

- Víctor Leiva and Carolina Marchant
Chilean Journal of Statistics: An open-access, indexed, and free forum for statistical publications from worldwide 123
- Roberto Vila, Helton Saulo, and Jamer Roldan
On some properties of the bimodal normal distribution and its bivariate version 125
- Omar Fetitah, Mohammed K. Attouch, Salah Khardani, and Ali Righi
Nonparametric relative error regression for functional time series data under random censorship 145
- Esra Polat
Robust Hotelling T^2 control chart using adaptive reweighted minimum covariance determinant estimator 171
- Moizés da S. Melo, Laís H. Loose, and Jhonnata B. de Carvalho
Lomax regression model with varying precision: Formulation, estimation, diagnostics, and application 189
- Magaly S. Moraga, Germán Ibacache-Pulgar, and Orietta Nicolis
On an elliptical thin-plate spline partially varying-coefficient model 205
- Bernardo B. de Andrade, Raul Y. Matsushita, Pushpa N. Rathie, Luan Ozelim, and Sandro B. de Oliveira
On a weighted Poisson distribution and its associated regression model 229

CHILEAN JOURNAL OF STATISTICS

Edited by Víctor Leiva and Carolina Marchant

A free open-access journal indexed by



Volume 12 Number 2
December 2021

ISSN: 0718-7912 (print)

ISSN: 0718-7920 (online)

Published by the
Chilean Statistical Society

SOCHÉ 
SOCIEDAD CHILENA DE ESTADÍSTICA

AIMS

The Chilean Journal of Statistics (ChJS) is an official publication of the Chilean Statistical Society (www.soche.cl). The ChJS takes the place of *Revista de la Sociedad Chilena de Estadística*, which was published from 1984 to 2000.

The ChJS covers a broad range of topics in statistics, as well as in artificial intelligence, big data, data science, and machine learning, focused mainly on research articles. However, review, survey, and teaching papers, as well as material for statistical discussion, could be also published exceptionally. Each paper published in the ChJS must consider, in addition to its theoretical and/or methodological novelty, simulations for validating its novel theoretical and/or methodological proposal, as well as an illustration/application with real data.

The ChJS editorial board plans to publish one volume per year, with two issues in each volume. On some occasions, certain events or topics may be published in one or more special issues prepared by a guest editor.

EDITORS-IN-CHIEF

Víctor Leiva *Pontificia Universidad Católica de Valparaíso, Chile*
Carolina Marchant *Universidad Católica del Maule, Chile*

EDITORS

Héctor Allende Cid *Pontificia Universidad Católica de Valparaíso, Chile*
Danilo Alvares *Pontificia Universidad Católica de Chile*
Robert G. Aykkroyd *University of Leeds, UK*
Narayanaswamy Balakrishnan *McMaster University, Canada*
Michelli Barros *Universidade Federal de Campina Grande, Brazil*
Carmen Batanero *Universidad de Granada, Spain*
Marcelo Bourguignon *Universidade Federal do Rio Grande do Norte, Brazil*
Márcia Branco *Universidade de São Paulo, Brazil*
Luis M. Castro *Pontificia Universidad Católica de Chile*
George Christakos *San Diego State University, US*
Enrico Colosimo *Universidade Federal de Minas Gerais, Brazil*
Gauss Cordeiro *Universidade Federal de Pernambuco, Brazil*
Francisco Cribari-Neto *Universidade Federal de Pernambuco, Brazil*
Francisco Cysneiros *Universidade Federal de Pernambuco, Brazil*
Mário de Castro *Universidade de São Paulo, São Carlos, Brazil*
Raul Fierro *Universidad de Valparaíso, Chile*
Jorge Figueroa-Zúñiga *Universidad de Concepción, Chile*
Isabel Fraga *Universidade de Lisboa, Portugal*
Manuel Galea *Pontificia Universidad Católica de Chile*
Diego Gallardo *Universidad de Atacama, Chile*
Christian Genest *McGill University, Canada*
Marc G. Genton *King Abdullah University of Science and Technology, Saudi Arabia*
Viviana Giampaoli *Universidade de São Paulo, Brazil*
Patricia Giménez *Universidad Nacional de Mar del Plata, Argentina*
Hector Gómez *Universidad de Antofagasta, Chile*
Yolanda Gómez *Universidad de Atacama, Chile*
Emilio Gómez-Déniz *Universidad de Las Palmas de Gran Canaria, Spain*
Eduardo Gutiérrez-Peña *Universidad Nacional Autónoma de México*
Nikolai Kolev *Universidade de São Paulo, Brazil*
Eduardo Lalla *University of Twente, Netherlands*
Shuangzhe Liu *University of Canberra, Australia*
Jesús López-Fidalgo *Universidad de Navarra, Spain*
Liliana López-Kleine *Universidad Nacional de Colombia*
Rosângela H. Loschi *Universidade Federal de Minas Gerais, Brazil*
Esam Mahdi *Qatar University, Qatar*
Manuel Mendoza *Instituto Tecnológico Autónomo de México*
Orietta Nicolis *Universidad Andrés Bello, Chile*
Ana B. Nieto *Universidad de Salamanca, Spain*
Teresa Oliveira *Universidade Aberta, Portugal*
Felipe Osorio *Universidad Técnica Federico Santa María, Chile*
Carlos D. Paulino *Instituto Superior Técnico, Portugal*
Fernando Quintana *Pontificia Universidad Católica de Chile*
Nalini Ravishanker *University of Connecticut, US*
Fabrizio Ruggeri *Consiglio Nazionale delle Ricerche, Italy*
José M. Sarabia *Universidad de Cantabria, Spain*
Helton Saulo *Universidade de Brasília, Brazil*
Pranab K. Sen *University of North Carolina at Chapel Hill, US*
Giovani Silva *Universidade de Lisboa, Portugal*
Prayas Sharma *National Rail and Transportation Institute, India*
Julio Singer *Universidade de São Paulo, Brazil*
Milan Stehlik *Johannes Kepler University, Austria*
Alejandra Tapia *Universidad Católica del Maule, Chile*
M. Dolores Ugarte *Universidad Pública de Navarra, Spain*

Chilean Journal of Statistics

VOLUME 12, NUMBER 2

DECEMBER 2021

CONTENTS

Víctor Leiva and Carolina Marchant <i>Chilean Journal of Statistics: An open-access, indexed, and free forum for statistical publications from worldwide</i>	123
Roberto Vila, Helton Saulo, and Jamer Roldan <i>On some properties of the bimodal normal distribution and its bivariate version</i>	125
Omar Fetitah, Mohammed K. Attouch, Salah Khardani, and Ali Righi <i>Nonparametric relative error regression for functional time series data under random censorship</i>	145
Esra Polat <i>Robust Hotelling T^2 control chart using adaptive reweighted minimum covariance determinant estimator</i>	171
Moizés da S. Melo, Laís H. Loose, and Jhonnata B. de Carvalho <i>Lomax regression model with varying precision: Formulation, estimation, diagnostics, and application</i>	189
Magaly S. Moraga, Germán Ibacache-Pulgar, and Orietta Nicolis <i>On an elliptical thin-plate spline partially varying-coefficient model</i>	205
Bernardo B. de Andrade, Raul Y. Matsushita, Pushpa N. Rathie, Luan Ozelim, and Sandro B. de Oliveira <i>On a weighted Poisson distribution and its associated regression model</i>	229

TWELFTH VOLUME – SECOND NUMBER
EDITORIAL PAPER

Chilean Journal of Statistics: An open-access, indexed, and free forum for statistical publications from worldwide

VÍCTOR LEIVA¹ and CAROLINA MARCHANT²

¹School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

²Faculty of Basic Sciences, Universidad Católica del Maule, Talca, Chile

The second issue of the twelfth volume of the Chilean Journal of Statistics (ChJS) was published on 31 December 2021. The ChJS has more than a decade of life in its current version published in English and almost four decades from its origins. We have published four issues in times of trial due to the COVID-19 pandemic, which has been very relevant for statistics because its use has permitted different governments to establish regulations stopping its spread.

The scientific and editorial production of this volume would not have been achieved without the valuable contributions of many people. We are pleased to inform the international community that outstanding researchers have honored us by publishing their interesting work in our journal. We also thank all the anonymous reviewers who have contributed to maintaining ChJS' high-quality standards. Furthermore, we feel obliged and pleased to thank our prestigious editorial board listed in <http://soche.cl/chjs/board.html>. Of course, we must also thank the President and the Board of Directors of the Chilean Statistics Society (listed in <https://soche.cl/quienes-somos>) and the entire Chilean statistical community for placing on us, the Editors-In-Chief of the ChJS, their confidence in our work.

The second issue of the twelfth volume of the ChJS comprises six articles as follows:

- (i) Our first paper is based on some properties of the bimodal normal distribution and its bivariate version, which was authored by Roberto Vila, Helton Saulo, and Jamer Roldan from Brazil and Peru.
- (ii) In the second paper, nonparametric relative error regression for functional time series data under random censorship was proposed by Omar Fetitah, Mohammed K. Attouch, Salah Khardani, and Ali Righi from Algeria and Tunisia.
- (iii) The third paper is authored by Esra Polat from Turkey and it is based on a robust Hotelling control chart using adaptive reweighted minimum covariance determinant estimator.
- (iv) Based on a Lomax regression model with varying precision, Moizés da S. Melo, Laís H. Loose, and Jhonnata B. de Carvalho, from Brazil, formulated the model, estimated its parameters, proposed model diagnostics, and applied their results to real-world data.
- (v) The fifth paper is authored by Magaly S. Moraga, Germán Ibacache-Pulgar, and Orietta Nicolis, from Chile and Italy, which proposes an elliptical thin-plate spline partially varying-coefficient model and its numerical application.
- (vi) In the sixth paper, Bernardo B. de Andrade, Raul Y. Matsushita, Pushpa N. Rathie, Luan Ozelim, and Sandro B. de Oliveira, from Brazil, postulated a weighted Poisson distribution and its associated regression model.

As the Chilean Statistics Society, we are proud because we continue to provide, by means of the ChJS, an open-access forum, publishing high-quality works free of any article processing charges (APC). In addition, we are indexed to the Elsevier Scopus and Clarivate ISI WoS systems. We are very motivated because, during 2021, we received 68 submissions from different countries.

Finally, we would like the international statistical and data-science communities, our editorial board, and our collaborators to champion the ChJS as a twelve-year, international, free of charges, and open-access forum, with fair and high-quality reviews. We encourage the international scientific community to submit their works to the ChJS.

Víctor Leiva and Carolina Marchant

Editors-in-Chief

Chilean Journal of Statistics

<http://soche.cl/chjs>

DISTRIBUTION THEORY
RESEARCH PAPER

On some properties of the bimodal normal distribution and its bivariate version

ROBERTO VILA^{1,*}, HELTON SAULO¹, and JAMER ROLDAN²

¹Department of Statistics, Universidade de Brasília, Brasília, Brazil

²Department of Mathematics, Instituto Federal de Goiás, Goiás, Brazil

(Received: 01 June 2021 · Accepted in final form: 23 December 2021)

Abstract

In this work, we derive some novel properties of the bimodal normal distribution. Some of its mathematical properties are examined. We provide a formal proof for the bimodality, present a stochastic representation, and assess identifiability. We also provide a closed formula for the moments of the bimodal normal distribution. We then discuss the maximum likelihood estimates as well as the existence of these estimates, and also some asymptotic properties of the estimator of the parameter that controls the bimodality. A bivariate version of the bimodal normal distribution is derived and some characteristics such as covariance and correlation are analyzed. We study stationarity and ergodicity and a triangular array central limit theorem. Finally, a Monte Carlo study is carried out for evaluating the performance of the maximum likelihood estimators empirically.

Keywords: Bimodality · Bivariate distribution · Central limit theorem · Ergodicity · Identifiability · Maximum likelihood method · Stationarity.

Mathematics Subject Classification: Primary 60E99 · Secondary 62E99.

1. INTRODUCTION

Bimodal distributions play an important role in the applied statistical literature; see, for example, Eugene et al. (2002), Hassan and El-Bassiouni (2016), and Alizadeh et al. (2017). The use of mixture-free bimodal distributions is very important as often real-world data are better modeled by these models, and in general, mixtures of distributions may suffer from identifiability problems in the parameter estimation; see Vila et al. (2020).

Recently, Gómez-Déniz et al. (2021) introduced a family of continuous distributions appropriate to describe the behavior of bimodal data. This family can accommodate any symmetric distribution and includes the bimodal normal (BN) as a special case. Bivariate distributions are of interest; see, for example, Saulo et al. (2020).

In this work, we derive some novel properties of the BN distribution. Particularly, in Section 2, we describe some preliminary properties, including the behavior of the density and hazard functions, median, moment generating function, mean, variance, among others. In Section 3, we obtain some results on the bimodality property of the BN distribution,

*Corresponding author. Email: rovig161@gmail.com

and the stochastic representation and moments are derived in Section 4. Also, in this section, we study some aspects of identifiability. In Section 5, we discuss maximum likelihood (ML) estimation, existence of the ML estimates, and some asymptotic properties of the ML estimators. A bivariate version of the BN distribution is derived and some characteristics such as covariance and correlation are analyzed in Section 6. In Section 7, the concepts of stationarity and ergodicity of a BN random process are studied. Ergodicity is an important ingredient to study functions of the distributional characteristics of the process when we have one realization of it. We find out that the BN random process is non-stationary. This result allows us to study, in Section 8, the triangular array central limit theorem, which is of vital importance in statistics. All theoretical results in this paper are new in the literature. In Section 9, we carry out Monte Carlo simulations. Finally, in Section 10, we discuss conclusions.

2. PRELIMINARY PROPERTIES

The random variable X follows a BN distribution if its probability density function (PDF) is given by

$$f_{\alpha,\zeta}(x) = \sqrt{2\pi} \operatorname{sech}(\zeta\alpha)\phi(\alpha)\phi(x)\cosh[\alpha(x - \zeta)], \quad x \in \mathbb{R}, \quad (1)$$

where $\zeta \in \mathbb{R}$ and $\alpha \in \mathbb{R}$ are shape and location parameters, respectively, ϕ is the standard normal PDF, and $\operatorname{sech}(z) = 1/\cosh(z)$, with $\cosh(z) = [\exp(z) + \exp(-z)]/2$. The parameter ζ presented in Equation (1) controls the skewness and the parameter α is related to the bimodality; see [Gómez-Déniz et al. \(2021\)](#).

In this work, we derive some novel properties of a special case of Equation (1), more specifically when $\zeta = 0$. Then, we say that a real-valued random variable X has a BN distribution with parameter vector parameter $\boldsymbol{\theta} = (\mu, \sigma, \alpha)^\top$, with $\mu \in \mathbb{R}$, $\sigma > 0$, and $\alpha \in \mathbb{R}$, denoted by $X \sim \text{BN}(\boldsymbol{\theta})$, if its PDF is expressed as

$$f(x; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2 - \frac{\alpha^2}{2}\right] \cosh\left[\alpha\left(\frac{x - \mu}{\sigma}\right)\right], \quad x \in \mathbb{R}, \quad (2)$$

where μ is a location parameter, σ is a scale parameter, and α is a parameter that controls the unimodality or bimodality of the distribution. When α approaches zero (that is, $|\alpha| \leq 1$) the distribution becomes unimodal and when α grows (that is, $|\alpha| > 1$) the bimodality becomes more accentuated. When $\alpha = 0$, we have the known normal distribution. For more details, see [Theorem 3.1](#).

Let $X \sim \text{BN}(\boldsymbol{\theta})$ with PDF $f(x; \boldsymbol{\theta})$ given in Equation (2). Then, the behavior of $f(x; \boldsymbol{\theta})$ with $x \rightarrow 0$ or $x \rightarrow \pm\infty$ is stated as

$$\lim_{x \rightarrow 0} f(x; \boldsymbol{\theta}) = \sqrt{2\pi} \phi_{\mu,\sigma^2}(0)\phi(\alpha) \cosh\left(\frac{\alpha\mu}{\sigma}\right) \quad \text{and} \quad \lim_{x \rightarrow \pm\infty} f(x; \boldsymbol{\theta}) = 0, \quad (3)$$

where $\phi_{\mu,\sigma^2}(x)$ is the PDF of the normal distribution with mean μ and variance σ^2 , and then we denote it by $\phi(x)$ instead of $\phi_{0,1}(x)$.

Observe that the cumulative distribution function (CDF) of $X \sim \text{BN}(\boldsymbol{\theta})$ is given by

$$F(x; \boldsymbol{\theta}) = \frac{1}{4} \left[2 + \operatorname{erf}\left(\frac{x - \mu - \alpha\sigma}{\sigma\sqrt{2}}\right) + \operatorname{erf}\left(\frac{x - \mu + \alpha\sigma}{\sigma\sqrt{2}}\right) \right], \quad (4)$$

where $\operatorname{erf}(x) = 2 \int_0^x \exp(-t^2) dt / \sqrt{\pi}$ is the error function. Note that $\lim_{\alpha \rightarrow 0} F(x; 0, 1, \alpha) = (1/2)[1 + \operatorname{erf}(x/\sqrt{2})] = \Phi(x)$, with Φ being the CDF of the standard normal distribution.

The CDF presented in Equation (4) is a special case of the bimodal skewed symmetric distribution of Hassan and El-Bassiouni (2016).

The hazard function $h(x; \boldsymbol{\theta}) = f(x; \boldsymbol{\theta})/[1 - F(x; \boldsymbol{\theta})]$ has the following behavior when $x \rightarrow 0$ or $x \rightarrow \pm\infty$: $\lim_{x \rightarrow -\infty} h(x; \boldsymbol{\theta}) = 0$, $\lim_{x \rightarrow +\infty} h(x; \boldsymbol{\theta}) = +\infty$ and

$$\lim_{x \rightarrow 0} h(x; \boldsymbol{\theta}) = \frac{4\sqrt{2\pi} \phi_{\mu, \sigma^2}(0) \phi(\alpha) \cosh(\alpha\mu/\sigma)}{2 - \operatorname{erf}\left(\frac{-\mu - \alpha\sigma}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(\frac{-\mu + \alpha\sigma}{\sigma\sqrt{2}}\right)}.$$

From the above limits, it can be concluded that the hazard function is non-decreasing.

A routine calculation shows that, if $X \sim \text{BN}(\boldsymbol{\theta})$, then:

(P.1) (PDF) The random variable $Z = (X - \mu)/\sigma$, where $\mu \in \mathbb{R}$ and $\sigma > 0$, has PDF given by $f(z; 0, 1, \alpha) = (1/\sqrt{2\pi}) \exp[-(z^2 + \alpha^2)/2] \cosh(\alpha z)$, for $z \in \mathbb{R}$, that is, $Z \sim \text{BN}(0, 1, \alpha)$;

(P.2) If f is a Borel measurable function then

$$\mathbb{E}[f((X - \mu)/\sigma)] = \exp(-\alpha^2/2) \mathbb{E}_{\Phi}[f(Z) \cosh(\alpha Z)],$$

where $Z \sim N(0, 1)$ and \mathbb{E}_{Φ} denotes the expectation with respect to distribution function Φ ;

(P.3) (Symmetry) $f(\mu - x; \boldsymbol{\theta}) = f(\mu + x; \boldsymbol{\theta})$ for all real numbers x ;

(P.4) (Median) The median m satisfies that

$$\operatorname{erf}[(m - \mu - \alpha\sigma)/(\sigma\sqrt{2})] = \operatorname{erf}[(-m + \mu - \alpha\sigma)/(\sigma\sqrt{2})], \quad \text{so that } m = \mu;$$

(P.5) (Moment generating function) $M_X(t) = \exp(\mu t + \sigma^2 t^2/2) \cosh(\alpha\sigma t)$, $t \in \mathbb{R}$;

(P.6) (Characteristic function) $\phi_X(t) = \exp(i\mu t - \sigma^2 t^2/2) \cosh(i\alpha\sigma t)$, for $t \in \mathbb{R}$;

(P.7) (Mean) $\mathbb{E}(X) = \mu$;

(P.8) (Variance) $\operatorname{Var}(X) = \sigma^2(1 + \alpha^2)$;

(P.9) (Skewness) $v = 0$, that is, the distribution is approximately symmetrical;

(P.10) (Kurtosis) $\kappa = \alpha^2(\alpha^2 + 6) + 3$;

(P.11) (Mean absolute deviation) $\text{MAD} = [2\phi(\alpha) + \alpha \operatorname{erf}(\alpha/\sqrt{2})]\sigma$;

(P.12) (Shannon entropy) $\text{SE} = \log(\sqrt{2\pi\sigma^2}) + 2\alpha^2 + 1/2 - \exp(-\alpha^2/2)[\exp(2\alpha^2) + 1]/2$.

3. UNIMODALITY AND BIMODALITY OF THE BN DISTRIBUTION

In this section, we provide unimodal and bimodal features of the BN distribution.

THEOREM 3.1 The PDF of the BN distribution given in Equation (2) is unimodal when $|\alpha| \leq 1$ and is bimodal when $|\alpha| > 1$.

PROOF Let us suppose that $\alpha \neq 0$ because for the case $\alpha = 0$ the unimodality is well known. The derivative of $f(x; \boldsymbol{\theta})$ with respect to x is given by

$$f'(x; \boldsymbol{\theta}) = \frac{f(x; \boldsymbol{\theta})}{\sigma} \left\{ \alpha \tanh\left[\alpha \left(\frac{x - \mu}{\sigma}\right)\right] - \left(\frac{x - \mu}{\sigma}\right) \right\}.$$

Then, $f'(x; \boldsymbol{\theta}) = 0$ if and only if

$$\tanh\left[\alpha\left(\frac{x - \mu}{\sigma}\right)\right] = \frac{x - \mu}{\alpha\sigma}. \quad (5)$$

Let $g(x; \boldsymbol{\theta}) = \tanh[\alpha(x - \mu)/\sigma] - (x - \mu)/(\alpha\sigma)$. Note that, for all $\alpha \neq 0$, $x = \mu$ is a root of $g(x; \boldsymbol{\theta})$. We divide the proof in the following two steps:

- (i) First step: proving unimodality. Note that, $g'(x; \boldsymbol{\theta}) = (1/\sigma)\{\alpha \operatorname{sech}^2[\alpha(x - \mu)/\sigma] - 1/\alpha\} < 0$ on $(-\infty, +\infty)$ when $0 < \alpha \leq 1$, and $g'(x; \boldsymbol{\theta}) > 0$ on $(-\infty, +\infty)$ when $-1 \leq \alpha < 0$, because $\operatorname{sech}^2(x) \leq 1$. Since the function $g(x; \boldsymbol{\theta})$ has opposite signs at the extremes of the interval (that is, $\lim_{x \rightarrow -\infty} g(x; \boldsymbol{\theta}) = +\infty$, $\lim_{x \rightarrow +\infty} g(x; \boldsymbol{\theta}) = -\infty$ when $0 < \alpha \leq 1$, and $\lim_{x \rightarrow -\infty} g(x; \boldsymbol{\theta}) = -\infty$, $\lim_{x \rightarrow +\infty} g(x; \boldsymbol{\theta}) = +\infty$ when $-1 \leq \alpha < 0$) and it is monotonic, it will have a single zero at $x = \mu$. Then, since $\lim_{x \rightarrow \pm\infty} f(x; \boldsymbol{\theta}) \stackrel{(3)}{=} 0$, the unimodality of the BN distribution with PDF stated in Equation (2) is guaranteed.
- (ii) Second step: proving bimodality. Without loss of generality, now we assume that $\alpha > 1$ because the other case $\alpha < -1$ is verified using similar arguments. For this case, note that $g(x; \boldsymbol{\theta}) > 0$ when $x \leq \mu - \sigma\alpha$ and $g(x; \boldsymbol{\theta}) < 0$ when $x \geq \mu + \sigma\alpha$. Then, there is no root of $g(x; \boldsymbol{\theta})$ outside of the interval $(\mu - \sigma\alpha, \mu + \sigma\alpha)$. Using Intermediate value theorem, $g(\mu - \sigma\alpha; \boldsymbol{\theta}) = 1 - \tanh(\alpha^2) > 0$, $\varepsilon^- = \lim_{x \rightarrow \mu^-} g(x; \boldsymbol{\theta}) < 0$, and $\varepsilon^+ = \lim_{x \rightarrow \mu^+} g(x; \boldsymbol{\theta}) > 0$, $g(\alpha; \boldsymbol{\theta}) = \tanh(\alpha^2) - 1 < 0$. Thus, there are $c_1 \in (\mu - \sigma\alpha, \varepsilon^-)$ and $c_3 \in (\varepsilon^+, \mu + \sigma\alpha)$: $g(c_i; \boldsymbol{\theta}) = 0$ for $i = 1, 3$. Now, we prove uniqueness of root on $(\mu - \sigma\alpha, \varepsilon^-)$. Indeed, assume that $g(x; \boldsymbol{\theta})$ has two solutions $g(a; \boldsymbol{\theta}) = g(b; \boldsymbol{\theta}) = 0$, $\mu - \sigma\alpha < a < b < \varepsilon^-$, then according to the Rolle theorem there is $c^* \in (a, b)$: $g'(c^*; \boldsymbol{\theta}) = 0$. But $g'(x; \boldsymbol{\theta}) = (1/\sigma)[\alpha \operatorname{sech}^2(\alpha x) - 1/\alpha] < 0$ on $(\mu - \sigma\alpha, \varepsilon^-)$ with $\alpha > 1$, and has no solutions, contradiction. Therefore, $g(x; \boldsymbol{\theta})$ has exactly one real solution on $(\mu - \sigma\alpha, \varepsilon^-)$. Similarly, it is verified that on $(\varepsilon^+, \mu + \sigma\alpha)$, $g(x; \boldsymbol{\theta})$ has exactly one real solution. Thus, for $\alpha > 1$, $g(x; \boldsymbol{\theta})$ has exactly three real roots, denoted by x_1, x_2, x_3 , such that $x_1 < x_2 = \mu < x_3$. Therefore, since $\lim_{x \rightarrow \pm\infty} f(x; \boldsymbol{\theta}) \stackrel{(3)}{=} 0$, the bimodality of the BN distribution expressed in (2) follows. ■

Remark 1 The modes of the BN distribution belong to the interval $(\mu - \sigma\alpha, \mu + \sigma\alpha)$. By symmetry, there is $\delta = \delta(\sigma, \alpha) \in (0, \sigma\alpha)$ so that $x_1 = \mu - \delta$ and $x_3 = \mu + \delta$. Moreover, when $|\alpha| > 1$ and $|x|$ is sufficiently large, the modes of the BN distribution are given by $x_1 \approx \mu - \sigma\alpha$ and $x_3 \approx \mu + \sigma\alpha$, because $\lim_{x \rightarrow \pm\infty} \tanh[\alpha(x - \mu)/\sigma] = \pm 1$.

COROLLARY 3.2 The modal point $x_0 = x_0(\boldsymbol{\theta})$ is a non-decreasing function of μ whenever $|\alpha| \leq 1$.

PROOF By Equation (5), a modal point x_0 of the BN distribution satisfies

$$x_0 = \alpha\sigma \tanh\left[\alpha\left(\frac{x_0 - \mu}{\sigma}\right)\right] + \mu. \quad (6)$$

Differentiating x_0 with respect to μ , we obtain

$$\frac{\partial x_0}{\partial \mu} = 1 - \alpha^2 \operatorname{sech}^2\left[\alpha\left(\frac{x_0 - \mu}{\sigma}\right)\right] \geq 0,$$

whenever $|\alpha| \leq 1$. Hence, x_0 is a non-decreasing function of μ . ■

COROLLARY 3.3 The modal point $x_0 = x_0(\boldsymbol{\theta})$ is a non-decreasing function of σ (respectively of α) whenever $x_0 \geq \mu$ and a non-increasing function of σ (respectively of α) whenever

$x_0 < \mu$.

PROOF Differentiating x_0 in Equation (6) with respect to σ and α , we get

$$\frac{\partial x_0}{\partial \sigma} = \alpha \tanh \left[\alpha \left(\frac{x_0 - \mu}{\sigma} \right) \right] - \alpha^2 \left(\frac{x_0 - \mu}{\sigma} \right) \operatorname{sech}^2 \left[\alpha \left(\frac{x_0 - \mu}{\sigma} \right) \right]$$

and

$$\frac{\partial x_0}{\partial \alpha} = \sigma \left\{ \tanh \left[\alpha \left(\frac{x_0 - \mu}{\sigma} \right) \right] + \alpha \left(\frac{x_0 - \mu}{\sigma} \right) \operatorname{sech}^2 \left[\alpha \left(\frac{x_0 - \mu}{\sigma} \right) \right] \right\}.$$

From the above equations, it follows that $\partial x_0 / \partial \sigma \geq 0$ (respectively $\partial x_0 / \partial \alpha \geq 0$) whenever $x_0 \geq \mu$ and $\partial x_0 / \partial \sigma < 0$ (respectively $\partial x_0 / \partial \alpha < 0$) whenever $x_0 < \mu$. ■

4. STOCHASTIC REPRESENTATION, MOMENTS, AND IDENTIFIABILITY

In this section, we provide the stochastic representation, moments, and identifiability of the BN distribution.

PROPOSITION 4.1 Suppose Z_{μ, σ^2} has a normal distribution with expected value μ and variance σ^2 . Let W have the Bernoulli distribution, so that $W = \alpha\sigma$ or $W = -\alpha\sigma$, each with probability $1/2$, and assume W is independent of Z_{μ, σ^2} . If $X = Z_{\mu, \sigma^2} + W$ then $X \sim \text{BN}(\boldsymbol{\theta})$. Conversely, if $X \sim \text{BN}(\boldsymbol{\theta})$, then $X = Z_{\mu, \sigma^2} + W$.

PROOF By law of total probability and by independence, we get

$$\begin{aligned} \mathbb{P}(X \leq x) &= \mathbb{P}(Z_{\mu, \sigma^2} + \alpha\sigma \leq x) \mathbb{P}(W = \alpha\sigma) + \mathbb{P}(Z_{\mu, \sigma^2} - \alpha\sigma \leq x) \mathbb{P}(W = -\alpha\sigma) \\ &= \mathbb{P}(Z_{\mu, \sigma^2} + \alpha\sigma \leq x) \frac{1}{2} + \mathbb{P}(Z_{\mu, \sigma^2} - \alpha\sigma \leq x) \frac{1}{2} \\ &= \Phi \left(\frac{x - \mu - \alpha\sigma}{\sigma} \right) \frac{1}{2} + \Phi \left(\frac{x - \mu + \alpha\sigma}{\sigma} \right) \frac{1}{2}. \end{aligned}$$

By using the identity $\Phi(x) = (1/2)[1 + \operatorname{erf}(x/\sqrt{2})]$, the above expression is equal to

$$\frac{1}{4} \left[2 + \operatorname{erf} \left(\frac{x - \mu - \alpha\sigma}{\sigma\sqrt{2}} \right) + \operatorname{erf} \left(\frac{x - \mu + \alpha\sigma}{\sigma\sqrt{2}} \right) \right] \stackrel{(4)}{=} F(x; \boldsymbol{\theta}), \quad x \in \mathbb{R}.$$

Then, we have completed the proof. ■

PROPOSITION 4.2 Let $X \sim \text{BN}(\boldsymbol{\theta})$. Then,

$$\mathbb{E}(X^n) = \begin{cases} \sigma^n 2^{\frac{n-2}{2}} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi}} \left[{}_1F_1 \left(-\frac{n}{2}, \frac{1}{2}; -\frac{\{\mu+\alpha\sigma\}^2}{2\sigma^2} \right) + {}_1F_1 \left(-\frac{n}{2}, \frac{1}{2}; -\frac{\{\mu-\alpha\sigma\}^2}{2\sigma^2} \right) \right], & n \text{ even,} \\ \sigma^{n-1} 2^{\frac{n-1}{2}} \frac{\Gamma(\frac{n}{2}+1)}{\sqrt{\pi}} \left[(\mu + \alpha\sigma) {}_1F_1 \left(\frac{1-n}{2}, \frac{3}{2}; -\frac{\{\mu+\alpha\sigma\}^2}{2\sigma^2} \right) + (\mu - \alpha\sigma) {}_1F_1 \left(\frac{1-n}{2}, \frac{3}{2}; -\frac{\{\mu-\alpha\sigma\}^2}{2\sigma^2} \right) \right], & n \text{ odd,} \end{cases}$$

where ${}_1F_1(a, b; x) = [\Gamma(b)/\Gamma(a)] \sum_{k=0}^{\infty} [\Gamma(a+k)/\Gamma(b+k)] (x^k/k!)$ is the Kummer confluent hypergeometric function; see Winkelbauer (2014).

PROOF By Proposition 4.1, we have

$$\mathbb{E}(X^n) = \frac{1}{2} [\mathbb{E}_{\Phi_{\mu+\alpha\sigma,\sigma^2}}(X^n) + \mathbb{E}_{\Phi_{\mu-\alpha\sigma,\sigma^2}}(X^n)],$$

where $\mathbb{E}_{\Phi_{\mu+\alpha\sigma,\sigma^2}}$ denotes the expectation with respect to distribution function $\Phi_{\mu+\alpha\sigma,\sigma^2}$. By combining the above equality with the following known identity (Winkelbauer, 2014), for $Y \sim N(\mu, \sigma^2)$,

$$\mathbb{E}(Y^n) = \begin{cases} \sigma^n 2^{n/2} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi}} {}_1F_1(-\frac{n}{2}, \frac{1}{2}; -\frac{\mu^2}{2\sigma^2}), & n \text{ even,} \\ \mu \sigma^{n-1} 2^{(n+1)/2} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi}} {}_1F_1(\frac{1-n}{2}, \frac{3}{2}; -\frac{\mu^2}{2\sigma^2}), & n \text{ odd,} \end{cases}$$

the proof follows. ■

PROPOSITION 4.3 Let $X \sim \text{BN}(\boldsymbol{\theta})$. Then,

$$\mathbb{E}\left[\left(\frac{X - \mu}{\sqrt{\text{Var}(X)}}\right)^n\right] = \begin{cases} \frac{1}{(1+\alpha^2)^{n/2}} \sum_{\substack{0 \leq k \leq n \\ k \text{ even}}} \binom{n}{k} \alpha^{n-k} 2^{-\frac{k}{2}} \frac{k!}{(k/2)!}, & n \text{ even,} \\ 0, & n \text{ odd.} \end{cases}$$

PROOF By using Proposition 4.1 and that $\text{Var}(X) = \sigma^2(1 + \alpha^2)$, we get

$$\mathbb{E}\left[\left(\frac{X - \mu}{\sqrt{\text{Var}(X)}}\right)^n\right] = \frac{1}{2(1 + \alpha^2)^{n/2}} \left\{ \mathbb{E}_{\Phi_{\mu+\alpha\sigma,\sigma^2}}\left[\left(\frac{X - \mu}{\sigma}\right)^n\right] + \mathbb{E}_{\Phi_{\mu-\alpha\sigma,\sigma^2}}\left[\left(\frac{X - \mu}{\sigma}\right)^n\right] \right\},$$

where $\mathbb{E}_{\Phi_{\mu+\alpha\sigma,\sigma^2}}$ denotes the expectation with respect to distribution function $\Phi_{\mu+\alpha\sigma,\sigma^2}$. Taking the changes of variable $z = (x - \mu)/\sigma$ and $dz = dx/\sigma$, and a binomial expansion, we have

$$\begin{aligned} \mathbb{E}\left[\left(\frac{X - \mu}{\sqrt{\text{Var}(X)}}\right)^n\right] &= \frac{1}{2(1 + \alpha^2)^{n/2}} \{ \mathbb{E}_{\Phi}[(Z + \alpha)^n] + \mathbb{E}_{\Phi}[(Z - \alpha)^n] \} \\ &= \frac{1}{2(1 + \alpha^2)^{n/2}} \sum_{k=0}^n \binom{n}{k} [1 + (-1)^{n-k}] \alpha^{n-k} \mathbb{E}_{\Phi}(Z^k). \end{aligned} \quad (7)$$

A simple observation shows that, when n is even,

$$\frac{1}{2(1 + \alpha^2)^{n/2}} \sum_{k=0}^n \binom{n}{k} [1 + (-1)^{n-k}] \alpha^{n-k} \mathbb{E}_{\Phi}(Z^k) = \frac{1}{(1 + \alpha^2)^{n/2}} \sum_{\substack{0 \leq k \leq n \\ k \text{ even}}} \binom{n}{k} \alpha^{n-k} \mathbb{E}_{\Phi}(Z^k), \quad (8)$$

and, when n is odd,

$$\frac{1}{2(1 + \alpha^2)^{n/2}} \sum_{k=0}^n \binom{n}{k} [1 + (-1)^{n-k}] \alpha^{n-k} \mathbb{E}_{\Phi}(Z^k) = \frac{1}{(1 + \alpha^2)^{n/2}} \sum_{\substack{0 \leq k \leq n \\ k \text{ odd}}} \binom{n}{k} \alpha^{n-k} \mathbb{E}_{\Phi}(Z^k). \quad (9)$$

Thus, by combining the known identities, $\mathbb{E}_{\Phi}(Z^k) = 0$ for k odd, and

$$\mathbb{E}_{\Phi}(Z^k) = 2^{-\frac{k}{2}} \frac{k!}{(k/2)!},$$

for k even, considering Equations (7), (8) and (9), the proof follows. ■

As a consequence of Proposition 4.1, we know that the BN PDF $f(x; \boldsymbol{\theta})$ given in Equation (2), with parameter vector $\boldsymbol{\theta} = (\mu, \sigma, \alpha)^\top$, can be written as a finite mixture of two normal distributions with different location parameters, that is, given by

$$f(x; \boldsymbol{\theta}) = \frac{1}{2} [\phi_{\mu+\alpha\sigma, \sigma^2}(x) + \phi_{\mu-\alpha\sigma, \sigma^2}(x)]. \quad (10)$$

Let \mathcal{N} be the family of normal distributions stated as

$$\mathcal{N} = \left\{ F: F(x; \mu, \sigma) = \int_{-\infty}^x \phi_{\mu, \sigma^2}(y) dy, \mu \in \mathbb{R}, \sigma > 0, x \in \mathbb{R} \right\}.$$

In addition, let $\mathcal{H}_{\mathcal{N}}$ be the class of all finite mixtures of \mathcal{N} . It is well-known that the class $\mathcal{H}_{\mathcal{N}}$ is identifiable; see Teicher (1963). The following result proves the identifiability of the BN distribution.

PROPOSITION 4.4 The mapping $\boldsymbol{\theta} \mapsto f(x; \boldsymbol{\theta})$, for all $x \in \mathbb{R}$, is one-to-one.

PROOF Let us suppose that $f(x; \boldsymbol{\theta}) = f(x; \boldsymbol{\theta}')$ for all $x \in \mathbb{R}$. Thus, by Equation (10), we have that

$$\frac{1}{2} [\phi_{\mu+\alpha\sigma, \sigma^2}(x) + \phi_{\mu-\alpha\sigma, \sigma^2}(x)] = \frac{1}{2} [\phi_{\mu'+\alpha'\sigma', \sigma'^2}(x) + \phi_{\mu'-\alpha'\sigma', \sigma'^2}(x)].$$

Since $\mathcal{H}_{\mathcal{N}}$ is identifiable, we have $\mu \pm \alpha\sigma = \mu' \pm \alpha'\sigma'$ and $\sigma^2 = \sigma'^2$. From where immediately follows that $\mu = \mu'$, $\sigma = \sigma'$ and $\alpha = \alpha'$. Therefore, $\boldsymbol{\theta} = \boldsymbol{\theta}'$, and the identifiability of distribution follows. ■

5. ASYMPTOTIC PROPERTIES

Let X be a random variable with BN distribution that depends on a parameter vector $\boldsymbol{\theta} = (\mu, \sigma, \alpha)^\top$, with $\boldsymbol{\theta}$ being an open subset of \mathbb{R}^3 , where distinct values of $\boldsymbol{\theta}$ yield distinct distributions for X (see Section 4). Let $\mathbf{X} = (X_1, \dots, X_n)^\top$ be a random sample of X . Then, the log-likelihood function for $\boldsymbol{\theta}$ is given by

$$l(\boldsymbol{\theta}; \mathbf{X}) \propto -n \log(\sigma) - \frac{n\alpha^2}{2} - \frac{1}{2} \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 + \sum_{i=1}^n \log \left\{ \cosh \left[\alpha \left(\frac{X_i - \mu}{\sigma} \right) \right] \right\}.$$

A simple computation shows that

$$\frac{\partial l(\boldsymbol{\theta}; \mathbf{X})}{\partial \mu} = \frac{n}{\sigma} \left(\frac{\bar{X} - \mu}{\sigma} \right) - \frac{\alpha}{\sigma} \sum_{i=1}^n \tanh \left[\alpha \left(\frac{X_i - \mu}{\sigma} \right) \right], \quad (11)$$

$$\frac{\partial l(\boldsymbol{\theta}; \mathbf{X})}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 - \frac{\alpha}{\sigma} \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right) \tanh \left[\alpha \left(\frac{X_i - \mu}{\sigma} \right) \right], \quad (12)$$

$$\frac{\partial l(\boldsymbol{\theta}; \mathbf{X})}{\partial \alpha} = -\alpha n + \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right) \tanh \left[\alpha \left(\frac{X_i - \mu}{\sigma} \right) \right]. \quad (13)$$

The log-likelihood equations for the estimators $\hat{\mu}$, $\hat{\sigma}$, $\hat{\alpha}$ are given by

$$\begin{aligned} \hat{\mu} &= \bar{X} - \frac{\hat{\alpha}}{n} \sum_{i=1}^n \tanh \left[\hat{\alpha} \left(\frac{X_i - \hat{\mu}}{\hat{\sigma}} \right) \right], \\ \hat{\sigma}^2 &= \frac{1}{(1 + \hat{\alpha}^2)n} \sum_{i=1}^n (X_i - \hat{\mu})^2, \\ \hat{\alpha} &= \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \hat{\mu}}{\hat{\sigma}} \right) \tanh \left[\hat{\alpha} \left(\frac{X_i - \hat{\mu}}{\hat{\sigma}} \right) \right]. \end{aligned}$$

In the following two propositions, we study the existence of the ML estimator when two parameters are assumed to be known.

PROPOSITION 5.1 If the parameters σ and α are known, then Equation (11) has at least one root on the interval $(-\infty, +\infty)$.

PROOF One can readily verify that $\lim_{\mu \rightarrow \mp\infty} \partial l(\boldsymbol{\theta}; \mathbf{X}) / \partial \mu = \pm\infty$. Hence, by Intermediate value theorem, there exists at least one solution on the interval $(-\infty, +\infty)$. ■

PROPOSITION 5.2 If the parameters μ and σ are known, then Equation (13) has at least one root on the interval $(-\infty, +\infty)$.

PROOF Since $\lim_{\alpha \rightarrow \mp\infty} \partial l(\boldsymbol{\theta}; \mathbf{X}) / \partial \alpha = \pm\infty$, the proof follows the same reasoning as the proofs of Proposition 5.1. ■

Now, we calculate the expectation of the score defined by Equations (11), (12) and (13) when $n = 1$. Indeed, by using the partial derivatives in Equations (11)-(13), with $n = 1$, and the fact that $x \mapsto x \cosh(\alpha x)$ and $x \mapsto \sinh(\alpha x)$ are odd functions, we obtain

$$\begin{aligned} \mathbb{E} \left[\frac{\partial \log \{f(X; \boldsymbol{\theta})\}}{\partial \mu} \right] &= \frac{n}{\sigma} \mathbb{E} \left(\frac{X - \mu}{\sigma} \right) - \frac{\alpha}{\sigma} \mathbb{E} \left\{ \tanh \left[\alpha \left(\frac{X - \mu}{\sigma} \right) \right] \right\} \\ &= \exp \left(-\frac{\alpha^2}{2} \right) \left\{ \frac{n}{\sigma} \mathbb{E}_{\Phi} [Z \cosh(\alpha Z)] - \frac{\alpha}{\sigma} \mathbb{E}_{\Phi} [\sinh(\alpha Z)] \right\} = 0, \end{aligned}$$

where in the second line, the changes of variables $z = (x - \mu)/\sigma$, $dz = dx/\sigma$ were considered.

Analogously, since $\mathbb{E}_{\Phi} [Z^2 \cosh(\alpha Z)] = (\alpha^2 + 1) \exp(\alpha^2/2)$ and $\mathbb{E}_{\Phi} [Z \sinh(\alpha Z)] =$

$\alpha \exp(\alpha^2/2)$, we get

$$\begin{aligned} \mathbb{E}\left[\frac{\partial \log \{f(X; \boldsymbol{\theta})\}}{\partial \sigma}\right] &= -\frac{1}{\sigma} + \frac{1}{\sigma} \mathbb{E}\left[\left(\frac{X - \mu}{\sigma}\right)^2\right] - \frac{\alpha}{\sigma} \mathbb{E}\left\{\left(\frac{X - \mu}{\sigma}\right) \tanh\left[\alpha\left(\frac{X - \mu}{\sigma}\right)\right]\right\} \\ &= -\frac{1}{\sigma} + \frac{1}{\sigma} \exp\left(-\frac{\alpha^2}{2}\right) \mathbb{E}_{\Phi}[Z^2 \cosh(\alpha Z)] - \frac{\alpha}{\sigma} \exp\left(-\frac{\alpha^2}{2}\right) \mathbb{E}_{\Phi}[Z \sinh(\alpha Z)] \\ &= 0 \end{aligned}$$

and let

$$\begin{aligned} \mathbb{E}\left[\frac{\partial \log \{f(X; \boldsymbol{\theta})\}}{\partial \alpha}\right] &= \mathbb{E}\left\{\left(\frac{X - \mu}{\sigma}\right) \tanh\left[\alpha\left(\frac{X - \mu}{\sigma}\right)\right]\right\} - \alpha \\ &= \exp\left(-\frac{\alpha^2}{2}\right) \mathbb{E}_{\Phi}[Z \sinh(\alpha Z)] - \alpha = 0. \end{aligned} \quad (14)$$

In what remains of this section, for the sake of simplicity of presentation, we will assume that μ and σ are known parameters and α is unknown. We are interested in knowing the large sample properties of ML estimator $\hat{\alpha}$ of the parameter α that generates uni- or bimodality in the BN distribution. We emphasize that similar results can be studied for μ and σ when the other parameters are known. Since

$$\frac{\partial^2 f(x; \boldsymbol{\theta})}{\partial \alpha^2} = \left[\alpha^2 + \left(\frac{x - \mu}{\sigma}\right)^2 - 1\right] f(x; \boldsymbol{\theta}) - 2\alpha \left(\frac{x - \mu}{\sigma}\right) \tanh\left[\alpha\left(\frac{x - \mu}{\sigma}\right)\right] f(x; \boldsymbol{\theta}),$$

$\mathbb{E}_{\Phi}[Z^2 \cosh(\alpha Z)] = (\alpha^2 + 1) \exp(\alpha^2/2)$ and $\mathbb{E}_{\Phi}[Z \sinh(\alpha Z)] = \alpha \exp(\alpha^2/2)$, for $Z \sim N(0, 1)$, we have

$$\begin{aligned} \int_{-\infty}^{+\infty} \frac{\partial^2 f(x; \boldsymbol{\theta})}{\partial \alpha^2} dx &= \mathbb{E}\left[\alpha^2 + \left(\frac{X - \mu}{\sigma}\right)^2 - 1\right] - 2\alpha \mathbb{E}\left\{\left(\frac{X - \mu}{\sigma}\right) \tanh\left[\alpha\left(\frac{X - \mu}{\sigma}\right)\right]\right\} \\ &= \alpha^2 + \exp\left(-\frac{\alpha^2}{2}\right) \mathbb{E}_{\Phi}[Z^2 \cosh(\alpha Z)] - 1 - 2\alpha \exp\left(-\frac{\alpha^2}{2}\right) \mathbb{E}_{\Phi}[Z \sinh(\alpha Z)] \\ &= 0, \end{aligned} \quad (15)$$

where in the second line, the changes of variables $z = (x - \mu)/\sigma$ and $dz = dx/\sigma$ were considered. In addition,

$$\frac{\partial^2 \log \{f(x; \boldsymbol{\theta})\}}{\partial \alpha^2} = \left(\frac{x - \mu}{\sigma}\right)^2 \operatorname{sech}^2\left[\alpha\left(\frac{x - \mu}{\sigma}\right)\right] - 1. \quad (16)$$

Then, by Equations (15) and (16), the Fisher information may also be written as

$$\begin{aligned}
\mathcal{I}(\alpha) &= \mathbb{E} \left[\frac{\partial \log \{f(X; \boldsymbol{\theta})\}}{\partial \alpha} \right]^2 = -\mathbb{E} \left[\frac{\partial^2 \log \{f(X; \boldsymbol{\theta})\}}{\partial \alpha^2} \right] + \int_{-\infty}^{+\infty} \frac{\partial^2 f(x; \boldsymbol{\theta})}{\partial \alpha^2} dx \\
&= 1 - \mathbb{E} \left\{ \left(\frac{X - \mu}{\sigma} \right)^2 \operatorname{sech}^2 \left[\alpha \left(\frac{X - \mu}{\sigma} \right) \right] \right\} \\
&= 1 - \exp \left(-\frac{\alpha^2}{2} \right) \mathbb{E}_{\Phi} [Z^2 \operatorname{sech}(\alpha Z)]. \tag{17}
\end{aligned}$$

THEOREM 5.3 Let us suppose that μ and σ are known parameters and α unknown, and $\Theta = \{\alpha \in \mathbb{R}: |\alpha| > 0\}$ be the parameter space. Then, with probability approaching one, as $n \rightarrow +\infty$, the log-likelihood equation $\partial l(\boldsymbol{\theta}; \mathbf{X})/\partial \alpha = 0$ has a consistent solution, denoted by $\hat{\alpha}$.

PROOF Since $\partial \log \{f(x; \boldsymbol{\theta})\}/\partial \alpha$, $\partial^2 \log \{f(x; \boldsymbol{\theta})\}/\partial \alpha^2$, $\partial^3 \log \{f(x; \boldsymbol{\theta})\}/\partial \alpha^3$ exist for all $\alpha \in \Theta$ and every x , by [Cramér \(1946\)](#) it is sufficient to prove that:

- (i) $\mathbb{E}[\partial \log \{f(X; \boldsymbol{\theta})\}/\partial \alpha] = 0$ for all $\alpha \in \Theta$;
- (ii) $-\infty < \mathbb{E}[\partial^2 \log \{f(X; \boldsymbol{\theta})\}/\partial \alpha^2] < 0$ for all $\alpha \in \Theta$;
- (iii) There exists a function $H(x)$ such that for all $\alpha \in \Theta$, $l|\partial^3 \log \{f(x; \boldsymbol{\theta})\}/\partial \alpha^3| < H(x)$ and $\mathbb{E}[H(X)] < \infty$.

In what follows we show the validity of items (i), (ii) and (iii) above. By Equation (14), the statement of item (i) follows. In order to verify item (ii), note that $\exp(-\alpha^2/2)\mathbb{E}_{\Phi}[Z^2 \operatorname{sech}(\alpha Z)] \leq \mathbb{E}_{\Phi}(Z^2) = 1$ for all $\alpha \in \Theta$. Moreover, the two sides are equal if and only if $\alpha = 0$. Since $\alpha \in \Theta$ (that is, $\alpha \neq 0$), it follows that $\exp(-\alpha^2/2)\mathbb{E}_{\Phi}[Z^2 \operatorname{sech}(\alpha Z)] < 1$. Hence,

$$-1 \leq \mathbb{E} \left[\frac{\partial^2 \log \{f(X; \boldsymbol{\theta})\}}{\partial \alpha^2} \right] \stackrel{(17)}{=} \exp \left(-\frac{\alpha^2}{2} \right) \mathbb{E}_{\Phi} [Z^2 \operatorname{sech}(\alpha Z)] - 1 < 0. \tag{18}$$

Then, item (ii) is valid. Thus, since $|\operatorname{sech}^2(\alpha x)| \leq 1$ and $|\tanh(\alpha x)| \leq 1$,

$$\begin{aligned}
\left| \frac{\partial^3 \log \{f(x; \boldsymbol{\theta})\}}{\partial \alpha^3} \right| &= \left| 2 \left(\frac{X - \mu}{\sigma} \right)^3 \operatorname{sech}^2 \left[\alpha \left(\frac{X - \mu}{\sigma} \right) \right] \tanh \left[\alpha \left(\frac{X - \mu}{\sigma} \right) \right] \right| \\
&\leq 2 \left| \left(\frac{x - \mu}{\sigma} \right) \right|^3 = H(x), \tag{19}
\end{aligned}$$

with $\mathbb{E}[H(X)] < \infty$. Therefore, we have completed the proof. ■

The following result supports the intuitive appeal of the ML estimator ([Bahadur, 1971](#)).

PROPOSITION 5.4 Under hypothesis of Theorem 5.3, we have that

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{R}^n} \mathbb{1}_{\{\mathbf{y} \in \mathbb{R}^n : \exp[l(\boldsymbol{\theta}'; \mathbf{y})] > \exp[l(\boldsymbol{\theta}; \mathbf{y})]\}}(\mathbf{x}) \exp[l(\boldsymbol{\theta}'; \mathbf{x})] d\mathbf{x} = 1,$$

for any $\boldsymbol{\theta} = (\mu, \sigma, \alpha)^\top$, $\boldsymbol{\theta}' = (\mu, \sigma, \alpha')^\top \in \Theta$ with $\alpha \neq \alpha'$. Here, $\mathbb{1}_A(x)$ is the indicator function of a set A having the value 1 for all x in A and the value 0 for all x not in A .

PROOF Since $\mathbf{X} = (X_1, \dots, X_n)^\top$ is a random sample of $X \sim \text{BN}(\boldsymbol{\theta})$, X_1, \dots, X_n are independent and identically distributed random variables with PDF $f(x; \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. Therefore, as the BN distribution is identifiable (see Section 4), by Bahadur (1971), the proof follows. ■

Next, we state a central limit theorem for the ML estimator $\hat{\alpha}$, which is important for studying confidence intervals and hypothesis tests, for example. Note that, under hypothesis of Theorem 5.3, the following conditions are satisfied:

- (A.1) The mapping $\alpha \mapsto f(x; \boldsymbol{\theta})$ is three times continuously differentiable on Θ , $\forall x \in \mathbb{R}$;
- (A.2) By Equation (14), $\int_{-\infty}^{+\infty} \partial f(x; \boldsymbol{\theta}) / \partial \alpha \, dx = \mathbb{E}[\partial \log\{f(X; \boldsymbol{\theta})\} / \partial \alpha] = 0$ and, by Equation (15), $\int_{-\infty}^{+\infty} \partial^2 f(x; \boldsymbol{\theta}) / \partial \alpha^2 \, dx = 0$;
- (A.3) By Equations (17) and (18), $0 < \mathcal{I}(\alpha) = 1 - \mathbb{E}[X^2 \text{sech}^2(\alpha X)] \leq 1$, $\forall \alpha \in \Theta$;
- (A.4) By Equation (19), there exists a function $H(x)$ such that for all $\alpha \in \Theta$,

$$\left| \frac{\partial^3 \log\{f(x; \boldsymbol{\theta})\}}{\partial \alpha^3} \right| < H(x), \quad \mathbb{E}[H(X)] < \infty;$$

- (A.5) By Theorem 5.3, the log-likelihood equation $\partial l(\boldsymbol{\theta}; \mathbf{X}) / \partial \alpha = 0$ has a consistent solution $\hat{\alpha}$.

Since conditions (A.1)-(A.5) are satisfied, by Cramér (1946), we have the following result.

THEOREM 5.5 Under hypothesis of Theorem 5.3, it holds that, $\sqrt{n}(\hat{\alpha} - \alpha)$ converges in distribution to $N(0, 1/\mathcal{I}(\alpha))$ as $n \rightarrow +\infty$.

6. THE BIVARIATE BN DISTRIBUTION

We said that a real random vector $\mathbf{X} = (X_1, X_2)^\top$ has bivariate BN (BBN) distribution with parameter vector parameter $\boldsymbol{\psi} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \alpha)^\top$, $\mu_i \in \mathbb{R}$, $\sigma_i > 0$, $\alpha \in \mathbb{R}$, denoted by $\mathbf{X} \sim \text{BBN}(\boldsymbol{\psi})$, if its PDF is given by, for each $\mathbf{x} = (x_1, x_2)^\top \in \mathbb{R}^2$,

$$f(\mathbf{x}; \boldsymbol{\psi}) = \frac{\exp[\alpha^2(\rho^2 - 2)/2]}{\sigma_1 \sigma_2} \phi\left(\frac{x_1 - \mu_1}{\sigma_1}, \frac{x_2 - \mu_2}{\sigma_2}; \rho\right) \cosh\left[\alpha\left(\frac{x_1 - \mu_1}{\sigma_1}\right) + \alpha(1 - \rho)\left(\frac{x_2 - \mu_2}{\sigma_2}\right)\right],$$

where $\rho \in (-1, 1)$ and

$$\phi(\mathbf{z}; \rho) = \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left[-\frac{1}{2(1 - \rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\right], \quad \mathbf{z} = (z_1, z_2)^\top \in \mathbb{R}^2,$$

is the PDF of the standard bivariate normal distribution with correlation coefficient ρ .

A simple algebraic manipulation shows that

$$\int_{-\infty}^{\infty} f(\mathbf{x}; \boldsymbol{\psi}) \, dx_1 = f(x_2; \boldsymbol{\theta}_2) \quad \text{and} \quad \int_{-\infty}^{\infty} f(\mathbf{x}; \boldsymbol{\psi}) \, dx_2 = f(x_1; \boldsymbol{\theta}_1),$$

where $f(x_i; \boldsymbol{\theta}_i)$ is the PDF of the BN distribution stated in Equation (2) with parameter vector $\boldsymbol{\theta}_i = (\mu_i, \sigma_i, \alpha)^\top$, for $i = 1, 2$. Thus, if $\mathbf{X} = (X_1, X_2)^\top \sim \text{BBN}(\boldsymbol{\psi})$ then $X_1 \sim \text{BN}(\boldsymbol{\theta}_1)$ and $X_2 \sim \text{BN}(\boldsymbol{\theta}_2)$.

