



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT

Statistical Journal



Catálogo 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

CREDITS

- | | |
|--|--|
| <ul style="list-style-type: none">- EDITOR-IN-CHIEF<ul style="list-style-type: none">- <i>Isabel Fraga Alves</i>- CO-EDITOR<ul style="list-style-type: none">- <i>Giovani L. Silva</i>- ASSOCIATE EDITORS<ul style="list-style-type: none">- <i>Marília Antunes</i>- <i>Barry Arnold</i>- <i>Narayanaswamy Balakrishnan</i>- <i>Jan Beirlant</i>- <i>Graciela Boente</i>- <i>Paula Brito</i>- <i>Vanda Inácio de Carvalho</i>- <i>Valérie Chavez-Demoulin</i>- <i>David Conesa</i>- <i>Charmaine Dean</i>- <i>Fernanda Figueiredo</i>- <i>Jorge Milhazes Freitas</i>- <i>Alan Gelfand</i>- <i>Stéphane Girard</i>- <i>Marie Kratz</i>- <i>Victor Leiva</i>- <i>Artur Lemonte</i>- <i>Shuangzhe Liu</i>- <i>Maria Nazaré Mendes-Lopes</i>- <i>Fernando Moura</i>- <i>John Nolan</i>- <i>Paulo Eduardo Oliveira</i>- <i>Pedro Oliveira</i>- <i>Carlos Daniel Paulino (2019-2021)</i>- <i>Arthur Pewsey</i>- <i>Gilbert Saporta</i>- <i>Alexandra M. Schmidt</i>- <i>Julio Singer</i> | <ul style="list-style-type: none">- <i>Manuel Scotto</i>- <i>Lisete Sousa</i>- <i>Milan Stehlik</i>- <i>María Dolores Ugarte</i>- FORMER EDITOR-IN-CHIEF<ul style="list-style-type: none">- <i>M. Ivette Gomes</i>- FORMER CO-EDITOR<ul style="list-style-type: none">- <i>M. Antónia Amaral Turkman</i>- EXECUTIVE EDITOR<ul style="list-style-type: none">- <i>José A. Pinto Martins</i>- FORMER EXECUTIVE EDITOR<ul style="list-style-type: none">- <i>Maria José Carrilho</i>- <i>Ferreira da Cunha</i>- SECRETARIAT<ul style="list-style-type: none">- <i>José Cordeiro</i>- <i>Olga Bessa Mendes</i>- PUBLISHER<ul style="list-style-type: none">- <i>Instituto Nacional de Estatística, I.P. (INE, I.P.)</i>Web site: http://www.ine.pt- COVER DESIGN<ul style="list-style-type: none">- <i>Mário Bouçadas, designed on the stain glass window at INE by the painter Abel Manta</i>- LAYOUT AND GRAPHIC DESIGN<ul style="list-style-type: none">- <i>Carlos Perpétuo</i>- PRINTING<ul style="list-style-type: none">- <i>Instituto Nacional de Estatística, I.P.</i>- EDITION<ul style="list-style-type: none">- <i>140 copies</i>- LEGAL DEPOSIT REGISTRATION<ul style="list-style-type: none">- <i>N.º 191915/03</i>- PRICE [VAT included]<ul style="list-style-type: none">- <i>€ 9,00</i> |
|--|--|

© INE, Lisbon. Portugal, 2021

Statistical data made available by Statistics Portugal might be used according to Creative Commons Attribution 4.0

International (CC BY 4.0), however the source of the information must be clearly identified



INDEX

Statistical Inference for a General Class of Noncentral Elliptical Distributions <i>Jimmy Reyes, Diego I. Gallardo, Filidor Vilca and Héctor W. Gómez</i>	161
The Use of Control Charts to Monitor Air Plane Accidents of the Hellenic Air Force <i>Vasileios Alevizakos, Christos Koukouvinos and Petros E. Maravelakis</i>	187
On the Estimation for Compound Poisson INARCH Processes <i>E. Gonçalves, N. Mendes-Lopes and F. Silva</i>	207
Parameter Estimation for the Two-Parameter Maxwell Distribution under Complete and Censored Samples <i>Talha Arslan, Sukru Acitas and Birdal Senoglu</i>	237
Semiparametric Additive Beta Regression Models: Inference and Local Influence Diagnostics <i>Germán Ibacache-Pulgar, Jorge Figueroa-Zuñiga and Carolina Marchant</i>	255
Variance Estimation in the Presence of Measurement Errors under Stratified Random Sampling <i>Neha Singh, Gajendra K. Vishwakarma and Raj K. Gangele</i>	275
Concomitants of Order Statistics and Record Values from Iterated FGM Type Bivariate-Generalized Exponential Distribution <i>H.M. Barakat, E.M. Nigm, M.A. Alawady and I.A. Husseiny</i>	291

STATISTICAL INFERENCE FOR A GENERAL CLASS OF NONCENTRAL ELLIPTICAL DISTRIBUTIONS

Authors: JIMMY REYES

– Departamento de Matemáticas, Facultad de Ciencias Básicas,
Universidad de Antofagasta, Antofagasta, Chile
jimmy.reyes@uantof.cl

DIEGO I. GALLARDO

– Departamento de Matemática, Facultad de Ingeniería,
Universidad de Atacama, Copiapó, Chile
diego.gallardo@uda.cl

FILIDOR VILCA

– Departamento de Estatística, IMECC, Universidade Estadual de Campinas,
Campinas, Brasil
fily@ime.unicamp.br

HÉCTOR W. GÓMEZ

– Departamento de Matemáticas, Facultad de Ciencias Básicas,
Universidad de Antofagasta, Antofagasta, Chile
hector.gomez@uantof.cl

Received: December 2017

Revised: August 2018

Accepted: December 2018

Abstract:

- In this paper we introduce a new family of noncentral elliptical distributions. This family is generated as the quotient of two independent random variables, one with noncentral standard elliptical distribution and the other the power of a $U(0,1)$ random variable. For this family of distributions, we derive general properties, including the moments and discuss some special cases based on the family of scale mixtures of normal distributions, where the main advantage is easy simulation and nice hierarchical representation facilitating the implementation of an EM algorithm for maximum likelihood estimation. This new family of distributions provides a robust alternative for parameter estimation in asymmetric distributions. The results and methods are applied to three real datasets, showing that this new distribution fits better than other models reported in the recent statistical literature.

Keywords:

- *noncentral slash-elliptical distribution; elliptical distribution; moments; kurtosis; EM-algorithm.*

AMS Subject Classification:

- 62E10, 62F10.

1. INTRODUCTION

Many univariate or multivariate distributions have been generalized to noncentral versions. These include numerous continuous univariate (Student- t , chi-squared, gamma, beta) distributions. The noncentral Student- t (NCt) distribution is a skewed distribution that has received attention in the statistical inference context. When the mean of a normal distribution is tested, the noncentral distribution describes how a test statistic t is distributed when the null hypothesis is false. That is

$$t_\nu(\lambda) = \frac{Z + \lambda}{\sqrt{U/\nu}},$$

where $Z \sim N(0, 1)$ and $U \sim \chi_\nu^2$ are independent random variables. Lahiri and Teigland ([18]), and Dasgupta and Lahiri ([7]) found the NCt distribution is useful in analyzing survey data and forecasting record data. Tsionas ([29]) used the NCt distribution in linear regression models and applied it to stock market data. Applications of the NCt distribution have been limited by the fact that the probability density function is not expressible in closed form, making the maximum likelihood (ML) estimation difficult. On the other hand, the symmetric Student- t (t) distribution has a long history in statistics to model data with outliers as does as the elliptical (EL) distribution; see for example, Lange *et al.* ([19]), Fang *et al.* ([10]), and Cambanis *et al.* ([6]). A random variable X is said to have an EL distribution with location μ and scale parameter σ , denoted as $X \sim \text{EL}(\mu, \sigma^2; g)$ if its probability density function (pdf) is given by

$$(1.1) \quad f_X(x) = \frac{1}{\sigma} g\left(\left(\frac{x - \mu}{\sigma}\right)^2\right),$$

for some nonnegative function $g(u)$, $u \geq 0$, referred to as the density generator which satisfies $\int_0^\infty u^{-\frac{1}{2}} g(u) du = 1$. Based on this family of EL distributions, Gómez *et al.* ([13]) and Gómez and Venegas ([15]) introduced the slash-elliptical (SEL) family of distributions. These distributions originate from the ratio between two independent random variables, one the standard EL distribution and the other a uniform $(0, 1)$ distribution,

$$(1.2) \quad Y = \frac{Z}{U^{\frac{1}{q}}},$$

where $Z \sim \text{EL}(0, 1; g)$ and $U \sim U(0, 1)$ are independent random variables with $q > 0$. The resulting distribution is denoted by $Y \sim \text{SEL}(0, 1, q)$, and has heavier tails than the standard normal distribution. On the other hand, when q tends to ∞ , the resulting distribution is the standard EL distribution. For example, if $Z \sim N(0, 1)$ and $q = 1$, one obtains the canonic slash distribution,

$$(1.3) \quad f(y) = \begin{cases} \frac{\phi(0) - \phi(y)}{y}, & \text{if } y \neq 0, \\ \frac{\phi(0)}{2}, & \text{if } y = 0, \end{cases}$$

where $\phi(\cdot)$ is the pdf of the standard normal distribution. This distribution has heavier tails than the normal distribution, that is, it has higher kurtosis. Properties of this family are discussed in Rogers and Tukey ([28]), Mosteller and Tukey ([24]) and Johnson *et al.* ([16]).

ML estimators for location and scale parameters are discussed in Kafadar ([17]). Wang and Genton ([31]) described multivariate symmetrical and skew-multivariate extensions of the slash (S) distribution. Arslan and Genc ([3]) discussed a symmetric extension of the multivariate slash distribution and Genc ([12]) discussed a symmetric generalization of the slash distribution.

The aim of this paper is to provide an extension of the family of SEL distributions to a family of noncentral (NC) distributions. We derive its properties and method of estimating the model parameters. Also, we present a multivariate extension.

The paper is organized as follows: In Section 2, we present the pdf of the noncentral slash-elliptical (NCSEL) distribution, and some of its properties. Also, moments of order r are obtained, including the asymmetry and kurtosis coefficients. In Section 3, we discuss derivation of moment method and maximum likelihood estimation and report results of using the proposed model in three real applications. Section 4 reports examples using both simulated and real data to illustrate the performance of the proposed method. Section 5 presents a discussion of the multivariate case. Finally, some concluding remarks are given in Section 6.

2. NONCENTRAL SLASH-ELLIPTICAL DISTRIBUTIONS

In this section, we introduce a family of NCSEL distributions, which is defined through the following stochastic representation. A random variable Y represented as

$$(2.1) \quad Y = \frac{W + \lambda}{U^{\frac{1}{q}}}, \quad \lambda \in \mathbb{R}, \quad q > 0,$$

where $W \sim \text{EL}(0, 1; g)$ and $U \sim \text{U}(0, 1)$ are independent random variables, is said to have a NCSEL distribution, with λ being the non-centrality parameter and q the kurtosis parameter. This distribution will be denoted by $Y \sim \text{NCSEL}(1, q, \lambda; g)$. Before presenting some of its important properties, we present two special cases. If $W \sim \text{N}(0, 1)$, then Y follows a noncentral slash (NCS) distribution, denoted by $Y \sim \text{NCS}(1, q, \lambda)$, while if W follows a t distribution, $t(0, 1; \nu)$, then the resulting distribution is a noncentral slash-Student- t (NCSt) distribution, denoted by $Y \sim \text{NCSt}(1, q, \lambda; \nu)$. For the special case of $q = 1$, this distribution is called the canonical NCSEL distribution.

2.1. Density function

The stochastic representation in (2.1) is useful to obtain the pdf of Y , as shown in the following result.

Proposition 2.1. *Let $Y \sim \text{NCSEL}(1, q, \lambda; g)$. Then, the pdf of Y is given by*

$$f_Y(y; 1, q, \lambda) = \begin{cases} \frac{q}{y^{q+1}} \int_{-\lambda}^{y-\lambda} (u + \lambda)^q g(u^2) du, & \text{if } y \neq 0, \\ \frac{q}{q + 1} g(\lambda^2), & \text{if } y = 0. \end{cases}$$

Proof: From (2.1), using the fact that U and W are independent and standard calculations (based on the Jacobian of the appropriate transformation), we obtain

$$f_{Y,U}(y, u) = u^{\frac{1}{q}} g\left((yu^{\frac{1}{q}} - \lambda)^2\right), \quad y \in \mathbb{R}, \quad 0 < u < 1.$$

Hence, the marginal pdf of Y is given by

$$f_Y(y; 1, q, \lambda) = \int_0^1 u^{\frac{1}{q}} g\left((yu^{\frac{1}{q}} - \lambda)^2\right) du.$$

Now, by substituting u for $u = yt^{\frac{1}{q}} - \lambda$, we have the required results for $y \neq 0$. For $y = 0$, the result is immediate. \square

Corollary 2.1. For the special case $q = 1$, the pdf reduces to the form

$$f_Y(y; 1, 1, \lambda) = \begin{cases} \frac{1}{y^2} \int_{-\lambda}^{y-\lambda} (u + \lambda) g(u^2) du, & \text{if } y \neq 0, \\ \frac{1}{2} g(\lambda^2), & \text{if } y = 0. \end{cases}$$

Corollary 2.2. If $W \sim N(0, 1)$, then

i) The pdf of Y is

$$f_Y(y; 1, q, \lambda) = \frac{1}{\sqrt{2\pi}} \int_0^1 u^{\frac{1}{q}} e^{-\frac{1}{2}(yu^{\frac{1}{q}} - \lambda)^2} du;$$

ii) For $q = 1$, the pdf of Y can be expressed as

$$f(y; 1, 1, \lambda) = \begin{cases} \frac{1}{2} \left\{ \phi(\lambda) - \phi(y - \lambda) + \lambda \left(\Phi(y - \lambda) + \Phi(\lambda) - 1 \right) \right\}, & \text{if } y \neq 0, \\ \frac{\phi(\lambda)}{2}, & \text{if } y = 0, \end{cases}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and the cumulative distribution function (cdf) of the standard normal distribution, respectively.

Proof: Both parts are direct consequences of Proposition 2.1. In Part i) consider $g(u) = (1/\sqrt{2\pi}) \exp(-u/2)$, and in Part ii), for $y \neq 0$, we have

$$f_Y(y; \lambda) = \int_0^1 \frac{u}{\sqrt{2\pi}} e^{-\frac{1}{2}(yu - \lambda)^2} du.$$

Letting $w = yu - \lambda$, $f_Y(y; \lambda)$ can be expressed as

$$\begin{aligned} f_Y(y; \lambda) &= \frac{1}{\sqrt{2\pi} y^2} \int_0^1 (w + \lambda) e^{-\frac{w^2}{2}} dw \\ &= \frac{1}{y^2} \left[\frac{1}{\sqrt{2\pi}} \left(e^{-\frac{\lambda^2}{2}} + e^{-\frac{(y-\lambda)^2}{2}} \right) + \int_{-\lambda}^{y-\lambda} \phi(w) dw \right] \\ &= \frac{1}{y^2} \left\{ \phi(\lambda) - \phi(y - \lambda) + \lambda \left(\Phi(y - \lambda) + \Phi(\lambda) - 1 \right) \right\}. \end{aligned}$$

Finally, for $y = 0$, the result is direct. \square

Figure 1 illustrates some possible shapes of the pdf of Y for some parameter values of λ . It can be seen that the parameter λ controls the skewness of the distribution. It is also possible to observe that, as $|\lambda|$ increases, the density becomes more skewed. Figure 2 displays some possible shapes of the pdf of Y for some parameter values of q and $\sigma = 1$. From this figure, we note that the parameter q controls the kurtosis of the distribution. Moreover, for smaller values of q we have a heavy-tailed distribution.

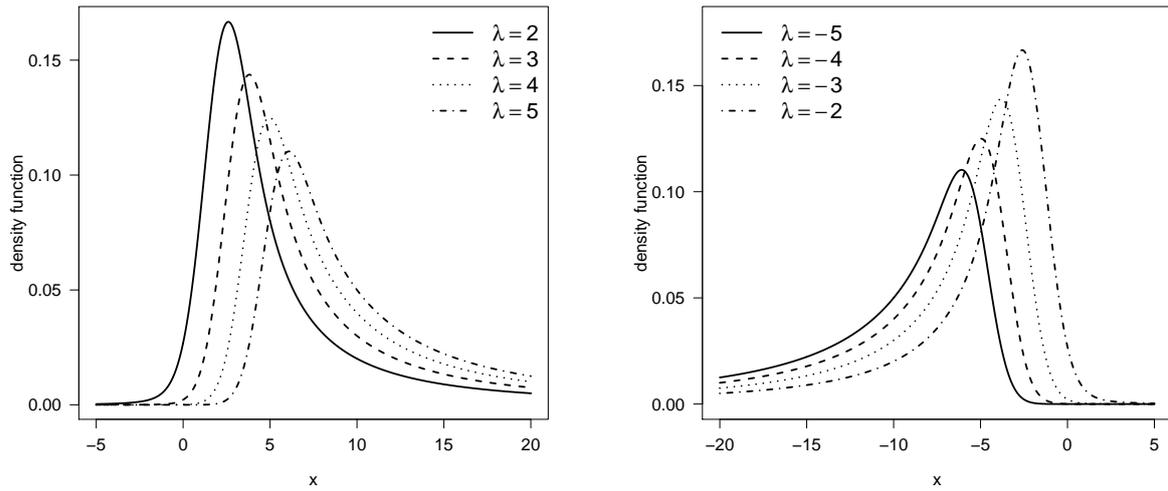


Figure 1: NCS pdf plots for $q = 1$ and different values of λ .

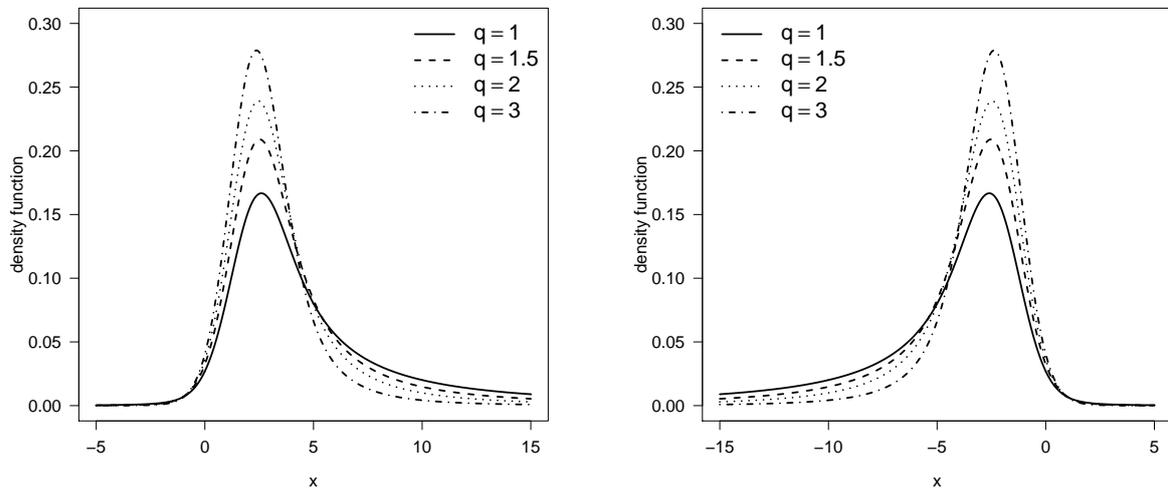


Figure 2: NCS pdf plots for $\lambda = 2$ (left panel) and $\lambda = -2$ (right panel) and different values of q .

A slight extension of the NCSEL distribution is obtained by introducing a scale parameter through the representation

$$(2.2) \quad Y = \frac{\sigma W + \lambda}{U^{\frac{1}{q}}} = \sigma \frac{W + \delta}{U^{\frac{1}{q}}},$$

where $\delta = \lambda/\sigma$, $W \sim \text{EL}(0, 1; g)$ and $U \sim \text{U}(0, 1)$ are independent, and σ is a scale parameter.

This distribution is denoted by $\text{NCSEL}(\sigma, q, \lambda; g)$, and its pdf is given by

$$f_Y(y; \sigma, q, \lambda) = \frac{1}{\sigma} \int_0^1 u^{\frac{1}{q}} g\left(\left(\frac{y u^{\frac{1}{q}} - \lambda}{\sigma}\right)^2\right) du.$$

An important class of symmetric distributions is the family of normal/independent (NI) (or scale mixture of normal) distributions, which contains many important unimodal distributions such as the contaminated normal (CN), S, t and Laplace (L) distributions, among others, all possessing heavier tails than the normal. For more information on this family of distributions, see for example, Andrews and Mallows ([2]) and Lange and Sinsheimer ([20]). A random variable W is said to have a standard NI distribution, if it can be related to the normal distribution through the stochastic representation $W = V^{-1/2}Z_0$, where $Z_0 \sim N(0, 1)$ is independent of the positive random variable V . The pdf of W can be expressed as

$$(2.3) \quad \phi_{\text{NI}}(w) = \int_0^\infty \frac{v^{1/2}}{\sqrt{2\pi}} \exp\left\{-\frac{v}{2} w^2\right\} dH_V(v; \boldsymbol{\nu}),$$

where $H_V(\cdot; \boldsymbol{\nu})$ is the cdf of V , indexed by a scalar or vector of parameters $\boldsymbol{\nu}$. The distribution of W is denoted by $W \sim \text{NI}(0, 1; H_V)$. In the EL distribution context, the generator function $g(\cdot)$ for an NI distribution is

$$(2.4) \quad g(u) = \int_0^\infty \frac{v^{1/2}}{\sqrt{2\pi}} \exp\left\{-\frac{v}{2} u\right\} dH_V(v), \quad v > 0.$$

Some special cases of the family of NI distributions are for example:

- 1) *The CN distribution:* Here V has pdf given by $h_V(v) = \nu \mathbb{I}_{\{\gamma\}}(v) + (1 - \nu) \mathbb{I}_{\{1\}}(v)$, $0 < \nu < 1$, $0 < \gamma < 1$, where $\mathbb{I}_A(\cdot)$ denotes the indicator function of the set A and $\boldsymbol{\nu} = (\nu, \gamma)^\top$. Then, the pdf of W is

$$\phi_{\text{NI}}(w) = \left[\nu \sqrt{\gamma} \phi(\sqrt{\gamma} w) + (1 - \nu) \phi(w) \right], \quad y \in \mathbb{R}.$$

- 2) *The S distribution:* Here $V \sim \text{Beta}(\nu, 1)$ and the pdf of W is

$$\phi_{\text{NI}}(w) = \nu \int_0^1 v^{\nu-1} \phi(w; 0, v^{-1}) dv, \quad w \in \mathbb{R}.$$

- 3) *The t distribution:* Here $V \sim \text{Gamma}(\nu/2, \nu/2)$, so the t distribution has as special cases the Cauchy model for $\nu = 1$ and the normal model as $\nu \rightarrow \infty$, and the pdf of W is

$$\phi_{\text{NI}}(w) = k(\nu) \nu^{\nu/2} (\nu + w^2)^{-\left(\frac{\nu+1}{2}\right)}, \quad w \in \mathbb{R},$$

where $k(\nu) = \Gamma\left(\frac{\nu+1}{2}\right) / [\sqrt{\pi} \Gamma\left(\frac{\nu}{2}\right)]$.

Remark 2.1. The special case $Y \sim \text{NCS}(\sigma, 1, \lambda)$, i.e. $q = 1$, will be called as the canonical NCS and its pdf is

$$f(y, \sigma, \lambda) = \begin{cases} \frac{\sigma^2}{y^2} \left[\phi\left(\frac{\lambda}{\sigma}\right) - \phi\left(\frac{y - \lambda}{\sigma}\right) + \frac{\lambda}{\sigma} \left(\Phi\left(\frac{y - \lambda}{\sigma}\right) + \Phi\left(\frac{\lambda}{\sigma}\right) - 1 \right) \right], & \text{if } y \neq 0, \\ \frac{\phi\left(\frac{\lambda}{\sigma}\right)}{2}, & \text{if } y = 0, \end{cases}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively.

2.2. Properties

In this section, we present some properties of the NCSEL distribution.

Proposition 2.2. *Let $Y \sim \text{NCSEL}(\sigma, q, \lambda; g)$. Then,*

- i) *If $\lambda = 0$ and $q \rightarrow \infty$, then $Y \sim \text{EL}(0, \sigma^2; g)$;*
- ii) *If $\lambda = 0$, then $Y \sim \text{SEL}(0, \sigma^2, q; g)$;*
- iii) *If $U_1 = U^{1/q}$ in (2.2), then the conditional pdf of $U_1 = u$, given $Y = y$, is*

$$f_{U_1|Y}(u|y) = \frac{qu^{q-1}}{f_Y(y)} f_{Y|U_1}(y|u) I_{(0,1)}(u),$$

where $f_{Y|U_1}(\cdot|u)$ is the pdf of $\text{EL}(\frac{\lambda}{u}, \frac{\sigma^2}{u^2}; g)$ distribution;

- iv) *If $W = V^{-1/2}Z_0 \sim \text{NI}(0, 1; H_V)$ in (2.2), then the conditional mean of U^rV^s for $r \geq 0, s \geq 0$, given $Y = y$, is*

$$E[U^rV^s|y] = \frac{q}{\sigma f_Y(y)} \int_0^1 u^{r+q} \left[\int_0^\infty \frac{v^{s+1/2}}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{v}{2\sigma^2}(uy - \lambda)^2\right\} f_V(v) dv \right] du.$$

For the special case $V \sim \text{Gamma}(\nu/2, \nu/2)$,

$$E[U^rV^s|y] = \frac{qd(\nu, s)}{\sigma f_Y(y)} \int_0^1 u^{r+q} \left[\nu + \left(\frac{uy - \lambda}{\sigma}\right)^2 \right]^{-\frac{2s+\nu+1}{2}} du,$$

where

$$d(\nu, s) = 2^s \nu^{\nu/2} \Gamma\left(\frac{2s + \nu + 1}{2}\right) / \left(\sqrt{\pi} \Gamma\left(\frac{\nu}{2}\right)\right).$$

Remark 2.2. We now present some comments on the usefulness of the results proposed in Proposition 2.2:

- 1) Parts i) and ii) state that the NCSEL distribution contains the elliptical distribution as a special case as $q \rightarrow \infty$ and the noncentral parameter is zero ($\lambda = 0$). Moreover, the NCSEL distribution contains as special case the SEL distribution when $\lambda = 0$.
- 2) Letting $U_2 = U^{-1/q}$ in the representation in (2.2), we can get the following model

$$(2.5) \quad Y = \mu + \lambda U_2 + U_2 W,$$

where $W \sim \text{EL}(0, \sigma^2; g)$ and U_2 are independent and $\mu \in \mathbb{R}$. We note that the conditional distribution of Y , given $U_2 = u$ follows a $Y | (U_2 = u) \sim \text{EL}(\mu + \lambda u, u\sigma^2; g)$ for some density generator $g(\cdot)$.

- 3) The distribution in (2.5) is like a variance-mean mixture of the EL distribution proposed by Barndorff-Nielsen ([5]), in which W follows a normal distribution, which has been used in financial empirical studies.
- 4) Finally, Part iv) is useful to implement the EM-algorithm in ML estimation.

2.3. Moments

In this section, we discuss distributional moments of the NCSEL distribution, an important need in any statistical analysis. Some of the important characteristics of a distribution can be studied through moments, which are used to derive moment estimators, and skewness and kurtosis coefficients.

Proposition 2.3. *Let $Y \sim \text{NCSEL}(\sigma, q, \lambda; g)$ such that*

$$Y = \frac{\sigma W + \lambda}{U^{\frac{1}{q}}} = \sigma \frac{W + \delta}{U^{\frac{1}{q}}},$$

where $\delta = \frac{\lambda}{\sigma}$. Then, for $r = 1, 2, 3, \dots$ and $q > r$, $E[Y^r] = \sigma^r \mu_r$, where

$$\mu_r = E[X^r] = \frac{q}{q-r} \sum_{k=0}^r \binom{r}{k} \delta^{r-k} a_{k/2},$$

with $X \sim \text{NCSEL}(1, q, \delta; g)$ and $a_{k/2} = \int_{-\infty}^{\infty} x^k g(x^2) dx$.

Proof: Using the stochastic representation of X and Y , and the independence of W and U , we have

$$\mu_r = E[X^r] = E\left[\left(\frac{W + \delta}{U^{\frac{1}{q}}}\right)^r\right] = E[(W + \delta)^r] E[U^{-\frac{r}{q}}].$$

Using the binomial theorem for $(W + \delta)^r$ and applying expectation, we have

$$E[(W + \delta)^r] = \sum_{k=0}^r \binom{r}{k} \delta^{r-k} E[W^k],$$

where $E[W^k] = a_{k/2} = \int_{-\infty}^{\infty} x^k g(x^2) dx$. Since $E[U^{-\frac{r}{q}}] = \frac{q}{q-r}$, $q > r$, we obtain the required result. \square

Corollary 2.3. *Let $Y \sim \text{NCSEL}(\sigma, q, \lambda; g)$. Then, the mean and variance of Y are given by*

$$E[Y] = \frac{\lambda q}{q-1}, \quad q > 1, \quad \text{and} \quad \text{Var}(Y) = \frac{\sigma^2 q}{q-2} \left(\left(\frac{\lambda}{\sigma(q-1)} \right)^2 + a_1 \right), \quad q > 2.$$

Proposition 2.4. *Let $Y \sim \text{NCSEL}(\sigma, q, \lambda; g)$. Then, the asymmetry and kurtosis coefficients of Y are respectively*

$$\gamma_1 = \frac{\frac{q}{q-3} (\delta^3 + 3a_1) - \frac{3\delta q}{(q-1)(q-2)} (\delta^2 + a_1) + \frac{2\delta^3 q^3}{(q-1)^3}}{\left[\frac{q}{q-2} \left(\frac{\delta^2}{(q-1)^2} + a_1 \right) \right]^{\frac{3}{2}}}, \quad q > 3,$$

$$\beta_2 = \frac{\frac{q}{q-4} (\delta^4 + 6\delta a_1 + a_2) - \frac{4\delta q^2}{(q-1)(q-3)} (\delta^3 + 3a_1) + \frac{6\delta^2 q^3}{(q-1)^2(q-2)} (\delta^2 + a_1) - \frac{3\delta^4 q^4}{(q-1)^4}}{\left[\frac{q}{q-2} \left(\frac{\delta^2}{(q-1)^2} + a_1 \right) \right]^2}, \quad q > 4.$$

Proof: The proof follows by using the formulas of asymmetry and kurtosis coefficients given respectively by

$$\gamma_1 = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{\frac{3}{2}}} \quad \text{and} \quad \beta_2 = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2},$$

where $\mu_k, k = 1, \dots, 4$, as given in Proposition 2.3. □

Figure 3 displays graphs for the asymmetry coefficient and kurtosis coefficient of the NCS distribution.

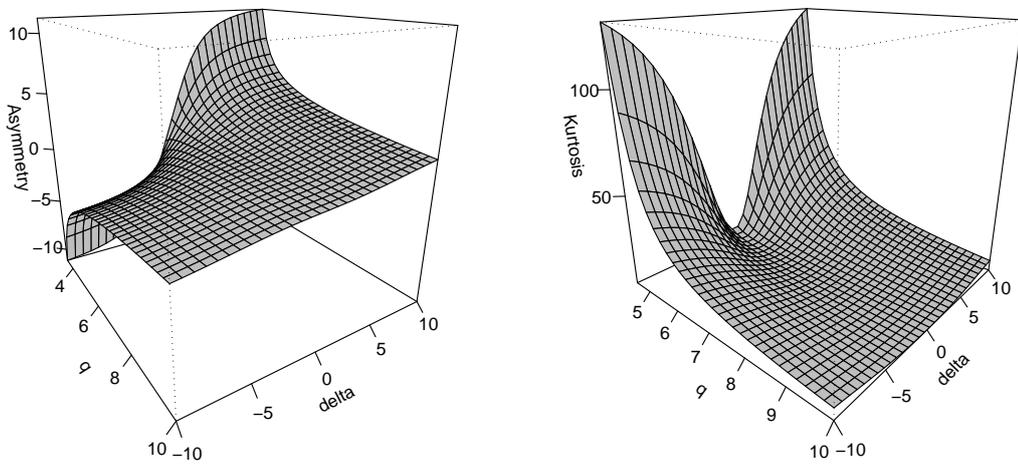


Figure 3: Graphs for the asymmetry coefficient (left) and the kurtosis coefficient (right) of the NCS distribution.

3. INFERENCE

Here, we discuss the moment method (MM) and ML estimation for parameters λ, σ and q of the NCSEL distribution based on a random sample Y_1, \dots, Y_n of $Y \sim \text{NCSEL}(\sigma, q, \lambda; g)$. We present the MM estimation and then the ML estimation.

3.1. Method of moment estimation

We discuss an MM estimation based on the distributional moments which are presented in the following result.

Proposition 3.1. *The moment estimators of λ and σ are*

$$\widehat{\lambda}_M(\widehat{q}_M) = \frac{\overline{Y}(\widehat{q}_M - 1)}{\widehat{q}_M} \quad \text{and} \quad \widehat{\sigma}_M(\widehat{q}_M) = \sqrt{\frac{1}{a_1} \left(\frac{S^2(\widehat{q}_M - 2)}{\widehat{q}_M} - \frac{\widehat{\lambda}_M^2}{(\widehat{q}_M - 1)^2} \right)},$$

where $a_1 = E[W^2]$, whereas the moment estimator of q is the solution in the interval $(3, \infty)$ for the nonlinear equation

$$(q - 3)\overline{Y}^3 - q(\sigma_M(q))^3 \left(\left(\frac{\lambda_M(q)}{\sigma_M(q)} \right)^3 + 3a_1 \right) = 0.$$

Proof: These equations follow from Proposition 2.3 and Corollary 2.3. \square

3.2. Maximum likelihood estimation

We now discuss the ML estimation for a sample of size n . The log-likelihood function for the parameters σ , q and λ can be written as

$$(3.1) \quad \ell(\sigma, q, \lambda) = -n \log(\sigma) + \sum_{i=1}^n \log G(y_i),$$

where $G(y_i) = G(y_i; \sigma, q, \lambda) = \int_0^1 v^{\frac{1}{q}} g\left(\left(\frac{y_i v^{\frac{1}{q}} - \lambda}{\sigma}\right)^2\right) dv$ and hence the likelihood equations are given by

$$(3.2) \quad \sum_{i=1}^n \frac{G_\sigma(y_i)}{G(y_i)} = \frac{n}{\sigma}, \quad \sum_{i=1}^n \frac{G_q(y_i)}{G(y_i)} = 0, \quad \sum_{i=1}^n \frac{G_\lambda(y_i)}{G(y_i)} = 0,$$

where $G_\sigma(y_i) = \frac{\partial}{\partial \sigma} G(y_i)$, $G_q(y_i) = \frac{\partial}{\partial q} G(y_i)$, $G_\lambda(y_i) = \frac{\partial}{\partial \lambda} G(y_i)$, which can be expressed as

$$\begin{aligned} G_\sigma(y_i) &= -\frac{2}{\sigma} \int_0^1 u^{\frac{1}{q}} g'(t_i^2) t_i^2 du, \\ G_q(y_i) &= -\frac{1}{\sigma q^2} \int_0^1 u^{\frac{1}{q}} \log(u) \left(\sigma g(t_i^2) + 2 t_i y_i g'(t_i^2) \right) du, \\ G_\lambda(y_i) &= -\frac{2}{\sigma} \int_0^1 u^{\frac{1}{q}} g'(t_i^2) t_i du, \end{aligned}$$

where $t_i = (y_i u^{\frac{1}{q}} - \lambda)/\sigma$. Solutions for equations in (3.2) can be obtained using numerical procedures such as the Newton–Raphson procedure. This procedure requires the maximization of the log-likelihood function which involves integrals that make the maximization difficult, especially when the NCSEL model is based on a bimodal elliptical distribution. But when the NCSEL model is based on the family of the NI distributions, an EM algorithm can be implemented to obtain the ML estimates of the model parameters, as we show next.

3.3. EM algorithm

The EM-algorithm is a well known technique for the ML estimation when unobserved (or missing) data or latent variables are present while modeling. This estimation algorithm enables computationally efficient determination of the ML estimates when iterative methods are required. For a random sample of size n of the NCSEL($\sigma, q, \lambda; \boldsymbol{\nu}$) model, let $\mathbf{y} = (y_1, \dots, y_n)^\top$ be observed data, and let $\mathbf{u} = (u_1, \dots, u_n)^\top$ and $\mathbf{v} = (v_1, \dots, v_n)$ be unobserved data, so the complete dataset is $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{v}^\top, \mathbf{u}^\top)^\top$. In what follows, we describe the implementation of the EM-algorithm for the ML estimation of the parameters of the NCSEL model. For this purpose, we first present the NCSEL model in an incomplete-data framework, where the model can be written hierarchically as

$$(3.3) \quad \begin{aligned} \mathbf{Y} \mid U_i = u_i, V_i = v_i &\sim N(u_i^{-1}\lambda, \sigma^2 u_i^{-2} v_i^{-1}), \\ U_i \mid V_i = v_i &\sim \text{Beta}(q, 1), \\ V_i &\sim h(\cdot). \end{aligned}$$

The complete-data log-likelihood function for $\boldsymbol{\theta} = (\sigma, q, \lambda)^\top$ given \mathbf{y}_c (without the additive constant) is given by

$$\ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c) = -\frac{n}{2} \log \sigma^2 + \frac{1}{2} \sum_{i=1}^n \log(u_i^2 v_i) - \frac{1}{2\sigma^2} \sum_{i=1}^n (u_i^2 v_i y_i^2 - 2 u_i v_i \lambda y_i + \lambda^2 v_i) + \ell_c(q \mid \mathbf{y}_c),$$

where $\ell_c(q \mid \mathbf{y}_c) = \sum_{i=1}^n \ell_{ci}(q \mid \mathbf{y}_c)$, with $\ell_{ci}(q \mid \mathbf{y}_c) = \log q + (q - 1) \log u_i$. Letting $\widehat{u_i v_i} = E(U_i V_i \mid \mathbf{y}_i, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}})$, $\widehat{u_i^2 v_i} = E(U_i^2 V_i \mid \mathbf{y}_i, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}})$ and $\widehat{v_i} = E(V_i \mid \mathbf{y}_i, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}})$. The conditional expectation of the complete-data log-likelihood function (without the additive constant) is given by $Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}) = E[\ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c) \mid \mathbf{y}, \widehat{\boldsymbol{\theta}}] = \sum_{i=1}^n Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}})$, where $Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}})$ has the form

$$Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}) = -\frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (\widehat{u_i^2 v_i} y_i^2 - 2 \lambda \widehat{u_i v_i} y_i + \lambda^2 \widehat{v_i}) + Q_{ci}(q \mid \widehat{\boldsymbol{\theta}}),$$

where $Q_{ci}(q \mid \widehat{\boldsymbol{\theta}}) = \log q + (q - 1) S_i$, with $S_i = E[\log U_i \mid \mathbf{y}_i]$, $i = 1, \dots, n$. Since the quantity S_i does not have closed form, it must be computed numerically. We follow the idea from Lee and Xu ([21]) and Reyes *et al.* ([27]) to compute $Q_{ci}(q \mid \widehat{\boldsymbol{\theta}})$. Specifically, let $\{u_r; r = 1, \dots, R\}$ be a sample randomly drawn from the conditional distribution $U_i \mid (Y_i = y_i, \boldsymbol{\theta} = \widehat{\boldsymbol{\theta}})$, so the quantity $Q_{ci}(q \mid \widehat{\boldsymbol{\theta}})$ can be approximated as follows:

$$Q_{ci}(q \mid \widehat{\boldsymbol{\theta}}) \approx \frac{1}{R} \sum_{r=1}^R \ell_{ci}(q \mid u_r).$$

We then have the EM-algorithm for the ML estimation of the parameters of the NCSEL model as follows:

E-Step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\sigma}^{(k)}, \widehat{q}^{(k)}, \widehat{\lambda}^{(k)})^\top$, compute $\widehat{u_i v_i}^{(k)}$, $\widehat{u_i^2 v_i}^{(k)}$ and $\widehat{v_i}^{(k)}$, for $i = 1, \dots, n$;

CM-step I: Update $\widehat{\lambda}^{(k)}$ and $\widehat{\sigma}^{(k)}$ and maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ over λ and σ , which leads to the expressions:

$$\widehat{\lambda}^{(k+1)} = \frac{\sum_{i=1}^n \widehat{u}_i \widehat{v}_i^{(k)} y_i}{\sum_{i=1}^n \widehat{v}_i^{(k)}},$$

$$\widehat{\sigma}^{2(k+1)} = \frac{1}{n} \sum_{i=1}^n \left(\widehat{u}_i^2 \widehat{v}_i^{(k)} y_i^2 - 2 \widehat{\lambda}^{(k+1)} \widehat{u}_i \widehat{v}_i^{(k)} y_i + \widehat{\lambda}^{2(k+1)} \widehat{v}_i^{(k)} \right);$$

CM-step II: Fix $\lambda = \widehat{\lambda}^{(k)}$ and $\sigma^2 = \widehat{\sigma}^{2(k)}$, update $q^{(k)}$ by

$$\widehat{q}^{(k+1)} = \arg \max_q Q\left(\widehat{\lambda}^{(k)}, \widehat{\sigma}^{2(k)}, q | \widehat{\boldsymbol{\theta}}^{(k)}\right).$$

The iterations are repeated until a suitable convergence rule is satisfied, say $\|\boldsymbol{\theta}^{(l+1)} - \boldsymbol{\theta}^{(l)}\|$ sufficiently small. Useful starting values are required to implement this algorithm, and the moment estimates can be used effectively as initial values in the iterative procedure for computing the ML estimates.

3.4. Estimation of standard errors

To compute the standard errors of the ML estimates, we follow the information-based method exploited by Louis ([22]) and Meilijson ([23]), who proposed using of empirical information matrix, which is computed as

$$I_c(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^n s(y_i|\boldsymbol{\theta}) s(y_i|\boldsymbol{\theta})^\top - \frac{1}{n} S(\mathbf{y}|\boldsymbol{\theta}) S(\mathbf{y}|\boldsymbol{\theta})^\top,$$

where $S(\mathbf{y}|\boldsymbol{\theta}) = \sum_{i=1}^n s(y_i|\boldsymbol{\theta})$, with $s(y_i|\boldsymbol{\theta}) = \mathbb{E}[(\partial \ell(\boldsymbol{\theta}|\mathbf{y}_{ci})/\partial \boldsymbol{\theta}) | y_i, \boldsymbol{\theta}]$ being the empirical score function for the i -th individual, which can be written as

$$s(y_i|\boldsymbol{\theta}) = \left(\partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial \sigma, \partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial q, \partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial \lambda \right)^\top,$$

whose elements are given by

$$\begin{aligned} \partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial \sigma &= -\frac{1}{\sigma} + \frac{1}{\sigma^3} \left(\widehat{u}_i^2 \widehat{v}_i y_i^2 - 2 \lambda \widehat{u}_i \widehat{v}_i y_i + \lambda^2 \widehat{v}_i \right), \\ \partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial q &= \frac{1}{q} + \mathbb{E}[\log U_i | y_i], \\ \partial Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})/\partial \lambda &= \frac{1}{\sigma^2} \left(\widehat{u}_i \widehat{v}_i y_i - \lambda \widehat{v}_i \right). \end{aligned}$$

Now, replacing $\boldsymbol{\theta}$ by its ML estimates $\widehat{\boldsymbol{\theta}}$ in $I_c(\boldsymbol{\theta}|\mathbf{y})$, we obtain

$$I_c(\widehat{\boldsymbol{\theta}}|\mathbf{y}) = \sum_{i=1}^n s(y_i|\widehat{\boldsymbol{\theta}}) s(y_i|\widehat{\boldsymbol{\theta}})^\top - \frac{1}{n} S(\mathbf{y}|\widehat{\boldsymbol{\theta}}) S(\mathbf{y}|\widehat{\boldsymbol{\theta}})^\top,$$

which is used to compute the standard errors of the ML estimates.

4. ILLUSTRATIVE EXAMPLES

4.1. Simulation study

For each scenario, we simulate data based on the stochastic representation of the model presented in (2.1). The objective of this simulation study is to evaluate if the estimation algorithm developed in Section 3.3 can recover the parameters with which the simulation is performed. We consider two special cases of NCSEL models based on the NCS distribution (Table 1) and the NCSt distribution with $\nu = 5$ (Table 2), while for $\nu = 10$ the result is reported in the Appendix (see Table 9). We consider three cases for λ : $-0.5, 0.5$ and 1.0 ;

Table 1: Simulation for the NCS distribution.

true values			$\hat{\theta}$	$n = 50$			$n = 100$			$n = 200$		
λ	σ	q		mean	s.e.	$\sqrt{\text{MSE}}$	mean	s.e.	$\sqrt{\text{MSE}}$	mean	s.e.	$\sqrt{\text{MSE}}$
-0.5	0.5	1	$\hat{\lambda}$	-0.5223	0.1201	0.1238	-0.5137	0.0837	0.0899	-0.5128	0.0584	0.0678
			$\hat{\sigma}$	0.5250	0.1203	0.1333	0.5127	0.0826	0.1046	0.5129	0.0579	0.0730
			\hat{q}	1.1230	0.2760	0.4144	1.0610	0.1693	0.1810	1.0391	0.1143	0.1229
	3	$\hat{\lambda}$	-0.5237	0.1067	0.1125	-0.5160	0.0739	0.0762	-0.5054	0.0502	0.0518	
		$\hat{\sigma}$	0.5150	0.1054	0.1062	0.5118	0.0733	0.0794	0.5032	0.0494	0.0494	
		\hat{q}	4.4343	3.7121	3.1171	3.7622	1.7486	2.0391	3.2533	0.7639	0.9762	
1.0	1	$\hat{\lambda}$	-0.5244	0.1922	0.1874	-0.5143	0.1328	0.1406	-0.5105	0.0926	0.0965	
		$\hat{\sigma}$	1.0459	0.2533	0.2627	1.0377	0.1737	0.2160	1.0269	0.1207	0.1564	
		\hat{q}	1.1265	0.3081	0.4225	1.0654	0.1807	0.1958	1.0447	0.1222	0.1346	
3	$\hat{\lambda}$	-0.5162	0.1782	0.1732	-0.5184	0.1230	0.1256	-0.5065	0.0843	0.0843		
	$\hat{\sigma}$	1.0549	0.2290	0.2290	1.0349	0.1585	0.1624	1.0150	0.1068	0.1120		
	\hat{q}	4.9179	5.0027	3.6413	3.9935	2.3857	2.4501	3.3853	1.0160	1.3326		
0.5	0.5	1	$\hat{\lambda}$	0.5257	0.1206	0.1239	0.5151	0.0838	0.0896	0.5104	0.0588	0.0669
			$\hat{\sigma}$	0.5261	0.1207	0.1270	0.5149	0.0831	0.0982	0.5139	0.0583	0.0812
			\hat{q}	1.1091	0.2599	0.2906	1.0582	0.1691	0.1826	1.0365	0.1147	0.1271
	3	$\hat{\lambda}$	0.5201	0.1061	0.1006	0.5133	0.0741	0.0755	0.5085	0.0504	0.0504	
		$\hat{\sigma}$	0.5185	0.1053	0.1077	0.5140	0.0741	0.0804	0.5071	0.0497	0.0508	
		\hat{q}	4.4930	3.6785	3.1798	3.8143	1.8557	2.1013	3.2819	0.7595	0.8955	
1.0	1	$\hat{\lambda}$	0.5206	0.1930	0.2006	0.5186	0.1325	0.1336	0.5085	0.0924	0.0941	
		$\hat{\sigma}$	1.0612	0.2536	0.2761	1.0364	0.1736	0.1792	1.0254	0.1204	0.1304	
		\hat{q}	1.1279	0.3017	0.3844	1.0627	0.1793	0.1973	1.0433	0.1217	0.1272	
3	$\hat{\lambda}$	0.5269	0.1779	0.1758	0.5128	0.1230	0.1304	0.5086	0.0848	0.0842		
	$\hat{\sigma}$	1.0445	0.2312	0.2270	1.0398	0.1593	0.1701	1.0184	0.1082	0.1115		
	\hat{q}	4.8398	5.0577	3.5896	4.0944	2.5002	2.5839	3.3558	0.9935	1.2306		
1.0	0.5	1	$\hat{\lambda}$	1.0273	0.1600	0.1719	1.0230	0.1120	0.1271	1.0211	0.0785	0.1279
			$\hat{\sigma}$	0.5068	0.1157	0.1313	0.5051	0.0797	0.1007	0.5123	0.0554	0.1152
			\hat{q}	1.0760	0.2167	0.2387	1.0495	0.1457	0.1567	1.0396	0.1010	0.1253
	3	$\hat{\lambda}$	1.0303	0.1318	0.1380	1.0150	0.0899	0.0939	1.0087	0.0620	0.0610	
		$\hat{\sigma}$	0.5076	0.0980	0.1059	0.5045	0.0670	0.0690	0.5015	0.0460	0.0435	
		\hat{q}	3.8927	2.1092	2.2969	3.3113	0.8944	1.1456	3.1370	0.5151	0.5628	
1.0	1	$\hat{\lambda}$	1.0566	0.2452	0.2488	1.0311	0.1686	0.1752	1.0307	0.1186	0.1524	
		$\hat{\sigma}$	1.0597	0.2465	0.2595	1.0353	0.1674	0.1953	1.0326	0.1172	0.1834	
		\hat{q}	1.1176	0.2673	0.3278	1.0721	0.1720	0.1955	1.0404	0.1156	0.1274	
3	$\hat{\lambda}$	1.0501	0.2142	0.2096	1.0324	0.1478	0.1531	1.0138	0.1011	0.1031		
	$\hat{\sigma}$	1.0414	0.2118	0.2181	1.0255	0.1468	0.1573	1.0130	0.0999	0.1032		
	\hat{q}	4.5624	3.8165	3.2397	3.7511	1.7692	2.0399	3.2569	0.7691	0.9233		

two for σ : 0.5 and 1.0; two for q : 1 and 3; and three for the sample size: $n = 50$, $n = 100$ and $n = 200$. Each combination of parameters and sample size was replicated 1000 times. We present the mean of the obtained estimators, the mean of the standard deviations calculated based on the observed information matrix and the root mean square error. Note that the bias of the estimators is acceptable and decreases as the sample size increases. Additionally, when the sample size increases, the mean of the estimated deviations approximates the term $\sqrt{\text{MSE}}$, suggesting consistent estimators.

Table 2: Simulation for the NCSt distribution with $\nu = 5$ degrees of freedom.

true values			$\hat{\theta}$	$n = 50$			$n = 100$			$n = 200$		
λ	σ	q		mean	s.e.	$\sqrt{\text{MSE}}$	mean	s.e.	$\sqrt{\text{MSE}}$	mean	s.e.	$\sqrt{\text{MSE}}$
-0.5	0.5	1	$\hat{\lambda}$	-0.5299	0.1315	0.1346	-0.5174	0.0906	0.0894	-0.5121	0.0632	0.0601
			$\hat{\sigma}$	0.5361	0.1395	0.1441	0.5207	0.0953	0.0946	0.5191	0.0665	0.0620
			\hat{q}	1.1435	0.2982	0.3739	1.0772	0.1830	0.2004	1.0560	0.1239	0.1254
	3	$\hat{\lambda}$	-0.5174	0.1245	0.1151	-0.5128	0.0872	0.0849	-0.5091	0.0599	0.0603	
		$\hat{\sigma}$	0.5196	0.1336	0.1200	0.5149	0.0946	0.0902	0.5086	0.0651	0.0656	
		\hat{q}	4.7202	5.7034	3.4817	4.1062	3.3006	2.6697	3.4869	1.4701	1.5523	
1.0	1	$\hat{\lambda}$	-0.5331	0.2151	0.2202	-0.5146	0.1443	0.1431	-0.5173	0.1017	0.1055	
		$\hat{\sigma}$	1.0902	0.2921	0.3318	1.0422	0.1947	0.1942	1.0312	0.1363	0.1294	
		\hat{q}	1.1882	0.4338	0.6419	1.0734	0.1946	0.2034	1.0517	0.1320	0.1383	
3	$\hat{\lambda}$	-0.5181	0.2047	0.1882	-0.5214	0.1453	0.1395	-0.5163	0.1008	0.1009		
	$\hat{\sigma}$	1.0402	0.2914	0.2394	1.0381	0.2103	0.1863	1.0252	0.1457	0.1435		
	\hat{q}	5.0573	8.1198	3.8670	4.2968	4.4232	2.9118	3.7668	2.3982	2.1330		
0.5	0.5	1	$\hat{\lambda}$	0.5302	0.1315	0.1326	0.5180	0.0898	0.0904	0.5146	0.0632	0.0608
			$\hat{\sigma}$	0.5426	0.1403	0.1523	0.5208	0.0945	0.0893	0.5174	0.0660	0.0619
			\hat{q}	1.1508	0.3073	0.3942	1.0856	0.1845	0.2021	1.0604	0.1246	0.1299
	3	$\hat{\lambda}$	0.5195	0.1260	0.1164	0.5142	0.0882	0.0832	0.5108	0.0608	0.0609	
		$\hat{\sigma}$	0.5219	0.1358	0.1246	0.5165	0.0962	0.0910	0.5113	0.0661	0.0672	
		\hat{q}	4.8069	6.2872	3.5639	4.0846	3.3858	2.5787	3.5687	1.7137	1.7158	
1.0	1	$\hat{\lambda}$	0.5380	0.2131	0.2281	0.5261	0.1464	0.1445	0.5123	0.1011	0.0956	
		$\hat{\sigma}$	1.0749	0.2865	0.3011	1.0518	0.1959	0.2048	1.0345	0.1359	0.1298	
		\hat{q}	1.1655	0.3513	0.4978	1.0810	0.1974	0.2185	1.0543	0.1324	0.1372	
3	$\hat{\lambda}$	0.5269	0.2062	0.1976	0.5206	0.1462	0.1400	0.5160	0.1004	0.0992		
	$\hat{\sigma}$	1.0444	0.2898	0.2434	1.0442	0.2135	0.1956	1.0208	0.1455	0.1376		
	\hat{q}	4.9896	7.5749	3.8015	4.3185	4.5664	2.9438	3.7512	2.3581	2.0826		
1.0	0.5	1	$\hat{\lambda}$	1.0530	0.1734	0.1826	1.0363	0.1197	0.1211	1.0298	0.0855	0.0804
			$\hat{\sigma}$	0.5341	0.1403	0.1475	0.5235	0.0952	0.0956	0.5187	0.0676	0.0653
			\hat{q}	1.1284	0.2458	0.3075	1.0859	0.1590	0.1768	1.0677	0.1109	0.1215
	3	$\hat{\lambda}$	1.0383	0.1514	0.1523	1.0149	0.1026	0.1044	1.0084	0.0698	0.0718	
		$\hat{\sigma}$	0.5177	0.1227	0.1225	0.5076	0.0838	0.0851	0.5061	0.0574	0.0583	
		\hat{q}	4.1493	3.1966	2.6803	3.4830	1.4792	1.5935	3.2059	0.6775	0.7802	
1.0	1	$\hat{\lambda}$	1.0635	0.2631	0.2795	1.0414	0.1807	0.1830	1.0265	0.1257	0.1239	
		$\hat{\sigma}$	1.0739	0.2784	0.3209	1.0503	0.1903	0.1866	1.0335	0.1317	0.1249	
		\hat{q}	1.1522	0.3307	0.5153	1.0766	0.1817	0.1832	1.0615	0.1248	0.1327	
3	$\hat{\lambda}$	1.0436	0.2494	0.2362	1.0352	0.1751	0.1659	1.0115	0.1193	0.1195		
	$\hat{\sigma}$	1.0400	0.2683	0.2404	1.0316	0.1897	0.1813	1.0202	0.1299	0.1308		
	\hat{q}	4.7107	5.8531	3.4608	4.0555	3.2062	2.5689	3.5171	1.5251	1.6419		

4.2. Applications to real data

In this section, we use two real datasets to show the flexibility and applicability of the proposed NCSEL model. In these applications, we present analyses of the two real datasets to show the flexibility and applicability of the proposed NCSEL model by illustrating the fit of the proposed model and the use of the proposed EM-algorithm. We compare the results of these fits with other models that have been used. All the computations were done using the R package.

4.2.1. Nickel dataset

In this application, we consider a dataset consisting details regarding of Nickel (Ni) concentrations in 86 soil samples analyzed at the Mining Department of the University of Atacama, Chile. We report the ML estimates obtained under other models such as the Epsilon Skew-Normal (ESN) distribution (Mudholkar and Hutson ([25])) and Skew-Normal (SN) distribution (Azzalini ([4])), and compare them with our NCS model. A descriptive summary of this dataset is displayed in Table 3 where b_1 and b_2 are sample asymmetry and kurtosis coefficients, respectively.

Table 3: Nickel data: Descriptive summary of the mineral data.

n	\bar{X}	S	b_1	b_2
86	21.3372	16.6392	2.4483	12.0429

We observe that the data have positive asymmetry and high kurtosis. For this dataset, the NCS model moment estimators are given by $\hat{\lambda}_M = 15.340$, $\hat{\sigma}_M = 9.234$ and $\hat{q}_M = 3.558$, which were used as initial values to start the EM algorithm. The ML estimates of the parameters of the ESN, SN and NCS models are found in Table 4. The AIC values Akaike ([1]) are given in Table 4. The model that provides the best fit for these data is the NCS model, which is supported by results in Figure 4 and the Q-Q plot in Figure 5.

Table 4: Nickel data: ML estimates and corresponding standard error (SE) for ESN, SN and NCS models.

Parameter	ESN	SN	NCS
μ	4.006 (1.249)	2.626 (2.066)	
σ	13.398 (1.022)	24.975 (2.454)	5.329 (0.735)
q			2.190 (0.398)
λ		10.259 (9.603)	12.030 (1.044)
ϵ	-0.856 (0.057)		
AIC	696.419	695.523	680.363

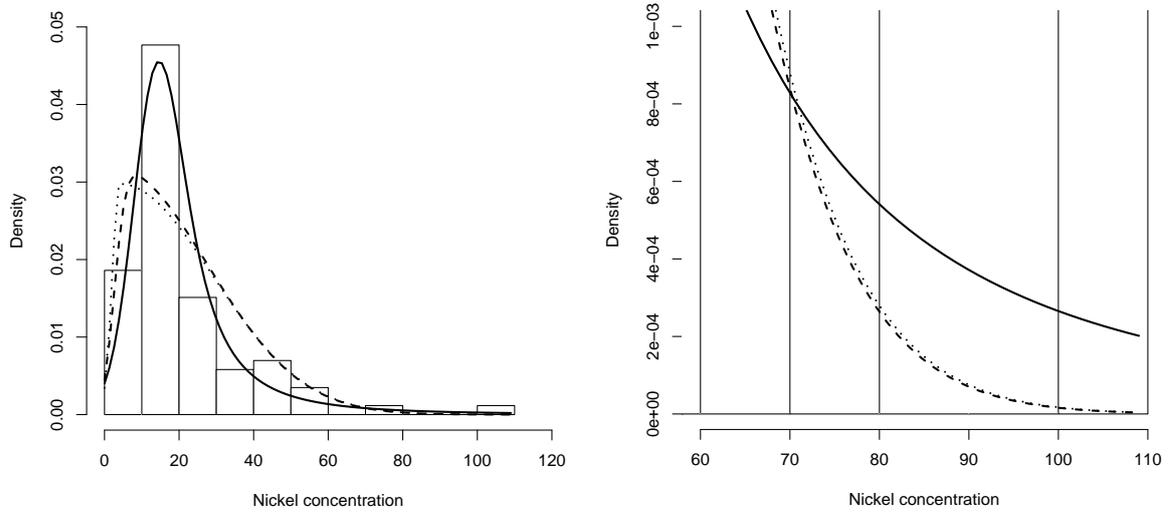


Figure 4: Nickel data: Fitted models, ESN (dotted line), estimated SN (dashed line) and estimated NCS (solid line) (Left panel). Upper tail of histogram with estimated ESN (dotted line), estimated SN (dashed line) and estimated NCS (solid line) (Right panel).

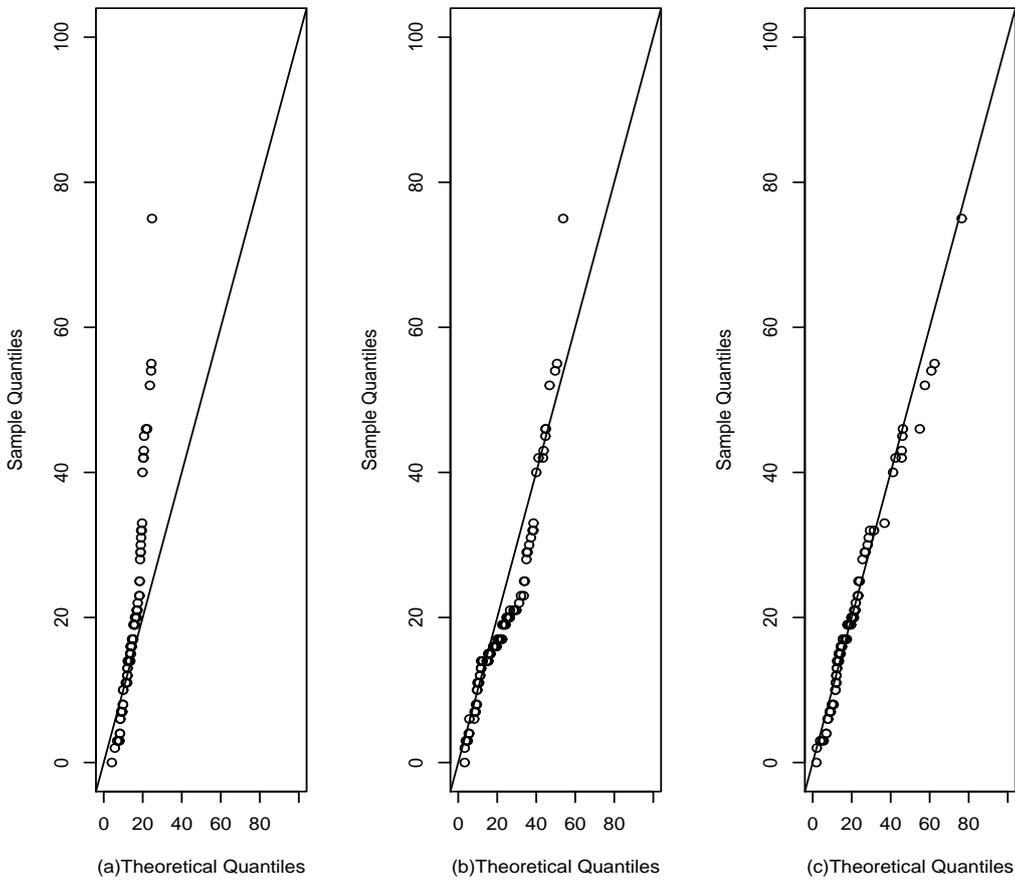


Figure 5: Nickel data: Q-Q plots; ESN model (a), SN model (b) and NCS model (c).

