quantitative Finance*, **15**(12), 1943–1962.

[37] R Development Core Team (2017). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/

[38] Schmit, J.T. and Roth, K. (1990). Cost effectiveness of risk management practices, *The Journal of Risk and Insurance*, **57**(3), 455–470.

[39] Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**(2), 461–464.

[40] Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*, Springer-Verlag, New York.

[41] Stute, W.; González Manteiga, W. and Presedo Quindimil, M. (1993). Bootstrap based goodness-of-fit tests, *Metrika*, **40**(3-4), 243–256.

[42] Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica*, **57**(2), 307–333.

REVSTAT – Statistical Journal
Volume 18, Number 4, October 2020, 437–452

PARAMETERS ESTIMATION FOR CONSTANT-STRESS PARTIALLY ACCELERATED LIFE TESTS OF GENERALIZED HALF-LOGISTIC DISTRIBUTION BASED ON PROGRESSIVE TYPE-II CENSORING

Author: Abdullah M. Almarashi
– Statistics Department, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia
aalmarashi@kau.edu.sa

Received: October 2017 Revised: May 2018 Accepted: May 2018

Abstract:

- In product-life testing experiments, the accelerated life testing (ALT) is applied to reduce the time and cost of tests. We consider the constant-stress partially ALT model when the lifetime of units under normal conditions follow the generalized half-logistic lifetime distribution based on progressive Type-II censored schemes. The likelihood functions of the parameters are derived and solved to present the maximum likelihood estimators of the model parameters. The approximate and two bootstrap confidence intervals are also proposed. The performance of the different methods were measured and compared through Monte Carlo simulation study. Finally, the results of a numerical example are discussed.

Keywords:

- *constant-stress partially accelerated life tests; generalized half-logistic distribution; maximum likelihood estimation; bootstrap confidence intervals.*

1. INTRODUCTION

According to [22, 18, 3, 4], there are different methods of accelerated life testing (ALT): the constant-stress ALT, in which the stress on the life test product remains at a constant level, the progressive-stress ALT, in which the stress applied to the product units in the test increases with time [7], and the step-stress ALT, in which the test condition changes for a given time or a specified number of failures [21, 7]. For more recent research on the constant-stress partially ALT, see [2, 1].

In product-life test experiments, censoring has played an important role. Different types of censoring are available. Type-I and Type-II censoring schemes (CSs) are commonly applied, both of which do not allow the removal of any units other than at the terminal point of the test. General CSs that allow units to be removed at any point during the test are called progressive Type-II right censoring. For important reviews of the literature on progressive censoring, see [9].

Let n be the number of units tested in a product-life testing experiment and T_1, T_2, ..., T_n, be the corresponding lifetimes. Assume that the T_i, $i = 1, 2, ..., n$ are independent and identically distributed (i.i.d.) with probability density function (PDF) $f(.)$ and cumulative distribution function (CDF) $F(.)$. In the progressive Type-II CS prior to the experiment, the effective sample size m and the corresponding CS $\mathbf{R} = \{R_1, R_2, ..., R_m\}$ are determined; then $T^{\mathbf{R}}_{i;m,n}$, $i = 1, 2, ..., m$ is the corresponding random variable of the progressive Type-II censored sample.

The joint likelihood function of the observed progressive Type-II censored sample $\underline{t} = (t^{\mathbf{R}}_{1;m,n}, t^{\mathbf{R}}_{2;m,n}, ..., t^{\mathbf{R}}_{m;m,n})$ is given by

$$f(\underline{t}, \theta) = Q \prod_{i=1}^{m} f(t^{\mathbf{R}}_{i;m,n})[1 - F(t^{\mathbf{R}}_{i;m,n})]^{Ri}, \tag{1.1}$$

where the observed progressive Type-II censored sample $\underline{t}$ satisfies $0 < t_{1;m,n} < t_{2;m,n} < ... < t_{m;m,n} < \infty$,and

$$Q = \prod_{i=0}^{m-1} \left(n - \sum_{j=0}^{i} R_j - i \right), \ R_0 = 0. \tag{1.2}$$

Balakrishnan [8] has considered the half-logistic distribution as the distribution of the absolute standard logistic variate. Important properties of a generalized version of the logistic distribution are discussed by Balakrishnan and Hossain [10]. The point estimation of the stress–strength reliability of generalized half-logistic distribution (GHLD) is presented by Ramakrishnan [23]. The shape parameter of the GHLD was estimated under Type-I progressive censoring in Arora et al. [5]. The Bayesian approach with a GHLD was discussed in Kim et al. [20]. Recently, testing procedures for the reliability functions of the GHLD were considered in Chaturvedi et al. [14] and in a Type-I generalized half-logistic survival model in Awodutire et al. [6].

Let T be a random variable of a GHLD with shape parameter β; the PDF and CDF are given respectively by

$$f(t) = \frac{\beta}{1+\exp(-t)}\left(\frac{2\exp(-t)}{1+\exp(-t)}\right)^{\beta},\ t>0,\ \beta>0, \tag{1.3}$$

and

$$F(x)=1-\left(\frac{2\exp(-t)}{1+\exp(-t)}\right)^{\beta}, \tag{1.4}$$

The reliability function $S(t)$ and the hazard rate function $H(t)$ are expressed as

$$S(t) = \left(\frac{2\exp(-t)}{1+\exp(-t)}\right)^{\beta}, t>0,\ \beta>0, \tag{1.5}$$

and

$$H(t) = \frac{\beta}{1+\exp(-t)}. \tag{1.6}$$

This GHLD is considered as a special probability distribution with a location parameter and a scale parameter, defined by $F(x)=1-\left(\frac{2\exp(\frac{-t}{\sigma})}{1+\exp(\frac{-t}{\sigma})}\right)^{\beta}$ with $\sigma=1$. The best linear unbiased estimator of the location and scale parameters as well as the values of the variance and covariance of these estimators is presented in [11]. Ref. [13] discusses the estimator as an approximation of the likelihood functions based on a Type-II censoring sample. The estimation of the parameter of the half-logistic distribution under progressive Type-II censored sample is presented in [19].

The aim of this paper is to estimate the GHLD under constant-stress partially ALT with progressive Type-II CS. The maximum likelihood estimator (MLE) and the bootstrap estimator of each unknown GHLD parameter and the acceleration factor are presented. The point estimates of the MLE and bootstrap estimator mainly assess and compare their biases and mean-squared errors (MSE's), as well as the approximate interval estimation and bootstrap confidence intervals (CIs), with respect to coverage percentage and the mean of interval lengths using extensive simulation studies.

In this article, the assumptions and model are described in Section 2. The MLEs and the corresponding approximate confidence intervals (ACIs) are given in Section 3. Two bootstrap CIs are discussed in Section 4. We assess and compare the results of Monte Carlo studies in Section 5. A numerical example of a simulated data set is presented in Section 6. Finally, some comments about the results of the simulation studies are presented in Section 7.

2. ASSUMPTIONS AND MODEL

In the experiment design for the constant-stress partially ALTs, n_1 units from n testing units are randomly chosen to be tested under normal conditions; the remaining units $n_2 = n - n_1$ are tested under accelerated conditions. The model for the progressive Type-II censoring with constant-stress partially ALTs is described as follows. The subscript label $j = 1, 2$ signify the two conditions, normal and accelerated; when the first failure $T^{\mathbf{R}j}_{j1;m_j,n_j}$ is recorded, R_{j1} units are randomly removed from the remaining $n_j - 1$ surviving units. Also at the second failure, $T^{\mathbf{R}j}_{j2;m_j,n_j}$ is recorded and R_{j2} units from the remaining $n_j - 2$-R_{j1} units are randomly removed. The test continues until the m_j -th $T^{\mathbf{R}j}_{jm_j;m_j,n_j}$ failure and the remaining $R_{jmj} = n_j$-$m_j - \sum_{k=1}^{m_j-1} R_{jk}$ units are removed, for $j = 1, 2$. In this model, each of the R_{ji} and $m_j < n_j$ are fixed prior to beginning the test. If the times of failure of the n_j units originally in the test are from a continuous population with a distribution function $F_j(t)$ and probability density function $f_j(t)$, the joint probability density function for $T^{\mathbf{R}j}_{j1;m_j,n_j} < T^{\mathbf{R}j}_{j2;m_j,n_j} < ... < T^{\mathbf{R}j}_{jm_j;m_j,n_j}$ and $j = 1, 2$ is given as follows.

The joint likelihood function for $\underline{t} = (T^{\mathbf{R}j}_{j1;m_j,n_j}, T^{\mathbf{R}j}_{j2;m_j,n_j}, ..., T^{\mathbf{R}j}_{jm_j;m_j,n_j})$ for $j = 1, 2$, is given by

$$L(\beta, \lambda|\underline{t}) = \prod_{j=1}^{2} Q_j \left\{ \prod_{i=1}^{m_j} f_j(t^{\mathbf{R}j}_{ji;m_j,n_j}) \left(S_j(t^{\mathbf{R}j}_{ji;m_j,n_j}) \right)^{R_{ji}} \right\}, \tag{2.1}$$

where $Q_j = \prod_{i=0}^{m_j-1} \left(n_j - \sum_{l=0}^{i} R_{lj} - i \right)$, $R_{0j} = 0$. In the accelerated lifetime model, assuming that $S_2(t) = S_1(\lambda t)$. Let T be a random variable under normal conditions, then the lifetime of the unit under accelerated conditions can be defined by $Y = \frac{T}{\lambda}$, where λ is the acceleration factor. Hence, the probability density and cumulative distribution functions of the GHLD with observed lifetime under the accelerated condition are given by

$$f_2(y) = \frac{\lambda\beta}{1 + \exp(-\lambda y)} \left(\frac{2\exp(-\lambda y)}{1 + \exp(-\lambda y)} \right)^{\beta}, y > 0,\ \beta,\ \lambda > 0. \tag{2.2}$$

and

$$F_2(y) = 1 - \left(\frac{2\exp(-\lambda y)}{1 + \exp(-\lambda y)} \right)^{\beta}. \tag{2.3}$$

Also, the reliability function $S(y)$ and hazard rate function $H(y)$ are given, respectively, by

$$S_2(y) = \left(\frac{2\exp(-\lambda y)}{1 + \exp(-\lambda y)} \right)^{\beta}, \tag{2.4}$$

and

$$H_2(y) = \frac{\lambda\beta}{1 + \exp(-\lambda y)}. \tag{2.5}$$

3. MAXIMUM LIKELIHOOD ESTIMATION

3.1. Point estimation

Let $\underline{T} = \left(T^{\mathbf{R}j}_{j1;m_j,n_j} < T^{\mathbf{R}j}_{j2;m_j,n_j} < ... < T^{\mathbf{R}j}_{jm_j;m_j,n_j}\right)$, $j = 1, 2$ denote two progressively Type-II censored samples from two populations for which the PDFs and CDFs are as given in (1.3), (1.4), (2.2), and (2.3), with $\mathbf{R}_j = (R_{j1}, R_{j2}, ..., R_{j1})$. The log-likelihood function $\ell(\beta, \lambda|\underline{t}) = \log L(\beta, \lambda|\underline{t})$ without normalized constant is then given by

$$\begin{aligned}\ell(\beta,\lambda|\underline{t}) &= (m_1+m_2)\log\beta + m_2\log\lambda + n\beta\log 2 - \sum_{i=1}^{m_1}\log\left[1+\exp(-t_{1i})\right)] \\ &- \sum_{i=1}^{m_2}\log\left[1+\exp(-\lambda t_{2i})\right)] - \beta\sum_{i=1}^{m_1}(R_{1i}+1)\log\left(1+\exp(t_{1i})\right) \\ &-\beta\sum_{i=1}^{m_2}(R_{2i}+1)\log\left(1+\exp(\lambda t_{2i})\right).\end{aligned} \tag{3.1}$$

The likelihood equation is obtained by calculating the first partial derivatives of (3.1) with respect to β and λ, and then equating each to zero:

$$\begin{aligned}\frac{\partial\ell(\beta,\lambda|\underline{t})}{\partial\beta} &= \frac{m_1+m_2}{\beta} + n\log 2 - \sum_{i=1}^{m_1}(R_{1i}+1)\log\left(1+\exp(t_{1i})\right) \\ &- \sum_{i=1}^{m_2}(R_{2i}+1)\log\left(1+\exp(\lambda t_{2i})\right) = 0,\end{aligned} \tag{3.2}$$

giving

$$\begin{aligned}\beta(\lambda) = -(m_1+m_2)\Bigg[&\sum_{i=1}^{m_1}(R_{1i}+1)\log\left(1+\exp(t_{1i})\right) \\ &+ \sum_{i=1}^{m_2}(R_{2i}+1)\log\left(1+\exp(\lambda t_{2i})\right) - n\log 2\Bigg]^{-1},\end{aligned} \tag{3.3}$$

and

$$\begin{aligned}\frac{\partial\ell(\beta,\lambda|\underline{t})}{\partial\lambda} &= \frac{m_2}{\lambda} + \sum_{i=1}^{m_2} t_{2i}\left(1+\exp(\lambda t_{2i})\right)^{-1} + \beta\sum_{i=1}^{m_2}(R_{2i}+1) \\ &\times t_{2i}\left(1+\exp(-\lambda t_{2i})\right)^{-1} = 0,\end{aligned} \tag{3.4}$$

giving

$$\frac{m_2}{\lambda} + \sum_{i=1}^{m_2} t_{2i}\left(1+\exp(-\lambda t_{2i})\right)^{-1} + \beta\sum_{i=1}^{m_2}(R_{2i}+1)t_{2i}\left(1+\exp(-\lambda t_{2i})\right)^{-1} = 0\,. \tag{3.5}$$

The likelihood equation reduce to the single nonlinear equation (3.5), which can be solved numerically using the fixed point method or the quasi-Newton Raphson to obtain the MLE of λ say $\hat{\lambda}$, and hence $\hat{\beta}$ using (3.3).

3.2. Approximate interval estimation

The asymptotic normality theory is applied to construct asymptotic CIs of the MLEs. The Fisher information matrix requires the second partial derivatives of (3.1) with respect to β and λ:

$$\frac{\partial^2 \ell(\alpha,\beta,\lambda|\underline{t})}{\partial\beta^2} = \frac{-(m_1+m_2)}{\beta}, \tag{3.6}$$

$$\frac{\partial^2 \ell(\beta,\lambda|\underline{t})}{\partial\beta\partial\lambda} = \frac{\partial^2 \ell(\beta,\lambda|\underline{t})}{\partial\lambda\partial\beta} = -\beta\sum_{i=1}^{m_2}(R_{2i}+1)t_{2i}\left(1+\exp(-\lambda t_{2i})\right)^{-1}, \tag{3.7}$$

and

$$\frac{\partial^2 \ell(\beta,\lambda|\underline{t})}{\partial\lambda^2} = \frac{-m_2}{\lambda^2} - \sum_{i=1}^{m_2} t_{2i}^2\left(1+\exp(-\lambda t_{2i})\right)^{-2} + \beta\sum_{i=1}^{m_2}(R_{2i}+1)t_{2i}^2\exp(\lambda t_{2i}) \times\left(1+\exp(\lambda t_{2i})\right)^{-2}. \tag{3.8}$$

Then, the expectation of the difference of equations (3.6) and (3.8) is defined as the Fisher information matrix $I(\beta,\lambda)$. The MLEs $(\hat{\beta},\hat{\lambda}\ )$ with some mild regularity conditions follows the approximately bivariate normal distribution with mean (β,λ) and covariance matrix $[I(\beta,\lambda)]^{-1}$. Usually, in practice, the estimate of $[I(\beta,\lambda)]^{-1}$ is used by $\left[I_0\left(\hat{\beta},\hat{\lambda}\right)\right]^{-1}$. A simpler and equally valid procedure is to use the approximation

$$(\hat{\beta},\hat{\lambda}) \sim N\left((\beta,\lambda), \left[I_0(\hat{\beta},\hat{\lambda})\right]^{-1}\right), \tag{3.9}$$

where $I_0(\beta,\lambda)$ is the observed information matrix

$$\left[\begin{array}{cc} -\frac{\partial^2 \ell(\beta,\lambda|\underline{x})}{\partial\beta^2} & -\frac{\partial^2 \ell(\beta,\lambda|\underline{x})}{\partial\beta\partial\lambda} \\ -\frac{\partial^2 \ell(\beta,\lambda|\underline{x})}{\partial\lambda\partial\beta} & -\frac{\partial^2 \ell(\beta,\lambda|\underline{x})}{\partial\lambda^2} \end{array}\right]^{-1}_{(\hat{\beta},\hat{\lambda})}. \tag{3.10}$$

The approximate CIs for the parameters β and λ are obtained from the bivariate normal distribution with mean $(\beta,\ \lambda)$ and covariance matrix $\left[I_0(\hat{\beta},\hat{\lambda})\right]^{-1}$. Thus, the $100(1-2\alpha)\%$ ACIs for β and λ are

$$\hat{\beta} \mp z_\alpha\sqrt{v_{11}} \text{ and } \hat{\lambda} \mp z_\alpha\sqrt{v_{22}}, \tag{3.11}$$

respectively, where v_{11} and v_{22} are the elements on the diagonal of the covariance matrix $I_0^{-1}(\hat{\beta},\hat{\lambda})$ and z_α is the percentile of the standard normal distribution with the right-tail probability α.

4. BOOTSTRAP CONFIDENCE INTERVALS

In some cases, if the objective of the study is to determine the estimators, CIs, bias, and variance of an estimator or to calibrate hypothesis tests, then the bootstrap technique plays an important role. Different types of bootstrap techniques are available, such as those called parametric [15] and nonparametric [17]. In this section the parametric bootstrap technique is adopted to construct the percentile bootstrap CI (PBCI) (see [16] for more details) and the bootstrap-t CI (BTCI) (see [15]). The following algorithm is used to differentiate the two types of bootstrap techniques:

1. Based on the observed original progressively Type-II sample, $(t_{j1;m_j,n_j} < t_{j2;m_j,n_j} < ... < t_{jm_j;m_j,n_j})$, obtain $\hat{\beta}$, and $\hat{\lambda}$, $j = 1, 2$.
2. Based on the values of n_j and m_j $(1 < m_j < n_j)$ with the same values of R_{ji}, $(i = 1, 2, ..., m_j)$, $j = 1, 2$, generate two independent random samples of sizes m_1 and m_2 from the GHLD, $\underline{t}^* = (t^*_{j1;m_j,n_j} < t^*_{j2;m_j,n_j} < ... < t^*_{jm_j;m_j,n_j})$ using the algorithm described in [12].
3. As in step 1 based on t^* compute the bootstrap sample estimates of $\hat{\beta}$, and $\hat{\lambda}$ denoted here as $\hat{\beta}^*$ and $\hat{\lambda}^*$.
4. Steps 2 and 3 are repeated N times, thereby N different bootstrap samples are represented. The value of N may be taken as 1000.
5. The values of $\hat{\beta}^*$ and $\hat{\lambda}^*$ are arranged all in ascending order to obtain the bootstrap sample $(\hat{\theta}_l^{*[1]}, \hat{\theta}_l^{*[2]}, ..., \hat{\theta}_l^{*[N]})$, $l = 1, 2$ where $(\theta_1^* = \beta^*, \theta_2^* = \lambda^*)$.

Percentile bootstrap CIs

For given $H(y) = P(\hat{\theta}_k^* \leqslant y)$ the cumulative distribution function of $\hat{\theta}_k^*$. Define $\hat{\theta}^*_{kboot} = H^{-1}(y)$ for given y. The approximate bootstrap $100(1-2\alpha)\%$ CI of $\hat{\theta}_l^*$ is given by

$$\left[\hat{\theta}^*_{l\text{boot}}(\alpha), \hat{\theta}^*_{l\text{boot}}(1-\alpha)\right]. \tag{4.1}$$

Bootstrap-t CIs

First, we present the order statistics $\omega_k^{*[1]} < \omega_k^{*[2]} < ... < \omega_k^{*[N]}$,

$$\omega_k^{*[j]} = \frac{\hat{\theta}_l^{*[j]} - \hat{\theta}_l}{\sqrt{\text{var}\left(\hat{\theta}_l^{*[j]}\right)}},\ j = 1, 2, ..., N,\ l = 1, 2, \tag{4.2}$$

where $\hat{\theta}_1 = \hat{\beta}$, $\hat{\theta}_2 = \hat{\lambda}$.

For given $H(y) = P(\omega_l^* < y)$ the cumulative distribution function of ω_l^*, and given y, is defined as

$$\hat{\theta}_{l\text{boot}-t} = \hat{\theta}_l + \sqrt{\text{Var}(\hat{\theta}_l)} H^{-1}(y). \tag{4.3}$$

The approximate $100(1-2\alpha)\%$ CIs of $\hat{\theta}_k$ is given by

$$\left(\hat{\theta}_{l\text{boot}-t}(\alpha), \hat{\theta}_{l\text{boot}-t}(1-\alpha)\right). \tag{4.4}$$

5. SIMULATION STUDIES

We now adopted undertake simulation studies with the help of the Mathematica program Ver. 8.0 to illustrate the theoretical results of the estimation problem. The performance of the different point estimators of the shape parameter of the GHLD and the acceleration factor are measured and compared with the average of the estimates (AVG), absolute relative bias (RAB), and mean square error (MSE); specifically,

$$\text{AVG}(\hat{\theta}_l) = \frac{1}{M}\sum_{i=1}^{M}\hat{\theta}_l^{(i)},\ (\theta_1 = \beta, \theta_2 = \lambda), \tag{5.1}$$

$$\text{RAB}(\hat{\theta}_l) = \frac{|\overline{\hat{\theta}_l} - \theta_l|}{\theta_l}, \tag{5.2}$$

and

$$\text{MSE}(\hat{\theta}_l) = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{\theta}_l^{(i)} - \theta_l\right)^2. \tag{5.3}$$

For each of the CIs, the ACIs and the different bootstrap CIs can be measured and compared using the average confidence lengths (AC) as well as the coverage percentages (CP). For the generated sample, we computed the 90% CIs, recorded AC, and checked whether the true value lay within the interval (CP). In simulation studies, this step is repeated 1000 times. The estimated CP was computed as the number of CIs that covered the true values divided by 1000 whereas the estimated expected width of the CI was computed as the sum of the lengths for all intervals divided by 1000. Now, we present the definitions of the different CSs that are used in our simulation studies:

CS I: $R_{ji} = 0$ for $i < m$ and $R_{jm} = n - m$,

CS II: $R_{ji} = 0$ for $i > 1$ and $R_{j1} = n - m$,

CS III: for odd m, $R_{ji} = 0$ for $i > \frac{m+1}{2}$ or $i < \frac{m+1}{2}$ and $R_{j\frac{m+1}{2}} = n - m$.

Also, for even m, $R_{ji} = 0$ for $i > \frac{m}{2}$ or $i < \frac{m}{2}$ and $R_{j\frac{m}{2}} = n - m$:

CS IV: $R_{j\frac{2m-n}{2}+1} = \ldots = R_{j\frac{n}{2}} = 1$, other $R_{ji} = 0$.

In our simulation studies, we consider two separate cases:

(**1**) The model parameter values ($\beta = 0.5, \lambda = 2.0$), the sample sizes ($n_1 = n_2 = \mathbf{n}$) and observed failure times ($m_1 = m_2 = \mathbf{m}$); results are listed in Tables 1 and 2.

(**2**) The model parameter values ($\beta = 2.5, \lambda = 1.5$), the sample sizes ($n_2 = 2n_1 = 2\mathbf{n}$) and observed failure times ($m_2 = 2m_1 = 2\mathbf{m}$); results are listed in Tables 3 and 4.

Table 1: AVG and RABs (MSEs) of ML and Bootstrap estimates for the parameters ($\beta = 0.5$ and $\lambda = 2.0$).

$(\mathbf{n}, \mathbf{m})$	CS	MLE		Bootstrap	
		β	λ	β	λ
(30,15)	I	0.5370 0.0507 (0.147)	1.8242 0.098 (0.471)	0.5410 0.055 (0.248)	1.8109 0.145 (0.645)
	II	0.5300 0.0481 (0.126)	1.8950 0.079 (0.410)	0.5312 0.049 (0.210)	1.8229 0.140 (0.584)
	III	0.5361 0.0497 (0.133)	1.8720 0.090 (0.425)	0.5347 0.053 (0.229)	1.8198 0.142 (0.609)
	IV	0.5457 0.0487 (0.123)	1.8889 0.087 (0.419)	0.5317 0.049 (0.219)	1.8301 0.142 (0.601)
(30,25)	I	0.5204 0.0413 (0.101)	1.889 0.052 (0.394)	0.5229 0.043 (0.131)	1.8740 0.085 (0.451)
	II	0.5154 0.039 (0.099)	1.9241 0.048 (0.289)	0.5201 0.040 (0.120)	1.8654 0.074 (0.325)
	III	0.5224 0.042 (0.102)	1.9094 0.049 (0.317)	0.5244 0.041 (0.135)	1.8741 0.081 (0.377)
	IV	0.5208 0.041 (0.100)	1.9107 0.050 (0.314)	0.5232 0.040 (0.124)	1.8841 0.080 (0.364)
(50,25)	I	0.5215 0.041 (0.098)	1.920 0.050 (0.378)	0.5240 0.045 (0.128)	1.9014 0.083 (0.440)
	II	0.5109 0.031 (0.081)	1.951 0.045 (0.326)	0.5217 0.041 (0.119)	1.9241 0.079 (0.420)
	III	0.5122 0.035 (0.093)	1.936 0.044 (0.331)	0.5217 0.040 (0.131)	1.9288 0.074 (0.426)
	IV	0.5220 0.034 (0.090)	1.944 0.043 (0.338)	0.5200 0.039 (0.130)	1.9233 0.071 (0.415)
(50,40)	I	0.5100 0.022 (0.052)	1.9821 0.033 (0.208)	0.5107 0.031 (0.101)	1.9621 0.036 (0.401)
	II	0.5102 0.020 (0.040)	1.9800 0.022 (0.109)	0.5099 0.022 (0.081)	1.9751 0.027 (0.265)
	III	0.5133 0.022 (0.042)	1.9741 0.025 (0.119)	0.5118 0.024 (0.094)	1.9751 0.029 (0.377)
	IV	0.5201 0.023 (0.041)	1.9788 0.024 (0.112)	0.5122 0.021 (0.090)	1.9788 0.030 (0.372)

Table 2: The (AC) and (CP) of 90% CIs (β, λ) at (0.5, 2.0).

$(\mathbf{n}, \mathbf{m})$	CS	MLE		Boot-P		Boot-t	
		β	λ	β	λ	β	λ
(30,15)	I	2.1214 (0.88)	3.2145 (0.87)	3.1354 (0.87)	5.2336 (0.86)	2.1019 (0.89)	3.2100 (0.88)
	II	2.1110 (0.88)	3.1177 (0.89)	3.1123 (0.93)	5.2210 (0.88)	2.1007 (0.89)	3.2006 (0.91)
	III	2.1133 (0.87)	3.1224 (0.88)	3.1209 (0.92)	5.2319 (0.88)	2.1016 (0.89)	3.2055 (0.90)
	IV	2.1125 (0.88)	3.1233 (0.88)	3.1212 (0.88)	5.2400 (0.87)	2.1109 (0.89)	3.2107 (0.91)
(30,25)	I	2.1009 (0.89)	3.2010 (0.88)	3.1210 (0.87)	5.2221 (0.88)	2.1000 (0.89)	3.2009 (0.90)
	II	2.0789 (0.92)	3.0166 (0.91)	3.1000 (0.92)	4.6215 (0.93)	1.9524 (0.91)	3.1612 (0.89)
	III	2.1087 (0.89)	3.0198 (0.89)	3.1017 (0.89)	5.1017 (0.92)	2.0041 (0.90)	3.2008 (0.89)
	IV	2.1108 (0.91)	3.1010 (0.90)	3.1205 (0.92)	5.1003 (0.89)	2.0000 (0.90)	3.2107 (0.919)
(50,25)	I	2.1023 (0.89)	3.1077 (0.88)	3.1187 (0.89)	5.2119 (0.88)	2.0139 (0.88)	3.1748 (0.90)
	II	2.0742 (0.93)	3.0142 (0.89)	2.9811 (0.92)	4.7217 (0.88)	1.9541 (0.90)	3.1752 (0.92)
	III	2.1102 (0.88)	3.1100 (0.89)	3.1107 (0.91)	5.1009 (0.89)	2.0051 (0.91)	3.2012 (0.89)
	IV	2.1111 (0.88)	3.1009 (0.91)	3.1217 (0.91)	5.1014 (0.89)	2.0021 (0.90)	3.2112 (0.89)
(50,40)	I	1.9821 (0.89)	3.0087 (0.89)	3.0584 (0.92)	5.0472 (0.88)	1.7742 (0.89)	3.1010 (0.910)
	II	1.7490 (0.88)	2.9874 (0.89)	2.6511 (0.89)	4.1145 (0.93)	1.7120 (0.89)	3.0770 (0.91)
	III	1.8890 (0.89)	3.1120 (0.89)	2.6742 (0.89)	4.1246 (0.92)	1.7331 (0.90)	3.1070 (0.90)
	IV	1.8741 (0.91)	3.10820 (0.92)	2.6662 (0.89)	4.1195 (0.92)	1.7320 (0.91)	3.1040 (0.89)

Table 3: AVG and RABs (MSEs) of ML and Bootstrap estimates for the parameters ($\beta = 2.5$ and $\lambda = 1.5$).

$(\mathbf{n}, \mathbf{m})$	CS	MLE		Bootstrap	
		β	λ	β	λ
(20,10)	I	2.5390 0.120 (0.521)	1.4522 0.109 (0.471)	2.5561 0.1324 (0.641)	1.4522 0.111 (0.499)
	II	2.5211 0.115 (0.446)	1.4745 0.087 (0.406)	2.5423 0.125 (0.549)	1.4642 0.099 (0.408)
	III	2.5341 0.120 (0.498)	1.4624 0.109 (0.450)	2.5450 0.129 (0.587)	1.4602 0.101 (0.470)
	IV	2.5327 0.118 (0.487)	1.4631 0.105 (0.450)	2.5462 0.131 (0.591)	1.4611 0.105 (0.465)
(20,15)	I	2.5220 0.101 (0.521)	1.4842 0.087 (0.328)	2.5325 0.101 (0.554)	1.4740 0.084 (0.332)
	II	2.5201 0.087 (0.421)	1.4892 0.060 (0.301)	2.5288 0.099 (0.511)	1.4884 0.050 (0.311)
	III	2.5213 0.099 (0.460)	1.4811 0.080 (0.317)	2.5485 0.110 (0.522)	1.4811 0.070 (0.328)
	IV	2.5217 0.097 (0.455)	1.4804 0.082 (0.322)	2.5477 0.108 (0.518)	1.4814 0.069 (0.331)
(30,20)	I	2.5198 0.100 (0.515)	1.4811 0.086 (0.312)	2.5311 0.099 (0.44)	1.4720 0.081 (0.311)
	II	2.5190 0.060 (0.400)	1.4893 0.055 (0.280)	2.5288 0.070 (0.500)	1.4870 0.046 (0.287)
	III	2.5196 0.090 (0.454)	1.4814 0.076 (0.312)	2.5462 0.101 (0.511)	1.4900 0.065 (0.314)
	IV	2.5211 0.097 (0.455)	1.4774 0.079 (0.318)	2.5477 0.106 (0.519)	1.4855 0.062 (0.325)
(30,25)	I	2.5101 0.089 (0.256)	1.4954 0.050 (0.214)	2.5210 0.060 (0.265)	1.4894 0.042 (0.266)
	II	2.5121 0.051 (0.202)	1.4998 0.020 (0.148)	2.5109 0.052 (0.215)	1.4899 0.040 (0.200)
	III	2.5111 0.060 (0.215)	1.4974 0.023 (0.201)	2.5109 0.069 (0.261)	1.4864 0.045 (0.212)
	IV	2.5113 0.059 (0.212)	1.4982 0.021 (0.212)	2.5110 0.067 (0.242)	1.4870 0.046 (0.209)

Table 4: The (AC) and (CP) of 90% CIs (β, λ) at (2.5, 1.5).

$(\mathbf{n}, \mathbf{m})$	CS	MLE		Boot-t		Boot-P	
		β	λ	β	λ	β	λ
(20,10)	I	4.1147 (0.87)	3.1231 (0.88)	5.2414 (0.85)	3.5421 (0.86)	4.1009 (0.89)	3.1037 (0.89)
	II	3.9544 (0.88)	3.0032 (0.88)	3.9881 (0.87)	3.2131 (0.88)	3.7542 (0.89)	3.0011 (0.91)
	III	3.9654 (0.88)	3.0172 (0.89)	3.9991 (0.92)	3.2321 (0.88)	3.8045 (0.90)	3.0712 (0.92)
	IV	3.9622 (0.88)	3.0161 (0.89)	3.9970 (0.93)	3.2300 (0.92)	3.8039 (0.91)	3.0702 (0.91)
(20,15)	I	3.7541 (0.91)	3.1001 (0.89)	3.7865 (0.88)	3.1124 (0.89)	3.7111 (0.89)	3.0099 (0.91)
	II	3.1542 (0.89)	2.8570 (0.88)	3.7742 (0.89)	2.899 (0.89)	3.1421 (0.90)	2.8110 (0.91)
	III	3.1588 (0.88)	2.8598 (0.89)	3.7760 (0.91)	2.9200 (0.88)	3.1441 (0.91)	2.8132 (0.91)
	IV	3.1570 (0.89)	2.8592 (0.90)	3.7755 (0.92)	2.9136 (0.91)	3.1432 (0.92)	2.8127 (0.90)
(30,20)	I	3.7531 (0.92)	3.0991 (0.89)	3.7854 (0.89)	3.1118 (0.89)	3.7101 (0.90)	3.0088 (0.92)
	II	3.1522 (0.90)	2.8550 (0.88)	3.7720 (0.91)	2.8965 (0.89)	3.1400 (0.91)	2.8094 (0.91)
	III	3.1573 (0.89)	2.8585 (0.89)	3.7750 (0.88)	2.9199 (0.88)	3.1432 (0.92)	2.8124 (0.91)
	IV	3.1555 (0.89)	2.8580 (0.88)	3.7742 (0.92)	2.9127 (0.93)	3.1421 (0.91)	2.8118 (0.91)
(30,25)	I	3.7014 (0.91)	3.0665 (0.89)	3.7116 (0.89)	3.0772 (0.92)	3.6542 (0.91)	3.0545 (0.92)
	II	3.5124 (0.901)	3.0256 (0.89)	3.5198 (0.91)	3.0281 (0.89)	3.5111 (0.90)	3.0231 (0.91)
	III	3.5321 (0.88)	3.0290 (0.89)	3.5221 (0.89)	3.0321 (0.88)	3.5185 (0.91)	3.0287 (0.91)
	IV	3.5314 (0.89)	3.0282 (0.898)	3.5214 (0.90)	3.0307 (0.92)	3.5172 (0.901)	3.0281 (0.92)

6. NUMERICAL EXAMPLE

For demonstration purposes, the estimation procedure described in the previous section is applied to the set of simulated progressive Type-II censoring data under the constant-stress partially ALT. The MLEs and the two bootstrap CIs are computed for model parameters β and λ with the real parameters are equal to 1.5 and 2.0, respectively. In this example, we simulate samples of size ($m_1 = m_2 = 15$ of $n_1 = n_2 = 30$) from the GHLD under the two progressive CSs $R_1 = R_2 = \{1, 0, 0, 0, 2, 0, 0, 2, 0, 0, 1, 0, 1, 0, 0, 0, 0, 2, 1, 0\}$ using the algorithm described in Balakrishnan and Sandhu [12]. The simulated data are presented in Table 5.

Table 5: Simulated progressively censored samples with constant PALTs.

Normal conditions	0.13901	0.22961	0.26912	0.47032	0.51005	0.52645	0.53583
	0.56987	0.65999	0.79289	0.80636	0.89349	1.56115	1.63822
	1.66079						
Accelerated conditions	0.00274	0.02767	0.06181	0.06717	0.12004	0.14341	0.25042
	0.27614	0.31457	0.42484	0.54109	0.54112	0.75652	1.13610
	1.41038						

In Figure 1, the two probability density functions show the effect of an acceleration factor.

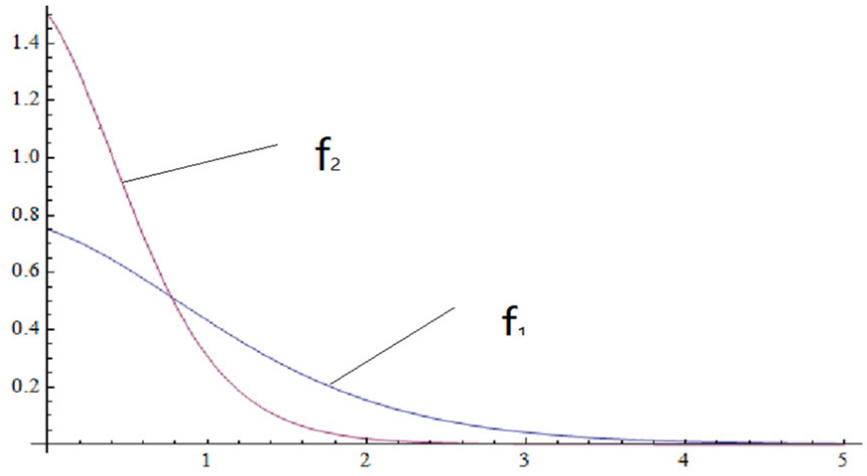

Figure 1: Probability density under normal and accelerated condition.

The iteration procedure of the MLE needs the initial value of parameter obtained from the profile log-likelihood function (Figure 2) as 1.8. The point estimates and related RABs and MSEs of the parameters as well as the 90% and 95% ACIs are listed in Table 6. Also, the point estimates and the relate RABs and MSEs of the parameters as well as the 90% and 95% PBCIs and BTCIs are presented in Table 7. We observed that the BTCIs and approximate MLE intervals are narrower than the PBCIs and always include the population parameter values.

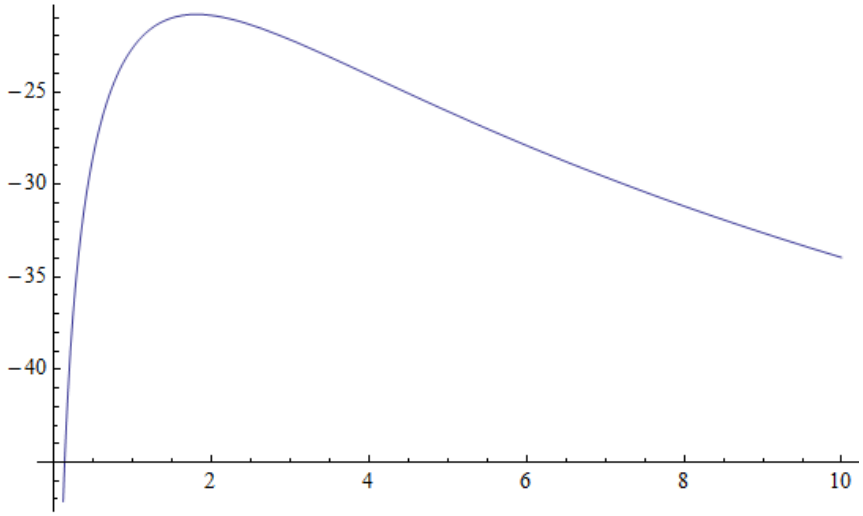

Figure 2: Profile log likelihood function of λ.

Table 6: MLEs, MSEs, RABs and (90%-95%) approximate confidence intervals.

$(.)_{\mathrm{ML}}$	RAB	MSE	90%	95%
1.5495	0.0330	0.0495	(0.7769, 2.3221)	(0.9011, 2.1979)
1.8034	0.0983	0.1966	(0.7231, 2.8837)	(0.8968, 2.7100)

Table 7: Percentile bootstrap CIs and Bootstrap-t CIs based on 500 replications.

$(.)_{\mathrm{Boot}}$	RAB	MSE	90%		95%	
			BPCI	BTCI	BPCI	BTCI
1.7421	0.1614	0.2421	(0.3241, 3.1205)	(0.7981, 2.2954)	(0.6581, 2.6325)	(0.8881, 2.1472)
2.3415	0.1707	0.3415	(0.5213, 3.2140)	(0.7751, 2.7098)	(0.4578, 2.6590)	(0.7922, 2.5213)

7. CONCLUDING REMARKS

In product-life testing experiments, reducing the time and cost, especially for units with high reliability, illustrates the importance of ALTs. Different types of ALTs are available, one of the types most suitable for different situations is the constant-stress partially ALTs. Also, the experimenter in some situations is unable to obtain complete information of failure times for all experimental units or is in need to remove some units other than the final point of the experiment. The conventional Type-I and Type-II CSs do not have the flexibility of allowing to remove any units at points other than the final point of the experiment.

Hence, in this paper, we adopted a more general CS with the constant-stress partially ALT, known as progressive Type-II censoring. Simulation studies were presented to assess and compare the performance of the proposed methods. From the results, we observed the following:

1. For fixed values of sample size **n** and with increasing effected sample size **m**, the MSEs and RABs of the considered parameters decrease.
2. For fixed values of the sample and failure time sizes, CS II, in which the censoring occurs after the first observed failure, gives more accurate results through the MSEs and RABs than the other schemes..
3. Results for the CS III and CS IV are more similar.
4. The bootstrap-t credible intervals give more accurate results than the ACIs than the bootstrap CIs because the lengths of the former are less than the lengths of the latter, for different sample sizes, observed failures, and schemes.
5. For fixed sample sizes and observed failures, CS II moreover gives lower lengths for the three methods to obtain the CIs compared with the other three schemes.

ACKNOWLEDGMENTS

This paper was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under Grant No. **G-611-130-36**. The author, therefore, acknowledge with thanks DSR for technical and financial support.

REFERENCES

[1] Abd-Elmougod, G.A. and Emad, E.M. (2016). Parameters estimation of compound Rayleigh distribution under an adaptive type-II progressively hybrid censored data for constant partially accelerated life tests, *Global Journal of Pure and Applied Mathematics*, **12**, 3253–3273.

[2] Abushala, T.A. and Soliman, A.A. (2013). Estimating the Pareto parameters under progressive censoring data for constant-partially accelerated life tests, *J. Statist. Comput. Simul.*, **85**, 917–934.

[3] Al-Hussaini, E.K. and Abdel-Hamid, A.H. (2004). Bayesian estimation of the parameters, reliability and hazard rate functions of mixtures under accelerated life tests, *Communication in Statistics – Simulation and Computation*, **33**, 963–982.

[4] Al-Hussaini, E.K. and Abdel-Hamid, A.H. (2006). Accelerated life tests under finite mixture models, *J. Statist. Comput. Simul.*, **76**, 673–690.

[5] Arora, S.H.; Bhimani, G.C. and Patel, M.N. (2010). Some results on maximum likelihood estimators of parameters of generalized half logistic distribution under Type-I progressive censoring with changing, *International Journal of Contemporary Mathematical Sciences*, **5**, 685–698.

[6] Awodutire, P.O.; Olapade, A.K. and Kolawole, O.A. (2016). The Type I generalized half logistic survival model, *International Journal of Theoretical and Applied Mathematics*, **2**, 74–78.

[7] Bai, D.S. and Chung, S.W. (1992). Optimal design of partially accelerated life tests for the exponential distribution under Type-I censoring, *IEEE Trans. Reliab.*, **41**, 400–406.

[8] Balakrishnan, N. (1985). Order statistics from the half logistic distribution, *J. Statist. Comput. Simul.*, **20**, 287–309.

[9] Balakrishnan, N. and Aggarwala, R. (2000). *Progressive Censoring – Theory, Methods, and Applications*, Birkhäuser, Boston.

[10] Balakrishnan, N. and Hossain, A. (2007). Inference for the Type-II generalized logistic distribution under progressive Type-II censoring, *J. Statist. Comput. Simul.*, **77**, 1013–1031.

[11] Balakrishnan, N. and Puthenpura, S. (1986). Best linear estimators of location and scale parameters of the half logistic distribution, *J. Statist. Comput. Simul.*, **25**, 193–204.

[12] Balakrishnan, N. and Sandhu, R.A. (1995). A simple simulation algorithm for generating progressively type-II censored samples, *The American Statistician*, **49**, 229–230.

[13] Balakrishnan, N. and Wong, K.H.T. (1991). Approximate MLEs for the location and scale parameters of the halflogistic distribution with type-II right censoring, *IEEE Trans. Reliab.*, **40**, 140–145.

[14] Chaturvedi, A.; Suk-Bok Kang, S.B. and Pathak, A. (2016). Estimation and testing procedures for the reliability functions of generalized half logistic distribution, *Journal of the Korean Statistical Society*, **45**, 314–328.

[15] Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Applications*, 2nd ed., Cambridge University Press, Cambridge, United Kingdom.

[16] Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans.* In "CBMS-NSF Regional Conference Series in Applied Mathematics", SIAM, Philadelphia, PA, 38.

[17] Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, New York.

[18] Kim, C.M. and Ba, D.S. (2002). Analysis of accelerated life test data under two failure modes, *Int. J. Reliability, Quality and Safety Engineering*, **9**, 111–125.

[19] Kim, C. and Han, K. (2010). Estimation of the scale parameter of the halflogistic distribution under progressively type II censored sample, *Stat. Papers*, **51**, 375–387.

[20] Kim, Y.; Kang, S.B. and Seo, J.I. (2011). Bayesian estimation in the generalized half logistic distribution under progressively Type II censoring, *Journal of the Korean Data and Information Science Society*, **22**, 977–987.

[21] Miller, R. and Nelson, W.B. (1983). Optimum simple step-stress plans for accelerated life testing, *IEEE Trans. Reliab.*, **32**, 59–65.

[22] Nelson, W. (1990). *Accelerated Testing: Statistical Models, Test Plans and Data Analysis*, Wiley, New York.

[23] Ramakrishnan, V. (2008). *Generalizations to half logistic distribution and related inference* (Ph.D. thesis), India: Acharya Nagarjuna University (AP).