By using previous results, a laborious algebraic calculation gives

$$\mathbb{E}(X_1|X_2 = x_2) = \mu_1 + \rho\sigma_1\left(\frac{x_2 - \mu_2}{\sigma_2}\right) + (1 - \rho^2)\sigma_1 \tanh\left[\alpha\left(\frac{x_2 - \mu_2}{\sigma_2}\right)\right],$$

that is,

$$\mathbb{E}(X_1|X_2) = \mu_1 + \rho\sigma_1\left(\frac{X_2 - \mu_2}{\sigma_2}\right) + (1 - \rho^2)\sigma_1 \tanh\left[\alpha\left(\frac{X_2 - \mu_2}{\sigma_2}\right)\right] \quad \text{almost sure.}$$

In consequence,

$$\begin{aligned} \mathbb{E}(X_1X_2) &= \mathbb{E}[X_2\mathbb{E}(X_1|X_2)] \\ &= \mu_1\mathbb{E}(X_2) + \rho\sigma_1\mathbb{E}\left[X_2\left(\frac{X_2 - \mu_2}{\sigma_2}\right)\right] + (1 - \rho^2)\sigma_1\mathbb{E}\left\{X_2 \tanh\left[\alpha\left(\frac{X_2 - \mu_2}{\sigma_2}\right)\right]\right\}. \end{aligned}$$

Since $X_2 \sim \text{BN}(\boldsymbol{\theta}_2)$, we get

$$\mathbb{E}(X_1X_2) = \mu_1\mu_2 + \rho\sigma_1\sigma_2(1 + \alpha^2) + (1 - \rho^2)\sigma_1\sigma_2\alpha.$$

Hence, as $\mathbb{E}(X_i) = \mu_i$ and $\text{Var}(X_i) = \sigma_i^2(1 + \alpha^2)$ (see properties P.7 and P.8 in Section 2),

$$\text{Cov}(X_1, X_2) = \sigma_1\sigma_2[\rho(1 + \alpha^2) + (1 - \rho^2)\alpha]; \quad (20)$$

$$\rho(X_1, X_2) = \frac{\rho(1 + \alpha^2) + (1 - \rho^2)\alpha}{(1 + \alpha^2)}.$$

The covariance matrix is given by

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2(1 + \alpha^2) & \sigma_1\sigma_2[\rho(1 + \alpha^2) + (1 - \rho^2)\alpha] \\ \sigma_1\sigma_2[\rho(1 + \alpha^2) + (1 - \rho^2)\alpha] & \sigma_2^2(1 + \alpha^2) \end{bmatrix}.$$

Some immediate observations are the following:

- When $\alpha = 0$, we have the following known facts corresponding to bivariate normal distribution: $\text{Cov}(X_1, X_2) = \rho\sigma_1\sigma_2$ and $\rho(X_1, X_2) = \rho$.
- When $\rho = 0$, we have $\text{Cov}(X_1, X_2) = \sigma_1\sigma_2\alpha$ and $\rho(X_1, X_2) = \alpha/(1 + \alpha^2)$.
- When $\rho = \alpha = 0$, X_1 and X_2 are independent.

7. STATIONARITY AND ERGODICITY

In this section, we provide stationarity and ergodicity properties of the BN distribution.

DEFINITION 7.1 A process X_t is strict-sense stationary (SSS) if its finite-dimensional distributions at times $t_1 < \dots < t_n$, $\forall n \in \mathbb{N}$, are the same after any time interval of length time interval of length t_0 . Thus, for each $n \in \mathbb{N}$ and $t_1 < \dots < t_n$ and $(x_1, \dots, x_n)^\top \in \mathbb{R}^n$ we have

$$\mathbb{P}(X_{t_1+t_0} \leq x_1, \dots, X_{t_n+t_0} \leq x_n) = \mathbb{P}(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n),$$

for any time t_0 .

We say that a process X_t is a BN random process if $X_t \sim \text{BN}(\boldsymbol{\theta}_t)$, where $\boldsymbol{\theta}_t = (\mu_t, \sigma_t, \alpha)^\top$, with $\mu_t \in \mathbb{R}$, $\sigma_t > 0$ and $\alpha \in \mathbb{R}$.

PROPOSITION 7.2 The BN random process is not SSS when μ_t and σ_t are not independent of time.

PROOF If a random process is SSS, then all expected values of functions of the random process must also be stationary. Since $\mathbb{E}(X_t) = \mu_t$ and $\text{Var}(X_t) = \sigma_t^2(1 + \alpha^2)$ (see properties P.7 and P.8 in Section 2) change depend on t , we have that the PDF changes with time. Then, the non-stationarity of random process follows. ■

DEFINITION 7.3 A process X_t is weak-sense stationary (WSS) if:

- $\mathbb{E}(X_t) = \mu$ is independent over time;
- $\mathbb{E}(X_t^2) < \infty$;
- $C_X(t, s) = \text{Cov}(X_t, X_s)$ only depends on the distance between the times considered.

If X_t is a BN random process, it is known that $\mathbb{E}(X_t) = \mu_t$, $\mathbb{E}(X_t^2) = \sigma_t^2(1 + \alpha^2) + \mu_t^2$ (see Section 2) and that $C_X(t, s) \stackrel{(20)}{=} \sigma_t \sigma_s [\rho(1 + \alpha^2) + (1 - \rho^2)\alpha]$. Then, the next result follows.

PROPOSITION 7.4 The BN random process is not WSS when μ_t and σ_t are not independent of time.

Remark 1 In the case that μ_t and σ_t [or $\rho = \alpha = 0$] are independent of time, it is clear that the BN process is SSS and WSS.

In many real-life situations, it is not always possible to have many realizations of the random process available to estimate a population parameter (for example, the mean, variance and covariance function of process), as in classical estimation, but rather a single one. In this case, in order to study the process, we calculate the temporal characteristic of the process.

DEFINITION 7.5 Let X_t be a random process. Then, we define the temporal mean of X_t as

$$\langle m_X \rangle_T = \frac{1}{2T} \int_{-T}^T X_t dt, \quad T > 0.$$

DEFINITION 7.6 A process X_t with mean μ independent of time is mean ergodic if

$$\lim_{T \rightarrow \infty} \text{Var}(\langle m_X \rangle_T) = \lim_{T \rightarrow \infty} \mathbb{E}(\langle m_X \rangle_T - \mu)^2 = 0.$$

PROPOSITION 7.7 The BN random process with mean μ independent of time is mean ergodic whenever

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \sigma_t dt = 0. \quad (21)$$

For example, we can take $\sigma_t = \exp(-t^2)$.

PROOF A simple calculus shows that

$$\text{Var}(\langle m_X \rangle_T) = \frac{1}{4T^2} \int_{-T}^T \int_{-T}^T C_X(t, t') dt' dt.$$

Since $C_X(t, s) \stackrel{(20)}{=} \sigma_t \sigma_{t'} [\rho(1 + \alpha^2) + (1 - \rho^2)\alpha]$, it follows that

$$\text{Var}(\langle m_X \rangle_T) = [\rho(1 + \alpha^2) + (1 - \rho^2)\alpha] \left(\frac{1}{2T} \int_{-T}^T \sigma_t dt \right)^2.$$

Letting $T \rightarrow \infty$ in the above equality, from condition stated in Equation (21), the proof follows. ■

DEFINITION 7.8 A WSS process X_t is covariance-ergodic if

$$\lim_{T \rightarrow \infty} \text{Var} \left[\frac{1}{2T} \int_{-T}^T (X_t - \mu)(X_{t+s} - \mu) dt \right] = 0.$$

When $s = 0$ the WSS process is called variance ergodic.

In general, the BN random process X_t is not a WSS process (see Proposition 7.4). Then, it is clear that X_t is not a covariance ergodic process.

PROPOSITION 7.9 When μ_t is independent of time and $\rho = \alpha = 0$, the BN process is variance ergodic whenever

$$\lim_{T \rightarrow \infty} \frac{1}{4T^2} \int_{-T}^T \int_{-T}^T \text{Cov}(X_t^2, X_{t'}^2) dt' dt = \lim_{T \rightarrow \infty} \frac{1}{4T^2} \int_{-T}^T \int_{-T}^T \text{Cov}(X_t^2, X_{t'}) dt' dt = 0. \quad (22)$$

PROOF When $\rho = \alpha = 0$, $C_X(t, t') = 0$. A simple calculus shows that

$$\text{Var} \left[\frac{1}{2T} \int_{-T}^T (X_t - \mu)^2 dt \right] = \frac{1}{4T^2} \int_{-T}^T \int_{-T}^T \text{Cov}[(X_t - \mu)^2, (X_{t'} - \mu)^2] dt' dt.$$

Since $C_X(t, t') = 0$, the above expression is given by

$$= \frac{1}{4T^2} \int_{-T}^T \int_{-T}^T [\text{Cov}(X_t^2, X_{t'}^2) - 2\mu \text{Cov}(X_t^2, X_{t'}) - 2\mu \text{Cov}(X_t, X_{t'}^2)] dt' dt.$$

By using condition stated in Equation (22), the proof follows. ■

8. A TRIANGULAR ARRAY CENTRAL LIMIT THEOREM

In this section, we provide a triangular array central limit theorem for the BN distribution.

DEFINITION 8.1 Two random variables X and Y are said to be positively quadrant dependent (PQD) if, for all $x, y \in \mathbb{R}$,

$$G(x, y) = \mathbb{P}(X > x, Y > y) - \mathbb{P}(X > x)\mathbb{P}(Y > y) \geq 0.$$

It is usual to rewrite $G(x, y)$ using CDFs as

$$G(x, y) = \mathbb{P}(X \leq x, Y \leq y) - \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y). \quad (23)$$

Remark 1 If F is a CDF, for all $x, y \in \mathbb{R}^2$ and $\alpha \in \mathbb{R}$, then

$$F(\min\{x, y\} - \alpha) + F(\min\{x, y\} + \alpha) \geq \frac{1}{2} [F(x - \alpha) + F(x + \alpha)] [F(y - \alpha) + F(y + \alpha)].$$

Indeed, without loss of generality, assume that $x < y$. Thus,

$$\begin{aligned} F(\min\{x, y\} - \alpha) + F(\min\{x, y\} + \alpha) &= F(x - \alpha) + F(x + \alpha) \\ &\geq \frac{1}{2} [F(x - \alpha) + F(x + \alpha)] [F(y - \alpha) + F(y + \alpha)], \end{aligned}$$

because $0 \leq F(y - \alpha) + F(y + \alpha) \leq 2$.

By the stochastic representation of Proposition 4.1, if $X_j \sim \text{BN}(\boldsymbol{\theta}_j)$, there are $Z_j \sim \text{N}(0, 1)$ and $A_j \sim \text{Bernoulli}(1/2)$, with $A_j \in \{\pm\alpha\}$, so that $X_j = \sigma_j(Z_j + A_j) + \mu_j$. From now on, in this section, we assume that variables Z_j and A_j are independent of j , that is, we have

$$X_j = \sigma_j(Z + A) + \mu_j. \quad (24)$$

PROPOSITION 8.2 The random variables $X \sim \text{BN}(\boldsymbol{\theta}_X)$ and $Y \sim \text{BN}(\boldsymbol{\theta}_Y)$ are PQD, where $\boldsymbol{\theta}_X = (\mu_X, \sigma_X, \alpha)$, $\mu_X \in \mathbb{R}$, $\sigma_X > 0$ and $\alpha \in \mathbb{R}$.

PROOF By Equation (24), $X = \sigma_X(Z + A) + \mu_X$ and $Y = \sigma_Y(Z + A) + \mu_Y$. Then, we get

$$\begin{aligned} \mathbb{P}(X \leq x, Y \leq y) &= \mathbb{P}\left(Z \leq \frac{x - \mu_X}{\sigma_X} - A, Z \leq \frac{y - \mu_Y}{\sigma_Y} - A\right) \\ &= \mathbb{P}\left(Z \leq \min\left\{\frac{x - \mu_X}{\sigma_X}, \frac{y - \mu_Y}{\sigma_Y}\right\} - A\right) = \mathbb{P}(Z \leq \varphi_{n; t_0}(A)). \end{aligned}$$

Let $\widehat{\mathbb{E}}$ be the expectation over A . By the Fubini theorem, we have

$$\begin{aligned} \mathbb{P}(Z \leq \varphi_{n; t_0}(A)) &= \widehat{\mathbb{E}}\left[\Phi\left(\min\left\{\frac{x - \mu_X}{\sigma_X}, \frac{y - \mu_Y}{\sigma_Y}\right\} - A\right)\right] \\ &= \frac{1}{2} \left[\Phi\left(\min\left\{\frac{x - \mu_X}{\sigma_X}, \frac{y - \mu_Y}{\sigma_Y}\right\} - \alpha\right) + \Phi\left(\min\left\{\frac{x - \mu_X}{\sigma_X}, \frac{y - \mu_Y}{\sigma_Y}\right\} + \alpha\right)\right]. \end{aligned}$$

Therefore,

$$\mathbb{P}(X \leq x, Y \leq y) = \frac{1}{2} \left[\Phi\left(\min\left\{\frac{x - \mu_X}{\sigma_X}, \frac{y - \mu_Y}{\sigma_Y}\right\} - \alpha\right) + \Phi\left(\min\left\{\frac{x - \mu_X}{\sigma_X}, \frac{y - \mu_Y}{\sigma_Y}\right\} + \alpha\right)\right].$$

Now, by using the identity $\text{erf}(x/\sqrt{2}) = 2\Phi(x) - 1$, the CDF stated in Equation (4) of $X \sim \text{BN}(\boldsymbol{\theta}_X)$ is written as

$$\mathbb{P}(X \leq x) = \frac{1}{2} \left[\Phi\left(\frac{x - \mu_X}{\sigma_X} - \alpha\right) + \Phi\left(\frac{x - \mu_X}{\sigma_X} + \alpha\right)\right].$$

Hence, by Remark 1, we get

$$\begin{aligned}
G(x, y) &\stackrel{(23)}{=} \mathbb{P}(X \leq x, Y \leq y) - \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y) \\
&= \frac{1}{2} \left[\Phi \left(\min \left\{ \frac{x - \mu_X}{\sigma_X}, \frac{y - \mu_Y}{\sigma_Y} \right\} - \alpha \right) + \Phi \left(\min \left\{ \frac{x - \mu_X}{\sigma_X}, \frac{y - \mu_Y}{\sigma_Y} \right\} + \alpha \right) \right] \\
&\quad - \frac{1}{4} \left[\Phi \left(\frac{x - \mu_X}{\sigma_X} - \alpha \right) + \Phi \left(\frac{x - \mu_X}{\sigma_X} + \alpha \right) \right] \left[\Phi \left(\frac{y - \mu_Y}{\sigma_Y} - \alpha \right) + \Phi \left(\frac{y - \mu_Y}{\sigma_Y} + \alpha \right) \right] \\
&\geq 0.
\end{aligned}$$

This completes the proof. ■

DEFINITION 8.3 We define a sequence of random variables $\{X_j\}$ to be linearly positive quadrant dependent (LPQD) if for any disjoint A, B and positive $\{\lambda_j\}$, $\sum_{k \in A} \lambda_k X_k$ and $\sum_{l \in B} \lambda_l X_l$ are PQD.

A reasoning similar to the proof of Proposition 8.2 gives the following result.

PROPOSITION 8.4 The sequence of random variables $\{X_j\}$, with $X_j \sim \text{BN}(\boldsymbol{\theta}_j)$, is LPQD, where $\boldsymbol{\theta}_j = (\mu_j, \sigma_j, \alpha)$, $\mu_j \in \mathbb{R}$, $\sigma_j > 0$ and $\alpha \in \mathbb{R}$.

THEOREM 8.5 Let $S_n = \sum_{j=1}^{M_n} [X_{n,j} - \mathbb{E}(X_{n,j})]$ where for each n , $X_{n,j} \sim \text{BN}(\boldsymbol{\theta}_{n,j})$, with $\boldsymbol{\theta}_{n,j} = (\mu_{n,j}, \sigma_{n,j}, \alpha)$, $\mu_{n,j} \in \mathbb{R}$, $\sigma_{n,j} > 0$ and $\alpha \in \mathbb{R}$. Suppose there exist $c_1, c_2, c_3 \in (0, \infty)$ and a sequence $u_l \rightarrow 0$ so that, for all n, j, l , we have that

$$\sigma_{n,j}^2 \geq c_1, \quad \sigma_{n,j}^3 \leq c_2; \tag{25}$$

$$\sum_{k=1}^{M_n} \sigma_{n,j} \sigma_{n,k} \leq c_3; \tag{26}$$

$$\sum_{\substack{k=1 \\ |k-j| \geq l}}^{M_n} \sigma_{n,j} \sigma_{n,k} \leq u_l. \tag{27}$$

Then,

$$\lim_{n \rightarrow \infty} \mathbb{P}([\text{Var}(S_n)]^{-1/2} S_n \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-x^2/2) dx, \quad \forall x \in \mathbb{R}.$$

PROOF Since for each n , $\{X_{n,j}\}$ is LPQD (see Proposition 8.4) but not SSS (see Proposition 7.2), by Cox and Grimmett (1984), it is enough to verify that

$$\text{Var}(X_{n,j}) \geq \tilde{c}_1, \quad \mathbb{E}|X_{n,j} - \mathbb{E}(X_{n,j})|^3 \leq \tilde{c}_2; \tag{28}$$

$$\sum_{k=1}^{M_n} \text{Cov}(X_{n,j}, X_{n,k}) \leq \tilde{c}_3; \tag{29}$$

$$\sum_{\substack{k=1 \\ |k-j| \geq l}}^{M_n} \text{Cov}(X_{n,j}, X_{n,k}) \leq \tilde{u}_l; \tag{30}$$

where $\tilde{u}_l \rightarrow 0$. Indeed, since, by Equation (25), $\sigma_{n,j}^2 \geq c_1$ and $\text{Var}(X_{n,j}) = \sigma_{n,j}^2(1 + \alpha^2)$ (see property P.8 in Section 2), we have that $\text{Var}(X_{n,j}) \geq \sigma_{n,j}^2 \geq c_1 = \tilde{c}_1$. Moreover, using the representation in Equation (24) and the condition given in Equation (25), we obtain

$$\mathbb{E}|X_{n,j} - \mathbb{E}(X_{n,j})|^3 = \sigma_{n,j}^3 \mathbb{E}|Z + A|^3 \leq 6\sqrt{2/\pi} \sigma_{n,j}^3 \leq 5c_2 = \tilde{c}_2,$$

that is, Equation (28) is satisfied.

Now, since $\text{Cov}(X_{n,j}, X_{n,k}) \stackrel{(20)}{=} \sigma_{n,j}\sigma_{n,k}[\rho(1 + \alpha^2) + (1 - \rho^2)\alpha]$, by conditions given in Equations (26) and (27), the statements in Equations (29) and (30) follow by taking $\tilde{c}_3 = c_3[\rho(1 + \alpha^2) + (1 - \rho^2)\alpha]$ and $\tilde{u}_l = [\rho(1 + \alpha^2) + (1 - \rho^2)\alpha]u_l$, respectively. ■

Remark 2 The set of $\sigma_{n,k}$ satisfying conditions stated in Equations (25), (26) and (27) is non-empty. Indeed, let us take $M_n = n$ and $\sigma_{n,k} = r^{-k}$, with $k \geq 1$ and $r > 1$, for all n . Immediately, we have $\sigma_{n,k} > 0$ and $\sigma_{n,k} \leq 1$, that is, Equation (25) is valid. Moreover,

$$\sum_{k=1}^n \sigma_{n,j}\sigma_{n,k} \leq \sum_{k=1}^n \sigma_{n,k} \leq \sum_{k=1}^{\infty} \frac{1}{r^k} = \frac{1}{r-1}, \quad r > 1.$$

Then, Equation (26) is satisfied. Thus, since $r^{|k-j|} \leq r^{j+k}$ for $r > 1$, we have $\sigma_{n,j}\sigma_{n,k} = r^{-(j+k)} \leq r^{-|k-j|}$. Therefore,

$$\sum_{\substack{k=1 \\ |k-j| \geq l}}^n \sigma_{n,j}\sigma_{n,k} \leq \sum_{\substack{k=1 \\ |k-j| \geq l}}^n \frac{1}{r^{|k-j|}} \leq \sum_{\substack{k=1 \\ |k-j| \geq l}}^{\infty} \frac{1}{r^{|k-j|}} = \sum_{m=l}^{\infty} \frac{1}{r^m} \left[\sum_{\substack{k=1 \\ |k-j|=m}}^{\infty} 1 \right],$$

where in the last equality we rearrange the summation terms. Since $[\sum_{k:|k-j|=m} 1]$ is the number of vertices at the boundary of the one-dimensional ball of radius m centered at j , there is $C > 0$ independent of j such that $[\sum_{k:|k-j|=m} 1] = C$. Hence,

$$\sum_{\substack{k=1 \\ |k-j| \geq l}}^n \sigma_{n,j}\sigma_{n,k} \leq C \sum_{m=l}^{\infty} \frac{1}{r^m} = u_l.$$

As $\sum_{m=0}^{\infty} r^{-m} = r(r-1)^{-1} < \infty$, for $r > 1$, it follows that $u_l \rightarrow 0$, when $l \rightarrow \infty$. Then, Equation (27) follows.

9. NUMERICAL EVALUATION

In this section, a Monte Carlo simulation study was carried out to evaluate the performance of the ML estimators of the BN model; see Section 5. All numerical evaluations were done in the R software; see R-Team (2020).

The simulation scenario considers sample size $n \in \{10, 75, 250, 600\}$, location parameter $\mu = 0.5$, scale parameter $\sigma = 1.0$, location parameter $\alpha \in \{-2.0, -0.5, 0.8, 3.0\}$, with 1,000 Monte Carlo replications for each combination of above-given parameters and sample size. The values of the location parameter α have been chosen in order to study the performance under unimodality and bimodality.

The ML estimation results for the considered BN model are presented in Figures 1-2. The empirical bias and root mean squared error (RMSE) are reported. A look at the results in Figures 1-2 allows us to conclude that, as the sample size increases, the empirical bias and RMSE both decrease, as expected. Moreover, we note that the performance of the estimate of μ is better when $|\alpha| > 1$ under bimodality.

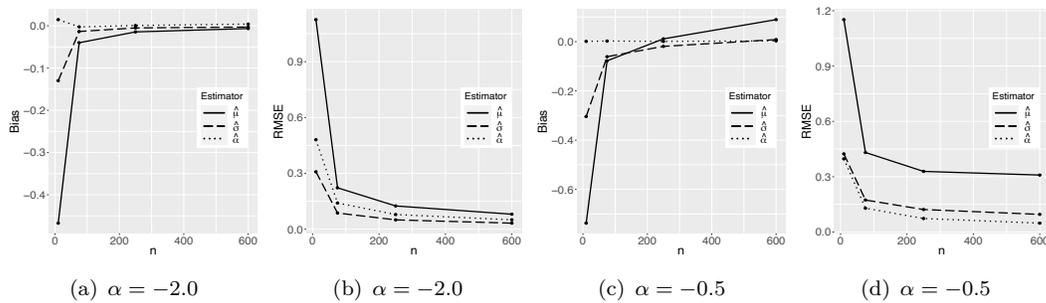


Figure 1. Empirical bias and RMSE from simulated data for the indicated ML estimates of the listed BN model parameters, n and α .

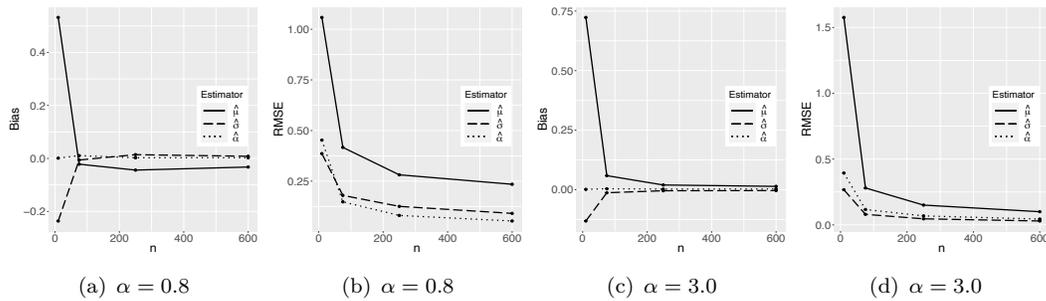


Figure 2. Empirical bias and RMSE from simulated data for the indicated ML estimates of the listed BN model parameters, n and α .

10. CONCLUDING REMARKS

We have stated novel properties of the bimodal normal distribution and discussed some mathematical properties, as well as proven its bimodality and identifiability. We have also analyzed some aspects related to the maximum likelihood estimation and its associated asymptotic properties. We have derived a bivariate version of the bimodal normal distribution and studied some of its characteristics such as covariance and correlation. We have considered stationarity and ergodicity as well as a triangular array central limit theorem. Finally, we have carried out Monte Carlo simulations to evaluate the behavior of the maximum likelihood estimators.

A possible limitation of our proposal might be associated with the moments. In this work, we have only derived the raw moments (moments of positive integer order). It would be interesting, if possible, to find the real moments. In addition, this work has studied consistency and a central limit theorem for one of the model parameters (since the others are known). Note that this parameter generates bimodality. It would be interesting that, using a more elaborate approach, considering an unknown parameter vector. As further research, one might explore the multivariate case and then obtain ergodicity and stationarity results.

AUTHOR CONTRIBUTIONS Conceptualization, R.V., H.S., J.R.; methodology, R.V., H.S., J.R.; software, R.V., H.S., J.R.; validation, R.V., H.S., J.R.; formal analysis, R.V., H.S., J.R.; investigation, R.V., H.S., J.R.; data curation, R.V., H.S., J.R.; writing-original draft preparation, R.V., H.S., J.R.; writing-review and editing, R.V., H.S., J.R.; visualization, R.V., H.S., J.R.; supervision, R.V., H.S., J.R. All authors have read and agreed the published version of the manuscript.

ACKNOWLEDGEMENTS The authors would like to thank the reviewers for helpful comments on our manuscript.

FUNDING The authors thank CNPq for the financial support. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) (Finance Code 001).

CONFLICTS OF INTEREST The authors declare no conflict of interest.

REFERENCES

- Alizadeh, M., MirMostafaei, S.M.T.K, and Ghosh, I., 2017. A new extension of power Lindley distribution for analyzing bimodal data. *Chilean Journal of Statistics*, 8, 67–86.
- Bahadur, R.R., 1971. *Some Limit Theorems in Statistics*. SIAM, Philadelphia.
- Cox, J.T. and Grimmett, G., 1984. Central limit theorems for associated random variables and the percolation model. *The Annals of Probability*, 12, 514–528.
- Cramér, H., 1946. *Mathematical Methods of Statistics*. Princeton University Press, New Jersey.
- Eugene, N., Lee, C., and Famoye, F., 2002. Beta-normal distribution and its applications. *Communications in Statistics: Theory and Methods*, 31, 497–512.
- Gómez-Déniz, E., Sarabia, J.M., and Calderón-Ojeda, E., 2021. Bimodal normal distribution: Extensions and applications. *Journal of Computational and Applied Mathematics*, 388, 113292.
- Hassan, M.Y. and El-Bassiouni, M.Y., 2016. Bimodal skew-symmetric normal distribution. *Communications in Statistics: Theory and Methods*, 45, 1527–1541.
- R-Team, 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Saulo, H., Leao, J., Vila, R., Leiva, V., and Tomazella, V., 2020. On mean-based bivariate Birnbaum-Saunders distributions: properties, inference and application. *Communications in Statistics: Theory and Methods* 49, 6032–6056.
- Teicher, H., 1963. Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 34, 1265–1269.

- Vila, R., Leão, J., Saulo, H., Shahzad, M.N., and Santos-Neto, M., 2020. On a bimodal Birnbaum-Saunders distribution with applications to lifetime data. *Brazilian Journal of Probability and Statistics*, 34, 495–518.
- Winkelbauer, A., 2014. Moments and absolute moments of the normal distribution. Preprint. ArXiv:1209.4340.

FUNCTIONAL AND NONPARAMETRIC STATISTICS
RESEARCH PAPER

Nonparametric relative error regression for functional time series data under random censorship

OMAR FETITAH¹, MOHAMMED K. ATTOUCH^{1,*}, SALAH K HARDANI², and ALI RIGHI¹

¹Sidi Bel Abbes University, Laboratoire de Probabilités Statistique Processus Stochastiques, Sidi Bel Abbes, Algeria

²Faculté des Sciences Tunis El-Manar, Laboratoire Modélisation Mathématique, Analyse Harmonique, Théorie de Potentiel, Tunisia

(Received: 15 March 2021 · Accepted in final form: 30 July 2021)

Abstract

In this paper, we investigate the asymptotic properties of a nonparametric estimator of the relative error regression given a dependent functional explanatory variable, in the case of a scalar censored response. We use the mean squared relative error as a loss function to construct a nonparametric estimator of the regression operator of these functional censored data. We establish the almost surely convergence (with rates) and the asymptotic normality of the proposed estimator. A simulation study and real data application are performed to lend further support to our theoretical results and to compare the quality of predictive performances of the relative error regression estimator than those obtained with standard kernel regression estimates.

Keywords: Almost surely convergence · α -mixing data · Censored data · Functional data analysis · Mean square relative error · Nonparametric estimation.

Mathematics Subject Classification: Primary 62G35 · Secondary 62G20.

1. INTRODUCTION

Functional data analysis is a branch of statistics that has gained popularity in recent years, either mathematically or in terms of applications. There are numerous practical applications for this data format, such as continuous phenomena (climatology, economics, linguistics, medicine, and so on). Since the publication of [Ramsay and Dalzell \(1991\)](#)'s work, numerous developments have been examined in order to produce theories and methodologies that are based on functional data ([Almanjahie et al., 2020](#)).

The monographs of [Ramsay and Silverman \(2005\)](#) provide an overview of both the theoretical and practical elements of functional data analysis, whereas the monographs of [Ferraty and Vieu \(2006\)](#) provide an overview of nonparametric techniques. Numerous nonparametric models have been developed. For example, [Ferraty and Vieu \(2004\)](#) established the strong consistency of the regression function when the explanatory variable is functional and the response is scalar, and their study extended to non-standard regression problems such as time series prediction and curve discrimination ([Ferraty et al., 2002](#); [Ferraty and Vieu, 2003](#)); for robust estimation, see also [Attouch et al. \(2009\)](#).

*Corresponding author. Email: attou_kadi@yahoo.fr.

Masry (2005) establishes the asymptotic normality of the same estimator under an α -mixing assumption. According to Dabo-Niang (2004), density estimation in a Banach space was investigated, as well as the density estimation of a diffusion process with respect to the Wiener measure. Ferraty and Vieu (2006) introduced the kernel type estimation (Azevedo et al., 2011) of some characteristics of the conditional cumulative distribution function (CDF) as well as the successive derivatives of the conditional density; the almost complete convergence (ACC) with rates for the kernel type estimates is established and illustrated by an application to El Niño data. It is common practice to estimate the regression function by minimizing the mean-squared loss function. When data contains outliers, this loss function is predicated on some restrictive constraints, such as the variance of the residual being equal for all observations. As a result, in order to overcome this limitation, we investigate an alternate strategy that allows us to create an effective predictor even when the data is influenced by the existence of outliers. As a result, the constraints of classical regression are addressed in this study by estimating the regression function with respect to the mean squared relative error (MSRE). The latter is a better indicator of a predictor's performance than the usual inaccuracy in the prediction.

The literature on the relative error regression in nonparametric functional data analysis (NFDA) is still limited. The first consistent results were obtained in by Campbell and Donner (1989), where relative regression was used as a classification tool. Jones et al. (2008) studied the nonparametric prediction via relative error regression. They investigated the asymptotic properties of an estimator minimizing the sum of the squared relative errors by considering both (kernel method and local linear approach). Recently, Mechab and Laksaci (2016) analyzed this regression model when the observations are weakly dependent. For spatial data, Attouch et al. (2017) proved the almost complete consistency and the asymptotic normality of this estimator. Fetitah et al. (2020) investigated the relative error in functional regression under random censorship when data are independent.

Nonparametric analysis of incomplete functional data, on the other hand, has a limited extensive literature. There are limited works on this issue (for example, Altandji et al. (2018) estimates the relative error in functional regression under the random left-truncation model). Carbonez et al. (1995) presented the kernel estimator of classical regression in the right censorship model, and improved it in Ould-Saïd and Guessoum (2008). To estimate the conditional quantile when regressors are functional, this approach was later employed by Horrigue and Ould-Saïd (2014). Additionally, using truncated data, Helal and Ould-Saïd (2016) used the same model.

In this paper we define and study a new estimator of the regression function when the interest random variable is subject to random right-censoring and the explanatory variable is functional. Notice that the main feature of our approach is to develop a prediction model alternative to the classical regression which is not sensitive to the presence of the outliers.

The paper is organized as follows. In Section 2 we define our parameter of interest and its corresponding estimators. In Section 3 we give some assumptions and state an almost sure (AS) consistency and asymptotic normality for the proposed estimator. A simulation study and real data application are performed in Section 4, whereas the technical details and the proofs are deferred to Section 4.2.

2. MODEL

2.1 BACKGROUND

Let consider that (Y_i, X_i) , for $i = 1 \dots n$, is a stationary α -mixing couples, where Y_i is real-valued and X_i takes values in some functional space \mathcal{F} . Assume that $\mathbb{E}|Y_i| < \infty$ and

define the regression functional as

$$r(x) = \mathbb{E}[Y_i | X_i = x], \quad x \in \mathcal{F}, \quad \forall i \in \mathbb{N}. \quad (1)$$

The model given in Equation (1) can be written as

$$Y_i = r(X_i) + \epsilon_i, \quad i = 1, \dots, n,$$

where ϵ_i is a random variable such that $\mathbb{E}[\epsilon_i | X_i] = 0$ and $\mathbb{E}[\epsilon_i^2 | X_i] = \sigma^2(X_i) < +\infty$.

Unlike to the multivariate case, there exists various versions of the functional regression estimate. However, all these versions are based on two common procedures. The first one is the functional operator which is supposed smooth enough to be locally well approximated by a polynomial. The second one is the use of the least square error given by

$$r(x) = \arg \min_{r^*} \left(\mathbb{E} \left[(Y - r^*(x))^2 | X = x \right] \right), \quad (2)$$

as a loss function to determine the estimates of r . In complete data, a typical kernel regression estimator based on Equation (2) (Ferraty et al., 2007) is given by

$$r_n(x) = \frac{\sum_{i=1}^n Y_i K(h^{-1}d(x, X_i))}{\sum_{i=1}^n K(h^{-1}d(x, X_i))},$$

where K is a kernel and $(h := h_n)$ is a sequence of bandwidths.

For results on both theoretical and application points of view considering independent or dependent case, we refer the reader to the studies of Attouch et al. (2017) and Chahad et al. (2017). Note that Amiri et al. (2014) analyzed the regression function of a real random variable with functional explanatory variable by using a recursive nonparametric kernel approach.

In the presence of right random censoring, the problem has been analyzed by Buckley and James (1979) using parametric methods. For nonparametric approaches, we refer to Amiri and Khardani (2018) and Stute (1993). Some asymptotic properties were established with a particular application to the conditional mode and quantile by Chaouch and Khardani (2015) and Khardani and Thiam (2016). Horrigue and Ould-Saïd (2014) considered a regression quantile estimation for dependent functional data. Nevertheless, the use of previous loss function given in Equation (2) as a measure of prediction performance may be not suitable in some situation. In particular, the presence of outliers can lead to unreasonable results since all variables have the same weight. Now, to overcome this limitation we propose to estimate the function r by an alternative loss function.

In the relative regression analysis, r is obtained by minimizing the MSRE, that is, $r(x)$ is the solution of the optimization problem:

$$r(x) = \arg \min_{r^*} \left(\mathbb{E} \left[\left(\frac{Y - r^*(X)}{Y} \right)^2 | X = x \right] \right).$$

As mentioned in Jones et al. (2008), where outlier data are present and the response variable of the model is positive, the MSRE is minimized.

It is clear that this criterion is a more meaningful measure of prediction performance than the least squares error, in particular when $Y > 0$, it often is that the ratio of prediction error to the response level, $(Y - r(X))/Y$, is of prime interest: the expected squared relative loss, $\mathbb{E}[\{(Y - r(X))/Y\}^2 | X]$, which is the MSRE, is minimized (specially in the presence of outliers). Moreover, the solution of this problem can be expressed by the ratio of first two

conditional inverse moments of Y given X . As discussed by [Park and Stefanski \(1998\)](#), for $Y > 0$

$$r(x) = \frac{\mathbb{E}[Y^{-1}|X = x]}{\mathbb{E}[Y^{-2}|X = x]} := \frac{g_1(x)}{g_2(x)}, \quad (3)$$

where $g_l(x) = \mathbb{E}[Y^{-l}|X = x]$, for $l = 1, 2$, with r being the best MSRE predictor of Y given $X = x$.

2.2 CONSTRUCTION OF THE ESTIMATOR

To construct our estimator, let us recall that in the case of complete data, a well-known estimator of the regression function is based on the Nadaraya-Watson weights. Let $\{Z_i = (X_i, Y_i)_{1 \leq i \leq n}\}$ be n pairs, identically distributed as $Z = (X, Y)$ and valued in $\mathcal{F} \times \mathbb{R}$, where (\mathcal{F}, d) is a semi-metric space (that is, X is a functional random variable (FRV) and d a semi-metric). Let x be a fixed element of \mathcal{F} . For the complete data, see [Demongeot et al. \(2016\)](#).

It is well known that the kernel estimator of Equation (3) is given by

$$\hat{r}(x) = \frac{\sum_{i=1}^n Y_i^{-1} K\left(\frac{d(x, X_i)}{h}\right)}{\sum_{i=1}^n Y_i^{-2} K\left(\frac{d(x, X_i)}{h}\right)} := \frac{\hat{g}_1(x)}{\hat{g}_2(x)},$$

where $\hat{g}_l(x) = \sum_{i=1}^n Y_i^{-l} K(d(x, X_i)/h) / (n \mathbb{E}(K(d(x, X_1)/h)))$, for $l = 1, 2$, with K is an asymmetrical kernel and $h = h_n$ (depending on n) is a strictly positive real. It is a functional extension of the familiar Nadaraya-Watson estimate. The main change comes from the semi-metric d which measures the proximity between functional objects.

In the censoring case, instead of observing the lifetimes Y , which has a continuous CDF F , we observe the censored lifetimes of items under study, that is, assuming that $(C_i)_{1 \leq i \leq n}$ is a sequence of independent and identically distributed censoring random variable (RV) with common unknown continuous CDF G . Then, in the right censorship model, we only observe the n pairs (T_i, δ_i) with $T_i = Y_i \wedge C_i$ and $\delta_i = \mathbb{1}_{\{Y_i \leq C_i\}}$, for $1 \leq i \leq n$, where $\mathbb{1}_A$ denotes the indicator function of the set A .

In what follows, we define the endpoints of F and G by $\tau_F = \sup\{t: \bar{F}(t) > 0\}$, and $\tau_G = \sup\{t: \bar{G}(t) > 0\}$ where $\bar{F}(x) = 1 - F(x)$ and $\bar{G}(x) = 1 - G(x)$. We assume that $\tau_F < \infty$ and $\bar{G}(\tau_F) > 0$, (this implies $\tau_F < \tau_G$).

In censorship model, only the $(X_i, T_i, \delta_i)_{1 \leq i \leq n}$ are observed. We define $\tilde{r}(x)$ as an estimate of $r(x)$ by

$$\tilde{r}(x) = \frac{\sum_{i=1}^n \frac{\delta_i T_i^{-1}}{\bar{G}(T_i)} K\left(\frac{d(x, X_i)}{h}\right)}{\sum_{i=1}^n \frac{\delta_i T_i^{-2}}{\bar{G}(T_i)} K\left(\frac{d(x, X_i)}{h}\right)} =: \frac{\tilde{g}_1(x)}{\tilde{g}_2(x)}, \quad (4)$$

where

$$\tilde{g}_l(x) = \frac{\sum_{i=1}^n \frac{\delta_i T_i^{-l}}{\bar{G}(T_i)} K\left(\frac{d(x, X_i)}{h}\right)}{n\mathbb{E}\left(K\left(\frac{d(x, X_1)}{h}\right)\right)}, \quad l = 1, 2.$$

In practice, G is unknown. We use the Kaplan-Meier estimator (Deheuvels and Einmahl, 2000) of \bar{G} given by

$$\bar{G}_n(t) = \begin{cases} \prod_{i=1}^n \left(1 - \frac{1-\delta_{(i)}}{n-i+1}\right)^{\mathbb{1}_{\{T_{(i)} \leq t\}}}, & \text{if } t \leq T_{(n)}, \\ 0, & \text{otherwise,} \end{cases}$$

where $T_{(1)} \leq \dots \leq T_{(n)}$ are the order statistics of $(T_i)_{1 \leq i \leq n}$ and $\delta_{(i)}$ is the concomitant of $T_{(i)}$. Therefore, the estimator of r (Fetitah et al., 2020) is stated as

$$\tilde{r}_n(x) = \frac{\sum_{i=1}^n \frac{\delta_i T_i^{-1}}{\bar{G}_n(T_i)} K\left(\frac{d(x, X_i)}{h}\right)}{\sum_{i=1}^n \frac{\delta_i T_i^{-2}}{\bar{G}_n(T_i)} K\left(\frac{d(x, X_i)}{h}\right)} := \frac{\tilde{g}_{1,n}(x)}{\tilde{g}_{2,n}(x)}, \quad (5)$$

where

$$\tilde{g}_{l,n}(x) = \frac{\sum_{i=1}^n \frac{\delta_i T_i^{-l}}{\bar{G}_n(T_i)} K\left(\frac{d(x, X_i)}{h}\right)}{n\mathbb{E}\left(K\left(\frac{d(x, X_1)}{h}\right)\right)}, \quad l = 1, 2.$$

Remark 2.1 In Equations (4) and (5), the sums are taken for the subscripts i , where $\bar{G}_n(T_i) \neq 0$ and $\bar{G}(T_i) \neq 0$. The same convention is followed in the forthcoming formulas. Note that, under the assumptions on the model, the sets $\{i, \bar{G}(T_i) = 0\}$ and $\{i, \bar{G}_n(T_i) = 0\}$ are \mathbb{P} -negligible.

3. ASSUMPTIONS AND MAIN RESULTS

3.1 GENERAL CONTEXT

Throughout this paper, x is a fixed element of the functional space \mathcal{F} . To formulate our assumptions, some notations are required. and we denote by \mathcal{N}_x a neighborhood of the point x . Hereafter, when no confusion is possible, we denote by c and c' some strictly positive generic constants.

Let $B(x, h)$ be the closed ball centered at x with radius h , and consider the CDF of $d(x, X)$ defined by

$$\varphi_x(h) = \mathbb{P}(X \in B(x, h)) = \mathbb{P}(d(x, X) \leq h),$$

with h being positive and satisfies $\varphi_x(0) = 0$ and $\varphi_x(h) \rightarrow 0$ when $h \rightarrow 0$. Let us consider the following definition.

Definition 3.1 Let $(Z_n)_{n \in \mathbb{N}}$ be a sequence of RVs. Given a positive integer n , set

$$\alpha(n) = \sup_k \sup \{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|, A \in \mathcal{F}_1^k(Z) \text{ and } B \in \mathcal{F}_{k+n}^\infty(Z) \},$$

where $\mathcal{F}_i^k(Z)$ denotes the σ -field generated by $\{Z_j, i \leq j \leq k\}$. The sequence is said to be α -mixing if the mixing coefficient $\alpha(n) \rightarrow 0$ when $n \rightarrow \infty$.

3.2 ASYMPTOTIC CONSISTENCY

Our main first result is the pointwise almost sure convergence. In order to state this result, we need some assumptions which are gathered together in order to make our results reading easier. In what follows, we assume that the following assumptions hold:

(H1) $\mathbb{P}(X \in B(x, h)) =: \varphi_x(h) > 0$, for all $h > 0$.

(H2) For all $(x_1, x_2) \in \mathcal{N}_x^2$, we have

$$|g_l(x_1) - g_l(x_2)| \leq cd^{k_l}(x_1, x_2) \text{ for an integer } k_l > 0 \text{ and } l = 1, 2.$$

(H3) The kernel K is a bounded and Lipschitzian function on its support $(0, 1)$ and satisfying:

$$0 < c \leq K(x) \leq c' < +\infty.$$

(H4) The bandwidth h satisfies $h \rightarrow 0$, $\log(n)/(n\varphi_x(h)) \rightarrow 0$ as $n \rightarrow \infty$.

(H5) The inverse moments of the response variable verify:

$$\text{for all } m \geq 2, \quad \mathbb{E}[Y^{-m}|X = x] < c_m < \infty.$$

where c_m is positive constant.

(H6)

(i) $(X_n, Y_n)_{n \geq 1}$ is a sequence of stationary α -mixing RVs with coefficient $\alpha(n) = O(n^{-a})$, for some $a \in (0, \infty)$.

(ii) $(C_n)_{n \geq 1}$ and $(X_n, Y_n)_{n \geq 1}$ are independent.

(H7) For all $i \neq j$, $\mathbb{E}[Y_i^{-1}Y_j^{-2}|(X_i, X_j)] \leq c < \infty$, and

$$0 < \sup_{i \neq j} \left\{ \mathbb{P}((X_i, X_j) \in B(x, h) \times B(x, h)) \right\} = O\left(\frac{(\varphi_x(h))^{(a+1)/a}}{n^{1/a}}\right).$$

(H8) There exists $\eta > 0$, such that, $cn^{\frac{3-a}{a+1}+\eta} \leq \varphi_x(h) \leq c'n^{\frac{1}{1-a}}$, with $a > 2$.

We are in state to give our main result.

Theorem 3.2 Under Assumptions (H1)-(H8), we have

$$|\tilde{r}_n(x) - r(x)| = O(h^{k_1}) + O(h^{k_2}) + O_{AS}\left(\sqrt{\frac{\log(n)}{n\varphi_x(h)}}\right).$$

3.3 ASYMPTOTIC NORMALITY

Here, we study of the asymptotic normality of $\tilde{r}_n(x)$. To do that, we replace assumptions (H1), (H3) and (H4) respectively by the following hypotheses:

(N1) The concentration property (H1) holds. Moreover, there exists a function $\chi_x(\cdot)$ such that,

$$\text{for all } s \in [0, 1], \quad \lim_{r \rightarrow 0} \frac{\varphi_x(sr)}{\varphi_x(r)} = \chi_x(s).$$

(N2) For $\gamma \in \{1, 2\}$, the functions $\Psi_\gamma(x) = \mathbb{E}[g_\gamma(X) - g_\gamma(x)|d(x, X) = x]$ are derivable at zero.

(N3) The kernel function K satisfies (H3) and is a differentiable function on $]0, 1[$ where its first derivative function K' is such that: $-\infty < c < K'(x) < c' < 0$.

(N4) The small ball probability satisfies: $n\varphi_x(h) \rightarrow \infty$.

(N5) For $m \in \{1, 2, 3, 4\}$, the functions $q_m(x) = \mathbb{E}[\bar{G}(Y)^{-1}Y^{-m}|X = x]$ are continuous in a neighborhood of x .

Assumption (H1) is the same as that used by [Ferraty and Vieu \(2006\)](#) which is linked to the functional structure of the functional covariate. Assumptions (H2), (H3) and (H7) deal with the functional aspect of the covariate and the associated small ball probability techniques used in this paper. Assumptions (H6) and (H8) specify the model and the rate of mixing coefficient. Condition (N5) stands as regularity condition that is useful to establish the asymptotic properties of the estimators. Assumptions (H3), (H4), (N3) and (N4) concern the kernel K and the smoothing parameter h and are technical conditions.

The fractal or geometric process is a family of infinite dimensional processes for which the small balls have the property $\varphi_x(t) = \mathbb{P}(\|x - X\| < t) \sim c_x t^\gamma$, where c_x and γ are positive constants. In this case, setting $h_n = An^{-a}$ with $0 < a < 1$ and $0 < A$ implies $\varphi_x(h) = c_x An^{-\gamma a}$. Thus, (H1), (H4) and (H8) hold when $\gamma < 1/a$.

Theorem 3.3 Under Assumptions (H6)-(H8)and (N1)-(N5), we have

$$\left(\frac{n\varphi_x(h)}{\sigma^2(x)}\right)^{1/2} \left(\tilde{r}_n(x) - r(x) - hB_n(x) - o(h)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \quad \text{as } n \rightarrow \infty.$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution,

$$B_n(x) = \frac{(\Psi'_1(0) - r(x)\Psi'_2(0))\beta_0}{\beta_1 g_2(x)}$$

and

$$\sigma^2(x) = \frac{(q_2(x) - 2r(x)q_3(x) + r^2(x)q_4(x))\beta_2}{\beta_1^2} \neq 0,$$

with $\beta_0 = K(1) - \int_0^1 (sK(s))' \chi_x(s) ds$ and $\beta_j = K^j(1) - \int_0^1 (K^j)'(s) \chi_x(s) ds \neq 0$, for $j = 1, 2$.

Remark 3.4 (Comeback to complete data). In absence of censoring ($\bar{G}(x) = 1$), the asymptotic variance becomes

$$\sigma^2(x) = \frac{(a_2(x) - 2r(x)a_3(x) + r^2(x)a_4(x))\beta_2}{\beta_1^2},$$

where $a_j(x) = \mathbb{E}[Y^{-j}|X = x]$, which is the result obtained by [Demongeot et al. \(2016\)](#).

4. SIMULATION AND APPLICATION

4.1 SIMULATION STUDY

In this section, we treat a simulation example to show the behaviour of our estimator $\tilde{r}_n(x)$ and to compare the sensitivity to outliers of the classical regression defined as the conditional expectation $m(x) = \mathbb{E}[Y|X = x]$ estimated by

$$\hat{m}(x) = \frac{\sum_{i=1}^n \frac{\delta_i T_i}{\bar{G}_n(T_i)} K(h^{-1}d(x, X_i))}{\sum_{i=1}^n K(h^{-1}d(x, X_i))},$$

and the relative error estimator $\tilde{r}_n(x)$ previously defined. To do this, we consider the classical nonparametric functional regression model stated as

$$Y = r(X) + \epsilon,$$

where the operator r is defined by $r(X) = 10/[1 + \int_0^1 X^2(t)dt]$. We consider two diffusion processes on the interval $[0, 1]$, $Z_1(t) = 2 - \cos(\pi tW)$ and $Z_2(t) = \cos(\pi tW)$, and we take $X(t) = AZ_1(t) + (1 - A)Z_2(t)$, where A is a Bernoulli distributed RV and W is an α -mixing process generated by the model expressed as

$$W_i = \frac{1}{\sqrt{2}}(W_{i-1} + \eta_i), \quad i = 1, \dots, 200,$$

with η_i being centered Gaussian distributed RVs with variance 0.5 and independent of η_i . We carried out the simulation with $n = 200$ sample of the curve X . The error variable $\epsilon_i \sim \mathcal{N}(0, 0.5)$. We also, simulate n independent and identically distributed RV C_i , for $i = 1, \dots, n$, with law $\mathcal{E}(\lambda)$ (that is, exponentially distributed with density $\lambda e^{-\lambda x} \mathbb{1}_{\{x \geq 0\}}$). Simulated data from our model are plotted in Figure 1. To compute our estimator based on the observed data (X_i, T_i, δ_i) , for $i = 1, \dots, n$, where $T_i = Y_i \wedge C_i$ and $\delta_i = \mathbb{1}_{\{Y_i \leq C_i\}}$.

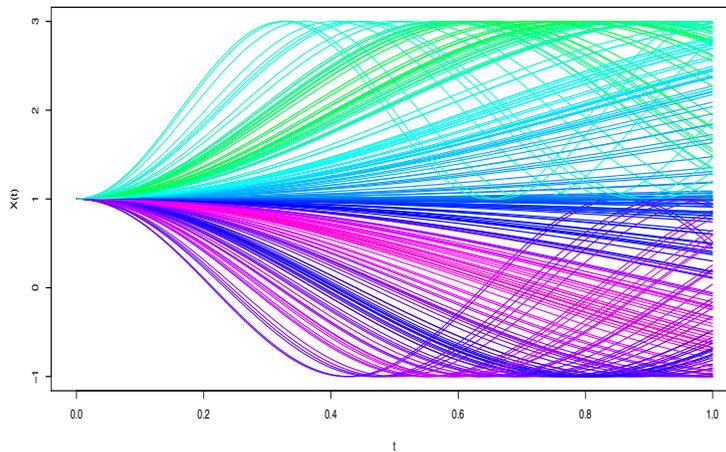


Figure 1. The curves $X_{i=1, \dots, 100}(t)$, for $t \in [0, 1]$.

We choose the quadratic kernel defined by

$$K(x) = \frac{3}{2} (1 - x^2) \mathbb{1}_{(0,1)}.$$

In practice, the semi-metric choice is based on the regularity of the curves X which are under analysis. In our case, we take the semi-metric based on the second derivatives of the curves X . More precisely, we take

$$d(X_i, X_j) = \left(\int_0^1 (X_i''(t) - X_j''(t))^2 dt \right)^{1/2} \quad \forall X_i, X_j \in \mathcal{F},$$

where X'' denotes the second derivative of the curve X . For the bandwidth, we choose the automatic selection with a cross validation procedure introduced by (Ferraty and Vieu, 2006, Ch.13).

We split the data generated from the model above into two subsets: a training sample (X_i, T_i, δ_i) , for $i = 1, \dots, 150$, and a test sample (X_j, T_j, δ_j) , for $j = 151, \dots, 200$. Then, we calculate the estimator $\tilde{r}(X_j)$ for any $j \in \{151, \dots, 200\}$.

The performance of both estimators is described by the mean squared error (MSE) formulated as

$$\text{MSE} = \frac{1}{50} \sum_{j=151}^{200} (r(X_j) - \tilde{r}(X_j))^2,$$

where $\tilde{r}(X_j)$ means the estimator of both regression models and $r(X_j)$ the response variable. We note that the result of our simulation study is evaluated over 100 independent replications.

The obtained results are shown in Figure 2 with the censorship rate $\text{CR} = 20.67\%$. It is clear that there is no meaningful difference between the two estimation methods: the classical kernel estimator (CKE) has an $\text{MSE}_{\text{CKE}} = 0.2209$, whereas the relative error estimator (REE) has an $\text{MSE}_{\text{REE}} = 0.1579$.

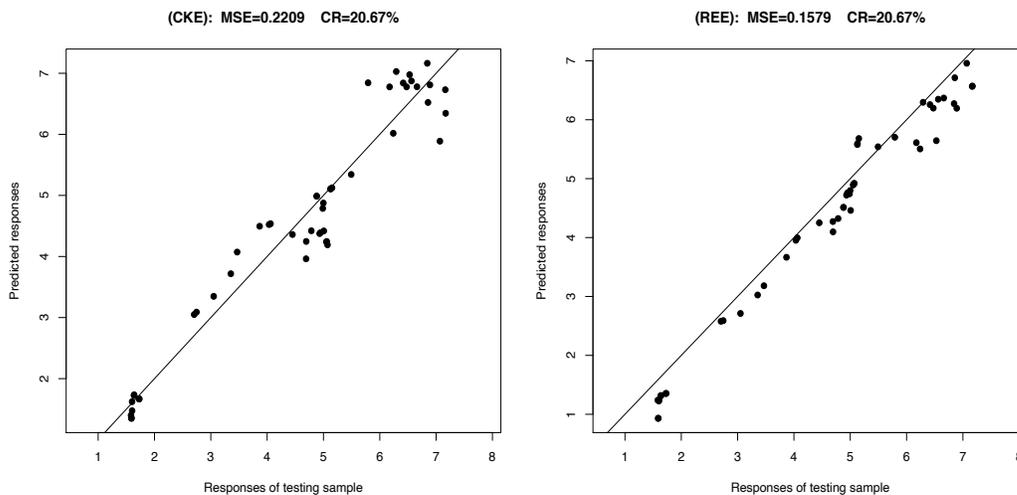


Figure 2. comparison between the CKE and the REE without outliers.

The results of a second illustration are given in Table 1, where from we observe that, in the presence of outliers $(0, 10, 20)$ with different values of $\text{CR} = 3\%, 30\%, 60\%$, the relative

error regression performs better than the classical method, even if the MSE of both methods increases substantially relatively to the number of the perturbed points and censorship rate, it remains very low in terms of the relative error.

Table 1. MSE of the CKE and REE according to numbers of introduced artificial outliers and different censorship rate.

Number of artificial outliers	→ CR ↓	0	10	20
Classical kernel estimator MSE _{CKE}	3%	0.0921	2856.646	6499.6945
	30%	0.8766	14126.2706	19358.5386
	60%	2.8038	32182.8188	56681.7038
Relative error estimator MSE _{REE}	3%	0.0551	0.0579	0.0665
	30%	0.0949	0.1048	0.1258
	60%	0.1455	0.1903	0.2712

Our main application of Theorem 3.3 is to build confidence intervals (CIs) for the true value of $r(x)$ given curve $X = x$. A plug-in estimate for the asymptotic standard deviation $(n\varphi_x(h)/\sigma^2(x))^{1/2}$ and the bias term $hB_n(x) + o(h)$. Precisely, we estimate $q_m(x)$ by means of

$$\tilde{q}_m(x) = \frac{\sum_{i=1}^n K_i \delta_i \bar{G}_n^{-2}(T_i) T_i^{-m}}{\sum_{i=1}^n K_i},$$

whereas we estimate empirically β_1 and β_2 by using

$$\hat{\beta}_1 = \frac{1}{n\varphi_x(h)} \sum_{i=1}^n K_i \quad \text{and} \quad \hat{\beta}_2 = \frac{1}{n\varphi_x(h)} \sum_{i=1}^n K_i^2.$$

Thus, the practical estimator of the normalized deviation is stated as

$$\tilde{\sigma}_n(x) = \left(\frac{(\sum_{i=1}^n K_i^2) (\tilde{q}_2(x) - 2\tilde{r}(x)\tilde{q}_3(x) + \tilde{r}^2(x)\tilde{q}_4(x))}{(\sum_{i=1}^n K_i)^2 \tilde{q}_2^2(x)} \right)^{1/2}.$$

We point out that the function φ_x do not intervene in the calculation of the CI by simplification. Hence, the approximate $(1 - \zeta/2) \times 100\%$ CI for $r(x)$, for any $x \in \mathcal{F}$, is given by

$$\left[\tilde{r}_n(x) - z_{1-\zeta/2} \tilde{\sigma}_n(x), \tilde{r}_n(x) + z_{1-\zeta/2} \tilde{\sigma}_n(x) \right],$$

where $z_{1-\zeta/2}$ denotes the $(1 - \zeta/2) \times 100$ th quantile of the standard normal distribution.