4.2.2. Copper data

This dataset refers to the soluble copper concentration of 1933 samples (Fuentes ([11])), for which the ML estimates are obtained for the Epsilon Skew-*t* (ESt) model (Gómez *et al.* ([14])), the Skew-*t* (St) model and for our NCSt model. A descriptive summary of this dataset is reported in Table 5. For this dataset, we observe positive asymmetry and kurtosis coefficients.

Table 5: Copper data: Descriptive statistics of the dataset.

n	\bar{y}	S	b_1	b_2
1933	0.591	0.302	1.196	4.633

Moreover, the moment estimates under the NCS model are given by $\hat{\lambda}_M = 0.441$, $\hat{\sigma}_M = 0.150$ and $\hat{q}_M = 3.950$, which were used as initial values to start the EM algorithm. Table 6 reports the estimates of the degrees of freedom, ν , for each model based on the Student-*t* distribution, which are obtained by maximizing the profile log-likelihood function, as in Vilca *et al.* ([30]). The estimates of ν is obtained for the ESt, St and NCSt models, as reported in Table 7. This table also includes the AIC values, revealing that the NCSt model fits the data well.

Table 6: Copper data: Estimation of ν for the St, ESt and NCSt models by maximizing the log-likelihood function.

ν	Log-likelihood	Log-likelihood	Log-likelihood
	St	ESt	NCSt
1	-359.599	-416.062	-709.297
2	-327.326	-266.335	-328.671
3	-209.663	-227.596	-223.877
4	-197.641	-213.163	-192.580
5	-191.688	-206.886	-188.485
6	-188.635	-203.988	-185.151
7	-187.067	-202.669	-189.139
8	-186.303	-202.151	-189.365
9	-185.992	-202.068	-190.621
10	-185.941	-202.195	-192.370
11	-186.042	-202.455	-193.973

Table 7: Copper data: ML estimates and the corresponding SE (in parentheses) for the St ($\nu = 10$), ESt ($\nu = 9$) and NCSt ($\nu = 6$) models.

Parameter	St	ESt	NCSt
μ	0.253 (0.008)	0.351 (0.013)	—
σ	0.404 (0.010)	0.2396 (0.004)	0.1355 (0.022)
q			5.991 (0.168)
λ	4.262 (0.360)		3.162 (0.475)
ϵ		-0.563 (0.029)	
AIC	377.881	410.115	376.301

Moreover, we present other results to show the performance of our approach. Figure 6 depicts plots of the fitted St, ESt and NCSt models using the ML estimates. We note that the fitted NCSt model presents heavier tails than the other models. Figure 7 shows the Q-Q plots for these fitted models. From all these summaries and plots, we can conclude that the NCSt model provides the best fit to the data.

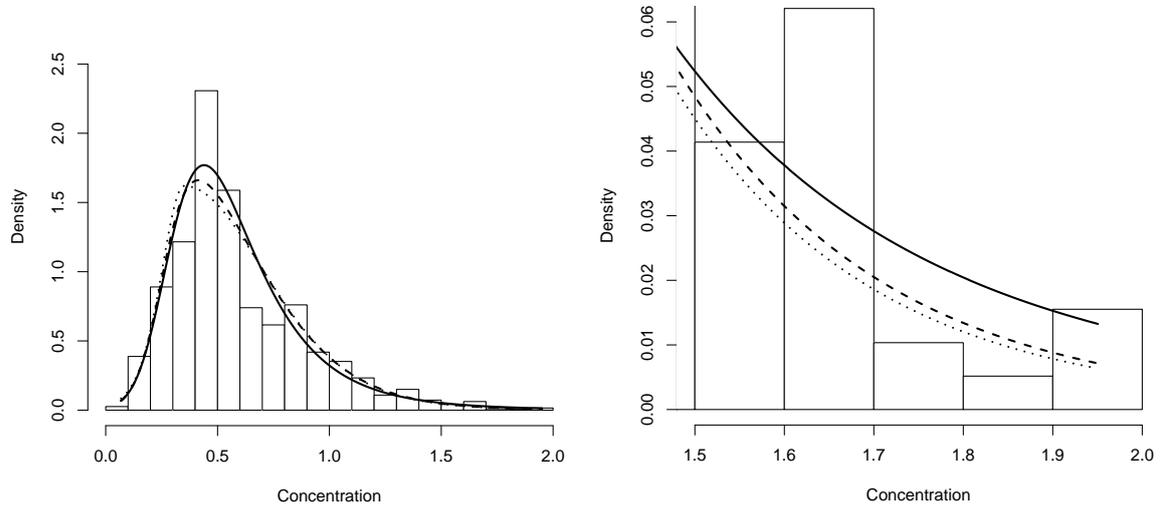


Figure 6: Copper data: Fitted models, NCSt (solid line), St (dashed line) and ESt (dotted line) (Left panel). Plots of the tails for the models (Right panel).

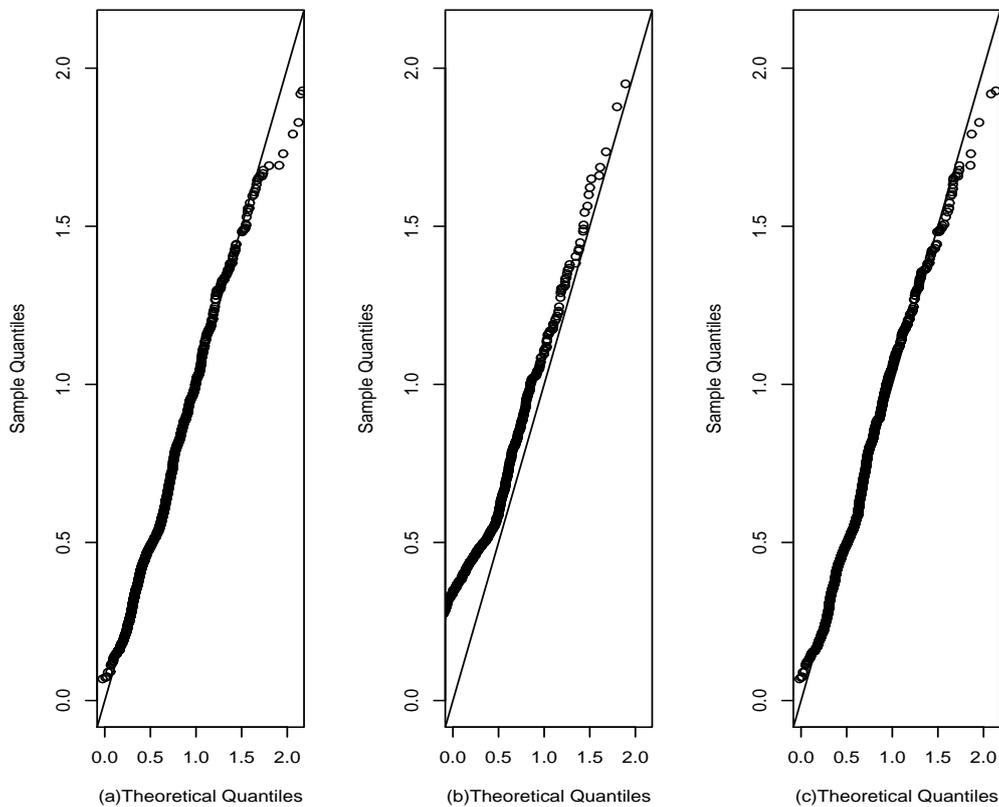


Figure 7: Copper data: Q-Q plots; St model (a), ESt model (b) and NCSt model (c).

4.2.3. Snack data

We consider in this application part of the data of an experiment performed in the Department of Nutrition of the Faculty of Public Health of the University of São Paulo, in which 5 different forms of a new type of snack, with low content of saturated fat and fatty acids, were compared over the course of 20 weeks. In this new product the hydrogenated vegetable fat has been replaced, in whole or in part, by canola oil. The forms are as follow: A (22% of fat 0% of canola oil), B (0% fat, 22% canola oil), C (17% fat, 5% canola oil), D (11% fat, 11% canola oil) and E (5% fat, 17% canola oil). The experiment was conducted so that in the even weeks 15 packs of each of the products A, B, C, D and E were analyzed in the laboratory and several variables were observed. In particular, we study the texture behavior of the products through the force necessary for shear (y). For more details on the study, see Paula ([26]), Section 2.8.1. The equation is

$$y_i = \beta_0 + \beta_1 x_{iB} + \beta_2 x_{iC} + \beta_3 x_{iD} + \beta_4 x_{iD} + \beta_5 x_{iE} + \beta_6 weeks_i + \varepsilon_i, \quad i = 1, \dots, n,$$

where $x_{iT} = 1$ if measurement i corresponds to a snack of type T , for $T = B, D, C, E$, and $weeks_i$ is the number of weeks that passed until measurement i was made.

We assume that $\varepsilon_i \sim \text{NCS}(\sigma, q, \lambda)$, where $\lambda = -\beta_0(q - 1)/q$, with $q > 1$. This condition is to obtain that $E(\varepsilon_i) = 0$, $i = 1, \dots, n$, with the purpose of comparing the fit under $\varepsilon_i \sim \text{ESN}(\sigma, \epsilon, \mu_1)$ and $\text{SN}(\sigma, \lambda, \mu_2)$ distributions. We also consider appropriate restrictions such as $\mu_1 = g(\sigma, \epsilon, \beta_0)$ and $\mu_2 = g(\sigma, \lambda, \beta_0)$, in order to obtain that $E(\varepsilon_i) = 0$, $i = 1, \dots, n$.

Results of the fit of the models are reported in Table 8. Note that, according to the AIC criterion, the best fit is provided by the NCS regression model. This is confirmed by the randomized quantile residuals, see Dunn and Smyth ([8]). If the model is correctly specified for the data, such residuals should be a random sample from the standard normal distribution. Figure 8 confirms that the NCS regression model provides a better fit than the ESN and SN regression models.

Table 8: Snack data: ML estimates and corresponding standard errors (SE) for ESN, SN and NCS regression models.

Parameter	ESN		SN		NCS	
	Estimate	SE	Estimate	SE	Estimate	SE
β_0	58.044	2.095	57.958	46.910	56.483	1.512
β_1	-10.907	1.755	-10.907	1.680	-8.167	1.626
β_2	-4.569	1.68	-4.569	1.680	-4.762	1.634
β_3	-15.174	1.84	-15.174	1.680	-11.708	1.632
β_4	-15.945	1.858	-15.944	1.680	-12.624	1.627
β_5	0.742	0.094	0.742	0.092	0.713	0.082
σ	14.550	0.399	14.551	0.379	9.233	0.527
ϵ	0.000	0.054	—	—	—	—
λ	—	—	0.007	4.038	—	—
q	—	—	—	—	6.444	0.614
AIC	6160.839		6160.839		6083.816	

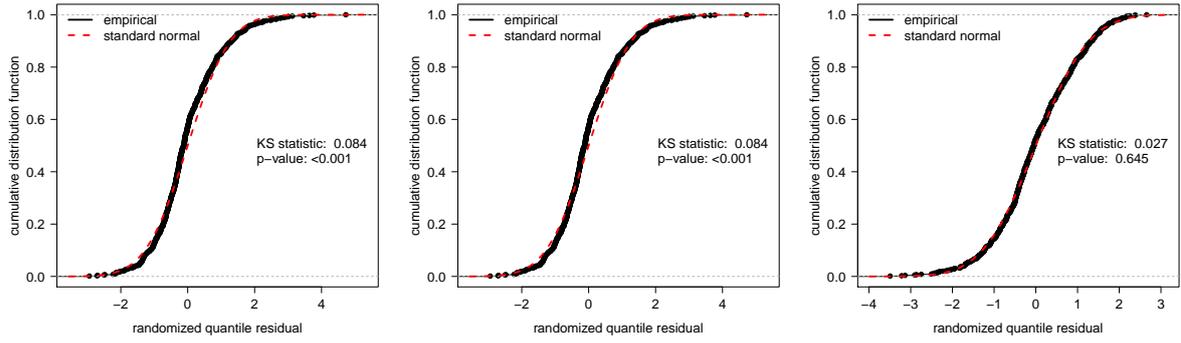


Figure 8: Snack data: Empirical cdf for randomized quantile residual versus cdf of standard normal distribution for ESN, SN and NCS regression models. Also provided are the statistics and p -values for the Kolmogorov–Smirnov (KS) test to compare both curves.

5. MULTIVARIATE NCSEL DISTRIBUTIONS

In this section the multivariate NCSEL distribution is introduced, its pdf is derived and some additional properties are studied.

In the multivariate setup, a k -dimensional random vector $\mathbf{Y} = (Y_1, \dots, Y_k)^\top$ follows an EL distribution with location parameter vector $\boldsymbol{\mu}$ and scale parameter matrix $\boldsymbol{\Sigma}$, which is positive definite, if its pdf is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1/2} g\left((\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right), \quad \mathbf{y} \in \mathbb{R}^k,$$

where g is the density generator function satisfying

$$\int_0^\infty u^{k-1} g(u^2) du < \infty.$$

We use the notation $\mathbf{Y} \sim \text{EL}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$. If the moments of each element of the random vector \mathbf{Y} are finite, then it follows that $E(\mathbf{Y}) = \boldsymbol{\mu}$ and $\text{Var}(\mathbf{Y}) = \alpha_g \boldsymbol{\Sigma}$, where α_g is a positive constant, as seen for example, in Fang *et al.* ([10]). Now a multivariate NCSEL distribution is proposed, where a k -variate vector \mathbf{Y} is said to have a multivariate noncentral slash-elliptical (MNCSEL) distribution with scale matrix $\boldsymbol{\Sigma}$ positive definite, $\boldsymbol{\lambda}$ being the non-centrality parameter and q the kurtosis parameter

$$(5.1) \quad \mathbf{Y} = \frac{\boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{X} + \boldsymbol{\lambda}}{U^{\frac{1}{q}}},$$

where $\mathbf{X} \sim \text{EL}_k(\mathbf{0}, \mathbf{I}_k; g)$ is independent of $U \sim U(0, 1)$. The resulting distribution is denoted by $\mathbf{Y} \sim \text{MNCSEL}_k(\boldsymbol{\Sigma}, q, \boldsymbol{\lambda}; g)$. The pdf of \mathbf{Y} is presented in the following result.

Proposition 5.1. *Let $\mathbf{Y} \sim \text{MNCSEL}_k(\boldsymbol{\Sigma}, q, \boldsymbol{\lambda}; g)$. Then, the pdf of \mathbf{Y} is given by*

$$(5.2) \quad f(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \int_0^1 z^{\frac{k}{q}} g\left[(\mathbf{y} z^{\frac{1}{q}} - \boldsymbol{\lambda})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} z^{\frac{1}{q}} - \boldsymbol{\lambda})\right] dz.$$

Proof: Using the fact that \mathbf{X} and U are independent and standard calculations of the Jacobian transformation of

$$\mathbf{Y} = \frac{\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{X} + \boldsymbol{\lambda}}{U^{\frac{1}{q}}} \quad \text{and} \quad Z = U,$$

we obtain the joint pdf of \mathbf{Y} and Z given by

$$f_{\mathbf{Y},Z}(\mathbf{y}, z) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} z^{\frac{k}{q}} g \left[(\mathbf{y} z^{\frac{1}{q}} - \boldsymbol{\lambda})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} z^{\frac{1}{q}} - \boldsymbol{\lambda}) \right].$$

The required result is obtained by integrating the above joint pdf with respect to z . \square

Remark 5.1. If $\boldsymbol{\lambda} = 0$, we obtain the family of distributions discussed by Gómez *et al.* ([13]) and Gómez and Venegas ([15]). On the other hand for $\boldsymbol{\lambda} = 0$ and under normality of \mathbf{X} , we obtain the slash multivariate introduced by Wang and Genton ([31]).

Another important property is that the MNCSEL distribution can be written as a scale mixture of an elliptical distribution and a uniform distribution in the unit interval.

Proposition 5.2. Let $\mathbf{Y} | (Z = z) \sim \text{EL}_k(z^{-\frac{1}{q}}\boldsymbol{\lambda}, z^{-\frac{2}{q}}\boldsymbol{\Sigma}; g)$ and $Z \sim \text{U}(0, 1)$. Then $\mathbf{Y} \sim \text{MNCSEL}_k(\boldsymbol{\Sigma}, q, \boldsymbol{\lambda}; g)$.

Proof: We can write

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \int_0^1 f_{\mathbf{Y}|Z}(\mathbf{y}) f_Z(z) dz \\ &= \int_0^1 |z^{-\frac{2}{q}}\boldsymbol{\Sigma}|^{-1/2} g \left[(\mathbf{y} - z^{-\frac{1}{q}}\boldsymbol{\lambda})^\top (z^{-\frac{2}{q}}\boldsymbol{\Sigma})^{-1} (\mathbf{y} - z^{-\frac{1}{q}}\boldsymbol{\lambda}) \right] dz. \end{aligned}$$

The result follows using properties of determinants. \square

Proposition 5.3. Let $\mathbf{Y} \sim \text{MNCSEL}_k(\boldsymbol{\Sigma}, q, \boldsymbol{\lambda}; g)$. Then,

$$\mathbb{E}[\mathbf{Y}] = \frac{q\boldsymbol{\lambda}}{q-1}, \quad q > 1, \quad \text{and} \quad \text{Var}(\mathbf{Y}) = \frac{q}{q-2} \left(\frac{\boldsymbol{\lambda}\boldsymbol{\lambda}^\top}{(q-1)^2} + \alpha_g \boldsymbol{\Sigma} \right), \quad q > 2.$$

Proof: Following the procedure in Proposition 5.2, we have $\mathbf{Y} | (Z = z) \sim \text{EL}_k(z^{-\frac{1}{q}}\boldsymbol{\lambda}, z^{-\frac{2}{q}}\boldsymbol{\Sigma})$. So, using the fact that $\mathbb{E}[Z^{-\frac{r}{q}}] = \frac{q}{q-r}$, $q > r$ and the conditional expectation properties:

$$\mathbb{E}[\mathbf{Y}] = \mathbb{E}[\mathbb{E}(\mathbf{Y}|Z)] = \mathbb{E}[Z^{-\frac{1}{q}}\boldsymbol{\lambda}] = \frac{q\boldsymbol{\lambda}}{q-1}, \quad q > 1.$$

Moreover, following the same idea we obtain the variance of \mathbf{Y} as follows:

$$\begin{aligned} \text{Var}(\mathbf{Y}) &= \text{Var}[\mathbb{E}(\mathbf{Y}|Z)] + \mathbb{E}[\text{Var}(\mathbf{Y}|Z)] \\ &= \text{Var}[Z^{-\frac{1}{q}}\boldsymbol{\lambda}] + \mathbb{E}[Z^{-\frac{2}{q}}\alpha_g\boldsymbol{\Sigma}] \\ &= \boldsymbol{\lambda} \text{Var}[Z^{-\frac{1}{q}}] \boldsymbol{\lambda}^\top + \alpha_g \mathbb{E}[Z^{-\frac{2}{q}}] \boldsymbol{\Sigma} \\ &= \boldsymbol{\lambda}\boldsymbol{\lambda}^\top \frac{q}{(q-2)(q-1)^2} + \alpha_g \boldsymbol{\Sigma} \frac{q}{q-2}, \quad q > 2 \\ &= \frac{q}{q-2} \left(\frac{\boldsymbol{\lambda}\boldsymbol{\lambda}^\top}{(q-1)^2} + \alpha_g \boldsymbol{\Sigma} \right), \quad q > 2. \end{aligned} \quad \square$$

6. CONCLUSION

Here we have introduced a new distribution called the NCSEL distribution. The main idea is to incorporate a non-centrality parameter in the usual SEL distribution. The resulting distribution is an asymmetric distribution that contains as special cases the EL and SEL distributions. For this family of distributions we point out some important characteristics and properties that allow us to obtain qualitatively robust ML estimates and efficiently compute them by using the EM-algorithm for a special class based on the family of NI distributions. We illustrate our results by using three numerical examples. They show the flexibility and inherent robustness of the estimation procedure in the NCSEL model.

Finally, the NCSEL can be used along the same lines as the skew distributions in the context of regression. This issue is currently under investigation, and we hope to report these findings in a future paper.

A. APPENDIX — Simulation study with $\nu = 10$ degrees of freedom

Table 9: Simulation for the NCSt distribution with $\nu = 10$ degrees of freedom.

true values			$\hat{\theta}$	$n = 50$			$n = 100$			$n = 200$		
λ	σ	q		mean	s.e.	$\sqrt{\text{MSE}}$	mean	s.e.	$\sqrt{\text{MSE}}$	mean	s.e.	$\sqrt{\text{MSE}}$
-0.5	0.5	1	$\hat{\lambda}$	-0.5237	0.1261	0.1303	-0.5150	0.0874	0.0862	-0.5085	0.0606	0.0594
			$\hat{\sigma}$	0.5286	0.1306	0.1324	0.5209	0.0901	0.0889	0.5125	0.0631	0.0603
			\hat{q}	1.0905	0.2628	0.2901	1.0666	0.1759	0.1848	1.0399	0.1178	0.1219
	0.5	3	$\hat{\lambda}$	-0.5195	0.1147	0.1113	-0.5132	0.0796	0.0803	-0.5078	0.0549	0.0556
			$\hat{\sigma}$	0.5171	0.1188	0.1110	0.5122	0.0825	0.0826	0.5100	0.0568	0.0611
			\hat{q}	4.4553	4.3511	3.1883	3.8335	2.2236	2.2079	3.4520	1.1176	1.3898
1.0	1	$\hat{\lambda}$	-0.5300	0.2026	0.2120	-0.5131	0.1377	0.1423	-0.5142	0.0974	0.0966	
		$\hat{\sigma}$	1.0666	0.2710	0.2854	1.0309	0.1827	0.1846	1.0346	0.129	0.1309	
		\hat{q}	1.1467	0.3292	0.4879	1.0635	0.1852	0.1984	1.0458	0.126	0.1296	
1.0	3	$\hat{\lambda}$	-0.5410	0.1955	0.1905	-0.5144	0.1340	0.1342	-0.5102	0.0924	0.0944	
		$\hat{\sigma}$	1.0590	0.2610	0.2388	1.0393	0.1832	0.1805	1.0285	0.1242	0.1291	
		\hat{q}	5.1102	6.5417	3.8243	4.2180	3.4874	2.8081	3.6203	1.5291	1.7002	
0.5	0.5	1	$\hat{\lambda}$	0.5294	0.1266	0.1269	0.5181	0.0871	0.0873	0.5099	0.0612	0.0581
			$\hat{\sigma}$	0.5319	0.1303	0.1420	0.5224	0.0895	0.0897	0.5146	0.0628	0.0603
			\hat{q}	1.1341	0.3178	0.4400	1.0661	0.1738	0.1850	1.0458	0.1194	0.1248
	0.5	3	$\hat{\lambda}$	0.5292	0.1171	0.1142	0.5147	0.0800	0.0802	0.5058	0.0543	0.053
			$\hat{\sigma}$	0.5273	0.1206	0.1170	0.5170	0.0828	0.0811	0.5061	0.056	0.0577
			\hat{q}	4.8062	4.9793	3.4829	3.8846	2.3178	2.2996	3.3449	0.9999	1.213
1.0	1	$\hat{\lambda}$	0.5249	0.2025	0.2049	0.5109	0.1383	0.1383	0.5146	0.0978	0.0995	
		$\hat{\sigma}$	1.0680	0.2701	0.2850	1.0363	0.1849	0.1837	1.0337	0.1295	0.1285	
		\hat{q}	1.1345	0.3277	0.4665	1.0620	0.1851	0.1953	1.0371	0.1244	0.1284	
1.0	3	$\hat{\lambda}$	0.5214	0.1928	0.1852	0.5266	0.1347	0.1325	0.5096	0.0917	0.094	
		$\hat{\sigma}$	1.0617	0.2578	0.2414	1.0418	0.1835	0.1784	1.0192	0.1231	0.1261	
		\hat{q}	5.1212	6.3880	3.8663	4.2130	3.3634	2.7218	3.5264	1.4173	1.5681	
1.0	0.5	1	$\hat{\lambda}$	1.0358	0.1656	0.1707	1.0212	0.1147	0.1091	1.0168	0.0801	0.0757
			$\hat{\sigma}$	0.5075	0.1261	0.1363	0.5079	0.0870	0.0862	0.5106	0.0607	0.0601
			\hat{q}	1.0947	0.2264	0.2533	1.0569	0.1496	0.1536	1.0458	0.1031	0.1122
	0.5	3	$\hat{\lambda}$	1.0376	0.1399	0.1438	1.0137	0.0952	0.0996	1.0064	0.0653	0.0656
			$\hat{\sigma}$	0.5153	0.1090	0.1133	0.5087	0.0747	0.0791	0.5028	0.051	0.0507
			\hat{q}	4.0107	2.4780	2.4694	3.4097	1.0933	1.3396	3.1600	0.5736	0.6723
1.0	1	$\hat{\lambda}$	1.0500	0.2525	0.2634	1.0242	0.1724	0.1668	1.0138	0.1207	0.1197	
		$\hat{\sigma}$	1.0579	0.2605	0.2797	1.0320	0.1775	0.1751	1.0235	0.1243	0.118	
		\hat{q}	1.1121	0.2715	0.3156	1.0594	0.1729	0.1839	1.0375	0.1172	0.1208	
1.0	3	$\hat{\lambda}$	1.0392	0.2312	0.2183	1.0212	0.1581	0.1585	1.0148	0.1085	0.1091	
		$\hat{\sigma}$	1.0405	0.2388	0.2193	1.0229	0.1638	0.1628	1.0136	0.1116	0.1131	
		\hat{q}	4.5632	4.5408	3.2450	3.8152	2.1794	2.1591	3.3253	0.9405	1.0617	

ACKNOWLEDGMENTS

The research of J. Reyes and H.W. Gómez was supported by Grant SEMILLERO UA-2017.

REFERENCES

- [1] AKAIKE, H. (1974). A new look at statistical model identification, *IEEE Transaction on Automatic Control*, **19**, 716–723.
- [2] ANDREWS, D.F. and MALLOWS, C.L. (1974). Scale mixtures of normal distributions, *Journal of the Royal Statistical Society B*, **36**, 99–102.
- [3] ARSLAN, O. and GENÇ, A.I. (2009). A generalization of the multivariate slash distribution, *Journal of Statistical Planning and Inference*, **139**(3), 1164–1170.
- [4] AZZALINI, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171–178.
- [5] BARNDORFF-NIELSEN, O.E. (1997). Normal inverse Gaussian distributions and stochastic volatility modeling, *Scandinavian Journal of Statistics*, **24**, 1–13.
- [6] CAMBANIS, S.; HUANG, S. and SIMONS, G. (1981). On the theory of elliptically contoured distributions, *J. Multivar. Anal.*, **11**, 365–385.
- [7] DASGUPTA, S. and LAHIRI, K. (1992). A comparative study of alternative methods of quantifying qualitative survey responses using NAPM data, *Journal of Business and Economic Statistics*, **10**, 391–400.
- [8] DUNN, P.K. and SMYTH, G.K. (1996). Randomized quantile residuals, *Journal of Computational and Graphical Statistics*, **5**, 236–244.
- [9] FAMA, E.F. and FRENCH, K.R. (1993). Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics*, **33**, 3–56.
- [10] FANG, K.T.; KOTZ, S. and NG, K.W. (1990). *Symmetric Multivariate and Related Distribution*, Chapman and Hall, London.
- [11] FUENTES, E. (2017). Recognition and Estimation of Resources for the Geological Unit Vetiform By Probing (RC) in the Tunnel of Exploration Underground, in the *Minera la Verdosa*, Valparaíso. Unpublished Thesis, Antofagasta University, Chile (in Spanish).
- [12] GENÇ, A.I. (2007). A generalization of the univariate slash by a scale-mixture exponential power distribution, *Communications in Statistics – Simulation and Computation*, **36**(5), 937–947.
- [13] GÓMEZ, H.W.; QUINTANA, F.A. and TORRES, F.J. (2007). A new family of slash-distributions with elliptical contours, *Statistics and Probability Letters*, **77**(7), 717–725.
- [14] GÓMEZ, H.W.; TORRES, F.J. and BOLFARINE, H. (2007). Large-sample inference for the epsilon-skew- t distribution, *Communications in Statistics – Theory and Methods*, **36**(1), 73–81.
- [15] GÓMEZ, H.W. and VENEGAS, O. (2008). Erratum to: A new family of slash-distributions with elliptical contours [Statist. Probab. Lett., **77** (2007), 717–725], *Statistics and Probability Letters*, **78**(14), 2273–2274.
- [16] JOHNSON, S.; KOTZ, S. and BALAKRISHNAN, N. (1995). *Continuous Univariate Distributions* (2nd Edn.), Wiley, New York.
- [17] KAFADAR, K. (1982). A biweight approach to the one-sample problem, *Journal of the American Statistical Association*, **77**(378), 416–424.
- [18] LAHIRI, K. and TEIGLAND, C. (1987). On the normality of probability distributions of inflation and GNP forecasts, *International Journal of Forecasting*, **3**, 269–279.
- [19] LANGE, K.L.; LITTLE, R.J.A. and TAYLOR, J.M.G. (1989). Robust statistical modeling using the t -distribution, *Journal of the American Statistical Association*, **84**, 881–896.
- [20] LANGE, K. and SINSHEIMER, J.S. (1993). Normal/independent distributions and their applications in robust regression, *Journal of Computational and Graphical Statistics*, **2**, 175–198.

- [21] LEE, S.Y. and XU, L. (2004). Influence analyses of nonlinear mixed-effects model, *Computational Statistics & Data Analysis*, **45**, 321–341.
- [22] LOUIS, T.A. (1982). Finding the observed information matrix when using the EM algorithm, *Journal of the Royal Statistical Society – Series B*, **44**(2), 226–233.
- [23] MEILIJON, I. (1989). A fast improvement to the EM algorithm on its own terms, *Journal of the Royal Statistical Society, Series B*, **51**(1), 127–138.
- [24] MOSTELLER, F. and TUKEY, J.W. (1977). *Data Analysis and Regression*, Addison-Wesley.
- [25] MUDHOLKAR, G.S. and HUTSON, A.D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data, *Journal Statistical Planning and Inference*, **83**, 291–309.
- [26] PAULA, G.A. (2013). *Modelos de Regressão com Apoio Computacional*, São Paulo, IME – USP.
- [27] REYES, J.; VILCA, F.; GALLARDO, D.I. and GÓMEZ, H.W. (2017). Modified slash Birnbaum–Saunders distribution, *Hacettepe Journal of Mathematics and Statistics*, **46**(5), 969–984.
- [28] ROGERS, W.H. and TUKEY, J.W. (1972). Understanding some long-tailed symmetrical distributions, *Statistica Neerlandica*, **26**(3), 211–226.
- [29] TSIONAS, E.G. (2002). Stochastic frontier models with random coefficients, *Journal of Applied Econometrics*, **17**, 127–147.
- [30] VILCA, F.L.; ZELLER, C.B. and CORDEIRO, G.M. (2015). The sinh-normal/independent nonlinear regression model, *Journal of Applied Statistics*, **42**, 1659–1676.
- [31] WANG, J. and GENTON, M.G. (2006). The multivariate skew-slash distribution, *Journal Statistical Planning and Inference*, **136**, 209–220.

THE USE OF CONTROL CHARTS TO MONITOR AIR PLANE ACCIDENTS OF THE HELLENIC AIR FORCE

- Authors: VASILEIOS ALEVIZAKOS
– Department of Mathematics, National Technical University of Athens,
Zografou, Athens, Greece
basalebiz@yahoo.gr
- CHRISTOS KOUKOUVINOS
– Department of Mathematics, National Technical University of Athens,
Zografou, Athens, Greece
ckoukouv@math.ntua.gr
- PETROS E. MARAVELAKIS
– Department of Business Administration, University of Piraeus,
Piraeus, Greece
maravel@unipi.gr

Received: March 2018

Revised: December 2018

Accepted: December 2018

Abstract:

- Accidents are unfortunate events that cause economic and/or human losses to individuals, organisations, companies or to the society. Air plane accidents cause both economic and human losses most of the times. In this paper we model the occurrences of air plane accidents using the Poisson distribution. We use control charts, which is the main tool of Statistical Process Control, to identify an out of control situation in the occurrence of air plane accidents. We propose the use of Shewhart and Exponentially Weighted Moving Average control charts and we apply them in real data from the Hellenic Air Force.

Keywords:

- *air plane accidents; Poisson; Shewhart; EWMA; air force.*

AMS Subject Classification:

- 62P30, 62-07.

1. INTRODUCTION

Air plane accidents are of major importance since they involve most of the times both human and economic losses. The last decades great effort has been imposed in the safety regulations in all the different aspects of commercial aviation. For example in a series of seven years (2010–2016) there was not any human loss in a crash on a United States-certificated scheduled airline operating anywhere according to official data.

In the case of military air forces things are a bit different. The continuous competitiveness of the air forces leads to the occurrence of air accidents. The accidents of air forces are not in the numbers of the previous decades but still they are a fact. However, in both commercial and military aviation few efforts have been made to monitor the air plane accidents.

Statistical Quality Control (SQC) is a well known collection of methods aiming to continuously improve the quality of a product or a process. Rockwell ([7]) initiated the use of statistical quality control techniques in the field of safety management. Specifically, Rockwell ([7]) dealt with the problem of safety performance measurement. The main tools of SQC methods that are used to monitor critical parameters of a process are the control charts.

The main objective of this paper is to demonstrate how we can use control charts to monitor the air plane accidents. To be more specific, in Section 2 we present the main points of the theory of control charts. We outline the Shewhart and Exponentially Weighted Moving Average (EWMA) Control Charts and the way they are used to monitor a process. In Section 3, we apply the techniques presented in Section 2 in real accident data from the Hellenic Air Force (HAF). Finally, in Section 4 we give some conclusions and guidelines for future research.

2. CONTROL CHARTS

One of the main objectives of a product or a process is to continuously improve its quality. This goal, in statistical terms, may be expressed as variability reduction. SQC is a popular collection of methods targeting at this purpose and control charts are known to be the main tools to detect shifts in a process. The most popular control charts are the Shewhart charts, the Cumulative Sum (CUSUM) charts and the Exponentially Weighted Moving Average charts (EWMA). Shewhart charts are used to detect large shifts in a process whereas CUSUM and EWMA charts have very good results for small to moderate shifts. Since the CUSUM and EWMA control charts have similar performance, in this paper we confine ourselves to the EWMA chart.

A control chart is a graphical representation of one or more characteristics of the process under investigation. It is the main tool to identify special causes of variability in a process. On the horizontal axis we plot the number of the sample drawn from the process or the time that the sample was inspected. On the vertical axis we plot the value of the characteristic or the characteristics measured for each sample or for the time of the horizontal axis. A straight line connects the successive points indicating the level of the characteristic in time or in successive samples. There are also three usually straight lines that stand for the upper control limit (UCL), the center line (CL) and the lower control limit (LCL).

We deduce that a process operates under control when the line connecting the sequence of points does not cross UCL or LCL. When a point plots outside these limits we conclude that the process is in an out-of-control state and corrective actions must be taken in order to remove the assignable cause that led to this problem.

In the literature, two distinct phases of control charting practice have been discussed (see, e.g. Woodall [12]). In Phase I, charts are used for retrospectively testing whether the process was in-control when the first subgroups were being drawn. In this phase, the charts are used as aids to the practitioner, in bringing a process into a state of statistical control. Once this is accomplished, the control chart is used to define what is meant by statistical control.

In Phase II, control charts are used to test if the process remains in-control when real time subgroups are drawn. In this phase, the control charts are used to monitor the process for a possible shift from the in-control state. The in-control characterization in this phase, is most of the times determined from the values of the process parameters. These values are usually estimated from historical data known to be under control. Usually these data are the ones from Phase I.

The design of a control chart must take into account two contradicting aims. The first one of them refers to the in-control state. In such a case, the control chart should signal (false alarm) as slow as possible. On the other hand, when a process is out-of-control the control chart must signal as soon as possible. The most popular measure to evaluate the performance of a chart concerning the previous two objectives is the average run length (ARL), which is based on the run length (RL) distribution. The number of observations when we plot individual data, or the number of samples when we plot data in subgroups, required for a control chart to signal is a run length (an observation of the RL distribution). The mean of the RL distribution is the ARL, and it can be defined as the average number of observations for a control chart to signal.

Since we deal with a parametric case of control charts we need to assume a distribution for the studied phenomenon. A detailed investigation is given in the following subsection.

2.1. Distribution of air plane accidents

A well known distribution used to model the occurrence of events in time is the Poisson distribution (Kjelln and Albrechtsen [2]). Assume that accidents occur at random points in time, let c be the average number of accidents per unit of time for example one year. Let x be the number of accidents occurring during t time periods. Then, the probability that x accidents will occur during t time periods is equal to

$$P(X=x) = \frac{(ct)^x}{x!} e^{-ct}, \quad x = 0, 1, 2, \dots$$

The control charts that will be presented in the following subsections assume that the air plane accidents are well modelled using the Poisson distribution.

2.2. The c chart

Assume that we want to monitor the number of accidents in a fixed time period and let $c > 0$ denote the parameter of the Poisson distribution for simplicity. If the true value of the parameter c is known, the Phase II three sigma control limits will be defined as:

$$\begin{aligned} \text{UCL} &= c + 3\sqrt{c}, \\ \text{CL} &= c, \\ \text{LCL} &= c - 3\sqrt{c}. \end{aligned}$$

If the computed value of LCL is less than zero, then we set $\text{LCL} = 0$.

When the true value of the parameter c is not known, then the average number of accidents in a preliminary sample (\bar{c}), is applied as an estimate of c . In this case, the Phase I control limits are defined as follows:

$$\begin{aligned} \text{UCL} &= \bar{c} + 3\sqrt{\bar{c}}, \\ \text{CL} &= \bar{c}, \\ \text{LCL} &= \bar{c} - 3\sqrt{\bar{c}}. \end{aligned}$$

The Phase I control limits are considered as trial control limits and the preliminary samples should be examined for lack of control. If there are observations that cross the estimated control limits due to common causes, usually these observations are excluded from the sample and the control limits are recalculated in the usual Phase I analysis (Montgomery [4]).

For the c chart, the probability of type I error (α) is calculated as

$$\begin{aligned} \alpha &= P\left(X \notin [\text{LCL}, \text{UCL}] \mid X \sim P(c)\right) \\ &= 1 - [F_x(\text{UCL}) - F_x(\text{LCL})] \\ &= 1 - \sum_{x=\lceil \text{LCL} \rceil}^{\lfloor \text{UCL} \rfloor} \frac{e^{-c} c^x}{x!} \end{aligned}$$

and the in-control ARL (ARL_0) is given by the formula

$$\text{ARL}_0 = \frac{1}{\alpha}.$$

The probability of type II error (β) is

$$\begin{aligned} \beta &= P\left(\text{LCL} \leq X \leq \text{UCL} \mid X \sim P(c^*)\right) \\ &= F_x(\text{UCL}) - F_x(\text{LCL}) \\ &= \sum_{x=\lceil \text{LCL} \rceil}^{\lfloor \text{UCL} \rfloor} \frac{e^{-c^*} c^{*x}}{x!}, \end{aligned}$$

where c^* is the average number of defects displayed in an inspection unit in an out of control process, $\lceil \text{LCL} \rceil$ denotes the smaller integer greater than or equal to LCL and $\lfloor \text{UCL} \rfloor$ denotes the largest integer less than or equal to UCL. The out-of-control ARL (ARL_1) is given by the formula

$$\text{ARL}_1 = \frac{1}{1 - \beta}.$$

We must note here that the same chart presented here can be used to monitor the number of nonconformities or defects in an inspection unit from a repetitive production process.

2.3. The ARL-unbiased c chart

The c chart with $3 - \sigma$ control limits has $\text{LCL} > 0$ if $c > 9$. In case $c \leq 9$, then $\text{LCL} < 0$ and as we mentioned before, we set it equal to zero and a downward shift of the process mean cannot be detected. Denoting as c_0 the in-control mean of the process, Paulino *et al.* [5] proved that for $c_0 > 9$, the ARL of a c chart with $3 - \sigma$ control limits takes its maximum value at the point

$$\delta^*(c_0) = \left[\frac{\text{UCL}!}{(\text{LCL} - 1)!} \right]^{\frac{1}{\text{UCL} - \text{LCL} + 1}} - c_0.$$

This means that the maximum of the ARL appears at a point $\delta^*(c_0)$ below the zero, i.e. some ARL_1 values that correspond to downward shifts are larger than the ARL_0 . In this case, we say that the chart is ARL-biased.

Many authors, such as Wetherill and Brown [11] and Ryan [8] used quantile-based control limits. In this case LCL and UCL are the largest and smallest non-negative integers, that satisfy

$$\begin{aligned} P(X < \text{LCL} \mid c = c_0) &\leq \alpha_{\text{LCL}}, \\ P(X > \text{UCL} \mid c = c_0) &\leq \alpha_{\text{UCL}}, \end{aligned}$$

where $\alpha_{\text{LCL}} + \alpha_{\text{UCL}} = \alpha$. Using the quantile-based control limits, we have $\text{ARL}_0 = 1/\alpha$.

Paulino *et al.* [5] proposed a c chart, named as ARL-unbiased c chart, with quantile-based control limits, that triggers a signal with probability one if the sample number of defects is below LCL or above UCL and probabilities γ_{LCL} and γ_{UCL} if the sample number of defects is equal to LCL and UCL, respectively. The values of probabilities γ_{LCL} and γ_{UCL} can be obtained by solving a system of linear equations. The solution of this system gives

$$(2.1) \quad \gamma_{\text{LCL}} = \frac{de - bf}{ad - bc},$$

$$(2.2) \quad \gamma_{\text{UCL}} = \frac{af - ce}{ad - bc},$$

where $a = P(X = \text{LCL} \mid c = c_0)$, $b = P(X = \text{UCL} \mid c = c_0)$, $c = \text{LCL} \cdot P(X = \text{LCL} \mid c = c_0)$, $d = \text{UCL} \cdot P(X = \text{UCL} \mid c = c_0)$, $e = \alpha - 1 + \sum_{x=\text{LCL}}^{\text{UCL}} P(X = x \mid c = c_0)$ and $f = \alpha \cdot c_0 - c_0 + \sum_{x=\text{LCL}}^{\text{UCL}} x \cdot P(X = x \mid c = c_0)$. A signal is triggered by the ARL-unbiased c chart with

probability

$$\xi(c^*) = \left[1 - \sum_{x=LCL}^{UCL} P(X=x | c=c^*) \right] + \gamma_{LCL} \cdot P(X=LCL | c=c^*) + \gamma_{UCL} \cdot P(X=UCL | c=c^*)$$

and $ARL_1 = 1/\xi(c^*)$.

Note that for the c chart, the probability of triggering a signal is equal to $\xi(c^*) = 1 - \sum_{x=LCL}^{UCL} P(X=x | c=c^*)$.

2.4. The classical Poisson EWMA control chart (PEWMA)

The EWMA control chart was introduced by Roberts [6]. Borror *et al.* [1] modified this chart to monitor Poisson data. Let X_1, X_2, \dots be i.i.d. Poisson random variables with mean c . When the process is in control, $c = c_0$. The EWMA statistics can be written as follows:

$$(2.3) \quad Z_t = \lambda X_t + (1 - \lambda)Z_{t-1}, \quad t = 1, 2, 3, \dots$$

where λ is the smoothing factor, $0 < \lambda \leq 1$ and the starting value is the process target, that is $Z_0 = c_0$. Values of λ in the interval $0.05 \leq \lambda \leq 0.25$ work well in practice, with $\lambda = 0.05$, $\lambda = 0.10$ and $\lambda = 0.20$ being popular choices (Montgomery [4]).

Using the abovementioned definition the mean value of Z_t is

$$E(Z_t) = c_0$$

and the variance of Z_t is

$$\text{Var}(Z_t) = \frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2t}] c_0.$$

Therefore, the PEWMA control chart is constructed by plotting Z_t versus the sample number i or time t . The center line and control limits for the PEWMA control chart are as follows:

$$(2.4) \quad \text{UCL} = c_0 + L \sqrt{\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2t}] c_0},$$

$$\text{CL} = c_0,$$

$$(2.5) \quad \text{LCL} = c_0 - L \sqrt{\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2t}] c_0},$$

where $L > 0$ can be chosen to provide a specified ARL_0 . If the computed value of LCL is less than zero, then we set $LCL = 0$. For large values of t , the control limits converge to the following values:

$$\text{UCL} = c_0 + L \sqrt{\frac{\lambda}{2 - \lambda} c_0},$$

$$\text{CL} = c_0,$$

$$\text{LCL} = c_0 - L \sqrt{\frac{\lambda}{2 - \lambda} c_0}.$$

It is recommended to use the exact control limits of Equations (2.4) and (2.5) for small values of λ (Montgomery [4]).

The PEWMA control chart raises an out-of-control signal when $Z_t < \text{LCL}$ or $Z_t > \text{UCL}$. The ARL values of the PEWMA chart are usually smaller than the ARLs for the c chart and the lower limit for the PEWMA is usually positive so that downward shifts in the process mean can be detected (Borror *et al.* [1]).

2.5. The Poisson Double EWMA (PDEWMA) control chart

Shamma and Shamma [9] developed a double EWMA control chart in an effort to increase the sensitivity of the EWMA control chart to detect small shifts and drifts in a process. Zhang *et al.* [13] extended the idea of the PEWMA chart to create the PDEWMA.

Let X_1, X_2, \dots be i.i.d. Poisson random variables with mean c . When the process is in control, $c = c_0$. The PDEWMA statistic can be written as follows:

$$(2.6) \quad \begin{aligned} Y_t &= \lambda X_t + (1 - \lambda)Y_{t-1}, \\ Z_t &= \lambda Y_t + (1 - \lambda)Z_{t-1}, \end{aligned}$$

where λ is the smoothing factor, $0 < \lambda \leq 1$ and $Y_0 = Z_0 = c_0$. It can be proved that the mean value of Z_t is

$$E(Z_t) = c_0$$

and the variance of Z_t is

$$\text{Var}(Z_t) = \lambda^4 \frac{1 + (1 - \lambda)^2 - (t + 1)^2(1 - \lambda)^{2t} + (2t^2 + 2t - 1)(1 - \lambda)^{2t+2} - t^2(1 - \lambda)^{2t+4}}{[1 - (1 - \lambda)^2]^3} c_0.$$

The PDEWMA control chart is constructed by plotting Z_t against t . The center line and control limits for the PDEWMA control chart are as follows:

$$(2.7) \quad \begin{aligned} \text{UCL} &= c_0 + L\sqrt{\text{Var}(Z_t)}, \\ \text{CL} &= c_0, \end{aligned}$$

$$(2.8) \quad \text{LCL} = c_0 - L\sqrt{\text{Var}(Z_t)},$$

where $L > 0$ can be chosen to provide a specified ARL_0 and when the computed value of LCL is less than zero, then we set $\text{LCL} = 0$. For large values of t , the control limits become (see the Appendix A for more details)

$$\begin{aligned} \text{UCL} &= c_0 + L\sqrt{\frac{\lambda(2 - 2\lambda + \lambda^2)}{(2 - \lambda)^3}} c_0, \\ \text{CL} &= c_0, \\ \text{LCL} &= c_0 - L\sqrt{\frac{\lambda(2 - 2\lambda + \lambda^2)}{(2 - \lambda)^3}} c_0. \end{aligned}$$

A process is considered to be out of control if a plotted point lies above the UCL or below the LCL.

Zhang *et al.* [13] concluded that for a PDEWMA chart a smaller value of λ makes the chart more sensitive (with smaller out-of-control ARLs). Furthermore, the PDEWMA chart gives out-of-control signals earlier than the classical PEWMA chart and in particular, the PDEWMA chart is more sensitive to small downward process mean changes than the PEWMA chart, a fact that compensates the complexity of PDEWMA in relation to PEWMA.

2.6. The Poisson EWMA control chart with Head-Start (HS PEWMA)

Lucas and Saccucci [3] introduced the Fast Initial Response (FIR) feature to the EWMA control charts. In this control chart an EWMA control scheme like the one presented in Subsection 2.4 is obtained by simultaneously implementing two one-sided EWMA, each with a head start (HS). The upper-sided HS PEWMA chart aims at detecting faster increases at the process mean whereas the lower-sided HS PEWMA chart aims at detecting faster decreases at the process mean.

Both the upper and the lower-sided HS PEWMA charts use Equation (2.3) to compute the HS PEWMA statistic. The difference with the PEWMA is the starting value. Specifically, the upper-sided HS PEWMA has a starting value larger than c_0 and lower than UCL (Equation (2.4)) whereas the lower-sided HS PEWMA has a starting value lower than c_0 and larger than LCL (Equation (2.5)).

The rationale of the HS PEWMA control chart is that if the process is initially out-of-control, then the HS PEWMA will give an out of control signal faster than the PEWMA chart. However, if the process is initially in control, HS PEWMA and PEWMA will tend to converge. In this paper, the starting value used in the HS PEWMA chart is the halfway between the mean of the process c_0 and the control limit (UCL and LCL for the upper and lower HS PEWMA control charts, respectively).

2.7. Fast Initial Response Poisson EWMA control chart (FIR PEWMA)

The FIR PEWMA control chart uses an exponentially decreasing adjustment method introduced by Steiner [10] to narrow the distance between the control limits. The control statistic of this chart is the same as in the classical PEWMA (Equation (2.3)) but its time-varying control limits are adjusted as follows:

$$(2.9) \quad \text{UCL} = c_0 + LF_{\text{adj}} \sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2t}] c_0},$$

$$\text{CL} = c_0,$$

$$(2.10) \quad \text{LCL} = c_0 - LF_{\text{adj}} \sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2t}] c_0},$$

where F_{adj} denotes the FIR adjustment factor and is expressed as

$$F_{\text{adj}} = 1 - (1 - f)^{1+a(t-1)},$$

$a > 0$ is the adjustment parameter and f is the distance from the starting value with $0 < f \leq 1$. The value of a is chosen, so that the FIR adjustment has a small effect when t gets a suitable (usually not large) value. Steiner [10] suggests to choose a so that the FIR has little effect after about 20 observations. This fact after some calculations leads to $a = (-2/\log(1 - f) - 1)/19$. In this paper, we use $f = 0.5$ and $a = 0.3$.

3. APPLICATION OF CONTROL CHARTS AT THE HELLENIC AIR FORCE (HAF) DATA

HAF is tasked with missions that, depending on the situation, conditions and environment, may involve acceptance of a significant and sometimes high risk. Daily challenges in the Aegean sea and many flight hours require continuous alertness for these missions to be performed safely. The cost of the accidents, both in the air and on the ground, and the high cost of acquiring new aircraft requires that every effort be made to minimize loss or damage in order to maintain the integrity of the aircraft and the flight ability of HAF. The implementation of this effort is achieved through the detection of risks and the monitoring of accidents.

The main aircraft included in the fleet of HAF is F-16. The annual F-16 accidents for HAF are presented in Table 1.

Table 1: Number of F-16 accidents (1988–2017).

Year	Accidents	Year	Accidents	Year	Accidents
1988	0	1998	0	2008	0
1989	0	1999	0	2009	1
1990	0	2000	1	2010	2
1991	0	2001	1	2011	0
1992	1	2002	0	2012	0
1993	1	2003	1	2013	0
1994	0	2004	2	2014	1
1995	3	2005	0	2015	2
1996	0	2006	1	2016	0
1997	1	2007	1	2017	0

The main objective of this application is to see if there is a shift in the F-16 accidents the last twenty years. For this reason, we use the first ten years to estimate the in-control mean of accidents. Since we have six total accidents the first ten years, we estimate c by

$$\bar{c} = \frac{6}{10} = 0.6.$$

Therefore, the Phase I (trial) control limits are given by

$$UCL = \bar{c} + 3\sqrt{\bar{c}} = 0.6 + 3\sqrt{0.6} = 2.92,$$

$$CL = \bar{c} = 0.6,$$

$$LCL = \bar{c} - 3\sqrt{\bar{c}} = 0.6 - 3\sqrt{0.6} = -1.72 \Rightarrow LCL = 0.$$

The control chart for the number of accidents of the first ten years is given in Figure 1.

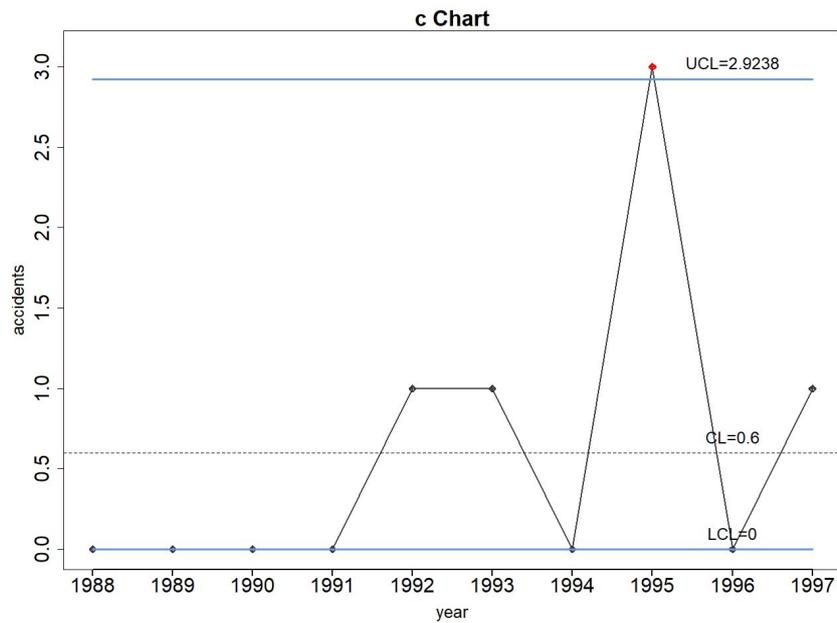


Figure 1: *c* chart for F-16 accidents (Phase I).

We may see in Figure 1 that there is one point that plots above the UCL (year 1995). We exclude this point and revise the trial control limits. The estimate of c is now computed as

$$\bar{c} = \frac{3}{9} = 0.3333.$$

Using the goodness of fit test ($\chi^2 = 0.7693$ with p value 0.6807), we conclude that the number of accidents from 1988 to 1997 (except of course year 1995) fits the Poisson distribution with parameter $c = 0.3333$. The revised control limits are

$$UCL = \bar{c} + 3\sqrt{\bar{c}} = 0.3333 + 3\sqrt{0.3333} = 2.0653,$$

$$CL = \bar{c} = 0.3333,$$

$$LCL = \bar{c} - 3\sqrt{\bar{c}} = 0.3333 - 3\sqrt{0.3333} = -1.73987 \Rightarrow LCL = 0.$$

Since all the points are between the control limits we assume that these are the final Phase I control limits that are to be used for the monitoring of the following time periods (years).

For the Phase II charts that follow, we assume that the parameter $\bar{c} = 0.3333$ is the true value of c . However, it is important to note that estimation error often exists in practice, which would result in negative effects on control charts performance.

Let $X_t, t = 1, 2, \dots, 20$, be the number of accidents from 1998 to 2017. Using the goodness of fit test ($\chi^2 = 0.8783$ with p value 0.8307), we observe that X_t fits the Poisson distribution with parameter $c^* = 0.65$. These points are plotted on the control chart (Phase II) in Figure 2. The c chart will never be able to detect a downward shift in the mean number of accidents since $LCL = 0$.

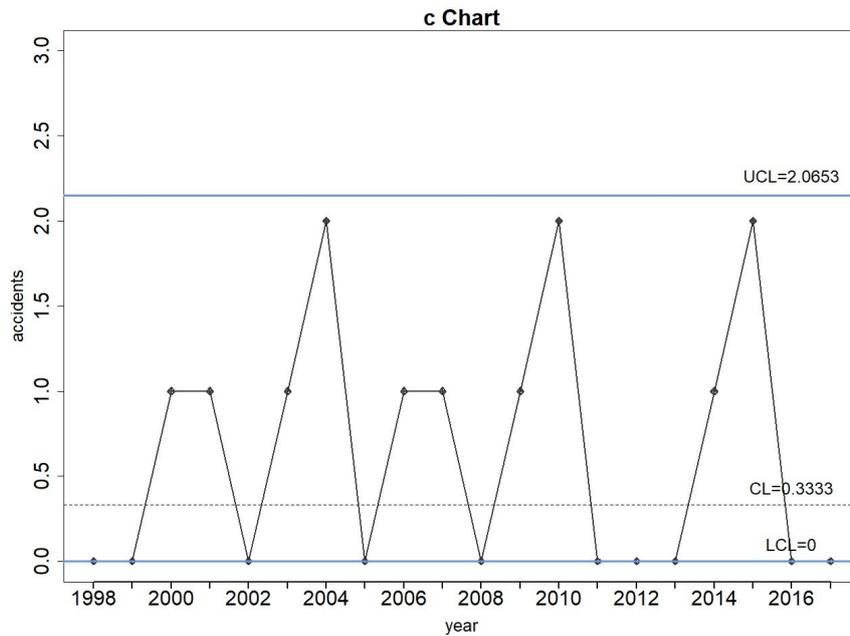


Figure 2: c chart for F-16 accidents (Phase II).

The in-control ARL for this c chart is

$$ARL_0 = \frac{1}{P(X_t > 2 \mid c = 0.3333)} = \frac{1}{0.0048163} \cong 207.63.$$

Therefore, if the process is really in-control, we will experience a false out-of-control signal about every 207–208 years. As the process shifts out of control to $c^* = 0.65$, the value of ARL_1 is

$$ARL_1 = \frac{1}{1 - P(0 \leq X_t \leq 2 \mid c^* = 0.65)} = \frac{1}{1 - 0.9716577} \cong 35.28$$

and it will take about 35 years to detect this shift with a point crossing the control limits.

In order to calculate the ARL for the PEWMA, PDEWMA, HS PEWMA and FIR PEWMA control charts, we perform Monte Carlo simulations using R. The simulation algorithm is explained as follows:

1. A combination of design parameters (λ, L) is selected and we also set $c_0 = 0.3333$. Then, the control limits of each control chart are calculated using Equations (2.4) and (2.5) for the PEWMA, (2.7) and (2.8) for the PDEWMA and (2.9) and (2.10) for the FIR PEWMA. The control limits for the HS PEWMA are calculated using the methodology described in Subsection 2.6.
2. 25,000 Poisson random numbers are generated with parameters from the previous step.
3. The statistics Z_t , $t = 1, \dots, 25,000$ are calculated for each control chart.
4. If $Z_t > LCL_t$ or $Z_t < UCL_t$, the process is considered to be in-control, but if $Z_t \leq LCL_t$ or $Z_t \geq UCL_t$, a signal is given and the process is considered to be out-of-control. When this event occurs, the simulations stop and the run-length (RL) is recorded.
5. Steps (2)–(4) are repeated 10,000 times. An approximation of the ARL is given by

$$ARL = \frac{\sum_{t=1}^N RL_t}{N}$$

where N is the number of simulations runs, i.e. in this article $N = 10,000$.

Table 2 shows the performance of various control charts for some combinations of (λ, L) . These combinations have been selected so that the ARL_0 of the control charts be close to 207.63. Moreover, the asymptotic control limits are presented in this table. The probabilities γ_{LCL} and γ_{UCL} of the ARL-unbiased c chart are calculated using Equations (2.1) and (2.2), respectively, and they are equal to 0.006171 and $6.523 \cdot 10^{-8}$. “—” is used to indicate that a downward shift cannot be detected, as in some control charts the asymptotic lower control limit is equal to zero. The same results are presented in Appendix B for the case that ARL_0 is close to 370.37. From Table 2, we conclude to the following:

1. PDEWMA control charts, as well as PEWMA, HS PEWMA and FIR PEWMA control charts with $\lambda = 0.05$ can detect a downward shift as they have $LCL > 0$. Moreover, the ARL-unbiased c chart can detect downward shifts although its LCL is equal to zero. However, these control charts, except from the ARL-unbiased c chart and PDEWMA chart with $\lambda = 0.05$, are ARL-biased, as some ARL_1 values are larger than the ARL_0 values. PDEWMA control chart with $\lambda = 0.05$ is suggested to be used in order to detect a downward shift as its ARL_1 values are smaller than the corresponding values of ARL-unbiased c chart.
2. For $\lambda = 0.05, 0.10$ and 0.15 , the PEWMA chart is more efficient than the PDEWMA chart in detecting upward shifts and vice versa for $\lambda = 0.20$. However, Zhang *et al.* [13] showed that PDEWMA chart performs similarly or slightly better than the PEWMA chart in detecting upward shifts considering the in-control mean equal to 4, 8, 12 or 20. We observe different performance of PEWMA and PDEWMA charts for processes where the in-control mean is small.

- For a specified value of λ , HS PEWMA and FIR PEWMA control charts are more efficient than c chart, ARL-unbiased c chart, PEWMA and PDEWMA control charts in detecting upward shifts. Furthermore, when $\lambda = 0.05$, the HS PEWMA performs similarly with the FIR PEWMA control chart, but when $\lambda = 0.10, 0.15$ or 0.20 , the FIR PEWMA is more efficient than the HS PEWMA. For example, when $c^* = 0.65$, the ARL_1 for a HS PEWMA chart with $\lambda = 0.05$ and $L = 2.331$ is 10.83, while the ARL_1 for a FIR PEWMA chart with $\lambda = 0.05$ and $L = 2.315$ is 10.89, the ARL_1 for a PEWMA chart with $\lambda = 0.05$ and $L = 2.261$ is 13.39 and the ARL_1 for a PDEWMA chart with $\lambda = 0.05$ and $L = 1.680$ is 14.95.

Table 2: ARL_1 values for various control charts with $ARL_0 \cong 207.63$.

	c chart	PEWMA				PDEWMA			
shift	$\lambda = 1$ $L = 3$	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	UCL = 2.065 LCL = 0	0.542 0.124	0.668 0	0.790 0	0.899 0	0.443 0.223	0.518 0.149	0.583 0.084	0.639 0.028
0.15	—	50.64	—	—	—	30.10	39.24	53.83	92.34
0.20	—	90.33	—	—	—	49.85	68.51	99.05	190.76
0.25	—	179.92	—	—	—	94.36	129.61	184.10	336.27
0.30	—	263.86	—	—	—	183.58	211.59	246.79	302.22
0.3333	207.63	207.49	207.56	207.70	207.69	207.44	207.38	207.67	207.48
0.40	126.16	86.94	83.64	89.31	96.31	99.88	102.20	99.29	91.70
0.45	91.92	49.43	49.78	54.28	60.95	55.49	58.30	58.05	54.63
0.50	69.50	31.72	32.96	36.52	41.38	35.69	37.57	37.62	36.20
0.55	54.16	22.23	23.35	26.05	30.02	24.97	26.28	26.36	25.62
0.60	43.26	16.79	17.69	19.65	22.97	18.78	19.60	19.65	19.14
0.65	35.28	13.39	14.14	15.44	17.98	14.95	15.60	15.66	15.26

	ARL-unbiased	HS PEWMA				FIR PEWMA			
shift		$\lambda = 0.05$ $L = 2.331$	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	UCL = 3 LCL = 0	0.549 0.118	0.671 0	0.791 0	0.902 0	0.547 0.119	0.679 0	0.801 0	0.913 0
0.15	187.61	49.11	—	—	—	52.35	—	—	—
0.20	195.73	91.92	—	—	—	94.68	—	—	—
0.25	202.46	190.56	—	—	—	188.82	—	—	—
0.30	206.73	280.12	—	—	—	272.15	—	—	—
0.3333	207.63	207.40	207.42	207.52	207.45	207.74	207.45	207.63	207.58
0.40	203.55	77.28	80.78	88.65	94.30	78.58	77.05	85.40	89.28
0.45	194.93	42.60	46.97	53.26	58.94	43.42	44.04	49.95	54.30
0.50	182.00	26.54	30.66	35.26	39.50	27.13	28.09	32.27	35.25
0.55	165.94	18.35	21.73	24.94	28.26	18.70	19.45	22.40	24.53
0.60	148.26	13.66	16.37	18.92	21.43	13.88	14.39	16.67	18.21
0.65	130.39	10.83	13.03	14.91	16.68	10.89	11.31	12.88	14.03

The PEWMA control charts for $\lambda = 0.05$ and $\lambda = 0.10$ are shown in Figures 3 and 4, respectively. These two control charts have the same performance since thirteen observations are needed to issue an out of control signal. Theoretically, the average number of observations needed to detect the shift is thirteen and fourteen, respectively (see Table 2).