REVSTAT – Statistical Journal
Volume 18, Number 4, October 2020, 453–460

AVERAGES FOR MULTIVARIATE RANDOM VECTORS WITH RANDOM WEIGHTS: DISTRIBUTIONAL CHARACTERIZATION AND APPLICATION

Authors: A.R. Soltani
– Department of Statistics and Operations Research, Faculty of Science, Kuwait University, Kuwait
Department of Statistics, Shiraz University, Shiraz, Iran
ar.soltani@ku.edu.kw

Rasool Roozegar
– Department of Statistics, Yazd University, Yazd, Iran
rroozegar@yazd.ac.ir

Received: January 2018 Revised: March 2018 Accepted: June 2018

Abstract:

- We consider a random weights average of n independent continuous random vectors $\mathbf{X}_1, ..., \mathbf{X}_n$, where random weights are cuts of $[0, 1]$ by an increasing sequence of the order statistics of a random sample from a uniform $[0, 1]$. We employ the multivariate Stieltjes transform and Watson [15] celebrated formula involving the multivariate B-spline functions for distributional identification of multivariate random weights averages. We show that certain classes of Dirichlet and random scale stable random vectors are random weightes averages.

Keywords:

- *multivariate weighted average with random weights; multivariate Cauchy Stieltjes transform; Dirichlet distribution; multivariate stable distributions.*

AMS Subject Classification:

- 62H05, 46F12, 65R10.

1. INTRODUCTION

An average $UX_1+(1-U)X_2$ of two independent continuous random variables X_1 and X_2, $U \sim$ uniform$(0,1)$, is the subject of Johnson and Kotz [5] expository article on the work of Van Assche [14]. Indeed, Johnson and Kotz noticed that the random variable *uniformly distributed between two random variables*, named by Van Assche, is a random weightes average, RWA in short. Soltani and Roozegar [13] consider RWA of a finite number of independent continuous random variables, where the weights are cuts of $(0,1)$ by an increasing selection of the order statistics of a uniform $(0,1)$ random sample. In his work, Van Assche [14] noticed that the Stieltjes transform is an appropriate tool for the distributional identification of random weightes averages, as the Fourier transform is for averages with non-random weights.

In this paper, we consider RWA of a number of independent continuous random vectors with values in $\mathbb{R}^d$, the d dimensional Euclidian space. The random weights are as in Soltani and Roozegar [13]: the cuts of $(0,1)$ by an increasing sequence $U_{(k_1)}, U_{(k_2)}, ..., U_{(k_{m-1})}$ of the order statistics $U_{(1)}, ..., U_{(n-1)}$ from a uniform $(0,1)$ sample $U_1, ..., U_{n-1}$; $1 \leq k_1 < k_2 < ... < k_{m-1} < k_m = n$, $U_{(n)} = 1$. We employ the multivariate Stieltjes, also called Cauchy-Stieltjes, transform (MCST in short) for the distributional identification of multivariate randomly weighted averages, MRWA. In this article, we prove that the MCST of order n, $\mathcal{S}[\mathbf{F}; n](\mathbf{z})$, of the distribution $\mathbf{F}$, the distribution of random weights averages of independent d-dimensional continuous random vectors $\mathbf{X}_1 \sim \mathbf{F}_1, ..., \mathbf{X}_m \sim \mathbf{F}_m$, is equal to the product of the corresponding MCST of $\mathbf{F}_1, ..., \mathbf{F}_m$, namely,

$$\mathcal{S}[\mathbf{F}; n](\mathbf{z}) = \mathcal{S}[\mathbf{F_1}; k_1](\mathbf{z})\mathcal{S}[\mathbf{F_2}; k_2 - k_1](\mathbf{z})...\mathcal{S}[\mathbf{F_m}; n - k_{m-1}](\mathbf{z}), \quad z \in \mathbb{C}^d. \tag{1.1}$$

Our approach is somewhat new and different from those applied in the references cited above. Van Assche [14] applies certain techniques from the differentiation of Schwartz distributions. Soltani and Roozegar [13] apply the divided differences and the theory of knots. In this article, we apply the pioneering formula of Watson [15] involving B-splines, discussed in Karlin, Micchelli and Rinott [6]. This approach is more direct and easily applied. It can be applied to the univariate RWA as well.

The notion of random weights averages in the literature may be attributed to the interesting observation of Galton, the founder of regression. He observed that, on average, a child's height is more mediocre (average) than his or her parent's height. Plausibly, the child's height is a RWA of his or her parents' heights. In contrast to the univariate RWA, multivariate RWA can be used for modeling when a finite number of characteristics are considered simultaneously.

Univariate and multivariate RWA have appeared in certain areas, such as sampling, density estimation, Bayesian and distributional characterizations, among others. In theory, general regression and neural networks, multivariate kernel density estimations and multivariate kernel regressions are all randomly weighted averages, see Nadaraya [8] and Watson [16]. An interesting example of averages of multivariate quantities with random weights is the random vector of the serial correlation coefficients, introduced by Watson [15], $\mathbf{r} = (r_1, r_2, ..., r_k)$, where

$$r_j = \frac{\lambda_0^{(j)} W_0 + \lambda_1^{(j)} W_1 + \cdots + \lambda_m^{(j)} W_m}{W_0 + \cdots + W_m}, \quad j = 1, 2, ..., k, \tag{1.2}$$

and $W_0, W_1, ..., W_m$ are independent gamma variables of integer order $\alpha_0, \alpha_1, ..., \alpha_m$, and $\boldsymbol{\lambda}_\ell = (\lambda_\ell^{(1)}, ..., \lambda_\ell^{(k)})$, $\ell = 0, ..., m$ are k-dimensional knots. In addition, Zeng [17] characterizes the multivariate stable distributions through the independence of the linear statistic $\mathbf{U} = \sum_{i=1}^n Y_i \mathbf{X}_i$ and the vector of random coefficients $\mathbf{Y} = (Y_1, ..., Y_n)^T \in \mathbb{R}^n$, where $\mathbf{X}_1, ..., \mathbf{X}_n$ are independent and identically distributed random vectors in $\mathbb{R}^k$, and are independent of $\mathbf{Y}$. A special form of $\mathbf{U}$ is a RWA of random vectors.

This article is organized as follows. We give preliminaries and the proof of (1.1) in Section 2. We proceed on to introduce and study interesting classes of distributions that are RWA of continuous random vectors. In particular, we prove that the RWA of independent Dirichlet random vectors is Dirichlet, and that the RWA of independent and identically symmetric stable random vectors is randomly scaled stable. We devote Section 3 to this issue.

2. PRELIMINARIES AND MAIN RESULT

Let us denote the RWA of m independent and continuous random vectors $\mathbf{X}_1, ..., \mathbf{X}_m$ in $\mathbb{R}^d$ by

$$\mathbf{S}_{m:n} = R_{1:n}\mathbf{X}_1 + R_{2:n}\mathbf{X}_2 + \cdots + R_{m:n}\mathbf{X}_m, \qquad m \geq 2, \tag{2.1}$$

where the random weights $R_{j:n}$ are assumed to be the m cuts of $[0, 1]$ by an increasing ordered array $U_{(k_1)}, ...U_{(k_{m-1})}$ of $U_{(1)}, ..., U_{(n-1)}$, the ordered statistics of $n-1$ independent and identically uniformly distributed random variables $U_1, ..., U_{n-1}$ on $[0, 1]$;

$$R_{j:n} = U_{(k_j)} - U_{(k_{j-1})},\ j = 1, 2, \cdots, m, \quad m \leq n,$$

where $k_0 = 0 < k_1 < ... < k_{m-1} < k_m = n$ are in $\{1, ..., n\}$ and $U_{(n)} = 1$.

The conditional density of $\mathbf{S}_{m:n}$ given $\mathbf{X}_1 = \mathbf{x}_1$, ..., $\mathbf{X}_m = \mathbf{x}_m$ is denoted by $M(\mathbf{t}|\mathbf{x}_1, ..., \mathbf{x}_m)$, $\mathbf{t} \in \mathbb{R}^d$. In the numerical analysis context, this density function is called "the Multivariate B-spline with knots $\{\mathbf{x}_1, ..., \mathbf{x}_m\}$", Karlin, Micchelli and Rinott [6]. The random vectors $\mathbf{X}_1, ..., \mathbf{X}_m \in \mathbb{R}^d$ have a convex hull with positive volume in $\mathbb{R}^d$. Our derivations very much rely on the fundamental result by Watson [15]:

$$\int_{\mathbb{R}^d} M(\mathbf{t}|\mathbf{x}_1, ..., \mathbf{x}_m) \frac{d\mathbf{t}}{(1 - < \mathbf{t}, \mathbf{x} >)^{\sum_{i=1}^m r_i}} = \prod_{i=1}^m (1 - < \mathbf{x}, \mathbf{x}_i >)^{-r_i}, \tag{2.2}$$

for $\max_i | < \mathbf{x}, \mathbf{x}_i > | < 1$, Karlin, Micchelli and Rinott [6].

The multivariate Cauchy-Stieltjes (or Stieltjes) transform (MCST) of a distribution $\mathbf{H}$ is defined by

$$\mathcal{S}[\mathbf{H}](\mathbf{z}) = \int_{\mathbb{R}^d} \frac{1}{1 - < \mathbf{z}, \mathbf{x} >} \mathbf{H}(d\mathbf{x}), \qquad \mathbf{z} \in \mathbb{C}^d \cap (\operatorname{supp} \mathbf{H})^c, \tag{2.3}$$

for $| < \mathbf{z}, \mathbf{x} > | < 1$, $< \mathbf{a}, \mathbf{b} >= \sum_{i=1}^k a_i b_i$, $\mathbb{C}$ is the set of complex numbers and $\operatorname{supp} \mathbf{H}$ stands for the support of $\mathbf{H}$, Kerov and Tsilevich [7] and Cuyt, Golub, Milanfar and Verdonk [1]. Similarly the MCST of order n of a distribution $\mathbf{H}$ is defined by

$$\mathcal{S}[\mathbf{H}; n](\mathbf{z}) = \int_{\mathbb{R}^d} \frac{1}{[1 - < \mathbf{z}, \mathbf{t} >]^n} \mathbf{H}(d\mathbf{t}), \qquad \mathbf{z} \in \mathbb{C}^d \cap (\operatorname{supp} \mathbf{H})^c, \tag{2.4}$$

for $| < \mathbf{z}, \mathbf{t} > | < 1$.

For $d = 1$, the MST is also called Markov transform, denoted by $\mathcal{M}_1[H](z)$. There is a relation between Markov transform and Stieltjes transform of a distribution H:

$$\mathcal{M}_1[H](z) = \frac{1}{z}\,\mathcal{S}[H](\frac{1}{z}),$$

where

$$\mathcal{S}[H](z) = \int_{\mathbb{R}} \frac{1}{z-x}\, H(dx),$$

for z in the set of complex numbers $\mathbb{C}$ which does not belong to the support of H, $z \in \mathbb{C} \cap (\operatorname{supp} H)^c$. For more on the Stieltjes transform see Debnath and Bhatta [2].

The following theorem is our main result in this section.

Theorem 2.1. *Let $\mathbf{X}_1, ..., \mathbf{X}_m$ be $m > 1$ independent and continuous random vectors in $\mathbb{R}^d$. Let $\mathbf{S}_{m:n}$ be the corresponding MRWA given by (2.1). Then*

$$\mathcal{S}[F_{\mathbf{S}_{m:n}}; n](\mathbf{z}) = \prod_{i=1}^{m} \mathcal{S}[F_i; r_i](\mathbf{z}), \qquad \mathbf{z} \in \mathbb{C}^d \bigcap_{i=1}^{m} (\operatorname{supp} F_i)^c, \tag{2.5}$$

where $r_i = k_i - k_{i-1}$; $\sum_{i=1}^{m} r_i = n$.

Proof: We note that

$$\begin{aligned}
F_{\mathbf{S}_{m:n}}(\mathbf{t}) &= E\left(I[\mathbf{S}_{m:n} \le \mathbf{t}]\right) \\
&= E\left(E\left(I[\mathbf{S}_{m:n} \le \mathbf{t}]|\mathbf{X}_1, ..., \mathbf{X}_m\right)\right) \\
&= \int_{\mathbb{R}^{md}} E\left(I[\mathbf{S}_{m:n} \le \mathbf{t}]|\mathbf{x}_1, ..., \mathbf{x}_m\right) \prod_{i=1}^{m} F_i(d\mathbf{x}_i) \\
&= \int_{\mathbb{R}^{md}} \int_{[\mathbf{s}<\mathbf{t}]} M(\mathbf{s}|\mathbf{x}_1, \cdots, \mathbf{x}_m) d\mathbf{s} \prod_{i=1}^{m} F_i(d\mathbf{x}_i) \\
&= \int_{[\mathbf{s}<\mathbf{t}]} \left\{ \int_{\mathbb{R}^{md}} M(\mathbf{s}|\mathbf{x}_1, \cdots, \mathbf{x}_m) \prod_{i=1}^{m} F_i(d\mathbf{x}_i) \right\} d\mathbf{s},
\end{aligned}$$

giving that

$$dF_{\mathbf{S}_{m:n}}(\mathbf{t}) = \int_{\mathbb{R}^{md}} M(\mathbf{t}|\mathbf{x}_1, \cdots, \mathbf{x}_m) \prod_{i=1}^{m} F_i(d\mathbf{x}_i).$$

Therefore,

$$\begin{aligned}
\mathcal{S}[F_{\mathbf{S}_{m:n}}; n](\mathbf{z}) &= \int_{\mathbb{R}^d} \frac{1}{[1- < \mathbf{z}, \mathbf{t} >]^n} dF_{\mathbf{S}_{m:n}}(\mathbf{t}) \\
&= \int_{\mathbb{R}^{md}} \left\{ \int_{\mathbb{R}^d} \frac{1}{[1- < \mathbf{z}, t >]^n} M(\mathbf{t}|\mathbf{x}_1, \cdots, \mathbf{x}_m) d\mathbf{t} \right\} \prod_{i=1}^{m} F_i(d\mathbf{x}_i) \\
&= \int_{\mathbb{R}^{md}} \prod_{i=1}^{m} [1- < \mathbf{z}, \mathbf{x_i} >]^{-r_i} \prod_{i=1}^{m} F_i(d\mathbf{x}_i) \qquad (2.6) \\
&= \prod_{i=1}^{m} \int_{\mathbb{R}^d} \frac{1}{[1- < \mathbf{z}, \mathbf{x}_i >]^{r_i}} F_i(d\mathbf{x}_i) \\
&= \prod_{i=1}^{m} \mathcal{S}[F_i; r_i](\mathbf{z}),
\end{aligned}$$

the third equality in (2.6) follows from (2.2). □

3. SOME CLASSES OF RWA OF RANDOM VECTORS

In this section we introduce two important classes of RWA of random vectors, Theorems 3.1 and 3.2.

In Theorem 3.1 below we assume $m = n$, $r_j = 1$ for $j = 1, 2, ..., m$.

Theorem 3.1. *Let $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ be independent random vectors such that $\mathbf{X}_i$ has a Dirichlet distribution with parameters $\alpha_i = (\alpha_{1i}, \alpha_{2i}, ..., \alpha_{di})'$, $\sum_{j=1}^{d} \alpha_{ji} = 1$, $i = 1, 2, ..., n$. Then the MRWA $S_{n:n}$ given by (2.1) has a Dirichlet distribution with parameters*

$$\sum_{i=1}^{n} \alpha_i = (\sum_{i=1}^{n} \alpha_{1i}, \sum_{i=1}^{n} \alpha_{2i}, ..., \sum_{i=1}^{n} \alpha_{di}).$$

Proof: The density and the Stieltjes transform of a Dirichlet distribution with parameters $\alpha = (\alpha_1, \alpha_2, ..., \alpha_d)'$ are given by

$$dF(\mathbf{x}) = f(x_1, x_2, ..., x_d) = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_d)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_d)} x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}, \quad \mathbf{x} \in \Delta^d, \tag{3.1}$$

and

$$\mathcal{S}[F](\mathbf{z}) = \int_{\Delta^d} \frac{F(d\mathbf{x})}{[1 - < \mathbf{z}, \mathbf{x} >]^{\sum_{i=1}^{d} \alpha_i}} = \prod_{j=1}^{d} \frac{1}{(1 - z_j)^{\alpha_j}}, \ \mathbf{z} = (z_1, ..., z_d)$$

respectively, $\Delta^d = \{(x_1, ..., x_d) \in \mathbb{R}^d, \ x_i > 0, \forall i, \sum_{i=1}^{d} x_i = 1\}$, Kerov and Tsilevich [7]. Let $\mathbf{X}_i \sim F_i$, $i = 1, 2, ..., n$. Then it follows from Theorem 2.1 that

$$\mathcal{S}[F_{\mathbf{S}_{n:n}}; n](\mathbf{z}) = \prod_{i=1}^{n} \prod_{j=1}^{d} \frac{1}{(1 - z_j)^{\alpha_{ji}}} = \prod_{j=1}^{d} \frac{1}{(1 - z_j)^{\sum_{i=1}^{n} \alpha_{ji}}}.$$

It is plain to show this function is the MCST, of order n, of a Dirichlet distribution with parameters $\sum_{i=1}^{n} \alpha_i = (\sum_{i=1}^{n} \alpha_{1i}, \sum_{i=1}^{n} \alpha_{2i}, ..., \sum_{i=1}^{n} \alpha_{di})$. Indeed for F', a Dirichlet distribution with parameters $(b_1, b_2, ..., b_d)$, with $\sum_{j=1}^{d} b_j = n$, we have

$$\mathcal{S}[F'; n](\mathbf{z}) = C(n; b_1, ..., b_d) \int_{\Delta^d} \frac{x_1^{b_1 - 1} x_2^{b_2 - 1} ... x_{d-1}^{b_{d-1} - 1} (1 - x_1 - ... - x_{d-1})^{b_d - 1}}{[1 - < \mathbf{z}, \mathbf{x} >]^n} dx_1 ... dx_{d-1},$$

where $C(n; b_1, ..., b_d) = \frac{\Gamma(n)}{\Gamma(b_1)\Gamma(b_2)...\Gamma(b_d)}$. Let $b_j = \sum_{i=1}^{n} \alpha_{ji}$, $j = 1, 2, ..., d$. Then the Euler type integral representation for the Lauricella function gives that

$$\begin{aligned} \frac{\Gamma(b_1)\Gamma(b_2)\cdots\Gamma(b_k)}{\Gamma(b_1 + b_2 + \cdots + b_k)} F_D^{(k)}(a, b_1, ..., b_k; b_1 + \cdots + b_k; z_1, ..., z_k) \ = \\ = \int ... \int_{\Delta^k} \frac{x_1^{b_1 - 1} x_2^{b_2 - 1} \cdots x_{k-1}^{b_{k-1} - 1} (1 - x_1 - \cdots - x_{k-1})^{b_k - 1}}{[1 - < \mathbf{z}, \mathbf{x} >]^a} dx_1 \cdots dx_{k-1}, \end{aligned} \tag{3.2}$$

where

$$F_D^{(k)}(a, b_1, ..., b_k; c; z_1, ..., z_k) = \sum_{m_1, ..., m_k \geq 0} \frac{(a)_{m_1 + \cdots + m_k} (b_1)_{m_1} \cdots (b_k)_{m_k}}{(c)_{m_1 + \cdots + m_k}} \frac{z_1^{m_1}}{m_1!} \cdots \frac{z_k^{m_k}}{m_k!} \tag{3.3}$$

is the Lauricella function, $(a)_m = a(a+1)...(a+m-1)$ the Pochhammer symbol and $c = \sum_{j=1}^{k} b_j$; Exton [3, 2.1.4, 2.3.5]. Therefore,

$$\mathcal{S}[F';n](\mathbf{z}) = F_D^{(d)}(n, b_1, ..., b_k; n; z_1, ..., z_d),$$

and

$$\begin{aligned}
\mathcal{S}[F';n](\mathbf{z}) &= \sum_{m_1,...,m_d \geq 0} (b_1)_{m_1} \cdots (b_d)_{m_d} \frac{z_1^{m_1}}{m_1!} \cdots \frac{z_d^{m_d}}{m_d!} \\
&= \prod_{j=1}^{d} \frac{1}{(1-z_j)^{b_j}} \\
&= \prod_{j=1}^{d} \frac{1}{(1-z_j)^{\sum_{i=1}^{n} \alpha_{ji}}} \\
&= \prod_{i=1}^{n} \prod_{j=1}^{d} \frac{1}{(1-z_j)^{\alpha_{ji}}}.
\end{aligned}$$

The proof of the theorem is complete. □

Theorem 3.2. *Let $\mathbf{X}_1, ..., \mathbf{X}_m$ be independent and identically distributed random vectors in $\mathbb{R}^d$ having a symmetric multivariate stable distribution of exponent $0 < \alpha \leq 2$. Let $\mathbf{S}_{m:n}$ be the corresponding MRWA given in (2.1). Then $\mathbf{S}_{m:n} \overset{d}{=} V_\alpha \mathbf{X}_1$, where $V_\alpha = (\sum_{j=1}^{m} R_{j:n}^{\alpha})^{1/\alpha}$.*

Proof: It is well known that if $\mathbf{X}_1, ..., \mathbf{X}_m$ are independent, identical and symmetrically distributed stable random vectors of exponent α, then $\sum_{j=1}^{m} a_j \mathbf{X}_j \overset{d}{=} (\sum_{j=1}^{m} a_j^{\alpha})^{1/\alpha} \mathbf{X}_1$, for any set of univariate positive constants $a_1, ..., a_m$, see Samorodnitsky and Taqqu [12]. Let $\mathcal{S}_{m:n}(\mathbf{z})$ stand for the Stieltjes transform of $\mathbf{S}_{m:n}$, then

$$\begin{aligned}
\mathcal{S}_{m:n}(\mathbf{z}) &= E\left(\frac{1}{1- < \mathbf{S}_{m:n}, \mathbf{z} >}\right) \\
&= E\left(E\left(\frac{1}{1- < \sum_{j=1}^{m} R_{j:n} \mathbf{X}_j, \mathbf{z} >} | R_{m:1}, ..., R_{m:n}\right)\right) \\
&= E\left(E\left(\frac{1}{1- < (\sum_{j=1}^{m} R_{j:n}^{\alpha})^{1/\alpha} \mathbf{X}_1, \mathbf{z} >} | R_{m:1}, ..., R_{m:n}\right)\right) \\
&= E\left(\frac{1}{1- < (\sum_{j=1}^{m} R_{j:n}^{\alpha})^{1/\alpha} \mathbf{X}_1, \mathbf{z} >}\right) \\
&= \mathcal{S}\{V_\alpha \mathbf{X}_1\}(\mathbf{z}),
\end{aligned}$$

giving the result, where $\mathcal{S}\{V_\alpha \mathbf{X}_1\}(\mathbf{z})$ stands for the Stieltjes transform of $V_\alpha \mathbf{X}_1$. □

Remark 3.1. Interestingly, it follows from Theorem 3.2 that the RWA of independently and identically distributed stable random vectors is not stable (unless $\alpha = 1$), but it is a certain *randomly scaled stable* random vector. Moreover it follows from the inequality $(a+b)^p < a^p + b^p$, $0 < p < 1$, that for $1 < \alpha \leq 2$, $V_\alpha < 1$. Consequently, the RWA $\mathbf{S}_{n:n}$ exhibits smaller variation than $\mathbf{X}$s.

Remark 3.2. We note that for $\alpha = 1$, $V_1 = 1$, and consequently $\mathbf{S}_{n:n} \stackrel{d}{=} \mathbf{X}_1$. If $\mathbf{h}(\mathbf{x})$ is the density function of a multivariate stable distribution of exponent 1, *call it multivariate Cauchy distribution*, then it follows from Theorems 2.1, and 3.2 that

$$\int_{\mathbb{R}^d} \frac{1}{[1- < \mathbf{z}, \mathbf{x} >]^n} \, \mathbf{h}(\mathbf{x}) d\mathbf{x} = \left[\int_{\mathbb{R}^d} \frac{1}{1- < \mathbf{z}, \mathbf{x} >} \, \mathbf{h}(\mathbf{x}) d\mathbf{x} \right]^n, \quad \text{for every } n \geq 1.$$

The density function $\mathbf{h}(\mathbf{x})$, in the context of stable random vectors with exponent $\alpha = 1$, in general does not assume a close formulation. The density function of a special class of multivariate Cauchy random vectors, called "multivariate Cauchy of order one" assumes the following formulation, given in Press [9], namely,

$$\mathbf{h}(\mathbf{x}) = K|\Sigma|^{-\frac{1}{2}}[1 + (\mathbf{x} - \mathbf{a})^T \Sigma^{-1} (\mathbf{x} - \mathbf{a})]^{-\frac{1+d}{2}},$$

where $K = \Gamma(\frac{1+d}{2})\pi^{-\frac{1+d}{2}}$, $\mathbf{a} \in \mathbb{R}^d$ and the $d \times d$ matrix Σ is positive definite.

Let us also record the following interesting symmetrical property of MRWA.

Theorem 3.3. *Let every $\mathbf{X}_i$ be symmetric about $\mathbf{a}_i$, for $i = 1, ..., m$. Then the MRWAs $\mathbf{D}_{m:n} - \mathbf{S}_{m:n}$ and $\mathbf{S}_{m:n} - \mathbf{D}_{m:n}$ have the same distribution, where $\mathbf{D}_{m:n} = \sum_{j=1}^m R_{j:n}\mathbf{a}_j$. In particular, if every $\mathbf{X}_i$ is symmetric about $\mathbf{a}$, then $\mathbf{S}_{m:n}$ will be symmetric about $\mathbf{a}$.*

The proof is straightforward, so it is omitted.

Let us call $\mathbf{D}_{m:n} = \sum_{j=1}^m R_{j:n}\mathbf{a}_j$ the centroid for MRWA $\mathbf{S}_m$. This is interesting; indeed it follows from this theorem that the centroid is random regresses of $\mathbf{a}_1, ..., \mathbf{a}_m$. According to Galton, see Hansen [4, page 40], the projected height of child on parent is a weighted average of the population mean height and the parents height with weights $(1/3, 2/3)$. Indeed if we let $ER_1 = 1/3$, then $E[S_2|X_2] = ER_1EX_1 + ER_2X_2 = (1/3)\mu + (2/3)X_2$; the right side is the equation reported in Hansen [4].

Conclusion. Averages for multivariate random vectors with random weights where the weights are spacings corresponding to a uniform $(0, 1)$ sample are introduced and studied in this article. Certain techniques for the their distributional studies are introduced. This study gives rise to new families of multivariate distributions. The statistics literature is quite rich about the sample mean and its applications. The topics that are studied for the sample mean, such as strong law of large numbers, asymptotic theory and its applications in inference, would be interesting subjects for further research work on randomly weighted average of random vectors. For further references, see also Roozegar and Soltani [10, 11].

ACKNOWLEDGMENTS

The authors would like to thank the Research Administration of the Kuwait University for funding this research under the Research Grant SS01/14.

The authors would like to thank a reviewer for the encouraging report and suggestions.

REFERENCES

[1] Cuyt, A.; Golub, G.; Milanfar, P. and Verdonk, B. (2005). Multidimensional integral inversion, with applications in shape reconstruction, *SIAM Journal on Scientific Computing*, **27**(3), 1058–1070.

[2] Debnath, L. and Bhatta, D. (2007). *Integral Transforms and Their Applications*, Chapman and Hall/CRC Press, London.

[3] Exton, H. (1976). *Multiple Hypergeometric Functions and Applications*, John Wiley & Sons Inc., New York.

[4] Hansen, B.E. (2015). *Econometrics*, University of Wisconsin, Free online access.

[5] Johnson, N.L. and Kotz, S. (1990). Randomly weighted averages: Some aspects and extensions, *The American Statistician*, **44**(3), 245–249.

[6] Karlin, S.; Micchelli, C.A. and Rinott, Y. (1986). Multivariate splines: A probabilistic perspective, *Journal of Multivariate Analysis*, **20**(1), 69–90.

[7] Kerov, S.V.E. and Tsilevich, N.V. (2004). The Markov–Krein correspondence in several dimensions, *Journal of Mathematical Sciences*, **121**(3), 2345–2359.

[8] Nadaraya, E.A. (1964). On estimating regression, *Theory of Probability & Its Applications*, **9**(1), 141–142.

[9] Press, S.J. (1972). Multivariate stable diatributions, *Journal of Multivariate Analysis*, **2**, 444–462.

[10] Roozegar, R. and Soltani, A.R. (2014). Classes of power semicircle laws that are randomly weighted average distributions, *Journal of Statistical Computation and Simulation*, **84**(12), 2636–2643.

[11] Roozegar, R. and Soltani, A.R. (2015). On the asymptotic behavior of randomly weighted averages, *Statistics & Probability Letters*, **96**, 262–279.

[12] Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, CRC Press, London.

[13] Soltani, A.R. and Roozegar, R. (2012). On distribution of randomly ordered uniform incremental weighted averages: Divided differences approach, *Statistics & Probability Letters*, **82**(5), 1012–1020.

[14] Van Assche, W. (1987). A random variable uniformly distributed between two independent random variables, *Sankhya A*, **49**, 207–211.

[15] Watson, G.S. (1956). On the joint distribution of the circular serial correlation coefficients, *Biometrika*, **4**, 161–168.

[16] Watson, G.S. (1964). Smooth regression analysis, *Sankhya A*, **26**, 359–372.

[17] Zeng, W.B. (1995). On characterization of multivariate stable distribution via random linear statistics, *Journal of Theoretical Probability*, **8**(1), 1–15.

REVSTAT – Statistical Journal
Volume 18, Number 4, October 2020, 461–481

NONPARAMETRIC CUSUM CHARTS FOR CIRCULAR DATA WITH APPLICATIONS IN HEALTH SCIENCE AND ASTROPHYSICS

Authors: F. Lombard
– Department of Statistics, University of Johannesburg, South Africa
fredl@uj.ac.za

Douglas M. Hawkins
– Scottsdale Scientific LLC, Scottsdale, AZ, United States of America
dhawkins@umn.edu

Cornelis J. Potgieter
– Department of Statistical Science, Southern Methodist University, Dallas, TX, United States of America
and
Department of Statistics, University of Johannesburg, South Africa
cpotgieter@smu.edu

Received: October 2017 Revised: April 2018 Accepted: June 2018

Abstract:

- This paper develops non-parametric rotation invariant CUSUMs suited to the detection of changes in the mean direction as well as changes in the concentration parameter of circular data. The properties of the CUSUMs are illustrated by theoretical calculations, Monte Carlo simulation and application to sequentially observed angular data from health science and astrophysics.

Keywords:

- *average run length; cumulative sum; directional data.*

AMS Subject Classification:

- 62L10, 62G99, 62H11.

1. INTRODUCTION

Sequential CUSUM methods for detecting parameter changes in distributions on the real line is a well developed field with an extensive literature. The same cannot be said about CUSUM methods to detect changes of location in non-Euclidean spaces such as the circle. Distributions on the circle generate data which cannot generally be treated in the same manner as linear data - see Fisher [3, Chapter 1 and Section 3.1], Mardia and Jupp [15, Chapter 1] and Jammalamadaka and SenGupta [9, Section 1.2.2]. One impediment to the application of linear CUSUM methods is the fact that a circle has no well separated beginning and end. Whichever point is selected as the beginning point, the distance between it and the endpoint is zero. A family of distributions with a fixed arc on the circle as support could in principle be treated as if the sample space were a finite fixed interval on the real line. However, the options involved in formulating a changepoint model would then be severely curtailed: a model involving shifts of arbitrary size in the location of the distribution would be out of the question. The distributions from which the data in our applications in Section 5 arise encompass the full circle and are therefore not amenable to analysis by linear CUSUM methods.

Lombard, Hawkins and Potgieter [13] reviewed the current state of change detection procedures for circular data. They also constructed distribution free CUSUMs for circular data in which the numerical value of an in-control mean direction is specified, the objective being to detect a change in mean direction away from this value. The situation is analogous to that in which the well known Page [18] CUSUM is applied, namely detection of a change away from a specified numerical value of the mean of a distribution on the real line. However, in the examples treated in Section 5 of the present paper, no in-control circular mean value is specified and the objective is to detect a change away from the unknown current circular mean value, whatever it may be. Such a CUSUM, unlike that proposed by Lombard, Hawkins and Potgieter [13], must be rotation invariant because the outcome of the analysis should not depend upon which point on the circle is chosen as the origin of angular measurement.

The main contribution of the present paper is the construction of such invariant CUSUMs for circular data. The CUSUMs we construct are non-parametric in the sense that their form is not dependent upon an underlying parametrically specified distribution. The in-control properties of the CUSUMs are shown in a Monte Carlo study to be quite robust over a wide class of circular distributions, which makes them near distribution free over this class. As far as we are aware, no CUSUMs of this nature for circular data have to date been treated in the statistical literature.

Section 2 of the paper focuses on mean direction. We provide justifications for the form of our CUSUM and discuss some computational details. In Section 3 we elaborate on its in-control and out-of-control properties. The results of an extensive Monte Carlo study are also reported. In Section 4 we briefly consider a CUSUM for detecting concentration changes. Section 5 demonstrates the application of the CUSUMs to two sets of data and Section 6 summarizes our results.

2. DETECTING DIRECTION CHANGE

2.1. Derivation of the CUSUM statistic

Initially the data $X_1, X_2, ...$ come from a non-uniform and unimodal continuous distribution F with unknown mean direction $\nu = \nu_0$ on the circle $[-\pi, \pi)$. This defines the in-control state. (Since mean direction is a vacuous concept in a uniform distribution, the latter is excluded from consideration. The CUSUM of Lombard and Maxwell [14], which is rotation invariant, can be used to detect a change from a uniform to a non-uniform distribution.) We estimate ν by

$$\hat{\nu}_n = \text{atan2}(S_n, C_n) \tag{2.1}$$

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

$$C_n = \sum_{j=1}^{n} \cos X_j, \qquad S_n = \sum_{j=1}^{n} \sin X_j, \tag{2.2}$$

and atan2 denotes the four-quadrant inverse tangent function

$$\text{atan2}(x, y) = \begin{cases} \tan^{-1}(x/y) & \text{if } y > 0 \\ \tan^{-1}(x/y) + \pi\text{sign}(x) & \text{if } y < 0 \\ (\pi/2)\text{sign}(x) & \text{if } y = 0, x \neq 0 \\ 0 & \text{if } y = x = 0, \end{cases}$$

the symbol $\tan^{-1}$ denoting the usual inverse tangent function with range restricted to $(-\pi/2, \pi/2)$. This non-parametric estimator is, in fact, also the maximum likelihood estimator of mean direction in a von Mises distribution, which is arguably the best known among circular distributions. The von Mises distribution with mean direction ν and concentration κ, has density function

$$f(x) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(x - \nu)], \quad -\pi \leq x < \pi,$$

where I_0 denotes the modified Bessel function of the first kind of order zero. The log-likelihood ratio based on observations $X_1 + \delta, ..., X_n + \delta$ is, apart from a factor not depending upon δ, given by

$$l(\delta) = \cos(X_n - \delta - \nu)$$

and a locally most powerful test of the hypothesis $H_0 : \delta = 0$ is therefore based on the derivative

$$\left.\frac{dl(\delta)}{d\delta}\right|_{\delta=0} = \sin(X_n - \nu).$$

Replacing ν by $\hat{\nu}_{n-1}$ leads to consideration of a CUSUM based on the statistic

$$V_n = \sin(X_n - \hat{\nu}_{n-1}). \tag{2.3}$$

Despite the fact that V_n originates from the von Mises distribution, it has at least two purely non-parametric origins that do not depend upon any assumption involving the type of the underlying distribution.

The first of these follows upon expanding the sine function and using the trigonometric relations

$$\sin(\hat{\nu}_{n-1}) = S_{n-1}/R_{n-1}, \qquad \cos(\hat{\nu}_{n-1}) = C_{n-1}/R_{n-1},$$

wherein

$$R_n^2 = C_n^2 + S_n^2. \tag{2.4}$$

This gives

$$V_n = (C_{n-1}/R_{n-1}) \sin X_n - (S_{n-1}/R_{n-1}) \cos X_n, \tag{2.5}$$

which is the (signed) area of the parallelogram spanned by the unit length vectors $(C_{n-1}, S_{n-1})/R_n$ and $(\sin X_n, \cos X_n)$. The former of these vectors points in the mean direction of the data $X_1, ..., X_{n-1}$ while the latter vector points in the direction of the new observation X_n.and the greater the angular distance between the two directions is, the larger will be the area of the parallelogram. Thus, if a change in mean direction ν occurs at index n, we can expect a succession of positive or negative values $V_n,\ n > \tau$.

A second non-parametric argument leading to consideration of V_n comes from considering the change $\hat{\nu}_n - \hat{\nu}_{n-1}$ in the estimate of ν effected by a change in mean direction from ν to $\nu + \delta$ occurring at index n. We have

$$\hat{\nu}_n = \text{atan2}\left[S_{n-1} + \sin(X_n + \delta), C_{n-1} + \cos(X_n + \delta)\right]$$

$$= \text{atan2}(S_{n-1}/n + \delta_{1,n}, C_{n-1}/n + \delta_{2,n})$$

where

$$n\delta_{1,n} = \sin(X_n + \delta) = \sin X_n + O(\delta),$$

$$n\delta_{2,n} = \cos(X_n + \delta) = \cos X_n + O(\delta).$$

Since both S_{n-1}/n and C_{n-1}/n converge as $n \to \infty$, and both $\delta_{1,n}$ and $\delta_{2,n}$ tend to zero, we can make a Taylor expansion around $(S_{n-1}/n, C_{n-1}/n)$. This gives

$$R_{n-1}(\hat{\nu}_n - \hat{\nu}_{n-1}) = n\delta_{1,n}\frac{C_{n-1}}{R_{n-1}} - n\delta_{2,n}\frac{S_{n-1}}{R_{n-1}} + O(n^{-1})$$

$$= \frac{C_{n-1}}{R_{n-1}} \sin X_n - \frac{S_{n-1}}{R_{n-1}} \cos X_n + O(\delta) + O(n^{-1})$$

$$= V_n + O(\delta) + O(n^{-1}),$$

which shows again the relevance of V_n for detecting changes in mean direction.

The most important property of V_n as far as motivation for the present paper is concerned is its rotation invariance: its numerical values are unaffected if *all* the data are rotated through the same fixed, but unknown, angle. Thus, a CUSUM based on V_n will be applicable in situations where no in-control direction is specified and the objective is merely to detect deviations from this arbitrary in-control direction. Both examples treated in Section 5 of the paper are of this nature. This contrasts with the distribution free CUSUMs in Lombard, Hawkins and Potgieter [13], which require a specified numerical value of the in-control mean direction.

2.2. Construction of the CUSUM

When the process is in control, that is, when $X_1, X_2, ...$ are independently and identically distributed (but with unknown mean direction), then

$$\xi_n := (V_n - \mathrm{E}_{n-1}[V_n])/\sqrt{\mathrm{Var}_{n-1}[V_n]},\ n \geq 2, \tag{2.6}$$

is a martingale difference sequence with conditional variance 1. Here and elsewhere, $\mathrm{E}_{n-1}[\cdot]$ and $\mathrm{Var}_{n-1}[\cdot]$ denote expected value and variance computed conditionally upon $X_1, ..., X_{n-1}$. Using standard martingale central limit theory, we can show that cumulative sums of the ξ_n will be asymptotically normally distributed regardless of the type of underlying distribution - see, e.g. Helland [8, Theorem 3.2]. Furthermore, if $\nu = \nu_0$ changes by an amount δ to $\nu = \nu_0 + \delta$ at observation $X_{\tau+1}$ (τ being the last in-control observation) then by either of the two arguments following (2.3), we can expect $\mathrm{E}_\tau[\xi_{\tau+1}]$ to be non-zero. Thus, a standard two-sided normal CUSUM for data on the real line, applied to the ξ_n sequence, could be expected to be effective in detecting a change away from the initial direction. Furthermore, the in-control behaviour should be *quantitatively* similar to that of a standard normal CUSUM.

The conditional mean and variance in (2.6) depend on the first two moments of $\sin X$ and $\cos X$, which are unknown parameters. Accordingly, given observations $X_1, ..., X_n$, we estimate the conditional mean and variance non-parametrically by

$$\hat{\mathrm{E}}_{n-1}[V_n] = \frac{1}{n-1}\sum\nolimits_{i=1}^{n-1} \sin(X_i - \hat{\nu}_{n-1}) = 0$$

and

$$\widehat{\mathrm{Var}}_{n-1}[V_n] = \frac{1}{n-1}\sum\nolimits_{i=1}^{n-1} \sin^2(X_i - \hat{\nu}_{n-1}) := B_{n-1}^2. \tag{2.7}$$

Then a computable CUSUM is obtained upon replacing ξ_n in (2.6) by

$$\hat{\xi}_n = V_n / B_{n-1}. \tag{2.8}$$

The CUSUM is started at observation $m+1$ by setting $D_i^{\pm} = 0$ for $i = 1, ..., m$ and

$$\begin{aligned} D_{m+n}^{+} &= \max\{0, D_{m+n-1}^{+} + \hat{\xi}_{m+n} - \zeta\} \\ D_{m+n}^{-} &= \min\{0, D_{m+n-1}^{-} + \hat{\xi}_{m+n} + \zeta\} \end{aligned} \tag{2.9}$$

for $n \geq 1$, where ζ is the reference value. The run length, N, is the first index n at which either $D_{m+n}^{+} \geq h$ or $D_{m+n}^{-} \leq -h$, where h is a control limit. The control limit is chosen to produce a specified in-control average run length (ARL), which we denote throughout by ARL_0. The first m observations serve to make an initial estimate of the population moments *after which the estimates are updated with the arrival of each new observation.* Since the the random variables $\sin X$ and $\cos X$ are bounded, convergence of sample moments to population moments would be quite rapid so that a relatively small number m of observations should suffice to initialize the CUSUM.

2.3. Implementation

Implementation of the CUSUM scheme requires an efficient method of updating the summand $\hat{\xi}_{n-1}$ upon arrival of a new observation X_n. For this, set

$$s_n = \sin X_n, \quad c_n = \cos X_n$$

and

$$C_n^{(2)} = \sum_{j=1}^{n} c_j^2, \quad S_n^{(2)} = \sum_{j=1}^{n} s_j^2, \quad A_n^{(2)} = \sum_{j=1}^{n} s_j c_j$$

and observe that

$$(n-1)B_{n-1}^2 = \frac{C_{n-1}^2}{R_{n-1}^2} S_{n-1}^{(2)} + \frac{S_{n-1}^2}{R_{n-1}^2} C_{n-1}^{(2)} - 2\frac{C_{n-1}S_{n-1}}{R_{n-1}^2} A_{n-1}^{(2)}. \tag{2.10}$$

In particular, we see that the R_{n-1} factors in V_n and B_{n-1} cancel, whence

$$\hat{\xi}_n = \frac{V_n^*}{B_{n-1}^*} := \frac{C_{n-1}\sin X_n - S_{n-1}\cos\ X_n}{\sqrt{\left(C_{n-1}^2 S_{n-1}^{(2)} + S_{n-1}^2 C_{n-1}^{(2)} - 2C_{n-1}S_{n-1}A_{n-1}^{(2)}\right)/(n-1)}}. \tag{2.11}$$

Next, note the simple recursions

$$S_{n-1} = S_{n-2} + s_{n-1}, \quad C_{n-1} = C_{n-2} + c_{n-1},$$

$$S_{n-1}^{(2)} = S_{n-2}^{(2)} + s_{n-1}^2, \quad C_{n-1}^{(2)} = C_{n-2}^{(2)} + c_{n-1}^2$$

and

$$A_{n-1}^{(2)} = A_{n-2}^{(2)} + s_{n-1}c_{n-1}.$$

To compute V_n^* in (2.11) given $S_{n-2}, C_{n-2}, c_{n-1}, c_n, s_{n-1}$ and s_n, use the first of these recursions. To compute B_{n-1}, given $S_{n-1}, C_{n-1}, S_{n-1}^{(2)}, C_{n-1}^{(2)}, A_{n-1}^{(2)}, c_{n-1}$, and s_{n-1}, use (2.10).

A rational basis for specifying a reference value ζ is also required. This aspect of the CUSUM design is considered in Section 3.3 of the paper.

3. In-control properties

While the proposed CUSUM is not distribution free, the asymptotic in-control normality of CUSUMs of $\hat{\xi}_n$ suggests that it may be nearly so. Then, use of standard normal distribution CUSUM control limits should lead to an in-control ARL sufficiently close to the nominal value to make the CUSUMs of practical use. The requisite control limit h can be obtained from the widely available software packages of Hawkins, Olwell and Wang, [7] or Knoth [12]. To check this expectation we estimated by Monte Carlo simulation the in-control ARL over a range of unimodal symmetric and asymmetric distributions on the circle. Among the multitude of possible distributions, the class of wrapped stable and Student t distributions, together with their skew versions, represent a wide range of unimodal distribution shapes on the circle. Simulated data from these distributions are easily obtained by generating random

numbers Y from the distribution on the real line and then wrapping these around the circle by the simple transformation $Y(\text{mod } 2\pi)$. Algorithms for generating the random numbers Y are given in Nolan [17] and in Azzalini and Capitanio [2]. The algorithms were implemented in Matlab and the relevant programs are included in the supplementary material to this paper.

Some simulations were also run on data from other types of distribution which are defined directly on the circle and not obtained by wrapping. Specifically, we used the sine-skewed distributions developed Umbach and Jammalamadaka [22] and by Abe and Pewsey [1]. In contrast to the wrapped stable and Student t distributions, the densities of these distributions have closed form expressions, which facilitates model fitting and parameter estimation. The various unimodal distribution shapes available in these classes of distributions are quite similar to those in the class of wrapped distributions. Since the behaviour of a non-parametric CUSUM depends more on the general shape of the underlying distribution than on the specific parameter values producing that shape, it comes as no surprise that the in-control behaviour of the CUSUMs proposed here is quite similar in the two classes (wrapped and directly constructed) of distributions. Since wrapped distributions are widely known and understood, we frame our discussion in the context of these distributions. Some simulation results for data from the sine-skewed distributions are included in the supplementary material to this paper. In the discussion that follows, $S_\alpha,\ 0 < \alpha \leq 2$, denotes a stable distribution with index α and $t_n,\ n \geq 1$ denotes a Student t-distribution with n degrees of freedom.