In order to compare our CI with that of the classical regression (Ferraty et al., 2007), we have

$$\sqrt{n\varphi_x(h)} \frac{\beta_1}{\sigma_\epsilon(x) \sqrt{\beta_2}} (\hat{m}(x) - m(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where $\sigma_\epsilon^2(x) = \mathbb{E}[(Y - m(x))^2 | X = x]$ and β_1, β_2 are define previously.

With simple calculus, we can estimate $\sigma_\epsilon^2(x)$ based on

$$\hat{\sigma}_\epsilon^2(x) = \hat{\rho}_2(x) - 2\hat{m}(x)\hat{\rho}_1(x) + \hat{m}^2(x),$$

where

$$\hat{\rho}_l(x) = \frac{\sum_{i=1}^n K_i \delta_i \bar{G}_n^{-1}(T_i) T_i^l}{\sum_{i=1}^n K_i}, \quad \forall l \in \{1, 2\}.$$

Therefore, the approximate $(1 - \zeta/2) \times 100\%$ CI for $m(x)$ (the classical regression), for any $x \in \mathcal{F}$, is formulated as

$$\left[\hat{m}(x) - z_{1-\zeta/2} \frac{\sqrt{\hat{\beta}_2 \hat{\sigma}_\epsilon(x)}}{\hat{\beta}_1}, \hat{m}(x) + z_{1-\zeta/2} \frac{\sqrt{\hat{\beta}_2 \hat{\sigma}_\epsilon(x)}}{\hat{\beta}_1} \right]$$

In order to construct confidence bands (for both CKE and REE), we proceed by the following algorithm:

- Step 1 Split our data into randomly chosen subsets: $(X_i, Y_i)_{i \in I}$ (training set) and $(X_j, Y_j)_{j \in J}$ (test set).
- Step 2 Calculate the estimator $\tilde{r}_n(X_i)$ for all $i \in I$ by using the training sample.
- Step 3 For each X_j in the test sample, set $i^* := \arg \min_{i \in I} d(X_j, X_i)$.
- Step 4 For all $j \in J$, define the confidence bands by means of

$$[\tilde{r}_n(X_{i^*}) - z_{0.975} \tilde{\sigma}_n(X_{i^*}), \tilde{r}_n(X_{i^*}) + z_{0.975} \tilde{\sigma}_n(X_{i^*})],$$

where $z_{0.975} \approx 1.96$ is the 97.5% quantile of a standard normal distribution.

- Step 5 We present our results by plotting the extremities of the predicted values versus the true values and the confidence bands.

Figures 3 and 4 shows clearly a good behaviour of our estimator compared to the classical regression, with censorship rate (CR = 30%) and in the presence of outliers. In these figures, the solid black curve connects the true values. The dashed blue curves connect the lower and upper predicted values. The solid red curve connects the crossed points which give the predicted values.

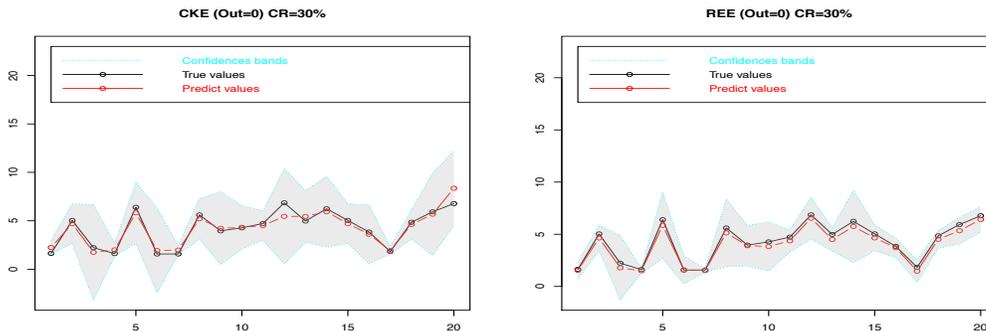


Figure 3. Extremities of the predicted values versus the true values and the confidence bands (simulation data without outliers).

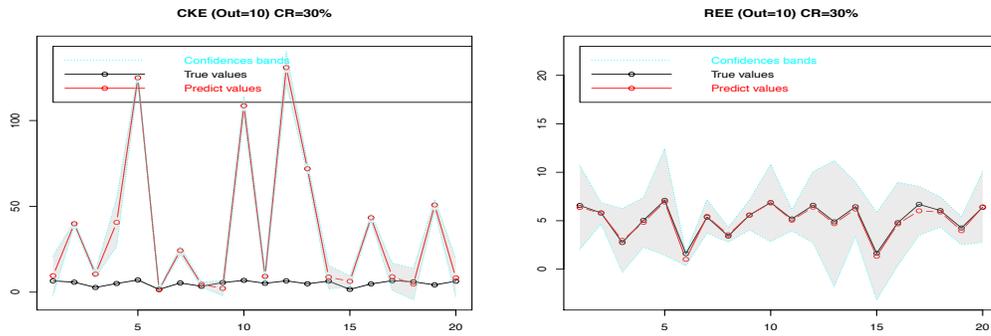


Figure 4. Extremities of the predicted values versus the true values and the confidence bands (simulation data in the presence of 10 outliers).

4.2 A REAL DATA APPLICATION

First, we have acquired a large dataset, consisting of number of 8784 records, containing the hourly energy consumption for the year 2016 (measured in MWh), retrieved from the smart metering device of a commercial center type of consumer (a large hypermarket). We have also acquired a dataset containing the historical hourly meteorological data regarding the temperature (measured in Celsius degrees). These data were recorded by the meteorological sensors of a specialized institute for the year 2016, consisting in a number of 8784 records; see [Pîrjan et al. \(2017\)](#) and [Mebsout et al. \(2020\)](#) for more description on this data set.

Now, we are interested in the estimation of interval prediction of peak consumption of energy. For a fixed day i , let us denote by $(E_i(t_j))_{j=1,\dots,24}$ the hourly measurements of some consumption of energy. The peak demand observed for the day i is defined as

$$P_i = \max_{j=1,\dots,24} E_i(t_j).$$

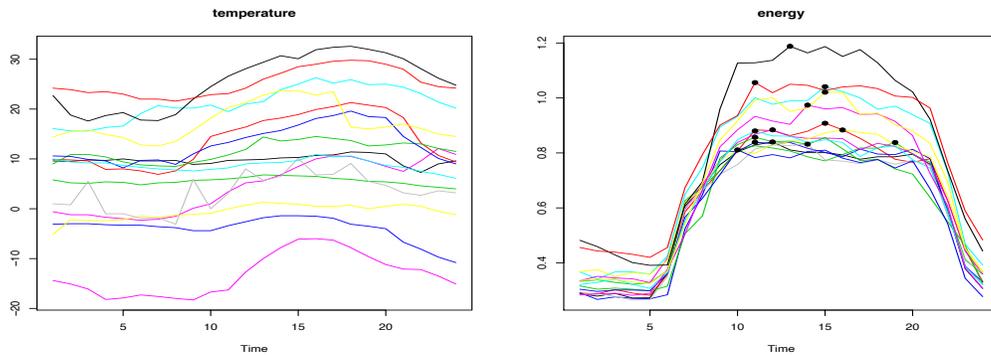


Figure 5. Sample of 15 daily temperature curves and the associated energy consumption curves.

It is well known that peak demand is very correlated with temperature measurements. Figure 5 provides a sample of 15 curves of hourly temperature measures and the associated electricity consumption curves. We split our sample of 366 days into a learning sample containing the first 300 days and a testing sample with the last 66 days. From the learning sample, we selected 30% of days within which we generated the censorship randomly. Figure 6 provides a sample of four censored daily load curves. For those days, we observe the electricity consumption until a certain time $t_c \in [1, 24]$, which corresponds to the time of

ensorship which is plotted in a dashed line in Figure 6. For a censored day, we define the censored random variable as

$$C_i = \max_{j=1, \dots, t_c} E_i(t_j).$$

Therefore, our sample is formed as follows $(X_i, Y_i, \delta_i)_{i=1, \dots, 300}$, where $\delta_i = 1$ if $Y_i = P_i$ and $\delta_i = 0$ if $Y_i = C_i$. In order to introduce the outliers in this sample, we randomly multiplies by 10 some response variable of a number of observations.

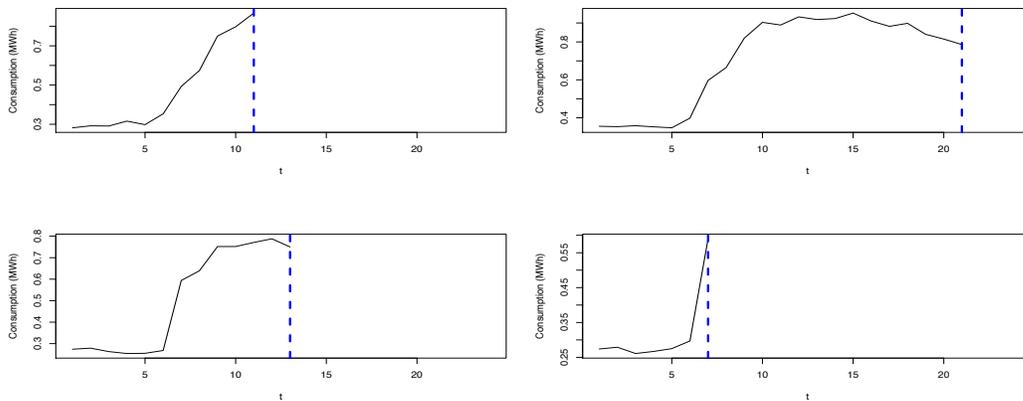


Figure 6. Sample of four censored daily load curves, the dashed line corresponds to the time of censorship t_c .

The selection of the bandwidth parameter is an important and basic problem in all kernel smoothing techniques. Another important point for ensuring a good behavior of the method is to use a semi-metric that is well adapted to the kind of data we have to deal with. Our data are based on the m eigen-functions of the empirical covariance operator associated with the m greatest eigenvalues (Ferraty and Vieu, 2006, Ch. 13). The estimators are obtained by choosing the optimal bandwidths by L^1 cross-validation method and the kernel K is the quadratic function defined by $K(x) = 3/2(1 - x^2) \mathbb{1}_{[0,1]}$. The error used is expressed by

$$\text{MSE}_{\text{CKE}} = \frac{1}{66} \sum_{i=301}^{366} (Y_i - \hat{m}(X_i))^2 \quad \text{and} \quad \text{MSE}_{\text{REE}} = \frac{1}{66} \sum_{i=301}^{366} (Y_i - \tilde{r}(X_i))^2.$$

The results are given in Figure 7, where two curves corresponding to the observed values (black curve) the predicted values (dashed curve green for the classical regression and red for the relative one) are drawn. Clearly, Figure 7 shows the good behavior of our procedure. We observe that the relative approach gives better results than the classical regression approach ($\text{MSE}_{\text{CKE}} = 0.0883$ and $\text{MSE}_{\text{REE}} = 0.0034$).

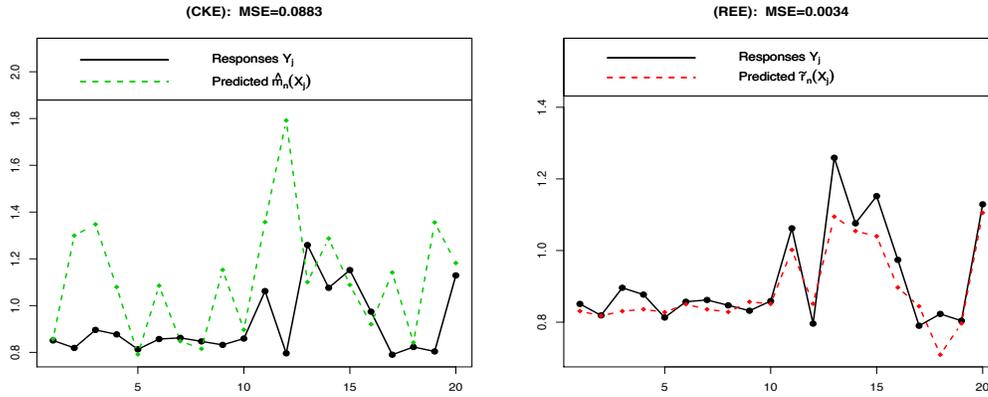


Figure 7. Prediction by classical and relative regression.

Now, we give in Table 2 the 90% predictive intervals of the concentrations for the peak load of the 20 last values in the sample test. This conclusion shows the good performance of our asymptotic normality.

Table 2. The 90% predictive intervals of the peak demand for the last 20 days.

True value	Predicted value	Predictive CI _{90%}	True value	Predicted value	Predictive CI _{90%}
0.851	0.8310	[0.6078, 1.0542]	1.062	1.0017	[0.8279, 1.1756]
0.819	0.8177	[0.7376, 0.8978]	0.796	0.8514	[0.7592, 0.9435]
0.896	0.8307	[0.7697, 0.8918]	1.259	1.0946	[0.9344, 1.2548]
0.877	0.8358	[0.4879, 1.1838]	1.076	1.0545	[0.8648, 1.2441]
0.813	0.8277	[0.4660, 1.1894]	1.152	1.0399	[0.9289, 1.1508]
0.857	0.8501	[0.5713, 1.1289]	0.974	0.8968	[0.7833, 1.0103]
0.862	0.8358	[0.7802, 0.8914]	0.790	0.8444	[0.7913, 0.8974]
0.847	0.8284	[0.3206, 1.3363]	0.823	0.7091	[0.0456, 1.3727]
0.832	0.8568	[0.7976, 0.9160]	0.804	0.7965	[0.6710, 0.9219]
0.859	0.8511	[0.7328, 0.9694]	1.129	1.1054	[0.8670, 1.3437]

CONCLUSIONS

In this paper, we have investigated the asymptotic properties of a nonparametric estimator of the relative error regression given a dependent functional explanatory variable, in the case of a scalar censored response. We have used the mean squared relative error as a loss function to construct a nonparametric estimator of the regression operator of these functional censored data. We have established the almost surely convergence and asymptotic normality of the proposed estimator. A simulation study and real data application were performed to support the theoretical results and to compare the quality of predictive performances of the relative error regression estimator than those obtained with standard kernel regression estimates. Our proposal provides interesting findings and is a tool that can be helpful to diverse practitioners. Our proposal has some limitations that open some doors for further research, which will be considered by the authors in future works.

APPENDIX

Proof of Theorem 3.2 From Equation (5), we have

$$\begin{aligned} |\tilde{r}_n(x) - r(x)| &\leq \frac{1}{|\tilde{g}_{2,n}(x)|} \left\{ |\tilde{g}_{1,n}(x) - \tilde{g}_1(x)| + |\tilde{g}_1(x) - \mathbb{E}(\tilde{g}_1(x))| \right. \\ &\quad \left. + |\mathbb{E}(\tilde{g}_1(x)) - g_1(x)| \right\} + \frac{|r(x)|}{|\tilde{g}_{2,n}(x)|} \left\{ |\tilde{g}_{2,n}(x) - \tilde{g}_2(x)| \right. \\ &\quad \left. + |\tilde{g}_2(x) - \mathbb{E}(\tilde{g}_2(x))| + |\mathbb{E}(\tilde{g}_2(x)) - g_2(x)| \right\}. \end{aligned}$$

Therefore, Theorem 3.2's result is a consequence of the following intermediate results.

Lemma 4.1 Under hypotheses (H2)-(H5), we have

$$|\tilde{g}_{l,n}(x) - \tilde{g}_l(x)| = O_{AS} \left(\sqrt{\frac{\log(\log(n))}{n}} \right),$$

for $l \in \{1, 2\}$.

Lemma 4.2 Under hypotheses (H1)-(H3) and (H5), we get

$$|\mathbb{E}(\tilde{g}_l(x)) - g_l(x)| = O(h^{k_l}),$$

for $l \in \{1, 2\}$.

Lemma 4.3 Under hypotheses (H1)-(H4) and (H6)-(H8), we obtain

$$|\tilde{g}_l(x) - \mathbb{E}(\tilde{g}_l(x))| = O_{ACC} \left(\sqrt{\frac{\log(n)}{n\varphi_x(h)}} \right),$$

for $l \in \{1, 2\}$.

Corollary 4.4 Under the hypotheses of Lemma 4.1 and 4.2, we have that

$$\text{there exists } \delta > 0; \text{ such that } \sum_{n=1}^{\infty} \mathbb{P}(|\tilde{g}_{2,n}(x)| < \delta) < \infty.$$

Let denote $K_i(x)$ by $K(d(x, X_i)/h)$.

Proof of Lemma 4.1

The proof is similar to Lemma 3.1 of Fetitah et al. (2020).

Proof of Lemma 4.2

For all $l = 1, 2$, we get that

$$\begin{aligned} |\mathbb{E}(\tilde{g}_l(x)) - g_l(x)| &= \left| \mathbb{E} \left(\frac{K_1(x)}{\mathbb{E}(K_1(x))} \mathbb{E} \left[\frac{\mathbb{E}(\mathbf{1}_{Y_1 \leq C_1} | Y_1) Y_1^{-l}}{\bar{G}(Y_1)} | X_1 \right] \right) - g_l(x) \right| \\ &= \frac{1}{\mathbb{E}(K_1(x))} \left| \mathbb{E} \left\{ \left[\mathbb{E}(Y_1^{-l} | X_1) - g_l(x) \right] \mathbf{1}_{B(x,h)}(X_1) K_1(x) \right\} \right|. \end{aligned}$$

Then, by the Hölder hypothesis (H2) we obtain that

$$|g_l(X_1) - g_l(x)| \leq ch^{k_l}.$$

Thus,

$$|\mathbb{E}(\tilde{g}_l(x)) - g_l(x)| \leq ch^{k_l}.$$

Proof of Lemma 4.3

For $l = 1, 2$ we put

$$\Delta_i(x) = \frac{\delta_i T_i^{-l}}{\bar{G}(T_i)} K\left(\frac{d(x, X_i)}{h}\right) - \mathbb{E}\left[\frac{\delta_i T_i^{-l}}{\bar{G}(T_i)} K\left(\frac{d(x, X_i)}{h}\right)\right].$$

The use of the Fuk-Nagaev inequality (Rio, 1999, p. 87, 6.19b), which is based on

$$\begin{aligned} S_n^2 &= \sum_{i=1}^n \sum_{j=1}^n |\text{Cov}(\Delta_i(x), \Delta_j(x))| \\ &= \sum_{i \neq j} |\text{Cov}(\Delta_i(x), \Delta_j(x))| + n \text{Var}(\Delta_1(x)). \end{aligned}$$

Now, by using (H5), we get

$$\begin{aligned} \text{Var}(\Delta_1(x)) &\leq \mathbb{E}\left[\frac{\delta_1 Y_1^{-2l}}{\bar{G}^2(Y_1)} K_1^2(x)\right] + \mathbb{E}^2\left[\frac{\delta_1 Y_1^{-l}}{\bar{G}(Y_1)} K_1(x)\right] \\ &\leq \mathbb{E}\left[K_1^2(x) \mathbb{E}\left(\frac{\mathbb{E}(\mathbf{1}_{Y_1 \leq C_1} | Y_1) Y_1^{-2l}}{\bar{G}^2(Y_1)} \middle| X_1\right)\right] \\ &\quad + \mathbb{E}^2\left[K_1(x) \mathbb{E}\left(\frac{\mathbb{E}(\mathbf{1}_{Y_1 \leq C_1} | Y_1) Y_1^{-l}}{\bar{G}(Y_1)} \middle| X_1\right)\right] \\ &\leq \frac{c}{\bar{G}(\tau_F)} \mathbb{E}[K_1^2(x) \mathbb{E}(Y_1^{-2l} | X_1)] + \mathbb{E}^2[K_1(x) \mathbb{E}(Y_1^{-l} | X_1)] \\ &\leq \frac{c}{\bar{G}(\tau_F)} \mathbb{E}[K_1^2(x)] + c\varphi_x^2(h) \\ &\leq c(\varphi_x(h) + \varphi_x^2(h)). \end{aligned}$$

In addition, for $i \neq j$, we have

$$\begin{aligned} |\text{Cov}(\Delta_i(x), \Delta_j(x))| &= |\mathbb{E}(\Delta_i(x)\Delta_j(x))| \\ &\leq c |\mathbb{E}(K_i(x)K_j(x)) + \mathbb{E}(K_i(x)) \mathbb{E}(K_j(x))|. \end{aligned}$$

Then, following Masry (1986), we define the sets given by $E_1 = \{(i, j), \text{ such that } 1 \leq |i-j| \leq \nu_n\}$ and $E_2 = \{(i, j) \text{ such that } \nu_n + 1 \leq |i-j| \leq n\}$, where $\nu_n \rightarrow \infty$ as $n \rightarrow \infty$. Then, we can write $\sum_{i \neq j} |\text{Cov}(\Delta_i(x), \Delta_j(x))| = J_{1,n} + J_{2,n}$, where $J_{1,n}$ and $J_{2,n}$ are the sums of the

covariances over E_1 and E_2 respectively. Therefore, under (H7), we get

$$\begin{aligned} J_{1,n} &= \sum_{E_1} |\text{Cov}(\Delta_i(x), \Delta_j(x))| \leq c \sum_{E_1} |\mathbb{E}(K_i(x)K_j(x)) + \mathbb{E}(K_1(x))^2| \\ &\leq c \sum_{E_1} |\mathbb{P}((X_i, X_j) \in B(x, h) \times B(x, h)) + \varphi_x(h)^2| \\ &\leq cn\nu_n\varphi_x(h) \left[\left(\frac{\varphi_x(h)}{n} \right)^{\frac{1}{a}} + \varphi_x(h) \right]. \end{aligned}$$

For the second term, we use the modified Davydov covariance inequality for mixing processes (Rio, 1999, p.10). Then, we have

$$\forall i \neq j, |\text{Cov}(\Delta_i(x), \Delta_j(x))| \leq c\alpha(|i - j|).$$

Thus, we get by (H6) that

$$J_{2,n} \leq \sum_{E_2} |\text{Cov}(K_i(x), K_j(x))| \leq n^2\nu_n^{-a}.$$

Hence, for $\nu_n = (\varphi_x(h)/n)^{-1/a}$, we have $\sum_{i \neq j} |\text{Cov}(\Delta_i(x), \Delta_j(x))| = O(n\varphi_x(h))$. Consequently, combining previous result, we obtain

$$S_n^2 = O(n\varphi_x(h)). \tag{6}$$

Using the Fuk-Nagaev inequality, we get, for all $l = 1, 2$, $\varepsilon > 0$ and $r > 1$, that

$$\begin{aligned} \mathbb{P} \left[\left| \mathbb{E}[\tilde{g}_l(x)] - \tilde{g}_l(x) \right| > \varepsilon \right] &= \mathbb{P} \left[\left| \frac{1}{n\mathbb{E}(K_1(x))} \sum_{i=1}^n \Delta_i(x) \right| > \varepsilon \right] \\ &= \mathbb{P} \left[\left| \sum_{i=1}^n \Delta_i(x) \right| > \varepsilon n\mathbb{E}(K_1(x)) \right] \\ &\leq c \left\{ \left(1 + \frac{\varepsilon^2 n^2 \mathbb{E}(K_1(x))^2}{r S_n^2} \right)^{-r/2} + nr^{-1} \left(\frac{r}{\varepsilon n\mathbb{E}(K_1(x))} \right)^{a+1} \right\} \\ &\leq c(A_1 + A_2), \end{aligned}$$

where

$$A_1 = \left(1 + \frac{\varepsilon^2 n^2 (\mathbb{E}[K_1(x)])^2}{r S_n^2} \right)^{-r/2} \quad \text{and} \quad A_2 = nr^{-1} \left(\frac{r}{\varepsilon n\mathbb{E}[K_1(x)]} \right)^{a+1}.$$

Therefore, by Equation (6) and putting

$$\varepsilon = \varepsilon_0 \sqrt{\frac{\log(n)}{n\varphi_x(h)}} \quad \text{and} \quad r = (\log(n))^2,$$

it follow that $A_2 \leq cn^{1-(a+1)/2}\varphi_x(h)^{-(a+1)/2}(\log(n))^{(3a-1)/2}$. Next, using the left side of (H8), we obtain $A_2 \leq cn^{-1-\eta(a+1)/2}(\log(n))^{(3a-1)/2}$. Hence, it exists some real $\nu > 0$ such

that

$$A_2 \leq cn^{-1-\nu}. \quad (7)$$

Because of $r = (\log(n))^2$, we show that

$$A_1 \leq \left(1 + \frac{\varepsilon_0^2}{\log(n)}\right)^{-\frac{(\log(n))^2}{2}} = e^{-\frac{(\log(n))^2}{2} \log\left(1 + \frac{\varepsilon_0^2}{\log(n)}\right)}.$$

Using the fact that $\log(1+x) = x - x^2/2 + o(x^2)$, when $x \rightarrow 0$, we get

$$A_1 \leq e^{-\frac{\varepsilon_0^2 \log(n)}{2}} = n^{-\frac{\varepsilon_0^2}{2}}.$$

The last result allows us to get directly that there exist some ε_0 and some ν' such that

$$A_1 \leq cn^{-1-\nu'}. \quad (8)$$

Therefore, by the results of Equations (8) and (7), we have

$$\sum_{n \geq 1} \mathbb{P} \left[\left| \mathbb{E}[\tilde{g}_l(x)] - \tilde{g}_l(x) \right| > \varepsilon_0 \sqrt{\frac{\log(n)}{n\varphi_x(h)}} \right] < \infty.$$

Proof of Corollary 4.4

The proof of this Corollary is analogous to Corollary 2 of [Demongeot et al. \(2016\)](#).

Proof of Theorem 3.3

From Equation (5), we adopt the decomposition stated as

$$\tilde{r}_n(x) - r(x) = \tilde{r}_n(x) - \tilde{r}(x) + \tilde{r}(x) - r(x) =: I_{1n}(x) + I_{2n}(x)$$

where

$$I_{1n}(x) =: \tilde{r}_n(x) - \tilde{r}(x) \quad \text{and} \quad I_{2n}(x) =: \tilde{r}(x) - r(x).$$

The proof is derived by showing first that $I_{1n}(x)$ is negligible whereas $I_{2n}(x)$ is asymptotically normal distributed.

From Lemma 4.1 and Corollary 4.4, we deduce that

$$I_{1n}(x) \xrightarrow{\mathbb{P}} 0. \quad (9)$$

Now, we can write that

$$I_{2n}(x) = \frac{1}{\tilde{g}_2(x)} \left[D_n + A_n \left(\mathbb{E}[\tilde{g}_2(x)] - \tilde{g}_2(x) \right) \right] + A_n, \quad (10)$$

where

$$A_n = \frac{1}{\mathbb{E}[\tilde{g}_2(x)]g_2(x)} \left\{ \mathbb{E}[\tilde{g}_1(x)]g_2(x) - \mathbb{E}[\tilde{g}_2(x)]g_1(x) \right\}$$

$$D_n = \frac{1}{g_2(x)} \left[V_{1n}(x)g_2(x) - V_{2n}(x)g_1(x) \right],$$

with $V_{ln}(x) = \tilde{g}_l(x) - \mathbb{E}[\tilde{g}_l(x)]$, for $l = 1, 2$.

Then, it follows from Equation (10) that

$$\begin{aligned} \tilde{r}(x) - r(x) - A_n &= \frac{1}{\tilde{g}_2(x)} \left[D_n + A_n \left(\mathbb{E}[\tilde{g}_2(x)] - \tilde{g}_2(x) \right) \right] \\ &=: \frac{D_n - A_n V_{2n}(x)}{\tilde{g}_2(x)}. \end{aligned}$$

Consequently, the proof of Theorem 3.3 can be deduced from the convergence in Equation (9) and the following intermediate results (cf. Lemmas 4.5, 4.6 and 4.7).

Lemma 4.5 Under hypotheses of Theorem 3.3, we have

$$\left(\frac{n\varphi_x(h)}{g_2^2(x)\sigma^2(x)} \right)^{1/2} \left(\left[\tilde{g}_1(x) - \mathbb{E}[\tilde{g}_1(x)] \right] g_2(x) - \left[\tilde{g}_2(x) - \mathbb{E}[\tilde{g}_2(x)] \right] g_1(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Lemma 4.6 Under hypotheses of Theorem 3.3, we obtain

$$A_n = hB_n + o(h).$$

Lemma 4.7 Under hypotheses of Theorem 3.3, we obtain

$$\tilde{g}_2(x) \xrightarrow{\mathbb{P}} g_2(x),$$

and

$$\left(\frac{n\varphi_x(h)}{g_2^2(x)\sigma^2(x)} \right)^{1/2} A_n \left(\mathbb{E}[\tilde{g}_2(x)] - \tilde{g}_2(x) \right) \xrightarrow{\mathbb{P}} 0.$$

Proof of Lemma 4.5

It is easy to see that

$$\sqrt{n\varphi_x(h)} \left[\left(\tilde{g}_2(x) - \mathbb{E}[\tilde{g}_2(x)] \right) g_1(x) - \left(\tilde{g}_1(x) - \mathbb{E}[\tilde{g}_1(x)] \right) g_2(x) \right] = \frac{1}{\sqrt{n}} \sum_{i=1}^n L_i(x),$$

where

$$L_i(x) := \frac{\sqrt{\varphi_x(h)}}{\mathbb{E}[K_1]} \left\{ \frac{\delta_i}{\bar{G}(T_i)} K_i \left(g_1(x)T_i^{-2} - g_2(x)T_i^{-1} \right) - \mathbb{E} \left[\frac{\delta_i}{\bar{G}(T_i)} K_i \left(g_1(x)T_i^{-2} - g_2(x)T_i^{-1} \right) \right] \right\}.$$

The proof of this lemma is based on the central limit theorem of Doukhan et al. (1994). We have then to consider the asymptotic behavior of the variance term and the following

assumption

$$\int_0^1 \alpha^{-1}(u) (Q_{L_1}(u))^2 du < +\infty,$$

where Q_{L_1} is the upper tail quantile function defined by

$$Q_{L_1}(u) = \inf \{t \geq 0 : \mathbb{P}(L_1 > t) \leq u\}$$

and $\alpha^{-1}(u) = \sum_{n \in \mathbb{N}} \mathbb{1}_{u < \alpha_n}$. Clearly,

$$\begin{aligned} \text{Var} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n L_i(x) \right) &= n\varphi_x(h) \text{Var} \left(\frac{g_1(x)}{n\mathbb{E}[K_1]} \sum_{i=1}^n \frac{\delta_i}{\bar{G}(T_i)} K_i T_i^{-2} - \frac{g_2(x)}{n\mathbb{E}[K_1]} \sum_{i=1}^n \frac{\delta_i}{\bar{G}(T_i)} K_i T_i^{-1} \right) \\ &= n\varphi_x(h) \left(\text{Var} [\tilde{g}_1(x)] g_2^2(x) + \text{Var} [\tilde{g}_2(x)] g_1^2(x) \right. \\ &\quad \left. - 2g_1(x)g_2(x) \text{Cov} [\tilde{g}_1(x), \tilde{g}_2(x)] \right). \end{aligned}$$

By definition of $\tilde{g}_l(x)$ for $l = 1; 2$, we have

$$\begin{aligned} n\varphi_x(h) \text{Var} [\tilde{g}_l(x)] &= \frac{\varphi_x(h)}{(\mathbb{E}[K_1])^2} \text{Var} \left[\frac{\delta_1}{\bar{G}(T_1)} K_1 T_1^{-l} \right] \\ &\quad + \frac{\varphi_x(h)}{n(\mathbb{E}[K_1])^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ |i-j|>0}}^n \text{Cov} \left[\frac{\delta_i}{\bar{G}(T_i)} K_i T_i^{-l}, \frac{\delta_j}{\bar{G}(T_j)} K_j T_j^{-l} \right] \\ &= \frac{\varphi_x(h)}{(\mathbb{E}[K_1])^2} J_{1,1} + \frac{\varphi_x(h)}{n(\mathbb{E}[K_1])^2} J_{2,n} \end{aligned}$$

where

$$\begin{aligned} J_{1,1} &= \text{Var} \left[\frac{\delta_1}{\bar{G}(T_1)} K_1 T_1^{-l} \right], \\ J_{2,n} &= \sum_{i=1}^n \sum_{\substack{j=1 \\ |i-j|>0}}^n \text{Cov} \left[\frac{\delta_i}{\bar{G}(T_i)} K_i T_i^{-l}, \frac{\delta_j}{\bar{G}(T_j)} K_j T_j^{-l} \right]. \end{aligned}$$

By conditioning on the random variable X_1 , by the same ideas in the proof of lemma 4.2, Lemma 4 in Ferraty et al. (2007) and by using hypotheses (H5), (N1) and (N4), we get

$$\begin{aligned} \mathbb{E} \left[\left(\frac{\delta_1 Y_1^{-l}}{\bar{G}(Y_1)} \right)^2 K_1^2(x) \right] &= \mathbb{E} \left[K_1^2(x) \mathbb{E} \left(\frac{\mathbb{E}(\mathbb{1}_{Y_1 \leq C_1} | Y_1) Y_1^{-2l}}{\bar{G}^2(Y_1)} | X_1 \right) \right] \\ &= \left[\mathbb{E} \left(\frac{Y_1^{-2l}}{\bar{G}(Y_1)} | X_1 = x \right) + o(1) \right] \mathbb{E} [K_1^2(x)] \\ &= \varphi_x(h) \mathbb{E} \left[\frac{Y_1^{-2l}}{\bar{G}(Y_1)} | X_1 = x \right] \left(K^2(1) - \int_0^1 (K^2(s))' \chi_x(u) du \right) + o(\varphi_x(h)) \end{aligned}$$

and

$$\mathbb{E} \left(\frac{\delta_1}{\bar{G}(Y_1)} Y_1^{-l} K_1 \right) = O(\varphi_x(h)).$$

Thus:

$$\begin{aligned} \text{Var} \left[\frac{\delta_1}{\bar{G}(T_1)} T_1^{-l} K_1 \right] &= \varphi_x(h) \mathbb{E} \left[\bar{G}^{-1}(Y_1) Y_1^{-2l} | X_1 = x \right] \left(K^2(1) - \int_0^1 (K^2(s))' \chi_x(u) du \right) \\ &\quad + O(\varphi_x^2(h)). \end{aligned}$$

We obtain

$$\frac{\varphi_x(h)}{(\mathbb{E}[K_1])^2} J_{1,1} \rightarrow \frac{q_{2l}(x)\beta_2}{\beta_1^2}. \tag{11}$$

Let us turn to $J_{2,n}$, for this we use the technique of Masry (1986). We define the same sets E_1 and E_2 in the proof of Lemma 4.3. Let $J_{2,n}^1$ and $J_{2,n}^2$ be the sums of covariances over E_1 and E_2 respectively. On the one hand, we have

$$J_{2,n}^1 = \sum_{E_1} \left| \text{Cov} \left[\frac{\delta_i}{\bar{G}(T_i)} K_i T_i^{-l}, \frac{\delta_j}{\bar{G}(T_j)} K_j T_j^{-l} \right] \right| \leq C \sum_{E_1} |\mathbb{E}[K_i K_j] - \mathbb{E}[K_i] \mathbb{E}[K_j]|.$$

Because of the assumptions of Lemma 4.3 we can write

$$J_{2,n}^1 \leq cn\nu_n \varphi_x(h) \left(\left(\frac{\varphi_x(h)}{n} \right)^{\frac{1}{a}} + \varphi_x(h) \right).$$

Hence, for the summation over E_2 , we use the Davydov-Rio inequality (Rio, 1999, p. 87), for mixing processes. This leads, for all $i \neq j$, to

$$|\text{Cov}(K_i, K_j)| \leq c\alpha(|i - j|),$$

Therefore,

$$\sum_{E_2} |\text{Cov}(K_i, K_j)| \leq n^2 \nu_n^{-a}.$$

The choice $\nu_n = 1/[\varphi_x(h) \log(n)]$, motivated by the upper bound in (H8), permits to get

$$\sum_{i \neq j}^n \text{Cov}(K_i, K_j) = o(n\varphi_x(h)),$$

then

$$\frac{\varphi_x(h)}{n(\mathbb{E}[K_1])^2} J_{2,n} = o(1) \text{ as } n \rightarrow \infty. \tag{12}$$

Thanks to Equations (11) and (12), we have

$$n\varphi_x(h) \text{Var}(\tilde{g}_l(x)) \longrightarrow \frac{\beta_2 q_{2l}(x)}{\beta_1^2} \text{ as } n \longrightarrow \infty. \quad (13)$$

Concerning the covariance term, we follow the same steps as for the variance given in Equation (13) then we get:

$$n\varphi_x(h) \text{Cov}(\tilde{g}_1(x), \tilde{g}_2(x)) \longrightarrow \frac{\beta_2 q_3(x)}{\beta_1^2} \text{ as } n \longrightarrow \infty. \quad (14)$$

Let us now prove the claimed result. Clearly, the function Q_{L_1} is nonincreasing, then

$$\sum_{n=1}^{\infty} \int_0^{\alpha_n} [Q_{L_1}(u)]^2 du \leq \sum_{n=1}^{\infty} \alpha_n Q_{L_1}^2(0).$$

By hypotheses (H1), (H3) and (H5) we can write

$$c \frac{1}{\sqrt{\varphi_x(h)}} \leq |L_1| \leq c' \frac{1}{\sqrt{\varphi_x(h)}}.$$

Then,

$$Q_{L_1}(0) \leq c' \frac{1}{\sqrt{\varphi_x(h)}}.$$

Therefore, we have

$$\sum_{n=1}^{\infty} \int_0^{\alpha_n} [Q_{L_1}(u)]^2 du \leq \sum_{n=1}^{\infty} \alpha_n (\varphi_x(h))^{-1}.$$

It follows from (H7) and (H8) that

$$\sum_{n=1}^{\infty} \int_0^{\alpha_n} [Q_{L_1}(u)]^2 du < \infty. \quad (15)$$

From Equations (13), (14) and by noting

$$\sigma^2(x) = \frac{(q_2(x) - 2r(x)q_3(x) + r^2(x)q_4(x))\beta_2}{\beta_1^2},$$

we conclude that

$$\text{Var}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n L_i(x)\right) \longrightarrow \sigma^2(x) \text{ as } n \longrightarrow \infty. \quad (16)$$

Now, the lemma can be easily deduced from Equations (15), (16) and the central limit

theorem of [Doukhan et al. \(1994\)](#) as

$$\frac{1}{\sqrt{ng_2^2\sigma^2(x)}} \sum_{i=1}^n L_i(x) = \left(\frac{n\varphi_x(h)}{g_2^2(x)\sigma^2(x)} \right)^{1/2} \times \left([\tilde{g}_1(x) - \mathbb{E}[\tilde{g}_1(x)]]g_2(x) - [\tilde{g}_2(x) - \mathbb{E}[\tilde{g}_2(x)]]g_1(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Proof of Lemma 4.6

As in [Ferraty et al. \(2007\)](#) we show that:

$$\mathbb{E}[\tilde{r}_n(x)] = \frac{\mathbb{E}[\tilde{g}_1(x)]}{\mathbb{E}[\tilde{g}_2(x)]} + O\left(\frac{1}{n\varphi_x(h)}\right).$$

So, it suffices to evaluate $\mathbb{E}[\tilde{g}_l(x)]$ for $l \in \{1, 2\}$, we obtain

$$\begin{aligned} \mathbb{E}[\tilde{g}_1(x)] &= \frac{1}{\mathbb{E}[K_1]} \mathbb{E}\left(K_1(x)\mathbb{E}[Y_1^{-1}|X_1]\right) \\ &= \frac{1}{\mathbb{E}[K_1]} \left(g_l(x)E[K_1] + E\left[K_1E\left(g_l(X_1) - g_l(x)|d(X_1, x)\right)\right]\right) \\ &= g_l(x) + \frac{E\left[K_1\left(\Psi_l(d(X_1, x))\right)\right]}{\mathbb{E}[K_1]} \\ &= g_l(x) + \frac{\int_0^1 K(t)\Psi_l(ht)d\mathbb{P}^{d(x,X)/h}(t)}{\int_0^1 K(t)d\mathbb{P}^{d(x,X)/h}(t)}. \end{aligned}$$

By using the first-order Taylor expansion for Ψ_l around 0, where $\Psi_l(0) = 0$, we have

$$E[\tilde{g}_l(x)] = g_l(x) + h\Psi'_l(0) \left[\frac{\int_0^1 tK(t)d\mathbb{P}^{d(x,X)/h}(t)}{\int_0^1 K(t)d\mathbb{P}^{d(x,X)/h}(t)} \right] + o(h).$$

According to *Lemma 2* of [Ferraty et al. \(2007\)](#) we get, under (N1)

$$\frac{\int_0^1 tK(t)d\mathbb{P}^{d(x,X)/h}(t)}{\int_0^1 K(t)d\mathbb{P}^{d(x,X)/h}(t)} \longrightarrow \frac{\beta_0}{\beta_1} \text{ and } \int_0^1 K(t)d\mathbb{P}^{d(x,X)/h}(t) \longrightarrow \beta_1.$$

Consequently

$$E[\tilde{g}_l(x)] = g_l(x) + h\Psi'_l(0)\frac{\beta_0}{\beta_1} + o(h)$$

then we deduce that:

$$A_n = \frac{\mathbb{E}[\tilde{g}_1(x)]}{\mathbb{E}[\tilde{g}_2(x)]} - r(x) = hB_n + o(h).$$

Proof of Lemma 4.7

The same idea in the proof of Lemma 3.6 of [Fetitah et al. \(2020\)](#).

AUTHOR CONTRIBUTIONS Conceptualization, investigation, methodology, software, writing-original draft preparation, writing review and editing, writing-original: all the authors participated equally. The authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGEMENTS The authors would like to thank the Editors-in-Chief and the anonymous reviewers for their valuable comments and suggestions which improved substantially the quality of this paper.

FUNDING The authors Fetitah, O. and Attouch, M. acknowledge the research grant under the project (Projets de Recherche Formation-Universitaire/Algeria) PRFU No C00L03UN220120210001. The author Khardani, S. benefited from the financial support of the PHC-Utique 20G1503.

CONFLICTS OF INTEREST The authors declare no conflict of interest.

REFERENCES

- Almanjahie, I.M., Attouch, M.K., Fetitah, O. and Louhab, H., 2020. Robust kernel regression estimator of the scale parameter for functional ergodic data with applications. *Chilean Journal of Statistics*, 11, 73-94.
- Altandji, B., Demangeot, J., Laksaci, A., and Rachedi, M., 2018. Functional data analysis: Estimation of the relative error in functional regression under random left-truncation model. *Journal of Nonparametric Statistics*, 30, 472-490.
- Amiri, A. and Khardani, S., 2018. Nonparametric regression estimation using recursive kernel for dependent censored data. *International Journal of Mathematics and Statistics*, 19, 52-82.
- Amiri, A., Crambes, C., and Thiam, B., 2014. Recursive estimation of nonparametric regression with functional covariate. *Computational Statistics and Data Analysis*, 69, 154-172.
- Attouch, M., Laksaci, A., and Ould Saïd, E., 2009. Asymptotic distribution of robust estimator for functional Nonparametric Models. *Communications in Statistics: Theory and Methods*, 38, 1317-1335.
- Attouch, M., Laksaci, A., and Messabihi, N., 2017. Nonparametric relative error regression for spatial random variables. *Statistical Papers*, 58, 987-1008.
- Azevedo, C.M. and Oliveira, P.E., 2011. On the kernel estimation of a multivariate distribution function under positive dependence. *Chilean Journal of Statistics*, 2(1), 99-113.
- Buckley, J. and James, I., 1979. Linear regression with censored data. *Biometrika*, 66, 429-436.
- Campbell, M. and Donner, A., 1989. Classification efficiency of multinomial logistic regression relative to ordinal logistic regression. *Journal of the American Statistical Association*, 84, 587-591.
- Carbonez, A., Gyorfi, L., and Edward, C., 1995. Partitioning-estimates of a regression function under random censoring. *Statistics and Decisions*, 13, 21-38.
- Chahad, A., Ait-Hennani, L., and Laksaci, A., 2017. Functional local linear estimate for functional relative-error regression. *Journal of Statistical Theory and Practice*, 11, 771-789.
- Chaouch, M. and Khardani, S., 2015. Randomly censored quantile regression estimation using functional stationary ergodic data. *Journal of Nonparametric Statistics*, 27, 65-87.

- Dabo-Niang, S., 2004. Kernel density estimator in an infinite-dimensional space with a rate of convergence in the case of diffusion process. *Applied Mathematics Letters*, 17, 381–386.
- Deheuvels, P. and Einmahl, J. H., 2000. Functional limit laws for the increments of Kaplan Meier product-limit processes and applications. *The Annals of Probability*, 28, 1301–1335.
- Demongeot, J., Hamie, A., Laksaci, A., and Rachdi, M., 2016. Relative-error prediction in nonparametric functional statistics: Theory and practice. *Journal of Multivariate Analysis*, 146, 261–268.
- Doukhan, P., Massart, P., and Rio, E., 1994. The functional central limit theorem for strongly mixing processes. *Annales de l’IHP Probabilités et Statistiques*, 30, 63–82.
- Ferraty, F. and Vieu, P., 2003. Curves discrimination: a nonparametric functional approach. *Computational Statistics and Data Analysis*, 44, 161–173.
- Ferraty, F. and Vieu, P., 2004. Nonparametric methods for functional data, with applications in regression, time-series prediction and curves discrimination. *Journal of Nonparametric Statistics*, 16, 111–125.
- Ferraty, F. and Vieu, P., 2006. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York, US.
- Ferraty, F., Goia, A., and Vieu, P., 2002. Functional nonparametric model for time series: a fractal approach for dimension reduction. *TEST*, 11, 317–344.
- Ferraty, F., Mas, A., and Vieu, P., 2007. Nonparametric regression on functional data: Inference and practical aspects. *Australian and New Zealand Journal of Statistics*, 49, 267–286.
- Fetitah, O., Almanjahie, I., Attouch, M., and Righi, A., 2020. Strong convergence of the functional nonparametric relative error regression estimator under right censoring. *Mathematica Slovaca*, 70, 1469–1490.
- Hellal, N. and Ould-Saïd, E., 2016. Kernel conditional quantile estimator under left truncation for functional regressors, *Opuscula Mathematica*, 36, 25–48.
- Horrigue, W. and Ould-Saïd, E., 2014. Nonparametric regression quantile estimation for dependent functional data under random censorship: Asymptotic normality. *Communications in Statistics: Theory and Methods*, 44, 4307–4332.
- Jones, M., Park, H., Shin, K.I., Vines, S., and Jeong, S.O., 2008. Relative error prediction via kernel regression smoothers. *Journal of Statistical Planning and Inference*, 138, 2887–2898.
- Khardani, S. and Thiam, B., 2016. Strong consistency result of a non parametric conditional mode estimator under random censorship for functional regressors. *Communications in Statistics: Theory and Methods*, 45, 1863–1875.
- Masry, E., 1986. Recursive probability density estimation for weakly dependent stationary processes. *IEEE Transactions on Information Theory*, 32, 254–267.
- Masry, E., 2005. Nonparametric regression estimation for dependent functional data: Asymptotic normality. *Stochastic Processes and their Applications*, 115, 155–177.
- Mebout, M., Attouch, M., and Fetitah, O., 2020. Nonparametric M-regression with Scale Parameter For Functional Dependent Data. *Applications and Applied Mathematics*, 15, 846–874.
- Mechab, W. and Laksaci, A., 2016. Nonparametric relative regression for associated random variables. *Metron*, 74, 75–97.
- Ould-Saïd, E. and Guessoum, Z., 2008. On nonparametric estimation of the regression function under random censorship model. *Statistics and Decisions*, 26, 159–177.
- Pîrjan, A., Oprea, S.V., Cărutasu, G., Petrosanu, D.M., Bâra, A., and Coculescu, C., 2017. Devising hourly forecasting solutions regarding electricity consumption in the case of commercial center type consumers. *Energies*, 10, 17–27.
- Park, H. and Stefanski, L., 1998. Relative-error prediction. *Statistics and Probability Letters*, 40, 227–236.

- Ramsay, J. and Dalzell, C., 1991. Some tools for functional data analysis. *Journal of the Royal Statistical Society B*, 53, 539–561.
- Ramsay, J. and Silverman, B., 2005. Principal components analysis for functional data. *Functional Data Analysis*, 147–172.
- Rio, E., 1999. *Théorie asymptotique des processus aléatoires faiblement dépendants*. Springer, New York, US.
- Stute, W., 1993. Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis*, 45, 89–103.

STATISTICAL QUALITY CONTROL
RESEARCH PAPER

Robust Hotelling T^2 control chart using adaptive reweighted minimum covariance determinant estimator

ESRA POLAT^{1,*}

¹Department of Statistics, Hacettepe University, Ankara, Turkey

(Received: 13 November 2020 · Accepted in final form: 14 May 2021)

Abstract

The classical Hotelling T^2 control chart using classical mean and covariance estimators is not efficient in case of outliers existence in data. To overcome this issue, robust mean and covariance estimators are used in literature. Hence, a robust Hotelling T^2 control chart is proposed based on the adaptive reweighted minimum covariance determinant estimator which is a good option to the classical multivariate T^2 chart in case of outliers presence. The new proposed chart's performance is evaluated by false alarm rates and probability of detection/percentage of outliers detection, later a comparison is made with the performance of the classical Hotelling T^2 chart and the chart obtained using the minimum covariance determinant estimator. Simulation and real data application results are indicated that proposed control chart has better performance in comparison to robust control chart based on the minimum covariance determinant especially in terms of false alarm rates and it performs better than classical chart in terms of probability of detection.

Keywords: Hotelling T^2 · Minimum covariance determinant · Multivariate control chart · Robust estimator · Statistical quality control.

Mathematics Subject Classification: Primary 62P30 · Secondary 62F35.

1. INTRODUCTION

In manufacturing process, multiple quality characteristics of a product are generally observed. Hence, multivariate control charts may be a suitable tool to observe the process. The Hotelling T^2 chart is the most commonly known one because its application is easy, it is flexible, it is sensitive to little process modifications and the software for its application is available. The Hotelling T^2 uses the classical mean and covariance matrix, is reactive to the outlier. Because in case of more quality characteristics are considered, the risk of multiple outliers existence is getting higher. In case of outliers existence, the classical control chart's performance decreases. Because of the masking effect, the classical method is not effective for multiple outliers case (Alfaro and Ortega, 2009). The masking effect in the monitoring process occurs as a result of the outlier, which cannot be detected by the control chart. To overcome the problem that arises, several robust methods have been proposed for reducing

*Corresponding author. Email: espolat@hacettepe.edu.tr

the effect of multiple outliers by substituting the existing estimators with the more robust ones. Furthermore, the performance of Hotelling T^2 control chart, in detecting the shift of mean vector, is increasing when the robust covariance matrix estimator is implemented (Williams et al., 2006; Ahsan et al., 2019).

Similar to control charts observing variability in a process, its structure arises from Phase I, Phase II (Alt, 1985), as well called retrospective and prospective analysis, in order of. The significant point of Phase I is the analysis of historical data for determining if the process is under control or not by estimation of the in-control parameters and control limits of the process. However, in case of Phase II, the focus is to monitor on-line data for rapidly finding shifts of process from the estimated in-control parameter values in Phase I. Outliers in Phase I may cause the increment of control limits and decrease of power for the detection of process changes in Phase II. Hence, Phase II analysis achievement based on a success in Phase I analysis in the estimation of in-control mean, variance and covariance parameters.

Ordinal Hotelling T^2 chart is a safe method when the underlying process data really has the normal distribution. In contrast to this, in case of outliers presence in data it is not a safe method for detecting out of control points properly. Because classical mean and covariance estimators in the original formulae cannot resist the outliers. Thereby, the classical Hotelling T^2 's chart ability for monitoring future process data is debatable. One of the way of getting rid of this issue is using control chart which is robust in case of outliers existence.

Up to now, many robustified versions of the Hotelling T^2 control chart have been proposed by utilizing from robust estimators. Abu-Shawiesh and Abdullah (2001) estimated the mean vector using Hodges-Lehmann and the variance-covariance matrix using Shamos-Bickel-Lehman. Vargas (2003) and Jensen et al. (2007) presented robust control charts using minimum covariance determinant (MCD) and minimum volume ellipsoid (MVE) estimators. They detected and omitted the outliers in Phase I data and later compute the traditional estimators using the remained clean observations in case of Phase II data. Although in this method, the breakdown point and calculation of the estimators become more significant, however, statistical efficiency does not become as critical as the extremely robust estimators change place with by classical estimators in Phase II case. When MVE and MCD are used in Phase I, they recognized some problems, such as T^2 obtained by using MVE performed badly under large sample size. However, T^2 obtained by using MCD requires more sample size if a lot of outlying observations is skeptical to guarantee that MCD estimator loses its ability especially in case of monitoring with higher dimensions (p) and it does not breakdown. Alfaro and Ortega (2008) introduced robust Hotelling T^2 control charts by changing the arithmetic mean with trimmed one and sample covariance with sample trimmed covariance. Chenouri et al. (2009) presented a robust chart based on reweighted MCD (RMCD) estimator that it is not overly affected by outlying observations and has better efficiency than MCD. The difference of their method from Vargas (2003) and Jensen et al. (2007) that they use RMCD estimators instead of traditional estimators in establishing Hotelling T^2 chart for Phase II data set. Alfaro and Ortega (2009) compared the performance of Hotelling T^2 control charts using robust MVE, trimmed, MCD and RMCD estimators. The result of this study was that the recommendation of the use of T^2 charts obtained by using RMCD and trimmed estimator in case of not many outlying observations in the production process since these two methods are able to control false alarm rates (FAR).

In the producing of products that concentrates mostly on determining the outlying observations compared to the false alarms, that is, a point outside the control limits for an in-control process (Da Silva et al, 2019), T^2 obtained by using MCD may be taken into consideration as the best option. Because Hotelling T^2 control charts based on MCD has a better performance in terms of probability of detection (POD). In theory, if the POD gets higher, the chart could also control the overall FAR α (Jensen et al., 2007). In spite of this, the results in Alfaro and Ortega (2009) revealed a discordance between the capability of

robust control chart in controlling the overall FAR and POD in case of specific situations. [Yañez et al. \(2010\)](#) constructed the T^2 control chart by using the biweight S estimators for mean and covariance estimators. Their chart outperformed the T^2 chart based on MVE for a small number of observations. [Yahaya et al. \(2011\)](#) presented the minimum variance vector (MVV) estimator in T^2 chart in order to observe the Phase II data. Overall, the robust control chart gave a quick detection in out-of-control status and at the same time, capable for controlling the overall FARs nevertheless as the p is increased. The only disadvantage was a large upper control limits (UCLs) in comparison to the classical T^2 chart. An improved version of the MVV chart was further suggested by [Ali et al. \(2013\)](#) to obtain desired UCLs whilst still it performs well in terms of FAR and POD. This was achieved by making the MVV estimators consistent at normal distribution as well as unbiased for finite samples. [Ali et al. \(2014\)](#) investigated the performance of reweighted version of MVV (RMVV) in constructing the Hotelling T^2 chart. [Yahaya et al. \(2019\)](#) introduced three robust Hotelling T^2 control charts using trimmed estimators. The modified Mahalanobis distance with median used as the location measure and one of the scale estimators MAD_n , S_n (mean absolute average) or T_n as the scale measure. As a consequence of these alternatives, three dissimilar trimmed estimators are introduced. The findings of their study revealed that their three control charts performance is moderate in terms of false alarms and magnificently for POD, outperform the classical control chart in any case of conditions. In case of outliers existence or samples deviation from normality, all of the studies revealed that the robust control charts surpass the classical Hotelling T^2 control chart.

In this study, following the robust Hotelling T^2 control chart literature, a robust Hotelling T^2 control chart is introduced that uses a robust adaptive reweighted minimum covariance determinant (ARWMCD) estimator. The new control chart's performance is evaluated by FARs and POD by doing a simulation and a real data application. Moreover, the performance of the new method is compared with robust control chart using MCD estimator and the classical chart.

The rest of the paper is organized as follows. Section 2 reviews the ARWMCD estimator. In Section 3, we present the new proposed Hotelling T^2 control chart. Section 4 contains a simulation study where the performance of the new robust Hotelling T^2 chart using ARWMCD estimator is compared to classical Hotelling T^2 chart and the robust Hotelling T^2 chart using MCD. In Section 5, we illustrate the performance of the new proposed robust Hotelling T^2 chart based on ARWMCD on the real data that is given in [Ali et al. \(2013\)](#). Finally, Section 6 collects some conclusions about the present study.

