ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Volume 19, No.3
July 2021

REVSTAT

Statistical Journal

Catalogação Recomendada

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

CREDITS

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Web site: http://www.ine.pt
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *140 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*
- **PRICE [VAT included]**
  - *€ 9,00*

EDITORIAL

#IYWSDS — International Year of Women in Statistics and Data Science

In this REVSTAT issue, we will be continuing the occasional publication of invited papers with the aim of diversifying the research topics of the its manuscripts, challenging internationally recognized researchers in Statistical Science.

We began this REVSTAT editorial policy in number 5 of volume 18 with the invited paper with discussion 'Statistics in Times of Pandemics: the Role of Statistical and Epidemiological Methods During the COVID-19 Emergency' by **Baltazar Nunes** and collaborators (2020). That issue was also a special issue in Celebration of WORLD STATISTICS DAY 20 October 2020. That was a year in which COVID-19 pandemic has become a current topic.

The second REVSTAT invited paper was published in number 1 of volume 19 entitled 'Skewed Probit Regression – Identifiability, Contraction and Reformulation' by Janet van Niekerk and **Harvard Rue** (2021). In addition to investigating possible skewness parameters and the penalizing complexity priors of these, they highlighted that methodology is available in the R-INLA package.

In this REVSTAT issue, the invited paper is 'Calibration of the bulk and extremes of spatial data' by **Maria Antónia Amaral-Turkman** and collaborators (2021). They concentrated on calibrating the bulk and the extremes of data, simultaneously, avoiding methods that rely on the choice of a threshold. As Amaral-Turkman is a pioneer in the development of Bayesian Statistics, this publication also aims to celebrate the International Year of Women in Statistics and Data Science (#IYWSDS) from May 2020 through July 2021. The IYWSDS was launched on 12 May 2020 with commemorating the *200th anniversary of Florence Nightingale's birth.*

July 13, 2021

Isabel Fraga Alves
Giovani Loiola da Silva

INDEX

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

REVSTAT – Statistical Journal
Volume 19, Number 3, July 2021, 309–325

CALIBRATION OF THE BULK AND EXTREMES OF SPATIAL DATA

(*Invited Paper*)

Authors: Maria Antónia Amaral Turkman *
– Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Portugal
maturkman@fc.ul.pt

Kamil Feridun Turkman
– Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Portugal
kfturkman@fc.ul.pt

Patrícia de Zea Bermudez
– Departamento de Estatística e Investigação Operacional and Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Portugal
pcbermudez@fc.ul.pt

Soraia Pereira
– Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Portugal
sapereira@fc.ul.pt

Paula Pereira
– Escola Superior de Tecnologia de Setúbal, Instituto Politécnico Setúbal and Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Portugal
sapereira@fc.ul.pt

Miguel de Carvalho
– School of Mathematics, University of Edinburgh, United Kingdom and Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Portugal
Miguel.deCarvalho@ed.ac.uk

Received: October 2020 Revised: April 2021 Accepted: May 2021

Abstract:

- In an environmental framework, extreme values of certain spatio-temporal processes, for example wind speeds, are the main cause of severe damage in property, such as electrical networks, transport and agricultural infrastructures. Typically, as is the case of wind speeds, data are available at few stations with many missing observations and consequently simulators are often used to augment information. However, simulated data often mismatch observed data, particularly at tails, therefore calibrating and bringing it in line with observed data may offer practitioners more reliable and richer data sources. In this work we will concentrate on calibrating the bulk and the extremes of data, simultaneously, avoiding methods that rely on the choice of a threshold. We propose and describe in detail a specific conditional quantile matching calibration method and exemplify it with wind speed data. We also briefly suggest how calibration should be extended specifically to data coming from the tails of simulated and observed data, using asymptotic models and methods suggested by extreme value theory.

*Corresponding author.

Keywords:

- *calibration; quantile matching approach; extended generalized Pareto distribution; spatial extremes; Bayesian hierarchical models.*

AMS Subject Classification:

- 62F15, 62H11, 62M30, 62P12.

1. INTRODUCTION

Extreme values of certain spatio-temporal processes, such as wind speeds, are the main cause of severe damage in property, from electricity distribution grid to transport and agricultural infrastructures. Accurate assessment of causal relationships between environmental processes and their effects on risk indicators, are highly important in risk analysis, which in return depends on sound inferential methods as well as on good quality informative data. Often, information on the relevant environmental processes comes from monitoring networks, as well as from numerical-physical models (simulators) that typically solve a large set of partial differential equations, capturing the essence of the physical process under study (see, for example, Skamarock *et al.* 2008 [18], Cardoso *et al.* 2013 [6]). In general, monitoring networks are formed by a sparse set of stations, whose instrumentation are vulnerable to disruptions, resulting in data sets with many missing observations. On the other hand, simulated data from numerical simulators typically supply average yield of the process in grid cells of pre-specified dimensions, often at high resolutions, spanning large spatial domains, with no missing observations. However, simulated data typically mismatch and misaligned observed data, therefore calibrating it and bringing it in line with observed data may supply modellers with more reliable and richer sources of data. Data assimilation methods, namely combining data from multiple sources, are well known in environmental studies, with data often being used to generate initial boundary conditions for the numerical simulators (Kalnay 2003 [11]). There is a very rich statistical literature on data assimilation and data fusion with the objective of enriching the information for inference (Fuentes and Raftery 2005 [9], Berrocal *et al.* 2012 [4], Zidek *et al.* 2012 [23], Berrocal *et al.* 2014 [5], McMillan *et al.* 2010 [12]). These statistical methods are often based on Bayesian hierarchical models for space-time data (see Banerjee *et al.* 2004 [1]) and are constructed around the idea of relating the monitoring station data and the simulated data using spatial linear models with spatially varying coefficients (see Berrocal, 2019 [3]). Since these relations involve data measured at different spatial resolutions, the models are often called downscaler models (see Berrocal *et al.* 2012 [4]). The principal objective of these downscaler models is to relate observations measured at different space resolutions using spatial linear models.

The motivation behind this present work has its roots in a consulting work done for a major electricity producer and distributor. The electricity grid constantly faces disruptions due to damages in the distribution system, with heavy economic losses. These damages and consequent disruptions occur due to a combination of many factors such as topography and precipitation, however extreme winds and storms are the main cause of such damages. Risk maps that indicate likely places of costly disruptions in electric grids are important decision support tools for administering the power grid and are particularly useful in deciding if costly corrective actions should be taken to improve structures. It is natural that these risk maps should be based primarily on observed wind speeds among other factors and it was decided that daily maximum wind speeds should be used as proxy information. Hence, such risk maps can be interpreted as vulnerability maps of electricity grid to extreme wind speeds, expressed in terms of probability. However generating such maps depends on reliable wind data at fairly high spatial and temporal resolutions.

The data set available for this particular study corresponds to simulated wind speeds from a simulator (The WRF model, version 3.1.1) at a regular grid of 81ksq grid cell size,

obtained at 10 minutes interval from 2006–2013; however only daily maximum wind speeds will be used. Observed daily maximum wind speeds are also available during the same period of time, from 117 stations in Portugal mainland, but missing observations reach to 90% in some stations. Only around one third of the stations have less than 30% missing observations. There is an additional challenge: although simulated and observed data match in the bulk of the distribution, they quite often mismatch at extreme values. Therefore, adequate methods of data fusion and calibration can be used to combine these two different sources of data, providing information which is more reliable from a spatial point of view and producing more accurate probability maps showing the spatial distribution of damage risks. Since electricity grid damages are ultimately caused by extreme wind speeds, the aim should be to develop statistical methods for data fusion and calibration that can extrapolate beyond the range of observed data into the tails of a distribution in line with extreme value theory.

We propose and describe, in detail, a specific conditional quantile matching calibration method for the bulk and the extreme observations of the data, based on models proposed by Naveau *et al.*, 2016 [14]. The outline of the paper is as follows: In Section 2, we report a new approach for calibration through a conditional quantile matching calibration method (Pereira *et al.*, 2019 [16]), using an extended generalized Pareto distribution (Papastathopoulos and Tawn 2013 [15], Naveau *et al.* 2016 [14]), adequate for calibrating simultaneously the bulk and the tails of the distribution. In Section 3, we built a Bayesian hierarchical model to implement this calibration strategy for spatio-temporal data. In Section 4, this method will be exemplified using a wind speed data. Finally, further discussion and conclusions are given in Section 5.

2. CALIBRATION METHODS FOR BULK AND TAILS

We denote by $Y(s,t)$ and $X(s,t)$, respectively the observed and simulated wind speeds at location $s \in \mathbb{R}^2$ and time t. To simplify notation, often we will use Y and X for observed and simulated wind speeds when data are used without any space-time reference. Typically X are simulated over a regular grid, say B, often represented by points s_B which correspond to the centroid of the grid cells, whereas Y are observed in stations located at different spatial points s.

For the time being, if we ignore totally space-time variations and dependence structures, calibration can be seen as a simple scaling making use of marginal distributions fitted corresponding to X and Y (CDF transform method, Michelangeli *et al.* 2009 [13]), as we explain below.

Suppose we have a set of n observed y_i and simulated x_i, $i = 1, ..., n$ data. Let F_Y and F_X be, respectively, the distribution functions of Y and X. Then the new calibrated (scaled) data x_i^* is defined as

$$x_i^* = F_Y^{-1}(F_X(x_i)), \quad i = 1, ..., n. \tag{2.1}$$

Since

$$P(X^* \leq z) = P(F_Y^{-1}(F_X(X) \leq z) = P(U \leq F_Y(z)) = F_Y(z), \qquad \text{with} \quad U \sim U(0,1),$$

calibrated data have the same distribution as the observed data. Note that if $F_X = F_Y$ then $x_i^* = x_i$. Figure 1 depicts the result of applying this calibration method when Y follows a Student distribution with 3 degrees of freedom, and X follows a standard normal distribution.

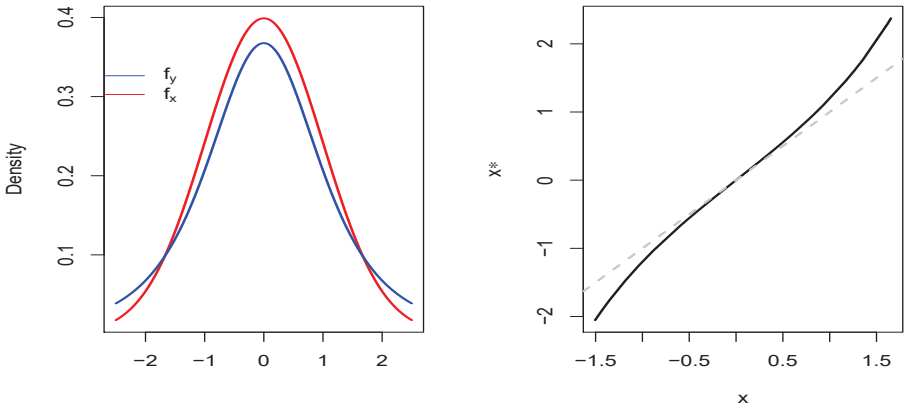


Figure 1: Illustration of the quantile matching approach.

This calibration method depends on the marginal distributions of the random variables involved Y and X and hence it does not take into consideration the expected strong dependence between the two sets of data. Thus, an ideal calibration should involve the joint distribution of Y and X defined in some way. One possibility is the use of a conditional quantile matching approach, which will be described in Section 2.2. Further, in the same section, we also introduce an extension to cover space-time non-homogeneity by scaling (calibrating) the data from

$$x^*(s,t) = F^{-1}_{Y(s,t)}(F_{X(s,t)}(x(s,t)), \tag{2.2}$$

assuming marginal distributions of $Y(s,t)$ and $X(s,t)$ for every s and t. This calibration method will take into consideration the strong space-time dependence structures expected in the data and consequently these distributions will be estimated by fitting them to data and considering the parameters as smooth functions of spatially and temporarily varying covariates and space-time latent processes as in Section 4. Notice that, in this case, $x^*(s,t)$ as defined in (2.2) will depend on unknown parameters and hence calibrated data have to be estimated.

Pereira *et al.* (2019) [16] develop a covariate-adjusted version of the quantile matching-based approach as in (2.1) where the distributions of simulated and real data change along a covariate. At the same time they suggest a regression method that simultaneously models the bulk and the (right) tail of the distributions involved, using the extended generalized Pareto distribution (EGPD) (Naveau *et al.*, 2016 [14]) as a model for both the simulated and observed data. In their approach, Pereira *et al.* (2019) [16] do not take into consideration any strong spatio-temporal variations and dependence structures that may exist both in the simulated as well as in the observed data. In what follows, we propose an extension of this conditional quantile matching calibration for the bulk and tails, taking into consideration spatio-temporal variations and dependence structures, thus extending their results significantly.

de Carvalho *et al.* 2020 [7] also work on covariate adjusted version of the extended generalized Pareto distribution (EGPD) (Naveau *et al.*, 2016 [14]) for the conditional bulk and conditional tail of a possibly heavy-tailed response. However, their objective, contrary to ours, is not calibration, but to learn the effect of covariates on an extreme value setting via a Lasso-type specification.

Under fairly general conditions, according to the asymptotic theory of extremes, the generalized Pareto distribution (GPD) appears as a natural model for the right tail of a distribution, by focusing on the excesses over a high but fixed threshold. Here, the choice of this threshold plays a very important role in inference, ignoring the part of the data that lie below this threshold. See, for example, Beirland *et al.* (2004) [2]. The EGPD modelling strategy suggested by Naveau *et al* (2016) [14] avoids this selection problem, as we will see in next section.

2.1. Naveau *et al.* (2016) EGPD models

Naveau *et al.* (2016) [14] suggest an extension of generalized Pareto model tailored for both the bulk and tails, and — contrarily to most methods for extremes — does not require a threshold to be selected. The objective of this extension is to generate a new class of distributions with GPD type tails consistent with extreme value theory, but also flexible enough to model efficiently the main bulk of the observed data without complicated threshold selection procedures.

Let Y be a positive random variable with cumulative distribution function defined as:

$$F_Y(y \mid \theta) = G\Big(H(y \mid \xi, \sigma)\Big),$$

where H is the cumulative distribution function of a generalized Pareto distribution (GPD) and G is a function obeying some general assumptions, so that a Pareto-type tail is ensured and the bulk is driven by the carrier G. (see Naveau *et al.* 2016 [14] and de Carvalho *et al.* 2020 [7]), that is

$$H(y \mid \xi, \sigma) = \begin{cases} 1 - \left(1 + \dfrac{\xi}{\sigma} y\right)_+^{-1/\xi}, & \xi \neq 0. \\ 1 - \exp\left(-\dfrac{y}{\sigma}\right), & \xi = 0\ . \end{cases}$$

with $a_+ = \max(a, 0)$, $\sigma > 0$, and $y > 0$ if $\xi \geq 0$ and $y < -\frac{\sigma}{\xi}$ if $\xi < 0$. The parameter σ is a dispersion parameter while ξ is a shape parameter controlling the rate of decay of the right tail of a distribution (e.g. de Zea Bermudez and Kotz [8]).

Naveau *et al.* [14] consider four forms of $G(u)$ resulting in four different classes of distributions. Although the theory below can be easily extended to any of the forms of the G function, in what follows we use one of the forms, namely, $G(u) = u^\kappa$, the canonical form of the EGPD (de Carvalho *et al.* 2020 [7]), where κ is a parameter controlling the shape of the lower tail. It is clear that smaller the κ more the distribution is concentrated near zero. The EGPD will have then three parameters, and we will refer to it as a $EGPD(\kappa, \xi, \sigma)$.

2.2. Spatio-temporal conditional quantile matching calibration for the bulk and tails

Let us assume that both random variables X and Y are space-time dependent and we want to calibrate X based on Y. The calibrated data are given as in (2.2). Now assume further that both random variables are distributed as a canonical $EGPD(\kappa, \xi, \sigma)$ where the parameters are indexed by the correspondent random variables. In order to better accommodate for the situation $\xi < 0$ we make a transformation $\delta = -\frac{\sigma}{\xi}$. Hence, for $\xi_x \neq 0$

$$F_{X(s,t)}(x(s,t) \mid \delta_x(s,t), \xi_x, \kappa_x) = \left(1 - \left(1 - \frac{1}{\delta_x(s,t)} x(s,t)\right)_+^{-1/\xi_x}\right)^{\kappa_x}, \tag{2.3}$$

for $x > 0$ if $\xi_x > 0$ and $x < \delta_x$ if $\xi_x < 0$.

Assuming as well $\xi_y \neq 0$

$$F_{Y(s,t)}(y(s,t) \mid \delta_y(s,t), \xi_y, \kappa_y) = \left(1 - \left(1 - \frac{1}{\delta_y(s,t)} y(s,t)\right)_+^{-1/\xi_y}\right)^{\kappa_y}, \tag{2.4}$$

for $y > 0$ if $\xi_y > 0$ and $y < \delta_y$ if $\xi_y < 0$.

Although it is assumed that these random variables are conditional independent, a dependence structure is introduced through the transformed space-time dependent parameters δ_x, δ_y by modelling them as a function of a common latent spatio-temporal process, in a Bayesian hierarchical modelling framework. Here, we are mainly interested in modelling the right tail as function of space and time. de Carvalho *et al.* 2020 [7], consider a Bayesian hierachical modelling of the EGPD for the case $\xi > 0$, where both bulk and tail are covariate adjusted.

As an exemplification of this modelling strategy, in the next section, we will built a Bayesian hierarchical model for the wind speed data.

3. BAYESIAN HIERARCHICAL MODEL FOR WIND SPEED DATA

A preliminary data analysis of the wind speed data used in this study, shows that observed and simulated data are consistent with the case $\xi < 0$ and hence, the distributions for X and Y will have an end-point characterized by the respective parameter δ.

Let T be length of the time period under study, N the number of stations with complete observed data during that period and N_s the total number of stations ($N < N_s$).

The observed data $\mathbf{Y}(\mathbf{s},\mathbf{t}) = \{Y(s_i, t_j),\, i = 1, ..., N;\, j = 1, ..., T\}$, are assumed to follow a distribution as in (2.4), with parameters κ_y, ξ_y and $\delta_y(i,j)$, such that $\delta_y(i,j) \sim \text{Exp}(\lambda_y(i,j))$, $\delta_y(i,j) > \max(y)$, i.e., follows a shifted exponential distribution, with

$$\log(\lambda_y(i,j)) = \beta_y + W(s_i) + Z(t_j), \tag{3.1}$$

where $W \sim MVN(0, \tau_W \Sigma_W)$ follows a Multivariate Gaussian process, defined on the space, as in Thomas *et al.* (2014) [21], with τ_W a precision parameter and the matrix Σ_W with diagonal elements equal to 1 and off-diagonal elements, $\Sigma_{i\ell} = f(d_{i\ell}; \alpha)$, where

$$f(d_{i\ell}; \alpha) = 2/\pi * \left\{ \cos^{-1}(d_{i\ell}/\alpha) - [(d_{i\ell}/\alpha)(1 - (d_{i\ell}^2/\alpha^2))]^{1/2} \right\}, \quad d_{i\ell} < \alpha,$$

is a function of $d_{i\ell}$, the centroids' distance of every two stations s_i and s_ℓ, and α a parameter representing the radius of the 'disc' centred at each s. The parameter α controls the rate of decline of correlation with distance.

For the temporal random process $\{Z(t_j)\}$ we assume a random walk process of order 1, $Z \sim MVN(0, \tau_Z \Sigma_Z)$, where τ_Z is a precision parameter and Σ_Z is a matrix with a structure reflecting the fact that any two increments $z_i - z_{i-1}$ are independent (Rue and Held, 2005 [17]).

For the simulated data $\mathbf{X}(\mathbf{s}, \mathbf{t}) = \{X(s_i, t_j), i = 1, ..., N_s; j = 1, ..., T\}$ we assume a distribution as in (2.3) with N_s total number of stations, where the model for $\delta_x(i, j) \sim \text{Exp}(\lambda_x(i, j))$, $\delta_x(i, j) > \max(x)$, shares the same latent processes W and Z as the model for the observed data (3.1), i.e.,

$$\log(\lambda_x(ij)) = \beta_x + W(s_i) + Z(t_j).$$

Let θ be a vector containing all model parameters including the latent Gaussian models W and Z. Assuming conditional independence, the likelihood $L(\theta \mid \mathbf{y}(\mathbf{s}, \mathbf{t}), \mathbf{x}(\mathbf{s}, \mathbf{t}))$ is a product of individual terms $L_{ij}(\theta \mid y(s_i, t_j))$ and $L_{\ell j}(\theta \mid x(s_\ell, t_j))$.

To complete the Bayesian hierarchical model we consider the following prior specification for the parameters and hyperparameters of the models not yet specified:

$$\begin{aligned} \beta_y, \beta_x &\text{ i.i.d. } N(0, 100), \\ \kappa_y, \kappa_x &\text{ i.i.d. } Ga(0.05, 0.05), \\ \xi_y, \xi_x &\text{ i.i.d. } U(-0.5, 0), \\ \tau_W, \tau_Z &\text{ i.i.d. } Ga(1, 0.1), \\ \alpha &\sim U(0.1, 0.9). \end{aligned} \tag{3.2}$$

Further, all these parameters are assumed to be *a priori* independent and hence the prior distribution $h(\theta)$ is the product of the individual priors. In Subsection 4.1 an explanation is given for the choice of these priors.

The posterior distribution

$$h(\theta \mid \mathbf{y}(\mathbf{s}, \mathbf{t}), \mathbf{x}(\mathbf{s}, \mathbf{t})) \propto L(\theta \mid \mathbf{y}(\mathbf{s}, \mathbf{t}), \mathbf{x}(\mathbf{s}, \mathbf{t}))\, h(\theta),$$

is analytically intractable and hence one has to resort to the use of computational methods, such as MCMC methods.

Now, let $F_X(x(s_i, t_j) \mid \kappa_x, \xi_x, \delta_x(i, j)) = p_{ij}$, for $i = 1, ..., N_s$ and $j = 1, ..., T$. Then, since, for any $p \in (0, 1)$ the inverse function of F_Y is

$$\delta_y \left[1 - \left(1 - p^{1/\kappa_y} \right)^{-\xi_y} \right],$$

the calibrated values defined in (2.2), as a function of the model parameters, are given by

$$x^*(s_i, t_j \mid \theta_{ij}) = \delta_y(s_i, t_j)\left[1 - \left(1 - p_{ij}^{1/\kappa_y}\right)^{-\xi_y}\right], \tag{3.3}$$

where $\theta_{ij} = (\kappa_x, \xi_x, \delta_x(i,j), \kappa_y, \xi_y, \delta_y(i,j))$.

Hence, following Bayesian methodology, calibrated data are estimated as the mean of the calibrated function, defined in (3.3), with respect to the posterior distribution $h(\theta \mid \mathbf{x}, \mathbf{y})$, that is

$$x^*(s_i, t_j) = \int \delta_y(s_i, t_j)\left[1 - \left(1 - p_{ij}^{1/\kappa_y}\right)^{-\xi_y}\right] h(\theta \mid \mathbf{x}, \mathbf{y}) d\theta.$$

In what follows we will call this posterior mean as the *calibrated data*.

Computation of calibrated data is achieved through MCMC methods, by simulating a sample of size M from the posterior distribution and approximating the integral as

$$x^*(s_i, t_j) \approx \frac{1}{M}\sum_{k=1}^{M} \delta_y(s_i, t_j)^{(k)}\left[1 - \left(1 - (p_{ij}^{(k)})^{1/\kappa_y^{(k)}}\right)^{-\xi_y^{(k)}}\right],$$

where $\delta_y(s_i, t_j)^{(k)}, -p_{ij}^{(k)}, \kappa_y^{(k)}, \xi_y^{(k)}$ are simulated values at the kth iterate after convergence is achieved. A sample of size M from the posterior distribution of the calibrated function given by

$$\delta_y(s_i, t_j)^{(k)}\left[1 - \left(1 - (p_{ij}^{(k)})^{1/\kappa_y^{(k)}}\right)^{-\xi_y^{(k)}}\right], \quad k = 1, ..., M$$

will allow the computation of any relevant summary statistics, such as, $\gamma\%$ credible intervals.

4. APPLICATION TO WIND SPEED DATA

We used observed and simulated wind speed data from 01/01/2013 to 28/02/2013, so $T = 59$. There are $N = 51$ stations where we have both observed and simulated daily maximum wind speeds. Additionally we have extra 66 stations with simulated values for the maximum daily wind speeds, so that $N_s = 117$. In Figure 2 we depict the median of observed and simulated wind speeds for the 51 stations together with the 2.5% and 97.5% empirical quantiles (referred to on the figure as the 95% IQR).

The model was implemented in `R2OpenBUGS` (Sturtz *et al.* 2005 [19]). In Table 1 we show the summary statistics for the marginal posterior distributions of the parameters of the model, based on a sample of size 40000, after a burn-in period of 20000 iterates. Convergence was assessed for the parameters of the model specified in Section 4.1.

We observe that the posterior mean of κ_y has a much smaller value than the posterior mean of κ_x which is consistent with the fact that, in general, simulated data are shifted to the right in relation to the observed data, indicating the possible existence of some bias in the simulated data. The posterior mean of the precision (inverse of the variance) parameters for the space model (τ_W) and for the temporal model (τ_Z) suggest that time dependence is stronger than space dependence. The posterior mean for β_y is slightly smaller than the posterior mean for β_x.

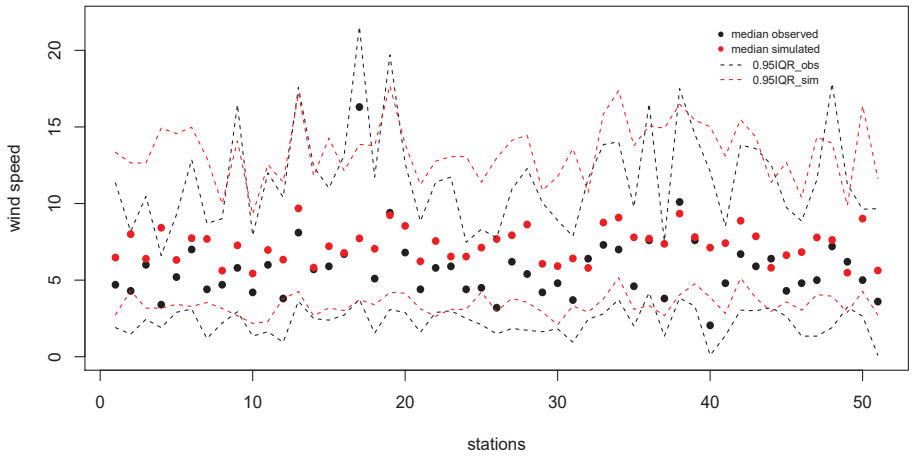


Figure 2: Median of observed and simulated wind speeds for the 51 stations, and the 95% IQR wind speeds by station (dashed lines).

Table 1: Summary statistics for the marginal posterior distributions.

Parameter	Mean	Standard deviation	2.5% quantile	Median	97.5% quantile
α	0.6691	0.1007	−0.4668	−0.5028	−0.8564
β_y	−1.1552	0.1702	−1.4760	−1.4250	−0.8063
β_x	−0.9274	0.1614	−1.2300	−1.1830	−0.5957
κ_y	−5.2951	0.1976	−4.9020	−4.9740	−5.7020
κ_x	18.7384	0.7467	17.2900	17.5100	20.3000
τ_W	−3.5699	0.7274	−2.3280	−2.4840	−5.1520
τ_Z	−0.4240	0.1081	−0.2457	−0.2678	−0.6675
ξ_y	−0.0703	0.0018	−0.0739	−0.0675	−0.0670
ξ_x	−0.0806	0.0014	−0.0834	−0.0782	−0.0777

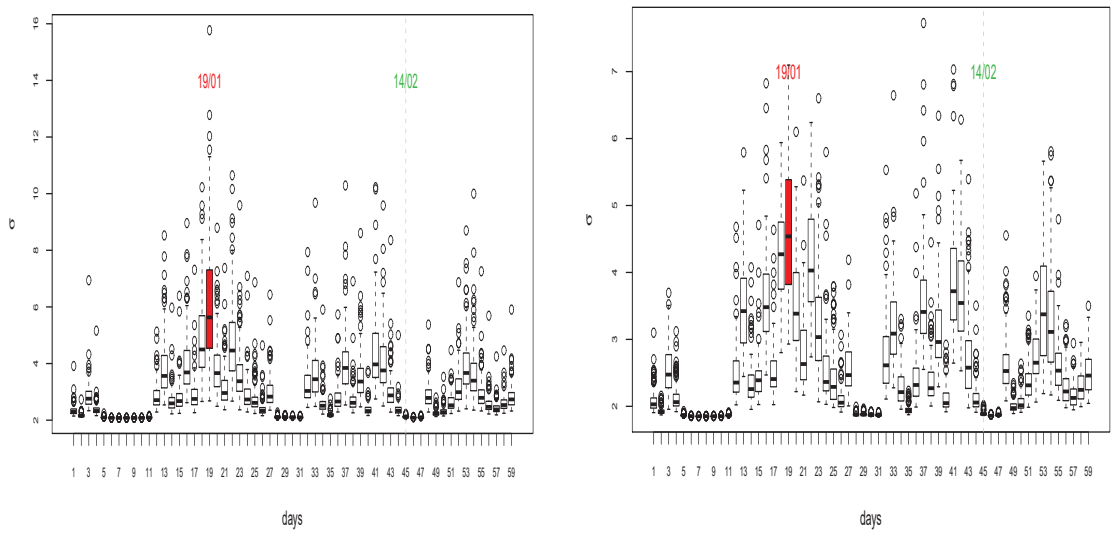


Figure 3: Boxplot of the posterior means of $\sigma_y(i,j)$ (left) and $\sigma_x(i,j)$ (right).

This naturally contributes for higher values for $\sigma_y(i,j)$ relatively to $\sigma_x(i,j)$ and with greater dispersion, as it can be seen in Figure 3 where we show daily boxplots of the posterior means of the parameters $\sigma(i,j), \forall j$ for both models. In that figure it is marked two dates, 19 of January, a day where it was observed a storm with heavy winds (storm GONG, maximum observed wind 29.6m/s), particularly in regions close to the littoral, and 14th of February, a very mild day all over the country (Valentine's day; maximum observed wind 8.20m/s). The variation observed along the days is consistent with the fact that on windy days the maximum wind speed along the stations varies much more than on mild days. Also the temporal dependence is clear in these pictures.

These two days were studied, in particular, for exemplification of the conditional quantile calibration method proposed. For the purpose of exemplification of the results we represent in Figures 4 and 5, on the left, a kernel density estimation (considering all the stations) for the observed and simulated maximum wind speed on that day, together with the mean of the posterior distribution of the calibrated data as defined in (3.3). On the right side, we represent the observed and simulated maximum wind speed on that day for each station, together with the calibrated data with the corresponding 95% credible interval.

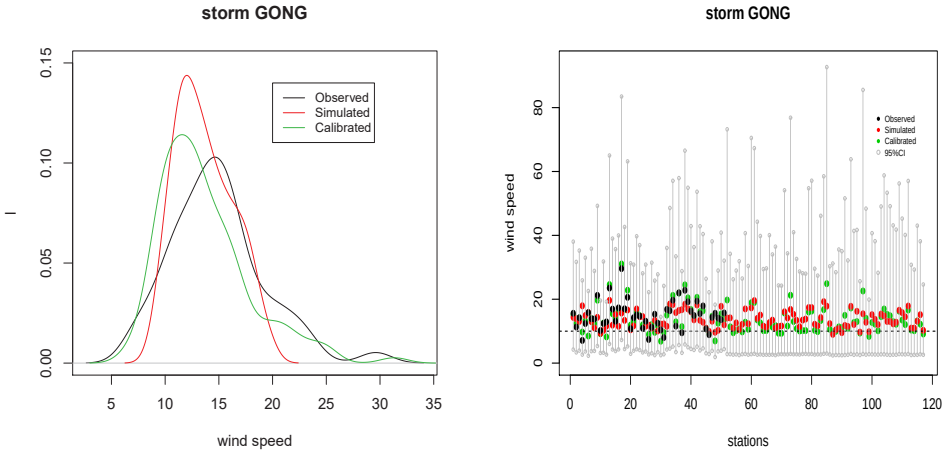


Figure 4: Kernel density estimation (left), observed and simulated maximum wind speed for each station, together with the mean of the posterior distribution for the calibrated data and 95% credible interval, for a storm day.

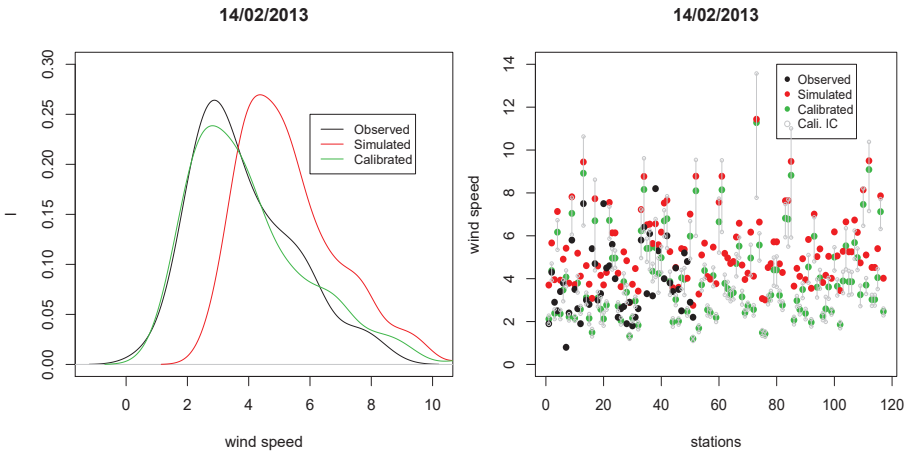


Figure 5: Kernel density estimation (left), observed and simulated maximum wind speed for each station, together with the mean of the posterior distribution for the calibrated data and 95% credible interval, for a mild day.

We observe that, on a storm day (Figure 4) the observed winds have longer tails than simulated winds. The calibration method was able to capture both tails of the distribution for the observed data, although it shifted the bulk of the distribution to the left. Notice that the 95% credible intervals are very large. This was to be expected due to the great variability of the simulated values from the posterior distribution of the shape parameter, as it can be observed in Figure 3. Regarding a mild windy day (Figure 5), the distribution of the simulated data is shifted to the right relatively to the distribution of the observed data with longer tails, as it was observed in a preliminary study. This bias is corrected with the calibration method. The 95% credible intervals are, in general, small.

In Figures 6 and 7 there is a spatial representation of the observed, simulated and calibrated values for each of these two days.

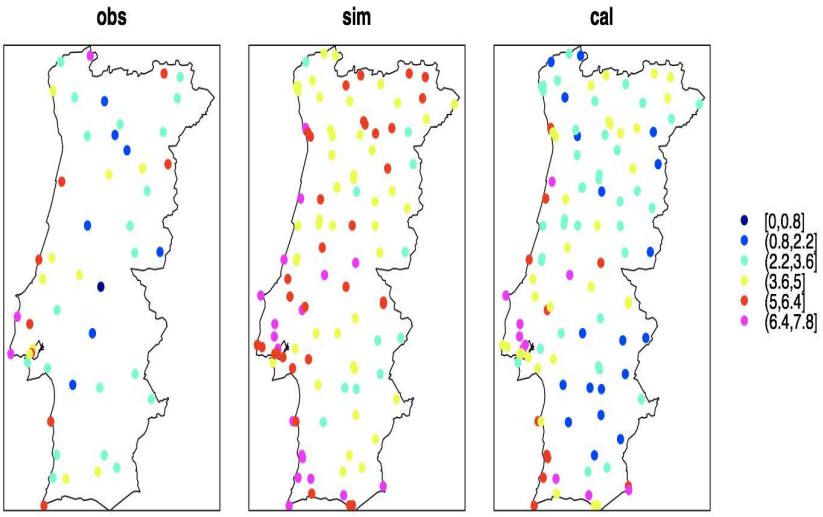


Figure 6: Storm day: observed, simulated and calibrated maximum wind speeds.

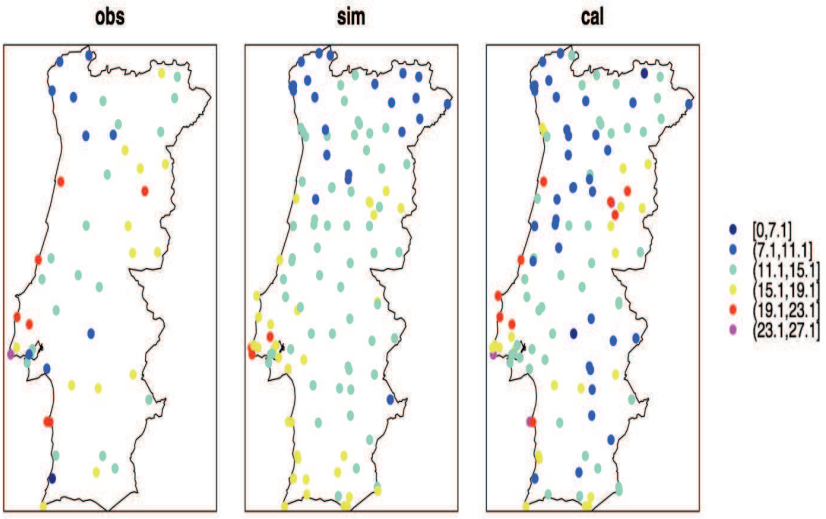


Figure 7: Mild day: observed, simulated and calibrated maximum wind speeds.

4.1. Choice of prior specification and a sensitivity study

Here a justification is given on the particular choice of priors used in this study. Apart from the parameters α and ξ, and in the absence of adequate prior information, we considered relatively vague priors for the other parameters involved, as described in (3.2).

A preliminary data analysis of the wind speed data showed that observed and simulated data were consistent with the case $\xi < 0$. Without further prior information on ξ, we considered a uniform prior distribution with support between -0.5 and 0, since it is known from the extreme value theory that estimators with good properties for ξ exist when $\xi > -0.5$.

In order to set a prior for the parameter α, that controls the rate of decline of correlation with distance, we follow the suggestion by Thomas *et al.*, 2014 [21]. Accordingly, in the absence of prior information, "a sensible 'default' choice is to consider an upper prior bound equal to a small multiple of the maximum distance in the study region" and the lower prior bound should be larger than the minimum distance between observations. Coordinates of the 117 stations were given in decimal degrees and transformed into km to compute distances.

Table 2: Summary statistics for the marginal posterior distributions.

Upper bound	Parameter	Mean	Standard deviation	2.5% quantile	Median	97.5% quantile
0.5	α	0.4552	0.0403	0.3522	0.3733	0.4988
0.9	α	0.6691	0.1007	0.4668	0.5028	0.8564
1.2	α	0.6744	0.1075	0.4677	0.5020	0.8821
0.5	β_y	−1.1510	0.1505	−1.4390	−1.3970	−0.8641
0.9	β_y	−1.1552	0.1702	−1.4760	−1.4250	−0.8063
1.2	β_y	−1.1356	0.1514	−1.4250	−1.3820	−0.8353
0.5	β_x	−0.9170	0.1381	−1.1820	−1.1400	−0.6549
0.9	β_x	−0.9274	0.1614	−1.2300	−1.1830	−0.5957
1.2	β_x	−0.9014	0.1412	−1.1590	−1.1240	−0.6171
0.5	κ_y	5.3136	0.1926	4.9500	5.0080	5.6970
0.9	κ_y	5.2951	0.1976	4.9020	4.9740	5.7020
1.2	κ_y	5.3056	0.1890	4.9390	5.0050	5.7110
0.5	κ_x	18.7734	0.8037	17.2000	17.4800	20.3100
0.9	κ_x	18.7384	0.7467	17.2900	17.5100	20.3000
1.2	κ_x	18.7919	0.7302	17.4500	17.6200	20.2500
0.5	τ_W	4.1859	0.7078	2.9510	3.1180	5.7120
0.9	τ_W	3.5699	0.7274	2.3280	2.4840	5.1520
1.2	τ_W	3.5600	0.7378	2.2620	2.4460	5.1570
0.5	τ_Z	0.4244	0.1034	0.2519	0.2729	0.6535
0.9	τ_Z	0.4240	0.1081	0.2457	0.2678	0.6675
1.2	τ_Z	0.4101	0.0981	0.2454	0.2665	0.6256
0.5	ξ_y	−0.0702	0.0018	−0.0738	−0.0673	−0.0668
0.9	ξ_y	−0.0703	0.0018	−0.0739	−0.0675	−0.0670
1.2	ξ_y	−0.0702	0.0017	−0.0736	−0.0674	−0.0670
0.5	ξ_x	−0.0805	0.0016	−0.0836	−0.0778	−0.0774
0.9	ξ_x	−0.0806	0.0014	−0.0834	−0.0782	−0.0777
1.2	ξ_x	−0.0804	0.0014	−0.0831	−0.0781	−0.0776

In order to avoid large numbers, we consider 100km as unit to compute distances. The maximum Euclidean distance among the 117 stations was computed as 5.65 and the minimum 0.01. Hence we set a uniform prior between 0.1 and 0.9. However we performed a sensitivity study considering as upper bound 0.5, 0.9 and 1.2, keeping the other priors unchanged. Summary results are displayed in Table 2 and in Figure 8.

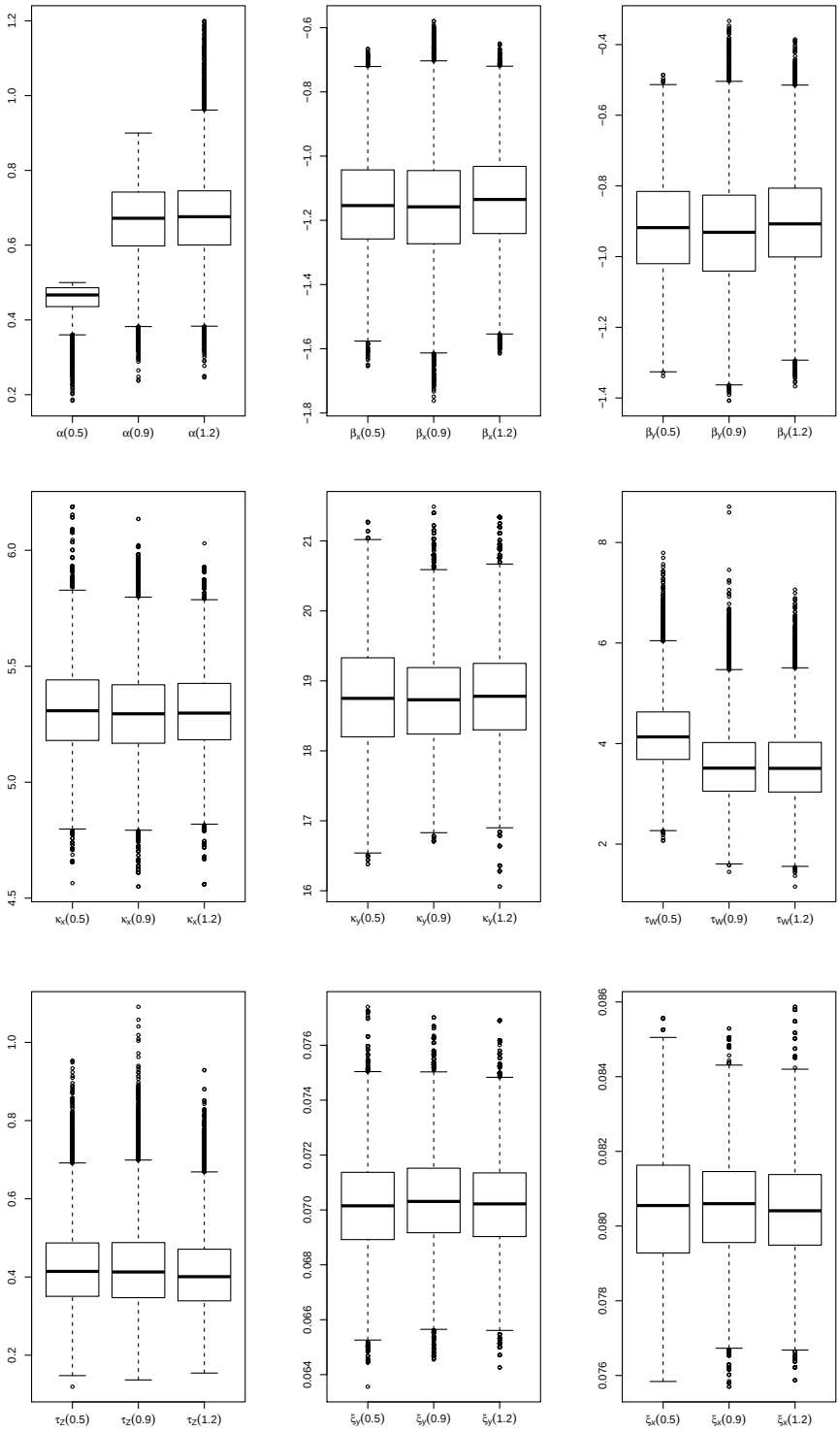


Figure 8: Boxplots of the posterior distribution of the parameters for different values for the upper limit of the prior for α.

As it can be seen, posterior distribution of α is sensitive to the its prior. The upper bound 0.5 seems not to be adequate since it is clear that there is a concentration of mass near the upper bound. However, when the upper bound is 0.9 or 1.2, the influence of the prior for α on its posterior is much less evident, particularly on the bulk of the posterior distribution. Influence of the prior on α on the other parameters is almost negligible. Basically there is only some influence on the posterior distribution of the parameter τ_W, although this influence is softened while comparing the prior with upper bound 0.9 and 1.2. More important it is that there is no influence of the prior for α regarding the calibrated data (posterior distribution of $F_Y^{-1}(F_X(x(s_i,t_j)))$ at $s_i, i = 1, ..., N_s$ and time $t_j, j = 1, ..., T$).

5. DISCUSSION AND FURTHER EXTENSIONS

In this article we proposed a hierarchical Bayesian approach to implement a conditional quantile matching calibration (CQCM) using a space-time extended generalized Pareto distribution for both the observed and simulated data.

The performance of the CQCM method was exemplified with two specific days, a storm day and a mild day. In both cases the calibrated data matched well the observed data on the tails, although on the storm day it did not capture well the bulk of the distribution. Also the 95% credible intervals were quite wide for the storm day, which may be an indication that appropriate methods to deal with extreme data should instead be considered to accommodate these rare situations.

Ideally this method should be extended to the grid level, since the simulator produces data at a fine grid level and this is much more interesting if the objective is the construction of a risk map. However this extension is not trivial and some assumptions regarding the model structure have to be assumed.

Damages in electricity grid are basically governed by extreme winds and primarily simulated and observed data coming from the right tail differ. Hence adequate calibration methods must be specifically adapted to extreme observations coming from the right tails and methods and models to be used in calibration should ideally be compatible with extreme value theory. A range of approaches for characterising the extremal behaviour of spatial process have been suggested and a brief comparison of these methods can be found in Tawn *et al.* (2018) [20]. Downscaling method described by Towe *et al.* (2017) [22] — based on the conditional extremes process — is more suitable, with adequate modifications, to calibrate extreme simulated data based on observed wind speeds. Work on this approach is under progress. Alternatively, calibration methods based on bivariate max stable processes (Genton *et al.*, 2015 [10]) can be devised, although this would require substantial computational complications.

ACKNOWLEDGMENTS

Research partially financed by national funds through FCT – Fundação para a Ciência e a Tecnologia, under the projects PTDC/MAT-STA/28649/2017 and UIDB/00006/2020. The authors also acknowledge the constructive suggestions from the anonymous reviewers.

REFERENCES

[1] Banerjee, S.; Carlin, B.P. and Gelfand, A.E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL: Chapman and Hall.

[2] Beirland, J.; Goegebeur, Y.; Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, J. Wiley, Chichester.

[3] Berrocal, V.J. (2019). *Data assimilation*. In "Handbook of Environmental and Ecological Statistics" (A.E. Gelfand; M. Fuentes; J.A. Hoeting and R.L. Smith, Eds.), Chapman and Hall/CRC, 133–151.

[4] Berrocal, V.J.; Gelfand, A.E. and Holland, D.M. (2012). Space-time data fusion under error in computer model output: An application to modeling air quality, *Biometrics*, **68**(3), 837–848.

[5] Berrocal, V.J.; Gelfand, A.E. and Holland, D.M. (2014). Assessing exceedance of ozone standards: A space-time downscaler for fourth highest ozone concentrations, *Environmetrics*, **25**, 279–291.

[6] Cardoso, R.M.; Soares, P.M.M.; Miranda, P.M.A. and Belo-Perira, M. (2013). WRF high resolution simulation of Iberian mean and extreme precipitation climate, *International Journal of Climatology*, **33**, 2591–2608.

[7] de Carvalho, M.; Pereira, S.; Pereira, P. and de Zea Bermudez, P. (2020). An extreme value Bayesian lasso for the conditional bulk and tail, https://arxiv.org/abs/2010.07164

[8] de Zea Bermudez, P. and Kotz, S. (2010). Parameter estimation of the generalized Pareto distribution, Part I, *Journal of Statistical Planning and Inference*, **140**, 1353–1373.

[9] Fuentes, M. and Raftery, A.E. (2005). Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models, *Biometrics*, **61**, 36–45.

[10] Genton, M.G.; Padoan, S.A. and Sang, H. (2015). Multivariate max-stable spatial processes, *Biometrika*, **102**, 215–230.

[11] Kalnay, E. (2003). *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge.

[12] McMillan, N.J.; Holland, D.M.; Morara, M. and Feng, J. (2010). Combining numerical model output and particulate data using Bayesian space-time modeling, *Environmetrics*, **21**, 48–65.

[13] Michelangeli, P.A.; Vrac, M. and Loukos, H. (2009). Probabilistic downscaling approaches: Application to wind cumulative distribution functions, *Geophysical Research Letters*, **36**, 1–6.

[14] Naveau, P.; Huser, R.; Ribereau, P. and Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection, *Water Resources Research*, **52**, 2753–2769.

[15] Papastathopoulos, I. and Tawn, J.A. (2013). Extended generalised Pareto models for tail estimation, *Journal of Statistical Planning and Inference*, **143**, 131–143.

[16] Pereira, S.; Pereira, P.; de Carvalho, M. and de Zea Bermudez, P. (2019). *Calibration of extreme values of simulated and real data*. In "Proceedings of International Workshop on Statistical Modelling, Volume I" (L. Meira-Machado and G. Soutinho, Eds.), INE, Guimarães, 147–150.

[17] Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, Monographs on Statistics and Applied Probability, vol. 104, Chapman & Hall, London.

[18] Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Duda, M.G.; Huang, X.Y.; Wang, W. and Powers, J.G. (2008). A description of the advanced research WRF version 3, *NCAR Technical Notes*, TN-475_STR, 1–125.

[19] Sturtz, S.; Ligges, U. and Gelman, A. (2005). R2WinBUGS: A Package for Running WinBUGS from R, *Journal of Statistical Software*, **12**, 1–16.

[20] Tawn, J.; Shooter, R.; Towe, R. and Lamb, R. (2018). Modelling spatial extreme events with environmental applications, *Spatial Statistics*, **28**, 39–58.

[21] Thomas, A.; Best, N.; Lunn, D.; Arnold, R. and Spiegelhalter, D. (2014). *GeoBUGS User Manual, Version 3.2.3*, MRC Biostatistics Unit, Cambridge.

[22] Towe, R.P.; Sherlock, E.F.; Tawn, J.A. and Jonathan, P. (2017). Statistical downscaling for future extreme wave heights in the North Sea, *Annals of Applied Statistics*, **11**, 2375–2403.

[23] Zidek, J.V.; Le, N.D. and Liu, Z. (2012). Combining data and simulated data for space-time fields: Application to ozone, *Environmental and Ecological Statistics*, **19**, 37–56.

REVSTAT – Statistical Journal
Volume 19, Number 3, July 2021, 327–362

ALTERNATIVE ESTIMATION OF THE COMMON MEAN OF TWO NORMAL POPULATIONS WITH ORDER RESTRICTED VARIANCES

Authors: Adarsha Kumar Jena
– Department of Mathematics, National Institute of Technology Rourkela, Rourkela-769008, India
jadarsha@gmail.com

Manas Ranjan Tripathy*
– Department of Mathematics, National Institute of Technology Rourkela, Rourkela-769008, India
manas@nitrkl.ac.in , manasmath@yahoo.co.in

Nabendu Pal
– Department of Mathematics, University of Louisiana at Lafayette, Lafayette, Louisiana, USA 70504
nabendu@louisiana.edu

Received: January 2019 Revised: March 2019 Accepted: April 2019

Abstract:

- The problem of estimating the common mean of two normal populations has been considered when it is known a priori that the variances are ordered. Under order restriction on the variances some new alternative estimators have been proposed including one that uses the maximum likelihood estimator (MLE) numerically. Further, it has been proved that each of these new estimators beats their unrestricted counterparts in terms of stochastic domination as well as Pitman measure of closeness criterion. Sufficient conditions for improving estimators in certain classes of equivariant estimators have been proved, and consequently improved estimators have been obtained under order restriction on the variances. A detailed simulation study has been done in order to evaluate the performances of all the proposed estimators using an existing estimator as a benchmark. From our simulation study, it has been established that the new alternative estimators improve significantly upon their unrestricted counterparts and compete well with an existing estimator under order restriction on the variances.

Keywords:

- *common mean; equivariant estimator; inadmissibility; maximum likelihood estimator; mean squared error; ordered variances; Pitman measure of closeness; percentage of relative risk; stochastic dominance.*

AMS Subject Classification:

- 62F10, 62F30, 62C15.

*Corresponding author.

1. INTRODUCTION

Suppose we have two normal populations with a common mean 'μ' and possibly different variances σ_1^2 and σ_2^2. More specifically, let $\underset{\sim}{X} = (X_1, X_2, ..., X_m)$ and $\underset{\sim}{Y} = (Y_1, Y_2, ..., Y_n)$ be independent random samples taken from two normal populations $N(\mu, \sigma_1^2)$ and $N(\mu, \sigma_2^2)$ respectively. The problem is to estimate the common mean 'μ' under the assumption that the variances follow the ordering $\sigma_1^2 \leq \sigma_2^2$. In order to evaluate the performance of an estimator the loss functions

$$L_1(d, \mu) = \left(\frac{d-\mu}{\sigma_1}\right)^2, \tag{1.1}$$

$$L_2(d, \mu) = |d - \mu|, \tag{1.2}$$

and

$$L_3(d, \mu) = (d - \mu)^2, \tag{1.3}$$

are typically used, where d is an estimator for estimating 'μ' and $\underset{\sim}{\alpha} = (\mu, \sigma_1^2, \sigma_2^2)$; $\sigma_1^2 \leq \sigma_2^2$. Furthermore the risk of an estimator d is defined by

$$R(d, \mu) = E_{\underset{\sim}{\alpha}}\{L_i(d, \mu)\};\ i = 1, 2, 3.$$

The problem of estimating the common mean of two or more normal populations, without considering the order restriction on the variances, is quite popular and has a long history in the literature of statistical inference. In fact, the origin of the problem has been in the recovery of inter-block information in the study of balanced incomplete block designs problem (see Yates [23]). Moreover, the problem has received considerable attention by several pioneer researchers in the last few decades due to its practical applications as well as the theoretical challenges involved in it. This well known problem arises in situations, where two or more measuring devices in a laboratory are used to measure certain quantity, several independent agencies are employed to test the effectiveness of certain new drugs produced by a developer, two or more different methods are used to evaluate certain characteristic etc.. Under these circumstances, if it is assumed that the samples drawn follow normal distributions, then the task boils down to draw inference on the common mean when the variances are unknown and unequal. We refer to some excellent papers by Chang and Pal [4], Lin and Lee [12] and Kelleher [10] for applications as well as examples of such nature. Probably, Graybill and Deal [8] were the first to consider this well-known common mean problem under normality assumption, without taking into account the order restriction on the variances. They proposed a combined estimator by taking convex combination of two sample means with weights as the functions of sample variances. Their combined estimator performs better than the individual sample means in terms of mean squared error when the sample sizes are at least 11. Since then a lot of attention has been paid in this direction by several researchers. In fact, the main goal has been to obtain either some competitors to Graybill-Deal estimator or some alternative estimators which may perform better than both the sample means. Also few attempts have been made to prove the admissibility or inadmissibility of the Graybill-Deal estimator. For a detailed literature review and recent updates on estimating the common mean of two or more normal populations without taking

into account the order restriction on the variances, we refer to Khatri and Shah [11], Brown and Cohen [3], Cohen and Sackrowitz [6], Moore and Krishnamoorthy [14], Pal and Sinha [16], Pal *et al.* [15], Tripathy and Kumar [20, 21] and the references cited therein.

On the other hand, relatively less attention has been paid in estimating the common mean 'μ' when it is known a priori, that the variances follow certain simple ordering, say, $\sigma_1^2 \leq \sigma_2^2$. As an application of the common mean estimation under two ordered variances one can cite the example of evaluating the octane level of a particular grade of gasoline by the state inspectors in the United States. Usually the inspectors evaluating the octane level of gasoline sold at a gasoline station take two types of samples. Multiple samples of gasoline are taken on spot and their octane levels are quickly evaluated by a hand held device which is less precise and hence have high variance. Another batch of gasoline samples is taken and sent to state labs for a detailed, time consuming analysis of the octane level which is more accurate and has a smaller variance. If the spot analysis shows the mean octane level within a certain margin of the declared octane level then the inspector gives the seller a pass. Otherwise, the results from the lab tests are combined with the spot tests to determine the mean octane level. Disciplinary actions against the seller can be taken only if the combined estimate of the mean octane level falls below the declared level by a substantial margin. Probably Elfessi and Pal [7] were the first to consider this model with some justification and proposed an estimator that performs better than the Graybill-Deal estimator. In fact, their proposed estimator performs better than the Graybill-Deal estimator in terms of stochastic domination as well as universal domination. Later on, their results have been extended to the case of $k(\geq 2)$ normal populations by Misra and van der Meulen [13]. Chang *et al.* [5] also considered the estimation of a common mean under order restricted variances. They proposed a broad class of estimators that includes estimator proposed by Elfessi and Pal [7]. In fact, their proposed estimators stochastically dominate the estimators which do not obey the order restriction on the variances. However, for practical applications purpose, it is essential to have the specific estimators. Moreover, it is also necessary to know the amount of risk reduction after using the prior information regarding the ordering of the variances. Also we note that, the problem of estimation of a common standard deviation of several normal populations when the means are known to follow a simple ordering has been considered by Tripathy *et al.* [22] from a decision theoretic point of view.

In view of the above, we have proposed certain alternative estimators for the common mean when it is known a priori that the variances are ordered. These new estimators, which utilize the information about variance ordering, are shown to dominate their unrestricted counterparts (proposed by Moore and Krishnamoorthy [14], Khatri and Shah [11], Brown and Cohen [3], Tripathy and Kumar [20]) stochastically, universally and in terms of Pitman nearness criterion. Moreover we have obtained a plug-in type restricted MLE which beats the unrestricted MLE with respect to a squared error loss function which has been shown numerically. In addition to these, we derive a sufficient condition for improving equivariant estimators using orbit-by-orbit improvement technique of Brewster and Zidek [2]. It is also interesting to see the performance of MLE with respect to other estimators (including the existing one proposed by Elfessi and Pal [7]), under order restriction on the variances, which is lacking in the literature. We also observe that a detailed and in-depth study to compare the performances of all the existing estimators for the common mean under order restricted variances is lacking in the literature. Therefore, we intend to study the performances of all the estimators, - both proposed as well as the existing ones, through a comprehensive simulation

study which may fill the knowledge gap and provide useful information to the researchers from an application point of view.

The rest of the work is organized as follows. In Section 2, certain basic results have been discussed and a new plug-in type restricted MLE for the common mean μ has been proposed. In Section 3, some alternative estimators for the common mean μ have been constructed under order restriction on the variances. It is shown that the proposed estimators dominate their old counterparts proposed by Moore and Krishnamoorthy [14], Khatri and Shah [11], Brown and Cohen [3], Tripathy and Kumar [20] in terms of stochastic domination as well as universal domination. Moreover, in Section 4, we have proved that these alternative estimators also dominate their respective unrestricted counterparts in terms of Pitman measure of closeness criterion (see Pitman [18]). Sufficient conditions for improving the estimators which are invariant under affine transformations have been proved in Section 5, and consequently improved estimators have been derived. Interestingly, these improved estimators turned out to be the same as obtained in Section 2. We note that a theoretical comparison of all these proposed estimators seems difficult, and hence a simulation study has been carried out in order to compare numerically the risk functions of all the proposed estimators in Section 6. Moreover, the percentage of risk improvements of all the improved estimators upon their unrestricted counterparts have been noted with respect to all the three loss functions (1.1)–(1.3), which are quite significant. The percentage of relative risk improvements of all the proposed estimators have been obtained with respect to the Graybill-Deal estimator (treated as a benchmark) and recommendations have been made there. Finally we conclude our remarks with some examples to compute the estimates in Section 7.

2. SOME BASIC RESULTS

In this section, we discuss the statistical model and propose some alternative estimators for the common mean μ, when it is known a priori that the variances follow the simple ordering, that is, $\sigma_1^2 \leq \sigma_2^2$.

Let $\underset{\sim}{X} = (X_1, X_2, ..., X_m)$ and $\underset{\sim}{Y} = (Y_1, Y_2, ..., Y_n)$ be independent random samples taken from two normal populations with a common mean μ and possibly different variances σ_1^2 and σ_2^2 respectively. Let $N(\mu, \sigma_i^2)$ be denote the normal population with mean μ and variance σ_i^2; $i = 1, 2$. The target is to derive certain estimators for μ, when it is known a priori that, the variances are ordered, that is, $\sigma_1^2 \leq \sigma_2^2$ or equivalently $\sigma_1 \leq \sigma_2$. We note, that a minimal sufficient statistics (not complete) for this model exists and is given by $(\bar{X}, \bar{Y}, S_1^2, S_2^2)$ where

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i,\ \bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j,\ S_1^2 = \sum_{i=1}^{m}(X_i - \bar{X})^2,\ S_2^2 = \sum_{j=1}^{n}(Y_j - \bar{Y})^2. \tag{2.1}$$

We further note that, $\bar{X} \sim N(\mu, \sigma_1^2/m)$, $\bar{Y} \sim N(\mu, \sigma_2^2/n)$, $S_1^2/\sigma_1^2 \sim \chi_{m-1}^2$, and $S_2^2/\sigma_2^2 \sim \chi_{n-1}^2$. When there is no order restrictions on the variances, a number of estimators have been proposed by several researchers in the recent past. Let us consider the following well known

estimators for the common mean μ when there is no order restriction on the variances:

$$d_{GD} = \frac{m(m-1)S_2^2\bar{X} + n(n-1)S_1^2\bar{Y}}{m(m-1)S_2^2 + n(n-1)S_1^2} \quad \text{(Graybill and Deal [8])},$$

$$d_{KS} = \frac{m(m-3)S_2^2\bar{X} + n(n-3)S_1^2\bar{Y}}{m(m-3)S_2^2 + n(n-3)S_1^2} \quad \text{(Khatri and Shah [11])},$$

$$d_{MK} = \frac{\bar{X}\sqrt{m(m-1)}S_2 + \bar{Y}\sqrt{n(n-1)}S_1}{\sqrt{m(m-1)}S_2 + \sqrt{n(n-1)}S_1} \quad \text{(Moore and Krishnamoorthy [14])},$$

$$d_{TK} = \frac{\bar{X}\sqrt{m}c_nS_2 + \bar{Y}\sqrt{n}c_mS_1}{\sqrt{m}c_nS_2 + \sqrt{n}c_mS_1} \quad \text{(Tripathy and Kumar [20])},$$

$$d_{BC1} = \bar{X} + \left\{\frac{(\bar{Y}-\bar{X})b_1S_1^2/m(m-1)}{S_1^2/m(m-1) + S_2^2/(n(n+2)) + (\bar{Y}-\bar{X})^2/(n+2)}\right\}$$

$$d_{BC2} = \bar{X} + (\bar{Y}-\bar{X})\left\{\frac{b_2n(n-1)S_1^2}{n(n-1)S_1^2 + m(m-1)S_2^2}\right\} \quad \text{(Brown and Cohen [3])},$$

$$d_{GM} = \frac{m\bar{X} + n\bar{Y}}{m+n} \quad \text{(grand sample mean)},$$

where $c_m = \Gamma(\frac{m-1}{2})/(\sqrt{2}\Gamma(\frac{m}{2}))$, $c_n = \Gamma(\frac{n-1}{2})/(\sqrt{2}\Gamma(\frac{n}{2}))$, $0 < b_1 < b_{\max}(m,n)$, $0 < b_2 < b_{\max}(m,n-3)$, and $b_{\max}(m,n) = 2(n+2)/nE(\max(1/V, 1/V^2))$. Here V is a random variable having F-distribution with $(n+2)$ and $(m-1)$ degrees of freedom.