In assessing the performance of the direction CUSUM under various symmetric in-control and out-of-control distributions, we standardize the observations to a common measure of concentration. The concentration parameter κ of the von Mises(ν, κ) distribution satisfies the relation

$$\kappa = A^{-1}(\mathrm{E}[\cos(X - \nu)]) \tag{3.1}$$

where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$ and I_1 denotes the modified Bessel function of the first kind of order 1. In view of the status of the von Mises distribution among circular distributions, which is much like that of the normal distribution among distributions on the real line, we use in this paper κ in (3.1) as a measure of the concentration of a unimodal circular distribution with mean direction ν. Thus, given κ and the density function of Y, the scale parameter σ is chosen to make the distribution of the wrapped random variable

$$X = (\sigma Y)_w := \sigma Y (\text{mod } 2\pi)$$

satisfy (3.1).

For instance, suppose Y has an S_α distribution with characteristic function

$$\phi(t; \alpha) = \mathrm{E}[\cos tY] = \exp(-|t|^\alpha).$$

Then (Jammallamadaka and SenGupta, [9, Proposition 2.1]),

$$\mathrm{E}[\cos(\sigma Y)_w] = \phi(\sigma; \alpha) = \exp(-\sigma^\alpha)$$

so that

$$\sigma = (-\log(\ A(\kappa)))^{1/\alpha}. \tag{3.2}$$

As another example, a Student t-distribution with α degrees of freedom has characteristic function

$$\phi(t;\alpha) = \frac{K_{\alpha/2}(\sqrt{\alpha}t)(\sqrt{\alpha}t)^{\alpha/2}}{2^{\alpha/2-1}\Gamma(\frac{\alpha}{2})}$$

where $K_{\alpha/2}$ denotes the modified Bessel function of the second kind order $\alpha/2$ and Γ denotes the gamma function. Thus, in this case,

$$\mathrm{E}[\cos(\sigma Y)_w] = \phi(\sigma;\alpha) = \frac{K_{\alpha/2}(\sqrt{\alpha}\sigma)(\sqrt{\alpha}\sigma)^{\alpha/2}}{2^{\alpha/2-1}\Gamma(\frac{\alpha}{2})},$$

and σ is the solution to the equation

$$K_{\alpha/2}(\sqrt{\alpha}\sigma)(\sqrt{\alpha}\sigma)^{\alpha/2} = 2^{\alpha/2-1}\Gamma(\frac{\alpha}{2})A(\kappa). \tag{3.3}$$

Some numerical values that were used in the simulation study which is reported next, are shown in Table 1.

Table 1: Scale parameter σ solving (3.2) and (3.3).

Distribution	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$
S_2	0.90	0.60	0.46
S_1	0.81	0.36	0.21
$S_{1/2}$	0.65	0.13	0.04
t_3	1.07	0.64	0.46
t_2	1.00	0.55	0.38

3.1. Symmetric distributions

We used standard normal control limits in 50,000 Monte Carlo realizations of the two-sided CUSUM in each of five underlying symmetric unimodal distributions: wrapped Student t-distributions with 2 and 3 degrees of freedom and three wrapped stable distributions with indexes $\alpha = 2$ (the wrapped normal distribution), $\alpha = 1$ (the wrapped Cauchy distribution, which is also the wrapped Student t-distribution with 1 degree of freedom) and $\alpha = 1/2$ (the wrapped symmetrized Lévy distribution). Except for the wrapped normal, these are wrapped versions of heavy-tailed symmetric distributions on the real line. Each of the distributions was standardized to concentrations of $\kappa = 1,\ 2$ and 3 by specifying the scale parameter σ (see Table 1) in accordance with (3.2) and (3.3). Two sets of simulations were run. In the first set, the CUSUMs were initiated at $n = 11$, the first $m = 10$ observations serving to establish initial estimates of the unknown parameters. In the second set we took $m = 25$, initiating the CUSUM at $n = 26$.

We present in Tables 2.1 and 2.2 aggregated sets of results representing the general picture. (Detailed tables are given in the supplementary material to this paper.) Each entry is the average of five estimated in-control ARLs, one from each of the five distributions.

The number in brackets shows the range of the five estimates. The tables show the results for reference values $\zeta = 0$ and $\zeta = 0.25$.

Table 2.1: Average in-control ARL of the non-parametric CUSUM in five symmetric distributions ($m = 10$). The number in brackets is the range of the five estimates.

ARL_0	$\zeta = 0$			$\zeta = 0.25$		
	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$
250	242 (2)	243 (5)	242 (4)	236*(4)	233* (2)	225* (20)
500	490 (3)	491 (6)	491 (10)	493 (9)	483 (8)	464* (52)
1000	1037 (9)	1039 (14)	1042 (20)	1018 (7)	997 (30)	958*† (117)

Table 2.2: Average in-control ARL of the non-parametric CUSUM in five symmetric distributions ($m = 25$). The number in brackets is the range of the five estimates.

ARL_0	$\zeta = 0$			$\zeta = 0.25$		
	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$
250	244 (2)	244 (6)	245 (7)	242 (4)	239 (3)	234* (8)
500	492 (4)	493 (7)	493 (10)	498 (9)	491 (7)	478 (28)
1000	1039 (11)	1041 (10)	1045 (17)	1024 (13)	1005 (26)	971 (82)

All but the four starred estimates shown in the tables lie within 5% of the nominal value. The exceptions, which all lie within 10%, occur at $\zeta = 0.25$ and predominantly at the smaller warmup $m = 10$. In the cell marked *† the five estimates were 874, 959, 976, 988 and 991, the outlier 874 coming from the very heavy tailed Lévy distribution. In fact, all three discrepancies in this column are attributable to a substantial underestimate from the Lévy distribution Clearly, the CUSUM is very near distribution free overall when a reference constant close to zero is used. With a larger reference constant, as the concentration increases so does the variation in true ARL between distributions. This behaviour can be explained to a large extent by reference to the martingale central limit theorem upon which the construction of the CUSUM rests. If the summand ξ_n is replaced by $\xi_n \mp \zeta$, the cumulative sums take the form $S_k \mp k\zeta$ where

$$S_k = \sum_{n=m+1}^{m+k} \hat{\xi}_n, \ k \geq 1 \tag{3.4}$$

and ζ is positive. The rationale behind the construction of the CUSUM consists essentially in replacing the discrete time process $S_k/h = \sum_{n=m+1}^{m+k} \hat{\xi}_n/h$, $k \geq 1$, where h is the control limit, by a continuous time Brownian motion process, $W(t)$, $t > 0$. This is effected by changing the time scale. We identify k with th^2 where h is the control limit, and then replace S_k/h by $W(th^2)/h$, which has the same distribution as $W(t)$. Similarly, $k\zeta$ is replaced by $th^2\zeta/h = th\zeta$.

Thus, $(S_k \mp k\zeta)/h,\ k \geq 1$, is replaced by $W(t) - th\zeta$. The validity of this procedure requires that h tends to ∞. Now, if ζ is positive and $h \to \infty$ then the drift term $th\zeta \to \infty$, which makes the resulting CUSUM useless. To avoid this effect, ζ must be chosen to be $O(1/h)$, which in practical terms means that ζ should be a small positive number or zero.

Next, the effect of any Phase I estimation on the in-control Phase II performance of the CUSUM needs to be considered. Given $\hat{\zeta}$, let $\hat{h}$ be the control limit which gives a standard normal CUSUM an in-control ARL value ARL_0. The simulation results in Tables 2.1, 2.2 and 3 together with the ensuing discussion indicate that the resulting Phase II CUSUM is near distribution free provided that the reference constant is suitably close to zero. Thus, regardless of the form of the underlying distribution, in such cases the true Phase II in-control ARL will be nearly constant and acceptably close to the nominal value ARL_0. This behaviour is in stark contrast to that of parametric CUSUMs where estimating unknown parameters from Phase I data and then pretending that the Phase I estimate is the true value, affects irrevocably the in-control ARL of the Phase II CUSUM. Then there is no guarantee that the in-control ARL will be equal to, or even near, the nominal value. This point has been made repeatedly in the published literature, most recently by Keefe, et al. [11, Introduction section] and Saleh et al. [20]. Hawkins and Olwell [6, pages 159–160] give a realistic example in which the true in-control ARL of a normal distribution CUSUM, with variance estimated from Phase I data, differs by two orders of magnitude from the nominal value.

In this connection, and to illustrate further the in-control behaviour of the nonparametric CUSUM, we present next a result that is representative of a general pattern. Consider a situation in which data arise from a wrapped t_3 distribution with concentration parameter κ - see (3.3). CUSUMs with reference constants $\zeta = 0$ and $\zeta = 0.25$ and nominal in-control ARL 500 are run at $\kappa = 1$ and $\kappa = 3$. A Phase I sample of size $m = 30$ is used in each case to obtain an initial value B_m^* of the sequence of denominators in the summands $\hat{\xi}_n$ see (2.11). The "true" in-control ARLs, estimated from 50,000 Monte Carlo trials in each instance, are shown in Table 3.

Table 3: Estimated in-control ARL of direction CUSUM for data from a wrapped t_3 distribution with concentration parameter κ. Warmup $m = 30$ and based on 50,000 Monte Carlo trials.

κ	$\zeta = 0$	$\zeta = 0.25$
$\kappa = 1$	492	499
$\kappa = 3$	492	482

In each of the six instances the 50,000 values of B_m^* were grouped into bins of unit length and the average of the corresponding run lengths in each bin calculated. Figure 1 shows plots of these average run lengths against the midpoints of the bins together with confidence intervals of width equal to three estimated standard errors (Bins containing fewer than 100 observations, which contain the less commonly occurring values of B_m^*, are not shown.) The figure thus provides a representation of the Phase II in-control ARL, *conditional upon* the Phase I estimate B_m^*. It is only at the combination $\kappa = 3,\ \zeta = 0.25$. that the Phase II in-control ARL exhibits substantial systematic variation away from the corresponding uncon-

ditional value in Table 3.

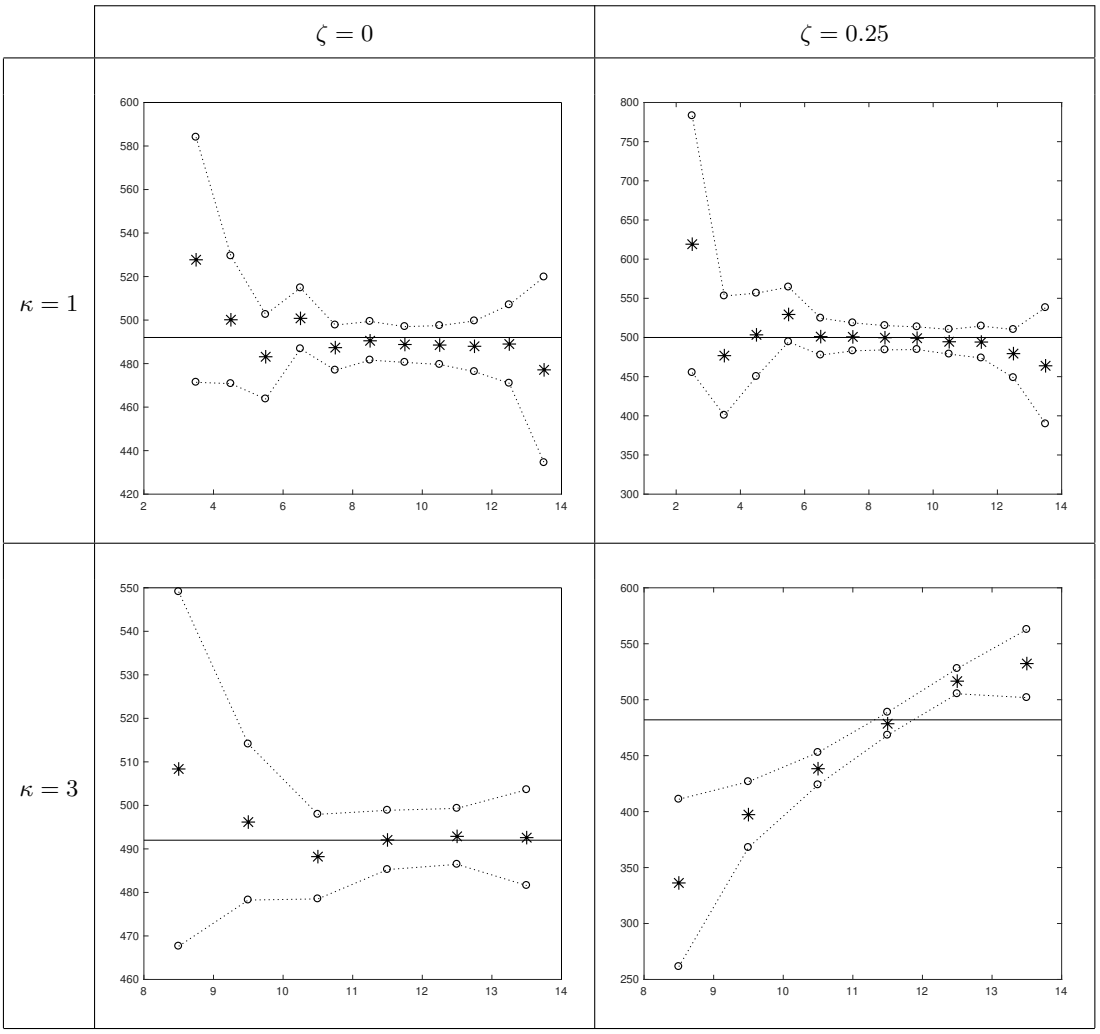

Figure 1: In-control ARL (on the vertical axis), conditional upon the value of B^*_{30} (on the horizontal axis), for two concentrations κ and two reference values ζ in wrapped t_3 distributions. The stars denote the ARL values and the dotted lines are 95% confidence intervals.

3.2. Asymmetric distributions

To assess the effect of skewness in the underlying distribution on the in-control ARL, we generated data from wrapped skew-normal distributions (Pewsey, [19]) with mean direction zero and skewness parameters $\lambda = 2$ (lightly skewed), $\lambda = 7$ (moderately skewed) and $\lambda = \infty$ (heavily skewed), wrapped skew-stable Cauchy- and Lévy distributions with skewness parameters $\beta = 0.75$ and 1.0 (Jammallamadaka and SenGupta, [9, Section 2.2.8]) and from wrapping skew-t distributions (Jones and Faddy, [10]) with 2 and 3 degrees of freedom and skewness parameters $\lambda = 2,\ 7$ and ∞. The aggregated results are in Tables 4.1 and 4.2. Comparing the results with those in Tables 2.1 and 2.2, we see that the general pattern is

the same. The main contributors to the apparent degradation seen at $\zeta = 0.25$, $\kappa = 3$ are the excessively skewed distributions, namely the wrapped skew-normal and t-distributions with skewness parameter $\lambda = \infty$ and the wrapped Lévy distribution with skewness parameter $\beta = 1$. These distributions produce estimates that are consistently substantially lower than the rest. This is perhaps not too surprising if one takes account of their shape. The supplementary material to this paper has a Figure showing a plot of a wrapped skew-t density with 2 degrees of freedom and skewness parameters $\lambda = 0,\ 2$ and 7 at $\kappa = 3$. The extreme skewness and high concentration at $\lambda = 7$ magnifies the deleterious effect that a large reference value has on the approximation to the nominal in-control ARL (Section 3.1, first paragraph after Table 2.2). The degradation noted above largely disappears when such highly skewed distributions are eliminated from consideration.

Table 4.1: Average in-control ARL of the non-parametric CUSUM in thirteen asymmetric distributions ($m = 10$). The number in brackets is the range of the thirteen estimates.

ARL_0	$\zeta = 0$			$\zeta = 0.25$		
	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$
250	241 (2)	240 (4)	238 (7)	235 (5)	228 (11)	217 (25)
500	489 (4)	487 (6)	484 (9)	490 (8)	474 (29)	448 (71)
1000	1039 (11)	1036 (9)	1031 (13)	1013 (13)	979 (61)	915 (178)

Table 4.2: Average in-control ARL of the non-parametric CUSUM in thirteen asymmetric distributions ($m = 25$). The number in brackets is the range of the thirteen estimates.

ARL_0	$\zeta = 0$			$\zeta = 0.25$		
	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$	$\kappa = 1$	$\kappa = 2$	$\kappa = 3$
250	243 (2)	242 (3)	242 (4)	240 (3)	235 (7)	229 (22)
500	491 (5)	490 (6)	489 (7)	494 (7)	484 (25)	463 (59)
1000	1039 (13)	1038 (10)	1038 (11)	1019 (17)	988 (65)	935 (161)

3.3. Choice of reference constant

We saw in Sections 3.1 and 3.2 that the CUSUM exhibits good in- and out-of-control behaviour throughout when a small positive reference constant ζ is used. In analogy with a normal distribution CUSUM, one would expect the CUSUM to then be quite adept at detecting small changes but less effective if the change is of substantial magnitude. In the latter case, efficient detection of a change requires use of a larger reference constant. Again in analogy with a normal distribution CUSUM, an appropriate choice of reference constant

for efficient detection of a rotation of size $\geq \delta_0$ *could* be

$$\zeta = \frac{\mathrm{E}[\sin(X + \delta_0 - \nu) - \sin(X - \nu)]}{\sqrt{\mathrm{Var}[\sin(X - \nu)]}},$$

which can be estimated from some in-control Phase I data $X_1, ..., X_m$ by

$$\hat{\zeta} = \frac{\delta_0}{2} \times \frac{m^{-1} \sum_{j=1}^{m} \sin(X_j + \delta_0 - \hat{\nu}_m)}{\sqrt{m^{-1} \sum_{j=1}^{m} \sin^2(X_j - \hat{\nu}_m)}}. \tag{3.5}$$

Clearly, the variability of the estimator $\hat{\zeta}$ will depend on both the size m of the in-control Phase I sample and on the type of the unknown underlying distribution. If $\hat{\zeta}$ turns out to be too large given the known limitations of the CUSUM, one could use a reference value $\hat{\zeta} \leq 0.25$, say, and solve for δ_0 from (3.5). This δ_0 would serve as an indication of the magnitude of change that the CUSUM could be expected to detect efficiently.

3.4. Out-of-control properties

While the in-control behaviour of the CUSUM is similar to that of a CUSUM for normal data on the real line, the same is not true in respect of its out-of-control behaviour. In fact, we show next that a consequence of the continual updating of the mean direction estimator $\hat{\nu}_n$ from (2.1) is that after a change of mean direction the CUSUM will return eventually to what appears to be an in-control state. This behaviour is similar to that of self-starting CUSUMs for linear data, and is a warning to users of the need for corrective action as soon as a change is diagnosed- see Hawkins and Olwell [6, Section 7.1].

Suppose there is a rotation of size δ from $n = \tau + 1$ onwards and set $Y_i = X_{i+\tau} + \delta,\ i \geq 1$. Then, using the approximations

$$\frac{1}{\tau + k} \approx 0 \ \textit{and} \ \frac{k}{\tau + k} \approx 1$$

for large k and fixed $\tau \geq m$, the mean direction estimated from the data $X_1, ..., X_\tau, Y_1, ..., Y_k$ is

$$\hat{\nu}_{\tau+k} = \mathrm{atan2}\left(\frac{S_\tau + \sum_{i=1}^{k} \sin Y_i}{\tau + k}, \frac{C_\tau + \sum_{i=1}^{k} \cos Y_i}{\tau + k} \right)$$

$$\approx \mathrm{atan2}\left(\frac{\sum_{i=1}^{k} \sin Y_i}{k}, \frac{\sum_{i=1}^{k} \cos Y_i}{k} \right) := \hat{\nu}_k(Y),$$

which is the estimated mean direction of $Y_i,\ 1 \leq i \leq k$. Thus, for sufficiently large k, $\hat{\nu}_{\tau+k}$ is in effect estimating the mean direction of the post-change observations $Y_1, ..., Y_k$. Consequently,

$$\hat{\xi}_{\tau+k+1} \approx \frac{\sin(Y_{k+1} - \hat{\nu}_k(Y))}{\sqrt{k^{-1} \sum_{i=1}^{k} \sin^2(Y_i - \hat{\nu}_k(Y))}}$$

which, because of its rotation invariance, has the same distribution as the in-control variable $\hat{\xi}_k$.

A further consequence of this behaviour is that, in the absence of a substantial amount of in-control Phase I data there is no simple manner in which to assess, a priori, the out-of-control ARL

$$\mathrm{E}[N - \tau | N > \tau]$$

of the CUSUM. Here $N - \tau$ is the time taken for an alarm to be raised after a change has occurred, the expected value being calculated upon an assumption of no false alarms prior to the change. Nevertheless, simulation results indicate that the out-of-control ARL of the two-sided CUSUM behaves in an appropriate manner, namely that the out-of-control ARL is less than the in-control ARL_0 and that it decreases as the size of the shift increases from 0 to $\pi/2$. For shifts of size in excess of $\pi/2$,the ARL starts increasing again. This behaviour is a result of the periodic nature of the CUSUM summand. Furthermore, that choosing $\zeta = 0$ leads to substantially larger out-of-control ARLs compared to those produced by small positive reference constants.

To illustrate that the general pattern of out-of-control ARL behaviour mimics that of a normal distribution CUSUM, Table 5 gives out-of-control ARL estimates from $10,000$ simulations involving in each case shifts δ of sizes ranging from $\pi/8$, to $7\pi/8$ in a wrapped Cauchy distribution with $\kappa = 2$, a warmup sample size $m = 25$ and reference constants $\zeta = 0$, $\zeta = 0.125$ and $\zeta = 0.25$. The in-control ARL was $1,000$ throughout. The results are for shifts induced respectively at observation $\tau = 100$ and at observation $\tau = 200$.

If a sufficiently large amount of in-control Phase I data are available to allow a non-trivial nonparametric estimate of the underlying density to be made (Taylor, [21]), the in-control and out-of-control properties of the CUSUM can be fathomed by sampling from the estimated density.

Table 5: Estimated out-of-control ARL of direction CUSUM for data from a wrapped Cauchy distribution with concentration parameter $\kappa = 2$. Warmup $m = 25$. Changepoints $\tau = 100$ and $\tau = 200$.

δ	$\tau = 100$			$\tau = 200$		
	$\zeta = 0$	$\zeta = 0.125$	$\zeta = 0.25$	$\zeta = 0$	$\zeta = 0.125$	$\zeta = 0.25$
$\delta = \pi/8$	123	49	82	82	37	39
$\delta = \pi/4$	50	17	14	40	17	13
$\delta = \pi/2$	31	11	8	28	11	8
$\delta = 3\pi/4$	38	16	12	37	15	12
$\delta = 7\pi/8$	58	29	26	61	31	28

3.5. Bimodal distributions

Thus far attention has focussed on unimodal distributions. However, many of the properties of the proposed CUSUM remain intact when the underlying distribution is multimodal.

Here, we restrict attention to bimodal densities of the form

$$f(\theta) = pg(\theta) + (1-p)g(\theta - \mu_0) \tag{3.6}$$

with $1/2 \leq p < 1$ and a unimodal density g on the circle. Since the concentration of f will be less than that of g, one finds that the approximation to the nominal in-control ARL often improves markedly, even at a reference constant 0.25. For instance, let g in (3.6) be a von Mises density with high concentration $\kappa = 3.42$ and mean 0. Then, if $p = 1$ (which is the unimodal case), and with $\zeta = 0.25$ and a nominal in-control ARL of 500, the estimated true in-control ARL is 461. On the other hand if $p = 1/3$ and $\mu_0 = -3\pi/4$, in which case f is bimodal with concentration equal to 1, the estimated true in-control ARL of 492 is much closer to the nominal value.

On the other hand, the ability of the CUSUM to detect a change of size $\delta \neq 0$ decreases as μ_0 in (3.6) nears $\pm\pi$ and vanishes when f in (3.6) is antipodal, that is, when $p = 1/2$ and $|\mu_0| = \pi$. Put another way, the CUSUM is then unable to distinguish between $f(\theta)$ and $f(\theta - \delta)$. The ostensible reason for this behaviour is that an antipodal distribution does not possess a well defined mean or median Nevertheless, a non-trivial CUSUM will result upon replacing the data X_i by $2X_i$. This replacement transforms $f(\theta)$ to $g(\theta/2)/2$, which is unimodal - see, for instance, Jammalamadaka and SenGupta [9, page 48].

4. CONCENTRATION CHANGE

For data $X_1, ..., X_n$ from a von Mises(ν, κ) distribution, locally most powerful tests of the hypothesis $\kappa = \kappa_0$ $(\neq 0)$ are based on the statistic $\sum_{i=1}^{n} \cos(X_i - \nu)$. However, the fact that κ is not a scale parameter of the distribution of X complicates matters. Hawkins and Lombard [5] showed that even if the mean direction ν is known, control limits for a specified in-control ARL in a von Mises CUSUM for detecting change away from κ_0 depend upon κ_0. Nonetheless, the locally most powerful test statistic suggests application of a CUSUM based on

$$V'_n = \cos(X_n - \hat{\nu}_{n-1}), \ n \geq 1.$$

Again, there are purely non-parametric interpretations of V'_n, devoid of any reference to a von Mises distribution. For instance, since

$$V'_n = (C_{n-1}/R_{n-1}) \cos X_n + (S_{n-1}/R_{n-1}) \sin X_n,$$

we see that V'_n is the (signed) length of the projection of the vector $y_n = (\sin X_n, \cos X_n)$ in the direction $\hat{\nu}_{n-1} \approx \nu$ of the unit vector $(S_{n-1}/R_{n-1}, C_{n-1}/R_{n-1})$. If the concentration increases (decreases) after $n = \tau$, the average of $V'_{\tau+1}, ..., V'_{\tau+k}$ will tend to be greater (smaller) than the average of $V'_1, ..., V'_\tau$. Another non-parametric interpretation rests on the fact that R_n^2 in (2.4) is a frequently used non-parametric measure of concentration in a sample $X_1, ..., X_n$. Simple algebra shows that the relative change in R_{n-1}^2 brought about by the next observation X_n is

$$\frac{R_n^2}{R_{n-1}^2} - 1 = \frac{2V'_n}{R_{n-1}} + \frac{1}{R_{n-1}^2},$$

again justifying consideration of V'_n.

Proceeding in much the same manner as in Section 2.2, a CUSUM of

$$\hat{\xi}'_n = \frac{\cos(X_n - \hat{\nu}_{n-1}) - R_{n-1}/(n-1)}{B'_{n-1}} \tag{4.1}$$

where

$$B'_n = \sqrt{n^{-1}\sum_{i=1}^{n}\cos^2(X_i - \hat{\nu}_n) - R_n^2/n^2},$$

is suggested to detect a change in concentration.

A change in the numerical value of κ has a much greater effect on the denominator B'_{n-1} in (4.1) than a change of direction has on the denominator B_{n-1} in (2.8). Furthermore, the distribution of V'_n is heavily skewed. Consequently, a CUSUM based on $\hat{\xi}'_n$ cannot be expected to have a near distribution free in-control ARL over a wide range of reference values. Indeed, simulation results indicate that one is essentially restricted to $\zeta = 0$ and a large (≥ 500) nominal in-control ARL if a satisfactory degree of in-control distribution freeness is to be had over the families of distributions considered in Section 3.

5. APPLICATIONS

In the two applications treated here we define the sample mean direction of data $X_1, ..., X_n$ by

$$\hat{\nu}_n = \text{atan2}\left(\sum_{i=1}^{n}\sin X_i, \sum_{i=1}^{n}\cos X_i\right)$$

and the sample concentration, by

$$\hat{\kappa}_n = A^{-1}\left(n^{-1}\sum_{i=1}^{n}\cos(X_i - \hat{\nu}_n)\right) = A^{-1}\left(\frac{R_n}{n}\right),$$

in analogy with (3.1). After a CUSUM signals, we estimate the changepoint τ in the conventional manner. That is, if the CUSUM signals with D^+ (D^-) at $n = N$, the changepoint estimate is the last index $n < N$ at which $D_n^+ = 0$ ($D_n^- = 0$). Both data sets are included in the supplementary material to the paper.

5.1. Acrophase data

The data, kindly provided by Dr. Germaine Cornelissen of the University of Minnesota Chronobiology Laboratory, come from ambulatory monitoring equipment worn by a patient suffering from episodes of clinical depression. The time at which systolic blood pressure reaches its maximum value on a given day is called the acrophase. Monitoring the acrophase can provide an automated early warning of a possible medical condition before it becomes clinically obvious. We show the results of a two-sided CUSUM analysis with reference constant $\zeta = 0.25$ (recommended reference value from (3.5) to enable detection of a 30 degree, i.e. $\pi/6 = 0.52$ radian, rotation) and control limits $h = \pm\ 8.59$, which leads to an in-control ARL of approximately 500. The first $m = 30$ observations are used to find initial estimates of the required parameters.

The left-hand panel in Figure 2 shows the CUSUM. The upper CUSUM D^+ signals at $n = 66$ and the changepoint estimate is $\hat{\tau} = 57$, that is, 27 observations after the warmup period. The right-hand panel in Figure 2 shows the CUSUM after restarting at $n = 88$, observations 58 through 87 serving as a warmup to estimate the new direction. A sustained decrease in the lower CUSUM D^- is evident. The CUSUM signals at $n = 120$, a changepoint being indicated at $n = 110$. Continuing in this manner produces the results in Table 6, which shows the progress of the CUSUMs as the data accrue. The estimate of the mean direction and concentration in each segment is shown in the third and fourth columns of the table.

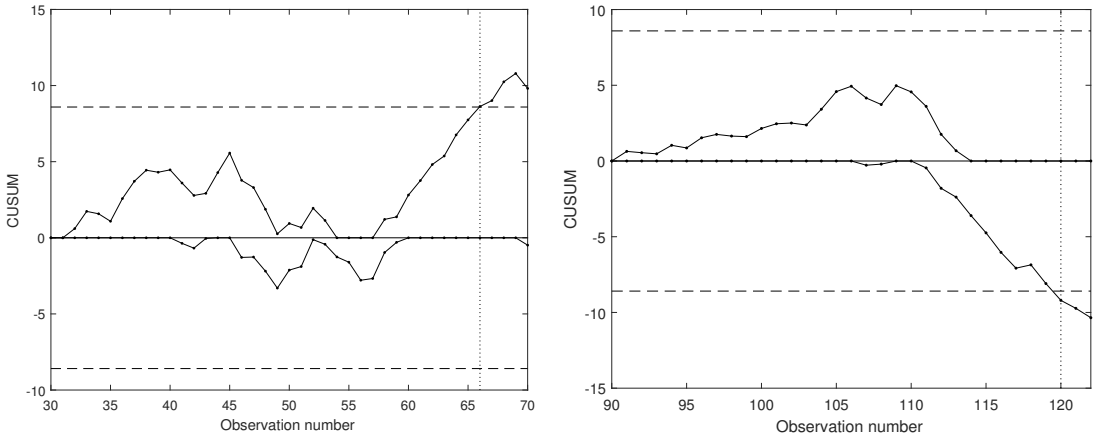

Figure 2: Direction CUSUMs of acrophase data. Left-hand panel: CUSUM after start at $n = 31$. Right-hand panel: CUSUM after restart at $n = 88$. The vertical dotted lines indicate the location of the estimated changepoints. The dashed horizontal lines indicate the control limits.

Table 6: Acrophase data: Progression of CUSUMs.

segment	signal at	$\hat{\nu}$	$\hat{\kappa}$
$1-57$	66	-1.70 (263°)	1.86
$58-110$	120	-0.76 (317°)	0.78
$111-140$	178	-1.90 (251°)	2.60
$141-241$	255	-1.19 (292°)	2.51
$242-282$	299	-0.90 (308°)	0.31
$283-306$	none	$-.007$ (360°)	1.68

Figure 3 shows dot plots, constructed after the fact, of the data in the six identified segments together with an indication of the mean in each segment. A noticeable feature in this plot is the first two increases followed by a sudden large decrease to more or less the original mean value. This is indicative of an external intervention in the treatment of the patient to reset the acrophase. After that, there follows a sustained increase, this time without any apparent external intervention. The figure also reveals some variation between the concentrations within the six segments — see the fourth column in Table 6. This does not affect the validity of the CUSUM since there is no assumption that the concentrations in the various segments must all be the same. In retrospect, it seems that the CUSUM has done a good job of identifying location changes.

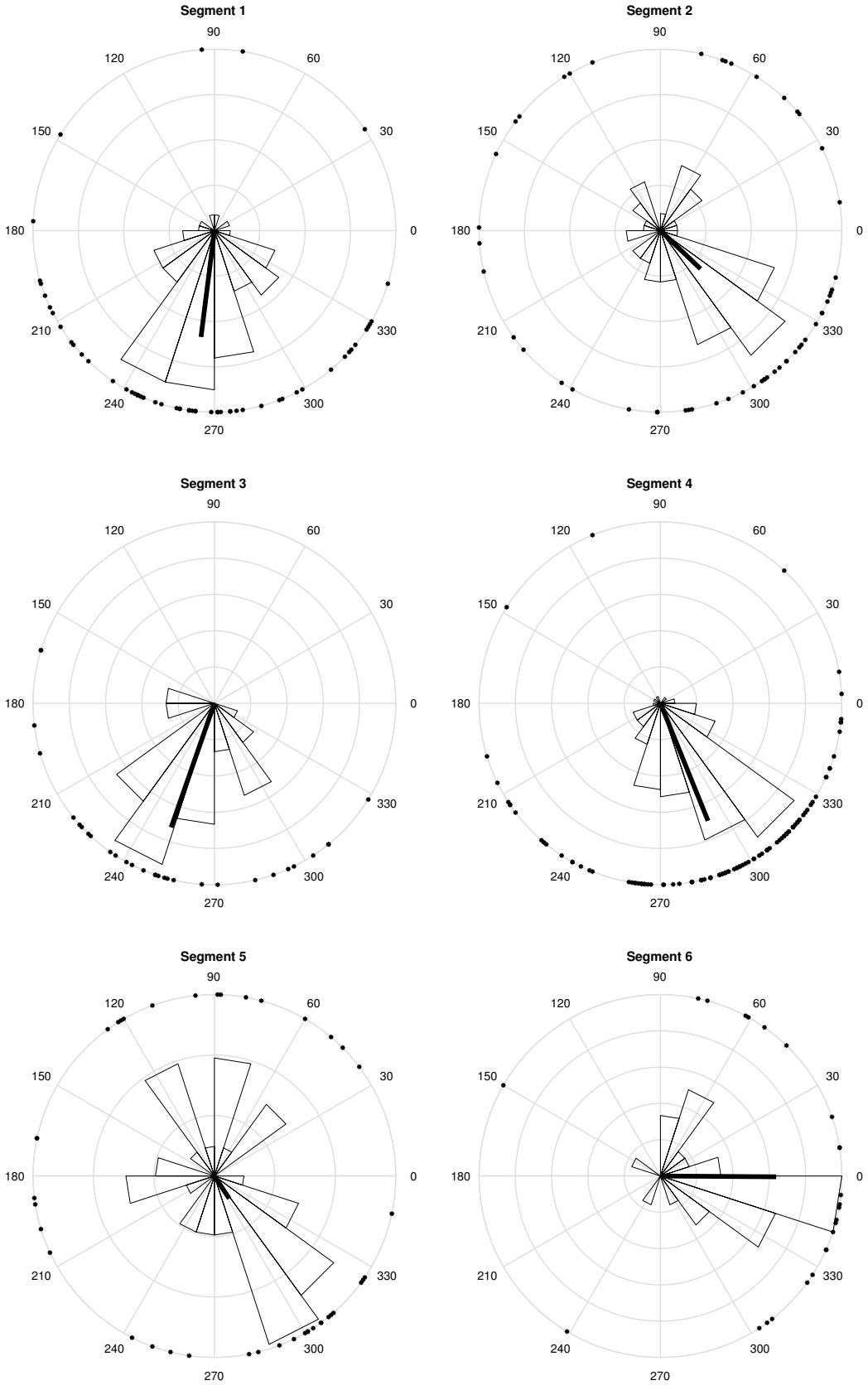

Figure 3: Rose plots of the data in each of the six identified segments of the acrophase data.

5.2. Pulsar data

Lombard and Maxwell [14] developed a rotation invariant cusum to detect deviation from a uniform distribution on the circle and applied it to some data consisting of arrival times of cosmic rays from the vicinity of a pulsar. The objective is to detect periods of sustained high energy radiation. Following a standard procedure in Astrophysics, the data were wrapped around a circle of circumference equal to the period of the pulsar. If no high energy radiation is present the wrapped data should be more or less uniformly distributed on the circumference of the circle, while a non-uniform distribution should manifest itself during periods of high energy radiation. They found that the first 190 observations could reasonably be assumed to have arisen from a uniform distribution. We now apply to observations 191 through 1250 the concentration CUSUMs from Section 4 of the present paper to detect further changes in concentration. The in-control ARL of the chart is set at 500 observations with reference value $\zeta = 0$ (again, the recommended reference value from (3.5) to enable detection of a 30 degree, i.e. $\pi/6 = 0.52$ radian, rotation) and control limits ± 30.46. The first $m = 50$ observations are used to obtain initial estimates of the required means, variances and covariance of $\sin X$ and $\cos X$.

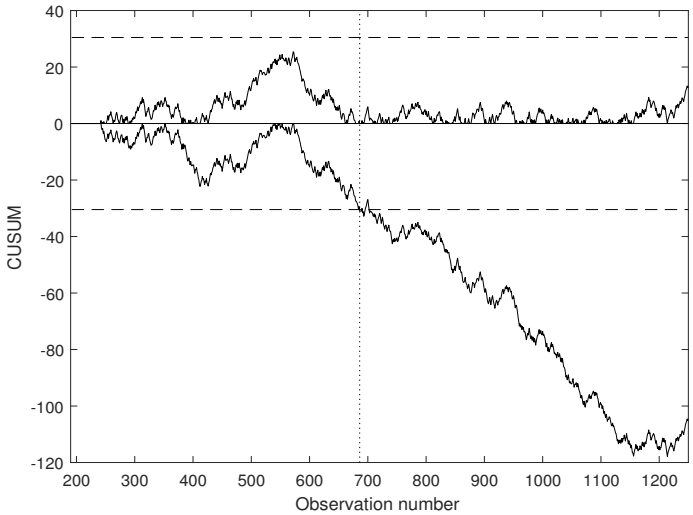

Figure 4: Concentration CUSUM of the pulsar data.

The full extent of the concentration CUSUM, without restarts, is shown in Figure 4. The first signal is at $n = 191 + 495 = 686$ and the changepoint is estimated at $n = 191 + 331 = 522$. The estimated concentration in the segment $[192, 522]$ is 0.35. Thereafter, the lower CUSUM D^- shows a sustained decrease to the end of the data series. In fact, if the CUSUM is restarted at $n = 523$, a changepoint is indicated at $n = 523$. Such a pattern is indicative of a more or less continuous decrease in concentration as the series progresses. The estimated concentration of the observations in the segment $[523, 1250]$ is 0.06, suggesting a uniform distribution in this segment. Hawkins and Lombard [4] applied a retrospective segmentation method to these data. Except for a short segment $[191 - 207]$, which falls within the warmup set used to initiate the CUSUM, the results of the CUSUM analysis agree quite well with their results. The numerical details are shown in Table 7.

Table 7: Pulsar data. Segments delineated by sequential CUSUM and retrospective segmentation.

Retrospective			CUSUM		
segment	$\hat{\nu}$	$\hat{\kappa}$	segment	$\hat{\nu}$	$\hat{\kappa}$
191 – 207	−0.41	1.89			
208 – 573	−1.58	0.35	191 – 522	−1.44	0.35
574 – 1250	—	0.0	523 – 1250	—	0.06

6. SUMMARY

We develop non-parametric rotation invariant CUSUMs for detecting changes in the mean direction and concentration of a circular distribution. The CUSUMs are designed for situations in which the initial mean direction and concentration are unspecified, the objective being to detect a change from the initial values, whatever the latter may be. Monte Carlo simulation results indicate that the CUSUMs have in-control average run lengths that are acceptably close to the nominal values over a wide class of symmetric and asymmetric circular distributions. Two applications of the methodology to data from Health Science and Astrophysics are discussed.

SUPPLEMENTARY MATERIAL

Supplementary material for this publication is available on GitHub at:
https://github.com/cpotgieter/nonparametric-cusums

The supplementary files consist of a pdf document with detailed simulation results, an Excel file with the datasets used in this paper, and the Matlab code for implementing the CUSUM procedures proposed here.

ACKNOWLEDGMENTS

The authors thank two referees for some valuable comments that led to an improved presentation of the results in the paper.

REFERENCES

[1] Abe, T. and Pewsey, A. (2011). Sine-skewed circular distributions, *Statistics Papers*, **52**, 683–707.

[2] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, *Journal of the Royal Statistical Society, Series B*, **65**, 367–389.

[3] Fisher, N. (1993). Statistical Analysis of Circular Data, Cambridge, Cambridge University Press.

[4] Hawkins, D.M. and Lombard, F. (2015). Segmentation of circular data, *Journal of Applied Statistics*, **42**(1), 88–97.

[5] Hawkins, D.M. and Lombard, F. (2017). CUSUM control for data following the von Mises distribution, *Journal of Applied Statistics*, **44**(8), 1319–1332.

[6] Hawkins, D.M. and Olwell, D.H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement*, Springer Verlag, New York.

[7] Hawkins, D.M.; Olwell, D.H. and Wang, B. (2016). http://cran.r-project.org/web/packages/CUSUMdesign/CUSUMdesign.pdf

[8] Helland, I. (1982). Central limit theorems for martingales with discrete or continuous time, *Scandinavian Journal of Statistics*, **9**, 79–94.

[9] Jammalamadaka, S.R. and SenGupta, A. (2001). *Topics in Circular Statistics*, World Scientific Publishing Company, Singapore.

[10] Jones, M.C. and Faddy, M.J. (2003). A skew extension of the t-distribution, with applications, *Journal of the Royal Statistical Society, Series B*, **65**, 159–174.

[11] Keefe, M.J.; Woodall, W.H. and Jones-Farmer, L.A. (2015). The conditional in-control performance of self-starting control charts, *Quality Engineering*, **27**, 488–499.

[12] Knoth, S. (2017). *spc: Statistical Process Control – Collection of Some Useful Functions*, R, package version 0.5.4. https://CRAN.R-project.org/package=spc

[13] Lombard, F.; Hawkins, D.M. and Potgieter, C.J. (2018). Sequential rank CUSUM charts for angular data, *Computational Statistics and Data Analysis*, **105**, 268–279.

[14] Lombard, F. and Maxwell, R.K. (2012). A CUSUM procedure to detect deviations from uniformity in angular data, *Journal of Applied Statistics*, **39**, 1871–1880.

[15] Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics*, John Wiley and Sons, Chichester.

[16] Mathworks: Matlab Version 2016b.

[17] Nolan, J.P. (2015). *Stable Distributions-Models for Heavy Tailed Data*, Birkhäuser, Boston. Note: In progress, Chapter 1 online at academic.2.american.edu/~jpnolan

[18] Page, E.S. (1954). Continuous Inspection Schemes, *Biometrika*, **41**, 100–115.

[19] Pewsey, A. (2000). The wrapped skew-normal distribution on the circle, *Communications in Statistics – Theory and Methods*, **29**(11), 2459–2472.

[20] Saleh, N.A.; Zwetsloot, I.M.; Mahmood, A.M. and Woodall, W.H. (2016). CUSUM charts with controlled conditional performance under estimated parameters, *Quality Engineering*, **28**, 402–425.

[21] Taylor, C.C. (2008). Automatic bandwidth selection for circular density estimation, *Computational Statistics and Data Analysis*, **52**, 3493–3500.

[22] Umbach, D. and Jammalamadaka, S.R. (2011). Building asymmetry into circular distributions, *Statistics and Probability Letters*, **79**, 659–663.

REVSTAT – Statistical Journal
Volume 18, Number 4, October 2020, 483–499

TEXT MINING AND RUIN THEORY: A CASE STUDY OF RESEARCH ON RISK MODELS WITH DEPENDENCE

Authors: Renata G. Alcoforado
– ISEG & CEMAPRE, Universidade de Lisboa,
Portugal
and
Department of Accounting and Actuarial Sciences,
Universidade Federal de Pernambuco,
Brazil
alcoforado.renata@ufpe.br

Alfredo D. Egídio dos Reis
– ISEG & CEMAPRE, Universidade de Lisboa
Portugal
alfredo@iseg.ulisboa.pt

Received: December 2017 Revised: June 2018 Accepted: June 2018

Abstract:

- This paper aims to analyze unstructured data using a text mining approach. The work was motivated in order to organize and structure research in Risk Theory. In our study, the subject to be analyzed is composed by 27 published papers of the risk and ruin theory topic, area of actuarial science. They were coded into 32 categories. For the purpose, all data was analyzed and figures were produced using the software *NVivo 11 plus*. Software *NVivo* is a specialized tool in analyzing unstructured data, although it is commonly used just for qualitative research. We used it for Quali-Quant analysis.

Keywords:

- *big data; unstructured data; text mining; risk theory; ruin theory; dependence modeling.*

AMS Subject Classification:

- 49A05, 78B26.

1. INTRODUCTION

As widely known, Big Data is an area of great development in statistics. We can define Big Data as "a phenomenon defined by the rapid acceleration in the expanding volume of high velocity, complex, and diverse types of data. Big Data is often defined along three dimensions — volume, velocity, and variety" (TechAmerica Foundation's Federal Big Data Commission [44], 2012).

According to Han *et al.* [23] (2012) data mining is the process of mining through large amount of data to extract meaningful information, knowledge. It is also treated by many people as a synonym for knowledge discovery from data, simply KDD.

Text mining in an analogous manner as data mining, aims to extract information from data, but in this case the data comprehend to texts and do it through identification and exploration of interesting patterns (Feldman and Sanger [16], 2006). Accordingly to Aggarwal and Zhai [2] (2012), the primary goal of text mining is analyzing information to discover patterns, going beyond information access to further help users analyze and digest information and facilitate decision making.

Text mining has been used as a form to extract knowledge from text, it has been applied to social media (see Corley *et al.* [14] (2010), Zeng *et al.* [45] (2010), Maynard *et al.* [33] (2012), He *et al.* [24] (2013), Mostafa [34] (2013)), health science (see Chen *et al.* [7] (2005), Cohen and Hersh [8] (2005), Collier *et al.* [9] (2008), Zweigenbaum *et al.* [46] (2007), Hirschman *et al.* [25] (2012)), in social sciences (see Peng *et al.* [37] (2012)) and other fields.

Francis and Flynn [17] (2010) show that text mining can be used to generate new information from the unstructured text data. Text mining can also be used to extract quantitative information, as Kim and Jun [28] (2015) did to obtain a Gaussian copula regression model.