2. ADAPTIVE REWEIGHTED MINIMUM COVARIANCE DETERMINANT ESTIMATOR

In addition to maximum robustness against to outliers, robust multivariate estimators must also propose a sensible efficiency for the normal distribution and a controllable asymptotic distribution. Nevertheless, MCD and MVE estimators do not satisfy that condition. [Gervini \(2003\)](#) expressed that considering the both of being robust and efficient, the best way utilizing a two-stage process. [Rousseeuw and Van Zomeren \(1990\)](#) also expressed that in this process, first of all, a tremendously robust nevertheless maybe not efficient estimator is calculated and it is used for observing outlying observations and calculating the sample location and covariance of the good data. This process comprises of omitting sample points whose Mahalanobis distances go beyond a certainly fixed threshold value. As beginning estimator for that processes, [Rousseeuw and Van Driessen \(1999\)](#) suggested an algorithm for computing MCD estimator, which does not ensure that the precise estimator is obtained, it is quicker and more precise than formerly obtained algorithms also for highly bigger data ($n \gg p$). The advantage of the $1/\sqrt{n}$ convergence rate, in addition to this truth, could indicate that the MCD technique uses the FAST-MCD algorithm is the best preference when compared to MVE for beginning estimator of a two-step process ([Gervini, 2003](#)).

MCD is investigating for those h observations for which the determinant of the traditional covariance matrix is minimum. Therefore, the MCD estimators are the location and covariance matrix of that h observations. The computation of MCD estimation is hard. The application of MCD estimator on data sets could merely be in case of the number of observations exceeds the number of variables ($n > p$). Because in case of $p > n$ then also $p > h$, and often the covariance matrix of any h observations is going to be singular, tends to zero determinant. Henceforth, each subset of h observations would tend to the minimum feasible determinant, resulting in a non-unique solution (Filzmoser et al., 2009). FAST-MCD algorithm can handle with larger sizes of sample such as tens of thousands. This algorithm obtains precise solution for small sizes of data and it is quicker and more precise than formerly proposed algorithms, yet for extremely big data sets. Since it is efficient and fast in calculation, Rousseeuw and Van Driessen (1999) proposed of using FAST-MCD algorithm for estimating mean and covariance. Since the raw MCD estimators of mean and covariance are reweighted for improving the finite sample efficiency, named as reweighted MCD (RMCD) estimators (Hubert and Vanden-Branden, 2003). Since it is very popular algorithm for robust literature, a brief information about FAST-MCD is given. Any interested reader could find for detailed information in Rousseeuw and Van Driessen (1999). The algorithms steps for p dimensional vector x_i , for $i = 1, \dots, n$, as follows.

Step 1: The MCD estimates could withstand $(n - h)$ outlying observations, therefore h (or equally the fraction $\alpha = h/n$) specifies the robustness of the estimator. $(1 - \alpha)$ measures the fraction of outliers the algorithm should resist. Any value between 0.5 and 1 may be specified (default = 0.75). In FAST-MCD algorithm by taking $[(n + p + 1)/2]$ as the accepted value of h , highly resist against outliers. Nevertheless, any integer h in the interval $[(n + p + 1)/2] \leq h < n$ could be used by researcher. In case of a huge fraction of outliers is assumed in data set, thereby, h must be selected as $h = [0.5n]$. Also, if it is correct that the data includes not much than 25% of outliers that is often the condition, a better balance between statistical efficiency and breakdown value is captured by choosing $h = [0.75n]$ (Rousseeuw and Van Driessen, 1999). In this study, we have also used the default value of $h = [0.75n]$.

Step 2: From here on $h < n$ and $p \geq 2$. If n is small (say, $n < 600$) then:

- Repeat (say) 500 times:
 - ✓ Construct an initial h -subset H_1 using method given in Rousseeuw and Van Driessen (1999), that is, starting from a random $(p + 1)$ -subset.
 - ✓ Carry out two C-steps described in Rousseeuw and Van Driessen (1999).
- For the 10 results with lowest $\det(\hat{\Sigma}_3)$:
 - ✓ Conduct C-steps until convergence
- Report the solution $(\hat{\mu}, \hat{\Sigma})$ with lowest $\det(\hat{\Sigma})$.

Step 3: If n is larger (say, $n \geq 600$), then:

- Construct up to five disjoint random subsets of size n_{sub} according to given in Rousseeuw and Van Driessen (1999) (say, subsets of size $n_{\text{sub}} = 300$).
- Inside each subset, repeat $500/5=100$ times:
 - ✓ Construct an initial subset H_1 of size $h_{\text{sub}} = [n_{\text{sub}}(h/n)]$.
 - ✓ Carry out two C-steps, using n_{sub} and h_{sub} .
 - ✓ Keep the 10 best results $(\hat{\mu}_{\text{sub}}, \hat{\Sigma}_{\text{sub}})$.
- Pool the subsets, yielding the merged set (say, of size $n_{\text{merged}} = 1500$).
- In the merged set, repeat for each of the 50 solutions $(\hat{\mu}_{\text{sub}}, \hat{\Sigma}_{\text{sub}})$:
 - ✓ Conduct two C-steps, using n_{merged} and $h_{\text{merged}} = [n_{\text{merged}}(h/n)]$.
 - ✓ Keep the 10 best results $(\hat{\mu}_{\text{merged}}, \hat{\Sigma}_{\text{merged}})$.

- In the full data set, repeat for the m_{full} best results:
 - ✓ Take several C-steps, using n and h .
 - ✓ Keep the best final result $(\hat{\boldsymbol{\mu}}_{\text{full}}, \hat{\boldsymbol{\Sigma}}_{\text{full}})$.

Here, m_{full} and the number of C-steps (preferably, until convergence) depend on how large the data set is (Rousseeuw and Van Driessen, 1999; Polat and Gunay, 2019). This algorithm is called as FAST-MCD. It is affine equivariant: when the data are translated or subjected to a linear transformation, the resulting $(\hat{\boldsymbol{\mu}}_{\text{full}}, \hat{\boldsymbol{\Sigma}}_{\text{full}})$ transforms accordingly. For convenience, the computer program contains two more steps (Rousseeuw and Van Driessen, 1999).

Step 4: In order to obtain consistency when the data come from a multivariate normal distribution, $\hat{\boldsymbol{\mu}}_{\text{MCD}} = \hat{\boldsymbol{\mu}}_{\text{full}}$ and $\hat{\boldsymbol{\Sigma}}_{\text{MCD}} = (\text{med}_i d_{(\hat{\boldsymbol{\mu}}_{\text{full}}, \hat{\boldsymbol{\Sigma}}_{\text{full}})}^2(i) / \chi_{p,0.5}^2) \hat{\boldsymbol{\Sigma}}_{\text{full}}$ are placed.

Step 5: One-step reweighted estimates could be obtained by reweighting each observation as

$$w_i = \begin{cases} 1, & \text{if } (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\text{MCD}})^\top \hat{\boldsymbol{\Sigma}}_{\text{MCD}}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\text{MCD}}) \leq \chi_{p,0.975}^2, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, using the weights w_i , the RMCD estimators are calculated as

$$\hat{\boldsymbol{\mu}}_{\text{RMCD}} = \frac{\sum_{i=1}^n w_i \mathbf{x}_i}{\sum_{i=1}^n w_i} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}_{\text{RMCD}} = \frac{\sum_{i=1}^n w_i (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\text{RMCD}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\text{RMCD}})^\top}{\sum_{i=1}^n w_i}.$$

If it is desirable that the estimator to be robust and efficient, a two-step process is suggested as a best preference. Gervini (2003) suggested basically enhancement above Rousseeuw and Van Zomeren (1990) that a reweighted one-stage estimator using adaptive threshold values. This adaptive reweighting system can keep the outlier robustness of the starting estimator in bias and breakdown, at the same time, reach 100% efficiency for the normal distribution. For the first time, Gervini and Yohai (2002) suggested this type of adaptive reweighting for the linear regression model. This conception is widened by Gervini (2003) that he suggested an adaptive technique for multivariate mean and covariance estimation.

Since x_1, \dots, x_n is a sample of under consideration in \mathfrak{R}^p and beginning robust estimators of mean and covariance are $(\hat{\boldsymbol{\mu}}_{0n}, \hat{\boldsymbol{\Sigma}}_{0n})$ (in our study, they are obtained by MCD estimator using FAST-MCD algorithm) then the Mahalanobis distances are stated as (Gervini, 2003; Polat and Gunay, 2019).

$$d_i := d(\mathbf{x}_i, \hat{\boldsymbol{\mu}}_{0n}, \hat{\boldsymbol{\Sigma}}_{0n}) = \left\{ (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{0n})^\top \hat{\boldsymbol{\Sigma}}_{0n}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{0n}) \right\}^{1/2}.$$

Under normality assumption, d_i^2 nearly have a χ_p^2 distribution and logically, being suspicious about data points with $d_i^2 \geq \chi_{p,0.975}^2$ as an outlier. Rousseeuw and Van Zomeren (1990) suggested to omit those outlying data points and calculated the sample mean and covariance matrix of left of the data set. Hence, by this method, they obtained new estimators $(\hat{\boldsymbol{\mu}}_{1n}, \hat{\boldsymbol{\Sigma}}_{1n})$; see Gervini (2003).

Gervini (2003) expressed that MCD estimators can be taken under consideration as the beginning robust estimators of mean and covariance in the adaptive reweighted procedure because the MCD technique computed using FAST-MCD algorithm is developed as a good option instead of MVE. Therefore, similar as in Polat and Gunay (2019), adaptive reweighted technique including the MCD estimators $(\hat{\boldsymbol{\mu}}_{\text{MCD}}, \hat{\boldsymbol{\Sigma}}_{\text{MCD}})$ is used as beginning robust estimators of mean and covariance $(\hat{\boldsymbol{\mu}}_{0n} = \hat{\boldsymbol{\mu}}_{\text{MCD}}, \hat{\boldsymbol{\Sigma}}_{0n} = \hat{\boldsymbol{\Sigma}}_{\text{MCD}})$. This technique had been named as ARWMCD and robust estimators, denoted as $(\hat{\boldsymbol{\mu}}_{\text{ARWMCD}}, \hat{\boldsymbol{\Sigma}}_{\text{ARWMCD}})$.

This reweighting stage rises up the beginning estimators efficiency and also keeps its robustness mostly. The threshold $\chi_{p,0.975}^2$ is a subjective value. Although they show a normal distribution, in case of big data sets noticeable number of data points must to be omitted out of analysis. For this issue, it is found that the best option constructing an adaptive threshold values, which gets higher related to n in case of the data are uncontaminated, however, stays bounded in case of outliers presence in the sample. The procedure of this method is as in follows. Note that the expression stated as

$$G_n(u) = \frac{1}{n} \sum_{i=1}^n I\left(d^2(\mathbf{x}_i, \hat{\boldsymbol{\mu}}_{\text{MCD}}, \hat{\boldsymbol{\Sigma}}_{\text{MCD}}) \leq u\right),$$

where $G_p(u)$ is the χ_p^2 distribution function, shows the experimental distribution of the squared Mahalanobis distances (Gervini, 2003; Polat and Gunay, 2019).

The approximation of G_n to G_p is assumed in case of the sample has normal distribution. Hence, comparing the tails of G_n with the tails of G_p is a technique of detection for outliers. In case of $\eta = \chi_{p,1-\alpha}^2$ for a fixed small α , for example $\alpha = 0.025$, we have (Gervini, 2003; Polat and Gunay, 2019)

$$\alpha_n = \sup_{u \geq \eta} \{G_p(u) - G_n(u)\}^+, \quad (1)$$

where $\{\cdot\}^+$ denotes the positive part. Note that α_n could be considered as an outlier measurement in the sample. It only allows positive differences in Equation (1) because a negative difference does not show existence of outliers. If $d_{(i)}^2$ shows the i th order statistic of the squared Mahalanobis distances and $i_0 = \max\{i: d_{(i)}^2 < \eta\}$, then Equation (1) comes down to as (Gervini, 2003; Polat and Gunay, 2019)

$$\alpha_n = \max_{i > i_0} \left\{ G_p(d_{(i)}^2) - \frac{i-1}{n} \right\}^+.$$

Those data points giving the largest $[\alpha_n n]$ distances are taken under consideration as outlying points and omitted in the reweighting stage, with $[a]$ showing the largest integer that is $\leq a$. The cut-off value is given by

$$c_n = G_n^{-1}(1 - \alpha_n),$$

where $G_n^{-1}(u) = \min\{s: G_n(s) \geq u\}$, $c_n = d_{(i_n)}^2$, with $i_n = n - [\alpha_n n]$ and that $i_n > i_0$ as a outcome of the description of α_n . Therefore, $c_n > \eta$. To describe the reweighted estimator, weights are stated as (Gervini and Yohai, 2002; Polat and Gunay, 2019)

$$w_{i,n} = w\left(\frac{d^2(\mathbf{x}_i, \hat{\boldsymbol{\mu}}_{\text{MCD}}, \hat{\boldsymbol{\Sigma}}_{\text{MCD}})}{c_n}\right). \quad (2)$$

The weight function defined as $w: [0, \infty) \rightarrow [0, 1]$ is non-increasing, with $w(u) = 0$ when $u \in [1, \infty)$ and $w(u) > 0$ when $u \in [0, 1)$, $w(0) = 1$. The simplest choice among those functions satisfying it is the hard-rejection function $w(u) = I(u < 1)$, which is the one most commonly used in practice.

Once the weights in Equation (2) are calculated, the one-stage reweighted $(\hat{\boldsymbol{\mu}}_{\text{ARWMD}}, \hat{\boldsymbol{\Sigma}}_{\text{ARWMD}})$ are given as

$$\hat{\boldsymbol{\mu}}_{\text{ARWMD}} = \frac{\sum_{i=1}^n w_{i,n} \mathbf{x}_i}{\sum_{i=1}^n w_{i,n}} \tag{3}$$

and

$$\hat{\boldsymbol{\Sigma}}_{\text{ARWMD}} = \frac{\sum_{i=1}^n w_{i,n} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\text{ARWMD}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\text{ARWMD}})^\top}{\sum_{i=1}^n w_{i,n}}. \tag{4}$$

3. THE NEW PROPOSED HOTELLING T^2 CONTROL CHART

The p dimensional random sample of n observations of prior data in case of Phase I is shown by $\mathbf{x}_i = \{x_1, \dots, x_n\}$, where \mathbf{x}_i are supposed to be independent and have a multivariate normal distribution with covariance matrix $\boldsymbol{\Sigma}$ and mean vector $\boldsymbol{\mu}$. In case of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are not known then the estimation of them utilizing an in-control data set is needed. The procedure of describing the in-control data set from \mathbf{x}_i is mentioned as Phase I action. Using preliminary data set, $\bar{\mathbf{x}}$ and \mathbf{S} are computed. These estimates are used to compute $T_{(i)}^2$ for $i = 1, \dots, n$ based on

$$T_{(i)}^2 = (\mathbf{x}_i - \bar{\mathbf{x}}) \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})^\top,$$

To obtain an in-control data set, describe outliers utilizing UCL established on the beta distribution given by

$$\text{UCL}_1 \sim \left[\frac{(n-1)^2}{n} \right] B_{(\alpha, \frac{p}{2}, \frac{n-p-1}{2})},$$

where $B(\alpha, p/2, (n-p-1)/2)$ is the $100 \times (1-\alpha)\%$ quantile of the beta distribution with $p/2$ and $(n-p-1)/2$ degrees of freedom, whereas α is the overall FAR.

The sample points where $T_{(i)}^2 > \text{UCL}_1$ are omitted that since they are outliers. The clean data set that the outlying observations are omitted (n_c) is then used for computing the new estimations, $\bar{\mathbf{x}}_N$ and \mathbf{S}_N . These estimations are used for computing $T_{(g)}^2$ statistic for Phase II observation, where $x_g \notin \mathbf{x}_i$, such that

$$T_{(g)}^2 = (\mathbf{x}_g - \bar{\mathbf{x}}_N) \mathbf{S}_N^{-1} (\mathbf{x}_g - \bar{\mathbf{x}}_N)^\top.$$

By using the desired values of α , p and n_c , compute the LCL and UCL using the F distribution as

$$\text{UCL} \sim \left[\frac{p(n_c+1)(n_c-1)}{n_c(n_c-p)} \right] F_{(\alpha, p, n_c-p)} \quad \text{and} \quad \text{LCL} = 0,$$

where $F_{(\alpha, p, n_c-p)}$ is the $100(1-\alpha)\%$ quantile of the F distribution with p and $n-p$ degrees of freedom and α is the overall FAR. Nevertheless, this classical procedure is merely effective in excluding very unusual outlying observations and observing large shift in the mean vector in small sample sizes, however, it is not successful for detecting more moderate outlying observations specifically when number of variables inflated (Vargas, 2003; Jensen et al., 2007; Chenouri et al., 2009). To overcome this issue of the process, in this study, ARWMD

estimator is used in Phase I data of \mathbf{x}_i . As it is known that ARWMCD gives robust estimators of covariance and mean, those are used as in-control estimators in Phase II, where the Phase II observations are $\mathbf{x}_g = \{x_{n+1}, x_{n+2}, \dots\}$, $\mathbf{x}_g \notin \mathbf{x}_i$.

The procedure for new robust chart is as follows. First, from the Phase I data set, \mathbf{x}_i , the ARWMCD location vector and covariance matrix estimators $\bar{\mathbf{x}}_{\text{ARWMCD}}(\hat{\boldsymbol{\mu}}_{\text{ARWMCD}})$ and $S_{\text{ARWMCD}}(\hat{\boldsymbol{\Sigma}}_{\text{ARWMCD}})$ are obtained as in Equations (3) and (4). Then, a robust Hotelling T^2 ($T_{\text{ARWMCD}(g)}^2$) for Phase II data, x_g , is defined based on these ARWMCD estimates (obtained from Phase I data) as

$$T_{\text{ARWMCD}(g)}^2 = (\mathbf{x}_g - \bar{\mathbf{x}}_{\text{ARWMCD}}) \mathbf{S}_{\text{ARWMCD}}^{-1} (\mathbf{x}_g - \bar{\mathbf{x}}_{\text{ARWMCD}})^\top. \quad (5)$$

The UCL, FAR and POD calculations are explained under Section 4 in detail.

4. SIMULATION STUDY

Robust estimators are used in place of the traditional mean and covariance in T^2 chart, which causes the replacing of distributional properties of the classical T^2 control chart (Williams et al., 2006). As the sampling distribution of the suggested Hotelling T^2 chart T_{ARWMCD}^2 is not known, the UCL is estimated with simulation. Moreover, as the distribution of T_{ARWMCD}^2 is not known, simulations were done for estimating the quantiles of the T_{ARWMCD}^2 , for few combines of dimensions and sample sizes as shown in Table 1. Even, the finite sample distribution of the MCD estimators is still questionable, thus, the distribution of T_{MCD}^2 is also unknown (Vargas, 2003; Jensen et al., 2007; Chenouri et al., 2009; Alfaro and Ortega, 2009). Therefore, quantile is also used in estimating the distribution of T_{MCD}^2 obtained via Monte Carlo method.

First of all, data sets are originated from the standard multivariate normal distribution $N_p(\mathbf{0}, \mathbf{I}_p)$. Then, robust estimators are computed from this distribution. Next, a new additional sample point from the standard multivariate normal distribution is generated and robust Hotelling T^2 statistic for this new sample point is computed. This process is repeated 5000 times and the 95th percentile of the 5000 robust Hotelling T^2 statistics considered as the UCL. For assessing the performance of T_{ARWMCD}^2 by comparison with classical T_0^2 and robust T_{MCD}^2 control charts, several conditions are generated by changing number of dimensions (p), observations (n) and percentage of outliers (ε) and a variety of mean shifts values ($\boldsymbol{\mu}_1$) as shown in Table 1.

Table 1. The Simulation settings

Variables	Values
Number of quality characteristics (p)	2, 5, 10
Proportion of contamination (ε)	0.1, 0.2
Mean shift ($\boldsymbol{\mu}_1$)	0 (no shift), 3, 5
Group size (n)	50, 100, 200

To estimate the 95% quantile of $T_{\text{ARWMCD}(g)}^2$ firstly, for a Phase I case with a sample size n and dimension p , $K = 5000$ samples of size n from a standard multivariate normal distribution $N_p(0, \mathbf{I}_p)$ are generated. For different sample sizes n , the ARWMCD mean vector and covariance matrix estimates are computed, $\boldsymbol{\mu}_{\text{ARWMCD}}(k)$ and $\mathbf{S}_{\text{ARWMCD}}(k)$, for $k = 1, \dots, K$. Additionally, for each data set, a new observation $X_{g,k}$ is randomly generated that it is handled as a Phase II sample point from $N_p(\mathbf{0}, \mathbf{I}_p)$ and the corresponding $T_{\text{ARWMCD}(g,k)}^2$ values are computed as given by Equation (5). The simulated values as

$T_{ARWMCD(g,1)}^2, \dots, T_{ARWMCD(g,K)}^2$ are used to obtain the empirical distribution function of $T_{ARWMCD(g)}^2$. Then, $T_{ARWMCD(g,K)}^2$ values are sorted in ascending order and the UCL is the 95th quantile of the 5000 statistics. The UCL values for classical and MCD control charts are also estimated by using this technique.

4.1 PERFORMANCE EVALUATION

The classical and two Robust Hotelling T^2 charts success is evaluated in terms of the FAR and POD for Phase II data. Hence, for Phase II sample points, 1000 new datasets were simulated from the standard normal distribution $N_p(\mathbf{0}, \mathbf{I}_p)$ of various sample sizes (n) and dimensions (p) as shown in Table 1. For deciding the FAR and POD, a Phase II sample point is randomly generated with in-control and out of control parameters respectively from Phase I and the robust Hotelling T^2 statistics are calculated. FAR is calculated using a new sample point from the in-control distribution, however, the POD is computed with a new sample point generated from the out-of-control distribution. The FAR or POD is predicted as the percentages of statistic values which are over the control limits of 1000 repetitions. In case of Phase I, several conditions of data sets are simulated by changing the number of observations, dimensions and proportions of contamination. By mixing normal distributions similar as in Alfaro and Ortega (2009), a contaminated model stated as

$$(1 - \varepsilon)N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) + \varepsilon N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \quad (6)$$

is used for investigating the effect of outliers on the charts success. Here, ε is the percentage of outlying observations, $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are the in-control parameters, however, $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$ are the out-of-control parameters. Contamination with shift in the mean, however, not any changes in covariance is assumed, henceforth, the covariance matrix $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ in Equation (6) are p dimensional identity matrices (\mathbf{I}_p). Four variables are changed to investigate the strengths and the weaknesses of the classical and robust Hotelling T^2 charts namely number of quality characteristics (p), proportion of contamination (ε), mean shift ($\boldsymbol{\mu}_1$) and sample size (n). The proportions of outliers as 0.1 or 0.2 and also the clean data set is taken under consideration. As for the POD a modification which is based on the shift in the mean vector $\boldsymbol{\mu}_1$ is a vector of size with value of 0 (in case of not any difference), 3 or 5 (in case of good leverage points) are considered. The setting values for the variables are listed in Table 1 following Alfaro and Ortega (2008), Vargas (2003) and Mohammadi et al. (2011). Changes on the mean shifts and proportions of outlying observations produce 5 dissimilar kinds of contaminated distributions stated as:

- $N_p(0, \mathbf{I}_p)$ – ideal case (clean data set);
- $(0.9)N_p(0, \mathbf{I}_p) + (0.1)N_p(3, \mathbf{I}_p)$ – slight contamination;
- $(0.8)N_p(0, \mathbf{I}_p) + (0.2)N_p(3, \mathbf{I}_p)$ – medium contamination;
- $(0.9)N_p(0, \mathbf{I}_p) + (0.1)N_p(5, \mathbf{I}_p)$ – medium contamination;
- $(0.8)N_p(0, \mathbf{I}_p) + (0.2)N_p(5, \mathbf{I}_p)$ – excessive contamination.

Later, in Phase II, the data are simulated from multivariate normal distribution $N_p(\boldsymbol{\mu}_1, \mathbf{I}_p)$, where $\boldsymbol{\mu}_1$ shows the shift in the mean vector such as the case in Phase I (that is, 0, 3, and 5). Then, the new control chart T_{ARWMCD}^2 is compared with robust Hotelling T^2 chart based on MCD (T_{MCD}^2) and the classical Hotelling T^2 control chart. For the classical chart T_0^2 , the method, which is without cleaning the outlying observations as stated in Alfaro and Ortega (2009), is considered. The programs and simulations were done using MATLAB. The FAST-MCD algorithm code named as `mcdcov` could be found in MATLAB LIBRA Toolbox (Verboven and Hubert, 2005). The features of computer used for simulation is Intel(R) Core(TM) i5-8250U CPU @ 1.60 GHz 1.80 Ghz.

4.2 SIMULATION RESULTS

Here, the results of performance of the classical T_0^2 and robust T_{MCD}^2, T_{ARWMCD}^2 charts are presented in terms of FARs and POD at $\alpha = 0.05$ in Tables 2 and 3.

4.2.1 FALSE ALARM RATES

The success of a chart cannot only be evaluated by its capability in diagnosing outliers, however, also in controlling the FAR, which is the probability of out-of-control signal in case of a process is under control. In case of the process instability, the value gets larger because of increment in variability. Expanded FAR could cause unrequired process regulations and loss of confidence in the control chart as an observing instrument (Chang and Bai, 2004). Therefore, a technique that could control the FAR to the wished level is essential. The Bradley liberal criterion of robustness is used for evaluating the robustness of the control charts. According to this criterion, a control chart is evaluated as robust in case of its empirical FAR is within the robust interval between 0.5α and 1.5α (Bradley, 1978). Henceforth, as the nominal value is accepted as $\alpha = 0.05$, the control chart is taken under consideration as robust if its FAR is within robust interval, 0.025 to 0.075. In Table 2, the FAR values lying within the robustness interval are bolded. A control chart, which is considered as best, the one has the ability of controlling the FARs within robust interval and also gives the closest FAR to nominal value, 0.05 (Jamaluddin et al., 2018). For every condition, the FARs given in Table 2 are presented in an ascending number of dimensions such as $p = 2, 5$ and 10, with $\alpha = 0.05$. The sample sizes are given in the first column of this table, in second column the proportions of outliers and in third column non-centrality values are provided.

Table 2. FAR values (%) of the three control charts in case of $\alpha = 0.05$.

n	ε	μ_1	$p = 2$			$p = 5$			$p = 10$			
			T_0^2	T_{MCD}^2	T_{ARWMCD}^2	T_0^2	T_{MCD}^2	T_{ARWMCD}^2	T_0^2	T_{MCD}^2	T_{ARWMCD}^2	
50	0	0	5.6	5.8	5.5	5.5	6.0	5.5	5.7	4.9	5.2	
		0.1	3	2.1	2.2	3.6	2.8	1.3	3.0	4.1	1.9	2.2
		5	1.6	2.3	3.8	2.6	1.4	3.1	4.0	1.8	2.0	
	0.2	3	2.1	0.8	2.2	2.8	0.5	1.8	4.3	1.9	2.2	
		5	1.9	0.6	3.1	2.7	0.2	1.6	4.2	0.8	1.1	
		5	4.7	4.5	4.2	5.0	3.3	4.2	5.1	4.3	5.3	
100	0	0	1.9	2.0	3.5	2.7	1.3	3.3	3.3	2.3	4.0	
		0.1	3	1.5	2.0	3.8	2.7	1.4	3.2	3.3	2.2	3.8
		5	2.0	0.5	2.3	2.8	0.2	3.1	3.1	0.8	3.4	
	0.2	3	1.5	0.4	3.3	2.7	0.2	3.1	3.2	0.5	3.2	
		5	6.3	5.7	6.0	4.2	4.1	4.4	5.0	5.6	5.5	
		5	2.3	3.1	4.9	2.5	2.1	3.4	3.8	2.3	4.1	
200	0	0	1.9	3.1	5.6	2.4	2.0	3.4	3.5	2.1	4.0	
		0.1	3	2.1	0.5	3.1	2.7	0.1	3.8	3.7	0.2	3.4
		5	1.9	0.2	5.2	2.7	0.1	3.9	3.6	0.2	3.3	

Table 2 shows robust T_{MCD}^2 and T_{ARWMCD}^2 charts that perform as good as the traditional chart in controlling FAR under ideal condition ($\varepsilon = 0, \mu_1 = 0$), regardless of the sample sizes n , outliers proportions ε and variable sizes p . However, the rates for all charts decrease when contamination exists with some results below the Bradley limit.

If $p = 2$, it is clear in Table 2 all results on FARs demonstrate that the T_{ARWMCD}^2 control chart has better performance than the T_0^2 and T_{MCD}^2 control charts. The T_{ARWMCD}^2 control chart has the capability in controlling FARs for nearly whole of the circumstances explored which is about 86% (13 out of 15) of the circumstances in comparison to T_0^2 and T_{MCD}^2 control charts, which are only effective for 20% (3 out of 15) and 33% (5 out of 15) of the circumstances, respectively. The T_{MCD}^2 control chart is affected bad with high proportion of outlying observations, $\varepsilon = 20\%$ for both moderate and high process mean shifts, that they are confirmed by the proportions of false alarm far below the significance value, $\alpha = 0.05$. Henceforth, T_{ARWMCD}^2 control chart performs superior to the traditional chart T_0^2 and robust control chart T_{MCD}^2 for bivariate case. In case of the dimensions raised to multivariate data, $p = 5$, the FARs for traditional chart (T_0^2) improved. In contrast, the FARs for T_{MCD}^2 chart

worsen with values as small as 0.001. In case of $p = 5$, the T_{ARWMCD}^2 control chart is still maintains its good performance that it is still effective for 86 % (13 out of 15) of conditions as compared to the T_{MCD}^2 control chart which is merely effective for 20 % (3 out of 15) of the conditions. The results of the FARs for the multivariate case of $p = 10$ shows that the T_{ARWMCD}^2 control chart has still a good performance in controlling FARs as it is effective for 73% (11 out of 15) of conditions. Nevertheless, the performance of T_{ARWMCD}^2 control chart diminishes in case of multivariate data in comparison with to bivariate data, where it capable in controlling FARs for only 11 simulated conditions as compared to 13 simulated conditions. An interesting result for T_0^2 , it is effective 93% (14 out of 15) of conditions for $p = 5$ and 100% (all of 15) of conditions for $p = 10$. T_{MCD}^2 control chart which is still merely effective for 20% (3 out of 15) of the conditions that means T_{MCD}^2 chart performs badly in controlling FAR in all cases.

4.2.2 PROBABILITY DETECTION OF OUTLIERS

The performance in terms of POD is recorded in Table 3. The results are also presented as graphs for a better visual and comparison, the values in Table 3 are translated into Figures 1, 2 and 3 based on the values of p . For each case, the performance of the control chart is considered as better in detecting changes in case of the probabilities value is nearer to 1.

Table 3. Percentage of detecting outliers at $\alpha = 0.05$.

n	ε	μ_1	$p = 2$			$p = 5$			$p = 10$		
			T_0^2	T_{MCD}^2	T_{ARWMCD}^2	T_0^2	T_{MCD}^2	T_{ARWMCD}^2	T_0^2	T_{MCD}^2	T_{ARWMCD}^2
50	0	0	5.6	5.8	5.5	5.5	6	5.5	5.7	4.9	5.2
		0.1	49.8	88.2	91.7	37.8	98.6	99.8	25	100	100
		5	74.8	100	100	46.4	100	100	26.7	100	100
	0.2	3	17.6	69.3	75.4	12.5	95.1	98.1	11	55.2	58.7
		5	17.2	100	100	11.8	100	100	10.9	84.2	85.0
		5	17.2	100	100	11.8	100	100	10.9	84.2	85.0
100	0	0	4.7	4.5	4.2	5	3.3	4.2	5.1	4.3	5.3
		0.1	49.6	90.3	92.3	38.8	99.8	100	28.1	100	100
		5	76.9	100	100	46.5	100	100	29.4	100	100
	0.2	3	16.4	72.8	73.0	10.5	98.6	99.9	10.2	84.4	84.9
		5	15.7	100	100	10.1	100	100	10.2	95	94.9
		5	15.7	100	100	10.1	100	100	10.2	95	94.9
200	0	0	6.3	5.7	6.0	4.20	4.1	4.4	5	5.6	5.5
		0.1	57.1	93.5	94.6	45.5	100	100	36.7	100	100
		5	84.6	100	100	54.6	100	100	39.7	100	100
	0.2	3	20.1	80.3	79.7	12.3	99.5	100	11.6	98.6	98.6
		5	21.2	100	100	12.6	100	100	12.2	99.9	99.9
		5	21.2	100	100	12.6	100	100	12.2	99.9	99.9

Once the values of p and n increased, that could be obviously seen in Figures 2 and 3, the line representing T_{ARWMCD}^2 consistently at the top location in the plots with the probability value of almost 1 and overlapping with T_{MCD}^2 line under most of the situations. Overall, the robust T_{ARWMCD}^2 and T_{MCD}^2 control charts steadily succeeded in high probability in diagnosing outlying observations. It is obvious that the line represents traditional T_0^2 charts is always at the lowest, producing a very large space between the other two lines (T_{ARWMCD}^2 and T_{MCD}^2). Across Figures 1-3, it is observed that for all of the conditions, the robust charts outperform the traditional chart by a large difference. The robust T_{ARWMCD}^2 chart under most conditions achieved the 100% detection with the lowest rate of 58.7% while the lowest rate for the robust T_{MCD}^2 chart is 55.2% and for traditional chart is 10.1%. Across different dimensions (p), there is no clear pattern of changes in performance among the charts. Generally, the robust charts as well as traditional chart show decrease in POD when ε increases, however, especially in case of the shift is $\mu_1 = 5$ and dimensions $p = 2$ or $p = 5$, POD values do not differ than the value of 100% for robust charts. The shift in mean (μ_1) shows positive effect on the POD performances of two robust charts regardless of the proportion of contamination (ε). However, for the traditional chart, positive effect only occurs when $\varepsilon = 0.1$. Moreover, the increase in sample sizes (n) brings some positive effect on the POD values for all the charts in some of the conditions.

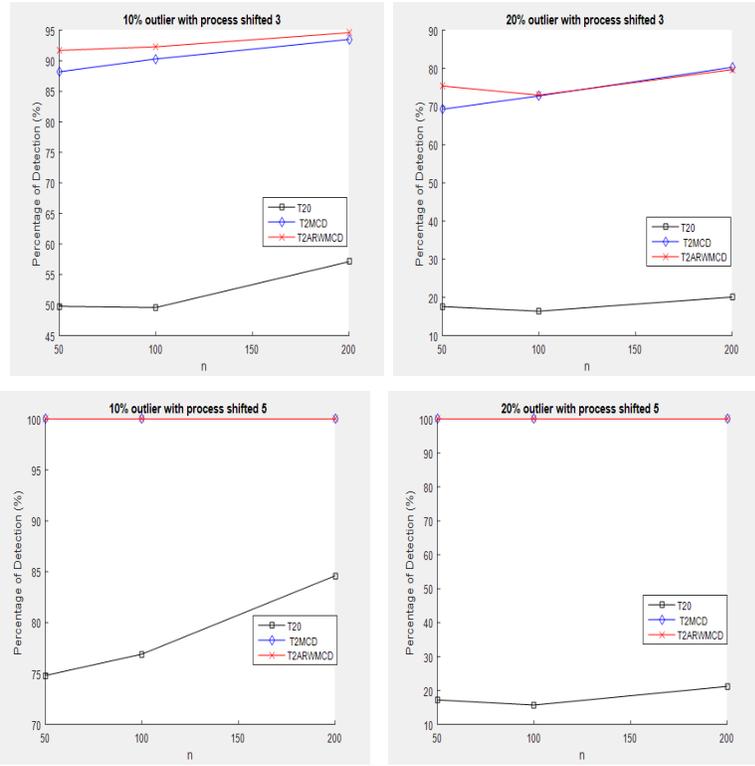


Figure 1. Percentages detection of outliers at $p = 2$.

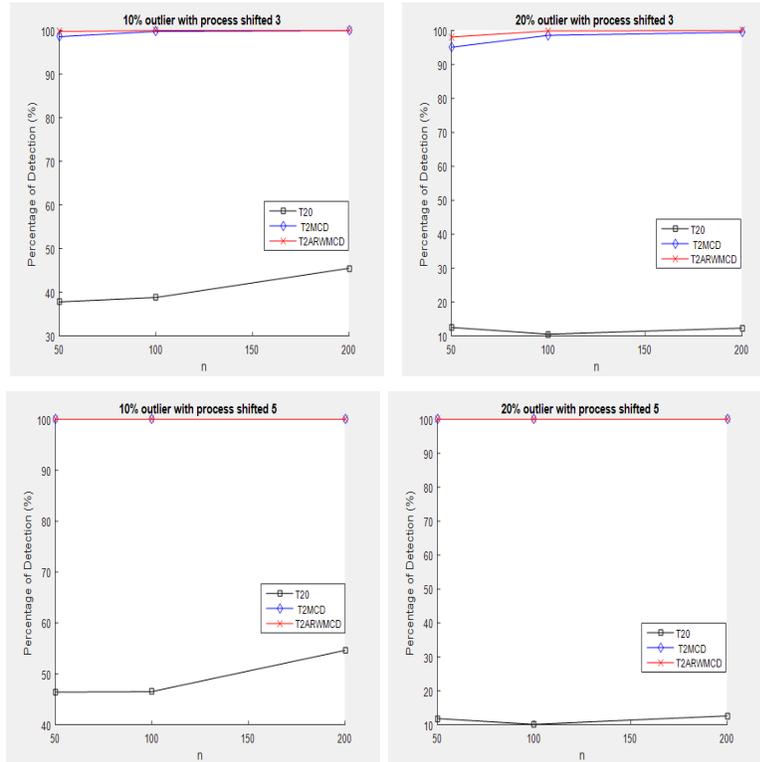


Figure 2. Percentages detection of outliers at $p = 5$.

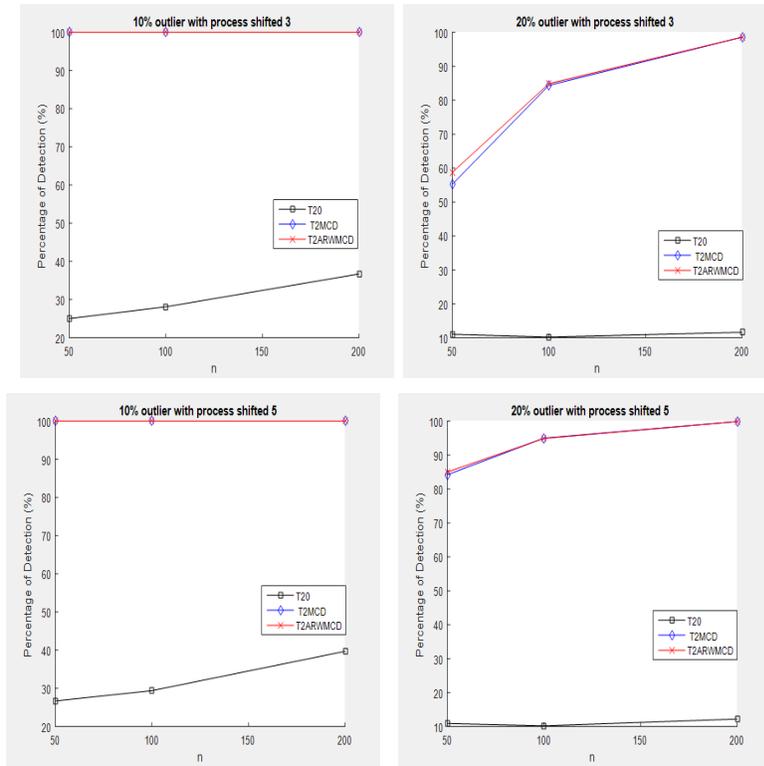


Figure 3. Percentages detection of outliers at $p = 10$.

5. REAL DATA ANALYSIS

The proposed robust control chart T^2_{ARWMCD} is applied on real data given by Asian Composites Manufacturing Sdn. Bhd. (ACM) that includes in the production of advanced composite panels for the aircraft industry. ACM produces flat and contoured primary (Aileron Skins, Spoilers and Spars) and secondary (Flat Panels, Leading Edges and MISC: Components) structure composite bond assemblies and subassemblies for aerospace industries (Ali et al., 2013). For demonstrating the Hotelling T^2 , the company that the part of the production of advanced for the aircraft industry has supplied the data on spoilers has shown in Table 4. The data set is used before in both Yahaya et al. (2011) and Ali et al. (2013) studies. Spoilers are critical instruments in an airplane which of them function is increasing lifts when the airplane is flying. The products are used in civilian, defense and space applications, which could not compromise any mistakes, albeit a minor one. Therefore, careful monitoring is needed to confirm that none of variation appears in the process. Any small error can risk a human life. A sample of 47 products ($n = 47$) that comprises of a few features like as trim edge (X_1), trim edge spar (X_2), and drill hole (X_3) was provided to Yahaya et al. (2011) by the firm. Note that 21 products were gathered in 2009, however, the rest had been gathered in 2010. Hence, they used the 2009 products data as Phase I historical data and they had taken under consideration the products from 2010 as future data. Hence, following these two studies, this data set is used in this study. The historical and future data are given in Tables 4 and 6, respectively. The products comprise of 3 quality variables (dimensions) as mentioned before known as trim edge, trim edge spar, and drill hole. The location vector (\bar{x}) and scatter matrix (S) estimations are given in Table 5. The calculations of the UCLs for $\alpha = 0.05$ based on the estimates are given in the last column of Table 5. The values of the T^2 statistics based on the classic, MCD and ARWMCD estimators are shown in the last three columns of Table 6. The graphical representations of the related control charts are shown in Figure 4.

Table 4. Historical data set (Phase I)

Product	Trim edge (X_1)	Trim edge spar (X_2)	Drill hole (X_3)
1	-0.0011	0.0003	0.0128
2	0.0011	0.0021	0.0246
3	0.0252	0.0308	0.0378
4	-0.0017	0.0109	0.0177
5	-0.0005	-0.0010	0.0106
6	0.0016	-0.0059	0.0128
7	0.0004	0.0001	0.0062
8	0.0078	0.0003	0.0159
9	0.0076	0.0089	0.0097
10	0.0020	0.0005	0.0071
11	0.0108	0.0011	0.0092
12	0.0039	0.0034	0.0425
13	0.0060	-0.0033	0.0160
14	0.0066	0.0100	0.0056
15	0.0045	-0.0067	0.0147
16	0.0110	-0.0207	0.0337
17	0.0047	0.0059	0.0065
18	0.0077	0.0003	0.0191
19	0.0015	0.0123	0.0124
20	0.0011	0.0038	0.0104
21	0.0056	0.0065	0.0063

Table 5. The location vector, covariance matrix and UCL estimations for real data

Types of control chart	Location vector (\bar{x})	Covariance matrix (S)	UCL
T_0^2	[0.00504 0.00284 0.01579]	$\begin{bmatrix} 0.000040 & 0.000200 & 0.000030 \\ 0.000020 & 0.000090 & 0.000010 \\ 0.000030 & 0.000010 & 0.000110 \end{bmatrix}$	11.035
T_{MCD}^2	[0.00414 0.00207 0.01096]	$\begin{bmatrix} 0.000022 & 0.000005 & 0.000004 \\ 0.000005 & 0.000053 & -0.000019 \\ 0.000004 & -0.000019 & 0.000030 \end{bmatrix}$	12.5435
T_{ARWMCD}^2	[0.00414 0.00207 0.01096]	$\begin{bmatrix} 0.000012 & 0.000003 & 0.000002 \\ 0.000003 & 0.000028 & -0.000010 \\ 0.000002 & -0.000010 & 0.000016 \end{bmatrix}$	13.0304

Table 6. The Hotelling T^2 values for the future (Phase II) data

Product	Trim edge	Trim edge spar	Drill hole	T_0^2	T_{MCD}^2	T_{ARWMCD}^2
1	0.0041	0.0087	0.0129	0.5582	1.76591	3.32673
2	0.0047	0.0109	0.0124	0.90026	2.46944	4.65208
3	0.0031	0.0057	0.0096	0.49916	0.34367	0.64743
4	0.0035	-0.0020	0.0101	0.54633	0.54563	1.02789
5	0.004	-0.0028	0.0125	0.45922	0.45797	0.86276
6	0.0031	0.0008	0.0061	0.90130	1.25274	2.35998
7	-0.0019	0.0101	0.0112	3.09329	4.44043	8.36515
8	0.0009	0.0039	0.0082	0.80608	0.68370	1.28799
9	-0.0052	0.0090	0.0203	7.36021	14.97663	28.2139
10	-0.0008	0.0110	0.0184	3.61976	9.74168	18.3520
11	-0.0021	0.0139	0.0170	5.38392	11.87166	22.3645
12	-0.0017	0.0092	0.0061	2.73870	2.97882	5.61168
13	-0.0010	0.0133	0.0138	3.80577	7.40398	13.9481
14	-0.003	0.0002	0.0053	2.05480	3.30863	6.23300
15	0.0016	0.0134	0.0151	2.50731	6.80538	12.8204
16	0.0027	0.0086	0.0070	1.19755	1.06789	2.01176
17	0.0004	0.0086	0.0087	1.57979	1.75966	3.31495
18	-0.0036	0.0136	0.0129	5.79103	9.28168	17.4854
19	-0.0028	0.0003	0.0078	1.83044	2.41775	4.55471
20	0.0120	0.0123	0.0768	38.1397	214.923	404.885
21	-0.0015	0.0004	0.0115	1.26507	1.54862	2.9174
22	0.0009	0.0232	0.0202	8.41812	24.6552	46.4468
23	-0.0035	0.0088	0.0107	3.75884	4.87934	9.19198
24	0.0016	0.0061	0.0066	1.06020	0.93200	1.75576
25	-0.0228	-0.0466	0.0231	42.8447	68.63065	129.290
26	0.0037	-0.0038	0.0147	0.4831	0.77959	1.46863

The comparisons of T^2 values in Table 6 with the related control limits in Table 5, it is seen that T_{MCD}^2 signals observations {9, 20, 22, 25} as out-of-control, the T_{ARWMCD}^2 signals observations {9, 10, 11, 13, 18, 20, 22, 25} as out-of-control, however, T_0^2 only signals 20 and 25 as out-of-control observations and it cannot signal other observations. The result for T_0^2 is not surprising as the analysis on the POD for simulated data revealed that T_0^2 is not as effective as the other two robust charts in diagnosing outliers. For a clearer visualization on the performance of the control charts in diagnosing out or control observations, graphical representation of the three control charts are shown in Figure 4.

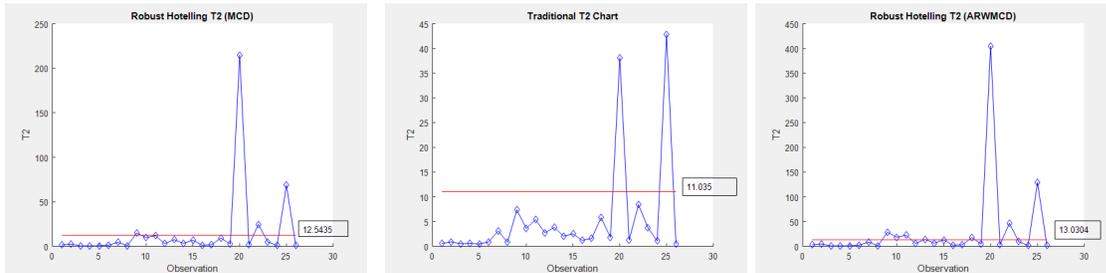


Figure 4. Hotelling T^2 control charts for real data.

6. CONCLUSION

In this study, an alternative to the classical Hotelling T^2 chart was proposed using a robust mean and covariance estimator called as adaptive reweighted minimum covariance determinant. The performance of proposed robust T_{ARWMCD}^2 chart was compared with the robust Hotelling T^2 chart using minimum covariance determinant (T_{MCD}^2) and the classical Hotelling T^2 chart (T_0^2) in terms of false alarm rates and probability of detection.

Simulation results showed that both the robust Hotelling T^2 charts, T_{MCD}^2 and T_{ARWMCD}^2 , provided the best performances in term of probability of detection when $p = 2$, $p = 5$ or $p = 10$. In terms of false alarms, the best performance was detected by the robust T_{ARWMCD}^2 chart when $p = 2$ and T_0^2 chart when $p = 5$ and $p = 10$. Furthermore, the T_{ARWMCD}^2 chart was the second one for these dimensions. Alfaro and Ortega (2009) revealed a confusing result between the probability of detection and the overall false alarm rates such that, for both T_0^2 and T_{MCD}^2 control charts when the probability of detection values increased, the false alarm rates inflated away from the nominal value. This situation was also observed in this study. Even though the classical T_0^2 control chart performed goodly in terms of false alarm rates, particularly when the number of dimensions is getting larger. However, it fails to achieve good probability of detection. In contrast to the T_0^2 chart, the robust Hotelling T_{MCD}^2 control chart performed perfectly in diagnosing outliers, despite that it fails badly in controlling false alarm rates. Nevertheless, the proposed T_{ARWMCD}^2 chart performed so well both in terms of diagnosing outliers and in controlling false alarm rates.

Real data analysis results showed that the proposed robust T_{ARWMCD}^2 control chart showed best performance in terms of diagnosing outliers and the T_{MCD}^2 control chart was the second one. Nonetheless, the classical T_0^2 control chart failed to detect most of the outliers. The real data application results showed consistency with the simulation results.

The overall findings reported that the performance of the robust T_{ARWMCD}^2 control chart in controlling false alarm rates was very good. However, the robust T_{MCD}^2 control charts performance in terms of controlling false alarm rates was not good. Nevertheless, both of these two robust charts were superior to the classical chart in detecting outliers regardless of the conditions imposed in this study. The traditional chart T_0^2 performed moderately in

lower dimension, but better in higher dimensions in controlling false alarm. In contrast, it reported inability to detect outliers. Overall, the proposed $T_{ARW MCD}^2$ control chart showed the best performance since this control chart produced good values for both false alarm rates and probability of detection.

AUTHOR CONTRIBUTIONS Conceptualization, investigation, methodology, software, writing-original draft preparation, writing review and editing, writing-original: E.P. The author has read and agreed to the published version of the manuscript.

ACKNOWLEDGEMENTS The author thanks Editors and Referees for this suggestions which allowed us to improve the presentation of this work.

FUNDING Not applicable

CONFLICTS OF INTEREST The authors declare no conflict of interest.

REFERENCES

- Abu-Shawiesh, M.O.A. and Abdullah, M.B, 2001. A new robust bivariate control chart for location. *Communication in Statistics: Simulation Computation*, 30(3), 513–529.
- Ahsan, M., Mashuri, M., Kuswanto, H., Prastyo, D.D., and Khusna, H., 2019. Multivariate T^2 control chart based on James-Stein and Successive Difference Covariance Matrix Estimators for intrusion detection. *Malaysian Journal of Science*, 38, 23–35.
- Alfaro, J.L. and Ortega, J.F., 2008. A robust alternative to Hotelling's T^2 control chart using trimmed estimators. *Quality and Reliability Engineering International*, 24, 601–611.
- Alfaro, J.L. and Ortega, J.F., 2009. A comparison of robust alternatives to Hotelling's T^2 control chart. *Journal of Applied Statistics*, 36(12), 1385–1396.
- Ali, H., Yahaya, S.S., and Omar, Z., 2013. Robust Hotelling T^2 control chart with consistent minimum vector variance. *Mathematical Problems in Engineering*, 2013, 401350.
- Ali, H., Yahaya, S.S., and Omar, Z., 2014. Robust Hotelling T^2 control chart using reweighted minimum vector variance estimators. *AIP Conference Proceedings*, 1635, 695–702 <https://doi.org/10.1063/1.4903658>.
- Alt, F.B., 1985. Multivariate quality control. In Kotz, S. and Johnson, N. (eds.), *Encyclopedia of Statistical Sciences Vol. 6*, pp. 110–122, Wiley, New York.
- Bradley, J.V., 1978. Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144–152.
- Chang, Y.S. and Bai, D.S., 2004. A multivariate T^2 control chart for skewed populations using weighted standard deviations. *Quality and Reliability Engineering International*, 20, 31–46.
- Chenouri, S., Steiner, S.H., and Mulayath, A., 2009. A multivariate robust control chart for individual observations. *Journal of Quality Technology*, 41(3), 259–271.
- Da Silva, G.V., Taconeli, C.A., Zeviani, W. M., and Do Nascimento, I.A.S., 2019. Performance of Shewhart control charts based on neoteric ranked set sampling to monitor the process mean for normal and non-normal processes. *Chilean Journal of Statistics*, 10(2), 131–154.
- Filzmoser, P., Serneels, S., Maronna, R., and Van Espen, P.J., 2009. Robust multivariate methods In Chemometrics. In Walczak, B., Ferre, R.T. and Brown, S. (eds.), *Comprehensive Chemometrics*, pp. 681–722. Elsevier, New York, US.

- Gervini, D. and Yohai, V.J., 2002. A class of robust and fully efficient regression estimators. *Annals of Statistics*, 30, 583–616.
- Gervini, D., 2003. A robust and efficient adaptive reweighted estimator of multivariate location and scatter. *Journal of Multivariate Analysis*, 84, 116–144.
- Hubert, M. and Vandenberg, K., 2003. Robust methods for partial least squares regression. *Journal of Chemometrics*, 17, 537–549.
- Jamaluddin, F., Ali, H.H., and Yahaya, S.S., 2018. The performance of robust multivariate EWMA control charts. *The Journal of Social Sciences Research*, 6, 52–58.
- Jensen, W.A., Birch, J.B., and Woodall, W.H., 2007. High breakdown estimation methods for phase I multivariate control charts. *Quality and Reliability Engineering International*, 23, 615–629.
- Mohammadi M., Midi H., Arasan J., and Al-Talib B., 2011. High breakdown estimators to robustify phase II control charts. *Applied Sciences*, 11, 503–511.
- Polat, E. and Gunay, S., 2019. A new robust partial least squares regression method based on a robust and an efficient adaptive reweighted estimator of covariance. *REVSTAT*, 17(4), 449–474.
- Rousseeuw, P.J. and Van Zomeren, B.C., 1990. Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85(411), 633–639.
- Rousseeuw, P.J. and Van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41, 212–224.
- Vargas, J.A., 2003. Robust estimation in multivariate control charts for individual observations. *Journal of Quality Technology*, 35(4), 367–376.
- Verboven, S. and Hubert, M., 2005. LIBRA: a MATLAB library for robust analysis. *Chemometrics and Intelligent Laboratory Systems*, 75, 127–136.
- Williams, J.D., Woodall, W.H., Birch, J.B., Sullivan, J. H., 2006. Distribution of Hotelling's T^2 statistic based on the successive differences estimator. *Journal of Quality Technology*, 38(3), 217–229.
- Yahaya, S.S., Ali, H., and Omar, Z., 2011. An alternative hotelling T^2 control chart based on minimum vector variance (MVV). *Modern Applied Science*, 5(4), 132–151.
- Yahaya, S.S., Haddad, F.S., Mahat, N.I., Rahman, A., and Ali, H., 2019. Robust hotelling T^2 charts with median based trimmed estimators. *Journal of Engineering and Applied Sciences*, 14(24), 9632–9638.
- Yañez, S., Gonzalez, N., and Vargas, J.A., 2010. Hotelling's T^2 control charts based on robust estimators. *Dyna*, 77(163), 239–247.

STATISTICAL MODELING
RESEARCH PAPER

Lomax regression model with varying precision: Formulation, estimation, diagnostics, and application

MOIZÉS DA S. MELO^{1,*}, LAÍS H. LOOSE², and JHONNATA B. DE CARVALHO³

¹Department of Statistics, Universidade Federal do Rio Grande do Norte, Natal, Brazil

²Department of Statistics, Universidade Federal de Santa Maria, Santa Maria, Brazil

³Department of Statistics, Universidade Federal do Amazonas, Manaus, Brazil

(Received: 13 July 2021 · Accepted in final form: 23 November 2021)

Abstract

In this paper, we propose a new regression model with varying precision based on the Lomax distribution with regression structures for both the mean and precision parameters. The structures contain unknown parameters, regressors, and a link function. We discuss methods for parameter estimation, hypothesis testing and diagnostic analysis, along with their asymptotic properties. We also provide the expressions for the score vector as well as for the observed and Fisher information matrices. We conduct a Monte Carlo simulation study to investigate the behavior of the estimators and evaluate their finite sample performance. Finally, we present and discuss an empirical application to illustrate the usefulness of the proposed model.

Keywords: Asymmetrical data · Maximum likelihood method · Monte Carlo simulation · Positive data · Reparametrization.

Mathematics Subject Classification: Primary 62J99 · Secondary 62F10.

1. INTRODUCTION

The Lomax distribution, also known as the Pareto type II model, belongs to the class of distributions with decreasing failure rate and was first introduced by [Lomax \(1954\)](#) for modeling business failure data. In the literature, the Lomax distribution has been applied in several fields. For example, [Harris \(1968\)](#) used this distribution for queue problems, [Atkinson and Harrison \(1978\)](#) used it for modeling business failure data, [Holland et al. \(2006\)](#) used it for modeling the distribution of the sizes of computer files on a server, [Corbellini et al. \(2007\)](#) used the Lomax distribution to model firm size distribution, and [Chandra and Khan \(2013\)](#) used it to determine the optimal time for level changes for stress plans in censored samples.

In the context of regression analysis, [Beirlant and Goegebeur \(2003\)](#) presented a regression model for random variables following a Lomax distribution, in which an exponential transformation is used to relate the response variable with covariates. [Stasinopoulos and Rigby \(2007\)](#) developed the package `gamlss` available in the software R ([R Development Core Team, 2021](#)), in which we can model the parameters using a regression structure and

*Corresponding author. Email: moizes.melo@ufrn.br

link functions. This package is an innovative proposal that makes it possible to consider regression structures in a wide range of probability distributions. In the approach presented by [Stasinopoulos and Rigby \(2007\)](#), a regression structure using a link function can be considered for modeling each of the distribution parameters. However, the modeling is not performed in terms of the mean of the distribution. This fact can make the interpretation of the parameters difficult, thus limiting the use of the model in practice. A possible approach to interpretation in terms of the mean is to use the invariance property of the estimators. This result can be applied when using some link functions such logarithmic ([Das et al., 2010](#)) or the square root. However, this is not possible when using the inverse link function.