Finally we consider the MLE of μ whose closed form does not exist (see Pal *et al.* [15]). The MLE of μ can be obtained numerically by solving the following system of three equations in three unknowns μ, σ_1^2, and σ_2^2:

$$\mu = \frac{\frac{m}{\sigma_1^2}\bar{x} + \frac{n}{\sigma_2^2}\bar{y}}{\frac{m}{\sigma_1^2} + \frac{n}{\sigma_2^2}}, \tag{2.2}$$

$$\sigma_1^2 = \frac{s_1^2}{m} + \left(\frac{n\sigma_1^2}{n\sigma_1^2 + m\sigma_2^2}\right)^2(\bar{x}-\bar{y})^2, \tag{2.3}$$

$$\sigma_2^2 = \frac{s_2^2}{n} + \left(\frac{m\sigma_2^2}{n\sigma_1^2 + m\sigma_2^2}\right)^2(\bar{x}-\bar{y})^2. \tag{2.4}$$

Here $(\bar{x}, \bar{y}, s_1^2, s_2^2)$ denotes the observed value of $(\bar{X}, \bar{Y}, S_1^2, S_2^2)$. Let the solution of the above system of equations be $\hat{\mu}_{ML}$, $\hat{\sigma}_{1ML}^2$ and $\hat{\sigma}_{2ML}^2$. These are the MLEs of μ, σ_1^2 and σ_2^2 respectively, when there is no order restriction on the variances.

Next, we discuss some results on estimating common mean when it is known a priori that the variances follow the simple ordering $\sigma_1^2 \le \sigma_2^2$. Let $\beta = (n(n-1)S_1^2)/(m(m-1)S_2^2 + n(n-1)S_1^2)$. Under order restriction on the variances, Elfessi and Pal [7] proposed a new estimator, call it $\hat{d}_{EP}$ which is given by

$$\hat{d}_{EP} = \begin{cases} (1-\beta)\bar{X} + \beta\bar{Y}, & \text{if } \frac{S_1^2}{m-1} \le \frac{S_2^2}{n-1} \\ \beta^*\bar{X} + (1-\beta^*)\bar{Y}, & \text{if } \frac{S_1^2}{m-1} > \frac{S_2^2}{n-1}, \end{cases}$$

where

$$\beta^* = \begin{cases} \beta, & \text{if } m = n \\ \dfrac{m}{m+n}, & \text{if } m \ne n. \end{cases}$$

In the above definition of $\hat{d}_{EP}$ for the case $m = n$, when $\beta^* = \beta$, we mean β as well as the conditions must be simplified for $m = n$.

It is well known that the estimator $\hat{d}_{EP}$ dominates d_{GD} stochastically as well as universally when $\sigma_1^2 \leq \sigma_2^2$. Further Misra and van der Meulen [13] extended these dominance results to the case of $k(\geq 2)$ normal populations and also proved that the estimator $\hat{d}_{EP}$ performs better than d_{GD} in terms of Pitman measure of closeness criterion. The MLE of μ has been obtained by solving the system of equations numerically as shown above (see equations (2.2) to (2.4))). When the variances are ordered, using the isotonic version of the MLEs of σ_i^2, we obtain plug-in type restricted MLEs (numerically) of σ_1^2 and σ_2^2 respectively as

$$\hat{\sigma}_{1R}^2 = \begin{cases} \hat{\sigma}_{1ML}^2, & \text{if } \hat{\sigma}_{1ML}^2 \leq \hat{\sigma}_{2ML}^2 \\ \dfrac{1}{2}(\hat{\sigma}_{1ML}^2 + \hat{\sigma}_{2ML}^2), & \text{if } \hat{\sigma}_{1ML}^2 > \hat{\sigma}_{2ML}^2, \end{cases}$$

and

$$\hat{\sigma}_{2R}^2 = \begin{cases} \hat{\sigma}_{2ML}^2, & \text{if } \hat{\sigma}_{1ML}^2 \leq \hat{\sigma}_{2ML}^2 \\ \dfrac{1}{2}(\hat{\sigma}_{1ML}^2 + \hat{\sigma}_{2ML}^2), & \text{if } \hat{\sigma}_{1ML}^2 > \hat{\sigma}_{2ML}^2 \end{cases}$$

(see Barlow *et al.* [1]). Substituting these estimators in (2.2), we get a plug-in type restricted MLE, (call it d_{RM}) for μ as

$$d_{RM} = \frac{m\hat{\sigma}_{2R}^2\bar{X} + n\hat{\sigma}_{1R}^2\bar{Y}}{m\hat{\sigma}_{2R}^2 + n\hat{\sigma}_{1R}^2}.$$

Further using the grand sample mean of the two populations, one gets another plug-in type restricted MLE of μ, call it $\hat{d}_{RM}$, and is given by

$$\hat{d}_{RM} = \begin{cases} \hat{\mu}_{ML}, & \text{if } \hat{\sigma}_{1ML}^2 \leq \hat{\sigma}_{2ML}^2 \\ \dfrac{m\bar{X} + n\bar{Y}}{m+n}, & \text{if } \hat{\sigma}_{1ML}^2 > \hat{\sigma}_{2ML}^2. \end{cases}$$

Through a simulation study, Tripathy and Kumar [20] concluded that the estimators d_{MK} and d_{TK} compete with each other and perform better than d_{GD} when the variances are not far away from each other. Authors also mentioned that for small values of the ratios of the variances, the estimator d_{KS} compete with d_{GD}. Hence, it is quite evident that one needs to find alternative estimators for μ which may compete with $\hat{d}_{EP}$ when $\sigma_1^2 \leq \sigma_2^2$ or equivalently $\sigma_1 \leq \sigma_2$. In the next sections to follow (Sections 3 and 4), we propose some new estimators which dominate their respective unrestricted counterparts stochastically as well as universally and may compete with $\hat{d}_{EP}$ in terms of risks. Now onwards for convenient we will denote $\hat{d}_{EP}$ as $\hat{d}_{GD}$.

Remark 2.1. One can construct another plug-in type estimator for μ by replacing the estimators $\hat{\sigma}_{1R}^2$ and $\hat{\sigma}_{2R}^2$ in d_{RM} by $\hat{\sigma}_{1R}^2 = \min(\hat{\sigma}_{1ML}^2, (m\hat{\sigma}_{1ML}^2 + n\hat{\sigma}_{2ML}^2)/(m+n))$ and $\hat{\sigma}_{2R}^2 = \max(\hat{\sigma}_{2ML}^2, (m\hat{\sigma}_{1ML}^2 + n\hat{\sigma}_{2ML}^2)/(m+n))$ respectively when $m \neq n$. It has been revealed from our numerical study (Section 6) that it acts as a competitor of d_{RM}.

Remark 2.2. The estimators d_{RM} and $\hat{d}_{RM}$ are seen to perform equally good, which has been checked from our simulation study in Section 6. Hence we include only $\hat{d}_{RM}$ in our numerical comparison in Section 6.

3. STOCHASTIC DOMINATION UNDER ORDER RESTRICTION ON THE VARIANCES

In this section we propose some alternative estimators for the common mean μ under order restriction on the variances that is when it is known a priori that $\sigma_1^2 \leq \sigma_2^2$ or equivalently $\sigma_1 \leq \sigma_2$. Further it will be shown that each of these alternative estimators dominate their unrestricted counterparts proposed by Moore and Krishnamoorthy [14], Tripathy and Kumar [20], Khatri and Shah [11] and Brown and Cohen [3] stochastically under order restriction on the variances.

To start with, let us define

$$\begin{aligned}
\beta_1 &= \frac{\sqrt{n(n-1)}S_1}{\sqrt{m(m-1)}S_2 + \sqrt{n(n-1)}S_1}, \\
\beta_2 &= \frac{\sqrt{n}c_m S_1}{\sqrt{m}c_n S_2 + \sqrt{n}c_m S_1}, \\
\beta_3 &= \frac{n(n-3)S_1^2}{m(m-3)S_2^2 + n(n-3)S_1^2}, \\
\beta_4 &= \frac{b_2 S_1^2}{S_1^2 + S_2^2}.
\end{aligned}$$

We propose the following estimators for the common mean μ, when the variances known to follow the simple ordering $\sigma_1^2 \leq \sigma_2^2$:

$$\hat{d}_{MK} = \begin{cases} (1-\beta_1)\bar{X} + \beta_1\bar{Y}, & \text{if } \frac{\sqrt{n-1}S_1}{\sqrt{m-1}S_2} \leq \sqrt{\frac{n}{m}}, \\ \beta_1^*\bar{X} + (1-\beta_1^*)\bar{Y}, & \text{if } \frac{\sqrt{n-1}S_1}{\sqrt{m-1}S_2} > \sqrt{\frac{n}{m}}, \end{cases}$$

$$\hat{d}_{TK} = \begin{cases} (1-\beta_2)\bar{X} + \beta_2\bar{Y}, & \text{if } \frac{S_1}{S_2} \leq \sqrt{\frac{n}{m}}\frac{c_n}{c_m}, \\ \beta_2^*\bar{X} + (1-\beta_2^*)\bar{Y}, & \text{if } \frac{S_1}{S_2} > \sqrt{\frac{n}{m}}\frac{c_n}{c_m}, \end{cases}$$

$$\hat{d}_{KS} = \begin{cases} (1-\beta_3)\bar{X} + \beta_3\bar{Y}, & \text{if } \frac{S_1^2}{S_2^2} \leq \frac{m-3}{n-3}, \\ \beta_3^*\bar{X} + (1-\beta_3^*)\bar{Y}, & \text{if } \frac{S_1^2}{S_2^2} > \frac{m-3}{n-3}, \end{cases}$$

where for $i = 1, 2, 3$ we denote

$$\beta_i^* = \begin{cases} \beta_i, & \text{if } m = n, \\ \dfrac{m}{m+n}, & \text{if } m \neq n. \end{cases}$$

Finally, we propose an estimator for the case of equal sample sizes as

$$\hat{d}_{BC2} = \begin{cases} (1-\beta_4)\bar{X} + \beta_4\bar{Y}, & \text{if } \frac{S_2^2}{S_1^2} \geq (2b_2 - 1), \\ \beta_4\bar{X} + (1-\beta_4)\bar{Y}, & \text{if } \frac{S_2^2}{S_1^2} < (2b_2 - 1). \end{cases}$$

In the above definitions of the estimators $\hat{d}_{MK}$, $\hat{d}_{TK}$, $\hat{d}_{KS}$ for the case $m = n$, when $\beta_i^* = \beta_i$; $i = 1, 2, 3$, we mean that both β_i and the corresponding conditions must be simplified by putting $m = n$.

To proceed further we need the following two definitions which will be used in developing the section. Let d_1 and d_2 be any two estimators of the unknown parameter say θ.

Definition 3.1. The estimator d_1 is said to dominate another estimator d_2 stochastically if $P_\theta[(d_2 - \mu)^2 \leq c] \leq P_\theta[(d_1 - \mu)^2 \leq c], \quad \forall\ c > 0$.

Definition 3.2. Let the loss function $L(d, \theta)$ in estimating θ by d be a non-decreasing function of the error $|d - \theta|$. An estimator d_1 is said to dominate another estimator d_2 universally if $EL(|d_1 - \theta|) \leq EL(|d_2 - \theta|)$, over the parameter space for all $L(.)$ non-decreasing. Further it was shown by Hwang [9] that d_1 dominates d_2 universally if and only if d_1 dominates d_2 stochastically.

Next, we prove the following results for estimating the common mean μ, under order restriction on the variances, which are immediate.

Theorem 3.1. *Let the loss function $L(.)$ be a non-decreasing function of the error $|d - \mu|$. Further assume that the variances are known to follow the ordering $\sigma_1^2 \leq \sigma_2^2$. Then for estimating the common mean μ we have the following dominance results.*

- (**i**) *The estimator $\hat{d}_{MK}$ dominates d_{MK} stochastically and hence universally.*
- (**ii**) *The estimator $\hat{d}_{TK}$ dominates d_{TK} stochastically and hence universally.*
- (**iii**) *The estimator $\hat{d}_{KS}$ dominates d_{KS} stochastically and hence universally.*
- (**iv**) *The estimator $\hat{d}_{BC2}$ dominates d_{BC2} stochastically and hence universally.*

Proof: Please see Appendix. □

4. PITMAN MEASURE OF CLOSENESS

In this section, we prove that the new proposed estimators $\hat{d}_{MK}$, $\hat{d}_{TK}$, $\hat{d}_{KS}$, and $\hat{d}_{BC2}$, perform better than their old counterparts in terms of Pitman measure of closeness criterion when it is known a priori that the variances follow the ordering $\sigma_1^2 \leq \sigma_2^2$ or equivalently $\sigma_1 \leq \sigma_2$. To prove the main results of this section, we need the following results. Let δ_1 and δ_2 be any two estimators of a real parametric function say $\psi(\theta)$. Pitman [18] proposed a measure of relative closeness to the parametric function $\psi(\theta)$ for comparing two estimators in the following fashions.

Definition 4.1. The estimator δ_1 should be preferred to δ_2 if for every θ, $PMC_\theta(\delta_1, \delta_2) = P_\theta(|\delta_1 - \psi(\theta)| < |\delta_2 - \psi(\theta)||\delta_1 \neq \delta_2) \geq \frac{1}{2}$, and with strict inequality for some θ.

The following lemma will be useful for proving the main results of this section, which was proposed by Peddada and Khatree [17].

Lemma 4.1. *Suppose the random vector (X, Y) has a bivariate normal distribution with $E(X) = E(Y) = 0$ and $E(X^2) < E(Y^2)$. Then $P(|X| < |Y|) > \frac{1}{2}$.*

Let $\underset{\sim}{\alpha} = (\mu, \sigma_1^2, \sigma_2^2)$ and $\Omega_R = \{\underset{\sim}{\alpha} = (\mu, \sigma_1^2, \sigma_2^2) : -\infty < \mu < \infty, 0 < \sigma_1^2 \leq \sigma_2^2 < \infty\}$. We prove the following theorem.

Theorem 4.1. *For estimating the common mean μ of two normal populations, when $\sigma_1^2 \leq \sigma_2^2$, we have the following dominance results:*

(**i**) $PMC(\hat{d}_{MK}, d_{MK}) > \frac{1}{2}, \forall\ \underset{\sim}{\alpha} \in \Omega_R.$

(**ii**) $PMC(\hat{d}_{TK}, d_{TK}) > \frac{1}{2}, \forall\ \underset{\sim}{\alpha} \in \Omega_R.$

(**iii**) $PMC(\hat{d}_{KS}, d_{KS}) > \frac{1}{2}, \forall\ \underset{\sim}{\alpha} \in \Omega_R.$

(**iv**) $PMC(\hat{d}_{BC2}, d_{BC2}) > \frac{1}{2}, \forall\ \underset{\sim}{\alpha} \in \Omega_R.$

Proof: The proof of the theorem is easy after using the Lemma 4.1, and hence has been omitted. □

In the next section we will introduce the concept of invariance to our problem and prove some inadmissibility results in the classes of equivariant estimators for the common mean.

5. INADMISSIBILITY RESULTS UNDER ORDER RESTRICTION ON THE VARIANCES

In this section we introduce the concept of invariance to the problem and derive some inadmissibility results for both affine and location equivariant estimators under order restriction on the variances. As a consequence, estimators dominating some of the existing well known estimators for the common mean have been derived, under order restriction on the variances.

5.1. Affine Class

Let us introduce the concept of invariance to our problem. More specifically, consider the affine group of transformations, $G_A = \{g_{a,b} : g_{a,b}(x) = ax + b, a > 0, b \in R\}$. Under the transformation $g_{a,b}$, $X_i \to aX_i + b; i = 1, 2, ..., m$, $Y_j \to aY_j + b; j = 1, 2, ..., n$ and consequently the sufficient statistics $\bar{X} \to a\bar{X} + b$, $\bar{Y} \to a\bar{Y} + b$, $S_i^2 \to a^2 S_i^2$, $\mu \to a\mu + b$, $\sigma_i^2 \to a^2\sigma_i^2$ and the family of distributions remains invariant. The problem remains invariant if we choose the loss function as (1.1). The form of an affine equivariant estimator for estimating μ, based on the sufficient statistic $(\bar{X}, \bar{Y}, S_1^2, S_2^2)$ is obtained as

$$d_\Psi = \bar{X} + S_1 \Psi(\underset{\sim}{T}), \tag{5.1}$$

where $\underset{\sim}{T} = (T_1, T_2)$, $T_1 = (\bar{Y} - \bar{X})/S_1$, $T_2 = S_2^2/S_1^2$ and Ψ is any real valued function.

Let us define a new function Ψ_0 for the affine equivariant estimator d_Ψ as

$$\Psi_0(\underset{\sim}{t}) = \begin{cases} \dfrac{n}{n+m}\min(t_1,0), & \text{if } \Psi(\underset{\sim}{t}) < \frac{n}{n+m}\min(t_1,0), \\ \Psi(\underset{\sim}{t}), & \text{if } \frac{n}{n+m}\min(t_1,0) \le \Psi(\underset{\sim}{t}) \le \frac{n}{n+m}\max(t_1,0), \\ \dfrac{n}{n+m}\max(t_1,0), & \text{if } \Psi(\underset{\sim}{t}) > \frac{n}{n+m}\max(t_1,0). \end{cases} \tag{5.2}$$

The following theorem gives a sufficient condition for improving estimators in the class of affine equivariant estimators of the form (5.1), under order restriction on the variances.

Theorem 5.1. *Let d_Ψ be an affine equivariant estimator of the form (5.1) for estimating the common mean μ and the loss function be the affine invariant loss (1.1). The estimator d_Ψ is inadmissible and is improved by d_{Ψ_0} if $P(\Psi(\underset{\sim}{T}) \neq \Psi_0(\underset{\sim}{T})) > 0$, for some choices of the parameters $\underset{\sim}{\alpha}$; $\sigma_1 \le \sigma_2$.*

Proof: Please see Appendix. □

Next we will apply Theorem 5.1, to obtain some improved estimators for the common mean μ, under the assumption that $\sigma_1^2 \le \sigma_2^2$. It is easy to observe that all the estimators discussed in Section 2, for the common mean μ without taking into account the order restriction on the variances, fall into the class $d_\Psi = \bar{X} + S_1\Psi(\underset{\sim}{T})$. We apply Theorem 5.1 to get their corresponding improved estimators under the assumption that $\sigma_1^2 \le \sigma_2^2$. Let us first consider the estimator $d_{GD} = \bar{X} + S_1\Psi(\underset{\sim}{T})$, where $\Psi(\underset{\sim}{T}) = (n(n-1)T_1)/(m(m-1)T_2 + n(n-1))$. Note that $\Psi(\underset{\sim}{t}) > (n/(m+n))\max(0,t_1)$, when $S_1^2/(m-1) > S_2^2/(n-1)$. Hence the estimator d_{GD} is inadmissible and is improved by the estimator

$$d_{GD}^a = \begin{cases} \dfrac{m(m-1)S_2^2\bar{X} + n(n-1)S_1^2\bar{Y}}{m(m-1)S_2^2 + n(n-1)S_1^2}, & \text{if } \frac{S_1^2}{m-1} \le \frac{S_2^2}{n-1} \\ \dfrac{m\bar{X} + n\bar{Y}}{m+n}, & \text{if } \frac{S_1^2}{m-1} > \frac{S_2^2}{n-1}, \end{cases}$$

under order restriction on the variances.

Similarly one can get the estimators which improve upon d_{KS}, d_{MK}, d_{TK}, d_{BC1}, and d_{BC2} respectively as

$$d_{KS}^a = \begin{cases} \dfrac{m(m-3)S_2^2\bar{X} + n(n-3)S_1^2\bar{Y}}{m(m-3)S_2^2 + n(n-3)S_1^2}, & \text{if } \frac{S_1^2}{m-3} \le \frac{S_2^2}{n-3} \\ \dfrac{m\bar{X} + n\bar{Y}}{m+n}, & \text{if } \frac{S_1^2}{m-3} > \frac{S_2^2}{n-3}, \end{cases}$$

$$d_{MK}^a = \begin{cases} \dfrac{\sqrt{m(m-1)}S_2\bar{X} + \sqrt{n(n-1)}S_1\bar{Y}}{\sqrt{m(m-1)}S_2 + \sqrt{n(n-1)}S_1}, & \text{if } \frac{S_1}{\sqrt{m-1}} \le \frac{S_2}{\sqrt{n-1}}, \\ \dfrac{m\bar{X} + n\bar{Y}}{m+n}, & \text{if } \frac{S_1}{\sqrt{m-1}} > \frac{S_2}{\sqrt{n-1}}, \end{cases}$$

$$d^a_{TK} = \begin{cases} \dfrac{\sqrt{m}c_n S_2 \bar{X} + \sqrt{n}c_m S_1 \bar{Y}}{\sqrt{m}c_n S_2 + \sqrt{n}c_m S_1}, & \text{if } \frac{S_1}{S_2} \le \sqrt{\frac{n}{m}}\frac{c_n}{c_m}, \\ \dfrac{m\bar{X} + n\bar{Y}}{m+n}, & \text{if } \frac{S_1}{S_2} \le \sqrt{\frac{n}{m}}\frac{c_n}{c_m}, \end{cases}$$

$$d^a_{BC1} = \begin{cases} d_{BC1}, & \text{if } \frac{S_2^2}{S_1^2} + n(\frac{\bar{Y}-\bar{X}}{S_1})^2 > \frac{n+2}{m(m-1)}[b_1(m+n) - n] \\ \dfrac{m\bar{X} + n\bar{Y}}{m+n}, & \text{if } \frac{S_2^2}{S_1^2} + n(\frac{\bar{Y}-\bar{X}}{S_1})^2 \le \frac{n+2}{m(m-1)}[b_1(m+n) - n], \end{cases}$$

and

$$d^a_{BC2} = \begin{cases} d_{BC2}, & \text{if } \frac{m(m-1)S_2^2}{n(n-1)S_1^2} \ge b_2(1+\frac{m}{n}) - 1 \\ \dfrac{m\bar{X} + n\bar{Y}}{m+n}, & \text{if } \frac{m(m-1)S_2^2}{n(n-1)S_1^2} < b_2(1+\frac{m}{n}) - 1. \end{cases}$$

Remark 5.1. It is interesting to note that, for the case of unequal sample sizes, that is for $m \neq n$, the estimators $d^a_{GD} = \hat{d}_{GD}$, $d^a_{KS} = \hat{d}_{KS}$, $d^a_{MK} = \hat{d}_{MK}$, $d^a_{TK} = \hat{d}_{TK}$, $d^a_{BC2} = \hat{d}_{BC2}$. However for equal sample sizes, application of the Theorem 5.1 produces different estimators.

Remark 5.2. We note that, though the MLE of μ can not be obtained in a closed form, however from (2.2) it is easy to write $\hat{\mu}_{ML} = \bar{X} + S_1\Psi_{ML}(\underset{\sim}{T})$, where $\Psi_{ML}(\underset{\sim}{T}) = T_1 n\hat{\sigma}^2_{1ML}/(m\hat{\sigma}^2_{2ML} + n\hat{\sigma}^2_{1ML})$, and $\hat{\sigma}^2_{1ML}$, $\hat{\sigma}^2_{2ML}$ are to be found by solving (2.3) and (2.4). Though $\Psi_{ML}(\underset{\sim}{T})$ does not have a closed form, for a given dataset(sample values), we can find the value of $\Psi_{ML}(\underset{\sim}{t})$. Therefor, we can find $\Psi_0^{ML}(\underset{\sim}{t})$ by using (5.2). Hence we can apply Theorem 5.1 and find the value of the improved estimator $d^a_{ML} = \bar{X} + S_1\Psi_0^{ML}(\underset{\sim}{t})$ which does not have a closed form. It has been observed in our simulation study that the improved version of the MLE appears to have the identical risk as the estimator d_{RM}.

5.2. Location Class

A larger class of estimators than the class considered above is the class of location equivariant estimators. Let $G_L = \{g_c : g_c(x) = x + c, -\infty < c < \infty\}$ be the location group of transformations. Under the transformation g_c, we observe that, $\bar{X} \to \bar{X} + c$, $\bar{Y} \to \bar{Y} + c$, $S_1^2 \to S_2^2$, $S_2^2 \to S_2^2$, and the parameters $\mu \to \mu + c$, $\sigma_1 \to \sigma_1$. The family of probability distributions is invariant and consequently the estimation problem is also invariant under the loss (1.1). Based on the minimal sufficient statistics $(\bar{X}, \bar{Y}, S_1^2, S_2^2)$ the form of a location equivariant estimator for estimating the common mean μ is thus obtained as

$$d_\psi = \bar{X} + \psi(\underset{\sim}{U}), \tag{5.3}$$

where $\underset{\sim}{U} = (T, S_1^2, S_2^2)$, $T = \bar{Y} - \bar{X}$, and ψ is a real valued function. Let us define a function ψ_0 for the location equivariant estimator d_ψ as

$$\psi_0(\underset{\sim}{t}) = \begin{cases} \dfrac{n}{n+m}\min\{t,0\}, & \text{if } \psi(\underset{\sim}{u}) < \frac{n}{m+n}\min\{t,0\}, \\ \psi(\underset{\sim}{u}), & \text{if } \frac{n}{n+m}\min\{t,0\} \le \psi(\underset{\sim}{u}) \le \frac{n}{n+m}\max\{t,0\}, \\ \dfrac{n}{n+m}\max\{t,0\}, & \text{if } \psi(\underset{\sim}{u}) > \frac{n}{n+m}\max\{t,0\}. \end{cases} \tag{5.4}$$

The following theorem gives a sufficient condition for improving location equivariant estimators under the condition that the variances follow the ordering $\sigma_1^2 \le \sigma_2^2$.

Theorem 5.2. *Let d_ψ be a location equivariant estimator for estimating the common mean μ and the loss function be (1.1). Let the function $\psi_0(\underset{\sim}{u})$ be as defined in (5.6). The estimator d_ψ is inadmissible and is improved by d_{ψ_0} if $P_{\underset{\sim}{\alpha}}(\psi(\underset{\sim}{U}) \neq \psi_0(\underset{\sim}{U})) > 0$ for some choices of the parameters $\underset{\sim}{\alpha} = (\mu, \sigma_1^2, \sigma_2^2)$; $\sigma_1^2 \le \sigma_2^2$.*

Proof: The proof is similar to the proof of the Theorem 5.1, and hence has been omitted for brevity. □

Remark 5.3. We also observe that all the estimators proposed in Section 2, including the MLE (whose closed form does not exist) belong to the class $d_\psi(\underset{\sim}{U}) = \bar{X} + \psi(\underset{\sim}{U})$. Hence as an application of Theorem 5.2, produces improved estimators. Further we note that, though location class produces larger class of estimators, the sufficient conditions in Theorem 5.2, does not help to obtain different improved estimators than those obtained by applying Theorem 5.1, under order restriction on the variances. In fact, the sufficient conditions in Theorem 5.1 and 5.2 for improving equivariant estimators produces the same improved estimators under order restricted variances.

Remark 5.4. The performances of all the improved estimators which has been proposed in Section 2 as well as in this section by applying Theorem 5.1, will be evaluated in Section 6, using the affine invariant loss function L_1. Further the percentage of risk improvements upon their respective old counterparts has been noted.

Remark 5.5. We note that the estimator d_{GM}, also belongs to the classes given in (5.1) and (5.5). However, the conditions in Theorem 5.1 and Theorem 5.2 for improving it, do not satisfy. Hence the estimator d_{GM} could not be improved by applying either Theorem 5.1 or Theorem 5.2, under the condition that the variances are ordered that is, $\sigma_1^2 \le \sigma_2^2$.

6. A SIMULATION STUDY

It should be noted that, in Section 2 we have constructed the plug-in type restricted MLE $\hat{d}_{RM}$ for the common mean μ, taking into account the order restriction on the variances.

Moreover, in Sections 3 and 4 we have also constructed some alternative estimators such as $\hat{d}_{MK}$, $\hat{d}_{TK}$, $\hat{d}_{KS}$, and $\hat{d}_{BC2}$ and proved that each of these estimators dominate their old unrestricted counterparts in terms of stochastic domination as well as Pitman measure of closeness criterion. Furthermore in Section 5, we have proposed some improved estimators namely d^a_{GD}, d^a_{KS}, d^a_{MK}, d^a_{TK}, d^a_{BC1}, d^a_{BC2} by an application of Theorem 5.1 and 5.2. In addition to all these estimators, we have also considered the improved estimator $\hat{d}_{GD}$ proposed by Elfessi and Pal [7]. In order to know the performances of all these improved estimators, one needs to compare the risk functions. We observe that an analytical comparison of all these estimators seems quite impossible, hence in this section we compare the risk functions of all the improved estimators numerically through Monte-Carlo simulation method. For this purpose we have generated $20,000$ random samples of sizes m and n respectively from $N(\mu, \sigma_1^2)$ and $N(\mu, \sigma_2^2)$, with the condition that $\sigma_1^2 \leq \sigma_2^2$. The accuracy of the simulation has been checked and the error has been checked which is seen up to 10^{-3}. To proceed further, we define the percentage of risk improvements of all the improved estimators upon each of their unrestricted counterparts as follows:

$$P1 = \left(1 - \frac{R(\hat{d}_{GD}, \mu)}{R(d_{GD}, \mu)}\right) \times 100, \quad P2 = \left(1 - \frac{R(\hat{d}_{KS}, \mu)}{R(d_{KS}, \mu)}\right) \times 100,$$

$$P3 = \left(1 - \frac{R(\hat{d}_{MK}, \mu)}{R(d_{MK}, \mu)}\right) \times 100, \quad P4 = \left(1 - \frac{R(\hat{d}_{TK}, \mu)}{R(d_{TK}, \mu)}\right) \times 100,$$

$$P5 = \left(1 - \frac{R(d^a_{GD}, \mu)}{R(d_{GD}, \mu)}\right) \times 100, \quad P6 = \left(1 - \frac{R(d^a_{KS}, \mu)}{R(d_{KS}, \mu)}\right) \times 100,$$

$$P7 = \left(1 - \frac{R(d^a_{MK}, \mu)}{R(d_{MK}, \mu)}\right) \times 100, \quad P8 = \left(1 - \frac{R(d^a_{TK}, \mu)}{R(d_{TK}, \mu)}\right) \times 100,$$

$$P9 = \left(1 - \frac{R(\hat{d}_{RM}, \mu)}{R(d_{ML}, \mu)}\right) \times 100.$$

In order to compare the performances of all the improved estimators among themselves we use the affine loss function (1.1). It is better to compare the risk functions of all the improved estimators with respect to a benchmark estimator which can be the Graybill-Deal (see Graybill and Deal [8]) estimator. We define the percentage of relative risk performances of all the improved estimators with respect to the benchmark estimator d_{GD} as follows:

$$R1 = \left(1 - \frac{R(\hat{d}_{GD}, \mu)}{R(d_{GD}, \mu)}\right) \times 100, \quad R2 = \left(1 - \frac{R(\hat{d}_{KS}, \mu)}{R(d_{GD}, \mu)}\right) \times 100,$$

$$R3 = \left(1 - \frac{R(\hat{d}_{MK}, \mu)}{R(d_{GD}, \mu)}\right) \times 100, \quad R4 = \left(1 - \frac{R(\hat{d}_{TK}, \mu)}{R(d_{GD}, \mu)}\right) \times 100,$$

$$R5 = \left(1 - \frac{R(d^a_{GD}, \mu)}{R(d_{GD}, \mu)}\right) \times 100, \quad R6 = \left(1 - \frac{R(d^a_{MK}, \mu)}{R(d_{GD}, \mu)}\right) \times 100,$$

$$R7 = \left(1 - \frac{R(d^a_{BC1}, \mu)}{R(d_{GD}, \mu)}\right) \times 100, \quad R8 = \left(1 - \frac{R(d^a_{BC2}, \mu)}{R(d_{GD}, \mu)}\right) \times 100,$$

$$R9 = \left(1 - \frac{R(\hat{d}_{RM}, \mu)}{R(d_{GD}, \mu)}\right) \times 100.$$

It is easy to observe that the risks of all the estimators are functions of τ with respect to the loss function L_1 as given in (1.1), where we denote $\tau = \sigma_1^2/\sigma_2^2$, $0 < \tau \leq 1$. We note that, when the sample sizes are unequal (that is when $m \neq n$) $\hat{d}_{GD} = d^a_{GD}$, $\hat{d}_{KS} = d^a_{KS}$, $\hat{d}_{MK} = d^a_{MK}$, and $\hat{d}_{TK} = d^a_{TK}$. Further we notice that for equal sample sizes (that is when $m = n$) $\hat{d}_{GD} = \hat{d}_{KS}$ and $\hat{d}_{MK} = \hat{TK}$. In our simulation study we have chosen $b_1 = \frac{1}{2}b_{\max}(m, n)$ and $b_2 = \frac{1}{2}b_{\max}(m, n-3)$, where the values of $b_{\max}(m, n)$ have been taken from the table given in Brown and Cohen [3]. Moreover we observe that for $b_2 = 1$, the estimator $d_{BC2} = d_{GD}$ also when $b_2 = 0$, it reduces to $\bar{X}$. The percentage of risk improvements of d^a_{BC1}, d^a_{BC2} and $\hat{d}_{BC2}$ upon their unrestricted counterparts are seen to be very marginal and hence have not been tabulated. The simulation study has been carried out for various combinations of sample sizes while the parameter $\tau \in (0, 1]$. For illustration purpose we have presented the percentage of risk improvements as well as the percentage of relative risk improvements of all the estimators for some choices of sample sizes in Tables 1–7. In Tables 1 and 2 we have presented the percentage of risk improvements of all the improved estimators upon their unrestricted counterparts for equal and unequal sample sizes respectively with respect to the loss function (1.1). Particularly, in Table 1, the percentage of risk improvements of all the improved estimators have been presented for the sample sizes $(5, 5)$, $(12, 12)$, $(20, 20)$ and $(30, 30)$. The first and the seventh column represent the values of τ and the rest of the columns represent the percentage of risk improvements of each of the improved estimators. The table consists of several cells. In each cell, corresponding to one choice of τ, there correspond four values of percentage of risk values for the sample sizes $(5, 5)$, $(12, 12)$, $(20, 20)$ and $(30, 30)$. Table 2, is divided into two parts, specifically the first half (column second to sixth) represents the percentage of risk performances for all the estimators with sample sizes $(5, 10)$, and $(12, 20)$. The second part (column seventh to eleventh) represents the percentage of risk improvements for the sample sizes $(10, 5)$ and $(20, 12)$. In this table the first column also represents the values of τ and the columns second onwards represent the percentage of risk improvements of all the estimators upon their unrestricted counterparts. In this table each cell contains two values of percentage of risk improvements. These two values correspond to one value of τ. In a very similar fashion the percentage of risk improvements of all the estimators have been presented in Tables 3 to 5 for equal and unequal sample sizes with respect to the loss functions (1.2) and (1.3).

The percentage of relative risk improvements of all the improved estimators with respect to the benchmark estimator d_{GD} (denoted as Ri; $i = 1, 2, 7$) have been presented in Tables 6 and 7 for equal and unequal sample sizes respectively. Specifically, in Table 6 we have presented the percentage of relative risk performances of all the improved estimators for the sample sizes $(5, 5)$, $(12, 12)$ and $(20, 20)$. The Table 6 consists of eight columns and each column have several cells. Corresponding to each value of τ there correspond three values of percentage of relative risks. These three values correspond to three sample sizes $(5, 5)$, $(12, 12)$ and $(20, 20)$ respectively. In a very similar way we have presented the percentage of relative risk improvements of all the improved estimators for the unequal sample sizes $(5, 10)$, $(12, 20)$, $(10, 5)$ and $(20, 12)$ in Table 7. Moreover, we have also plotted the risk values of all the improved estimators with respect to the loss function (1.1), against the choices of τ in Figure 1. Specifically, Figure 1 (a)–(b) presents the graph for equal sample sizes whereas Figure 1 (c)–(f) presents for unequal sample sizes. We note that the estimators $\hat{d}_{GD}$, $\hat{d}_{KS}$, $\hat{d}_{MK}$, $\hat{d}_{TK}$, $\hat{d}_{RM}$, d^a_{GD}, d^a_{MK}, d^a_{BC1}, d^a_{BC2} have been denoted by GDI, KSI, MKI, TKI, RML, GDA, MKA, BC1A and BC2A respectively in Figure 1 (a)–(f).

Table 1: Percentage of risk improvements of all the proposed estimators using the loss L_1 for the sample sizes $(m, n) = (5, 5), (12, 12), (20, 20), (30, 30)$.

$\tau \downarrow$	$P1$	$P5$	$P3$	$P6$	$P9$	$\tau \downarrow$	$P1$	$P5$	$P3$	$P6$	$P9$
0.05	1.93	1.17	0.72	0.41	17.78	0.55	9.76	8.91	5.93	4.43	11.82
	0.00	0.00	0.00	0.00	1.63		3.85	2.77	2.06	1.27	3.53
	0.00	0.00	0.00	0.00	0.00		1.42	0.94	0.73	0.42	1.04
	0.00	0.00	0.00	0.00	0.00		0.43	0.29	0.22	0.13	0.31
0.10	5.66	3.53	2.56	1.46	13.90	0.60	8.93	9.13	5.42	4.42	11.77
	0.00	0.00	0.00	0.00	0.31		3.95	2.92	2.11	1.33	3.55
	0.00	0.00	0.00	0.00	0.00		1.50	1.09	0.77	0.48	1.23
	0.00	0.00	0.00	0.00	0.00		0.83	0.53	0.42	0.24	0.57
0.15	6.52	4.19	3.19	1.87	11.42	0.65	8.46	9.08	5.14	4.34	11.39
	0.16	0.09	0.07	0.04	0.46		3.32	2.92	1.79	1.26	3.61
	0.00	0.00	0.00	0.00	0.00		1.77	1.38	0.92	0.59	1.55
	0.00	0.00	0.00	0.00	0.00		1.01	0.70	0.52	0.31	0.75
0.20	7.88	5.24	4.09	2.44	10.55	0.70	7.43	9.14	4.62	4.32	11.39
	0.51	0.30	0.23	0.13	0.67		4.07	3.59	2.23	1.57	4.32
	0.02	0.01	0.01	0.01	0.01		2.12	1.70	1.10	0.73	1.92
	0.00	0.00	0.00	0.00	0.00		1.30	0.93	0.67	0.41	0.99
0.25	8.30	5.65	4.49	2.73	10.63	0.75	6.29	9.04	3.80	4.14	11.61
	0.49	0.30	0.23	0.13	0.47		3.22	3.41	1.76	1.41	4.22
	0.06	0.04	0.03	0.02	0.04		2.09	1.88	1.10	0.78	2.11
	0.01	0.00	0.00	0.00	0.00		1.04	0.98	0.53	0.39	1.05
0.30	9.23	6.50	5.14	3.18	10.28	0.80	6.64	9.82	4.14	4.55	11.69
	0.88	0.53	0.43	0.24	0.80		2.96	3.81	1.61	1.49	4.80
	0.15	0.09	0.07	0.04	0.11		2.30	2.27	1.22	0.92	2.57
	0.07	0.04	0.03	0.02	0.04		1.51	1.31	0.78	0.54	1.4
0.35	10.02	7.35	5.70	3.63	11.26	0.85	3.39	8.52	2.11	3.65	10.86
	1.57	0.99	0.79	0.45	1.48		2.86	4.15	1.61	1.61	5.05
	0.26	0.15	0.13	0.07	0.19		2.21	2.57	1.17	0.99	2.92
	0.05	0.03	0.02	0.01	0.03		1.67	1.67	0.87	0.66	1.81
0.40	9.70	7.42	5.60	3.68	11.00	0.90	3.38	9.01	2.07	3.80	11.72
	2.11	1.35	1.07	0.62	1.81		1.47	3.65	0.80	1.26	4.67
	0.52	0.32	0.26	0.14	0.37		1.35	2.31	0.72	0.81	2.64
	0.04	0.02	0.02	0.01	0.02		1.25	1.66	0.65	0.60	1.79
0.45	10.91	8.78	6.49	4.41	12.00	0.95	1.82	8.51	1.19	3.53	10.70
	2.35	1.53	1.22	0.70	1.99		0.81	3.62	0.47	1.20	4.48
	0.55	0.36	0.27	0.16	0.41		1.27	2.60	0.69	0.89	2.94
	0.19	0.11	0.09	0.05	0.12		0.05	1.32	0.02	0.36	1.46
0.50	10.17	8.62	6.02	4.25	11.23	1.00	0.68	7.72	0.49	2.83	9.87
	3.47	2.36	1.82	1.08	3.06		0.55	3.34	0.31	0.93	4.21
	1.11	0.73	0.56	0.33	0.85		0.21	2.23	0.11	0.59	2.55
	0.32	0.21	0.16	0.09	0.23		0.78	2.09	0.42	0.67	2.29

Table 2: Percentage of risk improvements of all the proposed estimators using the loss L_1 for unequal sample sizes.

$\tau \downarrow$	$(m,n) = (5,10),(12,20)$					$(m,n) = (10,5),(20,12)$				
	$P1$	$P2$	$P3$	$P4$	$P9$	$P1$	$P2$	$P3$	$P4$	$P9$
0.05	0.00	0.01	0.00	0.00	0.02	2.18	0.89	2.19	1.98	43.23
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.86
0.10	0.12	0.47	0.00	0.00	0.26	4.21	1.94	4.54	4.12	39.23
	0.00	0.00	0.00	0.00	0.05	0.02	0.00	0.10	0.09	0.83
0.15	0.18	0.92	0.01	0.01	0.68	5.79	2.89	6.20	5.65	32.09
	0.00	0.01	0.00	0.00	0.03	0.02	0.01	0.19	0.18	0.73
0.20	0.61	1.80	0.06	0.07	1.26	6.32	3.12	7.25	6.60	26.89
	0.01	0.03	0.00	0.00	0.05	0.25	0.18	0.53	0.50	0.34
0.25	0.72	2.34	0.06	0.07	1.45	7.70	3.77	8.79	8.05	25.31
	0.04	0.07	0.00	0.00	0.11	0.43	0.30	1.06	1.00	0.80
0.30	1.06	3.18	0.08	0.10	2.17	8.27	4.10	9.36	8.56	19.80
	0.12	0.19	0.00	0.00	0.22	0.56	0.40	1.35	1.27	0.44
0.35	1.47	4.01	0.09	0.11	2.59	9.66	4.86	10.33	9.50	20.65
	0.22	0.34	0.00	0.00	0.41	1.01	0.76	2.03	1.93	0.97
0.40	1.74	4.60	0.13	0.16	2.83	10.28	5.06	11.21	10.32	19.58
	0.39	0.58	0.01	0.01	0.72	1.34	0.99	2.65	2.52	1.37
0.45	2.06	5.09	0.18	0.22	3.41	10.6	5.31	11.32	10.44	17.64
	0.55	0.81	0.01	0.02	0.98	1.77	1.30	3.30	3.15	1.51
0.50	2.56	6.05	0.21	0.26	4.08	11.26	5.93	11.62	10.75	17.39
	0.77	1.12	0.03	0.03	1.15	1.91	1.41	3.41	3.26	1.61
0.55	2.66	6.18	0.25	0.31	4.02	11.91	6.15	11.91	11.04	18.19
	1.08	1.53	0.05	0.06	1.66	2.31	1.71	3.82	3.65	1.97
0.60	2.69	6.22	0.25	0.30	4.07	11.77	6.11	11.73	10.88	17.98
	1.26	1.75	0.04	0.05	1.84	2.39	1.78	3.91	3.74	2.12
0.65	2.98	6.52	0.31	0.37	4.49	12.84	6.70	12.26	11.39	17.61
	1.36	1.88	0.06	0.07	2.01	3.24	2.49	4.56	4.38	2.95
0.70	2.86	6.28	0.31	0.36	4.21	12.64	6.65	11.67	10.83	16.90
	1.78	2.35	0.11	0.13	2.62	3.55	2.72	4.79	4.61	3.25
0.75	3.21	6.75	0.30	0.36	4.72	12.37	6.56	11.15	10.37	16.05
	1.93	2.53	0.12	0.13	2.66	3.41	2.63	4.37	4.20	3.27
0.80	3.75	7.44	0.46	0.54	5.39	12.47	6.46	11.25	10.45	15.33
	2.06	2.73	0.10	0.10	2.79	3.97	3.10	4.58	4.42	3.80
0.85	3.31	6.95	0.32	0.38	4.81	11.99	6.54	9.93	9.23	15.25
	2.43	3.09	0.18	0.19	3.30	4.00	3.10	4.51	4.34	3.83
0.90	3.41	6.87	0.39	0.45	4.89	11.40	5.77	9.40	8.70	14.64
	2.33	2.96	0.18	0.19	3.26	4.16	3.21	4.46	4.29	4.01
0.95	3.33	6.78	0.29	0.35	4.96	11.83	6.45	9.25	8.58	14.60
	2.26	2.86	0.15	0.16	3.02	3.95	3.05	3.90	3.76	3.91
1.00	2.99	6.27	0.24	0.28	4.32	11.40	6.12	8.47	7.85	13.82
	1.94	2.50	0.07	0.08	2.83	4.09	3.23	3.52	3.38	4.09

Table 3: Percentage of risk improvements of all the proposed estimators using the loss L_2 and L_3 loss.

$(m,n)\downarrow$	$(\sigma_1^2,\sigma_2^2)\downarrow$	L_2- Loss					L_3- Loss				
		$P1$	$P5$	$P3$	$P6$	$P9$	$P1$	$P5$	$P3$	$P6$	$P9$
(5, 5)	(0.05, 0.10)	4.62	3.96	2.81	1.99	5.37	10.73	9.21	6.55	4.69	11.72
	(0.05, 0.30)	2.27	1.33	1.20	0.65	3.28	6.48	4.25	3.23	1.91	11.12
	(0.05, 0.50)	1.26	0.70	0.63	0.33	3.06	3.94	2.41	1.72	0.97	13.05
	(0.05, 0.70)	0.77	0.42	0.36	0.19	3.54	3.59	2.22	1.48	0.83	16.24
	(0.05, 1.00)	0.61	0.32	0.27	0.14	3.27	2.79	1.75	1.07	0.61	19.27
	(1.00, 1.10)	1.19	4.25	0.76	1.78	5.45	1.99	8.39	1.21	3.43	10.50
	(1.00, 1.50)	3.71	4.17	2.27	1.96	5.45	8.26	9.29	5.10	4.45	11.27
	(1.00, 2.00)	4.53	3.81	2.74	1.89	5.15	10.33	8.67	6.21	4.37	11.48
	(1.00, 2.50)	4.71	3.43	2.76	1.74	4.91	9.80	7.52	5.66	3.72	10.99
	(1.00, 3.00)	4.31	2.91	2.48	1.48	4.35	9.40	6.78	5.28	3.34	10.71
	(2.00, 2.10)	0.21	3.94	0.18	1.43	5.03	1.61	8.57	0.98	3.49	10.69
	(2.00, 2.30)	1.43	4.41	0.92	1.86	5.64	3.37	8.83	2.06	3.75	11.11
	(2.00, 2.50)	3.16	4.87	2.06	2.25	6.19	6.11	9.37	3.86	4.33	12.05
	(2.00, 2.70)	3.06	4.37	1.87	1.99	5.64	6.34	8.90	3.96	4.13	11.08
	(2.00, 3.00)	4.26	4.39	2.67	2.14	5.74	8.38	9.18	5.21	4.46	11.75
(12, 12)	(0.05, 0.10)	1.56	1.04	0.84	0.49	1.24	2.61	1.90	1.37	0.85	2.49
	(0.05, 0.30)	0.04	0.02	0.02	0.01	0.09	0.17	0.09	0.07	0.04	0.26
	(0.05, 0.50)	0.01	0.00	0.00	0.00	0.04	0.01	0.00	0.00	0.00	0.48
	(0.05, 0.70)	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.29
	(0.05, 1.00)	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.38
	(1.00, 1.10)	1.39	2.19	0.83	0.88	2.65	0.98	3.51	0.53	1.16	4.47
	(1.00, 1.50)	1.99	1.68	1.12	0.76	2.01	3.94	3.35	2.13	1.47	4.21
	(1.00, 2.00)	1.41	0.98	0.78	0.46	1.19	2.77	1.98	1.46	0.89	2.59
	(1.00, 2.50)	0.89	0.59	0.46	0.27	0.75	1.73	1.16	0.88	0.52	1.50
	(1.00, 3.00)	0.61	0.36	0.31	0.17	0.47	1.35	0.82	0.67	0.37	1.18
	(2.00, 2.10)	0.26	1.84	0.15	0.57	2.31	1.37	3.98	0.76	1.37	4.99
	(2.00, 2.30)	0.92	1.80	0.46	0.63	2.22	1.80	3.61	1.00	1.31	4.45
	(2.00, 2.50)	1.18	1.74	0.64	0.66	2.13	3.37	4.06	1.86	1.64	4.95
	(2.00, 2.70)	2.08	1.93	1.15	0.83	2.33	4.04	3.90	2.21	1.66	4.77
	(2.00, 3.00)	1.85	1.55	1.00	0.69	1.87	3.78	3.30	2.04	1.43	4.16
(20, 20)	(0.05, 0.10)	0.56	0.34	0.29	0.16	0.38	1.12	0.72	0.57	0.33	0.81
	(0.05, 0.30)	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.01
	(0.05, 0.50)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.05, 0.70)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.05, 1.00)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(1.00, 1.10)	0.34	1.00	0.20	0.34	1.14	1.23	2.37	0.65	0.81	2.74
	(1.00, 1.50)	0.95	0.72	0.48	0.31	0.81	1.97	1.51	1.03	0.65	1.67
	(1.00, 2.00)	0.62	0.37	0.33	0.18	0.41	1.10	0.69	0.56	0.32	0.79
	(1.00, 2.50)	0.21	0.12	0.10	0.05	0.14	0.42	0.25	0.21	0.11	0.28
	(1.00, 3.00)	0.12	0.06	0.06	0.03	0.07	0.28	0.16	0.14	0.07	0.20
	(2.00, 2.10)	0.23	1.14	0.15	0.37	1.30	0.40	2.21	0.21	0.67	2.59
	(2.00, 2.30)	0.95	1.22	0.47	0.44	1.40	1.52	2.21	0.80	0.80	2.52
	(2.00, 2.50)	1.09	1.11	0.59	0.46	1.23	2.01	2.05	1.06	0.81	2.29
	(2.00, 2.70)	1.03	0.92	0.53	0.37	1.03	2.45	2.07	1.29	0.88	2.31
	(2.00, 3.00)	0.80	0.65	0.41	0.27	0.74	1.79	1.39	0.93	0.60	1.55

Table 4: Percentage of risk improvements of all the proposed estimators using the loss L_2 and L_3.

$(m,n)\downarrow$	$(\sigma_1^2,\sigma_2^2)\downarrow$	L_2- Loss					L_3- Loss				
		$P1$	$P2$	$P3$	$P4$	$P9$	$P1$	$P2$	$P3$	$P4$	$P9$
(5, 10)	(0.05, 0.10)	1.27	2.84	0.12	0.14	1.87	2.46	5.73	0.18	0.22	3.86
	(0.05, 0.30)	0.11	0.42	0.01	0.01	0.25	0.29	1.10	0.02	0.02	0.72
	(0.05, 0.50)	0.03	0.13	0.00	0.00	0.08	0.10	0.39	0.01	0.01	0.28
	(0.05, 0.70)	0.01	0.06	0.00	0.00	0.04	0.04	0.15	0.00	0.00	0.08
	(0.05, 1.00)	0.00	0.02	0.00	0.00	0.01	0.00	0.04	0.00	0.00	0.05
	(1.00, 1.10)	1.44	2.97	0.13	0.15	2.15	3.52	6.91	0.38	0.45	4.94
	(1.00, 1.50)	1.59	3.45	0.17	0.20	2.37	3.21	6.94	0.32	0.39	4.86
	(1.00, 2.00)	1.21	2.89	0.09	0.11	1.92	2.56	6.03	0.21	0.26	4.17
	(1.00, 2.50)	0.81	2.21	0.06	0.07	1.42	1.92	4.95	0.15	0.19	3.29
	(1.00, 3.00)	0.59	1.71	0.04	0.05	1.04	1.22	3.55	0.08	0.10	2.43
	(2.00, 2.10)	1.73	3.37	0.21	0.24	2.52	3.00	6.39	0.20	0.25	4.43
	(2.00, 2.30)	1.75	3.56	0.17	0.20	2.52	3.50	7.07	0.36	0.42	4.90
	(2.00, 2.50)	1.67	3.47	0.17	0.20	2.47	3.36	7.03	0.32	0.39	5.13
	(2.00, 2.70)	1.62	3.50	0.15	0.18	2.46	3.20	6.96	0.32	0.38	4.73
	(2.00, 3.00)	1.68	3.60	0.18	0.21	2.51	3.15	6.83	0.32	0.38	4.98
(12,20)	(0.05, 0.10)	0.45	0.64	0.02	0.03	0.68	0.77	1.13	0.01	0.01	1.24
	(0.05, 0.30)	0.00	0.01	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.02
	(0.05, 0.50)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.05, 0.70)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.05, 1.00)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(1.00, 1.10)	0.96	1.26	0.04	0.04	1.32	2.28	2.93	0.16	0.17	3.24
	(1.00, 1.50)	0.77	1.05	0.03	0.04	1.10	1.66	2.25	0.10	0.10	2.36
	(1.00, 2.00)	0.37	0.53	0.01	0.02	0.56	0.74	1.05	0.03	0.03	1.12
	(1.00, 2.50)	0.18	0.27	0.00	0.00	0.28	0.35	0.53	0.00	0.01	0.61
	(1.00, 3.00)	0.08	0.13	0.00	0.00	0.17	0.17	0.29	0.00	0.00	0.34
	(2.00, 2.10)	1.14	1.45	0.07	0.07	1.51	2.13	2.69	0.16	0.17	2.86
	(2.00, 2.30)	1.19	1.52	0.08	0.08	1.60	2.42	3.08	0.21	0.22	3.21
	(2.00, 2.50)	0.88	1.18	0.05	0.05	1.20	1.78	2.37	0.12	0.13	2.58
	(2.00, 2.70)	0.82	1.08	0.06	0.06	1.14	2.06	2.69	0.16	0.17	2.88
	(2.00, 3.00)	0.76	1.01	0.05	0.05	1.06	1.59	2.12	0.10	0.11	2.27

Table 5: Percentage of risk improvements of all the proposed estimators using the loss L_2 and L_3.

$(m,n)\downarrow$	$(\sigma_1^2,\sigma_2^2)\downarrow$	L_2-Loss					L_3-Loss				
		$P1$	$P2$	$P3$	$P4$	$P9$	$P1$	$P2$	$P3$	$P4$	$P9$
(10, 5)	(0.05, 0.10)	4.70	2.27	5.51	5.07	6.73	10.89	5.43	11.52	10.63	17.75
	(0.05, 0.30)	1.79	0.81	2.49	2.24	6.69	5.34	2.37	6.24	5.64	28.01
	(0.05, 0.50)	1.08	0.53	1.40	1.26	7.27	3.51	1.53	3.80	3.43	31.23
	(0.05, 0.70)	0.71	0.30	0.96	0.86	7.45	3.48	1.58	3.29	2.99	40.10
	(0.05, 1.00)	0.43	0.21	0.60	0.54	6.71	1.83	0.63	1.97	1.77	43.66
	(1.00, 1.10)	5.78	3.04	4.89	4.51	7.06	12.93	7.12	10.35	9.65	16.21
	(1.00, 1.50)	5.57	2.92	5.52	5.10	7.09	12.55	6.61	11.90	11.04	16.55
	(1.00, 2.00)	4.88	2.44	5.66	5.21	6.60	11.60	5.89	12.07	11.16	19.34
	(1.00, 2.50)	4.30	2.07	5.24	4.80	7.21	10.41	5.32	11.19	10.31	19.67
	(1.00, 3.00)	3.42	1.58	4.52	4.11	6.99	9.87	5.14	10.47	9.67	20.79
	(2.00, 2.10)	6.24	3.44	5.05	4.69	7.48	11.78	6.31	9.13	8.44	15.06
	(2.00, 2.30)	6.06	3.21	5.19	4.83	7.05	12.13	6.47	10.05	9.32	15.78
	(2.00, 2.50)	6.11	3.30	5.59	5.19	7.23	12.55	6.60	10.90	10.13	15.61
	(2.00, 2.70)	6.13	3.24	5.82	5.41	7.43	13.47	7.47	11.94	11.14	16.27
	(2.00, 3.00)	5.94	3.16	5.96	5.52	7.40	12.78	6.70	12.05	11.19	17.28
(20,12)	(0.05, 0.10)	0.88	0.64	1.71	1.63	0.74	1.91	1.42	3.46	3.29	1.71
	(0.05, 0.30)	0.03	0.02	0.10	0.10	0.11	0.03	0.22	0.20	0.00	1.16
	(0.05, 0.50)	0.00	0.00	0.02	0.02	0.06	0.00	0.00	0.03	0.03	2.40
	(0.05, 0.70)	0.00	0.00	0.00	0.00	0.04	0.01	0.00	0.02	0.02	1.01
	(0.05, 1.00)	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	1.66
	(1.00, 1.10)	2.08	1.64	2.19	2.12	2.04	4.26	3.38	4.20	4.06	4.16
	(1.00, 1.50)	1.71	1.31	2.47	2.38	1.55	3.02	2.30	4.33	4.16	2.72
	(1.00, 2.00)	0.91	0.68	1.73	1.65	0.74	2.03	1.51	3.62	3.46	1.75
	(1.00, 2.50)	0.53	0.39	1.18	1.12	0.43	1.27	0.93	2.61	2.48	1.13
	(1.00, 3.00)	0.34	0.24	0.83	0.78	0.28	0.64	0.43	1.67	1.58	0.50
	(2.00, 2.10)	1.71	1.26	1.72	1.65	1.65	3.90	3.06	3.70	3.56	3.89
	(2.00, 2.30)	1.80	1.40	1.98	1.91	1.74	3.83	2.97	4.11	3.96	3.83
	(2.00, 2.50)	1.80	1.40	2.20	2.12	1.66	3.52	2.69	4.35	4.18	3.27
	(2.00, 2.70)	1.74	1.33	2.38	2.29	1.62	3.63	2.80	4.68	4.51	3.47
	(2.00, 3.00)	1.64	1.25	2.39	2.30	1.45	3.53	2.73	4.80	4.62	3.34

Table 6: Percentage of relative risk improvements of all the proposed estimators using the loss L_1 for equal sample sizes.

$\tau \downarrow$	$(m,n) = (5,5),(12,12),(20,20)$						
	$R1$	$R5$	$R3$	$R6$	$R7$	$R8$	$R9$
0.05	1.93	1.17	−38.84	−39.28	13.40	12.32	−18.31
	0.00	0.00	−44.28	−44.28	1.63	1.52	−0.52
	0.00	0.00	−41.76	−41.76	0.31	0.29	0.26
0.15	6.52	4.19	−5.86	−7.31	11.43	7.96	−0.74
	0.16	0.09	−17.8	−17.84	2.18	1.85	0.76
	0.00	0.00	−19.5	−19.5	0.80	0.65	0.55
0.25	8.30	5.65	4.14	2.37	6.41	0.82	2.75
	0.49	0.30	−6.49	−6.60	−0.05	−0.59	0.26
	0.06	0.04	−8.94	−8.96	0.00	−0.12	0.15
0.35	10.02	7.35	10.34	8.37	−0.03	−8.24	4.55
	1.57	0.99	−1.19	−1.54	−0.14	−0.64	0.46
	0.26	0.15	−3.54	−3.59	−1.15	−1.15	0.07
0.45	10.91	8.78	13.07	11.15	−3.63	−13.69	6.22
	2.35	1.53	2.36	1.85	−3.04	−3.62	0.75
	0.55	0.36	−0.72	−0.83	−2.12	−1.98	0.06
0.55	9.76	8.91	14.54	13.18	−8.86	−20.58	6.29
	3.85	2.77	5.09	4.32	−4.05	−4.93	1.80
	1.42	0.94	1.64	1.34	−2.43	−2.08	0.53
0.65	8.46	9.08	14.31	13.59	−13.9	−27.3	6.72
	3.32	2.92	6.02	5.51	−7.72	−8.41	1.77
	1.77	1.38	2.88	2.56	−3.81	−3.38	0.92
0.75	6.29	9.04	13.72	14.02	−21.68	−37.8	6.07
	3.22	3.41	6.31	5.98	−9.55	−9.94	2.31
	2.09	1.88	3.85	3.53	−5.45	−4.59	1.40
0.85	3.39	8.52	12.52	13.89	−26.94	−44.56	5.66
	2.86	4.15	6.75	6.76	−11.21	−11.56	3.07
	2.21	2.57	4.70	4.53	−6.40	−5.48	2.09
0.95	1.82	8.51	12.45	14.52	−31.13	−50.45	5.42
	0.81	3.62	6.04	6.73	−15.23	−15.35	2.32
	1.27	2.60	4.20	4.40	−8.28	−7.02	2.20
1.00	0.68	7.72	10.26	13.22	−35.40	−55.92	4.94
	0.55	3.34	4.97	6.14	−17.04	−16.80	2.14
	0.21	2.23	3.34	4.02	−8.87	−7.50	1.83

Table 7: Percentage of relative risk improvements of all the proposed estimators using the loss L_1 for unequal sample sizes.

$\tau \downarrow$	$(m,n) = (5,10),(12,20),(10,5),(20,12)$						
	$R1$	$R2$	$R3$	$R4$	$R7$	$R8$	$R9$
0.05	0.00	−8.4	−23.32	−25.50	3.46	3.08	−1.69
	0.00	−0.53	−30.31	−31.00	0.79	0.67	0.54
	2.18	6.68	−63.88	−59.61	10.25	9.64	−4.87
	0.00	0.34	−57.97	−56.84	1.18	0.97	0.68
0.15	0.18	−6.53	−0.40	−1.37	0.01	−0.93	−2.03
	0.00	−0.71	−8.10	−8.45	0.13	0.00	−0.13
	5.79	9.87	−22.3	−20.38	13.2	11.27	5.33
	0.02	0.57	−30.61	−29.93	1.57	1.27	1.01
0.25	0.72	−2.63	5.58	5.09	−6.48	−7.87	−2.44
	0.04	−0.43	−0.79	−0.99	−2.23	−2.29	−0.66
	7.70	10.26	−6.39	−5.35	11.3	7.81	8.14
	0.43	0.98	−17.05	−16.61	1.79	1.42	1.25
0.35	1.47	1.49	8.63	8.48	−13.83	−15.54	−1.73
	0.22	0.01	2.81	2.70	−4.38	−4.29	−0.84
	9.66	10.74	2.29	2.88	9.38	4.71	9.32
	1.01	1.36	−8.73	−8.46	0.84	0.29	1.45
0.45	2.06	3.77	9.03	9.09	−22.4	−24.25	−1.3
	0.55	0.59	4.21	4.15	−6.82	−6.24	−0.50
	10.6	10.59	7.92	8.23	7.92	1.97	10.32
	1.77	2.04	−3.95	−3.77	1.31	0.78	1.98
0.55	2.66	5.76	9.55	9.74	−27.42	−29.41	−0.41
	1.08	1.52	5.02	5.02	−10.3	−9.21	0.26
	11.91	11.00	11.17	11.35	6.33	−0.93	11.03
	2.31	2.24	0.27	0.35	−0.60	−1.21	1.99
0.65	2.98	6.84	9.14	9.47	−35.76	−37.23	−0.34
	1.36	1.97	4.65	4.69	−13.33	−12.04	0.67
	12.84	11.37	13.59	13.68	4.43	−3.84	11.46
	3.24	3.00	3.15	3.20	−1.05	−1.75	2.65
0.75	3.21	7.98	8.47	8.93	−43.11	−44.95	−0.14
	1.93	2.67	4.58	4.65	−15.73	−13.78	1.42
	12.37	10.08	14.27	14.29	0.59	−9.52	10.97
	3.41	3.00	4.61	4.63	−2.65	−3.39	2.65
0.85	3.31	8.48	7.91	8.45	−51.29	−52.66	−0.84
	2.43	3.20	3.95	4.04	−17.53	−15.15	2.18
	11.99	9.02	14.76	14.74	−3.30	−14.75	10.02
	4.00	3.47	5.95	5.94	−3.31	−3.93	3.14
0.95	3.33	9.10	7.54	8.17	−56.73	−58.04	−0.47
	2.26	3.14	3.25	3.38	−21.16	−18.37	2.10
	11.83	8.84	14.60	14.56	−4.47	−16.63	10.04
	3.95	3.31	6.44	6.42	−4.8	−5.49	2.92
1.00	2.99	9.03	6.59	7.28	−62.71	−63.43	−1.96
	1.94	2.84	2.52	2.65	−24.06	−21.01	1.91
	11.4	8.07	14.64	14.59	−5.9	−19.2	9.47
	4.09	3.55	5.81	5.81	−4.87	−5.47	3.18

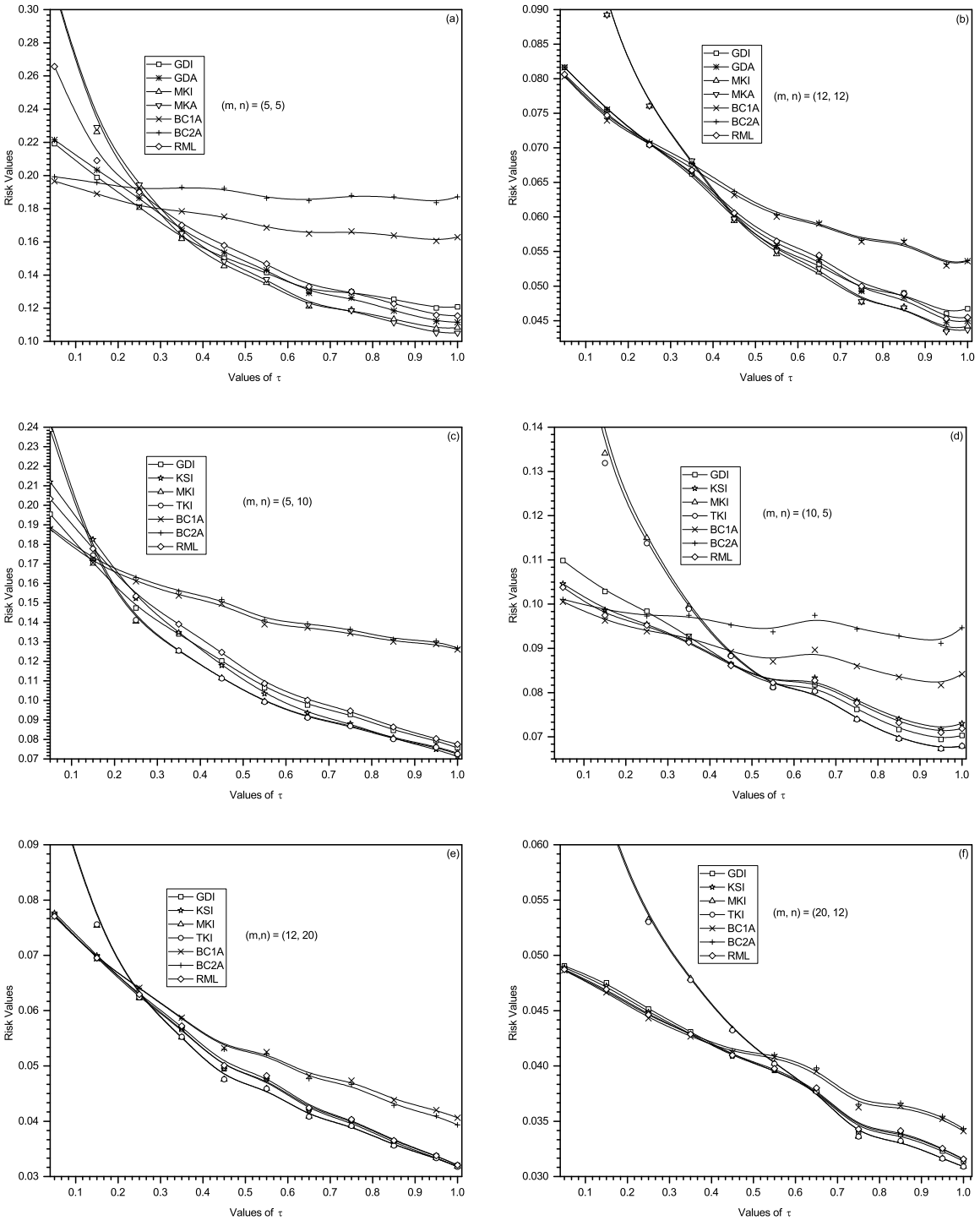


Figure 1: (a)–(f) Comparison of risk values of several estimators for common mean μ using the loss L_1 for sample sizes $(5, 5)$, $(12, 12)$, $(5, 10)$, $(10, 5)$, $(12, 20)$ and $(20, 12)$ respectively.