This paper was motivated to organize and structure our research in Risk Theory, the goal is to study this thematic in the most embracing, as well as profoundly, way. First, we need to know what has been studied in this topic so we selected the papers in the area and we aimed to extract knowledge from this database. We uploaded it in the software so it can be read for us.

The software can recognize patterns and present pertinent connections that otherwise we would miss and also spot the most pertinent papers in the area. The *NVivo* is usually used to qualitative analysis, but as Kim and Jun [28] (2015) did in their paper, we also did a quali-quant analysis that evidence the ability to use this software for quantitative analysis and we expect that other researchers will do the same.

This paper is organized as follows: In Section 2 we speak about the collected data under analysis. Section 3 is about the coding of the data, the coding matrix, the relationship between the nodes, that is, we plotted the nodes hierarchically. Then, we present the cluster analysis for the nodes and the sources (papers), the comparison diagrams and, to finalize, a structural matrix. To conclude, in Section 4 we write some final remarks.

2. THE DATA

Our data is composed by published scientific papers. For this particular study we chose a limited set, enough for our immediate purpose. Working with reference material in many aspects is no different from working with any other form of text. As it is in the form of research literature, it will contain author defined sections that can be compared across the references. Also, keywords are available. Therefore we can consider that this type of data is also likely to be more structured than information from an interview [see for example Bazeley and Jackson [5] (2013)].

We chose a set of 27 scientific papers to be analyzed. We uploaded these 27 papers in the platform, coded and then analyzed all data. These papers are references for a particular research project in development in risk theory. These papers are: Afonso *et al.* [1] (2017), Ammeter [3] (1948); Asmussen and Albrecher [4] (2010); Bergel and Egídio dos Reis [6] (2016); Constantinescu *et al.* [11] (2011); Constantinescu *et al.* [12] (2012); Constantinescu *et al.* [10] (2016); Czado *et al.* [15] (2011); Frees and Wang [20] (2006); Frees *et al.* [18] (2011); Frees *et al.* [19] (2016); Garrido *et al.* [21] (2016); Gschlößl and Czado [22] (2017); Jasiulewicz [26] (2001); Jørgensen and Paes De Souza [27] (1994); Krämer *et al.* [29] (2013); Kreer *et al.* [30] (2015); Li *et al.* [31] (2015); Maume-Deschamps *et al.* [32] (2017); Ni *et al.* [35] (2014a); Ni *et al.* [36] (2014b); Quijano Xacur and Garrido [38] (2015); Renshaw [39] (1994); Rolski *et al.* [40] (1999); Schulz [41] (2013); Shi *et al.* [42] (2015) and Song *et al.* [43] (2009).

Using the software, the first task we took was to build a *word cloud* composed by the the most pertinents words in our entire data base to use in our study. After removing all the verbs, articles and non-meaningful wording, the words are then gathered according to their *stem*, then search the frequency of words, making possible to obtain the *cloud* as shown in Figure 1. It is important to point out that the *word cloud* shows the essence of the data base, where the size matters.

In the coding we will present the figures in the order in which we elaborated them. First, as prior mentioned is the word cloud in Figure 1, which will contribute on the creation of the categories. Then, in Figure 2 is presented the Word Tree for the node "Aggregate claims model", that we obtain when coding the database.

In Figure 3 is a chart node coding for "Claim Severity", that derives from this specific category after coding the database. In sequence, we desire to see how each one of the categories fit hierarchically in the entire group of categories and also how they connect with one another, therefore we present them in Figure 4 and in Figure 5, respectively.

Then, we analyze first the categories and then the sources using cluster analysis, for the Cluster analysis of the categories we exhibit two figures, in Figure 6 is the circle graph and in Figure 7 is the dendrogram. As a result of the cluster analysis for the sources we display one dendrogram in Figure 8.

Posteriorly, we conclude from the cluster analysis and from the coding matrix the categories that are interesting to compare, hence we present in Figure 9 two comparison diagrams. Finally, we present a summarized framework matrix.

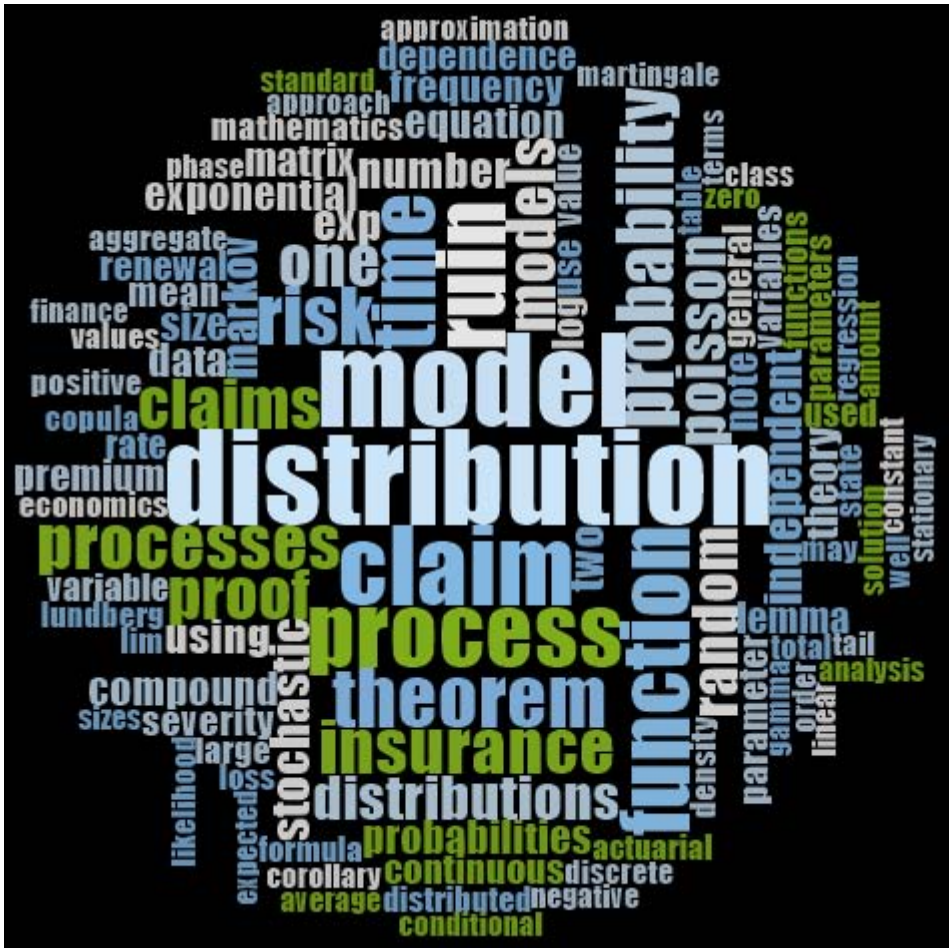

Figure 1: Word Cloud.

3. THE CODING AND ANALYSIS

A code is an abstract representation of a case. Corbin and Strauss [13] (2008) say that we can think of coding as "mining" the data, digging beneath the surface to discover the hidden treasures contained within data. Accordingly to Bazeley and Jackson [5] (2013), coding is one of several methods of working with and building knowledge about data.

Data mining assumes that the data is already stored in an structured way, whereas text mining assumes that the data is unstructured and still needs coding, accordingly to Feldman and Sanger [16] (2006).

In sequence, all the unstructured data was coded, building categories, that is, we put in each one of the categories the respective parts from text to be able to analyze it in a mathematical way. In other words, after coding we get a structure to be able to analyze with clusters and matrices. With that, we can plot the data now, this was not possible before. The categories were selected after extensive reading and observing the *word cloud*.

In our particular analysis the *codes* are: Actuarial; Aggregate Claims Model; Claim Frequency; Claim Severity; Compound Poisson; Conditional; Copulas; Covariates; Depen-

dence; Exponential; Formula; Function; Gamma; Independence; Insurance; Joint Distribution; Loss; Markov; Martingale; Mixed-Poisson; Parameters; Prediction; Premium; Randomness; Regression; Renewal; Risk Theory; Ruin Probability; Simulation; Spatial; Stationary and Stochastic Process.

After the code and data organization, for each category is plotted a *word tree* to see the connection from that word (or expression) in the sentence where it belongs. An example is given in Figure 2, we can observe how the "aggregate claims model" fits in the sentence. In this case, authors are talking mostly about the "dependent" and the "independent" aggregate claims model. They also talk about the "issue of dependence", "assumption of independence", "the marginal distributions", "the structure" and "the effect of extending" the aggregate claims model.

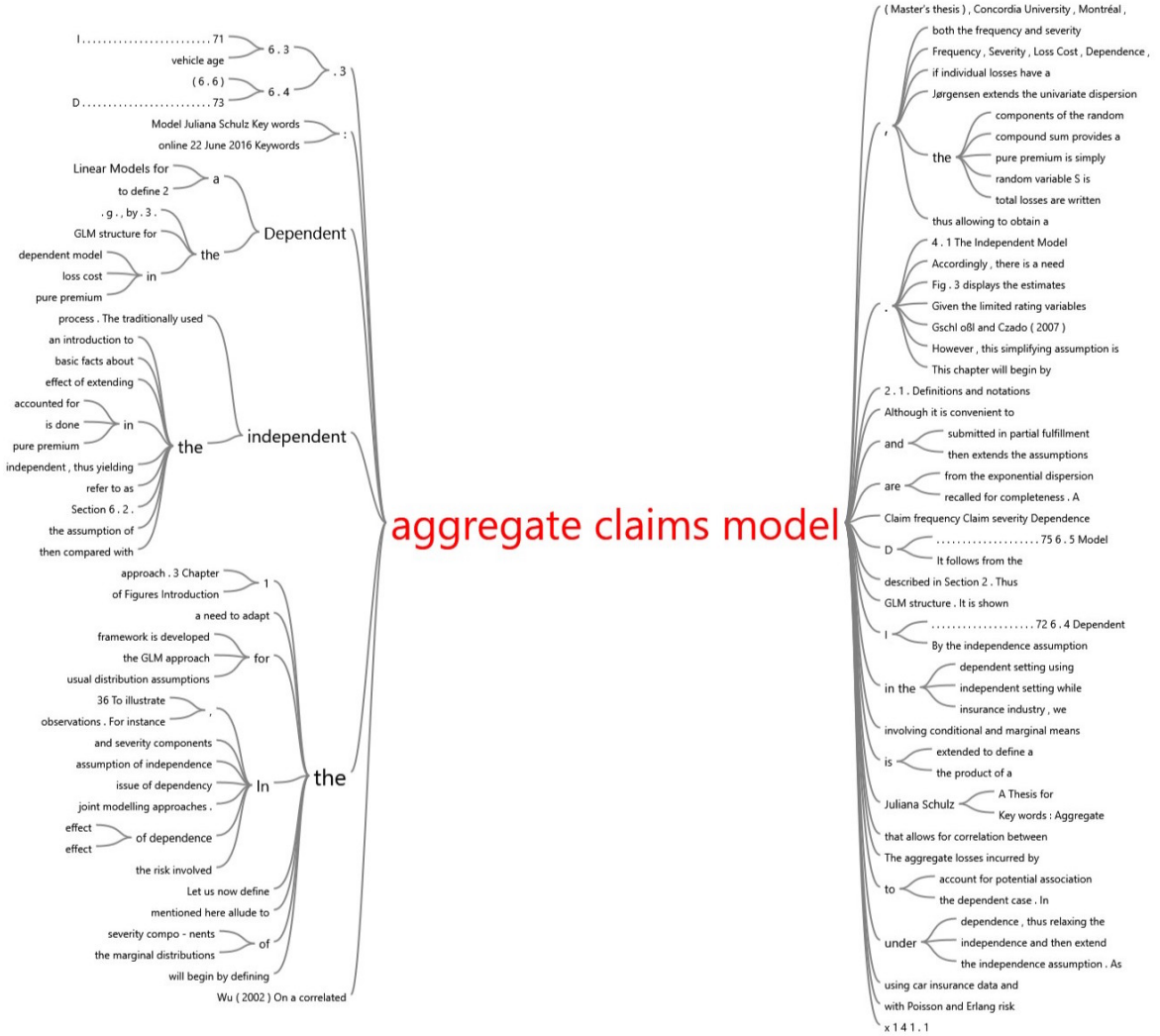

Figure 2: Word Tree – Aggregate Claims Model.

For every category is plotted a chart node coding that presents the sources from our database that address the most and the importance that each paper from the database gives to that code. In Figure 3 we can observe which authors and in which papers the category "Claim Severity" is included. So, we can distinguish the author Constantinescu from our database since four of the papers that address the most to "Claim Severity" are written by her, including the first one.

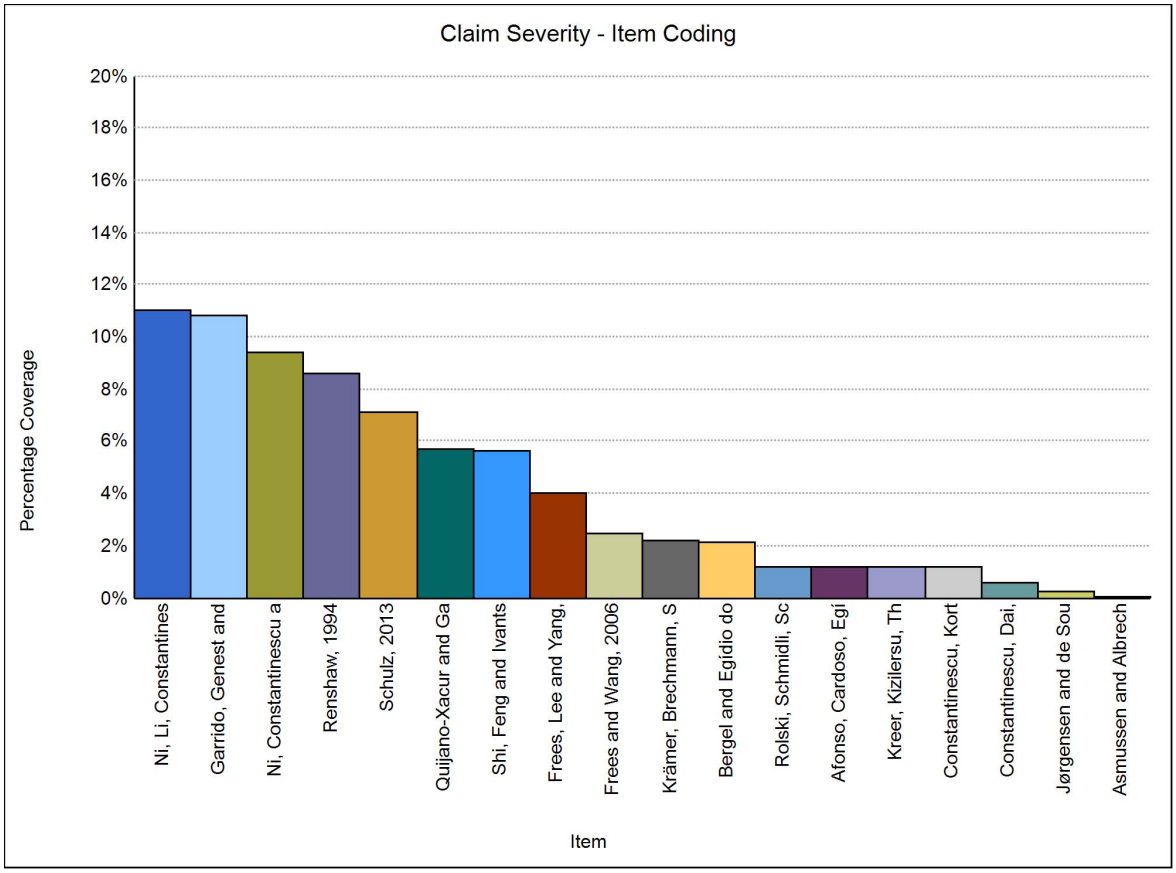

Figure 3: Chart Node Coding – Claim Severity.

We plotted the *nodes*, or categories, hierarchically presented in Figure 4 to observe which categories are most frequent, the most important among the data available.

In Figure 4 we can observe how the category that have the most importance is "Function", then followed by "Exponential" and then "Insurance". Another fact to point out is how "Claim frequency" is more important hierarchically than "Claim Severity", which is in line with the fact the most motor insurance models don't consider the Claim severity.

The authors when trying to capture the dependence between claim frequency and severity can use a "Regression" approach in which use one variable as a "Covariate" in the other regression or they can use a "Copula" approach. In the Figure 4 we can see how although they are almost the same size, "Regression" is still a bigger category.

Also, they can use a distribution to model when trying to capture the dependence, that distribution can be in hierarchically order: "Exponential", "Compound Poisson", "Gamma" and "Mixed Poisson". We can observe that stochastic processes are also very used. So, we can point out the following categories that fits into that description: "Markov", "Martingale", "Stationary" and "Stochastic Process" itself. As our database consists in authors that are trying to capture dependence between the two variables in some way, it is also important to mention how the code "Dependence" is more relevant then "Independence".

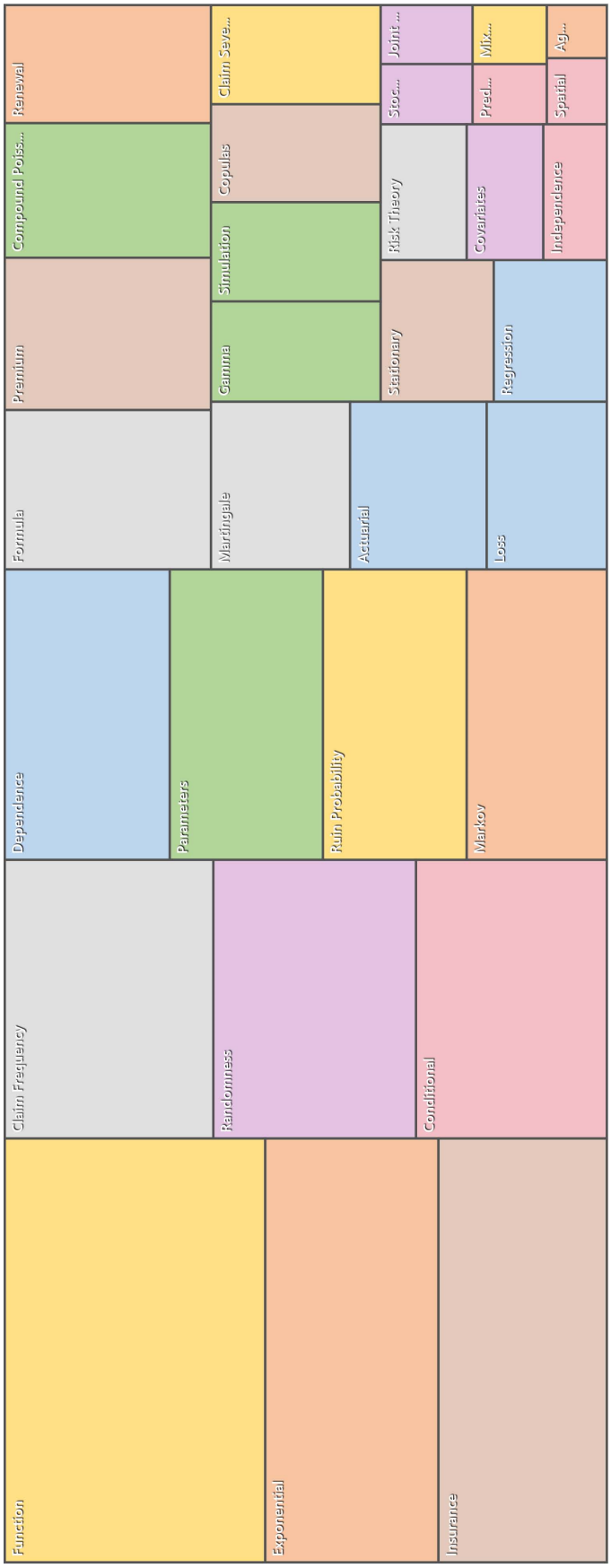

Figure 4: Nodes Hierarchically.

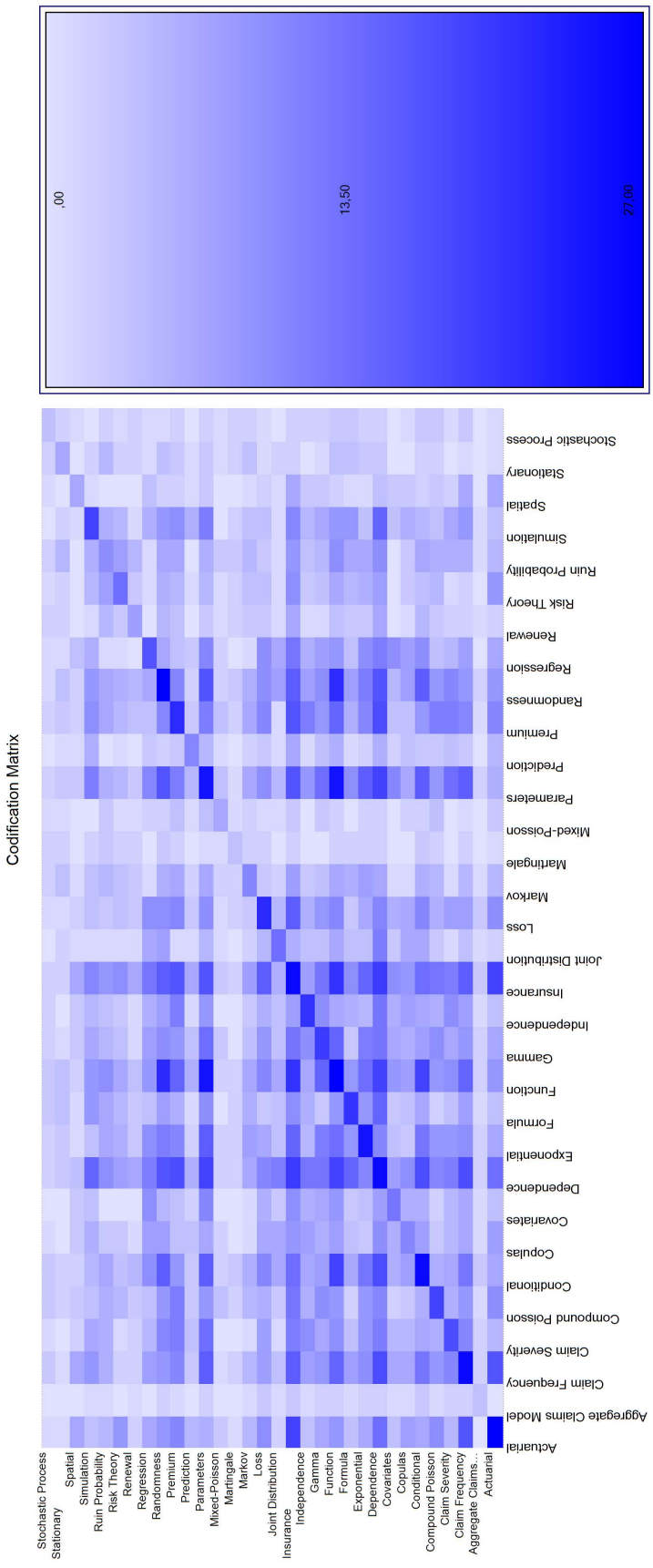

Figure 5: Coding Matrix – Heat.

Our target as a research topic is to be able to calculate "Premium" and "Ruin Probability", although both categories have almost the same size it is important to mention that in one hand 22 out of the 27 papers address "Premium", while 10 papers address to "Ruin Probability". On the other hand, from those 10, there are 614 coded references for "Ruin Probability" and in those 22 papers there are 464 coded references for "Premium". To conclude the analysis of Figure 4, the method used to calculate these two quantities "Ruin Probability" and "Premium" as mentioned above can also be theoretical through "Formulas" or numerical through "Simulation". The former is the one that is the most sought in this database.

After, we constructed the *coding matrix* presented on the Figure 5, which shows both in numbers and in graph what is the relationship between the coded categories. In this matrix the colors are meaningful, the darker the color, the more codes are represented in the other coded references. In percentage, until 1% is white, from 1% to 10% is light blue, from 10% to 20% is a shade darker as we can observe between "Claim Frequency" and "Spatial", from 20% to 30% is another shade darker as we can observe between the node "Premium" in the row and column. The darkest blue means that it is between 30% and 40% as in "Exponential" in the row and column.

Although we may think to be symmetric, this matrix coding is not symmetric. It would be if we used the numbers, but the numbers are not as important as the percentage of the total for that category. Each cell content is the column percentage of coded references, and it is not symmetric because of the way the data was collected. That is, the papers are about dependence between the claim frequency and severity random variables, as a consequence the codes are going to reference more dependence than the other way around.

So, for instance when we consider "Copulas" and "Dependence", dependence is in 10.88% of the coded references from Copulas, and Copulas are in only 4.68% coded references of the Dependence category. Another case is "Premium" and "Insurance", "Insurance" are represented in 10.41% of the "Premium" coded references while "Premium" are represented in 6.93% of the "Insurance" category.

The *cluster analysis* was afterwards performed in cluster by word similarity using Pearson's correlation coefficient as the similarity measure. We made it for both the categories and sources to see how they relate. The cluster analysis for the nodes is presented in Figures 6 and 7, a circle graph in Figure 6 and a dendrogram in Figure 7.

In the circle graph in Figure 6 the colors represent the clusters and the lines represent the connection between the nodes, the more and the thicker are the lines, the higher is Pearson's correlation coefficient. We can observe an asymmetry to the right that means that the nodes on the right have a higher correlation.

Referring now to Figure 7, in this dendrogram we can observe 10 clusters for the 32 nodes represented by the colors and the branches. The following categories before mentioned for stochastic processes are in one cluster together with "Claim Frequency", since the claim frequency is usually considered as an stochastic process. The coefficient between "Stochastic Process" and "Martingale" is 0.815.

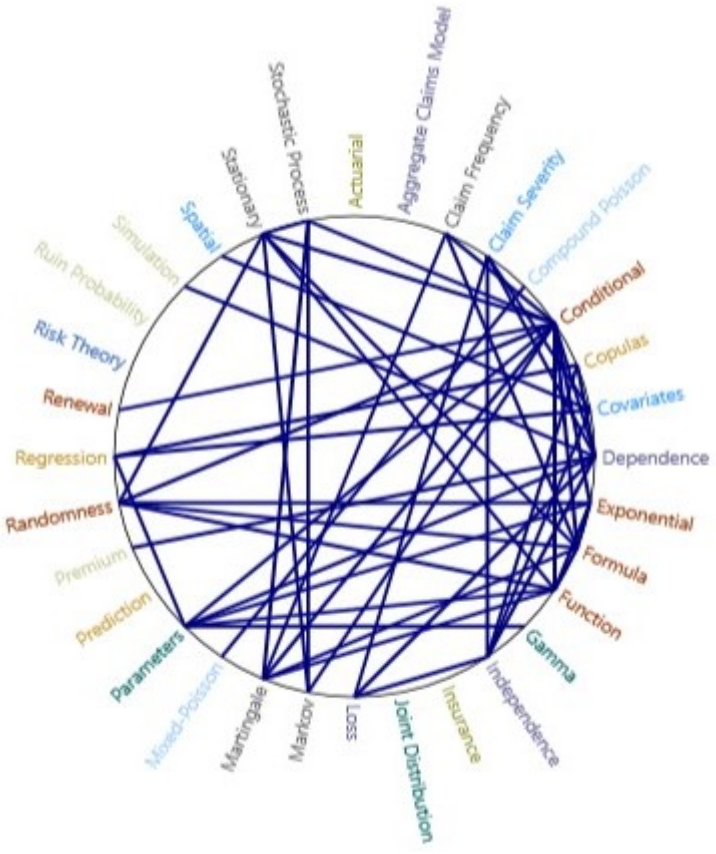

Figure 6: Cluster Analysis of the Nodes – Circle Graph.

The cluster with the highest similarity is the one that comprehends "Function", "Conditional", "Exponential", "Formula", "Randomness" and "Renewal", the coefficient between "Function" and "Conditional" is 0.848, between "Formula" and "Conditional" is 0.820, between "Exponential" and "Conditional" is 0.817.

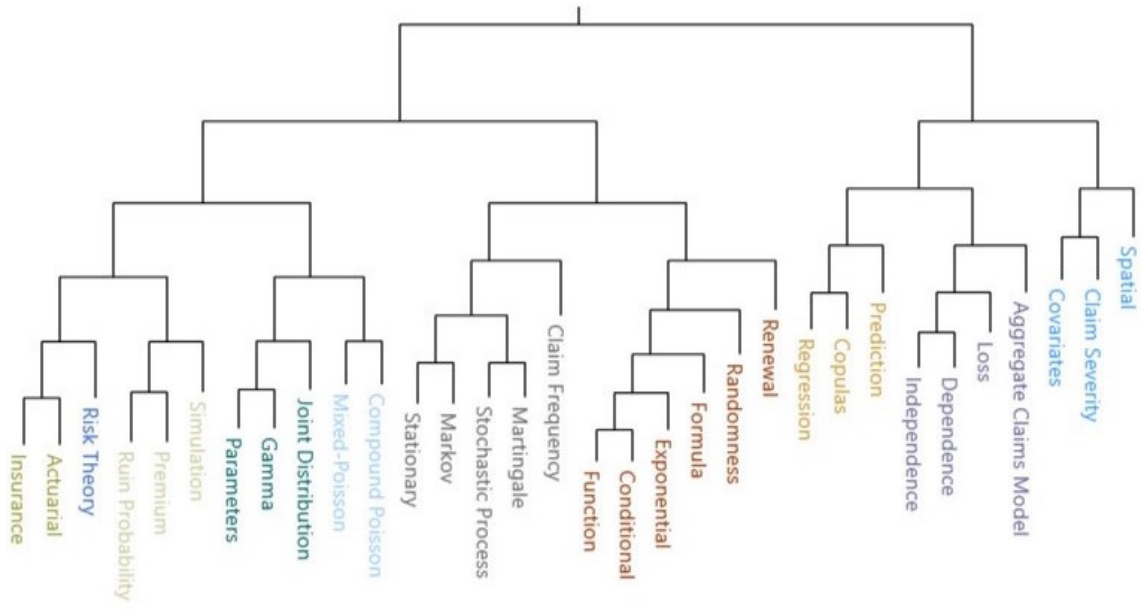

Figure 7: Cluster Analysis of the Nodes – Dendrogram.

The categories "Independence" and "Dependence" present a 0.806 coefficient and are clustered together. "Ruin Probability" and "Premium" present a 0.645, both are clustered with "Simulation". "Claim Severity" and "Claim Frequency" a 0.521, "Claim Severity" is in a cluster with "Covariates" and "Spatial" while "Claim Frequency" is in the first cluster mentioned. "Simulation" and "Formula" a 0.616 and are in different clusters. And finally, "Copulas" and "Regression" a 0.793 and both are in the same cluster (the yellow).

Cluster analysis for the sources from out database was also plotted and is presented in Figure 8. From Figure 8 we can observe the clusters accordingly to colors and branches. There are three big clusters that comprehend 18 papers. The clusters are build using the complete linkage hierarchical clustering algorithm, also known as farthest neighbor clustering.

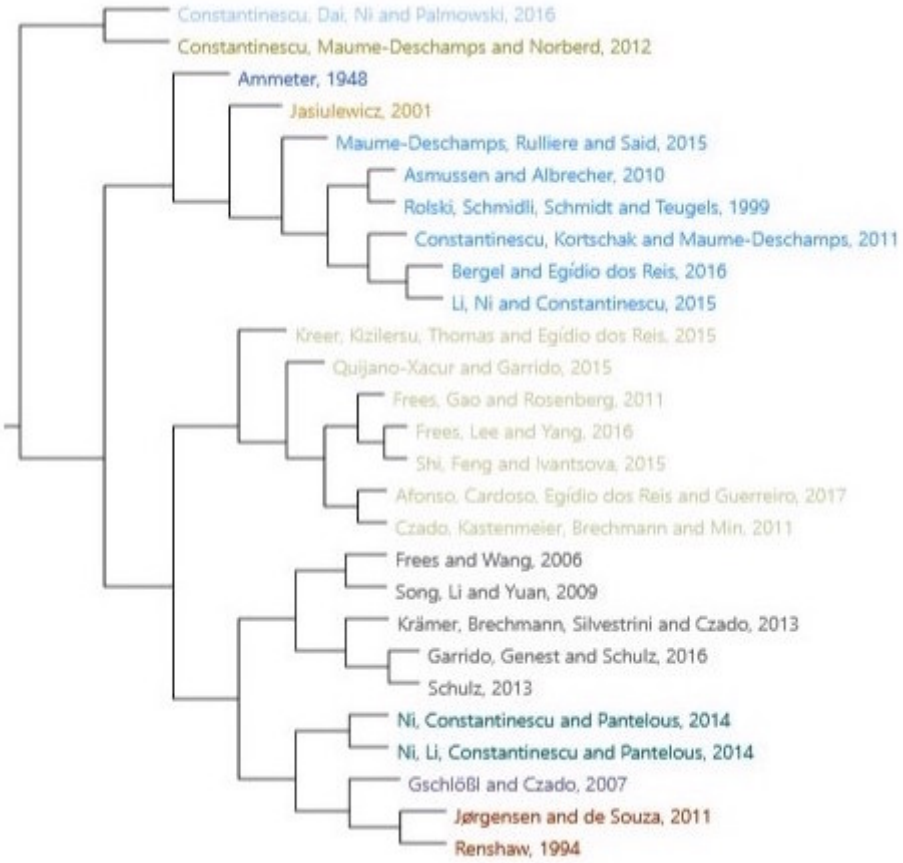

Figure 8: Cluster Analysis of the Sources – Dendrogram.

The higher correlation coefficients are in the middle cluster, the green one, between Frees *et al.* [19] (2016) and Shi *et al.* [42] (2015) is 0.91, between Shi *et al.* [42] (2015) and Czado *et al.* [15] (2011) is 0.88 and if we consider Frees *et al.* [18] (2011), Frees *et al.* [19] (2016), Shi *et al.* [42] (2015), Afonso *et al.* [1] (2017) and Czado *et al.* [15] (2011) the correlation coefficient between two of them at a time goes from 0.83 to 0.91.

The blue cluster groups the papers Maume-Deschamps *et al.* [32] (2017), Asmussen and Albrecher [4] (2010), Rolski *et al.* [40] (1999), Constantinescu *et al.* [11] (2011), Bergel and Egídio dos Reis [6] (2016) and Li *et al.* [31] (2015), the coefficients between those vary from 0.65 (the farthest sources, Maume-Deschamps *et al.* [32] (2017) and Li *et al.* [31] (2015)) to 0.86, coefficient between Rolski *et al.* [40] (1999) and Asmussen and Albrecher [4] (2010).

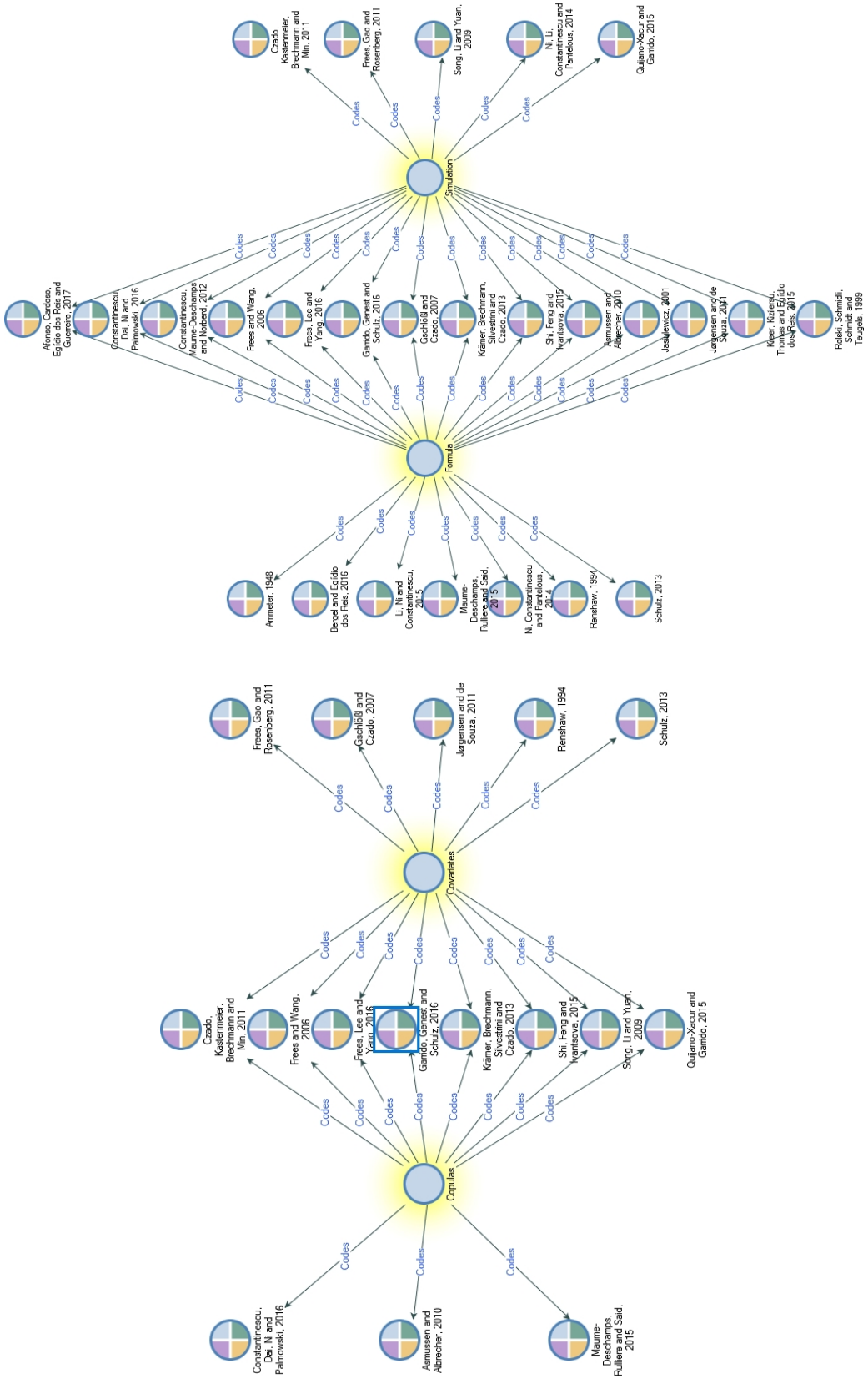

Figure 9: Comparison Diagrams.

We also plotted comparison diagrams. We present in Figure 9 the diagram comparing "Copulas and Covariates" on the left and the diagram comparing "Formula and Simulation" on the right.

From the comparison diagram between "Copulas" and "Covariates" presented on the left of Figure 9, we can observe that from the 16 references coded for this two categories, Constantinescu *et al.* [10] (2016), Asmussen and Albrecher [4] (2010) and Maume-Deschamps *et al.* [32] (2017) work only with copulas. Frees *et al.* [18] (2011), Gschlößl and Czado [22] (2017), Jørgensen and Paes De Souza [27] (1994), Renshaw [39] (1994) and Schulz [41] (2013) use covariates on their papers. The other eight papers use them both. Both copulas and covariates are methods to try to capture the dependence between the claim frequency and severity variables.

A second comparison diagram is presented on the right of Figure 9, comparing "Formula" and "Simulation". We can point out that there are 26 papers, since those are the approach that the authors can follow to calculate the ruin probability or/and premium. So some authors used Formula, Ammeter [3] (1948), Bergel and Egídio dos Reis [6] (2016), Li *et al.* [31] (2015), Maume-Deschamps *et al.* [32] (2017), Ni *et al.* [36] (2014b), Renshaw [39] (1994) and Schulz [41] (2013).

On the other hand the following authors from our database used "Simulation": Czado *et al.* [15] (2011), Frees *et al.* [18] (2011), Song *et al.* [43] (2009), Ni *et al.* [35] (2014a) and Quijano Xacur and Garrido [38] (2015). The remaining authors used both. It is worth to comment that Ni, Constantinescu and Pantelous published two papers in 2014, one using "Formula" and the other "Simulation".

To finalize, we built the *framework matrix* where each row shows each paper and each columns the category mentioned above, in order to identify subtle connections which can allow a thorough and rigorous study. In Table 1 is presented a summarized version of this framework matrix, in which the first column presents the name of the cases in study, the following columns are 12 different categories and we mark the cells with an "×" to represent the coded categories to each source.

The categories presented in the Table 1 can be shortly defined as:

A – *Actuarial/Actuaries:* Study of risk/ Scientist of risk;
B – *Aggregate Claims Model:* Model of claims that considers both and all together frequency and severity of claims;
C – *Claim Frequency:* Frequency or count of claims in the insurance company;
D – *Claim Severity:* Severity or amount of claims in the insurance company;
E – *Compound Poisson:* Distribution for the aggregate claim amounts used to model the frequency and severity of claims on aggregate;
F – *Copulas:* Is a multivariate probability tool used to capture dependence;
G – *Joint Distribution:* Is the distribution of the two or more variables calculated together, jointly;
H – *Premium:* Amount paid by the insured for the insurance policy;
I – *Regression:* Multiple Regression models, can also be GLM's;
J – *Ruin Probability:* Probability of ruin of an insurance portfolio or company;
K – *Simulation:* When simulating different scenarios on a software;
L – *Stochastic Process:* Random Processes used for the claim frequency, in this case it is divided into Markov and Martingale processes.

Table 1: Summarized Framework Matrix.

Source	A	B	C	D	E	F	G	H	I	J	K	L
Afonso *et al.* [1] (2017)			×	×	×			×		×	×	×
Ammeter [3] (1948)	×							×				
Asmussen and Albrecher [4] (2010)	×			×	×	×	×	×	×	×	×	×
Bergel and Egídio dos Reis [6] (2016)				×	×					×		
Constantinescu *et al.* [11] (2011)				×						×		×
Constantinescu *et al.* [12] (2012)	×		×		×			×		×		×
Constantinescu *et al.* [10] (2016)				×		×		×		×	×	×
Czado *et al.* [15] (2011)	×		×		×	×	×	×	×		×	
Frees and Wang [20] (2006)	×		×	×		×	×		×		×	×
Frees *et al.* [18] (2011)	×						×		×		×	
Frees *et al.* [19] (2016)	×	×	×	×		×	×	×	×		×	
Garrido *et al.* [21] (2016)	×	×	×	×	×			×	×		×	
Gschlößl and Czado [22] (2017)	×		×		×			×	×		×	×
Jasiulewicz [26] (2001)	×							×		×		
Jørgensen and Paes De Souza [27] (1994)				×	×		×	×	×		×	×
Krämer *et al.* [29] (2013)	×		×	×		×	×		×		×	
Kreer *et al.* [30] (2015)	×			×							×	
Li *et al.* [31] (2015)					×			×		×		×
Maume-Deschamps *et al.* [32] (2017)	×					×	×			×		
Ni *et al.* [35] (2014a)			×	×				×				
Ni *et al.* [36] (2014b)			×	×				×				
Quijano Xacur and Garrido [38] (2015)			×	×	×			×	×		×	
Renshaw [39] (1994)			×	×				×	×			
Rolski *et al.* [40] (1999)	×			×	×		×	×	×	×	×	×
Schulz [41] (2013)		×	×	×	×		×	×	×			
Shi *et al.* [42] (2015)	×		×	×	×	×	×	×	×		×	×
Song *et al.* [43] (2009)					×	×			×		×	

A: Actuarial; B: Aggregate Claims Model; C: Claim Frequency; D: Claim Severity; E: Compound Poisson; F: Copulas; G: Joint Distribution; H: Premium; I: Regression; J: Ruin Probability; K: Simulation; L: Stochastic Process

4. FINAL REMARKS

Our source intended to talk about the calculation of premiums and ruin probabilities for insurance application, also to associate the claim frequency with their severity. Some authors use copulas, other use covariates in a regression model, and other try to find a distribution that can capture that dependence.

We were motivated to organize and structure our research in Risk Theory and as presented in the paper, we were able to achieve this goal. And beyond that, after a deeper study we extracted quantitative knowledge from the database.

We obtained results that made possible to know which authors were the most important for each category as we saw in Figure 3. It was shown in Figure 4 which categories matters the most for this data base and in which ways, hierarchically, the authors approach the subject.

Additionally, in Figure 5 we presented in percentage the relationship between the nodes. At last, from the cluster analysis shown in Figures 6, 7 and 8, we captured relevant patterns among the nodes and the authors.

The result showed to be interesting to compare respective categories and plot comparison diagrams, for instance, comparing *Dependence* with *Independence*; *Simulation* with *Formula*; *Copulas* with *Covariates*; *Regression* with *Copulas*; *Claim Severity* with *Claim Frequency* among others.

To finalize, this text mining analysis presents a current overview of the knowledge in the field of Ruin Theory research. In addition, a conceptual framework was presented and the key categories for the dependency model were identified. It is presumed that this study will motivate future research on the impact of dependence between these two variables on risk models, bringing to light the categories and links that need further investigation.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support from FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through Project CEMAPRE – UID/MULTI/00491/2013 financed by FCT/ MCTES through national funds.

REFERENCES

[1] Afonso, L.B.; Cardoso, R.M.R. and Egídio dos Reis, A.D. (2017). Measuring the impact of a bonus-malus system in finite and continuous time ruin probabilities for large portfolios in motor insurance, *ASTIN Bulletin*, **47**(2), 417–435.

[2] Aggarwal, C.C. and Zhai, C., (Eds.) (2012). *Mining Text Data*, Springer Science & Business Media, 1st edition.

[3] Ammeter, H. (1948). A generalization of the collective theory of risk in regard to fluctuating basic-probabilities, *Scandinavian Actuarial Journal*, **1948**(1–2), 171–198.

[4] Asmussen, S. and Albrecher, H. (2010). *Ruin Probabilities*, World Scientific, Singapore, second edition.

[5] Bazeley, P. and Jackson, K. (2013). *Qualitative data analysis with Nvivo*, Sage Publications, second edition.

[6] Bergel, A.I. and Egídio dos Reis, A.D. (2016). Ruin problems in the generalized Erlang(n) risk model, *European Actuarial Journal*, **6**(1), 257–275.

[7] Chen, H.; Fuller, S.S.; Friedman, C. and Hersh, W. (Eds.) (2005). *Medical Informatics: Knowledge Management and Data Mining in Biomedicine*, "Integrated Series in Information Systems", volume 8.