When working with regression models for continuous positive variables, one possibility for modeling is to use transformations of the response variable. The most commonly transformation is the logarithmic. For example, [Fernández and De Andrade \(2020\)](#) proposed a log-erf-Frechet regression model and [Vigas et al. \(2017\)](#) proposed a regression model in the location-scale form based on the Poisson-Weibull distribution. In both approaches, the logarithm of the variable of interest is modeled. Nonetheless modeling the mean is the most common approach in regression models ([McCullagh and Nelder, 1989](#); [Ferrari and Cribari-Neto, 2004](#); [Fonseca et al., 2016](#); [Palm et al., 2019](#)). Regression models are usually proposed with a focus on constant dispersion or precision parameter. Some extensions for modeling parameters related to the variance of the distribution have been considered in the literature. Among them, we highlight the proposal for modeling the dispersion or precision, such as in generalized linear models ([Smyth, 1989](#)) and in the beta regression model ([Simas et al., 2010](#)). In more recent proposals, models that address the modeling of the two characteristics, mean and variance, have been introduced in seminal proposals. For example, [Santos-Neto et al. \(2016\)](#) proposed a reparameterized Birnbaum-Saunders regression model with varying precision and [Bourguignon and Gallardo \(2020\)](#) developed the reparameterized inverse gamma regression model with varying precision. In the context of the modal regression, [Bourguignon et al. \(2020\)](#) presented a parametric modal regression with varying precision where the response variable is gamma distributed. Additionally, [Altun \(2021\)](#) introduced a new Lomax regression model, in which the response's mean and shape parameter (α) are modeled by regression structures through the link functions. Recently, [Bourguignon and Nascimento \(2020\)](#) presented a Bayesian approach that considers a new parametrization that is indexed by mean and precision parameters, in which the response variable is a generalized Pareto distribution. The main advantage of this reparametrization is that it allows the mean and precision parameters to be modeled directly, allowing the construction of simple and interpretable models, such as in the context of generalized linear models ([McCullagh and Nelder, 1989](#)).

Based on the above discussion, this work has the objective of using the maximum likelihood (ML) approach to make inferences in the regression model with the same reparametrization used in [Bourguignon and Nascimento \(2020\)](#). The methods presented in this article differs from the described in [Bourguignon and Nascimento \(2020\)](#) in one main aspect; the approach to estimate parameters of the models. While [Bourguignon and Nascimento \(2020\)](#) used the Bayesian approach, we presented the ML inference approach. Additionally, in our proposal, the parametric support of the precision parameter is different from the one used by [Bourguignon and Nascimento \(2020\)](#), in this sense, other link functions are suggested and used for precision modeling. In this paper, the estimation of the parameters is performed using the ML method. We obtain analytical expressions for the score vector and Fisher information matrix, and also propose diagnostic measures and tools for model selection. We emphasize that obtaining the Fisher information matrix is possible due to the simplicity of the Lomax probability density function. Such expressions are impossible and/or very costly to obtain in some more complex distributions.

This paper is organized as follows. In Section 2, we present the Lomax distribution, the proposed reparameterization, reparametrized Lomax regression model, and log-likelihood function of the model. In Section 3, we present methods for the estimation and inferences, such as the score vector, the observed and Fisher information matrix, procedures for obtaining confidence intervals and hypothesis tests, additionally we introduce some diagnostic measures to check the goodness-of-fit of the proposed model. Monte Carlo simulation results are presented and discussed in Section 4. We also present and discuss an application. Finally, the conclusions and final remarks are presented in Section 5.

2. PROPOSED MODEL

In this section, we introduce the two-parameter Lomax distribution and its main characteristics such as mean, variance, cumulative distribution function, and quantile function. Furthermore, we present the reparameterization in terms of mean and precision parameters, the regression structures for modeling the mean and precision, as well as the log-likelihood function.

2.1 THE LOMAX DISTRIBUTION

Let Y be a random variable with Lomax distribution. Its probability density function is given by

$$f(y; \alpha, \lambda) = \frac{\alpha \lambda^\alpha}{(y + \lambda)^{(\alpha+1)}}, \quad y > 0, \quad (1)$$

where $\alpha > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. The mean and variance of Y are stated, respectively, by $E(Y) = \lambda/(\alpha - 1)$, for $\alpha > 1$, and $\text{Var}(Y) = \alpha \lambda^2 / ((\alpha - 1)^2(\alpha - 2))$, for $\alpha > 2$. The cumulative distribution function corresponding to Equation (1) is expressed by

$$F(y; \lambda, \alpha) = 1 - \left(1 + \frac{y}{\lambda}\right)^{-\alpha}.$$

2.2 A REPARAMETRIZED LOMAX DISTRIBUTION

In regression analysis, it is typically more useful and common to model the mean response, as it makes the model parameters easily interpretable. In order to obtain a regression structure for the mean of Y , we consider a new parameterization, which is obtained by taking $\mu = \lambda/(\alpha - 1)$ and $\phi = (\alpha - 2)/\alpha$ in Equation (1), that is, $\lambda = \mu(\alpha - 1)$ and $\alpha = 2/(1 - \phi)$. The mean-parametrized Lomax distribution with mean μ and precision ϕ is characterized by the probability density function expressed as

$$f(y; \mu, \phi) = \frac{\frac{2}{1-\phi} \left[\mu \left(\frac{2}{1-\phi} - 1 \right) \right]^{\frac{2}{1-\phi}}}{\left[y + \mu \left(\frac{2}{1-\phi} - 1 \right) \right]^{\frac{2}{1-\phi} + 1}}, \quad y > 0, \quad \mu > 0, \quad 0 < \phi < 1. \quad (2)$$

The mean and variance are given, respectively, by

$$E(Y) = \mu \quad \text{and} \quad \text{Var}(Y) = \frac{\mu^2}{\phi}.$$

The new cumulative distribution function is stated as

$$F(y; \mu, \phi) = 1 - \left[1 + \frac{y(1 - \phi)}{\mu(1 + \phi)} \right]^{-\frac{2}{1 - \phi}}. \quad (3)$$

2.3 THE REPARAMETRIZED LOMAX REGRESSION MODEL

Let Y_1, \dots, Y_n be independent random variables, where each $Y_t, t = 1, \dots, n$, follows the probability density function stated in Equation (2) with mean μ_t and precision ϕ_t . The regression structures for the mean and precision of Y_t are formulated, respectively, by

$$\eta_{1t} = g_1(\mu_t) = \mathbf{x}_t^\top \boldsymbol{\beta} \quad \text{and} \quad \eta_{2t} = g_2(\phi_t) = \mathbf{z}_t^\top \boldsymbol{\gamma}, \quad (4)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_r)^\top \in \mathbb{R}^{r+1}$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_q)^\top \in \mathbb{R}^{q+1}$ are vectors of unknown regression parameters assumed to be functionally independent ($r + q + 2 < n$), $\mathbf{x}_t = (1, x_{t1}, \dots, x_{tr})^\top$ and $\mathbf{z}_t = (1, z_{t1}, \dots, z_{tq})^\top$ are explanatory variables vectors, η_{1t} and η_{2t} are the mean and precision linear predictors, respectively, and g_1 and g_2 are twice-differentiable one-to-one monotonic functions called link functions, where $g_1: \mathbb{R}^+ \rightarrow \mathbb{R}$ and $g_2: (0, 1) \rightarrow \mathbb{R}$.

The proposed Lomax regression model is defined by Equations (2) and (4). Due to the restriction $\mu_t > 0$, the most common link function and that satisfies the conditions stated for g_1 is the logarithm, $g_1(\mu_t) = \log(\mu_t)$, because it provides non-negative values for $\mu_t = g_1^{-1}(\eta_t) = \exp(\eta_t)$ regardless the values assigned to η_t , is twice-differentiable one-to-one monotonic function. Other link functions are usual, but they do not satisfy all the conditions stated for g_1 , they are the square root, $g_1(\mu_t) = \sqrt{\mu_t}$, and inverse, $g_1(\mu_t) = 1/\mu_t$ (with special attention to the positivity of the estimates). For the restriction $0 < \phi_t < 1$, we can use the logit $g_2(\phi_t) = \log[\phi_t/(1 - \phi_t)]$, probit $g_2(\phi_t) = \Phi^{-1}(\phi_t)$, where Φ is the cumulative distribution function of a standard normal random variable, complementary log-log $g_2(\phi_t) = \log[-\log(1 - \phi_t)]$, and log-log $g_2(\phi_t) = -\log[-\log(\phi_t)]$ link functions (Ferrari and Cribari-Neto, 2004; Simas et al., 2010). For more details and a detailed discussion about link functions, see Atkinson (1985, Ch. 7) and McCullagh and Nelder (1989).

2.4 LIKELIHOOD FUNCTION

Let Y_1, \dots, Y_n be a sample from the proposed Lomax regression model, y_1, \dots, y_n its observations, and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ the corresponding regression parameter vector. The corresponding log-likelihood function for $\boldsymbol{\theta}$ is given by

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^n \ell_t(\mu_t, \phi_t), \quad (5)$$

where

$$\begin{aligned} \ell_t(\mu_t, \phi_t) = & \log(2) - \log(1 - \phi_t) + \frac{2}{1 - \phi_t} \log \left[\mu_t \left(\frac{2}{1 - \phi_t} - 1 \right) \right] \\ & - \left(\frac{2}{1 - \phi_t} + 1 \right) \log \left[y_t + \mu_t \left(\frac{1 + \phi_t}{1 - \phi_t} \right) \right]. \end{aligned} \quad (6)$$

3. ESTIMATION AND INFERENCE

In this section, we present details for performing point and interval estimation, and hypothesis testing. Initially, we present the score vector, the observed information matrix and Fisher information matrix, next present a test statistic to test hypotheses of interest and the formula for obtaining confidence intervals.

3.1 SCORE VECTOR

Taking first derivatives of the log-likelihood function with respect to each element of $\boldsymbol{\theta}$, we obtain the score vector $U(\boldsymbol{\theta}) = (\mathbf{U}_\beta(\boldsymbol{\theta})^\top, \mathbf{U}_\gamma(\boldsymbol{\theta})^\top)^\top$ given by

$$U_{\beta_i}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_i} = \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{d\mu_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \beta_i},$$

$$U_{\gamma_i}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma_i} = \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \frac{d\phi_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_i}.$$

From Equation (6), the derivative of $\ell_t(\mu_t, \phi_t)$ with respect to μ_t is defined by

$$\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} = \frac{2}{\mu_t(1 - \phi_t)} - \frac{(3 - \phi_t)(1 + \phi_t)}{(1 - \phi_t)^2 [y_t + c_t]} := b_t, \tag{7}$$

where $c_t = \mu_t(1 + \phi_t)/(1 - \phi_t)$. Note that $\eta_{1t} = g_1(\mu_t)$, then $d\mu_t/d\eta_{1t} = 1/g'_1(\mu_t)$, where g' is the first derivative of function g . We also have that $\partial \eta_{1t}/\partial \beta_i = x_{ti}$. Therefore, it follows that

$$U_{\beta_i}(\boldsymbol{\theta}) = \sum_{t=1}^n b_t \frac{1}{g'_1(\mu_t)} x_{ti}, \quad i = 0, 1, \dots, r,$$

where $x_{t0} = 1$. By taking derivative in Equation (6) with respect to ϕ_t , we define $\partial \ell_t(\mu_t, \phi_t)/\partial \phi_t := a_t$, it follows that

$$a_t = \left[\frac{1}{1 - \phi_t} + \frac{2 \log(c_t)}{(1 - \phi_t)^2} - \frac{2 \log(y_t + c_t)}{(1 - \phi_t)^2} + \frac{4}{(1 + \phi_t)(1 - \phi_t)^2} - \frac{2\mu_t(3 - \phi_t)}{(1 - \phi_t)^3(y_t + c_t)} \right]. \tag{8}$$

For γ_i , we have that $d\phi_t/d\eta_{2t} = 1/g'_2(\phi_t)$ and $\partial \eta_{2t}/\partial \gamma_i = z_{ti}$. Therefore, we obtain

$$U_{\gamma_i}(\boldsymbol{\theta}) = \sum_{t=1}^n a_t \frac{1}{g'_2(\phi_t)} z_{ti}, \quad i = 0, 1, \dots, q,$$

where $z_{t0} = 1$. The score vector can be expressed in matrix form as $\mathbf{U}_\beta(\boldsymbol{\theta}) = \mathbf{X}^\top \mathbf{M} \mathbf{b}$ and $\mathbf{U}_\gamma(\boldsymbol{\theta}) = \mathbf{Z}^\top \mathbf{M} \mathbf{a}$, where \mathbf{X} is an $n \times r$ matrix with the t -th row given by x_t , \mathbf{Z} is an $n \times q$ matrix with t -th row given by z_{ij} , $\mathbf{b} = (b_1, \dots, b_n)^\top$, $\mathbf{a} = (a_1, \dots, a_n)^\top$, $\mathbf{M} = \text{diag} \{1/g'_1(\mu_1), \dots, 1/g'_1(\mu_n)\}$, and $\mathbf{M} = \text{diag} \{1/g'_2(\phi_1), \dots, 1/g'_2(\phi_n)\}$.

The ML estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are obtained by solving the nonlinear system of equations expressed as $\mathbf{U}_\beta(\boldsymbol{\theta}) = \mathbf{0}$ and $\mathbf{U}_\gamma(\boldsymbol{\theta}) = \mathbf{0}$. Since the above system does not have analytical solution, the use of nonlinear optimization algorithms is required. In this work, we apply the Nelder-Mead simplex method (Nelder and Mead, 1965).

3.2 OBSERVED INFORMATION MATRIX

Taking second order derivatives of Equation (5) with respect to each element of $\boldsymbol{\theta}$, we have

$$\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_i \partial \beta_j} &= \sum_{t=1}^n \frac{\partial}{\partial \beta_i} \left[\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{d\mu_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \beta_j} \right] \\
&= \sum_{t=1}^n \left[\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_{1t}} + \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{\partial}{\partial \mu_t} \left(\frac{d\mu_t}{d\eta_{1t}} \right) \right] \frac{d\mu_t}{d\eta_{1t}} x_{tj} x_{ti}, \\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_i \partial \gamma_j} &= \sum_{t=1}^n \frac{\partial}{\partial \gamma_j} \left[\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{d\mu_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \beta_i} \right] = \sum_{t=1}^n \left[\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \mu_t \partial \phi_t} \frac{d\phi_t}{d\eta_{2t}} z_{tj} \right] \frac{d\mu_t}{d\eta_t} x_{ti}, \\
\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_i \partial \gamma_j} &= \sum_{t=1}^n \frac{\partial}{\partial \gamma_i} \left[\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \frac{d\phi_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_j} \right] \\
&= \sum_{t=1}^n \left[\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \phi_t^2} \frac{d\phi_t}{d\eta_{2t}} + \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \frac{\partial}{\partial \phi_t} \left(\frac{d\phi_t}{d\eta_{2t}} \right) \right] \frac{d\phi_t}{d\eta_{2t}} z_{tj} z_{ti}.
\end{aligned}$$

In addition, from Equation (7), we have

$$\begin{aligned}
\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \mu_t^2} &= \frac{\partial}{\partial \mu_t} \left[\frac{2}{\mu_t(1-\phi_t)} - \frac{(3-\phi_t)(1+\phi_t)}{(1-\phi_t)^2 [y_t + c_t]} \right] \\
&= -\frac{2}{\mu_t^2(1-\phi_t)} + \frac{(3-\phi_t)(1+\phi_t)^2}{(1-\phi_t)^3 (y_t + c_t)^2} := w_t, \\
\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \mu_t \partial \phi_t} &= \frac{\partial}{\partial \phi_t} \left[\frac{2}{\mu_t(1-\phi_t)} - \frac{(3-\phi_t)(1+\phi_t)}{(1-\phi_t)^2 [y_t + c_t]} \right] \\
&= \frac{2}{\mu_t(1-\phi_t)^2} - \frac{2\mu_t(1+\phi_t)^2 + 8y_t(1-\phi_t)}{(1-\phi_t)^4 (y_t + c_t)^2} := r_t.
\end{aligned}$$

And from Equation (8), we have

$$\begin{aligned}
\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \phi_t^2} &= \frac{\partial}{\partial \phi_t} \left[\frac{1}{1-\phi_t} + \frac{2 \log(c_t) - 2 \log(y_t + c_t)}{(1-\phi_t)^2} + \frac{4}{(1+\phi_t)(1-\phi_t)^2} \right. \\
&\quad \left. - \frac{2\mu_t(3-\phi_t)}{(1-\phi_t)^3 (y_t + c_t)} \right] \\
&= \frac{5 - \phi_t^2}{(1+\phi_t)(1-\phi_t)^3} + \frac{4(1+3\phi_t)}{(1+\phi_t)^2(1-\phi_t)^3} + \frac{4 \log(c_t)}{(1-\phi_t)^3} - \frac{4\mu}{(y+c_t)(1-\phi_t)^4} \\
&\quad - \frac{4 \log(y+c_t)}{(1-\phi_t)^3} + \frac{4\mu^2(3-\phi_t)}{(y+c_t)^2(1-\phi_t)^5} + \frac{4\mu(\phi_t-4)}{(y+c_t)(1-\phi_t)^4} := s_t.
\end{aligned}$$

Notice also that

$$\frac{\partial}{\partial \mu_t} \left(\frac{d\mu_t}{d\eta_{1t}} \right) = -\frac{g''(\mu_{1t})}{[g'(\mu_{1t})]^2} := m_t \quad \text{and} \quad \frac{\partial}{\partial \phi_t} \left(\frac{d\phi_t}{d\eta_{2t}} \right) = -\frac{g''(\phi_{2t})}{[g'(\phi_{2t})]^2} := o_t,$$

where g'' is the second derivative of function g .

Let $\mathbf{H} = \text{diag}\{h_1, \dots, h_n\}$ with $h_t = [w_t/g'(\mu_{1t}) + b_tm_t]/g'(\mu_{1t})$, $\mathbf{R} = (r_1, \dots, r_n)^\top$, and $\mathbf{P} = (p_1, \dots, p_n)^\top$ with $p_t = [s_t/g'(\phi_{2t}) + a_t o_t]/g'(\phi_{2t})$. The joint observed information matrix for $\boldsymbol{\theta}$ is given by

$$\mathbf{J}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{J}_{(\beta,\beta)} & \mathbf{J}_{(\beta,\gamma)} \\ \mathbf{J}_{(\gamma,\beta)} & \mathbf{J}_{(\gamma,\gamma)} \end{pmatrix},$$

where $\mathbf{J}_{(\beta,\beta)} = -\mathbf{X}^\top \mathbf{H} \mathbf{X}$, $\mathbf{J}_{(\beta,\gamma)} = \mathbf{J}_{(\gamma,\beta)}^\top = -\mathbf{X}^\top \mathbf{M} \mathbf{R} \mathbf{M} \mathbf{Z}$, and $\mathbf{J}_{(\gamma,\gamma)} = \mathbf{Z}^\top \mathbf{P} \mathbf{Z}$.

3.3 INFORMATION MATRIX, CONFIDENCE INTERVALS AND HYPOTHESIS TESTING

Before presenting to the important quantities of this subsection, we need some useful results used in obtaining of the Fisher information matrix provided by Lemma below.

Lemma 1: *Let Y_t be a random variable that follows a Lomax distribution with probability density function given in Equation (2). Then,*

$$\begin{aligned} \mathbb{E}\left(\frac{1}{Y_t + \mu_t (2/(1 - \phi_t - 1))}\right) &= \frac{2(1 - \phi_t)}{\mu_t(1 + \phi_t)(3 - \phi_t)}, \\ \mathbb{E}\left(\frac{1}{(Y_t + \mu_t (2/(1 - \phi_t - 1)))^2}\right) &= \frac{(1 - \phi_t)^2}{\mu_t^2(1 + \phi_t)^2(2 - \phi_t)}, \\ \mathbb{E}(\log(Y_t + \mu_t (2/(1 - \phi_t - 1)))) &= \log\left(\frac{\mu_t(1 + \phi_t)}{1 - \phi_t}\right) + \frac{1 - \phi_t}{2}. \end{aligned}$$

PROOF: Let f be the probability density function of Y_t . Then,

$$\mathbb{E}\left(\frac{1}{Y_t + \mu_t \left(\frac{2}{1 - \phi_t} - 1\right)}\right)^k = \int_0^\infty \frac{\frac{2}{1 - \phi_t} \left(\mu_t \left(\frac{2}{1 - \phi_t} - 1\right)\right)^{\frac{2}{1 - \phi_t}}}{\left(y_t + \mu_t \left(\frac{2}{1 - \phi_t} - 1\right)\right)^{\left(\frac{2}{1 - \phi_t} + 1 + k\right)}} dy_t,$$

where $k > 0$. Making the variable change $x_t = y_t + \mu_t(2/(1 - \phi_t - 1))$, then the above equation becomes

$$\begin{aligned} \mathbb{E}\left(\frac{1}{Y_t + \mu_t (2/(1 - \phi_t - 1))}\right)^k &= \frac{2(1 - \phi_t)^k}{\mu_t^k(1 + \phi_t)^k(2 + k - \phi_t k)}, \\ \mathbb{E}\left(\log\left(Y_t + \mu_t \left(\frac{2}{1 - \phi_t} - 1\right)\right)\right) &= \log\left(\frac{\mu_t(1 + \phi_t)}{1 - \phi_t}\right) + \frac{1 - \phi_t}{2}. \quad \blacksquare \end{aligned}$$

The Fisher information matrix is obtained by taking the expected value of the second order derivatives of the log-likelihood function, that is, $\mathbf{K}(\boldsymbol{\theta}) = \mathbb{E}[\mathbf{J}(\boldsymbol{\theta})]$. Since

$$\begin{aligned} \mathbb{E}\left(\frac{\partial \ell(\mu_t, \phi_t)}{\partial \mu_t}\right) &= \frac{2}{\mu_t(1 - \phi_t)} - \frac{(1 + \phi_t)(3 - \phi_t)}{(1 - \phi_t)^2} \mathbb{E}\left(\frac{1}{Y_t + c_t}\right) \\ &= \frac{2}{\mu_t(1 - \phi_t)} - \frac{2}{\mu_t(1 - \phi_t)} = 0, \end{aligned}$$

the expected value of the derivatives in Section 3.2 are given by

$$\begin{aligned} \mathbb{E} \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_i \partial \beta_j} \right) &= \sum_{t=1}^n \mathbb{E} \left(\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \mu_t^2} \right) \left(\frac{d\mu_t}{d\eta_{1t}} \right)^2 x_{tj} x_{ti}, \\ \mathbb{E} \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_i \partial \gamma_j} \right) &= \sum_{t=1}^n \mathbb{E} \left(\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \mu_t \partial \phi_t} \right) \frac{d\phi_t}{d\eta_{2t}} \frac{d\mu_t}{d\eta_{1t}} z_{tj} x_{ti}, \\ \mathbb{E} \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma_i \partial \gamma_j} \right) &= \sum_{t=1}^n \mathbb{E} \left(\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \phi_t^2} \right) \left(\frac{d\phi_t}{d\eta_{2t}} \right)^2 z_{tj} z_{ti}. \end{aligned}$$

Observe that taking the expected value in Equations (7) and (8), and substituting the results from this lemma, we have

$$\begin{aligned} \mathbb{E} \left(\frac{\partial \ell(\mu_t, \phi_t)}{\partial \phi_t} \right) &= \frac{1}{1 - \phi_t} + \frac{2 \log(c)}{(1 - \phi_t)^2} - \frac{2}{(1 - \phi_t)^2} \left(\log c + \frac{1 - \phi_t}{2} \right) + \frac{4}{(1 + \phi_t)(1 - \phi_t)^2} \\ &\quad - \frac{2\mu_t(3 - \phi_t)}{(1 - \phi_t)^3} \frac{2(1 - \phi_t)}{\mu_t(1 + \phi_t)(3 - \phi_t)} = 0, \end{aligned}$$

$$\begin{aligned} \mathbb{E} \left(\frac{\partial^2 \ell(\mu_t, \phi_t)}{\partial \mu_t^2} \right) &= -\frac{2}{\mu_t^2(1 - \phi_t)} + \frac{(1 + \phi_t)^2(3 - \phi_t)}{(1 - \phi_t)^3} \mathbb{E} \left(\frac{1}{Y_t + c_t} \right)^2 \\ &= -\frac{2}{\mu_t^2(1 - \phi_t)} + \frac{(1 + \phi_t)^2(3 - \phi_t)}{(1 - \phi_t)^3} \frac{(1 - \phi_t)^2}{\mu_t^2(1 + \phi_t)^2(2 - \phi_t)} := v_t, \end{aligned}$$

$$\begin{aligned} \mathbb{E} \left(\frac{\partial^2 \ell(\mu_t, \phi_t)}{\partial \mu_t \partial \phi_t} \right) &= \frac{2}{\mu_t(1 - \phi_t)^2} - \frac{2\mu_t(1 + \phi_t)^2}{(1 - \phi_t)^4} \mathbb{E} \left(\frac{1}{(Y_t + c_t)^2} \right) - \frac{8}{(1 - \phi_t)^3} \mathbb{E} \left(\frac{Y_t}{(Y_t + c_t)^2} \right) \\ &= \frac{2}{\mu_t(1 - \phi_t)^2} - \frac{2}{\mu_t(2 - \phi_t)(1 - \phi_t)^2} - \frac{8}{\mu_t(3 - \phi_t)(2 - \phi_t)(1 - \phi_t^2)} := d_t, \end{aligned}$$

$$\begin{aligned} \mathbb{E} \left(\frac{\partial^2 \ell(\mu_t, \phi_t)}{\partial \phi_t^2} \right) &= \frac{5 - \phi_t^2}{(1 + \phi_t)(1 - \phi_t)^3} + \frac{4(1 + 3\phi_t)}{(1 + \phi_t)^2(1 - \phi_t)^3} + \frac{4 \log(c_t)}{(1 - \phi_t)^3} \\ &\quad - \frac{4\mu_t}{(1 - \phi_t)^4} \mathbb{E} \left(\frac{1}{Y_t + c_t} \right) - \frac{4}{(1 - \phi_t)^3} \mathbb{E}(\log(Y_t + c_t)) + \frac{4\mu_t^2(3 - \phi_t)}{(1 - \phi_t)^5} \mathbb{E} \left(\frac{1}{Y_t + c_t} \right)^2 \\ &\quad + \frac{4\mu_t(\phi_t - 4)}{(1 - \phi_t)^4} \mathbb{E} \left(\frac{1}{Y_t + c_t} \right) \\ &= \frac{5 - \phi_t^2}{(1 + \phi_t)(1 - \phi_t)^3} + \frac{4(1 + 3\phi_t)}{(1 + \phi_t)^2(1 - \phi_t)^3} + \frac{4(3 - \phi_t)}{(1 - \phi_t)^3(1 + \phi_t)^2(2 - \phi_t)} \\ &\quad + \frac{8(\phi_t - 5)}{(1 - \phi_t)^3(1 + \phi_t)(3 - \phi_t)} - \frac{2}{(1 - \phi_t)^2} := q_t. \end{aligned}$$

Let $\mathbf{V} = \text{diag}\{v_1, \dots, v_n\}$, $\mathbf{D} = (d_1, \dots, d_n)^\top$, and $\mathbf{Q} = \text{diag}\{q_1, \dots, q_n\}$. The Fisher information matrix for $\boldsymbol{\theta}$ is given by

$$\mathbf{K}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{K}_{(\beta,\beta)} & \mathbf{K}_{(\beta,\gamma)} \\ \mathbf{K}_{(\gamma,\beta)} & \mathbf{K}_{(\gamma,\gamma)} \end{pmatrix},$$

where $\mathbf{K}_{(\beta,\beta)} = -\mathbf{X}^\top \mathbf{V} \mathbf{M}^2 \mathbf{X}$, $\mathbf{K}_{(\beta,\gamma)} = \mathbf{K}_{(\gamma,\beta)}^\top = -\mathbf{X}^\top \mathbf{M} \mathbf{D} \mathbf{M} \mathbf{Z}$, and $\mathbf{K}_{(\gamma,\gamma)} = -\mathbf{X}^\top \mathbf{Q} \mathbf{M}^2 \mathbf{X}$. Note that the parameters β and γ are not orthogonal.

Under usual regularity conditions, the ML estimators $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ are asymptotically consistent, having approximately normal distribution with mean vector $\boldsymbol{\theta}$ and variance-covariance matrix $\mathbf{K}(\boldsymbol{\theta})^{-1}$ in large samples (Pawitan, 2001), that is,

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \sim N_{r+q+2} \left(\begin{pmatrix} \beta \\ \gamma \end{pmatrix}, \mathbf{K}(\boldsymbol{\theta})^{-1} \right), \quad (9)$$

where N_{r+q+2} denotes the $(r + q + 2)$ -dimensional normal distribution and $\hat{\beta}$ and $\hat{\gamma}$ the ML estimators of β and γ , respectively.

Test statistics for hypothesis testing and confidence intervals can be obtained using the asymptotic result presented in Equation (9). Suppose the interest is to test the following hypotheses $\mathcal{H}_0: \theta_i = \theta_i^0$ versus $\mathcal{H}_1: \theta_i \neq \theta_i^0$, where θ_i^0 is a specified value for the unknown parameter θ_i . A useful statistic to test these hypotheses is the signed square root of the Wald statistic, given by $Z = (\hat{\theta}_i - \theta_i^0) / \sqrt{k^{ii}}$, where k^{ii} is the i -th diagonal element of $\mathbf{K}(\hat{\boldsymbol{\theta}})^{-1}$. This statistic is particularly convenient to test individual parameters (Pawitan, 2001). Under \mathcal{H}_0 and for large n , Z has a standard normal distribution. It is also possible to perform more general hypothesis testing inference using the likelihood ratio, Wald, and score statistics.

We can also use the result presented in Equation (9) to construct asymptotic confidence intervals for each parameter θ_i . An approximate $100(1 - \alpha)\%$ confidence interval for θ_i is defined as $(\hat{\theta}_i - z_{1-\alpha/2} \sqrt{k^{ii}}, \hat{\theta}_i + z_{1-\alpha/2} \sqrt{k^{ii}})$, where $\Phi(z_{1-\alpha/2}) = 1 - \alpha/2$.

3.4 DIAGNOSTIC MEASURES

In this subsection we suggest criteria for selecting the Lomax regression model and some diagnostic measures for examining the goodness-of-fit of the proposed model. For model selection, we consider the Akaike Information Criterion (AIC) (Akaike, 1974) and Bayesian Information Criterion (BIC) (Schwarz, 1978) given, respectively, by $\text{AIC} = -2\ell(\hat{\boldsymbol{\theta}}) + 2(q + r + 2)$ and $\text{BIC} = -2\ell(\hat{\boldsymbol{\theta}}) + \log(n)(q + r + 2)$.

For validating the proposed model, we perform residual analysis using the randomized quantile residuals (Dunn and Smyth, 1996), defined as $r_t^{(q)} = \Phi^{-1}(F(y_t; \hat{\mu}_t, \hat{\phi}_t))$, where $F(y_t; \hat{\mu}_t, \hat{\phi}_t)$ is the cumulative distribution function stated in Equation (3). If the model is correctly specified, these residuals should be independent and normally distributed, with zero mean and unit variance.

4. NUMERICAL RESULTS

In this section, we provide the simulation study in order to evaluate the performance of the ML estimators of the proposed model under different sample sizes. Also, we present and discuss an empirical application to illustrate the proposed framework.

4.1 SIMULATION STUDY

We conduct a Monte Carlo simulation study to evaluate the finite sample performance of the likelihood inference for the proposed Lomax regression model. We used 10,000 Monte Carlo replications and considered five sample sizes $n \in \{50, 100, 200, 500, 1000\}$. Performance measures for ML estimator evaluation are the mean, bias, relative bias (RB), standard deviation (SD), and root mean square error (RMSE).

We considered two scenarios with the following true parameter values: (i) Scenario 1: $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1) = (-3.0, 2.2, 1.5, 0.5, -0.3)$ and (ii) Scenario 2: $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \gamma_0, \gamma_1) = (0.5, 1.3, -0.7, 0.3, -0.5)$. In both scenarios, the covariates were generated independently from a standard uniform distribution, $\mathcal{U}(0, 1)$, and kept constant during all Monte Carlo replications. We considered the logarithmic and probit link functions for the mean and precision submodels, respectively. This scenario considers the link functions that provided the best fit in the application.

All simulations were performed using the R software. The maximization was obtained considering the `optim` function, available in R, using the Nelder-Mead method with first analytical derivatives; see [Nelder and Mead \(1965\)](#) for more details. As the iterative optimization algorithm requires a set of initial values for the parameters to be optimized, we suggest to use the following empirical approach to determine these values. The initial values $\boldsymbol{\beta}^{(0)}$ for $\boldsymbol{\beta}$ are obtained by the least squares estimates of $\boldsymbol{\beta}$ from the following linear regression model: $\log(y_t) = \mathbf{x}_t^\top \boldsymbol{\beta}$, while the starting values $\boldsymbol{\gamma}^{(0)}$ for $\boldsymbol{\gamma}$ are obtained by $\boldsymbol{\gamma}^{(0)} = (\bar{y}^2/S_y^2)\mathbf{1}_{q+1}^\top$, where \bar{y} and S_y^2 denote the sample mean and variance, respectively, and $\mathbf{1}_{q+1}$ denotes an $(q+1)$ -dimensional vector of 1's. We have tested the others methods, however, in our study, the Nelder-Mead method provided more robust estimates.

The simulation results are shown in [Tables 1 and 2](#). Based on the results presented, we can verify the good performance of the ML estimators of the Lomax regression model. We observe that the bias and RMSE of the ML estimators of $\boldsymbol{\beta}$ tend toward zero as the sample size increases, indicating the consistency property of the ML estimator. For the vector $\boldsymbol{\gamma}$, the ML estimators are biased in small samples, but the bias decreases as the sample size increases. This suggests that some procedure for inferential improvements can be considered to reduce the problem of biased ML estimator in small samples. We also highlight that this behavior of the ML estimators in precision modeling is recurrent in the literature ([Bourguignon and Nascimento, 2020](#); [Simas et al., 2010](#)).

4.2 EMPIRICAL APPLICATION

We illustrate the proposed model using dataset obtained from the United Nations Development Programme (available at <http://hdr.undp.org/en/data>). The response variable is the carbon dioxide emissions per capita (DEC, measured in tonnes) in 123 countries, including the autonomous territory of Hong-Kong and the United Kingdom collected in 2016. The covariates associated with this response variable are: forest area (FAR, measured in % of total land area), concentration index of exports (CIN, ranging from 0% to 100%, with a larger value denoting a higher concentration of exports), employment in agriculture (EAG, measured in % of total employment), and human development index (HDI). Some summary statistics of the response variable are given in [Table 3](#). [Figure 1](#) shows the dispersion plots between the response variable and covariates. After some adjustments, we consider only the set of regressors statistically significant at the level of 10% in the Lomax regression model. The HDI covariate was not significant for the mean submodel. Also, the FAR, CIN, and EAG covariates were not significant for the precision submodel. We use the observed information matrix obtained numerically using the `optim` function of R software because it provided lower variance estimates than the Fisher information matrix.

Table 1. Monte Carlo simulation results for likelihood inference, evaluation of point estimation, for the Lomax regression model - Scenario 1 - $g_1(\mu_t) = \log(\mu_t)$, $g_2(\phi_t) = \Phi^{-1}(\phi_t)$, $\beta_0 = -3.0$, $\beta_1 = 2.2$, $\beta_2 = 1.5$, $\gamma_0 = 0.5$ and $\gamma_1 = -0.3$.

n	Estimator	Mean	Bias	RB	SD	RMSE
50	$\hat{\beta}_0$	-3.030	-0.030	0.010	0.556	0.557
	$\hat{\beta}_1$	2.210	0.010	0.004	0.736	0.736
	$\hat{\beta}_2$	1.480	-0.020	-0.013	0.714	0.714
	$\hat{\gamma}_0$	2.161	1.661	3.322	9.454	9.599
	$\hat{\gamma}_1$	-0.704	-0.404	1.347	55.585	55.586
100	$\hat{\beta}_0$	-3.013	-0.013	0.004	0.357	0.358
	$\hat{\beta}_1$	2.198	-0.002	-0.001	0.434	0.434
	$\hat{\beta}_2$	1.508	0.008	0.005	0.502	0.502
	$\hat{\gamma}_0$	1.235	0.735	1.471	6.504	6.546
	$\hat{\gamma}_1$	-0.171	0.129	-0.431	13.708	13.709
200	$\hat{\beta}_0$	-3.008	-0.008	0.003	0.240	0.240
	$\hat{\beta}_1$	2.202	0.002	0.001	0.314	0.314
	$\hat{\beta}_2$	1.499	-0.001	-0.001	0.292	0.292
	$\hat{\gamma}_0$	0.805	0.305	0.610	2.413	2.432
	$\hat{\gamma}_1$	-0.214	0.086	-0.286	4.165	4.166
500	$\hat{\beta}_0$	-3.002	-0.002	0.001	0.135	0.135
	$\hat{\beta}_1$	2.199	-0.001	0.000	0.179	0.179
	$\hat{\beta}_2$	1.500	0.000	0.000	0.179	0.179
	$\hat{\gamma}_0$	0.641	0.141	0.281	0.777	0.789
	$\hat{\gamma}_1$	-0.368	-0.068	0.225	1.318	1.319
1000	$\hat{\beta}_1$	-3.001	-0.001	0.000	0.099	0.099
	$\hat{\beta}_2$	2.199	-0.001	0.000	0.130	0.130
	$\hat{\beta}_3$	1.501	0.001	0.001	0.129	0.129
	$\hat{\gamma}_0$	0.552	0.052	0.104	0.468	0.471
	$\hat{\gamma}_1$	-0.312	-0.012	0.039	0.786	0.786

Table 2. Monte Carlo simulation results for likelihood inference, evaluation of point estimation, for the Lomax regression model - Scenario 2 - $g_1(\mu_t) = \log(\mu_t)$, $g_2(\phi_t) = \Phi^{-1}(\phi_t)$, $\beta_0 = 0.5$, $\beta_1 = 1.3$, $\beta_2 = -0.7$, $\gamma_0 = 0.3$ and $\gamma_1 = -0.5$.

n	Estimator	Mean	Bias	RB	SD	RMSE
50	$\hat{\beta}_0$	0.443	-0.057	-0.114	0.578	0.581
	$\hat{\beta}_1$	1.310	0.010	0.008	0.673	0.673
	$\hat{\beta}_2$	-0.684	0.016	-0.023	0.613	0.614
	$\hat{\gamma}_0$	1.163	0.863	2.878	16.240	16.263
	$\hat{\gamma}_1$	-1.372	-0.872	1.744	34.205	34.216
100	$\hat{\beta}_0$	0.493	-0.007	-0.014	0.332	0.332
	$\hat{\beta}_1$	1.294	-0.006	-0.005	0.434	0.434
	$\hat{\beta}_2$	-0.705	-0.005	0.008	0.442	0.442
	$\hat{\gamma}_0$	0.669	0.369	1.232	6.398	6.408
	$\hat{\gamma}_1$	-0.774	-0.274	0.547	14.265	14.268
200	$\hat{\beta}_0$	0.498	-0.002	-0.004	0.200	0.200
	$\hat{\beta}_1$	1.298	-0.002	-0.001	0.263	0.263
	$\hat{\beta}_2$	-0.702	-0.002	0.003	0.271	0.271
	$\hat{\gamma}_0$	0.523	0.223	0.744	2.101	2.113
	$\hat{\gamma}_1$	-0.732	-0.232	0.465	3.488	3.496
500	$\hat{\beta}_0$	0.498	-0.002	-0.004	0.152	0.152
	$\hat{\beta}_1$	1.298	-0.002	-0.002	0.196	0.196
	$\hat{\beta}_2$	-0.699	0.001	-0.001	0.194	0.194
	$\hat{\gamma}_0$	0.410	0.110	0.365	0.694	0.702
	$\hat{\gamma}_1$	-0.612	-0.112	0.223	1.234	1.239
1000	$\hat{\beta}_0$	0.499	-0.001	-0.001	0.105	0.105
	$\hat{\beta}_1$	1.297	-0.003	-0.002	0.133	0.133
	$\hat{\beta}_2$	-0.699	0.001	-0.001	0.138	0.138
	$\hat{\gamma}_0$	0.344	0.044	0.148	0.397	0.400
	$\hat{\gamma}_1$	-0.524	-0.024	0.048	0.687	0.687

Table 3. Summary statistics of carbon dioxide emissions.

Min	1st Quantile	Median	Mean	3rd Quantile	Max	Variance
0.100	1.400	3.400	4.953	6.500	29.800	27.597

After testing different combinations of link functions, the link functions that provided the best fit were the logarithm and probit link functions for the mean and precision submodels, respectively, resulting in the regression structures stated as

$$\log(\mu_t) = \beta_1 \text{FAR}_t + \beta_2 \text{CIN}_t + \beta_3 \text{EAG}_t \quad \text{and} \quad \Phi^{-1}(\phi_t) = \gamma_1 \text{HDI}_t, \quad t = 1, 2, \dots, 123.$$

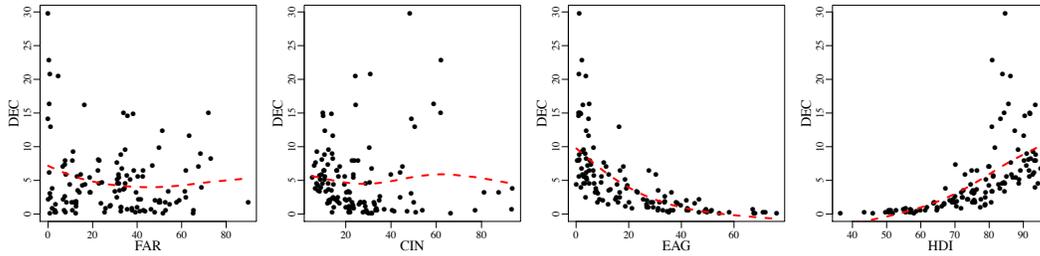


Figure 1. Plot for DEC versus FAR, CIN, EAG and HDI with corresponding smooth curves.

We compare the fitted Lomax regression model with the reparametrized gamma, reparametrized Weibull, and normal linear regression models using the `gamlss` package (Stasinopoulos and Rigby, 2007) in R. Some information about the regression structure of these models is summarized in Table 4.

Table 4. Regression structures for the gamma, Weibull, and normal models, with μ and σ^2 representing the mean and variance of the distribution, respectively.

Distribution	Reparametrization	Link function
$\text{gamma}(\theta_1, \theta_2)$	$\mu = \theta_1 \theta_2$ $\sigma = \theta_2 \sqrt{\theta_1}$	$g_1(\mu) = \log(\mu)$ $g_2(\sigma) = \log(\sigma)$
$\text{Weibull}(\theta_1, \theta_2)$	$\mu = \theta_1 \Gamma(1 + 1/\theta_2)$ $\sigma = \theta_2$	$g_1(\mu) = \log(\mu)$ $g_2(\sigma) = \log(\sigma)$
$\text{normal}(\theta_1, \theta_2^2)$	$\mu = \theta_1$ $\sigma = \theta_2$	$g_1(\mu) = \mu$ $g_2(\sigma) = \log(\sigma)$

Table 5 presents the parameter estimates, corresponding standard errors (SE), p -values associated with hypothesis testing based on the Wald square root statistic, and model selection criteria for the four fitted regression models. For comparison purposes, we fitted the reparametrized gamma, reparameterized Weibull, and normal linear regression models considering the same covariates. The two information criteria evaluated indicate that the Lomax regression model presented a better fit when compared to the other models. Considering the nature of the response variable, the Gamma and Weibull distributions are usual models for modeling continuous and positive data and that compete with the Lomax distribution, as their densities can assume a decreasing format. This is confirmed by the observed values of AIC and BIC. The normal model is one of the best known and most widely used in practice, but it is suitable for data with the supported in reals, and its probability density function does not assume a decreasing format. The values of AIC and BIC confirm that the normal model is not a competing model of the proposal presented here.

Table 5. Fit regression models for carbon emissions data.

Model	Effect	Parameter	Estimate (SE)	<i>p</i> -value	AIC	BIC
Lomax	FAR	β_1	0.0305 (0.0036)	< 0.01	650.94	662.19
	CIN	β_2	0.0466 (0.0045)	< 0.01		
	EAG	β_3	-0.0424 (0.0048)	< 0.01		
	HDI	γ_1	0.0174 (0.0090)	0.0530		
Gamma	FAR	β_1	0.0304 (0.0036)	< 0.01	652.01	663.26
	CIN	β_2	0.0468 (0.0044)	< 0.01		
	EAG	β_3	-0.0427 (0.0047)	< 0.01		
	HDI	γ_1	0.0002 (0.0007)	0.7658		
Weibull	FAR	β_1	0.0304 (0.0036)	< 0.01	651.97	663.22
	CIN	β_2	0.0467 (0.0044)	< 0.01		
	EAG	β_3	-0.0427 (0.0047)	< 0.01		
	HDI	γ_1	-0.0003 (0.0008)	0.7251		
Normal	FAR	β_1	0.0509 (0.0163)	< 0.01	742.36	753.60
	CIN	β_2	0.0959 (0.0189)	< 0.01		
	EAG	β_2	-0.0724 (0.0172)	< 0.01		
	HDI	γ_1	0.0211 (0.0008)	< 0.01		

Figure 2 presents the half-normal plots with simulated envelopes for the randomized quantile residuals based on 100 replicates for the considered models. From these plots, we note that, except for the normal regression model (Figure 2 (d)), almost all observations appear inside the envelope bands, indicating a good fit of the regression models to the the carbon dioxide emissions per capita (Atkinson, 1981). Figure 3 presents the residuals against the index and estimated probability density function of the residuals in a non-parametric way against the normal standard probability density function. As expected, the residuals seems to be oscillating around zero with constant variance and approximately normally distributed.

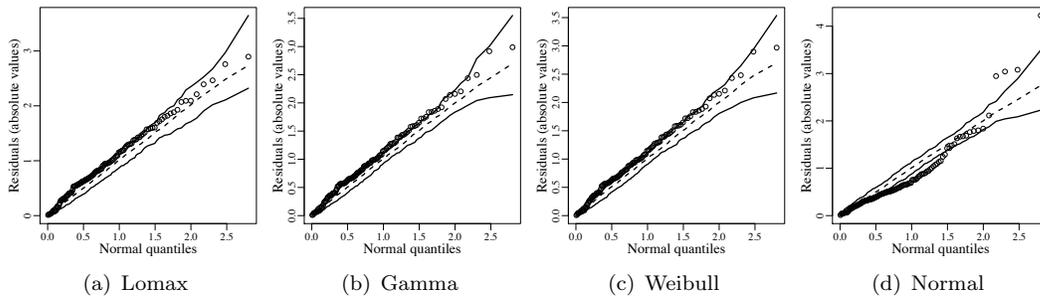


Figure 2. Half-normal plot of residuals for the fitted models in this study.

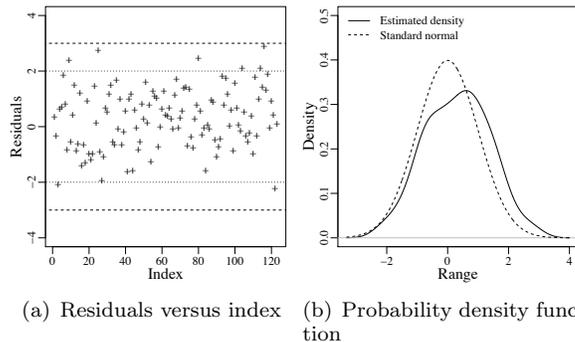


Figure 3. Residual plots for the proposed Lomax regression model.

The interpretation of the estimated parameters of the Lomax regression model are as follows:

- (i) For each 1% that on forest area increases, the mean of the carbon dioxide increases by 3.10% ($e^{\hat{\beta}_1} = 1.0310$).
- (ii) For each 1% that on concentration index of exports increases, the mean of the carbon dioxide increases by 4.77% ($e^{\hat{\beta}_2} = 1.0477$).
- (iii) For each 1% that the employment in agriculture increases, the mean of the carbon dioxide decreases by 4.15% ($e^{\hat{\beta}_3} = 0.9585$).
- (iv) The coefficient of γ_1 is 0.0174, so when the HDI increases, the precision increases.

5. CONCLUSION

In this paper, we proposed a frequentist approach for the mean-parameterized Lomax regression model with varying precision. The main advantage of this reparametrization is its ability to model the mean directly. This makes the interpretation of the regression coefficients easier in terms of the expectation of the response variable and the proposed model more comparable with other models in the class of generalized linear models. The estimation of the regression model parameters is based on the maximum likelihood approach. We provided closed-form expressions for the score vector, observed information matrix, and Fisher information matrix. Through Monte Carlo simulations, we evaluated the asymptotic properties of maximum likelihood estimators. The simulation results showed that these estimators present a good performance. Finally, we illustrated the practical applicability of the proposed framework through an empirical application.

SUPPLEMENTARY MATERIALS

The computational routine implemented in R is available online at <https://gist.github.com/moizesmelo/75a365ed957ae1ddbd9da9c3852597f3>.

AUTHOR CONTRIBUTIONS Conceptualization, M.S.M., L.H.L.; Methodology, M.S.M., L.H.L.; Software, M.S.M., L.H.L., J.B.C.; Validation, M.S.M., L.H.L., J.B.C.; Formal analysis, M.S.M., L.H.L., J.B.C.; Investigation, M.S.M., L.H.L., J.B.C.; Data Curation, M.S.M., L.H.L., J.B.C.; Visualization, M.S.M., L.H.L., J.B.C.; Supervision, M.S.M., L.H.L., J.B.C.; Writing - original draft, M.S.M., L.H.L., J.B.C.; Writing - review and editing, M.S.M., L.H.L., J.B.C. All authors have read and agreed the published version of the manuscript.

ACKNOWLEDGEMENTS We thank all the corrections and suggestions pointed out by the Editors-in-Chief and the reviewers.

FUNDING The authors gratefully acknowledge partial financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

DECLARATION OF CONFLICT OF INTEREST The authors declare no conflict of interest.

REFERENCES

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723
- Altun, E., 2021. The Lomax regression model with residual analysis: An application to insurance data. *Journal of Applied Statistics*, 48, 13–15.
- Atkinson, A.B. and Harrison, A.J., 1978. *Distribution of personal wealth in Britain*. Cambridge University Press, Cambridge, UK.
- Atkinson, A.C., 1981. Two graphical displays for outlying and influential observations in regression. *Biometrika*, 68, 13–20.
- Atkinson, A.C., 1985. *Plots, Transformation and Regression*. Clarendon Press, Oxford, UK.
- Beirlant, J. and Goegebeur, Y., 2003. Regression with response distributions of Pareto-type. *Computational Statistics and Data Analysis*, 42, 595–619.
- Bourguignon, M. and Gallardo, D.I., 2020. Reparameterized inverse gamma regression models with varying precision. *Statistica Neerlandica*, 74, 611–627.
- Bourguignon, M., Leão, J., and Gallardo, D.I., 2020. Parametric modal regression with varying precision. *Biometrical Journal*, 62, 202–220.
- Bourguignon, M. and Nascimento, F.F., 2020. Regression models for exceedance data: A new approach. *Statistical Methods and Applications*, 30, 1–17.
- Chandra, N. and Khan, M.A., 2013. Optimum plan for step-stress accelerated life testing model under type-i censored samples. *Journal of Modern Mathematics and Statistics*, 7, 58–62.
- Corbellini, A., Crosato, L., Ganugi, P., and Mazzoli, M., 2007. Fitting Pareto II distributions on firm size: Statistical methodology and economic puzzles. *Advances in Data*. Birkhauser, Boston, US, 26, 321–328.
- Das, S., Harel, O., Dey, D.K., Covault, J. and Kranzler, H.R., 2010. Analysis of extreme drinking in patients with alcohol dependence using Pareto regression. *Statistics in Medicine*, 29, 1250–1258.
- Dunn, P.K. and Smyth, G.K., 1996. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5, 236–244.
- Fernández, L.M.Z. and De Andrade, T.A.N., 2020. The erf-G family: New unconditioned and log-linear regression models. *Chilean Journal of Statistics*, 11, 3–23.
- Ferrari, S.L.P., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31, 799–815.
- Fonseca, R.V., Nobre, J.S., and Farias, R.B.A., 2016. Comparative inference and diagnostic in a reparameterized Birnbaum-Saunders regression model. *Chilean Journal of Statistics*, 7, 17–30.
- Harris, C.M., 1968. The Pareto distribution as a queue service discipline. *Operations Research*, 16, 307–313.
- Holland, O., Golaup, A., and Aghvami, A., 2006. Traffic characteristics of aggregated module downloads for mobile terminal reconfiguration. *IEE Proceedings Communications*, 153, 683–690.
- Lomax, K., 1954. Business failures: Another example of the analysis of failure data. *Journal of the American Statistical Association*, 49, 847–852.
- McCullagh, P. and Nelder, J., 1989. *Generalized Linear Models*. Chapman and Hall, Boca Raton, FL, US.
- Nelder, J.A. and Mead, R., 1965. A simplex method for function minimization. *The Computer Journal*, 7, 308–313.
- Palm, B.G., Bayer, F.M., Cintra, R.J., Pettersson, M.I., and Machado, R., 2019. Rayleigh regression model for ground type detection in SAR imagery. *IEEE Geoscience and Remote Sensing Letters*, 16, 1660–1664.

- Pawitan, Y., 2001. In all likelihood: Statistical modelling and inference using likelihood. Oxford University Press, Oxford, UK.
- R Development Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.
- Santos-Neto, M., Cysneiros, F.J.A., Leiva, V., and Barros, M., 2016. Reparameterized Birnbaum-Saunders regression models with varying precision. *Electronic Journal of Statistics*, 10, 2825–2855.
- Schwarz, G., 1978. Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464.
- Simas, A.B., Barreto-Souza, W., and Rocha, A.V., 2010. Improved estimators for a general class of beta regression models. *Computational Statistics and Data Analysis*, 54, 348–366.
- Smyth, G.K., 1989. Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society B*, 51, 47–60.
- 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.
- Vigas, V.P., Silva, G.O. and Louzada, F., 2017. The Poisson-Weibull regression model. *Chilean Journal of Statistics*, 8, 25–51.

STATISTICAL MODELING
RESEARCH PAPER

On an elliptical thin-plate spline partially varying-coefficient model

MAGALY S. MORAGA¹, GERMÁN IBACACHE-PULGAR^{2,3,*}, and ORIETTA NICOLIS⁴

¹Instituto de Estadística, Universidad Austral de Chile, Valdivia, Chile,

²Instituto de Estadística, Universidad de Valparaíso, Valparaíso, Chile,

³Centro Interdisciplinario de Estudios Atmosféricos y Astroestadística,
Universidad de Valparaíso, Valparaíso, Chile,

⁴Facultad de la Ingeniería, Universidad Andrés Bello, Viña del Mar, Chile

(Received: 03 August 2021 · Accepted in final form: 03 November 2021)

Abstract

In this work, we study the thin-plate spline partially varying-coefficient models with elliptical contoured errors in order to allow distributions with heavier and lighter tails than the normal ones, such as logistic, Pearson VII, power exponential, and Student-t, to be considered. We develop an estimation process for the parameters of the model based on the doubly penalized likelihood function and using smoothing splines. In addition, an explicit conditional solution for the double penalized maximum likelihood estimators is derived to obtain closed expressions for the variance-covariance matrix of the estimators, effective degrees of freedom of the smooth functions and surfaces, and hat matrix associated with the model. To show the proposed methodology, we analyze the Boston housing data utilizing-plate spline partially varying-coefficient model with normal and Student-t errors. This analysis suggests that the proposed model is helpful when we want to describe the effect of some covariates that vary smoothly as a function of other covariates, geographic referencing, and data with heavy-tailed indications.

Keywords: Maximum doubly penalized likelihood estimates · Partially varying-coefficient models · Robust estimates · Thin-plate spline models

Mathematics Subject Classification: Primary 62J02 · Secondary 62J12.

1. INTRODUCTION

Partially varying-coefficient models have received much attention in various research areas, due to its flexibility to explore the dynamic features which may exist the data and its easy interpretation. In the others words, this class of models allows to model the coefficients of the explanatory variables (or covariates) as smooth functions of other variables. These models are often used in research related to longitudinal, clustered, spatial and hierarchical sampling schemes, and are a natural alternative to the additive model introduced by [Breiman and Friedman \(1985\)](#); see also [Hastie and Tibshirani \(1993\)](#), [Fan and Zhang \(2008\)](#) and [Park et al. \(2015\)](#).

* Corresponding author: gibacachepulgar@gmail.com

Another aspect in the statistical literature, that has been developed in recent years, refers to the regression models under elliptical errors. These models suggest to replace the normal distribution by the elliptical one when the observations distributions are characterized by light-and heavy-tails. Savalli et al. (2006) proposed the elliptical linear mixed models, where the marginal model is also elliptical. Russo et al. (2009) extended the class given by Savalli et al. (2006) replacing linear fixed effects by a nonlinear fixed effect, creating the elliptical nonlinear mixed models, for which estimation procedures and diagnostic methods are developed. Galea and Vilca (2010) studied some hypothesis tests for the equality of variances and means in the context of univariate elliptical correlated data, with applications to portfolios data. Marciano et al. (2016) studied the calibration models for repeated measures considering a univariate elliptical distribution and developed a simulation study to evaluate the properties of the estimators. Ibacache-Pulgar and Paula (2011) presented a study on the existence and uniqueness of the maximum penalized likelihood estimate under the partially linear model with Student-t random error, Ibacache-Pulgar et al. (2012) developed influence diagnostics for elliptical semiparametric mixed models, where it is assumed that the non-parametric component is of type cubic spline, and Ibacache-Pulgar et al. (2013) studied semiparametric additive model under symmetric distributions. Recently, Ibacache-Pulgar and Reyes (2018) studied the elliptical partially varying-coefficient models and developed the technique of local influence to evaluate the sensitivity of the maximum penalized likelihood estimates.