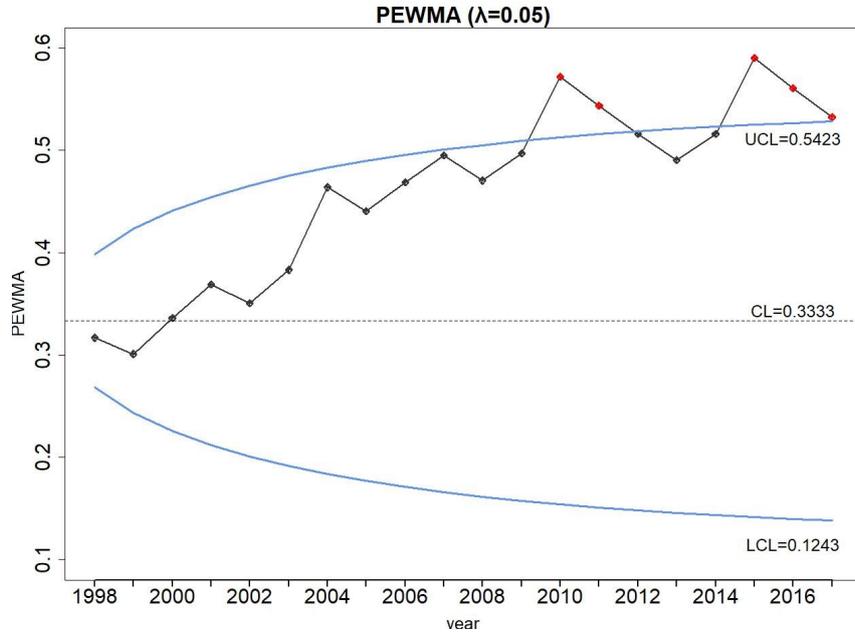


Figure 3: PEWMA ($\lambda = 0.05$).

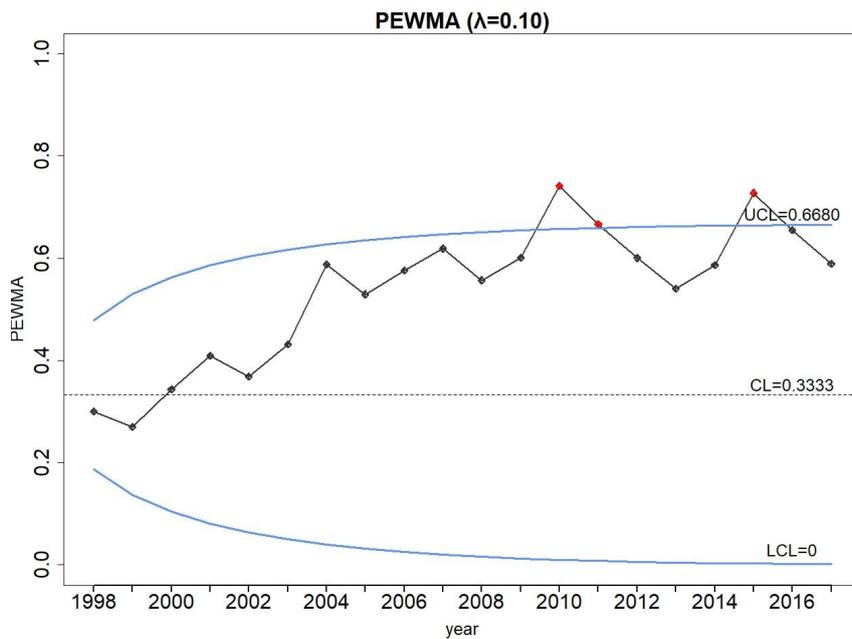


Figure 4: PEWMA ($\lambda = 0.10$).

The PDEWMA control charts for $\lambda = 0.05$ and $\lambda = 0.10$ are shown in Figures 5 and 6, respectively. These control charts have the same performance with the corresponding PEWMA charts as they also need thirteen observations to detect the shift. This value is close to the theoretically ARL_1 given in Table 2.

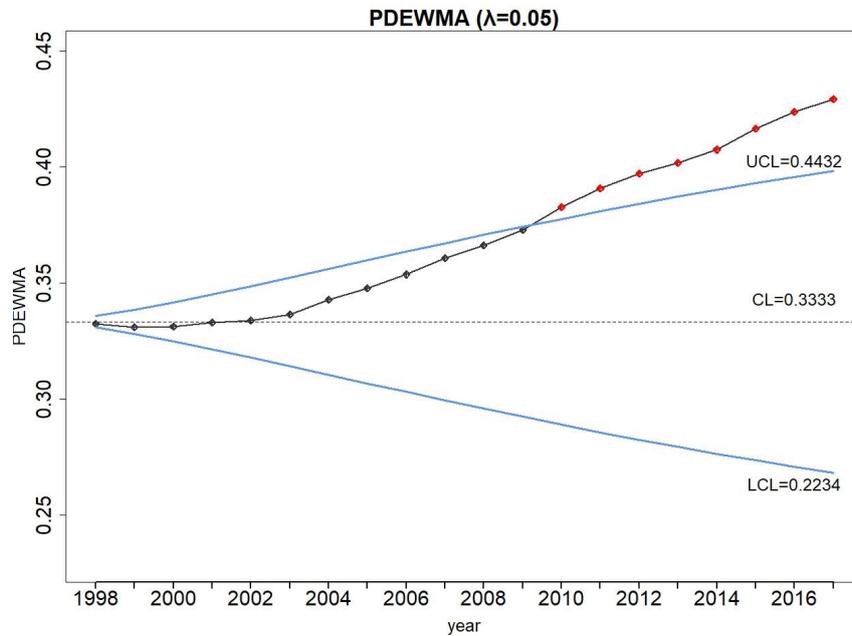


Figure 5: PDEWMA ($\lambda = 0.05$).

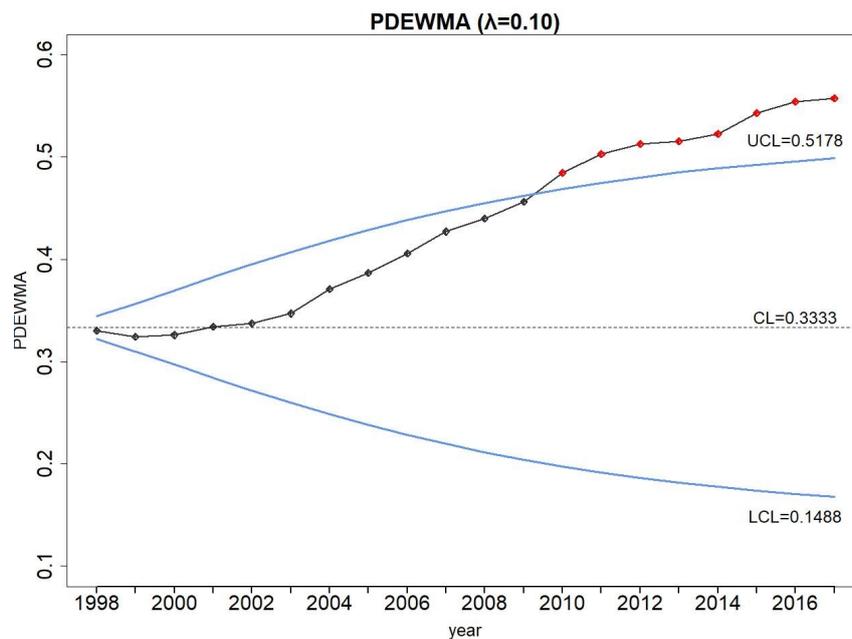


Figure 6: PDEWMA ($\lambda = 0.10$).

In Figure 7, we present the HS PEWMA control chart for $\lambda = 0.05$ when the starting value is halfway between the mean of the process c_0 and the control limit. We notice that the HS PEWMA control chart with $\lambda = 0.05$ detects the shift after ten observations and apparently its performance is much better than all the control charts already presented. Moreover, the theoretical ARL_1 value for this chart is 10.83 which is smaller than all the other competing charts. The HS PEWMA control chart with $\lambda = 0.10$ (Figure 8) detects the shift after thirteen observations, having similar performance to the corresponding PEWMA charts. Note also that as the value of λ increases, the two plotted statistics converge faster.

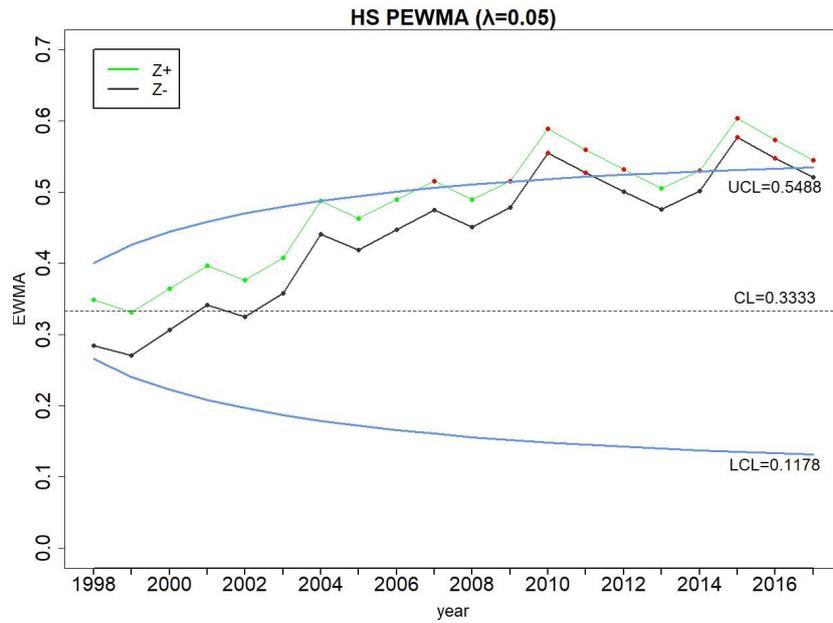


Figure 7: HS PEWMA ($\lambda = 0.05$).

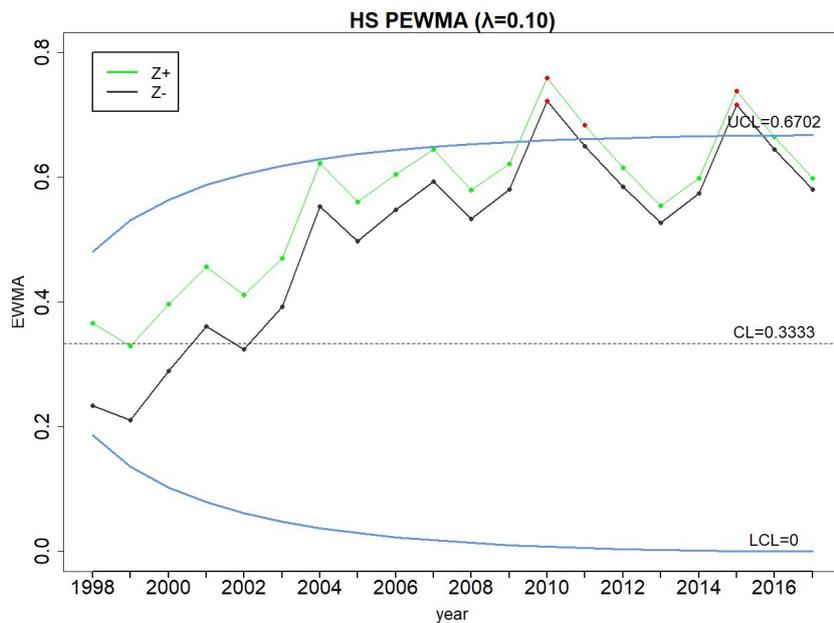


Figure 8: HS PEWMA ($\lambda = 0.10$).

In Figures 9 and 10 we present the FIR PEWMA for $\lambda = 0.05$ and $\lambda = 0.10$, respectively. We deduce that the FIR PEWMA chart with $\lambda = 0.05$ detects the shift after ten observations and therefore its performance is the same as HS PEWMA chart. On the other hand, FIR PEWMA control chart with $\lambda = 0.10$ detects the shift after thirteen observations and its performance is the same as the other three corresponding charts. In both Figures 9 and 10, the ARL_1 value is close to theoretical values given in Table 2.

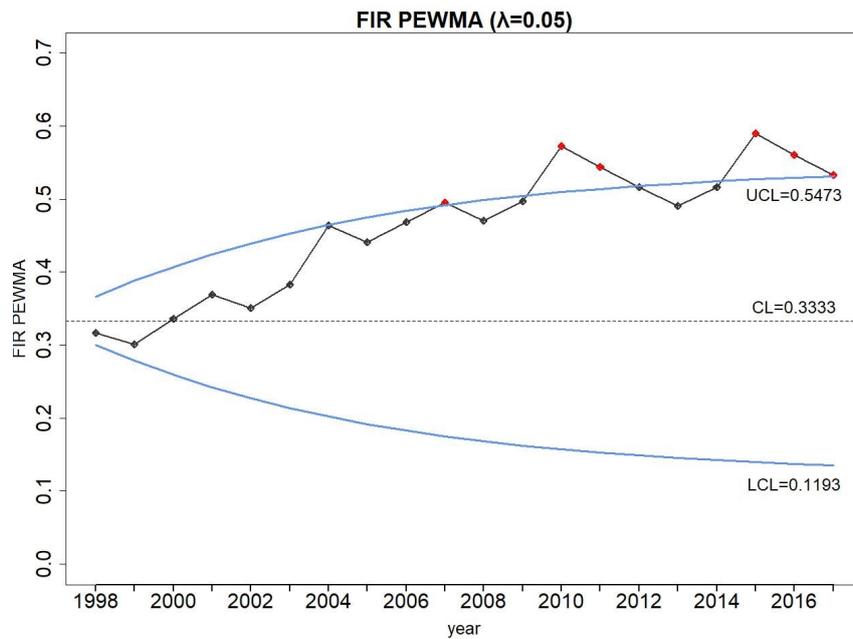


Figure 9: FIR PEWMA ($\lambda = 0.05$).

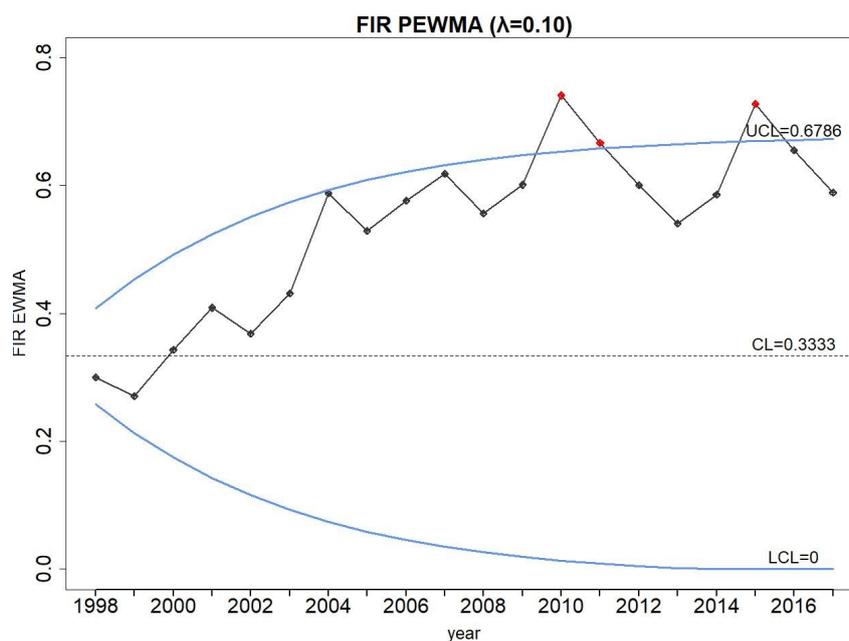


Figure 10: FIR PEWMA ($\lambda = 0.10$).

To sum up, the results show that PEWMA charts with HS or FIR feature and $\lambda = 0.05$ detect the shift more quickly than the other charts. Since, according to the HS PEWMA and FIR PEWMA with $\lambda = 0.05$ control charts, there is a shift in observation ten, management should search for an assignable cause at year 2007.

4. CONCLUSIONS

In this paper we model air plane accidents using the Poisson distribution and we monitor these accidents using Shewhart and EWMA control charts. We present several different control charts and we discuss their implementation both theoretically and practically. We apply these charts to the HAF Data and we draw useful conclusions.

Process monitoring with control charts is an important component within an overall process evaluation and improvement in air force industry. Future research will focus on more sophisticated control charts that can be applied in similar data taking into account the fact that less accidents occur as the air force industry incorporates new technologies.

A. APPENDIX

We have

$$\text{Var}(Z_t) = \lambda^4 \frac{1 + (1 - \lambda)^2 - (t + 1)^2(1 - \lambda)^{2t} + (2t^2 + 2t - 1)(1 - \lambda)^{2t+2} - t^2(1 - \lambda)^{2t+4}}{[1 - (1 - \lambda)^2]^3} c_0$$

and we will prove that $\lim_{t \rightarrow \infty} \text{Var}(Z_t) = \frac{\lambda(2 - 2\lambda + \lambda^2)}{(2 - \lambda)^3} c_0$.

First of all, for $\lambda = 1$, the PDEWMA, as well as the PEWMA, reduces to a c chart and $\text{Var}(Z_t) = c_0$.

For $\lambda < 1$ and applying L'Hospital's rule, we have

$$\begin{aligned} \lim_{t \rightarrow \infty} (t + 1)^2(1 - \lambda)^{2t} &= \lim_{t \rightarrow \infty} \frac{(t + 1)^2}{\left(\frac{1}{1 - \lambda}\right)^{2t}} = \lim_{t \rightarrow \infty} \frac{2t + 2}{2\left(\frac{1}{1 - \lambda}\right)^{2t} \ln\left(\frac{1}{1 - \lambda}\right)} \\ &= \lim_{t \rightarrow \infty} \frac{1}{2\left(\frac{1}{1 - \lambda}\right)^{2t} \left(\ln\left(\frac{1}{1 - \lambda}\right)\right)^2} = 0. \end{aligned}$$

In the same way, we have $\lim_{t \rightarrow \infty} (2t^2 + 2t - 1)(1 - \lambda)^{2t} = \lim_{t \rightarrow \infty} t^2(1 - \lambda)^{2t+4} = 0$.

$$\text{So, } \lim_{t \rightarrow \infty} \text{Var}(Z_t) = \lambda^4 \frac{1 + (1 - \lambda)^2}{[1 - (1 - \lambda)^2]^3} c_0 = \lambda^4 \frac{2 - 2\lambda + \lambda^2}{[\lambda(2 - \lambda)]^3} c_0 = \lambda \frac{2 - 2\lambda + \lambda^2}{(2 - \lambda)^3} c_0.$$

B. APPENDIX

Table 3: ARL_1 values for various control charts with $ARL_0 \cong 370.37$.

	<i>c</i> chart	PEWMA				PDEWMA			
shift	$\lambda = 1$ $L = 3$	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	UCL = 2.065 LCL = 0	0.566 0.100	0.716 0	0.845 0	0.966 0	0.460 0.207	0.542 0.125	0.613 0.054	0.680 0
0.15	—	75.31	—	—	—	39.57	55.48	93.30	—
0.20	—	158.08	—	—	—	68.18	105.69	207.60	—
0.25	—	392.14	—	—	—	141.56	233.07	457.80	—
0.30	—	585.13	—	—	—	322.63	423.92	552.79	—
0.3333	207.63	370.19	368.34	370.13	370.51	370.63	370.29	370.81	370.30
0.40	126.16	118.72	127.89	143.78	155.09	141.42	143.20	137.25	135.75
0.45	91.92	63.54	72.45	84.66	94.21	72.47	75.79	76.56	76.38
0.50	69.50	39.62	45.90	55.06	62.10	44.59	46.96	47.70	48.98
0.55	54.16	27.21	31.42	38.29	43.41	30.49	31.81	32.38	33.54
0.60	43.26	19.99	22.87	28.16	31.87	22.31	23.13	23.49	24.21
0.65	35.28	15.72	17.61	21.58	24.48	17.52	18.07	18.19	18.71

	ARL-unbiased	HS PEWMA				FIR PEWMA			
shift		$\lambda = 0.05$ $L = 2.551$	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	UCL = 4 LCL = 0	0.569 0.097	0.717 0	0.848 0	0.969 0	0.572 0.095	0.722 0	0.854 0	0.975 0
0.15	334.31	74.50	—	—	—	80.75	—	—	—
0.20	348.86	163.55	—	—	—	175.61	—	—	—
0.25	360.99	421.06	—	—	—	452.65	—	—	—
0.30	368.72	614.36	—	—	—	639.74	—	—	—
0.3333	370.37	370.87	370.71	370.65	370.80	370.43	370.60	370.37	370.12
0.40	362.87	113.36	127.20	139.00	151.09	109.94	121.22	135.30	145.38
0.45	346.93	58.86	70.27	80.53	90.26	56.24	65.95	76.19	84.75
0.50	322.95	35.78	43.58	50.99	58.22	33.96	40.61	47.42	53.40
0.55	293.14	24.27	29.73	34.93	39.97	22.72	27.12	31.68	35.96
0.60	260.38	17.80	21.65	25.23	28.91	16.37	19.43	22.68	25.53
0.65	227.39	13.86	16.69	19.18	21.91	12.70	14.75	17.05	19.11

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers and the Associate Editor for the useful suggestions and comments made, which helped us to improve the content of the paper.

REFERENCES

- [1] BORROR, C.M.; CHAMP, C.W. and RIGDON, S.E. (1998). Poisson EWMA control charts, *Journal of Quality Technology*, **30**, 352–361.
- [2] KJELLN, U. and ALBRECHTSEN, E. (2017). *Prevention of Accidents and Unwanted Occurrences, Theory, Methods, and Tools in Safety Management*, CRC Press.
- [3] LUCAS, J.M. and SACCUCCI, M.S. (1990). Exponentially weighted moving average control schemes: properties and enhancements, *Technometrics*, **32**, 1–12.
- [4] MONTGOMERY, D.C. (2013). *Introduction to Statistical Quality Control*, Wiley.
- [5] PAULINO, S.; MORAIS, M.C. and KNOTH, S. (2016). An ARL-unbiased c-chart, *Quality and Reliability Engineering International*, **32**, 2847–2858.
- [6] ROBERTS, S.W. (1959). Control chart tests based on geometric moving averages, *Technometrics*, **1**, 239–250.
- [7] ROCKWELL, T.H. (1959). Safety performance measurement, *Journal of Industrial Engineering*, **10**, 12–16.
- [8] RYAN, T.P. (2011). *Statistical Methods for Quality Improvement*, 3rd Edn., Wiley.
- [9] SHAMMA, S.E. and SHAMMA, A.K. (1992). Development and evaluation of control charts using double exponentially weighted moving averages, *International Journal of Quality & Reliability*, **9**, 18–25.
- [10] STEINER, S.H. (1999). EWMA control charts with time-varying control limits and fast initial response, *Journal of Quality Technology*, **31**, 75–86.
- [11] WETHERILL, G.B. and BROWN, D.W. (1991). *Statistical Process Control: Theory and Practice*, Chapman and Hall.
- [12] WOODALL, W.H. (2000). Controversies and contradictions in statistical process control, *Journal of Quality Technology*, **32**, 341–350.
- [13] ZHANG, L.; GOVINDARAJU, K.; LAI, C.D. and BEBBINGTON, M.S. (2003). Poisson DEWMA control chart, *Communication in Statistics – Simulation and Computation*, **32**, 1265–1283.

ON THE ESTIMATION FOR COMPOUND POISSON INARCH PROCESSES

Authors: E. GONÇALVES

– CMUC, Department of Mathematics, University of Coimbra,
Portugal
esmerald@mat.uc.pt

N. MENDES-LOPES

– CMUC, Department of Mathematics, University of Coimbra,
Portugal
nazare@mat.uc.pt

F. SILVA

– CMUC, Department of Mathematics, University of Coimbra,
Portugal
mat0504@mat.uc.pt

Received: January 2018

Accepted: February 2019

Abstract:

- Considering the wide class of discrete Compound Poisson INARCH models, introduced in [6], the main goal of this paper is to develop and compare parametric estimation procedures for first-order models, applicable without specifying the conditional distribution of the process. Therefore, two-step estimation procedures, combining either the conditional least squares (CLS) or the Poisson quasi-maximum likelihood (PQML) methods with that of the moment's estimation, are introduced and discussed. Specifying the process conditional distribution, we develop also within this class of models the conditional maximum likelihood (CML) methodology. A simulation study illustrates, particularly, the competitive performance of the two-step approaches regarding the more classical CML one which requires the conditional distribution knowledge. A final real-data example shows the relevance of this wide class of models, as it will be clear the better performance in the data fitting of some new models emerging in such class.

Keywords:

- *integer-valued time series; CP-INGARCH processes; estimation.*

AMS Subject Classification:

- 62M10, 60G12, 62F10.

1. INTRODUCTION

The family of discrete compound Poisson distributions, which includes as particular cases the Poisson, the Neyman type-A or the geometric Poisson laws, was recently used to define a new class of integer-valued GARCH models, the compound Poisson INGARCH ones [6], specified through the characteristic function of the conditional law of the process given its past. Namely, $X = (X_t, t \in \mathbb{Z})$ follows a CP-INGARCH process if the characteristic function of X_t conditioned on \underline{X}_{t-1} is such that

$$\begin{cases} \Phi_{X_t|\underline{X}_{t-1}}(u) = \exp \left\{ i \frac{\lambda_t}{\varphi_t'(0)} [\varphi_t(u) - 1] \right\}, & u \in \mathbb{R}, \\ E(X_t|\underline{X}_{t-1}) = \lambda_t = \alpha_0 + \sum_{j=1}^p \alpha_j X_{t-j} + \sum_{k=1}^q \beta_k \lambda_{t-k}, \end{cases}$$

where $\alpha_0 > 0$, $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q \geq 0$, \underline{X}_{t-1} represents the σ -field generated by $\{X_{t-s}, s \geq 1\}$ and $(\varphi_t, t \in \mathbb{Z})$ is a family of characteristic functions on \mathbb{R} , \underline{X}_{t-1} -measurables, associated to a family of discrete laws with support in \mathbb{N}_0 and finite mean. If $\beta_k = 0$, $k = 1, \dots, q$, the CP-INGARCH(p, q) model is simply denoted CP-INARCH(p). The functional form of the conditional characteristic function $\Phi_{X_t|\underline{X}_{t-1}}$ allows a wide flexibility of the class of CP-INGARCH models. In fact, as it is assumed that the family of discrete characteristic functions $(\varphi_t, t \in \mathbb{Z})$ is \underline{X}_{t-1} -measurable it means that its elements may be random functions or deterministic ones. Thus, this general model unifies and enlarges substantially the family of conditionally heteroscedastic integer-valued processes. In fact, it is possible to present new specific models with conditional distributions with interest in practical applications as, for instance, the geometric Poisson INGARCH ([6]) or the Neyman type-A INGARCH ([5]) ones, and also recover recent contributions such as the Poisson INGARCH ([4]), the generalized Poisson INGARCH ([15]), the negative binomial INGARCH ([14]) and the negative binomial DINARCH ([13]) processes (corresponding to random or deterministic functions φ_t , respectively). In addition to having the ability to describe different distributional behaviors and consequently different kinds of conditional heteroscedasticity, the CP-INGARCH model is able to incorporate simultaneously the overdispersion characteristic that has been recorded in real count data.

In this paper, we focus on the case where φ_t is deterministic and constant in time which still includes many of the particular cases referred above. For that reason, from now on we will refer these functions simply as φ . In this subclass of models, there exists a strictly stationary and ergodic solution with finite first and second order moments under $\sum_{j=1}^p \alpha_j + \sum_{k=1}^q \beta_k < 1$ ([6]). For $p = q = 1$, Gonçalves, Mendes-Lopes and Silva [7] stated that this simple coefficient condition is also necessary and sufficient to establish the existence of all the moments of X_t .

In this class of models we have, additionally to the usual estimation of the parameters of the conditional mean, the estimation of φ . We observe that a related problem with the knowledge of φ has been discussed in [12] in which a testing methodology was proposed to distinguish between a simple Poisson INARCH model ($\varphi(u) = \exp(iu)$) and a true CP-INARCH one ($\varphi(u) \neq \exp(iu)$). In order to analyse φ , in this paper we propose a two-step estimation procedure that lead us to its consistent estimation after estimating the conditional mean parameters.

The remainder of the paper proceeds as follows. In Section 2 we consider the subclass of CP-INARCH models of order one, with $\varphi_t = \varphi$ deterministic, and deduce its moments, central moments and cumulants up to the order 4. These results are particularly important in Section 3, devoted to estimation procedures, to deduce explicit expressions for the asymptotic distribution of the Conditional Least Squares (CLS) estimators of the conditional mean parameters, α_0 and α_1 . In a second step, the method of moments is used to estimate the additional parameter associated to the function φ . Another two-step estimation procedure, combining the Poisson Quasi Maximum Likelihood (PQML) and the moment methods, is also proposed in this section, followed by the Conditional Maximum Likelihood (CML) estimation for the NTA-INARCH(1) and GEOMP2-INARCH(1) models. Section 4 presents some simulation studies that illustrate and compare the performance of these three methodologies of estimation. In Section 5 an integer-valued time series related to the prices of electricity in Portugal and Spain between July 2016 and June 2017 is considered. The data is fitted by several CP-INARCH(1) models estimated by the three estimation approaches considered and the quality of the fitting is discussed using for the CML method, in particular, the values of the log likelihood function, Akaike and Bayesian information criteria. Detailed calculations are included in the [Appendices](#).

2. THE CP-INARCH(1) PROCESS

Let us consider now the subclass of CP-INARCH(1) models. Supposing $\varphi_t = \varphi$ constant in time and deterministic we recall that $\alpha_1 < 1$ is a necessary and sufficient condition to assure the existence of a strictly stationary and ergodic solution of the model. Moreover the process has moments of all the orders.

Setting $X = (X_t, t \in \mathbb{Z})$ a CP-INARCH(1) process we derive in the following closed-form expressions for the joint (central) moments and cumulants of the CP-INARCH(1) up to order 4. In fact, setting the notations below (used, for instance, by Weiß in [10]),

$$f_k = \frac{\alpha_0}{\prod_{j=1}^k (1 - \alpha_1^j)}, \quad k \in \mathbb{N},$$

$$(2.1) \quad \begin{aligned} \mu(s_1, \dots, s_{r-1}) &= E(X_t X_{t+s_1} \cdots X_{t+s_{r-1}}), \\ \tilde{\mu}(s_1, \dots, s_{r-1}) &= E\left((X_t - \mu)(X_{t+s_1} - \mu) \cdots (X_{t+s_{r-1}} - \mu)\right), \\ \kappa(s_1, \dots, s_{r-1}) &= \text{Cum}[X_t, X_{t+s_1}, \dots, X_{t+s_{r-1}}], \end{aligned}$$

with $r = 2, 3, 4$ and $0 \leq s_1 \leq \dots \leq s_{r-1}$, and

$$v_0 = -i \frac{\varphi''(0)}{\varphi'(0)}, \quad d_0 = -\frac{\varphi'''(0)}{\varphi'(0)}, \quad c_0 = i \frac{\varphi^{(iv)}(0)}{\varphi'(0)},$$

we establish the following results whose proofs may be found in [Appendices A](#) and [B](#), respectively.

Theorem 2.1 (Moments of a CP-INARCH(1) process). *We have:*

(a) For any $k \geq 0$, $\mu(k) = f_2(v_0 \alpha_1^k + \alpha_0(1 + \alpha_1))$.

(b) For any $l \geq k \geq 0$,

$$\begin{aligned} \mu(k, l) = & \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] f_3 \alpha_1^{l+k} + \frac{v_0(\alpha_0 + v_0)}{1 - \alpha_1} f_2 \alpha_1^l \\ & + v_0 f_1 f_2 \alpha_1^{l-k} + f_1 \mu(k). \end{aligned}$$

(c) For any $m \geq l \geq k \geq 0$,

$$\begin{aligned} \mu(k, l, m) = & \alpha_1^{m-l} \left[\left\{ (c_0 - 4v_0d_0 + 3v_0^3) + 3v_0(v_0^2 - d_0)\alpha_1 + (3v_0d_0 - c_0)\alpha_1^2 \right. \right. \\ & + (7v_0d_0 - 6v_0^3 - c_0)\alpha_1^3 + 3v_0(d_0 - 2v_0^2)\alpha_1^4 + (6v_0^3 - 6v_0d_0 + c_0)\alpha_1^5 \left. \right\} f_4 \alpha_1^{2l+k} \\ & + \frac{2v_0 + \alpha_0}{1 - \alpha_1} f_3 \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] \alpha_1^{2l} \\ & + \frac{v_0}{(1 - \alpha_1)(1 - \alpha_1^2)} f_2 \left[2v_0\alpha_0 + d_0(1 - \alpha_1) + v_0^2(2\alpha_1 - 1) \right] \alpha_1^{2l-k} \\ & + \frac{\alpha_0 f_3}{1 - \alpha_1} \left\{ d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right\} \alpha_1^{2(l-k)} + \frac{v_0 + \alpha_0}{1 - \alpha_1} \mu(k, l) \\ & \left. - f_2 \mu(k) \left[\alpha_0 + (v_0 + \alpha_0)\alpha_1 \right] \right] + f_1 \mu(k, l). \end{aligned}$$

Corollary 2.1 (Central Moments and Cumulants of a CP-INARCH(1) process).

We have:

(a) For any $s \geq 0$, $\tilde{\mu}(s) = \kappa(s) = v_0 \alpha_1^s f_2$.

(b) For any $l \geq s \geq 0$, we have

$$\tilde{\mu}(s, l) = \kappa(s, l) = f_3 \alpha_1^l \left[v_0^2(1 + \alpha_1 + \alpha_1^2) - \left\{ v_0^2(1 + \alpha_1 - 2\alpha_1^2) - d_0(1 - \alpha_1^2) \right\} \alpha_1^s \right].$$

(c) For any $m \geq l \geq s \geq 0$,

$$\begin{aligned} \kappa(s, l, m) = & \alpha_1^m f_4 \left[\left\{ c_0 + 3v_0^3 - 4v_0d_0 + 3v_0(v_0^2 - d_0)\alpha_1 + (3\alpha_0d_0 - c_0)\alpha_1^2 \right. \right. \\ & + (7v_0d_0 - 6v_0^3 - c_0)\alpha_1^3 + 3v_0(d_0 - 2v_0^2)\alpha_1^4 + (6v_0^3 - 6v_0d_0 + c_0)\alpha_1^5 \left. \right\} \alpha_1^{l+s} \\ & + v_0(1 + \alpha_1 + \alpha_1^2 + \alpha_1^3) \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] (2\alpha_1^l + \alpha_1^s) \\ & \left. + v_0(1 + \alpha_1 + \alpha_1^2)(1 + \alpha_1^2) \left[(1 + \alpha_1)v_0^2 + (d_0(1 - \alpha_1) + v_0^2(2\alpha_1 - 1))\alpha_1^{l-s} \right] \right], \end{aligned}$$

$$\tilde{\mu}(s, l, m) = \kappa(s, l, m) + v_0^2 f_2^2 (\alpha_1^{m-l+s} + 2\alpha_1^{m+l-s}).$$

From Theorem 2.1 we deduce, for instance,

$$(2.2) \quad E(X_t^2) = \mu(0) = \frac{\alpha_0(v_0 + \alpha_0(1 + \alpha_1))}{(1 - \alpha_1)(1 - \alpha_1^2)},$$

$$E(X_t^3) = \mu(0, 0) = \frac{\alpha_0}{(1 - \alpha_1)^3} \left[\frac{d_0 + (3v_0^2 - d_0)\alpha_1^2}{(1 + \alpha_1)(1 + \alpha_1 + \alpha_1^2)} + \frac{3v_0\alpha_0}{1 + \alpha_1} + \alpha_0^2 \right].$$

These results generalize those of Weiß [10] for the INARCH(1) model and the two last equalities are important to deduce explicit expressions for the asymptotic distribution of the CLS estimators of the parameters α_0 and α_1 provided in the next section. As we will take in our study some important particular cases concerning the process conditional law, we conclude this section recalling such cases and deducing the corresponding values of v_0 , d_0 and c_0 , previously introduced.

- a) The INARCH(1) model ([4]) corresponds to a CP-INARCH model considering φ the characteristic function of the Dirac's law concentrated in $\{1\}$, that is, with a Poisson conditional distribution; we denote it by Poisson-INARCH(1) model. So, we deduce that $v_0 = d_0 = c_0 = 1$.
- b) When φ is the characteristic function of the Poisson distribution with mean $\phi > 0$, $X_t | \underline{X}_{t-1}$ follows a Neyman type-A law with parameter $(\lambda_t/\phi, \phi)$, and we have the NTA-INARCH(1) model introduced in [5]. For this case, $v_0 = 1 + \phi$, $d_0 = 1 + 3\phi + \phi^2$ and $c_0 = 1 + 7\phi + 6\phi^2 + \phi^3$.
- c) Considering in the above expressions $v_0 = (2 - p^*)/p^*$, $d_0 = (6 - 6p^* + (p^*)^2)/(p^*)^2$ and $c_0 = ((2 - p^*)(12 - 12p^* + (p^*)^2))/(p^*)^3$, we obtain the expressions for the GEOMP2-INARCH(1) model ([6]). In fact, this process is defined considering φ the characteristic function of the geometric distribution with parameter $p^* \in]0, 1[$ and $X_t | \underline{X}_{t-1}$ following a geometric Poisson $(p^*\lambda_t, p^*)$ law.
- d) Another particular case of the CP-INARCH model is the NB2-INARCH (that is identical to the NB-DINARCH model proposed by Xu *et al.*, [13]), where $X_t | \underline{X}_{t-1}$ follows a negative binomial distribution with parameter $(\lambda_t/(\beta-1), 1/\beta)$ and $\beta > 0$. This process is stated when φ is the characteristic function of the logarithmic distribution with parameter $(\beta-1)/\beta$ and then we deduce $v_0 = \beta$, $d_0 = 2\beta^2 - \beta$ and $c_0 = 6\beta^2(\beta-1) + \beta$.
- e) When φ is the characteristic function of the Borel law with parameter $\kappa \in]0, 1[$, $X_t | \underline{X}_{t-1}$ follows a generalized Poisson distribution with parameter $((1-\kappa)\lambda_t, \kappa)$ and we recover the GP-INARCH model ([15]). So, $v_0 = (1-\kappa)^{-2}$, $d_0 = (2\kappa+1)(1-\kappa)^{-4}$ and $c_0 = (6\kappa^2 + 8\kappa + 1)(1-\kappa)^{-6}$.

3. ESTIMATION PROCEDURES

In this section, we focus on the estimation of the vector $\theta = (\alpha_0, \alpha_1, v_0)^\top$, where v_0 includes the additional parameter associated to the conditional distribution of the CP-INARCH(1) model (for example, $v_0 = 1 + \phi$ in the NTA-INARCH(1) model and $v_0 = (2 - p^*)/p^*$ in the GEOMP2-INARCH(1)). To estimate the true value of θ , we start by discussing a two-step approach using the conditional least squares and moment estimation methods; after we consider the combination of the Poisson Quasi-Maximum Likelihood and moments estimation methods and finally develop the conditional maximum likelihood estimation. For this purpose, let (x_1, \dots, x_n) be n particular values, arbitrarily fixed, of the process X .

3.1. Two-step estimation procedures

3.1.1. Conditional Least Squares and Moments estimation methods

In the first step, we discuss the conditional least squares (CLS) approach for the estimation of the conditional mean parameters α_0 and α_1 and, for parameter v_0 associated to the CP-INARCH(1) conditional distribution, an approach based on the moment estimation method is developed.

The CLS estimator of $\alpha = (\alpha_0, \alpha_1)$ is obtained by minimizing the sum of squares

$$Q_n(\alpha) = \sum_{t=2}^n \left[x_t - E(X_t | X_{t-1} = x_{t-1}) \right]^2 = \sum_{t=2}^n \left[x_t - \alpha_0 - \alpha_1 x_{t-1} \right]^2,$$

with respect to α . Solving the least squares equations

$$\begin{cases} \frac{\partial Q_n(\alpha)}{\partial \alpha_0} = -2 \sum_{t=2}^n (x_t - \alpha_0 - \alpha_1 x_{t-1}) = 0, \\ \frac{\partial Q_n(\alpha)}{\partial \alpha_1} = -2 \sum_{t=2}^n x_{t-1} (x_t - \alpha_0 - \alpha_1 x_{t-1}) = 0, \end{cases}$$

we obtain the following explicit expressions for the CLS estimator $\hat{\alpha}_n = (\hat{\alpha}_{0,n}, \hat{\alpha}_{1,n})$:

$$\begin{aligned} \hat{\alpha}_{1,n} &= \frac{\sum_{t=2}^n X_t X_{t-1} - \frac{1}{n-1} \cdot \sum_{t=2}^n X_t \cdot \sum_{s=2}^n X_{s-1}}{\sum_{t=2}^n X_{t-1}^2 - \frac{1}{n-1} (\sum_{t=2}^n X_{t-1})^2}, \\ \hat{\alpha}_{0,n} &= \frac{\sum_{t=2}^n X_t - \hat{\alpha}_{1,n} \sum_{t=2}^n X_{t-1}}{n-1}. \end{aligned} \tag{3.1}$$

The consistency and the asymptotic distribution of these estimators are stated in the next theorem. This theorem generalizes the results obtained in [11], Section 4.2, where the CLS estimators of α_0 and α_1 are obtained and studied in the particular case of a Poisson-INARCH model.

Theorem 3.1. *Let $\hat{\alpha}_n = (\hat{\alpha}_{0,n}, \hat{\alpha}_{1,n})$ be the CLS estimator of $\alpha = (\alpha_0, \alpha_1)$ given in (3.1). Then $\hat{\alpha}_n$ converges almost surely to α and*

$$\sqrt{n}(\hat{\alpha}_n - \alpha) \xrightarrow{d} N(\mathbf{0}_{2 \times 1}, \mathbf{V}^{-1} \mathbf{W} \mathbf{V}^{-1}),$$

as $n \rightarrow \infty$, where the entries of the matrix $\mathbf{V}^{-1} \mathbf{W} \mathbf{V}^{-1} = (b_{ij})$, $i, j = 1, 2$, are given by

$$b_{11} = \frac{\alpha_0}{1 - \alpha_1} \left(\alpha_0(1 + \alpha_1) + \frac{v_0^2 + (d_0 - v_0^2) \alpha_1(1 + \alpha_1 - \alpha_1^2) + (3v_0^2 - d_0) \alpha_1^4}{v_0(1 + \alpha_1 + \alpha_1^2)} \right),$$

$$b_{12} = b_{21} = v_0 \alpha_1 - \alpha_0(1 + \alpha_1) - \frac{\alpha_1(1 + \alpha_1)(d_0 + (3v_0^2 - d_0) \alpha_1^2)}{v_0(1 + \alpha_1 + \alpha_1^2)},$$

$$b_{22} = (1 - \alpha_1^2) \left(1 + \frac{\alpha_1(d_0 + (3v_0^2 - d_0) \alpha_1^2)}{v_0 \alpha_0(1 + \alpha_1 + \alpha_1^2)} \right),$$

and \xrightarrow{d} means convergence in distribution.

Proof: The results announced are proved using those of Klimko and Nelson [9, Section 3]. In fact, it is easily checked that the regularity conditions (i) to (iii) defined on [9, p. 634] are satisfied taking into account that $g(\alpha; X_{t-1}) = E(X_t | X_{t-1}) = \alpha_0 + \alpha_1 X_{t-1}$, and thus, by their Theorem 3.1, it follows that the CLS estimators are strongly consistent. Furthermore, the matrix \mathbf{V} is invertible as it is given by

$$\mathbf{V} = \begin{bmatrix} E\left(\frac{\partial g}{\partial \alpha_0} \frac{\partial g}{\partial \alpha_0}\right) & E\left(\frac{\partial g}{\partial \alpha_0} \frac{\partial g}{\partial \alpha_1}\right) \\ E\left(\frac{\partial g}{\partial \alpha_1} \frac{\partial g}{\partial \alpha_0}\right) & E\left(\frac{\partial g}{\partial \alpha_1} \frac{\partial g}{\partial \alpha_1}\right) \end{bmatrix} = \begin{bmatrix} E(1) & E(X_{t-1}) \\ E(X_{t-1}) & E(X_{t-1}^2) \end{bmatrix} = \begin{bmatrix} 1 & \frac{\alpha_0}{1-\alpha_1} \\ \frac{\alpha_0}{1-\alpha_1} & \frac{\alpha_0(v_0 + \alpha_0(1+\alpha_1))}{(1-\alpha_1)(1-\alpha_1^2)} \end{bmatrix},$$

considering the expressions stated in Theorem 2.1. Thus, Theorem 3.2 of [9] is satisfied implying the asymptotic normality of the CLS estimators. The entries of the covariance matrix of the asymptotic distribution $\mathbf{V}^{-1}\mathbf{W}\mathbf{V}^{-1}$ are derived in Appendix C. \square

To estimate the parameter v_0 we propose to use the moments estimation method. Taking into consideration the expression (2.2) of the second order moment of the CP-INARCH(1) model, an estimator for v_0 , whose strong consistence is a consequence from the strict stationarity and ergodicity of the process X , is given by solving the equation

$$\frac{\hat{\alpha}_{0,n}(v_0 + \hat{\alpha}_{0,n}(1 + \hat{\alpha}_{1,n}))}{(1 - \hat{\alpha}_{1,n})(1 - \hat{\alpha}_{1,n}^2)} = \frac{1}{n} \sum_{t=1}^n X_t^2$$

in order to v_0 . In this way we get the two-step CLS+M estimator for $(\alpha_0, \alpha_1, v_0)$.

We note that the estimation of v_0 doesn't involve the knowledge of the conditional law, as it is totally determined by the estimators of α_0 and α_1 and the empirical second order moment.

3.1.2. Poisson Quasi-Maximum Likelihood and Moments estimation methods

One of the advantages of using the above CLS+M approach is the fact that we do not need to specify entirely the conditional distribution of the CP-INARCH(1) model to estimate its parameters. We refer now another two-step approach where it is used the Poisson quasi-conditional maximum likelihood estimator (PQMLE) to estimate the conditional mean parameters α_0 and α_1 and, as previously, the moment estimation method for parameter v_0 . The resulting estimator is denoted PQML+M.

The PQMLE provides a general approach for estimating the conditional mean parameters of the CP-INARCH(1) models by maximizing a pseudo-likelihood function considering the conditional distribution the Poisson one, that is, the function

$$\tilde{L}_n(\theta|\mathbf{x}) = \sum_{t=2}^n \left(x_t \log(\lambda_t) - \lambda_t - \log(x_t!) \right).$$

Ahmad and Francq [1] found some regularity conditions to establish the consistency and asymptotic normality of the Poisson quasi-maximum likelihood estimator of the conditional mean parameters of a count time series. These regularity conditions are easily satisfied by a CP-INARCH(1) process with $\alpha_1 < 1$, and so the PQML estimator of (α_0, α_1) is consistent and asymptotically Gaussian. The almost sure convergence of the v_0 estimator follows as previously.

3.2. Conditional Maximum Likelihood Estimation

When the distribution of $X_t | \underline{X}_{t-1}$ is known, we can estimate its parameters using the conditional maximum likelihood estimation (CMLE) method. In this section, we discuss this procedure by considering NTA-INARCH(1) and GEOMP2-INARCH(1) models, as developed in [11], Section 4.1, for a Poisson-INARCH(1) model .

Starting by a NTA-INARCH(1) process, we have the conditional probability mass function of X_t ([8]) given by

$$P[X_t = x_t | \underline{X}_{t-1}] = \frac{e^{-\frac{\lambda_t}{\phi}} \phi^{x_t}}{x_t!} Z(\lambda_t, x_t, \phi), \quad Z(\lambda_t, X_t, \phi) = \sum_{j=0}^{\infty} \frac{(\lambda_t e^{-\phi})^j j^{X_t}}{\phi^j j!},$$

for $x_t = 0, 1, \dots$. The conditional likelihood function is then

$$L_n(\theta | \mathbf{x}) = \prod_{t=2}^n \frac{e^{-\frac{\lambda_t}{\phi}} \phi^{x_t}}{x_t!} Z(\lambda_t, x_t, \phi),$$

where for convenience $\theta = (\alpha_0, \alpha_1, \phi)$ as $v_0 = 1 + \phi$. So the log-likelihood function has the form

$$\log L_n(\theta | \mathbf{x}) = \sum_{t=2}^n l_t(\theta) = \sum_{t=2}^n \left\{ -\frac{\lambda_t}{\phi} + x_t \log(\phi) - \log(x_t!) + \log(Z(\lambda_t, x_t, \phi)) \right\}.$$

The first derivatives of l_t are given as

$$\begin{aligned} \frac{\partial l_t(\theta)}{\partial \phi} &= \frac{\lambda_t}{\phi^2} + \frac{x_t}{\phi} - \left(\frac{\phi + 1}{\phi} \right) \frac{Z(\lambda_t, x_t + 1, \phi)}{Z(\lambda_t, x_t, \phi)}, \\ \frac{\partial l_t(\theta)}{\partial \alpha_j} &= \left[-\frac{1}{\phi} + \frac{1}{\lambda_t} \frac{Z(\lambda_t, x_t + 1, \phi)}{Z(\lambda_t, x_t, \phi)} \right] \frac{\partial \lambda_t}{\partial \alpha_j}, \quad j = 0, 1, \end{aligned}$$

and the second derivatives of l_t are

$$\begin{aligned} \frac{\partial^2 l_t(\theta)}{\partial \phi^2} &= -\frac{2\lambda_t}{\phi^3} - \frac{x_t}{\phi^2} + \frac{Z(\lambda_t, x_t + 1, \phi)}{\phi^2 Z(\lambda_t, x_t, \phi)} + \left(\frac{\phi + 1}{\phi} \right)^2 \left[\frac{Z(\lambda_t, x_t + 2, \phi)}{Z(\lambda_t, x_t, \phi)} - \frac{Z^2(\lambda_t, x_t + 1, \phi)}{Z^2(\lambda_t, x_t, \phi)} \right], \\ \frac{\partial^2 l_t(\theta)}{\partial \phi \partial \alpha_j} &= \left[\frac{1}{\phi^2} - \frac{\phi + 1}{\phi \lambda_t} \left\{ \frac{Z(\lambda_t, x_t + 2, \phi)}{Z(\lambda_t, x_t, \phi)} - \frac{Z^2(\lambda_t, x_t + 1, \phi)}{Z^2(\lambda_t, x_t, \phi)} \right\} \right] \frac{\partial \lambda_t}{\partial \alpha_j}, \\ \frac{\partial^2 l_t(\theta)}{\partial \alpha_j \partial \alpha_k} &= \frac{1}{\lambda_t^2} \left[-\frac{Z(\lambda_t, x_t + 1, \phi)}{Z(\lambda_t, x_t, \phi)} + \frac{Z(\lambda_t, x_t + 2, \phi)}{Z(\lambda_t, x_t, \phi)} - \frac{Z^2(\lambda_t, x_t + 1, \phi)}{Z^2(\lambda_t, x_t, \phi)} \right] \frac{\partial \lambda_t}{\partial \alpha_j} \frac{\partial \lambda_t}{\partial \alpha_k}, \end{aligned}$$

for $0 \leq j, k \leq 1$, where the expressions for $\partial \lambda_t / \partial \alpha_j$ and $\partial^2 \lambda_t / \partial \alpha_j \partial \alpha_k$ are easily deduced.

Analogously, for the GEOMP2-INARCH(1) process we obtain the following expression:

$$\begin{aligned} \log L_n(\theta|\mathbf{x}) &= \sum_{t=2}^n l_t(\theta) \\ &= \sum_{t=2}^n \left\{ -\lambda_t + \log \left(I_{x_t=0} + \left[\sum_{n=1}^{x_t} \frac{\lambda_t^n}{n!} \binom{x_t-1}{n-1} (p^*)^n (1-p^*)^{x_t-n} \right] I_{x_t \neq 0} \right) \right\}, \end{aligned}$$

where $\theta = (\alpha_0, \alpha_1, p^*)$, as $v_0 = (2 - p^*)/p^*$ and taking into consideration that the conditional probability mass function of X_t is given by

$$\begin{aligned} P[X_t = 0 | \underline{X}_{t-1}] &= e^{-\lambda_t}, \\ P[X_t = x_t | \underline{X}_{t-1}] &= \sum_{n=1}^{x_t} e^{-\lambda_t} \frac{\lambda_t^n}{n!} \binom{x_t-1}{n-1} (p^*)^n (1-p^*)^{x_t-n}, \quad x_t = 1, 2, \dots \end{aligned}$$

Similarly to the previous case, the first and second derivatives of l_t in order to α_0, α_1 and p^* are deduced.

4. A SIMULATION STUDY

Some simulation studies are now developed to examine and compare the performance of the different estimators considered in Section 3 for the model parameters. We begin by illustrating the two-step approach based on CLS and moments estimation methods by computing the estimates and analyzing its performance. In the sequel, the several estimation procedures are discussed and compared. The study is developed considering the NTA-INARCH(1) and the GEOMP2-INARCH(1) models. So, after estimating α_0, α_1 and v_0 , we deduce the estimator of ϕ , in the first case, given by

$$\hat{\phi}_n = -1 - \hat{\alpha}_{0,n}(1 + \hat{\alpha}_{1,n}) + \frac{(1 - \hat{\alpha}_{1,n})(1 - \hat{\alpha}_{1,n}^2)}{n \hat{\alpha}_{0,n}} \sum_{t=1}^n X_t^2,$$

and, in the second one, that of p^* namely

$$\hat{p}_n^* = 2 \left[1 - \hat{\alpha}_{0,n}(1 + \hat{\alpha}_{1,n}) + \frac{(1 - \hat{\alpha}_{1,n})(1 - \hat{\alpha}_{1,n}^2)}{n \hat{\alpha}_{0,n}} \sum_{t=1}^n X_t^2 \right]^{-1}.$$

4.1. CLS estimators performance

4.1.1. NTA-INARCH(1) model

To illustrate the CLS method, we focus on the NTA-INARCH(1) model with true parameters $\alpha_0 = 2, \alpha_1 = 0.2$ and $\phi = 2$ and, for different sample sizes $n = 100, 250, 500, 750, 1000$, we present in Table 1 the expected values, variances and covariance of $\hat{\alpha}_{0,n}, \hat{\alpha}_{1,n}$ and $\hat{\phi}_n$,

considering 10 000 replications. In the last column of this table we present the true values of α_0 , α_1 and ϕ , as well as the entries of the asymptotic matrix $\mathbf{V}^{-1}\mathbf{W}\mathbf{V}^{-1}$, respectively b_{11} , b_{22} and b_{12} , given in Theorem 3.1. We verify that the asymptotic and the sample values are quite similar for large values of n .

Table 1: Means, variances and covariances for the CLS+M estimates of the NTA-INARCH(1) model with coefficients $\alpha_0 = 2$, $\alpha_1 = 0.2$, $\phi = 2$ and for different sample sizes n .

n	100	250	500	750	1 000	
$E_{\text{est}}(\widehat{\alpha}_0)$	2.0444	2.0161	2.0090	2.0090	2.0041	2
$E_{\text{est}}(\widehat{\alpha}_1)$	0.1797	0.1918	0.1956	0.1973	0.1981	0.2
$E_{\text{est}}(\widehat{\phi})$	1.9238	1.9670	1.9842	1.9899	1.9929	2
$n \cdot V_{\text{est}}(\widehat{\alpha}_0)$	12.2393	12.3125	12.3782	12.3133	12.3133	12.3774
$n \cdot V_{\text{est}}(\widehat{\alpha}_1)$	1.1793	1.1957	1.2227	1.2594	1.2776	1.2604
$n \cdot V_{\text{est}}(\widehat{\phi})$	21.9663	21.7000	21.3637	22.2183	22.1552	
$n \cdot \text{Cov}_{\text{est}}(\widehat{\alpha}_0, \widehat{\alpha}_1)$	-2.3311	-2.4081	-2.4814	-2.5270	-2.5911	-2.5510

Figure 1 displays a multiple boxplot for samples of length $n = 250$, 750 and 2000 of the CLS estimator of α_0 and α_1 based on 10 000 model replications as well as the histogram of the corresponding standardized values, for $n = 2000$, of a NTA-INARCH(1) model with $\alpha_0 = 2$, $\alpha_1 = 0.2$ and $\phi = 2$. These multiple boxplots show a significant stability and allow to infer a high rate of convergence to the limit distribution. In agreement with Theorem 3.1, the plots indicate the adequacy of the normal for the empirical marginal distributions of the estimators $\widehat{\alpha}_0$, $\widehat{\alpha}_1$. Let us observe that the Kolmogorov–Smirnov test for the sampling laws of the standardized CLS estimation gives large p -values for testing the standard normal distribution as, for instance, when we consider $n = 2000$ and 1000 replications we obtain 0.9454 and 0.4051.

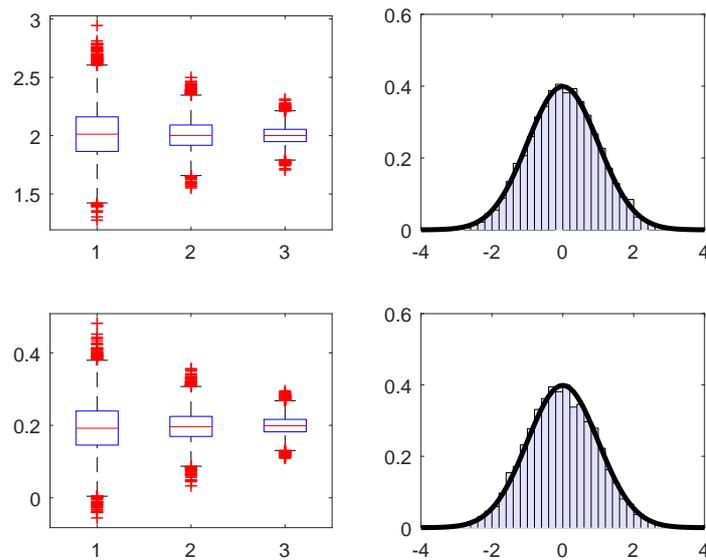


Figure 1: Boxplots for $n = 250$, 750 , 2000 (from left to right) and histogram for $n = 2000$ of the empirical law of $\widehat{\alpha}_0$ (on top) and $\widehat{\alpha}_1$ (below) for a NTA-INARCH(1) process with $\alpha_0 = 2$, $\alpha_1 = 0.2$ and $\phi = 2$. Superimposed is the standard normal density function. The results are based on 10 000 replications.

In Figure 2 we present now a multiple boxplot and the histogram of the distribution of $\sqrt{n}(\hat{\phi}_n - \phi)$. Figure 3 shows the similarity between the empirical cumulative distribution function of $\sqrt{n}(\hat{\phi}_n - \phi)$ (represented in solid line) and the cumulative distribution function of the normal(0, 4.7) law (in dashed line), whose parameters are the sample mean and variance of $\sqrt{n}(\hat{\phi}_n - \phi)$. The stability previously observed appears also here and, once again, the p -value of the Kolmogorov–Smirnov test, namely 0.8231 when $n = 2\,000$ and for 1 000 replications, indicates the adequacy of the normal for the empirical distribution of $\sqrt{n}(\hat{\phi}_n - \phi)$.

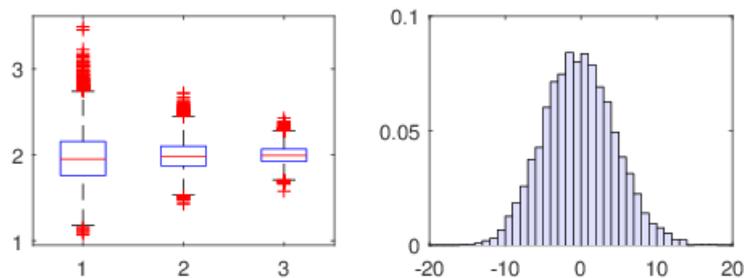


Figure 2: Boxplots for $n = 250, 750, 2\,000$ (from left to right) and histogram for $n = 2\,000$ of the empirical law of $\sqrt{n}(\hat{\phi}_n - \phi)$ when $\alpha_0 = 2, \alpha_1 = 0.2$ and $\phi = 2$ for a NTA-INARCH(1).

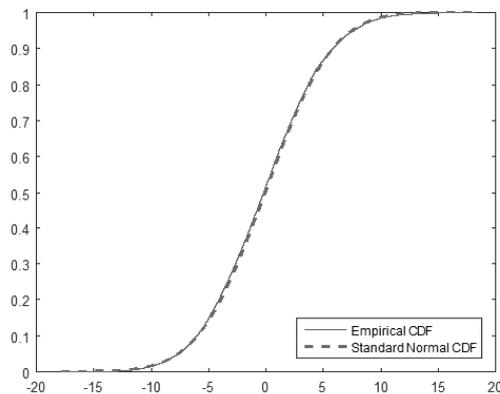


Figure 3: Empirical CDF of the law of $\sqrt{n}(\hat{\phi}_n - \phi)$ when $\alpha_0 = 2, \alpha_1 = 0.2$ and $\phi = 2$ for a NTA-INARCH(1) (in solid line) and the CDF of the normal(0, 4.7) law (in dashed line), for $n = 2\,000$.

From the empirical results presented in the two last lines of Table 2, we can presume that the estimators of α_0 (resp., α_1) and ϕ are asymptotically uncorrelated. In fact, for the NTA-INARCH(1) model in study, the empirical correlations $\rho_{\text{est}}(\hat{\alpha}_{0,n}, \hat{\phi}_n)$ and $\rho_{\text{est}}(\hat{\alpha}_{1,n}, \hat{\phi}_n)$ are significantly low. To support this statement we use the Monte Carlo method to determine confidence intervals for the mean of $\rho_{\text{est}}(\hat{\alpha}_{0,n}, \hat{\phi}_n)$ and for the mean of $\rho_{\text{est}}(\hat{\alpha}_{1,n}, \hat{\phi}_n)$ which we denote by $m_{0,n,\tilde{n}}$ and $m_{1,n,\tilde{n}}$, respectively. The confidence intervals are obtained considering $\tilde{n} = 35$ and $\tilde{n} = 50$ replications of n -dimensional samples ($n = 500$ and $n = 1\,000$) of a NTA-INARCH(1) model with $\alpha_0 = 2, \alpha_1 = 0.2$ and $\phi = 2$.