[8] Cohen, M.A. and Hersh, R.W. (2005). A survey of current work in biomedical text mining, *Brief Bioninform*, **6**(1), 57–71.

[9] Collier, N.; Doan, S.; Kawazoe, A.; Goodwin, R.M.; Conway, M.; Tateno, Y.; Ngo, Q.H.; Dien, D.; Kawtrakul, A.; Takeuchi, K.; Shigematsu, M. and Taniguchi, K. (2008). BioCaster: Detecting public health rumors with a Web-based text mining system, *Bioinformatics*, **24**(24), 2940–2941.

[10] Constantinescu, C.; Dai, S.; Ni, W. and Palmowski, Z. (2016). Ruin probabilities with dependence on the number of claims within a fixed time window, *Risks*, **4**(2), 17.

[11] Constantinescu, C.; Kortschak, D. and Maume-Deschamps, V. (2011). Ruin probabilities in models with a Markov chain dependence structure, *Scandinavian Actuarial Journal*, **1238**(December), 1–24.

[12] Constantinescu, C.; Maume-Deschamps, V. and Norberg, R. (2012). Risk processes with dependence and premium adjusted to solvency targets, *European Actuarial Journal*, **2**(1), 1–20.

[13] Corbin, J. and Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, Sage Publications, USA, 3rd edition.

[14] Corley, C.D.; Cook, D.J.; Mikler, A.R. and Singh, K.P. (2010). Text and structural data mining of influenza mentions in web and social media, *International Journal of Environmental Research and Public Health*, **7**(2), 596–615.

[15] Czado, C.; Kastenmeier, R.; Brechmann, E.C. and Min, A. (2011). A mixed copula model for insurance claims and claim sizes, *Scandinavian Actuarial Journal*, **1238**(January), 1–28.

[16] Feldman, R. and Sanger, J. (2006). *The Text Mining Handbook*, Cambridge University Press, New York, USA.

[17] Francis, L. and Flynn, M. (2010). Text mining handbook. In *Casualty Actuarial Society E-Forum*, 1–61.

[18] Frees, E.W.; Gao, J. and Rosenberg, M.A. (2011). Predicting the frequency and amount of health care expenditures, *North American Actuarial Journal*, **15**(3), 377–392.

[19] Frees, E.W.; Lee, G. and Yang, L. (2016). Multivariate frequency-severity regression models in insurance, *Risks*, **4**(1), 4.

[20] Frees, E.W. and Wang, P. (2006). Copula credibility for aggregate loss models, *Insurance: Mathematics and Economics*, **38**(2), 360–373.

[21] Garrido, J.; Genest, C. and Schulz, J. (2016). Generalized linear models for dependent frequency and severity of insurance claims, *Insurance: Mathematics and Economics*, **70**, 205–215.

[22] Gschlößl, S. and Czado, C. (2007). Spatial modelling of claim frequency and claim size in non-life insurance, *Scandinavian Actuarial Journal*, **2007**(3), 202–225.

[23] Han, J.; Kamber, M. and Pei, J. (2012). *Data Mining. Concepts and Techniques*, Elsevier, Waltham, USA, third edition.

[24] He, W.; Zha, S. and Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry, *International Journal of Information Management*, **33**, 464–472.

[25] Hirschman, L.; Burns, G.A.P.C.; Krallinger, M.; Arighi, C.; Cohen, K.B.; Valencia, A.; Wu, C.H.; Chatr-Aryamontri, A.; Dowell, K.G.; Huala, E.; Lourenço, A.; Nash, R.; Veuthey, A.L.; Wiegers, T. and Winter, A.G. (2012). Text mining for the biocuration workflow, *Database*, **2012**(November), 1–10.

[26] Jasiulewicz, H. (2001). Probability of ruin with variable premium rate in a Markovian environment, *Insurance: Mathematics and Economics*, **29**(2), 291–296.

[27] Jørgensen, B. and Paes De Souza, M.C. (1994). Fitting Tweedie's compound poisson model to insurance claims data, *Scandinavian Actuarial Journal*, **1994**(1), 69–93.

[28] Kim, J.M. and Jun, S. (2015). Graphical causal inference and copula regression model for apple keywords by text mining, *Advanced Engineering Informatics*, **29**(4), 918–929.

[29] Krämer, N.; Brechmann, E.C.; Silvestrini, D. and Czado, C. (2013). Total loss estimation using copula-based regression models, *Insurance: Mathematics and Economics*, **53**(3), 829–839.

[30] Kreer, M.; Kizilersü, A.; Thomas, A.W. and Egídio dos Reis, A.D. (2015). Goodness-of-fit tests and applications for left-truncated Weibull distributions to non-life insurance, *European Actuarial Journal*, **5**(1), 139–163.

[31] Li, B.; Ni, W. and Constantinescu, C. (2015). Risk models with premiums adjusted to claims number, *Insurance: Mathematics and Economics*, **65**(2015), 94–102.

[32] Maume-Deschamps, V.; Rullière, D. and Said, K. (2017). Impact of dependence on some multivariate risk indicators, *Methodology and Computing in Applied Probability*, **19**(2), 395–427.

[33] Maynard, D.; Bontcheva, K. and Rout, D. (2012). Challenges in developing opinion mining tools for social media, *LREC 2012*, pp. 15–22.

[34] Mostafa, M.M. (2013). More than words: Social networks' text mining for consumer brand sentiments, *Expert Systems with Applications*, **40**(10), 4241–4251.

[35] Ni, W.; Constantinescu, C. and Pantelous, A.A. (2014a). Bonus-Malus systems with Weibull distributed claim severities, *Annals of Actuarial Science*, **8**(02), 217–233.

[36] Ni, W.; Li, B.; Constantinescu, C. and Pantelous, A.A. (2014b). Bonus-Malus systems with hybrid claim severity distributions, *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*, pp. 1234–1244.

[37] Peng, T.-Q.; Zhang, L.; Zhong, Z.-J. and Zhu, J.J. (2012). Mapping the landscape of Internet Studies: Text mining of social science journal articles 2000–2009, *New Media & Society*, **15**(5), 644–664.

[38] Quijano Xacur, O.A. and Garrido, J. (2015). Generalised linear models for aggregate claims: to Tweedie or not? *European Actuarial Journal*, **5**(1), 181–202.

[39] Renshaw, A.E. (1994). Modelling the claims process in the presence of covariates, *ASTIN Bulletin*, **24**(2), 265–285.

[40] Rolski, T.; Schmidli, H.; Schmidt, V. and Teugels, J. (1999). *Stochastic Processes for Insurance and Finance*, John Wiley & Sons, West Sussex, England.

[41] Schulz, J. (2013). *Generalized Linear Models for a Dependent Aggregate Claims Model*, PhD thesis, Concordia University, Montréal, Quebec, Canada.

[42] Shi, P.; Feng, X. and Ivantsova, A. (2015). Dependent frequency-severity modeling of insurance claims, *Insurance: Mathematics and Economics*, **64**, 417–428.

[43] Song, P.X.-K.; Li, M. and Yuan, Y. (2009). Joint regression analysis of correlated data using Gaussian copulas, *Biometrics*, **65**(1), 60–68.

[44] TechAmerica Foundation's Federal Big Data Commission (2012). Demystifying Big Data: A Practical Guide to Transforming the Business of Government, Technical report, TechAmerica Foundation's. Retrieved July 10, 2017, from
https://www.attain.com/sites/default/files/take-aways-pdf/Solutions_Demystifying Big Data - A Practical Guide To Transforming The Business Of Government.pdf

[45] Zeng, D.; Chen, H.; Lusch, R. and Li, S.H. (2010). Social media analytics and intelligence, *IEEE Intelligent Systems*, **25**(6), 13–16.

[46] Zweigenbaum, P.; Demner-fushman, D.; Yu, H. and Cohen, K.B. (2007). Frontiers of biomedical text mining: Current progress, *Briefings in Bioinformatics*, **8**(5), 358–375.

REVSTAT – Statistical Journal
Volume 18, Number 4, October 2020, 501–520

DISSECTING THE MULTIVARIATE EXTREMAL INDEX AND TAIL DEPENDENCE

Authors: Helena Ferreira
– Centro de Matemática e Aplicações (CMA-UBI), Universidade da Beira Interior, Avenida Marquês d'Avila e Bolama, 6200-001 Covilhã, Portugal
helenaf@ubi.pt

Marta Ferreira
– Center of Mathematics of Minho University, Center for Computational and Stochastic Mathematics of University of Lisbon, Center of Statistics and Applications of University of Lisbon, Portugal
msferreira@math.uminho.pt

Received: April 2017 Revised: February 2018 Accepted: June 2018

Abstract:

- A central issue in the theory of extreme values focuses on suitable conditions such that the well-known results for the limiting distributions of the maximum of i.i.d. sequences can be applied to stationary ones. In this context, the extremal index appears as a key parameter to capture the effect of temporal dependence on the limiting distribution of the maxima. The multivariate extremal index corresponds to a generalization of this concept to a multivariate context and affects the tail dependence structure within the marginal sequences and between them. As it is a function, the inference becomes more difficult, and it is therefore important to obtain characterizations, namely bounds based on the marginal dependence that are easier to estimate. In this work we present two decompositions that emphasize different types of information contained in the multivariate extremal index, an upper limit better than those found in the literature and we analyse its role in dependence on the limiting model of the componentwise maxima of a stationary sequence. We will illustrate the results with examples of recognized interest in applications.

Keywords:

- *multivariate extreme values; multivariate extremal index; tail dependence; extremal coefficients; madogram.*

AMS Subject Classification:

- 60G70.

1. INTRODUCTION

Let F be a multivariate distribution function (df), with continuous marginal dfs, in the max-domain of attraction of a multivariate extreme values (MEV) df $\widehat{H}$ having unit Fréchet marginals, $\widehat{H}_j(x_j) \equiv \Phi_j(x_j) = \exp(-x_j^{-1})$, $x_j > 0$, $j = 1, ..., d$. Therefore, we have

$$F^n(u_{n1}(x_1), ..., u_{nd}(x_d)) \to \widehat{H}(x_1, ..., x_d), \tag{1.1}$$

where $u_{nj}(x_j) = a_{nj}x_j$ for some sequence $\{a_{nj} > 0\}$, $j = 1, ..., d$.

Consider $\{\mathbf{X}_n = (X_{n1}, ..., X_{nd})\}$ a stationary sequence such that $F_{\mathbf{X}_n} = F$ and let $\{\mathbf{M}_n = (M_{n1}, ..., M_{nd})\}$ be the componentwise maxima sequence generated from $\mathbf{X}_1, ..., \mathbf{X}_n$ and therefore $M_{nj} = \bigvee_{i=1}^n X_{ij}$, $j = 1, ..., d$. If

$$\lim_{n\to\infty} P(M_{n1} \le u_{n1}(x_1), ..., M_{nd} \le u_{nd}(x_d)) = H(x_1, ..., x_d), \tag{1.2}$$

for some MEV df H, we can relate $H(x_1, ..., x_d)$ and $\widehat{H}(x_1, ..., x_d)$ through the so called multivariate extremal index of $\{\mathbf{X}_n\}$. This is possible, even if the marginals $\widehat{H}_j$ are not unit Fréchet distributed, as considered for simplicity and without loss of generality. Indeed, to have (1.1) or *mutatis mutandis* (1.2), it is sufficient that, as $n \to \infty$, the sequence of copulas C_F^n, with $C_F(u_1, ..., u_d) = F(F_1^{-1}(u_1), ..., F_1^{-1}(u_d))$, converges to $C_{\widehat{H}}$, as well as, $F_j^n(u_{nj}(x_j)) \to \widehat{H}_j(x_j)$, $j = 1, ..., d$, which can be reduced to the case of convergence to the Fréchet without affecting the convergence of C_F^n.

We recall the definition of multivariate extremal index of $\{\mathbf{X}_n\}$ and its role in the relation between H and $\widehat{H}$ (Nandagopalan [18], 1994). The sequence $\{\mathbf{X}_n\}$ has multivariate extremal index $\theta(\boldsymbol{\tau}) \in (0, 1]$, $\boldsymbol{\tau} = (\tau_1, ..., \tau_d) \in \mathbb{R}_+^d$, when for each $\boldsymbol{\tau}$ there is a sequence of real levels $\{\mathbf{u}_n^{(\boldsymbol{\tau})} = (u_{n1}^{(\tau_1)}, ..., u_{nd}^{(\tau_d)})\}$ satisfying

$$nP(X_{1j} > u_{nj}^{(\tau_j)}) \to \tau_j,\ j \in D = \{1, ..., d\}, \tag{1.3}$$

$$P(\widehat{\mathbf{M}}_n \le \mathbf{u}_n^{(\boldsymbol{\tau})}) \to \widehat{\gamma}(\boldsymbol{\tau}) \text{ and} \tag{1.4}$$

$$P(\mathbf{M}_n \le \mathbf{u}_n^{(\boldsymbol{\tau})}) \to \gamma(\boldsymbol{\tau}) = (\widehat{\gamma}(\boldsymbol{\tau}))^{\theta(\boldsymbol{\tau})},$$

where $\widehat{\mathbf{M}}_n = (\widehat{M}_{n1}, ..., \widehat{M}_{nd})$, $\widehat{M}_{nj} = \bigvee_{i=1}^n \widehat{X}_{ij}$, $j = 1, ..., d$, and $\{\widehat{\mathbf{X}}_n\}$ is a sequence of independent vectors such that $F_{\widehat{\mathbf{X}}_n} = F_{\mathbf{X}_n}$.

Observe that

$$\widehat{\gamma}(\boldsymbol{\tau}) = \exp\left(-\lim_{n\to\infty} nP(\mathbf{X}_1 \not\le \mathbf{u}_n)\right) = \exp\left(-\Gamma(\boldsymbol{\tau})\right),$$

with

$$\begin{aligned}\Gamma(\boldsymbol{\tau}) &= \lim_{n\to\infty} nP\left(\bigcup_{j=1}^{d}\{X_{ij} > u_{nj}^{(\tau_j)}\}\right)\\ &= \sum_{\emptyset\neq J\subset D}(-1)^{|J|+1}\lim_{n\to\infty} nP\left(\bigcap_{j\in J}\{X_{ij} > u_{nj}^{(\tau_j)}\}\right)\\ &= \sum_{\emptyset\neq J\subset D}(-1)^{|J|+1}\Gamma_J^*(\boldsymbol{\tau}_J),\end{aligned}$$

where

$$\Gamma_J^*(\boldsymbol{\tau}_J) \equiv \Gamma^*(\tau_j,\, j\in J) = \lim_{n\to\infty} nP\left(\bigcap_{j\in J}\{X_{ij} > u_{nj}^{(\tau_j)}\}\right)$$

and, in particular, $\Gamma_{\{j\}}^*(\tau_j) = \tau_j$, $j \in D$. So, to say that $\Gamma(\boldsymbol{\tau})$ exists is equivalent to say that $\widehat{\gamma}(\boldsymbol{\tau})$ exists and we have

$$\gamma(\boldsymbol{\tau}) = \exp\Big(-\theta(\boldsymbol{\tau})\Gamma(\boldsymbol{\tau})\Big) = \exp\left(-\theta(\boldsymbol{\tau})\sum_{\emptyset\neq J\subset D}(-1)^{|J|+1}\Gamma_J^*(\boldsymbol{\tau}_J)\right).$$

In a one-dimensional setting, (1.3) and (1.4) are equivalent and $\{X_{nj}\}$ has extremal index $\theta_j \in [0,1]$ if, for all $\tau_j >$, there exists $u_{nj}^{(\tau_j)}$, such that $nP(X_{1j} > u_{nj}^{(\tau_j)}) \to \tau_j$ and $P(M_{nj} \le u_{nj}^{(\tau_j)}) \to \exp(-\theta_j\tau_j)$.

For the sake of simplicity, we will take $u_{nj}(x_j) = nx_j$, $j = 1, ..., d$. This assumption leads to levels $u_{nj}^{(\tau_j)}$ with $\tau_j = x_j^{-1}$ and $\hat{\gamma}(\boldsymbol{\tau}) = \hat{H}(\tau_1^{-1}, ..., \tau_d^{-1})$.

If $\{\mathbf{X}_n\}$ has multivariate extremal index $\theta(\boldsymbol{\tau})$ then any sequence of subvectors $\{(\mathbf{X}_n)_A\}$ with indexes in $A \subset \{1, ..., d\}$ has multivariate extremal index $\theta_A(\boldsymbol{\tau}_A)$, with

$$\theta_A(\boldsymbol{\tau}_A) = \lim_{\substack{\tau_i\to 0^+\\ i\notin A}} \theta(\tau_1, ..., \tau_d),\ \boldsymbol{\tau}_A \in \mathbb{R}_+^{|A|}.$$

In particular, for each $j = 1, ..., d$, $\{X_{nj}\}_{n\ge1}$ has extremal index θ_j.

If $\theta(\boldsymbol{\tau})$, $\boldsymbol{\tau} \in \mathbb{R}_+^d$, exists for $\{\mathbf{X}_n\}$ we have

$$H(x_1, ..., x_d) = \widehat{H}(x_1, ..., x_d)^{\theta(-\log \widehat{H}_1(x_1), ..., -\log \widehat{H}_d(x_d))} \tag{1.5}$$

and $H_j(x_j) = \widehat{H}_j(x_j)^{\theta_j}$, $j \in D$.

From inequalities (Galambos [7], 1987; Marshall and Olkin[15], 1983)

$$\prod_{j=1}^{d}\widehat{H}_j(x_j)^{\theta_j} \le \widehat{H}(x_1, ..., x_d)^{\theta(\tau_1(x_1), ..., \tau_d(x_d))} \le \min_{j=1,...,d}\widehat{H}_j(x_j)^{\theta_j},$$

we obtain

$$\frac{\bigvee_{j=1}^{d}\theta_j\tau_j}{\Gamma(\boldsymbol{\tau})} \le \theta(\boldsymbol{\tau}) \le \frac{\sum_{j=1}^{d}\theta_j\tau_j}{\Gamma(\boldsymbol{\tau})}. \tag{1.6}$$

Besides the relation between H and $\widehat{H}$, $\theta(\boldsymbol{\tau})$ also informs about the existence of clustering of events "at least some exceedance of $u_{nj}^{(\tau_j)}$ by X_{nj}, for some j", since (Nandagopalan [17], 1990)

$$\frac{1}{\theta(\boldsymbol{\tau})} = \lim_{n\to\infty} E\left(\sum_{i=1}^{r_n} \mathbb{1}_{\{\mathbf{X}_i \not\leq \mathbf{u}_n^{(\boldsymbol{\tau})}\}} \Big| \sum_{i=1}^{r_n} \mathbb{1}_{\{\mathbf{X}_i \not\leq \mathbf{u}_n^{(\boldsymbol{\tau})}\}} > 0\right), \tag{1.7}$$

for sequences $r_n = [n/k_n]$ and $k_n = o(n)$ provided that $\{\mathbf{X}_n\}$ satisfies condition strong-mixing.

The multivariate extremal index thus preserves, with the natural adaptations, the characteristics that made famous the univariate extremal index. Additionally to these similar characteristics to the univariate extremal index, it plays an unavoidable role in the tail dependence characterization of H. If the tail dependence coefficients applied to F remain unchanged when applied to $\widehat{H}$ (Li [14], 2009), we can not guarantee the same for H, as will be seen in Section 3. The presence of serial dependence within each marginal sequence and between marginal sequences, makes it impossible to approximate the dependence coefficients in the tail of $\mathbf{M}_n$ to those of F.

The dependence modeling between the marginals of F has received considerably more attention in literature than the dependence between the marginals of $F_{\mathbf{M}_n}$, which differs from $F_{\widehat{\mathbf{M}}_n} = F^n$ for being affected by $\theta(\boldsymbol{\tau})$. The need to characterize this dependence appears, for instance, when we have a random field $\{\mathbf{X}_{\mathbf{i},n}, \mathbf{i} \in \mathbb{Z}^2, n \geq 1\}$ and we consider random vectors $(X_{i_1,n}, ..., X_{i_s,n})$ corresponding to locations $(i_1, ..., i_s)$ at time instant n. Sequence $\{(X_{i_1,n}, ..., X_{i_s,n})\}_{n\geq 1}$ presents in general a multivariate extremal index $\theta_{i_1,...,i_s}(\boldsymbol{\tau})$ encompassing information about dependence in the space of locations $i_1, ..., i_s$ and when the time n varies (Pereira *et al.* [21], 2017). Relation (1.5) applied to MEV distributions $\widehat{H}$ and functions $\theta(x_1, ..., x_d)$ compatible with the properties of a multivariate extremal index, provide a means of constructing MEV distributions (Martins and Ferreira [16], 2005).

Notwithstanding all these challenges posed by and for the multivariate extremal index, the literature proves that it remained on the theoretical shelves of the study of extreme values.

The main difficulty of applying the multivariate extremal index lies in the fact that it is a function, unlike what happens with the marginal univariate extremal indexes, for which we have several estimation methods in the literature (see, e.g.: Hsing [9], 1993; Gomes *et al.* [8], 2008; Northrop [20], 2015; Ferreira and Ferreira [6], 2018; and references therein).

Since it remains present the need to estimate the tendency to form clusters in a context of multivariate sequences, we propose in this work:

- (**a**) decompose it, highlighting different types of information contained in it;
- (**b**) bound it in order to obtain a better upper limit than those available in the literature;
- (**c**) enhance its role in the dependence of the tail of H;
- (**d**) apply it to models of recognized interest in applications.

2. CO-MOVEMENTS POINT PROCESSES

Based on (1.7), the multivariate extremal index can be seen as the number of the limiting mean dimension of clustering of events counted by the point process

$$N_n = \sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{X}_i \not\leq \mathbf{u}_n^{(\boldsymbol{\tau})}\}}.$$

We are going to consider two point processes of more restricted events, corresponding to joint exceedances for various marginals of $\mathbf{X}_i$ and enhance the contribution of the extremal indexes of these events in the value of $\theta(\boldsymbol{\tau})$.

Let, for each $\emptyset \neq J \subset D = \{1, ..., d\}$,

$$N_{n,J}^{*} = \sum_{i=1}^{n} \mathbb{1}_{\{\bigcap_{j\in J}\{X_{ij} > u_{nj}\}\}},\ n \geq 1,$$

and

$$N_{n,J}^{**} = \sum_{i=1}^{n} \mathbb{1}_{\{\bigwedge_{j\in J} X_{ij} > \bigvee_{j\in J} u_{nj}\}},\ n \geq 1,$$

where notations $\wedge$ and $\vee$ stand for minimum and maximum, respectively.

We denote the respective limiting mean number of occurrences by

$$\Gamma_J^{*}(\boldsymbol{\tau}_J) = \lim_{n\to\infty} nP\left(\bigcap_{j\in J}\{X_{ij} > u_{nj}\}\right)$$

and

$$\Gamma_J^{**}(\boldsymbol{\tau}_J) = \lim_{n\to\infty} nP\left(\bigcap_{j\in J}\{X_{ij} > \bigvee_{j\in J} u_{nj}\}\right).$$

Observe that

$$\Gamma_J^{**}(\boldsymbol{\tau}_J) = \lim_{n\to\infty} nP\left(\bigcap_{j\in J}\left\{X_{ij} > \frac{n}{\bigwedge_{j\in J}\tau_j}\right\}\right).$$

Thus

$$\Gamma_J^{**}(\boldsymbol{\tau}_J) = \tau_J^{**}\left(\bigwedge_{j\in J}\tau_j\right),$$

with τ_J^{**} an increasing function in $\bigwedge_{j\in J}\tau_j$ and homogeneous of order 1. Therefore, we have

$$\tau_J^{**}\left(\frac{\bigwedge_{j\in J}\tau_j}{s}\right) = \frac{\tau_J^{**}\left(\bigwedge_{j\in J}\tau_j\right)}{s}, \tag{2.1}$$

for all $s \neq 0$, a relation that will be fundamental for the independence of θ^{**} from τ.

In case $J = D$, we will omit the index J in notation.

For each of these processes, we can define an index of clustering of occurrences, which we will also call extremal indexes, $\theta_J^*(\boldsymbol{\tau}_J)$ and $\theta_J^{**}(\boldsymbol{\tau}_J)$, being the latter a constant independent of $\boldsymbol{\tau}_J$, as we will see.

Let us assume that sequence $\{\mathbf{X}_n\}_{n\geq 1}$ satisfies the strong-mixing condition (Leadbetter *et al.* [11], 1983) and, as consequence, we have, as $n \to \infty$,

$$P\left(N_{n,J} = 0\right) - P^{k_n}\left(N_{[n/k_n],J} = 0\right) \to 0,$$

$$P\left(N_{n,J}^* = 0\right) - P^{k_n}\left(N_{[n/k_n],J}^* = 0\right) \to 0$$

and

$$P\left(N_{n,J}^{**} = 0\right) - P^{k_n}\left(N_{[n/k_n],J}^{**} = 0\right) \to 0,$$

for any integers sequence $\{k_n\}$, such that, $k_n \to \infty$, $k_n\alpha_n(l_n) \to 0$ and $k_n l_n/n \to 0$, as $n \to \infty$, where $\alpha_n(\cdot)$ and l_n are the sequences of the strong-mixing condition. Thus

$$P\left(N_{n,J} = 0\right) \to \exp\left(-\theta_J(\boldsymbol{\tau}_J)\Gamma_J(\boldsymbol{\tau}_J)\right),$$

$$P\left(N_{n,J}^* = 0\right) \to \exp\left(-\theta_J^*(\boldsymbol{\tau}_J)\Gamma_J^*(\boldsymbol{\tau}_J)\right)$$

and

$$P\left(N_{n,J}^{**} = 0\right) \to \exp\left(-\theta_J^{**}(\boldsymbol{\tau}_J)\tau_J^{**}\left(\bigwedge_{j\in J}\tau_j\right)\right), \tag{2.2}$$

with

$$\theta_J(\boldsymbol{\tau}_J) = \lim_{n\to\infty} k_n P\left(N_{[n/k_n],J} > 0\right)/\Gamma_J(\boldsymbol{\tau}_J),$$

$$\theta_J^*(\boldsymbol{\tau}_J) = \lim_{n\to\infty} k_n P\left(N_{[n/k_n],J}^* > 0\right)/\Gamma_J^*(\boldsymbol{\tau}_J),$$

$$\theta_J^{**}(\boldsymbol{\tau}_J) = \lim_{n\to\infty} k_n P\left(N_{[n/k_n],J}^{**} > 0\right)/\tau_J^{**}\left(\bigwedge_{j\in J}\tau_j\right)$$

and

$$\theta_J^{**}(\boldsymbol{\tau}_J)\tau_J^{**}\left(\bigwedge_{j\in J}\tau_j\right) \leq \theta_J^*(\boldsymbol{\tau}_J)\Gamma_J^*(\boldsymbol{\tau}_J) \leq \bigvee_{j\in J}\theta_j\tau_j \leq \theta_J(\boldsymbol{\tau}_J)\Gamma_J(\boldsymbol{\tau}_J).$$

In the following we present relations between $\theta_J^{**}(\boldsymbol{\tau}_J)$, $\theta_J^*(\boldsymbol{\tau}_J)$ and $\theta_J(\boldsymbol{\tau}_J)$, which will allow us a detailed interpretation of the information contained in $\theta(\boldsymbol{\tau})$ and an upper bound better than the one in (1.6). But first, we start by proving that $\theta_J^{**}(\boldsymbol{\tau}_J) = \theta_J^{**}$, i.e., these extremal indexes are independent of $\boldsymbol{\tau}$, which is already known for $J = \{j\}$ (Leadbetter *et al.* [11], 1983), $j = 1, ..., d$, since $\theta_{\{j\}}^{**} = \theta_j$. Indeed the proof runs along the same lines.

Proposition 2.1. *For stationary sequences* $\{\mathbf{X}_n\}$ *satisfying the strong-mixing condition, if there exists the limit* (2.2) *for some* $\boldsymbol{\tau}$*, then it exists for any* $\boldsymbol{\tau} > 0$ *and there exists a constant* $\theta_A^{**} \in [0,1]$ *such that*

$$P\left(N_{n,A}^{**}=0\right) \to \exp\left(-\theta_A^{**}\tau_A^{**}\left(\bigwedge_{j\in A}\tau_j\right)\right).$$

Proof: From the strong-mixing condition, we have

$$\begin{aligned}
\liminf_{n\to\infty} P\left(N_{n,A}^{**}=0\right) &= \liminf_{n\to\infty} P^{k_n}\left(N_{[n/k_n],A}^{**}=0\right) \\
&= \liminf_{n\to\infty}\left(1-\frac{k_n P\left(N_{[n/k_n],A}^{**}>0\right)}{k_n}\right)^{k_n} \\
&\geq \liminf_{n\to\infty}\left(1-\frac{nP\left(\bigwedge_{j\in A}X_{1j}>\bigvee_{j\in A}u_{nj}\right)}{k_n}\right)^{k_n} = \left(1-\frac{\tau_A^{**}\left(\bigwedge_{j\in A}\right)}{k_n}\right)^{k_n}.
\end{aligned}$$

Thus, if there exists $\Psi(\tau_A^{**}) = \limsup_{n\to\infty} P\left(N_{n,A}^{**}=0\right)$, we have $\Psi\left(\tau_A^{**}\left(\bigwedge_{j\in A}\right)\right) \geq \exp\left(-\tau_A^{**}\left(\bigwedge_{j\in A}\right)\right)$, and so $\Psi(\tau_A^{**})$ is a strictly positive function.

We also have that function $\Psi(\tau_A^{**})$ would have to satisfy $\Psi(\tau_A^{**}/k) = \Psi^{1/k}(\tau_A^{**})$, for all $\tau_A^{**}>0$ and $k=1,2,...$, since, representing $\sum_{i=1}^n \mathbb{1}_{\left\{\bigwedge_{j\in A}X_{ij}>m\right/\bigwedge_{j\in A}\tau_j\}}$ by $N_n^{**}\left(\mathbf{u}_m^{\left(\tau_A^{**}\left(\bigwedge_{j\in A}\tau_j\right)\right)}\right)$ and applying (2.2), it holds

$$\begin{aligned}
&\left|P\left(N_{[n/k_n],A}^{**}\left(\mathbf{u}_n^{\left(\tau_A^{**}\left(\bigwedge_{j\in A}\tau_j\right)\right)}\right)=0\right)-P\left(N_{[n/k_n],A}^{**}\left(\mathbf{u}_{[n/k_n]}^{\left(\tau_A^{**}\left(\bigwedge_{j\in A}\tau_j\right)/k_n\right)}\right)=0\right)\right| = \\
&\qquad \leq \left[\frac{n}{k_n}\right]\left|P\left(\bigwedge_{j\in A}X_{1j}>\frac{n}{\bigwedge_{j\in A}\tau_j}\right)-P\left(\bigwedge_{j\in A}X_{1j}>\frac{[n/k_n]}{\bigwedge_{j\in A}\tau_j/k_n}\right)\right| \\
&\qquad = \left[\frac{n}{k_n}\right]\left|\frac{\bigwedge_{j\in A}\tau_j}{n}(1+o(1))-\frac{\bigwedge_{j\in A}\tau_j/k_n}{[n/k_n]}(1+o(1))\right| = o(1)
\end{aligned}$$

and thus we would have

$$\begin{aligned}
\Psi\left(\frac{\tau_A^{**}}{k_n}\right) &= \limsup_{n\to\infty} P\left(N_{[n/k_n],A}^{**}\left(\mathbf{u}_{[n/k_n],A}^{\left(\tau_A^{**}/k_n\right)}\right)=0\right) = \limsup_{n\to\infty} P\left(N_{n,A}^{**}\left(\mathbf{u}_{n,A}^{\left(\tau_A^{**}\right)}\right)=0\right) \\
&= \Psi\left(\tau_A^{**}\right)^{1/k_n}.
\end{aligned}$$

On the other hand, $\Psi\left(\tau_A^{**}\right)$ would have to be a non increasing function because if, for some $\boldsymbol{\tau}_0 = (\tau_{0,1},...,\tau_{0,d})$, we have

$$\begin{aligned}
\tau_{0,A}^{**}\left(\bigwedge_{j\in A}\tau_{0,j}\right) &= \lim_{n\to\infty} nP\left(\bigwedge_{j\in A}X_{1j}>\frac{n}{\bigwedge_{j\in A}\tau_{0,j}}\right) > \tau_A^{**}\left(\bigwedge_{j\in A}\tau_j\right) \\
&= \lim_{n\to\infty} nP\left(\bigwedge_{j\in A}X_{1j}>\frac{n}{\bigwedge_{j\in A}\tau_j}\right)
\end{aligned}$$

and $\tau_A^{**}\left(\bigwedge_{j\in A}\tau_j\right)$ is increasing in $\bigwedge_{j\in A}\tau_j$, then for all n large,

$$\left\{\bigwedge_{j\in A} X_{1j} > \frac{n}{\bigwedge_{j\in A}\tau_j}\right\} \subset \left\{\bigwedge_{j\in A} X_{1j} > \frac{n}{\bigwedge_{j\in A}\tau_{0,j}}\right\}$$

and thus

$$\left\{N_{n,A}^{**}\left(\mathbf{u}_n^{\left(\tau_{0,A}^{**}\right)}\right)=0\right\} \subset \left\{N_{n,A}^{**}\left(\mathbf{u}_n^{\left(\tau_A^{**}\right)}\right)=0\right\}$$

and $\Psi\left(\tau_{0,A}^{**}\right) \le \Psi\left(\tau_A^{**}\right)$. If $\Psi\left(\tau_A^{**}\right)$ is a strictly positive function, non increasing and such that $\Psi\left(\tau_A^{**}/k\right) = \Psi\left(\tau_A^{**}\right)^{1/k}$, then $\Psi\left(\tau_A^{**}\right) = \exp\left(-\theta_A^{**}\tau_A^{**}\right)$, with θ_A^{**} a non negative constant. Since $\Psi\left(\tau_A^{**}\right) > \exp\left(-\tau_A^{**}\right)$, it also comes $\theta_A^{**} \le 1$. For the lower limit, we can make the same reasoning to obtain the result. □

Let us start by emphasizing that, to $\theta(\boldsymbol{\tau})\Gamma(\boldsymbol{\tau})$, we have the contribution of the clustering of the joint exceedances of all levels by the respective marginals, including the particular case of the clustering of exceedances of the largest level by the lower marginal, as well as, the clustering of exceedances of one or more levels by the respective marginals without joint exceedances of all levels.

Proposition 2.2. *Let $\{\mathbf{X}_n\}$ be a stationary sequence satisfying the strong-mixing condition and $\{\mathbf{u}_n^{(\boldsymbol{\tau})} = (u_n^{(\tau_1)}, ..., u_n^{(\tau_d)})\}$ a sequence of normalized real levels for which there exists $\Gamma(\boldsymbol{\tau})$. Then*

(**a**) $\theta(\boldsymbol{\tau})\Gamma(\boldsymbol{\tau}) = \theta^{**}\tau^{**}\left(\bigwedge_{j=1}^{d}\tau_j\right) + \theta^*(\boldsymbol{\tau})\Gamma^*(\boldsymbol{\tau})\beta^{(1)}(\boldsymbol{\tau}) + \sum_{\emptyset\neq J\subset D}(-1)^{|J|+1}\Theta_J(\boldsymbol{\tau}_J)$,

where

$$\beta^{(1)}(\boldsymbol{\tau}) = \lim_{n\to\infty} P\left(N_{r_n}^{**}=0 | N_{r_n}^{*}>0\right)$$

and

$$\Theta_J(\boldsymbol{\tau}_J) = \lim_{n\to\infty} k_n P\left(\bigcap_{j\in J}\{N_{r_n,\{j\}}>0\} | N_{r_n}^{*}=0\right);$$

(**b**) $\sum_{\emptyset\neq J\subset D}(-1)^{|J|+1}\Theta_J(\boldsymbol{\tau}_J) \le \sum_{j=1}^{d}\theta_j\tau_j.$

Proof: We have

$$\begin{aligned}
k_n P\left(N_{r_n}>0\right) =& \; k_n P\left(N_{r_n}^{**}>0\right) + k_n P\left(N_{r_n}^{*}>0, N_{r_n}^{**}=0\right) \\
&+ k_n P\left(N_{r_n}>0, N_{r_n}^{*}=0\right) \\
=& \; k_n P\left(N_{r_n}^{**}>0\right) + k_n P\left(N_{r_n}^{*}>0\right) P\left(N_{r_n}^{**}=0 | N_{r_n}^{*}>0\right) \\
&+ k_n P\left(\bigcup_{j=1}^{d}\{N_{r_n,\{j\}}>0\}, N_{r_n}^{*}=0\right) \\
=& \; k_n P\left(N_{r_n}^{**}>0\right) + k_n P\left(N_{r_n}^{*}>0\right) P\left(N_{r_n}^{**}=0 | N_{r_n}^{*}>0\right) \\
&+ \sum_{\emptyset\neq J\subset D}(-1)^{|J|+1} k_n P\left(\bigcap_{j\in J}\{N_{r_n,\{j\}}>0\}, N_{r_n}^{*}=0\right).
\end{aligned}$$

In what concerns the last term, observe that

$$\sum_{j=1}^{d} k_n P\left(N_{r_n,\{j\}} > 0, N^*_{r_n} = 0\right) = \sum_{j=1}^{d} k_n P\left(N_{r_n,\{j\}} > 0\right) - \sum_{j=1}^{d} k_n P\left(N_{r_n,\{j\}} > 0, N^*_{r_n} > 0\right)$$

and since $\lim_{n\to\infty} P\left(N^*_{r_n} = 0\right) = 1$, we have the result in (a). □

Observe that $\beta^{(1)}(\boldsymbol{\tau})$ reduces $\theta^*(\boldsymbol{\tau})$ from the joint exceedances of $\bigvee_{j=1}^{d} n/\tau_j$ accounted for θ^{**}. We can say that in the last term of representation of $\theta(\boldsymbol{\tau})\Gamma(\boldsymbol{\tau})$ we are accounting the tendency of one or more marginals to form clusters, without joint exceedances of all the marginals.

We illustrate the previous result with a bivariate sequence with unit Fréchet marginals and such that the joint tail is regularly varying at ∞ with index $\eta \in (0, 1]$ measuring a penultimate tail dependence, as the (sub)model presented in Ledford and Tawn ([12], 1996).

Example 2.1. Suppose that $d = 2$ and $\{(X_{n1}, X_{n2})\}_{n\geq 1}$ is a strong-mixing stationary sequence, with unit Fréchet marginals and such that X_{n1} and X_{n2} are asymptotically independent, i.e.,

$$nP(X_{n1} > nx, X_{n2} > ny) \to 0, \tag{2.3}$$

as $n \to \infty$, for x, y positive. Then

$$\begin{aligned}\theta^{**} &= k_n P\left(N^{**}_{r_n} > 0\right) \leq nP\left(X_{n1} > \frac{n}{\tau_1 \wedge \tau_2}, X_{n2} > \frac{n}{\tau_1 \wedge \tau_2}\right)\\ &\sim n\left(\frac{n}{\tau_1 \wedge \tau_2}\right)^{-1/\eta} L\left(\frac{n}{\tau_1 \wedge \tau_2}\right) \to 0,\end{aligned}$$

$$\begin{aligned}\theta^*(\tau_1, \tau_2) &\leq nP\left(X_{n1} > \frac{n}{\tau_1}, X_{n2} > \frac{n}{\tau_2}\right)\\ &\leq nP\left(X_{n1} > \frac{n}{\tau_1 \wedge \tau_2}, X_{n2} > \frac{n}{\tau_1 \wedge \tau_2}\right) \to 0.\end{aligned}$$

Therefore, regardless of additional conditions on the serial dependence, the validity of (2.3) implies

$$\theta(\boldsymbol{\tau})\Gamma(\boldsymbol{\tau}) = \sum_{\emptyset \neq J \subset \{1,2\}} (-1)^{|J|+1} \lim_{n\to\infty} k_n P\left(\bigcap_{j\in J} \{N_{r_n,\{j\}} > 0\}, N^*_{r_n} = 0\right)$$

and $\Gamma(\boldsymbol{\tau}) = \tau_1 + \tau_2$. Since $k_n P\left(N^*_{r_n} > 0\right) \to 0$ we can thus write in this model

$$\theta(\boldsymbol{\tau}) = \frac{1}{\tau_1+\tau_2} \lim_{n\to\infty} k_n\Big(P\big(N_{r_n,\{1\}} > 0\big) + P\big(N_{r_n,\{2\}} > 0\big) - P\big(N_{r_n,\{1\}} > 0, N_{r_n,\{2\}} > 0\big)\Big). \tag{2.4}$$

We now consider several particular situations.

(**a**) In the case of independent vectors (X_{n1}, X_{n2}), $n \geq 1$, we have

$$\begin{aligned}\theta(\boldsymbol{\tau}) &= \frac{1}{\tau_1+\tau_2}\left(\theta_1\tau_1+\theta_2\tau_2\right.\\ &\quad -\lim_{n\to\infty} k_n P\left(\bigcup_{1\leq i<i'\leq r_n}\left\{\left\{X_{i1}>\frac{n}{\tau_1}, X_{i2}\leq\frac{n}{\tau_2}, X_{i'1}\leq\frac{n}{\tau_1}, X_{i'2}>\frac{n}{\tau_2},\right\}\right.\right.\\ &\quad \left.\left.\left.\cup\left\{X_{i1}\leq\frac{n}{\tau_1}, X_{i2}>\frac{n}{\tau_2}, X_{i'1}>\frac{n}{\tau_1}, X_{i'2}\leq\frac{n}{\tau_2}\right\}\right\}\right)\right)\\ &= \frac{\tau_1+\tau_2}{\tau_1+\tau_2} = 1.\end{aligned}$$

It will then come $P\left(M_{n1}\leq n/\tau_1, M_{n2}\leq n/\tau_2\right)\to\exp(-\Gamma(\boldsymbol{\tau}))=\exp(-\tau_1)\exp(-\tau_2)$, that is, M_{n1} and M_{n2} are also asymptotically independent.

(**b**) Suppose that $\{(X_{n1}, X_{n2})\}_{n\geq 1}$, satisfies condition $D^{(m)}_{\{1,2\}}$ defined by

$$\lim_{n\to\infty} n\sum_{j=m+1}^{[n/k_n]} P\left(X_{11}>n/\tau_1, X_{j2}>n/\tau_2\right)=0,$$

which extends $D'_{\{1,2\}}$ of Davis ([2], 1982), satisfied by i.i.d. sequences. Then

$$\theta(\boldsymbol{\tau}) = \frac{1}{\tau_1+\tau_2}\left(\theta_1\tau_1+\theta_2\tau_2-\lim_{n\to\infty} n\sum_{i=2}^{m} P\left(X_{11}>n/\tau_1, X_{i2}>n/\tau_2\right)\right),$$

where the last part reflects the cross dependence.

(**c**) If we assume an analogous hypothesis of (2.3) for (X_{11}, X_{i2}) with different η_i, we will also obtain asymptotic independence between M_{n1} and M_{n2}, since the last term has null limit. We have $P\left(M_{n1}\leq n/\tau_1, M_{n2}\leq n/\tau_2\right)\to\exp(-\Gamma(\boldsymbol{\tau})\theta(\boldsymbol{\tau}))=\exp(-\theta_1\tau_1)\exp(-\theta_2\tau_2)$.

(**d**) If $\theta(\boldsymbol{\tau})=\theta$, $\forall\boldsymbol{\tau}\in\mathbb{R}^2_+$, then $\theta_1=\theta_2=\theta$ and, from (2.4),

$$\theta=\theta-\lim_{n\to\infty} k_n P\left(N_{r_n,\{1\}}>0, N_{r_n,\{2\}}>0\right),$$

which implies that this limit is null and thus $P\left(M_{n1}\leq n/\tau_1, M_{n2}\leq n/\tau_2\right)\to\exp(-\theta(\tau_1+\tau_2))$ $=\exp(-\theta\tau_1)\exp(-\theta\tau_2)$.

We present below a relation between $\theta(\boldsymbol{\tau})$ and the extremal indexes $\theta^{**}_{\{j,\ldots,d\}}$ and $\theta^{*}_{\{j,\ldots,d\}}\left(\boldsymbol{\tau}_{\{j,\ldots,d\}}\right)$, $j=1,\ldots,d$, which discriminates different informations contained in function $\theta(\boldsymbol{\tau})$ and provides an upper bound for $\theta(\boldsymbol{\tau})$ better than the one in (1.6). In Example 2.2 we show that the proposed upper bound for the M4 processes, can be better than the one presented in Ehlert and Schlather ([3], 2008). The new upper bound has also the advantage of depending only on constant extremal indexes which can be estimated by known methods of literature.