The following observations have been made during our simulation study as well as from the tables, which we discuss separately for equal and unequal sample sizes.

Case I: $m = n$.

1. The percentage of risk improvements as well as the risk values of all the new estimators upon their respective unrestricted counterparts decreases as the sample sizes increase for fixed values of the parameter, with respect to the loss functions L_1, L_2 and L_3.

2. Let the loss function be L_1. The percentage of risk improvement of $\hat{d}_{GD}$ (see $P1$) is seen maximum up to 12%. The percentage of risk improvement of d^a_{GD} (see $P5$) is seen maximum up to 10%. The percentage of risk improvement of $\hat{d}_{MK}$ (see $P3$) is seen maximum up to 7%. The percentage of risk improvement of d^a_{MK} (see $P6$) is seen maximum up to 6%, where the percentage of risk improvement of $\hat{d}_{RM}$ (see $P9$) is seen maximum up to 20%.

3. Let the loss function be L_2. The maximum percentage of risk improvement of $\hat{d}_{GD}$, d^a_{GD}, $\hat{d}_{MK}$, d^a_{MK}, and $\hat{d}_{RM}$ upon their respective unrestricted counterparts are seen near to 6%, 5%, 4%, 3% and 7% respectively. The maximum percentage of risk improvement is seen in the case of $\hat{d}_{RM}$ for small sample sizes and when σ_1^2 and σ_2^2 are close to each other.

4. Let the loss function be L_3. The maximum percentage of risk improvement of $\hat{d}_{GD}$, d^a_{GD}, $\hat{d}_{MK}$, d^a_{MK}, and $\hat{d}_{RM}$ upon their respective unrestricted counterparts are seen respectively as 11%, 10%, 7%, 5% and 20%. The maximum percentage of risk improvement of each of the estimators has been noticed for small sample sizes and when the variances are close to each other.

5. Here we note that, the percentage of risk improvements of all the new estimators upon their respective unrestricted counterparts are approximated values only which have been obtained numerically and hence it may vary with sample sizes.

6. The above numerical results (2)–(4) validates the theoretical findings in Sections 3, 4, and 5.

7. The risk values of all the estimators such as $\hat{d}_{GD}$, d^a_{GD}, $\hat{d}_{MK}$, d^a_{MK}, d^a_{BC1}, d^a_{BC2}, and $\hat{d}_{RM}$, decrease as the sample sizes increase. Further for fixed sample sizes, as the values of τ varies from 0 to 1, the risk values of all the estimators decrease. It has been noticed that, for small values of τ (say $0 < \tau < 0.25$), the percentage of relative risk improvement of ${d^a}_{BC1}$ is maximum and seen up to 15%. For the values of τ near to 1, (say for the range $0.50 < \tau < 1$) the estimators $\hat{d}_{MK}$ and d^a_{MK} have almost same percentage of relative risk improvements. For moderate values of τ (say $0.50 < \tau < 0.75$), the estimators $\hat{d}_{MK}$ and d^a_{MK} perform equally well, however as the sample sizes increases from moderate to large, the performance of these two estimators decrease and compete well with $\hat{d}_{GD}$. In fact, the dominance regions of $\hat{d}_{MK}$ and d^a_{MK} upon $\hat{d}_{GD}$ decrease. It has also been noticed that the estimators $\hat{d}_{GD}$, $\hat{d}_{MK}$, and d^a_{BC1} compete with d^a_{GD}, d^a_{MK} and d^a_{BC2} respectively.

Case II: $m \neq n$.

1. The percentage of risk improvements of all the improved estimators decrease as the sample sizes increase for fixed values of σ_1^2 and σ_2^2 with respect to the loss functions L_1, L_2 and L_3.

2. Let us first consider the loss function L_1. The percentage of risk improvement of $\hat{d}_{GD}$ upon d_{GD} (denoted as $P1$) is seen maximum up to 16%, the maximum percentage of risk improvement of $\hat{d}_{KS}$ upon d_{KS} (denoted as $P2$) is seen near to 8%. The maximum percentage of risk improvement of $\hat{d}_{MK}$ and $\hat{d}_{TK}$ over their corresponding unrestricted counterparts are seen near to 14% and 13% respectively. The maximum percentage of risk improvement of $\hat{d}_{RM}$ upon d_{ML} is seen up to 15%. We also note that, these maximum risk improvements have been noticed when $m > n$ for all the estimators.

3. Let us consider the loss function L_2. The maximum percentage of risk improvement of $\hat{d}_{GD}$ upon d_{GD} is seen up to 7%. The maximum percentage of risk improvement of $\hat{d}_{KS}$ over d_{KS} is seen near to 4%. The maximum percentage of risk improvement of $\hat{d}_{MK}$ upon d_{MK} is seen near to 7%. The maximum percentage of risk improvement of $\hat{d}_{TK}$ upon d_{TK} is seen near to 7%. The maximum percentage of risk improvement of $\hat{d}_{RM}$ upon d_{ML} is seen near to 13%.

4. Consider the loss function L_3. The maximum percentage of risk improvement of $\hat{d}_{GD}$ upon d_{GD} is seen up to 13%. The maximum percentage of risk improvement of $\hat{d}_{KS}$ upon d_{KS} is seen near to 8%. The maximum percentage of risk improvement of $\hat{d}_{MK}$ upon d_{MK} is seen near to 13%. The maximum percentage of risk improvement of $\hat{d}_{TK}$ upon d_{TK} is seen near to 13%. The maximum percentage of risk improvement of $\hat{d}_{RM}$ upon d_{ML} is seen near to 36%.

5. Here we note that, the percentage of risk improvements of all the improved estimators upon their respective unrestricted counterparts are approximated values only which have been obtained numerically and hence it may vary with sample sizes, however the trends remain the same.

6. The above numerical results $(2) - -(4)$ also validates the theoretical findings in Sections 3, 4, and 5.

7. The risk values of all the estimators, such as $\hat{d}_{GD}$, $\hat{d}_{KS}$, $\hat{d}_{MK}$, $\hat{d}_{TK}$, d^a_{BC1}, d^a_{BC2}, and $\hat{d}_{RM}$, decrease as the sample sizes increase. It has been noticed that for small values of τ (say $0 < \tau < 0.15$), the percentage of relative risk improvements of d^a_{BC1} and d^a_{BC2} are maximum and seen up to 12%. For the values of τ near to 1, (say $0.75 < \tau < 1$) the estimator $\hat{d}_{KS}$ (for $m < n$) and $\hat{d}_{MK}$, $\hat{d}_{TK}$ (when $m > n$) has maximum percentage of relative risk improvements. For moderate values of τ, the estimators $\hat{d}_{MK}$ and $\hat{d}_{TK}$ perform equally well, however as the sample sizes increase from moderate to large the performance of these two estimators decrease and in this case the estimators $\hat{d}_{GD}$ and $\hat{d}_{KS}$ perform better.

From the above discussions and also from our simulation study the following conclusions can be drawn regarding the use of the proposed estimators in practice:

1. Let us consider that the sample sizes are equal, that is $m = n$. When the variance of the first population is much smaller compare to the second, we recommend to use d^a_{BC1}. When the variance of both the populations are close to each other, we recommend to use either $\hat{d}_{MK}$ or d^a_{MK}, as they compete with each other. In other cases, that is neither the variances differ too much nor close to each other, the estimators $\hat{d}_{MK}$ and d^a_{MK} can be used for small sample sizes (say $m, n \leq 10$), and $\hat{d}_{MK}$ or $\hat{d}_{GD}$ for moderate to large sample sizes.

2. Consider that the sample sizes are unequal, that is $m \neq n$. When the variance of the first population is much smaller than the second, we recommend to use either the estimator d^a_{BC1} or d^a_{BC2}. When the variances of both the populations are close to each other, the estimators $\hat{d}_{KS}$ or $\hat{d}_{TK}$ (for $m < n$) and $\hat{d}_{TK}$ or $\hat{d}_{MK}$ (for $m > n$) can be recommended for use. However for moderate ranges of τ, the estimators $\hat{d}_{MK}$ or $\hat{d}_{TK}$ (for $m < n$) and the estimators $\hat{d}_{KS}$, $\hat{d}_{GD}$, $\hat{d}_{RM}$ or $\hat{d}_{TK}$ (for $m > n$) can be preferred as they all perform equally well.

7. CONCLUDING REMARKS AND EXAMPLES

In this paper, we have re-investigated the problem of estimating common mean of two normal populations when the variances are known to follow the ordering $\sigma_1^2 \leq \sigma_2^2$. It should be noted that, Elfessi and Pal [7] considered this model and obtained an estimator which dominates the well known Graybill-Deal (see Graybill and Deal [8]) estimator in terms of stochastic domination as well as Pitman measure of closeness criterion. We have proposed some new estimators for the common mean under order restricted variances which beat their unrestricted counterparts (previously proposed by Khatri and Saha [11], Moore and Krishnamoorthy [14], Tripathy and Kumar [20], Brown and Cohen [3]) stochastically, universally and in terms of Pitman measure of closeness criterion and compete well with the estimator proposed by Elfessi and Pal [7]. Moreover, we have obtained a plug-in type restricted MLE which beats the unrestricted MLE with respect to a squared error loss function. In addition to these, we have derived a sufficient condition for improving equivariant estimators using orbit-by-orbit improvement technique of Brewster and Zidek [2]. To the best of our knowledge, the performance of the MLE of the common mean has not been discussed under order restricted variances in the literature which was also lacking and we have tried to answer up to some extent. We have carried out a detailed and in-depth simulation study in order to compare the performances of both proposed as well as existing estimators with that of the plug-in type restricted MLE which was lacking in the literature. Under order restriction on the variances we have recommended estimators that can be used in practice. We hope that the current study may fill the knowledge gap and provide useful information to the researchers from an application point of view.

Example 7.1. (Simulated Data): The following two data sets each of size 10 from normal distributions have been generated using the software R, with a common mean $\mu = 25$ and with the condition that $\sigma_1^2 \leq \sigma_2^2$. Data Set A: 24.28, 25.94, 25.76, 29.14, 28.39, 23.51, 23.43, 22.60, 22.29, 28.26. Data Set B: 24.61, 23.70, 26.25, 29.11, 26.13, 23.52, 25.57, 22.34, 26.19, 23.04. The sufficient statistics can be computed as $\bar{x} = 25.36$, $\bar{y} = 25.04$, $s_1^2 = 57.69$, $s_2^2 = 36.46$. Based on the summery data it is seen that $s_1^2 > s_2^2$. This is a case where the improved estimators can be obtained. The various estimators are computed as $d_{GD} = 25.17$, $\hat{d}_{GD} = 25.24$, $d^a_{GD} = 25.20$, $d_{KS} = 25.17$, $\hat{d}_{KS} = 25.24$, $d^a_{KS} = 25.20$, $d_{MK} = 25.18$, $\hat{d}_{MK} = 25.22$, $d^a_{MK} = 25.20$, $d_{TK} = 25.18$, $\hat{d}_{TK} = 25.22$, $d^a_{TK} = 25.20$, $d_{BC1} = 25.25$, $d^a_{BC1} = 25.25$, $d_{BC2} = 25.18$, $\hat{d}_{BC2} = 25.20$, $d^a_{BC2} = 25.23$, $d_{ML} = 25.17$, $\hat{d}_{RM} = 25.20$, $d^a_{ML} = 25.20$. Depending upon the variance ratios, the improved estimators can be used.

Example 7.2. Rohatgi and Saleh [19], (p.515) discussed one example regarding the mean life time (in hours) of light bulbs. Suppose a random sample of 9 bulbs has sample mean 1309 hours with standard deviation of 420 hours. A second sample of 16 bulbs chosen from a different batch has sample mean 1205 hours and standard deviation 390 hours. A two sample t-test fails to reject the hypothesis that the means are equal. Suppose it is known a priori that the variance of first sample is smaller than the second one. This is a situation where our model will be useful. On the basis of these samples, we have $m = 9$, $n = 16$, $\bar{x} = 1309$, $\bar{y} = 1205$, $s_1 = (\sqrt{m-1})420$, $s_2 = (\sqrt{n-1})390$. The various estimators for the common mean are obtained as $d_{GD} = 1269.27$, $\hat{d}_{GD} = 1271.24$, $d_{KS} = 1267.44$, $\hat{d}_{KS} = 1271.24$, $d_{MK} = 1274.22$, $\hat{d}_{MK} = 1274.22$, $d_{TK} = 1274.01$, $\hat{d}_{TK} = 1274.01$, $d_{BC1} = 1284.37$, $d^a_{BC1} = 1284.37$, $d_{BC2} = 1287.32$, $d^a_{BC2} = 1287.32$, $d_{ML} = 1269.87$, $\hat{d}_{RM} = 1271.24$, and $d^a_{ML} = 1271.24$. In this situation we recommend to use either of the estimators $\hat{d}_{GD}$, $\hat{d}_{KS}$, or d^a_{ML}.

A. APPENDIX

Proof of Theorem 3.1: (i) First we prove the dominance result for equal sample sizes that is $m = n$. Consider the estimator $\hat{d}_{MK}$ which is given by

$$\hat{d}_{MK} = \begin{cases} (1-\beta_1)\bar{X} + \beta_1\bar{Y}, & \text{if } S_1 \le S_2, \\ \beta_1\bar{X} + (1-\beta_1)\bar{Y}, & \text{if } S_1 > S_2. \end{cases}$$

Our target is to show that,

$$P[(\hat{d}_{MK} - \mu)^2 \le c] \le P[(d_{MK} - \mu)^2 \le c], \quad \forall\ c > 0. \tag{A.1}$$

Which is equivalent to show that

$$P[(d_{MK} - \mu)^2 \le c | S_1 > S_2] \le P[(\hat{d}_{MK} - \mu)^2 \le c | S_1 > S_2], \quad \forall\ c > 0.$$

Denoting $X_1^* = (1-\beta_1)\bar{X} + \beta_1\bar{Y}$ and $X_2^* = \beta_1\bar{X} + (1-\beta_1)\bar{Y}$, we observe that $X_1^* - \mu \sim N(0, \sigma^2)$, $X_2^* - \mu \sim N(0, \sigma_*^2)$, where $\sigma^2 = (1-\beta_1)^2\frac{\sigma_1^2}{m} + \beta_1^2\frac{\sigma_2^2}{m}$ and $\sigma_*^2 = \beta_1^2\frac{\sigma_1^2}{m} + (1-\beta_1)^2\frac{\sigma_2^2}{m}$. Thus incorporating all these information the above inequality reduces to,

$$\Phi\Big(\frac{\sqrt{c}}{\sigma}\Big) \le \Phi\Big(\frac{\sqrt{c}}{\sigma_*}\Big), \ \forall\ c > 0 \text{ and } S_1 > S_2, \tag{A.2}$$

where $\Phi(.)$ is the cumulative distribution function of a standard normal random variable.

The inequality (A.2) is equivalent to show that, $\sigma^2 > \sigma_*^2$, when $S_1 > S_2$. This is true as, $\sigma^2 - \sigma_*^2 > 0$ when $\sigma_1^2 \le \sigma_2^2$ and $S_1 > S_2$. This proves the case of equal sample sizes.

Next we prove the result for the case of unequal sample sizes that is $m \ne n$. Denoting $V_1 = \sqrt{\frac{m}{m-1}}S_1$ and $V_2 = \sqrt{\frac{n}{n-1}}S_2$, the estimator $\hat{d}_{MK}$ can be written as,

$$\hat{d}_{MK} = \begin{cases} d_{MK}, & \text{if } V_1 \le V_2 \\ \frac{m\bar{X}+n\bar{Y}}{m+n}, & \text{if } V_1 > V_2. \end{cases}$$

Proceeding as before, one needs to show that,

$$P[(d_{MK} - \mu)^2 \le c | V_1 > V_2] \le P[(\hat{d}_{MK} - \mu)^2 \le c | V_1 > V_2], \quad \forall\ c > 0.$$

Which is further equivalent to show that

$$\Phi\Big(\frac{\sqrt{c}}{\nu}\Big) \le \Phi\Big(\frac{\sqrt{c}}{\nu_*}\Big), \ \forall\ c > 0,$$

where $\nu^2 = \frac{(m-1)s_2^2\sigma_1^2 + (n-1)s_1^2\sigma_2^2}{(\sqrt{m(m-1)}s_2 + \sqrt{n(n-1)}s_1)^2}$ and $\nu_*^2 = \frac{m\sigma_1^2 + n\sigma_2^2}{(m+n)^2}$. This is further equivalent to show that $\nu^2 > \nu_*^2$, $\forall\ c > 0$ when $V_1 > V_2$.

This is equivalent to show that,

$$\frac{\sigma_1^2 + \sigma_2^2\lambda^2}{(\sqrt{m} + \sqrt{n}\lambda)^2} > \frac{m\sigma_1^2 + n\sigma_2^2}{(m+n)^2}, \ \forall\ c > 0, \sigma_1 \le \sigma_2,$$

where $\lambda = \frac{\sqrt{n-1}S_1}{\sqrt{m-1}S_2}$. Let $h(\lambda) = \frac{\sigma_1^2 + \sigma_2^2\lambda^2}{(\sqrt{m}+\sqrt{n}\lambda)^2}$. To show that $h(\lambda) > h(\sqrt{\frac{n}{m}})$, for $\lambda > \sqrt{\frac{n}{m}}$. We observe that $\frac{dh}{d\lambda} \le 0$, if $\lambda \le \sqrt{\frac{n}{m}}\frac{\sigma_1^2}{\sigma_2^2} \le \sqrt{\frac{n}{m}}$, as $\sigma_1^2/\sigma_2^2 \le 1$. Further $\frac{dh}{d\lambda} > 0$, when $\lambda > \sqrt{\frac{n}{m}}$. Hence $h(\lambda)$ is increasing in the interval $[\sqrt{\frac{n}{m}}, \infty)$. Universal domination follows from Definition 3.2. This proves (i). The proofs of (ii)–(iv) are very much similar to the proof of (i) and hence have been omitted. This completes the proof of the Theorem 3.1. □

Proof of Theorem 5.1: The theorem can be proved by using a well known technique for improving equivariant estimators proposed by Brewster and Zidek [2]. To proceed, let us consider the conditional risk function of d_Ψ given $\underset{\sim}{T} = \underset{\sim}{t}$:

$$R(\underset{\sim}{\alpha}, d_\Psi|\underset{\sim}{t}) = \frac{1}{\sigma_1^2} E\{(\bar{X} + S_1\Psi(\underset{\sim}{T}) - \mu)^2|\underset{\sim}{T} = \underset{\sim}{t}\}.$$

The above risk function is convex in $\Psi(\underset{\sim}{t})$ and attains minimum at

$$\Psi(\underset{\sim}{t}, \underset{\sim}{\alpha}) = \frac{E\{(\mu - \bar{X})S_1|\underset{\sim}{T} = \underset{\sim}{t}\}}{E\{S_1^2|\underset{\sim}{T} = \underset{\sim}{t}\}}. \tag{A.3}$$

To evaluate the conditional expectations involved in the above expression, we use the following transformations. Let us define $V_1 = (\sqrt{m}(\bar{X} - \mu))/\sigma_1$, $V_2 = (\sqrt{m}(\bar{Y} - \mu))/\sigma_1$, $W_1 = S_1^2/\sigma_1^2$ and $W_2 = S_2^2/\sigma_2^2$. With these substitution the expression for $\Psi(\underset{\sim}{t}, \underset{\sim}{\alpha})$ then reduces to,

$$\Psi(\underset{\sim}{t}, \underset{\sim}{\alpha}) = -\frac{E(V_1 W_1^{\frac{1}{2}}|\underset{\sim}{T} = \underset{\sim}{t})}{\sqrt{m}E(W_1|\underset{\sim}{T} = \underset{\sim}{t})}. \tag{A.4}$$

These conditional expectations have been evaluated in [20] and are given by,

$$E(W_1|\underset{\sim}{T} = \underset{\sim}{t}) = \frac{m + n - 1}{\lambda},$$

and

$$E(V_1 W_1^{\frac{1}{2}}|\underset{\sim}{T} = \underset{\sim}{t}) = -\frac{n\sqrt{m}(m + n - 1)t_1}{(n + m\rho)\lambda},$$

where

$$\lambda = \frac{mnt_1^2}{n + m\rho} + \frac{t_2}{\rho} + 1, \quad \rho = \frac{\sigma_2^2}{\sigma_1^2} \geq 1.$$

Substituting these expressions in (A.4), we get the minimimizing choice of $\Psi(\underset{\sim}{t}, \underset{\sim}{\alpha})$ as

$$\hat{\Psi}(\underset{\sim}{t}, \rho) = \frac{nt_1}{n + m\rho}.$$

In order to derive the inadmissibility condition of the theorem, we need the supremum and infimum values of $\hat{\Psi}(\underset{\sim}{t}, \rho)$ with respect to $\rho \in [1, \infty)$ for fixed values of $\underset{\sim}{T} = \underset{\sim}{t}$. We consider the following two cases to obtain the supremum and infimum of $\hat{\Psi}(\underset{\sim}{t}, \rho)$:

Case-I: Let $t_1 \geq 0$. Now the function $\hat{\Psi}(\underset{\sim}{t}, \rho)$ is decreasing with respect to $\rho \in [1, \infty)$. Hence we obtain

$$\inf_{\rho \geq 1} \hat{\Psi}(\underset{\sim}{t}, \rho) = \lim_{\rho \to \infty} \hat{\Psi}(\underset{\sim}{t}, \rho) = 0 \text{ and } \sup_{\rho \geq 1} \hat{\Psi}(\underset{\sim}{t}, \rho) = \lim_{\rho \to 1} \hat{\Psi}(\underset{\sim}{t}, \rho) = \frac{nt_1}{n + m}.$$

Case-II: Let $t_1 < 0$. The function $\hat{\Psi}(\underset{\sim}{t}, \rho)$ is an increasing function of ρ. So for this case we obtain

$$\inf_{\rho \geq 1} \hat{\Psi}(\underset{\sim}{t}, \rho) = \lim_{\rho \to 1} \hat{\Psi}(\underset{\sim}{t}, \rho) = \frac{nt_1}{n + m} \text{ and } \sup_{\rho \geq 1} \hat{\Psi}(\underset{\sim}{t}, \rho) = \lim_{\rho \to \infty} \hat{\Psi}(\underset{\sim}{t}, \rho) = 0.$$

Combining the Case-I and Case-II, it is easy to define the function $\Psi_0(\underset{\sim}{t})$ as given in (5.2). Utilizing the function $\Psi_0(\underset{\sim}{t})$ and as an application of Theorem 3.1 (in Brewster and Zidek [2]), we get $R(d_{\Psi_0}, \underset{\sim}{\alpha}) \leq R(d_\Psi, \underset{\sim}{\alpha})$, when $\sigma_1 \leq \sigma_2$. This completes the proof of the theorem. □

A Sample Program Code of the Simulation Study

As suggested by an anonymous reviewer, below we present a sample program code of the simulation study for equal sample sizes.

```
library(MASS)
library(nleqslv)
M=20000
n1=30
n2=30
b=1.5434/2.0
c=1.5045/2.0
sd2=1.0
mu=0.0
cm=gamma((n1-1)/2)/(sqrt(2)*gamma(n1/2))
cn=gamma((n2-1)/2)/(sqrt(2)*gamma(n2/2))
for(sd1 in seq(0.05,1.0,0.05))
{x1=matrix(0,n1,M)
x2=matrix(0,n2,M)
m1=array(0,M)
m2=array(0,M)
s1=array(0,M)
s2=array(0,M)
d=array(0,M)
a1=array(0,M)
a2=array(0,M)
a3=array(0,M)
a4=array(0,M)
a5=array(0,M)
a6=array(0,M)
a7=array(0,M)
a8=array(0,M)
a9=array(0,M)
a10=array(0,M)
a11=array(0,M)
a12=array(0,M)
a13=array(0,M)
a14=array(0,M)
a15=array(0,M)
a16=array(0,M)
b1=array(0,M)
b2=array(0,M)
b3=array(0,M)
b4=array(0,M)
b5=array(0,M)
b6=array(0,M)
b7=array(0,M)
b8=array(0,M)
b9=array(0,M)
b10=array(0,M)
b11=array(0,M)
b12=array(0,M)
b13=array(0,M)
b14=array(0,M)
b15=array(0,M)
b16=array(0,M)
c1=array(0,M)
c2=array(0,M)
c3=array(0,M)
c4=array(0,M)
c5=array(0,M)
c6=array(0,M)
```

```
c7=array(0,M)
c8=array(0,M)
c9=array(0,M)
c10=array(0,M)
c11=array(0,M)
c12=array(0,M)
c13=array(0,M)
c14=array(0,M)
c15=array(0,M)
c16=array(0,M)
c17=array(0,M)
c18=array(0,M)
c19=array(0,M)
c20=array(0,M)
e1=array(0,M)
e2=array(0,M)
GD=array(0,M)
GDI=array(0,M)
GDA=array(0,M)
PsiGD=array(0,M)
Psi1=array(0,M)
Psi2=array(0,M)
PsiKS=array(0,M)
PsiMK=array(0,M)
PsiTK=array(0,M)
KS=array(0,M)
KSI=array(0,M)
KSA=array(0,M)
MK=array(0,M)
MKI=array(0,M)
MKA=array(0,M)
TK=array(0,M)
TKI=array(0,M)
TKA=array(0,M)
ML=array(0,M)
T1=array(0,M)
T2=array(0,M)
T3=array(0,M)
T4=array(0,M)
V1R=array(0,M)
V2R=array(0,M)
MLR=array(0,M)
PsiBC1=array(0,M)
BC1A=array(0,M)
BC1=array(0,M)
BC2=array(0,M)
PsiBC2=array(0,M)
BC2A=array(0,M)
g1=array(0,M)
g2=array(0,M)
g3=array(0,M)
g4=array(0,M)
RML=array(0,M)
beta1=array(0,M)
beta2=array(0,M)
beta3=array(0,M)
beta4=array(0,M)
beta5=array(0,M)
BC2I=array(0,M)
PsiML=array(0,M)
t1=array(0,M)
MLA=array(0,M)
for(j in 1:M)
```

$\{x1[,j] = rnorm(n1, mean = mu, sd = sqrt(sd1))$
$x2[,j] = rnorm(n2, mean = mu, sd = sqrt(sd2))$
$m1[j] = mean(x1[,j])$
$m2[j] = mean(x2[,j])$
$d[j] = m2[j] - m1[j]$
$s1[j] = sum((x1[,j] - m1[j])^2)$
$s2[j] = sum((x2[,j] - m2[j])^2)$
$t1[j] = d[j]/sqrt(s1[j])$
$beta1[j] = (n2 * (n2-1) * s1[j])/((n2 * (n2-1) * s1[j]) + (n1 * (n1-1) * s2[j]))$
$a1[j] = ((1 - beta1[j]) * m1[j]) + (beta1[j] * m2[j])$
$GD[j] = ((a1[j] - mu)/sqrt(sd1))^2$
$a2[j] = s1[j]/(n1-1)$
$a3[j] = s2[j]/(n2-1)$
$if(a2[j] <= a3[j])$
$\{a4[j] = a1[j]\}$
$else\{a4[j] = (beta1[j] * m1[j]) + ((1 - beta1[j]) * m2[j])\}$
$GDI[j] = ((a4[j] - mu)/sqrt(sd1))^2$
$a5[j] = (n2 * (n2-1)) * (d[j]/sqrt(s1[j]))$
$a6[j] = (n1 * (n1-1) * (s2[j]/s1[j])) + (n2 * (n2-1))$
$PsiGD[j] = a5[j]/a6[j]$
$Psi1[j] = (n2/(n1+n2)) * min((d[j]/sqrt(s1[j])), 0)$
$Psi2[j] = (n2/(n1+n2)) * max((d[j]/sqrt(s1[j])), 0)$
$if(PsiGD[j] < Psi1[j])$
$\{a7[j] = Psi1[j]\}$
$elseif(PsiGD[j] > Psi2[j])$
$\{a7[j] = Psi2[j]\}$
$else\{a7[j] = PsiGD[j]\}$
$a8[j] = m1[j] + (sqrt(s1[j]) * a7[j])$ $GDA[j] = ((a8[j] - mu)/sqrt(sd1))^2$
$beta2[j] = (n2 * (n2-3) * s1[j])/((n2 * (n2-3) * s1[j]) + (n1 * (n1-3) * s2[j]))$
$a9[j] = (beta2[j] * m2[j]) + ((1 - beta2[j]) * m1[j])$
$KS[j] = ((a9[j] - mu)/sqrt(sd1))^2$
$a10[j] = s1[j]/(n1-3)$
$a11[j] = s2[j]/(n2-3)$
$if(a10[j] <= a11[j])$
$\{a12[j] = a9[j]\}$
$else\{a12[j] = (beta2[j] * m1[j]) + ((1 - beta2[j]) * m2[j])\}$
$KSI[j] = ((a12[j] - mu)/sqrt(sd1))^2$
$a13[j] = (n2 * (n2-3)) * (d[j]/sqrt(s1[j]))$
$a14[j] = (n1 * (n1-3) * (s2[j]/s1[j])) + (n2 * (n2-3))$
$PsiKS[j] = a13[j]/a14[j]$
$if(PsiKS[j] < Psi1[j])$
$\{a15[j] = Psi1[j]\}$
$elseif(PsiKS[j] > Psi2[j])$
$\{a15[j] = Psi2[j]\}$
$else\{a15[j] = PsiKS[j]\}$
$a16[j] = m1[j] + (sqrt(s1[j]) * a15[j])$
$KSA[j] = ((a16[j] - mu)/sqrt(sd1))^2$
$beta3[j] = sqrt(n2 * (n2-1) * s1[j])/(sqrt(n2 * (n2-1) * s1[j]) + sqrt(n1 * (n1-1) * s2[j]))$
$b1[j] = ((1 - beta3[j]) * m1[j]) + (beta3[j] * m2[j])$
$MK[j] = ((b1[j] - mu)/sqrt(sd1))^2$
$b2[j] = sqrt(n1/(n1-1)) * sqrt(s1[j])$
$b3[j] = sqrt(n2/(n2-1)) * sqrt(s2[j])$
$if(b2[j] <= b3[j])$
$\{b4[j] = b1[j]\}$
$else\{b4[j] = (beta3[j] * m1[j]) + ((1 - beta3[j]) * m2[j])\}$
$MKI[j] = ((b4[j] - mu)/sqrt(sd1))^2$
$b5[j] = sqrt(n2 * (n2-1)) * (d[j]/sqrt(s1[j]))$
$b6[j] = (sqrt(n1 * (n1-1)) * sqrt(s2[j]/s1[j])) + (sqrt(n2 * (n2-1)))$
$PsiMK[j] = b5[j]/b6[j]$
$if(PsiMK[j] < Psi1[j])$
$\{b7[j] = Psi1[j]\}$
$elseif(PsiMK[j] > Psi2[j])$
$\{b7[j] = Psi2[j]\}$

$else\{b7[j] = PsiMK[j]\}$
$b8[j] = m1[j] + (sqrt(s1[j]) * b7[j])$
$MKA[j] = ((b8[j] - mu)/sqrt(sd1))^2$
$beta4[j] = (sqrt(n2 * s1[j]) * cm)/((sqrt(n2 * s1[j]) * cm) + (sqrt(n1 * s2[j]) * cn))$
$b9[j] = ((1 - beta4[j]) * m1[j]) + (beta4[j] * m2[j])$
$TK[j] = ((b9[j] - mu)/sqrt(sd1))^2$
$b10[j] = cm * sqrt(s1[j]) * sqrt(n1)$
$b11[j] = cn * sqrt(s2[j]) * sqrt(n2)$
$if(b10[j] <= b11[j])$
$\{b12[j] = b9[j]\}$
$else\{b12[j] = (beta4[j] * m1[j]) + ((1 - beta4[j]) * m2[j])\}$
$TKI[j] = ((b12[j] - mu)/sqrt(sd1))^2$
$b13[j] = sqrt(n2) * cm * (d[j]/sqrt(s1[j]))$
$b14[j] = (sqrt(n1) * cn * sqrt(s2[j]/s1[j])) + (sqrt(n2) * cm)$
$PsiTK[j] = b13[j]/b14[j]$
$if(PsiTK[j] < Psi1[j])$
$\{b15[j] = Psi1[j]\}$
$elseif(PsiTK[j] > Psi2[j])$
$\{b15[j] = Psi2[j]\}$
$else\{b15[j] = PsiTK[j]\}$
$b16[j] = m1[j] + (sqrt(s1[j]) * b15[j])$
$TKA[j] = ((b16[j] - mu)/sqrt(sd1))^2$
$c1[j] = (d[j] * b * s1[j])/(n1 * (n1 - 1))$
$c2[j] = (s1[j]/(n1 * (n1 - 1))) + (s2[j]/(n2 * (n2 + 2))) + ((d[j] * d[j])/(n2 + 2))$
$c3[j] = c1[j]/c2[j]$
$c4[j] = m1[j] + c3[j]$
$BC1[j] = ((c4[j] - mu)/sqrt(sd1))^2$
$c5[j] = (d[j] * b)/(n1 * (n1 - 1) * sqrt(s1[j]))$
$c6[j] = 1/(n1 * (n1 - 1))$
$c7[j] = s2[j]/(s1[j] * n2 * (n2 + 2))$
$c8[j] = ((d[j]/sqrt(s1[j]))^2)/(n2 + 2)$
$PsiBC1[j] = c5[j]/(c6[j] + c7[j] + c8[j])$
$c9[j] = m1[j] + (sqrt(s1[j]) * PsiBC1[j])$
$if(PsiBC1[j] < Psi1[j])$
$\{c10[j] = Psi1[j]\}$
$elseif(PsiBC1[j] > Psi2[j])$
$\{c10[j] = Psi2[j]\}$
$else\{c10[j] = PsiBC1[j]\}$
$c11[j] = m1[j] + (sqrt(s1[j]) * c10[j])$
$BC1A[j] = ((c11[j] - mu)/sqrt(sd1))^2$
$c12[j] = d[j] * c * s1[j] * (n2 * (n2 - 1))$
$c13[j] = (n2 * (n2 - 1) * s1[j]) + (n1 * (n1 - 1) * s2[j])$
$c14[j] = c12[j]/c13[j]$
$c15[j] = m1[j] + c14[j]$
$BC2[j] = ((c15[j] - mu)/sqrt(sd1))^2$
$c16[j] = (d[j]/sqrt(s1[j])) * c * n2 * (n2 - 1)$
$c17[j] = (n2 * (n2 - 1)) + (n1 * (n1 - 1) * (s2[j]/s1[j]))$
$PsiBC2[j] = c16[j]/c17[j]$
$if(PsiBC2[j] < Psi1[j])$
$\{c18[j] = Psi1[j]\}$
$elseif(PsiBC2[j] > Psi2[j])$
$\{c18[j] = Psi2[j]\}$
$else\{c18[j] = PsiBC2[j]\}$
$c19[j] = m1[j] + (sqrt(s1[j]) * c18[j])$
$BC2A[j] = ((c19[j] - mu)/sqrt(sd1))^2$
$beta5[j] = (c * n2 * (n2 - 1) * s1[j])/((n2 * (n2 - 1) * s1[j]) + (n1 * (n1 - 1) * s2[j]))$
$if((2 * c) <= (1 + (s2[j]/s1[j])))$
$\{c20[j] = ((1 - beta5[j]) * m1[j]) + (beta5[j] * m2[j])\}$
$else\{c20[j] = (beta5[j] * m1[j]) + ((1 - beta5[j]) * m2[j])\}$
$BC2I[j] = ((c20[j] - mu)/sqrt(sd1))^2$
$fnewton < -function(x)$
$\{y < -numeric(3)$
$d11 = n2 * x[1] * d[j]$

$d22 = n1 * d[j] * x[2]$
$d33 = (n2 * x[1]) + (n1 * x[2])$
$d44 = (n1 * m1[j]/x[1]) + (n2 * m2[j]/x[2])$
$d55 = (n1/x[1]) + (n2/x[2])$
$y[1] < -x[1] - (s1[j]/n1) - (d11/d33)^2$
$y[2] < -x[2] - (s2[j]/n2) - (d22/d33)^2$
$y[3] < -x[3] - (d44/d55)$
$y\}$
$xstart < -c(s1[j]/(n1-1), s2[j]/(n2-1), m1[j])$
$T1[j] = nleqslv(xstart, fnewton, control = list(btol = 0.0001), method = "Newton")$x[1]
$T2[j] = nleqslv(xstart, fnewton, control = list(btol = 0.0001), method = "Newton")$x[2]
$T3[j] = nleqslv(xstart, fnewton, control = list(btol = 0.0001), method = "Newton")$x[3]
$ML[j] = ((T3[j] - mu)/sqrt(sd1))^2$
$V1R[j] = min(T1[j], ((n1 * T1[j]) + (n2 * T2[j]))/(n1 + n2))$
$V2R[j] = max(T2[j], ((n1 * T1[j]) + (n2 * T2[j]))/(n1 + n2))$
$g1[j] = (n1 * V2R[j] * m1[j]) + (n2 * V1R[j] * m2[j])$
$g2[j] = (n1 * V2R[j]) + (n2 * V1R[j])$
$g3[j] = g1[j]/g2[j]$
$MLR[j] = ((g3[j] - mu)/sqrt(sd1))^2$
$if(T1[j] <= T2[j])$
$\{g4[j] = T3[j]\}$
$else\{g4[j] = ((n1 * m1[j]) + (n2 * m2[j]))/(n1 + n2)\}$
$RML[j] = ((g4[j] - mu)/sqrt(sd1))^2$
$PsiML[j] = (n2 * T1[j] * t1[j])/((n1 * T2[j]) + (n2 * T1[j]))$
$if(PsiML[j] < Psi1[j])$
$\{e1[j] = Psi1[j]\}$
$elseif(PsiML[j] > Psi2[j])$
$\{e1[j] = Psi2[j]\}$
$else\{e1[j] = PsiML[j]\}$
$e2[j] = m1[j] + (sqrt(s1[j]) * e1[j])$
$MLA[j] = ((e2[j] - mu)/sqrt(sd1))^2$
$\}\ tau = sd1/sd2$
$R1 = sum(GD)/M$
$R2 = sum(GDI)/M$
$R3 = sum(GDA)/M$
$R4 = sum(KS)/M$
R5=sum(KSI)/M
R6=sum(KSA)/M
R7=sum(MK)/M
R8=sum(MKI)/M
R9=sum(MKA)/M
R10=sum(TK)/M
R11=sum(TKI)/M
R12=sum(TKA)/M
R13=sum(BC1)/M
R14=sum(BC1A)/M
R15=sum(BC2)/M
R16=sum(BC2A)/M
R222=sum(BC2I)/M
R17=sum(ML)/M
R18=sum(MLR)/M
R19=sum(RML)/M
R20=sum(MLA)/M
P1=round(((R1-R2)/R1)*100,2)
P2=round(((R1-R3)/R1)*100,2)
P3=round(((R4-R5)/R4)*100,2)
P4=round(((R4-R6)/R4)*100,2)
P5=round(((R7-R8)/R7)*100,2)
P6=round(((R7-R9)/R7)*100,2)
P7=round(((R10-R11)/R10)*100,2)
P8=round(((R10-R12)/R10)*100,2)
P9=round(((R13-R14)/R13)*100,2)
P10=round(((R15-R16)/R15)*100,2)

P11=round(((R17-R18)/R17)*100,2)
P12=round(((R17-R19)/R17)*100,2)
P13=round(((R15-R222)/R15)*100,2)
P14=round(((R17-R20)/R17)*100,2)
PR1=round(((R1-R2)/R1)*100,2)
PR2=round(((R1-R3)/R1)*100,2)
PR3=round(((R1-R8)/R1)*100,2)
PR4=round(((R1-R9)/R1)*100,2)
PR5=round(((R1-R14)/R1)*100,2)
PR6=round(((R1-R16)/R1)*100,2)
PR7=round(((R1-R222)/R1)*100,2)
PR8=round(((R1-R19)/R1)*100,2)
PR9=round(((R1-R20)/R1)*100,2)
}}

ACKNOWLEDGMENTS

The second author (Manas Ranjan Tripathy) would like to thank Department of Science and Technology (SERB), [EMR/2017/003078] New Delhi, India for providing some financial support. The authors would like to thank an anonymous reviewer and the editor for their valuable suggestions and comments which have helped in improving the presentation of the paper.

REFERENCES

[1] Barlow, R.E.; Bartholomew, D.J.; Bremner, J.M. and Brunk, M.D. (1972). *Statistical Inference Under Order Restrictions*, Wiley, New York.

[2] Brewster, J.F. and Zidek, J.V. (1974). Improving on equivariant estimators, *Annals of Statistics*, **2**, 21–38.

[3] Brown, L.D. and Cohen, A. (1974). Point and confidence estimation of a common mean and recovery of inter-block information, *Annals of Statistics*, **2**(5), 963–976.

[4] Chang, C. and Pal, N. (2008). Testing on the common mean of several normal distributions, *Computational Statistics and Data Analysis*, **53**, 321–333.

[5] Chang, Y.T.; Oono, Y. and Shinozaki, N. (2012). Improved estimators for the common mean and ordered means of two normal distributions with ordered variances, *Journal of Statistical Planning and Inference*, **142**(9), 2619–2628.

[6] Cohen, A. and Sackrowitz, H. (1974). On estimating the common mean of two normal distributions, *Annals of Statistics*, **2**(6), 1274–1282.

[7] Elfessi, A. and Pal, N. (1992). A note on the common mean of two normal populations with order restricted variances, *Communications in Statistics – Theory and Methods*, **21**(11), 3177–3184.

[8] Graybill, F.A. and Deal, R.B. (1959). Combining unbiased estimators, *Biometrics*, **15**, 543–550.

[9] Hwang, J.T. (1985). Universal domination and stochastic domination: Estimation simultaneously under a broad class of loss functions, *Annals of Statistics*, **13**(1), 295–314.

[10] Kelleher, T. (1996). A Bayes method for estimating the common mean of two normal populations, *Communications in Statistics – Theory and Methods*, **25**(9), 2141–2157.

[11] Khatri, C.G. and Shah, K.R. (1974). Estimation of location parameters from two linear models under normality, *Communications in Statistics – Theory and Methods*, **3**, 647–663.

[12] Lin, S. and Lee, J.C. (2005). Generalized inferences on the common mean of several normal populations, *Journal of Statistical Planning and Inference*, **134**, 568–582.

[13] Misra, N. and van der Meulen, E.C. (1997). On estimation of the common mean of $k\,(\geq 2)$ normal populations with order restricted variances, *Statistics and Probability Letters*, **36**, 261–267.

[14] Moore, B. and Krishnamoorthy, K. (1997). Combining independent normal sample means by weighting with their standard errors, *Journal of Statistical Compution and Simulation*, **58**, 145–153.

[15] Pal, N.; Lin, J.J.; Chang, C.H. and Kumar, S. (2007). A revisit to the common mean problem: Comparing the maximum likelihood estimator with the Graybill-Deal estimator, *Computational Statistics and Data Analysis*, **51**(12), 5673–5681.

[16] Pal, N. and Sinha, B.K. (1996). Estimation of a common mean of several normal populations: A review, *Far East Journal of Mathematical Sciences*, Special Volume, Part I, 97–110.

[17] Peddada, S.D. and Khatree, R. (1986). On Pitman nearness and variance of estimators, *Communications in Statistics – Theory and Methods*, **15**, 3005–3018.

[18] Pitman, E.J.G. (1937). The closest estimates of statistical parameters, *Mathematical Proceedings of the Cambridge Philosophical Society*, **33**(2), 212–222.

[19] Rohatgi, V.K. and Saleh, A.K.Md.E. (2003). An Introduction to Probability and Statistics, John Wiley, 2nd Edition, New Work.

[20] Tripathy, M.R. and Kumar, S. (2010). Estimating a common mean of two normal populations, *Journal of Statistical Theory and Applications*, **9**(2), 197–215.

[21] Tripathy, M.R. and Kumar, S. (2015). Equivariant estimation of common mean of several normal populations, *Journal of Statistical Computation and Simulation*, **85**(18), 3679–3699.

[22] Tripathy, M.R.; Kumar, S. and Pal, N. (2013). Estimating common standard deviation of two normal populations with ordered means, *Statistical Methods and Applications*, **22**(3), 305–318.

[23] Yates, F. (1940). The recovery of inter-block information in balanced incomplete block designs, *Annals of Eugenics*, **10**, 317–325.

REVSTAT – Statistical Journal
Volume 19, Number 3, July 2021, 363–398

NONPARAMETRIC SMOOTHING FOR EXTREMAL QUANTILE REGRESSION WITH HEAVY TAILED DATA

Author: Takuma Yoshida
– Graduate School of Science and Engineering, Kagoshima University, Japan
yoshida@sci.kagoshima-u.ac.jp

Received: September 2017 Revised: April 2019 Accepted: April 2019

Abstract:

- In several different fields, it is interested in analyzing the upper or lower tail quantile of the underlying distribution rather than mean or center quantile. However, the investigation of the tail quantile is somewhat difficult because of data sparsity. This paper challenges to develop the nonparametric quantile regression for extremal quantile level. In the extremal quantile regression, there are two situation of technical conditions of order of convergence of the quantile level that intermediate order or extreme order. For the intermediate order quantile, the ordinary nonparametric estimator is used. On the other hand, for the extreme order quantile, we provide the new estimator by the extrapolating the intermediate order quantile estimator. The performance of the estimator is guaranteed by the asymptotic theory and the extreme value theory. As the result, we show the asymptotic normality and the rate of convergence of the nonparametric quantile regression estimator for both intermediate and extreme order quantile. Simulation is addressed to confirm the behavior of the proposed estimator. The data application is also assessed.

Keywords:

- *asymptotic normality; extrapolation; extremal quantile regression; extreme value theory; nonparametric estimator.*

AMS Subject Classification:

- 62G08, 62G20, 62G32.

1. INTRODUCTION

In a wide variety of areas, such as in the study of heavy rainfall, low birth weight, and high-risk finance, the tail behavior of the distribution of the target variable is of interest rather than the average or median. In these cases, we often investigate the upper or lower quantile of the data. However, the estimation of the tail quantile is difficult because of data sparsity. Therefore, the development of the mathematical properties of the tail quantile would be welcome. The theoretical performance of the tail behavior of the distribution function is provided by extreme value theory (EVT). The fundamental properties of EVT were surveyed by Coles (2001) [10], Beirlant *et al.* (2004a) [3], and de Haan and Ferreira (2006) [13]. On the other hand, the performance of the estimator is often guaranteed by a large sample or asymptotic theory in statistics. Thus, the mathematical properties of the estimator of the tail quantile are analyzed using a hybrid of EVT and asymptotic theory. In many cases, it is important to research the target variable along with the information of other variable. Then we should analyze the data in the literature of regression. In this paper, we consider the estimation of the extremal conditional quantile of the response Y given $X = x$.

Many researchers have developed the extremal conditional quantile estimation. Beirlant and Goegebeur (2004) [2] developed a Pareto distribution approach. Gardes *et al.* (2010) [20] and Gardes and Girard (2010) [19] studied the nearest-neighbor estimation. Daouia *et al.* (2011, 2013) [12, 11], El Methni *et al.* (2014) [18], and Girard and Louhichi (2015) [21] investigated the extremal quantile of the nonparametric estimator of the conditional distribution function of Y given $X = x$. The local-moment-type methods were studied by Goegebeur *et al.* (2017) [22]. Durrieu *et al.* (2015) [15] have developed the weighted quasi-log-likelihood method. On the other hand, quantile regression, which was pioneered by Koenker and Bassett (1978) [31], is a typical approach used to investigate the conditional quantile. For the center quantile, several authors have developed quantile regression methods. These fundamental developments have been summarized by Koenker (2005) [29]. However, much less work has been done on quantile regression for the extremal quantile. Chernozhukov (2005) [6], Chernozhukov and Fernández-val (2011) [7], Wang *et al.* (2012) [44], and He *et al.* (2016) [24] studied extremal quantile regression, but they focused only on linear models. For the tail quantile, the linear structure assumption is strong and its assumption is violated in several cases. Therefore, a nonparametric approach should be used in extremal quantile regression in such situations. Beirlant *et al.* (2004b) [4] studied extremal nonparametric quantile regression, but the theoretical property was not investigated. In this paper, we develop nonparametric quantile regression for the extremal quantile and mathematical properties.

Before we describe our study, we review extremal quantile regression with linear models in more detail. For extremal quantile regression, the quantile level τ approaches 0 or 1 as the sample size n increases. This paper treats only the upper quantile and, hence, $\tau \to 1$ as $n \to \infty$. Thus, there are two important types of the order of τ: the intermediate order quantile and the extreme order quantile. The former means that $\tau \to 1$ and $n(1-\tau) \to \infty$ as $n \to \infty$, whereas in the latter $\tau \to 1$ and $n(1-\tau) \to c \in [0, \infty)$ as $n \to \infty$. If τ is fixed, it is the so-called center quantile. According to Chernozhukov (2005) [6] and Chernozhukov and Fernández-val (2011) [7], the quantile regression estimator with linear models has asymptotic normality for the intermediate order quantile but it converges to a non-degenerated distribution (not normal) for the extreme order quantile. Thus, the extreme order quantile is difficult to handle.

Wang *et al.* (2012) [44] provided a nice approach to obtain the extreme order quantile estimator by extrapolation from the intermediate order quantile estimator. As a result, this extrapolated estimator has asymptotic normality. It seems that above results should be extended to the nonparametric quantile regression for many applications.

In this paper, we first construct the ordinary nonparametric estimator for the intermediate order quantile. We then use the B-spline method with ℓ_2 penalty. This approach was originally considered by O'Sullivan (1986) [37] and Eilers and Marx (1996) [16] in mean regression. Pratesi *et al.* (2009) [39], Reiss and Huang (2012) [40], and Yoshida (2013) [46] used the quantile regression for only the center quantile. We show the asymptotic bias and variance as well as the asymptotic normality of the penalized spline estimator. Next, the extrapolated estimator is obtained for the extreme order quantile. Similar to the approach of Wang *et al.* (2012) [44], we use the Weissman-type extrapolation method (see Weissman 1978 [45]). The asymptotic normality and the optimal rate of convergence of the extreme order quantile estimator are shown. To the best of our knowledge, this is the first time that the rate of convergence of the nonparametric estimator is revealed in the extremal quantile regression.

This paper is organized as follows. In Section 2, we coordinate the conditions of the true conditional quantile by EVT in nonparametric extremal quantile regression. Section 3 presents the nonparametric estimator and its asymptotic property for both intermediate and extreme order quantiles. In Section 4 the Monte Carlo simulation is conducted to confirm the performance of the proposed estimator. Section 5 addresses the application to Beijing's $PM_{2.5}$ pollution data. The conclusions and future research are described in Section 6. In Appendices A and B, the computational aspects of the penalized spline estimator and the proofs of the mathematical results that appear in this paper are presented. Throughout the paper and without loss of generality, we focus on the conditional high quantile because a low quantile of the response can be viewed as a high quantile of the inverted sign of the response.

2. CONDITIONAL EXTREMAL QUANTILES

2.1. Extreme value theory

Let $\{(X_i, Y_i);\ i = 1, ..., n\}$ be the independent copies of a random pair $(X, Y) \in \mathbb{R} \times \mathbb{R}$. We assume that the support of X is bounded on $[a, b]$, where $-\infty < a < b < \infty$. The conditional distribution function of Y given $X = x$ is denoted by $F_Y(y|x) = P(Y \leq y|X = x)$. Then the conditional $100\tau\%$ quantile of Y given $X = x$ is

$$q_Y(\tau|x) \equiv F_Y^{-1}(\tau|x) = \inf\{t, F_Y(t|x) \geq \tau\}.$$

The main purpose of this study is to estimate $q_Y(\tau|x)$ for a high quantile level $\tau \simeq 1$. The tail behavior of the distribution or quantile function can be characterized by EVT. To analyze the conditional high quantile of Y given $X = x$, we introduce the EVT conditions of $F_Y(\cdot|x)$ and $q_Y(\cdot|x)$. Define $F(y)$ and $q(\tau)$ as the marginal distribution function and $100\tau\%$ quantile of Y. Throughout the paper, we assume that F and $F(\cdot|x)$ belong to the maximum domain of attraction of an extreme value distribution G_γ, denoted by $F,\ F_Y(\cdot|x) \in D(G_\gamma)$.

The distribution function Q belongs to the maximum domain of attraction, which means that for the random sample $Z_1, ..., Z_n$ from Q, there exists a constant $\alpha_n > 0$ and $\beta_n \in \mathbb{R}$ such that for $1 + \gamma z \geq 0$,

$$P\left(\frac{\max_{1\leq i\leq n} Z_i - b_n}{a_n} \leq z\right) \to G_\gamma(z) = \exp[-(1+\gamma z)^{-1/\gamma}]$$

as $n \to \infty$. Here, $\gamma \in \mathbb{R}$ is the extreme value index (EVI). The EVI is very important since this generally controls the tail behavior of the distribution function. For $Q \in D(G_\gamma)$, if $\gamma = 0$ or $\gamma < 0$, Q has a light tail or short tail. When $\gamma > 0$, Q has a heavy tail. This paper only discusses the heavy-tail case and, hence, we assume that $\gamma > 0$ from now on. The maximum domain of attraction is a very weak condition. For example, uniform, beta, Gaussian, t, Pareto, Cauchy, and many other distributions belong to the maximum domain of attraction with appropriately specified $\gamma \in \mathbb{R}$. The details of the maximum domain of attraction and EVI are given in fundamental books such as that by de Haan and Ferreira (2006) [13].

We now state the conditions to connect the tail behavior of F and $F(\cdot|x)$. For this, we need an additional definition. Let $RV(a) = \{A \in \mathbb{R}^+ \to \mathbb{R}^+ | A(mt)/A(t) \to m^a$ as $t \to \infty$, $m > 0\}$ be the set of regularly varying functions, where $\mathbb{R}^+ = (0, \infty)$. When $A \in RV(0)$, A is the so-called slowly varying function.

Conditions A

A1. There exists $L \in RV(0)$ such that the distribution function F satisfies $1 - F(y) = y^{-1/\gamma} L(y)\{1 + o(1)\}$ as $y \to \infty$.

A2. We have $q_0(\tau) = \partial q(t)/\partial t|_{t=1-\tau}$ regularly varying at 0 with index $-\gamma - 1$. That is, for $x > 0$,

$$\lim_{\tau\to 0} \frac{q_0(x\tau)}{q_0(\tau)} = x^{-\gamma-1}.$$

A3. There exists an auxiliary function $f(x)$ such that $V \equiv Y - f(x)$ has the distribution function $F_V(y|x)$ satisfying, as $y \to \infty$,

$$1 - F_V(y|x) = H(x)\{1 - F(y)\}(1 + o(1)),$$

where $H(x) > 0$ is a positive, continuous, and bounded function on $[a, b]$ and has $E[H(X)] = 1$.

A4. For $q_V(\cdot|x) = F_V^{-1}(\cdot|x)$, $\partial q_V(\tau|x)/\partial\tau = \partial q(1 - (1-\tau)/H(x))/\partial\tau\{1 + o(1)\}$ uniformly in $x \in [a, b]$ as $\tau \to 1$.

Conditions A may not hold if either F or $F(\cdot|x)$ is not included in $D(G_\gamma)$. In other words, if F, $F(\cdot|x) \in D(G_\gamma)$, Conditions A are natural. Condition A1 is the formal notation of a Pareto-type tail (see Chernozhukov and Fernández-val 2011 [7]). The equivalent to condition A1 is

$$q(\tau) = (1-\tau)^{-\gamma} \bar{L}(1/(1-\tau))\{1 + o(1)\} \quad \text{as } \tau \to 1 \tag{2.1}$$

with $\bar{L} \in RV(0)$. Therefore, if the distribution F is continuous and $\partial \bar{L}(1/(1-\tau))/\partial\tau \to 0$ as $\tau \to 1$, condition A2 holds from condition A1. Actually, $q_0(\tau) = \partial q(t)/\partial t|_{t=1-\tau} = \gamma\tau^{-\gamma-1}\bar{L}(1/\tau) + \tau^{-\gamma}\partial\bar{L}(1/(1-t))/\partial t|_{t=1-\tau} = \gamma\tau^{-\gamma-1}\bar{L}(1/\tau)\{1 + o(1)\}$ as $\tau \to 0$. Therefore,

we have $\partial q(t)/\partial t|_{t=1-\tau} = \gamma\tau^{-\gamma-1}\bar{L}(1/\tau)\{1+o(1)\}$. Thus, condition A2 is weak. From A3, it is easy to show that $q_V(\tau|x) = H(x)^\gamma q(\tau)(1+o(1))$ as $\tau \to 1$. Furthermore, since $\tau = F_Y(q_Y(\tau|x)|x) = P(Y < q_Y(\tau|x)|x) = P(Y - f(x) < q_Y(\tau|x) - f(x)|x) = F_V(q_Y(\tau|x) - f(x)|x)$, we obtain $q_V(\tau|x) = q_Y(\tau|x) - f(x)$. Consequently, we have

$$q_Y(\tau|x) = f(x) + h(x)q(\tau)\{1+o(1)\} \quad \text{as } \tau \to 1, \tag{2.2}$$

where $h(x) = H(x)^\gamma$. Chernozhukov (2005) [6], Chernozhukov and Fernández-val (2011) [7] and Wang *et al.* (2012) [44] also provided this type of condition in multiple linear models. That is, they further assumed that $f(x) = x^T\beta$ and $h(x) = x^T c$ for $x = (x_1, \cdots, x_p)$, where β and c are unknown p-dimensional parameter vectors. Thus, A3 is the nonparametric model version of the above previous studies. Einmahl *et al.* (2016) [17] also considered the similar model to the survival function for more simple situation where time is the covariate. Condition A4 guarantees the existence of a conditional quantile density function (the derivative of the quantile function). Furthermore, the conditional quantile density function also behaves like a Pareto-type function by condition A4. Assumption A3 is strengthened by condition A4.

Remark 1. Let $U(t) = q(1-1/t) = \inf\{z|F(z) \geq 1-1/t\}$ and let $U_V(t|x) = q_V(1-1/t|x)$. In several articles (see, for example, de Haan and Ferreira 2006 [13]), the conditions of EVT are applied to $U(t)$ and $U_V(t|x)$ as $t \to \infty$. Since $q(\tau) = U(1/(1-\tau))$, the condition (2.1) is similar to $U(t) = t^\gamma L(t)\{1+o(1)\}$ with $t = 1/(1-\tau)$. Condition A4 can also be expressed as $\partial U_V(t|x)/\partial t = \partial U(tH(x))/\partial t$ with $t = 1/(1-\tau)$. Thus, we can reconsider the EVT conditions for quantiles by using U and U_V. In particular, the use of U is appropriate when using the second-order condition of EVT (see Section 3.2).

2.2. B-spline model

The conditional quantile $q_Y(\tau|x)$ can be written as

$$q_Y(\tau|x) = \text{argmin}_a \, E[\rho_\tau(Y-a)|X=x], \tag{2.3}$$

where $\rho_\tau(u) = u(\tau - I(u<0))$ is Koenker's check function (Koenker 2005 [29]) and I is the indicator function. The estimator of q_Y is often obtained along with an empirical version of (2.3). To estimate $q_Y(\tau|x)$, we use the B-spline regression method as the nonparametric technique in this paper. Let $\{B_k(x) : k = 1, ..., K+p\}$ be the pth degree of the B-spline basis with knots $a = \kappa_0 < \kappa_1 < ... < \kappa_K = b$. In addition, other sets of $2p$ knots are defined as $\kappa_{-p} = ... = \kappa_{-1} = a$ and $\kappa_{K+1} = ... = \kappa_{K+p} = b$. We then define the B-spline model as

$$s(x) = \sum_{k=1}^{K+p} B_k(x)b_k = \boldsymbol{B}(x)^T\boldsymbol{b},$$

where $\boldsymbol{B}(x) = (B_1(x), ..., B_{K+p}(x))^T$ and $\boldsymbol{b} = (b_1, ..., b_{K+p})^T$ is an unknown parameter vector. We now describe the relationship between the B-spline model and EVT discussed in the previous section. Let $W^m[a,b] = \{g|g^{(k)}$ is continuous, $k = 0, 1, ..., m-1$ and $\int_a^b \{g^{(m)}(x)\}^2 dx < \infty\}$ be the mth-order Sobolev space. From Barrow and Smith (1978) [1], for any function $g \in W^m([a,b])$, there exists $\boldsymbol{b}_g \in \mathbb{R}^{K+p}$ such that $g(x) - \boldsymbol{B}(x)^T\boldsymbol{b}_g = K^{-d}g^{(d)}(x)O(1)$ as $K \to \infty$, where $d = \min\{m, p+1\}$. For simplicity, we assume that $m \leq p+1$, that is $d = m$. Actually, $(p,m) = (3,2)$ is the standard condition of B-spline smoothing.

For $\tau \in (0,1)$, let

$$\boldsymbol{b}_0(\tau) = \text{argmin}_{\boldsymbol{b} \in \mathbb{R}^{K+p}} E[\rho_\tau(Y - s(x))|X = x]$$

and let $s_0(\tau|x) = \boldsymbol{B}(x)^T \boldsymbol{b}_0(\tau)$. We then found that $q_Y(\tau|x) = s_0(\tau|x) + K^{-m}\{\partial^m q_Y(\tau|x)/\partial x^m\}O(1)$ for $q_Y(\cdot|x) \in W^m[a,b]$. If $h \in W^m[a,b]$, Conditions A and (2.2) yield that $\{\partial^m q_Y(\tau|x)/\partial x^m\} = h^{(m)}(x)q(\tau)(1+o(1)) = O((1-\tau)^{-\gamma})$ and, hence, $s_0(\tau|x) - q_Y(\tau|x) = O(K^{-m}(1-\tau)^{-\gamma})$ as $\tau \to 1$ and $K \to \infty$, which indicates that the condition B4 below is required.