Table 2: Empirical correlations for the CLS+M estimates of the NTA-INARCH(1) model with coefficients $\alpha_0 = 2$, $\alpha_1 = 0.2$, $\phi = 2$ and for different sample sizes n .

n	250	750	1 000	5 000	10 000
$\rho_{\text{est}}(\hat{\alpha}_{0,n}, \hat{\alpha}_{1,n})$	-0.6276	-0.6417	-0.6385	-0.6482	-0.6402
$\rho_{\text{est}}(\hat{\alpha}_{0,n}, \hat{\phi}_n)$	0.0883	0.0962	0.1139	0.1059	0.0911
$\rho_{\text{est}}(\hat{\alpha}_{1,n}, \hat{\phi}_n)$	0.0272	0.0192	0.0078	0.0246	0.0438

Such intervals with confidence level 0.99 are presented in Table 3, where we stress the lower values when n or \tilde{n} increase. So we have estimated (α_0, α_1) and ϕ separately likely without loss of efficiency.

Table 3: Confidence intervals for the mean of $\rho_{\text{est}}(\hat{\alpha}_{0,n}, \hat{\phi}_n)$ and for the mean of $\rho_{\text{est}}(\hat{\alpha}_{1,n}, \hat{\phi}_n)$, with confidence level $\gamma = 0.99$ and for different sample sizes n and \tilde{n} .

	$\tilde{n} = 35$		$\tilde{n} = 50$	
	$n = 500$	$n = 1\,000$	$n = 500$	$n = 1\,000$
$m_{0,n,\tilde{n}}$	[0.0917, 0.1180]	[0.0883, 0.1162]	[0.0940, 0.1160]	[0.0814, 0.1064]
$m_{1,n,\tilde{n}}$	[0.0113, 0.0412]	[0.0165, 0.0412]	[0.0137, 0.0354]	[0.0132, 0.0397]

4.1.2. GEOMP2-INARCH(1) model

Let us consider now the GEOMP2-INARCH(1) model with true parameters $\alpha_0 = 2$, $\alpha_1 = 0.4$ and $p^* = 0.1$. As in the previous section, for different sample sizes n , we compute the expected values, variances and covariances of $\hat{\alpha}_{0,n}$, $\hat{\alpha}_{1,n}$ and \hat{p}_n^* (see Table 4, where in the last column we present the true values of α_0 , α_1 and p^* as well as the entries b_{11} , b_{22} and b_{12} of the asymptotic matrix $\mathbf{V}^{-1}\mathbf{W}\mathbf{V}^{-1}$) and for samples of length $n = 250, 750$ and $2\,000$ we plot a multiple boxplot and for $n = 2\,000$ the histograms for 10 000 values of the CLS+M estimators (in Figure 4) and similar conclusions to the previous case may be deduced.

Table 4: Expected values, variances and covariances for the CLS+M estimates of the GEOMP2-INARCH(1) model with $\alpha_0 = 2$, $\alpha_1 = 0.4$, $p^* = 0.1$ and different sample sizes n .

n	100	250	500	750	1 000	
$E_{\text{est}}(\hat{\alpha}_0)$	2.1401	2.0705	2.0381	2.0265	2.0241	2
$E_{\text{est}}(\hat{\alpha}_1)$	0.3267	0.3655	0.3803	0.3875	0.3900	0.4
$E_{\text{est}}(\hat{p}^*)$	0.1171	0.1068	0.1038	0.1025	0.1019	0.1
$n \cdot V_{\text{est}}(\hat{\alpha}_0)$	54.8720	54.9255	57.1036	57.6511	58.7167	61.5325
$n \cdot V_{\text{est}}(\hat{\alpha}_1)$	2.7975	3.2809	3.6923	3.8768	3.9021	4.3979
$n \cdot V_{\text{est}}(\hat{p}^*)$	0.2011	0.0879	0.0867	0.0884	0.0886	
$n \cdot \text{Cov}_{\text{est}}(\hat{\alpha}_0, \hat{\alpha}_1)$	-1.4509	-3.4576	-4.8393	-5.3491	-5.5056	-7.0598

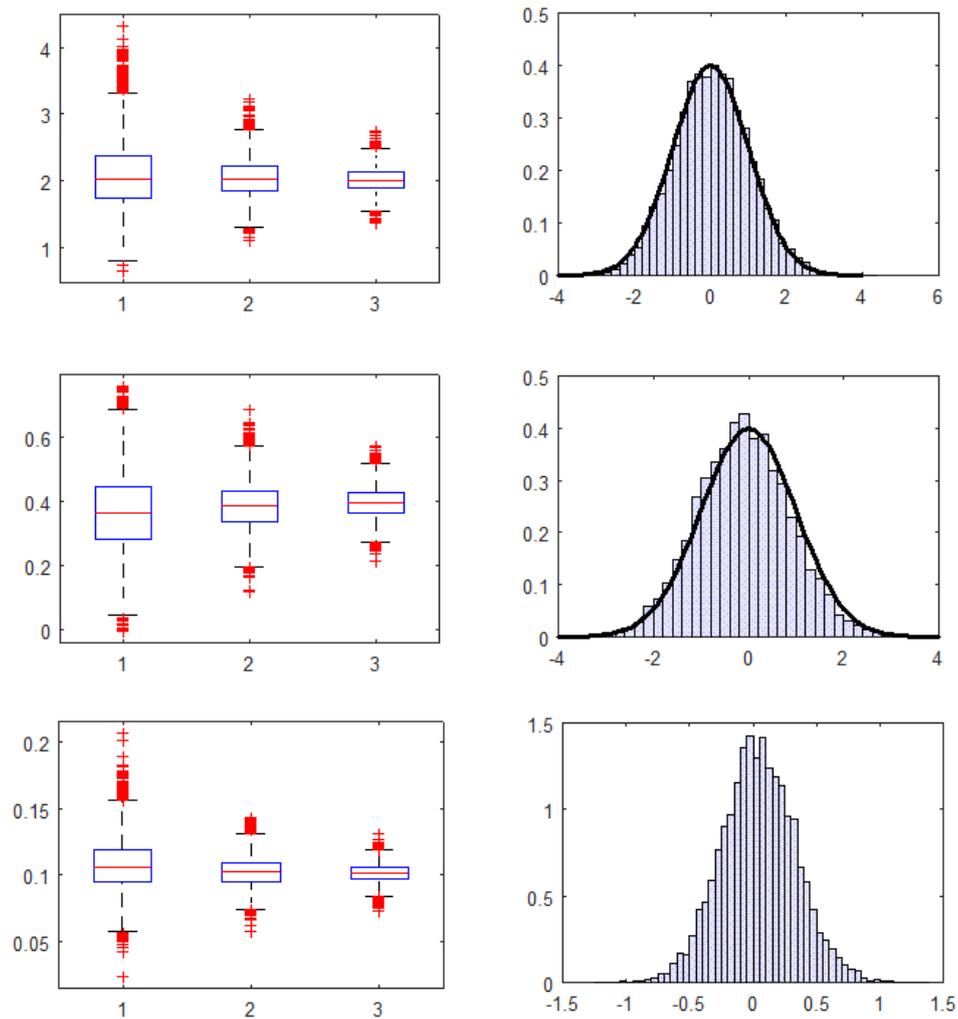


Figure 4: Boxplots for $n = 250, 750, 2000$ (from left to right) and histogram for $n = 2000$ of the empirical law of $\hat{\alpha}_0$ (on top), $\hat{\alpha}_1$ (in the middle) and \hat{p}^* (below) when $\alpha_0 = 2$, $\alpha_1 = 0.4$ and $p^* = 0.1$ for a GEOMP2-INARCH(1) process. Superimposed is the standard normal density function. The results are based on 10 000 replications.

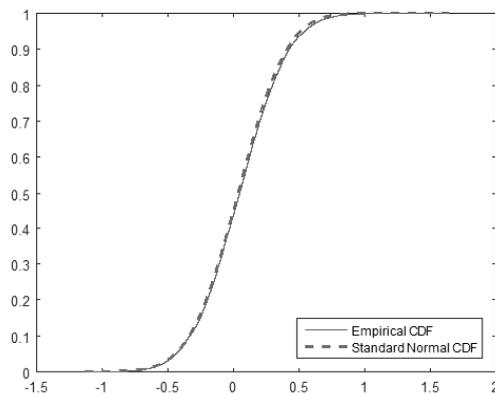


Figure 5: Empirical CDF of the law of $\sqrt{n}(\hat{p}_n^* - p^*)$ when $\alpha_0 = 2$, $\alpha_1 = 0.4$ and $p^* = 0.1$ for a GEOMP2-INARCH(1) model (in solid line) and the CDF of the normal(0, 0.3) law (in dashed line).

To show the adequacy of the normal for the empirical distribution of $\sqrt{n}(\hat{p}_n^* - p^*)$, in Figure 5 we present the empirical cumulative distribution function of $\sqrt{n}(\hat{p}_n^* - p^*)$ (represented in solid line) and the cumulative distribution function of the normal(0, 0.3) law (in dashed line). Analogously to the previous study, we can also presume that the estimators of α_0 , (resp., α_1) and p^* are asymptotically uncorrelated.

4.2. Comparative analysis of the estimation procedures

To examine and compare the finite sample performances of the CLS+M, PQML+M and CML methods, we consider two different NTA-INARCH(1) models with parameter values $\alpha_0 = 2, \alpha_1 = 0.2, \phi = 2$ and $\alpha_0 = 5, \alpha_1 = 0.3, \phi = 1$, and two different GEOMP2-INARCH(1) models with parameter values $\alpha_0 = 2, \alpha_1 = 0.2, p^* = 0.1$ and $\alpha_0 = 5, \alpha_1 = 0.3, p^* = 0.6$. The sample sizes considered are $n = 500$ and $1\,000$ and the number of replications $m = 10\,000$.

For the maximization of the log-likelihood functions, we use the MATLAB function `fmincon` where the estimates obtained using the CLS+M method were used as the initial values and the constrained conditions are $\alpha_0 > 0, 0 < \alpha_1 < 1, \phi > 0$ (for the NTA) and $0 < p^* < 1$ (for the GEOMP2). The performance of the estimators is evaluated by the mean square error, i.e.,

$$\frac{1}{m} \sum_{k=1}^m (\hat{\theta}_{j,k} - \theta_j)^2, \quad j = 1, 2, 3.$$

The results of the simulation experiments are presented in Tables 5 and 6 where the smallest values of the mean square errors are highlighted in italics.

Table 5: Mean estimates (in bold) and mean square errors (within parentheses) for the NTA-INARCH(1) model with different sample sizes n .

n	Method	$\alpha_0 = 2$	$\alpha_1 = 0.2$	$\phi = 2$	$\alpha_0 = 5$	$\alpha_1 = 0.3$	$\phi = 1$
500	CLS+M	2.0071 (0.0248)	0.1967 (0.0025)	1.9832 (0.0458)	5.0288 (0.1169)	0.2956 (0.0021)	0.9915 (0.0180)
	PQML+M	2.0061 (0.0239)	0.1971 (0.0023)	1.9831 (0.0459)	5.0259 (0.1123)	0.2960 <i>(0.0020)</i>	0.9912 (0.0181)
	CML	2.0047 <i>(0.0233)</i>	0.1977 <i>(0.0022)</i>	1.9937 <i>(0.0174)</i>	5.0249 <i>(0.1115)</i>	0.2961 <i>(0.0020)</i>	0.9928 <i>(0.0141)</i>
1 000	CLS+M	2.0023 (0.0124)	0.1982 (0.0013)	1.9906 (0.0219)	5.0117 (0.0582)	0.2979 <i>(0.0010)</i>	0.9946 (0.0089)
	PQML+M	2.0020 (0.0120)	0.1983 (0.0012)	1.9907 (0.0221)	5.0103 (0.0558)	0.2981 <i>(0.0010)</i>	0.9945 (0.0090)
	CML	2.0017 <i>(0.0116)</i>	0.1985 <i>(0.0011)</i>	1.9960 <i>(0.0085)</i>	5.0105 <i>(0.0552)</i>	0.2981 <i>(0.0010)</i>	0.9948 <i>(0.0072)</i>

From this study we may conclude that the three methods seem to perform quite well, although the CML gives slightly smaller mean square errors in most cases.

Table 6: Mean estimates (in bold) and mean square errors (within parentheses) for the GEOMP2-INARCH(1) model with different sample sizes n .

n	Method	$\alpha_0 = 2$	$\alpha_1 = 0.2$	$p^* = 0.1$	$\alpha_0 = 5$	$\alpha_1 = 0.3$	$p^* = 0.6$
500	CLS+M	2.0142 (0.0964)	0.1898 (0.0058)	0.1035 (0.0002)	5.0269 (0.1219)	0.2963 (0.0021)	0.6033 (0.0009)
	PQML+M	2.0070 (0.0913)	0.1926 (0.0052)	0.1036 (0.0002)	5.0250 (0.1173)	0.2966 (0.0020)	0.6033 (0.0009)
	CML	1.9967 (0.0807)	0.1968 (0.0036)	0.1013 (0.0001)	5.0240 (0.1141)	0.2967 (0.0020)	0.6027 (0.0007)
1000	CLS+M	2.0072 (0.0481)	0.1959 (0.0030)	0.1017 (0.0001)	5.0100 (0.0600)	0.2985 (0.0011)	0.6020 (0.0004)
	PQML+M	2.0032 (0.0450)	0.1975 (0.0026)	0.1018 (0.0001)	5.0084 (0.0578)	0.2988 (0.0010)	0.6020 (0.0004)
	CML	1.9995 (0.0397)	0.1989 (0.0018)	0.1006 (0.0000)	5.0080 (0.0566)	0.2988 (0.0010)	0.6016 (0.0003)

5. REAL DATA EXAMPLE — COUNTS OF DIFFERENCES IN THE PRICES OF ELECTRICITY IN PORTUGAL AND SPAIN

OMIE (<http://www.omie.es>) is the company that manages the wholesale electricity market on the Iberian Peninsula. Electricity prices in Europe are set on a daily basis (every day of the year) at 12 noon, for the twenty-four hours of the following day, known as daily market. The market splitting is the mechanism used for setting the price of electricity on the daily market. When the price of electricity is the same in Portugal and Spain, which corresponds to the desired situation, it means that the integration of the Iberian market is working properly.

In the following, we consider the time series that represents the number of hours in a day in which the prices of electricity for Portugal and Spain are different. The data presented in Figure 6 consists of 365 observations, starting from July 2016 and ending in June 2017.

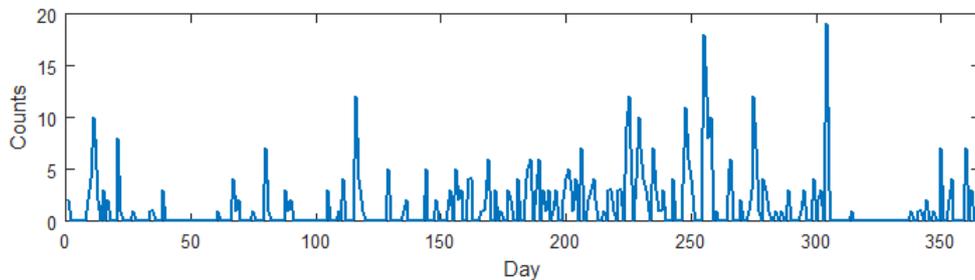


Figure 6: Daily number of hours in which the price of electricity of Portugal and Spain are different, starting from July 2016 and ending in June 2017.

Empirical mean and variance of the data are 1.4082 and 7.3027, respectively, indicating that the true marginal distribution is overdispersed. Let us observe that this time series exhibits also volatility clusters suggesting characteristics of conditional heteroscedasticity.

The partial autocorrelation function presented in Figure 7, suggests an order 1 dependence and so a CP-INARCH(1) model may be a reasonable choice to fit the data within the CP-INGARCH class. Despite the support bounding of this variable, the empirical analysis of the data set observed allows us to infer that its distributional characteristics (see histogram in Figure 7) are compatible with some compound Poisson laws.

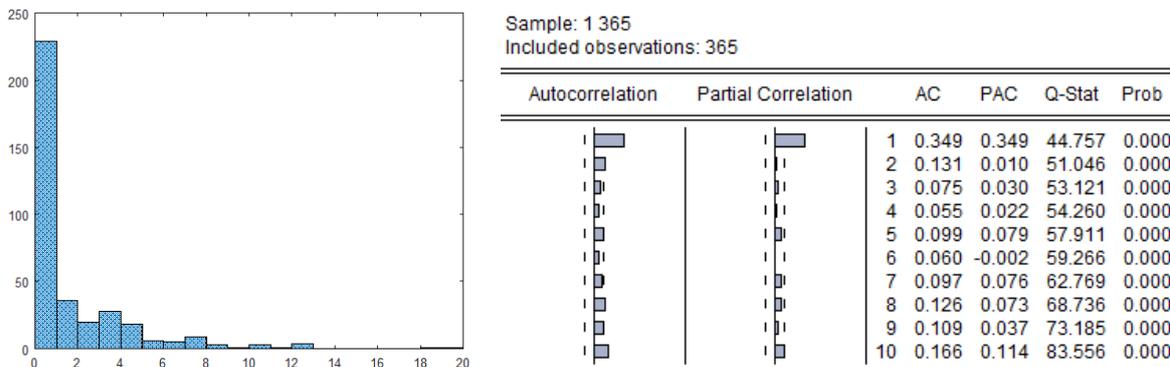


Figure 7: Sample histogram, autocorrelations and partial autocorrelations.

Trying to obtain a suitable model for this count time series, we present a comparative study between five CP-INARCH(1) processes, namely those associated to the Poisson ([4]), the generalized Poisson ([15]), the Neyman type-A, the geometric Poisson and the negative binomial ([13]) laws. Considering the slightly better performance observed in Section 3 for the CML estimator, we begin by using this methodology to estimate the models parameters and take a decision on the model fitting. The results, obtained with the help of MATLAB software, are displayed in Table 7. So, based on the values of the log likelihood function, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), we conclude that the GEOMP2-INARCH(1) model gives better fit than the other CP-INARCH(1) models considered. The NB2 model follows closely and the Poisson model shows the worst adequacy.

Table 7: CML parameters estimates for several CP-INARCH(1) models. Standard errors are shown in parentheses. The best values of the criteria $-\text{Log L}$, AIC and BIC are emphasised in italics.

	Model	$\hat{\alpha}_{0,365}$	$\hat{\alpha}_{1,365}$	Additional parameter	$-\text{Log L}$	AIC	BIC
INARCH(1)	Poisson	0.9751 (0.0008)	0.3055 (0.0018)		786.3	1576.5	1584.3
	GP	0.8971 (0.0012)	0.3608 (0.0018)	$\hat{\kappa}_{365} = 0.3736$ (0.0073)	524.6	1055.2	1066.9
	NTA	0.9558 (0.0051)	0.3192 (0.0125)	$\hat{\phi}_{365} = 2.4368$ (0.0502)	524.7	1055.4	1067.1
	GEOMP2	0.9338 (0.0060)	0.3349 (0.0022)	$\hat{p}_{365}^* = 0.3599$ (0.0024)	<i>516.2</i>	<i>1038.4</i>	<i>1050.1</i>
	NB2	0.9129 (0.0078)	0.3496 (0.0031)	$\hat{\beta}_{365} = 5.5659$ (0.0968)	519.8	1045.6	1057.3

The mean, variance and the first-order autocorrelation coefficient (FOAC) for the fitted CP-INARCH(1) models are summarized in Table 8. The results are in accordance with the previous conclusion as, although the similarity of the mean values, the variance and FOAC values point to a GEOMP2 or NB2-INARCH(1) choice. The two other methodologies are also considered to estimate the previous models and it should be noted in Table 9 the close proximity between each of the three parameters and those obtained by the CML method in the case of the GEOMP2 and also NB2 models. This conclusion is validated by the values referred in Table 10 for the sample and estimated means, variances and FOAC values under the two methods, particularly for the CLS+M one. Thus these methodologies seem to capture the same models as the powerful but distribution-demanding CML approach, which is in line with the previous conclusions of the simulation study.

Table 8: Sample and estimated means, variances and FOACs under CP-INARCH(1) models.

Method	Model	Sample	Poisson	GP	NTA	GEOMP2	NB2
CML	Mean	1.4082	1.4040	1.4034	1.4039	1.4040	1.4036
	Variance	7.3027	1.5485	4.1125	5.3723	7.2064	8.9001
	FOAC	0.349	0.3055	0.3608	0.3192	0.3349	0.3496

Table 9: Estimated parameters of several CP-INARCH(1) models based on CLS+M and PQML+M approaches. Standard errors are shown in parentheses. (a) means all models.

Method	Model	Additional parameter		
CLS+M	(a)	$\hat{\alpha}_{0,365} = 0.9138$	$\hat{\alpha}_{1,365} = 0.3490$	—
	Poisson	(0.0862)	(0.0564)	—
	GP	(0.1493)	(0.0928)	0.5319
	NTA	(0.1337)	(0.0799)	3.5645
	GEOMP2	(0.1392)	(0.0845)	0.3594
	NB2	(0.1445)	(0.0889)	4.5645
PQML+M	(a)	$\hat{\alpha}_{0,365} = 0.9751$	$\hat{\alpha}_{1,365} = 0.3055$	—
	GP	(0.0008)	(0.0018)	0.5392
	NTA	(0.0008)	(0.0018)	3.7096
	GEOMP2	(0.0008)	(0.0018)	0.3503
	NB2	(0.0008)	(0.0018)	4.7096

Table 10: Sample and estimated means, variances and FOACs under CP-INARCH(1) models.

Method	Model	Sample	GEOMP2	NB2
CLS+M	Mean	1.4082	1.4037	1.4037
	Variance	7.3027	7.2964	7.2958
	FOAC	0.349	0.3490	0.3490
PQML+M	Mean	1.4082	1.4040	1.4040
	Variance	7.3027	7.2926	7.2929
	FOAC	0.349	0.3055	0.3055

The statistical study that was developed in this Section was naturally circumscribed to the class of CP-INARCH(1) models considered here. However, this observed time series has characteristics that can also be taken into account if the adjustment is done in other classes of models, namely, in view of its histogram, the zero-inflated CP-INGARCH models ([7]).

6. CONCLUSION

The class of integer-valued GARCH models, specified through the characteristic function of the compound Poisson law and denoted CP-INGARCH ([6]) unifies and enlarges substantially the family of conditionally heteroscedastic integer-valued processes. With this new class, we may capture simultaneously different kinds of conditional volatility and the overdispersion characteristic often recorded in real count data. The probabilistic analysis of these models, concerning stationarity and ergodicity properties as well as moments studies, was the goal of previous works among which we may refer those established in [5] and [6]. The aim of this paper is to develop some statistical studies, regarding the parametric estimation of the CP-INARCH models, that allow the use of this general class with real data and show its true practical usefulness. We concentrate our study on the CP-INARCH models of order one, and a two-step estimation methodology, involving the conditional least squares or the Poisson quasi-maximum likelihood methods in a first step, and the moment's estimation method in the second one, has been introduced and developed. We point out the great advantage of this procedure regarding the more classical conditional maximum likelihood one, as its application is independent from the specific conditional distribution of the process. In fact, the simulation study presented allows concluding that the two-step methodology performance is strongly competitive with that of the conditional maximum likelihood estimation. We should also stress that the practical relevance of this wide class is clearly shown with the real-data example presented which illustrates the better quality of the fitting performed by new models emerged from that class.

Future developments of the present study should concern, particularly, the establishment of the conjectured Gaussian asymptotic distribution of the additional parameter estimator. The development of the parametric estimation of a more general CP-INGARCH model should also be considered.

A. APPENDIX — Proof of Theorem 2.1

To establish the results present in Theorem 2.1 let us begin by recalling the expression of the following conditional moments:

$$\begin{aligned}
 E(X_t | \underline{X}_{t-1}) &= \lambda_t = \alpha_0 + \alpha_1 X_{t-1}, \\
 \text{(A.1)} \quad E(X_t^2 | \underline{X}_{t-1}) &= v_0 \lambda_t + \lambda_t^2 = \alpha_1^2 X_{t-1}^2 + \alpha_1(2\alpha_0 + v_0) X_{t-1} + \alpha_0(\alpha_0 + v_0),
 \end{aligned}$$

$$\begin{aligned}
 \text{(A.2)} \quad E(X_t^3 | \underline{X}_{t-1}) &= i \Phi'''_{X_t | \underline{X}_{t-1}}(0) \\
 &= d_0 \lambda_t + 3v_0 \lambda_t^2 + \lambda_t^3 \\
 &= \alpha_1^3 X_{t-1}^3 + 3\alpha_1^2(v_0 + \alpha_0) X_{t-1}^2 + \alpha_1(3\alpha_0^2 + 6v_0\alpha_0 + d_0) X_{t-1} \\
 &\quad + \alpha_0(d_0 + 3v_0\alpha_0 + \alpha_0^2).
 \end{aligned}$$

(a) Using the fact that for $k \geq 0$, $\Gamma(k) = \alpha_1^k f_2$, we get

$$\text{(A.3)} \quad \mu(k) = E(X_t X_{t+k}) = \text{Cov}(X_t, X_{t+k}) + E(X_t)^2 = f_2 \left(v_0 \alpha_1^k + \alpha_0(1 + \alpha_1) \right).$$

(b) To derive $\mu(k, l)$, $0 \leq k \leq l$, we distinguish the following three cases:

Case 1: $l > k$. We have

$$\begin{aligned}
 \mu(k, l) &= E(X_t X_{t+k} X_{t+l}) \\
 &= E \left[X_t X_{t+k} E(X_{t+l} | \underline{X}_{t+l-1}) \right] \\
 &= \alpha_0 E(X_t X_{t+k}) + \alpha_1 E(X_t X_{t+k} X_{t+l-1}) \\
 &= \alpha_0 \mu(k) + \alpha_1 \mu(k, l-1) \\
 &= \alpha_0 \mu(k) + \alpha_1 \left[\alpha_0 \mu(k) + \alpha_1 \mu(k, l-2) \right] \\
 &= \dots \\
 &= \alpha_1^{l-k} \left[\mu(k, k) - f_1 \mu(k) \right] + f_1 \mu(k).
 \end{aligned}$$

Case 2: $l = k > 0$. We have

$$\begin{aligned}
 \mu(k, k) &= E \left[X_t E(X_{t+k}^2 | \underline{X}_{t+k-1}) \right] \\
 &= \alpha_1^2 E(X_t X_{t+k-1}^2) + \alpha_1(2\alpha_0 + v_0) E(X_t X_{t+k-1}) + \alpha_0(\alpha_0 + v_0) E(X_t) \\
 &= \alpha_1^2 \mu(k-1, k-1) + \alpha_1(2\alpha_0 + v_0) \mu(k-1) + \alpha_0(\alpha_0 + v_0) f_1 \\
 &= \dots \\
 &= \alpha_1^{2k} \left[\mu(0, 0) - \frac{v_0(2\alpha_0 + v_0) f_2}{1 - \alpha_1} - f_1 \mu(0) \right] + \frac{v_0(2\alpha_0 + v_0) f_2 \alpha_1^k}{1 - \alpha_1} + f_1 \mu(0).
 \end{aligned}$$

Case 3: $l = k = 0$. According to the relations between the moments and the cumulants (e.g., formula (15.10.4) in [3, p. 186]) and Theorem 4.2 of [7], we have

$$\begin{aligned}
 \mu(0, 0) &= E(X_t^3) \\
 &= \kappa_3 + 3\kappa_2 \mu + \mu^3 \\
 &= f_3 \left[d_0(1 - \alpha_1^2) + 3v_0^2 \alpha_1^2 \right] + 3v_0 f_2 f_1 + f_1^3 \\
 &= \left[d_0(1 - \alpha_1^2) + 3v_0^2 \alpha_1^2 \right] f_3 + \frac{2\alpha_0 v_0}{1 - \alpha_1} f_2 + f_1 \mu(0),
 \end{aligned}$$

since $f_1 = (1 - \alpha_1^2)f_2$. So the above formula for $\mu(k, k)$ simplifies to

$$\begin{aligned}\mu(k, k) &= \alpha_1^{2k} \left[\left[d_0(1 - \alpha_1^2) + 3v_0^2\alpha_1^2 \right] f_3 - \frac{v_0^2}{1 - \alpha_1} f_2 \right] + \frac{v_0(2\alpha_0 + v_0)}{1 - \alpha_1} f_2 \alpha_1^k + f_1 \mu(0) \\ &= \alpha_1^{2k} f_3 \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] + \frac{v_0(2\alpha_0 + v_0)}{1 - \alpha_1} f_2 \alpha_1^k + f_1 \mu(0),\end{aligned}$$

which also holds for $k = 0$. Replacing this expression in $\mu(k, l)$ above, it follows that

$$\begin{aligned}\mu(k, l) &= \alpha_1^{l-k} \left[\left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] f_3 \alpha_1^{2k} + \frac{v_0(2\alpha_0 + v_0)}{1 - \alpha_1} f_2 \alpha_1^k \right. \\ &\quad \left. + f_1 \mu(0) - f_1 \mu(k) \right] + f_1 \mu(k).\end{aligned}$$

As

$$f_1 \mu(0) - f_1 \mu(k) = v_0 f_1 f_2 - \frac{v_0 \alpha_0}{1 - \alpha_1} f_2 \alpha_1^k,$$

we finally obtain, for any $0 \leq k \leq l$,

$$\begin{aligned}\mu(k, l) &= \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] f_3 \alpha_1^{l+k} + \frac{v_0(\alpha_0 + v_0)}{1 - \alpha_1} f_2 \alpha_1^l \\ &\quad + v_0 f_1 f_2 \alpha_1^{l-k} + f_1 \mu(k).\end{aligned}$$

- (c) In what concerns the fourth-order moments $\mu(k, l, m)$ with $0 \leq k \leq l \leq m$, we proceed in a similar way as above and distinguish the following four cases:

Case 1: $m > l$. As above we have

$$\begin{aligned}\mu(k, l, m) &= E(X_t X_{t+k} X_{t+l} X_{t+m}) \\ &= \alpha_1^{m-l} \left[\mu(k, l, l) - f_1 \mu(k, l) \right] + f_1 \mu(k, l).\end{aligned}$$

Case 2: $m = l > k$. For this case, using formula (A.1), we obtain

$$\begin{aligned}\mu(k, l, l) &= E \left[X_t X_{t+k} E(X_{t+l}^2 | \underline{X}_{t+l-1}) \right] \\ &= \alpha_1^2 \mu(k, l-1, l-1) + \alpha_1(v_0 + 2\alpha_0) \mu(k, l-1) + \alpha_0(v_0 + \alpha_0) \mu(k).\end{aligned}$$

Replacing $\mu(k, l-1)$, using $\mu(0) = (v_0 + \alpha_0(1 + \alpha_1))f_2$ and replacing $\mu(k)$, we obtain

$$\begin{aligned}\mu(k, l, l) &= \alpha_1^{2(l-k)} \mu(k, k, k) + \mu(k) \mu(0) \\ &\quad - f_2 v_0 \left[f_2 (v_0 + \alpha_0(1 + \alpha_1)) + \frac{(v_0 + 2\alpha_0)(v_0 + \alpha_0)}{(1 - \alpha_1)^2} \right] \alpha_1^{2l-k} \\ &\quad - f_1 \left[f_1 \mu(0) + \frac{v_0(v_0 + 2\alpha_0)}{1 - \alpha_1} f_2 \right] \alpha_1^{2(l-k)} + \frac{v_0 + 2\alpha_0}{1 - \alpha_1} \left[\mu(k, l) - f_1 \mu(k) \right] \\ &\quad - \frac{v_0 + 2\alpha_0}{1 - \alpha_1} \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] f_3 \alpha_1^{2l}.\end{aligned}$$

So, replacing $\mu(0)$, recalling $\mu(0, 0)$ and taking into account that $\frac{f_1}{1-\alpha_1} = (1 + \alpha_1)f_2$, we get

$$\begin{aligned}
 \mu(k, l, l) &= \alpha_1^{2(l-k)} \mu(k, k, k) - \mu(k) f_2 [\alpha_0 + (v_0 + \alpha_0)\alpha_1] \\
 &\quad - \frac{f_2 v_0}{(1 - \alpha_1)(1 - \alpha_1^2)} [v_0^2(1 + \alpha_1) + v_0\alpha_0(4 + 3\alpha_1) + 3\alpha_0^2(1 + \alpha_1)] \alpha_1^{2l-k} \\
 &\quad - f_1 \left\{ \mu(0, 0) - [d_0(1 - \alpha_1^2) + 3v_0^2\alpha_1^2] f_3 + \frac{v_0^2 f_2}{1 - \alpha_1} \right\} \alpha_1^{2(l-k)} \\
 \text{(A.4)} \quad &\quad + \frac{v_0 + 2\alpha_0}{1 - \alpha_1} \mu(k, l) - \frac{v_0 + 2\alpha_0}{1 - \alpha_1} [d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2)] f_3 \alpha_1^{2l}.
 \end{aligned}$$

Case 3: $m = l = k > 0$. From formula (A.2) we have

$$\begin{aligned}
 \mu(k, k, k) &= E[X_t E(X_{t+k}^3 | \underline{X}_{t+k-1})] \\
 &= \alpha_1^3 \mu(k - 1, k - 1, k - 1) + 3\alpha_1^2 (v_0 + \alpha_0) \mu(k - 1, k - 1) \\
 &\quad + \alpha_1 (d_0 + 6v_0\alpha_0 + 3\alpha_0^2) \mu(k - 1) + \alpha_0 (d_0 + 3v_0\alpha_0 + \alpha_0^2) \mu.
 \end{aligned}$$

Replacing $\mu(k - 1, k - 1)$ and thereafter $\mu(k - 1)$, we deduce

$$\begin{aligned}
 \mu(k, k, k) &= \alpha_1^3 \mu(k - 1, k - 1, k - 1) \\
 &\quad + 3(v_0 + \alpha_0) [d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2)] f_3 \alpha_1^{2k} \\
 &\quad + \frac{v_0 f_2}{1 - \alpha_1} [3\alpha_1(v_0 + \alpha_0)^2 + 3\alpha_1(v_0 + \alpha_0)\alpha_0 + (d_0 + 6v_0\alpha_0 + 3\alpha_0^2)(1 - \alpha_1)] \alpha_1^k \\
 &\quad + f_1 f_2 \left\{ 3\alpha_1^2 (v_0 + \alpha_0) (v_0 + \alpha_0(1 + \alpha_1)) + (d_0 + 6v_0\alpha_0 + 3\alpha_0^2) \alpha_1 (1 - \alpha_1) (1 + \alpha_1) \right. \\
 &\quad \left. + (d_0 + 3v_0\alpha_0 + \alpha_0^2) (1 - \alpha_1) (1 - \alpha_1^2) \right\}.
 \end{aligned}$$

Making some calculations and then recalling the expression of $\mu(0, 0)$, we obtain

$$\begin{aligned}
 \mu(k, k, k) &= \alpha_1^3 \mu(k - 1, k - 1, k - 1) \\
 &\quad + 3(v_0 + \alpha_0) [d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2)] f_3 \alpha_1^{2k} \\
 &\quad + \frac{v_0 f_2}{1 - \alpha_1} [3\alpha_0^2(1 + \alpha_1) + 3v_0\alpha_0(2 + \alpha_1) + d_0(1 - \alpha_1) + 3v_0^2\alpha_1] \alpha_1^k \\
 &\quad + f_1 (1 - \alpha_1^3) \mu(0, 0).
 \end{aligned}$$

Replacing successively the expression of $\mu(k - j, k - j, k - j)$, $j = 1, \dots, k - 1$, it remains

$$\begin{aligned}
 \mu(k, k, k) &= \alpha_1^{3k} \left\{ \mu(0, 0, 0) - 3(v_0 + \alpha_0) [d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2)] \frac{f_3}{1 - \alpha_1} \right. \\
 &\quad \left. - \frac{v_0 f_2}{(1 - \alpha_1)(1 - \alpha_1^2)} [3\alpha_0^2(1 + \alpha_1) + 3v_0\alpha_0(2 + \alpha_1) + d_0(1 - \alpha_1) + 3v_0^2\alpha_1] - f_1 \mu(0, 0) \right\} \\
 &\quad + \frac{3(v_0 + \alpha_0) f_3 \alpha_1^{2k}}{1 - \alpha_1} [d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2)] \\
 \text{(A.5)} \quad &\quad + \frac{v_0 f_2 \alpha_1^k}{(1 - \alpha_1)(1 - \alpha_1^2)} [3\alpha_0^2(1 + \alpha_1) + 3v_0\alpha_0(2 + \alpha_1) + d_0(1 - \alpha_1) + 3v_0^2\alpha_1] + f_1 \mu(0, 0).
 \end{aligned}$$

Replacing $\mu(0, 0)$, highlighting $\frac{f_3}{1-\alpha_1^2}$, noting that $f_2 = (1 - \alpha_1^3)f_3$ and $\frac{f_3}{1-\alpha_1^2} = f_4(1 + \alpha_1^2)$ and developing the calculations, we finally get

$$\begin{aligned}
 \mu(k, k, k) &= \left\{ \mu(0, 0, 0) - f_4 \left[4v_0d_0 - 3v_0^3 + 3v_0(d_0 - v_0^2)\alpha_1 + v_0(3v_0^2 + d_0)\alpha_1^2 \right. \right. \\
 &\quad + v_0(6v_0^2 - d_0)\alpha_1^3 + 3v_0(2v_0^2 - d_0)\alpha_1^4 + v_0(9v_0^2 - 4d_0)\alpha_1^5 \\
 &\quad + \alpha_0(1 + \alpha_1^2) \left[3v_0^2 + 4d_0 + (3v_0^2 + 4d_0)\alpha_1 + (15v_0^2 - 4d_0)\alpha_1^2 + (12v_0^2 - 4d_0)\alpha_1^3 \right] \\
 &\quad \left. \left. + 6v_0\alpha_0^2(1 + \alpha_1^2)(1 + \alpha_1)(1 + \alpha_1 + \alpha_1^2) + \alpha_0^3(1 + \alpha_1^2)(1 + \alpha_1)^2(1 + \alpha_1 + \alpha_1^2) \right] \right\} \alpha_1^{3k} \\
 &\quad + 3 \frac{v_0 + \alpha_0}{1 - \alpha_1} f_3 \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] \alpha_1^{2k} + f_1 \mu(0, 0) \\
 \text{(A.6)} \quad &\quad + \frac{v_0}{(1 - \alpha_1)(1 - \alpha_1^2)} f_2 \left[3\alpha_0^2(1 + \alpha_1) + 3v_0\alpha_0(2 + \alpha_1) + d_0(1 - \alpha_1) + 3v_0^2\alpha_1 \right] \alpha_1^k.
 \end{aligned}$$

Case 4: $m = l = k = 0$. Once again, according to the relations between the moments and the cumulants, we obtain

$$\begin{aligned}
 \mu(0, 0, 0) &= E(X_t^4) \\
 &= \kappa_4 + 3\kappa_2^2 + 6\kappa_2\mu^2 + 4\kappa_3\mu + \mu^4 \\
 &= f_4 \left\{ c_0 + (3v_0^3 + 4v_0d_0 - c_0)\alpha_1^2 + (6v_0d_0 - c_0)\alpha_1^3 + (15v_0^3 - 10v_0d_0 + c_0)\alpha_1^5 \right. \\
 &\quad + \alpha_0(1 + \alpha_1^2) \left[3v_0^2 + 4d_0 + (3v_0^2 + 4d_0)\alpha_1 + (15v_0^2 - 4d_0)\alpha_1^2 + (12v_0^2 - 4d_0)\alpha_1^3 \right] \\
 &\quad \left. + 6v_0\alpha_0^2(1 + \alpha_1)(1 + \alpha_1^2)(1 + \alpha_1 + \alpha_1^2) + \alpha_0^3(1 + \alpha_1)^2(1 + \alpha_1^2)(1 + \alpha_1 + \alpha_1^2) \right\}.
 \end{aligned}$$

So the formula (A.6) for $\mu(k, k, k)$ studied in case 3 simplifies to

$$\begin{aligned}
 \mu(k, k, k) &= f_4 \left\{ c_0 - 4v_0d_0 + 3v_0^3 + 3v_0(v_0^2 - d_0)\alpha_1 + (3v_0d_0 - c_0)\alpha_1^2 \right. \\
 &\quad + (7v_0d_0 - 6v_0^3 - c_0)\alpha_1^3 + 3v_0(d_0 - 2v_0^2)\alpha_1^4 + (6v_0^3 - 6v_0d_0 + c_0)\alpha_1^5 \left. \right\} \alpha_1^{3k} \\
 &\quad + 3 \frac{v_0 + \alpha_0}{1 - \alpha_1} f_3 \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] \alpha_1^{2k} + f_1 \mu(0, 0) \\
 &\quad + \frac{v_0}{(1 - \alpha_1)(1 - \alpha_1^2)} f_2 \left[3\alpha_0^2(1 + \alpha_1) + 3v_0\alpha_0(2 + \alpha_1) + d_0(1 - \alpha_1) + 3v_0^2\alpha_1 \right] \alpha_1^k.
 \end{aligned}$$

Inserting into the formula (A.4) for $\mu(k, l, l)$ stated in case 2, we obtain

$$\begin{aligned}
 \mu(k, l, l) &= f_4 \left\{ c_0 - 4v_0d_0 + 3v_0^3 + 3v_0(v_0^2 - d_0)\alpha_1 + (3v_0d_0 - c_0)\alpha_1^2 \right. \\
 &\quad + (7v_0d_0 - 6v_0^3 - c_0)\alpha_1^3 + 3v_0(d_0 - 2v_0^2)\alpha_1^4 + (6v_0^3 - 6v_0d_0 + c_0)\alpha_1^5 \left. \right\} \alpha_1^{2l+k} \\
 &\quad + \frac{2v_0 + \alpha_0}{1 - \alpha_1} f_3 \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] \alpha_1^{2l} \\
 &\quad + \left\{ \frac{\alpha_0 f_3}{1 - \alpha_1} \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] \right\} \alpha_1^{2(l-k)} \\
 &\quad + \frac{v_0}{(1 - \alpha_1)(1 - \alpha_1^2)} f_2 \left[2v_0\alpha_0 + d_0(1 - \alpha_1) + v_0^2(2\alpha_1 - 1) \right] \alpha_1^{2l-k} \\
 &\quad + \frac{v_0 + 2\alpha_0}{1 - \alpha_1} \mu(k, l) - f_2 \mu(k) \left[\alpha_0 + (v_0 + \alpha_0)\alpha_1 \right].
 \end{aligned}$$

So it follows that we have

$$\begin{aligned}
 \mu(k, l, m) &= \alpha_1^{m-l} \left[\mu(k, l, l) - f_1 \mu(k, l) \right] + f_1 \mu(k, l) \\
 &= \alpha_1^{m-l} \left[f_4 \left\{ c_0 - 4v_0 d_0 + 3v_0^3 + 3v_0(v_0^2 - d_0)\alpha_1 + (3v_0 d_0 - c_0)\alpha_1^2 \right. \right. \\
 &\quad \left. \left. + (7v_0 d_0 - 6v_0^3 - c_0)\alpha_1^3 + 3v_0(d_0 - 2v_0^2)\alpha_1^4 + (6v_0^3 - 6v_0 d_0 + c_0)\alpha_1^5 \right\} \alpha_1^{2l+k} \right. \\
 &\quad \left. + \frac{2v_0 + \alpha_0}{1 - \alpha_1} f_3 \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] \alpha_1^{2l} \right. \\
 &\quad \left. + \left\{ \frac{\alpha_0 f_3}{1 - \alpha_1} \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] \right\} \alpha_1^{2(l-k)} \right. \\
 &\quad \left. + \frac{v_0}{(1 - \alpha_1)(1 - \alpha_1^2)} f_2 \left[2v_0 \alpha_0 + d_0(1 - \alpha_1) + v_0^2(2\alpha_1 - 1) \right] \alpha_1^{2l-k} \right. \\
 &\quad \left. + \frac{v_0 + \alpha_0}{1 - \alpha_1} \mu(k, l) - f_2 \mu(k) \left[\alpha_0 + (v_0 + \alpha_0)\alpha_1 \right] \right] + f_1 \mu(k, l),
 \end{aligned}$$

which holds for all $0 \leq k \leq l \leq m$.

B. APPENDIX — Proof of Corollary 2.1

To establish the results present in Corollary 2.1 we use the general relations between joint moments and joint cumulants (see [2], p. 5):

- (a) The second-order central moments and cumulants of X , for any $s \geq 0$, are given by

$$\tilde{\mu}(s) = \kappa(s) = \text{Cov}(X_t, X_{t+s}) = v_0 \alpha_1^s f_2.$$

- (b) The third-order central moments and cumulants, for any $l \geq s \geq 0$, are given by

$$\begin{aligned} \tilde{\mu}(s, l) &= \kappa(s, l) \\ &= f_3 \alpha_1^l \left[v_0^2 (1 + \alpha_1 + \alpha_1^2) - \left\{ v_0^2 (1 + \alpha_1 - 2\alpha_1^2) - d_0 (1 - \alpha_1^2) \right\} \alpha_1^s \right]. \end{aligned}$$

- (c) In what concerns the fourth-order cumulants we have, for $m \geq l \geq s \geq 0$,

$$\begin{aligned} \kappa(s, l, m) &= \alpha_1^{m-l} \left[\alpha_1^{2l+s} f_4 \left\{ c_0 - 4v_0 d_0 + 3v_0^3 + 3v_0(v_0^2 - d_0)\alpha_1 + (3v_0 d_0 - c_0)\alpha_1^2 \right. \right. \\ &\quad \left. \left. + (7v_0 d_0 - 6v_0^3 - c_0)\alpha_1^3 + 3v_0(d_0 - 2v_0^2)\alpha_1^4 + (6v_0^3 - 6v_0 d_0 + c_0)\alpha_1^5 \right\} \right. \\ &\quad \left. + \frac{2v_0 + \alpha_0}{1 - \alpha_1} f_3 \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] \alpha_1^{2l} \right. \\ &\quad \left. + \left\{ \frac{\alpha_0 f_3}{1 - \alpha_1} \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] \right\} \alpha_1^{2(l-s)} \right. \\ &\quad \left. + \frac{v_0}{(1 - \alpha_1)(1 - \alpha_1^2)} f_2 \left[2v_0 \alpha_0 + d_0(1 - \alpha_1) + v_0^2(2\alpha_1 - 1) \right] \alpha_1^{2l-s} \right. \\ &\quad \left. + \frac{v_0 + \alpha_0}{1 - \alpha_1} \mu(s, l) - f_2 \mu(s) \left[\alpha_0 + (v_0 + \alpha_0)\alpha_1 \right] \right] + \mathbf{f_1 \mu(s, l)} - \mathbf{f_1 \mu(s, l)} \\ &\quad - \mathbf{f_1} \left(\left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] \mathbf{f_3} \alpha_1^{m+l-2s} + \frac{v_0(v_0 + \alpha_0)}{1 - \alpha_1} f_2 \alpha_1^{m-s} \right. \\ &\quad \left. + v_0 f_1 f_2 \alpha_1^{m-l} + \mathbf{f_1 \mu(l-s)} - \mathbf{f_1 f_2} (v_0 \alpha_1^{l-s} + \alpha_0(1 + \alpha_1)) \right. \\ &\quad \left. + \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] \mathbf{f_3} \alpha_1^{m+l} \right. \\ &\quad \left. + \frac{v_0(v_0 + \alpha_0)}{1 - \alpha_1} \mathbf{f_2} \alpha_1^m + \mathbf{f_1 f_2} v_0 \alpha_1^{m-l} + \mathbf{f_1 \mu(l)} - \mathbf{f_1 \mu(l)} + \frac{v_0(v_0 + \alpha_0)}{1 - \alpha_1} \mathbf{f_2} \alpha_1^m \right. \\ &\quad \left. + \left[d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2) \right] \mathbf{f_3} \alpha_1^{m+s} + v_0 \mathbf{f_1 f_2} \alpha_1^{m-s} + \mathbf{f_1 \mu(s)} - \mathbf{f_1 \mu(s)} \right) \\ &\quad - \left(f_2 \left[v_0 \alpha_1^s + \alpha_0(1 + \alpha_1) \right] - f_1^2 \right) \left(f_2 \left[v_0 \alpha_1^{m-l} + \alpha_0(1 + \alpha_1) \right] - f_1^2 \right) \\ &\quad - \left(f_2 \left[v_0 \alpha_1^l + \alpha_0(1 + \alpha_1) \right] - f_1^2 \right) \left(f_2 \left[v_0 \alpha_1^{m-s} + \alpha_0(1 + \alpha_1) \right] - f_1^2 \right) \\ &\quad - \left(f_2 \left[v_0 \alpha_1^m + \alpha_0(1 + \alpha_1) \right] - f_1^2 \right) \left(f_2 \left[v_0 \alpha_1^{l-s} + \alpha_0(1 + \alpha_1) \right] - f_1^2 \right) \\ &\quad + f_1^2 \left(f_2 \left[v_0 \alpha_1^m + \alpha_0(1 + \alpha_1) \right] + f_2 \left[v_0 \alpha_1^{m-s} + \alpha_0(1 + \alpha_1) \right] \right. \\ &\quad \left. + f_2 \left[v_0 \alpha_1^{m-l} + \alpha_0(1 + \alpha_1) \right] - 3f_1^2 \right), \end{aligned}$$

where we highlight, using bold, expressions whose sum equals zero.

So, taking into account that

$$-f_2 \mu(s) [\alpha_0 + (v_0 + \alpha_0)\alpha_1] \alpha_1^{m-l} = \left[-f_1 \frac{\alpha_0 + v_0}{1 - \alpha_1} \mu(s) + v_0 f_2 \mu(s) \right] \alpha_1^{m-l}$$

and

$$\begin{aligned} & - \left(f_2 [v_0 \alpha_1^s + \alpha_0(1 + \alpha_1)] - f_1^2 \right) \left(f_2 [v_0 \alpha_1^{m-l} + \alpha_0(1 + \alpha_1)] - f_1^2 \right) \\ & - \left(f_2 [v_0 \alpha_1^l + \alpha_0(1 + \alpha_1)] - f_1^2 \right) \left(f_2 [v_0 \alpha_1^{m-s} + \alpha_0(1 + \alpha_1)] - f_1^2 \right) \\ & - \left(f_2 [v_0 \alpha_1^m + \alpha_0(1 + \alpha_1)] - f_1^2 \right) \left(f_2 [v_0 \alpha_1^{l-s} + \alpha_0(1 + \alpha_1)] - f_1^2 \right) \\ & + f_1^2 \left(f_2 [v_0 \alpha_1^m + \alpha_0(1 + \alpha_1)] + f_2 [v_0 \alpha_1^{m-s} + \alpha_0(1 + \alpha_1)] \right. \\ & \left. + f_2 [v_0 \alpha_1^{m-l} + \alpha_0(1 + \alpha_1)] - 3f_1^2 \right) = \\ & = -v_0^2 f_2^2 [\alpha_1^{m-l+s} + 2\alpha_1^{m+l-s}] + v_0 f_1^2 f_2 [\alpha_1^{m-l} + \alpha_1^{m-s} + \alpha_1^m] \end{aligned}$$

we obtain, by replacing $\mu(s, l)$,

$$\begin{aligned} \kappa(s, l, m) = \alpha_1^m f_4 \left[\right. & \left\{ c_0 - 4v_0 d_0 + 3v_0^3 + 3v_0(v_0^2 - d_0)\alpha_1 + (3\alpha_0 d_0 - c_0)\alpha_1^2 \right. \\ & + (7v_0 d_0 - 6v_0^3 - c_0)\alpha_1^3 + 3v_0(d_0 - 2v_0^2)\alpha_1^4 + (6v_0^3 - 6v_0 d_0 + c_0)\alpha_1^5 \left. \right\} \alpha_1^{l+s} \\ & + v_0(1 + \alpha_1 + \alpha_1^2 + \alpha_1^3) [d_0(1 - \alpha_1^2) - v_0^2(1 + \alpha_1 - 2\alpha_1^2)] (2\alpha_1^l + \alpha_1^s) \\ & \left. + v_0(1 + \alpha_1 + \alpha_1^2)(1 + \alpha_1^2) [(1 + \alpha_1)v_0^2 + (d_0(1 - \alpha_1) + v_0^2(2\alpha_1 - 1)) \alpha_1^{l-s}] \right], \end{aligned}$$

for any $m \geq l \geq s \geq 0$.

Finally, the fourth-order central moments of X are given by

$$\begin{aligned} \tilde{\mu}(s, l, m) & = \kappa(s, l, m) + v_0 \alpha_1^s f_2 v_0 \alpha_1^{m-l} f_2 + v_0 \alpha_1^l f_2 v_0 \alpha_1^{m-s} f_2 + v_0 \alpha_1^{l-s} f_2 v_0 \alpha_1^m f_2 \\ & = \kappa(s, l, m) + v_0^2 f_2^2 \alpha_1^{m-l+s} + 2v_0^2 f_2^2 \alpha_1^{m+l-s}. \end{aligned}$$

C. APPENDIX — Covariance matrix of the asymptotic distribution of CLS estimators in CP-INARCH model

To obtain the entries of the covariance matrix $\mathbf{V}^{-1}\mathbf{W}\mathbf{V}^{-1}$, let us begin by deducing the inverse of \mathbf{V} :

$$\mathbf{V}^{-1} = \frac{(1-\alpha_1)(1-\alpha_1^2)}{v_0\alpha_0} \begin{bmatrix} \frac{\alpha_0(v_0 + \alpha_0(1 + \alpha_1))}{(1-\alpha_1)(1-\alpha_1^2)} & -\frac{\alpha_0}{1-\alpha_1} \\ -\frac{\alpha_0}{1-\alpha_1} & 1 \end{bmatrix} = \begin{bmatrix} 1 + \frac{\alpha_0}{v_0}(1 + \alpha_1) & -\frac{1}{v_0}(1 - \alpha_1^2) \\ -\frac{1}{v_0}(1 - \alpha_1^2) & \frac{(1-\alpha_1)(1-\alpha_1^2)}{v_0\alpha_0} \end{bmatrix}.$$

Furthermore, considering $u_t(\alpha) = X_t - g(\alpha, X_{t-1})$,

$$\begin{aligned} E\left[f(X_{t-1}) \cdot u_t^2(\alpha)\right] &= E\left[f(X_{t-1}) \cdot E\left[(X_t - \alpha_0 - \alpha_1 X_{t-1})^2 \mid X_{t-1}\right]\right] \\ &= E\left[f(X_{t-1}) \cdot V\left[X_t - \alpha_0 - \alpha_1 X_{t-1} \mid X_{t-1}\right] + 0\right] \\ &= E\left[f(X_{t-1}) \cdot V\left[X_t \mid X_{t-1}\right]\right] = E\left[f(X_{t-1}) \cdot v_0(\alpha_0 + \alpha_1 X_{t-1})\right], \end{aligned}$$

because of the conditional compound Poisson distribution, and then

$$\begin{aligned} \mathbf{W} &= \begin{bmatrix} E\left(u_t^2 \frac{\partial g}{\partial \alpha_0} \frac{\partial g}{\partial \alpha_0}\right) & E\left(u_t^2 \frac{\partial g}{\partial \alpha_0} \frac{\partial g}{\partial \alpha_1}\right) \\ E\left(u_t^2 \frac{\partial g}{\partial \alpha_1} \frac{\partial g}{\partial \alpha_0}\right) & E\left(u_t^2 \frac{\partial g}{\partial \alpha_1} \frac{\partial g}{\partial \alpha_1}\right) \end{bmatrix} \\ &= \begin{bmatrix} E\left[1 \cdot v_0(\alpha_0 + \alpha_1 X_{t-1})\right] & E\left[X_{t-1} \cdot v_0(\alpha_0 + \alpha_1 X_{t-1})\right] \\ E\left[X_{t-1} \cdot v_0(\alpha_0 + \alpha_1 X_{t-1})\right] & E\left[X_{t-1}^2 \cdot v_0(\alpha_0 + \alpha_1 X_{t-1})\right] \end{bmatrix} \\ &= \frac{v_0\alpha_0}{1-\alpha_1} \begin{bmatrix} 1 & \frac{v_0\alpha_1 + \alpha_0(1 + \alpha_1)}{1-\alpha_1^2} \\ \frac{v_0\alpha_1 + \alpha_0(1 + \alpha_1)}{1-\alpha_1^2} & \frac{v_0\alpha_0(1 + 2\alpha_1)}{(1-\alpha_1)(1-\alpha_1^2)} + \frac{\alpha_0^2}{(1-\alpha_1)^2} + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{(1-\alpha_1^2)(1-\alpha_1^3)} \end{bmatrix}, \end{aligned}$$

since

$$\begin{aligned} E\left[v_0(\alpha_0 + \alpha_1 X_{t-1})\right] &= v_0 \left[\alpha_0 + \alpha_1 \frac{\alpha_0}{1-\alpha_1}\right] = \frac{v_0\alpha_0}{1-\alpha_1}, \\ E\left[X_{t-1} v_0(\alpha_0 + \alpha_1 X_{t-1})\right] &= v_0 \left[\frac{\alpha_0^2}{1-\alpha_1} + \frac{\alpha_1\alpha_0(v_0 + \alpha_0(1 + \alpha_1))}{(1-\alpha_1)(1-\alpha_1^2)}\right] \\ &= \frac{v_0\alpha_0}{1-\alpha_1} \cdot \frac{v_0\alpha_1 + \alpha_0(1 + \alpha_1)}{1-\alpha_1^2}, \end{aligned}$$

$$\begin{aligned}
 E[X_{t-1}^2 \cdot v_0(\alpha_0 + \alpha_1 X_{t-1})] &= \\
 &= v_0 \left[\frac{\alpha_0^2(v_0 + \alpha_0(1 + \alpha_1))}{(1 - \alpha_1)(1 - \alpha_1^2)} + \frac{\alpha_1 \alpha_0}{(1 - \alpha_1)^3} \left(\frac{d_0 + (3v_0^2 - d_0)\alpha_1^2}{(1 + \alpha_1)(1 + \alpha_1 + \alpha_1^2)} + \frac{3v_0 \alpha_0}{1 + \alpha_1} + \alpha_0^2 \right) \right] \\
 &= \frac{v_0 \alpha_0}{1 - \alpha_1} \left[\frac{v_0 \alpha_0(1 - \alpha_1) + 3v_0 \alpha_0 \alpha_1}{(1 - \alpha_1)^2(1 + \alpha_1)} + \frac{\alpha_0^2(1 - \alpha_1) + \alpha_0^2 \alpha_1}{(1 - \alpha_1)^2} + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{(1 - \alpha_1^2)(1 - \alpha_1^3)} \right] \\
 &= \frac{v_0 \alpha_0}{1 - \alpha_1} \left[\frac{v_0 \alpha_0(1 + 2\alpha_1)}{(1 - \alpha_1)(1 - \alpha_1^2)} + \frac{\alpha_0^2}{(1 - \alpha_1)^2} + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{(1 - \alpha_1^2)(1 - \alpha_1^3)} \right],
 \end{aligned}$$

using again the expressions stated in Theorem 2.1.