Proposition 2.3. *Let $\{\mathbf{X}_n\}$ be a stationary sequence satisfying the strong-mixing condition and $\{\mathbf{u}_n^{(\boldsymbol{\tau})}=(u_n^{(\tau_1)},\ldots,u_n^{(\tau_d)})\}$ a sequence of normalized real levels for which there exists $\Gamma(\boldsymbol{\tau})$. Then*

$$\begin{aligned}(\mathbf{a})\quad \theta(\boldsymbol{\tau})\Gamma(\boldsymbol{\tau}) = \lim_{n\to\infty} k_n P\left(N_{r_n}>0\right) &= \sum_{j=1}^{d}\theta_j\tau_j - \sum_{j=1}^{d-1}\theta^{**}_{\{j,\ldots,d\}}\tau^{**}_{\{j,\ldots,d\}}\left(\bigwedge_{i=j}^{d}\tau_i\right)\\ &-\sum_{j=1}^{d-1}\theta^{*}_{\{j,\ldots,d\}}\left(\boldsymbol{\tau}_{\{j,\ldots,d\}}\right)\Gamma^{*}_{\{j,\ldots,d\}}\left(\boldsymbol{\tau}_{\{j,\ldots,d\}}\right)\beta_j^{(1)}\left(\boldsymbol{\tau}_{\{j,\ldots,d\}}\right)\\ &-\sum_{j=1}^{d-1}\sum_{J\subset\{j+1\ldots,d\}}(-1)^{|J|+1}\beta^{(2)}_{\{j\}\cup J}\left(\boldsymbol{\tau}_{\{j\}\cup J}\right),\end{aligned}$$

where we have $\beta_j^{(1)}\left(\boldsymbol{\tau}_{\{j,\ldots,d\}}\right)=\lim_{n\to\infty}P\left(N^{**}_{r_n,\{j,\ldots,d\}}=0|N^{*}_{r_n,\{j,\ldots,d\}}>0\right)$ and $\beta^{(2)}_{\{j\}\cup J}\left(\boldsymbol{\tau}_{\{j\}\cup J}\right)=\lim_{n\to\infty}k_nP\left(\bigcap_{i\in\{j\}\cup J}\{N_{r_n,\{i\}}>0\}|N^{*}_{r_n,\{j,\ldots,d\}}=0\right)$, provided that the limiting constants exist.

$$(\mathbf{b})\quad \theta(\boldsymbol{\tau})\le\frac{1}{\Gamma(\boldsymbol{\tau})}\left(\sum_{j=1}^{d}\theta_j\tau_j-\sum_{j=1}^{d-1}\theta^{**}_{\{j,\ldots,d\}}\tau^{**}_{\{j,\ldots,d\}}\left(\bigwedge_{i=j}^{d}\tau_i\right)\right).$$

Proof: We have

$$\begin{aligned}k_nP\left(N_{r_n}>0\right) &= k_nP\left(\bigcup_{j=1}^{d}\{N_{r_n,\{j\}}>0\}\right)\\ &= \sum_{j=1}^{d-1}k_nP\left(N_{r_n,\{j\}}>0,\bigcap_{i=j+1}^{d}\{N_{r_n,\{i\}}=0\}\right)+k_nP\left(N_{r_n,\{d\}}>0\right)\\ &= \sum_{j=1}^{d}k_nP\left(N_{r_n,\{j\}}>0\right)-\sum_{j=1}^{d-1}k_nP\left(N_{r_n,\{j\}}>0,\bigcup_{i=j+1}^{d}\{N_{r_n,\{i\}}>0\}\right).\end{aligned}$$

Regarding the second term, we can also say that

$$\sum_{j=1}^{d-1}k_nP\left(N_{r_n,\{j\}}>0,\bigcup_{i=j+1}^{d}\{N_{r_n,\{i\}}>0\}\right)=$$

$$
\begin{aligned}
&= \sum_{j=1}^{d-1} k_n P\left(N_{r_n,\{j\}} > 0, \bigcup_{i=j+1}^{d} \{N_{r_n,\{i\}} > 0\}, N^*_{r_n,\{j,\ldots,d\}} > 0\right) \\
&\quad + \sum_{j=1}^{d-1} k_n P\left(N_{r_n,\{j\}} > 0, \bigcup_{i=j+1}^{d} \{N_{r_n,\{i\}} > 0\}, N^*_{r_n,\{j,\ldots,d\}} = 0\right) \\
&= \sum_{j=1}^{d-1} k_n P\left(N^*_{r_n,\{j,\ldots,d\}} > 0, N^{**}_{r_n,\{j,\ldots,d\}} > 0\right) \\
&\quad + \sum_{j=1}^{d-1} k_n P\left(N^*_{r_n,\{j,\ldots,d\}} > 0, N^{**}_{r_n,\{j,\ldots,d\}} = 0\right) \\
&\quad + \sum_{j=1}^{d-1} k_n P\left(N_{r_n,\{j\}} > 0, \bigcup_{i=j+1}^{d} \{N_{r_n,\{i\}} > 0\}, N^*_{r_n,\{j,\ldots,d\}} = 0\right) \\
&= \sum_{j=1}^{d-1} k_n P\left(N^{**}_{r_n,\{j,\ldots,d\}} > 0\right) \\
&\quad + \sum_{j=1}^{d-1} k_n P\left(N^*_{r_n,\{j,\ldots,d\}} > 0, N^{**}_{r_n,\{j,\ldots,d\}} = 0\right) \\
&\quad + \sum_{j=1}^{d-1} k_n P\left(N_{r_n,\{j\}} > 0, \bigcup_{i=j+1}^{d} \{N_{r_n,\{i\}} > 0\}, N^*_{r_n,\{j,\ldots,d\}} = 0\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\theta(\boldsymbol{\tau})\,\Gamma(\boldsymbol{\tau}) = \lim_{n\to\infty} k_n P\left(N_{r_n} > 0\right) &= \sum_{j=1}^{d} \theta_j \tau_j - \sum_{j=1}^{d-1} \theta^{**}_{\{j,\ldots,d\}} \tau^{**}_{\{j,\ldots,d\}} \left(\bigwedge_{i=j}^{d} \tau_i\right) \\
&\quad - \sum_{j=1}^{d-1} \theta^*_{\{j,\ldots,d\}}\left(\boldsymbol{\tau}_{\{j,\ldots,d\}}\right) \Gamma^*_{\{j,\ldots,d\}}\left(\boldsymbol{\tau}_{\{j,\ldots,d\}}\right) \lim_{n\to\infty} P\left(N^{**}_{r_n,\{j,\ldots,d\}} = 0 \,\middle|\, N^*_{r_n,\{j,\ldots,d\}} > 0\right) \\
&\quad - \sum_{j=1}^{d-1} \lim_{n\to\infty} k_n P\left(\bigcup_{i=j+1}^{d} \{N_{r_n,\{j\}} > 0, N_{r_n,\{i\}} > 0\} | N^*_{r_n,\{j,\ldots,d\}} = 0\right),
\end{aligned}
$$

since $P\left(N^*_{r_n,\{j,\ldots,d\}} = 0\right) \to 1$, as $n \to \infty$. □

The above result means that the values $\theta^*_{\{j,\ldots,d\}}\left(\boldsymbol{\tau}_{\{j,\ldots,d\}}\right)$, for each $j \in \{1,\ldots,d\}$, only contribute to $\theta(\boldsymbol{\tau})$ if it is not asymptotically almost surely the local occurrence of some joint exceedances of the largest level $u^{(\tau_j)}_{ni}$, $i \in \{1,\ldots,d\}$, among the joint exceedances of these levels. Otherwise, the joint exceedances clustering is considered only through the clustering of the joint exceedances of the largest level u_{ni}, $i \in \{j,\ldots,d\}$, and measured by $\theta^{**}_{\{j,\ldots,d\}}$, disappearing the third term. Therefore, the second and third terms together account for the clustering of two situations of joint exceedances. The fourth term measures the clustering of exceedances of u_{nj} and of one or more u_{ni}, $i \in \{j+1,\ldots,d\}$, in the absence of joint exceedances of levels u_{ni}, $i \in \{j,\ldots,d\}$, not accounted within the second and third terms. All these clustering situations were accounted by excess in the first term.

The function $\theta(\boldsymbol{\tau})$ is homogeneous of order zero and thus $\theta(\tau,...,\tau)=\theta(1,...,1)$, $\forall\tau\in\mathbb{R}$. The constant $\theta(1,...,1)$ has been used as a dependence coefficient of the marginals of H (see, e.g.: Martins and Ferreira [16], 2005; Ehlert and Schlather [3], 2008; Ferreira and Ferreira [5], 2015; and references therein).

We are going to analyze the consequences of the decompositions presented for $\theta(\boldsymbol{\tau})$ in the calculation of $\theta(\mathbf{1})$.

If $\tau_1=\cdots=\tau_d=\tau$, then $N_n^{**}=N_n^*$, $\beta_J^{(1)}(\boldsymbol{\tau})=0$, $\Gamma^*(\boldsymbol{\tau})=\tau^{**}(\boldsymbol{\tau})$ and $\Gamma(\boldsymbol{\tau})=\sum_{\emptyset\neq J\subset D}(-1)^{|J|+1}\tau_J^{**}(\boldsymbol{\tau}_J)$.

The first decomposition

$$\theta(\mathbf{1})\Gamma(\mathbf{1})=\theta^{**}\tau^{**}(\mathbf{1})+\lim_{n\to\infty}k_nP\left(\bigcup_{j=1}^{d}\{N_{r_n,\{j\}}>0\},N_{r_n}^*=0\right),$$

separates once again the contribution of the clustering of exceedances across all marginals from the contribution of the clustering of exceedances of one or more marginals without exceedances of all marginals.

In the next section, we will give an important utility to the boundary of $\theta(\boldsymbol{\tau})\Gamma^*(\boldsymbol{\tau})$. It will serve to delimitate the difference between the tail dependence coefficients of H and $\widehat{H}$.

The second decomposition allow us to obtain an upper bound for $\theta(\mathbf{1})$, which can be better than the one presented in (1.6). From the previous result, we have

$$\theta(\mathbf{1})\Gamma(\mathbf{1})\leq\sum_{j=1}^{d}\theta_j-\sum_{j=1}^{d-1}\theta_{\{j,...,d\}}^{**}\tau_{\{j,...,d\}}^{**}(\mathbf{1}). \tag{2.5}$$

From the proof of Proposition 2.3 we found that, instead of following the order $1,...,d$ to decompose initially the event $\{\bigcup_{j=1}^{d}N_{r_n,\{j\}}>0\}$ in a reunion of disjoint events $\{N_{r_n,\{j\}}>0,\bigcap_{i=j+1}^{d}\{N_{r_n,\{i\}}>0\}\}$, $j=1,...,d-1$ and $\{N_{r_n,\{d\}}>0\}$, we can consider any other permutation $(i_1,...,i_d)$ from $(1,...,d)$ and repeat the process. Therefore the previous upper limit can be improved in the following sense:

$$\theta(\mathbf{1})\Gamma(\mathbf{1})\leq\sum_{j=1}^{d}\theta_j-\bigvee_{(i_1,...,i_d)\in\mathcal{P}_d}\sum_{j=i_1}^{i_{d-1}}\theta_{\{j,...,i_d\}}^{**}\tau_{\{j,...,i_d\}}^{**}(\mathbf{1}),$$

where $\mathcal{P}_d$ denotes the set of all permutations of $(1,...,d)$.

Example 2.2. Consider the M4 process,

$$\begin{cases}X_{n1}=0.7Z_n\vee0.3Z_{n-2}\\ X_{n2}=0.7Z_{n-1}\vee0.1Z_{n-2}\vee0.5Z_{n-3},\end{cases}$$

with $\{Z_n\equiv Z_{1,n}\}$, where $\{Z_{l,n}\}$, $l\geq1$, $n\geq1$, is an array of independent unit Fréchet random variables. We have $\theta_1=0.7$, $\theta_2=0.5$ and $\theta(\mathbf{1})\Gamma(\mathbf{1})=0.7$. Since $\{X_n\}_{n\geq1}$ is 4-dependent,

representing $\{X_{i1} > n/\tau, X_{i2} > n/\tau\}$ by $A_{i,n}$ and $\tau_1 \wedge \tau_2 = \tau$, we have that

$$\begin{aligned}
\theta^{**}_{\{1,2\}}\tau^{**}_{\{1,2\}}(\tau) &= \lim_{n\to\infty} nP\left(A_{3,n} \cap \overline{A}_{4,n} \cap \overline{A}_{5,n} \cap \overline{A}_{6,n}\right) \\
&= \lim_{n\to\infty} nP\left(\{0.1Z_1 > n/\tau\} \cap \overline{A}_{4,n} \cap \overline{A}_{5,n} \cap \overline{A}_{6,n}\right) \\
&= \lim_{n\to\infty} nP\left(\{0.1Z_1 > n/\tau\} \cap \overline{A}_{4,n}\right) \\
&= \lim_{n\to\infty} nP\left(\{0.1Z_1 > n/\tau, 0.5Z_1 \le n/\tau\} \cup \{0.1Z_1 > n/\tau, 0.5Z_1 > n/\tau\}\right) \\
&= 0.1\tau = 0.1(\tau_1 \wedge \tau_2).
\end{aligned}$$

Therefore, Proposition 2.3 indicates that $\theta(\mathbf{1})\,\Gamma(\mathbf{1}) \le 0.7 + 0.5 - 0.1 = 1.1$. The upper limit in this type of processes has no great interest since we have the theoretical expression for $\theta(\boldsymbol{\tau})$. However, this example serves to show that our upper bound can be better than the one presented in Ehlert and Schlather ([3], 2008) for M4 processes. Indeed, by applying their Corollary 3, we obtain

$$\begin{aligned}
\theta(\mathbf{1})\Gamma(\mathbf{1}) &\le \left(\Gamma(\mathbf{1}) - \bigvee_{j=1}^{2}(1-\theta_j)\right) \wedge \sum_{j=1}^{d}\theta_j \\
&= ((0.7 + 0.4 + 0.3 + 0.5) - (0.3 \vee 0.5)) \wedge 1.2 \\
&= 1.4 \wedge 1.2 = 1.2.
\end{aligned}$$

In the cases where the number of non null signatures α_{lkj}, $l \ge 1$, $-\infty < k < \infty$, $j = 1, ..., d$, of an M4 process (Smith and Weissman [23], 1996; Zhang [24], 2002) exceeds the number d of marginals, examples are easily constructed in which the Ehlert and Schlather ([3], 2008) upper limit is reduced to $\sum_{j=1}^{d}\theta_j$, being in these cases the lower limit of (2.5) below this. Our upper bound still has the advantage of being applied to processes outside the max-stable class.

3. EFFECT OF THE EXTREMAL INDEX IN THE TAIL OF A BIVARIATE EXTREME VALUES DISTRIBUTION

For each pair (j, j'), $j < j'$ belonging to D, consider the bivariate (upper) tail dependence coefficient $\chi^F_{jj'} \in [0, 1]$ for random pair $(X_{nj}, X_{nj'})$ with df $F_{jj'}$, discussed in Sibuya ([22], 1960) and Joe ([10], 1997), defined by

$$\chi^F_{jj'} = \lim_{u\uparrow 1^+} P\left(F_j(X_{ij}) > u | F_{j'}(X_{ij'}) > u\right)$$

and coefficient $\overline{\chi}^F_{jj'} \in [-1, 1]$ of Coles *et al.* ([1], 1999), defined by

$$\overline{\chi}^F_{jj'} = \lim_{u\uparrow 1^+} \frac{2\log P\left(F_{j'}(X_{ij'}) > u\right)}{\log P\left(F_j(X_{ij}) > u, F_{j'}(X_{ij'}) > u\right)} - 1.$$

We can say that $\chi^F_{jj'}$ corresponds to the probability of one variable being high given that the other is high too. The case $\chi^F_{jj'} > 0$ means asymptotic dependence between X_{nj} and $X_{nj'}$ and whenever $\chi^F_{jj'} = 0$ the variables are said to be asymptotically independent.

Assuming $\chi^F_{jj'} > 0$ within asymptotically independent data may carry to an over-estimation of probabilities of extreme joint events (see, e.g., Ledford and Tawn [12, 13], 1996, 1997). Asymptotically independent models, i.e., having $\chi^F_{jj'} = 0$, may exhibit a residual tail dependence rendering different degrees of dependence at finite levels. Coefficient $\overline{\chi}^F_{jj'}$ is a suitable tail measure within this class. Thus the pair $(\chi^F_{jj'}, \overline{\chi}^F_{jj'})$ is a useful tool in characterizing the extremal dependence: under asymptotic dependence we have $\overline{\chi}^F_{jj'} = 1$ and $0 < \chi^F_{jj'} \leq 1$ quantifies the strength of dependence between the variables $(X_{nj}, X_{nj'})$ and, within the class of asymptotic independence, we have $\chi^F_{jj'} = 0$ and $-1 \leq \overline{\chi}^F_{jj'} < 1$ measures the strength of dependence of the random pair.

Observe that, both measures can be calculated from the copula $C_{F_{jj'}}(u,u) = F_{jj'}(F_j^{-1}(u), F_{j'}^{-1}(u))$, with

$$\chi^F_{jj'} = 2 - \lim_{u\uparrow 1^+} \frac{\log C_{F_{jj'}}(u,u)}{\log u}$$

and

$$\overline{\chi}^F_{jj'} = \lim_{u\uparrow 1^+} \frac{2\log(1-u)}{\log\left(1-2u+C_{F_{jj'}}(u,u)\right)} - 1.$$

If F belongs to the max-domain of attraction of $\widehat{H}$, then $\chi^F_{jj'} = \chi^{\widehat{H}}_{jj'}$ and $\overline{\chi}^F_{jj'} = \overline{\chi}^{\widehat{H}}_{jj'}$. This results from the uniform convergence of C^n_F to $C_{\widehat{H}}$ and from $C_{F^n_{jj'}}(u,u) = \left(C_{F_{jj'}}(u^{1/n}, u^{1/n})\right)^n$. We will then have

$$\lim_{u\uparrow 1^+}\lim_{n\to\infty} \frac{\left(C_{F_{jj'}}(u^{1/n},u^{1/n})\right)^n}{C_{F_{jj'}}(u,u)} = \lim_{n\to\infty}\lim_{u\uparrow 1^+} \frac{\left(C_{F_{jj'}}(u^{1/n},u^{1/n})\right)^n}{C_{F_{jj'}}(u,u)} = 1,$$

which guarantees the constancy of $\chi^{F^n}_{jj'}$ and $\overline{\chi}^{F^n}_{jj'}$, as $n \to \infty$.

The presence of dependence among the variables of $\{\mathbf{X}_n\}$ expressed by a function $\theta(\boldsymbol{\tau})$ with values less than one, may affect the limiting behavior of $\chi^{F_n}_{jj'}$ but not the limiting behavior of $\overline{\chi}^{F_n}_{jj'}$, where F_n denotes the df of $\mathbf{M}_n$.

Proposition 3.1. *For stationary sequences* $\{\mathbf{X}_n\}$, *with multivariate extremal index* $\theta(\boldsymbol{\tau})$, $\boldsymbol{\tau} \in \mathbb{R}^d_+$, *for any choice* $j < j'$ *in* D, *we have,* $\overline{\chi}^H_{jj'} = \overline{\chi}^{\widehat{H}}_{jj'}$.

Proof: Based on the spectral representation of MEV copulas (see, e.g., Falk *et al.* [4], 2010) and relation

$$C_{H_{jj'}}(u_j, u_{j'}) = \left(C_{\widehat{H}_{jj'}}\left(u_j^{1/\theta_j}, u_{j'}^{1/\theta_{j'}}\right)\right)^{\theta\left(-\frac{\log u_j}{\theta_j}, -\frac{\log u_{j'}}{\theta_{j'}}\right)}, \tag{3.1}$$

we have

$$\overline{\chi}^{\widehat{H}}_{jj'} = \lim_{u\uparrow 1^+} \frac{2\log(1-u)}{\log\left(1-2u-C_{\widehat{H}_{jj'}}(u,u)\right)} - 1 =$$

$$= \lim_{u\uparrow 1^+} \frac{2\log(1-u)}{\log\left(1-2u-\exp\left(-\int_0^1 \left(w(-\log u)\vee(1-w)(-\log u)\right) d\widehat{W}(w)\right)\right)} - 1$$

$$= \lim_{u\uparrow 1^+} \frac{2\log(1-u)}{\log\left(1-2u-u^{-\log C_{\widehat{H}_{jj'}}(e^{-1},e^{-1})}\right)} - 1$$

where $\widehat{W}$ is the spectral measure of $\widehat{H}$. On the other hand

$$\overline{\chi}^H_{jj'} = \lim_{u\uparrow 1^+} \frac{2\log(1-u)}{\log\left(1-2u-u^{\theta_{jj'}\left(\frac{1}{\theta_j},\frac{1}{\theta_{j'}}\right)\left(-\log C_{\widehat{H}_{jj'}}\left(\exp\left(-\theta_j^{-1}\right),\exp\left(-\theta_{j'}^{-1}\right)\right)\right)}\right)} - 1.$$

Therefore,

$$(1-\overline{\chi}^H_{jj'}) = (1-\overline{\chi}^{\widehat{H}}_{jj'})A$$

with

$$A = \lim_{u\uparrow 1^+} \frac{\log\left(1-2u-u^{\Gamma(1,1)}\right)}{\log\left(1-2u-u^{\theta_{jj'}\left(\frac{1}{\theta_j},\frac{1}{\theta_{j'}}\right)\Gamma\left(\frac{1}{\theta_j},\frac{1}{\theta_{j'}}\right)}\right)}$$

$$= \lim_{u\uparrow 1^+} \frac{\log\left(1-2u-u^a\right)}{\log\left(1-2u-u^b\right)} = \lim_{u\uparrow 1^+} \frac{-2+au^{a-1}}{-2+bu^{b-1}} \lim_{u\uparrow 1^+} \frac{1-2u+u^b}{1-2u+u^a} = 1,$$

with $a=\Gamma(1,1)$ and $b=\theta_{jj'}\left(\frac{1}{\theta_j},\frac{1}{\theta_{j'}}\right)\Gamma\left(\frac{1}{\theta_j},\frac{1}{\theta_{j'}}\right)$. □

Proposition 3.2. *For stationary sequences* $\{\mathbf{X}_n\}$*, with multivariate extremal index* $\theta(\boldsymbol{\tau})$*,* $\boldsymbol{\tau}\in\mathbb{R}^d_+$*, we have, for any choice* $j<j'$ *in* D*,*

(**a**) $\chi^H_{jj'} = 2-\theta_{jj'}\left(\frac{1}{\theta_j},\frac{1}{\theta_{j'}}\right)\Gamma_{jj'}\left(\frac{1}{\theta_j},\frac{1}{\theta_{j'}}\right)$;

(**b**) $\chi^H_{jj'} - \chi^{\widehat{H}}_{jj'} = \Gamma_{jj'}(1,1) - \theta_{jj'}\left(\frac{1}{\theta_j},\frac{1}{\theta_{j'}}\right)\Gamma_{jj'}\left(\frac{1}{\theta_j},\frac{1}{\theta_{j'}}\right)$.

Proof: Using the spectral representation of MEV copulas and relation (3.1), we have

$$\chi^H_{jj'} = 2-\theta_{jj'}\left(\frac{1}{\theta_j},\frac{1}{\theta_{j'}}\right)\lim_{u\uparrow 1^+}\frac{\int_0^1\left(-\frac{\log uw}{\theta_j}\vee-\frac{\log u(1-w)}{\theta_{j'}}\right)d\widehat{W}(w)}{-\log u}$$

$$= 2-\theta_{jj'}\left(\frac{1}{\theta_j},\frac{1}{\theta_{j'}}\right)\int_0^1\left(\frac{w}{\theta_j}\vee\frac{1-w}{\theta_{j'}}\right)d\widehat{W}(w)$$

$$= 2-\left(-\theta_{jj'}\left(\frac{1}{\theta_j},\frac{1}{\theta_{j'}}\right)\log C_{\widehat{H}_{jj'}}\left(\exp(-1/\theta_j),\exp(-1/\theta_{j'})\right)\right)$$

$$= 2-\theta_{jj'}\left(\frac{1}{\theta_j},\frac{1}{\theta_{j'}}\right)\Gamma_{jj'}\left(\frac{1}{\theta_j},\frac{1}{\theta_{j'}}\right),$$

where $\widehat{W}$ is the spectral measure of $\widehat{H}$. □

The previous proposition can be rewritten in terms of the extremal coefficients $\varepsilon^H_{jj'}$ and $\varepsilon^{\widehat{H}}_{jj'}$, such that, $C_{\widehat{H}_{jj'}}(u,u) = u^{\varepsilon^{\widehat{H}}_{jj'}}$ and $C_{H_{jj'}}(u,u) = u^{\varepsilon^{H}_{jj'}}$, since these satisfy the relations $\chi^H_{jj'} = 2 - \varepsilon^H_{jj'}$ and $\chi^{\widehat{H}}_{jj'} = 2 - \varepsilon^{\widehat{H}}_{jj'}$. From (a) we conclude that $\varepsilon^H_{jj'} = \theta_{jj'}\left(\frac{1}{\theta_j}, \frac{1}{\theta_{j'}}\right)\Gamma_{jj'}\left(\frac{1}{\theta_j}, \frac{1}{\theta_{j'}}\right)$. Consequently, for the measure of asymptotic independence called madogram (Naveau *et al.* [19], 2009), defined by

$$\nu^F_{jj'} = \frac{1}{2}E\left|F_j(X_{nj}) - F_{j'}(X_{nj'})\right|$$

and satisfying

$$\nu^F_{jj'} = \frac{1}{2}\frac{\varepsilon^F_{jj'} - 1}{\varepsilon^F_{jj'} + 1},$$

we have

(**a**) $\nu^F_{jj'} = \nu^{\widehat{H}}_{jj'} = \dfrac{1}{2}\dfrac{\Gamma_{jj'}(1,1) - 1}{\Gamma_{jj'}(1,1) + 1}$;

(**b**) $\nu^H_{jj'} = \dfrac{1}{2}\dfrac{\theta_{jj'}\left(\frac{1}{\theta_j}, \frac{1}{\theta_{j'}}\right)\Gamma_{jj'}\left(\frac{1}{\theta_j}, \frac{1}{\theta_{j'}}\right) - 1}{\theta_{jj'}\left(\frac{1}{\theta_j}, \frac{1}{\theta_{j'}}\right)\Gamma_{jj'}\left(\frac{1}{\theta_j}, \frac{1}{\theta_{j'}}\right) + 1}$.

Therefore, for large n, the madogram of $(M_{nj}, M_{nj'})$ can not be taken by the madogram of $(\widehat{M}_{nj}, \widehat{M}_{nj'})$.

From relation (b) in Proposition 2.3, we conclude that

$$\chi^H_{jj'} \geq \theta^{**}_{jj'}\tau^{**}_{jj'}\left(\frac{1}{\theta_j \vee \theta_{j'}}\right) \tag{3.2}$$

and we can establish the following consequence about the value of the difference between $\chi^H_{jj'}$ and $\chi^{\widehat{H}}_{jj'}$.

Corollary 3.1. *For stationary sequences* $\{\mathbf{X}_n\}$ *satisfying the strong-mixing condition, with multivariate extremal index* $\theta(\boldsymbol{\tau})$, $\boldsymbol{\tau} \in \mathbb{R}^d_+$, *we have, for any choice* $j < j'$ *in* D,

(**a**) $\theta(\boldsymbol{\tau}) = \theta$, $\forall \boldsymbol{\tau} \in \mathbb{R}^d_+$ *implies* $\chi^H_{jj'} = \chi^{\widehat{H}}_{jj'}$;

(**b**) $\left|\chi^H_{jj'} - \chi^{\widehat{H}}_{jj'}\right| \geq \max\left\{\theta^{**}_{jj'}\tau^{**}_{jj'}\left(\dfrac{1}{\theta_j \vee \theta_{j'}}\right) - 2 + \Gamma_{jj'}(1,1), 1 - \Gamma_{jj'}(1,1)\right\}$.

Proof: (**a**) If $\theta(\boldsymbol{\tau})$ is constant equal to θ, then $\theta_j = \theta_{j'} = \theta$ and, since Γ is homogeneous of order 1, from (b) of Proposition 3.2, we have $\chi^H_{jj'} - \chi^{\widehat{H}}_{jj'} = \Gamma_{jj'}(1,1) - \Gamma_{jj'}\left(\frac{\theta}{\theta}, \frac{\theta}{\theta}\right) = 0$; (**b**) The inequality follows from (b) of Proposition 3.2 and from (3.2). □

We emphasize that the quantity $\theta^{**}_{jj'}\tau^{**}_{jj'}\left(\frac{1}{\theta_j \vee \theta_{j'}}\right)$ that we find in (3.2) and in (b) of the previous proposition reflects a tendency to the appearance of clusters within $X_{nj} \wedge X_{nj'}$ through the extremal index $\theta^{**}_{jj'}$ and

$$\tau^{**}_{jj'}\left(\frac{1}{\theta_j \vee \theta_{j'}}\right) = \lim_{n\to\infty} nP\left(X_{nj} > n(\theta_j \vee \theta_{j'}), X_{nj'} > n(\theta_j \vee \theta_{j'})\right).$$

From this discussion we conclude that:

(**i**) The tail dependencies of $\left(\widehat{M}_{n1}, \widehat{M}_{n2}\right)$ and of (M_{n1}, M_{n2}), for large n, evaluated through coefficient χ, can be considered equal when the multivariate extremal index is constant, otherwise they differ in at least

$$\max\left\{\theta^{**}_{jj'}\tau^{**}_{jj'}\left(\frac{1}{\theta_j \vee \theta_{j'}}\right) - 2 + \Gamma_{jj'}(1,1), 1 - \Gamma_{jj'}(1,1)\right\},$$

where the previous quantities can be estimated from the existing methods in literature.

(**ii**) If we estimate the dependence $\chi^F_{jj'}$ on the tail of $\left(X_{nj}, X_{nj'}\right)$, we do not obtain the dependence on the tail of (M_{n1}, M_{n2}), unless we correct the result with an estimate of $\Gamma_{jj'}(1,1) - \theta_{jj'}\left(\frac{1}{\theta_j}, \frac{1}{\theta_{j'}}\right)\Gamma_{jj'}\left(\frac{1}{\theta_j}, \frac{1}{\theta_{j'}}\right)$.

In cases where $\widehat{H}$ has totally dependent marginals ($\chi^{\widehat{H}}_{jj'} = 1$) or has independent marginals ($\chi^{\widehat{H}}_{jj'} = 0$), the previous lower limit loses interest by triviality. We underline the expression of $\chi^H_{jj'}$ in these two cases in the next result, which is derived from (a) of Proposition 3.2.

Corollary 3.2. *For stationary sequences* $\{\mathbf{X}_n\}$, *with multivariate extremal index* $\theta(\boldsymbol{\tau})$, $\boldsymbol{\tau} \in \mathbb{R}^d_+$, *we have, for any choice* $j < j'$ *in* D,

(**a**) *If* H *has independent marginals, then* $\chi^H_{jj'} = 2 - \left(\frac{1}{\theta_j} + \frac{1}{\theta_{j'}}\right)\theta_{jj'}\left(\frac{1}{\theta_j}, \frac{1}{\theta_{j'}}\right)$;

(**b**) *If* H *has totally dependent marginals, then* $\chi^H_{jj'} = 2 - \left(\frac{1}{\theta_j} \vee \frac{1}{\theta_{j'}}\right)\theta_{jj'}\left(\frac{1}{\theta_j}, \frac{1}{\theta_{j'}}\right)$.

Now we construct some examples that illustrate the cases $\chi^H_{jj'} > \chi^{\widehat{H}}_{jj'}$ and $\chi^H_{jj'} < \chi^{\widehat{H}}_{jj'}$.

Example 3.1. We first consider the following bivariate M4 process with one moving pattern,

$$\begin{cases} X_{n1} = \frac{1}{8}Z_{n-1} \vee \frac{1}{8}Z_n \vee \frac{6}{8}Z_{n+1} \\ X_{n2} = \frac{2}{8}Z_{n-1} \vee \frac{1}{8}Z_n \vee \frac{5}{8}Z_{n+1}, \end{cases}$$

where $Z_n \equiv Z_{1,n}$, $\forall\, n \geq 1$. We have in this case

$$C_F(u_1, u_2) = \left(u_1^{1/8} \wedge u_2^{2/8}\right)\left(u_1^{1/8} \wedge u_2^{1/8}\right)\left(u_1^{6/8} \wedge u_2^{5/8}\right)$$

and

$$\chi^F = \chi^{\widehat{H}} = 2 - \left(\frac{2}{8} + \frac{1}{8} + \frac{6}{8}\right) = \frac{7}{8}.$$

Otherwise

$$H(x_1, x_2) = \exp\left(-\left(\frac{6x_1^{-1}}{8} \vee \frac{5x_2^{-1}}{8}\right)\right).$$

Therefore, $C_H(u_1, u_2) = u_1 \wedge u_2$ and $\chi^H = 1 > \chi^{\widehat{H}}$.

Example 3.2. Now consider a modification in the above example through the introduction of one more pattern,

$$\begin{cases} X_{n1} = \frac{1}{8}Z_{1,n} \vee \frac{6}{8}Z_{1,n+1} \vee \frac{1}{8}Z_{2,n} \\ X_{n2} = \frac{1}{8}Z_{1,n} \vee \frac{5}{8}Z_{1,n+1} \vee \frac{2}{8}Z_{2,n}\,. \end{cases}$$

We have the same C_F and $\chi^F = \frac{7}{8}$ as in the previous example, but here

$$H(x_1,x_2) = \exp\left(-\left(\frac{6x_1^{-1}}{8} \vee \frac{5x_2^{-1}}{8}\right)\right)\exp\left(-\left(\frac{x_1^{-1}}{8} \vee \frac{2x_2^{-1}}{8}\right)\right)$$

and therefore,

$$C_H(u_1,u_2) = \left(u_1^{6/7} \wedge u_2^{5/7}\right)\left(u_1^{1/7} \wedge u_2^{2/7}\right).$$

Then $\chi^H = 2 - \left(\frac{6}{7} + \frac{2}{7}\right) = \frac{6}{7} < \chi^{\widehat{H}}$.

ACKNOWLEDGMENTS

The authors wish to thank the reviewers for their significant comments that have enhanced this work. The first author's research was partially supported by the research unit UID/MAT/00212/2013. The second author was financed by Portuguese Funds through FCT-Fundação para a Ciência e a Tecnologia within the Projects UID/MAT/00013/2013, UID/MAT/00006/2013 and by the research center CEMAT (Instituto Superior Técnico, Universidade de Lisboa) through the Project UID/Multi/04621/2013.

REFERENCES

[1] Coles, S.; Heffernan, J. and Tawn, J. (1999). Dependence measures for extreme value analyses, *Extremes* **2**(4), 339–365.

[2] Davis, R.A. (1982). Limit laws for the maximum and minimum of stationary sequences, *Z. Wahrsch. verw. Gebiete*, **61**, 31–42.

[3] Ehlert, A. and Schlather, M. (2008). Capturing the multivariate extremal index: Bounds and interconnections, *Extremes*, **11**(4), 353–377.

[4] Falk, M.; Hüsler, J. and Reiss, R.-D. (2010). *Laws of Small Numbers: Extremes and Rare Events*, 3rd ed., revised and extended, Birkhäuser, Basel and Boston.

[5] Ferreira, H. and Ferreira, M. (2015). Extremes of scale mixtures of multivariate time series, *Journal of Multivariate Analysis*, **137**, 82–99.

[6] Ferreira, H. and Ferreira, M. (2018). Estimating the extremal index through local dependence, *Annales de l'Institut Henri Poincaré – Probabilités et Statistiques*, **54**(2), 587–605.

[7] Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed., Krieger, Melbourne, Florida.

[8] Gomes, M.I.; Hall, A. and Miranda, M.C. (2008). Subsampling techniques and the Jackknife methodology in the estimation of the extremal index, *Computational Statistics & Data Analysis*, **52**(4), 2022–2041.

[9] Hsing, T. (1993). Extremal index estimation for a weakly dependent stationary sequence, *Annals of Statistics*, **21**, 2043–2071.

[10] Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.

[11] Leadbetter, M.R.; Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, Berlin.

[12] Ledford, A.W. and Tawn, J.A. (1996). Statistics for near independence in multivariate extreme values, *Biometrika*, **83**, 169–187.

[13] Ledford, A.W. and Tawn, J.A. (1997). Modelling dependence within joint tail regions, *J. R. Statist. Soc. B*, **59**, 475–499.

[14] Li, H. (2009). Orthant tail dependence of multivariate extreme value distributions, *Journal of Multivariate Analysis*, **100**(1), 243–256.

[15] Marshall, A.W. and Olkin, I. (1983). Domains of attraction of multivariate extreme value distributions, *Annals of Probability*, **11**, 168–177.

[16] Martins, A.P. and Ferreira, H. (2005). The multivariate extremal index and the dependence structure of a multivariate extreme value distribution, *TEST*, **14**(2), 433–448.

[17] Nandagopalan, S. (1990). *Multivariate extremes and estimation of the extremal index*, PhD Thesis, Department of Statistics, University of North Carolina at Chapel Hill, NC, USA.

[18] Nandagopalan, S. (1994). On the multivariate extremal index, *J. of Research, National Inst. of Standards and Technology*, **99**, 543–550.

[19] Naveau, P.; Guillou, A.; Cooley, D. and Diebolt, J. (2009). Modelling pairwise dependence of maxima in space, *Biometrika*, **96**, 1–17.

[20] Northrop, P.J. (2015). An efficient semiparametric maxima estimator of the extremal index, *Extremes*, **18**(4), 585–603.

[21] Pereira, L.; Martins, A.P. and Ferreira, H. (2017). Clustering of high values in random fields, *Extremes*, **20**, 807–838.

[22] Sibuya, M. (1960). Bivariate extreme statistics, *Annals of the Institute of Statistical Mathematics*, **11**, 195–210.

[23] Smith, R.L. and Weissman, I. (1996). *Characterization and estimation of the multivariate extremal index*, Technical Report, University of North Carolina.

[24] Zhang, Z. (2002). *Multivariate Extremes, Max-Stable Process Estimation and Dynamic Financial Modeling*, PhD Thesis, Department of Statistics, University of North Carolina at Chapel Hill, NC, USA.

REVSTAT – Statistical Journal
Volume 18, Number 4, October 2020, 521–530

A NEW EXACT CONFIDENCE INTERVAL FOR THE DIFFERENCE OF TWO BINOMIAL PROPORTIONS

Author: Wojciech Zieliński
– Department of Econometrics and Statistics, Warsaw University of Life Sciences, Nowoursynowska 159, 02-776 Warszawa, Poland
wojciech_zielinski@sggw.edu.pl

Received: July 2017 Revised: June 2018 Accepted: July 2018

Abstract:

- We consider interval estimation of the difference between two binomial proportions. Several methods of constructing such an interval are known. Unfortunately those confidence intervals have poor coverage probability: it is significantly smaller than the nominal confidence level. In this paper a new confidence interval is proposed. The construction needs only information on sample sizes and sample difference between proportions. The coverage probability of the proposed confidence interval is at least the nominal confidence level. The new confidence interval is illustrated by a medical example.

Keywords:

- *confidence interval; binomial proportions.*

AMS Subject Classification:

- 62F25, 62P10, 62P20.

1. INTRODUCTION

Let ξ_1 and ξ_2 be two independent r.v.'s distributed as $\mathrm{Bin}(n_1, \theta_1)$ and $\mathrm{Bin}(n_2, \theta_2)$, respectively. We estimate the difference between the probabilities of success, i.e. $\vartheta = \theta_1 - \theta_2$. Construction of confidence intervals for the difference of proportions has a very long history and has been widely studied, due to its numerous applications in biostatistics and elsewhere; see e.g. Anbar [1], Newcomb [7], Zhou *et al.* [12]. In all those constructions, normal approximation to the binomial distribution is applied. As a consequence it may be observed that the coverage probabilities of the asymptotic confidence intervals are less than the nominal confidence level (for a single binomial proportion see for example Brown *et al.* [3]). This is in contradiction to Neyman's [8] definition of a confidence interval. In what follows, a new confidence interval is proposed. That confidence interval is based on the exact distribution of the difference of the observed numbers of successes. A similar method was applied in constructing a confidence interval for a linear combination of proportions (W. Zieliński [16]).

The paper is organized as follows. In the second section a new confidence interval is constructed. In the third section a medical example is discussed. Some remarks and conclusions are collected in the last section. In the first appendix there is given a short R-project program for calculating proposed confidence intervals. In the second appendix some known confidence intervals for the difference of probabilities are cited.

2. A NEW CONFIDENCE INTERVAL

Let $\xi_1 \sim \mathrm{Bin}(n_1, \theta_1)$ and $\xi_2 \sim \mathrm{Bin}(n_2, \theta_2)$ be independent binomially distributed random variables. The random variable $\hat{\vartheta} = \frac{\xi_1}{n_1} - \frac{\xi_2}{n_2}$ is the minimum variance unbiased estimator of $\vartheta = \theta_1 - \theta_2$.

The confidence intervals widely used in applications are constructed in the following statistical model:

$$\left(\{0, 1, ..., n_1\} \times \{0, 1, ..., n_2\}, \left\{\mathrm{Bin}(n_1, \theta_1) \cdot \mathrm{Bin}(n_2, \theta_2),\ 0 \le \theta_1, \theta_2 \le 1\right\}\right).$$

Since we are interested in estimating $\vartheta = \theta_1 - \theta_2$ on the basis of $\hat{\vartheta}$, we consider the new statistical model

$$\left(\mathcal{X}, \left\{\mathcal{P}(n_1, n_2, \vartheta),\ -1 \le \vartheta \le 1\right\}\right),$$

where

$$\mathcal{X} = \left\{\frac{k_1}{n_1} - \frac{k_2}{n_2} :\ k_1 \in \{0, 1, ..., n_1\},\ k_2 \in \{0, 1, ..., n_2\}\right\}.$$

The family $\{\mathcal{P}(n_1, n_2, \vartheta), -1 \le \vartheta \le 1\}$ of distributions is as follows. Since for a given $\vartheta \in (-1, 1)$ the probability θ_1 is a number from the interval $(a(\vartheta), b(\vartheta))$, where

$$a(\vartheta) = \max\{0, \vartheta\} \quad \text{and} \quad b(\vartheta) = \min\{1, 1 + \vartheta\},$$

the probability of the event $\{\hat{\vartheta} = u\}$ (for $u \in \mathcal{X}$) equals (simply apply the law of total probability and averaging with respect to θ_1)

$$P_\vartheta\{\hat{\vartheta} = u\} = P_\vartheta\left\{\frac{\xi_1}{n_1} - \frac{\xi_2}{n_2} = u\right\}$$
$$= \frac{1}{L(\vartheta)} \int_{a(\vartheta)}^{b(\vartheta)} \sum_{i_2=0}^{n_2} Q_{(\theta_1,n_1)}\left\{\xi_1 = n_1\left(u + \frac{i_2}{n_2}\right)\right\} Q_{(\theta_1-\vartheta,n_2)}\{\xi_2 = i_2\}\, d\theta_1.$$

Here $L(\vartheta) = b(\vartheta) - a(\vartheta)$ and $Q_{(\mu,m)}\{\zeta = k\} = \binom{m}{k}\mu^k(1-\mu)^{m-k}$ for $k = 0, 1, ..., m$.

Note that the family $\{\mathcal{P}(n_1, n_2, \vartheta), -1 \le \vartheta \le 1\}$ of distributions is decreasing in ϑ, i.e. for a given $u \in \mathcal{X}$,

$$P_{\vartheta_1}\{\hat{\vartheta} \le u\} \ge P_{\vartheta_2}\{\hat{\vartheta} \le u\} \quad \text{for} \quad \vartheta_1 < \vartheta_2.$$

It follows from that fact that the family of binomial distributions is decreasing in probability of a success and $P_\vartheta\{\hat{\vartheta} = u\}$ is a convex combination of binomial distributions.

Let $\hat{\vartheta} = u$ be observed. The (symmetric) confidence interval for ϑ at confidence level γ based on the exact distribution of $\hat{\vartheta}$ is $(\vartheta_L(u), \vartheta_U(u))$, where

$$(M) \qquad \vartheta_L(u) = \begin{cases} -1 & \text{for } u = -1, \\ \max\left\{\vartheta : P_\vartheta\{\hat{\vartheta} < u\} = \frac{1+\gamma}{2}\right\} & \text{for } u > -1, \end{cases}$$
$$\vartheta_U(u) = \begin{cases} 1 & \text{for } u = 1, \\ \min\left\{\vartheta : P_\vartheta\{\hat{\vartheta} \le u\} = \frac{1-\gamma}{2}\right\} & \text{for } u < 1. \end{cases}$$

Unfortunately, closed formulae for such confidence intervals are not available. Nevertheless, for given n_1, n_2 and observed u the confidence interval may be easily obtained with the standard mathematical software (for example R-project, Mathematica, MathLab etc.). Table 1 presents some 95% confidence intervals for $n_1 = n_2 = 10$ and Table 2 for $n_1 = 50$, $n_2 = 10$.

Table 1: Confidence intervals ($\gamma = 0.95$, $n_1 = n_2 = 10$).

$\hat{\vartheta}$	interval	$\hat{\vartheta}$	interval
−1.0	(−1.0000, −0.6733)	0.1	(−0.3319, 0.5171)
−0.9	(−0.9975, −0.5214)	0.2	(−0.2326, 0.6019)
−0.8	(−0.9751, −0.3940)	0.3	(−0.1291, 0.6813)
−0.7	(−0.9350, −0.2798)	0.4	(−0.0212, 0.7551)
−0.6	(−0.8832, −0.1745)	0.5	(0.0760, 0.8227)
−0.5	(−0.8227, −0.0760)	0.6	(0.1745, 0.8832)
−0.4	(−0.7551, 0.0212)	0.7	(0.2798, 0.9350)
−0.3	(−0.6813, 0.1291)	0.8	(0.3940, 0.9751)
−0.2	(−0.6019, 0.2326)	0.9	(0.5214, 0.9975)
−0.1	(−0.5171, 0.3319)	1.0	(0.6733, 1.0000)
0.0	(−0.4270, 0.4270)		

Table 2: Confidence intervals ($\gamma = 0.95$, $n_1 = 50, n_2 = 10$).

$\hat{\vartheta}$	interval	$\hat{\vartheta}$	interval
-1.0	$(-1.0000, -0.8346)$	0.1	$(-0.2103, 0.4135)$
-0.9	$(-0.9949, -0.6642)$	0.2	$(-0.1046, 0.5073)$
-0.8	$(-0.9563, -0.5302)$	0.3	$(\ 0.0023, 0.5962)$
-0.7	$(-0.8986, -0.4105)$	0.4	$(\ 0.0971, 0.6801)$
-0.6	$(-0.8322, -0.2998)$	0.5	$(\ 0.1957, 0.7590)$
-0.5	$(-0.7590, -0.1957)$	0.6	$(\ 0.2998, 0.8322)$
-0.4	$(-0.6801, -0.0971)$	0.7	$(\ 0.4105, 0.8986)$
-0.3	$(-0.5962, -0.0023)$	0.8	$(\ 0.5302, 0.9563)$
-0.2	$(-0.5073,\ \ 0.1046)$	0.9	$(\ 0.6642, 0.9949)$
-0.1	$(-0.4135,\ \ 0.2103)$	1.0	$(\ 0.8346, 1.0000)$
0.0	$(-0.3145,\ \ 0.3145)$		

For a given $\vartheta \in (-1, 1)$ the coverage probability, by construction, equals

$$\sum_{u=F_\vartheta^{-1}((1-\gamma)/2)}^{F_\vartheta^{-1}((1+\gamma)/2)} P_\vartheta\{\hat{\vartheta} = u\},$$

where $F_\vartheta^{-1}(\cdot)$ is the quantile function of the distribution of $\hat{\vartheta}$. Since the distribution of $\hat{\vartheta}$ is discrete, the coverage probability is at least γ. Figure 1 shows the coverage probability of the confidence interval (M) for $\gamma = 0.95$ (the coverage probability is calculated not simulated).