If f and h defined in (2.2) belong to $W^m[a,b]$, there exists $\boldsymbol{b}_f, \boldsymbol{b}_h \in \mathbb{R}^{K+p}$ such that $f(x) - \boldsymbol{B}(x)^T\boldsymbol{b}_f = O(K^{-m})$ and $h(x) - \boldsymbol{B}(x)^T\boldsymbol{b}_h = O(K^{-m})$. We then obtain $\boldsymbol{b}_0(\tau) = \boldsymbol{b}_f + q(\tau)\boldsymbol{b}_h + O(K^{-m}q(\tau))$ as $\tau \to 1$ and $K \to \infty$. Therefore, (2.1) and condition A4 indicate that $\partial s_0(\tau|x)/\partial\tau \sim \boldsymbol{B}(x)^T\boldsymbol{b}_h \partial q(\tau)/\partial\tau$ is satisfied since $\boldsymbol{b}_f$ and $\boldsymbol{b}_h$ are not dependent on τ. Thus, the B-spline model also holds (2.2) and condition A4 and, hence, the tail behavior of the B-spline model can be studied by using Conditions A. The following conditions are the fundamental assumptions for B-spline regression.

Conditions B

B1. For some constant $\nu > 0$, $E[|Y|^{2+\nu}|X = x] < \infty$.

B2. The functions f and h in (2.2) are included in $W^m[a,b]$.

B3. We have $\max_{1 \le j \le K}\{\kappa_{j+1} - \kappa_j\} = O(K^{-1})$.

B4. For some $\alpha \in (0,1)$, the number of knots $K = O(n^\alpha)$.

B5. As $\tau \to 1$ and $K \to \infty$, $K^m(1-\tau)^\gamma \to \infty$.

Condition B1 is needed to that the estimator satisfies the Lyapunov condition of central limit theorem. When condition B2 holds, the B-spline model can approximate to $q_Y(\tau|x)$. Conditions B3 and B4 are standard conditions for B-spline models. Together with condition B2, the B-spline model and EVT are connected for high quantile level. Condition B5 guarantees that the model bias between the conditional quantile and B-spline model converges to 0.

3. PENALIZED B-SPLINE ESTIMATOR FOR EXTREMAL QUANTILES

In this section, we define the nonparametric B-spline estimator and develop the asymptotic result. Then, we consider two scenario of extremal quantile rate: (i) intermediate order quantiles that $\tau \to 1$ and $(1-\tau)n \to \infty$ as $n \to \infty$ and (ii) extreme order quantiles that $\tau \to 1$ and $(1-\tau)n \to c \in [0,\infty)$ as $n \to \infty$. We denote the intermediate order quantile level by τ_I and the extreme order quantile level by τ_E, respectively. That is, as $n \to \infty$, $\tau_I, \tau_E \to 1$, $n(1-\tau_I) \to \infty$, $n(1-\tau_E) \to c \in [0,\infty)$, and $(1-\tau_I)/(1-\tau_E) \to \infty$.

3.1. Estimation of intermediate order quantiles

The ordinary B-spline quantile estimator for $\tau \in (0,1)$ is defined based on minimizing $\sum_{i=1}^{n} \rho_\tau(Y_i - s(x_i))$. However, it is known that the ordinary estimator tends to have a wiggly curve caused by data sparsity. To avoid this, we introduce the penalization method to control the behavior of the estimator. Although various types of penalties have been developed to prevent overfitting, we will use O'Sullivan's (1986) [37] penalty. For $\tau \in (0,1)$, the penalized spline estimator $\tilde{\boldsymbol{b}}(\tau) = (\tilde{b}_1(\tau), ..., \tilde{b}_{K+p}(\tau))^T$ of vector $\boldsymbol{b}(\tau) = (b_1(\tau), ..., b_{K+p}(\tau))^T$ is constructed by minimizing

$$\sum_{i=1}^{n} \rho_\tau(Y_i - s(x_i)) + \lambda \int_a^b \{s^{(m)}(x)\}^2 dx, \tag{3.1}$$

where $\lambda > 0$ is the smoothing parameter. Using $\tilde{\boldsymbol{b}}(\tau)$, for the intermediate order quantile level τ_I, we define

$$\tilde{q}_Y(\tau_I|x) = \boldsymbol{B}(x)^T \tilde{\boldsymbol{b}}(\tau_I).$$

We study the asymptotic theory for the penalized spline estimator $\tilde{q}_Y(\tau_I|x)$. Then, the conditions of the number of knots and the smoothing parameter included in $\tilde{q}_Y(\tau_I|x)$ are very important. The penalty $\int_a^b \{s^{(m)}(x)\}^2 dx$ can be written as $\boldsymbol{b}^T D_m^T R D_m \boldsymbol{b}$, where the $(K+p)$-matrix R has elements $R_{ij} = \int_a^b B_i(x) B_j(x) dx$ and the $(K+p-m) \times (K+p)$ matrix D_m satisfies $\boldsymbol{b}^{(m)} = D_m \boldsymbol{b}$, where $\boldsymbol{b}^{(m)} = (b_1^{(m)}, ..., b_{K+p-m}^{(b)})^T$, and for $m = 1, 2, ...,$

$$b_j^{(1)} = p \frac{b_j - b_{j-1}}{\kappa_{j+p} - \kappa_j}, \quad b_j^{(m)} = (p+1-m) \frac{b_j^{(m-1)} - b_{j-1}^{(m-1)}}{\kappa_{j+p+1-m} - \kappa_j}.$$

From now on, we use the symbols D_m and R. Let $G(h)$ be the $(K+p)$-matrix with elements $G_{ij} = \int_a^b h(x) B_i(x) B_j(x) dx$ and

$$\Lambda(h) = \Lambda(h, n, \tau_I) = \gamma^{-1} G(h^{-\gamma}) + \frac{\lambda q(\tau_I)}{(1-\tau_I)n} D_m^T R D_m.$$

Let $G = G(1)$ be $G(h)$ with $h(x) \equiv 1$. Define

$$K(m, \tau_I) = K \left(\frac{\lambda q(\tau_I)}{n(1-\tau_I)} \right)^{1/2m},$$

which controls the asymptotic scenario branch discussed in Remark 1 below.

Conditions C

C1. We have $K(m, \tau_I) \geq 1$.

C2. We have $K\{\lambda q(\tau_I)/n(1-\tau_I)\}^{1/2m} \to \infty$ as $n \to \infty$.

C3. We have $\lambda = o(q(\tau_I)^{-1}\{n(1-\tau_I)/q(\tau_I)\}^2)$ as $n \to \infty$.

Condition C concerns with the asymptotic property of the penalized spline estimator. C1 is detailed in Remark 2. C2 allows us to use the large K. If C3 fails, the asymptotic bias of the penalized spline estimator cannot be vanished. We now show the asymptotic distribution

of $\tilde{q}(\tau_I|x)$. First, we derive the two types of bias, model bias and shrinkage bias. Roughly speaking, the model bias is the bias between the B-spline model and the true function, and the shrinkage bias is the difference between the expectation of the penalized estimator and the unpenalized estimator. According to Section 2,2, the model bias is $b_a(\tau_I|x) = s_0(\tau_I|x) - q_Y(\tau_I|x) = O(K^{-m}q(\tau))$. This model bias becomes the negligible order from condition C2. That is, the bias is dominated by the shrinkage bias. Define

$$\begin{aligned} b_\lambda(\tau|x) &= \frac{\lambda q(\tau)}{(1-\tau)n}\boldsymbol{B}(x)^T\Lambda(H^{-\gamma})^{-1}D_m^T R D_m \boldsymbol{b}_0(\tau), \\ v^2(\tau|x) &= \frac{q(\tau)^2}{(1-\tau)n}\boldsymbol{B}(x)^T\Lambda(H^{-\gamma})^{-1}G\Lambda(H^{-\gamma})^{-1}\boldsymbol{B}(x). \end{aligned}$$

As a result, $b_\lambda(\tau_I|x)$ is the asymptotic shrinkage bias and $v(\tau_I|x)$ is the asymptotic variance of $\tilde{q}_Y(\tau_I|x)$. The following theorem shows the asymptotic order of the asymptotic bias and variance of the intermediate order quantile estimator.

Theorem 3.1. *Under Conditions A–C, as* $n \to \infty$,

$$b_\lambda(\tau_I|x) = O\left(q(\tau)\left(\frac{\lambda q(\tau_I)}{(1-\tau_I)n}\right)^{1/2}\right), \quad v^2(\tau_I|x) = O\left(\frac{q(\tau_I)^2}{(1-\tau_I)n}\left(\frac{\lambda q(\tau_I)}{(1-\tau_I)n}\right)^{-1/2m}\right).$$

From condition C3 and Theorem 3.1, we see that the shrinkage bias and variance converge to 0 as $n \to \infty$. Using the central limit theorem, Lyapunov's condition, and a Cramér–Wold device, the asymptotic normality of the estimator $\tilde{q}_Y(\tau_I|x)$ can be shown.

Theorem 3.2. *Suppose that Conditions A–C hold. As* $n \to \infty$, $b_\lambda(\tau|x)$ *and* $v^2(\tau|x)$ *are the asymptotic bias and variance of* $\tilde{q}_Y(\tau|x)$ *and*

$$\left(\frac{v(\tau_I|x)}{q_Y(\tau_Y|x)}\right)^{-1}\left\{\frac{\tilde{q}_Y(\tau_I|x)}{q_Y(\tau_Y|x)} - 1 - \frac{b_\lambda(\tau_I|x)}{q_Y(\tau_Y|x)}\right\} \xrightarrow{D} N(0,1).$$

Furthermore, under $\lambda = O(q(\tau_I)^2\{(1-\tau_I)n\}^{1/(2m+1)})$, *the optimal rate of convergence of the mean integrated squared error (MISE) of* $\tilde{q}_Y(\tau_I|x)/q_Y(\tau_I|x)$ *is*

$$E\left[\left\{\frac{\tilde{q}_Y(\tau_I|x)}{q_Y(\tau_I|x)} - 1\right\}^2\right] = O(\{(1-\tau_I)n\}^{-2m/(2m+1)}).$$

Theorems 3.1 and 3.2 yield that the trade-off between bias and variance is controlled by λ. Thus, this indicates that the careful choice of K is not important in the penalized spline methods. According to Yoshida (2013) [46], for the center quantile level τ, the MISE of the penalized spline quantile estimator has the order $O(n^{-2m/(2m+1)})$. Thus, the rate of convergence of the MISE of the penalized spline estimator for the intermediate quantile level is slower than that for the center quantile level. This result is not surprising in the context of the difficulties of the estimation for the tail quantile.

When $\boldsymbol{B}(x) = \boldsymbol{x}$ and $\lambda = 0$, the estimator is reduced to the ordinary quantile regression with the linear model. In the linear regression, the model bias is 0 and, hence, the bias term vanishes. On the other hand, since $G = E[XX^T]$ and $\Lambda(H^{-\gamma})^{-1} = \gamma^{-1}E[H(X)^{-\gamma}XX^T]^{-1}$, the asymptotic variance becomes

$$v^2(\tau_I|\boldsymbol{x}) = \frac{q(\tau_I)^2}{(1-\tau_I)n}\gamma^2\boldsymbol{x}^T E[H(X)^{-\gamma}XX^T]^{-1}E[XX^T]E[H(X)^{-\gamma}XX^T]^{-1}\boldsymbol{x},$$

which is similar in form to the asymptotic variance of the linear estimator of Lemma 3 of Wang *et al.* (2012) [44]. Then, the rate of convergence of the MISE of the linear estimator is $E[\{\tilde{q}_Y(\tau_I|x)/q_Y(\tau|x) - 1\}^2] = O(\{(1-\tau_I)n\}^{-1})$. Thus, it can be considered that Theorem 3.2 is the generalization of the asymptotic result of the linear-type parametric estimator.

Remark 2. Claeskens *et al.* (2009) [9] have studied the asymptotic properties of the penalized spline mean estimator in two scenarios: roughly speaking, case (a) small K scenario and case (b) large K scenario. In case (a), the asymptotic behavior of penalized splines is similar to that of regression splines, which have the unpenalized estimator ($\lambda \equiv 0$). Case (b) results in the penalized splines nearing the smoothing splines. We briefly describe the asymptotic scenario branch along with the result of Claeskens *et al.* (2009) [9]. The penalized spline mean estimator is obtained as $\hat{f}(x) = \boldsymbol{B}(x)^T(Z^TZ + (\lambda/n)D_m^TRD_m)^{-1}Z^T\boldsymbol{y}$, where $Z = (\boldsymbol{B}(x_1), ..., \boldsymbol{B}(x_n))^T$ is the design matrix and $\boldsymbol{y} = (y_1, ..., y_n)$. Then, the two asymptotic scenarios are divided by the asymptotic order of the maximum eigenvalue of $(Z^TZ + (\lambda/n)D_m^TRD_m)^{-1}$, which is obtained as $K(m)^{2m}$

$$K(m) = K\left(\frac{\lambda}{n}\right)^{1/2m}(1 + o(1)).$$

If $K(m) < 1$ for a sufficiently large n, K, and λ, we achieve case (a). When $K(m) > 1$ for a sufficiently large n, K, and λ, we achieve case (b). Although Claeskens *et al.* (2009) [9] focused only on mean regression, these two scenarios can also be discussed with respect to quantile regression. The asymptotic scenario branch discussed in this section is dependent on the asymptotic order of $\Lambda(h)^{-1}$. Similar to Claeskens *et al.* (2009) [9], the order of the maximum eigenvalue of $\Lambda(h)^{-1}$ can be obtained as $K(m, \tau_I)^{2m}$, which corresponds to $K(m)$ in mean regression. Consequently, condition C1 indicates that the large K scenario should be studied. We finally note why we focus on the large K scenario. Ruppert (2002) [41] recommended that one should first set the knots with a large K to obtain the overfitted estimator and control λ to achieve smoothness and fitness. Therefore, the large K scenario matches the concept of Ruppert (2002) [41] and this motivates us to consider the large K scenario.

3.2. Estimation of extreme order quantiles

For the extreme order quantile, the estimator $\tilde{q}_Y(\tau_E|x)$ discussed in the previous section would not have asymptotic normality (Chernozhukov and Fernández-val 2011 [7]). In this paper, we try to approximate the extreme conditional quantile from intermediate quantile. According to Weissman (1978) [45], the following holds:

$$q_Y(\tau_E|x) \approx \left(\frac{1-\tau_I}{1-\tau_E}\right)^{\gamma} q_Y(\tau_I|x).$$

From this, using the estimator of the intermediate order quantile $\tilde{q}_Y(\tau|x)$, we define the extrapolated estimator of the extreme order quantile. To achieve this, we need to estimate the EVI γ.

Let $\tau_1 > ... > \tau_k$ be the sequence of quantile levels, where $\tau_j = 1 - ([n^\eta] + j)/(n+1)$, $\eta \in (0, 1)$ and $[a]$ is the integer part of a. Then, since $(1-\tau_j)n = n([n^\eta] + j)/(n+1) \to \infty$ as

$n \to \infty$, all τ_j are intermediate order quantiles. Using this sequence, we define the Hill-types estimator of γ as

$$\hat{\gamma}(x) = \frac{1}{k-1} \sum_{j=1}^{k-1} \log \left(\frac{\tilde{q}_Y(\tau_j|x)}{\tilde{q}_Y(\tau_k|x)} \right).$$

In this paper, we assume that the tail behavior of $F_Y(\cdot|x)$ and $F(\cdot)$ are equivalent (see Condition A1). Therefore, it is somewhat unnatural that the estimator of γ varies with x. Nevertheless, we define the extrapolated estimator with $\hat{\gamma}(x)$ and investigate the mathematical property. For $x \in [a, b]$, using $\hat{\gamma}(x)$, we define the estimator of the extreme order quantile as

$$\hat{q}_Y(\tau_E|x) = \left(\frac{1-\tau_I}{1-\tau_E} \right)^{\hat{\gamma}(x)} \tilde{q}_Y(\tau_I|x).$$

We next consider the EVI estimator along with condition A1. Define the common index (pooled) estimator

$$\hat{\gamma}^C = \frac{1}{n} \sum_{i=1}^{n} \hat{\gamma}(x_i)$$

and the extrapolated estimator with common index estimator $\hat{\gamma}^C$ as

$$\hat{q}_Y^C(\tau_E|x) = \left(\frac{1-\tau_I}{1-\tau_E} \right)^{\hat{\gamma}^C} \tilde{q}_Y(\tau_I|x).$$

To investigate the asymptotic distribution of $\hat{q}_Y(\tau_E|x)$ and $\hat{q}_Y^C(\tau_E|x)$, we impose the second-order condition in Conditions A:

A5. The function $U(t) = F^{-1}(1-1/t)$ satisfies the second-order condition with (γ, ρ, A). That is, there exist $\rho < 0$ and $A(t) \in RV(\rho)$ such that as $t \to \infty$,

$$A(t)^{-1} \left\{ \frac{U(tz)}{U(t)} - z^\gamma \right\} \to \frac{(z^\gamma)(z^\rho - 1)}{\rho}.$$

Furthermore, $A(t) = \gamma d t^\rho$ with $d \neq 0$.

A3′. There exist $\delta > 0$ and positive, continuous and bounded function $H_1(x)$ such that as $y \to \infty$,

$$1 - F_V(y|x) = H(x)\{1 - F(y)\} + H_1(x)(1 - F(y))^{1+\delta}(1 + o(1))$$

Condition A5 is the standard second-order condition of EVT and is detailed in de Haan and Ferreira (2006) [13]. Condition A3′ provides the second order of tail behavior of $F(y)$. From conditions A5 and A3′, we see that $U_Y(\cdot|x)$ also satisfy the second-order condition with $(\gamma, \rho^* = \min\{\rho, -\delta\}, A^*(\cdot|x))$ and $A^*(t|x) = \gamma d^*(x) t^{\rho^*}$ with $d^*(x) \neq 0$, which were proven in Lemma 2 of Wang *et al.* (2012) [44]. Using this, we show the asymptotic property of the Hill-type estimator of the EVI in the following.

Theorem 3.3. *Suppose that the smoothing parameter included in* $\tilde{q}(\tau_I|x)$ *satisfies* $\lambda = O(q(\tau_I)^2\{(1-\tau_I)n\}^{1/(2m+1)})$. *Furthermore, suppose that* $k \to \infty$, $k/n \to 0$, $n^\eta \log(k)/k^{m/(2m+1)} \to 0$ *and* $k^{m/(2m+1)}(n/k)^{\max\{-\gamma,-\delta,\rho\}} \to 0$ *as* $n \to \infty$. *Under Conditions A–C, as* $n \to \infty$,

$$\frac{\hat{\gamma}(x) - \gamma - b(k|x)}{v(k|x)} \xrightarrow{D} N(0,1),$$

and

$$\frac{\hat{\gamma}^C - \gamma - E[b(k|X)]}{E[v(k|X)]} \xrightarrow{D} N(0,1),$$

where $b(k|x)$ *and* $v(k|x)$ *are defined in (B.4) of Appendix B and have an asymptotic order* $O(k^{-m/(2m+1)})$. *Furthermore,*

$$E[\{\hat{\gamma}(x) - \gamma\}^2] = O\left(k^{-2m/(2m+1)}\right)$$

and

$$E[\{\hat{\gamma}^C - \gamma\}^2] = O\left(k^{-2m/(2m+1)}\right).$$

Using Theorem 3.3, we obtain the asymptotic normality of the ratio of $\hat{q}_Y(\tau_E|x)$ and $\hat{q}_Y^C(\tau_E|x)$.

Theorem 3.4. *Suppose that the same conditions as Theorem 3.3. Furthermore, assume that* $k^{-m/(2m+1)} \log\{(1-\tau_I)/(1-\tau_E)\} \to 0$ *as* $k, n \to \infty$, $\tau_I, \tau_E \to 1$. *As* $n \to \infty$, $\tau_E \to 1$ *and* $n(1-\tau_E) \to c \in [0, \infty)$,

$$\frac{\frac{\hat{q}_Y(\tau_E|x)}{q_Y(\tau_E|x)} - 1 - \text{bias}(\tau_E|x)}{s(\tau_E|x)} \xrightarrow{D} N(0,1)$$

and

$$\frac{\frac{\hat{q}_Y^C(\tau_E|x)}{q_Y(\tau_E|x)} - 1 - \text{bias}^C(\tau_E|x)}{s^C(\tau_E|x)} \xrightarrow{D} N(0,1)$$

where $\text{bias}(\tau_E|x)$, $s(\tau_E|x)$, $\text{bias}^C(\tau_E|x)$ *and* $s^C(\tau_E|x)$ *are defined in (B.5), (B.6), (B.7) and (B.8) of Proof of Theorem 3.4 on Appendix B. Furthermore,*

$$E\left[\left\{\frac{\hat{q}_Y(\tau_E|x)}{q_Y(\tau_E|x)} - 1\right\}^2\right] = O\left(\max\left\{k^{-\frac{2m}{2m+1}} \log^2\left(\frac{1-\tau_I}{1-\tau_E}\right), \{(1-\tau_I)n\}^{-\frac{2m}{2m+1}}\right\}\right).$$

and

$$E\left[\left\{\frac{\hat{q}_Y^C(\tau_E|x)}{q_Y(\tau_E|x)} - 1\right\}^2\right] = O\left(\max\left\{k^{-\frac{2m}{2m+1}} \log^2\left(\frac{1-\tau_I}{1-\tau_E}\right), \{(1-\tau_I)n\}^{-\frac{2m}{2m+1}}\right\}\right).$$

For the asymptotic order in Theorem 3.4, the term $O(k^{-2m/(2m+1)} \log^2\{(1-\tau_I)/(1-\tau_E)\})$ is derived from $\hat{\gamma}(x)$ and the another term is derived from $\tilde{q}_Y(\tau_I|x)$. If we use $\tau_I = \tau_1 = (n - [n^\eta])/(n+1)$ or $\tau_I = O(\tau_1)$, we have $(1-\tau_I)n = O([n^\eta]) = o(k)$ since $[n^\eta]/k \to 0$.

That is, the asymptotic inference of $\hat{q}_Y(\tau_E|x)$ is dominated by that of $\tilde{q}_Y(\tau_I|x)$ and hence, the rate of convergence of the estimator is

$$E\left[\left\{\frac{\hat{q}_Y(\tau_E|x)}{q_Y(\tau_E|x)}-1\right\}^2\right]=O\left(\{(1-\tau_I)n\}^{-\frac{2m}{2m+1}}\right).$$

One may have sense of discomfort with this result since the extreme order quantile estimator and the intermediate order quantile estimator has same rate of convergence. Indeed, the leading terms of $\tilde{q}_Y(\tau_I|x)$ and $\hat{q}_Y(\tau_E|x)$ are similar. However, the convergence speed of the subsequent term of $\hat{q}_Y(\tau_E|x)$ is obviously slower than that of $\tilde{q}_Y(\tau_I|x)$ because of the influence of $\hat{\gamma}(x)$. Therefore, for the application with a finite sample, the behavior of $\tilde{q}_Y(\tau_I|x)$ would be more stable than $\hat{q}_Y(\tau_E|x)$. On the other hand, when $\tau_I=\tau_k$ or $\tau_I=O(\tau_k)$, which leads to $n(1-\tau_I)=O(k)$, is adopted, $\hat{q}_Y(\tau_E|x)$ is heavily affected by $\hat{\gamma}(x)$ but not by $\tilde{q}_Y(\tau_I|x)$. Actually, since $n(1-\tau_E)\to c\in[0,\infty)$ and $\log(\{1-\tau_I\}/\{1-\tau_E\})=\log(\{n(1-\tau_I)\}/\{n(1-\tau_E)\})=O(\log[k/\{n(1-\tau_E)\}])$, we have

$$\begin{aligned}E\left[\left\{\frac{\hat{q}_Y(\tau_E|x)}{q_Y(\tau_E|x)}-1\right\}^2\right] &= O\left(\max\left\{k^{-\frac{2m}{2m+1}}\log^2\left(\frac{k}{n(1-\tau_E)}\right),k^{-\frac{2m}{2m+1}}\right\}\right)\\ &= O\left(k^{-\frac{2m}{2m+1}}\log^2\left(\frac{k}{n(1-\tau_E)}\right)\right).\end{aligned}$$

For the common index quantile estimator $\hat{q}_Y^C(\tau_E|x)$ with $\tau_I=O(\tau_k)$, the asymptotic order of $\hat{q}_Y^C(\tau_E|x)/q_Y(\tau_E|x)$ are dominated by the term of $\hat{\gamma}^C$, and this do not vary with x. This result is quite unnatural in the quantile regression although $O(\log^2(k/\{n(1-\tau_E)\}))$, which is the difference between the asymptotic inference of $\hat{\gamma}^C$ and $\tilde{q}_Y(\tau_I|x)$, is quite small. Therefore, if the common index quantile estimator is mainly used, we may have to choose the baseline quantile τ_I so that $\tau_I>\tau_k$. Thus, the balance of τ_I and k controls the asymptotic behavior of $\hat{q}_Y(\tau_E|x)$. The same is true of $\hat{q}_Y^C(\tau_E|x)$.

Wang *et al.* (2012) [44] obtained the extrapolated estimator in the linear model with $\tau_I=\tau_k$. From their result, the rate of convergence of the MISE of the linear estimator is $E[\{\hat{q}_Y(\tau_E|x)/q_Y(\tau_E|x)-1\}^2]=O(k^{-1}\log^2(\{1-\tau_I\}/\{1-\tau_E\}))$. That is, the difference in the rate of convergence between the parametric estimator and the nonparametric estimator is k^{-1} and $k^{-2m/(2m+1)}$, which could be intuitively derived from the classical works on parametric and nonparametric regression.

Remark 3. The intermediate order quantile and the extreme order quantile are separated mathematically by the rate of the quantile level. However, in data analysis, the distinction between these two rates should be drawn for fixed n. Define $\xi=\xi(\tau,n)=(1-\tau)n$. Using ξ, Chernozhukov and Fernández-val (2011) [7] suggested the following rule of thumb. For the quantile level τ, if $\xi<30$, it is the extreme order inference, that is, $\tau=\tau_E$ and we should use $\hat{q}_Y(\tau|x)$. When $\xi\geq 30$, it is sufficient to use the intermediate order quantile estimator $\tilde{q}_Y(\tau|x)$. If the predictor is the continuous, this threshold is $\xi=15$–20. However, they noted that the above rule is conservative. In this paper, we treat $\tau_1=\frac{n-[n^\eta]}{n+1}$ as the intermediate order quantile. For example, when $n=200$, $\tau=0.925$ leads to $\xi=15$. Then, we have $\eta\approx 0.5$. For $n=1000$, $\tau=0.985$ and $\eta=0.4$ correspond to $(1-\tau_1)n\approx(1-\tau)n=15$.

On the other hand, Wang *et al.* (2012) [44] suggested to use $\eta = 0.1$. In their rule, we have $\xi(\tau_1, n) = 3$ for $n < 1025$. Thus, it seems that the rule of Chernozhukov and Fernández-val (2011) [7] is more conservative rather than that of Wang *et al.* (2012) [44]. In our experience, the rule of Wang *et al.* (2012) [44] worked well for $n \leq 1000$. Therefore, in the simulation study of the next section, we also use $\eta = 0.1$. However, the determination of the split of the intermediate order and the extreme order is still a difficult problem and further study would be welcomed.

4. SIMULATION

The practical performance of the proposed estimator is confirmed by Monte Carlo simulation. Define the true regression model as

$$Y_i = f(X_i) + \sigma(X_i)\varepsilon_i(X_i), \quad i = 1, ..., n, \tag{4.1}$$

where

$$f(x) = \sqrt{x(1-x)}\sin\left(\frac{2\pi(1+2^{-7/5})}{x+2^{-7/5}}\right)$$

and $\sigma(x) = 10^{-1}(1+x)$. The predictor X_i is independently generated from the standard uniform distribution. This setting was introduced by Daouia *et al.* (2013) [11]. We consider two types of error distribution: (a) $\varepsilon_i(X_i) = \varepsilon_i \sim t_5$ and (b) $\varepsilon_i(X_i) \sim t_{s(x)}$, where

$$s(x) = [\nu(x)] + 1, \quad \nu(x) = [\{1.1 - 0.5\exp[-64(x-0.5)^2]\}\{0.1 + \sin(\pi x)\}]^{-1}.$$

The error type (b) is also used by Daouia *et al.* (2013) [11]. For the t_ν distribution, the EVI is $\gamma = 1/\nu$ and hence, $\gamma = 0.2$ and $\gamma(x) = 1/s(x)$ for (a) and (b), respectively For both cases, the EVI is larger than 0, which indicates that the distribution of Y_i has a heavy tail. In (4.1), the conditional τth quantile of Y given $X = x$ is $q_Y(\tau|x) = f(x) + \sigma(x)q_\varepsilon(\tau|x)$, where $q_\varepsilon(\tau|x)$ is the τth quantile of $\varepsilon_i(x)$. For the case (a), $q_\varepsilon(\tau|x) = q(\tau)$ is the τth quantile of t_5 and is not dependent on x. Thus, the model (4.1) with (a) is the location-scale shifted model and is of the form of (2.2). In the case of (b), $\gamma(x) = 1/2$ for $x \in [0.12, 0.88]$ and $\gamma(x) \in (0, 1/3)$ otherwise. That is, the model (4.1) with (b) has high EVI at the center and low (but larger than 0) EVI at the boundary. The conditional quantile with (b) fail due to Conditions A. However, it is important to confirm the performance of the estimator under (b).

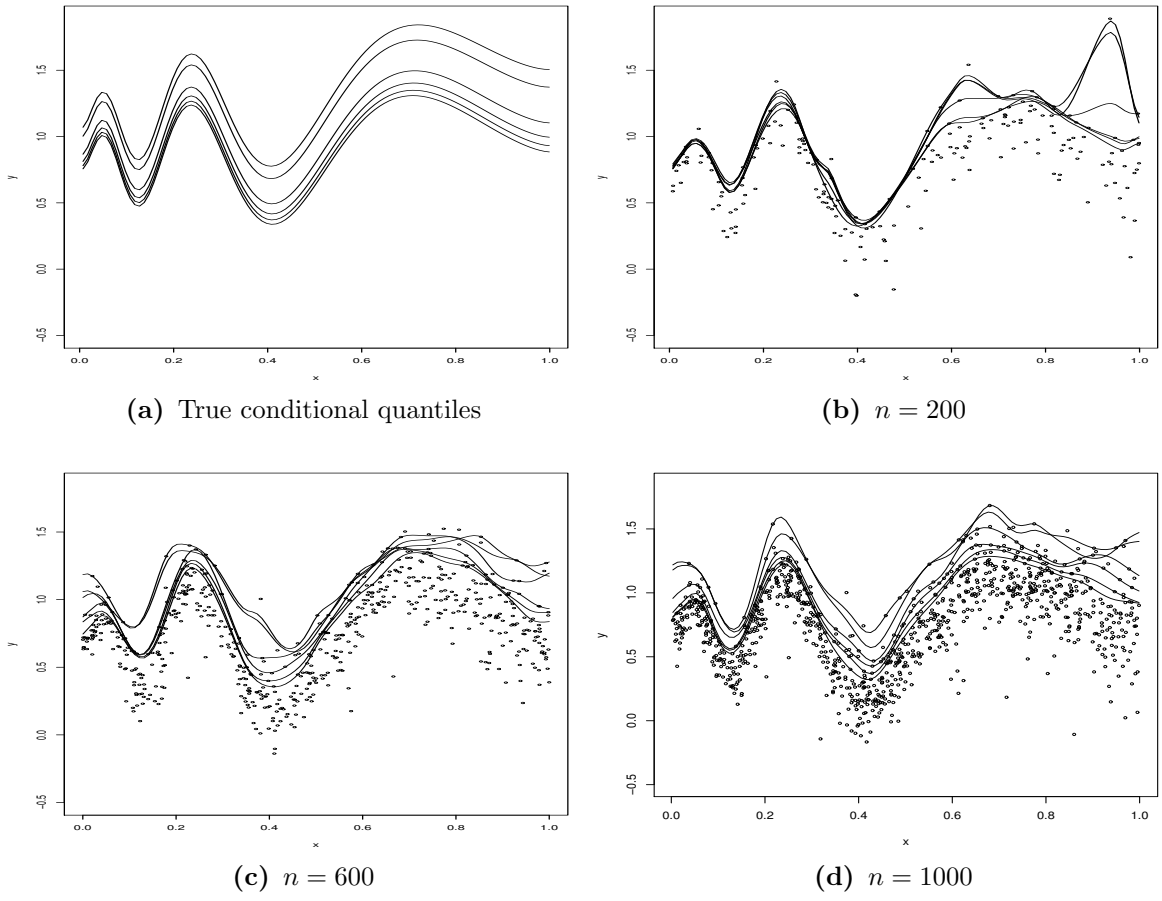
We construct the intermediate order quantile estimator $\tilde{q}_Y(\tau|x)$, the Hill-type estimator $\hat{\gamma}(x), \hat{\gamma}^C$, and the extreme order quantile estimator $\hat{q}_Y(\tau|x)$ and $\hat{q}_Y^C(\tau|x)$. For the intermediate order quantile estimator, we use the number of knots $K = 40$ and the smoothing parameter selected via generalized approximated cross-validation (Yuan 2006 [47]). To obtain $\hat{\gamma}$, $\hat{\gamma}^C$, $\hat{q}_Y$ and $\hat{q}_Y^C$, we need to determine $\tau_j = 1 - \frac{[n^\eta]+j}{n+1}$ $(j = 1, ..., k)$. In this simulation, η is chosen so that $(1-\tau_1)n = \xi = 3$ and $k = [7.5n^{1/3}]$. Such η and k are selected from a pilot study. Wang *et al.* (2012) [44] used $\eta = 0.1$ and $k = [4.5n^{1/3}]$ in the linear regression. Thus, our k is somewhat larger than that of Wang *et al.* (2012) [44].

For the estimator $\hat{f}(x)$ of the true function $f(x)$, the Mean Integrated Squared Error (MISE):

$$MISE(\hat{f}) = \int_0^1 E[\{\hat{f}(x) - f(x)\}^2]dx$$

is used as the accuracy measure of the estimator. We calculate the estimated MISE of $\tilde{q}_Y(\tau|x)$, $\hat{q}_Y(\tau|x)$ and $\hat{q}_Y^C(\tau|x)$ over 400 replications. The estimators $\tilde{q}_Y(\tau|x)$, $\hat{q}_Y(\tau|x)$ and $\hat{q}_Y^C(\tau|x)$ are denoted by PSE-I, PSE-E and PSE-Ep. As the competitors, we consider the functional nonparametric estimator (Gardes *et al.* 2010 [20]) and the kernel smoothing estimator (Daouia *et al.* 2013 [11]). The estimators $\hat{q}_1(\tau, x)$ and $\hat{q}_2(\tau, x)$ defined in Gardes *et al.* (2010) [20] are denoted by FNS-I and FNS-E, respectively. Furthermore, the estimators $\hat{q}_n(\tau|x)$ and $\tilde{q}_n^{\mathrm{RP}}(\tau|x)$ defined by Daouia *et al.* (2013) [11] are labeled by KSE-I and KSE-E in this section. Thus, FNS-I, FNS-E, KSE-I and KSE-E are also demonstrated in simulation.

We report the simulation results for the case (a). Figure 1 shows the true conditional quantiles and the intermediate order quantile estimators for one dataset with $n = 200, 600$ and 1000. For $\tau = 0.8, 0.85$ and 0.9, the estimator behaved well, but for $\tau \geq 0.95$, there was a significant difference between the true function and the intermediate order quantile estimator.

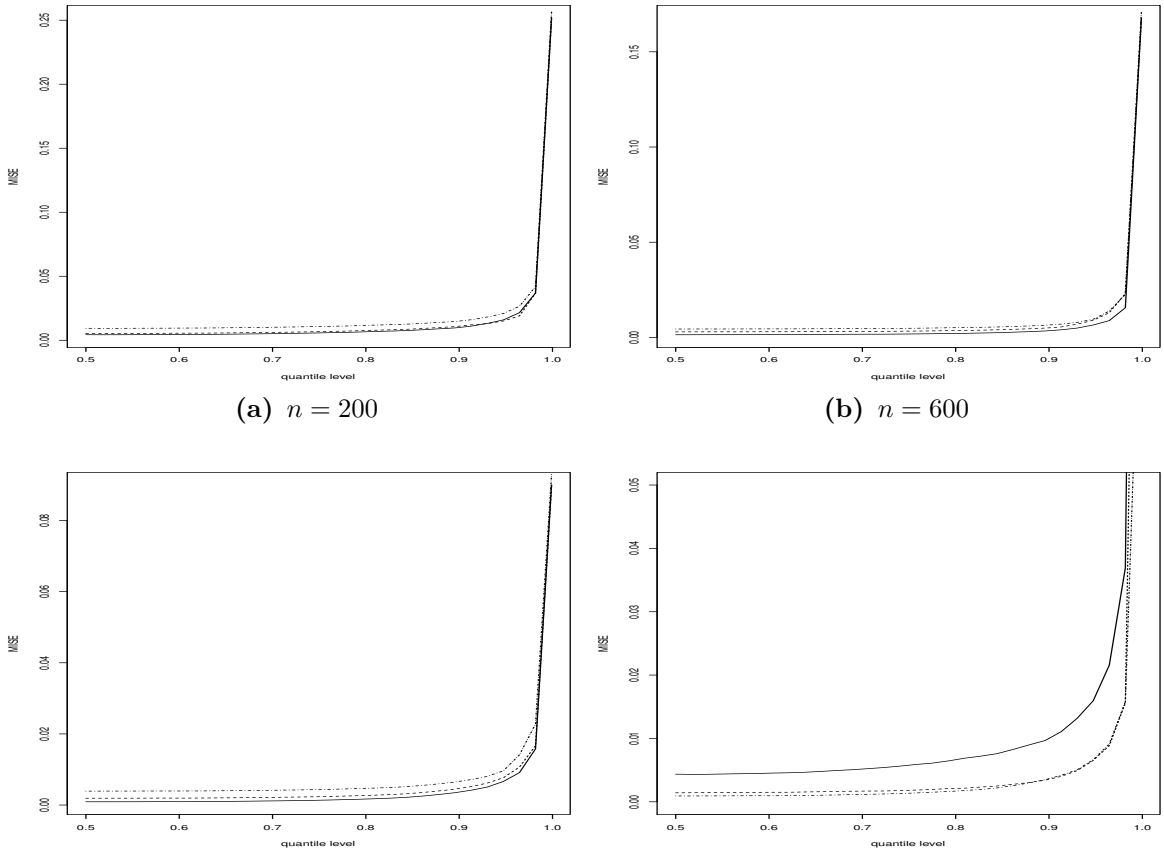


(a) True conditional quantiles **(b)** $n = 200$

(c) $n = 600$ **(d)** $n = 1000$

Figure 1: True conditional quantiles for $\tau = 0.8, 0.85, 0.9, 0.95, 0.99$ and 0.995, and these intermediate order quantile estimators for one dataset.

In Figure 2, the MISEs of the estimators for $\tau \in [0.5, 0.995]$ are illustrated. We can observe that the proposed estimator behaves better than the competitors. From Figure 2 (d), we can find that the estimator behaves well as n increases. This indicates that the estimator has a consistency property. However, as τ increases, the performance of the estimator becomes drastically decreases. Therefore, for $\tau \approx 1$, it is difficult to predict the conditional quantile using the intermediate order quantile estimator.



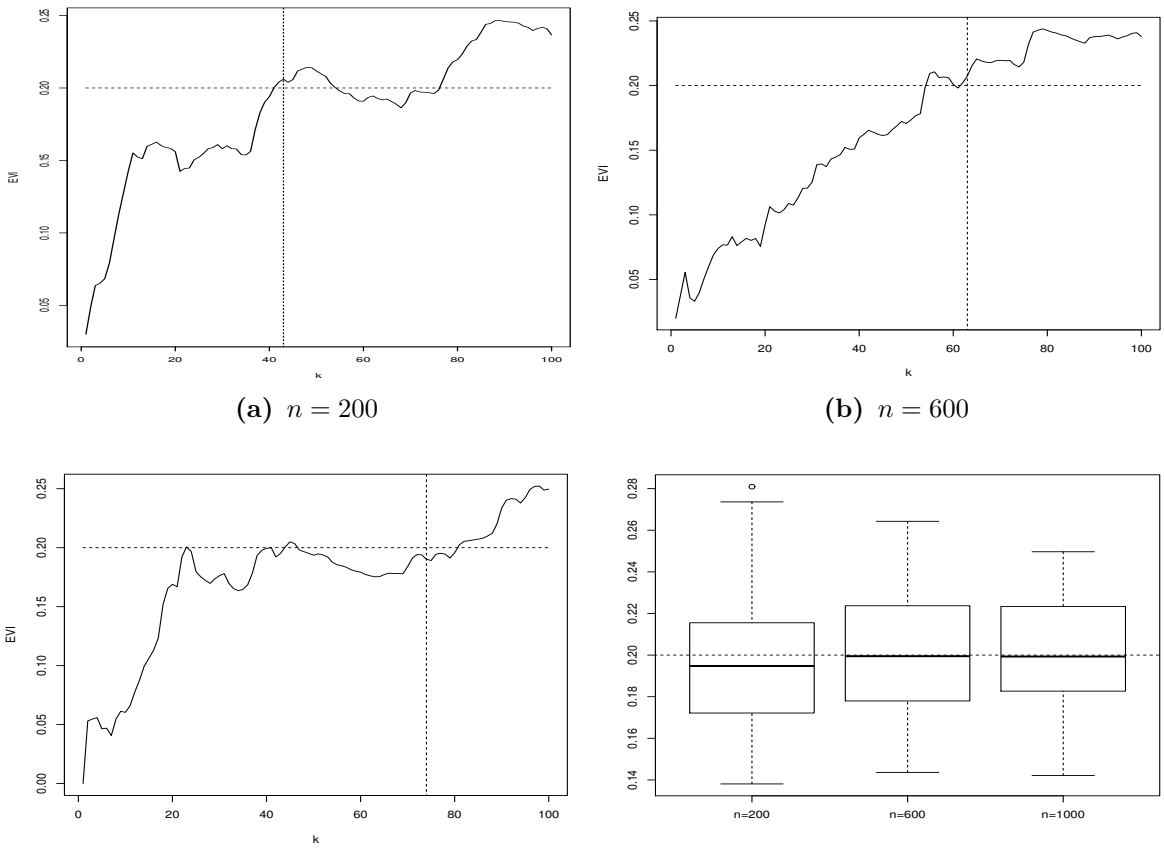
(a) $n = 200$ **(b)** $n = 600$

(c) $n = 1000$ **(d)** PSE-I

Figure 2: MISE of the intermediate order quantile estimators for $\tau \in [0.5, 0.995]$. In (a), (b) and (c), the solid line is PSE-I. The dashed line and dot-dashed lines are KSE-I and FNS-I, respectively. In (d), the solid, dashed and dot-dashed lines are PSE for $n = 200$, $n = 600$ and $n = 1000$.

We next show the performance of the EVI estimator. Figure 3 shows the behavior of $\hat{\gamma}^C$ over k for one dataset and the distribution of $\hat{\gamma}^C$ using $k = [7.5n^{1/3}]$ by Monte Carlo simulation. From the results, we see that the suggested $k = [7.5n^{1/3}]$ is good choice. When $n = 1000$, the behavior of $\hat{\gamma}^C$ is stable from (c) and (d).

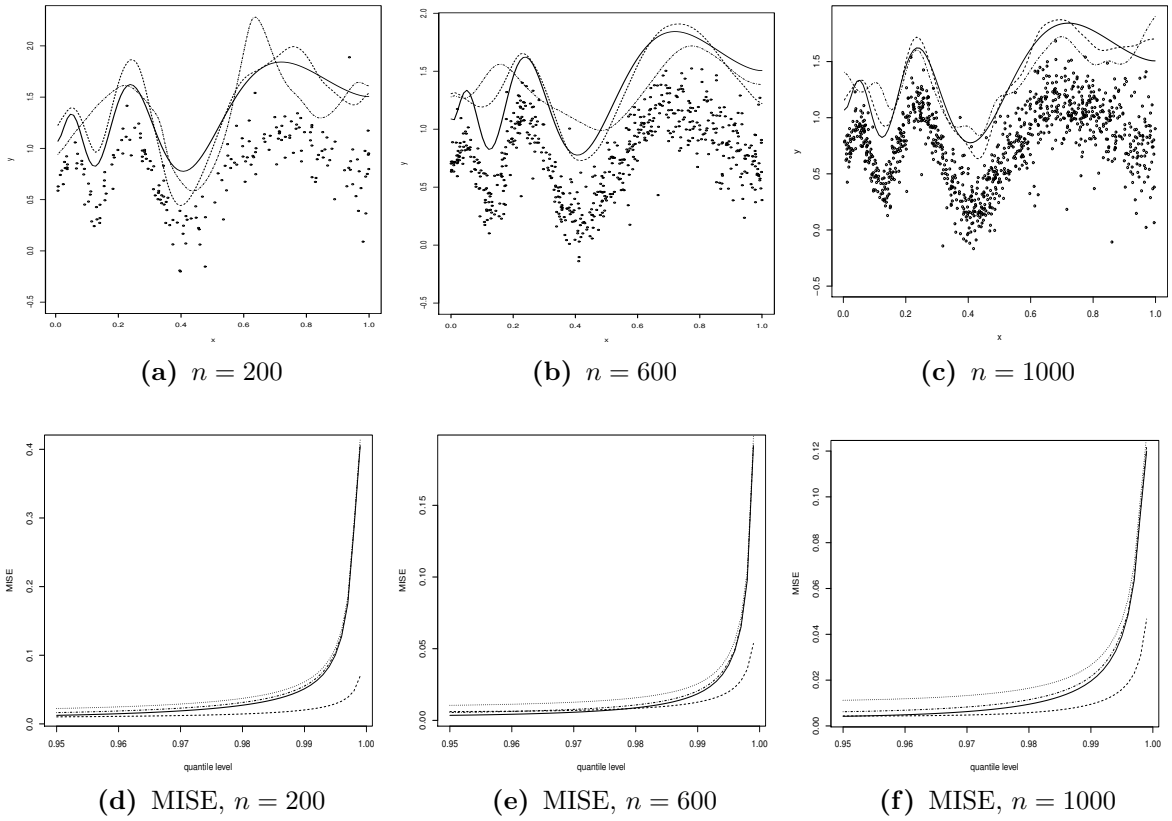
Figure 4 shows the extreme order quantile estimators PSE-E and PSE-Ep for one dataset and the MISE of the extreme order quantile estimators for $\tau \in [0.95, 0.999]$. From (a–c), we can observe that the estimator behaves well. We can see that the behavior of the PSE-Ep is stable than the PSE-E. This is not a surprising result since the estimator of EVI included in PSE-Ep is not dependent on x unlike PSE-E. It can be recognized from Figure 4 (d–f) that the proposed estimator has better behavior than the competitors although the differences are not large. Furthermore, the performance of the PSE-Ep was superior to that of PSE-E. We think that this is a result of the stability of $\hat{\gamma}^C$. It can be recognized from Figure 5 that the extrapolated estimator has consistency.



(a) $n = 200$ **(b)** $n = 600$

(c) $n = 1000$ **(d)** The distribution of $\hat{\gamma}^C$

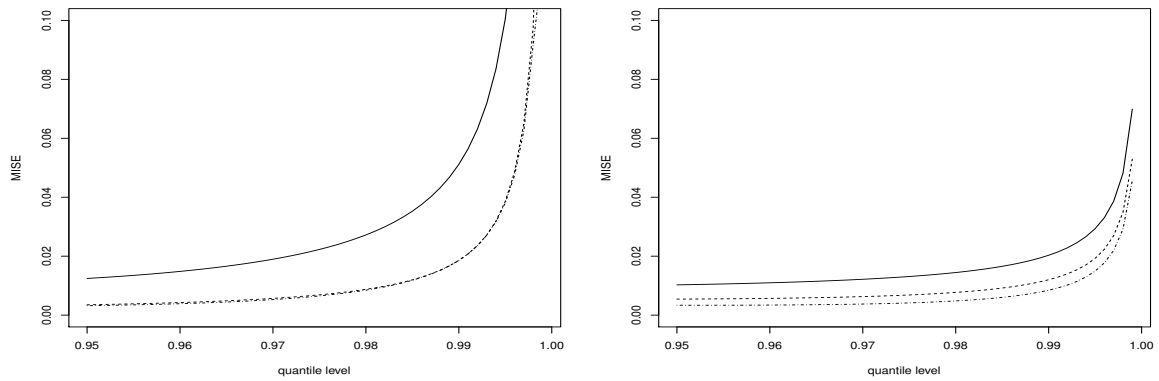
Figure 3: (a–c) The sample path of the common index estimator of EVI with k for one dataset. The dataset is similar to Figure 1 for each n. The dashed line is $k = [7.5n^{1/3}]$ and $\gamma = 0.2$. (d) Box plot of $\hat{\gamma}^C$ with $k = [7.5n^{1/3}]$ from 400 replications.



(a) $n = 200$ **(b)** $n = 600$ **(c)** $n = 1000$

(d) MISE, $n = 200$ **(e)** MISE, $n = 600$ **(f)** MISE, $n = 1000$

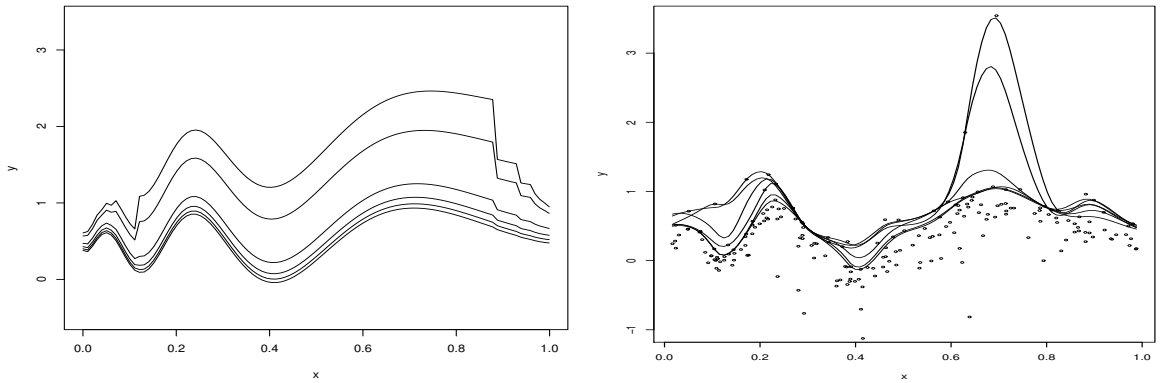
Figure 4: (a–c) The true conditional quantiles (solid) and the extreme order quantile estimator PSE-E (dot-dashed) and PSE-Ep (dashed) for $\tau = 0.995$ for one dataset. The dataset is similar to that given in Figure 1 for each n.
(d–f) MISE of the estimators for $\tau \in [0.95, 0.999]$. The solid, dashed, dotted and dot-dashed lines are PSE-E, PSE-Ep, FNS-E and KSE-E, respectively.



(a) MISE, PSE-E **(b)** MISE, PSE-Ep

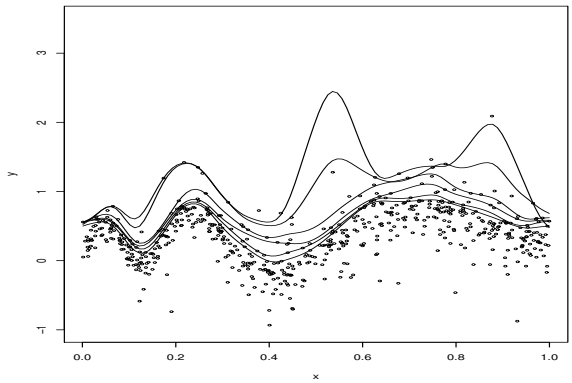
Figure 5: (a) and (b) are the MISE of the PSE-E and the PSE-Ep, respectively. The results are similar to Figure 4 (d–f). For both panels, solid, dashed and dot-dashed lines are for $n = 200$, 600 and 1000, respectively.

From now on, we describe the simulation results for the model (b). Figure 6 shows the true conditional quantiles and the intermediate order quantile estimators for $\tau \in [0.8, 0.995]$ for one dataset. It appears that for $\tau \in [0.8, 0.9]$, the estimator can capture the true conditional quantile even for $n = 200$. However, for $\tau \geq 0.95$, the estimator has a wiggly curve. In Figure 7, the results of MISE of the intermediate order quantile estimators for each n are illustrated. We found that the proposed estimator performs well for $\tau \in [0.5, 0.95)$. However, the MISE drastically grows as τ increases. The behaviors of PSE-E and PSE-Ep for one dataset are described in Figure 8 (a–c). It can be seen from Figures 6 and 8 (a–c) that the PSE-E and PSE-Ep performed better than PSE-I. Figure 8 (d)–(f) shows the MISE of the extreme order quantile estimators. It can be confirmed that the performance of PSE-E is slightly better than that of PSE-Ep. We see that the proposed estimators have better behavior than the competitors. Figure 9, the consistency of the PSE-E and the PSE-Ep can be observed in numerically. Although the performance of the proposed estimator is drastically superior to that of Daouia *et al.* (2013) [11], this simulation result indicates that our method is one of useful tools to the problem of extremal quantile regression.

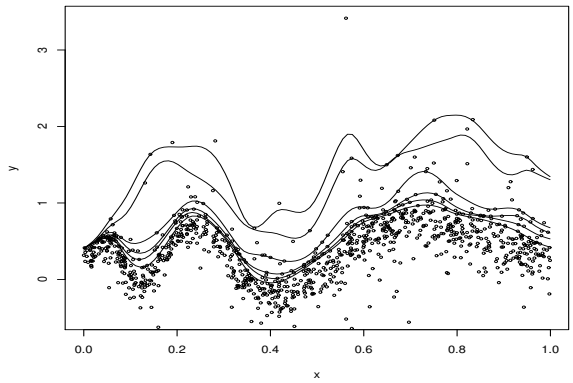


(a) True conditional quantiles

(b) $n = 200$

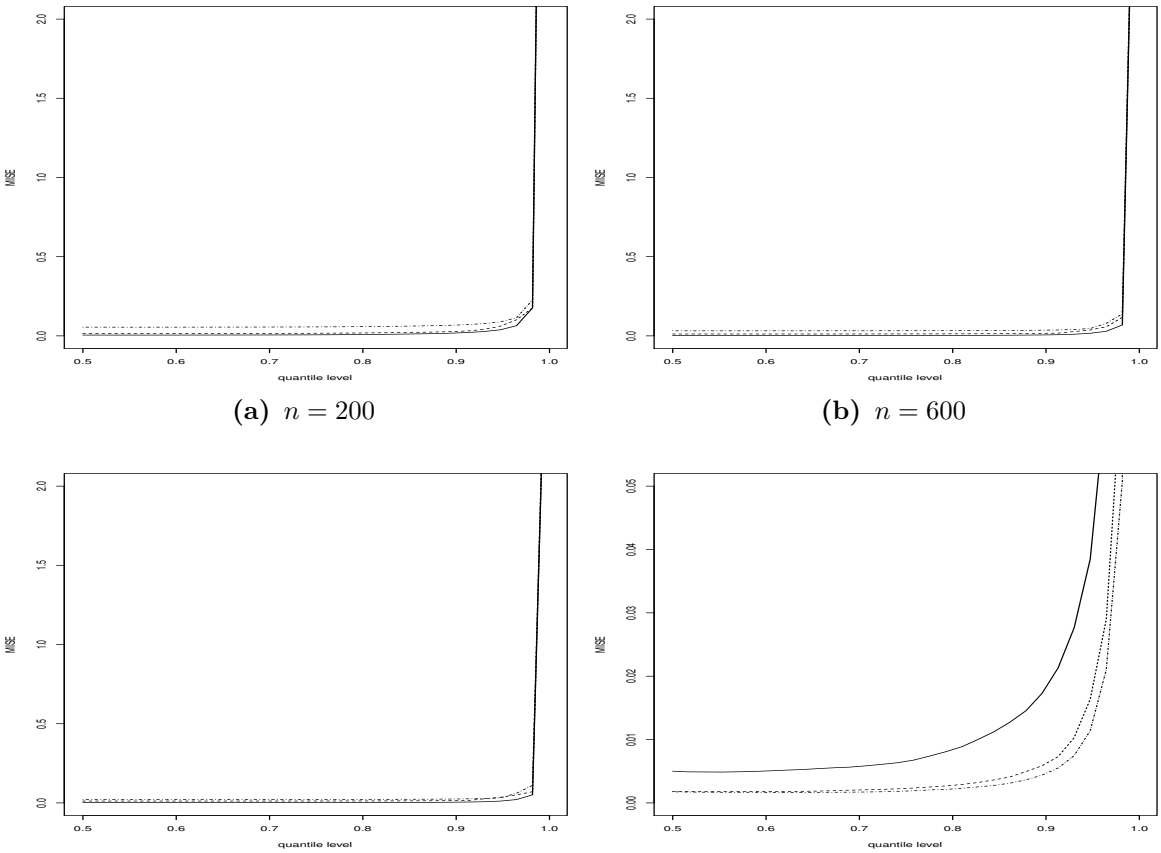


(c) $n = 600$



(d) $n = 1000$

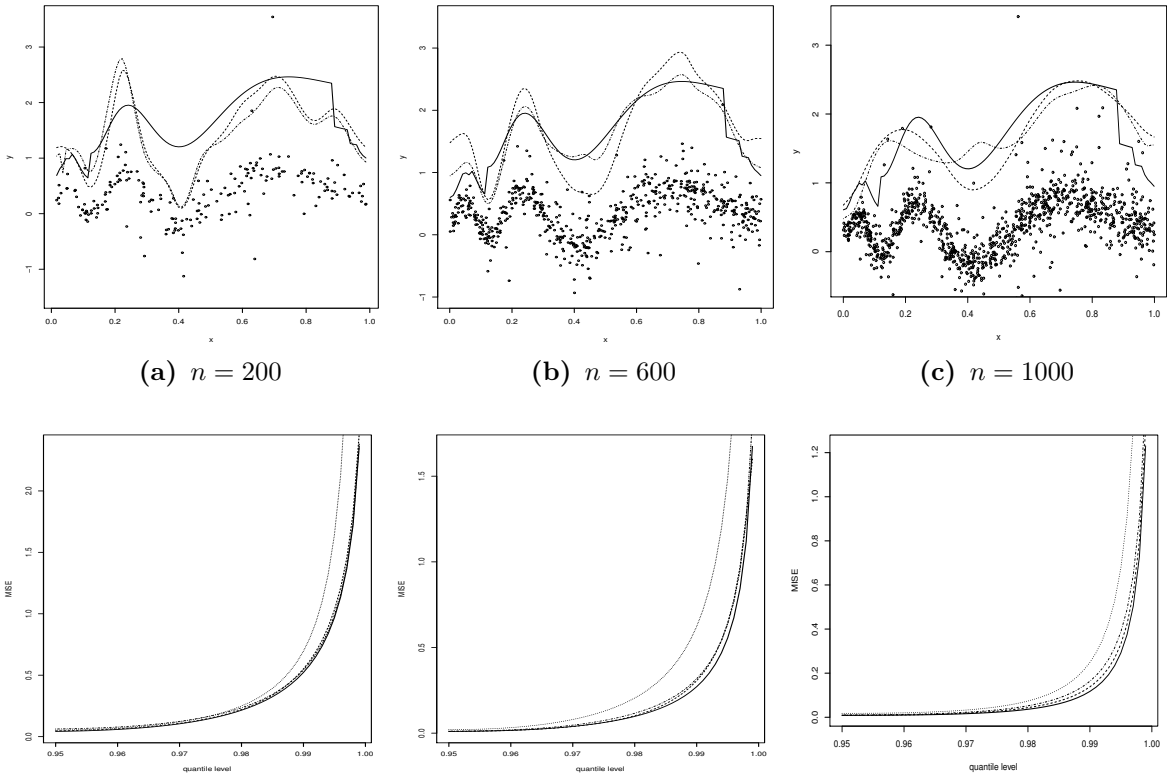
Figure 6: True conditional quantiles for $\tau = 0.8, 0.85, 0.9, 0.95, 0.99, 0.995$ and these intermediate order quantile estimators for one dataset with model (b).



(a) $n = 200$ **(b)** $n = 600$

(c) $n = 1000$ **(d)** PSE-I

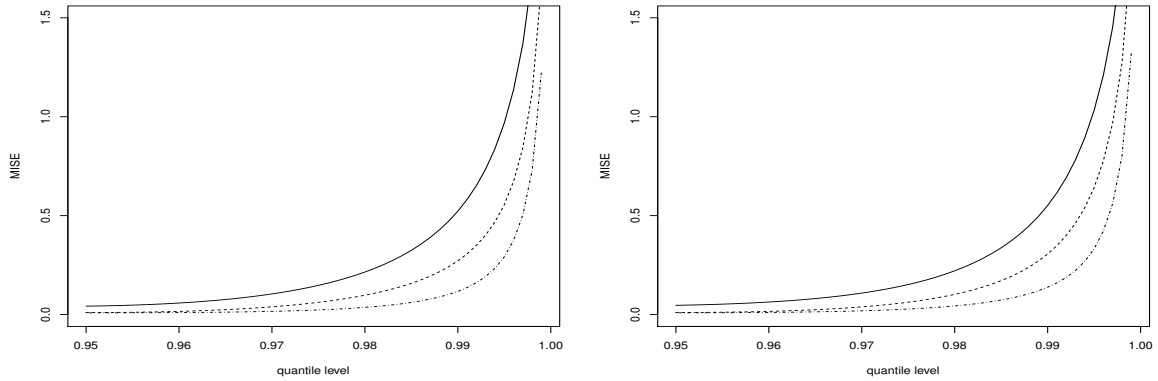
Figure 7: MISE of the intermediate order quantile estimators for $\tau \in [0.5, 0.995]$. The description is similar to Figure 2.



(a) $n = 200$ **(b)** $n = 600$ **(c)** $n = 1000$

(d) MISE, $n = 200$ **(e)** MISE, $n = 600$ **(f)** MISE, $n = 1000$

Figure 8: (a–c) The true conditional quantiles the extreme order quantile estimators for $\tau = 0.995$ for one dataset with model (b). The dataset is similar to that given in Figure 6 for each n. (d–f) MISE of the estimators for $\tau \in [0.95, 0.999]$. The description is similar to Figure 4.

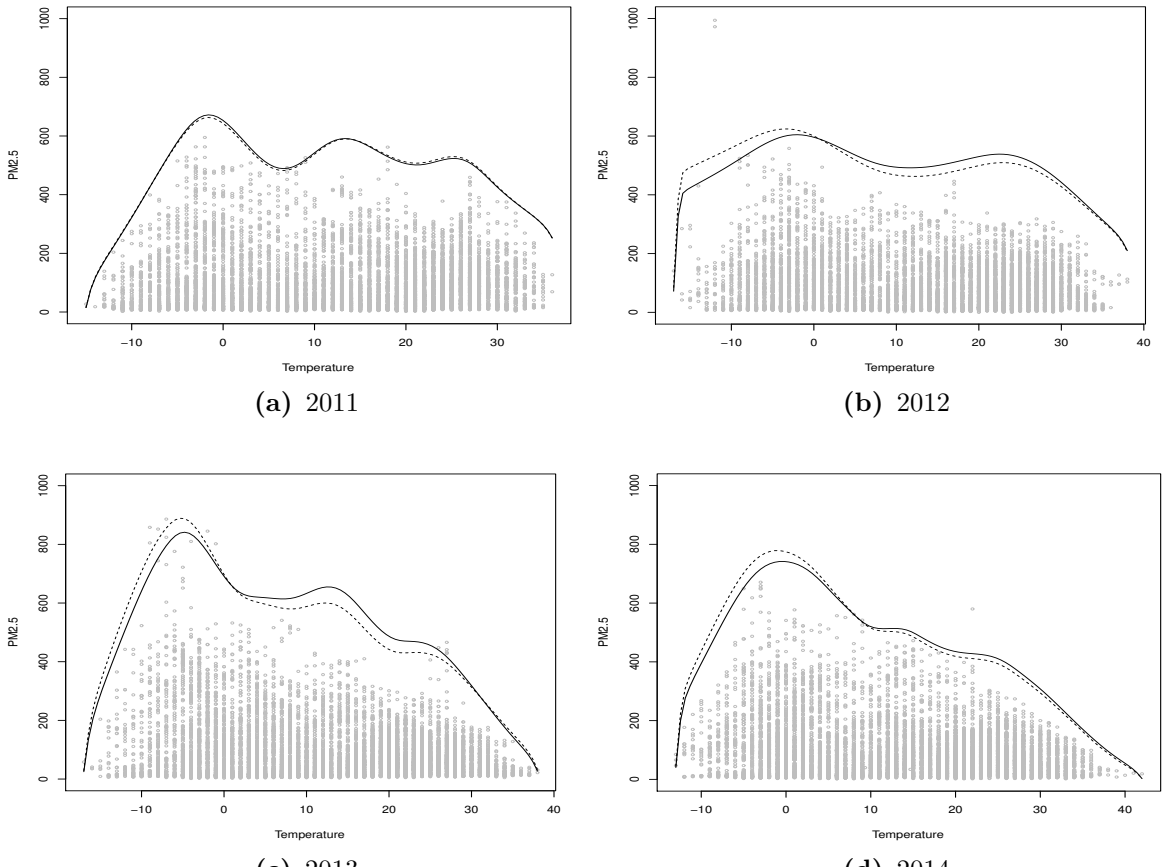


(a) MISE, PSE-E **(b)** MISE, PSE-Ep

Figure 9: (a) and (b) are the MISE of the PSE-E and the PSE-Ep, respectively. The description is similar to Figure 5.