In this paper, we extend the partially varying-coefficient model proposed by Ibacache-Pulgar and Reyes (2018) incorporating a component in its regression structure that allows us to model the effect of observations in two-dimensional space, such as, for example, coordinates. This structure is called thin-plate spline partially varying-coefficient model under elliptical errors. This model emerges as a powerful tool in statistical modeling because of its flexibility to model explanatory variables effects that can contribute parametric way and explanatory variables effects in which the coefficients are allowed to vary as smooth functions of other variables. Moreover, this class of models incorporate thin-plate spline (TPS) smoother, a spline-based technique which can be considered the natural generalization of cubic spline to any number of dimensions and almost any order of wiggleness penalty. The TPS smoother was initially introduced by Duchon (1975) and was later considered by many authors in the context of nonparametric and generalized linear models; see, Green and Silverman (1994) and Wood (2006) and the references therein. Since the TPS involves the estimation of many parameters (especially when the dimension is higher than one), Wood (2003) proposed a low rank smoother that use an approximate thin plate spline model based on the transformation and truncation of the basis that arises from the solution of the thin plate spline smoothing problem. The main advantage to include TSP in our model is that it allows to consider the effect of the geographical locations on the response variable.

This article is organized as follows. In Section 2, we formally introduce the thin-plate spline partially varying-coefficient model under elliptical distributions. Section 3 considers the problem of estimating the parameters and an application to a set of real data is considered in Section 4. Finally, in Section 5, we present some final conclusions derived from this study.

2. THE THIN-PLATE SPLINE PARTIALLY VARYING-COEFFICIENTS MODEL

In this section, we introduce the thin-plate spline partially varying-coefficient model (TP-SPVCM) under elliptical distributions. In addition, we introduce the doubly penalized likelihood function where the penalty term combines a $\mathcal{L}^2[a, b]$ penalty for each smooth varying-coefficient function with a second $\mathcal{L}^2[E^d]$ penalty for the smooth surface. Thus, we estimate the parameters and inference in the elliptical TPSPVCM.

2.1 MODEL SPECIFICATION

The study of varying-coefficients models (VCMs) does not necessarily arise from performing a mathematical extension of a particular class of models, but rather from the need to attend to real problems in areas as economics, finance, epidemiology, medical science, ecology, and environment. The TPSPVCM under study is given by

$$y_{ij} = \mathbf{z}_{ij}^\top \boldsymbol{\alpha} + \sum_{k=1}^s x_{ij}^{(k)} \beta_k(r_{k_{ij}}) + \boldsymbol{\ell}_i^\top \mathbf{g} + \varepsilon_{ij}, \quad i = 1, \dots, n, j = 1, \dots, m_i, \quad (1)$$

where y_{ij} denotes the j th measure associated with the i th cluster at point $r_{k_{ij}}$, \mathbf{z}_{ij} is $(p \times 1)$ vector of explanatory variable values, $\boldsymbol{\alpha}$ is a $(p \times 1)$ fixed parameter vector, β_k , for $k = 1, \dots, s$, are unknown smooth arbitrary functions of r_k , associated with the covariates $x_{ij}^{(k)}$, $\boldsymbol{\ell}_i$ is an $(n \times 1)$ vector with one in the i th position and zeros at the remaining positions, $\mathbf{g} = (g(\mathbf{t}_1), \dots, g(\mathbf{t}_n))^\top$, g is a smooth surface that depends of the vector $\mathbf{t}_i \in \mathcal{R}^2$, and ε_{ij} is a random error. Note that in this class of models the coefficients are allowed to vary as smooth functions of other variables.

To write the model given in Equation (1) in a matrix form, first consider the one-to-one linear transformation of the vector \mathbf{g} suggested by Green and Silverman (1994) stated as

$$\mathbf{g} = \begin{pmatrix} g(\mathbf{t}_1) \\ \vdots \\ g(\mathbf{t}_n) \end{pmatrix} = \mathbf{E}\boldsymbol{\delta} + \mathbf{T}^\top \mathbf{a},$$

where \mathbf{a} and $\boldsymbol{\delta}$ are vectors with components a_i and δ_i , \mathbf{E} is an $(n \times n)$ matrix defined by $E_{ij} = 1/(16\pi\|\mathbf{t}_i - \mathbf{t}_j\|^2 \log(\|\mathbf{t}_i - \mathbf{t}_j\|^2))$, with $E_{ii} = 0$ for each i , and \mathbf{T} is a $(3 \times n)$ matrix given by

$$\mathbf{T} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \mathbf{t}_1 & \mathbf{t}_2 & \dots & \mathbf{t}_n \end{pmatrix}.$$

Thus, the model model given in Equation (1) takes the form

$$\mathbf{y}_i = \tilde{\mathbf{Z}}_i \tilde{\boldsymbol{\alpha}} + \sum_{k=1}^s \tilde{\mathbf{N}}_{ki} \boldsymbol{\beta}_k + \tilde{\mathbf{E}}_i \boldsymbol{\delta} + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n, j = 1, \dots, m_i, \quad (2)$$

where \mathbf{y}_i is a $(m_i \times 1)$ random vector of observed responses from the i th cluster, $\tilde{\mathbf{Z}}_i = (\mathbf{Z}_i \ \tilde{\mathbf{T}}_i)$ is an $(m_i \times (p + 2))$ design matrix, \mathbf{Z}_i is an $(m_i \times p)$ design matrix with rows \mathbf{z}_{ij}^\top , $\tilde{\mathbf{T}}_i = \mathbf{F}_i \mathbf{T}^\top$ is an $(m_i \times 2)$ matrix, \mathbf{F}_i is an $(m_i \times n)$ matrix with an $(m_i \times 1)$ vector of ones in the i th column and zeros in the remaining positions, $\tilde{\boldsymbol{\alpha}}^\top = (\boldsymbol{\alpha}^\top, \mathbf{a}^\top)$, $\tilde{\mathbf{N}}_{ki} = \mathbf{X}_i^{(k)} \mathbf{N}_{ki}$, $\mathbf{X}_i^{(k)} = \text{diag}_{1 \leq j \leq m_i} (x_{ij}^{(k)})$, \mathbf{N}_{ki} is an $(m_i \times r_k)$ incidence matrix with the (j, l) th element equal to the indicator $I(r_{k_{ij}} = r_{k_l}^0)$, for $j = 1, \dots, m_i$, where $r_{k_l}^0$, for $l = 1, \dots, r_k$, denotes the distinct and ordered values of the explanatory variable $r_{k_{ij}}$, $\boldsymbol{\beta}_k = (\psi_{k_1}, \dots, \psi_{r_k})^\top$ is an $(r_k \times 1)$ vector of parameters with $\psi_{k_l} = \beta_k(r_{k_l}^0)$, for $l = 1, \dots, r_k$, $\tilde{\mathbf{E}}_i = \mathbf{F}_i \mathbf{E}$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{m_i})^\top$ is an $(m_i \times 1)$ vector of within-cluster errors. A compact way of writing model given in Equation (2) is formulated as

$$\mathbf{y} = \tilde{\mathbf{Z}} \tilde{\boldsymbol{\alpha}} + \tilde{\mathbf{N}}_1 \boldsymbol{\beta}_1 + \dots + \tilde{\mathbf{N}}_s \boldsymbol{\beta}_s + \tilde{\mathbf{E}} \boldsymbol{\delta} + \boldsymbol{\varepsilon}, \quad (3)$$

where $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$, $\tilde{\mathbf{Z}}$, $\tilde{\mathbf{N}}_k$, $\tilde{\mathbf{E}}$ and $\boldsymbol{\varepsilon}$ similarly.

2.2 DOUBLY PENALIZED LIKELIHOOD FUNCTION

Consider the model given by Equation (2) and assume that $\boldsymbol{\varepsilon}_i \sim \text{El}_{m_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$, with $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\tau})$ being a positive-definite matrix, with $\boldsymbol{\tau} = (\tau_1, \dots, \tau_d)^\top$. Thus, $\mathbf{y}_i \sim \text{El}_{m_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, with $\boldsymbol{\mu}_i = \tilde{\mathbf{Z}}_i \tilde{\boldsymbol{\alpha}} + \sum_{k=1}^s \tilde{\mathbf{N}}_{ki} \boldsymbol{\beta}_k + \tilde{\mathbf{E}}_i \boldsymbol{\delta}$, and density function stated as

$$f(\mathbf{y}_i) = |\boldsymbol{\Sigma}_i|^{-1/2} h(u_i), \quad i = 1, \dots, n, \quad (4)$$

where $u_i = \boldsymbol{\varepsilon}_i^\top \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i$ is the Mahalanobis distance, $\boldsymbol{\varepsilon}_i = \mathbf{y}_i - \boldsymbol{\mu}_i$, and h is a function of $\mathcal{R} \rightarrow [0, \infty]$ known as the density generator function (Fang et al., 1990). Then, the log-likelihood function of the model given in Equation (4) for $\boldsymbol{\theta} = (\tilde{\boldsymbol{\alpha}}^\top, \boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_s^\top, \boldsymbol{\delta}^\top, \boldsymbol{\tau}^\top)^\top$ is given by

$$L(\boldsymbol{\theta}) = \sum_{i=1}^n L_i(\boldsymbol{\theta}),$$

where $L_i(\boldsymbol{\theta}) = -(1/2) \log(|\boldsymbol{\Sigma}_i|) + \log(h(u_i))$ represents the individual contribution of the i th observation. Since the functions β_k 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 model. Therefore, we assume that the functions β_k (which are absolutely continuous) belong to the Sobolev function space stated as

$$\mathcal{W}_2^{(i)} = \{\beta_k: \beta_k, \beta_k^{(1)}, \dots, \beta_k^{(i-1)}, \beta_k^{(i)} \in \mathcal{L}^2[a_k, b_k]\}.$$

In addition, we assume that g belong to the functions space whose partial derivatives of total order m are in Hilbert space $\mathcal{L}^2[E^d]$ of square integrable functions on Euclidean d -space. Incorporating a penalty function over each function β_k and g , we have that the penalized log-likelihood function can be expressed as (Ibacache-Pulgar et al., 2013)

$$L_p(\boldsymbol{\theta}, \lambda_1, \dots, \lambda_s, \lambda_g) = L(\boldsymbol{\theta}) + \sum_{k=1}^s \lambda_k^* J(\beta_k) + \lambda_g^* J_m^d(g), \quad (5)$$

where $J(\beta_k)$ denotes the penalty functional over β_k , $J_m^d(g)$ is a penalty functional measuring the wiggleness of g , and $\lambda_k^* = \lambda^*(\lambda_k)$ and $\lambda_g^*(\lambda_g)$ are constants that depends on the smoothing parameters $\lambda_k \geq 0$ and $\lambda_g \geq 0$, respectively. In this paper, we consider as a measure of the curvature of β_k functions the squared norm expressed as

$$J(\beta_k) = \|\beta_k\|^2 = \int_{a_k}^{b_k} [\beta_k^{(i)}(r_k)]^2 dr_k,$$

where $\beta_k^{(i)}(r_k) = d^i \beta(r_k) / dr_k^i$, $r_{k_l}^0 \in [a_k, b_k]$, and

$$J_m^d(g) = \sum_{v_1 + \dots + v_d = m} \frac{m!}{v_1! \dots v_d!} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \left(\frac{\partial^m g}{\partial t_1^{\alpha_1} \dots \partial t_d^{\alpha_d}} \right)^2 \prod_{j=1}^d dt_j.$$

It is important mention that for $i = 2$, the estimation of β_k leads to a natural cubic spline with knots at the points $r_{k_l}^0$, for $l = 1, \dots, r_k$. In addition, for $d = 2, m = 2$ and

$g = g(t_1, t_2)$, that is,

$$J(g) = \int \int_{\mathcal{R}^2} \left\{ \left(\frac{\partial^2 g}{\partial t_1^2} \right)^2 + 2 \left(\frac{\partial^2 g}{\partial t_1 \partial t_2} \right)^2 + \left(\frac{\partial^2 g}{\partial t_2^2} \right)^2 \right\} dt_1 dt_2,$$

the estimation of g leads to a natural thin-plate spline. According to [Green and Silverman \(1994\)](#), we may express the penalty functional as

$$J(\beta_k) = \beta_k^\top \mathbf{K}_k \beta_k, \quad J(g) = \delta^\top \mathbf{E} \delta,$$

where \mathbf{K}_k is an $(q_k \times q_k)$ non-negative definite smoothing matrix associated with the k th explanatory variable that depends only on the knots. Then, if we consider $\lambda_k^* = -\lambda_k/2$ and $\lambda_g^* = -\lambda_g/2$, the penalized log-likelihood function given in Equation (5) can be expressed as

$$L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) = L(\boldsymbol{\theta}) - \sum_{k=1}^s \frac{\lambda_k}{2} \beta_k^\top \mathbf{K}_k \beta_k - \frac{\lambda_g}{2} \delta^\top \mathbf{E} \delta, \quad (6)$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_s, \lambda_g)^\top$ denotes an $((s+1) \times 1)$ vector of smoothing parameters that controls the tradeoff between goodness of fit and the smoothness estimated functions. Note that the first term in the right-hand side of Equation (6) measures the goodness of fit while the second and third terms penalizes the roughness of each β_k and g with a fixed parameter λ_k and λ_g , respectively. It should be noted that the choice of such parameters is crucial in the estimation process, since they controls the tradeoff between goodness of fit and the smoothness (regularity) estimated function. A more extensive discussion on the methods of selecting such parameters is presented later.

3. PARAMETERS ESTIMATION

The estimation problem in the context of TPSPVCM under elliptical distributions has not been discussed in the literature. However, several authors have considered this problem for some specific cases. For example, in the context of varying-coefficient model, [Cai et al. \(2000\)](#) estimated the coefficient functions based on local polynomial regression technique and proposed a method that involves solving hundreds of local likelihood equations through a one-one-step Newton-Raphson. [Chiang et al. \(2001\)](#) derived a componentwise smoothing spline procedure for the estimation of coefficient curves in a varying-coefficient model with repeatedly measured dependent variables; see also [Eubank et al. \(2004\)](#). [Krafty et al. \(2008\)](#) developed an estimation procedure of the coefficient functions when the within-subject covariance is unknown considering the criterion of iterative reweighted least squares. [Wang et al. \(2009\)](#) proposed an estimation method based on local ranks which is more efficient and robust compared to other methods such as local linear least squares method. [Liu and Li \(2015\)](#) estimated the coefficient curves in a varying-coefficient model for longitudinal data by using local polynomial smoothing method and showed that the resulting estimator is asymptotically more efficient than the ones which ignore the within-subject correlation structure. In this paper we propose to estimate the model parameters based on the work proposed by [Ibacache-Pulgar and Reyes \(2018\)](#), which consider to estimate the coefficient curves based on penalized likelihood criterion and smoothing spline.

3.1 ESTIMATION OF $\tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_s, \boldsymbol{\delta}$

To estimate the parameters $\tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_s, \boldsymbol{\delta}$ and $\boldsymbol{\tau}$ we propose to maximize the double penalized log-likelihood function assuming $\boldsymbol{\lambda}$ fixed, that is,

$$\max_{\tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_s, \boldsymbol{\delta}, \boldsymbol{\tau}} L_p(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \max_{\tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_s, \boldsymbol{\delta}, \boldsymbol{\tau}} L_p(\tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_s, \boldsymbol{\delta}, \boldsymbol{\tau}, \boldsymbol{\lambda}).$$

This procedure can be solved using the Fisher scoring algorithm (Ibacache-Pulgar and Reyes, 2018) stated as

$$\begin{pmatrix} \mathbf{I} & \mathbf{S}_0^{(u)} \tilde{\mathbf{N}}_1 & \dots & \mathbf{S}_0^{(u)} \tilde{\mathbf{N}}_s & \mathbf{S}_0^{(u)} \tilde{\mathbf{E}} \\ \mathbf{S}_1^{(u)} \tilde{\mathbf{N}}_0 & \mathbf{I} & \dots & \mathbf{S}_1^{(u)} \tilde{\mathbf{N}}_s & \mathbf{S}_1^{(u)} \tilde{\mathbf{E}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{S}_s^{(u)} \tilde{\mathbf{N}}_0 & \mathbf{S}_s^{(u)} \tilde{\mathbf{N}}_1 & \dots & \mathbf{I} & \mathbf{S}_s^{(u)} \tilde{\mathbf{E}} \\ \mathbf{S}_\delta^{(u)} \tilde{\mathbf{N}}_0 & \mathbf{S}_\delta^{(u)} \tilde{\mathbf{N}}_1 & \dots & \mathbf{S}_\delta^{(u)} \tilde{\mathbf{N}}_s & \mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_0^{(u+1)} \\ \boldsymbol{\beta}_1^{(u+1)} \\ \vdots \\ \boldsymbol{\beta}_s^{(u+1)} \\ \boldsymbol{\delta}^{(u+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_0^{(u)} \boldsymbol{\eta}^{(u)} \\ \mathbf{S}_1^{(u)} \boldsymbol{\eta}^{(u)} \\ \vdots \\ \mathbf{S}_s^{(u)} \boldsymbol{\eta}^{(u)} \\ \mathbf{S}_\delta^{(u)} \boldsymbol{\eta}^{(u)} \end{pmatrix}, \quad (7)$$

where $\boldsymbol{\beta}_0 = \tilde{\boldsymbol{\alpha}}$, $\tilde{\mathbf{N}}_0 = \tilde{\mathbf{Z}}$, $\boldsymbol{\eta}^{(u)} = \boldsymbol{\mu} + \mathbf{W}^{*-1} \mathbf{W}_v (\mathbf{y} - \boldsymbol{\mu})|_{\boldsymbol{\theta}^{(u)}}$ and $\mathbf{S}_k^{(u)} = \mathbf{S}_k|_{\boldsymbol{\theta}^{(u)}}$, with

$$\mathbf{S}_k^{(u)} = \begin{cases} (\tilde{\mathbf{N}}_0^\top \mathbf{W}^* \tilde{\mathbf{N}}_0)^{-1} \tilde{\mathbf{N}}_0^\top \mathbf{W}^*|_{\boldsymbol{\theta}^{(u)}}, & k = 0, \\ (\tilde{\mathbf{N}}_k^\top \mathbf{W}^* \tilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1} \tilde{\mathbf{N}}_k^\top \mathbf{W}^*, & k = 1, \dots, s, \end{cases}$$

and

$$\mathbf{S}_\delta^{(u)} = (\tilde{\mathbf{E}}^\top \mathbf{W}^* \tilde{\mathbf{E}} + \lambda_g \tilde{\mathbf{E}})^{-1} \tilde{\mathbf{E}}^\top \mathbf{W}^*,$$

where \mathbf{W}^* and \mathbf{W}_v are defined in the appendix. Then, the back-fitting (Gauss-Seidel) iterations that are used to solve the system stated in Equation (7) take the form

$$\boldsymbol{\beta}_k^{(u+1)} = \mathbf{S}_k^{(u)} \left(\boldsymbol{\eta}^{(u)} - \sum_{l=0, l \neq k}^s \tilde{\mathbf{N}}_l \boldsymbol{\beta}_l^{(u)} - \tilde{\mathbf{E}} \boldsymbol{\delta}^{(u)} \right), \quad k = 0, 1, \dots, s, \quad (8)$$

$$\boldsymbol{\delta}^{(u+1)} = \mathbf{S}_\delta^{(u)} \left(\boldsymbol{\eta}^{(u)} - \sum_{l=0}^s \tilde{\mathbf{N}}_l \boldsymbol{\beta}_l^{(u)} \right). \quad (9)$$

From the convergence of the iterative process given in Equation (8), we obtain the maximum double penalized likelihood estimator (MDPLE) of $\boldsymbol{\beta}_k$ and $\boldsymbol{\delta}$, which leads to a natural cubic spline estimate for $\boldsymbol{\beta}_k$ ($k = 1, \dots, s$).

It is important to note that in the iterative process above the parameter estimates depend on the smoothing matrices \mathbf{S}_k and \mathbf{S}_δ , the modified variable $\boldsymbol{\eta}$ and the partial residuals; see Equation 8. In addition, the weights v_i have an influence on the estimates of $\boldsymbol{\beta}_k$, for $k = 0, 1, \dots, s$, and $\boldsymbol{\delta}$. In particular, it can be shown that for the Student-t and power exponential distributions, for example, the current weight $v_i^{(r)} = v_i|_{\boldsymbol{\theta}^{(r)}}$ is inversely proportional to the Mahalanobis distance between the observed value \mathbf{y}_i and its current predicted value $\boldsymbol{\mu}_i^{(r)} = \boldsymbol{\mu}_i|_{\boldsymbol{\theta}^{(r)}}$, so that outlying observations tend to have small weights in the estimation process.

3.2 ESTIMATION OF τ

Regarding the MDPLE of τ , this can be obtained using the Fisher scoring algorithm formulated as

$$\tau^{(u+1)} = \tau^{(u)} - E \left\{ \frac{\partial^2 L_P(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^\top} \right\}^{-1} \frac{\partial L_P(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\tau}} \Big|_{\boldsymbol{\theta}^{(u)}}. \tag{10}$$

An iterative process to solve Equations (8) and (10) simultaneously is described in the appendix.

3.3 ESTIMATION OF THE SURFACE

In Section 2, we represent the surface g as a linear combination of the coefficient vectors $\boldsymbol{\delta}$ and \mathbf{a} . Considering the MDPLEs obtained through the iterative process described above, that is, $\widehat{\boldsymbol{\delta}}$ and $\widehat{\mathbf{a}}$, we have that the MDPLE, \widehat{g} , can be obtained as

$$\widehat{g} = \mathbf{E} \widehat{\boldsymbol{\delta}} + \mathbf{T}^\top \widehat{\mathbf{a}}. \tag{11}$$

Consequently, the estimator of the surface g is a natural thin-plate spline. Details of the conditions that guarantee this result are given, for example, in [Green and Silverman \(1994\)](#).

3.4 A CONDITIONAL EXPLICIT SOLUTION

Note that β_j , for $j = 0, 1, \dots, s$, and $\boldsymbol{\delta}$ can be estimated through the solutions to the set of normal equations ([Buja et al., 1989](#); [Opsomer and Ruppert, 1999](#)) derived from double penalized log-likelihood function. Indeed, taking partial derivatives of Equation (6) with respect to the parameter $\beta_0, \beta_1, \dots, \beta_s, \boldsymbol{\delta}$ and to equating zero, we obtain

$$\left. \begin{aligned} \mathbf{Z}^\top \mathbf{W}_v \boldsymbol{\varepsilon} &= \mathbf{0}, \\ \widetilde{\mathbf{T}} \mathbf{W}_v \boldsymbol{\varepsilon} &= \mathbf{0}, \\ \widetilde{\mathbf{N}}_1^\top \mathbf{W}_v \boldsymbol{\varepsilon} - \lambda_1 \mathbf{K}_1 \boldsymbol{\beta}_1 &= \mathbf{0}, \\ &\vdots \\ \widetilde{\mathbf{N}}_s^\top \mathbf{W}_v \boldsymbol{\varepsilon} - \lambda_1 \mathbf{K}_s \boldsymbol{\beta}_s &= \mathbf{0}, \\ \widetilde{\mathbf{E}} \mathbf{W}_v \boldsymbol{\varepsilon} - \lambda_g \mathbf{E} \boldsymbol{\delta} &= \mathbf{0}. \end{aligned} \right\} \tag{12}$$

From Equation (12) it is possible, at least conceptually, to derive an explicit expression for the estimates $\widehat{\boldsymbol{\beta}}_j$ ($j = 0, 1, \dots, s$) and $\widehat{\boldsymbol{\delta}}$ under some assumptions. For simplicity of notation consider $\boldsymbol{\beta}_{s+1} = \boldsymbol{\delta}$, $\mathbf{S}_{s+1} = \mathbf{S}_\delta$ and $p' = p + 2$, and assume $\boldsymbol{\lambda}$, \mathbf{W}_v and \mathbf{W}^* fixed, we can write the estimating equation system given in Equation (12) as

$$\begin{pmatrix} \mathbf{I}_{p'} & \mathbf{S}_0 \widetilde{\mathbf{N}}_1 & \dots & \mathbf{S}_0 \widetilde{\mathbf{N}}_s & \mathbf{S}_0 \widetilde{\mathbf{E}} \\ \mathbf{S}_1 \widetilde{\mathbf{N}}_0 & \mathbf{I}_{r_1} & \dots & \mathbf{S}_1 \widetilde{\mathbf{N}}_s & \mathbf{S}_1 \widetilde{\mathbf{E}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{S}_s \widetilde{\mathbf{N}}_0 & \mathbf{S}_s \widetilde{\mathbf{N}}_1 & \dots & \mathbf{I}_{r_s} & \mathbf{S}_s \widetilde{\mathbf{E}} \\ \mathbf{S}_{s+1} \widetilde{\mathbf{N}}_0 & \mathbf{S}_{s+1} \widetilde{\mathbf{N}}_1 & \dots & \mathbf{S}_{s+1} \widetilde{\mathbf{N}}_s & \mathbf{I}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_s \\ \boldsymbol{\beta}_{s+1} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_0 \\ \mathbf{S}_1 \\ \vdots \\ \mathbf{S}_s \\ \mathbf{S}_{s+1} \end{pmatrix} \mathbf{y}.$$

In practice, this system of equations is solved iteratively through a backfitting algorithm, and its backfitting estimators converge to the solution (Buja et al., 1989) stated as

$$\begin{pmatrix} \widehat{\beta}_0, \\ \widehat{\beta}_1, \\ \vdots \\ \widehat{\beta}_s, \\ \widehat{\beta}_{s+1}, \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{p'} & \mathbf{S}_0 \widetilde{\mathbf{N}}_1 & \dots & \mathbf{S}_0 \widetilde{\mathbf{N}}_s & \mathbf{S}_0 \widetilde{\mathbf{E}} \\ \mathbf{S}_1 \widetilde{\mathbf{N}}_0 & \mathbf{I}_{r_1} & \dots & \mathbf{S}_1 \widetilde{\mathbf{N}}_s & \mathbf{S}_1 \widetilde{\mathbf{E}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{S}_s \widetilde{\mathbf{N}}_0 & \mathbf{S}_s \widetilde{\mathbf{N}}_1 & \dots & \mathbf{I}_{r_s} & \mathbf{S}_s \widetilde{\mathbf{E}} \\ \mathbf{S}_{s+1} \widetilde{\mathbf{N}}_0 & \mathbf{S}_{s+1} \widetilde{\mathbf{N}}_1 & \dots & \mathbf{S}_{s+1} \widetilde{\mathbf{N}}_s & \mathbf{I}_n \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{S}_0 \\ \mathbf{S}_1 \\ \vdots \\ \mathbf{S}_s \\ \mathbf{S}_{s+1} \end{pmatrix} \mathbf{y} \equiv \mathbf{M}^{-1} \mathbf{S} \mathbf{y},$$

if the inverse of \mathbf{M} exists. Consequently, the backfitting estimator for $\widehat{\beta}_j$ ($j = 0, 1, \dots, s+1$) can be obtained directly as (Opsomer and Ruppert, 1999)

$$\widehat{\beta}_j = \mathcal{H}_j \mathbf{y}, \quad j = 0, 1, \dots, s+1, \quad (13)$$

where $\mathcal{H}_j = \mathbf{E}_j \mathbf{M}^{-1} \mathbf{S}$ is the smoother matrix obtained when fitting by smoothing spline the j th explanatory variable only, with \mathbf{E}_j being a partitioned matrix given by

$$\mathbf{E}_j = \begin{cases} \left(\mathbf{I}_{(p' \times p')} \mathbf{0}_{(p' \times r_1)} \dots \mathbf{0}_{(p' \times r_s)} \mathbf{0}_{(p' \times n)} \right), & j = 0, \\ \left(\mathbf{0}_{(r_1 \times p')} \mathbf{I}_{(r_1 \times r_1)} \dots \mathbf{0}_{(r_1 \times r_s)} \mathbf{0}_{(r_1 \times n)} \right), & j = 1, \\ \vdots & \vdots \\ \left(\mathbf{0}_{(r_s \times p')} \mathbf{0}_{(r_s \times r_1)} \dots \mathbf{I}_{(r_s \times r_s)} \mathbf{I}_{(r_s \times n)} \right), & j = s, \\ \left(\mathbf{0}_{(n \times p')} \mathbf{0}_{(n \times r_1)} \dots \mathbf{I}_{(n \times r_s)} \mathbf{I}_{(n \times n)} \right), & j = s+1. \end{cases}$$

The direct calculation of the MDPLEs from Equation (13) is rarely used in practice, because the backfitting algorithm is more efficient for obtaining $\widehat{\beta}_j$; it does not require high-dimensional matrices and their inverses. However, the above expressions can be useful if we wish to study some theoretical properties of the MDPLEs and carry out a diagnostic analysis based on the hat matrix associated with the model fit. Some closed expressions for the estimators in the context of the semiparametric additive models can be found, for example, in Ibacache-Pulgar et al. (2013).

3.5 ESTIMATION OF THE STANDARD ERRORS

We consider in this section the problem of how to derive the variance-covariance matrix of the MDPLE $\boldsymbol{\theta}$. According to Segal et al. (1994), the variance estimates for the MDPLEs developed by Wahba (1983) and Silverman (1985), under the Bayesian context, correspond to the inverse of the observed information matrix obtained by treating the penalized likelihood as a usual likelihood. Therefore, if we obtain the MDPLE of $\boldsymbol{\theta}$ through the Fisher scoring algorithm, it is reasonable to derive the variance-covariance matrix by using the inverse of the penalized Fisher information matrix. Thus, the asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\theta}}$ can be obtained from the inverse of the expected information matrix \mathcal{I}_p defined in the appendix, that is,

$$\widehat{\text{Cov}}_{\text{asymptotic}}(\widehat{\boldsymbol{\theta}}) \approx \mathcal{I}_p^{-1}(\widehat{\boldsymbol{\theta}}). \quad (14)$$

By using variance-covariance matrix given in Equation (14) we can construct an approximate pointwise standard error band (SEB) for β_k that allows us to assess how accurate the estimator $\widehat{\beta}_k$ at different locations within the range of interest. For example, we can consider the approximate pointwise SEB given by

$$\text{SEB}_{\text{approx}}(\beta_k(r_{k_l}^0)) = \widehat{\beta}_k(r_{k_l}^0) \pm 2\sqrt{\widehat{\text{Var}}(\widehat{\beta}_k(r_{k_l}^0))},$$

where $\widehat{\text{Var}}(\widehat{\beta}_k(r_{k_l}^0))$ is the l th principal diagonal element of the matrix given in Equation (14), for $l = 1, \dots, r_k$.

Note from Equations (13) and (11) that it is possible to obtain the covariance matrix for β_j ($j = 0, \dots, s + 1$) and $\widehat{\mathbf{g}}$, respectively. Indeed,

$$\widehat{\text{Cov}}(\widehat{\beta}_j) = \mathbf{H}_j \widehat{\text{Cov}}(\mathbf{y}) \mathbf{H}_j^\top$$

and

$$\widehat{\text{Cov}}(\widehat{\mathbf{g}}) = \mathbf{H}_g \widehat{\text{Cov}}(\mathbf{y}) \mathbf{H}_g^\top,$$

where $\mathbf{H}_g = \mathbf{E} \mathbf{H}_{s+1} + \mathbf{T}^\top \mathbf{H}'$, with \mathbf{H}_{s+1} defined above and \mathbf{H}' denoting the block of matrix \mathbf{H}_0 corresponding to vector \mathbf{a} , $\widehat{\text{Cov}}(\mathbf{y}) = \text{blockdiag}_{1 \leq i \leq n}(\xi_i \boldsymbol{\Sigma}_i)$ and $\xi_i > 0$ is a quantity that may be obtained from the derivatives of the characteristic function associated with elliptical distributions (Fang et al., 1990).

3.6 EFFECTIVE DEGREES OF FREEDOM

In general, in the literature concerning semiparametric models there are different definitions for the degrees of freedom (DF), depending on the context in which they are used. Here, the DF associated with the smooth varying-coefficient functions is defined as (Hastie and Tibshirani, 1990)

$$\begin{aligned} \text{DF}(\lambda_k) &= \text{tr}\{\widetilde{\mathbf{N}}_k \mathbf{S}_k\} \\ &= \text{tr}\left\{\widetilde{\mathbf{N}}_k^\top \mathbf{W}^* \widetilde{\mathbf{N}}_k \left(\widetilde{\mathbf{N}}_k^\top \mathbf{W}^* \widetilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k\right)^{-1}\right\}. \end{aligned}$$

In practice, it is desirable to have an approximation to this quantity. Let $\mathbf{Q}_{\mathbf{N}_k} = \widetilde{\mathbf{N}}_k^\top \mathbf{W}^* \widetilde{\mathbf{N}}_k$ and $\mathbf{Q}_{\lambda_k} = \lambda_k \mathbf{K}_k$. Since $\mathbf{W}^* > 0$ and $\text{rank}(\widetilde{\mathbf{N}}_k^\top) \leq r_k$, then $\mathbf{Q}_{\widetilde{\mathbf{N}}_k} \geq 0$. Therefore, there exists a matrix $\mathbf{Q}_{\widetilde{\mathbf{N}}_k}^{1/2} \geq 0$ such that $\mathbf{Q}_{\widetilde{\mathbf{N}}_k} = \mathbf{Q}_{\widetilde{\mathbf{N}}_k}^{1/2} \mathbf{Q}_{\widetilde{\mathbf{N}}_k}^{1/2}$. Thus, we can write $\text{tr}\{\widetilde{\mathbf{N}}_k \mathbf{S}_k\} = \text{tr}\{\widetilde{\mathbf{S}}_k\}$ as (Eilers and Marx, 1996)

$$\text{tr}\{\widetilde{\mathbf{S}}_k\} = \sum_{j=1}^{r_k} \frac{1}{1 + \lambda_k \ell_j},$$

where ℓ_j , for $j = 1, \dots, r_k$, are the eigenvalues of the matrix $\mathbf{Q}_{\widetilde{\mathbf{N}}_k}^{-1/2} \mathbf{Q}_{\lambda_k} \mathbf{Q}_{\widetilde{\mathbf{N}}_k}^{-1/2}$, for $k = 1, \dots, s$. Analogously to the selection of DFs associated with smooth varying-coefficient

functions, the DFs associated with smooth surface is given by

$$\begin{aligned} \text{DF}(\lambda_g) &= \text{tr}\{\tilde{\mathbf{E}}\mathbf{S}_\delta\} \\ &= \text{tr}\left\{\tilde{\mathbf{E}}\left(\tilde{\mathbf{E}}^\top \mathbf{W}^* \tilde{\mathbf{E}} + \lambda_g \tilde{\mathbf{E}}\right)^{-1} \tilde{\mathbf{E}}^\top \mathbf{W}^*\right\}. \end{aligned}$$

Thus, considering $\mathbf{Q}_{\tilde{\mathbf{E}}} = \tilde{\mathbf{E}}^\top \mathbf{W}^* \tilde{\mathbf{E}}$ and $\mathbf{Q}_{\lambda_g} = \lambda_g \mathbf{E}$, and since $\mathbf{W}^* > 0$ and $\text{rank}(\tilde{\mathbf{E}}^\top) \leq n$, then $\mathbf{Q}_{\tilde{\mathbf{E}}} \geq 0$. Therefore, there exists a matrix $\mathbf{Q}_{\tilde{\mathbf{E}}}^{1/2} \geq 0$ such that $\mathbf{Q}_{\tilde{\mathbf{E}}} = \mathbf{Q}_{\tilde{\mathbf{E}}}^{1/2} \mathbf{Q}_{\tilde{\mathbf{E}}}^{1/2}$. Thus, we can write $\text{tr}\{\tilde{\mathbf{E}}\mathbf{S}_\delta\} = \text{tr}\{\tilde{\mathbf{S}}_\delta\}$ as

$$\text{tr}\{\tilde{\mathbf{S}}_\delta\} = \sum_{j=1}^n \frac{1}{1 + \lambda_g \ell_j},$$

where ℓ_j , for $j = 1, \dots, n$, are the eigenvalues of the matrix $\mathbf{Q}_{\tilde{\mathbf{E}}}^{-1/2} \mathbf{Q}_{\lambda_g} \mathbf{Q}_{\tilde{\mathbf{E}}}^{-1/2}$. It is important to note that both $\text{DF}(\lambda_k)$ and $\text{DF}(\lambda_g)$ are inversely proportional to λ_k and λ_g , respectively. Alternatively, we can consider the backfitting estimators defined in Equation (13) and thus calculate the DFs associated with the smooth varying-coefficient functions as

$$\overline{\text{DF}}(\lambda_j) = \text{tr}\{\mathcal{H}_j\}, \quad j = 1, \dots, s,$$

with \mathcal{H}_j defined above. Similarly, the DFs associated with the smooth surface can be calculated from the representation $\hat{\mathbf{g}} = \mathbf{E}\hat{\boldsymbol{\delta}} + \mathbf{T}^\top \hat{\mathbf{a}} = \mathcal{H}_g \mathbf{y}$, whit $\mathcal{H}_g = \mathbf{E}\mathcal{H}_{s+1} + \mathbf{T}^\top \mathcal{H}'$, \mathcal{H}_{s+1} , \mathcal{H}' and \mathbf{y} defined in the previous sections. Thus, the DFs are given by

$$\overline{\text{DF}}(\lambda_g) = \text{tr}\{\mathcal{H}_g\}.$$

3.7 SELECTING AN APPROPRIATE MODEL

Under the elliptical TPSPVCM, we have a total of $2 + p + d + \text{DF}(\boldsymbol{\lambda})$ parameters to be estimated, with $\text{DF}(\boldsymbol{\lambda}) = \text{DF}(\lambda_g) + \sum_{k=1}^s \text{DF}(\lambda_k)$ denoting approximately the number of effective parameters involved in modeling of the smooth varying-coefficient functions and surface. In this case, the Akaike information criterion (AIC) (Akaike, 1973) or the Bayesian information criterion (BIC) (Schwarz et al., 1978) can be used for selecting an appropriate model. The idea is to minimize the function

$$\text{AIC}(\boldsymbol{\lambda}) = -2L_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda}) + 2\left[2 + p + d + \text{DF}(\boldsymbol{\lambda})\right],$$

where $L_p(\hat{\boldsymbol{\theta}}, \boldsymbol{\lambda})$ denotes the penalized log-likelihood function available at $\hat{\boldsymbol{\theta}}$ for a fixed $\boldsymbol{\lambda}$. It is important to mention that AIC is based on information theory and is useful for selecting an appropriate model given data with adequate sample size. An alternative version of the AIC, denoted by AICc, was proposed by Hurvich et al. (1998) in the context of parametric linear regression and autoregressive time series. Recently, Relvas (2016) adapted this criterion for the partially linear model with first-order autoregressive symmetric errors. Considering such proposals, we propose the AICc as an alternative for the selection of models under the elliptical TPSPVCM, which is given by

$$\text{AIC}_c(\boldsymbol{\lambda}) = \log \left\{ \frac{\|\sqrt{\widehat{\mathbf{W}}_v}(\mathbf{y} - \hat{\mathbf{y}})\|^2}{n} \right\} + \frac{2\left[\text{tr}(\widehat{\mathbf{H}}(\boldsymbol{\lambda})) + 1\right]}{n - \text{tr}(\widehat{\mathbf{H}}(\boldsymbol{\lambda})) - 2} + 1,$$

where $\widehat{\mathbf{y}} = \widehat{\mathbf{H}}(\boldsymbol{\lambda})\mathbf{y}$ and $\widehat{\mathbf{H}}(\boldsymbol{\lambda})$ corresponds to the smoother matrix, which is equivalent to the hat matrix defined in the class of parametric regression models. If we consider the matrix representation given in Equation (3) of our model and the backfitting estimators given in Equation (13), it is possible to obtain a closed expression for the matrix $\widehat{\mathbf{H}}(\boldsymbol{\lambda})$. Indeed, assuming that $\boldsymbol{\lambda}$, \mathbf{W}^* and \mathbf{W}_v are fixed, we have that

$$\mathbf{H}(\boldsymbol{\lambda}) = \sum_{j=0}^{s+1} \widetilde{\mathcal{H}}_j, \quad (15)$$

with $\widetilde{\mathcal{H}}_j = \mathbf{N}_j \mathcal{H}_j$. Note that the principal diagonal elements of $\mathbf{H}(\boldsymbol{\lambda})$ obtained in the last iteration of the iterative process, denoted here by $h_{ii}(\boldsymbol{\lambda})$, are called leverage points and play an important role in the construction of diagnostic techniques.

3.8 SMOOTHING PARAMETERS

The determination of the parameters λ_k and λ_g is a crucial part in the estimation process and different choice methods are available in the literature. For example, it is usual to consider the cross-validation method or the generalized cross-validation method (Craven and Wahba, 1978). Following Relvas (2016), an alternative to select smoothing parameters under the elliptical TPSPVCM is to consider a generalized cross-validation method defined by

$$\text{GCV}(\boldsymbol{\lambda}) = \frac{\| \sqrt{\widehat{\mathbf{W}}_v}(\mathbf{y} - \widehat{\mathbf{y}}) \|^2}{[1 - n^{-1} \text{tr}(\widehat{\mathbf{H}}(\boldsymbol{\lambda}))]}.$$

In this case, $\boldsymbol{\lambda}$ should be obtained by minimizing $\text{GCV}(\boldsymbol{\lambda})$ for a grid of $\boldsymbol{\lambda}$ values. Alternatively, these parameters may be selected by applying the AIC. In particular, we can consider the $\text{AIC}(\boldsymbol{\lambda})$ or $\text{AIC}_c(\boldsymbol{\lambda})$ criteria defined in the previous section, and use the effective DFs involved in nonparametric modeling to select appropriate smoothing parameters (Ibacache-Pulgar et al., 2013).

3.9 RESIDUAL ANALYSIS

We propose a standardized residual which can be used to detect error distribution misspecification as well as the presence of outlying observations. It follows from Equation (15) that the residuals vector is the difference between the observed data vector and estimated mean vector, that is,

$$\widehat{\mathbf{r}} = \mathbf{y} - \widehat{\mathbf{y}} = [\mathbf{I} - \mathbf{H}(\boldsymbol{\lambda})] \mathbf{y}. \quad (16)$$

Since that $\mathbf{H}(\boldsymbol{\lambda})$ is not a projection operator, this is, $\mathbf{H}^2(\boldsymbol{\lambda}) \neq \mathbf{H}(\boldsymbol{\lambda})$, we have that the approximate variance of the residual vector is given by

$$\text{Var}_{\text{approx}}(\widehat{\mathbf{r}}) = [\mathbf{I} - \mathbf{H}(\boldsymbol{\lambda})] \text{Cov}(\mathbf{y}) [\mathbf{I} - \mathbf{H}(\boldsymbol{\lambda})]^\top,$$

where $\text{Cov}(\mathbf{y}) = \text{blockdiag}_{1 \leq i \leq n}(\xi_i \boldsymbol{\Sigma}_i)$. Then, we have that the l th standardized residual takes the form

$$\hat{r}_l = \frac{\mathbf{d}_l^\top [\mathbf{I} - \mathbf{H}(\boldsymbol{\lambda})] \mathbf{y}}{\sqrt{\mathbf{d}_l^\top \widehat{\text{Var}}_{\text{approx}}(\hat{\mathbf{r}}) \mathbf{d}_l}},$$

where $\widehat{\text{Var}}_{\text{approx}}(\hat{\mathbf{r}}) = \text{Var}_{\text{approx}}(\hat{\mathbf{r}}) |_{\hat{\boldsymbol{\theta}}}$, with \mathbf{d}_l denoting an $(M \times 1)$ vector with 1 at the l th position and 0 elsewhere, for $l = 1, \dots, M$. Further details on the analysis of residuals in the semiparametric context can be found, for example, in [Ibacache-Pulgar et al. \(2013\)](#).

4. APPLICATION

In this section, we illustrate the applicability of the TPSPVCM through an application based on a set of real data. For comparative purposes, we consider random errors whose distribution belongs to the symmetric class; specifically, the normal and Student-t distributions.

4.1 DATA DESCRIPTION

In our application, we consider the house prices of Boston area reported by [Harrison and Rubinfeld \(1978\)](#) and analyzed by many authors; see, for example, [Belsley et al. \(1980\)](#), [Ibacache-Pulgar et al. \(2013\)](#) and, more recently, [Ibacache-Pulgar and Reyes \(2018\)](#). This data set contains a sample of 506 observations collected by the U.S Census Service concerning housing in the area of Boston. The variable LMV (logarithm of the median house price in USD 1000) is related with 14 explanatory variables, 6 of them are defined from census track and the remaining variables are defined for clusters. For simplicity, we consider four explanatory variables: LSTAT (logarithm of the proportion of the population that is lower status, ROOM (average number of rooms per dwelling), CRIM (per capita crime rate by town), TAX (full-value property-tax rate per USD 10000), and the geographical coordinates expressed in longitude and latitude. Similar to what observed by [Ibacache-Pulgar and Reyes \(2018\)](#), we see in [Figure 1\(a\)](#) that the relationship between LMV and the explanatory variable TAX is linear, whereas the relationship between LMV and LSTAT appear in nonlinear ways ([Figure 1\(b\)](#)). Also, [Figures 1\(c\) and 1\(d\)](#) suggests that the explanatory variables ROOM and CRIM might be interacting with the variable LSTAT in nonlinear fashion. [Figure 2](#) represents the spatial distribution of the LMV variable. From [Figure 2 \(right\)](#) we note that the lowest prices are concentrated between the latitudes 42.2 and 42.25 and longitudes between -71.0 and -71.1, while the highest prices are in the north part of the town. It is important to point out that [Ibacache-Pulgar and Reyes \(2018\)](#) analyzes this same set of variables using a partially varying-coefficient model but without considering the effect of the geographic coordinates associated with each of the households surveyed. We believe that including the effect of geographical coordinates can improve the fit of the model considered by [Ibacache-Pulgar and Reyes \(2018\)](#) in predictive terms, precision of the estimates and goodness of fit.

4.2 FITTING THE MODELS

Considering the analysis described above, we suggest the application of a partially varying-coefficient model that including the spatial variability. Specifically, we assume the following

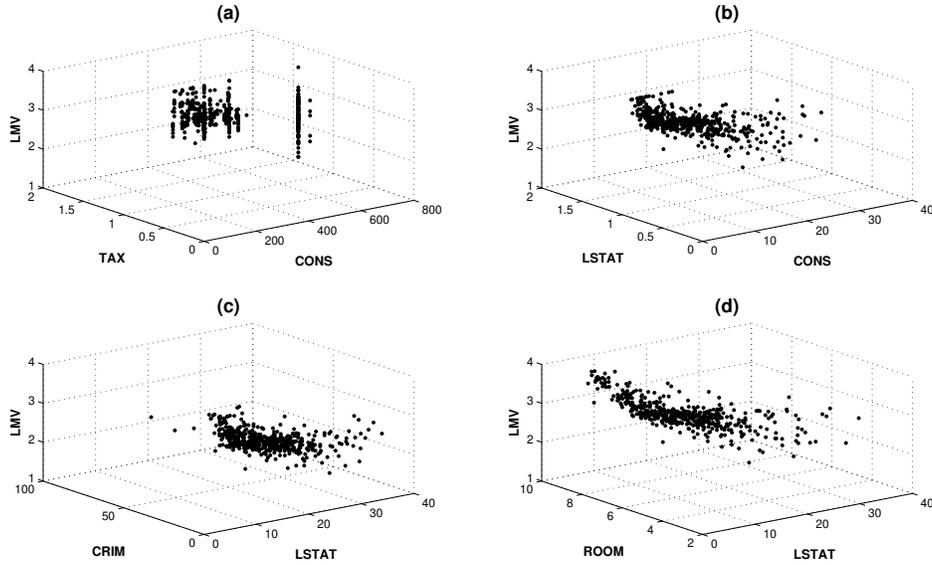


Figure 1. Three-dimensional graphics for house prices data. CONS denote an auxiliary variable defined as an $(n \times 1)$ ones vector.

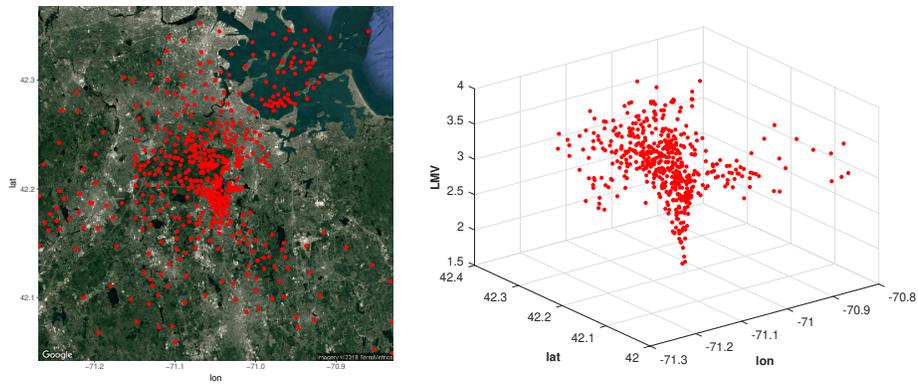


Figure 2. (Left) Google map of the Boston province. Red circles indicate the spatial distribution of the house prices data. (Right) Distributions of the LMV respect the longitude and the latitude.

thin-plate spline partially varying-coefficient model:

$$y_i = \alpha_0 + \alpha_1 z_i + \beta_1(r_i) x_i^{(1)} + \beta_2(r_i) x_i^{(2)} + g(\mathbf{t}_i) + \varepsilon_i, \quad i = 1, \dots, 506, \quad (17)$$

where y_i denotes the value of LMV in USD 1000, z_i the value of TAX, $x_i^{(1)}$ the value of CRIM, $x_i^{(2)}$ the value of ROOM, $\alpha = (\alpha_0, \alpha_1)^\top$ the parameters vector associated with parametric component, r_i the value of LSTAT from the i th experimental unit, β_k , for $k = 1, 2$, are unknown smooth functions, g is a smooth surface that depends of the vector of coordinates $\mathbf{t}_i = (t_{1i}, t_{2i}) \in \mathcal{R}^2$, and ε_i are independent random errors that follow a symmetric distribution whit location parameter 0, scale parameter ϕ and density generator function h . We compare in the sequel the fits based on normal and Student-t random errors. The DFs (ν) for the Student-t model was selected by the AIC, that is, by defining a grid of values for ν and choosing the one that minimize the AIC. Figure 3 shows the graph of AIC values for different DFs. We can see that this criterion is minimized for a value of $\nu = 4$.

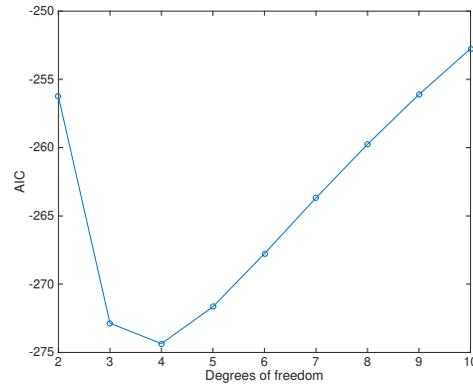


Figure 3. AIC values for different DFs, with $\nu = 2, \dots, 10$.

The MDPLE estimates, estimated standard errors and the corresponding AIC for the model of Equation (17) under normal and Student-t distributions are presented in Table 1. Comparing these results, we may notice a similarity between the estimates $\hat{\alpha}$ under both models, but the standard error for $\hat{\alpha}_1$ appears to be smaller under the Student-t model. Also, it can be seen that the scale parameters are different for the two fitted models, but the estimates are not comparable since they are on different scales. Additionally, we may notice that the AIC value under the Student-t model is smaller than the one under the normal model, indicating that the models with longer-than-normal tails seem to better fit the data, a fact that is also confirmed through the theoretical quantile versus empirical quantile (QQ) plots presented in Figure 4.

Table 1. Maximum penalized likelihood estimates, estimated standard errors (SE) and AIC values under normal and Student-t ($\nu = 4$) models fitted to house prices data.

	Normal		Student-t	
	Estimate	SE	Estimate	SE
α_1	3.0668	0.1145	3.0637	0.0964
α_2	-0.0003	0.0001	-0.0002	0.0001
ϕ	0.0344	0.0022	0.0172	0.0372
AIC	-218.76		-274.65	

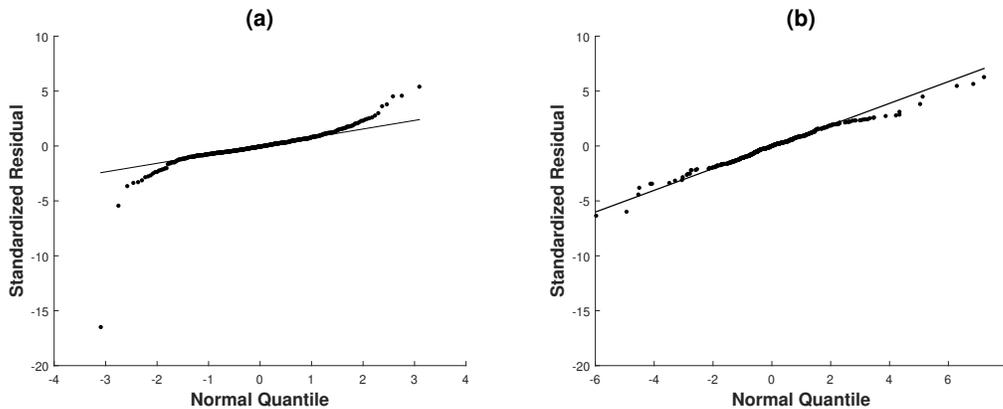


Figure 4. QQ plots fitted to house prices data: normal (a) and Student-t models with $\nu = 4$ DFs (b).

The standardized residual plot provide in Figure 5 is used to verify if there are outlying observations. In this case, the presence of some outlying observations for both models is clearly observed. Figure 6 displays the graphics of the LMV versus the fitted LMV from the two models. Although these plots indicate suitable fits for both models.

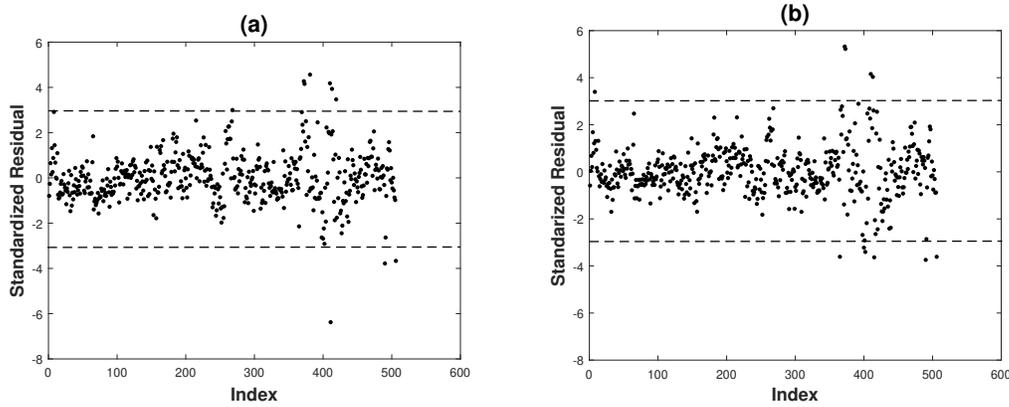


Figure 5. Index plots of standardized residuals to house prices data: normal (a) and Student-t models with $\nu = 4$ DFs (b).

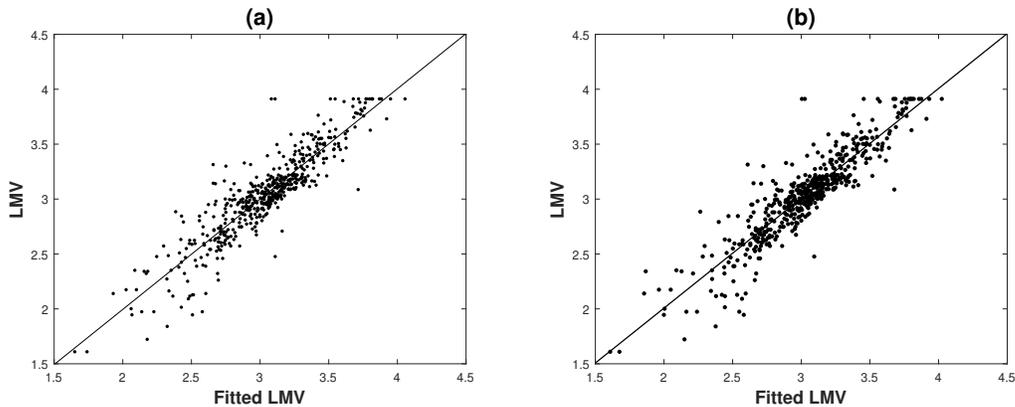


Figure 6. Scatter plots LMV versus fitted LMV to house prices data: normal (a) and Student-t models with $\nu = 4$ DFs (b).

The estimated coefficients functions β_1 and β_2 are computed using the smoothing parameters obtained by the method described in Subsection 3.8. Figures 7 and 8 show the estimated coefficient functions under both models and their corresponding approximate standard error band (dashed curves). The figures suggest that the coefficient curves vary with the explanatory variable LSTAT. In addition, it can be seen that the functions estimated under the normal model have a higher smoothness compared to those obtained from the Student-t model. It is important to remember that in this work we have incorporated the spatial variability of the data in the modeling process. Comparing with the results obtained by [Ibacache-Pulgar and Reyes \(2018\)](#), we can notice that the TPSPVCM model significantly improves the quality of the adjustment compared with the PVCVM model. For example, for normal TPSPVCM model, the AIC value is -218.7586 , while that for normal PVCVM model the AIC value is -139.4998 . Analogously, under Student-t TPSPVCM model with four DFs, the AIC value is -274.6546 , while that under Student-t PVCVM model with five DFs, the AIC value is -188.3909 . In addition, we can notice that for the normal model, the estimated functions differ significantly, while under the Student-t model, they retain the same tendency but with a greater degree of smoothness.