Now, the product of $\mathbf{V}^{-1}\mathbf{W}$ is given by

$$\begin{aligned}
 &\begin{bmatrix} 1 + \frac{\alpha_0}{v_0}(1 + \alpha_1) & -\frac{1}{v_0}(1 - \alpha_1^2) \\ -\frac{1}{v_0}(1 - \alpha_1^2) & \frac{(1 - \alpha_1)(1 - \alpha_1^2)}{v_0 \alpha_0} \end{bmatrix} \\
 &\cdot \begin{bmatrix} 1 & \frac{v_0 \alpha_1 + \alpha_0(1 + \alpha_1)}{1 - \alpha_1^2} \\ \frac{v_0 \alpha_1 + \alpha_0(1 + \alpha_1)}{1 - \alpha_1^2} & \frac{v_0 \alpha_0(1 + 2\alpha_1)}{(1 - \alpha_1)(1 - \alpha_1^2)} + \frac{\alpha_0^2}{(1 - \alpha_1)^2} + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{(1 - \alpha_1^2)(1 - \alpha_1^3)} \end{bmatrix} = \\
 &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 1 - \alpha_1 & \frac{v_0 \alpha_1}{1 - \alpha_1^2} - \frac{\alpha_0 \alpha_1}{1 - \alpha_1} - \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0(1 - \alpha_1^3)} \\ \frac{\alpha_1(1 - \alpha_1)}{\alpha_0} & 1 + \alpha_1 + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0 \alpha_0(1 + \alpha_1 + \alpha_1^2)} \end{bmatrix},
 \end{aligned}$$

since

$$\begin{aligned}
 a_{11} &= 1 + \frac{\alpha_0(1 + \alpha_1)}{v_0} - \frac{1 - \alpha_1^2}{v_0} \frac{v_0 \alpha_1 + \alpha_0(1 + \alpha_1)}{1 - \alpha_1^2} = 1 - \alpha_1, \\
 a_{12} &= \left(1 + \frac{\alpha_0}{v_0}(1 + \alpha_1) \right) \frac{v_0 \alpha_1 + \alpha_0(1 + \alpha_1)}{1 - \alpha_1^2} \\
 &\quad - \frac{1 - \alpha_1^2}{v_0} \left[\frac{v_0 \alpha_0(1 + 2\alpha_1)}{(1 - \alpha_1)(1 - \alpha_1^2)} + \frac{\alpha_0^2}{(1 - \alpha_1)^2} + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{(1 - \alpha_1^2)(1 - \alpha_1^3)} \right] \\
 &= \frac{v_0 \alpha_1}{1 - \alpha_1^2} + \frac{\alpha_0}{1 - \alpha_1} + \frac{\alpha_0 \alpha_1}{1 - \alpha_1} + \frac{\alpha_0^2(1 + \alpha_1)}{v_0(1 - \alpha_1)} - \frac{\alpha_0(1 + 2\alpha_1)}{1 - \alpha_1} - \frac{\alpha_0^2(1 + \alpha_1)}{v_0(1 - \alpha_1)} \\
 &\quad - \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0(1 - \alpha_1^3)} \\
 &= \frac{v_0 \alpha_1}{1 - \alpha_1^2} - \frac{\alpha_0 \alpha_1}{1 - \alpha_1} - \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0(1 - \alpha_1^3)}, \\
 a_{21} &= -\frac{(1 - \alpha_1^2)}{v_0} + \frac{(1 - \alpha_1)(1 - \alpha_1^2)(v_0 \alpha_1 + \alpha_0(1 + \alpha_1))}{v_0 \alpha_0(1 - \alpha_1^2)} \\
 &= -\frac{(1 - \alpha_1^2)}{v_0} + \frac{\alpha_1(1 - \alpha_1)}{\alpha_0} + \frac{(1 - \alpha_1^2)}{v_0} = \frac{\alpha_1(1 - \alpha_1)}{\alpha_0},
 \end{aligned}$$

$$\begin{aligned}
a_{22} &= -\frac{(1 - \alpha_1^2)(v_0\alpha_1 + \alpha_0(1 + \alpha_1))}{v_0(1 - \alpha_1^2)} \\
&\quad + \frac{(1 - \alpha_1)(1 - \alpha_1^2)}{v_0\alpha_0} \left[\frac{v_0\alpha_0(1 + 2\alpha_1)}{(1 - \alpha_1)(1 - \alpha_1^2)} + \frac{\alpha_0^2}{(1 - \alpha_1)^2} + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{(1 - \alpha_1^2)(1 - \alpha_1^3)} \right] \\
&= -\alpha_1 - \frac{\alpha_0(1 + \alpha_1)}{v_0} + 1 + 2\alpha_1 + \frac{\alpha_0(1 + \alpha_1)}{v_0} + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0\alpha_0(1 + \alpha_1 + \alpha_1^2)} \\
&= 1 + \alpha_1 + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0\alpha_0(1 + \alpha_1 + \alpha_1^2)}.
\end{aligned}$$

So, the asymptotic covariance matrix is such that

$$\begin{aligned}
\mathbf{V}^{-1}\mathbf{W}\mathbf{V}^{-1} &= \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \\
&= \frac{v_0\alpha_0}{1 - \alpha_1} \begin{bmatrix} 1 - \alpha_1 & \frac{v_0\alpha_1}{1 - \alpha_1^2} - \frac{\alpha_0\alpha_1}{1 - \alpha_1} - \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0(1 - \alpha_1^3)} \\ \frac{\alpha_1(1 - \alpha_1)}{\alpha_0} & 1 + \alpha_1 + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0\alpha_0(1 + \alpha_1 + \alpha_1^2)} \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} 1 + \frac{\alpha_0}{v_0}(1 + \alpha_1) & -\frac{1}{v_0}(1 - \alpha_1^2) \\ -\frac{1}{v_0}(1 - \alpha_1^2) & \frac{(1 - \alpha_1)(1 - \alpha_1^2)}{v_0\alpha_0} \end{bmatrix},
\end{aligned}$$

where

$$\begin{aligned}
b_{11} &= \frac{\alpha_0}{1 - \alpha_1} \left(\alpha_0(1 + \alpha_1) + \frac{v_0^2 + (d_0 - v_0^2)\alpha_1(1 + \alpha_1 - \alpha_1^2) + (3v_0^2 - d_0)\alpha_1^4}{v_0(1 + \alpha_1 + \alpha_1^2)} \right), \\
b_{12} = b_{21} &= v_0\alpha_1 - \alpha_0(1 + \alpha_1) - \frac{\alpha_1(1 + \alpha_1)(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0(1 + \alpha_1 + \alpha_1^2)}, \\
b_{22} &= (1 - \alpha_1^2) \left(1 + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0\alpha_0(1 + \alpha_1 + \alpha_1^2)} \right).
\end{aligned}$$

In fact, we have

$$\begin{aligned}
b_{11} &= \frac{v_0\alpha_0}{1 - \alpha_1} \left[(1 - \alpha_1) \left(1 + \frac{\alpha_0}{v_0}(1 + \alpha_1) \right) \right. \\
&\quad \left. - \frac{1}{v_0}(1 - \alpha_1^2) \left(\frac{v_0\alpha_1}{1 - \alpha_1^2} - \frac{\alpha_0\alpha_1}{1 - \alpha_1} - \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0(1 - \alpha_1^3)} \right) \right] \\
&= \frac{\alpha_0}{1 - \alpha_1} \left[v_0(1 - \alpha_1) + \alpha_0(1 - \alpha_1^2) - v_0\alpha_1 + \alpha_0\alpha_1(1 + \alpha_1) \right. \\
&\quad \left. + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)(1 + \alpha_1)}{v_0(1 + \alpha_1 + \alpha_1^2)} \right] =
\end{aligned}$$

$$\begin{aligned}
 &= \frac{\alpha_0}{1 - \alpha_1} \left[\alpha_0(1 + \alpha_1) + \frac{v_0^2(1 - 2\alpha_1)(1 + \alpha_1 + \alpha_1^2) + \alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)(1 + \alpha_1)}{v_0(1 + \alpha_1 + \alpha_1^2)} \right] \\
 &= \frac{\alpha_0}{1 - \alpha_1} \left(\alpha_0(1 + \alpha_1) + \frac{v_0^2 + (d_0 - v_0^2)\alpha_1(1 + \alpha_1 - \alpha_1^2) + (3v_0^2 - d_0)\alpha_1^4}{v_0(1 + \alpha_1 + \alpha_1^2)} \right), \\
 b_{12} &= \frac{v_0\alpha_0}{1 - \alpha_1} \left[-\frac{(1 - \alpha_1)(1 - \alpha_1^2)}{v_0} \right. \\
 &\quad \left. + \frac{(1 - \alpha_1)(1 - \alpha_1^2)}{v_0\alpha_0} \left(\frac{v_0\alpha_1}{1 - \alpha_1^2} - \frac{\alpha_0\alpha_1}{1 - \alpha_1} - \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0(1 - \alpha_1^3)} \right) \right] \\
 &= -\alpha_0(1 - \alpha_1^2) + v_0\alpha_1 - \alpha_0\alpha_1(1 + \alpha_1) - \frac{\alpha_1(1 + \alpha_1)(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0(1 + \alpha_1 + \alpha_1^2)} \\
 &= v_0\alpha_1 - \alpha_0(1 + \alpha_1) - \frac{\alpha_1(1 + \alpha_1)(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0(1 + \alpha_1 + \alpha_1^2)}, \\
 b_{21} &= \frac{v_0\alpha_0}{1 - \alpha_1} \left[\frac{\alpha_1(1 - \alpha_1)}{\alpha_0} \left(1 + \frac{\alpha_0(1 + \alpha_1)}{v_0} \right) - \frac{1 - \alpha_1^2}{v_0} \left(1 + \alpha_1 + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0\alpha_0(1 + \alpha_1 + \alpha_1^2)} \right) \right] \\
 &= v_0\alpha_1 + \alpha_0\alpha_1(1 + \alpha_1) - \alpha_0(1 + \alpha_1) - \alpha_0\alpha_1(1 + \alpha_1) - \frac{\alpha_1(1 + \alpha_1)(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0(1 + \alpha_1 + \alpha_1^2)} \\
 &= v_0\alpha_1 - \alpha_0(1 + \alpha_1) - \frac{\alpha_1(1 + \alpha_1)(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0(1 + \alpha_1 + \alpha_1^2)}, \\
 b_{22} &= \frac{v_0\alpha_0}{1 - \alpha_1} \left[-\frac{\alpha_1(1 - \alpha_1)(1 - \alpha_1^2)}{v_0\alpha_0} + \frac{(1 - \alpha_1)(1 - \alpha_1^2)}{v_0\alpha_0} \left(1 + \alpha_1 + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0\alpha_0(1 + \alpha_1 + \alpha_1^2)} \right) \right] \\
 &= -\alpha_1(1 - \alpha_1^2) + \alpha_1(1 - \alpha_1^2) + (1 - \alpha_1^2) \left(1 + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0\alpha_0(1 + \alpha_1 + \alpha_1^2)} \right) \\
 &= (1 - \alpha_1^2) \left(1 + \frac{\alpha_1(d_0 + (3v_0^2 - d_0)\alpha_1^2)}{v_0\alpha_0(1 + \alpha_1 + \alpha_1^2)} \right).
 \end{aligned}$$

ACKNOWLEDGMENTS

We are deeply grateful to the reviewer for useful suggestions that have allowed us to improve upon an earlier version of the manuscript. This work was partially supported by the Centre for Mathematics of the University of Coimbra – UID/MAT/00324/2019, funded by the Portuguese Government through FCT/MEC and co-funded by the European Regional Development Fund through the Partnership Agreement PT2020.

REFERENCES

- [1] AHMAD, A. and FRANCO, C. (2016). Poisson QMLE of count time series models, *Journal of Time Series Analysis*, **37**, 291–314.
- [2] BAKOUCH, H. (2010). Higher-order moments, cumulants and spectral densities of the NGINAR(1) process, *Statistical Methodology*, **7**, 1–21.
- [3] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press, Princeton.
- [4] FERLAND, R.; LATOUR, A. and ORAICHI, D. (2006). Integer-valued GARCH process, *Journal of Time Series Analysis*, **27**, 923–942.
- [5] GONÇALVES, E.; MENDES LOPES, N. and SILVA, F. (2015a). A new approach to integer-valued time series modeling: The Neyman Type-A INGARCH model, *Lithuanian Mathematical Journal*, **55**(2), 231–242.
- [6] GONÇALVES, E.; MENDES LOPES, N. and SILVA, F. (2015b). Infinitely divisible distributions in integer valued GARCH models, *Journal of Time Series Analysis*, **36**, 503–527.
- [7] GONÇALVES, E.; MENDES LOPES, N. and SILVA, F. (2016). Zero-inflated compound Poisson distributions in integer-valued GARCH models, *Statistics*, **50**, 558–578.
- [8] JOHNSON, N.L.; KOTZ, S. and KEMP, A.W. (2005). *Univariate Discrete Distributions*, Wiley, New York, 3rd Edn..
- [9] KLIMKO, L.A. and NELSON, P.I. (1978). On conditional least squares estimation for stochastic processes, *Ann. Stat.*, **6**(3), 629–642.
- [10] WEISS, C.H. (2010a). INARCH(1) processes: higher-order moments and jumps, *Stat. Probab. Lett.*, **80**, 1771–1780.
- [11] WEISS, C.H. (2010b). The INARCH(1) model for overdispersed time series of counts, *Commun. Statist. Simul. Comp.*, **39**(6), 1269–1291.
- [12] WEISS, C.H.; GONÇALVES, E. and MENDES LOPES, N. (2017). Testing the compounding structure of the CP-INARCH model, *Metrika*, **80**, 571–603.
- [13] XU, H.-Y.; XIE, M.; GOH, T.N. and FU, X. (2012). A model for integer-valued time series with conditional overdispersion, *Computational Statistics and Data Analysis*, **56**, 4229–4242.
- [14] ZHU, FK. (2011). A negative binomial integer-valued GARCH model, *Journal of Time Series Analysis*, **32**, 54–67.
- [15] ZHU, FK. (2012). Modelling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models, *Journal of Mathematical Analysis and Applications*, **389**(1), 58–71.

PARAMETER ESTIMATION FOR THE TWO-PARAMETER MAXWELL DISTRIBUTION UNDER COMPLETE AND CENSORED SAMPLES

Authors: TALHA ARSLAN
– Department of Econometrics, Van Yuzuncu Yil University,
Turkey
mstalhaarslan@yyu.edu.tr

SUKRU ACITAS
– Department of Statistics, Eskisehir Technical University,
Turkey
sacitas@eskisehir.edu.tr

BIRDAL SENOGLU
– Department of Statistics, Ankara University,
Turkey
senoglu@science.ankara.edu.tr

Received: January 2018

Revised: February 2019

Accepted: February 2019

Abstract:

- The Maxwell distribution is one of the basic distributions in Physics besides being popular in Statistics for modeling lifetime data. This paper considers the parameter estimation of the Maxwell distribution via modified maximum likelihood (MML) methodology for both complete and censored samples. The MML estimators for the location and scale parameters of the Maxwell distribution have explicit forms and they are robust against the plausible deviations from the assumed model. A Monte Carlo simulation study is conducted to compare the performances of the MML estimators with the corresponding maximum likelihood (ML), least squares (LS) and method of moments (MoM) estimators.

Keywords:

- *efficiency; Maxwell distribution; modified likelihood; Monte Carlo simulation; Newton–Raphson; Type-II censoring.*

AMS Subject Classification:

- 62F10, 62F12, 62F35, 62N02, 62P30.

1. INTRODUCTION

The Maxwell distribution is widely used in many problems especially in Physics. For example, the speed of molecules in thermal equilibrium is modelled by using the Maxwell distribution (Maxwell [21]; Mathai and Princy [20]). Also note that there is a lot of literature about the Maxwell distribution in Statistics. It was firstly used by Tyagi and Battacharya [31, 32] for modeling the lifetime data. They used Bayes method to estimate the scale parameter of the distribution and obtain the minimum variance unbiased estimator for the reliability function. Dey and Maiti [10] obtained the Bayes estimators of the scale parameter of the Maxwell distribution under various different loss functions. Kazmi *et al.* [16] obtained the maximum likelihood (ML) estimators of the location and scale parameters of the mixture of the Maxwell distribution under Type-I censoring. Al-Baldawi [3] compared the efficiency of the ML estimator of the scale parameter of the Maxwell distribution with the corresponding Bayes estimator. Hossain and Huerta [13] used the Maxwell distribution in analysing the different data sets taken from the literature. Li [19] obtained the estimators of the scale parameter of the Maxwell distribution using the Minimax, Bayesian and ML methods. Fan [12] considered the Bayesian method to estimate the loss and risk function for the scale parameter of the Maxwell distribution. Dey *et al.* [9] obtained estimators of the location and scale parameters of the Maxwell distribution via different estimation methods. See also Arslan *et al.* [5], where the modified maximum likelihood (MML) estimators for the location and scale parameters of the Maxwell distribution are obtained.

The ML methodology is used to obtain the estimators of the parameters of the Maxwell distribution in most of the studies. However, the ML estimators of the location and scale parameters of the Maxwell distribution cannot be obtained explicitly. Therefore, iterative methods should be used. It is known that using iterative methods causes various problems such as (i) non-convergence of iterations, (ii) convergence to multiple roots, and (iii) convergence to the wrong root; see e.g. Barnett [7], Puthenpura and Sinha [23], and Vaughan [33].

The motivation of this study is to obtain the explicit estimators for the location and scale parameters of the Maxwell distribution. For this purpose, Tiku's [28, 29] MML methodology is used. The MML estimators are formulated for both complete and censored samples. An extensive Monte-Carlo (MC) simulation study is carried out to compare performances of the MML estimators with the well-known and widely-used ML, least squares (LS) and method of moments (MoM) estimators.

The rest of the paper is organized as follows. Maxwell distribution is reviewed in Section 2. Section 3 is reserved to the parameter estimation methodologies. The results of the MC simulation study are presented in Section 4. The ML and MML estimators are given under Type-II censoring scheme in Section 5. In Section 6, two real data sets are analyzed to show the implementation of the proposed methodology. The paper ends with some concluding remarks.

2. MAXWELL DISTRIBUTION

Traditionally, the probability density function (pdf) of the Maxwell distribution is given by

$$(2.1) \quad f(v) = 4\pi \left(\frac{m}{\pi 2kT} \right)^{3/2} v^2 \exp \left\{ - \left(\frac{m}{2kT} v^2 \right) \right\}, \quad v > 0,$$

where m is the molecular weight in kg/mol, T is the temperature in Kelvin, k is the constant J/K and v denotes the speed of the molecule. If the reparametrization $\sigma = \sqrt{2kT/m}$ is used and a location parameter μ is added into the Equation (2.1), then the resulting distribution is called as two-parameter Maxwell distribution.

The pdf and the corresponding cumulative distribution function (cdf) of the two-parameter Maxwell distribution are given by

$$(2.2) \quad f(x; \mu, \sigma) = \frac{4}{\sigma \Gamma(1/2)} \left(\frac{x - \mu}{\sigma} \right)^2 \exp \left\{ - \left(\frac{x - \mu}{\sigma} \right)^2 \right\}, \quad \mu \leq x < \infty, \quad \sigma \geq 0,$$

and

$$(2.3) \quad F(x; \mu, \sigma) = \frac{1}{\Gamma(3/2)} \Gamma \left[\left(\frac{x - \mu}{\sigma} \right)^2, 3/2 \right],$$

respectively. Here, μ is the location parameter and σ is the scale parameter. Also, $\Gamma(\cdot)$ and $\Gamma(\cdot, \cdot)$ stand for the gamma and incomplete gamma functions, respectively. See Figure 1 where the plots of the Maxwell distribution are illustrated for certain values of σ .

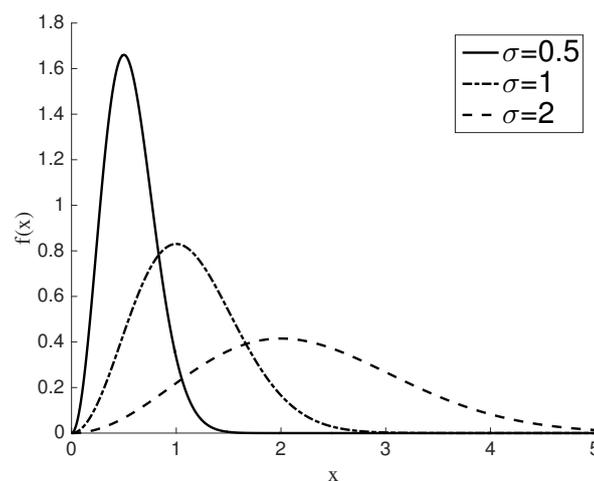


Figure 1: Plots of the Maxwell distribution for certain values of σ .

In the rest of the paper, we use the term Maxwell distribution instead of two-parameter Maxwell distribution for the sake of simplicity.

3. PARAMETER ESTIMATION UNDER COMPLETE SAMPLES

In this section, brief descriptions of the ML, MML, MoM and LS methodologies are provided.

3.1. The ML method

Let X_1, X_2, \dots, X_n be a random sample from the Maxwell distribution. Then, the log-likelihood ($\ln L$) function can be written as follows:

$$(3.1) \quad \ln L = n \ln C - n \ln \sigma + 2 \sum_{i=1}^n \ln z_i - \sum_{i=1}^n z_i^2,$$

where $C = 4/\Gamma(1/2)$ and $z_i = (x_i - \mu)/\sigma$ ($i = 1, 2, \dots, n$). The ML estimates of the parameters μ and σ are obtained as solutions of the following likelihood equations:

$$(3.2) \quad \frac{\partial \ln L}{\partial \mu} = -\frac{2}{\sigma} \sum_{i=1}^n g(z_i) + \frac{2}{\sigma} \sum_{i=1}^n z_i = 0$$

and

$$(3.3) \quad \frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} - \frac{2}{\sigma} \sum_{i=1}^n z_i g(z_i) + \frac{2}{\sigma} \sum_{i=1}^n z_i^2 = 0,$$

where $g(z) = z^{-1}$. Equations (3.2) and (3.3) cannot be solved explicitly since they contain the nonlinear $g(z) = z^{-1}$ function. In this study a Newton–Raphson (NR) method is utilized to obtain the solutions of Equations (3.2) and (3.3) simultaneously. The Hessian matrix,

$$(3.4) \quad \mathbf{H} = \begin{bmatrix} \frac{\partial^2 \ln L}{\partial \mu^2} & \frac{\partial^2 \ln L}{\partial \mu \partial \sigma} \\ \frac{\partial^2 \ln L}{\partial \sigma \partial \mu} & \frac{\partial^2 \ln L}{\partial \sigma^2} \end{bmatrix},$$

is used in the NR method. The elements of the Hessian matrix and Fisher Information matrix (**I**) are provided in the [Appendix](#) for the Maxwell distribution.

The following equations are used in the NR method to solve the likelihood equations in (3.2) and (3.3):

$$(3.5) \quad \begin{bmatrix} \frac{\partial^2 \ln L}{\partial \mu^2}(\mu^k, \sigma^k) & \frac{\partial^2 \ln L}{\partial \mu \partial \sigma}(\mu^k, \sigma^k) \\ \frac{\partial^2 \ln L}{\partial \sigma \partial \mu}(\mu^k, \sigma^k) & \frac{\partial^2 \ln L}{\partial \sigma^2}(\mu^k, \sigma^k) \end{bmatrix} \begin{bmatrix} \Xi \mu^k \\ \Xi \sigma^k \end{bmatrix} = \begin{bmatrix} \frac{\partial \ln L}{\partial \mu}(\mu^k, \sigma^k) \\ \frac{\partial \ln L}{\partial \sigma}(\mu^k, \sigma^k) \end{bmatrix},$$

where k denotes the iteration number and Ξ stands for the incremental values. See also Arslan and Senoglu [6], where a similar algorithm scheme has already been used for the one-way ANOVA model under Jones and Faddy's skew t distribution.

3.2. The MML method

As mentioned in the Subsection 3.1, the ML estimators of the location and scale parameters cannot be obtained in closed forms because of the nonlinear function $g(\cdot)$ in Equations (3.2) and (3.3). We here propose to use non-iterative MML methodology developed by Tiku [28, 29] to avoid the computational difficulties and/or problems mentioned in Section 1. The MML methodology also allows us to obtain closed forms of the estimators. There are three steps to obtain the MML estimators of the location parameter μ and scale parameter σ . They are given step by step as follows:

- Step 1.** Standardized observations $z_i = (x_i - \mu)/\sigma$ ($i = 1, 2, \dots, n$) are ordered in ascending way, i.e. $z_{(1)} \leq z_{(2)} \leq \dots \leq z_{(n)}$.
- Step 2.** The ordered observations are incorporated into likelihood equations, since complete sums are invariant to ordering, i.e. $\sum_{i=1}^n h(z_i) = \sum_{i=1}^n h(z_{(i)})$, where $h(\cdot)$ is any function.
- Step 3.** $g(z_{(i)})$ is linearized around the expected values of the standardized ordered observations, i.e. $t_{(i)} = E(z_{(i)})$, by using the first two terms of Taylor series expansion:

$$(3.6) \quad g(z_{(i)}) \cong \alpha_i - \beta_i z_{(i)}, \quad i = 1, \dots, n.$$

After incorporating Equation (3.6) into the likelihood equations, we obtain the following modified likelihood equations:

$$(3.7) \quad \frac{\partial \ln L^*}{\partial \mu} = -\frac{2}{\sigma} \sum_{i=1}^n (\alpha_i - \beta_i z_{(i)}) + \frac{2}{\sigma} \sum_{i=1}^n z_{(i)} = 0$$

and

$$(3.8) \quad \frac{\partial \ln L^*}{\partial \sigma} = -\frac{n}{\sigma} - \frac{2}{\sigma} \sum_{i=1}^n z_{(i)} (\alpha_i - \beta_i z_{(i)}) + \frac{2}{\sigma} \sum_{i=1}^n z_{(i)}^2 = 0.$$

The solutions of these equations are the following MML estimators:

$$(3.9) \quad \hat{\mu}_{\text{MML}} = \bar{x}_w - \frac{\Delta}{m} \hat{\sigma}_{\text{MML}} \quad \text{and} \quad \hat{\sigma}_{\text{MML}} = \frac{-B + \sqrt{B^2 + 4nC}}{2\sqrt{n(n-1)}},$$

where

$$\begin{aligned} \bar{x}_w &= \sum_{i=1}^n \delta_i x_{(i)} / m, & m &= \sum_{i=1}^n \delta_i, & \delta_i &= \beta_i + 1, & \beta_i &= t_{(i)}^{-2}, & \Delta &= \sum_{i=1}^n \alpha_i, \\ \alpha_i &= 2t_{(i)}^{-1}, & B &= 2 \sum_{i=1}^n \alpha_i (x_{(i)} - \bar{x}_w) & \text{and} & C &= 2 \sum_{i=1}^n \delta_i (x_{(i)} - \bar{x}_w)^2. \end{aligned}$$

Here, $x_{(i)}$ represents the i -th ordered observation. It should be noted that $t_{(i)} = E(z_{(i)})$ can be obtained approximately using the following equality:

$$t_{(i)} = F^{-1} \left(\frac{i}{n+1} \right), \quad i = 1, 2, \dots, n,$$

where $F^{-1}(\cdot)$ is the quantile function of the standard Maxwell distribution. The use of these approximate values does not affect the efficiency of the MML estimators adversely. It should also be noticed that the denominator of $\hat{\sigma}_{\text{MML}}$ is $2n$, however it is replaced by $2\sqrt{n(n-1)}$ for bias correction.

The MML estimators are derived in closed form since they are expressed as functions of the sample observations. Furthermore, they are asymptotically equivalent to the ML estimators. The MML estimators are also almost fully efficient, i.e. they have minimum variance bounds (MVBs). They also have very small bias or no bias even for small sample sizes. It should also be mentioned that the MML methodology gives small weight(s) to the outlying observation(s) in the direction of the longer tail(s). Therefore, the MML estimators are robust to the outlier(s), see e.g. Acitas *et al.* [1] and references given therein for further information. See also Figure 2 where plots of the weights for the Maxwell distribution, i.e. $\delta_i = t_{(i)}^{-2} + 1$, are illustrated.

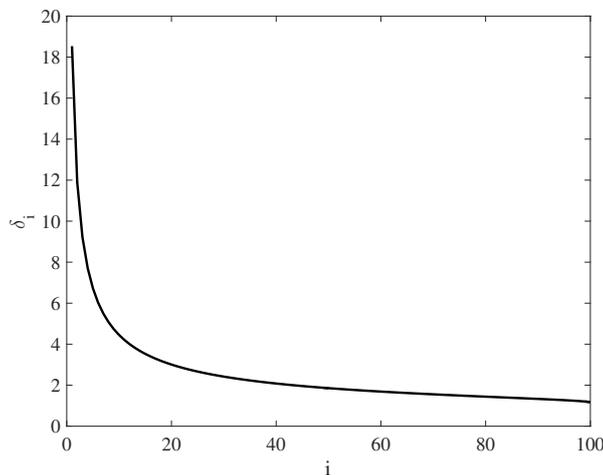


Figure 2: Plot of the weights for the Maxwell distribution, $n = 100$.

The asymptotic distributions of the $\hat{\mu}_{\text{MML}}$ and $\hat{\sigma}_{\text{MML}}$ are provided in Lemma 3.1 and Lemma 3.2.

Lemma 3.1. $\hat{\mu}_{\text{MML}}$ is normally distributed with mean μ and variance σ^2/m for $n \rightarrow \infty$.

Proof: The proof is done based on the following fact: The likelihood equation given in (3.2) and modified likelihood equation given in (3.7) are asymptotically equivalent. Furthermore, $\partial \ln L^*/\partial \mu$ can be written as

$$(3.10) \quad \frac{\partial \ln L^*}{\partial \mu} = \frac{m}{\sigma^2} \left[\left(\bar{x}_w - \frac{\Delta}{m} \hat{\sigma}_{\text{MML}} \right) - \mu \right] = \frac{m}{\sigma^2} (\hat{\mu}_{\text{MML}} - \mu);$$

see Kendall and Stuart [17]. $\hat{\mu}_{\text{MML}}$ is normally distributed since $E(\partial^r \ln L^*/\partial \mu^r) = 0$ for all $r \geq 3$; see Bartlett [8]. □

Lemma 3.2. *Conditional on μ known, $n\hat{\sigma}_{\text{MML}}^2/\sigma^2$ is asymptotically chi-square distributed with n degrees of freedom.*

Proof: This follows from the fact that $B_0/\sqrt{nC_0} \cong 0$ and thus,

$$(3.11) \quad \frac{\partial \ln L^*}{\partial \sigma} = \frac{n}{\sigma^3} \left(\frac{C_0}{n} - \sigma^2 \right),$$

where B_0 and C_0 are the same as B and C , respectively. See for example Tiku [30] and Senoglu [25] for further information. \square

3.3. The MoM method

MoM estimators of the location and scale parameters of the Maxwell distribution are obtained by equating the first two theoretical moments to the first two sample moments. Therefore, MoM estimators of μ and σ are given by

$$(3.12) \quad \hat{\mu}_{\text{MoM}} = \bar{x} - \frac{2}{\sqrt{\pi}} \hat{\sigma}_{\text{MoM}} \quad \text{and} \quad \hat{\sigma}_{\text{MoM}} = s \sqrt{\frac{2\pi}{3\pi - 8}},$$

respectively. Here,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{and} \quad s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

It is clear that MoM estimators are functions of the sample observations as in MML estimators.

3.4. The LS method

LS estimators of μ and σ are obtained by minimizing the following function

$$(3.13) \quad \sum_{i=1}^n \left(F(x_{(i)}) - \frac{i}{n+1} \right)^2, \quad i = 1, 2, \dots, n,$$

with respect to the parameters of interest (Swain *et al.* [27]). Here, $F(\cdot)$ is the cdf of the Maxwell distribution. It is clear that explicit forms of the LS estimators are not available. Therefore, we use the “fminunc” function which exists in the optimization toolbox of MATLAB2017a to obtain the LS estimates of μ and σ .

4. SIMULATION STUDY

In this section, the results of the simulation study in which the performances of the MML estimators are compared with the ML, MoM and LS estimators are presented.

In the simulation setup, we use the sample sizes $n = 10$ (small), $n = 20$, $n = 50$ (moderate) and $n = 120$ (large). Without loss of generality, the location parameter μ and scale parameter σ are taken to be 0 and 1, respectively. All the simulations are carried out for $\lfloor 100,000/n \rfloor$ MC runs where $\lfloor \cdot \rfloor$ denotes the floor function (also known as the greatest integer function) that takes integer part of the number. We use the MATLAB2017a software for all computations. In the ML estimation procedure, the initial values for $\hat{\mu}$ and $\hat{\sigma}$ are taken as $\mu^0 = \hat{\mu}_{\text{MML}}$ and $\sigma^0 = \hat{\sigma}_{\text{MML}}$, respectively.

It should be noted that estimates of μ may sometimes be greater than the smallest order statistics $x_{(1)}$ due to the computational problems. These estimators are referred as impermissible estimators (Dubey [11]). The problem is extinguished by reducing the impermissible estimators as $x_{(1)} - 10^{-4}$, see for example Kantar and Senoglu [15].

The performances of the ML, MML, MoM and LS estimators are compared by using bias, variance, mean square error (MSE) and deficiency (Def) criteria. Def is a natural measure of the joint efficiency of the estimators $\hat{\mu}$ and $\hat{\sigma}$ and is defined by

$$(4.1) \quad \text{Def}(\hat{\mu}, \hat{\sigma}) = \text{MSE}(\hat{\mu}) + \text{MSE}(\hat{\sigma});$$

see for example Akgul *et al.* [2]. The results of the simulation study are tabulated in Table 1. Following conclusions are drawn from Table 1.

Table 1: Simulated bias, variance, MSE and Def values of the ML, MML, MoM and LS estimators ($\mu = 0$ and $\sigma = 1$).

Sample size	Estimators	$\hat{\mu}$			$\hat{\sigma}$			Def
		Bias	Variance	MSE	Bias	Variance	MSE	
$n = 10$	ML	-0.108	0.050	0.062	0.094	0.042	0.051	0.113
	MML	-0.095	0.052	0.061	0.060	0.046	0.050	0.111
	MoM	-0.030	0.066	0.067	0.028	0.054	0.055	0.122
	LS	0.254	0.127	0.191	-0.338	0.131	0.245	0.436
$n = 20$	ML	-0.061	0.025	0.028	0.051	0.022	0.025	0.053
	MML	-0.061	0.026	0.029	0.038	0.023	0.025	0.054
	MoM	-0.018	0.033	0.034	0.016	0.028	0.028	0.061
	LS	0.193	0.058	0.096	-0.278	0.059	0.137	0.232
$n = 50$	ML	-0.030	0.009	0.010	0.022	0.009	0.009	0.019
	MML	-0.035	0.009	0.011	0.019	0.009	0.009	0.020
	MoM	-0.010	0.013	0.013	0.005	0.011	0.011	0.023
	LS	0.153	0.021	0.044	-0.241	0.020	0.078	0.122
$n = 120$	ML	-0.011	0.003	0.004	0.009	0.003	0.003	0.007
	MML	-0.015	0.004	0.004	0.009	0.003	0.003	0.007
	MoM	0.000	0.005	0.005	0.000	0.004	0.004	0.009
	LS	0.148	0.008	0.030	-0.232	0.008	0.062	0.092

Concerning the bias values, and for all sample sizes, the MoM estimator and LS estimator of μ have the smallest and the largest bias value, respectively. It can also be deduced from Table 1 that the bias values of the ML and MML estimators are very similar to each other as expected. The ML, MML and MoM estimators overestimate the location parameter μ while the LS estimator underestimates.

It is clear from Table 1 that the MoM estimator of σ has superiority over the ML, MML and LS estimators in terms of the bias criterion. For the small sample size, it is seen the MML estimator performs better than the ML estimator. However, the ML and MML estimators have more or less the same bias values for moderate and large sample sizes. The LS estimator of the σ has the biggest bias value among the all estimators.

Overall, all the estimators have negligible bias values except the LS estimators in what concerns the bias values of $\hat{\mu}$ and $\hat{\sigma}$.

Concerning the MSE values, the ML and MML estimator of μ have almost same the MSE values for all sample sizes. The LS estimator of the location parameter μ has the worst performance in terms of MSE among all other estimators.

Similar results are also obtained for the scale parameter σ . For example, the LS estimator does not perform well. The ML and MML estimators outperform the MoM estimator in most of the cases, however the MoM estimator has a considerably good performance. Table 1 also reveals that the ML and MML estimators are the most efficient.

To sum up, the ML and MML estimators are preferable among the other estimators according to the MSE criterion. The MSE values for $\hat{\mu}$ and $\hat{\sigma}$ decrease when the sample size n increases, as the theory says.

Concerning the Def values, the ML estimator has the smallest Def values among the other estimators for all cases. The Def values of the MML estimator are very close to those of the ML estimator except $n = 10$. The LS estimator shows the worst performance since it has the biggest Def values.

Finally, the ML and MML estimators are seen to be more efficient than the MoM and LS estimators. It is also clear that the performance of the ML and MML estimators are more or less the same as expected. As it is indicated previously, obtaining the ML estimates of the parameters requires iterative methods and this may cause some problems. On the other hand, the MML estimators are easily obtained from the sample observations without any iterative computations. As a result, the MML estimators may be preferable if our focus is to avoid the computational complexities besides having efficient estimators.

Robustness of the estimators

In this part of the simulation study, robustness properties of the ML, MML, MoM and LS estimators are investigated when there are plausible deviations from an assumed model. For this purpose, we assume that the underlying true model is Maxwell($\mu=0, \sigma=1$) and consider the following alternative models:

Outlier Model: $(n - r)$ Maxwell(0, 1) + r Maxwell(0, 2); $r = [0.5 + 0.1n]$.

Mixture Model: 0.80 Maxwell(0, 1) + 0.20 Maxwell(0, 2).

Contamination Model: 0.90 Maxwell(0, 1) + 0.10 Weibull(1, 0.8046).

Here, Weibull(1, 0.8046) denotes the Weibull distribution with scale parameter $\sigma = 1$ and shape parameter $p = 0.8046$. Simulated mean, variance, MSE and Def values for the ML, MML, MoM and LS estimators of μ and σ under the alternative models are given in Table 2.

Table 2: Simulated mean, variance, MSE and Def values of the ML, MML, MoM and LS estimators under the alternative models.

Sample size	Estimators	$\hat{\mu}$			$\hat{\sigma}$			Def
		Mean	Variance	MSE	Mean	Variance	MSE	
Model I: Outlier Model								
$n = 10$	ML	-0.032	0.090	0.091	1.142	0.108	0.128	0.220
	MML	-0.065	0.102	0.107	1.197	0.126	0.165	0.271
	MoM	-0.215	0.185	0.231	1.289	0.188	0.271	0.502
	LS	-0.215	0.178	0.326	1.535	0.198	0.484	0.810
$n = 20$	ML	-0.083	0.045	0.052	1.191	0.055	0.092	0.144
	MML	-0.097	0.049	0.059	1.219	0.061	0.109	0.168
	MoM	-0.242	0.099	0.158	1.315	0.099	0.198	0.356
	LS	-0.242	0.077	0.173	1.459	0.084	0.295	0.468
$n = 50$	ML	-0.113	0.019	0.032	1.219	0.023	0.071	0.103
	MML	-0.116	0.019	0.033	1.231	0.025	0.078	0.111
	MoM	-0.264	0.045	0.115	1.334	0.044	0.156	0.271
	LS	-0.264	0.029	0.094	1.402	0.030	0.191	0.285
Model II: Mixture Model								
$n = 10$	ML	-0.101	0.114	0.124	1.307	0.179	0.274	0.398
	MML	-0.142	0.129	0.150	1.372	0.206	0.344	0.494
	MoM	-0.321	0.231	0.334	1.483	0.288	0.521	0.856
	LS	-0.321	0.354	0.705	1.841	0.483	1.191	1.896
$n = 20$	ML	-0.175	0.058	0.089	1.380	0.094	0.239	0.328
	MML	-0.194	0.062	0.100	1.415	0.102	0.274	0.374
	MoM	-0.383	0.126	0.272	1.541	0.154	0.446	0.719
	LS	-0.383	0.143	0.370	1.715	0.198	0.710	1.081
$n = 50$	ML	-0.208	0.023	0.066	1.408	0.037	0.204	0.270
	MML	-0.211	0.023	0.068	1.422	0.038	0.216	0.284
	MoM	-0.410	0.050	0.218	1.561	0.060	0.375	0.593
	LS	-0.410	0.048	0.212	1.632	0.066	0.465	0.677
Model III: Contamination Model								
$n = 10$	ML	-0.096	0.167	0.177	1.095	0.217	0.226	0.402
	MML	-0.114	0.184	0.197	1.138	0.250	0.269	0.466
	MoM	-0.221	0.351	0.400	1.197	0.377	0.416	0.816
	LS	-0.221	0.194	0.363	1.478	0.199	0.428	0.791
$n = 20$	ML	-0.167	0.103	0.131	1.157	0.135	0.160	0.291
	MML	-0.167	0.109	0.137	1.177	0.151	0.182	0.319
	MoM	-0.266	0.255	0.326	1.236	0.261	0.317	0.643
	LS	-0.266	0.078	0.193	1.400	0.076	0.236	0.429
$n = 50$	ML	-0.215	0.049	0.096	1.207	0.069	0.111	0.207
	MML	-0.209	0.051	0.094	1.214	0.075	0.121	0.215
	MoM	-0.313	0.162	0.259	1.282	0.158	0.237	0.496
	LS	-0.313	0.029	0.113	1.354	0.028	0.154	0.266

It can be seen from the Table 2 that the ML and MML estimators outperform the MoM and LS estimators according to the MSE and Def criteria. This result implies that the ML and MML estimators of parameters μ and σ are more robust to the data anomalies given above.

5. PARAMETER ESTIMATION UNDER THE TYPE-II CENSORING

Analysis of censored samples are usually encountered in different fields of science such as agriculture, social sciences, medicine, and so on (Senoglu and Tiku [26]). Therefore, we consider a Type-II censoring scheme. Type-II censoring arises if a predetermined number of lower and upper observations are censored (Senoglu and Tiku [26]; Arslan and Senoglu [6]).

According to the simulation results related with the robustness issue in Section 4, we concentrated on the ML and MML estimators of μ and σ under censoring. Let

$$z_{(r_1)} \leq z_{(r_1+1)} \leq \dots \leq z_{(n-r_2-1)} \leq z_{(n-r_2)}$$

be a Type-II censored samples where r_1 and r_2 , with $r_1, r_2 \geq 0$ and $0 < r_1 + r_2 < n$, stand for the number of censored observations from the below and above, respectively. Then, the likelihood (L) function of the Maxwell distribution under the Type-II censored sample can be written as

$$(5.1) \quad L = \left[1 - F(z_{(r_1+1)})\right]^{r_1} \prod_{i=r_1+1}^{n-r_2} f(z_{(i)}) \left[F(z_{(n-r_2)})\right]^{r_2},$$

where $f(\cdot)$ and $F(\cdot)$ are the pdf and cdf of the Maxwell distribution given in Equations (2.2) and (2.3), respectively.

5.1. The ML method

The ML estimates of the parameters μ and σ under the Type-II censored samples are obtained by solving the following likelihood equations:

$$(5.2) \quad \frac{\partial \ln L}{\partial \mu} = -\frac{r_1}{\sigma} g_1(z_{r_1+1}) - \frac{2}{\sigma} \sum_{i=r_1+1}^{n-r_2} g_2(z_i) + \frac{2}{\sigma} \sum_{i=r_1+1}^{n-r_2} z_i + \frac{r_2}{\sigma} g_3(z_{n-r_2}) = 0$$

and

$$(5.3) \quad \begin{aligned} \frac{\partial \ln L}{\partial \sigma} = & -\frac{n - r_1 - r_2}{\sigma} - \frac{r_1}{\sigma} z_{r_1+1} g_1(z_{r_1+1}) - \frac{2}{\sigma} \sum_{i=r_1+1}^{n-r_2} z_i g_2(z_i) + \frac{2}{\sigma} \sum_{i=r_1+1}^{n-r_2} z_i^2 \\ & + \frac{r_2}{\sigma} z_{n-r_2} g_3(z_{n-r_2}) = 0, \end{aligned}$$

where $g_1(z_{r_1+1}) = \frac{f(z_{r_1+1})}{F(z_{r_1+1})}$, $g_2(z_i) = z_i^{-1}$ and $g_3(z_{n-r_2}) = \frac{f(z_{n-r_2})}{1 - F(z_{n-r_2})}$.

Similar to the complete sample case, the likelihood equations in (5.2) and (5.3) are nonlinear functions of the unknown parameters. Therefore, they cannot be obtained explicitly. The NR algorithm is also used here to solve the likelihood equations simultaneously.

5.2. The MML method

The MML estimators for the location μ and scale σ parameters of the Maxwell distribution are obtained under the Type-II censored samples by using an algorithm similar to the one given in Subsection 3.2.

Nonlinear functions are linearized around the expected values of the standardized ordered observations, i.e. $t_{(i)} = E(z_{(i)})$, by using the first two terms of a Taylor series expansion:

$$(5.4) \quad \begin{aligned} g_1(z_{(r_1+1)}) &\cong \alpha_{1r_1+1} - \beta_{1r_1+1} z_{(r_1+1)}, & g_2(z_{(i)}) &\cong \alpha_{2i} - \beta_{2i} z_{(i)}, \\ g_3(z_{(n-r_2)}) &\cong \alpha_{3n-r_2} - \beta_{3n-r_2} z_{(n-r_2)}, & i &= r_1 + 1, \dots, n - r_2. \end{aligned}$$

After replacing nonlinear functions with their linearized versions in the likelihood equations, the following MML estimators are obtained:

$$(5.5) \quad \hat{\mu}_{\text{MML}} = \bar{x}_w - \frac{\Delta}{m} \hat{\sigma}_{\text{MML}} \quad \text{and} \quad \hat{\sigma}_{\text{MML}} = \frac{-B + \sqrt{B^2 + 4AC}}{2\sqrt{A(A-1)}},$$

where

$$\begin{aligned} m &= r_1 \beta_{1r_1+1} + 2 \sum_{i=r_1+1}^{n-r_2} (\beta_{2i} + 1) - r_2 \beta_{3n-r_2}, & A &= n - r_1 - r_2, \\ \bar{x}_w &= \frac{r_1 \beta_{1r_1+1} x_{(r_1+1)} + 2 \sum_{i=r_1+1}^{n-r_2} (\beta_{2i} + 1) x_{(i)} - r_2 \beta_{3n-r_2} x_{(n-r_2)}}{m}, \\ \Delta &= r_1 \alpha_{1r_1+1} + 2 \sum_{i=r_1+1}^{n-r_2} (\alpha_{2i}) - r_2 \alpha_{3n-r_2}, \\ B &= r_1 \beta_{1r_1+1} (x_{(r_1+1)} - \bar{x}_w)^2 + 2 \sum_{i=r_1+1}^{n-r_2} (\beta_{2i} + 1) (x_{(i)} - \bar{x}_w)^2 - r_2 \beta_{3n-r_2} (x_{(n-r_2)} - \bar{x}_w)^2, \\ C &= r_1 \alpha_{1r_1+1} (x_{(r_1+1)} - \bar{x}_w)^2 + 2 \sum_{i=r_1+1}^{n-r_2} (\alpha_{2i} + 1) (x_{(i)} - \bar{x}_w)^2 - r_2 \alpha_{3n-r_2} (x_{(n-r_2)} - \bar{x}_w)^2, \\ \alpha_{1r_1+1} &= g_1(t_{(r_1+1)}) + \beta_{1r_1+1} t_{(r_1+1)}, & \beta_{1r_1+1} &= \frac{f'(t_{(r_1+1)})}{F(t_{(r_1+1)})} - \left[\frac{f(t_{(r_1+1)})}{F(t_{(r_1+1)})} \right]^2, \\ \alpha_{2i} &= 2t_{(i)}^{-1}, & \beta_{2i} &= t_{(i)}^{-2}, \\ \alpha_{3n-r_2} &= g_3(t_{(n-r_2)}) + \beta_{3n-r_2} t_{(n-r_2)}, & \beta_{3n-r_2} &= \frac{f'(t_{(n-r_2)})}{1 - F(t_{(n-r_2)})} - \left[\frac{f(t_{(n-r_2)})}{1 - F(t_{(n-r_2)})} \right]^2. \end{aligned}$$

It should be noticed that the denominator $2A$ is replaced by $2\sqrt{A(A-1)}$ in $\hat{\sigma}_{\text{MML}}$ as a bias correction.

We conducted a MC simulation study for this case and obtained similar results with those obtained in the complete sample case. Therefore, we would not give the results here for the sake of brevity. However, they can be provided upon request from the authors.

6. APPLICATIONS

In this section, two real data sets are modelled by using the Maxwell distribution. The unknown parameters are estimated via the ML and MML methods since the MoM and LS methods fail to exhibit a good performance (see Section 4).

6.1. Example 1: Breaking stress of carbon fibres data

In this subsection, observations on the breaking stress of carbon fibres (in Gba) are used to show the implementation of the proposed methodology. The data set is given in Table 3. Further information about the data set can be found in Nicolas and Padgett [22]. See also Qian [24] and Al-Sobhi and Soliman [4], where the breaking stress of carbon fibres data are modelled using the exponentiated exponential (EE) and exponentiated Weibull (EW) distributions.

Table 3: Observations on breaking stress of carbon fibres, $n = 100$.

0.39	0.81	0.85	0.98	1.08	1.12	1.17	1.18	1.22	1.25	1.36	1.41	1.47	1.57
1.57	1.59	1.59	1.61	1.61	1.69	1.69	1.71	1.73	1.8	1.84	1.84	1.87	1.89
1.92	2.00	2.03	2.03	2.05	2.12	2.17	2.17	2.17	2.35	2.38	2.41	2.43	2.48
2.48	2.5	2.53	2.55	2.55	2.56	2.59	2.67	2.73	2.74	2.76	2.77	2.79	2.81
2.81	2.82	2.83	2.85	2.87	2.88	2.93	2.95	2.96	2.97	2.97	3.09	3.11	3.11
3.15	3.15	3.19	3.19	3.22	3.22	3.27	3.28	3.31	3.31	3.33	3.39	3.39	3.51
3.56	3.6	3.65	3.68	3.68	3.68	3.70	3.75	4.2	4.38	4.42	4.7	4.9	4.91
5.08	5.56												

In this study, Maxwell distribution is considered for modelling purposes. The modelling performance of the Maxwell distribution is compared with the performances of EE and EW distributions using well-known criteria such as Akaike Information Criterion (AIC) and corrected AIC (AICc). The smaller value of the AIC and AICc imply better fitting.

The parameter estimates along with $\ln L$, AIC and AICc values are given in Table 4. The results show that the Maxwell distribution performs a better modeling performance than its rivals in terms of considered criteria.

Table 4: Parameter estimates for breaking stress of carbon fibres data.

		$\hat{\mu}$	$\hat{\sigma}$	$\ln L$	AIC	AICc	
Maxwell Distribution	ML	0.1402	2.1869	-141.6621	287.3242	287.4479	
	MML	0.1816	2.1636	-141.7226	287.4452	287.5689	
		$\hat{\alpha}_{ML}$	$\hat{\beta}_{ML}$	$\hat{\sigma}_{ML}$	$\ln L$	AIC	AICc
Exponentiated Weibull		1.3169	2.4091	2.6824	-141.3320	288.6640	288.9140
Exponentiated Exponential		7.7883	—	0.9870	-146.1823	296.3646	296.4883

It is also clear from the $\ln L$ values given in Table 4 that the ML estimates are preferable over the MML estimates. However, the ML estimates are obtained via the iterative method. On the other hand, the MML estimates are obtained easily since they are formulated explicitly. Furthermore, $\ln L$ values based on the ML and MML estimates do not differ so much. Therefore, the MML estimates can also be preferable for this data. It should be also noted that the Maxwell distribution provides better modelling performance than the EW distribution in spite of the fact that it has a lower number of parameters.

6.2. Example 2: Windmill data

The windmill data, in Table 5, was first considered by Joglekar *et al.* [14]. See also Kotb and Raqab [18], where the modified Weibull distribution is used for modelling this data set.

Table 5: Observations on windmill data, $n = 25$.

0.123	0.5	0.558	0.653	1.057	1.137	1.144	1.194	1.501	1.562
1.582	1.737	1.800	1.822	1.866	1.930	2.088	2.112	2.166	2.179
2.236	2.294	2.303	2.310	2.386					

In this study, the Maxwell distribution is used to model the windmill data. Its modelling performance is also compared with the modelling performance of the modified Weibull distribution. The results are given in Table 6.

Table 6: Parameter estimates for windmill data.

		$\hat{\mu}$	$\hat{\sigma}$	$\ln L$	AIC	AICc	
Maxwell Distribution	ML	-0.1640	1.5393	-25.9676	55.9351	56.4806	
	MML	-0.0905	1.5103	-26.0949	56.1898	56.7353	
		$\hat{\alpha}_{ML}$	$\hat{\beta}_{ML}$	$\hat{\theta}_{ML}$	$\ln L$	AIC	AICc
Modified Weibull		0.2249	6.4644	0.0080	-25.7511	57.5022	58.6451

It can be concluded from Table 6 that the Maxwell distribution is preferable over the modified Weibull distribution according to the AIC and AICc criteria. The MML estimates can also be used as an alternative to the ML estimates here since the results are similar. Furthermore, the MML estimators have closed forms unlike the ML estimators.

7. CONCLUSION

In this study, estimation of the location and scale parameters of the Maxwell distribution is considered. Since the ML estimators cannot be obtained explicitly, the MML estimators having closed forms are derived. The MML estimators are asymptotically equivalent to the ML estimators. They are also fully efficient. We conducted a MC simulation study to compare the performance of the MML estimators with the ML, MoM and LS estimators. Simulation results show that the performance of the ML estimators is better than the other estimators. Furthermore, the MML and ML estimators have more or less the same performance. However, the ML estimators are obtained based on iterative methods. It is well known that using iterative methods causes some problems as mentioned in the text. On the other hand, the MML estimators are easily obtained from the sample observations without any iterative computations. It is concluded that the MML estimators may be preferable as an alternative to the ML estimators, if our focus is to avoid the computational complexities whilst high efficiency.

A. APPENDIX

Elements of the Hessian matrix

$$\frac{\partial^2 \ln L}{\partial \mu^2} = -\frac{2n}{\sigma^2} - \frac{2}{\sigma^2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^{-2},$$

$$\frac{\partial^2 \ln L}{\partial \mu \partial \sigma} = -\frac{4}{\sigma^2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right),$$

$$\frac{\partial^2 \ln L}{\partial \sigma^2} = \frac{3n}{\sigma^2} - \frac{6}{\sigma^2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2.$$

Fisher Information (I) matrix of the Maxwell distribution

$$\mathbf{I} = \begin{bmatrix} -\mathbf{E} \left(\frac{\partial^2 \ln L}{\partial \mu^2} \right) & -\mathbf{E} \left(\frac{\partial^2 \ln L}{\partial \mu \partial \sigma} \right) \\ -\mathbf{E} \left(\frac{\partial^2 \ln L}{\partial \sigma \partial \mu} \right) & -\mathbf{E} \left(\frac{\partial^2 \ln L}{\partial \sigma^2} \right) \end{bmatrix} = \frac{n}{\sigma^2} \begin{bmatrix} 6 & \frac{8}{\sqrt{\pi}} \\ \frac{8}{\sqrt{\pi}} & 6 \end{bmatrix}.$$

REFERENCES

- [1] ACITAS, S.; KASAP, P.; SENOGLU, B. and ARSLAN, O. (2013). One-step M -estimators: Jones and Faddy's skewed t -distribution, *Journal of Applied Statistics*, **40**(7), 1545–1560.
- [2] AKGUL, F.G.; SENOGLU, B. and ARSLAN, T. (2016). An alternative distribution to Weibull for modeling the wind speed data: Inverse Weibull distribution, *Energy Conversion and Management*, **114**(7), 1234–1240.
- [3] AL-BALDAWI, T.H.K. (2013). Comparison of maximum likelihood and some Bayes estimators for Maxwell distribution based on non-informative priors, *Baghdad Science Journal*, **10**(2), 480–488.
- [4] AL-SOBHI, M.M. and SOLIMAN, A.A. (2016). Estimation for the exponentiated Weibull model with adaptive Type-II progressive censored schemes, *Applied Mathematical Modelling*, **40**, 1180–1192.
- [5] ARSLAN, T.; ACITAS, S. and SENOGLU, B. (2017). *Estimating the location and scale parameters of the Maxwell distribution*, Data Science, Statistics and Visualisation Conference, Lisboa.
- [6] ARSLAN, T. and SENOGLU, B. (2018). Type II censored samples in experimental design under Jones and Faddy's skew t distribution, *Iranian Journal of Science and Technology Transaction A: Science*, **42**, 2145–2157.
- [7] BARNETT, V.D. (1966). Evaluation of the maximum likelihood estimator when the likelihood equation has multiple roots, *Biometrika*, **53**, 151–165.
- [8] BARTLETT, M.S. (1953). Approximate confidence intervals, *Biometrika*, **40**, 12–19.
- [9] DEY, S.; DEY, T.; ALI, S. and MULEKAR, M.S. (2016). Two-parameter Maxwell distribution: Properties and different methods of estimation, *Journal of Statistical Theory and Practice*, **10**(2), 291–300.
- [10] DEY, S. and MAITI, S.S. (2010). Bayesian estimation of the parameter of Maxwell distribution under the different loss functions, *Journal of Statistical Theory and Practice*, **40**(2), 279–287.
- [11] DUBEY, S.D. (1967). On some permissible estimators of the location parameter of the Weibull and certain other distributions, *Technometrics*, **9**(2), 293–307.
- [12] FAN, G. (2016). Estimation of the loss and risk functions of parameter of Maxwell distribution, *Science Journal of Applied Mathematics and Statistics*, **4**(4), 129–133.
- [13] HOSSAIN, A.M. and HUERTA, G. (2016). Bayesian estimation and prediction for the Maxwell failure distribution based on Type II censored data, *Open Journal of Statistics*, **6**, 49–60.
- [14] JOGGLEKAR, G.; SCHUENEMEYER, J.H. and LARICCIA, V. (1989). Lack-of-fit testing when replicates are not available, *Amer. Statist.*, **43**, 135–143.
- [15] KANTAR, Y.M. and SENOGLU, B. (2008). A comparative study for the location and scale parameters of the Weibull distribution with given shape parameter, *Computers & Geosciences*, **34**, 1900–1909.
- [16] KAZMI, S.M.A.; ASLAM, M. and ALI, S. (2011). A note on the maximum likelihood estimators for the mixture of Maxwell distributions using Type-I censored scheme, *The Open Statistics and Probability Journal*, **3**, 31–35.
- [17] KENDALL, M.G. and STUART, A. (1961). *The Advanced Theory of Statistics*, Vol. 2, Charles Griffin and Co., London.
- [18] KOTB, M.S. and RAQAB, M.Z. (2017). Inference and prediction for modified Weibull distribution based on doubly censored samples, *Mathematics and Computers in Simulation*, **132**, 195–207.

- [19] LI, L. (2016). Minimax estimation of the parameter of Maxwell distribution under the different loss functions, *American Journal of Theoretical and Applied Statistics*, **5**(4), 202–207.
- [20] MATHAI, A.M. and PRINCY, T. (2017). Multivariate and matrix-variate analogues of Maxwell–Boltzmann and Rayleigh densities, *Pyhsica A*, **468**, 668–676.
- [21] MAXWELL, J.C. (1860). Illustrations of the dynamical theory of gases. Part I. On the motions and collisions of perfectly elastic spheres, *Philosophical Magazine*, 4th series, **19**, 19–32.
- [22] NICOLAS, M.D. and PADGETT, W.J. (2006). A bootstrap control chart for Weibull percentiles, *Qual. Rehab. Eng. Int.*, **22**, 141–151.
- [23] PUTHENPURA, S. and SINHA, N.K. (1986). Modified maximum likelihood method for the robust estimation of system parameters from very noise data, *Automatica*, **22**, 231–235.
- [24] QIAN, L. (2012). The Fisher information matrix for three-parameter exponentiated Weibull distribution under type II censoring, *Statistical Methodology*, **9**, 320–329.
- [25] SENOGLU, B. (2007). Estimating parameters in one-way analysis of covariance model with short-tailed symmetric error distributions, *Journal of Computational and Applied Mathematics*, **201**, 275–283.
- [26] SENOGLU, B. and TIKU, M.L. (2004). Censored and truncated samples in experimental design under non-normality, *Statistical Methods*, **6**(2), 173–199.
- [27] SWAIN, J.; VENKATRAMAN, S. and WILSON, J. (1988). Least squares estimation of distribution functions in Johnson’s translation system, *Journal of Statistical Computation and Simulation*, **29**, 271–297.
- [28] TIKU, M.L. (1967). Estimating the mean and standard deviation from a censored normal sample, *Biometrika*, **54**, 155–165.
- [29] TIKU, M.L. (1968). Estimating the parameters of Normal and Logistic distributions form censored samples, *Aust. J. Stat.*, **10**, 64–74.
- [30] TIKU, M.L. (1982). Testing linear contrast of means in experimental design without assuming normality and homogeneity of variances, *Biometrical Journal*, **24**(6), 613–627.
- [31] TYAGI, R.K. and BATTACHARYA, S.K. (1989a). Bayes estimation of the Maxwell’s velocity distribution function, *Statistica*, **4**, 563–567.
- [32] TYAGI, R.K. and BATTACHARYA, S.K. (1989b). A note on the MVU estimation of the Maxwell’s failure distribution, *Estadistica*, **41**, 73–79.
- [33] VAUGHAN, D.C. (1992). On the Tiku–Suresh method of estimation, *Commun. Stat. – Theory Meth.*, **21**(2), 451–469.

SEMIPARAMETRIC ADDITIVE BETA REGRESSION MODELS: INFERENCE AND LOCAL INFLUENCE DIAGNOSTICS

Authors: GERMÁN IBACACHE-PULGAR

- Institute of Statistics, Universidad de Valparaíso,
Valparaíso, Chile
- Centro Interdisciplinario de Estudios Atmosféricos y Astroestadística,
Universidad de Valparaíso, Valparaíso, Chile
german.ibacache@uv.cl

JORGE FIGUEROA-ZUÑIGA

- Department of Statistics, Universidad de Concepción,
Concepción, Chile
jfigueroaz@udec.cl

CAROLINA MARCHANT

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

Received: September 2018

Revised: April 2019

Accepted: April 2019

Abstract:

- In this paper, we study a semiparametric additive beta regression model using a parameterization based on the mean and a dispersion parameter. This model is useful for situations where the response variable is continuous and restricted to the unit interval, in addition to being related to other variables through a semiparametric regression structure. First, we formulate the model and then estimation of its parameters is discussed. A back-fitting algorithm is derived to attain the maximum penalized likelihood estimates by using natural cubic smoothing splines. We provide closed-form expressions for the score function, Fisher information matrix and its inverse. Local influence methods are derived as diagnostic tools. Finally, a practical illustration based on real data is presented and discussed.

Keywords:

- *beta distribution; diagnostic techniques; maximum penalized likelihood estimates; penalized likelihood function; semiparametric additive models.*

AMS Subject Classification:

- 62D05, 62F99, 62J99.

1. INTRODUCTION

Since the beta regression (BR) model was introduced by Ferrari and Cribari-Neto [13] (2004), it has become an excellent tool for modeling continuous data in the unit interval. This is mainly because of the flexibility of its probability density function (PDF), which can cover a wide range of shapes (symmetric, asymmetric, unimodal and bimodal) depending on different values of its parameters. In addition, the beta distribution can be parameterized in terms of its mean; see more details in Cribari-Neto and Zeileis [5] (2010), Figueroa-Zuñiga *et al.* [16] (2013), Zhao *et al.* [39] (2014), Queiroz da-Silva and Migon [30] (2016) and Huerta *et al.* [21] (2018).

Inclusion of nonparametric functions enhances the modeling flexibility for accommodating nonlinear effects of covariates. Semiparametric models have been successfully used to describe nonlinear components. Semiparametric additive beta regression (SABR) models emerge as a useful tool to describe situations where the response variable is continuous, restricted to the unit interval and related to covariates through a semiparametric regression structure (Zhu and Lee [11], 2003).

Note that parameter estimation in BR models, and consequently in SABR models, can be influenced by outlying observations. For this reason, diagnostic is a fundamental stage in the modeling of data. Diagnostic techniques used in a regression model can be divided into global influence (elimination of cases) and local influence. The main idea of the local influence technique, proposed by Cook [4] (1986), is to evaluate the sensitivity of the parameter estimators when small perturbations are introduced in the assumptions of the model or in the data (for example, in the response and covariates). This technique has the advantage, with respect to elimination of cases, that does not need to calculate the parameters estimates for each case eliminated. The following works are related to the local influence technique: Zhu and Lee [40] (2003) considered it in generalized linear mixed models; Zhu *et al.* [41] (2003) and Ibacache-Pulgar and Paula [22] (2011) provided local influence measures to evaluate the sensitivity of the maximum penalized likelihood (MPL) estimates in normal and Student-t partially linear models, respectively; Osorio *et al.* [28] (2007) derived it in elliptical linear models for longitudinal data; Cao and Lin [3] (2011) applied it to elliptical linear mixed models with first-order autoregressive errors; Ibacache-Pulgar *et al.* [24, 23] (2012, 2013) analyzed it in elliptical semiparametric mixed and symmetric semiparametric additive models, respectively; Uribe-Opazo *et al.* [36] (2012) and Garcia-Papani *et al.* [17] (2018) used it to evaluate sensitivity in spatial models; Zhang *et al.* [38] (2015) and Ibacache-Pulgar and Reyes [25] (2018) developed it for normal and elliptical partially varying-coefficient models, respectively; Emami [8] (2017) utilized it in Liu penalized least squares estimators; Marchant *et al.* [27] (2016) considered it in multivariate regression models; Ferreira and Paula [15] (2017) extended it for different perturbation schemes considering a skew-normal partially linear model; Leao *et al.* [26] (2018) derived it in cure rate models with frailties; Cysneiros *et al.* [6] (2019) implemented it in Cobb–Douglas type models; and Tapia *et al.* [33, 34] (2019a, b) applied it to mixed effects logistic and longitudinal count regression models. In the case of BR models, Espinheria *et al.* [9, 10] (2008a, b) derived the local influence technique under different perturbation schemes; Ferrari [12] (2011) derived it in BR models with varying dispersion; and Rocha and Simas [31] (2011) applied it to a general class of the BR models.

Note that the SABR model is a particular case of the GAMLSSS models (Stasinopoulos and Rigby [32], 2007). To the best of our knowledge, local influence diagnostics in SABR models have been no analyzed to the date. Therefore, the aim of this paper is to study the parameters estimation and to apply the approach of local influence in the SABR model.