5. DATA EXAMPLE

In this section, we apply the proposed methods to Beijing's $PM_{2.5}$ Pollution data. The data is available from the website of the UCI Machine Learning Repository and Liang *et al.* (2015) [34] provided several analyses for this data. One of fundamental purposes of this data is to analyze the relationship between $PM_{2.5}$ concentration and other meteorological variables. Our particular interest here is the prediction of high conditional quantiles of Y, $PM_{2.5}$ concentration ($\mu g/m^3$), with the predictor x, temperature (degrees Celsius). We can observe from the scatter plot of y and x (see Figure 10) that this relationship is not linear for the upper quantile. Therefore, the nonparametric approach is suitable for this data. We demonstrate the analysis for each year from 2011 to 2014. We then omit the missing data and hence the sample size is $n = 8032$, 8295, 8678, and 8661 in 2011, 2012, 2013, and 2014. We construct the extreme order quantile estimator for $\tau = 0.999$. The quantile level $\tau = 0.999$ indicates that about only eight events occur each year.

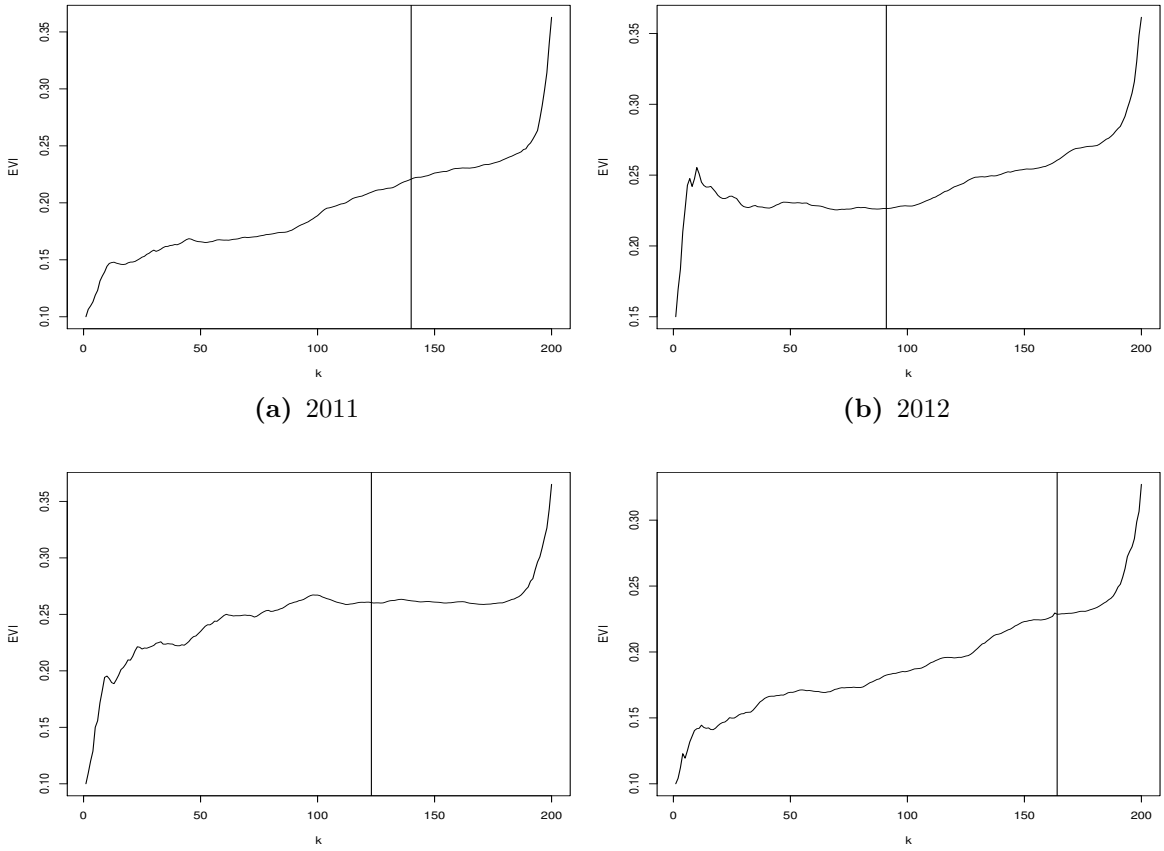


(a) 2011 **(b)** 2012

(c) 2013 **(d)** 2014

Figure 10: The proposed extreme order quantile estimator for $\tau = 0.999$ for years from 2011 to 2014. The solid line is $\hat{q}_Y^C(0.999|x)$ and the dashed line is $\hat{q}_Y(0.999|x)$ for each panel.

Figure 10 shows the extreme order quantile estimators for $\tau = 0.999$. It seems that the tail behavior is stable in 2011 and 2012. In 2013 and 2014, the estimator of the conditional quantile at $x \in [0, 10]$ has a large value compared with those in 2011 and 2012. Thus, in the cold season of 2013 and 2014, the risk of $PM_{2.5}$ of pollution was increased. We can observe that $\hat{q}_Y(0.999|x)$ and $\hat{q}_Y^C(0.999|x)$ are quite similar. This indicates that the estimator of EVI $\hat{\gamma}(x)$ hardly changes with x. We report the EVI estimator used for constructing the extrapolated estimator $\hat{q}_Y$ and $\hat{q}_Y^{pool}$. To obtain the estimator of EVI, we utilized $\eta = 0.4$ as $\tau_1 = n - [n^\eta]/(n+1)$ so that about $\tau_1 = 0.995$. Then, EVI is estimated by using the intermediate order quantiles estimators for $\tau \in (0.978, 0.995)$. This choice leads to $\xi = n(1 - \tau_1) \approx 37$ and, hence, this is a very conservative situation in the study of Chernozhukov and Fernández-val (2011) [7]. Furthermore, we then adopted to use k so that the sample path of the pooled EVI estimator is stable in each year. Figure 11 shows the sample path of $\hat{\gamma}^C$ and selected k. As a result, $\hat{\gamma} = 0.220$, 0.226, 0.260, and 0.231 in 2011, 2012, 2013, and 2014. Thus, the pooled EVI estimators are the same in 2011, 2012, and 2014, and it is only slightly larger in 2013.

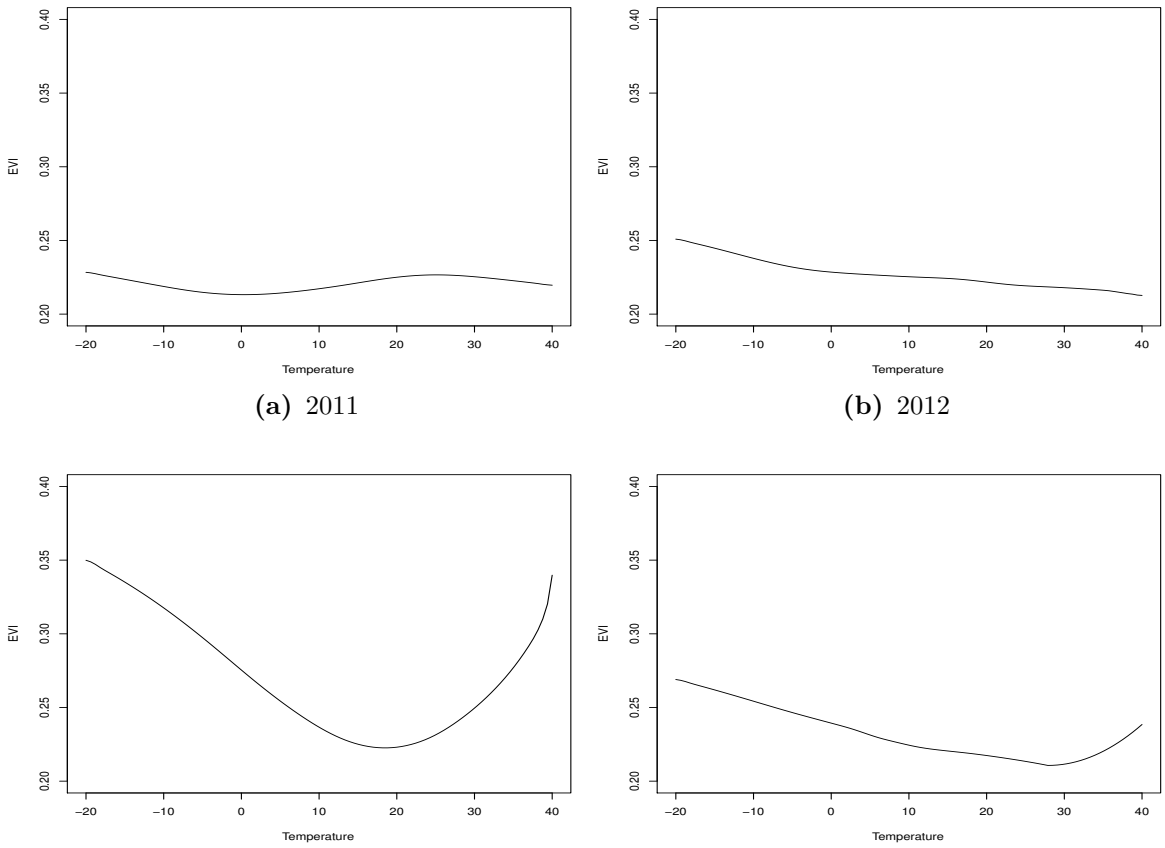


(a) 2011 **(b)** 2012

(c) 2013 **(d)** 2014

Figure 11: The sample path of EVI estimates $\hat{\gamma}^C$ with k for each year.

Figure 12 illustrates $\hat{\gamma}(x)$ with selected k for each year. We can observe that $\hat{\gamma}(x)$ has a narrow curve with x in 2011, 2012, and 2014. On the other hand, in 2013, $\hat{\gamma}(x)$ at the boundary is rather larger than at the center. Indeed, it can be seen from Figure 10 that the extreme point can be observed at $x < 0$ in 2013.



(a) 2011 **(b)** 2012

(c) 2013 **(d)** 2014

Figure 12: The EVI estimates $\hat{\gamma}(x)$ with selected k versus x for each year.

6. CONCLUSION

We have developed the nonparametric extremal quantile regression methods for heavy-tailed data. To show the mathematical property of the proposed estimator, we have used the hybrid techniques of asymptotic theory for the nonparametric regression and EVT for the tail behavior of the conditional distribution. We then considered two quantile rates: (i) the intermediate order quantile that $(1-\tau)n \to \infty$ as $n \to \infty$; and (ii) the extreme order quantile that $(1-\tau)n \to \xi < \infty$. For the intermediate order quantile, the penalized spline estimator and its asymptotic normality have been developed. On the other hand, for the extremal order quantile, we have studied the Weissman-type extrapolated estimator using the intermediate order quantile estimator and its asymptotic normality. For the both intermediate and extreme order quantile, we show the asymptotic normality and the optimal rate of convergence of the proposed estimator. In particular, we found that the convergence speed of the estimator for extremal quantile is slower than that for center quantile. This result would be intuitively correct.

We now discuss some future directions of study. First, for technical reasons, we assumed that the tail behavior of the conditional distribution of Y given $X = x$ is equivalent across the predictor x (see Conditions A1). Since the estimation of the tail behavior is difficult due to data sparsity, this assumption is helpful in data analysis. However, if this assumption is violated, additional research is needed to explicate the performance of the estimator.

Second, in this paper, we focused on the spline smoothing with ℓ_2 penalty. On the other hand, Koenker *et al.* (1994) [32] and Koenker (2011) [30] studied the smoothing spline with the ℓ_1-type penalty. That is, the penalty is defined as $\int_a^b |s^{(m)}(x)|dx$ instead of $\int_a^b \{s^{(m)}(x)\}^2 dx$. It is known that the estimator with ℓ_1 penalty has local adaptiveness. Therefore, for some cases, the performance of the estimator with ℓ_1 penalty would be better than that with ℓ_2 penalty. Recently, the ℓ_1-type penalty has been rapidly developed in mean regression (Kim *et al.* 2009 [27]; Tibshirani 2014 [43]; Sadhanala and Tibshirani 2017 [42]). Although it is difficult to show the asymptotic distribution of an ℓ_1 penalized estimator, the developments of the ℓ_1 penalized smoothing to the extremal quantile regression is an interesting problem.

Finally, we can consider extending the proposed method to the multidimensional case. In particular, it is important to use the additive models (Hastie and Tibshirani 1990 [23]) that for $x = (x_1, ..., x_d) \in \mathbb{R}^d$, the true function is can be decomposed as

$$f(x) = f_1(x_1) + ... + f_d(x_d),$$

where each f_j is the univariate function. The additive model is known to enables us to avoid the problem of dimensionality. The nonparametric additive quantile regression (for center quantile) was studied by Lu and Yu (2004) [35], Horowitz and Lee (2005) [26], Cheng *et al.* (2011) [8], Koenker (2011) [30], Lee *et al.* (2010) [33] and references therein. However, the extremal inference of the additive quantile regression has not yet been studied until now. It seems that the developments of the extremal quantile regression with the additive model is an important issue.

A. APPENDIX – Computation of the intermediate order quantile estimator

We describe here the approximation algorithm to solve (3.1). Nychka *et al.* (1995) [36] and Reiss and Huang (2012) [40] proposed the penalized iteratively reweighted least squares algorithm. We use here the modified version of Nychka *et al.* (1995) [36]. Nychka *et al.* (1995) [36] proposed the following optimization:

$$\tilde{\boldsymbol{b}} = \underset{\boldsymbol{b}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \rho_{\tau,\alpha}(y_i - \boldsymbol{B}(x_i)^T \boldsymbol{b}) + \lambda \boldsymbol{b}^T D_m^T R D_m \boldsymbol{b} \right\}, \tag{A.1}$$

where

$$\rho_{\tau,\alpha}(u) = \begin{cases} \rho_\tau(u), & |u| > \alpha, \\ \tau u^2/\alpha, & 0 \le u \le \alpha, \\ (1-\tau)u^2/\alpha, & -\alpha \le u \le 0 \end{cases}$$

Obviously, the loss function $\rho_{\tau,\alpha}(u)$ tends to ρ_τ as $\alpha \to 0$. However, for the tail quantile ($\tau \approx 0, 1$), the above algorithm will not converge in our implementation. Therefore, we suggest using the slightly modified version of (A.1). The idea is similar to the proximal gradient method and it is very simple. The modified algorithm is defined as given the tth iteration estimate $\tilde{\boldsymbol{b}}^{(t)}$,

$$\tilde{\boldsymbol{b}}^{(t+1)} = \underset{\boldsymbol{b}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \rho_{\tau,\alpha_t}(y_i - \boldsymbol{B}(x_i)^T \boldsymbol{b}) + \lambda \boldsymbol{b}^T D_m^T R D_m \boldsymbol{b} + \eta_t ||\boldsymbol{b} - \tilde{\boldsymbol{b}}^{(t)}||^2 \right\}, \tag{A.2}$$

where a_t and η_t are the step sizes. For a_t, it is sufficient to use some sequence $\{\alpha_t\}$ satisfying $\alpha_t \to 0$ as $t \to \infty$. On the other hand, the choice of η_t is more important than a_t. If η_t is large, $\boldsymbol{b}^{(t+1)} \approx \boldsymbol{b}^{(t)}$ and hence the speed of convergence is very fast. When η_t is small, on the other hand, there is almost no difference between (A.2) and (A.1), that is, the algorithm does not converge in many cases. In Sections 4 and 5, we used $\alpha_t = 0.1 \times 2^{-t}$ and $\eta_{t+1} = 1.2\eta_t (\eta_0 = 1)$.

B. APPENDIX – Proof of theorems

Let Z be the $n\times(K+p)$ design matrix having elements $B_j(x_i)$, $G_n = n^{-1}Z^T\mathrm{diag}[H(x_i)^{-\gamma}]Z$ and let $\Lambda_n = G_n + (\lambda q(\tau)/((1-\tau)n))D_m^T R D_m$. For any matrix $A = (a_{ij})_{ij}$, we denote $||A||_\infty = \max_{i,j}|a_{ij}|$. We first state the technical lemmas to prove theorems in this paper.

Lemma B.1. *Suppose that $K(m,\tau)\geq 1$. Under Conditions A–C, the following statements holds:* $||G(1)||_\infty = O(K^{-1})$, $||G(H^{-\gamma})||_\infty = O(K^{-1})$, $||\Lambda_n^{-1}||_\infty = O(K(1+K(m,\tau)^{2m})^{-1})$, $||\Lambda(H^{-\gamma})^{-1}||_\infty = O(K(1+K(m,\tau)^{2m})^{-1})$, $\max_{i,j}\{|\{G_n^{-1} - G(H^{-\gamma})^{-1}\}_{ij}|\} = o(K)$ *and* $\max_{i,j}\{|\{\Lambda_n^{-1} - \Lambda(H^{-\gamma})^{-1}\}_{ij}|\} = o(K(1+K(m,\tau)^{2m})^{-1})$.

Lemma B.2. *Suppose that $K(m,\tau)\geq 1$. Under Assumptions 2–3, for any $K+p$ square matrix A satisfying $||A||_\infty = O(n^a)$ for $a\in\mathbb{R}$, $||G(H^{-\gamma})A||_\infty = O(n^a K^{-1})$ and $||\Lambda(H^{-\gamma})^{-1}A||_\infty = O(n^a K(1+K(m,\tau)^{2m})^{-1})$.*

Lemma B.1 is proved by Lemma 6.3 and 6.4 of Zhou *et al.* (1998) [48] and Lemma A1 and A2 of Claeskens *et al.* (2009) [9]. Lemma B.2 says that the order of product of matrices is dependent only on the order of element of these matrices although the each element of GA and $\Lambda(h)^{-1}A$ is infinite sum as $n\to\infty$. The proof of Lemma B.2 can also be shown by Lemma A1 of Claeskens *et al.* (2009) [9]. Therefore, we only describe the outline here.

Proof of Lemma B.2: For any continuous and bounded function h, the matrix $G(h)$ is the band matrix from property of B-splines and hence $||GA||_\infty = O(n^a K^{-1})$ is obvious. Next, $\Lambda(h)^{-1}$ is the inverse of the band matrix. From Demko (1977) [14], there exists $c>0$ and $r\in(0,1)$ such that $|\{\Lambda(h)^{-1}\}_{ij}| < cr^{|i-j|}K(1+K(m,\tau)^{2m})^{-1}$. Thus, straightforward calculation yields that the infinite sum of each element of $\Lambda(h)^{-1}A$ is bounded by order of $\Lambda(h)^{-1}$ and the absolute of the maximum of element of A. □

Proof of Theorem 3.1: Define $h(x) = H(x)^{-\gamma}$. We write $\tau\equiv\tau_I$ and hence $\tau\to 1$ and $(1-\tau)n\to\infty$ as $n\to\infty$. By the fundamental property of B-spline basis, $s_0^{(m)}(\tau|x) = d^m s_0(\tau|x)/dx^m$ can be written as $s_0^{(m)}(\tau|x) = \boldsymbol{B}^{[p-m]}(x)^T D_m \boldsymbol{b}_0(\tau)$, where $\boldsymbol{B}^{[p-m]}(x)$ is the vector having element $\{B_1^{[p-m]}(x),...,B_{K+p-m}^{[p-m]}(x)\}$ and $B_k^{[p-m]}(x)$'s are $(p-m)$th degree B-spline bases. Therefore, the shrinkage bias can be expressed as

$$b_\lambda(\tau|x) = \frac{\lambda q(\tau)}{(1-\tau)n}\boldsymbol{B}(x)^T\Lambda(H^{-\gamma})^{-1}D_m^T\int_a^b \boldsymbol{B}^{[p-m]}(x)s_0^{(m)}(\tau|x)dx.$$

From Conditions A–B, we have $s_0(\tau|x) = q_Y(\tau|x)(1+o(1)) = h(x)q(\tau)(1+o(1))$ and $s_0^{(m)}(\tau|x) = h^{(m)}(x)q(\tau)(1+o(1))$ as $\tau\to 1$. Since h is bounded function, we get $\sup_{x\in[a,b]}|s_0^{(m)}(\tau|x)| = O(q(\tau))$. From the property of B-spline basis, we also obtain $\int_a^b B_k(x)ds = O(K^{-1})$. Thus, each element of $\int_a^b \boldsymbol{B}^{[p-m]}(x)s_0^{(m)}(\tau|x)dx$ has the order $O(K^{-1}q(\tau))$ Furthermore, the result of Cardot (2000) [5] provides $||D_m||_\infty = O(K^m)$. Therefore, Lemmas B.1, B.2 and the fact

that $K(m,\tau)^m/(1+K(m,\tau)^{2m}) < 1/2$ yield that

$$\begin{aligned} b_\lambda(\tau|x) &= O\left(q(\tau)\frac{\lambda q(\tau)K^m}{(1-\tau)n}(1+K(m,\tau)^{2m})^{-1}\right) \\ &= O\left(q(\tau)\left(\frac{\lambda q(\tau)}{(1-\tau)n}\right)^{1/2}\frac{K(m,\tau)^m}{1+K(m,\tau)^{2m}}\right) \\ &\le O\left(q(\tau)\left(\frac{\lambda q(\tau)}{(1-\tau)n}\right)^{1/2}\right). \end{aligned}$$

Next, we show the asymptotic order of $v(x|\tau)$. Since $||G||_\infty = O(K^{-1})$, from Lemmas B.1–B.2, we have

$$\begin{aligned} v(\tau|x) &= O\left(\frac{Kq(\tau)^2}{(1-\tau)n}\{1+K(m,\tau)^{2m}\}^{-2}\right) \\ &= O\left(\frac{q(\tau)}{(1-\tau)n}\left(\frac{\lambda q(\tau)}{(1-\tau)n}\right)^{-1/2m}K(m,\tau)\{1+K(m,\tau)^{2m}\}\right) \\ &\le O\left(\frac{q(\tau)}{(1-\tau)n}\left(\frac{\lambda q(\tau)}{(1-\tau)n}\right)^{-1/2m}\right), \end{aligned}$$

which completes the proof. □

Proof of Theorem 3.2: We write $\tau \equiv \tau_I$ and hence $\tau \to 1$ and $(1-\tau)n \to \infty$ as $n \to \infty$. Let $U_i = Y_i - \boldsymbol{B}(x)^T\boldsymbol{b}_0(\tau)$, $a_n = \sqrt{(1-\tau)n}/q(\tau)$ and

$$\begin{aligned} Q_n(\boldsymbol{\delta}|\tau) &= \frac{a_n}{\sqrt{n(1-\tau)}}\sum_{i=1}^n\left\{\rho_\tau(U_i - \boldsymbol{B}(x_i)^T\boldsymbol{\delta}/a_n) - \rho_\tau(U_i)\right\} \\ &\quad + \frac{a_n\lambda}{2\sqrt{n(1-\tau)}}(\boldsymbol{b}_0(\tau)+\boldsymbol{\delta}/a_n)^T D_m^T R D_m(\boldsymbol{b}_0(\tau)+\boldsymbol{\delta}/a_n). \end{aligned}$$

Then the minimizer of Q_n is obtained as

$$\tilde{\boldsymbol{\delta}} = a_n(\tilde{\boldsymbol{b}}(\tau) - \boldsymbol{b}_0(\tau)).$$

Using Knight's identity (Knight, 1998 [28]), we have

$$\rho_\tau(u-v) - \rho_\tau(u) = -v(\tau - I(u<0)) + \int_0^v \{I(u\le s) - I(u\le 0)\}ds$$

and writing

$$Q_n(\boldsymbol{\delta}|\tau) = W_n(\tau)^T\boldsymbol{\delta} + G_n(\boldsymbol{\delta}|\tau) + \frac{a_n\lambda}{2\sqrt{(1-\tau)n}}(\boldsymbol{b}_0(\tau)+\boldsymbol{\delta}/a_n)^T D_m^T R D_m(\boldsymbol{b}_0(\tau)+\boldsymbol{\delta}/a_n),$$

where

$$\begin{aligned} W_n(\tau) &\equiv \frac{-1}{\sqrt{(1-\tau)n}}\sum_{i=1}^n(\tau - I(Y_i < \boldsymbol{B}(x_i)^T\boldsymbol{b}_0))\boldsymbol{B}(x_i), \\ G_n(\boldsymbol{\delta}|\tau) &\equiv \sum_{i=1}^n\int_0^{\boldsymbol{B}(x)^T\boldsymbol{\delta}/a_n} I(U_i\le s) - I(U_i\le 0)ds. \end{aligned}$$

Since $\tau = P(Y < q_Y(\tau|x)|X = x) = P(Y < \boldsymbol{B}(x)^T\boldsymbol{b}_0(\tau)|X = x) + o(K^{-m})$ and $E[I(Y < \boldsymbol{B}(x)^T\boldsymbol{b}_0(\tau))] = P(Y < q_Y(\tau|x)|X = x) + o(K^{-m}(1-\tau)^{-\gamma})$, we obtain $E[W_n(\tau)] = o(1)$. The variance of $W_n(\tau)$ can be evaluated as

$$V[W_n(\tau)^T\boldsymbol{\delta}] = \frac{\tau(1-\tau)}{1-\tau}\boldsymbol{\delta}^T\left(\frac{1}{n}Z^TZ\right)\boldsymbol{\delta} \xrightarrow{P} \boldsymbol{\delta}^TG\boldsymbol{\delta}$$

as $n \to \infty$ and $\tau \to 1$. Lyapnov's condition for the central limit theorem and Cramèr-Wold device yield that $W_n(\tau)$ is asymptotically distributed as $\boldsymbol{W}$, which is the normal with mean $\mathbf{0}$ and variance G.

Next, we show that as $n \to \infty$ and $\tau \to 1$,

$$G_n(\boldsymbol{\delta}|\tau) \xrightarrow{P} \frac{1}{2}\gamma^{-1}\boldsymbol{\delta}^TG(H^{-\gamma})\boldsymbol{\delta}. \tag{B.1}$$

Before that, we provide some differential results. Let $f_Y(y)$ and $f_Y(y|x)$ be the marginal density of Y and conditional density of Y given $X = x$, respectively. From A3, $f_Y(y|x) = f_V(y|x)(1+o(1)) = H(x)f_Y(y)(1+o(1))$ as $y \to \infty$. In addition, since $1 = \partial F_V(q_V(\tau|x)|x)/\partial\tau = f_V(q_V(\tau|x)|x)\partial q_V(\tau|x)/\partial\tau$, we have $f_V(q_V(\tau|x)|x) = \{\partial q_V(\tau|x)/\partial\tau\}^{-1}$. Meanwhile, A4 and $q((1-\tau)/H(x)) = \{(1-\tau)/H(x)\}^{-\gamma}L(H(x)/(1-\tau))(1+o(1))$ yield that $\partial q((1-\tau)/H(x))/\partial\tau = \gamma(1-\tau)^{-\gamma-1}H(x)^{\gamma}L(H(x)/(1-\tau))(1+o(1))$. Consequently, as $\tau \to 1$,

$$f_Y(q_Y(\tau|x)|x) = \gamma^{-1}H(x)^{-\gamma}(1-\tau)^{\gamma+1}L(H(x)/(1-\tau))^{-1}(1+o(1)).$$

Furthermore, by $L \in RV(0)$, $q(\tau)f_Y(q_Y(\tau|x)|x)/\{1-\tau\} = \gamma^{-1}H(x)^{-\gamma}(1+o(1))$.

We return to show (B.1). Since

$$G_n(\boldsymbol{\delta}|\tau) = \frac{1}{\sqrt{(1-\tau)n}}\sum_{i=1}^{n}\left(\int_0^{\boldsymbol{\delta}^T\boldsymbol{B}(x_i)} I(U_i \le s/a_n) - I(U_i \le 0)ds\right),$$

we obtain

$$\begin{aligned}
E[G_n(\boldsymbol{\delta}|\tau)] &= \frac{1}{\sqrt{(1-\tau)n}}\sum_{i=1}^{n}E\left[\int_0^{\boldsymbol{\delta}^T\boldsymbol{B}(x_i)} I(U_i \le s/a_n) - I(U_i \le 0)ds\right] \\
&= \frac{1}{\sqrt{(1-\tau)n}}\sum_{i=1}^{n}\left[\int_0^{\boldsymbol{\delta}^T\boldsymbol{B}(x_i)} F_Y(q_Y(\tau|x_i) + s/a_n|x_i) - F_Y(q_Y(\tau|x_i)|x_i)ds\right] \\
&= \sum_{i=1}^{n}\int_0^{\boldsymbol{\delta}^T\boldsymbol{B}(x_i)} \frac{f_Y(q_Y(\tau|x_i))}{a_n\sqrt{(1-\tau)n}}sds(1+o(1)) \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\boldsymbol{\delta}^T\boldsymbol{B}(x_i)} \frac{q(\tau)f_Y(q_Y(\tau|x_i))}{(1-\tau)}sds(1+o(1)) \\
&= 2^{-1}\gamma^{-1}\boldsymbol{\delta}^T\left(\frac{1}{n}\sum_{i=1}^{n}H(x_i)^{-\gamma}\boldsymbol{B}(x_i)\boldsymbol{B}(x_i)^T\right)\boldsymbol{\delta}.
\end{aligned}$$

From the simple but tedious calculation, $V[G_n(\boldsymbol{\delta}|\tau)] = o(1)$ can be evaluated. These results yield that

$$E[G_n(\boldsymbol{\delta}|\tau)] \xrightarrow{P} \frac{1}{2}\gamma^{-1}\boldsymbol{\delta}^TG(H^{-\gamma})\boldsymbol{\delta}.$$

Thus, $Q_n(\boldsymbol{\delta}|\tau)$ is asymptotically equivalent to

$$\begin{aligned} Q_0(\boldsymbol{\delta}|\tau) \;=\; & \left(W^T + \frac{\lambda}{a_n\sqrt{(1-\tau)n}} \boldsymbol{b}_0(\tau)^T D_m^T R D_m \right) \boldsymbol{\delta} \\ & + \frac{1}{2}\boldsymbol{\delta}^T \left(\gamma^{-1} G(H^{-\gamma}) + \frac{\lambda}{2a_n\sqrt{(1-\tau)n}} D_m^T R D_m \right) \boldsymbol{\delta}. \end{aligned}$$

By the convexity lemma (see Pollard 1991 [38] and Knight 1998 [28]), the minimizer of Q_n and Q_0 are asymptotically equivalent and hence we have

$$\tilde{\boldsymbol{\delta}} = \left(\gamma^{-1} G(H^{-\gamma}) + \frac{\lambda}{2a_n\sqrt{(1-\tau)n}} D_m^T R D_m \right)^{-1} \left(W - \frac{\lambda}{\sqrt{a_n(1-\tau)n}} D_m^T R D_m \boldsymbol{b}_0(\tau) \right) + o_P(1).$$

Since $\tilde{q}_Y(\tau|x) - s_0(\tau|x) = \boldsymbol{B}(x)^T(\tilde{\boldsymbol{b}}(\tau) - \boldsymbol{b}_0(\tau))$, we obtain from $a_n = \sqrt{(1-\tau)n}/q(\tau)$ that

$$\begin{aligned} & \frac{\sqrt{(1-\tau)n}}{q(\tau)} (\tilde{q}_Y(\tau|x) - s_0(\tau|x)) \;= \\ & \quad = \; \boldsymbol{B}(x)^T \left(\gamma^{-1} G(H^{-\gamma}) + \frac{\lambda q(\tau)}{2(1-\tau)n} D_m^T R D_m \right)^{-1} W \\ & \qquad - \frac{\lambda}{\sqrt{(1-\tau)n}} \boldsymbol{B}(x)^T \left(\gamma^{-1} G(H^{-\gamma}) + \frac{\lambda q(\tau)}{2(1-\tau)n} D_m^T R D_m \right)^{-1} D_m^T R D_m \boldsymbol{b}_0(\tau) \\ & \qquad + o_P(1). \end{aligned} \tag{B.2}$$

The second term of right hand side of (B.2) is the shrinkage bias. Consequently, as $n \to \infty$,

$$\frac{\sqrt{(1-\tau)n}}{q(\tau)} \frac{\tilde{q}_Y(\tau|x) - q_Y(\tau|x) - b_\lambda(\tau|x)}{\sqrt{\boldsymbol{B}(x)^T \Lambda(H^{-\gamma})^{-1} G \Lambda(H^{-\gamma})^{-1} \boldsymbol{B}(x)}} \xrightarrow{D} N(0,1).$$

Finally, we obtain

$$\begin{aligned} E[\{\hat{q}_Y(\tau|x) - q_Y(\tau|x)\}^2] \;&=\; b_\lambda(\tau|x)^2 + v(\tau|x) \\ &=\; O\left(q(\tau)^2 \frac{\lambda q(\tau)}{(1-\tau)n} \right) + O\left(\frac{q(\tau)^2}{(1-\tau)n} \left(\frac{\lambda q(\tau)}{(1-\tau)n} \right)^{-1/2m} \right). \end{aligned} \tag{B.3}$$

We now derive the optimal rate of convergence of MISE of $\tilde{q}_Y(\tau|x)$. For the constant $C_1 > 0$, $C_2 > 0$, the solution of

$$C_1 q(\tau)\lambda - C_2 \left(\frac{q(\tau)}{(1-\tau)n} \right)^{-1/2m} \lambda^{-1/2m} = 0$$

is $\lambda = C q(\tau_I)^{-1} \{n(1-\tau_I)\}^{1/(2m+1)}$ for $C > 0$. By applying this λ in (B.3), we obtain

$$E\left[\left\{ \frac{\hat{q}_Y(\tau|x)}{q_Y(\tau|x)} - 1 \right\}^2 \right] = O(\{(1-\tau)n\}^{-2m/(2m+1)}),$$

which completes the proof. □

To improve the outlook, we now describe about the asymptotic bias and variance of $\hat{\gamma}(x)$ before prove Theorem 3.3. Define

$$\text{(B.4)} \quad b(k|x) = \frac{1}{k-1}\sum_{j=1}^{k-1}\frac{b_\lambda(\tau_j|x)}{q_Y(\tau_j|x)} - \frac{b_\lambda(\tau_k|x)}{q_Y(\tau_k|x)}, \quad \boldsymbol{v}(k|x) = \frac{1}{k-1}\sum_{j=1}^{k-1}\frac{\boldsymbol{\nu}(\tau_j|x)}{q_Y(\tau_j|x)} - \frac{\boldsymbol{\nu}(\tau_k|x)}{q_Y(\tau_k|x)},$$

where

$$\boldsymbol{\nu}(\tau|x) = \frac{q(\tau)}{\sqrt{(1-\tau)n}}G^{1/2}\Lambda(H^{-\gamma})^{-1}\boldsymbol{B}(x).$$

As the result, $b(k|x)$ and $\sqrt{\boldsymbol{v}(k|x)^T\boldsymbol{v}(k|x)}$ are the asymptotic bias and standard deviation of $\hat{\gamma}(x)$. We here get the asymptotic order of $b(k|x)$ and $v(k|x)$ from easy calculation.

Since $b_\lambda(\tau_j|x)/q_Y(\tau_j|x) = O(\{(1-\tau_j)n\}^{-m/(2m+1)})$ and each element of $\boldsymbol{\nu}(\tau_j|x)$ has $O(\{(1-\tau_j)n\}^{-m/(2m+1)})$, we have

$$b(k|x) = O\left(\frac{1}{k-1}\sum_{j=1}^{k-1}\{(1-\tau_j)n\}^{-\frac{m}{2m+1}} - \{(1-\tau_k)n\}^{-\frac{m}{2m+1}}\right)$$

and

$$\boldsymbol{v}(k|x) = O\left(\frac{1}{k-1}\sum_{j=1}^{k-1}\{(1-\tau_j)n\}^{-\frac{m}{2m+1}} - \{(1-\tau_k)n\}^{-\frac{m}{2m+1}}\right).$$

We then have from $[n^\eta]/k \to 0 (n, k \to \infty)$ that

$$\begin{aligned}
\frac{1}{k}\sum_{j=1}^{k}\{(1-\tau_j)n\}^{-\frac{m}{2m+1}} &= \frac{1}{k}\sum_{j=1}^{k}\left\{\frac{[n^\eta]+j}{n+1}n\right\}^{-\frac{m}{2m+1}} \\
&= k^{-\frac{m}{2m+1}}\frac{1}{k}\sum_{j=1}^{k}\left\{\frac{[n^\eta]+j}{k+1}\right\}^{-\frac{m}{2m+1}} \\
&= k^{-\frac{m}{2m+1}}\int_0^1 u^{-\frac{m}{2m+1}}du(1+o(1)) \\
&= \frac{2m+1}{m}k^{-\frac{m}{2m+1}}(1+o(1))
\end{aligned}$$

and

$$\{(1-\tau_k)n\}^{-\frac{m}{2m+1}} = \left\{\frac{[n^\eta]+k}{n+1}n\right\}^{-\frac{m}{2m+1}} = k^{-\frac{m}{2m+1}}(1+o(1)).$$

This indicates that $b(k|x) = O(k^{-\frac{m}{2m+1}})$ and $\boldsymbol{v}(k|x) = O(k^{-\frac{m}{2m+1}})$.

Proof of Theorem 3.3: Theorem 3.1 indicates that $\tilde{q}_Y(\tau_k|x) - q_Y(\tau_k|x) = o_P(1)$ as $n \to \infty$. Therefore, the proposed estimator can be calculated as for $x \in \mathbb{R}$,

$$\begin{aligned}
\hat{\gamma}(x) &= \frac{1}{k-1}\sum_{j=1}^{k-1} \log \frac{\hat{q}_Y(\tau_j|x)}{\hat{q}_Y(\tau_k|x)} \\
&= \frac{1}{k-1}\sum_{j=1}^{k-1} \log \frac{q_Y(\tau_j|x)\left\{1+\frac{\hat{q}_Y(\tau_j|x)-q_Y(\tau_j|x)}{q_Y(\tau_j|x)}\right\}}{q_Y(\tau_k|x)\left\{1+\frac{\hat{q}_Y(\tau_k|x)-q_Y(\tau_k|x)}{q_Y(\tau_k|x)}\right\}} \\
&= \frac{1}{k-1}\sum_{j=1}^{k-1} \log \frac{q_Y(\tau_j|x)}{q_Y(\tau_k|x)} + \frac{1}{k-1}\sum_{j=1}^{k-1} \log \frac{\left\{1+\frac{\hat{q}_Y(\tau_j|x)-q_Y(\tau_j|x)}{q_Y(\tau_j|x)}\right\}}{\left\{1+\frac{\hat{q}_Y(\tau_k|x)-q_Y(\tau_k|x)}{q_Y(\tau_k|x)}\right\}} \\
&= \frac{1}{k-1}\sum_{j=1}^{k-1} \log \frac{q_Y(\tau_j|x)}{q_Y(\tau_k|x)} + \frac{1}{k-1}\sum_{j=1}^{k-1} \log \left\{1+\frac{\hat{q}_Y(\tau_j|x)-q_Y(\tau_j|x)}{q_Y(\tau_j|x)}\right\} \\
&\quad - \frac{1}{k-1}\sum_{j=1}^{k-1} \log \left\{1+\frac{\hat{q}_Y(\tau_k|x)-q_Y(\tau_k|x)}{q_Y(\tau_k|x)}\right\} \\
&= \frac{1}{k-1}\sum_{j=1}^{k-1} \log \frac{q_Y(\tau_j|x)}{q_Y(\tau_k|x)} \\
&\quad + \frac{1}{k-1}\sum_{j=1}^{k-1} \left\{\frac{\hat{q}_Y(\tau_j|x)-q_Y(\tau_j|x)}{q_Y(\tau_j|x)} - \frac{\hat{q}_Y(\tau_k|x)-q_Y(\tau_k|x)}{q_Y(\tau_k|x)}\right\}(1+o_P(1)) \\
&\equiv D_{1n} + D_{2n}.
\end{aligned}$$

We then note that D_{1n} is not random variable. Similar to the proof of Theorem 2.3 of Wang *et al.* (2012) [44], we have as $k \to \infty$,

$$D_{1n} = \gamma + O(k^{-1/2}) = \gamma + o(k^{-m/(2m+1)})$$

for $m \geq 1$. Next, we consider D_{2n}. Under the conditions for Theorem 3.3, using the result of Theorems 3.1–3.2 and the property of B-spline basis, we have

$$\frac{\tilde{q}_Y(\tau_j|x) - q_Y(\tau_j|x)}{q_Y(\tau_j|x)} = \frac{b_\lambda(\tau_j|x) + \boldsymbol{\nu}(\tau_j|x)^T \boldsymbol{W}}{q_Y(\tau_j|x)}(1+o_P(1)),$$

where $\boldsymbol{W} \sim N_{K+p}(\mathbf{0}, I)$. Therefore, D_{2n} can be evaluated as

$$D_{2n} = \{b(k|x) + \boldsymbol{v}(k|x)^T \boldsymbol{W}\}(1+o_P(1))$$

That is,

$$\hat{\gamma}(x) = \gamma + b(k|x) + \boldsymbol{v}(k|x)^T \boldsymbol{W} + o(k^{-m/(2m+1)}),$$

where $b(k|x) = O(k^{-m/(2m+1)})$ and $\boldsymbol{v}(k|x) = O(k^{-m/(2m+1)})$. Consequently, we get

$$\frac{\hat{\gamma}(x) - \gamma - b(k|x)}{\sqrt{\boldsymbol{v}(k|x)^T \boldsymbol{v}(k|x)}} \xrightarrow{D} N(0,1)$$

and $E[\{\hat{\gamma}(x) - \gamma\}^2] = O(k^{-2m/(2m+1)})$. For the common index version $\hat{\gamma}^C$, similar to above, the straightforward calculation yields that

$$\hat{\gamma}^C = \frac{1}{n}\sum_{i=1}^{n} \hat{\gamma}(x_i) = \gamma + E[b(k|X)] + E[\boldsymbol{v}(k|X)]^T \boldsymbol{W} + o(k^{-m/(2m+1)}).$$

This completes the proof. □

Proof of Theorem 3.4: First, the second order condition for $U_Y(1/(1-\tau)|x) = q_Y(\tau|x)$ yields that

$$\frac{q_Y(\tau_I|x)}{q_Y(\tau_E|x)} = \left(\frac{U_Y(1/(1-\tau_E)|x)}{U_Y(1/(1-\tau_I)|x)}\right)^{-1} = \left(\frac{1-\tau_I}{1-\tau_E}\right)^{-\gamma}\{1+o(k^{-m/(2m+1)})\}.$$

Furthermore, the result of Theorem 3.3 indicates that

$$\begin{aligned}\left(\frac{1-\tau_I}{1-\tau_E}\right)^{\hat{\gamma}(x)-\gamma} &= \exp\left[(\hat{\gamma}(x)-\gamma)\log\left(\frac{1-\tau_I}{1-\tau_E}\right)\right] \\ &= 1+(\hat{\gamma}(x)-\gamma)\log\left(\frac{1-\tau_I}{1-\tau_E}\right)(1+o_P(1)) \\ &= 1+\{b(k|x)+\boldsymbol{v}(k|x)^T\boldsymbol{W}\}\log\left(\frac{1-\tau_I}{1-\tau_E}\right)(1+o_P(1)).\end{aligned}$$

Meanwhile, we obtain

$$\frac{\hat{q}_Y(\tau_I|x)}{q_Y(\tau_I|x)} = 1+\frac{b_\lambda(\tau_I|x)}{q_Y(\tau_I|x)}+\frac{\boldsymbol{\nu}(\tau_I|x)^T}{q_Y(\tau_I|x)}\boldsymbol{W}+o_P(k^{-m/(2m+1)}),$$

where $\boldsymbol{W}$ is that given in the proof of Theorem 3.3. Using above, we have

$$\begin{aligned}\frac{\hat{q}_Y(\tau_E|x)}{q_Y(\tau_E|x)} &= \left(\frac{1-\tau_I}{1-\tau_E}\right)^{\hat{\gamma}(x)}\frac{\hat{q}_Y(\tau_I|x)}{q_Y(\tau_E|x)} \\ &= \left(\frac{1-\tau_I}{1-\tau_E}\right)^{\hat{\gamma}(x)}\frac{\hat{q}_Y(\tau_I|x)}{q_Y(\tau_I|x)}\frac{q_Y(\tau_I|x)}{q_Y(\tau_E|x)} \\ &= \left(\frac{1-\tau_I}{1-\tau_E}\right)^{\hat{\gamma}(x)}\frac{\hat{q}_Y(\tau_I|x)}{q_Y(\tau_I|x)}(1+o(k^{-m/(2m+1)})) \\ &= \left\{1+(b(k|x)+\boldsymbol{v}(k|x)^T)\boldsymbol{W})\log\left(\frac{1-\tau_I}{1-\tau_E}\right)(1+o_P(1))\right\} \\ &\quad\times\left[1+\left\{\frac{b_\lambda(\tau_I|x)}{q_Y(\tau_I|x)}+\frac{\boldsymbol{\nu}(\tau_I|x)^T}{q_Y(\tau_I|x)}\boldsymbol{W}\right\}(1+o_P(1))\right](1+o(k^{-m/(2m+1)})) \\ &= 1+\left\{\log\left(\frac{1-\tau_I}{1-\tau_E}\right)\boldsymbol{v}(k|x)+\frac{\boldsymbol{\nu}(\tau_I|x)}{q_Y(\tau_I|x)}\right\}^T\boldsymbol{W} \\ &\quad+b(k|x)\log\left(\frac{1-\tau_I}{1-\tau_E}\right)+\frac{b_\lambda(\tau_I|x)}{q_Y(\tau_I|x)} \\ &\quad+o\left(k^{-m/(2m+1)}\log\left(\frac{1-\tau_I}{1-\tau_E}\right)\right)+o_P(\{n(1-\tau_I)\}^{-m/(2m+1)}).\end{aligned}$$

Consequently, we obtain

$$\frac{\frac{\hat{q}_Y(\tau_E|x)}{q_Y(\tau_E|x)}-1-\text{bias}(\tau_E|x)}{s(\tau_E|x)} \xrightarrow{D} N(0,1),$$

where

$$\text{bias}(\tau_E|x) = b(k|x)\log\left(\frac{1-\tau_I}{1-\tau_E}\right)+\frac{b_\lambda(\tau_I|x)}{q_Y(\tau_I|x)} \tag{B.5}$$

and

$$s(\tau_E|x) = \left\|\log\left(\frac{1-\tau_I}{1-\tau_E}\right)\boldsymbol{v}(k|x)+q_Y(\tau_I|x)^{-1}\boldsymbol{\nu}(\tau_I|x)\right\|. \tag{B.6}$$

Here, for a vector $\boldsymbol{a}$, $||\boldsymbol{a}||$ means the ℓ_2-norm of $\boldsymbol{a}$. Furthermore, we get

$$E\left[\left\{\frac{\hat{q}_Y(\tau_E|x)}{q_Y(\tau_E|x)}-1\right\}^2\right]=O\left(\max\left\{k^{-\frac{2m}{2m+1}}\log^2\left(\frac{1-\tau_I}{1-\tau_E}\right),\{(1-\tau_I)n\}^{-\frac{2m}{2m+1}}\right\}\right).$$

Similarly, for the common index estimator $\hat{q}_Y^C(\tau_E|x)$, we have

$$\begin{aligned}\frac{\hat{q}_Y^C(\tau_E|x)}{q_Y(\tau_E|x)} &= 1+E[b(k|X)]\log\left(\frac{1-\tau_I}{1-\tau_E}\right)+\frac{b_\lambda(\tau_I|x)}{q_Y(\tau_I|x)}\\ &\quad+\left\{E[\boldsymbol{v}(k|X)]\log\left(\frac{1-\tau_I}{1-\tau_E}\right)+\frac{\boldsymbol{\nu}(\tau_I|x)}{q_Y(\tau_I|x)}\right\}^T\boldsymbol{W}\\ &\quad+o_P\left(k^{-m/(2m+1)}\log\left(\frac{1-\tau_I}{1-\tau_E}\right)\right)+o_P(\{(1-\tau_I)n\}^{-m/(2m+1)}).\end{aligned}$$

Accordingly,

$$\frac{\frac{\hat{q}_Y^C(\tau_E|x)}{q_Y(\tau_E|x)}-1-\mathrm{bias}^C(x)}{s^C(\tau_E|x)}\xrightarrow{D}N(0,1),$$

where

$$\mathrm{bias}^C(\tau_E|x)=E[b(k|X)]\log\left(\frac{1-\tau_I}{1-\tau_E}\right)+\frac{b_\lambda(\tau_I|x)}{q_Y(\tau_I|x)} \tag{B.7}$$

and

$$s^C(\tau_E|x)=\left|\left|E[\boldsymbol{v}(k|X)]\log\left(\frac{1-\tau_I}{1-\tau_E}\right)+\frac{\boldsymbol{\nu}(\tau_I|x)}{q_Y(\tau_I|x)}\right|\right|. \tag{B.8}$$

Finally, we obtain the optimal rate of convergence of MISE of the common index estimator as

$$E\left[\left\{\frac{\hat{q}_Y^C(\tau_E|x)}{q_Y(\tau_E|x)}-1\right\}^2\right]=O\left(\max\left\{k^{-\frac{2m}{2m+1}}\log^2\left(\frac{1-\tau_I}{1-\tau_E}\right),\{(1-\tau_I)n\}^{-\frac{2m}{2m+1}}\right\}\right). \qquad \square$$

ACKNOWLEDGMENTS

The authors are grateful to Editor, Associate Editor and anonymous referees for their valuable comments and suggestions, which led to improvements of the paper. The research of the author was partially supported by JSPS-KAKENHI 26730019 and 18K18011.

REFERENCES

[1] Barrow, D.L. and Smith, P.W. (1978). Asymptotic properties of best $L_2[0,1]$ approximation by spline with variable knots, *Quart. Appl. Math.*, **36**, 293–304.

[2] Beirlant, J. and Goegebeur, Y. (2004). Local polynomial maximum likelihood estimation for Pareto-type distributions, *J. Multivar. Anal.*, **89**, 97–118.

[3] Beirlant, J.; Goegebeur, Y.; Segers, J. and Teugels, J. (2004a). *Statistics of Extremes: Theory and Applications*, John Wiley & Sons.

[4] Beirlant, J.; de Wet, T. and Goegebeur, Y. (2004b). Nonparametric estimation of extreme conditional quantiles, *J. Statist. Comput. Simul.*, **74**, 567–580.

[5] Cardot, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions, *J. Nonparam. Statist.*, **12**, 503–538.

[6] Chernozhukov, V. (2005). Extremal quantile regression, *Ann. Statist.*, **33**, 806–839.

[7] Chernozhukov, V. and Fernández-val (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks, *The Review of Economic Studies*, **78**, 559–589.

[8] Cheng, Y.; de Gooijer, J.G. and Zerom, D. (2011). Efficient estimation of an additive quantile regression model, *Scandinavian Journal of Statistics*, **38**, 46–62.

[9] Claeskens, G.; Krivobokova, T. and Opsomer, J.D. (2009). Asymptotic properties of penalized spline estimators, *Biometrika*, **96**, 529–544.

[10] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag.

[11] Daouia, A.; Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression, *Bernoulli*, **19**, 2557–2589.

[12] Daouia, A.; Gardes, L; Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves, *Test*, **20**, 311–333.

[13] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, New York, Springer-Verlag.

[14] Demko, S. (1977). Inverses of band matrices and local convergence of spline projections, *SIAM. J. Numer. Anal.*, **14**, 616–619.

[15] Durrieu, G.; Grama, I.; Pham, Q.K. and Tricot, J.M. (2015). Nonparametric adaptive estimation of conditional probabilities of rare events and extreme quantiles, *Extremes*, **18**, 437–478.

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

[17] Einmahl, J.H.J.; de Haan, L. and Zhou, C. (2016). Statistics of heteroscedastic extremes, *J. Roy. Statist. Soc. Ser. B*, **78**, 31–51.

[18] El Methni, J.; Gardes, L. and Girard, S. (2014). Non-parametric estimation of extreme risk measures from conditional heavy-tailed distributions, *Scand. J. Statist.*, **41**, 988–1012.

[19] Gardes, L. and Girard, S. (2010). Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels, *Extremes*, **13**, 177–204.

[20] Gardes, L.; Girard, S. and Lekina, A. (2010). Functional nonparametric estimation of conditional extreme quantiles, *Journal of Multivariate Analysis*, **101**, 419–433.

[21] Girard, S. and Louhichi, S. (2015). On the strong consistency of the kernel estimator of extreme conditional quantiles, *Functional Statistics and Applications*, 59–77.

[22] Goegebeur, Y.; Guillou, A. and Osmann, M. (2017). A local moment type estimator for an extreme quantile in regression with random covariates, *Communications in Statistics – Theory and Methods*, **46**, 319–343.

[23] Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*, Chapman and Hall.

[24] He, F.; Cheng, Y. and Tong, T. (2016). Estimation of extreme conditional quantiles through an extrapolation of intermediate regression quantiles, *Statistics and Probability Letters*, **113**, 30–37.

[25] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Ann. Statist.*, **13**, 331–341.

[26] Horowitz, J. and Lee, S. (2005). Nonparametric estimation of an additive quantile regression model, *J. Amer. Statist. Assoc.*, **100**, 1238–1249.

[27] Kim, S.J.; Koh, K.; Boyd, S. and Gorinevsky, D. (2009). ℓ_1 trend filtering, *SIAM Review*, **51**, 339–360.

[28] Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions, *Ann. Statist.*, **26**, 755–770.

[29] Koenker, R. (2005). *Quantile Regression*, Cambridge University Press, Cambridge.

[30] Koenker, R. (2011). Additive models for quantile regression: Model selection and confidence bandaids, *Brazilian Journal of Probability and Statistics*, **25**, 239–262.

[31] Koenker, R. and Bassett, G. (1978). Regression quantiles, *Econometrica*, **46**, 33–50.

[32] Koenker, R.; Ng, P. and Portnoy, S. (1994). Quantile smoothing splines, *Biometrika*, **81**, 673–680.

[33] Lee, Y.K.; Mammen, E. and Park, B.U. (2010). Backfitting and smooth backfitting for additive quantile models, *Ann. Statist.*, **38**, 2857–2883

[34] Liang, X.; Zou, T.; Guo, B.; Li, S.; Zhang, H.; Zhang, S.; Huang, H. and Chen, S.X. (2015). Assessing Beijing's PM2.5 pollution: severity, weather impact, APEC and winter heating, *Proceedings of the Royal Society A*, **471**.

[35] Lu, Z. and Yu, K. (2004). Local linear additive quantile regression, *Scand. J. Statist.*, **31**, 333–346.

[36] Nychka, D.; Gray, G.; Haaland, P.; Martin, D. and O'Connell, M. (1995). A nonparametric regression approach to syringe grading for quality improvement, *J. Amer. Statist. Assoc.*, **90**, 1171–1178.

[37] O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems (with discussion), *Statist. Sci.*, **1**, 505–27.

[38] Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators, *Econometric Theory*, **7**, 186–199.

[39] Pratesi, M.; Ranalli, M.G. and Salvati, N. (2009). Nonparametric M-quantile regression using penalised splines, *J. Nonparam. Statist.*, **21**, 287–304.

[40] Reiss, P.T. and Huang, L. (2012). Smoothness selection for penalized quantile regression splines, *The International Journal of Biostatistics*, **8.1**.

[41] Ruppert, D. (2002). Selecting the number of knots for penalized splines, *J. Comp. Graph. Statist.*, **11**, 735–757.

[42] Sadhanala, V. and Tibshirani, R.J. (2017). Additive models with trend filtering, *arXiv*, arXiv:1702.05037v2.

[43] Tibshirani, R.J. (2014). Adaptive piecewise polynomial estimation via trend filtering, *Ann. Statist.*, **42**, 285–323.

[44] Wang, H.J.; Li, D. and He, X. (2012). Estimation of high dimensional conditional quantiles for heavy-tailed distributions, *J. Amer. Statist. Assoc.*, **107**, 1453–1464.

[45] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *J. Amer. Statist. Assoc.*, **73**, 812–815.

[46] Yoshida, T. (2013). Asymptotics for penalized spline estimators in quantile regression, *Communications in Statistics – Theory and Methods*, DOI:10.1080/03610926.2013.765477.

[47] Yuan, M. (2006). GACV for quantile smoothing splines, *Computational Statistics & Data Analysis*, **50**, 813–829.

[48] Zhou, S.; Shen, X. and Wolfe, D.A. (1998). Local asymptotics for regression splines and confidence regions, *Ann. Statist.*, **26**, 1760–1782.

REVSTAT – Statistical Journal
Volume 19, Number 3, July 2021, 399–420

AN ASYMMETRIC AREA MODEL-BASED APPROACH FOR SMALL AREA ESTIMATION APPLIED TO SURVEY DATA

Authors: Marcelo Rodríguez
– Facultad de Ciencias Básicas, Universidad Católica del Maule, Chile
mrodriguez@ucm.cl

Víctor Leiva *
– Escuela de Ingeniería Industrial, Pontificia Universidad Católica de Valparaíso, Chile
victorleivasanchez@gmail.com , www.victorleiva.cl

Mauricio Huerta
– Escuela de Ingeniería Industrial, Pontificia Universidad Católica de Valparaíso, Chile
mauricio.huerta.a@gmail.com

Camilo Lillo
– Center of Experience and Services (CES), Universidad Adolfo Ibáñez, Santiago, Chile
camilo.lillof@gmail.com

Alejandra Tapia
– Facultad de Ciencias Básicas, Universidad Católica del Maule, Chile
alejandraandreatapiasilva@gmail.com

Fabrizio Ruggeri
– CNR-IMATI, Italy
fabrizio@mi.imati.cnr.it

Received: March 2019 Revised: April 2019 Accepted: April 2019

Abstract:

- The Birnbaum–Saunders distribution is asymmetrical and has received considerable attention due to its properties and its relationship with the normal distribution. In this paper, we propose a methodology for estimating the mean of small areas based on a Birnbaum–Saunders distribution which is reparameterized in terms of its mean, similarly to the normal distribution, but in an asymmetric framework. In addition, the variance of the reparameterized Birnbaum–Saunders distribution is a function of its mean, similarly to the gamma distribution, which allows a GLM type modeling to be conducted. The Birnbaum–Saunders area model has properties that are unavailable in its competing models, as describing the mean in the original scale, unlike the existing models which employ a logarithmic transformation that reduces the test power and complicates the interpretation of results. The Birnbaum–Saunders area model can be formulated similarly as the Gaussian area model, permitting us to capture the essence of the small area estimation based on sample means and variances obtained from the areas. The methodology includes a formulation based on the Fay–Herriot model, estimation of model parameters with the maximum likelihood and Bayes empirical methods, as well as diagnostics using residuals. We illustrate the methodology with real-world survey data and compare the results with those obtained by the standard Fay–Herriot model.

Keywords:

- *empirical best linear unbiased predictor; R software; random effects; variance components.*

AMS Subject Classification:

- 62D05, 62F99, 62J99.

*Corresponding author.

1. INTRODUCTION

In sample surveys, it is of interest to obtain estimates for some parameters of the population from which the data are collected (Lumley and Scott, 2017 [31]). These estimates can be obtained not only for the target population, but also for sub-populations usually named small areas or domains. The small area estimation is a statistical technique used to estimate parameters in small sub-populations (Rao, 2003 [39]; Avila *et al.*, 2020 [3]), which may consist of geographical areas or socio-demographic groups, as a country, region, county, municipality or neighborhood.

Due to the high acceptance in relation to small area estimation, several models have been derived, used and analyzed. A summary of design-based small area estimation methodologies is presented in the book of Särndal *et al.* (2003) [46], whereas reviews of model-based small area estimation methodologies are provided in Ghosh and Rao (1994) [21], Rao (2003) [39], Datta (2009) [12], Lehtonen and Veijanen (2009) [25] and, more recently, in Coelho and Casimiro (2008) [9], Coelho and Pereira (2011) [10], Pereira and Coelho (2012) [36], Avila *et al.* (2020) [3] and Rueda *et al.* (2019) [43].

For small area estimation, the area model was first proposed by Fay and Herriot (1979) [19]. The Fay–Herriot (FH) model is considered as a generalization of the model formulated by Carter and Rolph (1974) [7], incorporating auxiliary variables (covariates). The FH model proposes an adaptation to the Carter-Rolph and James-Stein estimators, which was applied to income estimates in small areas during the population and housing census of the United States in 1970. The FH model assumes normality and incorporates linear regression in the context of heterogeneity of variances, so that it can be considered as a mixed model. To estimate the components of variance, different methods have been considered. Fay and Herriot (1979) [19] used weighted residual square sums and the moment method. Prasad and Rao (1990) [37] proposed an ordinary least square estimator. Datta and Lahiri (2000) [13] used the maximum likelihood (ML) and restricted maximum likelihood (REML) estimators.

When estimating means of small areas based on sampling design, there are desirable properties, such as unbiasedness and consistency, at country and region levels, but at lower levels (for example municipalities), the consistency property of the estimator is not fulfilled (Rao, 2003 [39]). Small area estimation is often based on the FH model, which allows for results in a more reliable way in order to produce statistics at lower levels than countries or regions. The FH model has good properties at low geographic levels when combining survey data with data from other sources, such as administrative or census records. In particular, the Chilean government has used the FH model since 2010 to estimate small areas (Casas-Cordero *et al.*, 2016 [8]). However, one of the drawbacks of the FH model is the assumption of normality for the response variable and random effect, because often this assumption is not fulfilled, due to asymmetry in the data distribution (Berg and Chandra, 2014 [5]). A solution to solve the problem of asymmetrical patterns present in the data is working with their log-transformations. However, data analyses performed under a wrong transformation reduces the power of the study (Huang and Qu, 2006 [22]; Dreassi *et al.*, 2014 [15]). Therefore, the research question is whether there is a gain in modifying the distributional assumption in terms of the accuracy of the estimator for producing statistics at a small area level or not.

Small-area estimation in non-normal models has been studied by few authors, even though this was postulated by Rao (2003) [39, Chap. 9] as an open problem. Fabrizi and Trivisano (2010) [18] extended the FH model assuming that the random effects follow power exponential distributions. Berg and Chandra (2014) [5] presented an empirical Bayes (EB) estimator for small area estimation based on a log-normal model and Fabrizi *et al.* (2016) [17] used the beta model for small area estimation.

The Birnbaum–Saunders (BS) distribution is asymmetrical and it has good properties (Ferreira *et al.*, 2012 [20]; Santos-Neto *et al.*, 2014 [44]; Bourguignon *et al.*, 2017 [6]). Statistical modeling based on the BS distribution has received much attention because of its relationship with the normal distribution and other properties. Rieck and Nedelman (1991) [41] were the pioneers in deriving BS regression models, whereas Villegas *et al.* (2011) [47] extended this regression model considering mixed effects and using an EB estimator to predict the random effects. Leiva *et al.* (2014) [29] and Santos-Neto *et al.* (2016) [45] focused on a reparameterized BS (RBS) distribution to model the response with no transformations following the idea of generalized linear models (McCullagh and Nelder, 1989 [32]). This modeling approach was based on fixed effects and no studies were reported using random effects. One of the parameters of the RBS distribution is its mean, such as the normal distribution, but in an asymmetric framework. In addition, the variance of the RBS distribution is a function of its mean, such as the gamma distribution. In Balakrishnan and Kundu (2019) [4] and Leiva *et al.* (2019) [27], detailed information is reviewed for these models. However, to the best of our knowledge, no area models for small area estimation based on BS, gamma and log-normal distributions have been reported in the literature.

In small area estimation, an alternative solution to solve the problem of asymmetric data is considering generalized linear models and, in particular, the RBS distribution (Leiva *et al.*, 2014 [29]). This solution provides some advantages over the log-transformation solution. First, the mean is modeled directly, making inference straightforward and avoiding the need of re-transformations back to the original scale. Second, this solution enables us to go beyond exponential family and allows some flexibility through the choice of a link function (for example, logarithmic, inverse or logit) and a distribution for the response through its mean-variance relationship. Moreover, the use of the the RBS distribution permits us to capture the essence of the small area estimation problem based on sample means and variances obtained from the areas, because it is possible to express its precision parameter as a function of these area means and variances, such as in the normal case; see Santos-Neto *et al.* (2014) [44] and Subsection 2.2 for more details about this important aspect. Therefore, the RBS distribution seems to be a good alternative to the FH type models for small area estimation.

The main objective of this work is to estimate the mean of small areas based on an RBS area model. The specific objectives are: (i) to establish an algorithm for estimating parameters from an RBS area model; (ii) to propose a residual for this model, allowing the examination of the model assumptions; and (iii) to illustrate the proposed methodology with survey data and to compare its results to the standard FH model. This methodology is implemented in the R software (`www.r-project.org` and R Core Team, 2016 [38]).