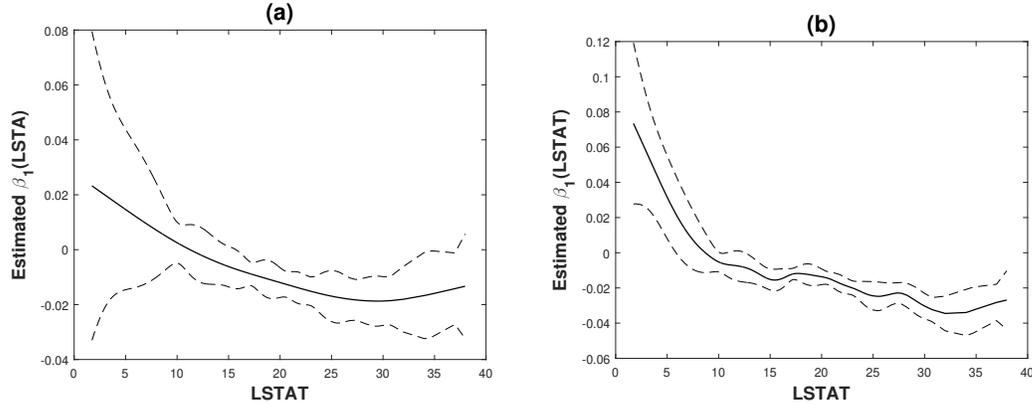


Figure 7. Plots of estimated coefficient function β_1 for the house prices data and its approximate pointwise standard error bands denoted by the dashed lines: Normal (a) and Student-t with $\nu = 4$ DFs (b) models.

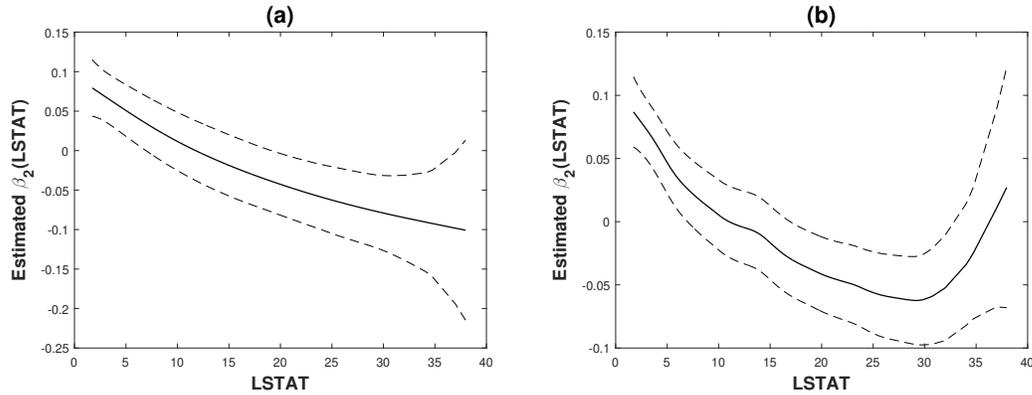


Figure 8. Plots of estimated coefficient function β_2 for the house prices data and its approximate pointwise standard error bands denoted by the dashed lines: Normal (a) and Student-t with $\nu = 4$ DFs (b).

4.3 ROBUSTNESS ASPECTS OF THE MDPLEs

It is important to note that for univariate Student-t distribution the current weight $v_i^{(u)} = (\nu + 1)/(\nu + u_i^{(u)})$, with $u_i^{(u)} = (y_i - \mu_i^{(u)})^2/\phi^{(u)}$, is inversely proportional to the distance between the observed value y_i and its current predicted value $\mu_i^{(u)}$, so that outlying observations tend to have small weights in the estimation process. Therefore, we may expect that the MDPLEs from the Student-t TPSPVCMs are less sensitive to outlying observations than the MDPLEs from normal models. Figure 9 shows the plot between the standardized residual defined in Equation (16) and estimated weights under Student-t model. We can be seen that observation #411 has a very small residual and a high estimated weight, but its removal from the data set did not generate significant changes in the estimation of the parameters. For this reason the summary of the fit without this observation is omitted. Finally, it is important to note that the iterative process under Student-t model generates a reduction in the weights associated with the observations detected as discrepant. Hence such estimators present some characteristics of robustness similar to the associated with the weight function described by Huber (1981).

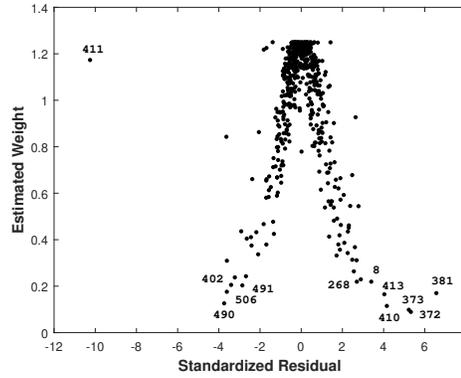


Figure 9. Plot of the standardized residual and estimated weights for the house prices data under Student-t model.

5. CONCLUDING REMARKS

The elliptical thin-plate spline partially varying-coefficient models proposed in this paper have special characteristics compared to other types of models existing in the literature. Specifically, these models allow describing the mean of the data in those cases in which there are explanatory variables that are related to the response variable through a regression structure that depends on a parametric component (usual linear predictor), a non-parametric component (explanatory variables effects in which the coefficients are allowed to vary as smooth functions of other variables) and a spatial component (thin-plate spline). In addition, the distributional assumption established on random errors allows us to model datasets in which the assumption of normality is not appropriate. We derive a reweighted iterative process for obtaining the maximum doubly penalized likelihood estimators based on the Score Fisher and back-fitting methods. Closed-form expressions are obtained for the penalized observed and expected information matrices, and expressions for the standard errors of the maximum doubly penalized likelihood estimators are also available. We propose a way to estimate the smoothing parameters based on generalized cross-validation and a method for the selection of models by using the AIC. A real dataset previously analyzed under normal errors is reanalyzed under Student-t errors by including a smoothing surface for the spatial variability. By comparing the AIC values of the two models, the Student-t showed the better fitting. Thus, we can recommend Student-t thin-plate spline partially varying-coefficient models as an option to fit datasets with indications of heavy tails. The computational implementation of all our results was carried out in MATLAB software, and the codes can be requested from the authors to the email german.ibacache@uv.cl.

APPENDIX

Here, we show the score function, the observed information matrix and the expected information matrix for the elliptical TPSPCVM base on the doubly penalized log-likelihood function given in Equation (6).

PENALIZED SCORE FUNCTION

Let $\mathbf{W}_v = \text{blockdiag}_{1 \leq i \leq n}(v_i \mathbf{W}_i)$, with $\mathbf{W}_i = \boldsymbol{\Sigma}_i^{-1}$, $v_i = -2\zeta_h(u_i)$, $\zeta_h(u_i) = d \log h(u_i)/du_i$, $\boldsymbol{\Sigma}_i^* = \boldsymbol{\Sigma}_i^{-1} \partial \boldsymbol{\Sigma}_i / \partial \ell$, $\boldsymbol{\Upsilon} = \text{blockdiag}_{1 \leq i \leq n}(\boldsymbol{\Upsilon}_i)$, with $\boldsymbol{\Upsilon}_i =$

$v_i \Sigma_i^{-1} (\partial \Sigma_i / \partial \ell) \Sigma_i^{-1}$. Assuming that Equation (6) is regular with respect to all elements of $\boldsymbol{\theta}$, we have that the penalized score function of $\boldsymbol{\theta}$ under elliptical TPSPVCM is given by

$$U_p(\boldsymbol{\theta}) = \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} U_p(\boldsymbol{\theta})^{\tilde{\alpha}} \\ U_p(\boldsymbol{\theta})^{\beta_1} \\ \vdots \\ U_p(\boldsymbol{\theta})^{\beta_s} \\ U_p(\boldsymbol{\theta})^{\delta} \\ U_p(\boldsymbol{\theta})^{\tau} \end{pmatrix},$$

where $U_p^{\tilde{\alpha}}(\boldsymbol{\theta}) = (\mathbf{Z}^\top \mathbf{W}_v \boldsymbol{\varepsilon} - \tilde{\mathbf{T}} \mathbf{W}_v \boldsymbol{\varepsilon})^\top$, $U_p^{\beta_k}(\boldsymbol{\theta}) = \tilde{\mathbf{N}}_k^\top \mathbf{W}_v \boldsymbol{\varepsilon} - \lambda_1 \mathbf{K}_k \boldsymbol{\beta}_k$, for $k = 1, \dots, s$, $U_p^{\delta}(\boldsymbol{\theta}) = \tilde{\mathbf{E}} \mathbf{W}_v \boldsymbol{\varepsilon} - \lambda_g \mathbf{E} \boldsymbol{\delta}$ and $U_p^{\tau}(\boldsymbol{\theta}) = -(1/2) \sum_{i=1}^n \text{tr}(\Sigma_i^*) + (1/2) \boldsymbol{\varepsilon}^\top \boldsymbol{\Upsilon} \boldsymbol{\varepsilon}$.

PENALIZED OBSERVED INFORMATION MATRIX

For simplicity, let $\Psi_i = 2\Psi_{1i} + \Psi_{2i}$, $\Psi_i^* = \Psi_{1i} + \Psi_{2i}$ and $\Psi_i^{**} = \Psi_{1i} + 2\Psi_{2i}$, with $\Psi_{1i} = v_i' \Sigma_i^{-1} \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^\top \Sigma_i^{-1}$ and $\Psi_{2i} = v_i \Sigma_i^{-1}$. In addition, let $\Psi = \text{diag}_{1 \leq i \leq n} (\Psi_i)$ and $\Omega = \text{diag}_{1 \leq i \leq n} (\Omega_i)$, with $\Omega_i = \Psi_i^* (\partial \Sigma_i / \partial \eta) \Sigma_i^{-1}$. The L_p ($p^* \times p^*$) Hessian matrix under elliptical TPSPVCM is defined as

$$L_p(\boldsymbol{\theta}) = \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} L_p^{\tilde{\alpha}\tilde{\alpha}}(\boldsymbol{\theta}) & L_p^{\tilde{\alpha}\beta}(\boldsymbol{\theta}) & L_p^{\tilde{\alpha}\delta}(\boldsymbol{\theta}) & L_p^{\tilde{\alpha}\tau}(\boldsymbol{\theta}) \\ L_p^{\beta\tilde{\alpha}}(\boldsymbol{\theta}) & L_p^{\beta\beta}(\boldsymbol{\theta}) & L_p^{\beta\delta}(\boldsymbol{\theta}) & L_p^{\beta\tau}(\boldsymbol{\theta}) \\ L_p^{\delta\tilde{\alpha}}(\boldsymbol{\theta}) & L_p^{\delta\beta}(\boldsymbol{\theta}) & L_p^{\delta\delta}(\boldsymbol{\theta}) & L_p^{\delta\tau}(\boldsymbol{\theta}) \\ L_p^{\tau\tilde{\alpha}}(\boldsymbol{\theta}) & L_p^{\tau\beta}(\boldsymbol{\theta}) & L_p^{\tau\delta}(\boldsymbol{\theta}) & L_p^{\tau\tau}(\boldsymbol{\theta}) \end{pmatrix},$$

whose elements are given by

$$L_p^{\tilde{\alpha}\tilde{\alpha}}(\boldsymbol{\theta}) = \begin{pmatrix} -\mathbf{Z}^\top \Psi \mathbf{Z} & -\mathbf{Z}^\top \Psi \tilde{\mathbf{E}} \\ -\tilde{\mathbf{E}}^\top \Psi \mathbf{Z} & -\tilde{\mathbf{T}}^\top \Psi \tilde{\mathbf{T}} \end{pmatrix}, \quad L_p^{\tilde{\alpha}\beta}(\boldsymbol{\theta}) = \begin{pmatrix} -\mathbf{Z}^\top \Psi \tilde{\mathbf{N}}_1 & \dots & -\mathbf{Z}^\top \Psi \tilde{\mathbf{N}}_s \\ -\tilde{\mathbf{T}}^\top \Psi \tilde{\mathbf{N}}_1 & \dots & -\tilde{\mathbf{T}}^\top \Psi \tilde{\mathbf{N}}_s \end{pmatrix},$$

$$L_p^{\tilde{\alpha}\delta}(\boldsymbol{\theta}) = \begin{pmatrix} -\mathbf{Z}^\top \Psi \tilde{\mathbf{E}} \\ -\mathbf{Z}^\top \Psi \tilde{\mathbf{T}} \end{pmatrix}, \quad L_p^{\tilde{\alpha}\tau}(\boldsymbol{\theta}) = \begin{pmatrix} -\mathbf{Z}^\top \Omega \boldsymbol{\varepsilon} \\ -\tilde{\mathbf{T}}^\top \Omega \boldsymbol{\varepsilon} \end{pmatrix},$$

$$L_p^{\beta\beta}(\boldsymbol{\theta}) = \begin{pmatrix} -\tilde{\mathbf{N}}_1^\top \Psi \tilde{\mathbf{N}}_1 - \lambda_1 \mathbf{K}_1 & \dots & -\tilde{\mathbf{N}}_1^\top \Psi \tilde{\mathbf{N}}_s \\ -\tilde{\mathbf{N}}_2^\top \Psi \tilde{\mathbf{N}}_1 & \dots & -\tilde{\mathbf{N}}_2^\top \Psi \tilde{\mathbf{N}}_s \\ \vdots & \ddots & \vdots \\ -\tilde{\mathbf{N}}_s^\top \Psi \tilde{\mathbf{N}}_1 & \dots & -\tilde{\mathbf{N}}_s^\top \Psi \tilde{\mathbf{N}}_s - \lambda_s \mathbf{K}_s \end{pmatrix}, \quad L_p^{\beta\delta}(\boldsymbol{\theta}) = \begin{pmatrix} -\tilde{\mathbf{N}}_1^\top \Psi \tilde{\mathbf{E}} \\ -\tilde{\mathbf{N}}_2^\top \Psi \tilde{\mathbf{E}} \\ \vdots \\ -\tilde{\mathbf{N}}_s^\top \Psi \tilde{\mathbf{E}} \end{pmatrix},$$

$$L_p^{\beta\tau}(\boldsymbol{\theta}) = \begin{pmatrix} -\tilde{\mathbf{N}}_1^\top \Omega \boldsymbol{\varepsilon} \\ -\tilde{\mathbf{N}}_2^\top \Omega \boldsymbol{\varepsilon} \\ \vdots \\ -\tilde{\mathbf{N}}_s^\top \Omega \boldsymbol{\varepsilon} \end{pmatrix}, \quad L_p^{\delta\delta}(\boldsymbol{\theta}) = -\tilde{\mathbf{E}}^\top \Psi \tilde{\mathbf{E}} - \lambda_g \mathbf{E}, \quad L_p^{\delta\tau}(\boldsymbol{\theta}) = -\tilde{\mathbf{E}}^\top \Omega \boldsymbol{\varepsilon}$$

and

$$\begin{aligned} \mathbf{L}_p^{\tau\tau}(\boldsymbol{\theta}) &= \sum_{i=1}^n \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \left[\frac{\partial \boldsymbol{\Sigma}_i}{\partial \ell} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial j} - \frac{\partial^2 \boldsymbol{\Sigma}_i}{\partial j \partial \ell} \right] \right) - \frac{1}{2} \boldsymbol{\varepsilon}_i^\top \boldsymbol{\Sigma}_i^{-1} \times \\ &\quad \left[\frac{\partial \boldsymbol{\Sigma}_i}{\partial \ell} \boldsymbol{\Psi}_i^{**} \frac{\partial \boldsymbol{\Sigma}_i}{\partial j} - v_i \frac{\partial^2 \boldsymbol{\Sigma}_i}{\partial j \partial \ell} \right] \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\varepsilon}_i. \end{aligned}$$

PENALIZED EXPECTED INFORMATION MATRIX

Let $d_{g_i} = E(\zeta_g^2(\gamma_i)\gamma_i)$ and $f_{g_i} = E(\zeta_g^2(\gamma_i)\gamma_i^2)$, with $\gamma_i = \mathbf{e}_i^\top \mathbf{e}_i$, $\mathbf{e}_i \sim \text{El}_{m_i}(\mathbf{0}, \mathbf{I}_{m_i})$, and $\mathbf{W}^* = \text{blockdiag}_{1 \leq i \leq n}((4d_{g_i}/m_i) \mathbf{W}_i)$. By calculating the expectation of the matrix $-\mathbf{L}_p$, we obtain the $(p^* \times p^*)$ penalized expected information matrix given by

$$\mathcal{I}_p(\boldsymbol{\theta}) = -E \left(\frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right).$$

Following Lange et al. (1989), we have that the (j^*, ℓ^*) -element of the matrix \mathcal{I}_p for i th cluster, with respect to the parameters $\theta_{j^*}^*$ and $\theta_{\ell^*}^*$, can be obtained as

$$\mathcal{I}_{p_i}(\boldsymbol{\theta}) = E \left(\frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \theta_{j^*}^*} \frac{\partial L_{p_i}(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \theta_{\ell^*}^*} \right).$$

After some algebraic manipulations we find that the $\mathcal{I}_p(\boldsymbol{\theta})$ matrix have a block-diagonal structure of the form

$$\mathcal{I}_p(\boldsymbol{\theta}) = \text{blockdiag}(\mathcal{I}_p^{11}(\boldsymbol{\theta}), \mathcal{I}_p^{22}(\boldsymbol{\theta})),$$

where

$$\mathcal{I}_p^{11}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}_p^{\tilde{\alpha}\tilde{\alpha}}(\boldsymbol{\theta}) & \mathcal{I}_p^{\tilde{\alpha}\beta}(\boldsymbol{\theta}) & \mathcal{I}_p^{\tilde{\alpha}\delta}(\boldsymbol{\theta}) \\ \mathcal{I}_p^{\beta\tilde{\alpha}}(\boldsymbol{\theta}) & \mathcal{I}_p^{\beta\beta}(\boldsymbol{\theta}) & \mathcal{I}_p^{\beta\delta}(\boldsymbol{\theta}) \\ \mathcal{I}_p^{\delta\tilde{\alpha}}(\boldsymbol{\theta}) & \mathcal{I}_p^{\delta\beta}(\boldsymbol{\theta}) & \mathcal{I}_p^{\delta\delta}(\boldsymbol{\theta}) \end{pmatrix},$$

whose elements of the matrix are given by

$$\mathcal{I}_p^{\tilde{\alpha}\tilde{\alpha}}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{Z}^\top \mathbf{W}^* \mathbf{Z} & \mathbf{Z}^\top \mathbf{W}^* \tilde{\mathbf{E}} \\ \tilde{\mathbf{E}}^\top \mathbf{W}^* \mathbf{Z} & \tilde{\mathbf{T}}^\top \mathbf{W}^* \tilde{\mathbf{T}} \end{pmatrix}, \quad \mathcal{I}_p^{\tilde{\alpha}\beta}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{Z}^\top \mathbf{W}^* \tilde{\mathbf{N}}_1 \dots \mathbf{Z}^\top \mathbf{W}^* \tilde{\mathbf{N}}_s \\ \tilde{\mathbf{T}} \mathbf{W}^* \tilde{\mathbf{N}}_1 \dots \tilde{\mathbf{T}} \mathbf{W}^* \tilde{\mathbf{N}}_s \end{pmatrix},$$

$$\mathcal{I}_p^{\tilde{\alpha}\delta}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{Z}^\top \mathbf{W}^* \tilde{\mathbf{E}} \\ \mathbf{Z}^\top \mathbf{W}^* \tilde{\mathbf{T}} \end{pmatrix}, \quad \mathcal{I}_p^{\delta\delta}(\boldsymbol{\theta}) = \tilde{\mathbf{E}}^\top \mathbf{W}^* \tilde{\mathbf{E}} + \lambda_g \tilde{\mathbf{E}},$$

$$\mathcal{I}_p^{\beta\beta}(\boldsymbol{\theta}) = \begin{pmatrix} \widetilde{\mathbf{N}}_1^\top \mathbf{W}^* \widetilde{\mathbf{N}}_1 + \lambda_1 \mathbf{K}_1 & \dots & \widetilde{\mathbf{N}}_1^\top \mathbf{W}^* \widetilde{\mathbf{N}}_s \\ \widetilde{\mathbf{N}}_2^\top \mathbf{W}^* \widetilde{\mathbf{N}}_1 & \dots & \widetilde{\mathbf{N}}_2^\top \mathbf{W}^* \widetilde{\mathbf{N}}_s \\ \vdots & \ddots & \vdots \\ \widetilde{\mathbf{N}}_s^\top \mathbf{W}^* \widetilde{\mathbf{N}}_1 & \dots & \widetilde{\mathbf{N}}_s^\top \mathbf{W}^* \widetilde{\mathbf{N}}_s + \lambda_s \mathbf{K}_s \end{pmatrix}, \quad \mathcal{I}_p^{\beta\delta}(\boldsymbol{\theta}) = \begin{pmatrix} \widetilde{\mathbf{N}}_1^\top \mathbf{W}^* \widetilde{\mathbf{E}} \\ \widetilde{\mathbf{N}}_2^\top \mathbf{W}^* \widetilde{\mathbf{E}} \\ \vdots \\ \widetilde{\mathbf{N}}_s^\top \mathbf{W}^* \widetilde{\mathbf{E}} \end{pmatrix}$$

and

$$\mathcal{I}_p^{22} = \sum_{i=1}^n \mathcal{I}_{p_i}^{\tau\tau},$$

where the (j, ℓ) th element of $\mathcal{I}_{p_i}^{\tau\tau}$ is given by

$$\mathcal{I}_{p_{i_{j\ell}}} = \left[\frac{b_{i_{j\ell}}}{4} \left(\frac{4f_{g_i}}{m_i(m_i + 2)} - 1 \right) + \frac{2f_{g_i}}{m_i(m_i + 2)} \text{tr} \left(\boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial j} \boldsymbol{\Sigma}_i^{-1} \frac{\partial \boldsymbol{\Sigma}_i}{\partial \ell} \right) \right],$$

where $b_{i_{j\ell}} = \text{tr}(\boldsymbol{\Sigma}_i^{-1} \partial \boldsymbol{\Sigma}_i / \partial j) \text{tr}(\boldsymbol{\Sigma}_i^{-1} \partial \boldsymbol{\Sigma}_i / \partial \ell)$.

JOINT ITERATIVE PROCESS

Since parameters $(\tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_s, \boldsymbol{\delta})$ and $\boldsymbol{\tau}$ are orthogonal, the estimation process is simplified, so we can consider the simultaneous estimation of $(\tilde{\boldsymbol{\alpha}}, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_s, \boldsymbol{\delta})$ and $\boldsymbol{\tau}$ through process of two independent stages. Specifically, the solution of the estimating equation system given in Equation (7) to obtain the MDPLE of $\boldsymbol{\theta}$ 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 $\boldsymbol{\tau}$, which is equivalent to the following iterative process:

- (i) Initialize:
 - (a) Fitting a TPSPVCM under normal errors to get $\boldsymbol{\beta}_j^{(0)}$ ($j = 0, 1, \dots, s$) and $\boldsymbol{\delta}_0$.
 - (b) Get starting value for $\boldsymbol{\tau}$ by using the fitted values from (a).
 - (c) From the current value $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}_0^{(0)\top}, \boldsymbol{\beta}_1^{(0)\top}, \dots, \boldsymbol{\beta}_s^{(0)\top}, \boldsymbol{\delta}^0, \boldsymbol{\tau}^{(0)})^\top$ obtaining $\boldsymbol{\Sigma}_i^{(0)} = \boldsymbol{\Sigma}_i|_{\boldsymbol{\theta}^{(0)}}$, $\mathbf{W}^{*(0)}$, $v_i^{(0)} = v_i|_{\boldsymbol{\theta}^{(0)}}$ and $\mathbf{W}_v^{(0)} = \text{blockdiag}_{1 \leq i \leq n} (v_i^{(0)} \mathbf{W}_i^{(0)})$, with $\mathbf{W}_i^{(0)} = \boldsymbol{\Sigma}_i^{(0)-1}$. Then, we obtain

$$\begin{aligned} \boldsymbol{\eta}^{(0)} &= \boldsymbol{\mu}^{(0)} + \mathbf{W}^{*(0)-1} \mathbf{W}_v^{(0)} (\mathbf{y} - \boldsymbol{\mu}^{(0)}), \\ \mathbf{S}_0^{(0)} &= (\widetilde{\mathbf{N}}_0^\top \mathbf{W}^{*(0)} \widetilde{\mathbf{N}}_0)^{-1} \widetilde{\mathbf{N}}_0^\top \mathbf{W}^{*(0)}, \\ \mathbf{S}_k^{(0)} &= (\widetilde{\mathbf{N}}_k^\top \mathbf{W}^{*(0)} \widetilde{\mathbf{N}}_k + \lambda_k \mathbf{K}_k)^{-1} \widetilde{\mathbf{N}}_k^\top \mathbf{W}^{*(0)}, \quad k = 1, \dots, s, \\ \mathbf{S}_\delta^{(0)} &= (\widetilde{\mathbf{E}}^\top \mathbf{W}^{*(0)} \widetilde{\mathbf{E}} + \lambda_g \widetilde{\mathbf{E}})^{-1} \widetilde{\mathbf{E}}^\top \mathbf{W}^{*(0)}. \end{aligned}$$

(ii) Step 1: Iterate repeatedly by cycling between the equations stated as

$$\begin{aligned} \beta_0^{(u+1)} &= \mathbf{S}_0^{(u)} \left(\boldsymbol{\eta}^{(u)} - \sum_{l=1}^s \widetilde{\mathbf{N}}_l \beta_l^{(u)} - \widetilde{\mathbf{E}} \boldsymbol{\delta}^{(u)} \right), \\ \beta_1^{(u+1)} &= \mathbf{S}_1^{(u)} \left(\boldsymbol{\eta}^{(u)} - \widetilde{\mathbf{N}}_0 \beta_0^{(u+1)} - \sum_{k=2}^s \widetilde{\mathbf{N}}_k \beta_k^{(u)} - \widetilde{\mathbf{E}} \boldsymbol{\delta}^{(u)} \right), \\ &\vdots \\ \beta_s^{(u+1)} &= \mathbf{S}_s^{(u)} \left(\boldsymbol{\eta}^{(u)} - \sum_{k=0}^{s-1} \widetilde{\mathbf{N}}_k \beta_k^{(u+1)} - \widetilde{\mathbf{E}} \boldsymbol{\delta}^{(u)} \right), \\ \boldsymbol{\delta}_s^{(u+1)} &= \mathbf{S}_s^{(u)} \left(\boldsymbol{\eta}^{(u)} - \sum_{k=0}^s \widetilde{\mathbf{N}}_k \beta_k^{(u+1)} \right), \end{aligned}$$

for $u = 0, 1, \dots$

Repeat (ii) replacing $\beta_j^{(u)}$ by $\beta_j^{(u+1)}$, for $j = 0, 1, \dots, s$, and $\boldsymbol{\delta}_j^{(u)}$ by $\boldsymbol{\delta}_j^{(u+1)}$ until convergence criterion $\Delta_u^\beta(\beta_j^{(u+1)}, \beta_j^{(u)}) = \sum_{j=0}^s \|\beta_j^{(u+1)} - \beta_j^{(u)}\| / \sum_{j=0}^s \|\beta_j^{(u)}\|$ and $\Delta_u^\delta(\boldsymbol{\delta}^{(u+1)}, \boldsymbol{\delta}^{(u)}) = \|\boldsymbol{\delta}^{(u+1)} - \boldsymbol{\delta}^{(u)}\| / \|\boldsymbol{\delta}^{(u)}\|$ is below some small threshold (Hastie and Tibshirani, 1990).

(iii) Step 2: For current values $\beta_j^{(u+1)}$, for $j = 0, 1, \dots, s$, and $\boldsymbol{\delta}^{(u+1)}$, obtaining $\boldsymbol{\tau}^{(u+1)}$ by using

$$\boldsymbol{\tau}^{(u+1)} = \boldsymbol{\tau}^{(u)} - \mathbf{E} \left\{ \frac{\partial^2 L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}^\top} \right\}^{-1} \frac{\partial L_p(\boldsymbol{\theta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\tau}} \Big|_{\boldsymbol{\theta}^{(u)}}.$$

(iv) Iterating between steps (ii) and (iii) by replacing $\beta_j^{(0)}$, for $j = 0, 1, \dots, s$, $\boldsymbol{\delta}^{(0)}$ and $\boldsymbol{\tau}^{(0)}$ by $\beta_j^{(u+1)}$, $\boldsymbol{\delta}^{(u+1)}$ and $\boldsymbol{\tau}^{(u+1)}$, respectively, until convergence.

AUTHOR CONTRIBUTIONS Conceptualization, M.M., G.I.P., O.N.; methodology, M.M., G.I.P., O.N.; software, M.M., G.I.P., O.N.; investigation, M.M., G.I.P., O.N.; writing-original draft preparation, M.M., G.I.P., O.N., writing review and editing, M.M., G.I.P., O.N., writing-original M.M., G.I.P., O.N. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGEMENTS The authors thanks Editors and Referees for their suggestions which allowed us to improve the presentation of this work.

FUNDING M. Moraga acknowledges financial support from Universidad Austral de Chile by grant DID S-2017-32. G. Ibacache-Pulgar acknowledges funding support by grant FONDECYT 11130704. O. Nicolis acknowledges funding support by grant FONDECYT 1201478.

CONFLICTS OF INTEREST The authors declare no conflict of interest.

REFERENCES

- Akaike, H., 1973. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723.
- Belsley, D. A., Kuh, E., and Welsch, R. E., 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley, New York, US.
- Breiman, L. and Friedman, J. H., 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580–598.
- Buja, A., Hastie, T., and Tibshirani, R., 1989. Linear smoothers and additive models. *The Annals of Statistics*, 17, 453–510.
- Cai, Z., Fan, J., and Li, R., 2000. Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95, 888–902.
- Chiang, C.T., Rice, J.A., and Wu, C.O., 2001. Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *Journal of the American Statistical Association*, 96, 605–619.
- Craven, P. and Wahba, G., 1978. Smoothing noisy data with spline functions. *Numerische Mathematik*, 31, 377–403.
- Duchon, J., 1975. Fonctions splines et vecteurs aleatoires. Séminaire d'Analyse Numérique, Univ. Sci. Médicale, Grenoble, France, Tech. Rep, 213.
- Eilers, P.H., and Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89–121.
- Eubank, R.L., 2004. Smoothing spline estimation in varying-coefficient models. *Journal of the Royal Statistical Society B*, 66, 653–667.
- Fan, J. and Zhang, W., 2008. Statistical methods with varying coefficient models. *Statistics and its Interface*, 1, 179–195.
- Fang, K.T., Kotz, S., and Ng, K.W. 1990. *Symmetric Multivariate and Related Distributions*. Chapman and Hall, London, UK.
- Galea, M. and Vilca, F., 2010. Hypotheses testing for comparing means and variances of correlated response in the symmetric non-normal case. *Chilean Journal of Statistics*, 1(2), 113–123.
- Green, P.J. and Silverman, B.W., 1994. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall, London, UK.
- Harrison, D. and Rubinfeld, D.L., 1978. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5, 81–102.
- Hastie, T. and Tibshirani, R., 1990. *Generalized Additive Models*. Wiley, New York, USA.
- Hastie, T. and Tibshirani, R., 1993. Varying-coefficient models. *Journal of the Royal Statistical Society B*, 55, 757–796.
- Huber, P.J., 1981. *Robust Statistics*. Wiley, New York, USA.
- Hurvich, C.M., Simonoff, J.S., and Tsai, C.L., 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society B*, 60, 271–293.
- Ibache-Pulgar, G. and Paula, G.A., 2011. Local influence for Student-t partially linear models. *Computational Statistics and Data Analysis*, 55, 1462–1478.
- Ibache-Pulgar, G., Paula, G.A., and Cysneiros, F.J.A., 2013. Semiparametric additive models under symmetric distributions. *TEST*, 22, 103–121.
- Ibache-Pulgar, G., Paula, G.A., and Galea, M., 2012. Influence diagnostics for elliptical semiparametric mixed models. *Statistical Modelling*, 12, 165–193.
- Ibache-Pulgar, G. and Reyes, S., 2018. Local influence for elliptical partially varying-coefficient model. *Statistical Modelling*, 2, 149–174.
- Krafty, R.T., 2008. Varying coefficient model with unknown within-subject covariance for analysis of tumor growth curves. *Biometrics*, 64, 1023–1031.

- Lange, K.L., Little, R.J., and Taylor, J.M., 1989. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84, 881–896.
- Liu, S. and Li, G., 2015. Varying-coefficient mean–covariance regression analysis for longitudinal data. *Journal of Statistical Planning and Inference*, 160, 89–106.
- Marciano, F.W., Blas, B.G., and Cysneiros, F.J.A., 2013. Symmetrical linear calibration model to replicated data. *Chilean Journal of Statistics*, 7(1), 3–16.
- Opsomer, J.D. and Ruppert, D., 1999. A root- n consistent backfitting estimator for semi-parametric additive modeling. *Journal of Computational and Graphical Statistics*, 8, 715–732.
- Park, B.U., Mammen, E., Lee, Y.K., and Lee, E.R., 2015. Varying coefficient regression models: A review and new developments. *International Statistical Review*, 83, 36–64.
- Relvas, C.E.M. and Paula, G.A., 2016. Partially linear models with first-order autoregressive symmetric errors. *Statistical Papers*, 57, 795–825.
- Russo, C.M., Paula, G.A., and Aoki, R., 2009. Influence diagnostics in nonlinear mixed-effects elliptical models. *Computational Statistics and Data Analysis*, 53, 4143–4156.
- Savalli, C., Paula, G.A., and Cysneiros, F.J.A., 2006. Assessment of variance components in elliptical linear mixed models. *Statistical Modelling*, 6, 59–76.
- Schwarz, G., 1978. Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464.
- Segal, M.R., Bacchetti, P., and Jewell, N.P., 1994. Variances for maximum penalized likelihood estimates obtained via the EM algorithm. *Journal of the Royal Statistical Society B*, 56, 345–352.
- Silverman, B.W., 1985. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society*, 47, 1–21.
- Wahba, G., 1983. Bayesian “confidence intervals” for the cross-validated smoothing spline. *Journal of the Royal Statistical Society B*, 45, 133–150.
- Wang, L., Kai, B., and Li, R., 2009. Local rank inference for varying coefficient models. *Journal of the American Statistical Association*, 104, 1631–1645.
- Wood, S.N., 2003. Thin plate regression splines. *Journal of the Royal Statistical Society B*, 65, 95–114.
- Wood, S.N., 2006. *Generalized additive models: an introduction with R*. Chapman and Hall, New York, US.

STATISTICAL MODELING
RESEARCH PAPER

On a weighted Poisson distribution and its associated regression model

BERNARDO B. DE ANDRADE^{1,*}, RAUL Y. MATSUSHITA¹, PUSHPA N. RATHIE¹, LUAN OZELIM²,
and SANDRO B. DE OLIVEIRA³

¹Departamento de Estatística, Universidade de Brasília, Brasília, Brazil

²Departamento de Engenharia Civil e Ambiental, Universidade of Brasília, Brasília, Brazil

³Rede Sarah de Hospitais de Reabilitação, Brasília, Brazil

(Received: 20 November 2020 · Accepted in final form: 03 April 2021)

Abstract

Count data emerge naturally within the biomedical and economic sciences, in engineering and in industrial applications. The benchmark Poisson distribution is seldom an appropriate statistical model for counts but none of the more flexible distributions available are universally accepted as an alternative. Among such flexible models, the class of weighted Poisson distributions has recently been studied in theoretical investigations but their application is still incipient. This article investigates a particular weighted Poisson model, providing the associated statistical tools for analyzing count data. We make comparisons with other flexible models using public available datasets. For the weighted Poisson model under investigation, we have developed estimation by maximum likelihood and method of moments, random number generation, visual tools for univariate analysis and finally, regression modeling. Results indicate that weighted Poisson distributions are very flexible and capable of modeling count responses in different scenarios.

Keywords: Generalized linear models · Overdispersion · Quasi-likelihood · Touchard · Underdispersion

Mathematics Subject Classification: Primary MSC 62J12 · Secondary MSC 62Fxx

1. INTRODUCTION

The Poisson model is the default for analyzing statistically independent counts, such as number of insurance claims and days of hospitalization, and it should provide an adequate fit when data come from a population with mean equal to the variance (equidispersion). However, count data often exhibit over or underdispersion and several distributions have been proposed for modeling counts. The most documented alternatives to the Poisson distribution are the quasi-Poisson (QP) model, the negative binomial (NB) and, to a lesser extent, the generalized Poisson (GP) (Cameron and Trivedi, 1998; Hilbe, 2014). Less popular models include the Poisson-inverse Gaussian, the compound Poisson, the hyper Poisson, the Poisson-Lindley, the Conway-Maxwell-Poisson (CMP) and the weighted Poisson (WP); see

*Corresponding author. Email: bbandrade@umb.br

Matsushita et al. (2018) and references therein. More complex models include the Poisson-Goncharov (Denuit, 1997), the Hinde-Demétrio family by Kokonendji et al. (2004) and a four-parameter extension of the CMP by Chakraborty and Imoto (2016). It is a long, yet incomplete, list and no single model dominates all others since there are multiple interrelated criteria for judging a given model. One might consider, in his own order of importance: (i) the model's ability to address varying levels in both directions of dispersion; (ii) mathematical and computational tractability; (iii) whether the model generalizes the Poisson; (iv) if the model arises naturally in some observable process like natural phenomena or in connection with a stochastic process of broad applicability; (v) if the model is a member of the exponential family (EF); (vi) availability of estimation and visualization tools for data analysis; (vii) availability of associated regression tools and interpretability of coefficients; (viii) applicability in a specific field (say, actuarial modeling of claim counts or econometric modeling of health insurance, etc.) and (ix) applicability across many disciplines.

Theoretical aspects of WP distributions, covering much of (i)–(v), have been studied by several authors (del Castillo and Pérez-Casany, 1998; del Castillo and Pérez-Casany, 2005; Kokonendji et al., 2008; Matsushita et al., 2018). However, the statistical toolbox for WP models still lacks several important tools. This paper focuses on a particular WP model which we call the Touchard model introduced by del Castillo and Pérez-Casany (1998). Ho et al. (ress) present simulation showing the effect of misspecifying the model (Poisson instead of Touchard) in the context of control charts.

In Section 2 we describe the methodological background of WP distributions including the Touchard. New results for the Touchard model are given in the Appendix A-D. Section 3 develops tools for univariate estimation, inference and visualization. Finally, regression modeling is addressed in Section 4 where we develop estimation, inference and diagnostic tools similar to those in generalized linear models (GLMs). Section 5 concludes the paper.

All computations and graphics presented here were done in the R system and all the required code will be included in the next release of the publicly available package (Andrade and Oliveira, 2019) which, so far, only deals with fixed a .

2. WEIGHTED POISSON MODELS

Weighted distributions date back to Fisher (1934) and have been used to adjust a given benchmark model relative to the way the data are ascertained (Rao, 1985). The adjusted distribution is used to model observational data recorded without a suitable sampling frame including situations such as size-biased sampling, damage models, nonresponse and visibility bias (Patil and Rao, 1978). Under idealized conditions, an observed value y would be a realization of a benchmark random variable Y^* . (For instance, one can think that the number of insurance claims would be Poisson distributed if there were no hunger for bonus (deliberate non-reporting of accidents to save bonus on next premium), if all drivers were insured (no selection bias), if all (insured) drivers were subjected to the same routes and same driving distances, etc. However, not only policies are not randomly drawn with a proper sampling frame but also the conditions just listed are not met. The result is a distribution of claims biased towards 0 and 1.) However, the benchmark may need to be adjusted so that its support is reweighted according to the belief that when the event $Y = y$ is realized, the probability of ascertaining it is $w(y)$. Thus, the realization y is, in fact, from a weighted version Y with probability density function $f(y) = w(y)f^*(y)/\tau$, where f^* is the benchmark density and τ is the normalization constant.

When modeling counts, the benchmark density is often Poisson, $f^*(y) = e^{-\lambda}\lambda^y/y!$, leading to the general class of WP distributions which we denote by $WP(\lambda; w(y))$. The model $WP(\lambda; y)$ is known as the size-biased Poisson model and its distribution is simply $1 + Poi(\lambda)$. Another important model is $WP(\lambda; (y!)^\nu)$, for $\nu > 0$, which is the CMP distribution.

We also note that important models such as the NB and the GP are not members of the $WP(\lambda; w(y))$ class. Interestingly, the NB distribution is a Poisson–Gamma mixture, in which the parameter space of the Poisson is weighted, rather than its support.

A family within the $WP(\lambda; w(y))$ class is obtained with $w(y) = \exp[\delta t(y)]$, where t is a convex function and $\delta \in \mathbb{R}$ provides overdispersion ($\delta < 0$), Poissonness ($\delta = 0$) or underdispersion ($\delta > 0$). We denote this family by $WP(\lambda, \delta; t)$. [Kokonendji et al. \(2008\)](#) showed, among other theoretical aspects, that such WP models are pointwise dual in the sense that the entire range of δ is guaranteed to account for either over or underdispersion of the same magnitude.

An important member in the $WP(\lambda, \delta; t(y))$ family is the one with $t(y) = \log(y + a)$, for $a \geq 0$ ([del Castillo and Pérez-Casany, 1998](#)). Also, [del Castillo and Pérez-Casany \(2005\)](#) fitted this model (with fixed $a = 7$) to the counts of car accidents in a year among 9461 drivers. We have fitted the $WP(\lambda, \delta; \log(y + a))$ model to different datasets and results indicate that it is a strong competitor for both under and overdispersed cases; see [Table 2](#).

2.1 THE TOUCHARD DISTRIBUTION REDEFINED

The $WP(\lambda, \delta; \log(y + 1))$ model has been recently studied by [Matsushita et al. \(2018\)](#) where the choice of $a = 1$ was inspired by the Touchard polynomials, rather than by weighting schemes. Those authors have labeled it the Touchard distribution, which we redefine. Note that [del Castillo and Pérez-Casany \(1998\)](#) denoted the $WP(\lambda, \delta; \log(y + a))$ model by WPD_a , since a is a fixed tuning parameter. The resulting two-parameter model is more tractable both mathematically and computationally but it is not justifiable for actual data analysis across different disciplines. In order to distinguish the WPD_a model from the general case, which allows for choices of $w(y)$ other than $(y + a)^\delta$, and also to avoid confusion with the use of weights (in the context of regression), we hereafter call the $WP(\lambda, \delta; \log(y + a))$ model the Touchard model and denote $Y \sim \text{Tu}(\lambda, \delta, a)$. Its probability density function is given by

$$f(y, \lambda, \delta, a) = \frac{\lambda^y (y + a)^\delta}{y! \tau(\lambda, \delta, a)}, \quad y = 0, 1, \dots, \quad (1)$$

with $a, \lambda > 0$ and $\delta \in \mathbb{R}$. Note that if $\delta \neq 0$ then $a = 0$ can be considered, in which case $\mathbb{P}(Y = 0) = 0$. We do not consider this case here.

Numerical evaluation of the normalization constant $\tau(\lambda, \delta, a)$ can be done by truncation of its defining sum. The number of terms required to reach a given precision depends on the parameter values. Even though 50 terms suffice in most cases ([Matsushita et al., 2018, Table 1](#)), τ can be computed with a pre-specified relative precision, without fixing the truncation point, using a recursive expression similar to the formula given in (10) of [Matsushita et al. \(2018\)](#); see also [Appendix A](#).

The values of a and δ jointly determine the shape of the distribution. With $\delta < 0$, the smaller a is, the more the distribution is inflated at zero and the resulting model becomes an alternative to zero-inflated and hurdle models. Larger values of a provide milder zero inflation for given δ . With $\delta > 0$, the smaller a is, the more the distribution is deflated at zero relative to the Poisson. Larger values of a decrease the zero deflation, making the distribution closer to Poisson. [Figure 1](#) illustrates these facts for some selected values of the parameters.

From now on, we define $\tau_j(\lambda, \delta, a) = \tau(\lambda, \delta + j, a) / \tau(\lambda, \delta, a)$. Statistical moments $m_k = \mathbb{E}(Y^k)$ are given by (del Castillo and Pérez-Casany, 1998)

$$m_k = \sum_{j=0}^k \binom{k}{j} (-1)^{k-j} \tau_j(\lambda, \delta, a), \quad (2)$$

yielding the mean given by

$$\mu = \tau_1(\lambda, \delta, a) - a, \quad (3)$$

and the variance stated as

$$\sigma^2 = \tau_2(\lambda, \delta, a) - \tau_1^2(\lambda, \delta, a). \quad (4)$$

Touchard quantiles may be calculated with an initial approximation based on the Cornish-Fisher expansion (using up to m_3) followed by a search in the appropriate direction as implemented in Andrade and Oliveira (2019).

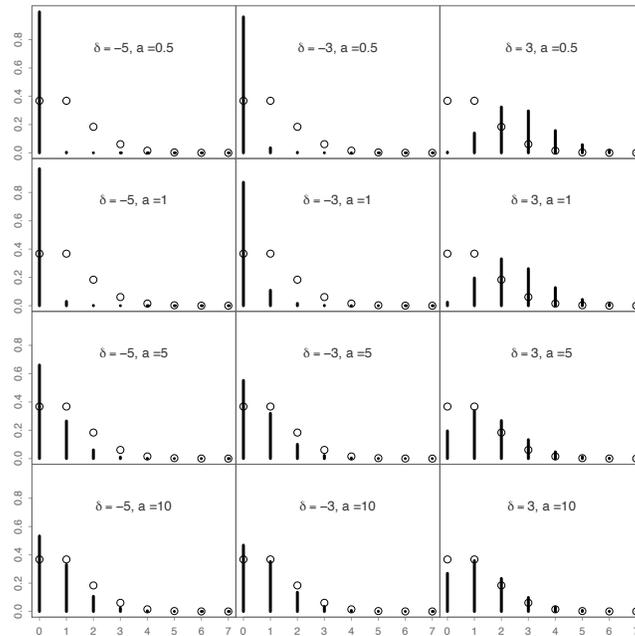


Figure 1. Examples of WP (Touchard) probabilities with $\lambda = 1$, $\delta = \pm 3$ and $a = 0.5, 1, 5$ and 10 . Circles represent Poisson probabilities with $\lambda = 1$.

Exact (or approximate) results for the variance as a function of the mean are given in Table 1 for four models of interest besides the Touchard: the two usual benchmarks, QP and NB, and the CMP and GP which are well documented in the literature and which have been implemented for data analysis in the R system, including regression. The QP, NB and GP exhibit polynomial relations between their means and variances with linear, quadratic and cubic relations, respectively. As opposed to these models, the difference between the mean and the variance in the Touchard and in the CMP models are constant for large enough μ . Figure 2 illustrates the mean-variance relation for the Touchard model with different parameter values. The larger $|\delta|$, the further away the curve (μ, σ^2) is from the 45° diagonal (equidispersion). Larger values of a bring the curve closer to the diagonal and it also decreases the initial curvature. The curvature is a lot more sensible to the value of a

and the magnitude of δ in the overdispersed cases ($\delta < 0$). Therefore, it must be noted that even though the mean and variance are not explicit parameters, the Touchard model can be implicitly reparametrized by the mean and the variance (del Castillo and Pérez-Casany, 2005).

Table 1. Mean-variance relations for selected models.

	QP(μ, φ)	NB(μ, α)	GP(μ, α)	Tou(λ, δ, a)	CMP(λ, ν)
σ^2	$\varphi\mu$	$\mu + \alpha\mu^2$	$\mu(1 + \alpha\mu)^2$	$\approx \mu - \delta \left[\frac{\lambda}{\lambda+a} \right]^2$	$\approx \frac{\mu + (\nu-1)/(2\nu)}{\nu}$

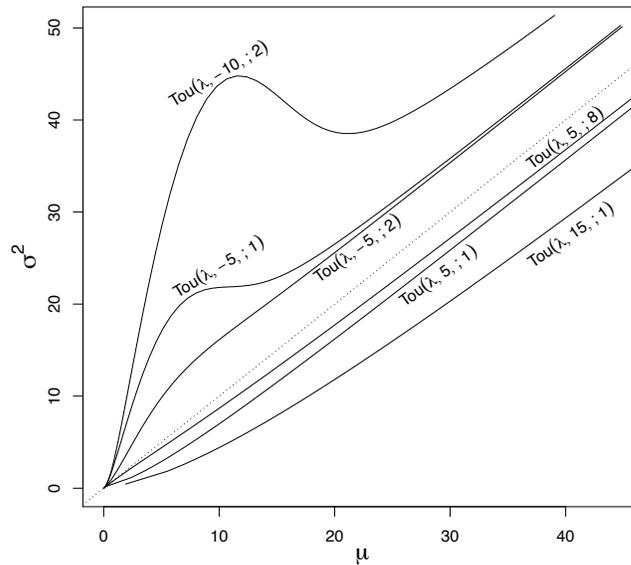


Figure 2. Variance as a function of the mean for Touchard models with arbitrary parameter values. Dotted line on diagonal represents the Poisson $\sigma^2 = \mu$.

3. ESTIMATION, TESTING AND VISUALIZATION

Here, we describe estimation and testing for the Touchard parameters via maximum likelihood (ML) method of moment method (MM). Conditions for consistency and normal-based large-sample inference require the support to be independent of the parameters, identifiability and bounds on third derivatives (Lehmann and Casella, 1998, Sec. 6.5). We provide a proof of identifiability in Appendix B. With a fixed, the Touchard distribution is a member of the two-dimensional EF with sufficient statistics Y and $Z = \log(Y + a)$ and natural parameters $\log(\lambda)$ and δ . For instance, we already know that $\partial^3 \log(f(y))/(\partial\lambda^2\partial\delta)$ must be bounded by some $M(Y)$ with finite expectation. Therefore, we only need to check third-order derivatives of the log-likelihood involving a . These derivates involve (finite) moments of Y, Z and $(Y + a)^{-k}$, for $k \leq 3$, and therefore satisfy the necessary conditions.

3.1 MAXIMUM LIKELIHOOD ESTIMATION

Given a random sample (Y_1, \dots, Y_n) with observations (y_1, \dots, y_n) , the ML estimates $\hat{\lambda}$, $\hat{\delta}$ and \hat{a} must satisfy the conditions stated as

$$\begin{cases} \mu(\hat{\lambda}, \hat{\delta}, \hat{a}) - \bar{y} & = 0, \\ \kappa(\hat{\lambda}, \hat{\delta}, \hat{a}) - \bar{z} & = 0, \\ \theta_1(\hat{\lambda}, \hat{\delta}, \hat{a}) - \bar{w}_1 & = 0, \end{cases}$$

where $\kappa(\lambda, \delta, a) \equiv \mathbb{E}(Z)$, $W_j = (Y + a)^{-j}$ and $\theta_j(\lambda, \delta, a) \equiv \mathbb{E}(W_j)$.

Using results from Appendix C, it can be shown that the expected Fisher information is given by

$$\mathbf{I}(\lambda, \delta, a) = n \begin{pmatrix} \frac{\sigma^2}{\lambda^2} & \frac{\text{Cov}(Y, Z)}{\lambda} & \frac{\delta \text{Cov}(Y, W_1)}{\lambda} \\ & \text{Var}(Z) & \delta \text{Cov}(Z, W_1) \\ & & \delta^2 \text{Var}(W_1) + \delta(W_2 - \theta_2) \end{pmatrix}.$$

Standard errors (SEs) of ML estimators are computed from the diagonal of $\mathbf{I}^{-1}(\hat{\lambda}, \hat{\delta}, \hat{a})$ or from the inverse of observed Fisher information which is often produced by (Newton-type) numerical maximization routines.

3.2 METHOD OF MOMENTS

Based on the expression given in Equation (2), the moment conditions for the MM estimators are expressed as

$$\begin{cases} \mu(\hat{\lambda}, \hat{\delta}, \hat{a}) - \bar{y} & = 0, \\ m_2(\hat{\lambda}, \hat{\delta}, \hat{a}) - \frac{1}{n} \sum y_i^2 & = 0, \\ m_3(\hat{\lambda}, \hat{\delta}, \hat{a}) - \frac{1}{n} \sum y_i^3 & = 0. \end{cases}$$

SEs are obtained from the diagonal of a sandwich estimate of the asymptotic variance, $n\hat{\mathbf{G}}^\top \hat{\mathbf{V}}^{-1} \hat{\mathbf{G}}$, where $\hat{\mathbf{G}}$ is a consistent estimator of the expected value of the gradient associated with the moment conditions and $\hat{\mathbf{V}}$ consistently estimates the associated expected Hessian (Cameron and Trivedi, 1998, Sec. 2.8.4).

3.3 SCORE TEST

A score test for $H_0: \delta = 0$ is based on fitting the Poisson model, which is the distribution under the null hypothesis. Define $\nu(\lambda, \delta, a) \equiv \text{Var}(Z)$ and $\rho(\lambda, \delta, a) \equiv \text{Corr}(Y, Z)$. Evaluation under H_0 , in which case $\hat{\lambda} = \bar{y}$ and $\delta = 0$, is represented by a subscript zero (0). The resulting test statistic is stated as

$$S = \frac{n(\bar{z} - \kappa_0)^2}{\nu_0(1 - \rho_0^2)}, \quad (5)$$

which is asymptotically $\chi^2(1)$ distributed under H_0 . We do not provide an explicit proof since this is a special case of the general result derived at the end of Section 4.1.

3.4 VISUALIZATION

Basic visualization of an observed versus a theoretical (or fitted) distribution of counts can be readily achieved by barplots or, preferably, in the form of a rootogram. The rootogram can be adapted to regression settings as seen in Section 4.3. Another visualization aimed at assessing the goodness of fit is the Touchardness plot which compares deviations between observed data and the Touchard model. These two plots are now described.

A clever visualization tool for count data is the Tukey hanging rootogram. It is a variation of the histogram with the vertical axis showing the square root of the frequencies to de-emphasize outlying values and right skewness (common for count data). The usual bars “hang” from the fitted values so that the discrepancies are visualized against a straight line (the axis) rather than against a curve. The bars are drawn from $\sqrt{\hat{y}}$ to $\sqrt{\hat{y}} - \sqrt{\bar{y}}$. A generalization of the rootogram for regression is shown in Figure 4 (right) for counts of crab satellites predicted by color and weight.

Hoaglin and Tukey (2009) devised a goodness-of-fit plot for count data based on the count metameter. The key idea is to compare the observed frequency of count y , denoted hereafter by n_y , with the expected frequency from a given model. Assuming a $\text{Tou}(\lambda, \delta, a)$ model and n data points, the count metameter, $\varphi(n_y)$, is obtained through the equation $n_y = nf(y, \lambda, \delta, a)$ which yields the expression given by

$$\varphi(n_y) \equiv \log\left(\frac{y!n_y}{n(y+a)^\delta}\right) = -\log[\tau(\lambda, \delta, a)] + \log(\lambda)y. \quad (6)$$

Deviations between the observed counts and the theoretical model are assessed by examining the relation $\varphi(n_y) \times y$. If the points $(\varphi(n_y), y)$ follow a straight line with $\exp(\text{slope})$ close to $\hat{\lambda}$, we have indication that the Touchard model is appropriate. In practice, δ is fixed at $\hat{\delta}$ in order to compute $\varphi(n_y)$ and the intercept and slope are obtained by least squares. In addition, Hoaglin and Tukey (2009) proposed an approximate confidence interval for the logarithm of the theoretical frequency. Figure 3 shows the Touchardness plot for data on counts of crab satellites, indicating an adequate fit.

We can obtain starting values $(\lambda^0, \delta^0, a^0)$ for numerical procedures associated with ML and MM based on Equation (6). Consider a in a grid, say $a \in \{0.1, 0.5, 1, 1.5, \dots, a_{\max}\}$, where a_{\max} is arbitrarily defined. One idea is to fit the linear model stated as

$$\log\left(\frac{y!n_y}{n}\right) = \beta_0 + \beta_1 y + \beta_2 \log(y+a)$$

and set $\lambda^0 = \exp(\hat{\beta}_1)$, $\delta^0 = \hat{\beta}_2$ and a^0 yielding the fit with the smallest sum of squared errors. Alternatively, one may re-interpret the model given in Equation (6) as a Poisson log-linear model with $\log(n/y!)$ as offset,

$$\log(n_y) = \beta_0 + \beta_1 y + \beta_2 \log(y+a) + \log(n/y!),$$

and again set $\lambda^0 = \exp(\hat{\beta}_1)$, $\delta^0 = \hat{\beta}_2$ and a^0 yielding the fit with the smallest deviance.

3.5 ILLUSTRATION WITH SELECTED DATASETS

Next, we illustrate the Touchard model and associated tools with univariate, public available, data of actuarial end economic interest. We consider the five models listed in Table 1. Except for the NB, the other models handle both under and overdispersion. The QP and the Touchard models are pointwise dual (in the sense that the entire range of δ accounts

for either over or underdispersion of the same magnitude) whereas the CMP is pointwise dual only in the region $0 < \nu < 2$ (Kokonendji et al., 2008). The QP, CMP and NB are members of the EF (thus enjoying the desirable asymptotic results for ML estimation), though only the QP is in the one-dimensional EF with dispersion. The support of the GP distribution depends on the parameter α and it violates standard conditions for consistency and asymptotic normality of ML estimators (Cameron and Trivedi, 1998).

We have fitted different count models, including the Touchard one, to 14 publicly available datasets. The datasets were not selected with any particular criteria other than being publicly available and having been used by the authors for class illustrations, mostly for audiences from Actuarial Sciences and Economics. Most cases exhibit high inflation of zeros ($\hat{f}_0 > 0.80$), dispersion index $d = S^2/\bar{y}$ greater than one and large sample size.

Eight datasets refer to the number of traffic accidents in a year in different locations labeled by country name and year such as `Zaire74`, `Belgium58`, etc. Some have become benchmarks in the actuarial literature (Denuit, 1997). Typical of such data, the relative frequency of zeros is very high.