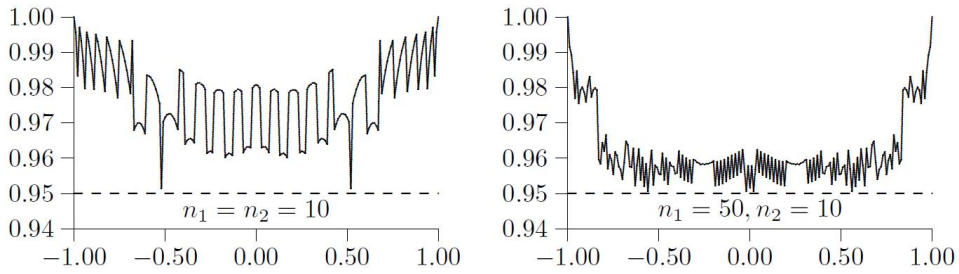

Figure 1: The probability of coverage, $\gamma = 0.95$.

The length of the confidence interval depends on the sample sizes n_1 and n_2. Suppose we may conduct n trials including n_1 trials with success probability θ_1 and $n_2 = n - n_1$ trials with probability θ_2. To find the optimal n_1, i.e. one minimizing the length, it is enough to minimize the distance between quantiles of orders $\frac{1+\gamma}{2}$ and $\frac{1-\gamma}{2}$ of the distribution of $\hat{\vartheta}$. It is easy to note that the distribution of $\hat{\vartheta}$ is unimodal, so it is enough to minimize the variance of $\hat{\vartheta}$. This variance equals

$$D_\vartheta^2(\hat{\vartheta}) = \frac{1}{L(\vartheta)} \int_{a(\vartheta)}^{b(\vartheta)} \left(D^2_{(\theta_1, n_1)}\left(\frac{\xi_1}{n_1}\right) + D^2_{(\theta_1 - \vartheta, n_2)}\left(\frac{\xi_2}{n_2}\right) \right) d\theta_1 = \frac{1 - 3\vartheta^2 + 2|\vartheta|^3}{6nf(1-f)},$$

where $f = n_1/n$. The variance $D_\vartheta^2(\hat{\vartheta})$ is (uniformly in ϑ) minimal for $f = 1/2$, i.e. half of the trials should be done with probability θ_1. Hence, to obtain the maximal precision of

estimation, i.e. the shortest (symmetric) confidence interval, the number of trials should be equally divided between the two groups. Of course this is possible in the case of a planned experiment. Unfortunately, in many real experiments (especially medical ones) it is not possible to have planned experiments.

3. A MEDICAL EXAMPLE

The aim of the investigation was to compare the frequencies of occurrence of the specific immunoglobulin E G6 (*Phleum pratense* L.) in two sites: urban (represented by the Polish town Lublin) and rural (represented by the Polish district Zamość). The investigation is part of the ECAP project (ecap.pl/eng_www/index_-home.html) conducted by Prof. Bolesław Samoliński (Warsaw Medical University). The data are presented by his courtesy.

Let θ_t and θ_c denote the percentages of people with high concentration of sIgE G6 (at least 0.35 IU/ml) in the town and in the country, respectively. We are interested in estimating the difference $\theta_t - \theta_c$ at confidence level 0.95. A sample of size $n_t = 743$ was drawn from the town, and a sample of size $n_c = 329$ from the country. The difference between the sample proportions equals 0.0603. The confidence interval for the difference of proportions $\theta_t - \theta_c$ at confidence level 0.95 is $(0.0052, 0.1154)$ (calculated from formula (M) with $u = 0.0603$). Since the lower end of the confidence interval is positive, we may conclude that the fraction of people with allergy to *Phleum pratense* L. is higher in the town than in the country.

In the above samples the level of the specific immunoglobulin E D1 (*Dermatophagoides pteronyssinus*) was also marked. The question is the same as in the previous investigation: what is the difference between percentages of people with allergy to *Dermatophagoides pteronyssinus* in urban and in rural areas. The difference between the observed proportions is 0.0292 and confidence interval, at confidence level 0.95, is $(-0.0276, 0.0853)$. Since the confidence interval covers 0, it may be supposed that the percentages of people with allergy to that allergen are the same.

4. DISCUSSION AND CONCLUSIONS

Estimating the difference of two binomial proportions is one of the crucial problems in medicine, biometrics etc. In this paper a new confidence interval for that difference is proposed. The confidence interval is based on the exact distribution of the sample difference, hence it works for large as well as for small samples. The coverage probability of that confidence interval is at least the nominal confidence level, in contrast to asymptotic confidence intervals known in the literature. It must be noted that the only information needed to construct the new confidence interval is sample sizes and sample difference between proportions, while for the confidence intervals appearing in the literature the knowledge of sample sizes as well as sample proportions in each sample is needed. Unfortunately it may lead to misunderstandings. Namely, suppose that seven experiments were conducted. In each experiment two samples of sizes fifty and ten respectively, were drawn ($n_1 = 50$, $n_2 = 10$). The resulting numbers of successes are shown in Table 3 (the first two columns).

Table 3: Confidence intervals in seven experiments.

ξ_1	ξ_2	$\hat{\vartheta}$	Wang c.i.	K_1 c.i.	K_2 c.i.
16	0	0.32	(0.04738; 0.47101)	(0.01975; 0.62025)	(0.19070; 0.44930)
21	1	0.32	(−0.00273; 0.50696)	(−0.00719; 0.64719)	(0.08915; 0.55085)
26	2	0.32	(−0.03047; 0.55617)	(−0.01873; 0.65873)	(0.03602; 0.60398)
31	3	0.32	(−0.02693; 0.58380)	(−0.01645; 0.65645)	(0.00571; 0.63429)
36	4	0.32	(−0.02108; 0.61329)	(−0.00007; 0.64007)	(−0.00816; 0.64816)
41	5	0.32	(0.00656; 0.62735)	(0.03283; 0.60717)	(−0.00769; 0.64769)
46	6	0.32	(0.03955; 0.63766)	(0.08920; 0.55080)	(0.00718; 0.63282)

It is seen that the sample difference between proportions (the third column) is the same in all experiments, but the confidence intervals are quite different (Table 3 gives results for three confidence intervals, but for other confidence intervals the results are similar). Moreover, for example application of (K_1) or Wang confidence intervals in the sixth experiment suggests that $\hat{\vartheta} = 0.32$ is a statistically significant difference while in the fourth one it is not. The confidence interval (M) we propose does not have this drawback: for observed $\hat{\vartheta}$ we obtain one confidence interval whatever ξ_1 and ξ_2 are (here it is $(0.02110; 0.61120)$).

Closed formulae for the new confidence interval are not available. But it is easy to calculate the confidence interval for given n_1, n_2 and an observed sample difference $\hat{\vartheta}$ (see Appendix 1 for an exemplary R code). Because the proposed confidence interval may be applied for small as well as for large sample sizes, it may be recommended for practical use.

The coverage probability of the proposed confidence interval is at least the nominal confidence level. The equality of the coverage probability and the confidence level may be obtained by an appropriate randomization. The idea of randomized confidence intervals is presented for example in R. Zieliński and W. Zieliński [13], W. Zieliński [15], [16]. The same idea may be applied to the proposed confidence interval; work on this is in progress.

APPENDIX 1

An exemplary R code for calculating the confidence interval is enclosed. I am grateful to Prof. Stanisław Jaworski for his help.

```
CI=function(uemp,n,gamma){
u=abs(uemp)
g=function(u,vartheta,lq=0){
f=function(theta,k){pbinom(n[1]*(u+k/n[2])-lq,n[1],theta)*dbinom(k,n[2],theta-vartheta)}
a=max(0,vartheta)
b=min(1,1+vartheta)
wynik=c()
for (k in 0:n[1]){wynik[k+1]=integrate(f,a,b,k=k)$value }
t=sum(wynik)/(b-a)
(t-(1+gamma*(-1+2*lq))/2)^2}
P=ifelse(u==1,1,optimize(g,c(u,1),u=u)$minimum) # upper
L=optimize(g,c(-1,u),u=u,lq=1)$minimum # lower
info=paste("at 1-alpha=",gamma,", where u=",uemp, ", n1=",n[1],", n2=",n[2],sep="")
if (uemp>0)
{paste("Confidence interval (",round(L,4),",",round(P,4),") ",info,sep="")}
else
{paste("Confidence interval (",round(-P,4),",",round(-L,4),") ",info,sep="")}
}
#Example of usage
n=c(10,10) # input n1 and n2
CI(-0.3,n,gamma=0.99) # input the observed difference and the confidence level
```

APPENDIX 2

Confidence intervals for $\vartheta = \theta_1 - \theta_2$ appearing in the literature are constructed for "large" sample sizes n_1 and n_2. It is assumed that ξ_1 and ξ_2 (and so $\xi_1 - \xi_2$) are normally distributed. In what follows, γ denotes the assumed confidence level and $z = z_{(1+\gamma)/2}$ denotes the quantile of order $(1+\gamma)/2$ of the standard normal distribution.

1. The approximate confidence interval based on the test statistic of the hypothesis $H\colon \theta_1 = \theta_2$ has the form

$$(K_1) \qquad \hat{\vartheta} \pm z\sqrt{\frac{\xi_1+\xi_2}{n_1+n_2}\left(1-\frac{\xi_1+\xi_2}{n_1+n_2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}.$$

This is one of the most common confidence intervals. It may be found in various statistical textbooks (https://onlinecourses.science.psu.edu/stat414/node/268 for example).

2. By the de Moivre-Laplace theorem, $\hat{\vartheta} \sim N\left(\theta, \frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}\right)$ asymptotically. A simple application of the asymptotic distribution gives

$$(K_2) \qquad \hat{\vartheta} \pm z\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}}$$

(for example stattrek.com/estimation/difference-in-proportions.aspx?Tutorial=AP). Mee and Anbar [5] expressed the above interval in terms of $\hat{\vartheta}$:

$$\hat{\vartheta} \pm z\sqrt{\frac{(\tilde{\psi}+\hat{\vartheta}/2)(1-\tilde{\psi}-\hat{\vartheta}/2)}{n_1} + \frac{(\tilde{\psi}-\hat{\vartheta}/2)(1-\tilde{\psi}+\hat{\vartheta}/2)}{n_2}},$$

where $\tilde{\psi} = (\hat{\theta}_1 + \hat{\theta}_2)/2$.

Miettinen and Nurminen [6] slightly modified the above confidence interval:

$$(K_2') \qquad \hat{\vartheta} \pm z\sqrt{\frac{n_1+n_2}{n_1+n_2-1}\left\{\frac{(\tilde{\psi}+\hat{\vartheta}/2)(1-\tilde{\psi}-\hat{\vartheta}/2)}{n_1} + \frac{(\tilde{\psi}-\hat{\vartheta}/2)(1-\tilde{\psi}+\hat{\vartheta}/2)}{n_2}\right\}}.$$

3. The binomial distribution is a discrete one and is approximated by a continuous distribution. Hence the so called continuity correction is introduced (Fleiss [4], p. 29):

$$(K_3) \qquad \hat{\vartheta} \pm z\sqrt{\frac{\xi_1(n_1-\xi_1)}{n_1^3} + \frac{\xi_2(n_2-\xi_2)}{n_2^3} + \frac{1}{2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}.$$

This confidence interval is very conservative: its coverage probability is significantly higher than the assumed confidence level.

4. Using the Haldane method, Beal [2] obtained the confidence interval

$$(K_4) \qquad \vartheta^* \pm w,$$

where

$$\vartheta^* = \frac{\hat{\vartheta} + z^2\nu(1-2\tilde{\psi})}{1+z^2u},$$

$$w = \frac{z}{1+z^2u}\sqrt{u\{4\tilde{\psi}(1-\tilde{\psi}) - \hat{\vartheta}^2\} + 2\nu(1-2\tilde{\psi})\hat{\vartheta} + 4z^2u^2(1-\tilde{\psi})\tilde{\psi} + z^2\nu^2(1-2\tilde{\psi})^2},$$

$$\tilde{\psi} = \frac{1}{2}\left(\hat{\theta}_1 + \hat{\theta}_2\right) \quad u = \frac{1}{4}\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \quad \nu = \frac{1}{4}\left(\frac{1}{n_1} - \frac{1}{n_2}\right).$$

Using the Jeffreys-Perks method he obtained a similar confidence interval with

$$(K_4') \qquad \tilde{\psi} = \frac{1}{2}\left(\frac{\xi_1 + 0.5}{n_1+1} + \frac{\xi_2 + 0.5}{n_2+1}\right).$$

5. The method based on the Wilson [11] score method for the single proportion gives the confidence interval

$$(K_5) \qquad L = \hat{\vartheta} - \delta_{12}, \quad U = \hat{\vartheta} + \delta_{21},$$

where

$$\delta_{ij} = \sqrt{(\hat{\theta}_i - l_i)^2 + (u_j - \hat{\theta}_j)^2} = z\sqrt{l_i(1-l_i)/n_i + u_j(1-u_j)/n_j}$$

and l_i and u_i are the roots of $|\hat{\theta}_i - \theta_i| = z\sqrt{\theta_i(1-\theta_i)/n_i}$. Note that $l_i = 0$ for $\xi_i = 0$ and $u_i = 1$ for $\xi_i = n_i$.

Using the continuity-correction score intervals, Fleiss [4] (pp. 13–14) obtained l_i and u_i as the solutions of

$$(K_5') \qquad \left|\hat{\theta}_i - \theta_i\right| - \frac{1}{2n_i} = z\sqrt{\frac{\theta_i(1-\theta_i)}{n_i}}.$$

6. Zhou *et al.* [12] proposed two new confidence intervals based on the asymptotic Edgworth expansion of $\hat{\theta}_1 - \hat{\theta}_2$. The first one is

$$(K_6) \qquad \left(\hat{\vartheta} - \frac{\hat{\sigma}}{\sqrt{n}}\left(z - \frac{\hat{Q}(z)}{\sqrt{n}}\right), \hat{\vartheta} + \frac{\hat{\sigma}}{\sqrt{n}}\left(z + \frac{\hat{Q}(z)}{\sqrt{n}}\right)\right),$$

where $(n = n_1 + n_2)$

$$\hat{Q}(t) = \frac{\hat{a} + \hat{b}t^2}{\hat{\sigma}}, \ \hat{\sigma} = \sqrt{n}\sqrt{\frac{\xi_1(n_1-\xi_1)}{n_1^3} + \frac{\xi_2(n_2-\xi_2)}{n_2^3}}, \ \hat{a} = \frac{\hat{\delta}}{6\hat{\sigma}^2}, \ \hat{b} = \frac{n(n_1-2\xi_1)}{2n_1^2} - \hat{a},$$

$$\hat{\delta} = \left(\frac{n}{n_1}\right)^2 \frac{\xi_1(n_1-\xi_1)(n_1-2\xi_1)}{n_1^3} - \left(\frac{n}{n_2}\right)^2 \frac{\xi_2(n_2-\xi_2)(n_2-2\xi_2)}{n_2^3}.$$

The second confidence interval has the form

$$(K_7) \qquad \left(\hat{\vartheta} - \frac{\hat{\sigma}}{\sqrt{n}}g^{-1}(z), \hat{\vartheta} - \frac{\hat{\sigma}}{\sqrt{n}}g^{-1}(-z)\right),$$

where

$$g^{-1}(u) = \frac{\sqrt{n}}{\hat{b}\hat{\sigma}}\left(\left(1 + 3(\hat{b}\hat{\sigma})\left(\frac{u}{\sqrt{n}} - \frac{\hat{a}}{\hat{\sigma}}n\right)\right)^{1/3} - 1\right).$$

The upper ends of the above mentioned confidence intervals may be greater than one (or their lower ends may be smaller than -1). It is customary to truncate such an interval at 1 (or -1 respectively), but such an operation results in a very low coverage probability for values of ϑ near 1 (or -1 respectively).

Wang [10] (see also Shan and Wang [9]) proposed a confidence interval which does not have the above disadvantage.

REFERENCES

[1] Anbar, D. (1983). On estimating the difference between two probabilities with special reference to clinical trials, *Biometrics*, **39**, 257–262.

[2] Beal, S.L. (1987). Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples, *Biometrics*, **43**, 941–950.

[3] Brown, L.D.; Cai, T.T. and DasGupta, A. (2001). Interval estimation for a binomial proportion, *Statistical Science*, **16**, 101–133.

[4] Fleiss, J.L. (1981). *Statistical Methods for Rates and Proportions*, 2nd ed., Wiley, New York.

[5] Mee, R.W. and Anbar, D. (1984). Confidence bounds for the difference between two probabilities, *Biometrics*, **40**, 175–176.

[6] Miettinen, O.S. and Nurminen, M. (1985). Comparative analysis of two rates, *Statistics in Medicine*, **4**, 213–226.

[7] Newcombe, R. (1998). Interval estimation for the difference between independent proportions: comparison of eleven methods, *Statistics in Medicine*, **17**, 873–890.

[8] Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection, *Journal of the Royal Statistical Society*, **97**, 558–625.

[9] Shan, G. and Wang, W. (2013). ExactCIdiff: An R package for computing exact confidence intervals for the difference of two proportions, *R Journal*, **5/2**, 62–70.

[10] Wang, W. (2010). On construction of the smallest one-sided confidence interval for the difference of two proportions, *The Annals of Statistics*, **38**, 1227–1243.

[11] Wilson, E.B. (1927). Probable inference, the law of succession, and statistical inference, *Journal of the American Statistical Association*, **22**, 209–212.

[12] Zhou, X-H.; Tsao, M. and Qin, G. (2004). New intervals for the difference between two independent binomial proportions, *J. Statist. Plann. Inference*, **123**, 97–115.

[13] Zieliński, R. and Zieliński, W. (2005). Best exact nonparametric confidence intervals for quantiles, *Statistics*, **39**, 67–71.

[14] Zieliński, W. (2014). The shortest randomized confidence interval for probability of success in a negative binomial model, *Applicationes Mathematicae*, **41**, 43–49.

[15] Zieliński, W. (2017). The shortest Clopper–Pearson randomized confidence interval for binomial probability, *REVSTAT – Statistical Journal*, **15**, 141–153.

[16] Zieliński, W. (2018). Confidence interval for the weighted sum of two binomial proportions, *Applicationes Mathematicae*, **45**, 53–60.

REVSTAT – Statistical Journal
Volume 18, Number 4, October 2020, 531–551

ON BAYESIAN ANALYSIS OF SEEMINGLY UNRELATED REGRESSION MODEL WITH SKEW DISTRIBUTION ERROR

Authors: Omid Akhgari
– Department of Statistics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran
o.akhgari@modares.ac.ir

Mousa Golalizadeh
– Department of Statistics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran
golalizadeh@modares.ac.ir

Received: May 2017 Revised: January 2018 Accepted: July 2018

Abstract:

- The simultaneous equation models (SEMs) are one of the standard statistical tools for analyzing multivariate regression when the errors are correlated with some covariates. A particular version of the SEMs is the Seemingly Unrelated Regression (SUR) models which consist of several regression equations with errors being correlated across the equations. There are many occasions in which the normality assumption for the error term might not hold in these models. Although transforming the error to comply with the normal density is a solution, the interpretation of the estimators for the parameters and the associated model might not be straightforward. However, taking into account the skew-normal distribution for the error might, sometimes, be a good alternative. In this paper such scenario is considered as well as a Bayesian framework to estimate the parameters, with a brief review of frequentist methodology. The full conditional posterior densities are derived and relevant statistical inferences are provided. A simulation study is conducted to evaluate the performance of the proposed method. Also, the utilized model is applied to fit relevant equations on Iran gross and income data collected in the year 2009.

Keywords:

- *Simultaneous Equation Model; skew-normal distribution; Bayesian inference; Markov chain Monte Carlo; gross and income.*

AMS Subject Classification:

- 62J20, 62J99.

1. INTRODUCTION

There are many examples in which a single equation can represent a causal relationship among variables. However, there is a case in which individual expression may not cover the desired effect or produce estimates with weak statistical properties. There are examples from many scientific fields such as econometrics where single equations are not enough. In such cases, Simultaneous Equations Models (SEMs) can appropriately represent a joint relationship among variables. The interested readers can, for example, consult [19] for more details on this topic.

Regarding the assumptions of the general structure of any linear model, the predictors are not only fixed but also independent of the error term in the model. However, there are numerous real-life examples in which some of the covariates in a model are correlated with the error term. According to [30] and [11], such variables are called endogenous.

One of the particular cases of SEMs is SUR models, first proposed by [31]. He also explained the procedures to estimate the parameters of these models using the generalized least square method. Amazingly, the literature on treating such models from the frequentist point of view is scarce. For instance, the well known maximum likelihood method of estimation for the parameters of the SUR was only tackled by [13]. However, the popularity of Bayesian approach was more than expected. To name some, we can refer to [28], [21], [29] and [12]. Historically, Bayesian statistical inference on the SUR model was first proposed by [32]. Also, Bayesian moment and direct Monte Carlo method were followed by [33]. Most literature shows that popular MCMC sampling technique was the central theme of study to treat the SUR model . The references include [24], [10], and [27]. Also, [35] proposed the implementation of hierarchical Bayes approach in this model using direct Monte Carlo and importance sampling techniques. Recently, [26] studied the topic of variable selection in the SUR models.

Another important aspect of the SUR models refers to a way one considers a distribution for the error term. It is quite common to choose it as normal. But, there are numerous examples in which the empirical density of response is often asymmetric in practice. One of the procedures to overcome this problem is to utilize some transformations. It might induce relatively normal distribution for the transformed response. However, this strategy has some drawbacks. First, the estimators are usually bias. Secondly, there is lack of proper interpretation for the estimators of the parameters based on the transformed response. Using some asymmetric distributions, which not only possess the same properties as the normal distribution but also can overcome the deficiencies mentioned above, has recently received considerable attention in the literature. The skew-normal density, initially proposed by [3], is one of the well-known distributions to tackle the asymmetric feature of the data. [5] have also discovered the properties of the multivariate skew normal distribution. Later on, [4] studied further features of this density. [7], [17], [18] and [2], among others, provided several generalizations of this distribution. Recently, [6] investigated some other properties of the skew-symmetric distribution.

Most of the research conducted to estimate parameters of a SUR is focused on the case in which the distribution of the variable under investigation is normal. Instead, in this paper

we consider the skew-normal distribution for the errors in the SUR and propose procedures to estimate the parameters using the Bayesian methodology. We also conduct some intensive simulation studies to evaluate the methods suggested in this article. Moreover, we show an application of the model in this paper on real-life data.

To present our results, we organized the paper as follows. First, a brief review of the Seemingly Unrelated Regression (SUR) and a Bayesian approach to treating the SUR model with normal distribution for the errors are presented in Section 2. Then, a Bayesian approach to treating the SUR model with skew-normal distribution for the errors is given in Section 3. The simulation study for evaluating the proposed models and the analyses of real-life data, related to the gross and income in the year 2009 in Iran, for illustration purpose, are presented in Sections 4 and 5. General conclusions are provided at the end. The proofs for some theoretical results are sketched in Appendix.

2. BAYESIAN INFERENCE ON PARAMETERS OF A SUR MODEL WITH ERROR DISTRIBUTED AS NORMAL

Econometric analysis of the linear models are usually classified into two scenarios which identify based on the numbers of the equations used to express the relationship among the variables. In the single equation methodology, a dependent variable is typically modeled as a function of one or more covariates. In many situations, such single equation may not cover the desired effect or may even produce estimates with poor statistical properties. The methods of SUR model have been proposed to eliminate the shortcomings obstacles involved in the former methodology. Statistical inference based on the normal response in this model is the object of the current section.

Let assume we aim to estimate the parameters of a SUR model. This objective can commonly be achieved via many parametric and nonparametric estimating procedures based on the frequentist inference including OLS[1], IOLS[2], FGLS[3], IGLS[4] and ML[5]. See, for example, [28] for more details on this topic. However, there are some problems to implement the ML method of estimation in a SUR model. First, there are not usually some explicit expressions for the estimators of the parameters. This fact leads, in turn, to a high cost of analytical computations to solve corresponding normal equations. Secondly, if there is any initial subjective information about the parameters it cannot be directly utilized in the frequentist inference methodology. To overcome these two problems, one can follow a Bayesian approach instead. This section describes the procedure to perform such inference along with general notations used throughout the current paper.

Suppose there are g equations with g endogenous variables associating with $y_1, y_2, ..., y_g$. Specifically, for $i = 1, 2, ..., g$, suppose $X^{(i)}$ is an $n \times k_i$ matrix of explanatory variables and $\beta^{(i)}$ is a k_i-vector of parameters. Then, the i-th equation of a linear simultaneous system can

[1]Ordinary Least Square
[2]Iterative Ordinary Least Square
[3]Feasible Generalized Least Squares
[4]Iterative Generalized Least Squares
[5]Maximum Likelihood

be written as

$$y_{ti} = \sum_{l=1}^{k_i} x_{tl}^{(i)} \beta_l^{(i)} + u_{ti} = X_{t\bullet}^{(i)} \boldsymbol{\beta}_{t\bullet}^{(i)} + u_{ti}, \quad t = 1, 2, ..., n, \tag{2.1}$$

where

$$X_{t\bullet}^{(i)} = (x_{t1}, x_{t2}, ..., x_{tk_i}),$$
$$\boldsymbol{\beta}_{t\bullet}^{(i)} = (\beta_1^{(i)}, \beta_2^{(i)}, ..., \beta_{k_i}^{(i)})^{\mathsf{T}},$$

and

$$E(u_{ti}) = 0, \qquad \mathrm{Var}(u_{ti}) = \sigma_{ii},$$
$$\mathrm{Cov}(u_{ti}, u_{tj}) = \sigma_{ij}, \quad i, j = 1, 2, ..., g, \quad t = 1, 2, ..., n. \tag{2.2}$$

Let us define, for fixed t, the g-vectors $y_{t\bullet}$ and $u_{t\bullet}$ consist of the y_{ti}'s and the u_{ti}'s, respectively, for $i = 1, ..., g$. Accordingly, the k-vector $\beta_\bullet$ is formed by stacking the $\boldsymbol{\beta}_{t\bullet}^{(i)}$ vertically. The matrix of $X_{t\bullet}$ is of dimension $g \times k$ and is defined to be a block-diagonal matrix with diagonal blocks $X_{t\bullet}^{(i)}$ also for fixed t, $k = \sum_{i=1}^{g} k_i$. Precisely, our new notations can be summarized as follows:

$$y_{t\bullet} = \begin{pmatrix} y_{t1} \\ y_{t2} \\ \vdots \\ y_{tg} \end{pmatrix}_{g\times 1}, \qquad u_{t\bullet} = \begin{pmatrix} u_{t1} \\ u_{t2} \\ \vdots \\ u_{tg} \end{pmatrix}_{g\times 1}, \qquad \beta_\bullet = \begin{pmatrix} \boldsymbol{\beta}_{t\bullet}^{(1)} \\ \boldsymbol{\beta}_{t\bullet}^{(2)} \\ \vdots \\ \boldsymbol{\beta}_{t\bullet}^{(g)} \end{pmatrix}_{k\times 1},$$

$$X_{t\bullet} = \begin{pmatrix} X_{t\bullet}^{(1)} & 0 & ... & 0 \\ 0 & X_{t\bullet}^{(2)} & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & X_{t\bullet}^{(g)} \end{pmatrix}_{g\times k}. \tag{2.3}$$

Hence, the linear simultaneous system (2.1) is rewritten as follows

$$y_{t\bullet} = X_{t\bullet}\beta_\bullet + u_{t\bullet}, \quad t = 1, 2, ..., n. \tag{2.4}$$

Based on the assumption for the first two moments of u's, let us consider the normal distribution for them. Then, following the new notations, $u_{t\bullet} \sim N(0_g, \Sigma)$ where

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & ... & \sigma_{1g} \\ \sigma_{21} & \sigma_{22} & ... & \sigma_{2g} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{g1} & \sigma_{g2} & ... & \sigma_{gg} \end{pmatrix}. \tag{2.5}$$

Now, recalling the expression (2.4) and distribution $u_{t\bullet}$, the likelihood function for the parameters $(\beta_\bullet, \Sigma)$, provide the data including those available in $y_\bullet$ and $X_\bullet$, represented by D, leads to

$$L\big((\beta_\bullet, \Sigma)|D\big) = \frac{1}{(2\pi)^{ng/2}|\Sigma|^{n/2}} \exp\Big\{ -\frac{1}{2} \mathrm{tr}\left(V\Sigma^{-1}\right)\Big\}, \tag{2.6}$$

where 'tr' denotes the trace of matrix and V is a $g \times g$ matrix given by

$$V = \sum_{t=1}^{n} (y_{t\bullet} - X_{t\bullet}\beta_\bullet)(y_{t\bullet} - X_{t\bullet}\beta_\bullet)^{\mathsf{T}}.$$

Now, suppose one prefers to follow a Bayesian methodology to estimate the parameters of the SUR model (2.4). As is common, one first should determine priors for the parameters. Both the noninformative and informative priors can be used here. Let us assume a uniform prior for $\beta_\bullet$ and Jeffreys prior for Σ, independent of each other [20]. Then, we have our joint prior, say $\pi_1(\,.\,)$, as

$$\pi_1(\beta_\bullet, \Sigma) = \pi(\beta_\bullet)\pi(\Sigma) \propto |\Sigma|^{-\frac{g+1}{2}}. \tag{2.7}$$

The joint posterior density function is then given by Bayes' theorem, i.e.

$$\pi(\beta_\bullet, \Sigma|D) \propto |\Sigma|^{-(n+g+1)/2} \exp\big[-\frac{1}{2}\operatorname{tr}\{V\Sigma^{-1}\}\big]. \tag{2.8}$$

Now, it is straightforward to compute the full conditional posterior distribution $\pi(\beta_\bullet|\Sigma, D)$ and $\pi(\Sigma|\beta_\bullet, D)$. They are given by

$$\begin{aligned} \beta_\bullet|(\Sigma, D) &\sim N(\hat{\beta}_\bullet, \hat{\Sigma}_{\beta_\bullet}) \\ \Sigma|(\beta_\bullet, D) &\sim IW(V, n), \end{aligned} \tag{2.9}$$

where

$$\begin{aligned} \widehat{\beta}_\bullet &= \widehat{\Sigma}_{\beta_\bullet}\big(\sum_{t=1}^{n} X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}y_{t\bullet}\big), \\ \widehat{\Sigma}_{\beta_\bullet} &= \big[\sum_{t=1}^{n} X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}X_{t\bullet}\big]^{-1}, \end{aligned} \tag{2.10}$$

and $IW(\cdot,\cdot)$ denotes the inverse Wishart distribution. As seen, both full conditional posterior distributions have closed forms. Hence, the standard SUR model is also amenable to a 2-block Gibbs sampling formulation. See, for example, [34], for more details.

In some circumstances, one might prefer an informative prior for the parameters $\beta_\bullet$. In such case, it is common to consider the normal density. Precisely, let assume $\beta_\bullet \sim N(\beta_\circ, A_{\beta_\bullet}^{-1})$. Further, suppose the same prior as before has been considered for Σ, i.e. $\pi\big(\Sigma\big) \propto \big|\Sigma\big|^{-\frac{g+1}{2}}$, independently from $\beta_\bullet$. Then, the joint posterior distribution has a closed form in this case as well. However, the conditional posterior distributions have relatively different structures. In particular, it can be shown that

$$\begin{aligned} \beta_\bullet|(\Sigma, D) &\sim N(\overline{\beta}_\bullet, \overline{\Sigma}_{\beta_\bullet}), \\ \Sigma|(\beta_\bullet, D) &\sim IW(V, n), \end{aligned} \tag{2.11}$$

where

$$\begin{aligned} \overline{\beta}_\bullet &= \overline{\Sigma}_{\beta_\bullet}\big[\big(\sum_{t=1}^{n} X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}y_{t\bullet}\big) + A_{\beta_\bullet}\beta_\circ\big], \\ \overline{\Sigma}_{\beta_\bullet} &= \big[\big(\sum_{t=1}^{n} X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}y_{t\bullet}\big) + A_{\beta_\bullet}\big]^{-1}. \end{aligned} \tag{2.12}$$

So far, the full conditional posterior distributions were derived using the assumption of the normal distribution of the errors. In the next section, we assume that the error term follows the skew-normal distribution and compute the posterior density and full conditional distributions.

It worths to mention here that one of the possible procedure to draw samples from the posterior density of the parameters is to follow an MCMC algorithm. Particularly, if the full conditional distributions of relevant parameters are available in closed forms a Gibbs sampling algorithm could be employed to draw samples from corresponding densities. The literature shows that such view to the SUR models [32], [25], [8], [29] and [24] has already investigated this topic.

3. BAYESIAN INFERENCE ON SUR MODELS USING THE SKEW-NORMAL DENSITY FOR ERROR

To consider a normal density for the distribution of the error while utilizing a SUR model is a standard procedure to make statistical inference. However, this assumption might not hold in some real-life example and so corresponding statistical inferences might not lead to feasible results. Instead, to use skew-normal distribution for the density of error is an alternative option. Having said that, to recall ML method of estimation is then one of the conventional parametric statistical inference methods to consider. However, similar to the situation mentioned in the case of considering the normal distribution for error (see initial discussions in Section 2), there are some problems to implement this method as well. Hence, we outline the Bayesian statistical inference on the parameters of a SUR model with the error comes from skew-normal in this section. Moreover, some important statistical features of this strategy are also highlighted.

To start, let us first briefly review the properties of a SUR model under assumption of an skew-normal density for the error in the model (2.4). Specifically, let write

$$u_{t\bullet} = \big(u_{t1}, ..., u_{tg}\big)^{\mathsf{T}} \sim SN\big(0_g,\ \Sigma,\ \lambda\big), \qquad t = 1, ..., n. \tag{3.1}$$

Following [3], the distribution of $u_{t\bullet}$, for $t = 1, ..., n$, is given by

$$f_{U_{t\bullet}}\big(u_{t\bullet}\big) = 2\phi_g\big(u_{t\bullet}; 0_g,\ \Sigma\big)\ \Phi_g\big(\lambda^{\mathsf{T}}\omega^{-1}u_{t\bullet}\big), \tag{3.2}$$

where $\phi_g(u_{t\bullet}; 0, \Sigma)$ is the g-dimensional normal density with zero mean vector and covariance matrix Σ, $\Phi_g(\cdot)$ is the cumulative distribution function of the standard normal density, and λ is a g-dimensional vector with constant values. Here, ω is a diagonal matrix whose components are the square root of the corresponding covariance matrix Σ. Now, we can write down either the likelihood function of the parameters or its logarithm. We prefer the later one, denoted here by $l(\lambda, \beta_\bullet, \Sigma)$, which is given by

$$\begin{aligned} l(\lambda, \beta_\bullet, \Sigma) \ =\ & n\ \log 2 - \frac{ng}{2}\ \log(2\pi) - \frac{n}{2}\ \log|\Sigma| \\ & - \frac{1}{2}\ \sum_{t=1}^{n}\big[(y_{t\bullet} - X_{t\bullet}\beta_\bullet)^{\mathsf{T}}\Sigma^{-1}(y_{t\bullet} - X_{t\bullet}\beta_\bullet)\big] + \sum_{t=1}^{n}\log\ \Phi_1\big(\lambda^{\mathsf{T}}\omega^{-1}u_{t\bullet}\big). \end{aligned} \tag{3.3}$$

If one is going to estimate the parameters directly using (3.3), there exist some problems. The main drawbacks are lack of convergence in employing any likelihood-based numerical algorithm such pseudo-Newton and the high cost of computations. To circumvent these issues, we propose to follow the Bayesian methodology instead.

Here, an integral part of specifying a Bayesian paradigm is the selection of some prior distributions for all unknown parameters, i.e., $\theta = (\beta_\bullet, \Sigma, \lambda)$. In the absence of prior information and to guarantee to have feasible properties for the posterior, we adopt proper but diffuse priors. Suppose elements of θ are independent a priori, and the following priors have been considered

$$\begin{aligned} &\beta_\bullet \sim N\big(\beta_\circ, \Sigma_{\beta_\circ}\big), \quad \pi(\Sigma) \propto |\Sigma|^{-\frac{g+1}{2}}, \\ &\lambda \sim N\big(\lambda_0, \Lambda_0\big), \quad z_0 \sim N\big(0_g, I_g\big). \end{aligned} \tag{3.4}$$

Then, the join posterior of all parameters is given by

$$\begin{aligned} \pi(\beta_\bullet, \Sigma, \lambda|y) \;\propto\; & \phi_k\big(\beta_\bullet; \beta_\circ, \Sigma_{\beta_\circ}\big) \\ & \times\; \phi_g\big(\lambda; \lambda_0, \Lambda_0\big) \\ & \times\; |\Sigma|^{-\frac{g+1}{2}} \\ & \times\; 2^n \prod_{t=1}^{n} \phi_g\big(y_{t\bullet}; X_{t\bullet}\beta_\bullet, \Sigma\big)\Phi_1(\lambda^{\mathsf{T}}\omega^{-1}(y_{t\bullet} - X_{t\bullet}\beta_\bullet)). \end{aligned} \tag{3.5}$$

As seen, this expression doesn't have a closed form, so we can not compute the join posterior analytically. To turn around this problem, we use the stochastic representation of the skew-normal distribution (see [1]), i.e. $y_{t\bullet} = \lambda \odot |z_0| + z_1$ where $\odot$ denotes Hadamard product, $z_0 \sim N\big(0_g, I_g\big)$, $z_1 \sim N\big(X_{\bullet t}\beta_\bullet, \Sigma\big)$. Moreover, it is assumed that z_0 and z_1 are independent. Then, it is expected that we could drive the full conditional distributions for each parameters. Below, we provide them in turn. More details on computing those expressions are given in Appendix. Note that we write the full conditional distribution in an deliberator order. The reason to do so is when one is going to update samples from corresponding densities for each parameter in an MCMC sampling algorithm the same order should be followed.

First, we have

$$\beta_\bullet|\big(\Sigma, \lambda, |z_0|, D\big) \sim N\big(\widetilde{\beta}_\bullet, \widetilde{\Sigma}_{\beta_\bullet}\big),$$

where $\widetilde{\beta}_\bullet = \widetilde{\Sigma}_{\beta_\bullet}\big(\sum_{t=1}^n X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}y_{t\bullet} + \Sigma_{\beta_\circ}^{-1}\beta_\circ - \sum_{t=1}^n X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}\Lambda|z_0|\big)$ and $\widetilde{\Sigma}_{\beta_\bullet} = \big(\sum_{t=1}^n X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}X_{t\bullet} + \Sigma_{\beta_\circ}^{-1}\big)^{-1}$.

Secondly, we have

$$\Sigma|\big(\beta_\bullet, \lambda, |z_0|, D\big) \sim IW(R, n), \tag{3.6}$$

where $R = \sum_{t=1}^n \big(y_{t\bullet} - [\lambda \odot |z_0| + X_{t\bullet}\beta_\bullet]\big)\big(y_{t\bullet} - [\lambda \odot |z_0| + X_{t\bullet}\beta_\bullet]\big)^{\mathsf{T}}$.

Next, the full conditional distributions of λ is given by

$$\lambda|\big(\beta_\bullet, \Sigma, |z_0|, D\big) \sim N\big(\widetilde{\lambda}, \widetilde{\Lambda}\big), \tag{3.7}$$

where $\widetilde{\Lambda} = \big(nZ_0^*\Sigma^{-1}Z_0^* + \Lambda_0^{-1}\big)^{-1}$, $\widetilde{\lambda} = \widetilde{\Lambda}(\sum_{t=1}^n Z_0^*\Sigma^{-1}y_{t\bullet} - \sum_{t=1}^n Z_0^*\Sigma^{-1}X_{t\bullet}\beta_\bullet + \Lambda_0^{-1}\lambda_0)$, and Z_0^* is an $g \times g$ diagonal matrix whose components are filled with elements of vector $|z_0|$.

Finally, at the last step, the density of $|z_0|$ should be derived. It is straightforward to show that

$$|z_0| \;\Big|\Big(\beta_\bullet, \Sigma, \lambda, D\Big) \sim TN\Big(\widetilde{z}_0, \Psi_{z_0}, (0, +\infty)\Big), \tag{3.8}$$

where $TN\Big(\mu, \Sigma, (a, b)\Big)$ stands for the multivariate truncated normal distribution $N(\mu, \Sigma)$ lying within the interval $(a, b), -\infty \leq a < b \leq +\infty$. Also $\Psi_{z_0} = \big(I_g + n\Delta\Sigma_{-1}\Delta\big)^{-1}$ and $\widetilde{z}_0 = \Psi_{z_0}\big(\sum_{t=1}^{n} \Delta\Sigma^{-1}[y_{t\bullet} - X_{t\bullet}^{\mathsf{T}}\beta_{\bullet}]\big)$ where $\Delta = \text{diag}(\lambda_1, ..., \lambda_g)$.

Now, we are at a position to conduct some simulation studies to evaluate the proposed models.

4. SIMULATION STUDIES

Here, we outline our simulation studies to evaluate the procedure in estimating the parameters of the SUR models given in Sections 2 and 3. Suppose the following simultaneous model is given:

$$\begin{cases} y_1 = \beta_0 + \beta_1 z_1 + \beta_2 x_1 + u_1, \\ y_2 = \gamma_0 + \gamma_1 z_1 + \gamma_2 x_2 + u_2. \end{cases} \tag{4.1}$$

The assumptions imposed for this model are the same as those proposed in Sections 2 and 3. Moreover, $u = (u_1, u_2)^{\mathsf{T}} \sim N(0, \Sigma)$ where y_1 and y_2 are endogenous variables. In addition, we assume variables z_1, x_1 and x_2 are exogenous. When we switch to the scenario in which the error terms follow the asymmetric distribution, we assume $u = (u_1, u_2)^{\mathsf{T}} \sim SN(0_2, \Sigma, \lambda)$ where λ is shape parameter vector.

To generate data from the model with equations in (4.1), we do need to fix the parameters. We are writing them all together either in regression equations or explicit expressions. They are given as follows:

$$\begin{aligned} y_1 &= 4 - 3z_1 - 4x_1 + u_1, \\ y_2 &= 7 + 3z_1 - 2x_2 + u_2, \end{aligned} \tag{4.2}$$

$$\begin{aligned} u &= \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \sim SN(0_2, \Sigma, \lambda), \\ \Sigma &= \begin{pmatrix} 3 & -1 \\ -1 & 4 \end{pmatrix}, \quad \lambda = \begin{pmatrix} 4 \\ 7 \end{pmatrix}. \end{aligned}$$

The sample size for each equation is fixed at 1000 cases. Consequently, due to having two equations in (4.1), the total number of available data is 2000. The Bayesian inference is conducted using the proper priors given in (3.4) with the hyperparameters fixed on some specific values. Particularly, we consider

$$\begin{aligned} \beta_{\bullet} \sim N(0_2, 100I_2), \qquad \lambda \sim N(0_2, 100I_2), \\ \Sigma \propto \begin{vmatrix} 0.01 & 0 \\ 0 & 0.01 \end{vmatrix}^{-\frac{1}{2}}. \end{aligned}$$

As mentioned earlier, one can follow the Gibbs sampling algorithm to draw samples and then estimate the parameters of the models while employing an MCMC algorithm. To do this, we fixed the simulated sample size at 100,000 iterations for each chain. Convergence of the

MCMC algorithm was confirmed by the Gelman and Rubin convergence measure [16], but not reported here. To get independent samples, the burn-in was set on 25,000 iterations for each chain and the last 75,000 iterations were used to make statistical inference on parameters. Then, with taking each 50-th observation, we were ultimately left with 1,500 samples. The summarized results are presented in Table 1.

As can be seen, the Table 1 includes two parts. The results using the normal and skew-normal distributions assumption for the errors are shown in the left and right panels, respectively. The regression coefficients and covariance elements are also estimated. As the values in the left panel of the table show, both intercepts for each equation of the model (4.1) are overestimated. This phenomenon is also the case for the elements of the covariance matrix. The other coefficients are estimated relatively as good as expected.

Table 1: The estimate of parameters and other measures after fitting the SUR model (4.1) under the assumptions of the skew-normal (right panel) and normal (left panel) distributions for the errors.

Parameter	N-MCMC					SN-MCMC				
	Mean	Sd	2.5%	97.5%	ES	Mean	Sd	2.5%	97.5%	ES
β_0	9.38	0.334	8.722	10.01	5.38	3.875	0.326	3.284	4.521	0.125
β_1	−3.009	0.017	−3.043	−2.973	0.009	−2.986	0.015	−3.027	−2.97	0.014
β_2	−3.981	0.024	−4.026	−3.936	0.009	−3.983	0.021	−4.029	−3.948	0.017
γ_0	16.79	0.562	15.7	17.86	9.79	7.510	0.391	5.619	7.139	0.510
γ_1	3.05	0.033	2.981	3.111	0.05	3.006	0.020	2.987	3.065	0.006
γ_2	−1.934	0.039	−2.009	−1.856	0.066	−2.000	0.025	−2.027	−1.927	0.000
σ_{11}	20.25	0.883	18.52	22.01	17.25	3.504	0.572	2.522	4.798	0.504
σ_{12}	1.123	1.215	−1.307	3.549	2.123	−1.298	0.643	−2.436	0.056	0.298
σ_{22}	74.53	3.334	68.23	81.09	70.53	3.211	0.769	1.839	4.887	0.789
λ_1	—	—	—	—	—	6.498	0.250	6.303	7.451	2.498
λ_2	—	—	—	—	—	12.136	0.371	12.01	12.43	5.136

The results using the skew-normal Bayesian approach is given on the right panel of Table 1. As seen, relatively small values of effect size (ES), defined as absolute bias, indicate that all parameters are well estimated. To consider all measures, a general result is that taking into account the skew-normal distribution instead of normal for the errors does better a job of fitting the model (4.1) while employing a Bayesian approach to making statistical inference.

To evaluate performance of the method in more details, we iterated the procedure of generating the data from model (4.1) with the sample size of $n = 1000$ for the numbers of 50 times and then computed the Mean Squared Error (MSE) criterion, i.e.

$$\text{MSE}(\hat{\alpha}) = \frac{1}{50}\sum_{i=1}^{50}(\hat{\alpha}_i - \alpha_{\text{True}})^2. \tag{4.3}$$

The results (not shown here) have confirmed that the MSE criterion in estimating the parameters using the skew-normal Bayesian approach is very close to zero, but this was not the case for the normal distribution assumption for the error in the model (4.1). It means that the estimators derived from the skew-normal case are relatively more accurate and precise than the normal assumption.