The paper is organized as follows. In Section 2, we present a background about the standard FH structure and a modeling approach based on the RBS distribution. Section 3 proposes the new RBS area model and its corresponding estimation, inference and residual analysis for its diagnostic. In Section 4, the methodology is illustrated with unpublished Chilean survey data, comparing it to a standard methodology. Section 5 gives our conclusions about this research.

2. BACKGROUND

In this section, we provide some preliminaries aspects related to the standard FH model and RBS regression.

2.1. The Fay–Herriot model

Fay and Herriot (1979) [19] proposed their model to improve the accuracy of the estimator $Y_i = \widehat{\theta}_i$ based on the sampling design (direct estimator) used to infer on the true small area mean θ_i, for $i = 1, ..., m$, where m is the number of areas. The FH model has a hierarchical structure consisting of the following two levels:

$$\begin{aligned} &\text{Level 1. Sampling model: } Y_i|\theta_i \overset{\text{IND}}{\sim} \text{N}(\theta_i, \psi_i), \text{ for } i = 1, ..., m, \\ &\text{Level 2. Linking model: } \quad \theta_i \overset{\text{IND}}{\sim} \text{N}(\underline{x}_i^\top \underline{\beta}, \sigma^2), \text{ for } i = 1, ..., m, \end{aligned} \tag{2.1}$$

where "IND" denotes "independent", ψ_i corresponds to the variance of the sampling error, $\underline{x}_i = (1, x_{1i}, ..., x_{(p-1)i})$ are the values of $p-1$ covariates for the area i, $\underline{\beta} = (\beta_0, \beta_1, ..., \beta_{p-1})^\top$ is a vector of unknown regression parameters, and σ^2 is the unknown variance of the area random effect, both to be estimated. Note that Level 1 describes the variability of the direct estimator $\widehat{\theta}_i$ of θ_i attributed to the sampling, whereas Level 2 links θ_i to the vector of $p-1$ known area covariates (Jiang and Lahiri, 2006 [23]; Li and Lahiri, 2010 [30]). Mixing the components of both models at Levels 1 and 2, we get the linear mixed model

$$Y_i|\theta_i = \underline{x}_i^\top \underline{\beta} + b_i + \varepsilon_i, \quad \varepsilon_i \overset{\text{IND}}{\sim} \text{N}(0, \psi_i), \quad i = 1, ..., m, \tag{2.2}$$

where $b_i \overset{\text{IID}}{\sim} \text{N}(0, \sigma^2)$ are independent and identically distributed (IID) area random effects with unknown σ^2 to be estimated from the data, whereas $\varepsilon_i \overset{\text{IND}}{\sim} \text{N}(0, \psi_i)$ are the sampling errors with known variances ψ_i. Furthermore, it is assumed that b_i and ε_i are independent random variables.

We want to estimate/predict the small area mean $\theta_i = \underline{x}_i^\top \underline{\beta} + b_i$, for $i = 1, ..., m$, and to obtain an uncertainty measurement related to this estimation/prediction. Considering the model defined in (2.2), the best predictor (BP) of θ_i (Rao and Molina, 2015 [40]), which minimizes the mean squared error, may be formulated as a weighted average of the direct estimator $\widehat{\theta}_i$ and the regression-synthetic estimator $\underline{x}_i^\top \underline{\beta}$ (Rao and Molina, 2015 [40]), expressed as

$$\widehat{\theta}_i^{\,\text{BP}} = (1 - B_i)\widehat{\theta}_i + B_i\, \underline{x}_i^\top \underline{\beta}, \quad i = 1, ..., m, \tag{2.3}$$

with the weight $0 < B_i < 1$ defined as $B_i = \psi_i/(\sigma^2 + \psi_i)$. Observe that $(1 - B_i)$ is function of the variance ratio σ^2/ψ_i and measures the uncertainty when θ_i is estimated in relation to the total variance $\sigma^2 + \psi_i$ (Rao and Molina, 2015 [40]). In addition, the parameter σ^2 is a homogeneity measure of the areas after accounting for the values $\underline{x}_i$ of covariates. If σ^2 is known, $\underline{\beta}$ may be approximated using the standard weighted least square estimator $\widetilde{\underline{\beta}}$ (Mert,

2015 [33]). Hence, by replacing it in (2.3), we obtain the best linear unbiased prediction (BLUP) of θ_i (Rao and Molina, 2015 [40]) by

$$\widehat{\theta}_i^{\,\mathrm{BLUP}} = (1 - B_i)\widehat{\theta}_i + B_i\, \underline{x}_i^{\top}\widetilde{\underline{\beta}}, \quad i = 1, ..., m, \tag{2.4}$$

where

$$\widetilde{\underline{\beta}} = \frac{\sum_{i=1}^{m} \underline{x}_i \widehat{\theta}_i/(\sigma^2 + \psi_i)}{\sum_{i=1}^{m} \underline{x}_i \underline{x}_i^{\top}/(\sigma^2 + \psi_i)}. \tag{2.5}$$

The BLUP of θ_i defined by (2.4) depends on σ^2 through of $\widetilde{\underline{\beta}}$, which is unknown in practice. From (2.4), we get the empirical best linear unbiased predictor (EBLUP) of θ_i as

$$\widehat{\theta}_i^{\,\mathrm{EBLUP}} = (1 - \widehat{B}_i)\widehat{\theta}_i + \widehat{B}_i\, \underline{x}_i^{\top}\widetilde{\underline{\beta}}, \tag{2.6}$$

where $\widehat{B}_i$ is the estimate of $B_i = \psi_i/(\sigma^2 + \psi_i)$ when σ^2 is replaced by an estimator $\widehat{\sigma}^{\,2}$, and $\widetilde{\underline{\beta}}$ is given in (2.5). Note that the model defined in (2.2) may be rewritten as matrix by

$$\underline{Y} = \boldsymbol{X}\underline{\beta} + \boldsymbol{I}_m \underline{b} + \underline{\varepsilon}, \tag{2.7}$$

where $\underline{Y} = (Y_1, ..., Y_m)^{\top}$, with $Y_i = \widehat{\theta}_i$, for $i = 1, ..., m$, $\boldsymbol{X} = (\underline{x}_1, ..., \underline{x}_m)^{\top}$ is of full rank, $\boldsymbol{I}_m$ is the $m \times m$ identity matrix, $\underline{\beta}$ is given below (2.1), $\underline{b} = (b_1, ..., b_m)^{\top}$ and $\underline{\varepsilon} = (\varepsilon_1, ..., \varepsilon_m)^{\top}$. Furthermore, $\underline{b}$ and $\underline{\varepsilon}$ are independently distributed with $\underline{b} \sim \mathrm{N}_m(\underline{0}_{m\times 1}, \boldsymbol{G})$, $\underline{\varepsilon} \sim \mathrm{N}_m(\underline{0}_{m\times 1}, \boldsymbol{R})$, where $\underline{0}_{m\times 1}$ is $m \times 1$ vector of zeros, $\boldsymbol{G} = \sigma^2 \boldsymbol{I}_m$ and $\boldsymbol{R}$ is a diagonal matrix defined as $\boldsymbol{R} = \mathrm{diag}\{\psi_1, ..., \psi_m\}$. The model defined in (2.7) is a particular case of a linear mixed model with its variance-covariance matrix assuming the form $\boldsymbol{V} = \boldsymbol{G} + \boldsymbol{R}$ (Datta *et al.*, 2005 [14]).

Observe that the EBLUP given in (2.6) depends on $\widehat{\sigma}^{\,2}$, with several methods being proposed in the literature for doing this estimation (Fay and Herriot, 1979 [19]; Prasad and Rao, 1990 [37]). The ML method has been widely used in small area estimation (Jiang and Lahiri, 2006 [23]; Rao and Molina, 2015 [40]), with Datta and Lahiri (2000) [13] using it in the context of the FH model. In this case, the log-likelihood function takes the form

$$\ell(\sigma^2, \underline{\beta}; \underline{y}) = c - \frac{1}{2}\log(|\boldsymbol{V}|) - \frac{1}{2}(\underline{y} - \boldsymbol{X}\underline{\beta})^{\top}\boldsymbol{V}^{-1}(\underline{y} - \boldsymbol{X}\underline{\beta}), \tag{2.8}$$

where c is a constant that is independent of σ^2 and $\underline{y}$ is the observed value of $\underline{Y}$. By taking derivatives of (2.8) with respect to $\underline{\beta}$ and σ^2, we obtain

$$\frac{\partial \ell(\sigma^2, \underline{\beta}; \underline{y})}{\partial \underline{\beta}} = \boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\underline{y} - \boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}\underline{\beta}, \tag{2.9}$$

$$\frac{\partial \ell(\sigma^2, \underline{\beta}; \underline{y})}{\partial \sigma^2} = \frac{1}{2}(\underline{y} - \boldsymbol{X}\underline{\beta})^{\top}\boldsymbol{V}^{-2}(\underline{y} - \boldsymbol{X}\underline{\beta}) - \frac{1}{2}\mathrm{tr}(\boldsymbol{V}^{-1}), \tag{2.10}$$

where $\mathrm{tr}(A)$ is the trace of the matrix $\boldsymbol{A}$. Thus, equating (2.9) and (2.10) to zero, and solving them simultaneously with respect to σ^2 and $\underline{\beta}$, we generate the corresponding ML estimators.

2.2. Birnbaum–Saunders statistical modeling

The BS distribution can be parameterized in terms of its mean μ and precision δ from its original parameterization by $\alpha = \sqrt{2/\delta}$ and $\beta = \delta\,\mu/(\delta+1)$ (Leiva, 2016 [26]). Thus, we have $\delta = 2/\alpha^2$ and $\mu = \beta\,(1+\alpha^2/2)$, where $\delta > 0$ and $\mu > 0$ (Santos-Neto *et al.*, 2016 [45]). Hence, if $Y \sim \mathrm{RBS}(\mu, \delta)$, its probability density function (PDF) is given by

$$f(y;\mu,\delta) = \frac{\exp(\delta/2)\,\sqrt{\delta+1}}{4\sqrt{\pi\mu}\,y^{3/2}}\left(y+\frac{\delta\mu}{\delta+1}\right)\exp\left(-\frac{\delta}{4}\left(\frac{(\delta+1)y}{\delta\mu}+\frac{\delta\mu}{(\delta+1)y}\right)\right),\ y>0. \tag{2.11}$$

The RBS PDF defined in (2.11) has diverse shapes as μ changes, when δ is fixed, and similarly as δ changes when μ is fixed. Note that the μ controls the scale of the RBS distribution but it is also its mean, which may be proved because $b\,Y \sim \mathrm{RBS}(b\mu, \delta)$, with $b > 0$. Notice that the parameter δ controls the shape of the RBS distribution, making it more platykurtic as δ increases. In addition, the RBS variance decreases when δ increases, converging to 5.0, as δ approaches zero, doing it to be a precision parameter, as mentioned. For more details about the graphical plots and shape analysis of the RBS distribution, see Leiva *et al.* (2014) [29], Balakrishnan and Kundu (2019) [4] and Leiva *et al.* (2019) [27].

Note that the random variables Y and Z with RBS and standard normal distributions, respectively, are related by

$$\begin{aligned} Y &= \frac{\delta\,\mu}{\delta+1}\left(\frac{Z}{\sqrt{2\,\delta}}+\sqrt{\left(\frac{Z}{\sqrt{2\,\delta}}\right)^2+1}\right)^2, \\ Z &= \sqrt{\frac{\delta}{2}}\left(\sqrt{\frac{(\delta+1)\,Y}{\mu\,\delta}}-\sqrt{\frac{\mu\,\delta}{(\delta+1)\,Y}}\right). \end{aligned} \tag{2.12}$$

Thus, from (2.12), the cumulative distribution function (CDF) and the quantile function (QF) of $Y \sim \mathrm{RBS}(\mu, \delta)$ are defined respectively as

$$\begin{aligned} F(y;\mu,\delta) &= \Phi\left(\sqrt{\frac{\delta}{2}}\left(\sqrt{\frac{(\delta+1)\,y}{\mu\,\delta}}-\sqrt{\frac{\mu\,\delta}{(\delta+1)\,y}}\right)\right), \quad y>0, \\ y(q;\mu,\delta) &= F^{-1}(q) = \frac{\delta\,\mu}{\delta+1}\left(\frac{z(q)}{\sqrt{2\,\delta}}+\sqrt{\left(\frac{z(q)}{\sqrt{2\,\delta}}\right)^2+1}\right)^2, \quad 0<q<1, \end{aligned} \tag{2.13}$$

where Φ and z are the standard normal CDF and QF, respectively, whereas F^{-1} is the inverse function of the RBS CDF. The mean and variance of $Y \sim \mathrm{RBS}(\mu, \delta)$ are given by $\mathrm{E}[Y] = \mu$ and $\mathrm{Var}[Y] = \psi = \mu^2(2\delta+5)/(\delta+1)^2$, respectively. Note the similarity of the variances of the RBS and gamma distributions, which allows the RBS distribution to model data analogously as in generalized linear models (Leiva *et al.*, 2014 [29]). Note also that, as mentioned, the RBS distribution has the mean as one of its parameters, which is an advantage on the gamma distribution. Note that, in small area estimation, one has available the sample mean and variance of each area, which is a natural aspect under normality. However, in the case of the RBS distribution, it is characterized by the mean (as in the normal case) but also by a precision parameter δ, which is different from the variance of the normal case. Santos-Neto *et al.* (2014) [44] proposed a moment estimator of δ through

$$\widehat{\delta} = \frac{\overline{Y} - S^2 + \sqrt{\overline{Y}^4 + 3\overline{Y}^2 S^2}}{S^2}, \tag{2.14}$$

where $\overline{Y}$ and S^2 represent the mean and sample variance of the random variable Y, respectively. Thus, (2.14) allows us to see the problem under the RBS perspective such as the normal framework.

Rieck and Nedelman (1991) [41] defined that if $Y \sim \text{BS}(\alpha, \beta)$, then $Z = \log(Y)$ follows a logarithmic BS distribution with shape parameter α and location parameter $\gamma = \log(\beta) \in \mathbb{R}$. In this regression model, the original response must be transformed to a logarithmic scale. Thus, although in this scale the mean $\gamma = \log(\beta)$ is modeled, in the natural scale $\beta = \exp(\gamma)$ is modeled, which in the BS case corresponds to the median. Leiva *et al.* (2014) [29] introduced a new approach for BS modeling, generalizing the existing works on the topic. In the estimation process, they considered $Y_1, ..., Y_m$ as independent $\text{RBS}(\mu_i, \delta)$ distributed random variables, for $i = 1, ..., m$. Then, the authors defined a statistical model based on the systematic component $\mu_i = g^{-1}(\underline{x}_i^\top \underline{\beta})$, where g^{-1} is the inverse function of the link function g, $\underline{\beta}$ is a vector of unknown parameters to be estimated, and $\underline{x}_i$ represents the values of the covariates. For the vector of parameters $(\underline{\beta}^\top, \delta)^\top$, simplifying the notation according to $\ell(\underline{\beta}, \delta; \underline{y}) = \ell(\underline{\beta}, \delta)$, $\ell_i(\mu_i, \delta; y_i) = \ell_i(\mu_i, \delta)$, and by using this same simplified notation from now on, the log-likelihood function of the model is given by $\ell(\underline{\beta}, \delta) = \sum_{i=1}^m \ell_i(\mu_i, \delta)$, where

$$\ell_i(\mu_i, \delta) = \frac{\delta}{2} - \frac{\log(16\pi)}{2} - \frac{1}{2}\log\left(\frac{(\delta+1)y_i^3\mu_i}{(\delta y_i + y_i + \delta\mu_i)^2}\right) - \frac{y_i(\delta+1)}{4\mu_i} - \frac{\delta^2\mu_i}{4(\delta+1)y_i}.$$

The score functions with first derivatives of β_l, for $l = 0, 1..., p-1$, and δ are respectively given by $\dot{\ell}_{\beta_l} = \partial\ell(\underline{\beta}, \delta)/\partial\beta_l$ and $\dot{\ell}_\delta = \partial\ell(\underline{\beta}, \delta)/\partial\delta$. Thus, the score vector is $\underline{\dot{\ell}}_{\underline{\beta},\delta} = (\underline{\dot{\ell}}_{\underline{\beta}}^\top, \dot{\ell}_\delta)^\top$; see details in Leiva *et al.* (2014) [29]. To estimate the model parameters by the ML method, the equation $\underline{\dot{\ell}}_{\beta,\delta} = \underline{0}_{p\times 1}$ must be solved. However, no closed-form expressions for these estimates are available. Then, an iterative approach is needed, such as the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm; see details in Nocedal and Wright (1999) [35]. This iterative approach is used for solving unconstrained non-linear optimization problems, belonging to the class of quasi-Newton methods.

3. THE NEW STATISTICAL MODEL

In this section, we propose a methodology based on the FH model and the RBS regression model. The methodology considers the formulation of the new RBS area model, the estimation algorithm and inference for the population mean, as well as a residual analysis for model diagnostics. The standard FH model defined in (2.1) assumes normality for random effects and errors. In this case, the EB estimator and the EBLUP coincide. Note that the distribution of the direct domain mean estimator comes from the survey design, which from design-based theory is known to be approximately normal (for large enough samples). The normal approximation is not necessarily good in small areas with very small sample sizes. We consider the RBS distribution to model small area mean, whereas the random effect distribution is also assumed RBS for computational and theoretical convenience. When non-normality is assumed in the response or in the random effects, Rao (2003) [39] proposed to use the EB estimator.

3.1. Formulation

Such as in the standard model defined in (2.1), the proposed model consists of the two following levels:

$$\begin{aligned} &\text{Level 1. Sampling model: } Y_i|\theta_i \overset{\text{IND}}{\sim} \text{RBS}(\theta_i, \delta_i), \text{ for } i = 1, ..., m, \\ &\text{Level 2. Linking model: } \quad \theta_i \overset{\text{IND}}{\sim} \text{RBS}(g^{-1}(\underline{x}_i^\top \underline{\beta}), \kappa), \text{ for } i = 1, ..., m, \end{aligned} \tag{3.1}$$

where θ_i is the mean of the area i, g^{-1} is the inverse of the link function g, $\underline{\beta}$ and $\underline{x}_i$ are as defined in (2.1), whereas κ is the unknown precision parameter of the area random effect to be estimated. Note that δ_i depends on known variances ψ_i of the area i which are related according to the results proposed by Santos-Neto *et al.* (2014) [44], from where the empirical relationship is given in (2.14). Therefore, from this relationship, we have

$$\delta_i = \frac{\theta_i - \psi_i + \sqrt{\theta_i^4 + 3\theta_i^2\psi_i}}{\psi_i}, \quad i = 1, ..., m. \tag{3.2}$$

Thus, from (3.2), we put the model proposed in (3.1) in a small area framework.

The proposed BS area models have properties that are unavailable in the models of this type existing in the literature. Specifically, the BS area models considered in this work allow us to describe the mean of the data in their original scale, unlike the existing models, which employ a logarithmic transformation of the data, provoking a possible reduction of the power of the study and difficulties of interpretation. In addition, these BS area models can be formulated in a similar form as the normal area models, permitting us to capture the essence of the small area estimation problem based on sample means and variances obtained from the areas.

3.2. EB estimation and quadrature methods

We consider the EB approach to estimate the small area mean. First, by considering the PDF given in (2.11), we obtain the marginal PDF from the conditional (sampling model) and prior (linking model) distributions. Second, we estimate the parameters $\underline{\beta}$ and κ based on the corresponding marginal likelihood function. Third, we obtain the posterior distribution by plugging it in the estimated value of $\underline{\lambda} = (\underline{\beta}^\top, \kappa)^\top$. Fourth, we find the EB estimator of the conditional expectation of a small area mean given the observed data with respect to the RBS area model. In order to calculate this expected value, we use the posterior distribution presented in (3.13). The EB approach described above is detailed in Algorithm 1.

Algorithm 1 – Empirical Bayes approach

1: Establish the conditional PDF of Y_i given θ_i, denoted by $f(y_i|\theta_i)$, for $i = 1, ..., m$.

2: Indicate the prior distribution $\pi(\theta_i; \underline{\lambda})$, for $i = 1, ..., m$.

3: Obtain the marginal PDF

$$m(y_i; \underline{\lambda}) = \int_{\mathrm{R}_{\theta_i}} f(y_i|\theta_i)\pi(\theta_i; \underline{\lambda})\, \mathrm{d}\theta_i, \quad i = 1, ..., m,$$

recalling that R_{θ_i} is the parameter space of θ_i.

4: Estimate the model parameter $\underline{\lambda}$ by maximizing the marginal likelihood function

$$L(\underline{\lambda}) = \prod_{i=1}^{m} \int_{\mathrm{R}_{\theta_i}} f(y_i|\theta_i)\pi(\theta_i; \underline{\lambda})\, \mathrm{d}\theta_i.$$

5: Calculate the posterior distribution

$$\pi(\theta_i|y_i; \widehat{\underline{\lambda}}) = \frac{f(y_i|\theta_i)\pi(\theta_i; \widehat{\underline{\lambda}})}{\int_{\mathrm{R}_{\theta_i}} f(y_i|\theta_i)\pi(\theta_i; \widehat{\underline{\lambda}})\, \mathrm{d}\theta_i}, \quad i = 1, ..., m,$$

to make inferences about θ_i, where $\widehat{\underline{\lambda}}$ is an estimator of $\underline{\lambda}$.

6: Determine the EB estimator of θ_i using

$$\widetilde{\theta}_i^{\mathrm{EB}} = \mathrm{E}(\theta_i|y_i; \widehat{\underline{\lambda}}) = \frac{\int_{\mathrm{R}_{\theta_i}} \theta_i f(y_i|\theta_i)\pi(\theta_i; \widehat{\underline{\lambda}})\, \mathrm{d}\theta_i}{\int_{\mathrm{R}_{\theta_i}} f(y_i|\theta_i)\pi(\theta_i; \widehat{\underline{\lambda}})\, \mathrm{d}\theta_i}, \quad i = 1, ..., m.$$

The conditional PDF (sampling model), for $i = 1, ..., m$, is given by

$$f(y_i|\theta_i) = \frac{\exp(\delta_i/2)\sqrt{\delta_i + 1}}{4\sqrt{\pi\theta_i}\, y_i^{3/2}} \left(y_i + \frac{\delta_i\theta_i}{\delta_i + 1}\right) \exp\left(-\frac{\delta_i}{4}\left(\frac{y_i(\delta_i + 1)}{\delta_i\theta_i} + \frac{\delta_i\theta_i}{y_i(\delta_i + 1)}\right)\right), \tag{3.3}$$

whereas the prior distribution, for $i = 1, ..., m$, is defined as

$$\pi(\theta_i; \underline{\lambda}) = \frac{\exp(\kappa/2)\sqrt{\kappa + 1}}{4\sqrt{\pi\, g^{-1}(\underline{x}_i^\top \underline{\beta})}\, \theta_i^{3/2}} \left(\theta_i + \frac{\kappa g^{-1}(\underline{x}_i^\top \underline{\beta})}{\kappa + 1}\right) \exp\left(-\frac{\kappa}{4}\left(\frac{\theta_i(\kappa + 1)}{\kappa g^{-1}(\underline{x}_i^\top \underline{\beta})} + \frac{\kappa g^{-1}(\underline{x}_i^\top \underline{\beta})}{\theta_i(\kappa + 1)}\right)\right). \tag{3.4}$$

Based on (3.3) and (3.4), the marginal PDF is obtained as

$$m(y_i; \underline{\lambda}) = \int_0^\infty f(y_i|\theta_i)\pi(\theta_i; \underline{\lambda})\, \mathrm{d}\theta_i, \quad i = 1, ..., m. \tag{3.5}$$

In order to calculate the integral given in (3.5), a Gaussian quadrature can be used. A quadrature rule is an approximation of the definite integral of a function, usually stated as a weighted sum of values at specified points within the domain of integration, which is conventionally taken as $[-1, 1]$. Thus, this rule may be stated as

$$\int_{-1}^{1} f(u)\, \mathrm{d}u = \sum_{j=1}^{n} w_j f(u_j). \tag{3.6}$$

Observe that the Gaussian quadrature given in (3.6) only produces good results if the function f is well approximated by a polynomial function within the range $[-1, 1]$. Then, the integration problem presented in (3.5) can be expressed in a more general way by introducing a positive weight function ω into the integrand, and allowing an interval other than $[-1, 1]$. In this way, the problem reduces to calculating

$$\int_a^b \omega(u)\, f(u)\, \mathrm{d}u, \tag{3.7}$$

for some choices of a, b and ω. Note that if $a = -1$, $b = 1$ and $\omega(u) = 1$, the integral given in (3.7) is the same as that given in (3.6). Some particular cases of the Gaussian quadrature are presented in Table 1.

Table 1: Intervals and forms for $\omega(u)$ of some Gaussian quadratures corresponding to the indicated orthogonal polynomial.

Interval	$\omega(u)$	Orthogonal polynomial
$[-1, 1]$	1	Legendre
$(-1, 1)$	$(1-u)^{\alpha}(1+u)^{\beta}, \quad \alpha, \beta > -1$	Jacobi
$(-1, 1)$	$1/\sqrt{1-u^2}$	Chebyshev
$[0, \infty)$	$\exp(-u)$	Laguerre
$(-\infty, \infty)$	$\exp(-u^2)$	Hermite

Note that the Gauss–Laguerre (GL) quadrature is an extension of the Gaussian quadrature method over the interval $[0, \infty)$ to approximate the integral obtained in (3.5) (Abramowitz and Stegun, 1972 [1]). Therefore, we approximate the marginal PDF presented in (3.5) by the GL quadrature by means of

$$m(y_i; \underline{\beta}, \kappa) = \sum_{j=1}^{n} w_j f(y_i|\theta_{ij}) \pi(\theta_{ij}; \underline{\lambda}) \exp(\theta_{ij}), \quad i = 1, ..., m, \tag{3.8}$$

where n is the number of quadrature points, m is the number of areas, θ_{ij} is the jth root of the Laguerre polynomial in the area i given by

$$L_n(\theta_{ij}) = \sum_{r=0}^{n} \binom{n}{r} \frac{(-1)^r}{r!} \theta_{ij}^r,$$

and the weight w_j is given by

$$w_j = \frac{\theta_{ij}}{(n+1)^2 (L_{n+1}(\theta_{ij}))^2}, \quad i = 1, ..., m, \quad j = 1, ..., n.$$

3.3. ML estimation and Fisher information

Once the marginal PDF presented in (3.5) is approximated by the GL quadrature, we can approximate the corresponding likelihood function to estimate the parameters of the model defined in (3.1) with the ML method. Recalling that $\underline{\lambda} = (\underline{\beta}^\top, \kappa)^\top$, the marginal likelihood function is given by

$$L(\underline{\lambda}) = \prod_{i=1}^{m} m(y_i; \underline{\lambda}).$$

Therefore, the corresponding log-likelihood function approximated by the GL quadrature is given by

$$\ell(\underline{\lambda}) = \sum_{i=1}^{m} \log \left(\sum_{j=1}^{n} w_j f(y_i|\theta_{ij}) \pi(\theta_{ij}; \underline{\lambda}) \exp(\theta_{ij}) \right). \tag{3.9}$$

The respective score vector, obtained by differentiating (3.9) with respect to $\underline{\lambda}$, is established as

$$\dot{\underline{\ell}}(\underline{\lambda}) = \frac{\partial \ell(\underline{\lambda})}{\partial \underline{\lambda}} = (\dot{\underline{\ell}}_{\underline{\beta}}(\underline{\lambda})^\top, \dot{\ell}_\kappa(\underline{\lambda}))^\top.$$

The ML estimates of $\underline{\beta}$ and κ, $\widehat{\underline{\beta}}$ and $\widehat{\kappa}$ namely, respectively, are the solution to the system of equations given by $\dot{\underline{\ell}}_{\underline{\beta}}(\underline{\lambda}) = \underline{0}_{p\times 1}$ and $\dot{\ell}_\kappa(\underline{\lambda}) = 0$. Since the corresponding ML estimates cannot be expressed in a closed form, we compute them by maximizing the log-likelihood function defined in (3.9) numerically with the BFGS algorithm. As starting values, the estimates obtained under an RBS regression model can be considered.

The second derivatives of $\ell(\underline{\lambda})$ defined in (3.9), with respect to $\underline{\beta}$ and κ, are expressed as

$$\frac{\partial^2 \ell(\underline{\lambda})}{\partial \beta_l \partial \beta_k}, \frac{\partial^2 \ell(\underline{\lambda})}{\partial \beta_l \partial \kappa}, \frac{\partial^2 \ell(\underline{\lambda})}{\partial \kappa^2}, \quad l = 0, 1, ..., p-1.$$

Consequently, the corresponding Hessian matrix is given by

$$\ddot{\boldsymbol{\ell}}(\underline{\lambda}) = \begin{pmatrix} \dfrac{\partial^2 \ell(\underline{\lambda})}{\partial \underline{\beta} \partial \underline{\beta}^\top} & \dfrac{\partial^2 \ell(\underline{\lambda})}{\partial \underline{\beta} \partial \kappa} \\ \dfrac{\partial^2 \ell(\underline{\lambda})}{\partial \kappa \partial \underline{\beta}^\top} & \dfrac{\partial^2 \ell(\underline{\lambda})}{\partial \kappa^2} \end{pmatrix}.$$

In addition, the expected Fisher information matrix is obtained as

$$\boldsymbol{K}(\underline{\lambda}) = -\mathrm{E}[\ddot{\boldsymbol{\ell}}(\underline{\lambda})]. \tag{3.10}$$

3.4. Inference

Regularity conditions (see Cox and Hinkley, 1974 [11]) must be fulfilled for an RBS area model if its parameters are within the parameter space. Then, the ML estimator $\widehat{\underline{\lambda}}$ is consistent and follows an asymptotic joint distribution, which is normal with asymptotic mean $\underline{\lambda}$, and an asymptotic variance-covariance matrix $\boldsymbol{\Sigma}(\underline{\lambda})$. Thus, as $m \to \infty$ and recalling that $\underline{\lambda} = (\underline{\beta}^\top, \kappa)^\top$, we have

$$\sqrt{n}\,(\widehat{\underline{\lambda}} - \underline{\lambda}) \quad \xrightarrow{\mathrm{D}} \quad \mathrm{N}_{p+1}(\underline{0}_{(p+1)\times 1}, \boldsymbol{\Sigma}(\underline{\lambda})), \tag{3.11}$$

where $\xrightarrow{\mathrm{D}}$ denotes convergence in distribution. Note that if $\boldsymbol{J}(\underline{\lambda}) = \lim_{n\to\infty}(1/n)\boldsymbol{K}(\underline{\lambda})$ exists and is non-singular, with $\boldsymbol{K}(\underline{\lambda})$ being the expected Fisher information matrix given in (3.10), then $\boldsymbol{\Sigma}(\underline{\lambda}) = \boldsymbol{J}(\underline{\lambda})^{-1}$. The diagonal elements of $\boldsymbol{K}(\underline{\lambda})^{-1}$, $k_{ll}^{-1}(\underline{\lambda})$ namely, may be used for approximating the corresponding asymptotic standard errors (SEs), that is, by using

$$\mathrm{SE}[\widehat{\underline{\lambda}}_l] = \sqrt{k_{ll}^{-1}(\underline{\lambda})}, \quad l = 1, ..., p+1. \tag{3.12}$$

Note that $\widehat{\boldsymbol{K}}(\underline{\lambda})^{-1} = \boldsymbol{K}(\widehat{\underline{\lambda}})^{-1}$ is a consistent estimator of $\boldsymbol{\Sigma}(\underline{\lambda})$ and then the associated asymptotic SEs given in (3.12) may be estimated as $\widehat{\mathrm{SE}}[\widehat{\underline{\lambda}}_l] = (k_{ll}^{-1}(\widehat{\underline{\lambda}}))^{1/2}$, for $l = 1, ..., p+1$. Asymptotic inference on parameters can be conducted using (3.11) and (3.12).

3.5. Estimating the small area mean and bootstrapping

To estimate a small area mean, we use the posterior PDF evaluated at the ML estimates given by

$$\pi(\theta_i | y_i; \widehat{\underline{\beta}}, \widehat{\kappa}) = \frac{f(y_i|\theta_i)\pi(\theta_i; \widehat{\underline{\beta}}, \widehat{\kappa})}{m(y_i; \widehat{\underline{\beta}}, \widehat{\kappa})}, \quad i = 1, ..., m, \tag{3.13}$$

where $m(y_i; \widehat{\underline{\beta}}, \widehat{\kappa})$ is presented in (3.8), and $\widehat{\underline{\beta}}, \widehat{\kappa}$ are the corresponding ML estimates. Therefore, the EB estimator for the mean of an RBS area model, based on the GL quadrature, is given by

$$\widetilde{\theta}_i^{\mathrm{EB}} = \mathrm{E}(\theta_i | y_i; \widehat{\underline{\beta}}, \widehat{\kappa}) = \frac{\sum_{j=1}^n w_j \theta_{ij} f(y_i|\theta_{ij}) \pi(\theta_{ij}; \widehat{\underline{\beta}}, \widehat{\kappa}) \exp(\theta_{ij})}{\sum_{j=1}^n w_j f(y_i|\theta_{ij}) \pi(\theta_{ij}; \widehat{\underline{\beta}}, \widehat{\kappa}) \exp(\theta_{ij})}, \quad i = 1, ..., m. \tag{3.14}$$

Suppose that we have a random sample from an unknown distribution function F, and we want to make statistical inference about a parameter θ_i, for $i = 1, ..., m$. Bootstrapping is a non-parametric approach which relies upon the assumption that the current sample is representative of the population, and therefore, the empirical CDF $\widehat{F}$ is a non-parametric estimate of the population CDF F. From the sample, the statistic of interest, $\widetilde{\theta}_i^{\mathrm{EB}}$ namely, can be calculated as an empirical estimate of the true parameter. To measure the accuracy of the estimator, a bootstrapped SE, defined as

$$\mathrm{SE}(\widetilde{\theta}_i^{\mathrm{EB}}) = \sqrt{\mathrm{Var}(\widetilde{\theta}_i^{\mathrm{EB}})}, \quad i = 1, ..., m,$$

can be calculated; see Algorithm 2.

Algorithm 2 – Bootstrap standard error

1: Collect a random sample of size m with replacement (bootstrap sample) from a matrix of data with m rows corresponding to the areas and three columns related to the response $Y_i = \widehat{\theta}_i$, which is based on the sampling design used to estimate the true small area mean θ_i, the variance of the sampling error ψ_i, and the covariates $\underline{x}_i$, for $i = 1, ..., m$.

2: Fit an RBS area model with the bootstrap sample of Step 1 and compute the statistic of interest $\widetilde{\theta}_i^{\mathrm{EB}}$, for $i = 1, ..., m$.

3: Repeat Steps 1–2 a large number of times (for example, $\mathrm{B} = 10,000$) and compute B bootstrap values of $\widetilde{\theta}_i^{\mathrm{EB}}$, which forms its empirical sampling distribution.

4: Calculate the sample standard deviation (SD) of the B bootstrap values of $\widetilde{\theta}_i^{\mathrm{EB}}$, which allows us to obtain the bootstrap SE of $\widetilde{\theta}_i^{\mathrm{EB}}$, for $i = 1, ..., m$.

3.6. Model selection

Models are often compared using selection measures as the log-likelihood function or Akaike information (AIC) and Bayesian information (BIC) criteria. Note that AIC and BIC are defined as

$$\mathrm{AIC} = -2\ell(\widehat{\underline{\lambda}}) + 2(p+1), \quad \mathrm{BIC} = -2\ell(\widehat{\underline{\lambda}}) + (p+1)\log(m), \tag{3.15}$$

where ℓ is the corresponding log-likelihood function given in (3.9), $p+1$ is the number of parameters and m the number of areas. AIC and BIC correspond to the log-likelihood function plus a component penalizing such a function, as the model has more parameters making it more complex. A model with a smaller AIC or BIC is better than another competing model (Ferreira *et al.*, 2012 [20]).

3.7. Diagnostic analysis

Residuals are frequently used to validate the assumptions of statistical models and may also be employed as tools for model selection. Based on Nobre and da Motta-Singer (2007) [34], we define a conditional residual which follows a standard normal distribution and accommodates the extra source of variability present in linear mixed models as $r_i^{(\mathrm{C})} = y_i - \widetilde{\theta}_i^{\mathrm{EB}}$, where $\widetilde{\theta}_i^{\mathrm{EB}}$ is given in (3.14) and y_i is an observed value of Y_i. We consider the randomized quantile (RQ) residual proposed by Dunn and Smyth (1996) [16], which is useful for asymmetric distributions. We use an index plot of the conditional RQ residual to verify homoscedasticity, whereas the distributional assumption is analyzed by simulated envelopes (Atkinson, 1985 [2]). For the RBS area model proposed in this work, the conditional RQ residual is defined as

$$r_i^{\mathrm{RQ(C)}} = \phi^{-1}(F(y_i; \widetilde{\theta}_i^{\mathrm{EB}}, \widehat{\kappa})) \quad i = 1, ..., m, \tag{3.16}$$

where F is the RBS CDF defined in (2.13). As F is continuous, then $F(Y_i)$ is uniformly distributed on the unit interval. In order to verify the normality of the conditional RQ residual based on the RBS area model, we utilize a theoretical quantile versus empirical quantiles (QQ) plot with simulated envelopes proposed by Atkinson (1985) [2]; see Algorithm 3.

Algorithm 3 – Goodness of fit to any distribution based on QQ plots with simulated envelopes.

1: Collect data $y_1, ..., y_m$.
2: Obtain the empirical quantiles $y_{i:m}$ as observed order statistics for $i = 1, ..., m$ from $y_1, ..., y_m$.
3: Estimate the parameters of the model by $\widehat{\lambda}$ with $y_1, ..., y_m$.
4: Compute $w_{i:m} = (i - 0.5)/m$, for $i = 1, ..., m$.
5: Calculate the theoretical quantiles $t_{i:m} = F^{-1}(w_{i:m})$, where F^{-1} is the inverse function of the CDF F.
6: Draw the QQ plot with points $y_{i:m}$ versus $t_{i:m}$, for $i = 1, ..., m$.
7: Specify an α level for the simulated envelopes.
8: Generate s samples of size m from a distribution with CDF F and estimated parameters $\widehat{\lambda}$.
9: Construct envelopes with limits given by $l_i = y_{i:m}(\alpha/2)$ and $u_i = y_{1:m}(1 - \alpha/2)$ for $i = 1, ..., m$.
10: Establish that the assumed distribution is adequate if all the points are inside of the envelope, otherwise it is not adequate.

4. SURVEY DATA ANALYSYS

In this section, we provide an illustrative example with a Chilean survey data set for analysis of service quality. Also, we compare the results obtained with the proposed methodology to a standard methodology based on the normal distribution.

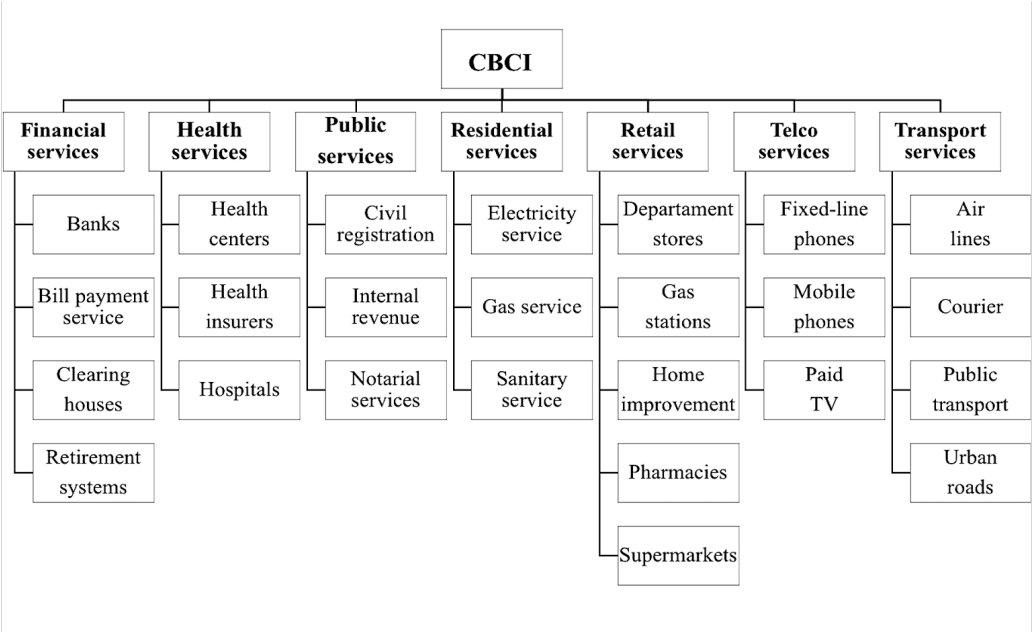


Figure 1: Structure of Chilean industries and sectors used to calculate the CBCI in 2017.

4.1. The data set

The data set under analysis was collected between January-2017 and November-2017 in 34 of 52 municipalities located at the Metropolitan region of Chile. In this data set, the response is the Chilean business confidence index (CBCI). This index is built from a sample survey which measures the confidence of customers towards the service provided by diverse companies. The CBCI is calculated by the Center of Experiences and Services (CES) of the Adolfo Ibáñez University (UAI), CES-UAI in short; see http://www.ces-uai.cl and more details of the CBCI in Leiva *et al.* (2018) [28]. Figure 1 shows the industrial sectors that allow us to estimate the CBCI. In this study, we consider as covariate the urban life quality index (ULQI) which allows us to model the CBCI. This covariate is obtained from the Institute of Urban and Territorial Studies of the Pontifical Catholic University of Chile (http://fadeu.uc.cl). The data set used in this illustration is presented in Table 2.

Table 2: CBCI (with variance and size sample) and UQLI values for the indicated municipality.

Municipality ID	$Y_i\|\theta_i$	ψ_i	n_i	x_i
1. Pedro Aguirre Cerda (PC)	30.11	83.93	382	26.45
2. Conchalí (CO)	30.32	81.32	508	30.74
3. Quinta Normal (QN)	31.17	82.77	401	30.18
4. Lo Espejo (LE)	31.49	82.69	416	24.11
5. Cerro Navia (CN)	31.80	82.34	522	26.98
6. La Granja (LG)	32.23	78.28	453	33.98
7. Renca (RN)	32.63	83.67	472	36.42
8. Independencia (IN)	34.41	80.64	529	30.05
9. Estación Central (EC)	34.81	81.91	497	33.41
10. Lo Prado (LP)	34.81	83.05	451	30.09
11. San Ramón (SR)	35.63	84.88	394	35.53
12. Quilicura (QU)	37.13	83.31	505	39.70
13. El Bosque (EB)	37.25	80.58	502	28.10
14. Pudahuel (PU)	37.28	80.74	566	36.27
15. Puente Alto (PA)	37.87	79.54	676	36.92
16. Huechuraba (HU)	38.46	78.78	559	37.26
17. La Pintana (LA)	38.99	79.32	477	24.29
18. San Joaquín (SJ)	39.18	79.05	462	38.29
19. La Cisterna (LC)	39.23	80.12	418	32.89
20. Recoleta (RE)	40.00	79.11	520	32.36
21. Cerrillos (CE)	42.25	79.10	426	32.65
22. San Miguel (SM)	42.66	78.59	511	43.42
23. Maipú (MP)	43.50	78.39	1016	46.43
24. San Bernardo (SB)	43.91	76.56	608	28.93
25. Santiago (SA)	44.00	78.14	759	40.55
26. Peñalolen (PE)	48.54	75.99	789	38.83
27. La Florida (LF)	49.22	74.69	963	38.95
28. Macul (MA)	49.50	79.59	605	47.87
29. La Reina (LR)	51.82	74.49	716	52.45
30. Ñuñoa (NU)	52.14	73.89	980	54.27
31. Lo Barnechea (LB)	56.08	73.62	658	57.67
32. Vitacura (VI)	65.60	72.21	643	57.93
33. Providencia (PR)	71.10	68.81	928	59.96
34. Las Condes (LN)	73.60	72.58	1099	63.61

4.2. Exploratory data analysis

Table 3 provides a descriptive summary of the CBCI in the different municipalities of the Chilean Metropolitan region, which includes $\overline{y}$, median (MD), SD, coefficients of variation (CV) of skewness (CS) and of kurtosis (CK), as well as the minimum ($y_{(1)}$) and maximum ($y_{(m)}$) values. Figure 2 presents the histogram, adjusted box-plot and standard box-plot of the CBCI, as well as the scatter-plot between CBCI and UQLI. Figure 3 displays the map of the municipalities (with their abbreviations detailed in Table 3) located in the Chilean Metropolitan region with their corresponding CBCI colored in gray according to an intensity related to the value of this index.

Table 3: Descriptive statistics for the CBCI in municipalities of the Chilean Metropolitan region.

$y_{(1)}$	MD	$\overline{y}$	$y_{(m)}$	SD	CV	CS	CK
30.11	39.09	42.32	73.6	11.12	26.27	1.36	4.33

Based on Figure 2 and Table 3, we conduct an exploratory data analysis (EDA). First, from Figure 2 (left and center), note that the CBCI follows a positive skew (asymmetric) distribution (CS $>$ 0). We use an adjusted boxplot for asymmetric data (see Rousseeuw *et al.*, 2016 [42]), from which we conclude that there are no atypical data. In addition, Figure 2 (right) presents a linear or logarithmic relationship between CBCI and UQLI. Furthermore, a non-constant variance is detected by this scatter-plot. Supported by this EDA, the RBS area model proposed in this work seems to be a good candidate to describe the data set under study.

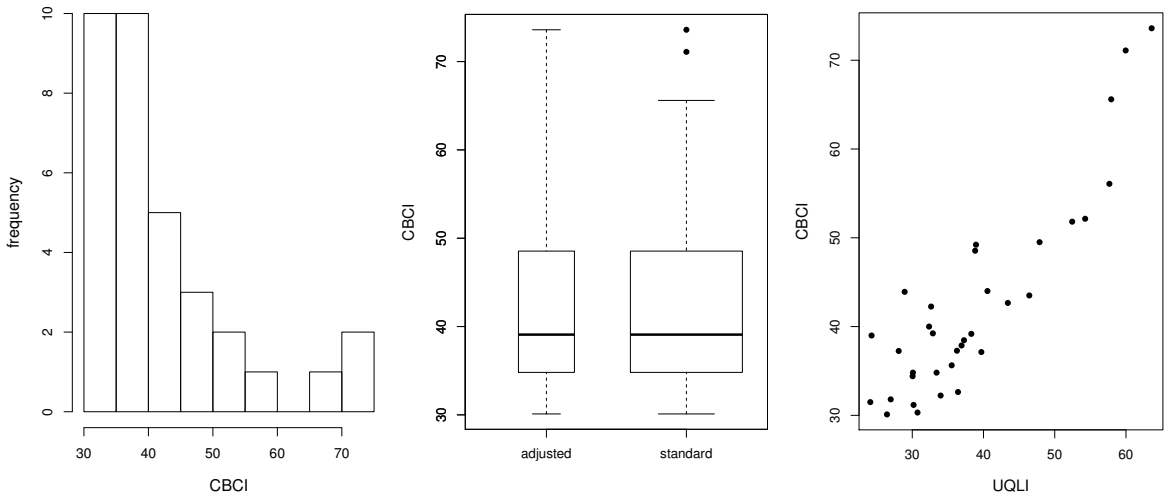


Figure 2: Histogram (left) and box-plot (center) of CBCI, and scatterplot between CBCI and UQLI (right).

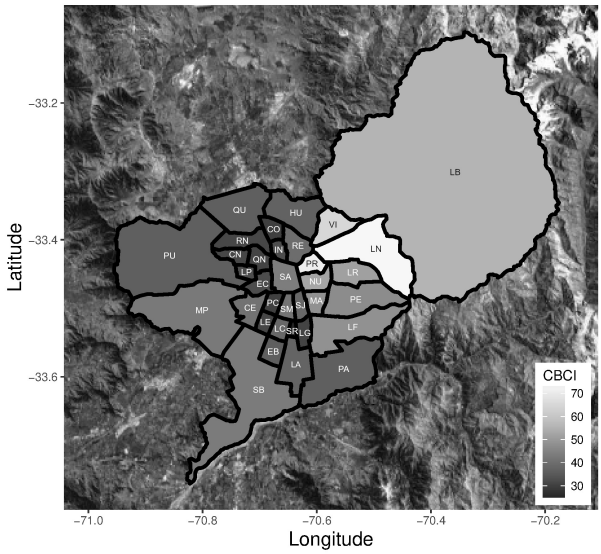


Figure 3: Map with CBCI of the indicated municipalities located at the Chilean Metropolitan region.

4.3. Modeling, estimation and inference

Based on the EDA above performed, we use the RBS area model defined in (3.1), with $i = 1, ..., 34$. In addition, δ_i can be obtained from (3.2) as $\delta_i = (y_i - \psi_i + (y_i^4 + 3y_i^2\psi_i)^{1/2})/\psi_i$, for $i = 1, ..., 34$, where ψ_i is the known variance of the municipality i. RBS area models with identity and logarithmic link functions, in short log-RBS, defined in (3.1) are compared to FH models with these same link functions. We use naive model selection tools such as AIC and BIC given in (3.15). Based on the values of AIC and BIC reported in Table 4, note that the RBS area model with logarithmic link function is the best one among the competing models to fit Chilean survey data. Once the RBS area model with logarithmic link function is selected, we estimate its parameters and the SE of the EB estimator using bootstrapping, denoted by $\widehat{\mathrm{SE}}(\widetilde{\theta}_i^{\mathrm{EB}}) = (\widehat{\mathrm{Var}}(\widetilde{\theta}_i^{\mathrm{EB}}))^{1/2}$; see Algorithm 2. Table 5 presents the values for the response variables $(Y_i|\theta_i)$, EB estimates $(\widetilde{\theta}_i^{\mathrm{EB}})$, estimated SE $(\widehat{\mathrm{SE}}(\widetilde{\theta}_i^{\mathrm{EB}}))$ and lower limit (LL) and upper limit (UP) of the 95% bootstrap confidence interval for $\widetilde{\theta}_i^{\mathrm{EB}}$.

Table 4: AIC and BIC values for the listed model and link by municipality ID with CBCI-UQLI data.

Criteria	RBS-log	RBS-identity	Normal-log	Normal-identity
$\ell(\widehat{\lambda})$	−119.807	−129.750	−130.250	−129.750
AIC	247.614	253.601	264.501	267.501
BIC	250.194	256.188	265.079	270.081

The ML estimates of the parameters β_0, β_1 and κ of the model given in (3.1) using a logarithmic link function, with the estimated SEs in parenthesis, are: $\widehat{\beta}_0 = 4.027(0.237)$, $\widehat{\beta}_1 = 0.063(0.006)$ and $\widehat{\kappa} = 163.505(6.401)$. From this information, note that all coefficients are significant at 5% based on the normal approximation of the distribution of the ML estimators.

Table 5: Estimates, SEs and 95% confidence intervals for the area small mean based on the RBS area model with logarithm link function using CBCI and UQLI data.

ID	$\widetilde{\theta}_i^{\mathrm{EB}}$	$\widehat{\mathrm{SE}}(\widetilde{\theta}_i^{\mathrm{EB}})$	LL	UL	ID	$\widetilde{\theta}_i^{\mathrm{EB}}$	$\widehat{\mathrm{SE}}(\widetilde{\theta}_i^{\mathrm{EB}})$	LL	UL
PC	30.59	1.26	28.11	33.06	SJ	39.16	1.19	36.81	41.50
CO	31.38	2.39	26.68	36.08	LC	38.94	0.99	36.99	40.88
QN	32.05	1.90	28.32	35.77	RE	39.46	1.42	36.66	42.25
LE	31.68	0.86	29.99	33.36	CE	41.50	2.18	37.22	45.77
CN	32.17	0.89	30.41	33.92	SM	43.03	1.49	40.09	45.96
LG	33.17	2.64	27.98	38.35	MP	43.32	2.20	39.00	47.63
RN	33.77	3.28	27.32	40.21	SB	43.34	3.85	35.78	50.89
IN	34.74	0.74	33.27	36.20	SA	43.85	0.58	42.70	44.99
EC	35.44	1.40	32.68	38.19	PE	48.68	2.62	43.54	53.81
LP	35.09	0.64	33.83	36.34	LF	48.68	2.77	43.24	54.11
SR	36.21	1.71	32.85	39.56	MA	48.68	0.77	47.15	50.20
QU	37.13	2.48	32.26	41.99	LR	52.50	1.25	50.04	54.95
EB	36.42	1.62	33.24	39.59	NU	52.49	1.71	49.13	55.84
PU	37.00	1.27	34.49	39.50	LB	56.37	1.63	53.16	59.57
PA	37.28	1.26	34.80	39.75	VI	65.71	2.16	61.45	69.96
HU	37.80	1.13	35.57	40.02	PR	71.44	3.17	65.22	77.65
LA	37.59	3.25	31.22	43.96	LN	73.87	2.85	68.27	79.47

4.4. Diagnostics and model checking

Based on Figure 4, we evaluate the assumptions of the RBS area model with logarithm link function by an analysis of the conditional QR residual defined in (3.16) based on Chilean service quality data. This figure shows on the left an index plot of the conditional RQ residual by municipality, whereas on the right, a QQ plot with simulated envelopes for this residual is sketched. Note that outliers are not detected in these figures. In addition, since in the RBS model the variance is a function of its mean, the RBS area model manages well the problem of non-constant variance detected in the EDA. Also, note that the simulated envelopes for the conditional RQ residual verify the distributional assumption for the RBS area model and the absence of outlying observations. Therefore, based on this residual analysis and such as conjectured in our EDA, the RBS area model with logarithm link function is an excellent formulation for describing the Chilean service quality data analyzed in this study.

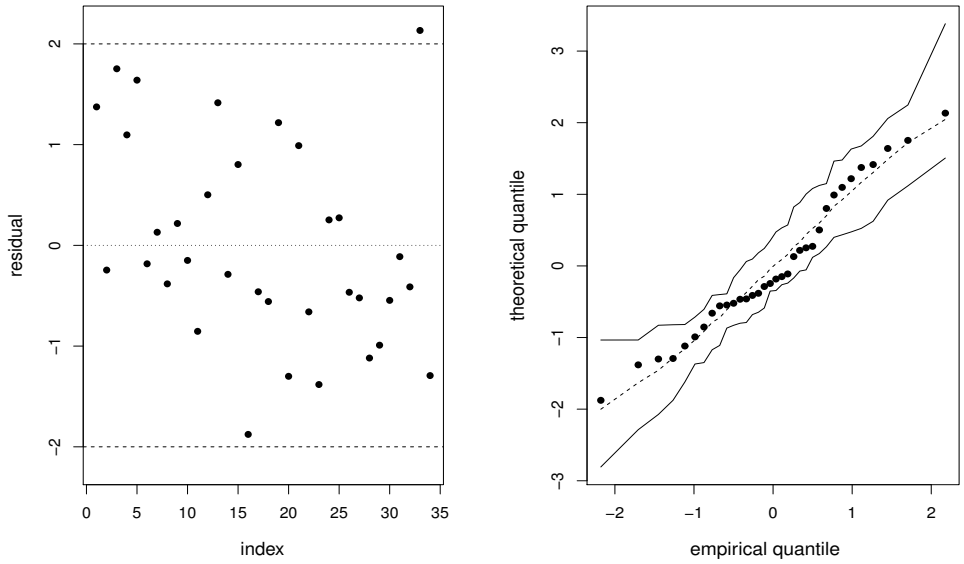


Figure 4: Index plot (left) of the conditional RQ residual and QQ plot with simulated envelopes (right) with CBCI-UQLI data.

5. CONCLUSIONS

The Birnbaum–Saunders area models proposed in this article have properties that are unavailable in the models of this type existing in the literature. Some of these properties are quite needed for describing small areas problems. Specifically, the Birnbaum–Saunders area models considered in this work allow us to describe the mean of the data in their original scale, unlike the existing models, which employ a logarithmic transformation of the data with the consequent problems. In addition, these Birnbaum–Saunders area models can be formulated in a similar form as the normal area models, permitting capturing the essence of the small area estimation problem based on sample means and variances obtained from the areas. Furthermore, the Birnbaum–Saunders area models considered in this study assume a link function, which enables for different structures present in the data. The proposed methodology allowed us to find the estimator of the small area mean based on the empirical Bayes estimator using Gaussian quadrature methods. We also considered a residual to evaluate the model assumptions and atypical data. Finally, we performed a statistical modeling for small area estimation with unpublished Chilean survey data by using the new approach proposed in the article, which have shown the applicability and scope of our proposal. The methodology introduced in this article has been implemented in the `R` software.

ACKNOWLEDGMENTS

The authors thank the editors and reviewers for their constructive comments on an earlier version of this manuscript. This work was supported partially by FONDECYT, grant number 1200525 (V. Leiva, M. Huerta) and fellowship "Becas-Conicyt" (M. Rodríguez) from the National Agency for Research and Development (ANID) of the Chilean government under the Ministry of Science, Technology, Knowledge and Innovation.

REFERENCES

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

[2] Atkinson, A. (1985). *Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Clarendon Press, Oxford, UK.

[3] Avila, J.L.; Huerta, M.; Leiva, V.; Riquelme, M. and Trujillo, L. (2020). The Fay–Herriot model in small area estimation: EM algorithm and application to official data, *REVSTAT*, **18**(5), 613–635.

[4] Balakrishnan, N. and Kundu, D. (2019). Birnbaum–Saunders distribution: A review of models, analysis and applications, *Applied Stochastic Models in Business and Industry*, **35**, 4–49.

[5] Berg, E. and Chandra, H. (2014). Small area prediction for a unit-level lognormal model, *Computational Statistics and Data Analysis*, **78**, 159–175.

[6] Bourguignon, M., Leao, J., Leiva, V., and Santos-Neto, M. (2017). The transmuted Birnbaum–Saunders distribution, *REVSTAT*, **5**, 601–628.

[7] Carter, G. and Rolph, J. (1974). Empirical Bayes methods applied to estimating fire alarm probabilities, *Journal of the American Statistical Association*, **69**, 880–885.

[8] Casas-Cordero, C.; Encina, J. and Lahiri, P. (2016). *Poverty mapping for the Chilean comunas*. In "Analysis of Poverty Data by Small Area Estimation" (M. Pratesi, Ed.), vol. 20, pp. 379–404, Wiley, Chichester, UK.

[9] Coelho, P.S. and Casimiro, F. (2008). Post Enumeration Survey of the 2001 Portuguese population and housing censuses, *REVSTAT*, **6**, 231–252.

[10] Coelho, P.S. and Pereira, L.N. (2011). A spatial unit level model for small area estimation, *REVSTAT*, **9**, 155–180.

[11] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*, Chapman and Hall, London, UK.

[12] Datta, G.S. (2009). *Model-based approach of small area estimation*. In "Handbook of Statistics. Sample Surveys: Inference and Analysis" (D. Pfeffermann and C.R. Rao, Eds.), vol. 29B, pp. 251–288, Elsevier, Oxford, UK.

[13] Datta, G.S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems, *Statistica Sinica*, **10**, 613–627.

[14] Datta, G.S.; Rao, J.N.K. and Smith, D.D. (2005). On measuring the variability of small area estimators under a basic area level model, *Biometrika*, **92**, 183–196.

[15] Dreassi, E.; Petrucci, A. and Rocco, E. (2014). Small area estimation for semicontinuous skewed spatial data: An application to the grape wine production in Tuscany, *Biometrical Journal*, **56**, 141–156.

[16] Dunn, P. and Smyth, G. (1996). Randomized quantile residuals, *Journal of Computational and Graphical Statistics*, **5**, 236–244.

[17] Fabrizi, E.; Ferrante, M.R. and Trivisano, C. (2016). *Bayesian beta regression model for the estimation of poverty and inequality parameters in small area*. In "Analysis of Poverty Data by Small Area Estimation" (M. Pratesi, Ed.), pp. 299–314, Wiley, Chichester, UK.

[18] Fabrizi, E. and Trivisano, C. (2010). Robust linear mixed models for small area estimation, *Journal of Statistical Planning and Inference*, **140**, 433–443.

[19] Fay, R.E. and Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data, *Journal of the American Statistical Association*, **74**, 269–277.

[20] Ferreira, M.; Gomes, M.I. and Leiva, V. (2012). On an extreme value version of the Birnbaum–Saunders distribution, *REVSTAT*, **10**, 181–210.

[21] Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal with discussion, *Statistical Science*, **9**, 55–76.

[22] Huang, S. and Qu, Y. (2006). The loss in power when the test of differential expression is performed under a wrong scale, *Journal of Computational Biology*, **13**, 786–797.

[23] Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation, *TEST*, **15**, 1–96.

[24] Kiaer, A.N. (1895). Observations et expériences concernant les dénombrements représentatifs, *Bulletin of the International Statistical Institute*, **9**, 176–183.

[25] Lehtonen, R. and Veijanen, A. (2009). *Design-based methods of estimation for domains and small areas*. In "Handbook of Statistics. Sample Surveys: Inference and Analysis" (D. Pefeffermann and C.R. Rao, Eds.), vol. 29B, pp. 219–249, Elsevier, Oxford, UK.

[26] Leiva, V. (2016). *The Birnbaum–Saunders Distribution*, Academic Press, New York, US.

[27] Leiva, V.; Lillo, C.; Gomes, M.I. and Ferreira, M. (2019). Discussion of Birnbaum–Saunders distribution: A review of models, analysis, and applications and a novel financial extreme value data analytics from natural disasters, *Applied Stochastic Models in Business and Industry*, **35**, 90–95.

[28] Leiva, V.; Lillo, C. and Morrás, R. (2018). *On a business confidence index and its data analytics: A Chilean case*. In "Recent Studies on Risk Analysis and Statistical Modeling" (T. Oliveira, C. Kitsos, A. Oliveira and L.M. Grilo, Eds.), pp. 67–85, Springer, Switzerland.

[29] Leiva, V.; Santos-Neto, M.; Cysneiros, F.J.A. and Barros, M. (2014). Birnbaum–Saunders statistical modelling: A new approach, *Statistical Modelling*, **14**, 21–48.

[30] Li, H. and Lahiri, P. (2010). An adjusted maximum likelihood method for solving small area estimation problems, *Journal of Multivariate Analysis*, **101**, 882–892.

[31] Lumley, T. and Scott, A. (2017). Fitting regression models to survey data, *Statistical Science*, **32**, 265–278.

[32] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, Chapman and Hall, London, UK.

[33] Mert, Y. (2015). Generalized least squares and weighted least squares estimation methods for distributional parameters, *REVSTAT*, **13**, 263–282.

[34] Nobre, J. and Singer, J. (2007). Residual analysis for linear mixed models, *Biometrical Journal*, **49**, 863–875.

[35] Nocedal, J. and Wright, S. (1999). *Numerical Optimization*, Springer, New York, US.

[36] Pereira, L.N. and Coelho, P.S. (2012). Small area estimation using a spatio-temporal linear mixed model, *REVSTAT*, **10**, 285–308.

[37] Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators, *Journal of the American Statistical Association*, **85**, 163–171.

[38] R Core Team (2016). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

[39] Rao, J.N.K. (2003). *Small Area Estimation*, Wiley, New Yersey, US.

[40] Rao, J.N.K. and Molina, I. (2015). *Small Area Estimation*, Wiley, New Yersey, US.

[41] Rieck, J.R. and Nedelman, J.R. (1991). A log-linear model for the Birnbaum–Saunders distribution, *Technometrics*, **3**, 51–60.

[42] Rousseeuw, P.J.; Croux, C.; Todorov, V.; Ruckstuhl, A.; Salibian-Barrera, M.; Verbeke, T.; Koller, M. and Maechler, M. (2016). *robustbase: Basic robust statistics*, R package version 0.92-6.

[43] Rueda, M.M.; Arcos, A.; Molina, D. and Trujillo, M. (2019). Model-assisted and model-calibrated estimation for class frequencies with ordinal outcomes, *REVSTAT*, **16**(3), 323–348.

[44] Santos-Neto, M.; Cysneiros, F.J.A.; Leiva, V. and Barros, M. (2014). A reparameterized Birnbaum–Saunders distribution and its moments, estimation and applications, *REVSTAT*, **12**, 247–272.

[45] 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.

[46] Särndal, C.E.; Swensson, B. and Wretman, J. (2003). *Model Assisted Survey Sampling*, Springer, New York, US.

[47] Villegas, C.; Paula, G.A. and Leiva, V. (2011). Birnbaum–Saunders mixed models for censored reliability data analysis, *IEEE Transactions on Reliability*, **60**, 748–758.