The `CrabSat` dataset features the highest level of overdispersion for the response among the cases studied with d close to 3. The counts are from well-known data on the number of satellites (male crabs gathered around the female attempting to fertilize her eggs) appearing in several textbooks (Agresti, 2013).

`MedVisits` consists of over 5000 counts of doctor visits in the past two weeks for a single-adult (Australian Health Survey 1977-78) and has been analyzed by several authors (see, for example, Cameron and Trivedi (1998) and references therein) in the study of health service utilization and health insurance choice. The sample variance is about twice the sample mean ($d = 2.1$). Zeros and ones correspond to 95% of the observations. These data are also used in Section 4.1 in the context of regression modeling.

The dataset `Strikes` records the number of outbreaks of strikes in the UK, in a 4-week period, during 1948-59 for the coal mining industry.

`Shells` brings the number of accidents in the manufacture of high-explosive shells in a British military factory at the time of World War I.

The dataset `Bids` contains the number of takeover bids received by 126 U.S. firms that were targets of tender offers, over a 52-week period following the initial bid.

`AZCardio` contains close to 2000 observations from the 1991 Arizona cardiovascular patient files. The counts refer to the length (days beyond the day of admission) of hospital stay (restricted to less than 9 days) for cardiovascular patients.

Table 2 summarizes the results from fitting five statistical models to the 14 datasets previously described. the Akaike information criterion (AIC) is reported along with data summaries. Qualitatively, similar results were obtained using the χ^2 metric (not shown) instead of the AIC. The Touchard model yields is among the best models all cases with the exception of `Bids` for which convergence was not achieved; these data seem to be well fit by the Poisson model, meaning that a three-parameter distribution is unnecessary.

In Section 4.1, we revisit the `CrabSat` data for a regression illustration. We thus provide more detail regarding its univariate mode fit. Parameter estimates under the Touchard model from ML and MM methods are shown in Table 3 and visualization of model fit by ML is depicted in Figure 3 showing an adequate fit. As expected, both the score and the likelihood ratio tests (not shown) strongly reject $\delta = 0$. The AIC values for different models are: 746 (Touchard), 771 (NB), 774 (CMP) and 782 (GP). The results place the Touchard as the primary candidate for modeling these data. We must note that some differences in AIC values are small for different models and that mere minimization of a given information criterion is not by itself a deal breaker in model selection. Several other points may be assessed including study domain knowledge and substantive interpretability of models. We remind the reader of the criteria (i)–(ix) listed in the Introduction. The main point to be

taken here is that this superficial examination of the applicability of the Touchard model across datasets from different domains strongly indicates that the Touchard model (and weighted Poisson models in general) are an important addition to the toolbox of count data models. Recall that the Touchard can address both under- and over-dispersion.

Table 2. Data summaries and AIC with selected count models fitted by ML to 14 datasets.

Dataset	n	y_{\max}	d	\hat{f}_0	AIC by model				
					Poisson	NB	Tou	CMP	GP
CrabSat	173	15	2.9	0.36	990	771	746	774	782
MedVisits	5,190	9	2.1	0.80	7,968	7,176	7,125	7,320	7,156
Shells	647	5	1.5	0.69	1,236	1,189	1,189	1,189	1,189
AZCardio	1,982	7	1.5	0.07	9,055	8,796	8,726	8,753	8,807
Zaire74	4,000	5	1.4	0.93	2,494	2,372	2,371	2,419	2,371
Belgium58	9,461	7	1.4	0.83	10,983	10,700	10,691	10,713	10,696
Switz61	119,853	6	1.2	0.88	109,225	109,234	109,226	109,235	109,230
Bids	126	10	1.2	0.07	405	406	NA	407	406
NewYork93	365	8	1.2	0.06	1,450	1,449	1,461	1,449	1,449
Belgium75	106,974	4	1.1	0.91	72,379	72,212	72,213	72,213	72,212
Belgium93	63,299	4	1.1	0.90	44,303	44,133	44,130	44,134	44,131
Belgium94	131,182	4	1.1	0.90	90,453	90,163	90,162	90,164	90,162
Germany60	23,589	6	1.1	0.87	20,598	20,451	20,449	20,451	20,450
Strikes	156	4	0.7	0.29	386	388	381	380	382

Table 3. Estimates (with SEs) from fitting the $Tou(\lambda, \delta, a)$ model to the counts in the CrabSat dataset.

	Method	
	ML	MM
λ	8.0 (1.2)	11.4 (3.3)
δ	-2.7 (0.83)	-5.6 (2.6)
a	0.4 (0.22)	1.3 (0.80)

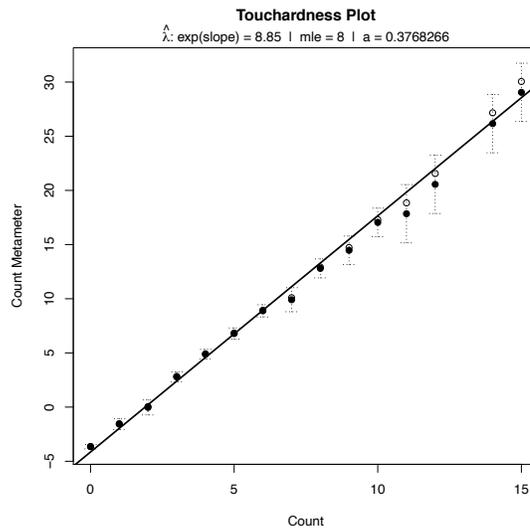


Figure 3. Goodness-of-fit (Touchardness) plots with the CrabSat dataset: circles are the observed count metameters; filled points show the CIs centers.

4. REGRESSION MODELING

As mentioned in the Introduction, there are several probabilistic models for count data but most have had limited applicability. Models for which regression tools are not readily available will not be considered by most data analysts. In this section we develop regression tools for the Touchard model, including ML estimation, large-sample inference, prediction, visualization and basic diagnostic tools.

Regression for count data is typically based on either the Poisson or the NB models. The (quasi) Poisson model with a dispersion parameter $\varphi \in \mathbb{R}$, $\text{QP}(\mu, \varphi)$, with expectation μ and variance $\varphi\mu$ can accommodate under/overdispersed data. QP regression is a GLM with a vast literature on estimation, inference and diagnostics. In the $\text{NB}(\mu, \alpha)$ model, the variance is a quadratic function of the mean, $\mu + \alpha\mu^2$, $\alpha \geq 0$, which can handle overdispersed data. The $\text{GP}(\mu, \alpha)$ distribution is another model which can be explicitly parametrized in terms of its mean and an extra parameter with the variance being a cubic function of μ . The QP model is a member of the one-dimensional EF with dispersion but neither the NB nor the GP is a member of the EF. The Touchard distribution is in three-dimensional EF but cannot be formulated in terms of the one-dimensional EF with dispersion. We recall that mean-variance relations for these models are reported in Table 1.

The first question posed by any WP regression model is the choice of parametrization for the systematic component. Given observed responses $Y_i \sim \text{Tu}(\lambda_i, \delta, a)$ and a vector of predictors $\mathbf{x}_i \in \mathbb{R}^p$, for $i = 1, \dots, n$, we can postulate that either. (We restrict our attention to the default log link but other links for count data could be considered.) Then, we have

$$\log(\lambda_i) = \mathbf{x}_i^\top \boldsymbol{\alpha}, \quad (7)$$

or

$$\log(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}. \quad (8)$$

We call the model given in Equation (7) is a direct regression model, which is computationally more convenient. This is the strategy adopted by [Sellers and Shmueli \(2010\)](#) in the context of the CMP regression model. In the GLM-type model given by Equation (8), λ_i must be treated as an implicit function of μ_i , δ and a which implies a computational cost related to solving Equation (3) for λ at every evaluation of the likelihood. We propose a quasi-likelihood approach to deal with the usual practice of linking the mean response to a linear predictor as in Equation (8) in the context of Touchard regression.

4.1 DIRECT REGRESSION MODEL

In this section we assume that λ is a function of the linear predictor $\mathbf{x}_i\boldsymbol{\alpha}$, as in Equation (7). The model log-likelihood for independent data points is given by $\ell(\boldsymbol{\alpha}, \delta, a) = \sum_i \ell_i$, where

$$\ell_i = y_i \mathbf{x}_i^\top \boldsymbol{\alpha} + \delta \log(y_i + a) - \log\left[\tau(\exp(\mathbf{x}_i^\top \boldsymbol{\alpha}), \delta, a)\right] - \log(y_i!).$$

Recall that we have defined $z_i = \log(y_i + a)$, $w_{1i} = (y_i + a)^{-1}$ and $w_{2i} = (y_i + a)^{-2}$. We also define $\kappa_i = \mathbb{E}(Z_i)$, $\gamma_i = \text{Cov}(Y_i, Z_i)$, $\theta_{ji} = \mathbb{E}(W_{ji})$ and $\nu_i = \text{Var}(Z_i)$. Denoting by $\mathbf{X}_{n \times p}$ the model matrix with i -th row \mathbf{x}_i^\top , the observed responses by the n -dimensional vector \mathbf{y} with

$\boldsymbol{\mu} = \mathbb{E}(\mathbf{y})$ and $\mathbf{V} = \text{diag}(\sigma_i^2)$, we can write the score vector as

$$\mathbf{s}(\boldsymbol{\alpha}, \delta, a) = \begin{pmatrix} \mathbf{X}^\top(\mathbf{y} - \boldsymbol{\mu}) \\ \sum_i (z_i - \kappa_i) \\ \delta \sum_i (w_{1i} - \theta_{1i}) \end{pmatrix} \tag{9}$$

and the Fisher information matrix is formulated as

$$\mathbf{F}(\boldsymbol{\alpha}, \delta, a) = - \begin{pmatrix} \mathbf{X}^\top \mathbf{V} \mathbf{X} & \mathbf{X}^\top \boldsymbol{\gamma} & \mathbf{X}^\top \mathbf{T} \mathbf{X} \\ & \sum_i \nu_i & \sum_i [\delta \text{Cov}(Z_i, W_{1i}) - (\theta_{1i} - W_{1i})] \\ & & \delta^2 \sum_i \text{Var}(W_{1i}) + \delta [\sum_i (W_{2i} - \theta_{2i})] \end{pmatrix}$$

where $\mathbf{T} = \text{diag}(t_i)$ with $t_i = 1 - \tau_1(\lambda_i, \delta, a)\tau_{-1}(\lambda_i, \delta, a)$. The usual inference based on the ML estimates uses asymptotic normality with

$$\widehat{\text{Var}}(\widehat{\boldsymbol{\alpha}}, \widehat{\delta}, \widehat{a}) = \widehat{\mathbf{F}}^{-1}.$$

The effect of the coefficients on the mean response is expressed as

$$\frac{\partial \mu_i}{\partial x_{ij}} = \sigma_i^2 \alpha_j. \tag{10}$$

Therefore, the marginal effect of α_j on μ_i is affected by the variability associated with the corresponding observation. Thus, over- or under-dispersion play a role that did not exist in the canonical Poisson model for which $\partial \mu_i / \partial x_{ij} = \mu_i \alpha_j$.

It is possible to obtain a score test for Poissonness. Calculations are facilitated by the fact that residuals and regressors are orthogonal, that is, $\mathbf{X}^\top(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0}$, when the score given in Equation (9) is zero. The score test allows us to test the null $H_0: \delta = 0$ without having to first fit the Touchard model.

Let $\boldsymbol{\alpha}_0$ be the coefficient estimates from H_0 (canonical Poisson GLM). We use the same notation as in Section 3.3, where the subscript 0 indicates evaluation under H_0 . Here, $\boldsymbol{\lambda}_0 = \exp(\mathbf{X}\boldsymbol{\alpha}_0)$ and the score vector given by Equation (9) becomes

$$\mathbf{s}_0 = (0, \dots, 0, K_0),$$

where $K_0 = \sum_i [z_i - \kappa_{i,0}]$. By partitioning \mathbf{F}_0 and writing its inverse (Graybill, 1983, Ch. 8), we have $F_{0,22}^{-1} = (F_{0,22} - \mathbf{F}_{0,21}\mathbf{F}_{0,11}^{-1}\mathbf{F}_{0,12})^{-1}$ and, after some algebra, the score statistic $\mathbf{S} = \mathbf{s}_0^\top \mathbf{F}_0^{-1} \mathbf{s}_0$ reduces to

$$\mathbf{S} = \frac{K_0^2}{F_{0,22}^{-1}} = \frac{K_0^2}{\mathbf{V}_0 - \boldsymbol{\gamma}_0^\top \mathbf{X}(\mathbf{X}^\top \mathbf{V}_0 \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\gamma}_0},$$

where $\mathbf{V}_0 = \sum_i \nu_{i,0}$. The special case with no covariates was given by Equation (5). The asymptotic distribution of S under H_0 is $\chi^2(1)$ given regularity conditions warranted by the EF.

4.2 QUASI-POISSON GLM WITH TOUCHARD VARIANCE

As mentioned before, a regression model based on Equation (8) is computationally more demanding since λ_i must be treated as an implicit function of μ_i , δ and a . Derivative calculations become a lot more involved. However, the coefficients in the usual (log-mean) parametrization are more easily interpreted since, as opposed to Equation (10), we have that

$$\frac{\partial \mu_i}{\partial x_{ij}} = \mu_i \beta_j,$$

so that β_j can be interpreted as a semi-elasticity, the proportionate change in the mean when x_j changes by one unit, all else constant.

A solution is available by the well established quasi-likelihood approach. The QP(μ, φ) regression estimates, $\hat{\beta}$, are robust to distributional assumptions in the sense that the Poisson model is used simply to motivate the estimating equations, and that $\hat{\beta}$ is consistent as long as the link function and linear predictor are correctly specified (Agresti, 2015, Ch. 8). The choice of $v(\mu, \varphi) = \varphi\mu$ corresponds to assuming an EF with dispersion but other forms of conditional variance may also be entertained (Cameron and Trivedi, 1998, Ch. 3) such as with the Touchard variance $v(\mu, \delta, a)$ given by Equation (4). We denote this model by QPT(μ, δ, a). The estimated regression coefficients are simply those from a Poisson GLM, $\hat{\beta}$, whereas the other parameters are estimated by the MM, obtained by numerically solving the system in (δ, a) stated as

$$\begin{cases} \sum_{i=1}^n v(\hat{\mu}_i, \delta, a) &= \sum_{i=1}^n (y_i - \hat{\mu}_i)^2, \\ \sum_{i=1}^n \kappa(\hat{\mu}_i, \delta, a) &= \sum_{i=1}^n z_i. \end{cases}$$

If the variance is believed to be correctly specified (in which case $\hat{\beta}$ is asymptotically efficient among estimators that are locally linear in \mathbf{y}) then

$$\widehat{\text{Var}}(\hat{\beta}) = \frac{n}{n-p} (\mathbf{X}^\top \widehat{\mathbf{M}} \mathbf{X})^{-1} (\mathbf{X}^\top \widehat{\mathbf{V}} \mathbf{X}) (\mathbf{X}^\top \widehat{\mathbf{M}} \mathbf{X})^{-1},$$

where $\widehat{\mathbf{M}}$ is diagonal with typical element $\hat{\mu}_i$ and $\widehat{\mathbf{V}}$ is diagonal with typical element $v(\hat{\mu}_i, \hat{\delta}, \hat{a})$.

4.3 DIAGNOSTICS

This section develops standard diagnostic tools to assess the fit of a Touchard regression model including an adaptation of the rootogram. We denote the predicted mean by $\hat{\mu}$. In the case of Equation (7), then $\hat{\mu}_i = \exp(\mathbf{x}_i^\top \hat{\beta})$. In the case of Equation (8), we have $\hat{\mu}_i = \tau_1(\hat{\lambda}_i, \hat{\delta}, \hat{a}) - \hat{a}$ with $\hat{\lambda}_i = \exp(\mathbf{x}_i^\top \hat{\alpha})$. For the predicted λ we develop the following notation: we write $\lambda = \Lambda(\mu, \delta, a)$ for λ satisfying Equation (3) with given values of μ , δ and a . The most intuitive residual measure in regression models is the Pearson residual given by

$$r_i = \frac{y_i - \hat{\mu}_i}{\hat{\sigma}_i}.$$

Pearson residuals may be visualized by a Q-Q normal plot of standardized r_i with a simulated envelope (Atkinson, 1985).

We use the usual definition of deviance in GLMs, $D(\mathbf{y}; \boldsymbol{\mu}) = 2(\tilde{\ell} - \hat{\ell})$, to define an approximate measure of deviance where we consider δ and a fixed (at estimated values). Here, $\hat{\ell}$ denotes the maximized log-likelihood and $\tilde{\ell}$ is the saturated log-likelihood. Saturation is achieved by setting $\mu_i = y_i$ since $\partial \ell_i / \partial \mu_i = 0 \Leftrightarrow (y_i - \mu_i) / \sigma_i^2 = 0$. As a result, deviance residuals, d_i , are defined by the signed square root of the components of $D(\mathbf{y}; \boldsymbol{\mu}) = \sum_1^n d_i^2$, as

$$d_i = \begin{cases} \text{sign}(y_i - \hat{\mu}_i) \sqrt{2 \left[y_i \log\left(\frac{\tilde{\lambda}_i}{\hat{\lambda}_i}\right) - \log\left(\frac{\tilde{\tau}_i}{\hat{\tau}_i}\right) \right]}, & y_i > 0, \\ \text{sign}(y_i - \hat{\mu}_i) \sqrt{2 \left[\log(\hat{\tau}_i) - \delta \log(\hat{a}) \right]}, & y_i = 0, \end{cases}$$

where $\tilde{\lambda}_i = \Lambda(y_i, \hat{\delta}, \hat{a})$ and $\hat{\lambda}_i$ is either $\exp(\mathbf{x}_i^\top \hat{\boldsymbol{\alpha}})$ or $\Lambda(\exp(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}), \hat{\delta}, \hat{a})$.

Observations with a large absolute value of either r_i or d_i are viewed as discrepant. Detectable patterns in the plot of residuals against the estimated linear predictor are indicative of misspecification.

Another diagnostic measure is the generalized leverage defined as the diagonal of $\mathbf{L}_{n \times n}(\boldsymbol{\theta}) = \partial \hat{\mathbf{y}} / \partial \mathbf{y}^\top$. The actual computation of \mathbf{L} is based on first derivatives of the mean vector, $\boldsymbol{\mu}_\theta$, and second derivatives of the likelihood, $\check{\ell}_{\theta\theta}$ and $\check{\ell}_{\theta\mathbf{y}}$, with

$$\mathbf{L} = \boldsymbol{\mu}_\theta (-\check{\ell}_{\theta\theta})^{-1} \check{\ell}_{\theta\mathbf{y}},$$

evaluated at $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}, \hat{\delta}, \hat{a})$ or $(\hat{\boldsymbol{\beta}}, \hat{\delta}, \hat{a})$ (Wei et al., 1998).

In the regression based on Equation (7), the computation of \mathbf{L} yields the expression stated as

$$\mathbf{L} = \begin{pmatrix} \mathbf{V}\mathbf{X} & \boldsymbol{\gamma} & \boldsymbol{\eta} \end{pmatrix} \mathbf{F}^{-1} \begin{pmatrix} \mathbf{X}^\top \\ \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \end{pmatrix},$$

where the vector $\boldsymbol{\eta}$ has i -th component $\delta \text{Cov}(Y_i, W_{1i})$, \mathbf{a}_1 has i -th component $1/(y_i + a)$ and \mathbf{a}_2 has i -th component $-\delta/(y_i + a)^2$. In the GLM-type model based on Equation (8), the computation of \mathbf{L} yields the usual projection matrix of a Poisson GLM with the weights adjusted for the Touchard variance, that is, we get

$$\mathbf{L} = \mathbf{W}^{-1/2} \mathbf{X} (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{-1/2}.$$

where $\mathbf{W}_{n \times n}$ is diagonal with typical element μ_i^2 / σ_i^2 .

With the diagonal of \mathbf{L} and the Pearson residuals, one can compute the approximate Cook distance given by

$$C_i = \frac{\mathbf{L}_{ii} r_i^2}{p(1 - \mathbf{L}_{ii})^2},$$

to measure the squared distance between $\hat{\boldsymbol{\alpha}}$ (or $\hat{\boldsymbol{\beta}}$) and the same estimate without i -th observation (Cook and Weisberg, 1982).

The extension of the rootogram (Section 3.4) to regression models has been proposed by Kleiber and Zeileis (2016) as a complement to residual diagnostics in order to visualize important features of count data such as dispersion, skewness, zero inflation and multimodality,

vis a vis fitted models.

Given regression estimates $\hat{\alpha}$ (or $\hat{\beta}$), $\hat{\delta}$ and \hat{a} , one obtains $\hat{\lambda}_1, \dots, \hat{\lambda}_n$ and the expected frequency of count y stated as

$$E_y = \sum_{i=1}^n f(y, \hat{\lambda}_i, \hat{\delta}, \hat{a}), \quad j = 0, 1, \dots$$

A rootogram can now be drawn with bars from $\sqrt{E_y} - \sqrt{n_y}$ up to $\sqrt{E_y}$, to visually assess the goodness of fit provided by the Touchard regression model. Discrepancies between the observed frequencies n_y and the E_y are visualized against a straight line; Figure 4 (right).

4.4 EXAMPLE: CRAB SATELLITES

As an illustration, we fit the Touchard regression models along with other count regression models to the well-known data on the number of satellites (male crabs gathered around the female attempting to fertilize her eggs). For predicting the number of satellites y , we consider `weight` (kg) and `color` (with two categories, light and dark (baseline)). Therefore, the linear predictor is formulated as $\beta_0 + \beta_1 \text{weight} + \beta_2 \text{color}$.

Table 4. Estimation results (with SEs) from different models fitted by ML to the `CrabSat` dataset ($n = 173$, $n_0 = 62$, $k = 2$, $y_{\max} = 15$).

	log(μ) = linear predictor				log(λ) = linear predictor	
	QP	QPT	NB	GP	Tou	CMP
intercept	-0.49 (0.32)	-0.49 (0.29)	-0.93 (0.40)	-1.11 (0.48)	1.34 (0.22)	-0.58 (0.08)
weight	0.54 (0.12)	0.54 (0.11)	0.71 (0.16)	0.63 (0.20)	0.21 (0.05)	0.13 (0.04)
color	0.27 (0.18)	0.27 (0.18)	0.29 (0.20)	0.37 (0.06)	0.09 (0.06)	0.10 (0.06)
dispersion	$\hat{\varphi} = 3.15$	$\hat{\delta} = -3.99$ $\hat{a} = 0.83$	$\hat{\alpha} = 0.96$	$\hat{\alpha} = 0.34$	$\hat{\delta} = -2.05$ $\hat{a} = 0.25$	$\hat{\nu} = 0.08$ (*)
AIC	738*	725 [†]	754	719	758	747
SSR _{raw}	1,522	1,522	1,692	1,882	1,480	3,773
SSR _{pear}	536	175	155	166	185	11,258
SSR _{dev}	553	214	197	NA	215	602
\hat{n}_0	40	56	50	45	63	42
AME _{wei}	1.6	1.6	2.1	2.4	1.7	0.4
IRR _{wei}	1.7	1.7	2.0	1.9	NA	NA

Approximate QP(μ, φ) log-likelihood (Nelder and Pregibon, 1987).

([†]) Touchard log-likelihood evaluated at QPT estimates.

NA = not available.

Estimation results are shown in Table 4. The estimated coefficients from the Touchard GLM-type regression are very close to those from QP. The value of a was chosen to yield the highest likelihood. The two Touchard models yield similar results with significant improvement over the GP and CMP as seen by different sums of residuals, log-likelihood and estimated number of zeros. The Touchard models yield closer estimates of the proportion of zeros than the other models considered. The Touchard regression models provide the estimates (56 and 63) closest to $n_0 = 62$ among the models considered. The sum of deviance residuals for the NB model is the lowest despite the higher likelihood achieved by

the Touchard models. The sums of residuals estimated by the CMP is much higher than those predicted by the other models which probably explains the huge values for the raw and Pearson residuals. Rootograms are shown in Figure 4 indicating some misfitting for the Touchard but noticeable improvement over the fit provided by the NB regression model.

All models agree qualitatively in terms of lack of marginal significance of **color** and a considerable effect of **weight**. However, different models give different effect sizes for **weight**. Marginal effects in count regression are often reported either as the average marginal effect (AME) or as the relative change in the conditional mean (incidence rate ratio - IRR). For instance, the AME associated with the predictor **weight** is given by

$$AME_{\text{wei}} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \mu_i}{\partial x_{i1}},$$

where $x_1 \equiv \text{weight}$. The AME_{wei} for models based on $\log(\mu_i)$ is simply $(1/n) \sum_i \hat{\mu}_i \hat{\beta}_1$. Therefore the AMEs for the GLM-type models are 1.59 (QP and QPT), 2.12 (NB) and 2.42 (GP). For both the direct Touchard model and the CMP model, which are based on $\log(\lambda_i)$, we have $\text{Var}(Y) = \partial \mu / \partial \log(\lambda)$. Thus, $AME_{\text{wei}} = (1/n) \sum_i \hat{\sigma}_i^2 \hat{\alpha}_1$ and which amounts to 1.73 (Touchard) and 0.36 (CMP). Taking a unit change in **weight** as the variation of interest, the marginal effect in terms of IRR is stated as

$$IRR_{\text{wei}} = \frac{\mathbb{E}(y | \text{weight} + 1, \text{color})}{\mathbb{E}(y | \text{weight}, \text{color})}.$$

For the GLM-type models, we have $IRR_{\text{wei}} = \exp(\beta_1)$ with values of 1.72 (QP and QPT), 1.65 (Tou), 2.03 (NB) and 1.88 (GP). For the direct models based on $\log(\lambda)$ the IRR does not reduce to a simple expression, varying across observations and values of **color**. Sellers and Shmueli (2010) suggest dividing the CMP coefficients by $\hat{\nu}$ as a crude approximation for comparison with Poisson coefficients. For the coefficient of **weight** this yields $0.13/0.08 = 1.63$ which is much higher than the Poisson estimate of 0.54.

Estimated marginal effects are thus higher under the NB and GP models and strikingly lower under the CMP model.

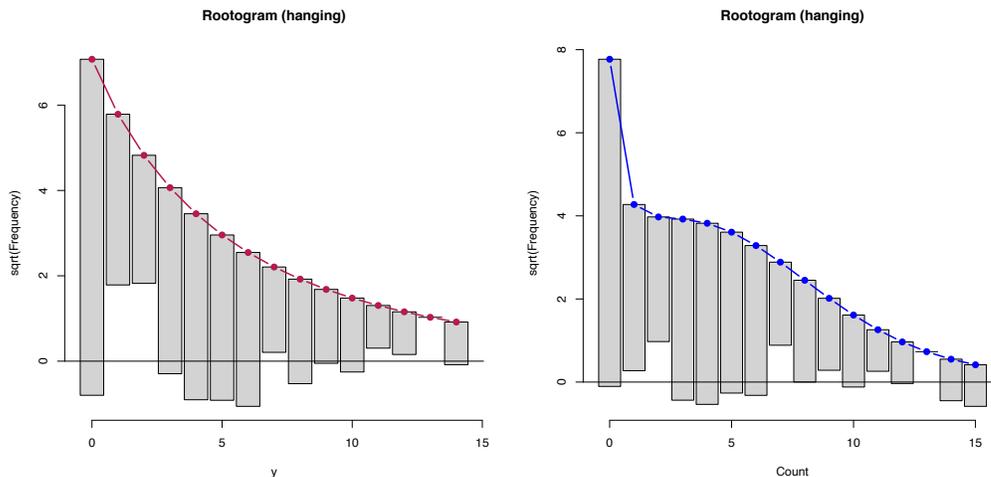


Figure 4. Rootgram associated with NB regression model (left) and Touchard GLM-type model (right) with **CrabSat** dataset.

5. CONCLUDING REMARKS

We have provided several tools for analyzing count data with a flexible weighted Poisson distribution (Touchard) including regression modeling. We have concluded that the Touchard is a viable and flexible alternative to model over or underdispersed count data. Data analyses presented here and in Matsushita et al. (2018) show that the Touchard is a competitive alternative to traditional models within the exponential family. The statistical tools developed here are based on classical methods including maximum likelihood, method of moments and quasi-likelihood. In terms of tractability, elegance and numerical implementation, the Touchard model is more flexible than the negative binomial and the Conway-Maxwell-Poisson, besides many other count models for which similar tools are not yet available. A major advantage of using the weighted Poisson with $(\lambda, \delta; \log(y+a))$ model is that one does not need to switch between highly different models for different datasets.

Regression modeling has been presented in terms of two variants and in the form of quasi-Poisson estimation with the Touchard variance. Direct modeling of $\log(\lambda)$ in terms of a linear predictor has been developed. When mean prediction and interpretability are wanted, regression based on $\log(\mu)$ is available in the form of a GLM-type model. The QPT methodology overcomes the computational burden associated with the GLM-type model and is a viable choice for large datasets and when directly modeling is either too slow computationally or the related maximization is unstable. Future work may investigate quantile-based regression and modeling of δ in terms of covariates. Bayesian modeling based on the Touchard model is also open for research.

APPENDIX A. FURTHER RESULTS ON THE NORMALIZING CONSTANT

Before providing new results, we state, for mere completion, the following result proved by del Castillo and Pérez-Casany (1998).

THEOREM A.1 The series $\tau(\lambda, \delta, a) = \sum_{i=0}^y [\lambda^i (y+a)^\delta] / i!$ converges for $\lambda, a > 0$ and $\delta \in \mathbb{R}$.

We now provide asymptotic expressions for τ and for the first two moments of $Y \sim \text{Tou}(\lambda, \delta, a)$.

THEOREM A.2 To first-order, the following approximations hold:

$$\tau \approx \exp(\lambda)(\lambda + a)^\delta,$$

$$\mu \approx \lambda + \frac{\lambda\delta}{\lambda + a}, \tag{A1}$$

and

$$\sigma^2 \approx \lambda \left(1 + \frac{a\delta}{(\lambda + a)^2} \right). \tag{A2}$$

Proof The first approximation may be obtained by replacing $(y+a)^\delta$ by $(\lambda+a)^\delta + (\lambda+a)^{\delta-1}(y-\lambda) + o(|(y-\lambda)|)$ into the series defining $\tau(\lambda, \delta, a)$. Alternatively, it is easy to see

that $\tau = \exp(\lambda)\mathbb{E}(Y^* + a)^\delta$, where $Y^* \sim \text{Poi}(\lambda)$. Therefore, by the Delta method, we have

$$\begin{aligned} \tau &\approx \exp(\lambda)[\mathbb{E}(Y^*) + a]^\delta \\ &= \exp(\lambda)(\lambda + a)^\delta, \end{aligned}$$

Using the above approximation for τ and the fact that $\mu = (\lambda/\tau)(\partial\tau/\partial\lambda)$ (see Appendix C) we obtain the approximation stated in Equation (A1) for μ . Since $\sigma^2 = \lambda(\partial\mu/\partial\lambda)$, we obtain approximation stated in Equation (A2) for σ^2 . ■

We observe that: (i) the above approximations are exact for $\delta = 0$ and $\delta = 1$; (ii) they are better the larger λ is relative to $|\delta|$ and the larger a is; (iii) for large λ , $\mu \rightarrow \lambda + \delta$ and $\sigma^2 \rightarrow \mu - \delta$; see Figure 2.

Having a compact notation for τ in terms of other special functions is of interest to study further properties and analytical characteristics. This has been explored, for example, by Castellares and Lemonte (2019), where a previous diverging series was re-derived in terms of the integro-exponential function to provide a correct converging result for the moments of the generalized Gompertz distribution. The following theorem provides a representation of τ in terms of the generalized hypergeometric function.

THEOREM A.3 The function $\tau(\lambda, \delta, a)$ can be represented in terms of the I generalized hypergeometric function as

$$\tau(\lambda, \delta, a) = I_{1,2}^{1,1} \left[-\lambda \left| \begin{matrix} (-a, 1, \delta) \\ (0, 1, 1), (1 - a, 1, \delta) \end{matrix} \right. \right], \tag{A3}$$

in which the I -function (Rathie, 1997) is defined as an contour complex integral which contain powers of Gamma functions in their integrands by

$$\begin{aligned} I_{p,q}^{m,n} \left[z \left| \begin{matrix} (a_1, \alpha_1, A_1), \dots, (a_n, \alpha_n, A_n), (a_{n+1}, \alpha_{n+1}, A_{n+1}), \dots, (a_p, \alpha_p, A_p) \\ (b_1, \beta_1, B_1), \dots, (b_m, \beta_m, B_m), (b_{m+1}, \beta_{m+1}, B_{m+1}), \dots, (b_q, \beta_q, B_q) \end{matrix} \right. \right] \\ = \frac{1}{2\pi i} \int_L \frac{\prod_{j=1}^m \Gamma^{B_j}(b_j - \beta_j s) \prod_{j=1}^n \Gamma^{A_j}(1 - a_j + \alpha_j s)}{\prod_{j=m+1}^q \Gamma^{B_j}(1 - b_j + \beta_j s) \prod_{j=n+1}^p \Gamma^{A_j}(a_j - \alpha_j s)} z^s ds, \end{aligned} \tag{A4}$$

in which α_j, A_j, β_j and B_j are assumed to be positive quantities and all the a_j and b_j are complex such that no singularity of $\Gamma^{B_j}(b_j - \beta_j s)$ coincides with any singularities of $\Gamma^{A_j}(1 - a_j + \alpha_j s)$. In general, these singularities are not poles.

There are three different contours L of integration stated as:

- L goes from from $\sigma - i\infty$ to $\sigma + i\infty$ (σ real) such that all the sigularities of $\Gamma^{B_j}(b_j - \beta_j s)$, $j = 1, \dots, m$ lie to the right of L and all the singularities of $\Gamma^{A_j}(1 - a_j + \alpha_j s)$, for $j = 1, \dots, n$, lie to the left of L .
- L is a loop beginning and ending at $+\infty$ and encircling all the singularities of $\Gamma^{B_j}(b_j - \beta_j s)$, for $j = 1, \dots, m$, once in the clock-wise direction, but none of the singularities of $\Gamma^{A_j}(1 - a_j + \alpha_j s)$, $j = 1, \dots, n$.
- L is a loop beginning and ending at $-\infty$ and encircling all the singularities of $\Gamma^{A_j}(1 - a_j + \alpha_j s)$, for $j = 1, \dots, n$, once in the anti-clockwise direction, but none of the singularities of $\Gamma^{B_j}(b_j - \beta_j s)$, for $j = 1, \dots, m$.

Proof Let one consider the I -function on the right-hand side of Equation (A3) and its

contour integral representation given by Equation (A4) as

$$I_{1,2}^{1,1} \left[-\lambda \left| \begin{matrix} (-a, 1, \delta) \\ (0, 1, 1), (1-a, 1, \delta) \end{matrix} \right. \right] = \frac{1}{2\pi i} \int_L \frac{\Gamma(-s)\Gamma^\delta(1+a+s)}{\Gamma^\delta(a+s)} (-\lambda)^s ds$$

Since none of the singularities of $\Gamma^\delta(1+a+s)$ coincide with the poles of $\Gamma(-s)$, the simple application of the residue theorem (Springer, 1979) to the poles of the latter imply that

$$\begin{aligned} I_{1,2}^{1,1} \left[-\lambda \left| \begin{matrix} (-a, 1, \delta) \\ (0, 1, 1), (1-a, 1, \delta) \end{matrix} \right. \right] &= \sum_{r=0}^{\infty} \lim_{s \rightarrow r} \frac{(-s+r)\Gamma(-s)\Gamma^\delta(1+a+s)}{\Gamma^\delta(a+s)} (-\lambda)^s \\ &= \sum_{r=0}^{\infty} \frac{\Gamma^\delta(1+a+r)(-\lambda)^r}{\Gamma^\delta(a+r)r!(-1)^r} \\ &= \sum_{r=0}^{\infty} \frac{(a+r)^\delta \lambda^r}{r!}, \end{aligned}$$

as desired. ■

APPENDIX B. IDENTIFIABILITY

By definition (Lehmann and Casella, 1998, Sec. 1.5), for a given statistical model $\mathcal{P} = \{P_\zeta: \zeta \in \mathcal{Z}\}$, where \mathcal{Z} denotes the parameter space, we say \mathcal{P} is identifiable if

$$P_{\zeta_1} = P_{\zeta_2} \implies \zeta_1 = \zeta_2 \quad \forall \zeta_1, \zeta_2 \in \mathcal{Z}.$$

With a fixed, the two-parameter $\text{Tou}(\lambda, \delta, a)$ model is clearly identifiable since the statistics Y and $\log(Y+a)$ are linearly independent (Lehmann and Casella, 1998, Sec. 1.5). A more general result with all parameters free is developed next. We begin by defining $\zeta = (\lambda, \delta, a)$ and $f_\zeta(y)$ in place of Equation (1). Thus, we define

$$\mathcal{P} = \left\{ f_\zeta(y) = \frac{\lambda^y (y+a)^\delta}{y! \tau(\lambda, \delta, a)} : \lambda, a > 0, \delta \in \mathbb{R} \right\}.$$

In order to prove identifiability, we must set $f_{\zeta_1}(y) = f_{\zeta_2}(y), \forall y$. However, we can avoid $\tau(\lambda, \delta, a)$ by working instead with

$$\frac{f_{\zeta_1}(y+1)}{f_{\zeta_1}(y)} = \frac{f_{\zeta_2}(y+1)}{f_{\zeta_2}(y)},$$

which reduces to

$$\frac{\lambda_1}{(y+1)} \left(\frac{y+1+a_1}{y+a_1} \right)^{\delta_1} = \frac{\lambda_2}{(y+1)} \left(\frac{y+1+a_2}{y+a_2} \right)^{\delta_2}.$$

In order to compare both sides of Equation (B), let us consider the series representation of the function $h(\delta, a, x) = ((x+a)/(x+1+a))^\delta$, which is can be seen as the product of the functions $h_a(\delta, a, x) = (x+a)^\delta$ and $h_b(\delta, a, x) = (x+1+a)^{-\delta}$. Thus, the MacLaurin Series

of $h(\delta, a, x)$ in terms of x is obtained as

$$h(\delta, a, x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} \frac{\partial^n}{\partial x^n} (h_a(\delta, a, x)h_b(\delta, a, x)) \Big|_{x=0}. \tag{B1}$$

The generalized Leibiniz rule states that

$$\frac{\partial^n}{\partial x^n} h_a(\delta, a, x)h_b(\delta, a, x) = \sum_{y=0}^n \frac{n!}{(n-y)!k!} \frac{\partial^{n-y}}{\partial x^{n-y}} h_a(\delta, a, x) \frac{\partial^y}{\partial x^k} h_b(\delta, a, x). \tag{B2}$$

The derivatives on the right hand side of Equation (B2) are quite simple and result in

$$\frac{\partial^{n-y}}{\partial x^{n-y}} h_a(\delta, a, x) = \frac{\Gamma(\delta + 1)}{\Gamma(\delta - n + y + 1)} (x + a)^{\delta-n+y} \tag{B3}$$

and

$$\frac{\partial^y}{\partial x^y} h_b(\delta, a, x) = (-1)^y \frac{\Gamma(\delta + y)}{\Gamma(\delta)} (x + 1 + a)^{-\delta-y}. \tag{B4}$$

By combining Equations (B2), (B3) and (B4), after some algebra we arrive at

$$\begin{aligned} \frac{\partial^n}{\partial x^n} h_a(\delta, a, x)h_b(\delta, a, x) = \\ \frac{\Gamma(\delta + 1)(x + a)^{\delta-n}(x + 1 + a)^{-\delta}}{\Gamma(\delta - n + 1)} \sum_{y=0}^n \binom{n}{y} \frac{(-1)^k \Gamma(\delta + y)\Gamma(\delta - n + 1) \left(\frac{x+a}{x+a+1}\right)^y}{\Gamma(\delta)\Gamma(\delta - n + y + 1)}. \end{aligned} \tag{B5}$$

By definition, the right hand side of Equation (B5) may be expressed in terms of the hypergeometric function ${}_2F_1$ and by means of Equation (B1), we obtain

$$h(\delta, a, x) = \sum_{n=0}^{\infty} \frac{\Gamma(\delta + 1)a^{\delta-n}(1 + a)^{-\delta}}{n!\Gamma(\delta - n + 1)} {}_2F_1\left(-n, \delta; 1 - n + \delta; \frac{a}{a + 1}\right) x^n. \tag{B6}$$

By applying the ratio test to the series, we get

$$R = \frac{x(a)^{-1}(\delta - n)}{(n + 1)} \frac{{}_2F_1\left(-n - 1, \delta; -n + \delta; \frac{a}{1+a}\right)}{{}_2F_1\left(-n, \delta; 1 - n + \delta; \frac{a}{1+a}\right)}. \tag{B7}$$

In general, the hypergeometric function in Equation (B7) vanishes for finite n and, in the limiting case, the continued fraction representation for the ratio of hypergeometric functions given by

$$\frac{{}_2F_1(a + 1, b; c + 1; z)}{{}_2F_1(a, b; c; z)} = \frac{1}{1 + \frac{\frac{(a-c)b}{c(c+1)}z}{1 + \frac{(b-c-1)(a+1)}{(c+1)(c+2)}z}} \tag{B8}$$

1 + ⋯

can be used. By noticing that whenever $(a - c - j)(b + j)$, $\forall j \in \{0, 1, 2, \dots\}$ is in the numerator of the ratio which multiplies z in Equation (B8), this ratio vanishes to 0 as $n \rightarrow \infty$, we get

$$R \xrightarrow{n \rightarrow \infty} \frac{-x}{(a)}.$$

Thus, Equation (B6) is valid for $x < a$. To account for other values of x , we can define

$$g(\delta, a, \tilde{x}) = \left(\frac{1 + a\tilde{x}}{1 + (1 + a)\tilde{x}} \right)^\delta,$$

where $\tilde{x} = 1/x$ and back to Equation (B) by following a similar procedure as above, the series representation for $g(\delta, a, \tilde{x})$ can be obtained as

$$g(\delta, a, x^{-1}) = \sum_{n=0}^{\infty} \frac{\Gamma(\delta + 1)(1 + a)^n}{n! \Gamma(\delta - n + 1)} {}_2F_1\left(-n, \delta; 1 - n + \delta; \frac{a}{a + 1}\right) x^{-n}. \tag{B9}$$

Equation (B9) is valid for $x > a + 1$ and it remains to be addressed the case $a < x < 1 + a$. For real values of a , since x is a positive integer in our case, the only possible value for x which falls into this interval is $x = \lceil a \rceil$. This special case is treated in the last paragraph. Now, without loss of generality, let $a_1 \leq a_2$. From Equation (B), we conclude that the identifiability problem reduces to the expression given by

$$\begin{cases} \lambda_1 h(\delta_2, a_2, y) = \lambda_2 h(\delta_1, a_1, y), & \text{for } 0 \leq y < a_1, \\ \lambda_1 h(\delta_2, a_2, y) = B_1, & \text{for } a_1 \leq y < \min(a_1 + 1, a_2), \\ B_2 = B_3, & \text{for } \min(a_1 + 1, a_2) < y < \max(a_1 + 1, a_2), \\ B_4 = \lambda_1 g(\delta_1, a_1, 1/y), & \text{for } \max(a_1 + 1, a_2) < y < a_2 + 1, \\ \lambda_1 g(\delta_1, a_1, 1/y) = \lambda_2 g(\delta_2, a_2, 1/y), & \text{for } y > a_2 + 1, \end{cases} \tag{B10}$$

where the functions B_i , $i = 1, 2, 3, 4$, depend on the max and min functions applications. For instance, if $\min(1 + a_1, a_2) = 1 + a_1$ then $B_1 = \lambda_2 ((\lceil a_1 \rceil + a_1) / (1 + (\lceil a_1 \rceil + a_1)))^\delta$. Regarding the first expression in Equation (B10), by performing a term-by-term matching procedure, the first three terms of the series in Equation (B6) indicate that

$$\begin{aligned} \text{(i)} \quad & \lambda_1 a_2^{\delta_2} (1 + a_2)^{-\delta_2} = \lambda_2 a_1^{\delta_1} (1 + a_1)^{-\delta_1}, \\ \text{(ii)} \quad & \lambda_1 \delta_2 a_2^{\delta_2 - 1} (1 + a_2)^{-\delta_2 - 1} = \lambda_2 \delta_1 a_1^{\delta_1 - 1} (1 + a_1)^{-\delta_1 - 1}, \\ \text{(iii)} \quad & \frac{\lambda_1}{2} \delta_2 a_2^{\delta_2 - 2} (1 + a_2)^{-\delta_2 - 2} (-1 + \delta_2 - 2a_2) = \\ & \frac{\lambda_2}{2} \delta_1 a_1^{\delta_1 - 2} (1 + a_1)^{-\delta_1 - 2} (-1 + \delta_1 - 2a_1). \end{aligned} \tag{B11}$$

The quotients $(ii)/(i)$ and $(iii)/(ii)$ imply

$$\begin{aligned} \text{(iv)} \quad & \frac{\delta_1}{a_1(1+a_1)} = \frac{\delta_2}{a_2(1+a_2)}, \\ \text{(v)} \quad & \frac{-1+\delta_1-2a_1}{2a_1(1+a_1)} = \frac{-1+\delta_2-2a_2}{2a_2(1+a_2)}. \end{aligned} \tag{B12}$$

By solving the system in Equation (B12), the two possible solutions for a_1 are $a_1 = a_2$ or $a_1 = -(1 + a_2)(1 + 2a_2)^{-1}$. Obviously, since $a_i \geq 0$, for $i = 1, 2$, the only possible solution is $a_1 = a_2$.

By using such solution back on (iv) of Equation (B12), $\delta_1 = \delta_2$. Thus, (i) of Equation (B11) implies that $\lambda_1 = \lambda_2$.

Now, for the last expression of Equation (B10), the term-by-term series comparison provides

- (vi) $\lambda_1 = \lambda_2,$
- (vii) $\lambda_1 \delta_1 = \lambda_2 \delta_2,$
- (viii) $\frac{\lambda_1}{2} \delta_1 (-1 + \delta_1 - 2a_1) = \frac{\lambda_2}{2} \delta_2 (-1 + \delta_2 - 2a_2).$

The system solution is quite straightforward, implying that: $\lambda_1 = \lambda_2, \delta_1 = \delta_2$ and $a_1 = a_2$. The remaining cases to be discussed are the other equations which have not been addressed yet in Equation (B10). Let S denote the support of the probability density function of the Touchard distribution. So far, it has been shown that the function is identifiable over $S \setminus \{[a_1, a_2 + 1]\}$. In order to prove the identifiability over the whole support, it is sufficient to check if the conditions found for $S \setminus \{[a_1, a_2 + 1]\}$ also work when $y \in [a_1, a_2 + 1]$. This easily follows by noticing that the identifiability of a probability density function boils down to a system of equations. Therefore, all the equations must be simultaneously satisfied in order to exist a solution. Thus, it has been shown that for $[f_{\zeta_1}(y) = f_{\zeta_2}(y)], \forall y \implies \zeta_1 = \zeta_2,$ which proves identifiability when all three parameters are free.

APPENDIX C. USEFUL DERIVATIVES

The following expressions were used to obtain several formulas involving derivatives presented in the article. Recall that we have defined $Z = \log(Y + a), W_1 = (Y + a)^{-1}, \mu = \mathbb{E}(Y), \kappa = \mathbb{E}(Z), \sigma^2 = \text{Var}(Y), \nu = \text{Var}(Z)$ and $\gamma = \text{Cov}(Y, Z)$. We also define $\kappa_2 = \mathbb{E}[Z^2]$ and $\kappa_3 = \mathbb{E}[YZ]$. We start with first and second derivatives of $\tau(\lambda, \delta)$ given by

$$\frac{\partial \tau}{\partial \lambda} = \frac{\tau \mu}{\lambda}; \quad \frac{\partial \tau}{\partial \delta} = \tau \kappa; \quad \frac{\partial \tau}{\partial a} = \tau(\lambda, \delta - 1, a); \quad \frac{\partial^2 \tau}{\partial \lambda^2} = \frac{\tau[m_2 - \mu]}{\lambda^2}; \quad \frac{\partial^2 \tau}{\partial \delta^2} = \tau \kappa_2 \quad \text{and} \quad \frac{\partial^2 \tau}{\partial \lambda \partial \delta} = \frac{\tau \kappa_3}{\lambda}.$$

Next, we list some partial derivatives of the expectations $\mu(\lambda, \delta)$ and $\kappa(\lambda, \delta)$ stated as

$$\frac{\partial \mu}{\partial \lambda} = \frac{\sigma^2}{\lambda}; \quad \frac{\partial \mu}{\partial \delta} = \gamma; \quad \frac{\partial \mu}{\partial a} = \delta \text{Cov}(Y, W_1), \quad \frac{\partial \kappa}{\partial \lambda} = \frac{\gamma}{\lambda}, \quad \frac{\partial \kappa}{\partial \delta} = \nu.$$

APPENDIX D. RANDOM NUMBER GENERATION (RNG) BY THE INVERSE TRANSFORMATION

In order to generate y from $Y \sim \text{Tou}(\lambda, \delta, a)$, we take $F(y) = \sum_{x=0}^y f(x)$ and $Y = \min\{y: F(y - 1) < U \leq F(y), U \sim \text{Unif}(0, 1)\}$. We note that the RNG can avoid costly computation of cumulative probabilities and factorial terms by using the ratio $c(y) = f(y)/f(y - 1)$.

Pseudo-code for the proposed RNG is shown in Algorithm D and it is implemented for the R system (Andrade and Oliveira, 2019).

The expected number of iterations in the while loop in Algorithm D is $1 + \mathbb{E}(Y)$ which, according to results from previous section, is approximately $1 + \lambda[1 + \delta/(\lambda + a)]$ or simply $1 + \lambda + \delta$ for $\lambda \gg \delta$. The generation time increases with λ and δ with δ being a stronger factor than λ (Table D1). The value of a is of least importance for running time. Generation of highly overdispersed data is faster. Efficient generators use multiple schemes taking into consideration the parameter values (Fishman, 2013) and this is still open for research in the context of Touchard RNG.

Pseudo-code for Touchard RNG.

Input: $p_0 = \tau(\lambda, \delta, a)^{-1}$.

Output: $Y \sim \text{Tou}(\lambda, \delta, a)$.

Define: $p_k := f(k; \lambda, \delta, a)$ and $c(k) := \frac{p_k}{p_{k-1}}$, $k = 1, 2, \dots$

initialization

$p \leftarrow p_0; q \leftarrow p_0; k \leftarrow 0$

$U \sim \text{Unif}(0, 1)$

While: $U > q$

$k \leftarrow k + 1$

$p \leftarrow p \cdot c(k)$

$q \leftarrow q + p$

Return: $Y = k$

Table D1. Time to generate 10^4 values using Algorithm D (in plain C) for different combinations of parameter values with $a = 1$. Times reported in 1/1000 seconds (Intel Pentium dual core i5 1200MHz running Linux Debian).

		δ					
		-10	-5	-1	1	5	10
λ	0.5	2	1	2	2	10	15
	2	2	1	2	3	19	27
	10	1	9	8	10	54	66
	20	1	55	15	18	92	109

AUTHOR CONTRIBUTIONS Conceptualization, investigation, methodology, software, writing-original draft preparation, writing review and editing, writing-original: all the authors participated equally. The authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGEMENTS The author thanks Editors and Referees for this suggestions which allowed us to improve the presentation of this work.

FUNDING Not applicable

CONFLICTS OF INTEREST The authors declare no conflict of interest.

REFERENCES

- Agresti, A., 2013. Categorical Data Analysis. Wiley, New York, US.
- Agresti, A., 2015. Foundations of Linear and Generalized Linear Models. Wiley, New York, US.
- Andrade, B. and Oliveira, S., 2019. touchard: Touchard Model and Regression. R package version 2.0.0.
- Atkinson, A., 1985. Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis. Clarendon Press, Oxford, UK.

- Cameron, A.C. and Trivedi, P.K., 1998. *Regression Analysis of Count Data*. Cambridge University Press, Cambridge, MA, US.
- Castellares, F. and Lemonte, A.J., 2019. On the moments of the generalized gompertz distribution. *Applied Mathematical Modelling*, 72, 420 – 424.
- Chakraborty, S. and Imoto, T., 2016. Extended Conway-Maxwell-Poisson distribution and its properties and applications. *Journal of Statistical Distributions and Applications*, 3, 1–19.
- Cook, R.D. and Weisberg, S., 1982. *Residuals and Influence in Regression*. Chapman and Hall, New York, US.
- del Castillo, J. and Pérez-Casany, M., 1998. Weighted poisson distributions for overdispersion and underdispersion situations. *Annals of the Institute of Statistical Mathematics*, 50, 567–585.
- del Castillo, J. and Pérez-Casany, M., 2005. Overdispersed and underdispersed poisson generalizations. *Journal of Statistical Planning and Inference*, 134, 486–500.
- Denuit, M., 1997. A new distribution of poisson-type for the number of claims. *ASTIN Bulletin*, 27, 229–242.
- Fisher, R.A., 1934. The effect of methods of ascertainment upon the estimation of frequencies. *Annals of Eugenics*, 6, 13–25.
- Fishman, G., 2013. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer, New York, US.
- Graybill, F.A., 1983. *Matrices with applications in statistics*. Wadsworth International Group, London, Uk.
- Hilbe, J.M., 2014. *Modeling Count Data*. Cambridge University Press, Cambridge, MA, US.
- Ho, L.L., Andrade, B., Bourguignon, M., and Fernandes, F.H., 2021. Monitoring count data with Shewhart control charts based on the Touchard model. *Quality and Reliability Engineering International*, 37, 1875–1893.
- Hoaglin, D.C. and Tukey, J.W., 2009. Checking the Shape of Discrete Distributions, Chapter 9, 345–416. Wiley, New York, US.
- Kleiber, C. and Zeileis, A., 2016. Visualizing count data regressions using rootograms. *The American Statistician*, 70, 296–303.
- Kokonendji, C.C., Mizère, D., and Balakrishnan, N., 2008. Connections of the Poisson weight function to overdispersion and underdispersion. *Journal of Statistical Planning and Inference*, 138, 1287–1296.
- Kokonendji, C.C., Simplicio, D.G., and Demétrio, C.G.B., 2004. Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes. *SORT*, 28, 201–214.
- Lehmann, E.L. and Casella, G., 1998. *Theory of Point Estimation*. Springer, New York, US.
- Matsushita, R., Pianto, D., Andrade, B., and Cançado, A., 2018. The touchard distribution and process. *Communications in Statistics: Theory and Methods*.
- Nelder, J.A. and Pregibon, D., 1987. An extended quasi-likelihood function. *Biometrika*, 74, 221–232.
- Patil, G.P. and Rao, C.R., 1978. Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, 34, 179–189.
- Rao, C.R., 1985. Weighted distributions arising out of methods of ascertainment: What population does a sample represent? In Atkinson, A.C. and Fienberg, S.E. (eds.) *A Celebration of Statistics*, pp. 543–569, Springer, New York, US.
- Rathie, A.K., 1997. A new generalization of generalized hypergeometric functions. *Le Mathematique*, 12, 297–310.
- Sellers, K.F. and Shmueli, G., 2010. A flexible regression model for count data. *The Annals of Applied Statistics*, 4, 943–961.

- Springer, M.D., 1979. *The Algebra of Random Variables*. Wiley, New York, US.
- Wei, B.C., Hu, Y.Q., and Fung, W.K., 1998. Generalized leverage and its applications. *Scandinavian Journal of Statistics*, 25, 25–37.

INFORMATION FOR AUTHORS

The editorial board of the Chilean Journal of Statistics (ChJS) is seeking papers, which will be refereed. We encourage the authors to submit a PDF electronic version of the manuscript in a free format to the Editors-in-Chief of the ChJS (E-mail: chilean.journal.of.statistics@gmail.com). Submitted manuscripts must be written in English and contain the name and affiliation of each author followed by a leading abstract and keywords. The authors must include a "cover letter" presenting their manuscript and mentioning: "We confirm that this manuscript has been read and approved by all named authors. In addition, we declare that the manuscript is original and it is not being published or submitted for publication elsewhere".

PREPARATION OF ACCEPTED MANUSCRIPTS

Manuscripts accepted in the ChJS must be prepared in Latex using the ChJS format. The Latex template and ChJS class files for preparation of accepted manuscripts are available at <http://soche.cl/chjs/files/ChJS.zip>. Such as its submitted version, manuscripts accepted in the ChJS must be written in English and contain the name and affiliation of each author, followed by a leading abstract and keywords, but now mathematics subject classification (primary and secondary) are required. AMS classification is available at <http://www.ams.org/mathscinet/msc/>. Sections must be numbered 1, 2, etc., where Section 1 is the introduction part. References must be collected at the end of the manuscript in alphabetical order as in the following examples:

Arellano-Valle, R., 1994. Elliptical Distributions: Properties, Inference and Applications in Regression Models. Unpublished Ph.D. Thesis. Department of Statistics, University of São Paulo, Brazil.

Cook, R.D., 1997. Local influence. In Kotz, S., Read, C.B., and Banks, D.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 1., Wiley, New York, pp. 380-385.

Rukhin, A.L., 2009. Identities for negative moments of quadratic forms in normal variables. Statistics and Probability Letters, 79, 1004-1007.

Stein, M.L., 1999. Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

Tsay, R.S., Peña, D., and Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika, 87, 789-804.

References in the text must be given by the author's name and year of publication, e.g., Gelfand and Smith (1990). In the case of more than two authors, the citation must be written as Tsay et al. (2000).

COPYRIGHT

Authors who publish their articles in the ChJS automatically transfer their copyright to the Chilean Statistical Society. This enables full copyright protection and wide dissemination of the articles and the journal in any format. The ChJS grants permission to use figures, tables and brief extracts from its collection of articles in scientific and educational works, in which case the source that provides these issues (Chilean Journal of Statistics) must be clearly acknowledged.