This paper is organized as follows. In Section 2, the SABR model is presented and a penalized log-likelihood function is considered for the parameters estimation. In this section, we present a weighted back-fitting algorithm to obtain MPL estimates and parameters smoothing selection. Section 3 discusses and derives the local influence curvatures and Section 4 illustrates the proposed methodology with a real data set. Finally, in Section 5 some concluding remarks are mentioned.

2. THE SEMIPARAMETRIC ADDITIVE BETA REGRESSION MODEL

2.1. Formulation

Let Y_1, \dots, Y_n be independent random variables following a beta distribution, where each Y_i has a PDF given by

$$(2.1) \quad f_Y(y_i; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu_i\phi)\Gamma([1-\mu_i]\phi)} y_i^{\mu_i\phi-1} [1-y_i]^{[1-\mu_i]\phi-1}, \quad i = 1, \dots, n,$$

with $0 < y_i < 1$, $0 < \mu_i < 1$ and $\phi > 0$. Then, the mean of Y_i can be written as

$$(2.2) \quad g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta},$$

where $g(\cdot)$ is a strictly monotonic and twice differentiable link function that maps $(0, 1)$ into real numbers set; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is a vector of unknown regression parameters; and $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ is a vector of observed covariates ($p < n$).

The SABR models are often used in research related to longitudinal, clustered and spatial sampling schemes. The mean of this model can be obtained from (2.2) as

$$(2.3) \quad g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + f_1(t_{1_i}) + \dots + f_s(t_{s_i}),$$

or alternatively as $g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{n}_{1_i}^\top f_1 + \dots + \mathbf{n}_{s_i}^\top f_s$, $g(\cdot)$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$, and $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ are such as in (2.2), but now we add the nonparametric structure by f_k 's, which are unknown smooth arbitrary functions on covariates t_k 's, for $k = 1, \dots, s$; where $\mathbf{n}_{k_i}^\top$ denotes the i -th row of the incidence matrix \mathbf{N}_k whose (i, l) -th element corresponds to the indicator function $I(t_{k_i} = t_{k_l}^0)$, with $t_{k_l}^0$, for $l = 1, \dots, r_k$, denoting the distinct and ordered values of the covariate t_k and $\mathbf{f}_k = (f_k(t_{k_1}), \dots, f_k(t_{k_{r_k}}))^\top$. There are several possible choices for the link function $g(\cdot)$. For instance, one can use the logit specification $g(\mu) = \log\{\mu/(1-\mu)\}$, the probit function $g(\mu) = \Phi^{-1}(\mu)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function, the complementary log-log link $g(\mu) = \log\{-\log(1-\mu)\}$, and the log-log link $g(\mu) = -\log\{-\log(\mu)\}$, among others. A particularly useful link function is the logit link, in which case we write

$$\mu_i = \frac{\exp\left[\mathbf{x}_i^\top \boldsymbol{\beta} + \sum_{k=1}^s \mathbf{n}_{k_i}^\top \mathbf{f}_k\right]}{1 + \exp\left[\mathbf{x}_i^\top \boldsymbol{\beta} + \sum_{k=1}^s \mathbf{n}_{k_i}^\top \mathbf{f}_k\right]}.$$

Since the functions f_k 's belong to the infinite dimensional space and are considered parameters with respect to the expected value of Y_i , some restricted subspace should be defined for the nonparametric functions to ensure identifiability of the parameters associated with the model. Therefore, we assume that the function f_k belongs to the Sobolev function space (Adams and Fournier [1], 2003) defined as $\mathcal{W}_2^{(2)} = \{f_k : f_k, f_k^{(1)} \text{ abs. cont.}, f_k^{(2)} \in \mathcal{L}^2[a_k, b_k]\}$, where $f_k^{(2)}(t_k) = \partial^2/\partial t_k^2 f_k(t_k)$, with $t_k^0 \in [a_k, b_k]$. Then, the log-likelihood function of the model defined in (2.1) and (2.3) is given by

$$(2.4) \quad \ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}),$$

where

$$\ell_i(\boldsymbol{\theta}) = \log(\Gamma(\phi)) - \log(\Gamma(\mu_i\phi)) - \log(\Gamma([1-\mu_i]\phi)) + [\mu_i\phi - 1] \log(y_i) + \{[1-\mu_i]\phi - 1\} \log[1-y_i],$$

with μ_i defined in (2.2) and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \mathbf{f}_1^\top, \dots, \mathbf{f}_s^\top, \phi)^\top \in \Theta \subseteq \mathbb{R}^{p^*}$, with $p^* = p + r + 1$ and $r = \sum_{k=1}^s r_k$. Incorporating a penalty function over each function f_k , we have that the penalized log-likelihood function can be expressed as

$$(2.5) \quad \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \ell(\boldsymbol{\theta}) - \sum_{k=1}^s \frac{\alpha_k}{2} \mathbf{f}_k^\top \mathbf{K}_k \mathbf{f}_k,$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_s)^\top$ denotes an $(s \times 1)$ vector of smoothing parameters and \mathbf{K}_k is a $(r_k \times r_k)$ nonnegative definite matrix that depends only on the knots $t_{k,l}^0$, for $l = 1, \dots, r_k$. Details about the construction of this matrix can be found in Green and Silverman [18] (1994). Note that direct maximization of the log-likelihood function, without imposing restrictions on smooth functions, can generate problems of identifiability or over-fitting. To correct these problems, it is suggested to incorporate a penalty term for each smooth function in the log-likelihood function. Then, the MPL estimates are obtained by maximizing this function. As the resulting estimation equations are non-linear, an iterative process is required to obtain the parameter estimates. Therefore, in the analysis of local influence presented in Section 3, the MPL estimate is replaced by an estimate obtained in the last iteration of the process, after reaching convergence.

2.2. Estimation

In order to define the penalized score function, consider \mathbf{X} being an $(n \times p)$ matrix whose i -th row is \mathbf{x}_i^\top , \mathbf{N}_k being an $(n \times r_k)$ matrix which i -th row is $\mathbf{n}_{k,i}^\top$, $\mathbf{T} = \text{diag}(1/g'(\mu_1), \dots, 1/g'(\mu_n))$, $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^\top$, $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_n^*)^\top$, $y_i^* = \log(y_i/[1-y_i])$ and $\mu_i^* = \psi(\mu_i\phi) - \psi([1-\mu_i]\phi)$, for $i = 1, \dots, n$, with $\psi(\cdot)$ denoting the digamma function, this is, $\psi(z) = d \log \Gamma(z)/dz$, for $z > 0$. Then, assuming that (2.5) is regular with respect to $\boldsymbol{\beta}, \mathbf{f}_1, \dots, \mathbf{f}_s$ and ϕ , the penalized score function of $\boldsymbol{\theta}$ is defined as

$$\mathbf{U}_p(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{\partial \ell_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta}}.$$

After some algebraic manipulations we have in matrix form that

$$\begin{aligned} \frac{\partial \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} &= \phi \mathbf{X}^\top \mathbf{T}[\mathbf{y}^* - \boldsymbol{\mu}^*], \\ \frac{\partial \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial f_k} &= \phi \mathbf{N}_k^\top \mathbf{T}[\mathbf{y}^* - \boldsymbol{\mu}^*] - \alpha_k \mathbf{K}_k f_k, \quad k = 1, \dots, s, \\ \frac{\partial \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \phi} &= \sum_{i=1}^n \left\{ \mu_i [y_i^* - \mu_i^*] + \log(1 - y_i) - \psi[(1 - \mu_i)\phi] + \psi(\phi) \right\}. \end{aligned}$$

2.3. Weighted back-fitting algorithm

To estimate $\boldsymbol{\theta}$ by the MPL method, we have to solve the $U_p(\boldsymbol{\theta}) = \mathbf{0}$. However, the estimating equations are nonlinear and require an iterative method. For example, the determination of the MPL estimates $\hat{\boldsymbol{\theta}}$ can be performed by using the Fisher scoring algorithm. Let $f_0 = \boldsymbol{\beta}$ and $\mathbf{N}_0 = \mathbf{X}$, and consider for simplicity $\boldsymbol{\alpha}$ and \mathbf{W} fixed, with \mathbf{W} defined in the [Appendix](#). Then, the Fisher scoring algorithm is given by (see [Ibacache-Pulgar et al. \[23\], 2013](#))

$$(2.6) \quad \begin{pmatrix} \mathbf{I} & \mathbf{S}_0^{(u)} \mathbf{N}_1 & \cdots & \mathbf{S}_0^{(u)} \mathbf{N}_s \\ \mathbf{S}_1^{(u)} \mathbf{N}_0 & \mathbf{I} & \cdots & \mathbf{S}_1^{(u)} \mathbf{N}_s \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_s^{(u)} \mathbf{N}_0 & \mathbf{S}_s^{(u)} \mathbf{N}_1 & \cdots & \mathbf{I} \end{pmatrix} \begin{pmatrix} f_0^{(u+1)} \\ f_1^{(u+1)} \\ \vdots \\ f_s^{(u+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_0^{(u)} \mathbf{z}^{(u)} \\ \mathbf{S}_1^{(u)} \mathbf{z}^{(u)} \\ \vdots \\ \mathbf{S}_s^{(u)} \mathbf{z}^{(u)} \end{pmatrix},$$

where $\mathbf{z}^{(u)} = \boldsymbol{\eta} + \mathbf{W}^{-1} \mathbf{T}[\mathbf{y}^* - \boldsymbol{\mu}^*] |_{\boldsymbol{\theta}^{(u)}}$ with $\boldsymbol{\eta} = (g(\mu_1), \dots, g(\mu_n))^\top$ and $g(\cdot)$ as given in [\(2.3\)](#), and $\mathbf{S}_k^{(u)} = \mathbf{S}_k |_{\boldsymbol{\theta}^{(u)}}$, with

$$\mathbf{S}_k = \begin{cases} (\mathbf{N}_0^\top \mathbf{W} \mathbf{N}_0)^{-1} \mathbf{N}_0^\top \mathbf{W}, & k = 0, \\ (\mathbf{N}_k^\top \mathbf{W} \mathbf{N}_k + \alpha_k \mathbf{K}_k)^{-1} \mathbf{N}_k^\top \mathbf{W}, & k = 1, \dots, s. \end{cases}$$

As it is known, the back-fitting algorithm is a simple iterative procedure used to fit a generalized additive model; see [Hastie et al. \[20\] \(2001\)](#). Then, in our case, the back-fitting (Gauss–Seidel) iterations ([Hastie and Tibshirani \[19\], 1990](#)) that are used to solve the equations system [\(2.6\)](#) take the form

$$(2.7) \quad f_k^{(u+1)} = \mathbf{S}_k^{(u)} \left[\mathbf{z}^{(u)} - \sum_{l=0, l \neq k}^s \mathbf{N}_l f_l^{(u)} \right].$$

Note that the system of equations given in [\(2.6\)](#) is consistent and the back-fitting algorithm given in [\(2.7\)](#) converges to a solution for any starting values if the weight matrix involved is symmetric and defined positive; see [Berhane and Tibshirani \[2\] \(1998\)](#). In addition, we have that this solution is unique under no concavity in the data. In particular, for a model with smooth terms f_1 and f_2 but without the constant terms, $\boldsymbol{\beta}$, we have the following considerations:

- (i) If $\|\mathbf{S}_1 \mathbf{S}_2\| < 1$, the estimating equations are consistent and have a unique solution, and the final iterations from the back-fitting algorithm are independent of the starting values and starting order.

- (ii) If $\|\mathbf{S}_1\mathbf{S}_2\| = 1$, this gives an indication of concurvity in the data (strict collinearity), and therefore the back-fitting algorithm converges to one of the solutions of estimating equations system, and the starting functions determine the final solutions.
- (iii) If the \mathbf{S}_k smoothers are not centered ($\mathbf{S}_k^\top \mathbf{1} = \mathbf{1}$), typically $\|\mathbf{S}_1\mathbf{S}_2\| = 1$, we can consider a centered smoother such that $\mathbf{S}_k^\top \mathbf{1} = \mathbf{0}$, with $\mathbf{1}$ denoting a $(r_k \times 1)$ vector of ones, which is defined as

$$\mathbf{S}_k = \left(\mathbf{I}_{(r_k, r_k)} - \frac{\mathbf{1}\mathbf{1}^\top}{r_k} \right) \left(\mathbf{N}_k^\top \mathbf{W} \mathbf{N}_k + \alpha_k \mathbf{K}_k \right)^{-1} \mathbf{N}_k^\top \mathbf{W}.$$

The MPL estimate of the scale parameter, $\hat{\phi}$, can be obtained by solving the following iterative process (see Ibacache-Pulgar and Reyes [25], 2018):

$$\phi^{(u+1)} = \phi^{(u)} - \mathbb{E} \left\{ \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \phi \partial \phi} \right\}^{-1} \frac{\partial \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \phi} \Big|_{\boldsymbol{\theta}^{(u)}}.$$

Note that Equation (2.7), which involves a diagonal matrix denoted by \mathbf{W} , leads to an iterative weighted back-fitting solution. The convergence of the iterative process is guaranteed by the diagonal structure of \mathbf{W} . Note also that this matrix must be updated in each iteration of the back-fitting iterative process and in each stage of the Fisher scoring algorithm.

Observe that the joint iterative procedure proposed to estimate the parameters of the model is based on the Fisher scoring and back-fitting algorithms. First, note that Equation (2.6) corresponds to the matrix equation of the Fisher scoring algorithm. Then, after algebraic manipulations, the solutions to this system correspond precisely to the back-fitting iterations. In addition, note also that the scale parameter is estimated by a Fisher scoring algorithm. In summary, the Fisher scoring algorithm allows us to estimate the parameter vector associated with our model and the back-fitting algorithm to update the estimates of the parameters associated with the parametric and nonparametric components of the model for each stage of the Fisher score algorithm.

2.4. Approximate covariance matrix

The covariance matrix of $\hat{\boldsymbol{\theta}}$ is obtained from the inverse of the expected information matrix \mathcal{I}_p defined in the Appendix. Therefore, the approximate covariance matrix of $\hat{\boldsymbol{\theta}}$ is given as $\widehat{\text{Cov}}(\hat{\boldsymbol{\theta}}) \approx \mathcal{I}_p^{-1} |_{\hat{\boldsymbol{\theta}}}$, where

$$\mathcal{I}_p^{-1} = \begin{pmatrix} \mathcal{J}_1^{-1} & -\mathcal{I}_{11}^{-1} \mathcal{I}_{12} \mathcal{J}_2^{-1} \\ -\mathcal{I}_{22} \mathcal{I}_{21} \mathcal{J}_1^{-1} & \mathcal{J}_2^{-1} \end{pmatrix},$$

with $\mathcal{J}_1 = \mathcal{I}_{11} - \mathcal{I}_{12} \mathcal{I}_{22}^{-1} \mathcal{I}_{21}$ and $\mathcal{J}_2 = \mathcal{I}_{22} - \mathcal{I}_{21} \mathcal{I}_{11}^{-1} \mathcal{I}_{12}$. An approximate pointwise standard error band ($\text{SEB}_{\text{approx}}$) for $f_k(\cdot)$, that allows us to assess how accurate the estimator $\hat{f}_k(\cdot)$, can be defined as

$$\text{SEB}_{\text{approx}}(f_k(t_l^0)) = \hat{f}_k(t_l^0) \pm 2\sqrt{\widehat{\text{Var}}(\hat{f}_k(t_l^0))}, \quad l = 1, \dots, r,$$

where $\widehat{\text{Var}}(\hat{f}_k(t_l))$ is the l -th principal diagonal element of the corresponding block-diagonal matrix from \mathcal{I}_p^{-1} .

2.5. Smoothing parameters and degrees of freedom

The determination of the parameter vector $\boldsymbol{\alpha}$ is a crucial aspect in the estimation process. Different choice methods are available in the literature for this purpose. For example, an alternative to select smoothing parameters under the SABR model is to consider the Akaike information criterion — AIC — (see details in Ferreira *et al.* [14], 2012; Ventura *et al.* [37], 2019) defined by

$$\text{AIC}(\boldsymbol{\alpha}) = -2\ell_p(\widehat{\boldsymbol{\theta}}, \boldsymbol{\alpha}) + 2[p + 1 + \text{df}(\boldsymbol{\alpha})],$$

where $\ell_p(\widehat{\boldsymbol{\theta}}, \boldsymbol{\alpha})$ denotes the penalized log-likelihood function available at $\widehat{\boldsymbol{\theta}}$ for a fixed $\boldsymbol{\alpha}$ and $\text{df}(\boldsymbol{\alpha}) = \sum_{k=1}^s \text{df}(\alpha_k)$ denotes approximately the number of effective parameters involved in modeling of the nonparametric effects. The idea is to minimize the function $\text{AIC}(\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$. Following Hastie and Tibshirani [19] (1990) and Eilers and Marx [7] (1996), the degrees of freedom (DF) associated with the k -th smooth function are given by

$$\text{df}(\alpha_k) = \text{tr}\{\widetilde{\mathbf{S}}_k\} = \sum_{j=1}^{r_k} \frac{1}{1 + \alpha_k \phi L_j}, \quad j = 1, \dots, r_k,$$

which measures the individual effect contribution of the k -th component, with $\widetilde{\mathbf{S}}_k = \mathbf{N}_k \mathbf{S}_k$ and \mathbf{S}_k defined previously, and L_j are the eigenvalues of the matrix $\mathbf{Q}_{\mathbf{N}_k}^{-1/2} \mathbf{Q}_{\alpha_k} \mathbf{Q}_{\mathbf{N}_k}^{-1/2}$, with $\mathbf{Q}_{\mathbf{N}_k} = \mathbf{N}_k^\top \mathbf{W} \mathbf{N}_k$ and $\mathbf{Q}_{\alpha_k} = \alpha_k \mathbf{K}_k$. Note that the AIC is based on information theory and is useful for selecting an appropriate model and smoothing parameters given data with adequate sample size; see Ferreira *et al.* [14] (2012) and Ventura *et al.* [37] (2019).

3. LOCAL INFLUENCE DERIVATION

3.1. General context

In general, local influence analysis can be developed jointly for the entire parameter vector. However, it is important to know the influence that the observations exert separately on the estimates of the parametric components, nonparametric components and the dispersion parameter. Some works related to the application of the method of local influence in semiparametric models have revealed empirical evidence that the observations that exert an influence on the estimates of the parametric component are not necessarily influential on the estimates of the non parametric component and viceversa.

To assess the influence of perturbations on the MPL estimates $\widehat{\boldsymbol{\theta}}$, we can consider the likelihood displacement defined by $\text{LD}(\boldsymbol{\omega}) = 2[\ell_p(\widehat{\boldsymbol{\theta}}, \boldsymbol{\alpha}) - \ell_p(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}, \boldsymbol{\alpha})] \geq 0$, where $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ is the MPL estimates of $\boldsymbol{\theta}$ for a perturbed model, whose perturbed penalized log-likelihood function is denoted by $\ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega})$, and $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$ is an n -dimensional vector of perturbations restricted to some open subset $\Omega \in \mathbb{R}^n$. It is assumed that exists $\boldsymbol{\omega}_0 \in \Omega$, a vector of no perturbation, such that $\ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{\omega}_0) = \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})$. Cook [4] (1986) suggested to study the local behavior of $\text{LD}(\boldsymbol{\omega})$ around $\boldsymbol{\omega}_0$. The normal curvature at the arbitrary direction \mathbf{l} , with

$\|\mathbf{l}\| = 1$ is given by $C_{\mathbf{l}}(\hat{\boldsymbol{\theta}}) = -2\{\mathbf{l}^\top \boldsymbol{\Delta}_p^\top \ddot{\ell}_p^{-1} \boldsymbol{\Delta}_p \mathbf{l}\}$, which is the objective function of the normal curvature, where $\ddot{\ell}_p$ is the penalized Hessian matrix of $\ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})$ evaluated at $\hat{\boldsymbol{\theta}}$, and $\boldsymbol{\Delta}_p$ is a penalized perturbation matrix, with elements

$$\Delta_{pj} = \left. \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \theta_l \partial \omega_j} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}, \quad l = 1, \dots, p^*, \quad j = 1, \dots, n,$$

and $\ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})$ being the penalized log-likelihood function corresponding to the model perturbed by $\boldsymbol{\omega}$. For the SABR model, the elements of $\ddot{\ell}_p$ are given in the [Appendix](#). We consider the direction $\mathbf{l} = \mathbf{e}_i$, called total local influence of the i -th individual, where \mathbf{e}_i is an n -dimensional vector with a one at the i -th position and zeros at the remaining positions. In this case, the normal curvature takes the form $C_{\mathbf{e}_i}(\boldsymbol{\theta}) = 2|c_{ii}|$, where c_{ii} is the i -th principal diagonal element of the matrix $\mathbf{C} = \boldsymbol{\Delta}_p^\top \ddot{\ell}_p^{-1} \boldsymbol{\Delta}_p$. The index plot of \mathbf{l} may reveal those observations that under small perturbations exert a notable influence on $\hat{\boldsymbol{\theta}}$. In order to have a curvature invariant under uniform change of scale, we consider the conformal normal curvature proposed by Poon and Poon [29] (1999). This normal curvature is defined as $B_{\ell}(\boldsymbol{\theta}) = -[\mathbf{l}^\top \boldsymbol{\Delta}_p^\top \ddot{\ell}_p^{-1} \boldsymbol{\Delta}_p \mathbf{l}] / [(\text{tr}(\boldsymbol{\Delta}_p^\top \ddot{\ell}_p^{-1} \boldsymbol{\Delta}_p)^2)^{1/2}]$, and is characterized to allow for any unitary direction \mathbf{l} , with $0 \leq B_{\ell}(\boldsymbol{\theta}) \leq 1$. A suggestion is to consider, for example, the direction $\mathbf{l} = \mathbf{e}_i$ and observing the index plot of $B_{\mathbf{e}_i}(\boldsymbol{\theta})$. If our interest lies in studying the local influence on a subvector of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\theta}_1$, the normal curvature for $\boldsymbol{\theta}_1$ at the unitary direction \mathbf{l} is given by $C_{\ell}(\hat{\boldsymbol{\theta}}_1) = -2[\mathbf{l}^\top \boldsymbol{\Delta}_p^\top (\ddot{\ell}_p^{-1} - \mathbf{G}_{22}) \boldsymbol{\Delta}_p \mathbf{l}]$, where

$$\mathbf{G}_{22} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\ell}_{p22}^{-1} \end{pmatrix},$$

with $\ddot{\ell}_{p22}$ obtained from the partition of $\ddot{\ell}_p$ according to the partition of $\boldsymbol{\theta}$. In this case, the index plot of the eigenvector $\mathbf{l} = \mathbf{l}_{\max}$, which corresponds to the largest absolute eigenvalue of the matrix $\mathbf{G} = \boldsymbol{\Delta}_p^\top (\ddot{\ell}_p^{-1} - \mathbf{G}_{22}) \boldsymbol{\Delta}_p$, may indicate the points with large influence on $\hat{\boldsymbol{\theta}}_1$.

3.2. Cases-weight perturbation

Let $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$ be a weight vector. In this case, perturbed penalized log-likelihood function is given by

$$\ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega}) = \ell(\boldsymbol{\theta} | \boldsymbol{\omega}) - \sum_{k=1}^s \frac{\alpha_k}{2} \mathbf{f}_k^\top \mathbf{K}_k \mathbf{f}_k,$$

where $\ell(\boldsymbol{\theta} | \boldsymbol{\omega}) = \sum_{i=1}^n \omega_i \ell_i(\boldsymbol{\theta})$, with $0 \leq \omega_i \leq 1$ and $\boldsymbol{\omega}_0 = (1, \dots, 1)^\top$. Hence, the elements of the penalized perturbation matrix are expressed as

$$\begin{aligned} \left. \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\boldsymbol{\phi}} \mathbf{X}^\top \hat{\mathbf{T}} \hat{\mathbf{E}}, \\ \left. \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \mathbf{f}_k \partial \boldsymbol{\omega}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\boldsymbol{\phi}} \mathbf{N}_k^\top \hat{\mathbf{T}} \hat{\mathbf{E}}, \quad k = 1, \dots, s, \\ \left. \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \phi \partial \boldsymbol{\omega}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\mathbf{u}}^\top, \end{aligned}$$

where $\hat{\mathbf{E}} = \text{diag}(y_i^* - \hat{\mu}_i^*)$, for $i = 1, \dots, n$, and $\hat{\mathbf{u}} = (\hat{u}_1, \dots, \hat{u}_n)^\top$, with $u_i = \mu_i [y_i^* - \mu_i^*] + \ln(1 - y_i) - \psi[(1 - \mu_i)\phi] + \psi(\phi)$.

3.3. Response perturbation

Consider now an additive perturbation on the i -th response by making $y_{i\omega} = y_i + \omega_i$, where $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$ is the vector of perturbations. Then, under the scheme of response perturbation, the perturbed penalized log-likelihood function is constructed from (2.5) with y_i replaced by $y_{i\omega}$ and $\boldsymbol{\omega}_0 = (0, \dots, 0)^\top$. Hence, the elements of the penalized perturbation matrix take the form

$$\begin{aligned} \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\boldsymbol{\phi}} \mathbf{X}^\top \hat{\mathbf{T}} \mathbf{M}, \\ \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial f_k \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\boldsymbol{\phi}} \mathbf{N}_k^\top \hat{\mathbf{T}} \mathbf{M}, \quad k = 1, \dots, s, \\ \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \phi \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\mathbf{a}}^\top, \end{aligned}$$

where $\mathbf{M} = \text{diag}_{1 \leq i \leq n} (m_i)$ and $\mathbf{a} = (a_1, \dots, a_n)^\top$, with $m_i = 1/[y_i(1 - y_i)]$ and $a_i = -[y_i - \mu_i]/[y_i(1 - y_i)]$.

Note that the perturbation $\boldsymbol{\omega}$ must be generated in the support $[-y, 1 - y]$ in order to guarantee that the perturbed response variable retains the original distribution support. It is important to mention that the space of $\boldsymbol{\omega}$ depends on the type of perturbation that is introduced in the response variable. In our case, we consider a perturbation of additive type, but eventually we could consider a perturbation of the multiplicative type.

3.4. Continuous covariate perturbation

Consider now an additive perturbation on a continuous covariate, namely $x_{id\omega}$, by making $x_{id\omega} = x_{id} + \omega_i$, with $\omega_i \in \mathbb{R}$. Then, under the scheme of covariate perturbation, the perturbed penalized log-likelihood function is constructed from (2.5) with x_{id} replaced by $x_{id\omega}$, $\mu_{i\omega} = g^{-1}(\eta_{i\omega})$ in the place of μ_i , for $\eta_{i\omega} = \mathbf{x}_{i\omega}^\top \boldsymbol{\beta} + \mathbf{n}_{1_i}^\top f_1 + \dots + \mathbf{n}_{s_i}^\top f_s$, and $\boldsymbol{\omega}_0 = (0, \dots, 0)^\top$. Hence, the elements of the penalized perturbation matrix assumes the form

$$\begin{aligned} \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\hat{\boldsymbol{\phi}} \hat{\boldsymbol{\beta}}_d \mathbf{X}^\top \hat{\mathbf{Q}} + \hat{\boldsymbol{\phi}} \mathbf{P} \hat{\mathbf{T}} \hat{\mathbf{E}}, \\ \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial f_k \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\hat{\boldsymbol{\phi}} \hat{\boldsymbol{\beta}}_d \mathbf{N}^\top \hat{\mathbf{Q}} + \hat{\boldsymbol{\phi}} \mathbf{P} \hat{\mathbf{T}} \hat{\mathbf{E}}, \quad k = 1, \dots, s, \\ \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \phi \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= -\hat{\boldsymbol{\beta}}_d \hat{\mathbf{h}}^\top \hat{\mathbf{T}}, \end{aligned}$$

where \mathbf{P} is a $(p \times n)$ matrix of zeros except for the p -th line, which contains ones, and $\mathbf{h} = (h_1, \dots, h_n)^\top$, with $h_i = c_i - (y_i^* - \mu_i^*)$ and c_i as defined in the [Appendix](#).

3.5. Scale perturbation

Under scale parameter perturbation scheme, it is assumed that $\phi_{i\omega} = \omega_i^{-1}\phi$, with $\omega_i > 0$. Then, the perturbed penalized log-likelihood function is constructed from (2.5) with ϕ replaced by $\phi_{i\omega}$ and $\boldsymbol{\omega}_0 = (1, \dots, 1)^\top$. Hence, the elements of the penalized perturbation matrix take the form

$$\begin{aligned} \left. \frac{\partial^2 \ell_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\boldsymbol{\phi}} \mathbf{X}^\top \hat{\mathbf{T}} \hat{\mathbf{F}}, \\ \left. \frac{\partial^2 \ell_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial f_k \partial \boldsymbol{\omega}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\boldsymbol{\phi}} \mathbf{N}_k^\top \hat{\mathbf{T}} \hat{\mathbf{F}}, \quad k = 1, \dots, s, \\ \left. \frac{\partial^2 \ell_{p_i}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{\omega})}{\partial \boldsymbol{\phi} \partial \boldsymbol{\omega}^\top} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} &= \hat{\boldsymbol{\phi}} \hat{\mathbf{d}}^\top - \hat{\mathbf{u}}^\top, \end{aligned}$$

where $\mathbf{F} = \text{diag}_{1 \leq i \leq n}(F_i)$, $\mathbf{d} = (d_1, \dots, d_n)^\top$, $\mathbf{u} = (u_1, \dots, u_n)^\top$, with $u_i = \mu_i[y_i^* - \mu_i^*] + \log(1 - y_i) - \psi[(1 - \mu_i)\phi] + \psi(\phi)$, $F_i = [c_i - (y_i^* - \mu_i^*)]$ and c_i as defined in the [Appendix](#).

4. EMPIRICAL ILLUSTRATION

4.1. Data and exploratory analysis

To illustrate our methodology, we consider the Australian athletes data set that has been reported by Telford and Cunningham [35] (1991). The purpose of this study is to investigate the relationships of hematological measures with various covariates, such as height and mass, among others, for a sample of 202 elite Australian athletes who trained at the Australian Institute of Sport. The objective of the present data analysis is to model the percent body fat through the SABR model. We consider as covariates: (i) sum of skin folds (SSF), (ii) hemaglobin concentration (HG), and (iii) lean body mass (LBM), whereas the percent body fat (BFAT) is the response variable. Figure 1 contains the scatter plots between the response and each covariate. From Figure 1(a), we observe that a linear relationship between BFAT and SSF, while that Figures 1(b)–(c) show no evidence of linear relationships between BFAT and the covariates HG and LBM.

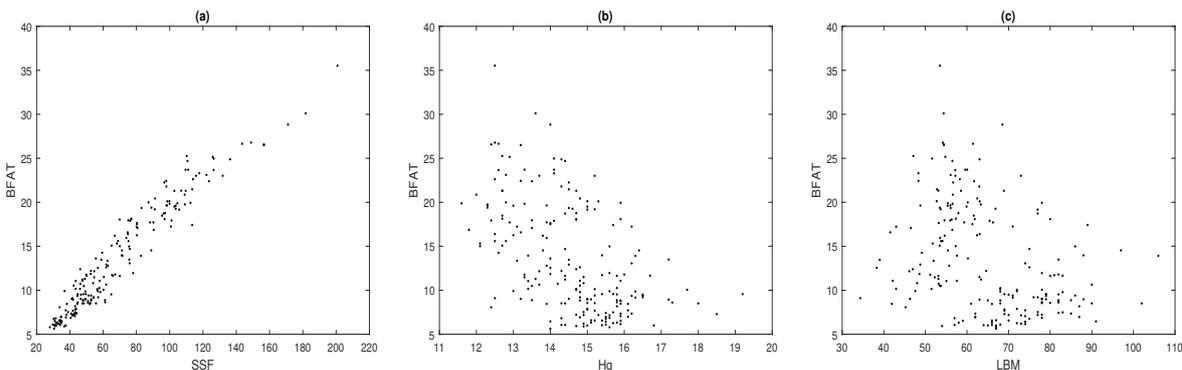


Figure 1: Scatter plots of BFAT versus SSF (a), HG (b) and LBM (c) with Australian athletes data.

4.2. Model fitting

First, we consider a BR model given by

$$g(\mu_i) = \beta_0 + \beta_1 \text{SSF}_i + \beta_2 \text{HG}_i + \beta_3 \text{LBM}_i, \quad i = 1, \dots, 202,$$

with logit link function. The maximum likelihood estimates and the corresponding approximate estimated standard error (in parenthesis) are reported in Table 1, with AIC = -1031.317. Note that under this model, we are assuming that the relationship of each covariate with the response is linear. However, as previously commented, we observe in Figures 1(b) and 1(c) that the relationships between BFAT and the covariates HG and LBM seem to be nonlinear, which suggests a SABR model with link function

$$g(\mu_i) = \beta_0 + \beta_1 \text{SSF}_i + f_1(\text{HG}_i) + f_2(\text{LBM}_i).$$

The MPL estimates and the corresponding approximate estimated standard error associated with the parametric component and scale parameter are also reported in Table 1.

Table 1: Maximum likelihood and MPL estimates and the standard error (in parenthesis) for indicated model with Australian athletes data.

Model	Parameters				
	β_0	β_1	β_2	β_3	ϕ
BR	-2.020 (0.1591)	0.012 (0.0003)	-0.027 (0.0122)	-0.005 (0.0012)	307.180 (30.5601)
SABR	-2.788 (0.0420)	0.012 (0.0004)	—	—	361.768 (35.9955)

Comparing the results reported in Table 1, we note a similarity between the estimates for $\hat{\beta}_0$ and $\hat{\beta}_1$ under both models, but the estimated standard error of $\hat{\beta}_0$ is smaller under the SABR model. However, the estimate $\hat{\phi}$ under the SABR model is larger (including its estimated standard error) than that obtained for the BR model. The estimates of the smoothing parameters α_1 and α_2 , as well as the corresponding DFs, are reported in Table 2.

Table 2: Smoothing components fitted under the SABR model to Australian athletes data.

	Smoothing function	
	$f_1(\text{HG})$	$f_2(\text{LBM})$
DF(α_k)	4.662	4.988
α_k	0.0014	0.9012

Figures 2 (a)–(b) show the estimated smooth functions under the SABR model and the corresponding approximate SEB (dashed curves). The estimated smooth functions are computed using the smoothing parameters obtained by the AIC as described in Subsection 2.5.

The graphical plots suggest nonlinear tendencies for HG and LBM. Then, we find a value of $AIC(\alpha_1, \alpha_2) = -1055.466$, which is less than that obtained under the BR model, indicating a superiority of the model that includes a nonparametric additive component.

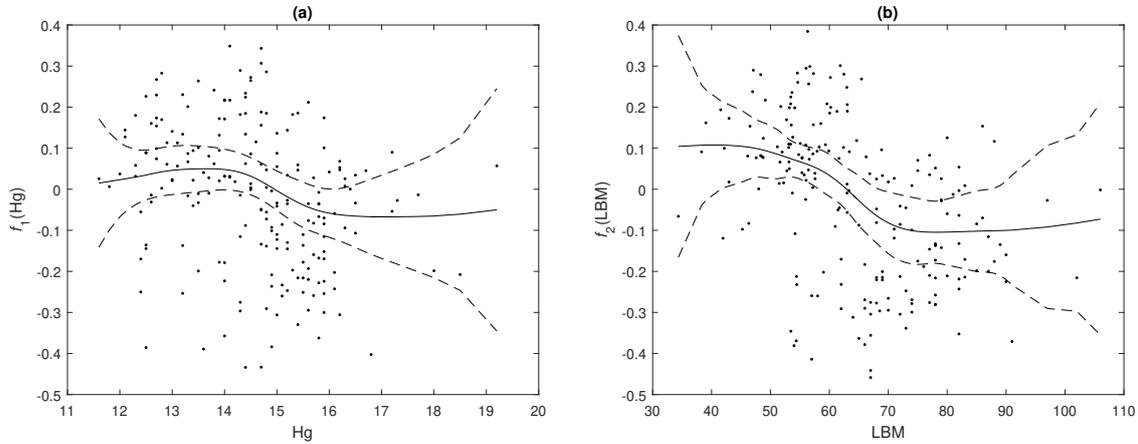


Figure 2: Plots estimated smooth functions f_1 (a) and f_2 (b) for the Australian athletes data and their approximate pointwise SEB denoted by the dashed lines.

4.3. Diagnostics

In this illustration, we consider residual proposed by Ferrari and Cribari-Neto [13] (2004) given by

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{\text{Var}}(y_i)}}, \quad i = 1, \dots, 202,$$

where $\widehat{\text{Var}}(y_i) = \hat{\mu}_i[1 - \hat{\mu}_i]/[1 + \hat{\phi}]$ and $\hat{\mu}_i = g^{-1}(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}} + \mathbf{n}_{1_i}^\top \hat{\mathbf{f}}_1 + \mathbf{n}_{2_i}^\top \hat{\mathbf{f}}_2)$, with $\hat{\boldsymbol{\beta}}, \hat{\mathbf{f}}_k$, for $k = 1, 2$, and $\hat{\phi}$ denoting the MPL estimates.

Figure 3 displays the graphical plot of the standardized ordinary residuals against the indices of the observations. We note that the residual are randomly scattered around zero and that observations #51, #53 and #56 are indicated as atypical cases. Note that a residual analysis permits us to detect deviations from the model assumptions, but also atypical cases.

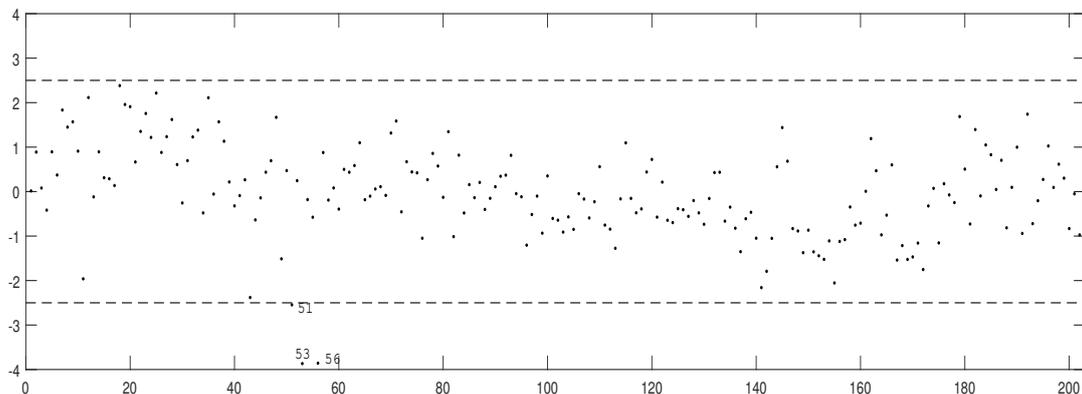


Figure 3: Plot standardized ordinary residuals versus the index of the observation.

An atypical case can be potentially influential or not, but from a scope of global influence. However, potentially influential cases detected by the displacement of likelihood functions are evaluated from a scope of local influence. In any case, this potential influence (global or local) must be studied by means of relative changes (RC) when the potentially influential case is removed from the full data set. This allows us to know whether inferential changes are generated or not.

Now, in order to identify potentially influential observations under the fitted model to Australian athletes data, we present index plots of $B_i = B_{e_i}(\boldsymbol{\lambda})$, for $\boldsymbol{\lambda} = \boldsymbol{\beta}, f_k, \phi$, with $k = 1, 2$. Figure 4 shows the index plot of B_i for the case-weight perturbation scheme under the fitted model. Looking at Figure 4, note that observations #51, #53 and #56 are more influential on the MPL estimate $\hat{\boldsymbol{\lambda}}$.

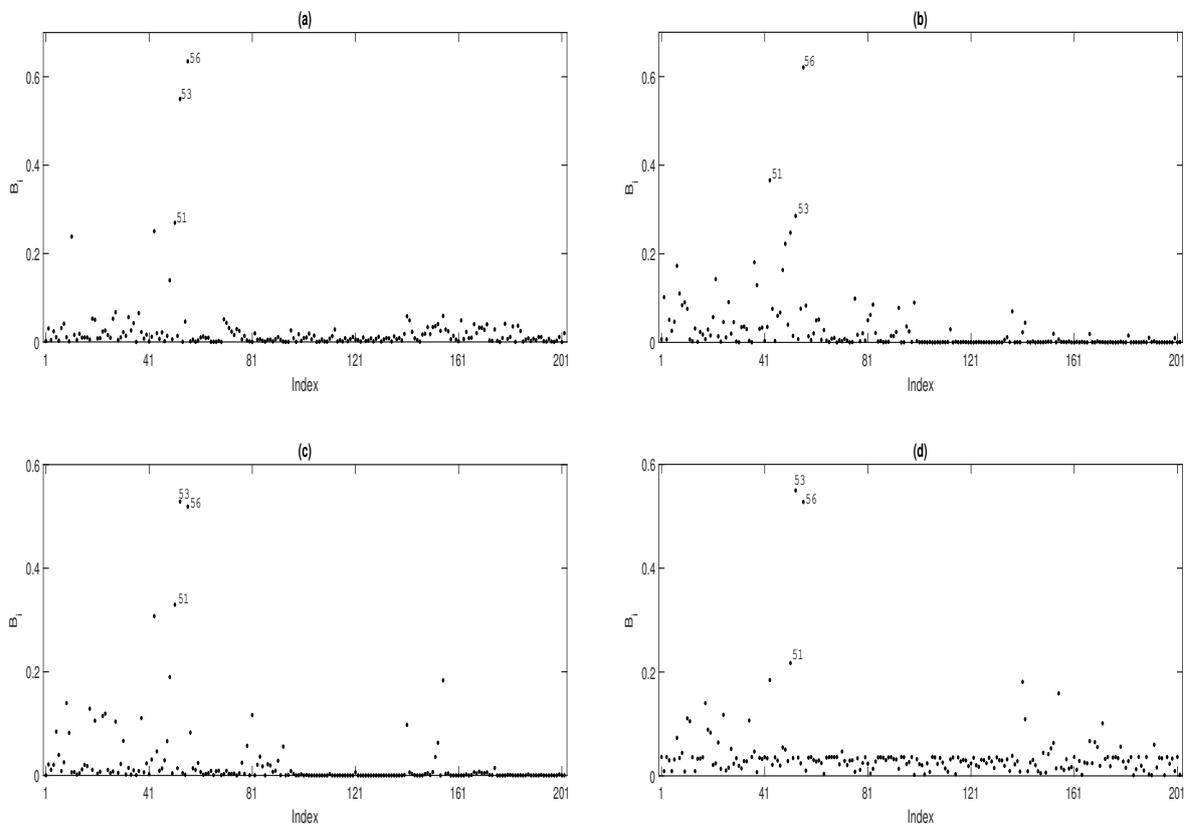


Figure 4: Index plots of B_i for assessing local influence on $\hat{\boldsymbol{\beta}}$ (a), \hat{f}_1 (b), \hat{f}_2 (c) and $\hat{\phi}$ (d) considering case-weight perturbation under model fitted to Australian athletes data.

Figure 5 presents the index plots of B_i , considering the response perturbation scheme under the fitted model. In Figure 5, observe that observations #160, #166, and #181 are more influential on MPL estimate $\hat{\boldsymbol{\lambda}} = \hat{f}_1$, whereas none observation is pointed out on the estimates remaining. The index plots of B_i for the scale parameter and covariate perturbation are omitted because the results are similar to those obtained under case-weight perturbation scheme. Note that observations #51, #53 and #56 are also detected as atypical according to the residual analysis.

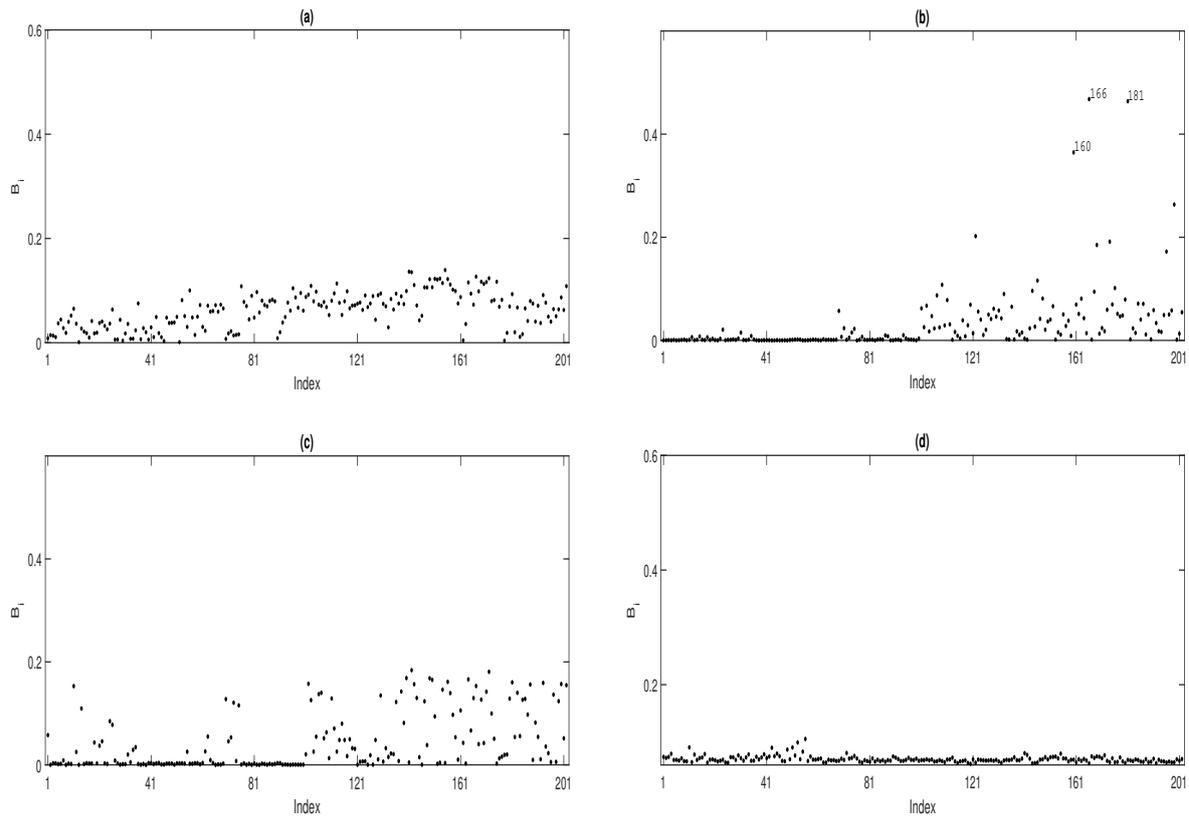


Figure 5: Index plots of B_i for assessing local influence on $\hat{\beta}$ (a), \hat{f}_1 (b), \hat{f}_2 (c) and $\hat{\phi}$ (d) considering response perturbation under the fitted model to Australian athletes data.

We now investigate the impact on the model inference when the observations detected as potentially influential in the diagnostic analysis are removed. Table 3 presents the RCs in % of the MPL estimates of β_j , for $j = 1, 2$, and ϕ after removing from the data set the pointed out observations in the local influence graphical plots under the SABR model.

Table 3: RC (in %) on the MPL estimate of β_j and ϕ under the SABR model fitted to Australian athletes data after removing the indicated cases.

Removed case	Parameters			Relative changes		
	β_0	β_1	ϕ	RC_{β_0}	RC_{β_1}	RC_{ϕ}
51	-2.8004	0.0126	374.3139	0.45	5.00	3.46
53	-2.8216	0.0129	393.3938	1.21	7.50	8.74
56	-2.8230	0.0130	395.2551	1.25	8.31	9.29
51-53	-2.8007	0.0126	372.6768	0.45	5.01	3.02
51-56	-2.8031	0.0127	374.6170	0.54	5.83	3.55
53-56	-2.8245	0.0130	394.0092	1.31	8.32	8.91
51-53-56	-2.8013	0.0126	371.6199	0.47	5.02	2.72

The RCs of each estimated parameter are defined as $RC_\psi = |(\hat{\psi} - \hat{\psi}_{(I)})/\hat{\psi}| \times 100\%$, where $\hat{\psi}_{(I)}$ denotes the MPL estimate of ψ , with $\psi = \beta_j, \phi$, after the corresponding observation(s) are removed according to the set I . Note that, although some RC are large, inferential changes are not detected. It is interesting to notice from Table 3 the coherence with the diagnostic graphical plots shown previously. For instance, elimination of the observations #51, #53 and #56, detected as potentially influential observations by local influence technique, leads to significant changes in the MPL estimate, mainly in $\hat{\beta}_1$ and $\hat{\phi}$. This indicates the need of a diagnostic examination.

5. CONCLUDING REMARKS

In this paper, we have proposed a methodology of inference and diagnostics for the semiparametric additive beta regression model. Specifically, we have derived a weighted back-fitting iterative process to estimate the parameters of the additive component of the model, that is, of regression coefficients and smooth functions. We have estimated the approximate variance-covariance matrix of maximum penalized likelihood estimates based on the Fisher information matrix obtained from the penalized log-likelihood function. Moreover, we have derived diagnostics for this model using the local influence technique to evaluate the sensitivity of the maximum penalized likelihood estimates by using several perturbation schemes in the model and data. Finally, we have performed a statistical modeling with real data set. The study has provided evidences on the advantage of incorporating a semiparametric additive term in those situations where there are covariates that contribute nonlinearly to the model. Thus, we recommend semiparametric additive beta regression models as an option to fit continuous data sets in the unit interval when covariates are present and that contribute nonlinearly to the model. The computational codes used in the illustration are available under request from the authors.

A. APPENDIX

Hessian matrix

Let $\ddot{\ell}_p(p^* \times p^*)$ be the Hessian matrix with the $(j^*; \ell^*)$ -element given by $\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha}) / \partial \theta_{j^*} \theta_{\ell^*}$, for $j^*, \ell^* = 1, \dots, p^*$. After some algebraic manipulations, we find

$$\begin{aligned} \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= -\phi \mathbf{X}^\top \mathbf{Q} \mathbf{X}, \\ \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial f_k \partial f_{k'}^\top} &= \begin{cases} -\phi \mathbf{N}_k^\top \mathbf{Q} \mathbf{N}_k - \alpha_k \mathbf{K}_k, & k = k', \\ -\phi \mathbf{N}_k^\top \mathbf{Q} \mathbf{N}_{k'}, & k \neq k', \end{cases} \\ \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial f_k^\top} &= -\phi \mathbf{X}^\top \mathbf{Q} \mathbf{N}_k, \quad k = 1, \dots, s, \end{aligned}$$

where $\mathbf{Q} = \text{diag}_{1 \leq i \leq n}(q_i)$, with

$$q_i = \left[\phi \left\{ \psi'(\mu_i \phi) + \psi'[(1 - \mu_i) \phi] \right\} + [y_i^* - \mu_i^*] \frac{g''(\mu_i)}{g'(\mu_i)} \right] \frac{1}{[g'(\mu_i)]^2}.$$

In addition, we have that the second derivative de $\ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})$ with respect to ϕ , and $\boldsymbol{\beta}$ and f_k , respectively, can be written by

$$\begin{aligned} \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \phi} &= \frac{2}{\phi^2} \mathbf{X}^\top \mathbf{b}, \\ \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial f_k \partial \phi} &= \frac{2}{\phi^2} \mathbf{N}_k^\top \mathbf{b}, \quad k = 1, \dots, s, \end{aligned}$$

where $\mathbf{b} = (b_1, \dots, b_n)^\top$, with

$$b_i = \left\{ [y_i^* - \mu_i^*] - \phi \frac{\partial \mu_i^*}{\partial \phi} \right\} \frac{1}{[g'(\mu_i)]}.$$

Furthermore, the second derivative de $\ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})$ with respect to ϕ is given by

$$\frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \phi^2} = \text{trace}(\mathbf{D}),$$

where $\mathbf{D} = \text{diag}_{1 \leq i \leq n}(d_i)$, with

$$d_i = -\left[\psi'[\mu_i \phi] \mu_i^2 + \psi'[(1 - \mu_i) \phi] [1 - \mu_i]^2 - \psi'(\phi) \right].$$

Expected information matrix

In general, by calculating the expectation of the matrix $-\ddot{\ell}_p$, we obtain the $(p^* \times p^*)$ penalized expected information matrix denoted by

$$\mathcal{I}_p = -\text{E} \left(\frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right).$$

Let $\mathbf{W} = \text{blockdiag}_{1 \leq i \leq n}(w_i)$ and $\mathbf{c} = (c_1, \dots, c_n)^\top$, with

$$w_i = \phi \left[\psi'[\mu_i \phi] + \psi'[1 - \mu_i] \phi \right] \frac{1}{[g'(\mu_i)]^2}, \quad c_i = \phi \left[\psi'[\mu_i \phi] \mu_i - \psi'[(1 - \mu_i) \phi] [1 - \mu_i] \right].$$

After some algebraic manipulations, we find

$$\mathcal{I}_p = \begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{pmatrix},$$

where

$$\mathcal{I}_{11} = \begin{pmatrix} \phi \mathbf{X}^\top \mathbf{W} \mathbf{X} & \phi \mathbf{X}^\top \mathbf{W} \mathbf{N}_1 & \cdots & \phi \mathbf{X}^\top \mathbf{W} \mathbf{N}_s \\ \phi \mathbf{N}_1^\top \mathbf{W} \mathbf{X} & \phi \mathbf{N}_1^\top \mathbf{W} \mathbf{N}_1 + \lambda_1 \mathbf{K}_1 & \cdots & \phi \mathbf{N}_1^\top \mathbf{W} \mathbf{N}_s \\ \vdots & \vdots & \ddots & \vdots \\ \phi \mathbf{N}_s^\top \mathbf{W} \mathbf{X} & \phi \mathbf{N}_s^\top \mathbf{W} \mathbf{N}_1 & \cdots & \phi \mathbf{N}_s^\top \mathbf{W} \mathbf{N}_s + \lambda_s \mathbf{K}_s \end{pmatrix},$$

$$\mathcal{I}_{12} = \begin{pmatrix} \mathbf{X}^\top \mathbf{T} \mathbf{c} \\ \mathbf{N}_1^\top \mathbf{T} \mathbf{c} \\ \vdots \\ \mathbf{N}_s^\top \mathbf{T} \mathbf{c} \end{pmatrix} = \mathcal{I}_{21}^\top,$$

and $\mathcal{I}_{22} = \text{trace}(\mathbf{D})$. Note that the parameters β , f_k , with $k = 1, \dots, n$, and ϕ are not orthogonal, in contrast to what is verified in the class of generalized linear regression models.

Iterative process

The solution of the estimating equation system given in (2.6) to obtain the MPL estimate of θ may be attained by iterating between a weighted back-fitting algorithm with weight matrix \mathbf{W} and a Fisher score algorithm to obtain maximum likelihood estimation of the parameter ϕ , which is equivalent to the following iterative process:

- (i) Initialize:
 - (a) By fitting a beta regression model considering $f_0^{(0)} = \beta^{(0)}$ and $\mathbf{N}_0 = \mathbf{X}$.
 - (b) By getting a starting value for ϕ by using the fitted values from (a).
 - (c) From the current value $\theta^{(0)} = (f_0^\top, f_1^{(0)\top}, \dots, f_s^{(0)\top}, \phi^{(0)})^\top$ by obtaining the weight matrix $\mathbf{W}^{(0)}$ and $\mathbf{T}^{(0)}$, with $w_i^{(0)} = w_i |_{\theta^{(0)}}$, and then by getting

$$\mathbf{z}^{(0)} = \boldsymbol{\eta}^{(0)} + \mathbf{W}^{(0)-1} \mathbf{T}^{(0)} (\mathbf{y}^* - \boldsymbol{\mu}^{*(0)}),$$

$$\mathbf{S}_0^{(0)} = (\mathbf{N}_0^\top \mathbf{W}^{(0)} \mathbf{N}_0)^{-1} \mathbf{N}_0^\top \mathbf{W}^{(0)},$$

$$\mathbf{S}_k^{(0)} = (\mathbf{N}_k^\top \mathbf{W}^{(0)} \mathbf{N}_k + \alpha_k \mathbf{K}_k)^{-1} \mathbf{N}_k^\top \mathbf{W}^{(0)}, \quad k = 1, \dots, s.$$

- (ii) Iterate repeatedly by cycling between the equations

$$\begin{aligned} f_0^{(u+1)} &= \mathcal{S}_0^{(u)} \left(\mathbf{z}^{(u)} - \sum_{l=1}^s \mathbf{N}_l f_l^{(u)} \right), \\ f_1^{(u+1)} &= \mathcal{S}_1^{(u)} \left(\mathbf{z}^{(u)} - \mathbf{N}_0 f_0^{(u+1)} - \sum_{l=2}^s \mathbf{N}_l f_l^{(u)} \right), \\ &\vdots \\ f_s^{(u+1)} &= \mathcal{S}_s^{(u)} \left(\mathbf{z}^{(u)} - \sum_{l=0}^{s-1} \mathbf{N}_l f_l^{(u+1)} \right), \end{aligned}$$

for $u = 0, 1, \dots$. Repeat step (ii) replacing $f_j^{(u)}$ by $f_j^{(u+1)}$ until convergence criterion $\Delta_u(f_j^{(u+1)}, f_j^{(u)}) = \sum_{j=0}^s \|f_j^{(u+1)} - f_j^{(u)}\| / \sum_{j=0}^s \|f_j^{(u)}\|$ is reached for a threshold value; see Hastie and Tibshirani [19] (1990).

- (iii) For current values $f_j^{(u+1)}$, with $j = 0, 1, \dots, s$, obtain $\phi^{(u+1)}$ by using

$$\phi^{(u+1)} = \phi^{(u)} - \mathbf{E} \left\{ \frac{\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \phi \partial \phi} \right\}^{-1} \frac{\partial \ell_p(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \phi} \Big|_{\boldsymbol{\theta}^{(u)}}.$$

- (iv) Iterate between steps (ii) and (iii) by replacing $f_j^{(0)}$, with $j = 0, 1, \dots, s$, and $\phi^{(0)}$ by $f_j^{(u+1)}$ and $\phi^{(u+1)}$, respectively, until reaching convergence.

ACKNOWLEDGMENTS

The authors would like to thank the editors and four reviewers very much for their constructive comments on an earlier version of this manuscript which resulted in this improved version. This work was supported by project grants “Fondecyt 11130704” (G. Ibacache-Pulgar) and “Fondecyt 11190636” (C. Marchant) from the National Agency for Research and Development (ANID) of the Chilean government and by ANID – Millennium Science Initiative Program – NCN17_059 (C. Marchant). J.I. Figueroa-Zuñiga acknowledges funding support by grant VRID Enlace No. 217.014.027-1.0, from the Universidad de Concepción, Chile. G. Ibacache-Pulgar acknowledges the support of the Centro Interdisciplinario de Estudios Atmosféricos y Astroestadística from the Universidad de Valparaíso, Chile.

REFERENCES

- [1] ADAMS, R.A. and FOURNIER, J. (2003). *Sobolev Spaces. Pure and Applied Mathematics*, Academic Press, Boston.
- [2] BERHANE, K. and TIBSHIRANI, J. (1998). Generalized additive models for longitudinal data, *The Canadian Journal of Statistics*, **26**, 517–535.
- [3] CAO, C.Z. and LIN, J.G. (2011). Diagnostics for elliptical linear mixed models with first-order autoregressive errors, *Journal of Statistical Computation and Simulation*, **81**, 1281–1296.
- [4] COOK, R.D. (1986). Assessment of local influence (with discussion), *Journal of the Royal Statistical Society B*, **48**, 133–169.
- [5] CRIBARI-NETO, F. and ZEILEIS, A. (2010). Beta regression in R, *Journal of Statistical Software*, **34**, Issue 2.
- [6] CYSNEIROS, F.J.A.; LEIVA, V.; LIU, S.; MARCHANT, C. and SCALCO, P. (2019). A Cobb–Douglas type model with stochastic restrictions: Formulation, local influence diagnostics and data analytics in economics, *Quality and Quantity*, **53**, 1693–1719.
- [7] EILERS, P.H.C. and MARX, B.D. (1996). Flexible smoothing with *B*-splines and penalties, *Statistical Science*, **11**, 89–121.
- [8] EMMAMI, H. (2017). Local influence for Liu estimators in semiparametric linear models, *Statistical Papers*, **19**, 529–544.
- [9] ESPINHERIA, P.; FERRARI, S. and CRIBARI-NETO, F. (2008a). On beta regression residuals, *Journal of Applied Statistics*, **35**, 407–419.
- [10] ESPINHERIA, P.; FERRARI, S. and CRIBARI-NETO, F. (2008b). Influence diagnostics in beta regression, *Computational Statistics and Data Analysis*, **52**, 4417–4431.
- [11] FANG, K. and YAO, Z. (2017). Semi-parametric additive beta regression model with its application, *CMS*, **25**, 116–124.
<http://www.zgglkx.com/CN/10.16381/j.cnki.issn1003-207x.2017.09.013>
- [12] FERRARI, S. (2011). Diagnostics tools in beta regression with varying dispersion, *Statistica Neerlandica*, **65**, 337–351.
- [13] FERRARI, S. and CRIBARI-NETO, F. (2004). Beta regression for modeling rates and proportions, *Journal of Applied Statistics*, **31**, 799–815.
- [14] FERREIRA, M.; GOMES, M.I. and LEIVA, V. (2012). On an extreme value version of the Birnbaum–Saunders distribution, *REVSTAT*, **10**, 181–210.
- [15] FERREIRA, C.S. and PAULA, G.A. (2017). Estimation and diagnostic for skew-normal partially linear models, *Journal of Applied Statistics*, **44**, 3033–3053.
- [16] FIGUEROA-ZUÑIGA, J.; ARELLANO-VALLE, R. and FERRARI, S. (2013). Mixed beta regression: A Bayesian perspective, *Computational Statistics and Data Analysis*, **61**, 137–147.
- [17] GARCIA-PAPANI, F.; LEIVA, V.; URIBE-OPAZO, M.A. and AYKROYD, R.G. (2018). Birnbaum–Saunders spatial regression models: Diagnostics and application to chemical data, *Chemometrics and Intelligent Laboratory Systems*, **177**, 114–128.
- [18] GREEN, P.J. and SILVERMAN, B.W. (1994). *Nonparametric Regression and Generalized Linear Models*, Chapman and Hall, Boca Raton.
- [19] HASTIE, T. and TIBSHIRANI, R. (1990). *Generalized Additive Models*, Chapman and Hall, London.
- [20] HASTIE, T.; TIBSHIRANI, R. and FRIEDMAN, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer.