REVSTAT – Statistical Journal
Volume 19, Number 3, July 2021, 421–442

OPTIMAL B-ROBUST ESTIMATION FOR THE PARAMETERS OF THE MARSHALL–OLKIN EXTENDED BURR XII DISTRIBUTION WITH AN APPLICATION TO PHARMACOKINETICS

Authors: Yeşim Güney
– Department of Statistics, Ankara University, Turkey
ydone@ankara.edu.tr

Şenay Özdemir
– Department of Statistics, Afyon Kocatepe University, Turkey
senayozdemir@aku.edu.tr

Yetkin Tuaç
– Department of Statistics, Ankara University, Turkey
ytuac@ankara.edu.tr

Olcay Arslan
– Department of Statistics, Ankara University, Turkey
oarslan@ankara.edu.tr

Received: September 2018 Revised: May 2019 Accepted: May 2019

Abstract:

- Parameters of Marshall–Olkin Extended Burr XII (MOEBXII) distribution are usually estimated using maximum likelihood (ML) and least squares (LS) estimation methods. However, these estimators are not robust to the outliers which are often encountered in practice. The purpose of this paper is to obtain robust estimators for the parameters of MOEBXII distribution using optimal B-robust estimation method. A simulation study is provided to show the performance of the proposed estimators over ML, LS and robust M estimators.Further, a real data example from a pharmacokinetics study is also given to illustrate the modeling capacity of the MOEBXII distribution when the parameters are properly estimated.

Keywords:

- *least squares estimator; Marshall–Olkin extended Burr XII (MOEBXII) distribution; maximum likelihood estimator; optimal B-robust estimator.*

AMS Subject Classification:

- 62F35, 65C60.

1. INTRODUCTION

Burr [4] introduced a family of continuous distributions that includes twelve types of cumulative distribution functions with different shapes. Since then, Burr XII distribution has attracted attention in many different fields [15, 20, 2, 22, 21, 16, 10]. The Burr distribution has relationship with several distributions and some of them are summarized by Rodriguez [23] and Tadikamalla [26]. Because of it is flexility for modeling data, several generalizations of the Burr XII distribution have been introduced in literature. One of these generalizations is based on the Marshall–Olkin transformation which further improves the flexibility of the Burr XII distribution. Marshall and Olkin [19] introduced a method of obtaining a family of distributions with an additional parameter α. Let $F(x)$ and $\bar{F}(x) = 1 - F(x)$ be the cumulative distribution function (cdf) and the survival function of the baseline distribution, respectively. Then, a Marshall–Olkin (MO) extended distribution can be defined with the following survival function

$$\bar{F}_\alpha(x) = \frac{\alpha \bar{F}(x)}{1 - \bar{\alpha}\bar{F}(x)} \tag{1.1}$$

where $\alpha > 0$ is an additional parameter and $\bar{\alpha} = 1 - \alpha$. When $\alpha = 1$, we get the baseline distribution. Using the transformation given in (1.1) several generalized distributions are defined in the literature. One of these generalizations is the Marshall–Olkin extended Burr type XII (MOEBXII) distribution introduced by Al-Saiari *et al.* [3].

Several researchers have considered parameter estimation of the Burr XII distribution. For instance, Wingo [30, 31] has considered estimating the parameters of the Burr XII distribution using the ML estimation method. Malinowska *et al.* [17] have provided the minimum variance linear unbiased estimators (MVLUE), the best linear invariant estimators (BLIE) and the ML estimators based on n-selected generalized order statistics for the parameters of the Burr XII distribution. Shao [24] has given a complete investigation on the behaviors of the ML estimates based on uncensored and right-censored data. Wang and Cheng [29] have used a robust regression method to estimate the parameters of the Burr XII distribution. Dogru and Arslan [6, 7] have proposed estimators based on the M estimation and the optimal B-robust (OBR) estimation methods to estimate the parameters of Burr XII distribution. However, concerning the MOEBXII distribution a small number of researchers have been considered to estimate the parameters of the MOEBXII distribution in the literature. For example, Al-Saiari *et al.* [3] have used the ML and Bayes estimation methods to estimate the parameters of the MOEBXII distribution. Since ML estimators may be spoiled when there are outliers in the data, robust estimation methods can be used to estimate the parameters of MOEBXII distribution. Recently, Guney and Arslan [9] and Ozdemir *et al.* have explored the robust estimation methods to estimate the parameters of MOEBXII distribution if robustness is a concern. The aim of this paper is twofold. First, alternative to the robust estimation methods used in the paper by Guney and Arslan [9] and Ozdemir *et al.*, we propose to use the OBR estimation method to estimate the parameters of the MOEBXII distribution. By doing this, we gain robustness against to the outliers in the data. The second aim of this study is to use the MOEBXII distribution to model the pharmacokinetics data using the robust estimators which has not been tried before.

Note that, the pharmacokinetics properties of the drug are among the most important drug characteristics for optimal treatment after the selection of the appropriate drug in the

treatment of a disease. The most appropriate daily dose to achieve the effective plasma level is determined by these features. Among these properties one of the most important pharmacokinetics property is plasma drug concentration. The maximum concentration (C_{max}) and the time taken to reach the maximum concentration (T_{max}) are also important variables for the pharmacokinetics studies. These variables can be easily estimated by using the right distribution. However, to obtain the reliable estimates of C_{max} and T_{max}, trustfully modeling of the plasma drug concentration is necessary (For more details see [1]).

The remainder of the paper is organized as follows. In Section 2, we briefly recall the MOEBXII distribution. In Section 3, we first summarize the ML, LS and robust M estimation methods, and then we give the OBR estimators for the parameters of the MOEBXII distribution. Section 4 and Section 5 are dedicated to the simulation study and a real data from pharmacokinetics study to compare the performance of the OBR estimation method with the ML, LS and robust M estimation methods. Finally, conclusions and discussions are given in Section 6.

2. MARSHALL–OLKIN EXTENDED BURR XII DISTRIBUTION

The probability density function (pdf) and the cdf of Burr XII distribution are

$$f(x;c,k) = ck\frac{x^{(c-1)}}{(1+x^c)^{k+1}}, x \geq 0, \tag{2.1}$$

$$F(x;c,k) = 1 - \frac{1}{(1+x^c)^k}, x \geq 0 \tag{2.2}$$

where c and $k > 0$ are the shape parameters. Substituting the cdf of the Burr XII distribution given in (2.2) into the transformation equation given in (1.1) the Marshall–Olkin Extended Burr XII distribution (MOEBXII(α, c, k)) is obtained with the following pdf and cdf, respectively

$$f(x;\alpha,c,k) = \alpha ck\frac{x^{(c-1)}\left(1+x^c\right)^{-(k+1)}}{\left[1-(1-\alpha)\left(1+x^c\right)^{-k}\right]^2}, x \geq 0, \tag{2.3}$$

$$F(x;\alpha,c,k) = \frac{1-\left(1+x^c\right)^{-k}}{1-(1-\alpha)\left(1+x^c\right)^{-k}}, x \geq 0 \tag{2.4}$$

where α, c and $k > 0$ are the shape parameters [3]. When $\alpha = 1$ the Burr XII distribution is recovered with two parameters c and k. The MOEBXII distribution contains distributions with different shapes for the different values of the parameters. For example we get, bell-shaped, right-skewed or L-shaped distributions when we set different values for α, c and k. This makes crucial advantage of flexibility for this distribution to fit data sets with several different shapes. One can see [3] for further details about the MOEBXII distribution.

3. PARAMETER ESTIMATION

In this section, the ML, LS, robust M and the OBR estimation methods to estimate the parameters of the MOEBXII distribution.

3.1. Maximum Likelihood Estimation

Let $x = (x_1, x_2, ..., x_n)$ be a random sample of size n from the MOEBXII(α, c, k) distribution with the unknown parameters α, c and k. The log-likelihood function is

$$l(\alpha, c, k) = n\log(\alpha ck) + (c-1)\sum_{i=1}^{n}\log x_i - (k+1)\sum_{i=1}^{n}\log(1+x_i^c) - 2\sum_{i=1}^{n}\log(1-(1-\alpha)(1+x_i^c)^{-k}). \tag{3.1}$$

Taking the derivatives of this function with respect to α, c and k, we get the following score functions:

$$s_\alpha = \frac{n}{\alpha} - 2\sum_{i=1}^{n}\frac{(1+x_i^c)^{-k}}{1-(1-\alpha)(1+x_i^c)^{-k}}, \tag{3.2}$$

$$s_c = \frac{n}{c} + \sum_{i=1}^{n}\log x_i - (k+1)\sum_{i=1}^{n}\frac{x_i^c\log(x_i)}{1+x_i^c} - 2k(1-\alpha)\sum_{i=1}^{n}\frac{(1+x_i^c)^{-(k+1)}x_i^c\log(x_i)}{1-(1-\alpha)(1+x_i^c)^{-k}}, \tag{3.3}$$

$$s_k = \frac{n}{k} - \sum_{i=1}^{n}\log(1+x_i^c) - 2(1-\alpha)\sum_{i=1}^{n}\frac{(1+x_i^c)^{-k}\log(1+x_i^c)}{1-(1-\alpha)(1+x_i^c)^{-k}}. \tag{3.4}$$

The ML estimators of the parameters can be obtained by setting the score functions to zero and solving them simultaneously with respect to α, c and k. Since the likelihood equations $(s_\alpha = 0, s_c = 0, s_k = 0)$ cannot be solved analytically, we need to use some numeric methods to obtain the estimates of the parameters.

3.2. Least Squares Estimation

LS estimation method was used to estimate the parameters of the Burr distribution [13] and the MOEBXII distribution [9]. The LS estimation method to estimate the parameters of the MOEBXII distribution can be summarized as follows. It is basically based on minimizing

the following function:

$$\begin{aligned} S(\alpha, c, k) &= \sum_{i=1}^{n} \left(\widehat{F}(x_i) - F(x_i)\right)^2 \\ &= \sum_{i=1}^{n} \left(\widehat{F}(x_i) - \frac{1-(1+x_i^c)^{-k}}{1-(1-\alpha)(1+x_i^c)^{-k}}\right)^2. \end{aligned} \tag{3.5}$$

Since the cdf of the MOEBXII distribution is a non-linear function, the minimization of equation (3.5) is not easy to obtain. To handle this problem, $\log\left(\frac{1}{1-F(x)}\right)$ transformation can be used.

Let $y_{(i)} = \log\left(\frac{1}{1-\widehat{F}(x_{(i)})}\right)$ and $u_{(i)} = \log\left(\frac{1}{1-F(x_{(i)})}\right)$ with

$$\widehat{F}\left(x_{(i)}\right) = \frac{i-0.5}{n}, i = 1, 2, ..., n. \tag{3.6}$$

Here $x_{(i)}$ denotes the $i.$ order statistics of the sample from the MOEBXII distribution. Thus, the LS estimates of the parameters can be obtained by minimizing the following objective function:

$$S(\alpha, c, k) = \sum_{i=1}^{n} \left(y_{(i)} - u_{(i)}\right)^2. \tag{3.7}$$

To obtain the LS estimates, the following equations should be solved with respect to α, c and k:

$$\sum_{i=1}^{n} \left(y_{(i)} - u_{(i)}\right) \frac{1-\left(1+x_{(i)}^c\right)^{-k}}{\alpha\left[1-(1-\alpha)\left(1+x_{(i)}^c\right)^{-k}\right]} = 0, \tag{3.8}$$

$$\sum_{i=1}^{n} \left(y_{(i)} - u_{(i)}\right) \frac{kx_{(i)}^c \log\left(x_{(i)}\right)}{\left(1+x_{(i)}^c\right)\left[1-(1-\alpha)\left(1+x_{(i)}^c\right)^{-k}\right]} = 0, \tag{3.9}$$

$$\sum_{i=1}^{n} \left(y_{(i)} - u_{(i)}\right) \frac{\log\left(1+x_{(i)}^c\right)}{\left[1-(1-\alpha)\left(1+x_{(i)}^c\right)^{-k}\right]} = 0. \tag{3.10}$$

3.3. M Estimation

Guney and Arslan [9] have been proposed to estimate the parameters of the MOEBXII distribution using M estimation method ([14]). The method is based on minimizing the following objective function with respect to the parameters of interest:

$$Q(\alpha, c, k) = \sum_{i=1}^{n} \rho\left(y_i - u_i\right). \tag{3.11}$$

Here ρ is more resistant than the square function in LS method to the outliers in data set. It is also non-negative, symmetric function and $\rho(0) = 0$. In this study, we consider the Tukey's ρ function given as

$$\rho(x) = \begin{cases} 1 - (1 - (x/b)^2)^3 \ , & |x| \leq b, \\ 1 \qquad\qquad\qquad\quad , & |x| > b, \end{cases} \tag{3.12}$$

$$\rho'(x) = \Psi(x) = \begin{cases} x(1 - (x/b)^2)^2 \ , & |x| \leq b, \\ 0 \qquad\qquad\qquad\quad , & |x| > b, \end{cases} \tag{3.13}$$

with the robustness tunning constant b (see Maronna *et al.*, [18], pp. 29). Here the tuning constant b determines if an observation is an outlier or not. Tukey's biweight function truncates the residuals that are larger than b. Therefore, small values of b imply higher robustness while large values of b provide higher efficiency. In literature the suggested choice of b is 4.685 to achieve 95% asymptotic efficiency at the standard normal distribution [18].

Since ρ is differentiable, M estimates can be obtained by solving the following non-linear equations based on the derivatives of objective function (3.11):

$$\log \widehat{\alpha} = \frac{\sum_{i=1}^{n} \omega_i \left(y_i - k \log\left(1 + x_i^c\right) - \log h_i\right) \left(\frac{1-\left(1+x_i^c\right)^{-k}}{\alpha h_i}\right)}{\sum_{i=1}^{n} \omega_i \left(\frac{1-\left(1+x_i^c\right)^{-k}}{\alpha h_i}\right)}, \tag{3.14}$$

$$\widehat{k} = \frac{\sum_{i=1}^{n} \omega_i \left(y_i + \log(\alpha) - \log h_i\right) \frac{\log(1+x_i^c)}{h_i}}{\sum_{i=1}^{n} \omega_i \frac{\left(\log(1+x_i^c)\right)^2}{h_i}}, \tag{3.15}$$

$$\sum_{i=1}^{n} \omega_i \left(y_i - u_i\right) \frac{x_i^c \log\left(x_i\right) \left(1 - (1 + x_i^c)^{-k}\right)}{(1 + x_i^c)^{-k} h_i} = 0, \tag{3.16}$$

where $h_i = 1 - (1 - \alpha)\,(1 + x_i^c)^{-k}$ and the weights are

$$\omega_i = \left(1 - \left(\frac{y_i - u_i}{b}\right)^2\right)^2 I(|y_i - u_i| \leq b)\,. \tag{3.17}$$

3.4. Optimal B-Robust Estimation

The class of the OBR estimators was defined by Hampel *et al.* [11]. The OBR estimation method is a robust alternative modification of M estimation method with bounded influence function. It is also the most efficient one in the class of robust M-estimators. In literature, Victoria-Feser [27] and Victoria-Feser and Ronchetti [28] introduced the OBR estimation method to estimate the parameters of the Pareto and the gamma distributions. Dogru and Arslan [7] introduced the OBR estimation method for the Burr XII distribution. Dogru and Arslan [8] also proposed robust estimators by using the OBR estimation method for the parameters of the generalized half-normal distribution.

According to Hampel *et al.* [11], there are two ways of defining the optimal B-robust estimation. The first one is the minimax approach defined by Huber [14]. The second one is called the infinitesimal approach introduced by Hampel *et al.* [11]. In this paper, we will use the second approach that aims to find M-estimators with bounded influence function (IF) and minimum asymptotic variance.

IF can be defined as follows. For a sample of n observations, $\underline{x} = (x_1, x_2, ..., x_n)$, the empirical distribution function $F_n(x)$ is

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}(x) \tag{3.18}$$

where δ_{x_i} denotes a point mass in x. For a parametric model $\{F_\theta : \theta \in \Theta \subset R^p\}$, estimator of θ; T_n can be represented as a statistical functional of the empirical distribution, i.e. $T_n(x_1, x_2, ..., x_n) = T_n(F_n)$. In our case $\theta = (\alpha, c, k)$. Then, the IF of T_n is given by

$$IF(x, T_n, F_\theta) = \lim_{\epsilon \to 0} \frac{T_n((1-\epsilon)F_\theta + \epsilon\delta_x) - T_n(F_\theta)}{\epsilon}. \tag{3.19}$$

The IF describes the relative influence of individual observations toward the value of an estimate [11]. When the IF is unbounded, an outlier can have an overriding influence on the estimate. The IF of the ML estimator is

$$IF = J(\theta)^{-1} s(x, \theta) \tag{3.20}$$

where $J(\theta)$ is the Fisher information matrix and $s(x, \theta) = \frac{\partial}{\partial \theta} \log f(x, \theta)$ is the vector of score functions. It is clear that the IF of the ML estimator will not be bounded if the score function is not bounded.

Concerning the score functions for the MOEBXII distribution given in (3.2)–(3.4), one can easily observe that the score function for α is bounded but the score functions for c and k are unbounded functions of x as in the Burr XII distribution. That is, we have $\lim_{x \to \infty} s_c = -\infty$ and $\lim_{x \to \infty} s_k = -\infty$. These unboundedness of score functions for the parameter c and k can also be easily observed in Figure 1.

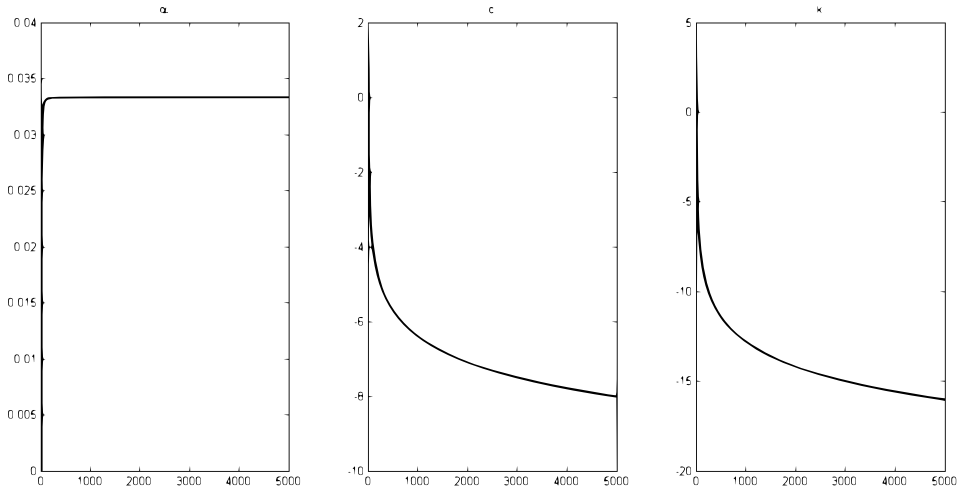


Figure 1: Plots of the score functions with $(\alpha, c, k) = (30, 2, 1)$.

If α, c and k are estimated by using the ML and LS estimation methods, these estimators may suffer from possible outliers. Therefore, instead of using the ML and LS methods we will propose to use the OBR estimation method in the presence of outliers.

Consider the following standardized OBR estimating equation

$$\sum_{i=1}^{n} \Psi_b(A(\theta)(s(\theta, x_i) - a(\theta))) = \sum_{i=1}^{n} W(\theta, x_i, c_B)(s(\theta, x_i) - a(\theta)) = 0 \tag{3.21}$$

where

$$W(\theta, x_i, c_B) = \min\left(1, \frac{c_B}{\|A(\theta)(s(\theta, x_i) - a(\theta))\|}\right), \tag{3.22}$$

Ψ_b is the derivative of ρ_b, is $c_B \geq \sqrt{\dim(\theta)}$ is a tuning parameter, $\|\cdot\|$ denoted the Euclidean norm, $s(\cdot)$ is the score function, $A(\theta)$ is a $\dim(\theta) \times \dim(\theta)$ scaling matrix and $a(\theta)$ is a $\dim(\theta)$ centering vector determined by

$$E\left[\Psi_b(x)\Psi_b(x)^T\right] = \left[A(\theta)^T A(\theta)\right]^{-1}, \tag{3.23}$$

$$E\left[\Psi_b(s(\theta, x) - a(\theta))\right] = 0. \tag{3.24}$$

The OBR estimates for the parameter θ will be the solution of this equation. The OBR estimator keeps a level of efficiency close to the ML estimator because of the score function. The constant c_B, robustness constant, is typically fixed by setting the amount of efficiency loss and a bound on the IF. For higher values of c_B the estimator gains efficiency, but lose robustness and vice versa. If the bound on the IF is removed, i.e, choose $c_B = \infty$ the OBR estimation method reduces to the ML estimation method. To compute the OBR estimates of the parameters, we follow an algorithm proposed by Victoria-Feser and Ronchetti [28].

OBRE Algorithm:

1. Fix the precision threshold η and the initial value for $\theta_{(0)}$ (we can take the ML estimates as the initial values).

Take initial values $\mathbf{a} = \mathbf{0}$, and $A = \left(\left[J^{-1}\right]^T\right)^{1/2}$ where

$$J = \int s(\theta, x)s(\theta, x)^T dF_\theta(x) \tag{3.25}$$

is the Fisher Information Matrix.

2. Solve the following equations with respect to $\mathbf{a}$ and A

$$A^T A = M_2^{-1} \tag{3.26}$$

$$\mathbf{a} = \frac{\int W(\theta, x, c_B)s(\theta, x)dF_\theta(x)}{\int W(\theta, x, c_B)dF_\theta(x)} \tag{3.27}$$

where

$$M_k = \int W(\theta, x, c_B)^k \left[s(\theta, x) - \mathbf{a}(\theta)\right]\left[s(\theta, x) - \mathbf{a}(\theta)\right]^T dF_\theta(x), \quad k = 1, 2. \tag{3.28}$$

The current values of θ, $\mathbf{a}$ and A are used as initial values to solve the given equations.

3. Now compute M_1 and

$$\Delta\theta = M_1^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} W(\theta, x_i, c_B)^k \left[s(\theta, x_i) - a(\theta)\right]\right). \tag{3.29}$$

4. If $\|\Delta\theta\| > \nu$ then $\theta \to \theta + \Delta\theta$ and return to step 2, otherwise terminate the algorithm.

Victoria-Feser and Ronchetti [27] mentioned that: "The algorithm is convergent provided the starting point is near to the solution" in their study. Therefore, we used different initial points for the first step of the algorithm. Then we observed that there are no significant differences between the estimates according to the different initial points. In this study, the ML estimates are used as an initial points.

4. SIMULATION STUDY

A Monte Carlo simulation study was conducted based on various scenarios for the number of observations and outliers to examine the performance of the estimation methods; the ML, LS, robust M estimation with Tukey and the OBR estimation methods. The superiority of the estimates was assessed by using the performance measures, bias and Root-mean-square error (RMSE) defined as

$$\text{Bias}\left(\hat{\theta}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_i - \theta\right), \tag{4.1}$$

$$\text{RMSE}\left(\hat{\theta}\right) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_i - \theta\right)^2}. \tag{4.2}$$

We generated $N = 100$ replications from the MOEBXII distribution with the sample sizes $n = 25$, $n = 50$ and $n = 100$. We consider the following parameter values $(\alpha, c, k) = (3, 1, 1)$, $(3, 1, 2)$, $(3, 2, 1)$, $(3, 2, 2)$, $(3, 3, 3)$, $(5, 1, 1)$, $(5, 1, 2)$, $(5, 2, 1)$ and $(5, 2, 2)$. (One can find the details for generating data set from the MOEBXII distribution in [9]). In this study, the outliers are generated by multiplying the largest observations in the data by 5.

To obtain the M estimations in the simulation study, we determine the tuning constant $b = 4.685$ for Tukey's ρ function. For the OBR estimation method, robustness parameter c_B and precision threshold ν were taken as 3 and 10^{-6} respectively.

The simulation results in all cases are summarized in Tables 1–8. In these tables, the bias and RMSE values calculated by using the equations (4.1)–(4.2) are reported for the ML, LS, M estimation with Tukey's ρ function and the OBR methods.

Tables 1–3 present the results from the case without outlier. From these tables, we can observe that the OBR estimation method has superiority in terms of bias and RMSE in all simulation scenarios for small sample sizes. For moderate sample size we can still observe the better performance of the OBR estimators in most of the cases. However, when sample size increases, the superiority of the ML estimation method in terms of RMSE can be easily observed form Table 3, which is an expected performance of the ML estimation method.

Table 1: The bias and RMSE (Parenthesis) for $n = 25$.

Parameter α	ML	LS	M (Tukey)	OBR
(3,1,1)	0.1081 (0.2221)	−0.4797 (0.6033)	−0.0864 (0.0623)	−0.0845 **(0.0464)**
(3,1,2)	0.0118 (0.2294)	0.3000 (0.3205)	−0.0938 (0.1003)	0.0069 **(0.0165)**
(3,2,1)	0.1761 (0.2195)	−0.0871 (0.4903)	−0.0431 (0.2034)	−0.0011 **(0.0032)**
(3,2,2)	0.1778 (0.2142)	0.1685 (0.2696)	0.1198 (0.1808)	0.0100 **(0.0036)**
(3,3,3)	0.0525 (0.1987)	0.1937 (0.1947)	−0.0778 (0.1537)	0.0101 **(0.0039)**
(5,1,1)	0.0328 (0.2173)	−0.0641 (0.6265)	−0.0990 (0.2846)	−0.0495 **(0.1965)**
(5,1,2)	−0.0668 (0.1981)	−0.4701 (0.1622)	−0.4683 (0.5792)	−0.0102 **(0.1415)**
(5,2,1)	−0.0606 (0.2083)	−0.4757 (0.5291)	0.0095 (0.1544)	0.0088 **(0.0039)**
(5,2,2)	−0.0269 (0.2111)	−0.0653 (0.7695)	−0.4989 (0.6140)	0.0087 **(0.0193)**

Parameter c	ML	LS	M (Tukey)	OBR
(3,1,1)	0.1229 (0.1026)	0.3267 (0.2359)	0.3239 (0.2059)	0.0690 **(0.0841)**
(3,1,2)	0.1073 (0.0555)	0.2684 (0.1663)	0.2711 (0.1606)	−0.0005 **(0.0150)**
(3,2,1)	0.1180 (0.1370)	0.4928 (0.4435)	0.4513 (0.3848)	−0.0012 **(0.0039)**
(3,2,2)	0.0801 (0.1123)	0.3417 (0.3475)	0.3760 (0.3220)	−0.0011 **(0.0001)**
(3,3,3)	0.0198 (0.1421)	0.3733 (0.4870)	0.3993 (0.4264)	−0.0034 **(0.0003)**
(5,1,1)	0.0645 (0.0888)	0.3310 (0.3067)	0.4119 (0.3607)	0.1541 **(0.0271)**
(5,1,2)	0.0947 (0.0531)	0.3728 (0.2557)	0.2709 (0.1452)	0.0010 **(0.0007)**
(5,2,1)	0.0333 (0.1558)	0.2107 (0.5419)	0.1893 (0.3722)	−0.0028 **(0.0002)**
(5,2,2)	0.1550 (0.1154)	0.1436 (0.5794)	0.4771 (0.3996)	−0.0016 **(0.0004)**

Parameter k	ML	LS	M (Tukey)	OBR
(3,1,1)	−0.0403 (0.0739)	0.4731 (0.3267)	−0.1742 (0.1155)	−0.0320 **(0.0635)**
(3,1,2)	−0.0305 (0.1011)	−0.3271 (0.4601)	−0.1788 (0.1281)	0.0026 **(0.0020)**
(3,2,1)	−0.0068 (0.0497)	0.2188 (0.3604)	−0.1080 (0.1292)	0.0006 **(0.0023)**
(3,2,2)	0.0312 (0.0941)	−0.2834 (0.3071)	−0.2902 (0.2239)	0.0025 **(0.0037)**
(3,3,3)	−0.0025 (0.1321)	−0.1886 (0.2640)	−0.0999 (0.2113)	0.0045 **(0.0709)**
(5,1,1)	−0.2456 (0.1743)	−0.3374 (0.4375)	−0.3855 (0.3544)	0.0067 **(0.0570)**
(5,1,2)	−0.0290 (0.0966)	−0.9483 (0.9097)	−0.9361 (0.8912)	−0.0017 **(0.0078)**
(5,2,1)	0.0013 (0.0479)	−0.5580 (0.3312)	−0.5309 (0.2920)	0.0018 **(0.0012)**
(5,2,2)	−0.0521 (0.0893)	−0.0616 (0.1308)	−0.0527 (0.1543)	0.0020 **(0.0009)**

It is obvious from Table 1 that the OBR estimation method has the best performance in terms of RMSE for all parameters for the small sample size ($n = 25$). The biases of the OBR estimates are lower than that of other methods for most of the values of the parameters. Table 2 shows that the OBR and the ML estimation methods are compatible according to the RMSE values under the assumption of moderate sample sizes ($n = 50$). According to the results given in Table 3, as the sample size increases, the ML estimation method seems superior to the other methods in terms of RMSE in most of the cases as expected.

Table 2: The bias and RMSE (Parenthesis) for $n = 50$.

Parameter α	ML	LS	M (Tukey)	OBR
(3,1,1)	0.0988 (0.2140)	−0.3154 (0.1141)	−0.0477 (0.0240)	−0.0127 **(0.0047)**
(3,1,2)	−0.0183 (0.2215)	−0.3803 (0.1631)	−0.1934 (0.0724)	−0.0066 **(0.0628)**
(3,2,1)	0.0435 **(0.0188)**	−0.2936 (0.1646)	−0.0262 (0.0495)	−0.0442 (0.0950)
(3,2,2)	−0.0083 (0.2069)	0.0970 (0.1246)	−0.1616 (0.0687)	0.0156 **(0.0666)**
(3,3,3)	−0.0537 **(0.0210)**	−0.0713 (0.1523)	−0.0840 (0.0774)	0.1114 (0.3523)
(5,1,1)	−0.0260 (0.2289)	−0.4694 (0.2311)	−0.4008 (0.2362)	−0.0178 **(0.0745)**
(5,1,2)	−0.0579 (0.2139)	−0.1342 (0.2093)	−0.4957 (0.2457)	−0.0263 **(0.0214)**
(5,2,1)	−0.0264 (0.2139)	0.0147 (0.2217)	−0.4142 (0.2188)	−0.0259 **(0.0013)**
(5,2,2)	−0.0481 (0.1915)	0.0584 (0.2048)	−0.4999 (0.2500)	−0.0331 **(0.0727)**

Parameter c	ML	LS	M (Tukey)	OBR
(3,1,1)	0.0466 (0.0479)	0.2047 (0.0652)	0.1930 (0.0645)	0.0015 **(0.0184)**
(3,1,2)	0.0267 (0.0423)	0.1867 (0.0529)	0.1461 (0.0529)	0.0011 **(0.0096)**
(3,2,1)	0.0477 (0.0710)	0.3217 (0.1453)	0.2983 (0.1346)	0.0054 **(0.0512)**
(3,2,2)	0.0501 (0.0359)	0.3102 (0.1353)	0.2588 (0.1226)	−0.0024 **(0.0021)**
(3,3,3)	0.0307 (0.0582)	0.3692 (0.1865)	0.3227 (0.1811)	−0.0287 **(0.0215)**
(5,1,1)	0.0505 **(0.0342)**	0.1857 (0.0528)	0.2200 (0.0804)	0.0016 (0.0469)
(5,1,2)	0.0227 **(0.0146)**	0.1026 (0.0547)	0.3426 (0.1504)	0.0012 (0.0648)
(5,2,1)	0.0592 (0.0816)	0.1588 (0.1661)	0.3253 (0.1614)	0.0043 **(0.0387)**
(5,2,2)	0.0613 (0.0893)	0.1208 (0.0954)	0.4289 (0.2160)	0.0024 **(0.0342)**

Parameter k	ML	LS	M (Tukey)	OBR
(3,1,1)	−0.0218 (0.0198)	−0.0982 (0.0281)	−0.2219 (0.0605)	−0.0026 **(0.0179)**
(3,1,2)	0.0469 (0.0505)	0.1787 (0.0472)	−0.1068 (0.0315)	−0.0021 **(0.0059)**
(3,2,1)	0.0032 (0.0659)	−0.0263 (0.0565)	−0.0805 **(0.0558)**	−0.0049 (0.0613)
(3,2,2)	0.0104 **(0.0461)**	−0.2862 (0.1539)	−0.1492 (0.0642)	0.0069 (0.0086)
(3,3,3)	−0.0239 (0.0933)	−0.1003 (0.1198)	−0.2754 (0.1172)	0.0456 **(0.0650)**
(5,1,1)	−0.0299 (0.0239)	−0.4980 (0.2481)	−0.4967 (0.2468)	−0.0023 **(0.0012)**
(5,1,2)	−0.0134 **(0.0330)**	−0.0235 (0.0946)	−0.4470 (0.2499)	−0.0049 (0.0771)
(5,2,1)	−0.0006 **(0.0210)**	0.0010 (0.0487)	−0.4959 (0.2462)	−0.0033 (0.0246)
(5,2,2)	−0.0033 **(0.0242)**	0.0197 (0.0909)	−0.4989 (0.2489)	−0.0051 (0.0279)

Table 3: The bias and RMSE (Parenthesis) for $n = 100$.

Parameter α	ML	LS	M (Tukey)	OBR
(3,1,1)	0.0097 **(0.0355)**	−0.2899 (0.1072)	−0.0607 (0.0397)	−0.1376 (0.2175)
(3,1,2)	−0.0191 **(0.0368)**	−0.3663 (0.1805)	−0.2242 (0.0980)	0.0879 (0.2121)
(3,2,1)	0.0325 **(0.0347)**	−0.3331 (0.1830)	−0.0721 (0.0711)	0.0540 (0.1974)
(3,2,2)	0.0249 **(0.0365)**	0.0373 (0.1226)	−0.1339 (0.0878)	−0.0457 (0.2264)
(3,3,3)	−0.0006 **(0.0328)**	−0.1143 (0.1465)	−0.1815 (0.1187)	−0.0452 (0.2125)
(5,1,1)	−0.0487 **(0.0357)**	−0.4458 (0.2158)	−0.3646 (0.2165)	−0.0300 (0.2173)
(5,1,2)	−0.0168 **(0.0349)**	−0.1446 (0.1883)	−0.4709 (0.2435)	−0.0504 (0.2126)
(5,2,1)	0.0293 **(0.0354)**	0.0118 (0.2107)	−0.3618 (0.2227)	0.0597 (0.2211)
(5,2,2)	0.0137 **(0.0368)**	−0.0741 (0.2187)	−0.4960 (0.2468)	0.0701 (0.1997)

Parameter c	ML	LS	M (Tukey)	OBR
(3,1,1)	0.0315 **(0.0164)**	0.2066 (0.0980)	0.2044 (0.0986)	0.0557 (0.0492)
(3,1,2)	0.0254 **(0.0134)**	0.1471 (0.0566)	0.1561 (0.0598)	0.0115 (0.0177)
(3,2,1)	−0.0001 **(0.0282)**	0.2233 (0.1595)	0.2116 (0.1534)	0.0666 (0.1013)
(3,2,2)	0.0034 **(0.0203)**	0.3162 (0.1652)	0.3160 (0.1594)	0.1252 (0.0733)
(3,3,3)	0.0232 **(0.0288)**	0.3041 (0.1628)	0.2714 (0.1573)	0.1005 (0.1030)
(5,1,1)	0.2304 (0.0941)	0.1675 (0.0636)	0.0613 (0.0446)	0.0152 **(0.0211)**
(5,1,2)	0.0318 **(0.0154)**	0.0833 (0.0476)	0.3125 (0.1322)	0.0323 (0.0344)
(5,2,1)	0.3095 **(0.1688)**	0.0565 (0.1553)	0.0231 (0.0306)	0.1173 (0.1171)
(5,2,2)	0.0069 **(0.0290)**	0.0375 (0.0949)	0.3916 (0.1986)	0.0320 (0.0849)

Parameter k	ML	LS	M (Tukey)	OBR
(3,1,1)	0.0071 (0.0218)	−0.0794 (0.0507)	−0.2181 (0.0733)	−0.0462 **(0.0118)**
(3,1,2)	−0.0088 (0.0514)	0.1005 (0.0536)	−0.1030 (0.0447)	−0.0233 **(0.0287)**
(3,2,1)	0.0049 **(0.0128)**	0.0109 (0.0759)	−0.1993 (0.0722)	−0.0009 (0.0382)
(3,2,2)	−0.2288 (0.0843)	−0.2906 (0.1341)	−0.0071 **(0.0223)**	−0.0313 (0.0732)
(3,3,3)	−0.0237 **(0.0302)**	−0.1285 (0.1235)	−0.1397 (0.1128)	−0.0007 (0.0958)
(5,1,1)	0.0130 **(0.0226)**	−0.4924 (0.2428)	−0.4952 (0.2455)	−0.0218 (0.0363)
(5,1,2)	−0.0045 **(0.0230)**	−0.0171 (0.1301)	−0.4933 (0.2456)	−0.0490 (0.0512)
(5,2,1)	0.0153 **(0.0106)**	0.0231 (0.0478)	−0.4908 (0.2415)	−0.0143 (0.0306)
(5,2,2)	0.0152 **(0.0215)**	−0.0074 (0.0779)	−0.4995 (0.2495)	0.0207 (0.0566)

We recreated the simulation for the same scenarios with outliers and the results are summarized in Tables 4–7. We generate one outlier to see the performance of the estimators in case there is an outlier in the data, for all the sample sizes. Further, to see the behavior of the estimators under the condition that there are more than one outlier, we conduct an additional simulation which we use four outliers in sample size 50. It is already mentioned that the four outliers are generated by multiplying the four largest observation with 5.

Table 4 shows the simulation results for the sample size $n = 25$ with one outlier. We observe that outlier induces a large influence on the bias and RMSE of the ML and the LS estimators whereas it has a smaller impact on the robust estimators. If the M and the OBR estimation methods are compared with each other, the OBR estimation method is superior to the M estimation method in terms of the RMSE.

Table 5 shows the simulation results with one outlier with the sample size 50. When the data include outlier, the ML and the LS estimators are drastically worsen which is reflected to the higher RMSE and biases. However, the M and the OBR estimators still have better performance.

Table 4: The bias and RMSE (Parenthesis) for $n = 25$ with one outlier.

Parameter α	ML	LS	M (Tukey)	OBR
(3,1,1)	0.5260 (0.7583)	0.3937 (0.6005)	0.0648 (0.0974)	0.0675 **(0.0334)**
(3,1,2)	0.7677 (1.0733)	−0.2801 (0.9506)	0.2737 (0.3186)	0.1322 **(0.1359)**
(3,2,1)	0.6696 (0.8518)	0.7122 (0.5237)	0.0598 (0.0664)	0.0693 **(0.0384)**
(3,2,2)	0.9541 (0.9890)	−0.3073 (0.2022)	0.2807 (0.5148)	0.1792 **(0.1211)**
(3,3,3)	0.3557 (0.1608)	0.8745 (0.8261)	0.1740 (0.2903)	0.0327 **(0.0782)**
(5,1,1)	0.7769 (0.8408)	0.6287 (0.5721)	0.3778 (0.2233)	0.1630 **(0.0783)**
(5,1,2)	0.8666 (0.9375)	0.7299 (0.7325)	0.4934 (0.2456)	0.0133 **(0.0215)**
(5,2,1)	0.6980 (0.9422)	0.5848 (0.5159)	0.3630 (0.2123)	−0.0127 **(0.0257)**
(5,2,2)	0.9814 (0.9669)	0.8705 (0.8508)	0.4707 (0.7286)	0.2630 **(0.2471)**

Parameter c	ML	LS	M (Tukey)	OBR
(3,1,1)	−0.2184 (0.1522)	−0.3247 (0.2277)	−0.2399 (0.1012)	−0.0170 **(0.0016)**
(3,1,2)	−0.1275 (0.0787)	−0.3053 (0.2022)	−0.2481 (0.1208)	−0.0298 **(0.0158)**
(3,2,1)	−0.2308 (0.3010)	−0.4433 (0.4108)	−0.2805 (0.1529)	−0.0303 **(0.0039)**
(3,2,2)	0.0665 (0.1365)	−0.2382 (0.2869)	−0.1332 (0.1106)	−0.0491 **(0.0328)**
(3,3,3)	0.4780 (0.3820)	−0.2563 (0.3921)	−0.4035 (0.3110)	−0.1655 **(0.1587)**
(5,1,1)	−0.2185 (0.2043)	−0.4727 (0.3573)	−0.2399 (0.1081)	−0.0214 **(0.0015)**
(5,1,2)	−0.0734 (0.1097)	−0.3024 (0.1942)	−0.3290 (0.1483)	0.0227 **(0.0528)**
(5,2,1)	−0.2499 (0.4091)	−0.5456 (0.5562)	−0.1779 (0.1339)	−0.0251 **(0.0810)**
(5,2,2)	−0.0615 (0.2139)	−0.5764 (0.5222)	−0.3317 (0.1961)	−0.0728 **(0.0270)**

Parameter k	ML	LS	M (Tukey)	OBR
(3,1,1)	0.2541 (0.1131)	−0.3756 (0.3442)	0.3033 (0.1241)	0.0213 **(0.0048)**
(3,1,2)	0.4746 (0.3717)	0.4407 (0.5491)	0.2398 (0.1200)	0.0637 **(0.0899)**
(3,2,1)	0.2620 (0.1153)	−0.5760 (0.4492)	0.2638 (0.1155)	0.0170 **(0.0019)**
(3,2,2)	0.5505 (0.3801)	0.5809 (0.3841)	0.2605 (0.1061)	0.0583 **(0.0551)**
(3,3,3)	0.9466 (0.9067)	0.7017 (0.5734)	0.4727 (0.4533)	0.2293 **(0.1271)**
(5,1,1)	0.2190 (0.1047)	0.6239 (0.4033)	0.4945 (0.2448)	0.0280 **(0.0026)**
(5,1,2)	0.2445 (0.3798)	0.9613 (0.9299)	0.4976 (0.9480)	0.2095 **(0.2479)**
(5,2,1)	0.1730 (0.1975)	0.5925 (0.3678)	0.4706 (0.2271)	−0.0004 **(0.0538)**
(5,2,2)	0.4730 (0.3290)	0.9820 (0.9687)	0.4554 (0.2455)	0.0791 **(0.0385)**

Table 5: The bias and RMSE (Parenthesis) for $n = 50$ with one outlier.

Parameter α	ML	LS	M (Tukey)	OBR
(3,1,1)	0.7658 (0.7925)	0.2757 (0.5942)	0.0844 (0.0336)	0.0972 **(0.0192)**
(3,1,2)	0.9122 (0.9377)	−0.2170 (0.1263)	0.2961 (0.1128)	0.0691 **(0.0969)**
(3,2,1)	0.8448 (0.8977)	0.7232 (0.5303)	0.0696 (0.0452)	0.0194 **(0.0203)**
(3,2,2)	0.9860 (0.9818)	−0.2744 (0.1613)	0.1150 (0.1290)	0.3149 **(0.0935)**
(3,3,3)	0.9464 (0.9216)	0.2705 (0.2294)	0.3848 (0.1637)	−0.0039 **(0.0482)**
(5,1,1)	0.5430 (0.8791)	0.6820 (0.5746)	0.3746 (0.2256)	0.1248 **(0.0341)**
(5,1,2)	0.8018 (0.9470)	0.9043 (0.8835)	0.4865 (0.2431)	0.1075 **(0.2400)**
(5,2,1)	0.8957 (0.9413)	0.6427 (0.5411)	0.4286 (0.2280)	0.0585 **(0.1521)**
(5,2,2)	0.9518 (0.9853)	0.9194 (0.8797)	0.4780 (0.2481)	0.1677 **(0.2322)**

Parameter c	ML	LS	M (Tukey)	OBR
(3,1,1)	−0.1187 (0.1086)	−0.2660 (0.1783)	−0.1984 (0.0917)	−0.0161 **(0.0004)**
(3,1,2)	−0.0095 (0.0199)	−0.1356 (0.0502)	−0.1181 (0.0473)	−0.0036 **(0.0008)**
(3,2,1)	−0.1724 (0.2018)	−0.2910 (0.2963)	−0.1612 (0.1244)	−0.0281 **(0.0024)**
(3,2,2)	−0.1955 (0.1236)	−0.2882 (0.2294)	0.0101 (0.0581)	−0.0290 **(0.0045)**
(3,3,3)	0.5491 (0.3680)	−0.2338 (0.2683)	−0.0770 (0.1276)	−0.0786 **(0.0182)**
(5,1,1)	−0.1099 (0.0960)	−0.3937 (0.2707)	−0.2269 (0.1000)	−0.0159 **(0.0004)**
(5,1,2)	−0.0081 (0.0340)	−0.3363 (0.1934)	−0.3118 (0.1338)	−0.0102 **(0.0010)**
(5,2,1)	−0.1651 (0.2883)	−0.5914 (0.5327)	−0.3109 (0.1790)	−0.0341 **(0.0034)**
(5,2,2)	0.1245 (0.1309)	−0.5378 (0.4875)	−0.3568 (0.1874)	−0.0336 **(0.0051)**

Parameter k	ML	LS	M (Tukey)	OBR
(3,1,1)	0.2419 (0.1065)	−0.3498 (0.3793)	0.2529 (0.0928)	0.0223 **(0.0009)**
(3,1,2)	0.5264 (0.3272)	0.5314 (0.4231)	0.2287 (0.0853)	0.0217 **(0.0074)**
(3,2,1)	0.2877 (0.1012)	−0.5686 (0.3724)	0.2880 (0.1090)	0.0161 **(0.0011)**
(3,2,2)	0.6017 (0.4020)	0.6141 (0.4017)	0.2857 (0.1095)	0.0346 **(0.0075)**
(3,3,3)	0.9529 (0.9132)	0.6530 (0.4790)	0.2459 (0.1222)	0.0912 **(0.0250)**
(5,1,1)	0.1709 (0.0692)	0.6144 (0.3888)	0.4969 (0.2470)	0.0198 **(0.0008)**
(5,1,2)	0.3460 (0.1854)	0.9919 (0.9857)	0.4437 (0.2444)	0.0243 **(0.0083)**
(5,2,1)	0.2067 (0.0796)	0.6092 (0.3788)	0.4968 (0.2470)	0.0118 **(0.0081)**
(5,2,2)	0.4108 (0.2310)	0.9936 (0.9879)	0.4991 (0.2491)	0.0377 **(0.0080)**

Table 6 represents the simulation results with one outlier with the sample size 100. According to Table 6, the OBR estimation method outperforms in terms of bias and RMSE values for the most values of the parameters among the others.

The results given in Table 7 are similar to the results reported in Tables 4–6. The OBR estimator seems superior to the other estimators in terms of bias and RMSE values.

To sum up, all of these results show that the amount of efficiency we lose by using the OBR estimation method is negligible in comparison to the other estimation methods in most of the cases.

Table 6: The bias and RMSE (Parenthesis) for $n = 100$ with one outlier.

Parameter α	ML	LS	M (Tukey)	OBR
(3,1,1)	0.8537 (0.9116)	0.7543 (0.6559)	0.0566 (0.0316)	0.0104 **(0.0018)**
(3,1,2)	0.9586 (0.9905)	−0.3791 (0.2258)	0.2690 (0.1110)	0.0130 **(0.0016)**
(3,2,1)	0.9976 (0.9954)	0.8250 (0.6824)	0.0300 (0.0309)	0.0137 **(0.0290)**
(3,2,2)	0.4103 (0.1894)	−0.2510 (0.1249)	0.3399 (0.1359)	0.0192 **(0.0018)**
(3,3,3)	0.9669 (0.9403)	−0.2345 (0.1484)	0.4405 (0.1998)	0.0393 **(0.0753)**
(5,1,1)	0.9206 (0.9225)	0.8001 (0.8085)	0.4078 (0.2375)	0.0153 **(0.0003)**
(5,1,2)	0.9836 (0.9710)	0.9266 (0.8986)	0.4604 (0.2496)	0.0318 **(0.0386)**
(5,2,1)	0.9471 (1.0099)	0.7665 (0.6802)	0.4633 (0.2418)	0.0180 **(0.0407)**
(5,2,2)	0.4346 (0.2033)	0.9036 (0.8763)	0.4800 (0.2500)	0.0268 **(0.0034)**

Parameter c	ML	LS	M (Tukey)	OBR
(3,1,1)	−0.0742 (0.4301)	−0.1888 (0.6320)	−0.1622 (0.0572)	−0.0013 **(0.0231)**
(3,1,2)	−0.0144 (0.0093)	−0.1600 (0.4116)	−0.1358 (0.0390)	−0.0015 **(0.0012)**
(3,2,1)	−0.2041 (0.1376)	−0.4140 (0.3012)	−0.2573 (0.1460)	−0.0035 **(0.0155)**
(3,2,2)	0.0852 (0.0584)	−0.2921 (0.1768)	−0.1749 (0.0913)	−0.0036 **(0.0451)**
(3,3,3)	0.5257 (0.3139)	−0.3673 (0.3133)	−0.1820 (0.1206)	−0.0102 **(0.0600)**
(5,1,1)	−0.0490 (0.0431)	−0.3632 (0.2194)	−0.2476 (0.1015)	−0.0014 **(0.0245)**
(5,1,2)	0.0414 (0.0126)	−0.2780 (0.1118)	−0.2837 (0.1087)	−0.0020 **(0.0012)**
(5,2,1)	−0.1072 (0.1860)	−0.6308 (0.5946)	−0.3055 (0.1529)	−0.0031 **(0.0117)**
(5,2,2)	0.3992 (0.0753)	−0.5101 (0.3795)	−0.3966 (0.1878)	−0.0032 **(0.0343)**

Parameter k	ML	LS	M (Tukey)	OBR
(3,1,1)	0.2783 (0.0949)	−0.9356 (1.0459)	0.2708 (0.0819)	0.0020 **(0.0576)**
(3,1,2)	0.5343 (0.3174)	0.6186 (0.4029)	0.2308 (0.1867)	0.0040 **(0.0915)**
(3,2,1)	0.3315 (0.1188)	−1.1387 (1.3363)	0.3016 (0.0988)	0.0025 **(0.0857)**
(3,2,2)	0.5936 (0.3715)	0.5878 (0.3550)	0.2493 (0.1933)	0.0050 **(0.1096)**
(3,3,3)	0.9831 (0.9681)	0.8888 (0.8074)	0.2834 (0.1353)	0.0124 **(0.0084)**
(5,1,1)	0.1908 (0.0588)	0.6082 (0.3774)	0.4993 (0.2493)	0.0019 **(0.0458)**
(5,1,2)	0.3245 (0.1361)	0.9955 (0.9914)	0.4971 (0.2472)	0.0055 **(0.0107)**
(5,2,1)	0.2006 (0.0638)	0.6048 (0.3712)	0.4987 (0.2488)	0.0021 **(0.0549)**
(5,2,2)	0.3428 (0.1376)	0.9992 (0.9985)	0.3428 (0.1376)	0.0041 **(0.0679)**

Table 7: The bias and RMSE (Parenthesis) for $n = 50$ with four outliers.

Parameter α	ML	LS	M (Tukey)	OBR
(3,1,1)	1.8192 (1.6599)	0.8780 (0.7723)	0.1795 (0.0634)	0.0834 **(0.0116)**
(3,1,2)	2.3995 (2.9045)	0.4393 (0.2599)	0.4573 (0.2555)	0.1753 **(0.0405)**
(3,2,1)	2.1213 (2.6922)	0.8396 (0.7130)	0.2445 (0.1116)	0.1012 **(0.0123)**
(3,2,2)	2.6898 (2.2763)	0.3345 (0.1911)	0.4635 (0.2822)	0.1896 **(0.0679)**
(3,3,3)	2.9292 (2.5814)	0.3273 (0.1249)	0.5968 (0.3730)	0.2891 **(0.1190)**
(5,1,1)	2.9800 (3.0851)	0.8503 (0.7336)	0.2737 (0.1275)	0.1125 **(0.0160)**
(5,1,2)	3.9945 (4.3873)	0.9993 (0.9986)	0.9991 (0.9982)	0.2696 **(0.1125)**
(5,2,1)	3.7069 (3.1097)	0.9545 (0.9885)	0.3372 (0.1666)	0.1335 **(0.0220)**
(5,2,2)	3.5945 (4.1334)	0.9853 (0.9761)	0.9940 (0.9898)	0.2896 **(0.1202)**

Parameter c	ML	LS	M (Tukey)	OBR
(3,1,1)	0.2709 (0.1274)	0.2205 (0.0778)	0.2232 (0.0878)	0.0150 **(0.0005)**
(3,1,2)	0.3592 (0.1972)	0.1978 (0.0636)	0.1986 (0.0697)	0.0167 **(0.0004)**
(3,2,1)	0.7857 (1.1572)	0.5197 (0.6674)	0.3933 (0.2462)	0.0329 **(0.0015)**
(3,2,2)	0.9072 (1.1683)	0.3664 (0.2268)	0.3465 (0.1961)	0.0417 **(0.0044)**
(3,3,3)	1.3173 (2.0957)	0.3082 (0.1525)	0.3728 (0.2224)	0.0800 **(0.0098)**
(5,1,1)	0.3561 (0.2606)	0.3810 (0.2948)	0.3490 (0.2166)	0.0128 **(0.0002)**
(5,1,2)	0.4625 (0.3055)	0.2670 (0.1102)	0.2925 (0.1343)	0.0162 **(0.0004)**
(5,2,1)	0.9675 (1.4720)	0.8075 (1.0989)	0.5793 (0.4692)	0.0352 **(0.0017)**
(5,2,2)	0.9994 (1.2173)	0.4066 (0.2439)	0.4501 (0.3097)	0.0402 **(0.0025)**

Parameter k	ML	LS	M (Tukey)	OBR
(3,1,1)	0.5637 (0.3516)	1.8390 (1.6041)	0.3359 (0.1325)	0.0196 **(0.0008)**
(3,1,2)	1.4464 (2.1420)	0.8282 (0.7011)	0.3686 (0.2293)	0.0474 **(0.0031)**
(3,2,1)	0.6755 (0.4794)	1.4009 (2.7706)	0.3795 (0.1661)	0.0210 **(0.0006)**
(3,2,2)	1.6785 (2.8440)	0.8418 (0.7197)	0.5169 (0.3583)	0.0521 **(0.0067)**
(3,3,3)	2.8503 (3.1275)	0.8355 (0.7122)	0.3288 (0.1913)	0.0926 **(0.0129)**
(5,1,1)	0.5399 (0.3318)	0.6267 (0.4037)	0.6583 (0.4414)	0.0159 **(0.0003)**
(5,1,2)	1.4151 (2.0741)	0.9999 (0.9998)	0.7133 (0.7656)	0.0454 **(0.0032)**
(5,2,1)	0.6832 (0.4851)	0.6485 (0.4277)	0.6753 (0.4605)	0.0205 **(0.0006)**
(5,2,2)	1.6970 (1.8872)	0.9983 (0.9967)	0.1379 (0.1287)	0.0497 **(0.0038)**

According to a anonymous referee's suggestion, we conduct an additional simulation study to confirm the results of real data example considered in the next section. In this simulation design we generate 50 observations from the MOEBXII distribution by using the following initial parameters $(\alpha, c, k) = (30, 2, 1)$. We consider two outlier cases about this simulation design, first we add one outlier and then we add four outliers. The results of this simulation are given in Table 8. According to Table 8, the OBR and the ML methods show similar performances when the data set has no outliers. Considering the RMSE values, the OBR and ML estimators show better performance than the LS and the M estimators. On the other hand, when we create one outlier in the data, the performances of the ML and LS estimators are drastically worsen in terms of the RMSE and the bias values. Unlike the ML and the LS estimates, M estimates do not affected from the outlier. Considering the OBR estimator, we observe that it has the best performance among all the estimators we considered. If the data set has four outliers, then the OBR estimator has the best performances and it is followed by the M estimator. In this case, the ML and the LS estimators are worse according to bias and RMSE values.

In summary, when there are potential outliers in the data the OBR estimation method outperforms among the others in terms of the bias and RMSE values.

Table 8: The bias and RMSE (Parenthesis) for $n = 50$ with $(\alpha, c, k) = (30, 2, 1)$.

No outlier	ML	LS	M (Tukey)	OBR
α	−0.0911 **(0.0189)**	0.1513 (0.1769)	0.0941 (0.1980)	0.0728 (0.0185)
c	0.0297 (0.0177)	−0.2540 (0.2389)	−0.1555 (0.2224)	0.0064 **(0.0083)**
k	−0.0355 (0.0334)	−0.1736 (0.2788)	−0.2290 (0.1051)	0.0034 **(0.0027)**

One outlier	ML	LS	M (Tukey)	OBR
α	0.4326 (0.5467)	0.9932 (0.9774)	0.2994 (0.1121)	0.1264 **(0.0404)**
c	−0.3290 (0.2696)	0.9540 (0.9310)	0.2066 (0.0912)	0.0828 **(0.0024)**
k	0.2182 (0.1020)	0.6785 (0.5012)	0.1924 (0.0469)	0.0478 **(0.0079)**

Four outliers	ML	LS	M (Tukey)	OBR
α	0.6399 (0.9130)	1.0653 (1.8696)	0.5462 (0.3189)	0.1545 **(0.0821)**
c	0.7921 (1.1059)	0.7132 (0.8086)	0.5152 (0.2921)	0.0095 **(0.0003)**
k	0.3678 (0.2183)	0.5893 (0.5048)	0.1863 (0.0461)	0.0547 **(0.0095)**

5. REAL DATA EXAMPLE

In this section the application of the MOEBXII distribution to a real data set is discussed to illustrate the performance of the proposed parameter estimation method. We use a data set from a pharmacy study of Canaparo et *et al.* [5]. The sample size is $n = 65$. The data is related to the ibuprofen which is widely available as an over-the-counter treatment for pain and fever. It represents the mean plasma concentration–time profile of Ibuprofen (S) in all healthy subjects after a single 400 mg oral dose of racemic Ibuprofen. Ibuprofen blood plasma levels are were computed at various time points using data from pharmacokinetics trials.

We use the MOEBXII distribution to fit the data. We consider the ML, LS, M and the OBR estimators to obtain the parameter estimates. The following steps are used to obtain the OBR estimates of the parameters:

- **(i)** Obtain the ML estimate.
- **(ii)** Take $c_B = 3$, the ML estimate as an initial estimate and calculate the OBR estimate.
- **(iii)** Take $c_B = 3$, the OBR estimate obtained in step (ii) as a new initial estimate and calculate the OBR estimate again [11].

Note that one can see [11] and [28] for further details about the selection of the robustness tuning constant [8].

To further see the performance of the estimator, we consider adding one and four outliers to the data. The parameter estimates for the real data are given in Table 9. In this table, we summarize the results for the cases outliers and without outliers. The fitted densities obtained from the ML, the LS, the robust M and the OBR estimates in case of outliers and without outliers, and histogram of the ibuprofen data are shown in Figure 2.

Table 9: The ML, LS, M and OBR estimates for ibuprofen data.

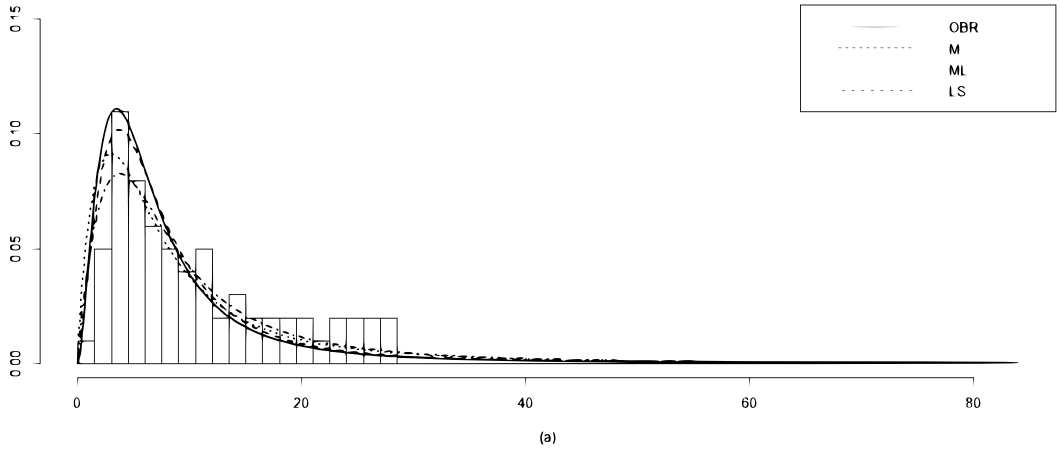
Estimates	without outlier			with one outlier			with four outliers		
	$\widehat{\alpha}$	$\widehat{c}$	$\widehat{k}$	$\widehat{\alpha}$	$\widehat{c}$	$\widehat{k}$	$\widehat{\alpha}$	$\widehat{c}$	$\widehat{k}$
ML	23.7002	1.7654	0.9243	24.7562	1.5853	0.9899	20.1377	2.2483	0.4986
LS	37.002	1.8365	0.9726	38.1051	1.3580	1.2221	35.8070	3.8365	0.3769
M(Tukey)	41.1842	2.2900	0.8721	40.1748	2.3886	0.8160	41.8742	2.9751	0.6129
OBR	34.5757	2.5723	0.7726	34.7176	2.5421	0.7853	34.8430	2.4720	0.8013

Figure 2(a) illustrates the fitted densities when there is no outlier in the data. From Figure 2(a), it can be seen that the MOEBXII distribution is suitable to model the mean plasma concentration of ibuprofen. All of the mentioned estimators are in good agreement in terms of fitting data in the tail. However, the ML and LS are not provided a good fit in the central portion of the data. The fitted density obtained from the robust estimator based on Tukey's ρ_b function shows better fit than the ML and LS fits in the central portion of the data. In particularly, the model obtained from the OBR estimates performs fairly well to describe the central part of the data set. The fitted densities obtained from the ML, LS estimates don't seem catch C_{max}, the pick of the data. Therefore these estimators can not give reasonable estimate for T_{max}, the time taken to reach the maximum concentration.

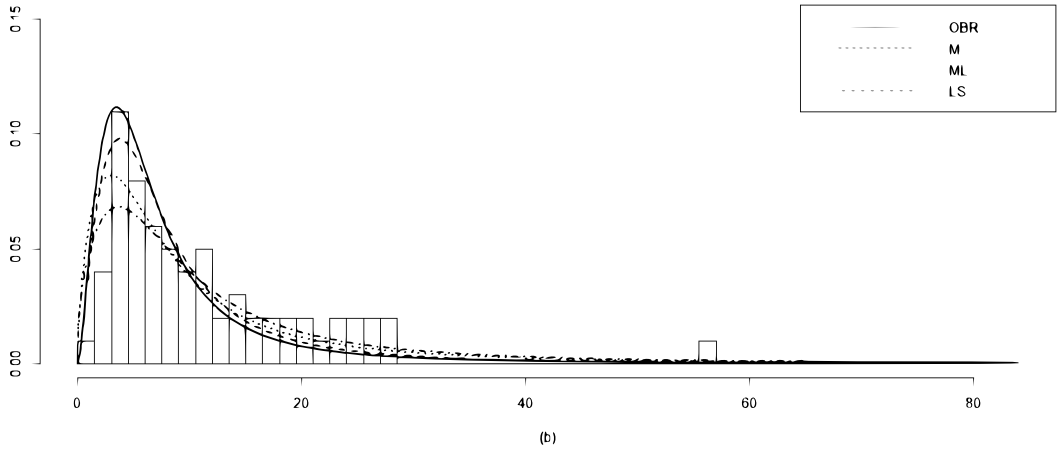
Figure 2(b) shows the fitted densities when there is one outlier in the data set. From this figure, the OBR and M estimators seem not to be affected from one outlier. In addition, from Table 9, it is clear that the estimates obtained from the OBR and M estimation with one outlier is closer to the estimates obtained without outlier. Similar comments can be made for the ML estimates. Adding one outlier causes a small difference on the ML estimation. However, it does not still provides better fit than the OBR and M estimators do. The fitted density obtained from the ML estimates seems not catching the pick of the data. Concerning the LS estimator, it can be seen that only one outlier has an significant effect on LS estimator. This can also be observed from Table 9.

Finally, in Figure 2(c) we display the histogram of the data with four outliers along with the fitted densities. From this figure, we can clearly see that the best fitted density is obtained from the OBR estimation method. The OBR is followed by the M estimator. This figure demonstrates how outliers could potentially distort the ML and the LS estimates. The performance of the ML and the LS estimators are worse than the OBR and M estimator. Results from Table 9 is also supported this outcome.

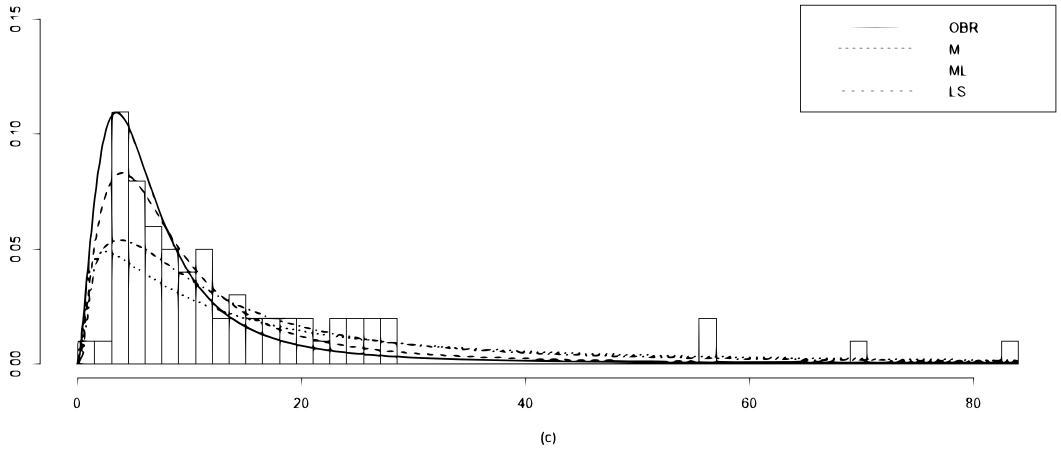
From these results, we can conclude that if we have some outliers in the data, the OBR estimation method can be used safely because the OBR estimation method are not affected from the outliers as the other methods do. To sum up, we can clearly observe that the OBR estimation method can be used to find better fits for the data sets that may have some outliers.