5. REAL APPLICATION

We are interested in applying the proposed models in this paper on real-life data. To do this, we used the cost and income data collected on year 2009 in Iran. There are about 13,345 families from 32 provinces. Here, the main goal is on survey effects of some variables on gross cost (GH) and income (D). In this study, both of these quantities are considered as endogenous variables and other covariates are set as exogenous. A general description of the considered variables are reported in Table 2. Also, Figures 1, 2 and 3 (upper panel) provide some geometric displays of some exogenous and two endogenous variables, i.e. GH, D.

Table 2: A general description of variables utilized in real application.

Variable names	Abbreviation signs	Variable Type	Codes
Gross cost	GH	Quantitative	—
Income	D	Quantitative	—
Family size	C_1	Quantitative	—
Number of literate	C_2	Quantitative	—
Number of employees	C_3	Quantitative	—
Number of people with income	C_4	Quantitative	—
Age	A	Quantitative	—
Location Area	B_1	Quantitative	—
Private car	B_2	Qualitative	1: Use, 0: Nonuse
Internet	B_3	Qualitative	1: Use, 0: Nonuse
Gas	B_4	Qualitative	1: Use, 0: Nonuse
Mobile	B_5	Qualitative	1: Use, 0: Nonuse
Incomes of agricultural free businesses	D_1	Quantitative	—
Incomes of nonfarm free businesses	D_2	Quantitative	—
Other Incomes	D_3	Quantitative	—
Other Non-monetary Incomes	D_4	Quantitative	—

To initiate a statistical analysis based upon a common linear regression model, we are concerned about the accuracy of considering the normality assumption for the response variables, i.e. GH, and D, here. We used the Kolmogorov- Smirnov (KS) test for this purpose. Based on this test, normality assumption has not been confirmed for both endogenous variables with the p-value < 0.05. Further, to have visual tools, the quantile-quantile plots for each of the income and gross cost were also drawn. They appeared on the lower panel of Figure 3. As seen, both plots are confirming a lack of the normal distributions fitting for each variable. Moreover, the contour plot based on these endogenous variables, which is appeared on the upper panel Figure 3, also shows a departure from the bivariate normal distribution assumption. It might be argued here that a logarithm transformation of the endogenous variables might solely lead to a better fit of normality assumption for these variables. However, based on our investigation (not reported here), we did not reach to such conclusion. Hence, we preferred to invoke some asymmetric densities, particularly skew-normal distribution, to proceed our analysis. However, we also utilized the normal distribution for the endogenous variables to make a comparison, similar to our simulation studies.

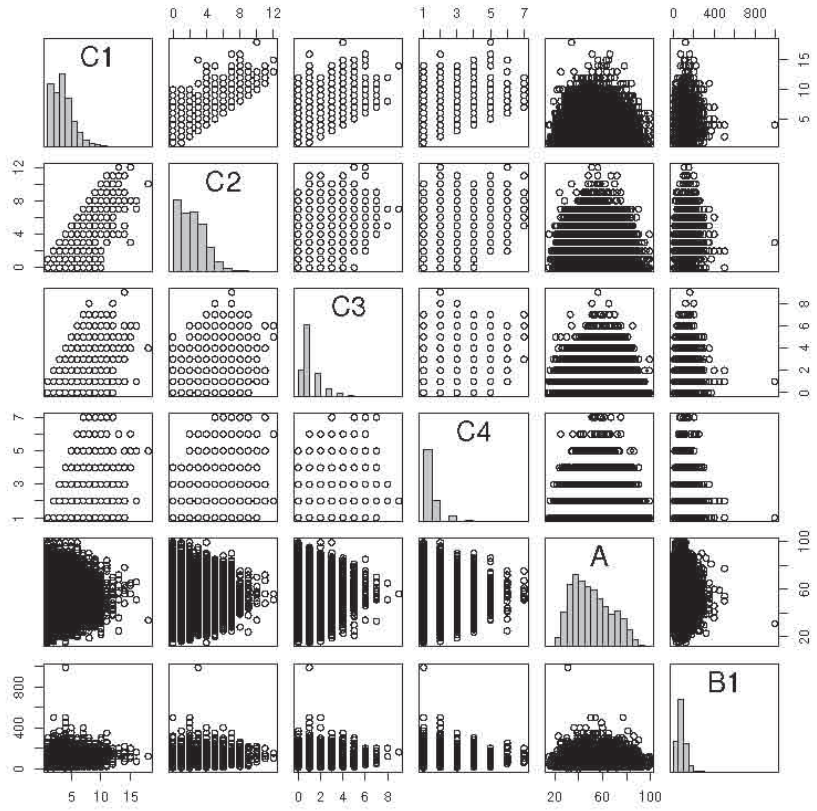

Figure 1: The pairs plot of quantitative variables described in Table 2.

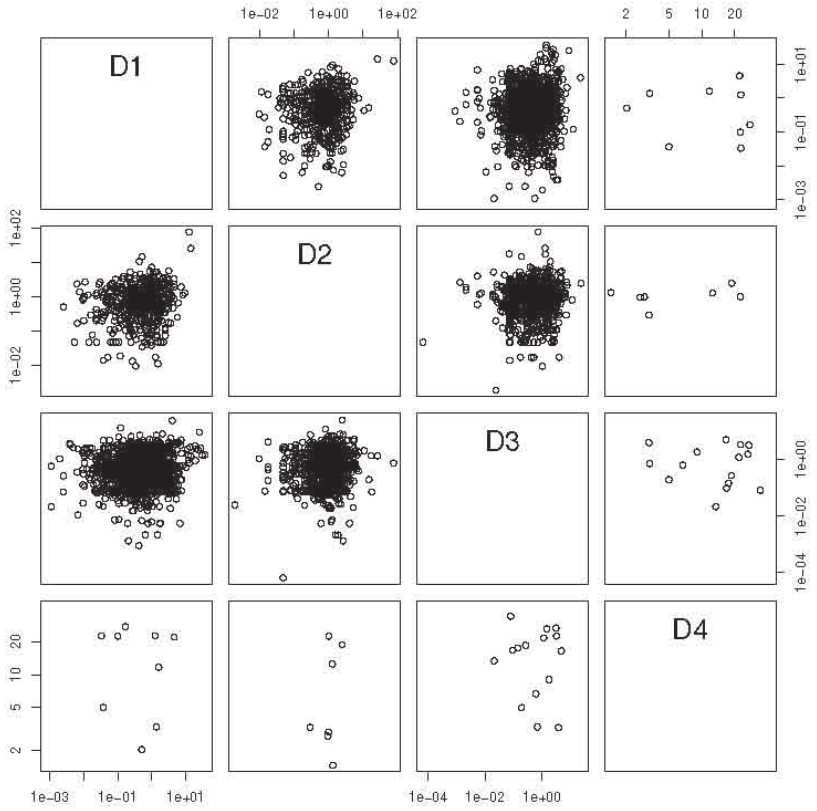

Figure 2: The pairs plot for several types of incomes described in Table 2.

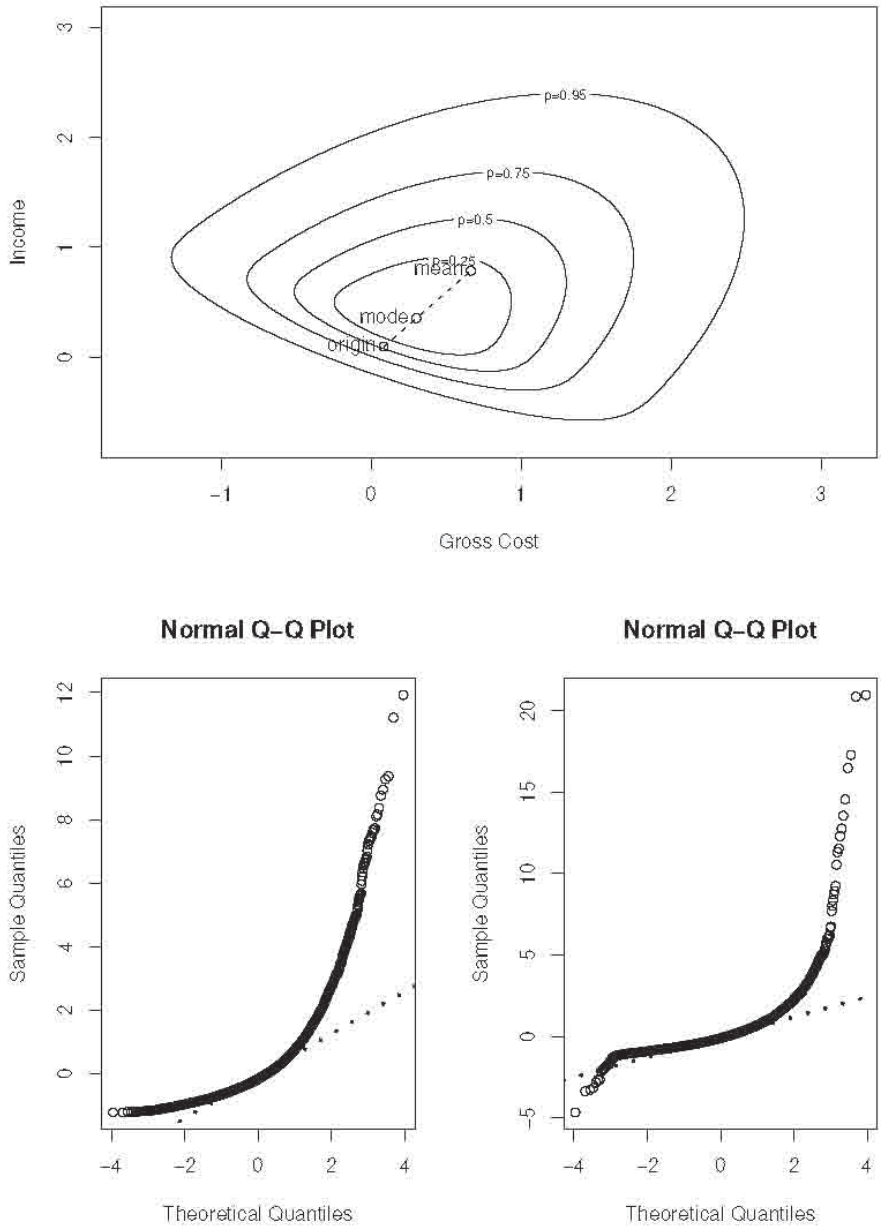

Figure 3: The contour plot of gross cost (GH) against income (D) (upper panel) along with the quantile-quantile plot (lower panel) for each of them. Lack of normal distribution fitting, either jointly or marginally, using endogenous variables are apparent from both plots.

Based upon a general view of the data and also after consulting the subjects with some econometrics experts in Statistical Center of Iran, the following SUR model was utilized to express the inter-relationship between endogenous and exogenous variables:

$$\begin{aligned} GH &= \beta_0 + \sum_{i=1}^{4} \beta_{C_i} C_i + \sum_{i=1}^{5} \beta_{B_i} B_i + \beta_A A + \epsilon_1, \\ D &= \gamma_0 + \sum_{i=1}^{4} \gamma_{D_i} D_i + \epsilon_2. \end{aligned} \tag{5.1}$$

This model has been fitted through both frequentist and Bayesian approaches as well as under the assumption of the normal (N) and skew-normal (SN) distributions for the errors. Table 3 shows the estimates along with standard errors of the estimates. As seen, the table is divided in three parts. The first row panel represents the quantities mentioned above for the parameters of the first equation in (5.1). Similarly, those for the second equation appear in the second row panel. Finally, the last row panel constitutes the estimates and their standard errors for the components of the covariance matrix of the errors in (5.1) as well as those values for the skewness parameters, if they are required. The important point to emphasize is that we have only reported those estimates which were significant at %5 level. Hence, one does not see some of the coefficients from the SUR model (5.1) in Table 3.

Table 3: The estimates along with standard errors of the estimates for the parameters of the SUR model fitted through both frequentist and Bayesian approaches as well as under assumption of the normal (N) and skew-normal (SN) distributions for the errors in (5.1) for the Iranian cost and income data collected in year 2009.

Parameter	Bayesian				Frequentist			
	Estimation		Standard error		Estimation		Standard error	
	N	SN	N	SN	N	SN	N	SN
β_0	−1.338	−1.581	0.030	0.017	−1.50	−1.34	0.047	0.006
β_{C_1}	0.048	0.026	0.006	0.003	0.036	0.040	0.009	0.001
β_{C_2}	0.056	0.045	0.006	0.004	0.082	0.046	0.010	0.002
β_{C_3}	0.070	0.047	0.007	0.004	0.106	0.059	0.011	0.004
β_{C_4}	−0.015	0.014	0.009	0.005	0.061	−0.024	0.014	0.004
β_{B_1}	0.006	0.001	0.001	0.001	0.003	0.003	0.0006	0.0001
β_{B_2}	0.004	0.002	0.001	0.001	0.004	0.002	0.0002	0.0005
β_{B_3}	0.531	0.304	0.016	0.009	0.649	0.531	0.024	0.013
β_{B_4}	0.509	0.281	0.032	0.018	0.689	0.490	0.051	0.032
β_{B_5}	0.011	0.049	0.012	0.007	0.064	0.031	0.018	0.0085
β_A	0.236	0.180	0.016	0.009	0.276	0.137	0.025	0.0064
γ_0	0	−0.514	0.004	0.0001	0	−0.103	0.025	0.0038
γ_{D_1}	0.500	0.562	0.004	0.001	0.544	0.499	0.013	0.0039
γ_{D_2}	0.467	0.498	0.004	0.002	0.503	0.487	0.013	0.0039
γ_{D_3}	0.352	0.375	0.004	0.001	0.412	0.352	0.013	0.0039
γ_{D_4}	0.030	0.030	0.004	0.002	0.033	0.030	0.013	0.0038
σ_{11}	0.671	0.018	0.007	0.009	0.656	0.051	-	-
σ_{21}	0.129	0.001	0.004	0.002	0.084	0.009	-	-
σ_{22}	0.317	0.003	0.003	0.001	0.315	0.018	-	-
λ_1	-	1.194	-	0.007	-	1.181	-	-
λ_2	-	0.764	-	0.003	-	0.869	-	-

Based upon results given at first row panel in Table 3, the number of literate, employees and family size have a direct effect on total family gross cost. The usage of facilities, including Private car, internet, gas and mobile, also has the positive impact on family gross cost. In other words, the utilization of these services leads to an increase in family gross cost. However, if the families are not using these items still there is an increment on cost too. The rationale behind this surprising result comes from the SUR model in which those families would then pay for other luxuries items. Finally, the age of the people who are in charge of the family cost and also the area where the families live both leada to the positive effect on the gross cost.

Now, let us analyze the result at the second row panel in Table 3. As seen, the incomes from the agricultural and non-farm free businesses, other incomes and non-monetary gains have direct effect on the family incomes. Furthermore, other non-monetary earning incomes have less effect on the family incomes subject to other variables. Also, the effect of the incomes from agricultural free businesses on the family incomes are high.

Some interesting results appear in the second row panel of Table 3. First, it is related to the estimate of the intercept (γ_0). Unlike the case for the skew-normal distribution, its estimation is zero when assuming a normal density for the errors in the SUR model (5.1). Second, as seen, the estimates for the components of the covariance matrix based on normality assumption are somewhat bigger than those in the skew-normal case. Albeit, this needs more considerations.

So far, the reader probably discovered a proper strategy to fit the SUR model (5.1) based on the presented results. However, we are interested in selecting one of two methodologies and distributional assumption for the errors through utilizing a sensible statistical measure. There are several methods to choose the appropriate model among two possible candidates while employing either Bayesian or frequentist statistical inferences methodologies. It is usually accepted among statisticians that the Bayes factor criterion is a proper measure to compare the performance of different candidate models while implementing a Bayesian methodology. However, in utilizing the frequentist methodology, the researchers consider the log likelihood and AIC criteria. To compare two candidate models L_1 and L_2 the Bayes factor is represented by a quantity which is simply a ratio (see [22]). Precisely, suppose $\pi(L_1)$ and $\pi(L_2)$ are priors for two models. Then, given the data D, the Bayes factor of model L_1 w.r.t L_2 is written as

$$B_{12} = \frac{Pr(D|L_1)}{Pr(D|L_2)} = \frac{\frac{\pi(L_1|D)}{\pi(L_2|D)}}{\frac{\pi(L_1)}{\pi(L_2)}}, \tag{5.2}$$

where $Pr(D|L) = \int f(D|L,\theta)\pi(\theta|D)d\theta$ and $\theta = (\beta, \gamma, \Sigma, \lambda)$. However, because we are not able to compute the joint posterior analytically this criteria cannot be employed here. [23] proposed a method when there is no closed form for the posterior density. Following them, if $\{\theta\}_{i=1}^{m}$ are samples from the posterior distribution of $\pi(\theta|D, L)$, we can write:

$$f^{(j+1)}(D|L) = \frac{\frac{km}{1-k} + \sum_{j=1}^{m} \frac{f(D|\theta^{(j)},L)}{kf^{(j)}(D|L)+(1-k)f(D|\theta^{(j)},L)}}{\frac{km}{(1-k)f^{(j)}(D|L)} + \sum_{j=1}^{m} \frac{1}{kf^{(j)}(D|L)+(1-k)f(D|\theta^{(j)},L)}}, \tag{5.3}$$

where k is a small value being in the interval $(0, 1)$. To derive this quantity, we repeated our analysis till achieving a reasonable convergence. In some small-scale numerical experiments, we have discovered that the quantity (5.3) performed well for k as small as 0.01.

The logarithm of pseudo-marginal likelihood (LPML) is another criterion to select between two candidate models (see [14]). It is derived from predictive considerations, particularly Conditional Predictive Ordinate (CPO), and leads to pseudo-Bayes factors for choosing an optimal model. It is popular mainly due in part to its relative ease of computation making the LPML a stable estimate base on the samples derived from any MCMC algorithm. Following the [15] and assuming availability of the samples $\theta^{(1)}, ..., \theta^{(s)}$, obtained from corresponding posterior, the i-th CPO_i and $LPML$ are, respectively, estimated as

$$\frac{1}{CPO_i} = \frac{1}{s}\sum_{k=1}^{s} \frac{1}{f_i(y_i|\theta^{(k)}, M)}, \tag{5.4}$$

and

$$LPML = \sum_{i=1}^{n} \log(CPO_i). \tag{5.5}$$

The LPML and BF quantities for both cases (N and SN) are reported in Table 4. As seen, the value of LPML for SN is greater than that for N. Moreover, the ratio of BF for SN in compare with the N model is relatively bigger and so indicating the superiority of SN again. Although there are some debates on using these criteria under the frequentist view, we also reported the estimates of the parameters for both N and SN cases just to have a basis for seeing difference on utilizing two methodologies.

Following the results gained in analyzing this example, our recommendation is to consider a skew-normal rather than the normal density for the error while using a SUR model to analyze the Iran gross and income data collected in year 2009.

Table 4: The performance criteria of the SUR model (5.1) fitted using the skew-normal distribution on the Iranian cost and income data collected in year 2009.

Model	Bayesian		Frequentist	
	BF	LPML	AIC	Log likelihood
N	0.917	−59274.01	80925.15	−40443.57
SN	1.091	−34585.61	−124982.3	62512.15

6. CONCLUSION

When dealing with simultaneous relationship among variables, the SUR model is a particular case of SEM. The frequentist inference utilized for the SUR model under the skew-normal assumption or the error is very time consuming and also challenging to tackle. Hence, a Bayesian inference implemented in the SUR model under skew-normal distribution assumption for errors is developed in this paper. Regarding the model selection, the BF and LPML criteria had some superiorities in choosing better model using our real data set as well as in our simulation studies. Based on the results in this paper we can stat when data are not symmetric, the SUR model accompanied with considering a skew-normal distribution for the error performs well on fitting the data at least in comparison with invoking the normal density.

In future study, we aim to investigate how the endogenous variables can improve the estimation of the parameters in the SEMs while the errors follow the skew-normal distributions. Moreover, to plug in these structures into multilevel models is the way we are going to extend our current research.

A. APPENDIX

According to formula (2.8), full conditional posterior is given by

$$
\begin{aligned}
g(\beta_\bullet|\Sigma, D) &\propto L(D|\beta_\bullet, \Sigma)\pi_1(\beta_\bullet) \propto \exp\big\{-1/2\ \mathrm{tr}(V\Sigma^{-1})\big\} \\
&\propto \exp\Big\{-1/2\sum_{t=1}^{n}(y_{t\bullet}-X_{t\bullet}\beta_\bullet)^{\mathsf{T}}\Sigma^{-1}(y_{t\bullet}-X_{t\bullet}\beta_\bullet)\Big\} \\
&\propto \exp\Big\{-1/2\sum_{t=1}^{n}\big[-\beta_\bullet^{\mathsf{T}}X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}+\beta_\bullet^{\mathsf{T}}X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}X_{t\bullet}\beta_\bullet-y_{t\bullet}^{\mathsf{T}}\Sigma^{-1}X_{t\bullet}\beta_\bullet\big]\Big\} \\
&\propto \exp\Big\{-1/2\big[\beta_\bullet-\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}X_{t\bullet}\big)^{-1}\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}y_{t\bullet}\big)\big]^{\mathsf{T}}\big[\sum_{t=1}^{n}X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}X_{t\bullet}\big] \\
&\qquad \big[\beta_\bullet-\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}X_{t\bullet}\big)^{-1}\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}y_{t\bullet}\big)\big]\Big\}. \qquad \text{(A.1)}
\end{aligned}
$$

So $\beta_\bullet$ is multivariate normal with mean $\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}X_{t\bullet}\big)^{-1}\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}y_{t\bullet}\big)$ and covariance matrix $\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}X_{t\bullet}\big)^{-1}$. The full conditional posterior is derived as follows:

$$
g(\Sigma|\beta_\bullet, D) \propto L(D|\beta_\bullet, \Sigma)\pi_1(\Sigma) \propto |\Sigma|^{-(n+g+1)/2}\exp\big[-1/2\ \mathrm{tr}\{V\Sigma^{-1}\}\big]. \qquad \text{(A.2)}
$$

It is straightforward to check that this expression is proportional to an inverse Wishart distribution with degrees of freedom n and scale covariance V.

Based on the prior density $\beta_\bullet \sim N(\beta_\circ, A_{\beta_\circ}^{-1})$ and $\pi(\Sigma) \propto |\Sigma|^{-\frac{g+1}{2}}$, the posterior distribution can easily be computed. However, it doesn't have a closed form. Instead, the full conditional posterior can be obtained using the expressions

$$
\begin{aligned}
g(\beta_\bullet|\Sigma, D) &\propto L(D|\beta_\bullet, \Sigma)\pi_2(\beta_\bullet) \\
&\propto |\Sigma|^{-\frac{n}{2}}\exp\Big\{-1/2\ \Big[\sum_{t=1}^{n}(y_{t\bullet}-X_{t\bullet}\beta_\bullet)^{\mathsf{T}}\Sigma^{-1}(y_{t\bullet}-X_{t\bullet}\beta_\bullet) \\
&\quad +(\beta_\bullet-\beta_\circ)^{\mathsf{T}}A_{\beta_\bullet}(\beta_\bullet-\beta_\circ)\Big]\Big\} \\
&\propto \exp\Big\{-1/2\ \Big[\beta_\bullet^{\mathsf{T}}\Big\{\Big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}X_{t\bullet}\Big)+A_{\beta_\bullet}\Big\}\beta_\bullet \\
&\quad -2\beta_\bullet^{\mathsf{T}}\Big\{\Big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}y_{t\bullet}\Big)+A_{\beta_\bullet}\beta_\circ\Big\}\Big]\Big\} \\
&\propto \exp\big\{-1/2\ (\beta_\bullet-\overline{\beta}_\bullet)^{\mathsf{T}}\overline{\Sigma}_{\beta_\bullet}^{-1}(\beta_\bullet-\overline{\beta}_\bullet)\big\}. \qquad \text{(A.3)}
\end{aligned}
$$

Consequently $\beta_\bullet$ given (Σ, D) is multivariate normal with mean

$$
\overline{\beta}_\bullet = \overline{\Sigma}_{\beta_\bullet}\big[\big(\sum_{t=1}^{n}X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}y_{t\bullet}\big)+A_{\beta_\bullet}\beta_\circ\big]
$$

and covariance matrix

$$\overline{\Sigma}_{\beta_\bullet} = \big[\big(\sum_{t=1}^{n} X_{t\bullet}^{\mathsf{T}}\Sigma^{-1}X_{t\bullet}\big) + A_{\beta_\bullet}\big]^{-1}.$$

Similarity, the full conditional posterior $\Sigma|(\beta_\bullet, D)$ is given by

$$\pi(\Sigma|\beta_\bullet, D) \propto L(D|\beta_\bullet, \Sigma)\pi_2(\Sigma) \propto |\Sigma|^{-(n+g+1)/2} \exp\big[-1/2\ \operatorname{tr}\{V\Sigma^{-1}\}\big]. \tag{A.4}$$

Seeing similarity of this expression to (A.2), the full conditional distribution $\Sigma|(\beta_\bullet, D)$ is inverse Wishart with degrees of freedom n and scale covariance V.

Recall: The probability density function for the random matrix X $(n \times p)$ that follows the matrix normal distribution $MN_{n\times p}\big(M, U, V\big)$ has the form

$$p\big(X|M, U, V\big) = \frac{\exp\big(-\frac{1}{2}\operatorname{tr}\big[V^{-1}(X-M^{\mathsf{T}})U^{-1}(X-M)\big]\big)}{(2\pi)^{np/2}|V|^{n/2}|U|^{p/2}}, \tag{A.5}$$

where M is $n \times p$, U is $n \times n$ and V is $p \times p$ matrices. Note that the matrix normal is linked to the multivariate normal distribution in the following way:

$$X \sim MN_{n\times p}\big(M, U, V\big) \tag{A.6}$$

if and only if

$$\operatorname{Vec}(X) \sim N\big(\operatorname{Vec}(M)_{np}, V \otimes U\big), \tag{A.7}$$

where $\operatorname{Vec}(M)$ denotes the vectorization of M.

Suppose that $y_{t\bullet} \sim SN\big(X_{t\bullet}\beta_\bullet\big)$. According to stochastic representations of multivariate skew-normal distribution (see [1]), we have

$$y_{t\bullet} = \lambda \odot |z_0| + z_1, \tag{A.8}$$

where $\odot$ denotes Hadamard product, $z_0 \sim N\big(0_g, I_g\big)$, $z_1 \sim N\big(X_{\bullet t}\beta_\bullet, \Sigma\big)$ and also z_0 and z_1 are independent. Thus,

$$\begin{pmatrix} z_0 \\ z_1 \end{pmatrix} \sim N\Big(\begin{pmatrix} 0 \\ X_{t\bullet}\beta_\bullet \end{pmatrix}, \begin{pmatrix} I_g & 0 \\ 0 & \Sigma \end{pmatrix}\Big). \tag{A.9}$$

Thus, the conditional distribution $y_{t\bullet}$ given z_0 leads to

$$y_{t\bullet}|z_0 \sim N\big(\lambda \odot |z_0| + X_{t\bullet}\beta_\bullet, \Sigma\big). \tag{A.10}$$

The full conditional posterior distribution of all parameters are determined based on

(A.10). So, for $\beta_\bullet|\big(\Sigma,\lambda,z_0,D\big)$, we have

$$\begin{aligned}\pi\big(\beta_\bullet|\Sigma,\lambda,z_0,D\big) &\propto L(y_{t\bullet}|z_0,\Sigma,\lambda,\beta_\bullet)\pi(\beta_\bullet)\\ &\propto |\Sigma|^{-\frac{n}{2}}\exp\Big\{-\frac{1}{2}\sum_{t=1}^{n}\big(y_{t\bullet}-[\lambda\odot|z_0|+X_{t\bullet}\beta_\bullet]\big)^{\mathsf{T}}\Sigma^{-1}\\ &\quad\big(y_{t\bullet}-[\lambda\odot|z_0|+X_{t\bullet}\beta_\bullet]\big)\Big\}\\ &\quad\times|\Sigma_{\beta_\circ}|^{-\frac{1}{2}}\exp\big\{(\beta_\bullet-\beta_\circ)^{\mathsf{T}}\Sigma_{\beta_\circ}^{-1}(\beta_\bullet-\beta_\circ)\big\}\\ &\propto \exp\Big\{-\frac{1}{2}\big[\beta_\bullet^{\mathsf{T}}(\sum_{t=1}^{n}X_\bullet^{\mathsf{T}}\Sigma^{-1}X_\bullet^{\mathsf{T}}+\Sigma_{\beta_\circ}^{-1})\beta_\bullet\\ &\quad-2\beta_\bullet^{\mathsf{T}}(\sum_{t=1}^{n}X_\bullet^{\mathsf{T}}\Sigma^{-1}y_\bullet-\sum_{t=1}^{n}X_\bullet^{\mathsf{T}}\Sigma^{-1}\lambda\odot|z_0|+\Sigma_{\beta_\circ}^{-1}\beta_\circ)\big]\Big\}\\ &\propto \exp\Big\{-\frac{1}{2}(\beta_\bullet-\widetilde{\beta}_\bullet)^{\mathsf{T}}\widetilde{\Sigma}_{\beta_\bullet}^{-1}(\beta_\bullet-\widetilde{\beta}_\bullet)\Big\}. \qquad \text{(A.11)}\end{aligned}$$

Consequently, $\beta_\bullet|\big(\Sigma,\lambda,z_0,D\big)$ is multivariate normal with mean

$$\widetilde{\beta}_\bullet=\widetilde{\Sigma}_{\beta_\bullet}(\sum_{t=1}^{n}X_\bullet^{\mathsf{T}}\Sigma^{-1}y_\bullet-\sum_{t=1}^{n}X_\bullet^{\mathsf{T}}\Sigma^{-1}\lambda\odot|z_0|+\Sigma_{\beta_\circ}^{-1}\beta_\circ)$$

and covariance

$$\widetilde{\Sigma}_{\beta_\bullet}=(\sum_{t=1}^{n}X_\bullet^{\mathsf{T}}\Sigma^{-1}X_\bullet^{\mathsf{T}}+\Sigma_{\beta_\circ}^{-1})^{-1}.$$

Similarly, the full conditional posterior distribution $\Sigma|\big(\beta_\bullet,\lambda,z_0,D\big)$ is computed: i.e.

$$\begin{aligned}\pi\big(\Sigma|\beta_\bullet,\lambda,z_0,D\big) &\propto L(y_{t\bullet}|z_0,\Sigma,\lambda,\beta_\bullet)\pi(\Sigma)\\ &\propto |\Sigma|^{-\frac{n}{2}}\exp\Big\{-\frac{1}{2}\sum_{t=1}^{n}\big(y_{t\bullet}-[\lambda\odot|z_0|+X_{t\bullet}\beta_\bullet]\big)^{\mathsf{T}}\Sigma^{-1}\\ &\quad\big(y_{t\bullet}-[\lambda\odot|z_0|+X_{t\bullet}\beta_\bullet]\big)\Big\}\\ &\quad\times|\Sigma|^{-\frac{g+1}{2}}\\ &\propto |\Sigma|^{-\frac{n+g+1}{2}}\exp\Big\{-\frac{1}{2}\operatorname{tr}(R\Sigma^{-1})\Big\}, \qquad \text{(A.12)}\end{aligned}$$

where $R=\sum_{t=1}^{n}\big(y_{t\bullet}-[\lambda\odot|z_0|+X_{t\bullet}\beta_\bullet]\big)\big(y_{t\bullet}-[\lambda\odot|z_0|+X_{t\bullet}\beta_\bullet]\big)^{\mathsf{T}}$. So

$$\Sigma|\big(\beta_\bullet,\lambda,z_0,D\big)\sim IW(R,n),$$

where $IW(\cdot,\cdot)$ denotes the inverse Wishart distribution.

Suppose

$$Z_0^*=\begin{pmatrix}|Z_{0_1}| & 0 & \dots & 0\\ 0 & |Z_{0_2}| & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & |Z_{0_g}|\end{pmatrix}_{g\times g}.$$

Now, we can write $\lambda^{\mathsf{T}} \odot |z_0|^{\mathsf{T}} = \lambda^{\mathsf{T}} Z_0^*$. Then, the full conditional $\lambda|\big(\beta_\bullet, \Sigma, z_0, D\big)$ is given by

$$\begin{aligned}
\pi\big(\lambda|\Sigma, \beta_\bullet, z_0, D\big) &\propto L(y_{t\bullet}|z_0, \Sigma, \lambda, \beta_\bullet)\pi(\lambda) \\
&\propto |\Sigma|^{-\frac{n}{2}} \exp\Big\{ -\frac{1}{2}\sum_{t=1}^{n} \big(y_{t\bullet} - [\lambda \odot |z_0| + X_{t\bullet}\beta_\bullet]\big)^{\mathsf{T}} \Sigma^{-1} \\
&\quad \big(y_{t\bullet} - [\lambda \odot |z_0| + X_{t\bullet}\beta_\bullet]\big)\Big\} \\
&\quad \times |\Lambda_0|^{-\frac{1}{2}} \exp\big\{(\lambda - \lambda_0)^{\mathsf{T}} \Lambda_0^{-1} (\lambda - \lambda_0)\big\} \\
&\propto \exp\Big\{ -\frac{1}{2}\big[\lambda^{\mathsf{T}}(nZ_0^*\Sigma^{-1}Z_0^* + \Lambda_0^{-1})\lambda \\
&\quad - 2\lambda^{\mathsf{T}}(\sum_{t=1}^{n} Z_0^*\Sigma^{-1}y_{t\bullet} - \sum_{t=1}^{n} Z_0^*\Sigma^{-1}X_{t\bullet}\beta_\bullet + \Lambda_0^{-1}\lambda_0)\big]\Big\} \\
&\propto \exp\Big\{ -\frac{1}{2}(\lambda - \widetilde{\lambda})^{\mathsf{T}} \widetilde{\Lambda}^{-1} (\lambda - \widetilde{\lambda})\Big\}. \qquad \text{(A.13)}
\end{aligned}$$

As a result, the full conditional distribution $\lambda|\big(\beta_\bullet, \Sigma, z_0, D\big)$ is multivariate normal with mean

$$\widetilde{\lambda} = \widetilde{\Lambda}(\sum_{t=1}^{n} Z_0^*\Sigma^{-1}y_{t\bullet} - \sum_{t=1}^{n} Z_0^*\Sigma^{-1}X_{t\bullet}\beta_\bullet + \Lambda_0^{-1}\lambda_0)$$

and covariance

$$\widetilde{\Lambda} = (nZ_0^*\Sigma^{-1}Z_0^* + \Lambda_0^{-1})^{-1}.$$

Suppose $\Delta = \mathrm{diag}(\lambda_1, ..., \lambda_g)$, such that $\lambda \odot |z_0| = \Delta|z_0|$. Then, the full conditional distribution $|z_0|$ given $\big(\beta_\bullet, \Sigma, \lambda, D\big)$ is determined as

$$\begin{aligned}
\pi\Big(|z_0| \,\Big|\, \Sigma, \beta_\bullet, \lambda, D\Big) &\propto L(y_{t\bullet}|z_0, \Sigma, \lambda, \beta_\bullet)\pi(|z_0|) \\
&\propto |\Sigma|^{-\frac{n}{2}} \exp\Big\{ -\frac{1}{2}\sum_{t=1}^{n} \big(y_{t\bullet} - [\Delta|z_0| + X_{t\bullet}\beta_\bullet]\big)^{\mathsf{T}} \Sigma^{-1} \\
&\quad \big(y_{t\bullet} - [\Delta|z_0| + X_{t\bullet}\beta_\bullet]\big)\Big\} \\
&\quad \times \exp\Big\{ -\frac{n}{2}|z_0|^{\mathsf{T}} I_g |z_0|\Big\} \\
&\propto \exp\Big\{ -\frac{1}{2}\big[|z_0|^{\mathsf{T}}(n\Delta\Sigma^{-1}\Delta + I_g)|z_0| \\
&\quad - 2|z_0|^{\mathsf{T}}(\sum_{t=1}^{n} \Delta\Sigma^{-1}y_{t\bullet} - \sum_{t=1}^{n} \Delta\Sigma^{-1}X_{t\bullet}\beta_\bullet)\big]\Big\} \\
&\propto \exp\Big\{ -\frac{1}{2}(|z_0| - \widetilde{z}_0)^{\mathsf{T}} \Psi_z^{-1} (|z_0| - \widetilde{z}_0)\Big\}. \qquad \text{(A.14)}
\end{aligned}$$

Consequently the full conditional posterior distribution is $TN(\widetilde{z}_0, \Psi_z, (0, +\infty))$ where

$$\widetilde{z}_0 = \Psi_z(\sum_{t=1}^{n} \Delta\Sigma^{-1}y_{t\bullet} - \sum_{t=1}^{n} \Delta\Sigma^{-1}X_{t\bullet}\beta_\bullet)$$

and

$$\Psi_z = (n\Delta\Sigma^{-1}\Delta + I_g).$$

Here, $TN\big(\mu, \Sigma; (a, b)\big)$ stands for the multivariate truncated normal distribution $N(\mu, \Sigma)$ lying within the interval (a, b), $-\infty \leq a < b \leq +\infty$.

Hadmard Product:

For two matrices, A, B, of the same dimension, $m \times n$ the Hadamard product, $A \odot B$, is a matrix, of the same dimension as the operands, with elements given by $(A \odot B)_{i,j} = (A)_{i,j} \cdot (B)_{i,j}$, writing as [9]

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \odot \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{pmatrix}. \tag{A.15}$$

ACKNOWLEDGMENTS

The authors also acknowledge the valuable suggestions from the referees.

REFERENCES

[1] Arellano-Valle, R.B.; Bolfarine, H. and Lachos, V.H. (2007). Bayesian inference for skew-normal linear mixed models, *Journal of Applied Statistics*, **34**, 663–682.

[2] Arellano-Valle, R.B.; Cortes, M.A. and Gomez, H.W. (2010). An extension of the epsilon-skew-normal distribution, *Communications in Statistics Theory and Methods*, **39**, 912–922.

[3] Azzalini, A. (1986). Further results on a class of distribution which includes the normal ones, *Statistica*, **46**, 199–208.

[4] Azzalini, A. and Capitanio, A. (1999). Statistical application of the multivaritae skew-normal distribution, *Journal of the Royal Statistical Society, Series B*, **61**, 579–602.

[5] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution, *Biometrika*, **83**, 715–726.

[6] Azzalini, A. and Regoli, G. (2012). Some properties of skew-symmetric distributions, *Annals of the Institute of Statistical Mathematics*, **64**, 857–879.

[7] Balakrishnan, N. (2002). Discussion of skewed multivariate models related to hidden truncation and/or selective reporting, *Test*, **11**, 37–39.

[8] Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Cambridge.

[9] Chandler, D. (1962). The norm of the Schur product operation, *Numerische Mathematik*, **4**, 343–344.

[10] Chib, S. and Greenberg, E. (1995). Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models, *Journal of Econometrics*, **68**, 339–360.

[11] Epple, D. and Bennett, T.M. (2006). Simultaneous equation econometrics: the missing example, *Economic Inquiry*, **44**, 374–384.

[12] Fiebig, D. (2001). *Seemingly unrelated regression*. In: "Companion in Econometrics" (B. Baltagi, Ed.), Basil Blackwell, London.

[13] Fraser, D.A.S.; Rekkasb, M. and Wong, A. (2005). Highly accurate likelihood analysis for the seemingly unrelated regression problem, *Journal of Econometrics*, **127**, 17–33.

[14] Geisser, S. and Eddy, W.F. (1979). A predictive approach to model selection, *Journal of American Statistical Association*, **74**, 153–160.

[15] Gelfand, A.E. and Dey, D.K. (1994). Bayesian model choice: asymptotics and exact calculations, *Journal of Royal Statistical Society, Series B*, **56**, 501–514.

[16] Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, **4**, 457–472.

[17] Genton, G.G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Chapman & Hall, Boca Raton.

[18] Gupta, A.K.; González-Farías, G. and Domínguez-Molina, J.A. (2004). A multivariate skew-normal distribution, *Journal of Multivariate Analysis*, **89**, 181–190.

[19] Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations, *Econometrica*, **11**, 1–12.

[20] Jeffreys, H. (1961). *Theory of Probability* (3rd ed.), Oxford University Press, Oxford.

[21] Judge, G.G.; Griffiths, W.E.; Hill, R.C.; Lutkepohl, H. and Lee, T.C. (1985). *The Theory and Practice of Econometrics* (Second edition), John Wiley and Sons, New York.

[22] Lee, P.M. (2012). *Bayesian Statistics: An Introduction*, John Wiley and Sons, New York.

[23] Newton, M.A. and Raftery, A.E. (1994). Approximate Bayesian inference by the weighted likelihood bootstrap, *Journal of the Royal Statistical Society, Series B*, **56**, 3–48.

[24] Percy, D.F. (1992). Predictions for seemingly unrelated regressions, *Journal of the Royal Statistical Society, Series B*, **54**, 243–252.

[25] Press, S.J. (1972). *Applied Multivariate Analysis*, John Wiley and Sons, New York.

[26] Puelz, D.; Hahn, P.R. and Carvalho, C. (2017). Variable selection in seemingly unrelated regressions with random predictors, *Bayesian Analysis*, **12**, 969–989 (available at SSRN).

[27] Smith, M. and Kohn, R. (2000). Nonparametric seemingly unrelated regression, *Journal of Econometrics*, **98**, 257–282.

[28] Srivastava, V.K. and Dwivedi, T.D. (1979). Estimation of seemingly unrelated regression equations: a brief survey, *Journal of Econometrics*, **10**, 15–32.

[29] Srivastava, V.K. and Giles, D.E.A. (1987). *Seemingly Unrelated Regression Equations Models*, Dekker, New York.

[30] Theil, H. (1954). Estimation of parameters in econometric models, *Bulletin of the International Statistical Institute*, **34**, 122–129.

[31] Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias, *Journal of American Statistical Association*, **57**, 500–509.

[32] Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons, New York.

[33] Zellner, A. and Ando, T. (2010a). Approach for Bayesian analysis of the seemingly unrelated regression mode, *Journal of Econometrics*, **159**, 33–45.

[34] Zellner, A. and Ando, T. (2010b). Bayesian and non-Bayesian analysis of the seemingly unrelated regression model with student-t errors, and its application for forecasting, *International Journal of Forecasting*, **26**, 413–434.

[35] Zellner, A.; Ando, T.; Basturk, N.; Hoogerheide, L. and Van Dijk, H. (2014). Bayesian analysis of instrumental variable models: acceptance-rejection within direct Monte Carlo, *Econometric Reviews*, **33**, 3–35.

REVSTAT – STATISTICAL JOURNAL

Background

Statistics Portugal (INE, I.P.), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of a scientific statistical journal called Revista de Estatística. The original language used in this publication was Portuguese and the idea behind it was to publish it, three times a year, containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided that the publication should also include papers in English. This step was taken to achieve a broader dissemination, and to encourage foreign contributors to submit their work for publication.

At the time, the Editorial Board was mainly comprised of Portuguese university professors. It is now comprised of international university faculties and this has been the first step aimed at changing the character of Revista de Estatística from a national to an international scientific journal.

We have also initiated a policy of publishing special volumes that may be thematic highlighting areas of interest or associated with scientific events in Statistics. For example, in 2001, a special issue of Revista de Estatística was published containing three volumes of extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

In 2003, the name of the Journal has been changed to REVSTAT - STATISTICAL JOURNAL, now fully published in English, with a prestigious international editorial board, aiming to become a reference scientific journal that promotes the dissemination of relevant research results in Statistics.

The editorial policy of REVSTAT Statistical Journal is mainly placed on the originality and importance of the research.

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.

The only working language allowed is English.

Abstract and Indexing Services

The REVSTAT is covered by the following abstracting/indexing services:

- Current Index to Statistics
- Google Scholar
- Mathematical Reviews® (MathSciNet®)
- Science Citation Index Expanded
- Zentralblatt für Mathematic
- Scimago Journal & Country Rank
- Scopus

Instructions to Authors

Articles must be written in English and will be submitted according to the following guidelines:

The corresponding author sends the manuscript in PDF format to the Executive Editor (revstat@ine.pt) with the Subject "New Submission to REVSTAT"; a MS#REVSTAT reference will be assigned later.

Optionally, in a mail cover letter, authors are welcome to suggest one of the Editors or Associate Editors, whose opinion may be considered suitable to be taken into account.

The submitted manuscript should be original and not have been previously published nor about to be published elsewhere in any form or language, avoiding concerns about self-plagiarism'.

Content published in this journal is peer-reviewed (Single Blind).

All research articles will be refereed by at least two researchers, including one from the Editorial Board unless the submitted manuscript is judged unsuitable for REVSTAT or does not contain substantial methodological novelty, in which case is desk rejected.

Manuscripts should be typed only in black, in double-spacing, with a left margin of at least 3 cm, with numbered lines, and with less than 25 pages. Figures (minimum of 300dpi) will be reproduced online in colours, if produced this way; however, authors should take into account that the printed version is always in black and grey tones.

The first page should include the name, ORCID iD (optional), Institution, country, and mail-address of the author(s) and a summary of fewer than one hundred words, followed by a maximum of six keywords and the AMS 2000 subject classification.

Authors are encouraged to submit articles using LaTeX, in the REVSTAT style, which is available at the LaTeX2e MACROS webpage.

References about the format and other useful information on the submission are available in the LaTeX2e Templates page.

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.

Supplementary files (in REVSTAT style) 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).

Any contact with REVSTAT must always contain the assigned REVSTAT reference number.

Accepted papers

Authors of accepted papers are requested to provide the LaTex files to the Secretary of the REVSTAT revstat@ine.pt. The authors should also mention if figure files were included, and submit electronic figures separately in .gif, .jpg, .png or .pdf format. Figures must be a minimum of 300dpi.

Copyright and reprints

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, Statistics Portugal, in order to ensure the widest possible dissemination of information, namely through the Statistics Portugal website (http://www.ine.pt).

After assigning copyright, authors may use their own material in other publications provided that REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.

Editorial Board