- [21] HUERTA, M.; LEIVA, V.; LILLO, C. and RODRIGUEZ, M. (2018). A beta partial least squares regression model: Diagnostics and application to mining industry data, *Applied Stochastic Models in Business and Industry*, **34**, 305–321.
- [22] IBACACHE-PULGAR, G. and PAULA, G.A. (2011). Local influence for Student- t partially linear models, *Computational Statistics and Data Analysis*, **55**, 1462–1478.
- [23] IBACACHE-PULGAR, G.; PAULA, G.A. and CYSNEIROS, F.J.A. (2013). Semiparametric additive models under symmetric distributions, *Test*, **22**, 103–121.
- [24] IBACACHE-PULGAR, G.; PAULA, G.A. and GALEA, M. (2012). Influence diagnostics for elliptical semiparametric mixed models, *Statistical Modelling*, **12**, 165–193.
- [25] IBACACHE-PULGAR, G. and REYES, S. (2018). Local influence for elliptical partially varying-coefficient model, *Statistical Modelling*, **18**, 149–174.
- [26] LEAO, J.; LEIVA, V.; SAULO, H. and TOMAZELLA, V. (2018). Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data, *Statistics in Medicine*, **37**, 4421–4440.
- [27] MARCHANT, C.; LEIVA, V.; CYSNEIROS, F.J.A. and VIVANCO, J.F. (2016). Diagnostics in multivariate Birnbaum–Saunders regression models, *Journal of Applied Statistics*, **43**, 2829–2849.
- [28] OSORIO, F.; PAULA, G.A. and GALEA, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure, *Computational Statistics and Data Analysis*, **51**, 4354–4368.
- [29] POON, W. and POON, Y.S. (1999). Conformal normal curvature and assessment of local influence, *Journal of the Royal Statistical Society B*, **61**, 51–61.
- [30] QUEIROZ DA-SILVA, C. and MIGON, H. (2016). Hierarchical Beta model, *REVSTAT*, **14**, 49–73.
- [31] ROCHA, A. and SIMAS, A. (2011). Influence diagnostics in a general class of beta regression models, *TEST*, **20**, 95–119.
- [32] STASINOPOULOS, D.M. and RIGBY, R.A. (2007). Generalized additive models for location scale and shape (GAMLSS) in R, *Journal of Statistical Software*, **23**, 1–46.
- [33] TAPIA, H.; GIAMPAOLI, V.; DIAZ, M.P. and LEIVA, V. (2019a). Sensitivity analysis of longitudinal count responses: A local influence approach and application to medical data, *Journal of Applied Statistics*, **46**, 1021–1042.
- [34] TAPIA, H.; LEIVA, V.; DIAZ, M.P. and GIAMPAOLI, V. (2019b). Influence diagnostics in mixed effects logistic regression models, *TEST*, **28**, 920–942.
- [35] TELFORD, R.D. and CUNNINGHAM, R.B. (1991). Sex, sport and body-size dependency of hematology in highly trained athletes, *Medicine and Science in Sports and Exercise*, **23**, 788–794.
- [36] URIBE-OPAZO, M.A.; BORSSOI, J.A. and GALEA, M. (2012). Influence diagnostics in Gaussian spatial linear models, *Journal of Applied Statistics*, **3**, 615–630.
- [37] VENTURA, M.; SAULO, H.; LEIVA, V. and MONSUETO, S. (2019). Log-symmetric regression models: Information criteria, application to movie business and industry data with economic implications, *Applied Stochastic Models in Business and Industry*, **35**(4), 963–977.
- [38] ZHANG, J.; ZHANG, X.; MA, H. and ZHIYA, C. (2015). Local influence analysis of varying coefficient linear model, *Journal of Interdisciplinary Mathematics*, **3**, 293–306.
- [39] ZHAO, W.; ZHANG, R.; LV, Y. and LIU, J. (2014). Variable selection for varying dispersion beta regression model, *Journal of Applied Statistics*, **41**, 95–108.
- [40] ZHU, H.T. and LEE, S.Y. (2003). Local influence for generalized linear mixed models, *The Canadian Journal of Statistics*, **31**, 293–309.
- [41] ZHU, Z.Y.; HE, X. and FUNG, W.K. (2003). Local influence analysis for penalized Gaussian likelihood estimators in partially linear models, *Scandinavian Journal of Statistics*, **30**, 767–780.

VARIANCE ESTIMATION IN THE PRESENCE OF MEASUREMENT ERRORS UNDER STRATIFIED RANDOM SAMPLING

Authors: NEHA SINGH

– Department of Mathematics & Computing, Indian Institute of Technology (ISM),
Dhanbad-826004, Jharkhand, India
nehaismstats@gmail.com

GAJENDRA K. VISHWAKARMA

– Department of Mathematics & Computing, Indian Institute of Technology (ISM),
Dhanbad-826004, Jharkhand, India
vishwagk@rediffmail.com

RAJ K. GANGELE

– Department of Mathematics and Statistics, Dr. Harisingh Gour University
(Central University),
Sagar-470003, M.P., India
rkgangele23@gmail.com

Received: August 2018

Revised: April 2019

Accepted: April 2019

Abstract:

- This study focuses on the estimation of population variance of study variable in stratified random sampling using auxiliary information when the observations are contaminated by measurement errors. Three classes of estimators of variance under measurement error are proposed by using the approach of Srivastava and Jhajj [18] for the study variable. The properties of the estimator viz. bias and mean square error of the proposed classes of estimators are provided. The conditions for which proposed estimators are more efficient compared to usual estimators are discussed. It is shown that the proposed classes of estimators include a large number of estimators of the population variance of stratified random sampling and their bias and mean square error can be easily derived.

Keywords:

- *variance; auxiliary variable; mean square error; measurement error; stratified random sampling.*

AMS Subject Classification:

- 49A05, 78B26.

1. INTRODUCTION

In survey sampling, the auxiliary information is mainly used in order to gain efficiency for the estimation. The literature on estimating the population variance by using auxiliary variable is substantial and widely discussed. Some authors including, Das and Tripathi [4], Srivastava and Jhajj [18], Isaki [5], Upadhyay and Singh [19, 20], Singh *et al.* [13], Prasad and Singh [8], Biradar and Singh [2], Singh and Biradar [11] have paid their attention towards the estimation of population variance of study variable using auxiliary information in simple random sampling. While dealing with planning surveys, in case of heterogeneous population, stratified random sampling has more importance in precise estimates over the simple random sampling. Singh and Vishwakarma [14] discussed a general method for the estimation of the variance of the stratified random sample mean by using auxiliary information.

The theories of survey sampling assume that the observations recorded during data collection are always free from measurement error. However, this assumption does not meet in many real-life situations and the data is contaminated with errors. The mean square error and other properties of the estimator obtained with significant measurement error may lead to serious fallacious results. Cochran [3], has discussed the source of measurement errors in survey data. Many authors such as Shalabh [9], Srivastava and Shalabh [16], Maneesha and Singh [7], Allen *et al.* [1], Shalabh and Tsai [10], Singh and Vishwakarma [15] have studied the impacts of measurement errors in the ratio, product and regression methods of estimation under simple random sampling.

Let us consider a finite heterogeneous population of size N , partitioned into L non-overlapping strata of sizes N_h , $h = 1, 2, \dots, L$, where $\sum_{h=1}^L N_h = N$. Let (y_{hj}, x_{hj}) be the pair of observed values instead of true pair values (Y_{hj}, X_{hj}) of the study character y and the auxiliary character x respectively of the j -th unit ($j = 1, 2, \dots, N_h$) in the h -th stratum. Also, let (y_{hj}, x_{hj}) be the pair of values on (y, x) drawn from the h -th stratum ($j = 1, 2, \dots, n_h$; $h = 1, 2, \dots, L$). It is familiar that in stratified random sampling an unbiased estimator of the population mean ($\mu_Y = \sum_{h=1}^L W_h \mu_{Yh}$; $W_h = \frac{N_h}{N}$) of the variable y is given by

$$(1.1) \quad \bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h,$$

where $\bar{y}_h = \frac{1}{n_h} \sum_{j=1}^{n_h} y_{hj}$ is the sample mean of h -th stratum and $\mu_{Yh} = \frac{1}{N_h} \sum_{j=1}^{N_h} y_{hj}$ is the population mean of h -th stratum. Let the observational errors be

$$(1.2) \quad u_{hj} = y_{hj} - Y_{hj}, \quad v_{hj} = x_{hj} - X_{hj},$$

which are normally distributed with mean zero and variances σ_{uh}^2 and σ_{vh}^2 respectively. Also let ρ_h be the population correlation coefficient between Y and X in h -th stratum. For simplicity in exposition, it is assumed that the variables u_{hj} and v_{hj} are uncorrelated although (Y_{hj}, X_{hj}) are correlated.

To obtain the bias and mean square error we define

$$\hat{\sigma}_{Yh}^2 = \sigma_{Yh}^2(1 + \varepsilon_{0h}), \quad \hat{\sigma}_{Xh}^2 = \sigma_{Xh}^2(1 + \varepsilon_{1h}), \quad \bar{x}_h = \mu_{Xh}(1 + \varepsilon_{2h}),$$

such that $E(\varepsilon_{ih}) = 0$, $\forall i = 0, 1, 2$;

$$E(\varepsilon_{0h}^2) = \frac{A_{Yh}}{n_h}, \quad E(\varepsilon_{1h}^2) = \frac{A_{Xh}}{n_h}, \quad E(\varepsilon_{2h}^2) = \frac{C_{Xh}^2}{n_h \theta_{Xh}}, \quad E(\varepsilon_{0h} \varepsilon_{1h}) = \frac{1}{n_h} (\delta_{22h} - 1),$$

$$E(\varepsilon_{1h} \varepsilon_{2h}) = \frac{1}{n_h} (\delta_{03h} C_{Xh}), \quad E(\varepsilon_{0h} \varepsilon_{2h}) = \frac{1}{n_h} (\delta_{21h} C_{Xh}),$$

where

$$A_{Yh} = \gamma_{2Yh} + \gamma_{2Uh} \frac{\sigma_{Uh}^4}{\sigma_{Yh}^4} + \frac{2}{\theta_{Yh}^2}, \quad \beta_{2h}(Y) = \delta_{40h} = \frac{\mu_{40h}}{\mu_{20h}^2}, \quad C_{Xh} = \frac{\sigma_{Xh}}{\mu_{Xh}},$$

$$A_{Xh} = \gamma_{2Xh} + \gamma_{2Vh} \frac{\sigma_{Vh}^4}{\sigma_{Xh}^4} + \frac{2}{\theta_{Xh}^2}, \quad \beta_{2h}(X) = \delta_{04h} = \frac{\mu_{04h}}{\mu_{02h}^2},$$

$$\gamma_{2Yh} = \beta_{2h}(Y) - 3, \quad \gamma_{2Xh} = \beta_{2h}(X) - 3, \quad \gamma_{2Uh} = \beta_{2h}(U) - 3,$$

$$\gamma_{2Vh} = \beta_{2h}(V) - 3, \quad \theta_{Yh} = \frac{\sigma_{Yh}^2}{\sigma_{Yh}^2 + \sigma_{Uh}^2}, \quad \theta_{Xh} = \frac{\sigma_{Xh}^2}{\sigma_{Xh}^2 + \sigma_{Vh}^2},$$

$$\delta_{rsh} = \frac{\mu_{rsh}}{(\mu_{20h}^r \mu_{02h}^s)^{\frac{1}{2}}}, \quad \mu_{rsh} = \frac{1}{N_h} \sum_{j=1}^{N_h} (y_{hj} - \mu_{Yh})^r (x_{hj} - \mu_{Xh})^s.$$

(r, s) are positive integers, μ_{Yh} and μ_{Xh} are the h -th stratum population mean of study and auxiliary variable respectively. C_{Xh} is the coefficient of variation of h -th stratum, θ_{Yh} and θ_{Xh} are the reliability ratio of h -th stratum of study and auxiliary variable respectively and lying between zero and one.

The variance of the stratified random sample mean is given by

$$(1.3) \quad V(\bar{y}_{st}) = \sum_{h=1}^L W_h^2 \frac{\sigma_{Yh}^2}{n_h} = \sigma_{st}^2,$$

where $\sigma_{Yh}^2 = \frac{1}{N_h} \sum_{j=1}^{N_h} (y_{ij} - \bar{\mu}_{Yh})^2$ is the population variance of h -th stratum.

The unbiased estimator of σ_{st}^2 , i.e. $V(\bar{y}_{st})$, is given by

$$(1.4) \quad \hat{\sigma}_{st}^2 = \sum_{h=1}^L W_h^2 \frac{s_{yh}^2}{n_h},$$

where $s_{yh}^2 = \frac{1}{(n_h-1)} \sum_{j=1}^{n_h} (y_{hj} - \bar{y}_h)^2$ is an unbiased estimator of σ_{st}^2 . But in the presence of measurement error s_{yh}^2 is not an unbiased estimator for σ_{st}^2 . In the measurement error case the unbiased estimator of σ_{st}^2 is given by $\hat{\sigma}_{st}^2 = \sum_{h=1}^L W_h^2 \frac{\hat{\sigma}_{Yh}^2}{n_h}$, where $\hat{\sigma}_{Yh}^2 = (s_{yh}^2 - \sigma_{uh}^2)$.

The variance of $\hat{\sigma}_{st}^2$ in the presence of measurement error is given by

$$(1.5) \quad V(\hat{\sigma}_{st}^2) = \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} [A_{Yh}] = \text{MSE}(\hat{\sigma}_{st}^2).$$

Singh and Karpe [12] have studied the impact of measurement error on separate ratio and product also combined ratio as well as product estimators for the population mean under stratified random sampling. We have considered the problem of estimating population variance using information on the auxiliary variable by adopting Srivastava and Jhajj [18]

method in stratified random sampling in the presence of measurement error. Three classes of estimators for the estimation of population variance are proposed under stratified random sampling when both the study and auxiliary variables are commingled with measurement errors as:

- i) Estimator of variance σ_{st}^2 when the mean μ_{Xh} of the auxiliary variable x in the h -th stratum of the population is known.
- ii) Estimation of variance σ_{st}^2 when the variance σ_{Xh}^2 of the auxiliary variable x in the h -th stratum of the population is known.
- iii) Estimation of variance σ_{st}^2 when the mean μ_{Xh} and the variance σ_{Xh}^2 of the auxiliary variable x in the h -th stratum of the population are known.

The crux of this study is to exhibit the effect of measurement errors on the estimates of the variance of the stratified random sample mean while using auxiliary information.

2. THE PROPOSED CLASSES OF ESTIMATORS

2.1. Estimation of population variance σ_{st}^2 of the stratified simple random sample mean when mean μ_{Xh} of h -th stratum of the auxiliary variable x in the population is known

By using information on population mean μ_{Xh} of the h -th stratum of auxiliary variable, a class of estimators of population variance σ_{st}^2 for the study variable is proposed as

$$(2.1) \quad \hat{\sigma}_a^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \hat{\sigma}_{Yh}^2 a_h(l_h),$$

where $l_h = \bar{x}_h / \mu_{Xh}$ and $a_h(\cdot)$ is a function of l_h such that $a_h = 1$. It satisfies conditions given by Srivastava [17] viz. function are continuous and bounded also the first as well as second order partial derivatives of the function exist. Expanding the function about the point 'unity' in a second order Taylor's series, we have

$$(2.2) \quad a_h(l_h) = a_h(1) + (l_h - 1)a_{1h}(1) + \frac{1}{2}(l_h - 1)^2 a_{2h}(1),$$

where a_{1h} , a_{2h} are first order and second order derivative with respect to l_h about point unity.

$$(2.3) \quad \hat{\sigma}_a^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \sigma_{Yh}^2 (1 + \varepsilon_{0h}) \left[1 + (l_h - 1)a_{1h}(1) + \frac{1}{2}(l_h - 1)^2 a_{2h}(1) \right],$$

$$\hat{\sigma}_a^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \sigma_{Yh}^2 (1 + \varepsilon_{0h}) \left[1 + \varepsilon_{2h} a_{1h}(1) + \frac{1}{2} \varepsilon_{2h}^2 a_{2h}(1) \right],$$

$$(2.4) \quad (\hat{\sigma}_a^2 - \sigma_{st}^2) = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) (\sigma_{Yh}^2) \left[\varepsilon_{0h} + \varepsilon_{2h} a_{1h}(1) + \varepsilon_{0h} \varepsilon_{2h} a_{1h}(1) + \frac{1}{2} \varepsilon_{2h}^2 a_{2h}(1) + \frac{1}{2} \varepsilon_{0h} \varepsilon_{2h}^2 a_{2h}(1) \right].$$

Taking expectation on both sides of (2.4) we get

$$(2.5) \quad \text{Bias}(\hat{\sigma}_a^2) = \sum_{h=1}^L \left(\frac{W_h^2}{n_h^2} \right) \sigma_{Yh}^2 \left[\delta_{21h} C_{Xh} a_{1h}(1) + \frac{1}{2} \frac{C_{Xh}^2}{\theta_{Xh}} a_{2h}(1) \right].$$

For the mean square error we have

$$(2.6) \quad (\hat{\sigma}_a^2 - \sigma_{st}^2)^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \sigma_{Yh}^4 \left\{ \varepsilon_{0h} + \varepsilon_{2h} a_{1h}(1) \right\}^2,$$

$$(2.7) \quad (\hat{\sigma}_a^2 - \sigma_{st}^2)^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right)^2 \sigma_{Yh}^4 \left\{ \varepsilon_{0h}^2 + \varepsilon_{2h}^2 a_{1h}^2(1) + 2\varepsilon_{0h} \varepsilon_{2h} a_{1h}(1) \right\}.$$

Taking expectation up to terms of order n_h^{-3} , we get the mean square error of $\hat{\sigma}_a^2$ as

$$(2.8) \quad \text{MSE}(\hat{\sigma}_a^2) = \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[A_{Yh} + \frac{C_{Xh}^2}{\theta_{Xh}} a_{1h}^2(1) + 2\delta_{21h} C_{Xh} a_{1h}(1) \right].$$

The MSE in (2.8) is minimized for

$$(2.9) \quad a_{1h}(1) = - \left(\frac{\delta_{21h} \theta_{Xh}}{C_{Xh}} \right).$$

Thus, the resultant minimum MSE of $\hat{\sigma}_a^2$ is given by

$$(2.10) \quad \min.\text{MSE}(\hat{\sigma}_a^2) = \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[A_{Yh} - \delta_{21h}^2 \theta_{Xh} \right].$$

Hence, a theorem can be established as follows.

Theorem 2.1. *Up to terms of the order n_h^{-3} ,*

$$\min.\text{MSE}(\hat{\sigma}_a^2) \geq \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[A_{Yh} - \delta_{21h}^2 \theta_{Xh} \right],$$

with equality holding if $a_{1h}(1) = - \left(\frac{\delta_{21h} \theta_{Xh}}{C_{Xh}} \right)$.

The following estimators

$$\begin{aligned} \hat{\sigma}_{a1}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 l_h^{\alpha_{1h}}, & \hat{\sigma}_{a2}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 [2 - l_h^{\alpha_{1h}}], \\ \hat{\sigma}_{a3}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 \left[\frac{\alpha_{1h} + l_h}{1 + \alpha_{1h} l_h}\right], & \hat{\sigma}_{a4}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 [\alpha_{1h} + (1 - \alpha_{1h}) l_h], \\ \hat{\sigma}_{a5}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 [\alpha_{1h} + (1 - \alpha_{1h}) l_h^{-1}], \\ \hat{\sigma}_{a6}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 [\alpha_{1h} + (1 - \alpha_{1h}) l_h^{-\alpha_{2h}}], \\ \hat{\sigma}_{a7}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 [\alpha_{1h} + (1 - \alpha_{1h}) l_h]^{-1}, \end{aligned}$$

are some of the members of the proposed class of estimators $\hat{\sigma}_a^2$. The optimum values of the scalars α_{1h} and α_{2h} can be derived from the right-hand side (2.9) of and the minimum mean square error of the listed estimators can be derived from (2.8). The lower bound of the MSE of estimators $\hat{\sigma}_{ai}^2, (i = 1 \text{ to } 7)$ is the same as given by (2.10).

Following by [17] and Srivastava and Jhajj [18] we have proposed a wider class of estimators of σ_{st}^2 as

$$(2.11) \quad \hat{\sigma}_D^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) D_h(\hat{\sigma}_{Yh}^2, l_h),$$

where function $D_h(\cdot, \cdot)$ satisfies

$$D_h(\sigma_{Yh}^2, 1) = \sigma_{Yh}^2 \Rightarrow D_{1h}(\sigma_{Yh}^2, 1) = \frac{\partial D_h(\cdot)}{\partial \hat{\sigma}_{Yh}^2} \Big|_{(\sigma_{Yh}^2, 1)} = 1.$$

It can be shown that the minimum MSE of $\hat{\sigma}_D^2$ and the minimum MSE of $\hat{\sigma}_a^2$ are equal. We can state that the difference type estimator

$$(2.12) \quad \hat{\sigma}_{std_1}^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \left\{ \hat{\sigma}_{Yh}^2 + d_{1h}(l_h - 1) \right\},$$

is a member of class $\hat{\sigma}_D^2$ where d_{1h} is a suitably chosen constant.

2.2. Estimation of population variance σ_{st}^2 of the stratified simple random sample mean when variance σ_{Xh}^2 of h -th stratum of the auxiliary variable x in the population is known

A class of estimators of the variance σ_{st}^2 of the stratified simple random sample mean when the variance σ_{Xh}^2 of the auxiliary variable x of the h -th stratum in the population is known, is defined as

$$(2.13) \quad \hat{\sigma}_b^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 b_h(m_h),$$

where $m_h = \frac{\hat{\sigma}_{Yh}^2}{\sigma_{Xh}^2}$, and $b_h(m_h)$ is a function of m_h such that $b_h(1) = 1$. The function is continuous and bounded in \mathbb{R} and its first as well as the second order partial derivatives exist. Now expanding the function at point ‘unity’ in a second order Taylor’s series, we can write

$$(2.14) \quad b_h(m_h) = b_h(1) + (m_h - 1)b_{1h}(1) + \frac{1}{2}(m_h - 1)^2b_{2h}(1),$$

where $b_{1h}(1)$ and $b_{2h}(1)$ are the first order and second order derivative with respect to m_h of the function $b_h(m_h)$ about the point ‘unity’.

$$(2.15) \quad \hat{\sigma}_b^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \hat{\sigma}_{Yh}^2 \left[b_h(1) + (m_h - 1)b_{1h}(1) + \frac{1}{2}(m_h - 1)^2b_{2h}(1) \right],$$

$$(2.16) \quad \hat{\sigma}_b^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \sigma_{Yh}^2 (1 + \varepsilon_{0h}) \left[1 + \varepsilon_{1h}b_{1h}(1) + \frac{1}{2}\varepsilon_{1h}^2b_{2h}(1) \right].$$

To calculate the bias and the MSE of the estimator we can write

$$(2.17) \quad \begin{aligned} (\hat{\sigma}_b^2 - \sigma_{st}^2) &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \sigma_{Yh}^2 \left[\varepsilon_{0h} + \varepsilon_{1h}b_{1h}(1) + \varepsilon_{0h}\varepsilon_{1h}b_{1h}(1) \right. \\ &\quad \left. + \frac{1}{2}\varepsilon_{1h}^2b_{2h}(1) + \frac{1}{2}\varepsilon_{0h}\varepsilon_{1h}^2b_{2h}(1) \right]. \end{aligned}$$

Taking expectation on both sides of (2.17) we get the bias of $\hat{\sigma}_b^2$ as

$$(2.18) \quad \text{Bias}(\hat{\sigma}_b^2) = \sum_{h=1}^L \left(\frac{W_h^2}{n_h^2} \right) \sigma_{Yh}^2 \left[(\delta_{22h} - 1)b_{1h}(1) + \frac{1}{2}A_{Xh}b_{2h}(1) \right].$$

For the mean square error we have

$$(2.19) \quad (\hat{\sigma}_b^2 - \sigma_{st}^2)^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right)^2 \sigma_{Yh}^4 \left[\varepsilon_{0h} + \varepsilon_{1h}b_{1h}(1) \right]^2,$$

$$(2.20) \quad (\hat{\sigma}_b^2 - \sigma_{st}^2)^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right)^2 \sigma_{Yh}^4 \left[\varepsilon_{0h}^2 + \varepsilon_{1h}^2b_{1h}^2(1) + 2\varepsilon_{0h}\varepsilon_{1h}b_{1h}(1) \right].$$

Taking expectation up to terms of order n_h^{-3} , we get the mean square error of $\hat{\sigma}_b^2$ as

$$(2.21) \quad \text{MSE}(\hat{\sigma}_b^2) = \sum_{h=1}^L \frac{(W_h\sigma_{Yh})^4}{n_h^3} \left[A_{Yh} + A_{Xh}b_{1h}^2(1) + 2(\delta_{22h} - 1)b_{1h}(1) \right],$$

which is minimized for

$$(2.22) \quad b_{1h}(1) = - \left(\frac{\delta_{22h} - 1}{A_{Xh}} \right).$$

Thus, the resultant minimum MSE of $\hat{\sigma}_b^2$ can be written as:

$$(2.23) \quad \min.\text{MSE}(\hat{\sigma}_b^2) = \sum_{h=1}^L \frac{(W_h\sigma_{Yh})^4}{n_h^3} \left[A_{Yh} - \frac{(\delta_{22h} - 1)^2}{A_{Xh}} \right].$$

Hence, a theorem can be established as follows.

Theorem 2.2. Up to terms of order n_h^{-3} ,

$$\min.\text{MSE}(\hat{\sigma}_b^2) \geq \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[A_{Yh} - \frac{(\delta_{22h} - 1)^2}{A_{Xh}} \right],$$

with equality holding if $b_{1h}(1) = -\left(\frac{\delta_{22h}-1}{A_{Xh}}\right)$.

The listed estimators

$$\begin{aligned} \hat{\sigma}_{b1}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 m_h^{\eta_{1h}}, & \hat{\sigma}_{b2}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 [2 - m_h^{\eta_{1h}}], \\ \hat{\sigma}_{b3}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 \left[\frac{\eta_{1h} + m_h}{1 + \eta_{1h} m_h}\right], & \hat{\sigma}_{b4}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 [\eta_{1h} + (1 - \eta_{1h}) m_h], \\ \hat{\sigma}_{b5}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 [\eta_{1h} + (1 - \eta_{1h}) m_h^{-1}], \\ \hat{\sigma}_{b6}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 [\eta_{1h} + (1 - \eta_{1h}) m_h^{\eta_{2h}}], \\ \hat{\sigma}_{b7}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \hat{\sigma}_{Yh}^2 [\eta_{1h} + (1 - \eta_{1h}) m_h]^{-1}, \end{aligned}$$

are some of the members of the proposed class of estimators $\hat{\sigma}_b^2$. The optimum values of the scalars η_{1h} and η_{2h} from $\hat{\sigma}_{b1}^2$ to $\hat{\sigma}_{b7}^2$ can be derived from (2.22) and the minimum mean square errors of each of the listed estimators can be derived from (2.21). The lower bound of the MSE of the estimators $\hat{\sigma}_{bi}^2$ ($i = 1$ to 7) is given by (2.23).

A wider class of estimators of σ_{st}^2 than $\hat{\sigma}_b^2$ is

$$(2.24) \quad \hat{\sigma}_e^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) e_h(\hat{\sigma}_{Yh}^2, m_h),$$

where $e_h(\hat{\sigma}_{Yh}^2, m_h)$ is a function of $(\hat{\sigma}_{Yh}^2, m_h)$ and

$$e_h(\sigma_{Yh}^2, 1) = \sigma_{Yh}^2 \Rightarrow e_{1h}(\sigma_{Yh}^2) = 1 \quad \text{with} \quad e_{1h}(\sigma_{Yh}^2, 1) = \frac{\partial e_h(\cdot)}{\partial \hat{\sigma}_{Yh}^2} \Big|_{(\sigma_{Yh}^2, 1)}.$$

It can be exhibited that up-to third order, the optimum mean square error of $\hat{\sigma}_e^2$ and $\hat{\sigma}_b^2$ is the same as given by (2.23). It can also be shown that the difference-type estimator

$$(2.25) \quad \hat{\sigma}_{std_2}^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h}\right) \left\{ \hat{\sigma}_{Yh}^2 + d_{2h}(m_h - 1) \right\}$$

is a specific member of the class of estimator $\hat{\sigma}_e^2$ but not of the $\hat{\sigma}_b^2$ class, where d_{2h} is an appropriately chosen constant.

2.3. Estimation of population variance σ_{st}^2 of the stratified simple random sample mean when the population mean μ_{Xh} and the variance σ_{Xh}^2 of h -th stratum of the auxiliary variable x in the population are known

We define a class of estimators $\hat{\sigma}_c^2$, for the estimation of variance of the stratified simple random sample mean when the mean μ_{Xh} and the variance σ_{Xh}^2 of the auxiliary variable x of the h -th stratum in the population are known, as

$$(2.26) \quad \hat{\sigma}_c^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \hat{\sigma}_{Yh}^2 c_h(l_h, m_h),$$

where $c_h(l_h, m_h)$ is a function of $l_h = \bar{x}_h/\mu_{Xh}$ and $m_h = \hat{\sigma}_{Xh}^2/\sigma_{Xh}^2$, such that $c_h(1, 1) = 1$ and it also satisfies similar conditions as mentioned in [18]:

$$(2.27) \quad c_h(l_h, m_h) = \left[c_h(1, 1) + (l_h - 1)c_{1h}(1, 1) + (m_h - 1)c_{2h}(1, 1) \right. \\ \left. + \frac{1}{2} \left\{ (l_h - 1)^2 c_{11h}(1, 1) + 2(l_h - 1)(m_h - 1)c_{12h}(1, 1) + (m_h - 1)^2 c_{22h}(1, 1) \right\} \right. \\ \left. + \frac{1}{6} \left\{ (l_h - 1)^3 c_{111h}(l_h^*, m_h^*) + 3(l_h - 1)^2(m_h - 1)c_{112h}(l_h^*, m_h^*) \right. \right. \\ \left. \left. + 3(l_h - 1)(m_h - 1)^2 c_{122h}(l_h^*, m_h^*) + (m_h - 1)^3 c_{222h}(l_h^*, m_h^*) \right\} \right],$$

where

$$l_h^* = 1 + \phi(l_h - 1), \quad m_h^* = 1 + \phi(m_h - 1), \quad 0 < \phi < 1; \\ \left\{ c_{1h}(1, 1), c_{2h}(1, 1) \right\}, \left\{ c_{11h}(1, 1), c_{12h}(1, 1), c_{22h}(1, 1) \right\}, \\ \left\{ c_{111h}(l_h^*, m_h^*), c_{112h}(l_h^*, m_h^*), c_{122h}(l_h^*, m_h^*), c_{222h}(l_h^*, m_h^*) \right\},$$

respectively denote the first, second and third order partial derivatives of the function $c_h(l_h, m_h)$. Expressing (2.27) in terms of ε_{0h} , ε_{1h} and ε_{2h} using $c_h(1, 1) = 1$ we have

$$(2.28) \quad \hat{\sigma}_c^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \sigma_{Yh}^2 (1 + \varepsilon_{0h}) \left[\left\{ 1 + \varepsilon_{2h}c_{1h}(1, 1) + \varepsilon_{1h}c_{2h}(1, 1) \right\} \right. \\ \left. + \frac{1}{2} \left\{ \varepsilon_{2h}^2 c_{11h}(1, 1) + 2\varepsilon_{1h}\varepsilon_{2h}c_{12h}(1, 1) + \varepsilon_{1h}^2 c_{22h}(1, 1) \right\} \right. \\ \left. + \frac{1}{6} \left\{ \varepsilon_{2h}^3 c_{111h}(l_h^* m_h^*) + 3\varepsilon_{1h}\varepsilon_{2h}^2 c_{112h}(l_h^* m_h^*) \right. \right. \\ \left. \left. + 3\varepsilon_{1h}^2 \varepsilon_{2h} c_{122h}(l_h^* m_h^*) + \varepsilon_{1h}^3 c_{222h}(l_h^* m_h^*) \right\} \right].$$

To calculate the bias and the MSE of the estimator we can write (2.28) as

$$(2.29) \quad \hat{\sigma}_c^2 - \sigma_{st}^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \sigma_{Yh}^2 \left[\left\{ \varepsilon_{0h} + \varepsilon_{2h}c_{1h}(1, 1) + \varepsilon_{1h}c_{2h}(1, 1) \right. \right. \\ \left. \left. + \varepsilon_{0h}\varepsilon_{2h}c_{1h}(1, 1) + \varepsilon_{0h}\varepsilon_{1h}c_{2h}(1, 1) \right\} \right. \\ \left. + \frac{1}{2} \left\{ \varepsilon_{2h}^2 c_{11h}(1, 1) + 2\varepsilon_{1h}\varepsilon_{2h}c_{12h}(1, 1) + \varepsilon_{1h}^2 c_{22h}(1, 1) \right\} \right].$$

Taking expectation of both sides of (2.29) we get the bias of the estimator $\hat{\sigma}_c^2$:

$$\begin{aligned}
 \text{Bias}(\hat{\sigma}_c^2) &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h^2} \right) \sigma_{Yh}^2 \left[\delta_{21h} C_{Xh} c_{1h}(1, 1) + (\delta_{22h} - 1) c_{2h}(1, 1) \right. \\
 (2.30) \quad &\quad \left. + \frac{1}{2} \left(\frac{C_{Xh}^2}{\theta_{Xh}} c_{11h}(1, 1) + 2\delta_{03h} C_{Xh} c_{12h}(1, 1) + A_{Xh} c_{22h}(1, 1) \right) \right].
 \end{aligned}$$

For the mean square error we have

$$(2.31) \quad (\hat{\sigma}_c^2 - \sigma_{st}^2)^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right)^2 \sigma_{Yh}^4 \left[\varepsilon_{0h} + \varepsilon_{2h} c_{1h}(1, 1) + \varepsilon_{1h} c_{2h}(1, 1) \right]^2,$$

$$\begin{aligned}
 (2.32) \quad (\hat{\sigma}_c^2 - \sigma_{st}^2)^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right)^2 \sigma_{Yh}^4 \left[\varepsilon_{0h}^2 + \varepsilon_{2h}^2 c_{1h}^2(1, 1) + \varepsilon_{1h}^2 c_{2h}^2(1, 1) \right. \\
 &\quad \left. + 2\varepsilon_{0h} \varepsilon_{2h} c_{1h}(1, 1) + 2\varepsilon_{1h} \varepsilon_{2h} c_{1h}(1, 1) c_{2h}(1, 1) + 2\varepsilon_{0h} \varepsilon_{1h} c_{2h}(1, 1) \right].
 \end{aligned}$$

Taking expectation up-to order n_h^{-3} , we get the mean square error of $\hat{\sigma}_c^2$ as

$$\begin{aligned}
 (2.33) \quad \text{MSE}(\hat{\sigma}_c^2) &= \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[A_{Yh} + \frac{C_{Xh}^2}{\theta_{Xh}} c_{1h}^2(1, 1) + A_{Xh} c_{2h}^2(1, 1) \right. \\
 &\quad \left. + 2\delta_{21h} C_{Xh} c_{1h}(1, 1) + 2\delta_{03h} C_{Xh} c_{1h}(1, 1) c_{2h}(1, 1) + 2(\delta_{22h} - 1) c_{2h}(1, 1) \right],
 \end{aligned}$$

where $c_{1h}(1, 1)$ and $c_{2h}(1, 1)$ denote the first order partial derivatives of $c_h(l_h, m_h)$ with respect to l_h and m_h respectively about the point $(1, 1)$:

$$(2.34) \quad \begin{bmatrix} C_{Xh}^2 \theta_{Xh} & \delta_{03h} C_{Xh} \\ \delta_{021h} C_{Xh} & A_{Xh} \end{bmatrix} \begin{bmatrix} c_{1h}(1, 1) \\ c_{2h}(1, 1) \end{bmatrix} = - \begin{bmatrix} \delta_{21h} C_{Xh} \\ \delta_{22h} - 1 \end{bmatrix}.$$

By solving (2.34) we can determine the minimum values of $c_{1h}(1, 1)$ and $c_{2h}(1, 1)$ respectively as

$$\begin{aligned}
 (2.35) \quad c_{1h}(1, 1) &= \frac{[\delta_{03h}(\delta_{22h} - 1) - \delta_{21h} A_{Xh}]}{[C_{Xh}(A_{Xh}/\theta_{Xh}) - \delta_{03h}^2]}, \\
 c_{2h}(1, 1) &= \frac{[\delta_{03h} \delta_{21h} - (\delta_{22h} - 1)/\theta_{Xh}]}{[(A_{Xh}/\theta_{Xh}) - \delta_{03h}^2]}.
 \end{aligned}$$

Putting (2.35) in (2.33) we obtain minimum MSE of $\hat{\sigma}_c^2$ as

$$(2.36) \quad \min.\text{MSE}(\hat{\sigma}_c^2) = \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[A_{Yh} - \frac{(\delta_{22h} - 1)^2 + \delta_{21h}^2 \theta_{Xh} A_{Xh} - 2\theta_{Xh} \delta_{21h} \delta_{03h} (\delta_{22h} - 1)}{(A_{Xh} - \delta_{03h}^2 \theta_{Xh})} \right],$$

$$(2.37) \quad \min.\text{MSE}(\hat{\sigma}_c^2) = \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[A_{Yh} - \delta_{21h}^2 \theta_{Xh} - \frac{\{\delta_{21h} \theta_{Xh} \delta_{03h} - (\delta_{22h} - 1)\}^2}{(A_{Xh} - \delta_{03h}^2 \theta_{Xh})} \right].$$

Hence, a theorem can be established as follows.

Theorem 2.3. Up to terms of order n_h^{-3} ,

$$\min.\text{MSE}(\hat{\sigma}_c^2) \geq \sum_{h=1}^L \left(\frac{(W_h \sigma_{Yh})^4}{n_h^3} \right) \left[A_{Yh} - \delta_{21h}^2 \theta_{Xh} - \frac{\{\delta_{21h} \theta_{Xh} \delta_{03h} - (\delta_{22h} - 1)\}^2}{(A_{Xh} - \delta_{03h}^2 \theta_{Xh})} \right],$$

with equality holding if

$$(2.38) \quad \begin{aligned} c_{1h}(1, 1) &= \frac{[\delta_{03h}(\delta_{22h} - 1) - \delta_{21h} A_{Xh}]}{[C_{Xh}(A_{Xh}/\theta_{Xh}) - \delta_{03h}^2]}, \\ c_{2h}(1, 1) &= \frac{[\delta_{03h} \delta_{21h} - (\delta_{22h} - 1)/\theta_{Xh}]}{[(A_{Xh}/\theta_{Xh}) - \delta_{03h}^2]}. \end{aligned}$$

The present estimators

$$\begin{aligned} \hat{\sigma}_{c1}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \hat{\sigma}_{Yh}^2 l_h^{\alpha_{1h}} m_h^{\alpha_{2h}}, \\ \hat{\sigma}_{c2}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \hat{\sigma}_{Yh}^2 [\alpha_{1h} l_{1h} + (1 - \alpha_{1h}) m_{1h}^{\alpha_{2h}}], \\ \hat{\sigma}_{c3}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \hat{\sigma}_{Yh}^2 [\alpha_{3h} l_{1h}^{\alpha_{1h}} + (1 - \alpha_{3h}) m_{2h}^{\alpha_{2h}}], \\ \hat{\sigma}_{c4}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \hat{\sigma}_{Yh}^2 [\exp\{\alpha_{1h}(l_h - 1) + \alpha_{2h}(m_h - 1)\}], \\ \hat{\sigma}_{c5}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \frac{\hat{\sigma}_{Yh}^2}{[1 + \alpha_{1h}\{l_{1h}^{\alpha_{2h}} m_{2h}^{\alpha_{3h}} - 1\}]}, \\ \hat{\sigma}_{c6}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \hat{\sigma}_{Yh}^2 [1 - \alpha_{1h}(l_h - 1) + \alpha_{2h}(m_h - 1)], \\ \hat{\sigma}_{c7}^2 &= \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \hat{\sigma}_{Yh}^2 [1 - \alpha_{1h}(l_h - 1) + \alpha_{2h}(m_h - 1)]^{-1}, \end{aligned}$$

are some particular members of the suggested class of estimator $\hat{\sigma}_c^2$. The mean square error of these estimators can be obtained from (2.33) by choosing the suitable value for the constants. The lower bound of the MSE of the estimators $\hat{\sigma}_{ci}^2$ ($i = 1$ to 7) is the same as given by (2.37).

A class of estimators for σ_{st}^2 wider than $\hat{\sigma}_c^2$ is proposed as

$$(2.39) \quad \hat{\sigma}_f^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) f_h(\hat{\sigma}_{Yh}^2, l_h, m_h),$$

where $f_h(\hat{\sigma}_{Yh}^2, l_h, m_h)$ is a function of $(\hat{\sigma}_{Yh}^2, l_h, m_h)$ such that

$$f_h(\sigma_{Yh}^2, 1, 1) = \sigma_{Yh}^2 \Rightarrow f_{1h}(\sigma_{Yh}^2, 1, 1) = \frac{\partial f_h(\sigma_{Yh}^2, l_h, m_h)}{\partial \hat{\sigma}_{Yh}^2} |_{(\sigma_{Yh}^2, 1, 1)} = 1.$$

It can unveil that up-to order n_h^{-3} the optimum MSE of $\hat{\sigma}_f^2$ is same as the optimum MSE of $\hat{\sigma}_c^2$ at (2.36) or (2.37) and is not reduced. The difference-type estimator

$$(2.40) \quad \hat{\sigma}_{std_3}^2 = \sum_{h=1}^L \left(\frac{W_h^2}{n_h} \right) \left\{ \hat{\sigma}_{Yh}^2 + d_{3h}(l_h - 1) + d_{4h}(m_h - 1) \right\}$$

is a specific member of the class (2.39) but not (2.26), where d_{3h} and d_{4h} are acceptable constants.

3. EFFICIENCY COMPARISONS

To obtain the conditions for which the proposed classes of estimators $\hat{\sigma}_a^2$, $\hat{\sigma}_b^2$ and $\hat{\sigma}_c^2$ perform better than usual unbiased estimator $\hat{\sigma}_{st}^2$, from (1.5), (2.10), (2.23) and (2.37) we can write

$$(3.1) \quad \text{MSE}(\hat{\sigma}_{st}^2) - \min.\text{MSE}(\hat{\sigma}_a^2) = \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \delta_{21h}^2 \theta_{Xh} \geq 0,$$

$$(3.2) \quad \text{MSE}(\hat{\sigma}_{st}^2) - \min.\text{MSE}(\hat{\sigma}_b^2) = \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \frac{(\delta_{22h} - 1)^2}{A_{Xh}} \geq 0,$$

$$(3.3) \quad \text{MSE}(\hat{\sigma}_{st}^2) - \min.\text{MSE}(\hat{\sigma}_c^2) = \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[\delta_{21h}^2 \theta_{Xh} + \frac{\{\theta_{Xh} \delta_{03h} \delta_{21h} - (\delta_{22h} - 1)\}^2}{(A_{Xh} - \delta_{h03}^2 \theta_{Xh})} \right] \geq 0.$$

Remarks: To exhibit the impact of measurement error on MSE of the estimators, let the observation for both the study variable and auxiliary variable be recorded without error. Now the MSE of the proposed class of estimators $\hat{\sigma}_a^2$, to the third degree of approximation is given as

$$(3.4) \quad \text{MSE}^*(\hat{\sigma}_a^2) = \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[(\delta_{40h} - 1) + C_{Xh}^2 a_{1h}^2(1) + 2\delta_{21h} C_{Xh} a_{1h}(1) \right],$$

which is the same as the obtained by Singh and Vishwakarma [14].

From (2.8) and (3.4) we have

$$\begin{aligned} \text{MSE}(\hat{\sigma}_a^2) - \text{MSE}^*(\hat{\sigma}_a^2) &= \sum_{h=1}^L \frac{(W_h \sigma_{Yh}^4)}{n_h^3} \left[(\gamma_{2Uh} + 2) \left(\frac{1 - \theta_{Yh}}{\theta_{Yh}} \right)^2 + 4 \left(\frac{1 - \theta_{Yh}}{\theta_{Yh}} \right) \right. \\ &\quad \left. + C_{Xh}^2 (1 - \theta_{Xh}) a_{1h}^2(1) \right]. \end{aligned}$$

The difference is always positive in nature, thus we can infer that the presence of measurement error incorporates larger mean square error than the absence of measurement error.

To obtain the optimum value of the constant differentiating partially (3.4) with respect to a_{1h} and equate to zero we get

$$a_{1h}(1) = - \left(\frac{\delta_{21h}}{C_{Xh}} \right).$$

Thus, the resultant minimum mean square error is

$$(3.5) \quad \min.\text{MSE}^*(\hat{\sigma}_a^2) = \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} [(\delta_{40h} - 1) - \delta_{21h}^2].$$

Now the impact of measurement error can be obtained (2.10) and (3.5) from and as

$$(3.6) \quad \begin{aligned} \min.\text{MSE}(\hat{\sigma}_a^2) - \min.\text{MSE}^*(\hat{\sigma}_a^2) &= \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[(\gamma_{2Uh} + 2) \left(\frac{1 - \theta_{Yh}}{\theta_{Yh}} \right)^2 \right. \\ &\quad \left. + 4 \left(\frac{1 - \theta_{Yh}}{\theta_{Yh}} \right) + \delta_{21h}^2 (1 - \theta_{Xh}) \right]. \end{aligned}$$

The MSE of another proposed class of estimators $\hat{\sigma}_b^2$ for the estimation of σ_{st}^2 in the absence of measurement error is given in (3.7) and is similar to Singh and Vishwakarma [14]:

$$(3.7) \quad \text{MSE}^*(\hat{\sigma}_b^2) = \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[\delta_{40h} - 1 + (\delta_{04h} - 1)b_{1h}^2(1) + 2(\delta_{22h} - 1)b_{1h}(1) \right].$$

From equation (2.21) and (3.7) we can write

$$(3.8) \quad \begin{aligned} \text{MSE}(\hat{\sigma}_b^2) - \text{MSE}^*(\hat{\sigma}_b^2) &= \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[(\gamma_{2Uh} + 2) \left(\frac{1 - \theta_{Yh}}{\theta_{Yh}} \right)^2 + 4 \left(\frac{1 - \theta_{Yh}}{\theta_{Yh}} \right) \right. \\ &\quad \left. + (\gamma_{2Vh} + 2) \left(\frac{1 - \theta_{Xh}}{\theta_{Xh}} \right)^2 + 4 \left(\frac{1 - \theta_{Xh}}{\theta_{Xh}} \right) b_{1h}^2(1) \right]. \end{aligned}$$

The right-hand side of (3.8) is always positive in nature, thus we can infer that mean square error of the proposed estimator is always larger when observation is recorded with error. $\text{MSE}^*(\hat{\sigma}_b^2)$ is minimized for

$$(3.9) \quad b_{1h}(1) = - \left(\frac{\delta_{22h} - 1}{\delta_{04} - 1} \right).$$

Thus, the resultant minimum mean square error is

$$(3.10) \quad \min.\text{MSE}^*(\hat{\sigma}_b^2) = \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[(\delta_{40h} - 1) - \frac{(\delta_{22h} - 1)^2}{(\delta_{04h} - 1)} \right].$$

From (2.23) and (3.10) we can derive the impact of measurement error as

$$(3.11) \quad \begin{aligned} \min.\text{MSE}(\hat{\sigma}_b^2) - \min.\text{MSE}^*(\hat{\sigma}_b^2) &= \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[(\gamma_{2Uh} + 2) \left(\frac{1 - \theta_{Yh}}{\theta_{Yh}} \right)^2 + 4 \left(\frac{1 - \theta_{Yh}}{\theta_{Yh}} \right) \right. \\ &\quad \left. + \frac{(\delta_{22h} - 1)^2}{A_{Xh}(\delta_{40h} - 1)} (\gamma_{2Vh} + 2) \left(\frac{1 - \theta_{Xh}}{\theta_{Xh}} \right)^2 + 4 \left(\frac{1 - \theta_{Xh}}{\theta_{Xh}} \right) \right]. \end{aligned}$$

The mean square error of the third proposed class of estimators $\hat{\sigma}_c^2$ for the estimation of σ_{st}^2 in the absence of measurement error is given by Singh and Vishwakarma [14] as

$$(3.12) \quad \begin{aligned} \text{MSE}^*(\hat{\sigma}_c^2) &= \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[(\delta_{40h} - 1) + C_{Xh}^2 c_{1h}^2(1, 1) + (\delta_{04h} - 1) c_{2h}^2(1, 1) \right. \\ &\quad \left. + 2\delta_{21h} C_{Xh} c_{1h}(1, 1) + 2\delta_{03h} C_{Xh} c_{1h}(1, 1) c_{2h}(1, 1) \right. \\ &\quad \left. + 2(\delta_{22h} - 1) c_{2h}(1, 1) \right]. \end{aligned}$$

From we can write as

$$\begin{aligned}
 \text{MSE}(\hat{\sigma}_c^2) - \text{MSE}^*(\hat{\sigma}_c^2) &= \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[(\gamma_{2U_h} + 2) \left(\frac{1 - \theta_{Yh}}{\theta_{Yh}} \right)^2 + 4 \left(\frac{1 - \theta_{Yh}}{\theta_{Yh}} \right) \right. \\
 (3.13) \quad &+ C_{Xh}^2 \left(\frac{1 - \theta_{Xh}}{\theta_{Xh}} \right) c_{1h}^2(1) + (\gamma_{2V_h} + 2) \left(\frac{1 - \theta_{Xh}}{\theta_{Xh}} \right)^2 \\
 &\left. + 4 \left(\frac{1 - \theta_{Xh}}{\theta_{Xh}} \right) c_{2h}^2 \right]
 \end{aligned}$$

and $\text{MSE}^*(\hat{\sigma}_c^2)$ is minimum for

$$\begin{aligned}
 (3.14) \quad c_{1h}(1, 1) &= \frac{[\delta_{03h}(\delta_{22h} - 1) - \delta_{21h}(\delta_{04h} - 1)]}{C_{Xh} [\delta_{04h} - \delta_{03h}^2 - 1]}, \\
 c_{2h}(1, 1) &= \frac{[\delta_{03h}\delta_{21h} - (\delta_{22h} - 1)]}{[\delta_{04h} - \delta_{03h}^2 - 1]}.
 \end{aligned}$$

Thus we can write the resultant minimum MSE as

$$(3.15) \quad \min.\text{MSE}^*(\hat{\sigma}_c^2) = \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[(\delta_{40h} - 1) - \delta_{21h}^2 - \frac{\{\delta_{03h}\delta_{21h} - (\delta_{22h} - 1)\}^2}{(\delta_{04h} - \delta_{03h}^2 - 1)} \right],$$

which can be easily obtained from (2.37) by putting $\sigma_{U_h}^2 = \sigma_{V_h}^2 = 0$.

Hence, we can derive the impact of measurement error on the mean square error of the estimator $(\hat{\sigma}_c^2)$ as

$$\begin{aligned}
 (3.16) \quad \min.\text{MSE}(\hat{\sigma}_c^2) - \min.\text{MSE}^*(\hat{\sigma}_c^2) &= \\
 &= \sum_{h=1}^L \frac{(W_h \sigma_{Yh})^4}{n_h^3} \left[(\gamma_{2U_h} + 2) \left(\frac{1 - \theta_{Yh}}{\theta_{Yh}} \right)^2 + 4 \left(\frac{1 - \theta_{Yh}}{\theta_{Yh}} \right) + (1 - \theta_{Xh})\delta_{21h}^2 \right. \\
 &\quad \left. - \left\{ \frac{A_1(\delta_{04h} - \delta_{03h}^2 - 1) - B_1(A_{Xh} - \delta_{h03}^2 \theta_{Xh})}{(\delta_{04h} - \delta_{03h}^2 - 1)(A_{Xh} - \delta_{h03}^2 \theta_{Xh})} \right\} \right],
 \end{aligned}$$

where

$$\begin{aligned}
 A_1 &= \left\{ \delta_{03h}\delta_{21h}\theta_{Xh} - (\delta_{22h} - 1) \right\}^2, \\
 B_1 &= \left\{ \delta_{03h}\delta_{21h} - (\delta_{22h} - 1) \right\}^2.
 \end{aligned}$$

The right-hand side of (3.16) is the effect of measurement error in the mean square error of the estimator which is always positive in nature. Thus, the proposed classes of estimators have larger MSE in the presence of measurement errors in both study and auxiliary variables than in the absence of measurement errors. When the measurement error is insignificant, the inference based on these data may remain valid. Nevertheless, when the amount of error is more significant in observed data, the inference may be invalid and inaccurate and often may lead to unexpected and undesirable consequences.

4. DISCUSSION AND CONCLUSION

The data available for statistical analysis are always contaminated with measurement error and may lead to fallacious inference results. When data contains heterogeneity among units in terms of value, survey users are advised to form several homogeneous groups, and the sampling design is well known as stratified sampling. To the best of our knowledge, the study of measurement error for the estimation of variance in stratified random sampling has not been addressed yet. Singh and Karpe [12] have studied the effect of measurement error on estimation of population mean in stratified random sampling. Estimation of variance has vital importance as it has practical uses in real-life. It is discussed by Lee [6], Srivastava and Jhaji [18], Wu [21] and Singh and Vishwakarma [14] without the measurement error framework.

The present study deals with the problem of estimation of variance by using auxiliary information under the stratified sampling framework when observations are contaminated by measurement errors. Three wider classes of estimators have been proposed. The theoretical comparisons show that the proposed classes of estimators ($\hat{\sigma}_a^2$, $\hat{\sigma}_b^2$ and $\hat{\sigma}_c^2$) in the presence of measurement error are more efficient than usual unbiased estimators. Since the proposed estimators are defined as a class, a large number of estimators become the members of this class. So the impact of measurement error on the bias and the mean square error of these estimators can be obtained easily. We can also conclude that the MSE in the presence of measurement error is larger than in the absence of it. Thus, the present study for the estimation of variance under measurement error for the stratified random sampling is useful and may attract others to carry out some work of practical use in this direction.

ACKNOWLEDGMENTS

Authors are deeply indebted to the editor-in-chief Prof. Maria Ivette Gomes and learned referee for their valuable suggestions leading to improvement the quality of contents and presentation of the original manuscript. Authors are also thankful to Dr. (Md.) Mojibur Rahman, Associate Professor of English Language, Department of Humanities and Social Sciences, Indian Institute of Technology (ISM) Dhanbad, India, for his kind support in improving the grammar of this manuscript.

REFERENCES

- [1] ALLEN, J.; SINGH, H.P. and SMARANDACHE, F. (2003). A family of estimators of population mean using multi-auxiliary information in presence of measurement errors, *International Journal of Social Economics*, **30**(7), 837–849.
- [2] BIRADAR, R.S. and SINGH, H.P. (1994). An alternative to ratio estimator for population variance, *Assam Statistical Review*, **8**(2), 18–33.

- [3] COCHRAN, W.G. (1968). Errors of measurement in statistics, *Technometrics*, **10**(4), 637–666.
- [4] DAS, A.K. and TRIPATHI, T.P. (1978). Use of auxiliary information in estimating the finite population variance, *Sankhya*, **40**(C), 139–148.
- [5] ISAKI, C.T. (1983). Variance estimation using auxiliary information, *Journal of the American Statistical Association*, **78**(381), 117–123.
- [6] LEE, K.H. (1973). Variance estimation in stratified sampling, *Journal of the American Statistical Association*, **68**(342), 336–342.
- [7] MANEESHA and SINGH, R.K. (2002). Role of regression estimator involving measurement errors, *Brazilian Journal of Probability and Statistics*, **16**(1), 39–46.
- [8] PRASAD, B. and SINGH, H.P. (1990). Some improved ratio-type estimators of population variance using auxiliary information in sample surveys, *Communication in Statistics – Theory and Methods*, **21**(5), 1367–1376.
- [9] SHALABH (1997). Ratio method of estimation in the presence of measurement errors, *Journal of the Indian Society of Agricultural Statistics*, **50**(2), 150–155.
- [10] SHALABH and TSAI, J.R. (2017). Ratio and product methods of estimation of population mean in the presence of correlated measurement errors, *Communications in Statistics – Simulation and Computation*, **46**(7), 5566–5593.
- [11] SINGH, H.P. and BIRADAR, R.S. (1994). Estimation of finite population variance using auxiliary information, *Journal of the Indian Society of Statistics and Operation Research*, **15**(1), 47–63.
- [12] SINGH, H.P. and KARPE, N. (2010). Effect of measurement errors on the separate and combined ratio and product estimators in stratified random sampling, *Journal of Modern Applied Statistical Methods*, **9**(2), 388–402.
- [13] SINGH, H.P.; UPADHYAYA, L.N. and NAMJOSHI, U.D. (1988). Estimation of finite population variance, *Current Science*, **57**(27), 1331–1334.
- [14] SINGH, H.P. and VISHWAKARMA, G.K. (2008). Some families of estimators of variance of stratified random sample mean using auxiliary information, *Journal of Statistical Theory and Practice*, **2**(1), 21–43.
- [15] SINGH, N. and VISHWAKARMA, G.K. (2019). A generalised class of estimator of population mean with the combined effect of measurement errors and non-response in sample survey, *Revista Investigacion Operacional*, **40**(2), 275–285.
- [16] SRIVASTAVA, A.K. and SHALABH (2001). Effect of measurement errors on the regression method of estimation in survey sampling, *Journal of Statistical Research*, **35**(2), 35–44.
- [17] SRIVASTAVA, S.K. (1971). A generalized estimator for the mean of a finite population using multiauxiliary information, *Journal of the American Statistical Association*, **66**, 404–407.
- [18] SRIVASTAVA, S.K. and JHAJJ, H.S. (1980). A class of estimators using auxiliary information for estimating finite population variance, *Sankhya*, **42**(C), 87–96.
- [19] UPADHYAYA, L.N. and SINGH, H.P. (1983). Use of auxiliary information in the estimation of population variance, *Mathematical Forum*, **6**(2), 33–36.
- [20] UPADHYAYA, L.N. and SINGH, H.P. (1986). On a dual to ratio estimator for estimating finite population variance, *Nepalese Mathematical Scientific Report*, **11**(1), 37–42.
- [21] WU, C.F.J. (1985). Variance estimation for the combined ratio and combined regression estimators, *Journal of the Royal Statistical Society, Series B (Methodological)*, **47**(1), 147–154.

CONCOMITANTS OF ORDER STATISTICS AND RECORD VALUES FROM ITERATED FGM TYPE BIVARIATE-GENERALIZED EXPONENTIAL DISTRIBUTION

Authors: H.M. BARAKAT
– Mathematical Department, Zagazig University – Faculty of Science,
Zagazig, Egypt
hbarakat2@hotmail.com

E.M. NIGM
– Mathematical Department, Zagazig University – Faculty of Science,
Zagazig, Egypt
s_nigm@yahoo.com

M.A. ALAWADY
– Mathematical Department, Zagazig University, Faculty of Science,
Zagazig, Egypt
Ma_alawady@yahoo.com

I.A. HUSSEINY
– Mathematical Department, Zagazig University – Faculty of Science,
Zagazig, Egypt
ishusseiny@gmail.com

Received: January 2018

Revised: December 2018

Accepted: April 2019

Abstract:

- We introduce the successive iterations in the original FGM type bivariate-generalized exponential distribution. Some distributional properties of concomitants of order statistics as well as record values for this family are studied. Recurrence relations between single as well as product moments of concomitants are obtained. Finally, we give some different applications of this study.

Keywords:

- *concomitants; order statistics; record values; generalized exponential distribution; iterated FGM family.*

AMS Subject Classification:

- 62B10, 62G30.

1. INTRODUCTION

Ordered random variables (rv's) have attracted many researchers due to their applicability in many practical areas, like order statistics (os's) and record values. Both of os's and record values are used extensively in statistical models and inference, where they describe rv's arranged in order magnitude. The os's occur as a natural choice when dealing with floods, drought, earthquakes, etc. Also, record values arise naturally in many real life applications involving data related to sport, weather, and life testing studies. Actually, there is strong relation between the os and the record value models. For example, the record values provide the information about the maximum (minimum) value among all previously recorded observations, for more detail, see Arnold *et al.* [5] (1998). The concept of concomitant os's, also called induced os's, is related to the ordering bivariate rv's. The concomitant os's arise when one sorts the members of a random sample according to corresponding values of an other random sample. The term concomitant of os's was first induced and applied extensively by David [13] (1973). According to Hanif [22] (2007) in collecting any data for an observation, several characteristics are often recorded, some of them are considered as primary and others can be observed from the primary data automatically. The latter one is called concomitant, for more detail see David and Nagaraja [14, 15] (1998, 2003). The most important use of concomitants of record values arises in experiments, in which specified characteristic's measurements of an individual are made sequentially. Moreover, only values that exceed or fall below the current extreme value are recorded, so that only observations are bivariate record values, i.e., records and their concomitants. Some properties of concomitants of record values are discussed by Ahsanullah [1] (2009) and Ahsanullah and Shakil [2] (2013). Clearly, both concomitants of os's and record values are strongly relevant with a bivariate data that has a common bivariate distribution function (df). One of the most useful and popular bivariate df is the so-called Farlie–Gumbel–Morgenstern (FGM). The FGM df is defined by $H(x, y) = F_X(x)F_Y(y)[1 + \alpha F_X(x)F_Y(y)]$, where F_X and F_Y are the marginals df's, while \bar{F}_X and \bar{F}_Y are the survival function of F_X and F_Y , respectively, and $-1 \leq \alpha \leq 1$. The FGM distribution is a flexible family useful in applications provided that the correlation between the variables is not too large. It can be utilized for arbitrary continuous marginals. The FGM df was originally introduced by Morgenstern [29] (1956) for Cauchy marginals. In 1960 Gumbel [20] investigated the same structure for exponential marginals. Also, in 1960, Farlie [18], in connection with his investigations of the correlation coefficient, suggested a generalization of the bivariate form studied by Morgenstern and Gumbel. Huang and Kotz [25] (1984) used successive iterations in the original FGM distribution to increase the correlation between components. As a particular case, the bivariate FGM with a single iteration is defined by

$$(1.1) \quad F_{X,Y}(x, y) = F_X(x)F_Y(y) \left[1 + \lambda \bar{F}_X(x)\bar{F}_Y(y) + \gamma F_X(x)F_Y(y)\bar{F}_X(x)\bar{F}_Y(y) \right],$$

denoted by FGM(λ, γ). The corresponding probability density functions (pdf) is given by:

$$(1.2) \quad \begin{aligned} f_{X,Y}(x, y) = \\ = f_X(x)f_Y(y) \left[1 + \lambda(1 - 2F_X(x))(1 - 2F_Y(y)) + \gamma F_X(x)F_Y(y)(2 - 3F_X(x))(2 - 3F_Y(y)) \right], \end{aligned}$$

where $F_X(x)$ and $F_Y(y)$ are df's, while $f_X(x)$ and $f_Y(y)$ are the pdf's of the rv's X and Y , respectively. When the two marginals $F_X(x)$ and $F_Y(y)$ are continuous, Huang and Kotz [25] (1984) showed that the natural parameter space Ω (the admissible set of the parameters λ

and γ that makes $F_{X,Y}(x, y)$ is a df) is convex, where $\Omega = \{(\lambda, \gamma): -1 \leq \lambda \leq 1; \lambda + \gamma \geq -1; \gamma \leq \frac{3-\lambda+\sqrt{9-6\lambda-3\lambda^2}}{2}\}$. Moreover, when the marginals are uniform then, the correlation coefficient is $\rho = \frac{\lambda}{3} + \frac{\gamma}{12}$ (cf. Huang and Kotz [26], 1999). Finally, the maximal correlation coefficient attained for this family is $\max \rho = 0.434$ versus $\max \rho = \frac{1}{3} = 0.333$ achieved for $\lambda = 1$ in the original FGM version. This fact gives a satisfactory motivation to deal with the model FGM(λ, γ) rather than the classical model FGM. The model FGM(λ, γ) provides a very general expression of a bivariate distribution from which members can be derived by substituting expressions of any desired set of marginal distributions. On the other hand, since both the bivariate df's and density are given in terms of marginals, it is easy to generate a random sample from the model FGM(λ, γ). Thus members of this family can be used in simulation studies. Moreover, a number of properties results from the simple analytic form of the model FGM(λ, γ), for example, rv's having a FGM(λ, γ) are exchangeable whenever the marginal distributions are identical. Also, the model FGM(λ, γ) is closed with respect to monotonic increasing functions of rv's. Moreover, the system is closed with respect to mixtures of bivariate FGM(λ, γ) df's having the same marginal distributions. the bivariate FGM(λ, γ) df's are specially suited to data situations describing weak dependence between the rv's X and Y . Measures of dependence vary over a larger range than for the classical FGM df's.

In this paper, we study the family FGM(λ, γ), with generalized exponential (GE) marginals. The generalized exponential distribution (GE), a most attractive generalization of the exponential distribution, introduced by Gupta and Kundu [21] (1999), has widespread interest and applications, e.g., it can be used quite effectively in analyzing many lifetime data, particularly in place of two-parameter gamma and two-parameter Weibull distributions. Many authors studied various properties of the GE, see for example, Ahsanullah *et al.* [3] (2013) and AL-Hussaini and Ahsanullah [4] (2015).

A continuous rv is said to be has the GE with scale parameter $\theta > 0$ and shape parameter $\alpha > 0$ (denoted by GE($\theta; \alpha$)), if the df and the corresponding pdf are given, for $x > 0$, respectively, by

$$F_X(x) = (1 - \exp(-\theta x))^\alpha$$

and

$$(1.3) \quad f_X(x) = \alpha\theta(1 - \exp(-\theta x))^{\alpha-1} \exp(-\theta x).$$

Gupta and Kundu [21] (1999) showed that the k -th moment of GE($\theta; \alpha$) is

$$\mu_k = \frac{\alpha k!}{\theta^k} \sum_{i=0}^{\aleph(\alpha-1)} \frac{(-1)^i}{(i+1)^{k+1}} \binom{\alpha-1}{i},$$

where $\aleph(x) = \infty$, if x is non-integer and $\aleph(x) = x$, if x is integer. Furthermore, the mean, variance and moment generating function of GE($\theta; \alpha$) are given by $\mu_1 = E(X) = \frac{B(\alpha)}{\theta}$, $\text{Var}(X) = \frac{C(\alpha)}{\theta^2}$ and $M_X(t) = \alpha\beta(\alpha, 1 - \frac{t}{\theta})$, respectively, where $B(\alpha) = \Psi(\alpha + 1) - \Psi(1)$, $C(\alpha) = \Psi'(1) - \Psi'(\alpha + 1)$, $\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ and $\Psi(\cdot)$ is the digamma function, while $\Psi'(\cdot)$ is its derivation ($\Psi'(\cdot)$ is known as the trigamma function). Tahmasebi and Jafari [38] (2015) studied some properties of the classical FGM type bivariate GE df. Moreover, Tahmasebi and Jafari [38] (2015) studied some distributional properties of concomitants of os's as well as record values of this df.

In this paper, the result of Tahmasebi and Jafari [38] (2015) is extended to FGM(λ, γ) family with two marginals F_X and F_Y , where $X \sim \text{GE}(\theta_1; \alpha_1)$ and $Y \sim \text{GE}(\theta_2; \alpha_2)$ (denoted by FGM($\lambda, \gamma; \theta_1, \alpha_1; \theta_2, \alpha_2$)). Moreover, some new results, which were not obtained by Tahmasebi and Jafari [38] (2015) for FGM family, are given such as recurrence relations for the single, as well as the product, moments of bivariate concomitants of os's, the concomitant rank-os's, and the asymptotic behavior of the concomitants of os's. It is worth mentioning that, the same problem tackled by Barakat *et al.* [7, 6] (2019, 2018) for the Huang–Kotz FGM and Bairamov–Kotz–Becki FGM with GE marginals, respectively. Moreover, the FGM($\lambda, \gamma; \theta_1, \alpha_1; \theta_2, \alpha_2$) is not a special case of any of the latter models. Nowadays, we can find several recent relevant works on this subject. Among these works are Tahmasebi and Behboodian [36] (2012), Tahmasebi and Jafari [37] (2014) and Tahmasebi *et al.* [39, 40] (2015, 2016).

2. THE FGM($\lambda, \gamma; \theta_1, \alpha_1; \theta_2, \alpha_2$) FAMILY AND SOME OF ITS PROPERTIES

The joint df and pdf of (X, Y) are defined by (1.1) and (1.2), respectively, where $X \sim \text{GE}(\theta_1; \alpha_1)$ and $Y \sim \text{GE}(\theta_2; \alpha_2)$. Thus, it is easy to show that the (n, m) -th joint moments the of FGM($\lambda, \gamma; \theta_1, \alpha_1; \theta_2, \alpha_2$) family is given by

$$(2.1) \quad \begin{aligned} E(X^n Y^m) &= E(X^n)E(Y^m) + \lambda(E(X^n) - E(U_1^n))(E(Y^m) - E(V_1^m)) \\ &\quad + \gamma(E(U_1^n) - E(U_2^n))(E(V_1^m) - E(V_2^m)), \quad n, m = 1, 2, \dots, \end{aligned}$$

where $U_1 \sim \text{GE}(\theta_1; 2\alpha_1)$, $U_2 \sim \text{GE}(\theta_1; 3\alpha_1)$, $V_1 \sim \text{GE}(\theta_2; 2\alpha_2)$ and $V_2 \sim \text{GE}(\theta_2; 3\alpha_2)$. Thus, by combining (2.1) and (1.3), we get

$$E(XY) = \frac{B(\alpha_1)B(\alpha_2) + \lambda D(2\alpha_1)D(2\alpha_2) + \gamma D(3\alpha_1)D(3\alpha_2)}{\theta_1 \theta_2},$$

where $D((k + 1)\alpha) = B((k + 1)\alpha) - B(k\alpha)$, $k = 1, 2$. Therefore, the coefficient of correlation between X and Y is

$$\rho_{X,Y} = \frac{\lambda D(2\alpha_1)D(2\alpha_2) + \gamma D(3\alpha_1)D(3\alpha_2)}{\sqrt{C(\alpha_1)C(\alpha_2)}} = \lambda g_1(\alpha_1, \alpha_2) + \gamma g_2(\alpha_1, \alpha_2).$$

Clearly, the function $g_1(\alpha_1, \alpha_2)$ and $g_2(\alpha_1, \alpha_2,)$ is increasing and positive function with respect to each of $\alpha_i, i = 1, 2$. Therefore, if $\lambda, \gamma > 0$, then $\rho_{X,Y}$ is increasing and positive function and if $\lambda, \gamma < 0$, then $\rho_{X,Y}$ is decreasing and negative function with respect to each of α_1 and α_2 . Moreover, we can show that $\lim_{\alpha_1 \rightarrow \infty} g_1(\alpha_1, \alpha_2) = \frac{6(\log(2))^2}{\pi^2}$, $\lim_{\alpha_2 \rightarrow \infty} g_2(\alpha_1, \alpha_2,) = \frac{6(\log(\frac{3}{2}))^2}{\pi^2}$, $\lim_{\alpha_1 \rightarrow 0^+} g_1(\alpha_1, \alpha_2) = 0$ and $\lim_{\alpha_2 \rightarrow 0^+} g_2(\alpha_1, \alpha_2) = 0$. Therefore, $\max \rho_{X,Y} = 0.392$ at corner point $(\lambda, \gamma) = (1, 1)$ and $\min \rho_{X,Y} = -0.292$ at corner point $(\lambda, \gamma) = (-1, 0)$.

The conditional df of Y given $X = x$ is given by

$$(2.2) \quad \begin{aligned} F_{Y|X}(y|x) &= F_Y(y) \left[1 + \lambda(1 - F_Y(y))(1 - 2F_X(x)) \right. \\ &\quad \left. - \gamma F_X(x)F_Y(y)(1 - F_Y(y))(2 - 3F_X(x)) \right]. \end{aligned}$$

Therefore, the regression curve of Y given $X = x$ for $\text{FGM}(\lambda, \gamma; \theta_1, \alpha_1; \theta_2, \alpha_2)$ is

$$\begin{aligned}
 \text{E}(Y|X = x) &= \text{E}(Y) + \lambda(1 - 2F_X(x))(\text{E}(Y) - \text{E}(V_1)) \\
 &\quad + \gamma F_X(x)(2 - 3F_X(x))(\text{E}(V_1) - \text{E}(V_2)) \\
 (2.3) \qquad &= \frac{1}{\theta_2} \left[B(\alpha_2) + \lambda D(2\alpha_2)(2F_X(x) - 1) + \gamma F_X(x) D(3\alpha_2)(3F_X(x) - 2) \right],
 \end{aligned}$$

where $V_1 \sim \text{GE}(\theta_2; 2\alpha_2)$ and $V_2 \sim \text{GE}(\theta_2; 3\alpha_2)$ and the conditional expectation is non-linear with respect to x .

3. CONCOMITANTS OF OS'S BASED ON $\text{FGM}(\lambda, \gamma; \theta_1, \alpha_1; \theta_2, \alpha_2)$

Suppose (X_i, Y_i) , $i = 1, 2, \dots, n$, is a random sample from a bivariate df $F_{X,Y}(x, y)$. If we order the sample by the X -variate, and obtain the os's, $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$, for the X sample, then the Y -variate associated with the r -th order statistic $X_{r:n}$ is called the concomitant of the r -th order statistic, and is denoted by $Y_{[r:n]}$.

Let $X \sim \text{GE}(\theta_1; \alpha_1)$ and $Y \sim \text{GE}(\theta_2; \alpha_2)$. Since the conditional pdf of $Y_{[r:n]}$ given $X_{[r:n]} = x$ is $f_{Y_{[r:n]}|X_{r:n}}(y|x) = f_{Y|X}(y|x)$ (cf. Galambos [19], 1987, see also Tahmasebi and Jafari [38], 2015), then the pdf of $Y_{[r:n]}$ is given by

$$\begin{aligned}
 f_{[r:n]}(y) &= f_Y(y) + \left[\lambda(f_Y(y) - f_{V_1}(y)) + \gamma(f_{V_2}(y) - f_{V_1}(y)) \right] \Delta_{r,n}^{(1)} \\
 (3.1) \qquad &\quad + \left[\gamma(f_{V_1}(y) - f_{V_2}(y)) \right] \Delta_{r,n}^{(2)},
 \end{aligned}$$

where

$$\Delta_{r,n}^{(i)} = \frac{\beta(r, n - r + 1) - (i + 1)\beta(r + i, n - r + 1)}{\beta(r, n - r + 1)}, \quad i = 1, 2.$$

Therefore, the moment generating function of $Y_{[r:n]}$ is given by

$$\begin{aligned}
 M_{[r:n]}(t) &= \alpha_2 \left[\beta \left(\alpha_2, 1 - \frac{t}{\theta_2} \right) + \lambda \Delta_{r,n}^{(1)} \left(\beta \left(\alpha_2, 1 - \frac{t}{\theta_2} \right) - \beta \left(2\alpha_2, 1 - \frac{t}{\theta_2} \right) \right) \right. \\
 &\quad \left. + \gamma \Delta_{r,n}^{(2)} \left(\beta \left(2\alpha_2, 1 - \frac{t}{\theta_2} \right) - \beta \left(3\alpha_2, 1 - \frac{t}{\theta_2} \right) \right) \right].
 \end{aligned}$$

Consequently, the k -th moment of $Y_{[r:n]}$ is given by

$$\begin{aligned}
 \mu_{[r:n]}^{(k)} &= \text{E}[Y_{[r:n]}^k] = \text{E}[Y^k] + \Delta_{r,n}^{(1)} \left(\gamma(\text{E}[V_2^k] - \text{E}[V_1^k]) - \lambda(\text{E}[V_1^k] - \text{E}[Y^k]) \right) \\
 &\quad - \gamma \Delta_{r,n}^{(2)} (\text{E}[V_2^k] - \text{E}[V_1^k]).
 \end{aligned}$$

Moreover, the mean of $Y_{[r:n]}$:

$$(3.2) \quad \mu_{[r:n]} = \mu_{[r:n]}^{(1)} = \frac{1}{\theta_2} \left[B(\alpha_2) + \Delta_{r,n}^{(1)} (\gamma D(3\alpha_2) - \lambda D(2\alpha_2)) - \gamma \Delta_{r,n}^{(2)} D(3\alpha_2) \right].$$

Theorem 3.1. For any $1 \leq r \leq n - 3$, we get

$$\begin{aligned} & \left[(n + 2)A(\lambda, \gamma) - 3(r + 1)\gamma D(3\alpha_2) \right] \mu_{[r+2:n]} = \\ = & \left[2(n + 2)A(\lambda, \gamma) - 3(2r + 3)\gamma D(3\alpha_2) \right] \mu_{[r+1:n]} - \left[(n + 2)A(\lambda, \gamma) - 3(r - 2)\gamma D(3\alpha_2) \right] \mu_{[r:n]}. \end{aligned}$$

Moreover, for all $n > 2$, we get

$$\begin{aligned} & \left[A(\lambda, \gamma)(2 - n(n + 1)) - 3(r + 1)(n - 1)\gamma D(3\alpha_2) \right] \mu_{[r:n]} = \\ & = (n + 2) \left[A(\lambda, \gamma)(n + 1) + 3(r + 1) + 3(r + 1)\gamma D(3\alpha_2) \right] \mu_{[r:n-2]} \\ & \quad - \left[2A(\lambda, \gamma)(n + 2) + 3(r + 1)(2n + 1)\gamma D(3\alpha_2) \right] \mu_{[r:n-1]}, \end{aligned}$$

where $A(\lambda, \gamma) = \gamma D(3\alpha_2) - \lambda D(2\alpha_2)$.

Proof: It is easy, after some algebra, to show that the mean $\mu_{[r:n]}$, defined by (3.2), satisfies the following relation:

$$(3.3) \quad \frac{\mu_{[r+2:n]} - \mu_{[r:n]}}{\mu_{[r+1:n]} - \mu_{[r:n]}} = \frac{A(\lambda, \gamma) \left(\Delta_{r+2,n}^{(1)} - \Delta_{r,n}^{(1)} \right) + \gamma D(3\alpha_2) \left(\Delta_{r+2,n}^{(2)} - \Delta_{r,n}^{(2)} \right)}{A(\lambda, \gamma) \left(\Delta_{r+1,n}^{(1)} - \Delta_{r,n}^{(1)} \right) + \gamma D(3\alpha_2) \left(\Delta_{r+1,n}^{(2)} - \Delta_{r,n}^{(2)} \right)}.$$

On the other hand, we can check that $\Delta_{r+2,n}^{(1)} - \Delta_{r,n}^{(1)} = \frac{-4}{n+1}$, $\Delta_{r+1,n}^{(1)} - \Delta_{r,n}^{(1)} = \frac{-2}{n+1}$, $\Delta_{r+2,n}^{(2)} - \Delta_{r,n}^{(2)} = \frac{-12r-18}{(n+1)(n+2)}$ and $\Delta_{r+1,n}^{(2)} - \Delta_{r,n}^{(2)} = \frac{-6r-6}{(n+1)(n+2)}$. Thus, by combining the last four relations with (3.3), we get first recurrence relation in Theorem 3.1. Also, we can easily check that $\Delta_{r,n}^{(1)} - \Delta_{r,n-2}^{(1)} = \frac{4r}{(n-1)(n+1)}$, $\Delta_{r,n-1}^{(1)} - \Delta_{r,n-2}^{(1)} = \frac{2r}{n(n-1)}$, $\Delta_{r,n}^{(2)} - \Delta_{r,n-2}^{(2)} = \frac{6r(r+1)(2n+1)}{n(n-1)(n+1)(n+2)}$ and $\Delta_{r,n-1}^{(2)} - \Delta_{r,n-2}^{(2)} = \frac{6r(r+1)}{n(n-1)(n+1)}$. The last four relation and the relation (3.2) imply the second recurrence relation of the theorem. This completes the proof. \square

Remark 3.1. By putting $\gamma = 0$ in the two recurrence relations defined in Theorem 3.1 (note that $A(\lambda, 0) = 0$), we get the two corresponding recurrence relations defined in Theorem 3.1 of Barakat *et al.* [7] (2019), at $p = 1$.

By multiplying the both sides of (3.1) by $(y - \mu_{[r:n]})^2$ and integrating, we obtain the variance of $Y_{[r:n]}$ as

$$(3.4) \quad \begin{aligned} \sigma_{[r:n]}^2 = & \frac{1}{\theta_2^2} \left[(1 + \pi_1)(C(\alpha_2) - \pi_1 B^2(2\alpha_2)) + (\pi_2 - \pi_1)(C(2\alpha_2) + B^2(2\alpha_2)) \right. \\ & - B^2(2\alpha_2)(\pi_1 + \pi_2)^2 + \pi_2(C(3\alpha_2) - B^2(3\alpha_2)(1 + \pi_2)) - 2B(\alpha_2)B(2\alpha_2)\pi_3 \\ & \left. - 2B(\alpha_2)B(3\alpha_2)\pi_2(1 + \pi_1) - 2B(2\alpha_2)B(3\alpha_2)\pi_2(\pi_1 + \pi_2) \right], \end{aligned}$$

where $\pi_1 = \lambda \Delta_{r,n}^{(1)}$, $\pi_2 = \gamma \Delta_{r,n}^{(2)}$ and $\pi_3 = \pi_1(1 + \pi_1) + \pi_2(1 + \pi_1)$.

3.1. Joint df of concomitants of os's based on FGM($\lambda, \gamma: \theta_1, \alpha_1; \theta_2, \alpha_2$)

The joint pdf of concomitants $Y_{[r:n]}$ and $Y_{[s:n]}, r < s$, is (cf. Tahmasebi and Jafari [38], 2015)

$$f_{[r,s:n]}(y_1, y_2) = \int_0^\infty \int_0^{x_2} f_{Y|X}(y_1|x_1)f_{Y|X}(y_2|x_2)f_{r,s:n}(x_1, x_2) dx_1 dx_2,$$

where $\beta(a, b, c) = \frac{\Gamma(a)\Gamma(b)\Gamma(c)}{\Gamma(a+b+c)}$ and

$$f_{r,s:n}(x_1, x_2) = \frac{1}{\beta(r, s - r, n - s + 1)} F_X^{r-1}(x_1) \times (F_X(x_2) - F_X(x_1))^{s-r-1} (1 - F_X(x_2))^{n-s} f_X(x_1) f_X(x_2), \quad x_1 < x_2.$$

Therefore,

$$\begin{aligned} f_{[r,s:n]}(y_1, y_2) &= \int_0^\infty \int_0^{x_2} f_Y(y_1) \left[1 + \lambda(1 - 2F_X(x_1))(1 - 2F_Y(y_1)) \right. \\ &\quad \left. + \gamma F_X(x_1)F_Y(y_1)(2 - 3F_X(x_1))(2 - 3F_Y(y_1)) \right] \left[f_Y(y_2) \left[1 + \lambda(1 - 2F_X(x_2)) \right. \right. \\ (3.5) \quad &\quad \left. \left. \times (1 - 2F_Y(y_2)) + \gamma F_X(x_2)F_Y(y_2)(2 - 3F_X(x_2))(2 - 3F_Y(y_2)) \right] \right] \\ &\quad \times \left[\frac{F_X^{r-1}(x_1)(F_X(x_2) - F_X(x_1))^{s-r-1}(1 - F_X(x_2))^{n-s}}{\beta(r, s - r, n - s + 1)} f_X(x_1) f_X(x_2) \right] dx_1 dx_2. \end{aligned}$$

On the other hand, after some algebra we can write the joint pdf $f_{[r,s:n]}(y_1, y_2)$, defined by (3.5), in the following compact form:

$$\begin{aligned} f_{[r,s:n]}(y_1, y_2) &= f_Y(y_1)f_Y(y_2) \left[1 + \lambda(1 - 2F_Y(y_1))I_1 + \lambda(1 - 2F_Y(y_2))I_2 \right. \\ &\quad + \lambda^2(1 - 2F_Y(y_1))(1 - 2F_Y(y_2))I_3 + \gamma F_Y(y_1)(2 - 3F_Y(y_1))I_4 \\ &\quad + \gamma F_Y(y_2)(2 - 3F_Y(y_2))I_5 + \gamma^2 F_Y(y_1)F_Y(y_2)(2 - 3F_Y(y_1))(2 - 3F_Y(y_2))I_6 \\ &\quad + \lambda\gamma F_Y(y_2)(1 - 2F_Y(y_1))(2 - 3F_Y(y_2))I_7 \\ &\quad \left. + \lambda\gamma F_Y(y_1)(1 - 2F_Y(y_2))(2 - 3F_Y(y_1))I_8 \right], \end{aligned}$$

where $I_1 = \Delta_{r,s,n}^{(1)}, I_2 = \Delta_{r,s,n}^{(2)}, I_3 = \Delta_{r,s,n}^{(1)} + \Delta_{r,s,n}^{(2)} - \Delta_{r,s,n}^{(3)}, I_4 = \Delta_{r,s,n}^{(4)} - \Delta_{r,s,n}^{(1)}, I_5 = \Delta_{r,s,n}^{(5)} - \Delta_{r,s,n}^{(2)}, I_6 = (\Delta_{r,s,n}^{(6)} + \Delta_{r,s,n}^{(7)}) - (\Delta_{r,s,n}^{(3)} - \Delta_{r,s,n}^{(8)}), I_7 = \Delta_{r,s,n}^{(5)} - \Delta_{r,s,n}^{(2)} + \Delta_{r,s,n}^{(3)} - \Delta_{r,s,n}^{(7)}$ and $I_8 = \Delta_{r,s,n}^{(4)} - \Delta_{r,s,n}^{(1)} + \Delta_{r,s,n}^{(3)} - \Delta_{r,s,n}^{(6)}$. Moreover,

$$\begin{aligned} \Delta_{r,s,n}^{(i)} &= \frac{\beta(r, s - r, n - s + 1) - (p_i + 1)\beta(r + p_i, s - r, n - s + 1)}{\beta(r, s - r, n - s + 1)}, \quad i = 1, 4, \\ \Delta_{r,s,n}^{(i)} &= \frac{\beta(r, s - r, n - s + 1) - (p_i + 1)\beta(s + p_i, n - s + 1)\beta(r, s - r)}{\beta(r, s - r, n - s + 1)}, \quad i = 2, 5, \\ \Delta_{r,s,n}^{(i)} &= \frac{\beta(r, s - r, n - s + 1) - (p_i + 1)^2\beta(s + 2p_i, n - s + 1)\beta(r + p_i, s - r)}{\beta(r, s - r, n - s + 1)}, \quad i = 3, 8, \\ \Delta_{r,s,n}^{(i)} &= \frac{\beta(r, s - r, n - s + 1) - 6\beta(s + 3, n - s + 1)\beta(r + p_i, s - r)}{\beta(r, s - r, n - s + 1)}, \quad i = 6, 7, \end{aligned}$$

where $p_1 = p_2 = p_3 = p_7 = 1$ and $p_4 = p_5 = p_6 = p_8 = 2$. Therefore, the product moment $E[Y_{[r:n]}Y_{[s:n]}]$ is obtained directly as

$$(3.6) \quad \begin{aligned} \mu_{[r,s:n]} = \frac{1}{\theta^2} & \left[B^2(\alpha_2)\xi_1(\lambda, r, s, n) - B(\alpha_2)B(2\alpha_2)\xi_2(\gamma, \lambda, r, s, n) \right. \\ & \left. + B^2(2\alpha_2)\xi_3(\gamma, \lambda, r, s, n) - B(\alpha_2)B(2\alpha_2)\xi_4(\gamma, \lambda, r, s, n) + \gamma^2 B^2(3\alpha_2)I_6 \right], \end{aligned}$$

where

$$\begin{aligned} \xi_1(\lambda, r, s, n) &= 1 + \lambda(I_1 + I_2 + I_3), \\ \xi_2(\gamma, \lambda, r, s, n) &= \lambda(I_1 + I_2 + 2\lambda I_3) - \gamma(I_4 + I_5) - \lambda\gamma(I_7 + I_8), \\ \xi_3(\gamma, \lambda, r, s, n) &= \lambda^2 I_3 + \gamma^2 I_6 - \lambda\gamma(I_7 + I_8) \end{aligned}$$

and

$$\xi_4(\gamma, \lambda, r, s, n) = \gamma(I_4 + I_5) + \lambda\gamma(I_7 + I_8).$$

Therefore, by using (3.2) and (3.6) we can after some algebra calculate the covariance between $Y_{[r:n]}$ and $Y_{[s:n]}$ as

$$(3.7) \quad \begin{aligned} \sigma_{[r,s:n]} = \frac{1}{\theta^2} & \left[B^2(\alpha_2)\delta_{r,s,n}^{(1)} - B(\alpha_2)B(2\alpha_2)\delta_{r,s,n}^{(2)} \right. \\ & \left. + B^2(2\alpha_2)\delta_{r,s,n}^{(3)} - B(\alpha_2)B(3\alpha_2)\delta_{r,s,n}^{(4)} + B^2(3\alpha_2)\delta_{r,s,n}^{(5)} \right]. \end{aligned}$$

where

$$\begin{aligned} \delta_{r,s,n}^{(1)} &= 1 + \lambda(I_1 + I_2 + \lambda I_3 - \Delta_{r,n}^{(1)} - \Delta_{s,n}^{(1)}), \\ \delta_{r,s,n}^{(2)} &= \lambda(I_1 + I_2 + 2\lambda I_3 - \Delta_{r,n}^{(1)} - \Delta_{s,n}^{(1)}) - \gamma(I_4 + I_5 - \Delta_{r,n}^{(2)} - \Delta_{s,n}^{(2)}) - \lambda\gamma(I_7 + I_8), \\ \delta_{r,s,n}^{(3)} &= \lambda^2(I_3 + \Delta_{r,n}^{(1)}\Delta_{s,n}^{(1)}) + \gamma^2(I_6 + \Delta_{r,n}^{(2)}\Delta_{s,n}^{(2)}) - \lambda\gamma(I_7 + I_8), \\ \delta_{r,s,n}^{(4)} &= \gamma(I_4 + I_5 - \Delta_{r,n}^{(2)}\Delta_{s,n}^{(2)}) + \lambda\gamma(I_7 + I_8) \end{aligned}$$

and

$$\delta_{r,s,n}^{(5)} = \gamma^2(I_6 + \Delta_{r,n}^{(2)}\Delta_{s,n}^{(2)}).$$

We can now use (3.7) and (3.4) to obtain the coefficient of correlation between $Y_{[r:n]}$ and $Y_{[s:n]}$ as $\rho_{[r,s:n]} = \frac{\sigma_{[r,s:n]}}{\sigma_{[r:n]}\sigma_{[s:n]}}$. By putting $\gamma = 0$ in (3.4) and (3.7), we can easily check that the $\rho_{[r,s:n]}$ is exactly the coefficient of correlation between $Y_{[r:n]}$ and $Y_{[s:n]}$ calculated by Barakat *et al.* [7] (2019), at $p = 1$.

Theorem 3.2. For any $1 \leq r \leq n - 3$, we get

$$(3.8) \quad \mu_{[r+2,s:n]} = 2\mu_{[r+1,s:n]} - \mu_{[r,s:n]} - \tau_n(s; \lambda, \gamma; \alpha_2),$$

where

$$\tau_n(s; \lambda, \gamma; \alpha_2) = \frac{6A_1(n+3)(n+4) + 12A_2(s+2)(n+4) + 18A_3(s+2)(s+3)}{(n+1)(n+2)(n+3)(n+4)}.$$

Moreover, for any $1 \leq s \leq n - 3$, we get

$$(3.9) \quad \mu_{[r,s+2:n]} = 2\mu_{[r,s+1:n]} - \mu_{[r,s:n]} - \omega_n(r; \lambda, \gamma; \alpha_2),$$

where

$$\omega_n(r; \lambda, \gamma; \alpha_2) = \frac{6A_4(n+3)(n+4) + 12rA_5(n+4) + 18A_3r(r+1)}{(n+1)(n+2)(n+3)(n+4)}.$$

Finally, for all $n > 2$, we get

$$(3.10) \quad (n + 1)\mu_{[r,s;n]} = 2n\mu_{[r,s;n-1]} - (n - 1)\mu_{[r,s;n-2]} + \zeta_n(r, s; \lambda, \gamma; \alpha_2),$$

where

$$\begin{aligned} \zeta_n(r, s; \lambda, \gamma; \alpha_2) = & \frac{3A_4s(s + 1)(n + 3)(n + 4) + 36rA_2r(r + 1)(s + 2)(n + 4)}{n(n + 1)(n + 2)(n + 3)(n + 4)} \\ & + \frac{36A_5(s + 1)(s + 2)(n + 4) + 108A_3(s + 2)(s + 3)r(r + 1)}{n(n + 1)(n + 2)(n + 3)(n + 4)} \\ & + \frac{8A_6r(s + 1)(n + 3)(n + 4) + 6A_1r(r + 1)(n + 3)(n + 4)}{n(n + 1)(n + 2)(n + 3)(n + 4)}, \end{aligned}$$

$$A_1 = \frac{1}{\theta_2^2} \left[\lambda\gamma(B^2(2\alpha_2) + B(\alpha_2)B(2\alpha_2) - B(\alpha_2)B(3\alpha_2)) - \gamma B(\alpha_2)B(2\alpha_2) \right],$$

$$A_2 = \frac{1}{\theta_2^2} \left[\gamma^2(B^2(2\alpha_2) + B^2(3\alpha_2)) + \lambda\gamma(B(\alpha_2)B(3\alpha_2) - B(\alpha_2)B(2\alpha_2)) \right],$$

$$A_3 = \frac{-\gamma^2}{\theta_2^2} (B^2(2\alpha_2) + B^2(3\alpha_2)),$$

$$A_4 = \frac{1}{\theta_2^2} \left[\gamma(B(\alpha_2)B(2\alpha_2) - B(\alpha_2)B(3\alpha_2)) + \lambda\gamma(B(\alpha_2)B(2\alpha_2) + B(\alpha_2)B(3\alpha_2) - B^2(\alpha_2)) \right],$$

$$A_5 = \frac{1}{\theta_2^2} \left[\gamma^2(B^2(2\alpha_2) + B^2(3\alpha_2)) + \lambda\gamma(B(\alpha_2)B(3\alpha_2) - B(\alpha_2)B(2\alpha_2) + B^2(\alpha_2)) \right]$$

and

$$A_6 = \frac{1}{\theta_2^2} \left[\lambda^2(B(2\alpha_2) - B(\alpha_2))^2 + 2\lambda\gamma B(\alpha_2)(B(2\alpha_2) - B(3\alpha_2)) - \gamma^2 B^2(3\alpha_2) \right].$$

Proof: It is easy to check that

$$(3.11) \quad \Delta_{r+2,s,n}^{(i)} - \Delta_{r,s,n}^{(i)} = 2(\Delta_{r+1,s,n}^{(i)} - \Delta_{r,s,n}^{(i)}), \quad i = 1, 3, 6,$$

$$(3.12) \quad \Delta_{r+2,s,n}^{(i)} - \Delta_{r,s,n}^{(i)} = (\Delta_{r+1,s,n}^{(i)} - \Delta_{r,s,n}^{(i)}) \frac{2r + 3}{r + 1}, \quad i = 4, 7, 8,$$

and

$$(3.13) \quad \Delta_{r,s,n}^{(2)} = \Delta_{r+1,s,n}^{(2)} = \Delta_{r+2,s,n}^{(2)}, \quad \Delta_{r,s,n}^{(5)} = \Delta_{r+1,s,n}^{(5)} = \Delta_{r+2,s,n}^{(5)}.$$

The recurrence relation (3.8) is now followed by combining (3.11) and (3.12) with (3.13). Now, we turn to prove (3.9). First, we notice that

$$(3.14) \quad \Delta_{r,s,n}^{(1)} = \Delta_{r,s+1,n}^{(1)} = \Delta_{r,s+2,n}^{(1)}$$

and

$$(3.15) \quad \Delta_{r,s,n}^{(4)} = \Delta_{r,s+1,n}^{(4)} = \Delta_{r,s+2,n}^{(4)}.$$

Moreover, it is easy to check that

$$(3.16) \quad \Delta_{r,s+2,n}^{(i)} - \Delta_{r,s,n}^{(i)} = 2(\Delta_{r,s+1,n}^{(i)} - \Delta_{r,s,n}^{(i)}), \quad i = 2, 3, 6,$$

and

$$(3.17) \quad \Delta_{r,s+2,n}^{(i)} - \Delta_{r,s,n}^{(i)} = (\Delta_{r,s+1,n}^{(i)} - \Delta_{r,s,n}^{(i)}) \frac{2s + 2p_i + 1}{s + p_i},$$

where $i = 5, 7, 8$ and $p_5 = 1, p_7 = 2, p_8 = 3$. Therefore, the recurrence relation (3.9) is followed by combining (3.14), (3.15), (3.16) and (3.17). In order to prove the recurrence relation (3.10), we first notice that

$$(3.18) \quad \Delta_{r,s,n-2;p}^{(i)} - \Delta_{r,s,n;p_i}^{(i)} = (\Delta_{r,s,n-1;p_i}^{(i)} - \Delta_{r,s,n;p_i}^{(i)}) \frac{2n + p_i - 1}{n - 1},$$

where $i = 1, 2, \dots, 8$ and $p_1 = p_2 = 1, p_3 = p_4 = p_5 = 2, p_6 = p_7 = 3, p_8 = 4$. The recurrence relation (3.10) is now followed by using (3.18). The proof is completed. \square

Remark 3.2. By putting $\gamma = 0$ in (3.8), (3.9) and (3.10), we get (3.24), (3.25) and (3.26) in Theorem 3.3 of Barakat *et al.* [7] (2019), at $p = 1$.

4. CONCOMITANTS OF RECORD VALUES BASED ON FGM($\lambda, \gamma; \theta_1, \alpha_1; \theta_2, \alpha_2$)

Let $(X_i, Y_i), i = 1, 2, \dots,$ be a random sample from FGM($\lambda, \gamma; \theta_1, \alpha_1; \theta_2, \alpha_2$). When the experimenter interests in studying just the sequence of records of the first component X_i 's, the second component associated with the record value of the first one is termed as the concomitant of that record value. The concomitants of record values arise in a wide variety of practical experiments, e.g., see Bdair and Raqab [8] (2014) and Arnold *et al.* [5] (1998). Let $\{R_n, n \geq 1\}$ be the sequence of record values in the sequence of X 's, while $R_{[n]}$ be the corresponding concomitant. Houchens [24] (1984) has obtained the pdf of concomitant of n -th record value for $n \geq 1$, as $h_{[n]}(y) = \int_0^\infty f_Y(y|x)g_n(x)dx$, where $g_n(x) = \frac{1}{\Gamma(n)}(-\log(1 - F_X(x)))^{n-1}f_X(x)$ is the pdf of R_n . Therefore, after some algebra, we get

$$(4.1) \quad h_{[n]}(y) = (1 + \lambda\Upsilon_{n:1})f_Y(y) + (\gamma\Upsilon_{n:2} - \lambda\Upsilon_{n:1})f_{V_1}(y) - \gamma\Upsilon_{n:2}f_{V_2}(y),$$

where $V_1 \sim \text{GE}(\theta_2; 2\alpha_2), V_2 \sim \text{GE}(\theta_2; 3\alpha_2)$ and

$$\Upsilon_{n:p} = \left[1 - (1 + p) \sum_{i=0}^{n(p)} \frac{(-1)^i \binom{p}{i}}{(i + 1)^n} \right]$$

(clearly, $\Upsilon_{n:1} = (2^{-(n-1)} - 1)$). The representation (4.1) enables us to derive the mean and the variance of $R_{[n]}$ as

$$\mu_{[R_n]} = \frac{1}{\theta_2} \left[B(\alpha_2) - \lambda\Upsilon_{n:1}D(2\alpha_2) - \gamma\Upsilon_{n:2}D(3\alpha_2) \right]$$

and

$$(4.2) \quad \sigma_{[R_n]}^2 = \frac{1}{\theta_2^2} \left[C(\alpha_2) + \lambda\Upsilon_{n:1}(C(\alpha_2) - C(2\alpha_2)) + \gamma\Upsilon_{n:2}(C(2\alpha_2) - C(3\alpha_2)) - (1 + \lambda\Upsilon_{n:1})\lambda\Upsilon_{n:1}D^2(2\alpha_2) - (1 + \gamma\Upsilon_{n:2})\gamma\Upsilon_{n:2}D^2(3\alpha_2) - \lambda\gamma\Upsilon_{n:1}\Upsilon_{n:2}D(2\alpha_2)D(3\alpha_2) \right].$$

Again, by putting $\gamma = 0$, we get the mean and the variance of $R_{[n]}$ for the Huang–Kotz FGM family based on the GE marginals at $p = 1$ (cf. Barakat *et al.* [7], 2019).

The joint pdf of the concomitants $R_{[n]}$ and $R_{[m]}$, $n < m$, is given by

$$h_{[n,m]}(y_1, y_2) = \int_0^\infty \int_{x_1}^\infty f_{Y|X}(y_1|x_1)f_{Y|X}(y_2|x_2)g_{m,n}(x_1, x_2) dx_2dx_1,$$

where

$$g_{m,n}(x) = \frac{1}{\Gamma(n)\Gamma(m-n)} \left(-\log(1-F_X(x_1))\right)^{n-1} \left(-\log \frac{1-F_X(x_2)}{1-F_X(x_1)}\right)^{m-n-1} \frac{f_X(x_1)f_X(x_2)}{1-F_X(x_1)}$$

is the joint pdf of R_n and R_m . Therefore, after some algebra, we get

$$\begin{aligned} h_{[n,m]}(y_1, y_2) = f_Y(y_1)f_Y(y_2) & \left[1 + \lambda(1-2F_Y(y_1))J_1 + \lambda(1-2F_Y(y_2))J_2 \right. \\ & + \lambda^2(1-2F_Y(y_1))(1-2F_Y(y_2))J_3 + \gamma F_Y(y_1)(2-3F_Y(y_1))J_4 \\ (4.3) \quad & + \gamma F_Y(y_2)(2-3F_Y(y_2))J_5 + \gamma^2 F_Y(y_1)F_Y(y_2)(2-3F_Y(y_1))(2-3F_Y(y_2))J_6 \\ & + \lambda\gamma F_Y(y_2)(1-2F_Y(y_1))(2-3F_Y(y_2))J_7 \\ & \left. + \lambda\gamma F_Y(y_1)(1-2F_Y(y_2))(2-3F_Y(y_1))J_8 \right], \end{aligned}$$

where $J_1 = \Upsilon_{n:1}$, $J_2 = \Upsilon_{m:1}$, $J_3 = 4\Upsilon_{n:1} + \Upsilon_{m:1} - \Upsilon_{n,m:1,1}$, $J_4 = \Upsilon_{n:2} - \Upsilon_{n:1}$, $J_5 = \Upsilon_{m:2} - \Upsilon_{m:1}$, $J_6 = \Upsilon_{n,m:2,1} + \Upsilon_{n,m:1,2} - \Upsilon_{n,m:1,1} - \Upsilon_{n,m:2,2}$, $J_7 = \Upsilon_{m:2} + \Upsilon_{n,m:1,1} - \Upsilon_{m:1} - \Upsilon_{n,m:1,2}$, $J_8 = \Upsilon_{n:2} + \Upsilon_{n,m:1,1} - \Upsilon_{n:1} - \Upsilon_{n,m:2,1}$ and

$$\Upsilon_{n,m:p,q} = \left[1 - (1+p)(1+q) \sum_{i=0}^{\aleph(p)} \sum_{j=0}^{\aleph(q)} \frac{(-1)^{i+j} \binom{p}{i} \binom{q}{j}}{(i+j+1)^n (j+1)^{m-n}} \right].$$

The representation (4.3) enables us to derive the product moment and the covariance of $R_{[n]}$ and $R_{[m]}$, respectively, as

$$\begin{aligned} \mu_{[R_n,R_m]:p} = \frac{1}{\theta_2^2} & \left[B^2(\alpha_2)\xi_1(\lambda, n, m) - B(\alpha_2)B(2\alpha_2)\xi_2(\gamma, \lambda, n, m) \right. \\ & \left. + B^2(2\alpha_2)\xi_3(\gamma, \lambda, n, m) - B(\alpha_2)B(2\alpha_2)\xi_4(\gamma, \lambda, n, m) + B^2(3\alpha_2)\gamma^2 J_6 \right], \end{aligned}$$

where $\xi_1(\lambda, n, m) = 1 + \lambda(J_1 + J_2 + J_3)$, $\xi_2(\gamma, \lambda, n, m) = \lambda(J_1 + J_2 + 2\lambda J_3) - \gamma(J_4 + J_5) - \lambda\gamma(J_7 + J_8)$, $\xi_3(\gamma, \lambda, n, m) = \lambda^2 J_3 + \gamma^2 J_6 - \lambda\gamma(J_7 + J_8)$ and $\xi_4(\gamma, \lambda, n, m) = \gamma(J_4 + J_5) + \lambda\gamma(J_7 + J_8)$ and

$$\begin{aligned} \sigma_{[R_n,R_m]} = \frac{1}{\theta_2^2} & \left[B^2(\alpha_2)\eta_{n,m}^{(1)} - B(\alpha_2)B(2\alpha_2)\eta_{n,m}^{(2)} \right. \\ (4.4) \quad & \left. + B^2(2\alpha_2)\eta_{n,m}^{(3)} - B(\alpha_2)B(3\alpha_2)\eta_{n,m}^{(4)} + B^2(3\alpha_2)\eta_{n,m}^{(5)} \right], \end{aligned}$$

where

$$\begin{aligned} \eta_{n,m}^{(1)} &= 1 + \lambda(J_1 + J_2 + \lambda J_3 - \Upsilon_{n:1} - \Upsilon_{m:1}), \\ \eta_{n,m}^{(2)} &= \lambda(J_1 + J_2 + 2\lambda J_3 - \Upsilon_{n:1} - \Upsilon_{m:1}) - \gamma(J_4 + J_5 - \Upsilon_{n:2} - \Upsilon_{m:2}) - \lambda\gamma(J_7 + J_8), \\ \eta_{n,m}^{(3)} &= \lambda^2(J_3 + \Upsilon_{n:1}\Upsilon_{m:1}) + \gamma^2(J_6 + \Upsilon_{n:2}\Upsilon_{m:2}) - \lambda\gamma(J_7 + J_8), \\ \eta_{n,m}^{(4)} &= \gamma(J_4 + J_5 - \Upsilon_{n:2}\Upsilon_{m:2}) + \lambda\gamma(J_7 + J_8) \\ \text{and} \quad \eta_{n,m}^{(5)} &= \gamma^2(J_6 + \Upsilon_{n:2}\Upsilon_{m:2}). \end{aligned}$$

Finally, by combining (4.2) and (4.4), we get the correlation coefficient of the concomitants $R_{[n]}$ and $R_{[m]}$ as

$$\rho_{[R_n, R_m]} = \frac{\sigma_{[R_n, R_m]}}{\sqrt{\sigma_{[R_n]}^2 \sigma_{[R_m]}^2}}.$$

Clearly, by putting $\gamma = 0$ in (4.2) and (4.4), we can easily check that the $\rho_{[R_n, R_m]}$ is exactly the coefficient of correlation between R_n and R_m calculated by Barakat *et al.* [7] (2019), at $p = 1$.

5. APPLICATIONS

Concomitants of os's and record values have received a continued remarkable attention in recent years due to their applicability in many problems. The most striking application of concomitants of os's arises in biological selection problems. For example, in choosing the top k out of n rams as judged by their genetic make up is selected for breeding, then $Y_{[n-k+1:n]}, \dots, Y_{[n:n]}$, might represent the quality of the wool of one of their female offspring. In such type of experiments a geneticist is more likely to choose the best set of offsprings with less number of trials than one in which all trials are undertaken which is much expensive and time consuming. Examples of such application can be found in Scaria and Thomas [34] (2014).

Estimation of the parameters associated with the df of the rv Y of primary interest using concomitants of os's or record values on the auxiliary rv X is an another important application, where extensive works are seen carried out. For example see, Begum and Khan [9] (2000), Scaria [33] (2003), Philip and Thomas [31] (2015), Veena and Thomas [42] (2015) and Domma and Giordano [16] (2016).

Another important application of concomitants of os's and record values is a method of sampling known as ranked set sampling. Namely, when we have an auxiliary rv X , which is easily measurable while the measurement of the rv Y of primary interest is hard and expensive. In order to achieve observational economy, we choose n^2 units randomly from the population and arrange them in n groups of n units each for measurement of the observed rv X . Therefore, based on the observations on X , units in each group are ranked among themselves and from the j -th group the unit ranked j is chosen for measurement of the variable Y of primary interest for $j = 1, 2, \dots, n$. Clearly the observations finally measured on Y are concomitants of os's. For some references in this area one may refer, Chen *et al.* [12] (2004), Chacko and Thomas [10, 11] (2008, 2009), Lesitha and Thomas [28] (2013), Paul and Thomas [30] (2017) and Philip and Thomas [32] (2017).

Moreover, some results on characterization of bivariate distributions by properties of concomitants of os's are available in Thomas and Veena [41] (2011). Besides the preceding applications, there are important other recent applications, For example, Jung *et al.* [27] (2008) presented an application of generalized FGM copula function in exchange markets using directional dependence concept. Hlubinka and Kotz [23] (2010) used the generalized FGM distribution and related copulas as bivariate models for the distribution of spheroidal characteristics. Sheikhi and Tata [35] (2013) modeled the joint distribution of a linear combination of concomitants of os's and linear combinations of their os's as a unified skew-normal

family assuming a multivariate normal distribution. Eryilmaz [17] (2016) has shown that the concomitants are potentially useful in reliability modeling.

Eryilmaz [17] (2016) has analysed the FGM with exponential marginals from a reliability point of view. We extend some of these results to the $FGM(\lambda, \gamma; \theta_1, \alpha_1; \theta_2, \alpha_2)$.

Let $X_i \sim GE(\theta_1; \alpha_1)$ and $Y_i \sim GE(\theta_2; \alpha_2)$ denote respectively the lifetime of the i -th component, and the utility of the i -th component during its lifetime, $i = 1, \dots, n$. Total utility of n components is defined by the rv $\sum_{i=1}^n Y_i$. Moreover, the residual performance after the first failure in the system is given by $\sum_{i=1}^n Y_i - Y_{[1:n]}$. Although the components are identical, they may have different contribution/utility to the performance of the whole system since the components may be located in different positions or they may be used by different operators. The utility of the component is positively correlated with its lifetime. Such a dependence can be modeled by $FGM(\lambda, \gamma; \theta_1, \alpha_1; \theta_2, \alpha_2)$.

The residual performance after time t is defined by the process (cf. Eryilmaz [17], 2016)

$$S(t) = \sum_{i=N(t)+1}^n Y_{[i:n]}, \quad t > 0,$$

where the process $N(t)$ denotes the number of failures up to time t , i.e., $P(N(t) = r) = \binom{n}{r} F_X^r(t) (1 - F_X(t))^{n-r}$, $r = 0, 1, \dots, n$, with $P(N(t) = 0) = 1$. Clearly, knowing the mean value of $S(t)$ may help to an engineer at various stages such as design, and preventive maintenance. By using Proposition 1 of Eryilmaz [17] (2016) and after some algebra, we can show that

$$E(S(t)) = \frac{n}{\theta_2} \left[B(\alpha_2)(1 - F_X(t)) + \lambda D(2\alpha_2)(F_X(t) - F_{U_1}(t)) + \gamma D(3\alpha_2)(F_{U_1}(t) - F_{U_2}(t)) \right],$$

where $U_1 \sim GE(\theta_1; 2\alpha_1)$ and $U_2 \sim GE(\theta_1; 3\alpha_1)$.

On the other hand, it is useful to know about the mean residual performance of the system when at a specific time there are exactly m working components. For this purpose, we consider the conditional mean residual performance defined by $\psi_m(t) = E(S(t) = j | M(t) = n - N(t) = m)$, where $M(t)$ is the number of working components at time t . Now, using Theorem 1 of Eryilmaz [17] (2016), we get after some algebra

$$\begin{aligned} \psi_m(t) &= \frac{m}{n} \frac{E(S(t))}{1 - F_X(t)} \\ &= \frac{m}{\theta_2} \left[B(\alpha_2) + \lambda D(2\alpha_2) \frac{F_X(t) - F_{U_1}(t)}{1 - F_X(t)} + \gamma D(3\alpha_2) \frac{F_{U_1}(t) - F_{U_2}(t)}{1 - F_X(t)} \right]. \end{aligned}$$

By using applying L'Hospital's rule, we get

$$\lim_{t \rightarrow \infty} \psi_m(t) = \frac{m}{\theta_2} \left[B(\alpha_2) + \lambda D(2\alpha_2) + \gamma D(3\alpha_2) \right] = \lim_{t \rightarrow \infty} E(Y | X = t)$$

($E(Y | X = t)$ is given by (2.3)).

Furthermore, we can consider the random time until the total output of the system first falls below the critical level k . Clearly, the waiting time until the total output first falls below k is of special importance in the analysis. The corresponding time is defined

by the rv $T(k) = \inf\{t: S(t) < k\}$. Since this waiting time corresponds to one of the failure time of the components, the two events $\{T(k) = X_{r:n}\}$ and $\{S(X_{r-1:n}) \geq k \text{ and } S(X_{r:n}) < k\}$ are equivalent, where the rv $S(X_{r:n}) = \sum_{i=r+1}^n Y_{[i:n]}$ defines the residual performance after the r -th failure in the system (cf. Eryilmaz [17], 2016). For a system consisting of $n = 3$ components, using Proposition 3 of Eryilmaz [17], 2016), we get

$$P(T(k) = X_{2:3}) = \int_0^\infty P(Y_1^* + Y_2^* \geq k) dF_{X_{1:3}}(x) - \int_0^\infty P(Y_1^* \geq k) dF_{X_{2:3}}(x),$$

$$P(T(k) = X_{3:3}) = \int_0^\infty P(Y_1^* \geq k) dF_{X_{2:3}}(x),$$

and $P(T(k) = X_{1:3}) = 1 - P(T(k) = X_{2:3}) - P(T(k) = X_{3:3})$, where $P(Y^* < y) = F_{Y|X}(y|x)$ is defined by (2.2). Thus,

$$\begin{aligned} P(Y_1^* + Y_2^* > k) &= \int_0^\infty P(Y_1^* + Y_2^* > k | Y_1^* = y) dy \\ &= \int_0^k P(Y_1^* + Y_2^* > k | Y_1^* = y) dy + \int_k^\infty P(Y_1^* + Y_2^* > k | Y_1^* = y) dy \\ &= \int_0^k (1 - P(Y_2^* \leq k - y)) f_{Y_1^*}(y) dy + 1 \\ &\quad - F_Y(k) \left[1 - \lambda F_X(x) \bar{F}_Y(k) - \gamma F_X^2(x) F_Y(k) \bar{F}_Y(k) \right]. \end{aligned}$$

By using the binomial theorem, the above integration can be easily explicitly evaluated. However, Eryilmaz [17] (2016) presented a simple Monte-Carlo simulation algorithm to compute the probability $P(T(k) = X_{r:n})$ for general bivariate df $F_{X,Y}$.

6. CONCLUDING REMARKS

While introducing the iterated FGM distribution by Huang and Kotz [25] (1984), and thereby showed that the maximum correlation is higher than was previously known. Moreover, Huang and Kotz [25] (1984) showed that just one single iteration can result in tripling the covariance for certain marginals. Other than this a systematic study (by Huang and Kotz [25], 1984) of the properties of this promising distribution and its application does not appear to have been discussed in literature. The present paper is an attempt in this direction. Some new distributional properties of concomitants of os's of the iterated FGM based on the GE df were presented in Section 2. Moreover, several new useful recurrence relations between single and product moments of concomitants were established. Finally, by relying of the results of Section 2, we gave an application of this model in reliability theory. Besides this application we reviewed some various applications for concomitants and the FGM distribution. Most probably, the utilization of the iterated FGM distribution instead FGM distribution for studying these applications will give more accurate results.

ACKNOWLEDGMENTS

The authors are grateful to the Editor in Chief, Professor M. Ivette Gomes, and the referees for suggestions and comments that improved the presentation substantially.

REFERENCES

- [1] AHSANULLAH, M. (2009). Records and concomitants, *Bull. Malays. Math.*, **32**(2), 101–117.
- [2] AHSANULLAH, M. and SHAKIL, M. (2013). Characterizations of Rayleigh distribution based on order statistics and record values, *Bull. Malays. Math. Sci. Soc.*, **36**(3), 625–635.
- [3] AHSANULLAH, M.; SHAKIL, M. and KIBRIA, G.M. (2013). A note on a characterization of Gompertz–Verhulst distribution, *J. of Statist. Theory and Appl.*, **13**(1), 17–26.
- [4] AL-HUSSAINI, E.K. and AHSANULLAH, M. (2015). *Exponentiated Distributions*, Atlantis Studies in Probability and Statistics, Atlantis Press, Paris.
- [5] ARNOLD, B.C.; BALAKRISHNAN, N. and NAGARAJA, H.N. (1998). *Records*, Wiley Series in Probability and Statistics, Wiley, New York.
- [6] BARAKAT, H.M.; NIGM, E.M. and SYAM, A.H. (2018). Concomitants of order statistics and record values from Bairamov–Kotz–Bekci FGM bivariate-generalized exponential distribution, *Filomat Journal*, **32**(9), 3313–3324.
- [7] BARAKAT, H.M.; NIGM, E.M. and SYAM, A.H. (2019). Concomitants of ordered variables from Huang–Kotz FGM type bivariate generalized exponential distribution, *Bull. Malays. Math. Sci. Soc.*, **42**, 337–353.
- [8] BDAIR, O.M. and RAQAB, M.Z. (2014). Mean residual life of k -th records under double monitoring, *Bull. Malays. Math. Sci. Soc.*, **37**(2), 457–464.
- [9] BEGUM, A.A. and KHAN, A.H. (2000). Concomitants of order statistics from Marshall and Olkin’s bivariate Weibull distribution, *Calcutta Statist. Assoc. Bul.*, **50**, 65–70.
- [10] CHACKO, M. and THOMAS, P.Y. (2008). Estimation of a parameter of Morgenstern type bivariate exponential distribution by ranked set sampling, *Ann. Ins. Statist. Math.*, **60**, 301–318.
- [11] CHACKO, M. and THOMAS, P.Y. (2009). Estimation of parameters of Morgenstern type bivariate logistic distribution by ranked set sampling, *J. of Indian Soc. of Agricultural Statist.*, **63**, 77–83.
- [12] CHEN, Z.; BAI, Z. and SINHA, B.K. (2004). *Ranked Set Sampling, Theory and Applications*, Lecture Notes in Statistics, Springer, New York.
- [13] DAVID, H.A. (1973). Concomitants of order statistics, *Bull. of Inter. Statist. Inst.*, **45**(2), 295–300.
- [14] DAVID, H.A. and NAGARAJA, H.N. (1998). *Concomitants of order statistics*. In “Order Statistics: Theory & Methods” (N. Balakrishnan and C.R. Rao, Eds.), Elsevier, Amsterdam, pp. 487–513.
- [15] DAVID, H.A. and NAGARAJA, H.N. (2003). *Order Statistics*, John Wiley Sons, Inc.
- [16] DOMMA, F. and GIORDANO, S. (2016). Concomitants of m -generalized order statistics from generalized Farlie–Gumbel–Morgenstern distribution family, *J. of Comp. Appl. Math.*, **294**, 413–435.

- [17] ERYILMAZ, S. (2016). On an application of concomitants of order statistics, *Comm. in Statist. – Theory and Meth.*, **45**(19), 5628–5636.
- [18] FARLIE, D.J.G. (1960). The performance of some correlation coefficients for a general bivariate distribution, *Biometrika*, **47**, 307–323.
- [19] GALAMBOS, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, Krieger, FL (2nd Edn.).
- [20] GUMBEL, E.J. (1960). Bivariate exponential distributions, *J. Amer. Statist. Assoc.*, **55**, 698–707.
- [21] GUPTA, R.D. and KUNDU, D. (1999). Generalized exponential distributions, *Austral. & New Zealand J. Statist.*, **41**, 173–188.
- [22] HANIF, S. (2007). *Concomitants of order random variables*, Ph.D. Thesis, National College of Business Administration & Economics, Lahore.
- [23] HLUBINKA, D. and KOTZ, S. (2010). The generalized FGM distribution and its application to stereology of extremes, *Appl. of Math.*, **55**, 495–512.
- [24] HOUCHEMS, R.L. (1984). *Record Value Theory and Inference*, ProQuest LLC, Ann Arbor.
- [25] HUANG, J.S. and KOTZ, S. (1984). Correlation structure in iterated Farlie–Gumbel–Morgenstern distributions, *Biometrika*, **71**(3), 633–636.
- [26] HUANG, J.S. and KOTZ, S. (1999). Modifications of the Farlie–Gumbel–Morgenstern distributions. A tough hill to climb, *Metrika*, **49**, 135–145.
- [27] JUNG, Y.S.; KIM, J.M. and KIM, J. (2008). New approach of directional dependence in exchange markets using generalized FGM copula function, *Comm. in Statist. – Sim. and Comp.*, **37**, 772–788.
- [28] LESITHA, G. and THOMAS, P.Y. (2013). Estimation of the scale parameter of a log-logistic distribution, *Metrika*, **76**, 427–448.
- [29] MORGENSTERN, D. (1956). Einfache Beispiele zweidimensionaler Verteilungen, *Mitt. Math. Stat*, **8**, 234–235.
- [30] PAUL, J. and THOMAS, P.Y. (2017). Concomitant record ranked set sampling, *Comm. in Statist. – Theory and Meth.*, **46**(19), 9518–9540.
- [31] PHILIP, A. and THOMAS, P.Y. (2015). On concomitants of order statistics arising from the extended Farlie–Gumbel–Morgenstern bivariate logistic distribution and its application in estimation, *Statistical Method.*, **25**, 59–73.
- [32] PHILIP, A. and THOMAS, P.Y. (2017). On concomitants of order statistics and its application in defining ranked set sampling from Farlie–Gumbel–Morgenstern bivariate Lomax distribution, *JIRSS*, **16**(2), 67–95.
- [33] SCARIA, J. (2003). *Concomitants of order statistics from Morgenstern family*, Ph.D. Thesis, Cochin University of Science and Technology, Cochin – 682 022, India.
- [34] SCARIA, J. and THOMAS, B. (2014). Second order concomitants from the Morgenstern family of distributions, *J. App. Statist. Science*, **21**, 63–76.
- [35] SHEIKHI, A. and TATA, M. (2013). The exact joint distribution of concomitants of order statistics and their order statistics under normality, *REVSTAT*, **11**(2), 121–134.
- [36] TAHMASEBI, S. and BEHBOODIAN, J. (2012). Shannon information for concomitants of generalized order statistics in Farlie–Gumbel–Morgenstern (FGM) family, *Bull. Malays. Math. Sci. Soc.*, **35**(4), 975–981.
- [37] TAHMASEBI, S. and JAFARI, A.A. (2014). Estimators for the parameter mean of Morgenstern type bivariate generalized exponential distribution using ranked set sampling, *SORT*, **38**(2), 161–180.

- [38] TAHMASEBI, S. and JAFARI, A.A. (2015). Concomitants of order statistics and record values from Morgenstern type bivariate-generalized exponential distribution, *Bull. Malays. Math. Sci. Soc.*, **38**, 1411–1423.
- [39] TAHMASEBI, S.; JAFARI, A.A. and AFSHARI, M. (2015). Concomitants of dual generalized order statistics from Morgenstern type bivariate generalized exponential distribution, *J. Statist. Theory and Appl.*, **14**(1), 1–12.
- [40] TAHMASEBI, S.; JAFARI, A.A. and AHSANULLAH, M. (2016). Properties on concomitants of generalized order statistics from a bivariate Rayleigh distribution, *Bull. Malays. Math. Sci. Soc.*.
- [41] THOMAS, P.Y. and VEENA, T.G. (2011). On an application of concomitants of order statistics in characterizing a family of bivariate distributions, *Comm. in Statist. – Theory and Meth.*, **40**(1), 1445–1452.
- [42] VEENA, T.G. and THOMAS, P.Y. (2015). Application of concomitants of order statistics of independent non-identically distributed random variables in estimation, *Comm. in Statist. – Theory and Meth.*, **44**, 2530–2545.
- [43] VEENA, T.G. and THOMAS, P.Y. (2017). Role of concomitants of order statistics in determining of parent bivariate distributions, *Comm. in Statist. – Theory and Meth.*, **46**(16), 7976–7997.

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

Editor-in-Chief

Isabel Fraga Alves, University of Lisbon, Portugal

Co-Editor

Giovani L. Silva, University of Lisbon, Portugal

Associate Editors

Marília Antunes, University of Lisbon, Portugal

Barry Arnold, University of California, USA

Narayanaswamy Balakrishnan, McMaster University, Canada

Jan Beirlant, Katholieke Universiteit Leuven, Belgium

Graciela Boente, University of Buenos Aires, Argentina

Paula Brito, University of Porto, Portugal

Vanda Inácio de Carvalho, University of Edinburgh, UK

Valérie Chavez-Demoulin, University of Lausanne, Switzerland

David Conesa, University of Valencia, Spain

Charmaine Dean, University of Waterloo, Canada

Fernanda Figueiredo, University of Porto, Portugal

Jorge Milhazes Freitas, University of Porto, Portugal

Alan Gelfand, Duke University, USA

Stéphane Girard, Inria Grenoble Rhône-Alpes, France

Marie Kratz, ESSEC Business School, France

Victor Leiva, Pontificia Universidad Católica de Valparaíso, Chile

Artur Lemonte, Federal University of Rio Grande do Norte, Brazil

Shuangzhe Liu, University of Canberra, Australia

Maria Nazaré Mendes-Lopes, University of Coimbra, Portugal

Fernando Moura, Federal University of Rio de Janeiro, Brazil

John Nolan, American University, USA

Paulo Eduardo Oliveira, University of Coimbra, Portugal

Pedro Oliveira, University of Porto, Portugal

Carlos Daniel Paulino (2019-2021), University of Lisbon, Portugal

Arthur Pewsey, University of Extremadura, Spain

Gilbert Saporta, Conservatoire National des Arts et Métiers, France

Alexandra M. Schmidt, McGill University, Canada

Julio Singer, University of Sao Paulo, Brazil

Manuel Scotto, University of Lisbon, Portugal

Lisete Sousa, University of Lisbon, Portugal

Milan Stehlík, University of Valparaíso, Chile and LIT-JK University Linz, Austria

María Dolores Ugarte, Public University of Navarre, Spain

Executive Editor

José A. Pinto Martins, Statistics Portugal

Secretariat

José Cordeiro, Statistics Portugal

Olga Bessa Mendes, Statistics Portugal