(a) without outlier



(b) one outlier



(c) four outliers

Figure 2: Histogram of Ibuprofen data and the fitted densities with the ML, LS, robust M and OBR estimates.

6. CONCLUSION

Two objectives have been considered in this study. First we have proposed to use the OBR estimation method to estimate the parameters of the MOEBXII distribution proposed by Al-Saiari et *et al.* [3] with the advantage of flexibility to fit the data sets with various shapes. Second, we have considered the modeling the data sets from pharmacokinetics studies represent the changes in plasma concentrations of drugs with the MOEBXII distribution. When the estimation problem is addressed, from both the simulation study and the real data example we observe that the OBR estimator exhibits strong robustness in presence of observations discordant with the assumed model. These results show that not only the OBR estimate achieves smaller RMSE for the small sample sizes but also its RMSE is smaller for the outlier cases for each sample sizes than those of the ML, LS and robust M estimators. The simulation results of the ML and LS estimators for the outlier cases are quite different from the cases without outlier. The existence of outliers in data results in striking differences in RMSE of ML and LS estimates, in contrast to robust estimates, especially the OBR estimates. A general inspection of the table shows that a comparison of the OBR with the ML, LS and robust M estimation methods reveals the superiority of the new estimate in the outlier case and/or small sample case. When we consider the real world data analysis, modeling pharmacokinetics data set with the MOEBXII distribution, from the real data example we can observe that the MOEBXII distribution with the OBR estimates can be a good choice for modeling the changes in plasma concentrations of drugs which is an important pharmacokinetics variable. Because estimating the parameters with the OBR estimation method would be more reliable in estimating other variables such as C_{max} and T_{max} other pharmacokinetics variables.

ACKNOWLEDGMENTS

The authors thank the anonymous referee, the editor and the associate editor for their careful reading, suggestions and encouragement about this paper.

REFERENCES

[1] Akkan, A.G. (2014). Definition of basic FK / FD parameters, *Journal of ANKEM*, **28**(2), 82–85.

[2] Al-Hussaini, E.K. and Jaheen, Z.F. (1992). Bayes estimation of the parameters, Reliability and Failure Rate Functions of the Burr Type XII Failure Model, *Journal of Statistical Computation and Simulation*, **41**(1–2), 31–40.

[3] Al-Saiari, A.Y.; Baharith, L.A. and Mousa, S.A. (2014). Marshall–Olkin extended Burr type XII distribution, *International Journal of Statistics and Probability*, **3**(1), 78–84.

[4] Burr, W. (1942). Cumulative frequency functions, *The Annals of Mathematical Statistics*, **1**(2), 215–232.

[5] Canaparo, R.; Muntoni, E.; Zara, G.P.; Della Pepa, C.; Berno, E.; Costa, M. and Eandi, M. (2000). Determination of Ibuprofen in human plasma by highperformance liquid chromatography: validation and application in pharmacokinetic study, *Biomed. Chromatogr.*, **14**, 219–226.

[6] Dogru, F.Z. and Arslan, O. (2015). Alternative robust estimators for the shape parameters of the Burr XII distribution, *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering*, **9**(5), 271–276.

[7] Dogru, F.Z. and Arslan, O. (2016). Optimal B-robust estimators for the parameters of the Burr XII distribution, *Journal of Statistical Computation and Simulation*, **86**(6), 1133–1149.

[8] Dogru, F.Z.; Bulut, Y.M. and Arslan, O. (2017). Optimal B-robust estimators for the parameters of the generalized half-normal distribution, *REVSTAT – Statistical Journal*, **15**(3), 455–471.

[9] Guney, Y. and Arslan, O. (2017). Robust parameter estimation for the Marshall–Olkin extended Burr XII distribution, *Commun. Fac. Sci. Univ. Ank. Series A1*, **66**(2), 141–161.

[10] Gupta, P.L.; Gupta, R.C. and Lvin, S.J. (1996). Analysis of failure time data by Burr distribution, *Communications in Statistics – Theory and Methods*, **25**, 2013–2024.

[11] Hampel, F.R.; Ronchetti, E.M.; Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Function*, John Wiley & Sons Inc., New York.

[12] Holland, P.W. and Welsch, R.E. (1977). Robust regression using iteratively reweighted least-squares, *Communications in Statistics – Theory and Methods*, **6**(9), 813–827.

[13] Hossain, A.M. and Nath, S.K. (1997). Estimation of parameters in the presence of outliers for a Burr XII distribution, *Communications in Statistics – Theory and Methods*, **26**(3), 637–652.

[14] Huber, P.J. (1964). Robust estimation of a location parameter, *Ann. Math. Statist.*, **35**, 73–101.

[15] Klugman, S.A. (1986). Loss Distributions, *Proceedings of Symposia in Applied Mathematics Actuarial Mathematics*, **35**, 31–55.

[16] Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data*, John Wiley, New York.

[17] Malinowska, I.; Pawlas, P. and Szynal, D. (2006). Estimation of location and scale parameters for the Burr XII distribution using generalized order statistics, *Linear Algebra and its Applications*, **417**(1), 150–162.

[18] Maronna, R.A.; Martin, R.D. and Yohai, V.J. (2006). textitRobust Statistics: Theory and Methods, John Wiley, New York.

[19] Marshall, A.W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, *Biometrika*, **84**(3), 641–652.

[20] McDonald, J.B. (1984). Some generalized function for the size distribution of income, *Econometrica*, **52**(3), 647–663.

[21] Moore, D. and Papadopoulos, A.S. (2000). The Burr Type XII distribution as a failure model under various loss functions, *Microelectronics Reliability*, **40**(12), 2117–2122.

[22] Papadopoulos, A.S. (1978). The Burr distribution as a life time model from a Bayesian approach, *IEEE Transactions on Reliability*, **27**(5), 369–371.

[23] Rodriguez, R.N. (1977). A guide to Burr XII distribution, *Biometrika*, **64**, 129–134.

[24] Shao, Q. (2004). Notes on maximum likelihood estimation for the three-parameter Burr XII distribution, *Computational Statistics & Data Analysis*, **45**, 675–687.

[25] Swain, J.; Venkatraman, S. and Wilson, J. (1988). Least squares estimation of distribution function in Johnson's translation system, *Journal of Statistical Computation and Simulation*, **29**, 271–297.

[26] Tadikamalla, P.R. (1980). A look at the Burr and related distributions, *Int. Stat. Rev.*, **48**(3), 337–344.

[27] Victoria-Feser, M.P. (1993). *Robust methods for personal income distribution models*, Dissertation, University of Geneva, Switzerland.

[28] Victoria-Feser, M.P. and Ronchetti, E. (1994). Robust methods for personal-income distribution models, *The Canadian Journal of Statistics*, **22**(3), 247–258.

[29] Wang, F.K. and Cheng, Y.F. (2010). Robust regression for estimating the Burr XII parameters with outliers, *Journal of Applied Statistics*, **37**(5), 807–819.

[30] Wingo, D.R. (1983). Maximum likelihood methods for fitting the Burr type XII distribution to life test data, *Biometrical Journal*, **25**(1), 77–84.

[31] Wingo, D.R. (1993). Maximum likelihood methods for fitting the Burr type XII distribution to multiply (progressively) censored life test data, *Metrika*, **40**, 203–210 .

REVSTAT – Statistical Journal
Volume 19, Number 3, July 2021, 443–462

PSEUDO-GAUSSIAN AND RANK-BASED TESTS FOR FIRST-ORDER SUPERDIAGONAL BILINEAR MODELS IN PANEL DATA

Authors: Aziz Lmakri
– Mathematics and Applications Laboratory, Faculty of Sciences and Techniques, Abdelmalek Essaadi University, Tangier, Morocco
lmakriaziz@gmail.com

Abdelhadi Akharif
– Mathematics and Applications Laboratory, Faculty of Sciences and Techniques, Abdelmalek Essaadi University, Tangier, Morocco
aakharif@gmail.com

Amal Mellouk
– Regional Center for Education and Training Trades, Tangier, Morocco
prmelloukamal@hotmail.com

Mohamed Fihri
– Laboratory of Mathematics, Computing and Applications – Information Security (LabMiA-SI), Department of Mathematics, Faculty of Sciences, Mohammed-V University in Rabat, Morocco
m.fihri@gmail.com

Received: March 2019 Revised: July 2019 Accepted: July 2019

Abstract:

- In this paper, locally asymptotically optimal (in the Hájek-Le Cam sense) parametric, pseudo-Gaussian and rank-based procedures are proposed for the problem of testing randomness against first-order superdiagonal bilinear panel dependence (in large n and small T panels). Local powers and asymptotic relative efficiencies are computed and show that the van der Waerden version of our rank-based tests uniformly dominates. Small-sample performances are investigated via simulations and confirm the theoretical findings, they also demonstrate the remarkable performances of rank procedures based on data-driven scores.

Keywords:

- *panel data; first-order superdiagonal bilinear model; local asymptotic normality; pseudo-Gaussian test; rank test; asymptotic relative efficiency.*

AMS Subject Classification:

- 62F03, 62G10.

1. INTRODUCTION

Nonlinear time series have attracted much attention in the last four decades. Many classes of models have been proposed and applied with great success in many important real-life problems; such as economics (Granger and Andersen [12]), demography (Subba Rao and Gabr [32]), environmental studies (Guegan [14]), etc. One of the most popular was the bilinear time series models $BL(p,q,P,Q)$[1]. In the first time, these models were proposed and developed by Granger and Andersen [12]; then becomes Phan and Tran [27], Subba Rao [31], Guegan [13], Liu and Brockwell [25]. Particularly to those models, we quote the first-order superdiagonal bilinear models $BL(0,0,2,1)$, who also recognized applications in many fields (see, for example, [26, 36, 5]).

This paper deals with the presence of a first-order superdiagonal bilinear model in ***panel data*** (a series of T observations made through time over a number n of individuals), denoted by $BLP(0,0,2,1)$ and defined, for $i = 1,2,...,n$ and $t = 1,2,...,T$, as:

$$X_{i,t} = bX_{i,t-2}\varepsilon_{i,t-1} + \varepsilon_{i,t}, \tag{1.1}$$

where $X_{i,t}$ is a panel observation (for individual i at time t) described by a nonlinear stochastic difference in time equation; $(\varepsilon_{i,t})$ is a white noise process, i.e. a sequence of independent, identically distributed random variables with mean zero, finite variance σ^2 and density distribution $\varepsilon \mapsto f(\varepsilon) := (1/\sigma)f_1(\varepsilon/\sigma)$ (where $f_1 \in \mathcal{F}_0$, see (2.1)) and b is a constant in $\mathbb{R}$. The probabilistic properties of a first-order superdiagonal time series model $BL(0,0,2,1)$ processes (such as invertibility and stationarity) have been studied by several references [28, 13]. These properties also remain valid under a *first-order superdiagonal panel model* $BLP(0,0,2,1)$. Let us denote by $\mathcal{F}_{i,t}(\varepsilon)$ and $\mathcal{F}_{i,t}(X)$ the σ-algebras generated by $\{\varepsilon_{i,s}|s \leqslant t\}$ and $\{X_{i,s}|s \leqslant t\}$, respectively. Then:

1. Equation (1.1) admits a unique stationary solution $(X_{i,t})$ (i.e., $\mathcal{F}_{i,t}(\varepsilon)$-measurable) iff $b^2\sigma^2 < 1$, in this case, one can write

$$X_{i,t} = \sum_{j=1}^{\infty} b^j \varepsilon_{i,t-2j} \prod_{k=1}^{j} \varepsilon_{i,t-2k+1} + \varepsilon_{i,t}; \tag{1.2}$$

2. Equation (1.1) is invertible (i.e., $\varepsilon_{i,t}$ is $\mathcal{F}_{i,t}(X)$-measurable) iff $2b^2\sigma^2 < 1$, in this case, one can write

$$\varepsilon_{i,t} = X_{i,t} + \sum_{j=1}^{\infty} (-b)^j X_{i,t-j} \prod_{k=1}^{j} X_{i,t-k-1}. \tag{1.3}$$

[1] These models are defined as:

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \sum_{j=1}^{q} c_j \varepsilon_{t-j} + \sum_{j=1}^{P}\sum_{k=1}^{Q} b_{jk}\varepsilon_{t-j}X_{t-k} + \varepsilon_t.$$

Several methods — such as the method of moments, the least squares method and the repeated residual method — have been established in the literature for estimating the parameters of bilinear models, see, for example, Pham and Tran [28], Sesay and Subba Rao [30], Grahn [11], Bouzaachane [5] and Tan and Wang [34].

Before turning to the problem of estimating the parameters of model (1.1), it is very important to know if it is indeed a $BLP(0,0,2,1)$, and how the test proposed for testing randomness against first-order superdiagonal bilinear panel dependence is efficient. Note that if $b=0$, $X_{i,t}$ reduces to white noise ($X_{i,t}=\varepsilon_{i,t}$), else $b\neq 0$, panel data follows a $BLP(0,0,2,1)$ (alternative hypothesis) — such a test is *bilateral*.

To start with, locally and asymptotically optimal parametric tests are constructed using the Local Asymptotic Normality LAN property. Then, the special case of the pseudo-Gaussian tests (optimal under Gaussian densities and valid under finite-variance non-Gaussian ones) is derived. Unfortunately, their local asymptotic power, under non-Gaussian g_1 (especially the skew and heavy-tailed ones), can be extremely poor. Which leads us to the construction of rank-based optimal tests (van der Waerden, Wilcoxon, Laplace, data-driven scores, etc.).

Asymptotic relative efficiencies with respect to the pseudo-Gaussian procedure show that the van der Waerden version of our rank-based tests uniformly dominates its pseudo-Gaussian countepart.

The paper is organized as follow: Section 2.1 provides the main definitions and assumptions. The local asymptotic normality, with respect to b and σ^2, in the vicinity of $b=0$, of the family of distributions associated with (1.1) (with specified f_1), is established in Section 2.2. In Section 3.1, we propose (still, for specified f_1) the optimal parametric test. The particular case of the pseudo-Gaussian test is proposed in Section 3.2. Section 4 proposes rank-based procedures that remain valid irrespective of f_1. Particular cases (van der Warden, Wilcoxon, Laplace scores, ...) are considered in Section 4.3. Asymptotic relative efficiencies with respect to the pseudo-Gaussian test is derived in Section 5. Section 6 provides some simulation results assessing the finite-sample performance of the various tests proposed. Finally, Section 7 concludes.

2. LOCAL ASYMPTOTIC NORMALITY

2.1. Notations and main technical assumptions

Denote by $\mathbf{P}^{(n)}_{\sigma^2,0;f_1}$ the probability distribution under the null $X_{i,t}=\varepsilon_{i,t}$. Under the alternative, the probability distribution is denoted by $\mathbf{P}^{(n)}_{\sigma^2,b;f_1}$ ($b\neq 0$), the observations $X^{(n)}:=(X_1^{(n)\prime},X_2^{(n)\prime},...,X_n^{(n)\prime})'$ with $X_i^{(n)}:=(X_{i,1},...,X_{i,T})'$ is generated by (1.1).

We suppose that the vector $X_0^{(n)}:=\{(X_{i,-1}^{(n)}\varepsilon_{i,0},X_{i,0}^{(n)}),i=1,2,...,n\}$ is observable for each individual i, and admits a density $h_\theta(.)$ continuous in θ. The influence of these starting values is asymptotically negligible (see Hallin and Werker (1999) [20] for a detailed discussion).

Throughout, we consider the class of standardized densities

$$\mathcal{F}_0 := \left\{ f_1 : \int_{-\infty}^{0} f_1(u)du = 0.5 = \int_{-1}^{1} f_1(u)du \right\}. \tag{2.1}$$

Under $f_1 \in \mathcal{F}_0$, the median and median absolute deviation are 0 and σ respectively; this standardization avoids all moment assumptions and has no impact on subsequent results.

Our derivation of locally asymptotically optimal tests at density f_1 will be based on the local asymptotic normality, with respect to $(\sigma^2, b)'$, of the families of distributions

$$\mathcal{P}_{f_1}^{(n)} := \left\{ \mathbf{P}_{\sigma^2,b;f_1}^{(n)} | (\sigma^2, b)' \in \mathbb{R}_+^* \times \mathbb{R} \text{ and } 2b^2\sigma^2 < 1 \right\} \tag{2.2}$$

at any $\theta := (\sigma^2, 0)'$.

This LAN property requires some technical assumptions on the innovation density f_1. Denote by $\mathcal{F}_A$ the class of all densities f_1 satisfying the following technical assumptions:

(**A.1**) $f_1 \in \mathcal{F}_0$;

(**A.2**) $f_1(u) > 0, \quad \forall u \in \mathbb{R}$;

(**A.3**) f_1 is absolutely continuous on bounded intervals, i.e., there exists f_1' such that

$$f_1(b) - f_1(a) = \int_a^b f_1'(u)du \quad \text{for all} \quad a < b,$$

and, letting $\Phi_{f_1} = -f_1'/f_1$, assume that

$$I(f_1) := \int_{\mathbb{R}} \Phi_{f_1}^2(u) f_1(u)du \qquad \text{and} \qquad J(f_1) := \int_{\mathbb{R}} u^2 \Phi_{f_1}^2(u) f_1(u)du$$

are finite.

For instance, interesting special cases of f_1 are obtained:

- The double-exponential or Laplace distribution, with standardized density
$$f_1(u) = f_{\mathcal{L}}(u) := (1/2d)\exp(-|u|/d),$$
with $I(f_1) = 1/d^2$ and $J(f_1) = 2$; the normalizing constant $d := 1/\ln(2) \simeq 1.4426$ is such that $f_{\mathcal{L}} \in \mathcal{F}_A$.
- The logistic distribution, with standardized density
$$f_1(u) = f_{Log}(u) := \sqrt{b}\exp(-\sqrt{b}u)/(1+\exp(-\sqrt{b}u))^2,$$
with $I(f_1) = b/3$ and $J(f_1) = (12+\pi^2)/9$; the normalizing constant $b := (\ln 3)^2) \simeq 1.2069$ is such that $f_{\mathcal{L}} \in \mathcal{F}_A$.
- The Student distributions (with $\nu > 2$ degrees of freedom), with standardized density
$$f_1(u) = f_{t_\nu}(u) := \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\sqrt{a_\nu/\pi\nu}(1+a_\nu u^2/\nu)^{-(\nu+1)/2},$$
with $I(f_1) = a_\nu(\nu+1)/(\nu+3)$ and $J(f_1) = 3(\nu+1)/(\nu+3)$; the normalizing constant $a_\nu > 0$ is such that $f_{t_\nu} \in \mathcal{F}_A$.
- The Gaussian distribution, with standardized density (with mean zero and variance $1/a$)
$$f_1(u) = f_{\mathcal{N}}(u) := \sqrt{a/2\pi}\exp(-au^2/2),$$
with $I(f_1) = a \simeq 0.4549$ and $J(f_1) = 3$.

2.2. LAN

Let us denote by $\theta^{(n)}$ the local sequences of perturbations of $\theta = (\sigma^2, 0)'$, where

$$\theta^{(n)} = \theta + n^{\frac{-1}{2}}\tau \quad \text{with} \quad \tau = \big(\tau_1, \tau_2\big)' \in \mathbb{R}^2.$$

The bilateral test is equivalent to:

$$\begin{cases} \mathbf{P}^{(n)}_{\theta;f_1} : \tau_2 = 0, \\ \mathbf{P}^{(n)}_{\theta^{(n)};f_1} : \tau_2 \neq 0. \end{cases}$$

Under the null, the likelihood function for $(X_0^{(n)}, X^{(n)})$ is

$$L_{\theta;f}(X_0^{(n)}, X^{(n)}) = h_\theta(X_0^{(n)}) \prod_{i=1}^{n}\prod_{t=1}^{T} f(X_{i,t}). \tag{2.3}$$

If $\tau_2 \neq 0$, the likelihood function for $(X_0^{(n)}, X^{(n)})$ in this case is

$$\begin{aligned} L_{\theta^{(n)};f}(X_0^{(n)}, X^{(n)}) &= h_{\theta^{(n)}}(X_0^{(n)}) \prod_{i=1}^{n}\prod_{t=1}^{T} f\big(X_{i,t} + \sum_{j=1}^{\infty}(-n^{\frac{-1}{2}}\tau_2)^j X_{i,t-j} \prod_{k=1}^{j} X_{i,t-k-1}\big) \\ &= h_{\theta^{(n)}}(X_0^{(n)}) \prod_{i=1}^{n}\prod_{t=1}^{T} f\big(X_{i,t} + \Upsilon_n(\tau_2)\big), \end{aligned} \tag{2.4}$$

where $\Upsilon_n(\tau_2) := \sum_{j=1}^{\infty}(-n^{\frac{-1}{2}}\tau_2)^j X_{i,t-j} \prod_{k=1}^{j} X_{i,t-k-1}$.

Denote by $\Lambda^{(n)}_{\theta^{(n)}/\theta;f}$ the logarithm of the likelihood ratio (conditional on $X_0^{(n)}$) for $\mathbf{P}^{(n)}_{\theta^{(n)};f}$ against $\mathbf{P}^{(n)}_{\theta;f}$:

$$\Lambda^{(n)}_{\theta^{(n)}/\theta;f} := \log\Big(L_{\theta^{(n)};f}(X_0^{(n)}, X^{(n)}) / L_{\theta;f}(X_0^{(n)}, X^{(n)})\Big). \tag{2.5}$$

It can be expressed as follows:

$$\Lambda^{(n)}_{\theta^{(n)}/\theta;f} = \sum_{i=1}^{n}\sum_{t=1}^{T}\Big(\log f(X_{i,t} + \Upsilon_n(\tau_2)) - \log f(X_{i,t})\Big) + o_p(1).$$

The $o_p(1)$ term (under $\mathbf{P}^{(n)}_{\theta;f}$, as $n \to \infty$) corresponds to the influence of the starting value $X_0^{(n)}$.

Write $Z_{i,t}$ for the standardized residual

$$Z_{i,t}(\sigma^2, b) := \sigma^{-1}\bigg(X_{i,t} + \sum_{j=1}^{\infty}(-b)^j X_{i,t-j} \prod_{k=1}^{j} X_{i,t-k-1}\bigg),$$

and note that, under $\mathbf{P}^{(n)}_{\theta;f_1}$, these residuals coincide with $\sigma^{-1}\varepsilon_{i,t}$. The local asymptotic normality result, with respect to σ^2 and the parameter of interest b for a fixed density f_1, is established in the next proposition.

Proposition 2.1. *Let $f_1 \in \mathcal{F}_A$. Then the family $\mathcal{P}_{f_1}^{(n)}$ is LAN at any $\theta = (\sigma^2, 0)'$, with central sequence*

$$\Delta_{f_1}^{(n)}(\theta) := \begin{pmatrix} \Delta_{f_1;1}^{(n)}(\theta) \\ \Delta_{f_1;2}^{(n)}(\theta) \end{pmatrix} := \begin{pmatrix} \frac{1}{2\sigma^2} n^{\frac{-1}{2}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[\Phi_{f_1}(Z_{i,t}) Z_{i,t} - 1\right] \\ n^{\frac{-1}{2}} \sigma \sum_{i=1}^{n} \sum_{t=3}^{T} \Phi_{f_1}(Z_{i,t}) Z_{i,t-1} Z_{i,t-2} \end{pmatrix}, \tag{2.6}$$

and information matrix

$$\Gamma_{f_1}(\theta) := \left(\Gamma_{f_1;ij}(\theta)\right)_{1 \le i,j \le 2} := \begin{pmatrix} \frac{T}{4\sigma^4}\big(J(f_1) - 1\big) & 0 \\ 0 & \sigma^2 (T-2) I(f_1) \sigma_{f_1}^4 \end{pmatrix}. \tag{2.7}$$

More precisely, for any $\tau = (\tau_1, \tau_2)' \in \mathbb{R}^2$, under $\boldsymbol{P}_{\theta;f_1}^{(n)}$, as $n \to \infty$ and fixed T, we have

$$\Lambda_{\theta^{(n)}/\theta;f_1}^{(n)} = \tau' \Delta_{f_1}^{(n)}(\theta) - \frac{1}{2} \tau' \Gamma_{f_1}(\theta) \tau + o_p(1), \tag{2.8}$$

and $\Delta_{f_1}^{(n)}(\theta)$ is asymptotically normal, with mean zero under $\boldsymbol{P}_{\theta;f_1}^{(n)}$, mean $\Gamma_{f_1}(\theta)\tau$ under $\boldsymbol{P}_{\theta^{(n)};f_1}^{(n)}$ and variance $\Gamma_{f_1}(\theta)$ under both.

Proof: The proof relies on Swensen's conditions 1.2 to 1.7 of lemma 1 in [33]. More precisely, the only delicate one is the condition 1.2. The main point consists in showing that

$$(\sigma^2, b) \mapsto q_{\sigma^2,b;f_1}^{\frac{1}{2}}(z) := \left[\frac{1}{\sigma} f_1\left(\frac{z + \sum_{j=1}^{\infty} (-b)^j x_j \prod_{k=1}^{j} x_{k-1}}{\sigma}\right)\right]^{\frac{1}{2}}$$

is differentiable in mean quadratic. It is established in the following lemma.

Lemma 2.1. *Let $f_1 \in \mathcal{F}_A$. Define, for $z \in \mathbb{R}$,*

$$\begin{aligned} D_{\sigma^2} q_{\sigma^2,0;f_1}^{\frac{1}{2}}(z) &= \frac{1}{4\sigma^2} q_{\sigma^2,0;f_1}^{\frac{1}{2}}(z)\big(\tfrac{z}{\sigma}\Phi_{f_1}\big(\tfrac{z}{\sigma}\big) - 1\big), \\ D_b q_{\sigma^2,b;f_1}^{\frac{1}{2}}(z)_{|b=0} &= \frac{1}{2\sigma} q_{\sigma^2,0;f_1}^{\frac{1}{2}}(z) \Phi_{f_1}\big(\tfrac{z}{\sigma}\big) x_1 x_0. \end{aligned}$$

Then, as s and $l \to 0$,

1. $\displaystyle\int_{\mathbb{R}} \left[q_{\sigma^2+s,l;f_1}^{\frac{1}{2}}(z) - q_{\sigma^2+s,0;f_1}^{\frac{1}{2}}(z) - l D_b q_{\sigma^2+s,b;f_1}^{\frac{1}{2}}(z)_{|b=0}\right]^2 dz = o(l^2),$
2. $\displaystyle\int_{\mathbb{R}} \left[q_{\sigma^2+s,0;f_1}^{\frac{1}{2}}(z) - q_{\sigma^2,0;f_1}^{\frac{1}{2}}(z) - s D_{\sigma^2} q_{\sigma^2,0;f_1}^{\frac{1}{2}}(z)\right]^2 dz = o(s^2),$
3. $\displaystyle\int_{\mathbb{R}} \left[q_{\sigma^2+s,l;f_1}^{\frac{1}{2}}(z) - q_{\sigma^2,0;f_1}^{\frac{1}{2}}(z) - (s,l) \begin{pmatrix} D_{\sigma^2} q_{\sigma^2,0;f_1}^{\frac{1}{2}}(z) \\ D_b q_{\sigma^2,b;f_1}^{\frac{1}{2}}(z)_{|b=0} \end{pmatrix}\right]^2 dz = o(\| (s,l)' \|^2).$

Proof of Lemma 2.1:

1. Let $\Upsilon(b) = \sum_{j=1}^{\infty}(-b)^j x_j \prod_{k=1}^{j} x_{k-1}$. Then 1 takes the form

$$\int_{\mathbb{R}} \left[\frac{1}{\sqrt{\sigma^2+s}} f_1^{\frac{1}{2}}\left(\frac{z+\Upsilon(l)}{\sqrt{\sigma^2+s}}\right) - \frac{1}{\sqrt{\sigma^2+s}} f_1^{\frac{1}{2}}\left(\frac{z}{\sqrt{\sigma^2+s}}\right) - l\frac{1}{2\sqrt{\sigma^2+s}} q^{\frac{1}{2}}_{\sigma^2+s,0;f_1}(z)\Phi_{f_1}\big(\frac{z}{\sqrt{\sigma^2+s}}\big)x_1x_0 \right]^2 dz = o(l^2),$$

is equivalent to

$$\int_{\mathbb{R}} \left[f^{\frac{1}{2}}\big(z+\Upsilon(l)\big) - f^{\frac{1}{2}}\big(z\big) - \frac{l}{2} f^{\frac{1}{2}}\big(z\big)\Phi_f\big(z\big)x_1x_0 \right]^2 dz = o(l^2),$$

which is equivalent to

$$\int_{\mathbb{R}} l^2 \left[\frac{f^{\frac{1}{2}}\big(z+\Upsilon(l)\big) - f^{\frac{1}{2}}\big(z\big)}{l} + \frac{1}{2}\frac{f'\big(z\big)}{f^{\frac{1}{2}}\big(z\big)}x_1x_0 \right]^2 dz = o(l^2),$$

hence, for proving that, it is sufficient to prove that

$$\lim_{l\to 0}\int_{\mathbb{R}} \left[\frac{f^{\frac{1}{2}}\big(z+\Upsilon(l)\big) - f^{\frac{1}{2}}\big(z\big)}{l} + \frac{1}{2}\frac{f'\big(z\big)}{f^{\frac{1}{2}}\big(z\big)}x_1x_0 \right]^2 dz = 0.$$

We have

$$\begin{aligned}
\lim_{l\to 0}\frac{f^{\frac{1}{2}}\big(z+\Upsilon(l)\big) - f^{\frac{1}{2}}\big(z\big)}{l} &= \lim_{l\to 0}\frac{f^{\frac{1}{2}}\big(z+\Upsilon(l)\big) - f^{\frac{1}{2}}\big(z\big)}{\Upsilon(l)} \times \frac{\Upsilon(l)}{l} \\
&= \big(f^{\frac{1}{2}}\big(z\big)\big)' \times (-x_1x_0) \\
&= -\tfrac{1}{2}\frac{f'\big(z\big)}{f^{\frac{1}{2}}\big(z\big)}x_1x_0.
\end{aligned}$$

And just show that $\int_{\mathbb{R}} \left[\frac{f^{\frac{1}{2}}\big(z+\Upsilon(l)\big) - f^{\frac{1}{2}}\big(z\big)}{l} \right]^2 dz \leqslant \int_{\mathbb{R}} \left[\frac{-1}{2}\frac{f'\big(z\big)}{f^{\frac{1}{2}}\big(z\big)}x_1x_0 \right]^2 dz < \infty.$

We know that $f^{\frac{1}{2}}\big(z+\Upsilon(l)\big) - f^{\frac{1}{2}}(z) = \int_z^{z+\Upsilon(l)} \frac{1}{2} f'(t) f^{\frac{-1}{2}}(t)dt$, then

$$\begin{aligned}
\int_{z=-\infty}^{+\infty} \left[\frac{f^{\frac{1}{2}}\big(z+\Upsilon(l)\big) - f^{\frac{1}{2}}\big(z\big)}{l} \right]^2 dz &= \int_{z=-\infty}^{+\infty} \frac{1}{l^2}\left[\int_{t=z}^{z+\Upsilon(l)} \frac{1}{2} f'(t) f^{\frac{-1}{2}}(t)dt \right]^2 dz \\
&\leqslant \frac{\Upsilon(l)}{l^2}\int_{z=-\infty}^{+\infty}\int_{t=z}^{z+\Upsilon(l)} \left[\frac{1}{2} f'(t) f^{\frac{-1}{2}}(t)\right]^2 dt\, dz \\
&\leqslant \frac{\Upsilon(l)}{l^2}\int_{t=-\infty}^{+\infty}\int_{z=t-\Upsilon(l)}^{t} \left[\frac{1}{2} f'(t) f^{\frac{-1}{2}}(t)\right]^2 dt\, dz \\
&\leqslant \left[\frac{\Upsilon(l)}{l}\right]^2 \int_{t=-\infty}^{+\infty} \left[\frac{1}{2} f'(t) f^{\frac{-1}{2}}(t)\right]^2 dt \\
&\leqslant (-x_1x_0)^2 \int_{t=-\infty}^{+\infty} \left[\frac{1}{2} f'(t) f^{\frac{-1}{2}}(t)\right]^2 dt \\
&\leqslant \int_{\mathbb{R}} \left[\frac{-1}{2} f'(t) f^{\frac{-1}{2}}(t) x_1x_0\right]^2 dt.
\end{aligned}$$

This completes the proof of part 1 of Lemma 2.1.

2. The problem here is reduced to the classical case of linear models considered by Swensen (1985) [33].

3. The result here follows from 1 and 2 above. This completes the proof of Lemma 2.1. □

The diagonal form of the information matrix confirms that σ^2 and b are not related, in the parametric family (2.2). They play distinct and well separated roles.

The Gaussian versions ($f_1 = f_{\mathcal{N}}$) of (2.6) and (2.7) are

$$\Delta_{\mathcal{N}}^{(n)}(\theta) = \begin{pmatrix} \frac{1}{2\sigma^2} n^{\frac{-1}{2}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[a Z_{i,t}^2 - 1 \right] \\ n^{\frac{-1}{2}} \sigma a \sum_{i=1}^{n} \sum_{t=3}^{T} Z_{i,t} Z_{i,t-1} Z_{i,t-2} \end{pmatrix} \quad \text{and} \quad \Gamma_{\mathcal{N}}(\theta) = \begin{pmatrix} \frac{T}{2\sigma^4} & 0 \\ 0 & \frac{\sigma^2}{a}(T-2) \end{pmatrix},$$

respectively.

The result of Proposition 2.1, implies that, under assumptions $\mathcal{F}_A$, as $n \to \infty$, the family of *first-order superdiagonal panel models* $BLP(0,0,2,1)$ possesses the LAN property in a neighbourhood of white noise. This result leads us to construct asymptotically optimal parametric tests under a specified f_1. Note that these tests are valid under a specified f_1, and thereafter we will propose more general tests such as Pseudo-Gaussian and Rank-based procedures which are valid under general densities.

3. OPTIMAL PARAMETRIC AND PSEUDO-GAUSSIAN TESTS

As mentioned above, the Le Cam theory of LAN experiments allows for constructing tests which are locally asymptotically optimal (namely, *most stringent*). The basic idea is the weak convergence concept of the sequence of local experiments to the Gaussian shift two-dimensional model $\boldsymbol{\Delta} \sim \mathcal{N}(\boldsymbol{\Gamma}\tau, \boldsymbol{\Gamma})$. For a general theory on locally asymptotically optimal testing in LAN families, the reader is referred to Le Cam (1986) [23] and van der Vaart (1998) [35].

We are interested in testing the null hypothesis $b = 0$ of randomness in (1.1), with unspecified standardized error density in $\mathcal{F}_0$. To do, let us start with the case when $f_1 \in \mathcal{F}_0$ is specified, i.e., the null hypothesis is such that

$$\mathcal{H}_0^{(n)}(f_1) := \bigcup_{\sigma^2>0} \{\mathbf{P}^{(n)}_{\sigma^2,0;f_1}\},$$

and parametric alternatives take the form

$$\mathcal{H}_1^{(n)}(f_1) := \bigcup_{\sigma^2>0} \bigcup_{b\in\mathbb{R}} \{\mathbf{P}^{(n)}_{\sigma^2,b;f_1}\}.$$

3.1. Optimal parametric tests

Since $\theta = (\sigma^2, 0)' = (1,0)'\sigma^2 =: \Omega\sigma^2$, then $\theta \in \mathcal{M}(\Omega)$, where $\mathcal{M}(\Omega)$ is the linear subspace of dimension 1 of $\mathbb{R}^2$ generated by the vector $\Omega := (1,0)'$. Recall that we are testing $\tau_2 = 0$ against $\tau_2 \neq 0$, which is equivalent to testing $\tau \in \mathcal{M}(\Omega)$ against $\tau \notin \mathcal{M}(\Omega)$. Such tests should be based on the asymptotically chi-square distribution (see S. Ghosh (1999) [10]) and therefore the test statistic takes the form

$$Q_{f_1}(\theta) := \Delta_{f_1}^{(n)'}(\theta)\left[\Gamma_{f_1}^{-1}(\theta) - \Omega\big(\Omega'\Gamma_{f_1}(\theta)\Omega\big)^{-1}\Omega'\right]\Delta_{f_1}^{(n)}(\theta). \tag{3.1}$$

By algebra calculations, one can write

$$Q_{f_1}(\theta) = \Gamma_{f_1;22}^{-1}(\theta)\Delta_{f_1;2}^{(n)2}(\theta) = \underline{\Delta}_{f_1}^{(n)2}/((T-2)I(f_1)\sigma_{f_1}^4) =: \underline{Q}_{f_1}, \tag{3.2}$$

with $\underline{\Delta}_{f_1}^{(n)} = n^{\frac{-1}{2}}\sum_{i=1}^{n}\sum_{t=3}^{T}\Phi_{f_1}(Z_{i,t})Z_{i,t-1}Z_{i,t-2}$.

The test based on (3.2) is locally asymptotically *most stringent* for the problem of detecting the $BLP(0,0,2,1)$ dependance in white noise process. The application of Le Cam's Third Lemma provides the asymptotic law of $\underline{Q}_{f_1}$ under $\mathbf{P}^{(n)}_{\theta^{(n)};f_1}$, so we have the following proposition.

Proposition 3.1. *Let $f_1 \in \mathcal{F}_A$. Then, for any $\tau = (\tau_1, \tau_2)' \in \mathbb{R}^2$,*

(**i**) *$\underline{Q}_{f_1}$ is asymptotically central chi-square with 1 degree of freedom under $\boldsymbol{P}^{(n)}_{\theta;f_1}$, and asymptotically noncentral chi-square, still with 1 degrees of freedom and with noncentrality parameter $\lambda_{f_1} := (T-2)I(f_1)\sigma^2\sigma_{f_1}^4\tau_2^2$ under $\boldsymbol{P}^{(n)}_{\theta^{(n)};f_1}$;*

(**ii**) *The sequence of tests rejecting the null hypothesis $\boldsymbol{P}^{(n)}_{\theta;f_1}$ whenever $\underline{Q}_{f_1} > \chi^2_{1,1-\alpha}$,* [2] *is locally asymptotically most stringent, at asymptotic level α, for $\bigcup_{\sigma^2}\{\boldsymbol{P}^{(n)}_{\sigma^2,0;f_1}\}$ against $\bigcup_{\sigma^2\in\mathbb{R}_+^*}\bigcup_{b\in\mathbb{R}}\left\{\boldsymbol{P}^{(n)}_{\sigma^2,b;f_1}\right\}$;*

(**iii**) *The asymptotic power under $\boldsymbol{P}^{(n)}_{\theta^{(n)};f_1}$ is $1 - F(\chi^2_{1,1-\alpha}, \lambda_{f_1})$.* [3]

Proof:

(**i**) From Proposition 2.1, one can write

$$Q_{f_1}(\theta) = \Gamma_{f_1;22}^{-1}(\theta)\Delta_{f_1;2}^{(n)2}(\theta), \tag{3.3}$$

with

$$\Delta_{f_1;2}^{(n)}(\theta) = n^{\frac{-1}{2}}\sigma\sum_{i=1}^{n}\sum_{t=3}^{T}\Phi_{f_1}(Z_{i,t})Z_{i,t-1}Z_{i,t-2} = \sigma\underline{\Delta}_{f_1}^{(n)},$$

[2] $\chi^2_{1,1-\alpha}$ is the $(1-\alpha)$-quantile of the central chi-square distribution with one degree of freedom.

[3] $F(.,\lambda_{f_1})$ is the noncentral chi-square distribution function with one degree of freedom and noncentrality parameter λ_{f_1}.

where $\underline{\Delta}_{f_1}^{(n)} := n^{\frac{-1}{2}} \sum_{i=1}^{n} \sum_{t=3}^{T} \Phi_{f_1}(Z_{i,t}) Z_{i,t-1} Z_{i,t-2}$, then

$$Q_{f_1}(\theta) = \left[\sigma^2 (T-2) I(f_1) \sigma_{f_1}^4\right]^{-1} \left[\sigma \underline{\Delta}_{f_1}^{(n)}\right]^2 = \underline{\Delta}_{f_1}^{(n)^2} / ((T-2) I(f_1) \sigma_{f_1}^4) = \underline{Q}_{f_1}.$$

(**ii**) Under $\mathbf{P}_{\theta;f_1}^{(n)} : \underline{\Delta}_{f_1}^{(n)} \sim \mathcal{N}\big(0, (T-2) I(f_1) \sigma_{f_1}^4\big)$, then $\underline{\Delta}_{f_1}^{(n)^2} / ((T-2) I(f_1) \sigma_{f_1}^4) = \underline{Q}_{f_1} \sim \chi_1^2$.

Under $\mathbf{P}_{\theta^{(n)};f_1}^{(n)}$, from Le Cam's Third Lemma, we have

$$\underline{\Delta}_{f_1}^{(n)} \sim \mathcal{N}\big((T-2) I(f_1) \sigma \sigma_{f_1}^4 \tau_2, (T-2) I(f_1) \sigma_{f_1}^4\big),$$

hence $\underline{\Delta}_{f_1}^{(n)^2} / ((T-2) I(f_1) \sigma_{f_1}^4) = \underline{Q}_{f_1} \sim \chi_1^2(\lambda_{f_1})$: noncentral chi-square of one degree of freedom and non-centrality parameter

$$\lambda_{f_1} := \big(\sqrt{(T-2) I(f_1) \sigma_{f_1}^4} \sigma \tau_2\big)^2 = (T-2) I(f_1) \sigma^2 \sigma_{f_1}^4 \tau_2^2.$$

(**iii**) We know that the power of the test is defined by

$$1 - \beta \ := \ \text{Prob}\left[\text{rejecting} \ \ \mathcal{H}_f^{(n)}(\theta) \ / \ \mathcal{H}_f^{(n)}(\theta^{(n)})\right] = \ \text{Prob}\left[\underline{Q}_{f_1} > \chi_{1,1-\alpha}^2 \ / \ \tau_2 \neq 0\right]$$

where β is the second species risk and defined by

$$\text{Prob}\left[\underline{Q}_{f_1} < \chi_{1,1-\alpha}^2 \ / \ \tau_2 \neq 0\right] = F(\chi_{1,1-\alpha}^2, \lambda_{f_1}).$$

Hence, the power of the test is $1 - F(\chi_{1,1-\alpha}^2, \lambda_{f_1})$. □

The Gaussian versions of $\underline{Q}_{f_1}$ is

$$\underline{Q}_{\mathcal{N}} = \frac{a^3}{T-2} \left[n^{\frac{-1}{2}} \sum_{i=1}^{n} \sum_{t=3}^{T} Z_{i,t} Z_{i,t-1} Z_{i,t-2}\right]^2. \tag{3.4}$$

Unfortunately, this test statistic needs f_1 to be specified as a standardized Gaussian one, so the parameter a also has to be given. In the next, we will show that an appropriate version remains asymptotically valid under arbitrary f_1 with finite variance and optimal under Gaussian one (pseudo-Gaussian test).

3.2. Pseudo-Gaussian tests

The Gaussian central sequence $\Delta_{\mathcal{N};2}^{(n)}(\theta)$ allows obtaining asymptotically optimal tests under $f_1 = f_{\mathcal{N}}$, as well as efficient detection of panel bilinear models, in the parametric Gaussian model characterized by Gaussian disturbances. Extending the validity of the Gaussian optimal test to general densities g_1 in a broad class of densities is of course highly desirable.

Let us show that this is indeed possible and that a slight modification, $\Delta^{*(n)}_{\mathcal{N};2}$, say, of the efficient central sequence $\Delta^{(n)}_{\mathcal{N};2}$ leads to a *pseudo-Gaussian test* which remaining valid when the actual density g_1 belongs to the class $\mathcal{F}^{(2)}_A$ of all densities in $\mathcal{F}_A$ with *finite* variance. Define

$$\Delta^{*(n)}_{\mathcal{N};2}(\theta) = n^{\frac{-1}{2}}\sigma a \sum_{i=1}^{n}\sum_{t=3}^{T}(Z_{i,t}-m_1^{(n)})(Z_{i,t-1}-m_1^{(n)})(Z_{i,t-2}-m_1^{(n)}),$$

where $m_1^{(n)}=\dfrac{1}{nT}\displaystyle\sum_{i=1}^{n}\sum_{t=1}^{T}Z_{i,t}$ is a $\sqrt{n}$-consistent estimator, under $\mathbf{P}^{(n)}_{\theta;g_1}$, of $\mu_1(g_1) := \displaystyle\int_{\mathbb{R}} z g_1(z)dz$. Decomposing $Z_{i,t}-m_1^{(n)}$ into $(Z_{i,t}-\mu_1(g_1))+(\mu_1(g_1)-m_1^{(n)})$, then, it is easy to check that under $\mathbf{P}^{(n)}_{\theta;g_1}$, as $n\to\infty$,

$$\Delta^{*(n)}_{\mathcal{N};2}(\theta) = n^{\frac{-1}{2}}\sigma a \sum_{i=1}^{n}\sum_{t=3}^{T}(Z_{i,t}-\mu_1(g_1))(Z_{i,t-1}-\mu_1(g_1))(Z_{i,t-2}-\mu_1(g_1)) + o_p(1).$$

Then, still under $\mathbf{P}^{(n)}_{\theta;g_1}$, $\Delta^{*(n)}_{\mathcal{N};2}(\theta)$ is asymptotically normal with zero mean and variance

$$\Gamma^{*}_{\mathcal{N};g_1;22} = a^2\sigma^2(T-2)\sigma^6_{g_1},$$

where $\sigma^2_{g_1} := \displaystyle\int_{\mathbb{R}}(z-\mu_1(g_1))^2 g_1(z)dz$.

On the other hand, it is easy to see that, under $\mathbf{P}^{(n)}_{\theta^{(n)};g_1}$, $\Delta^{*(n)}_{\mathcal{N};2}(\theta)$ and the log-likelihood $\Lambda^{(n)}_{\theta^{(n)}/\theta;g_1}$ are jointly binormal; the desired result then follows from a routine application of Le Cam's Third Lemma.

A pseudo-Gaussian test may then be based on a statistic of the form

$$\begin{aligned} Q^{*}_{\mathcal{N};g_1}(\theta) &:= (\Gamma^{*}_{\mathcal{N};g_1;22}(\theta))^{-1}\Delta^{*(n)^2}_{\mathcal{N};2}(\theta) \\ &:= \frac{1}{(T-2)\sigma^6_{g_1}}\left[n^{\frac{-1}{2}}\sum_{i=1}^{n}\sum_{t=3}^{T}(Z_{i,t}-m_1^{(n)})(Z_{i,t-1}-m_1^{(n)})(Z_{i,t-2}-m_1^{(n)})\right]^2. \end{aligned} \tag{3.5}$$

In practice, the pseudo-Gaussian test will be based on

$$Q^{\dagger}_{\mathcal{N}} := \frac{1}{(T-2)s^6}\left[n^{\frac{-1}{2}}\sum_{i=1}^{n}\sum_{t=3}^{T}(Z_{i,t}-m_1^{(n)})(Z_{i,t-1}-m_1^{(n)})(Z_{i,t-2}-m_1^{(n)})\right]^2,$$

where $s^2 = \dfrac{1}{nT}\displaystyle\sum_{i=1}^{n}\sum_{t=1}^{T}(Z_{i,t}-m_1^{(n)})^2$ is the empirical variance of the $(Z_{i,t}-m_1^{(n)})$'s.

Showing that, under $\mathbf{P}^{(n)}_{\theta;g_1}$, $Q^{\dagger}_{\mathcal{N}} - Q^{*}_{\mathcal{N};g_1}(\theta) = o_p(1).$, as $n\to\infty$, we thus have the following result.

Proposition 3.2. *Let $g_1\in\mathcal{F}^{(2)}_A$. Then,*

(**i**) *$Q^{\dagger}_{\mathcal{N}}$ is asymptotically central chi-square with 1 degree of freedom under $\boldsymbol{P}^{(n)}_{\theta;g_1}$, and asymptotically noncentral chi-square, still with 1 degree of freedom and with noncentrality parameter $\lambda_{\mathcal{N}} := (T-2)\sigma^2_g\tau^2_2$ under $\boldsymbol{P}^{(n)}_{\theta^{(n)};g_1}$;*

(**ii**) *The sequence of tests rejecting the null hypothesis* $\bigcup_{g_1 \in \mathcal{F}_A^{(2)}} \bigcup_{\sigma \in \mathbb{R}_+^*} \left\{ \boldsymbol{P}^{(n)}_{\sigma^2,0;g_1} \right\}$ *whenever* $Q^{\dagger}_{\mathcal{N}} > \chi^2_{1,1-\alpha}$, *is locally asymptotically most stringent, at asymptotic level* α, *against alternatives of the form* $\bigcup_{\sigma \in \mathbb{R}_+^*} \bigcup_{b \in \mathbb{R}} \left\{ \boldsymbol{P}^{(n)}_{\sigma^2,b;f_{\mathcal{N}}} \right\}$;

(**iii**) *The asymptotic power under* $\boldsymbol{P}^{(n)}_{\theta^{(n)};g_1}$ *is* $1 - F(\chi^2_{1,1-\alpha}, \lambda_{\mathcal{N}})$.

4. OPTIMAL RANK TESTS

We start by describing the group invariance structure of the testing problem considered. Then we introduce (and study the properties of) rank-based versions of the central sequences. This will allow us to develop the resulting (optimal) rank tests and to derive their asymptotic properties. The general results of Hallin and Werker (2003) indicate that semiparametrically efficient and rank-based procedures have been established in relation with ranks that are being maximal invariants under model-generating groups of transformations. It is clearly that the null hypothesis $\mathcal{H}_0^{(n)}$ is invariant under the group $(\mathcal{G}^{(nT)}, \star)$, such as for any transformation $\mathcal{G}_h$ of $\mathbb{R}^{nT}$ we define $\mathcal{G}_h(Y_{11}, ..., Y_{nT}) := (h(Y_{11}), ..., h(Y_{nT}))$, where $y \mapsto h(y)$ is continuous and monotone increasing and $\lim_{y \to \pm\infty} h(y) = \pm\infty$. The invariance principle therefore suggests restricting to tests that are invariant with respect to this group. The maximal invariant associated with $(\mathcal{G}^{(nT)}, \star)$ is the rank $R^{(n)} := \left(R^{(n)}_{1,1}, ..., R^{(n)}_{n,T}\right)$, where $R^{(n)}_{i,t}$ denotes the rank of $Z^{(n)}_{i,t}$ among $\left(Z^{(n)}_{1,1}, ..., Z^{(n)}_{n,T}\right)$. It is easy to check that $(\mathcal{G}^{(nT)}, \star)$ is actually a generating group for the null hypothesis $\mathcal{H}_0^{(n)}$. As a direct corollary, rank tests are distribution-free under the whole null hypothesis. This explains why rank tests will be validity-robust.

4.1. Rank-based versions of central sequences

According to Hallin and Werker (2003) [21] and under the LAN property with efficient central sequence $\Delta^{(n)}_{f_1;2}$, an efficient semiparametric inference obtained conditioning $\Delta^{(n)}_{f_1;2}$ by the rank vector $R^{(n)}$, under the null hypothesis

$$\underset{\sim}{\Delta}{}^{(n)}_{f_1;2} := E\left[\Delta^{(n)}_{f_1;2} \mid R^{(n)}\right]. \tag{4.1}$$

The conditional definition (4.1) of $\Delta^{(n)}_{f_1;2}$ gives a statistic based on the ranks of exact scores, thus *Hájek's projection theorem* establishes the asymptotic equivalence between a non-parametric statistic and its parametric counterpart (for more details, consult the book of Hájek, Šidák and Sen (1999) [16]).

To combine validity-robustness/invariance with Le Cam optimality at density f_1, we introduce rank-based versions of the central sequence that appear in the LAN property above (Proposition 2.1).

$$\underset{\sim}{\Delta}{}^{(n)}_{f_1;2} := n^{\frac{-1}{2}} \sigma \sum_{i=1}^{n} \sum_{t=3}^{T} \left\{ \varphi_{f_1}\left(\frac{R^{(n)}_{i,t}}{N+1}\right) F_1^{-1}\left(\frac{R^{(n)}_{i,t-1}}{N+1}\right) F_1^{-1}\left(\frac{R^{(n)}_{i,t-2}}{N+1}\right) - \overline{m}_{f_1} \right\} \tag{4.2}$$

with $N = n(T-2)$, $\varphi_{f_1} := \Phi_{f_1} \circ F_1^{-1}$ and

$$\overline{m}_{f_1} := \frac{1}{N(N-1)(N-2)} \sum_{1 \leqslant t_1 \neq t_2 \neq t_3 \leqslant N} \sum \sum \varphi_{f_1}\big(\frac{t_1}{N+1}\big) F_1^{-1}\big(\frac{t_2}{N+1}\big) F_1^{-1}\big(\frac{t_3}{N+1}\big).$$

Let

$$\begin{aligned} s_{f_1}^{(n)^2} :=\ & \frac{1}{N(N-1)(N-2)} \sum_{1 \leqslant t_1 \neq t_2 \neq t_3 \leqslant N} \sum \sum \Big[\varphi_{f_1}\big(\frac{t_1}{N+1}\big) F_1^{-1}\big(\frac{t_2}{N+1}\big) F_1^{-1}\big(\frac{t_3}{N+1}\big)\Big]^2 \\ & + \frac{2}{N(N-1)(N-2)(N-3)} \\ & \quad \times \sum_{1 \leqslant t_1 \neq t_2 \neq t_3 \neq t_4 \leqslant N} \sum \sum \sum \varphi_{f_1}\big(\frac{t_1}{N+1}\big) \varphi_{f_1}\big(\frac{t_2}{N+1}\big) F_1^{-1}\big(\frac{t_2}{N+1}\big) \Big[F_1^{-1}\big(\frac{t_3}{N+1}\big)\Big]^2 F_1^{-1}\big(\frac{t_4}{N+1}\big) \\ & + \frac{2}{N(N-1)(N-2)(N-3)(N-4)} \\ & \quad \times \sum_{1 \leqslant t_1 \neq t_2 \neq t_3 \neq t_4 \neq t_5 \leqslant N} \sum \sum \sum \sum \varphi_{f_1}\big(\frac{t_1}{N+1}\big) F_1^{-1}\big(\frac{t_2}{N+1}\big) F_1^{-1}\big(\frac{t_3}{N+1}\big) \varphi_{f_1}\big(\frac{t_3}{N+1}\big) \\ & \qquad \times F_1^{-1}\big(\frac{t_4}{N+1}\big) F_1^{-1}\big(\frac{t_5}{N+1}\big) \\ & + \frac{N-5}{N(N-1)(N-2)(N-3)(N-4)(N-5)} \\ & \quad \times \sum_{1 \leqslant t_1 \neq t_2 \neq t_3 \neq t_4 \neq t_5 \neq t_6 \leqslant N} \sum \sum \sum \sum \varphi_{f_1}\big(\frac{t_1}{N+1}\big) F_1^{-1}\big(\frac{t_2}{N+1}\big) F_1^{-1}\big(\frac{t_3}{N+1}\big) \\ & \qquad \times \varphi_{f_1}\big(\frac{t_4}{N+1}\big) F_1^{-1}\big(\frac{t_5}{N+1}\big) F_1^{-1}\big(\frac{t_6}{N+1}\big) - N\overline{m}_{f_1}^2. \end{aligned}$$

Define the cross-information coefficients $\mathcal{I}(f_1, g_1)$ and $\sigma(f_1, g_1)$ as

$$\mathcal{I}(f_1, g_1) := \int_0^1 \Phi_{f_1}(F_1^{-1}(u)) \Phi_{g_1}(G_1^{-1}(u)) du \quad \text{and} \quad \sigma(f_1, g_1) := \int_0^1 F_1^{-1}(v) G_1^{-1}(v) dv,$$

we then have, for the rank-based $\underset{\sim}{\Delta}{}_{f_1;2}^{(n)}$, the following asymptotic representation result.

Proposition 4.1. *Let f_1 and $g_1 \in \mathcal{F}_A$. Then, as $n \to \infty$ and fixed T,*

(**i**) *Under $\boldsymbol{P}_{\theta;g_1}^{(n)}$,*

$$\begin{aligned} \underset{\sim}{\Delta}{}_{f_1;2}^{(n)} &:= E_{g_1}^{(n)}\big[\Delta_{f_1;2}^{(n)} \mid R_{1,1}^{(n)}, ..., R_{n,T}^{(n)}\big] + o_{L^2}(1) \\ &= \Delta_{f_1,g_1;2}^{*(n)} + o_{L^2}(1), \end{aligned} \tag{4.3}$$

with (denoting by G_1 the distribution function associated with g_1)

$$\Delta_{f_1,g_1;2}^{*(n)} := n^{\frac{-1}{2}} \sigma \sum_{i=1}^{n} \sum_{t=3}^{T} \varphi_{f_1}\big(G_1(Z_{i,t})\big) F_1^{-1}\big(G_1(Z_{i,t-1})\big) F_1^{-1}\big(G_1(Z_{i,t-2})\big); \tag{4.4}$$

(**ii**) *Still under $\boldsymbol{P}_{\theta;g_1}^{(n)}$, $\underset{\sim}{\Delta}{}_{f_1;2}^{(n)}$ has zero mean and variance $\Gamma_{f_1;22}^{*(n)} := \sigma^2(T-2) s_{f_1}^{(n)^2} = \Gamma_{f_1;22}^{*} + o(1)$, where $\Gamma_{f_1;22}^{*} := (T-2) I(f_1) \sigma^2 \sigma_{f_1}^4$;*

(**iii**) *$\Delta_{f_1,g_1;2}^{*(n)}$ is asymptotically normal, with zero mean under $\boldsymbol{P}_{\theta;g_1}^{(n)}$, mean $(T-2)\mathcal{I}(f_1, g_1)\sigma^2(f_1, g_1)\sigma^2\tau_2$ under $\boldsymbol{P}_{\theta^{(n)};g_1}^{(n)}$ and variance $\Gamma_{f_1;22}^{*}$ under both.*

Proof: The proof of Part (i) of the proposition follows along the same lines as in Hallin *et al.* (1985) [18], and therefore is omitted. Part (ii) is obtained by direct computation. As for Part (iii), under $\mathbf{P}^{(n)}_{\theta;g_1}$, the result straightforwardly follows from classical central limit theorem. On the other hand, it is easy to see that, still under $\mathbf{P}^{(n)}_{\theta^{(n)};g_1}$, $\Delta^{*(n)}_{f_1,g_1;2}$ and the log-likelihood $\Lambda^{(n)}_{\theta^{(n)}/\theta;g_1}$ are jointly binormal; the desired result then follows from a routine application of Le Cam's Third Lemma. □

4.2. Optimal rank tests

The rank-based version of the quadratic statistic is given by

$$\begin{aligned} \underset{\sim f_1}{Q} &:= (\Gamma^{*(n)}_{f_1;22})^{-1}\underset{\sim f_1;2}{\Delta}{}^{(n)2} \\ &= \frac{1}{(T-2)s^{(n)2}_{f_1}}\left[n^{\frac{-1}{2}}\sum_{i=1}^{n}\sum_{t=3}^{T}\left\{\varphi_{f_1}\Big(\frac{R^{(n)}_{i,t}}{N+1}\Big)F_1^{-1}\Big(\frac{R^{(n)}_{i,t-1}}{N+1}\Big)F_1^{-1}\Big(\frac{R^{(n)}_{i,t-2}}{N+1}\Big)-\overline{m}_{f_1}\right\}\right]^2, \end{aligned} \tag{4.5}$$

we then have the following general result.

Proposition 4.2. *Let f_1 and $g_1 \in \mathcal{F}_A$. Then, for any $\tau = (\tau_1, \tau_2)' \in \mathbb{R}^2$, as $n \to \infty$ and for all fixed T,*

(**i**) $\underset{\sim f_1}{Q}$ *is asymptotically central chi-square with* 1 *degree of freedom under* $\boldsymbol{P}^{(n)}_{\theta;g_1}$, *and asymptotically noncentral chi-square, still with* 1 *degree of freedom and with noncentrality parameter*

$$\lambda_{f_1,g_1} := (T-2)\mathcal{I}^2(f_1,g_1)\sigma^4(f_1,g_1)\sigma^2\tau_2^2/I(f_1)\sigma^4(f_1)$$

under $\boldsymbol{P}^{(n)}_{\theta^{(n)};g_1}$;

(**ii**) *The sequence of tests rejecting the null hypothesis* $\bigcup_{g_1\in\mathcal{F}_A}\bigcup_{\sigma^2}\left\{\boldsymbol{P}^{(n)}_{\sigma^2,0;g_1}\right\}$ *whenever* $\underset{\sim f_1}{Q} > \chi^2_{1,1-\alpha}$, *is locally asymptotically most stringent, at asymptotic level* α, *against alternatives of the form* $\bigcup_{\sigma\in\mathbb{R}^*_+}\bigcup_{b\in\mathbb{R}}\left\{\boldsymbol{P}^{(n)}_{\sigma^2,b;f_1}\right\}$;

(**iii**) *The asymptotic power under* $\boldsymbol{P}^{(n)}_{\theta^{(n)};f_1}$ *is* $1-F(\chi^2_{1,1-\alpha},\lambda_{f_1,g_1})$.

4.3. Examples of non-parametric statistics

The quadratic statistic $\underset{\sim f_1}{Q}$ is a non-parametric statistic that depends only on the determining of the score function f_1 and provides general form for the optimal rank tests of the null hypothesis of randomness.

The three most important particular cases for the rank test statistic presented are the *van der Waerden* (normal score), the *Wilcoxon* (logistic score) and the *Laplace* (double-exponential score) test statistics, which are respectively optimal at normal, logistic and double-exponential distributions.

(**i**) The ***van der Waerden's*** test statistic is given by

$$\underset{\sim vdW}{Q} := \frac{a^2}{(T-2)s^{(n)^2}_{f_{\mathcal{N}}}} \underset{\sim vdW}{\Delta}^{(n)^2},$$

with

$$\underset{\sim vdW}{\Delta}^{(n)} = n^{\frac{-1}{2}} \sum_{i=1}^{n} \sum_{t=3}^{T} \left\{ \Psi^{-1}\left(\frac{R^{(n)}_{i,t}}{N+1}\right) \Psi^{-1}\left(\frac{R^{(n)}_{i,t-1}}{N+1}\right) \Psi^{-1}\left(\frac{R^{(n)}_{i,t-2}}{N+1}\right) - \overline{m}_{vdW} \right\} \tag{4.6}$$

and

$$\overline{m}_{f_{\mathcal{N}}} = \frac{1}{N(N-1)(N-2)} \sum\sum\sum_{1 \leqslant t_1 \neq t_2 \neq t_3 \leqslant N} \Psi^{-1}\left(\frac{t_1}{N+1}\right) \Psi^{-1}\left(\frac{t_2}{N+1}\right) \Psi^{-1}\left(\frac{t_3}{N+1}\right),$$

where Ψ is the standard normal distribution function.

(**ii**) The ***Wilcoxon's*** test statistic is given by

$$\underset{\sim W}{Q} := \frac{1}{(T-2)bs^{(n)^2}_{l}} \underset{\sim W}{\Delta}^{(n)^2},$$

with

$$\underset{\sim W}{\Delta}^{(n)} = n^{\frac{-1}{2}} \sum_{i=1}^{n} \sum_{t=3}^{T} \left\{ \left(2\frac{R^{(n)}_{i,t}}{N+1} - 1\right) \log\left(\frac{R^{(n)}_{i,t-1}}{N+1-R^{(n)}_{i,t-1}}\right) \log\left(\frac{R^{(n)}_{i,t-2}}{N+1-R^{(n)}_{i,t-2}}\right) - \overline{m}_{l} \right\} \tag{4.7}$$

and

$$\overline{m}_{l} = \frac{1}{N(N-1)(N-2)} \sum\sum\sum_{1 \leqslant t_1 \neq t_2 \neq t_3 \leqslant N} \left(\frac{2t_1}{N+1} - 1\right) \log\left(\frac{t_2}{N+1-t_2}\right) \log\left(\frac{t_3}{N+1-t_3}\right).$$

(**iii**) The ***Laplace's*** test statistic is given by

$$\underset{\sim \mathcal{L}}{Q} := \frac{d^2}{(T-2)s^{(n)^2}_{\mathcal{D}e}} \underset{\sim \mathcal{L}}{\Delta}^{(n)^2},$$

with

$$\underset{\sim \mathcal{L}}{\Delta}^{(n)} = n^{\frac{-1}{2}} \sum_{i=1}^{n} \sum_{t=3}^{T} \left\{ sign\left(F_1^{-1}\left(\frac{R^{(n)}_{i,t}}{N+1}\right)\right) F_1^{-1}\left(\frac{R^{(n)}_{i,t-1}}{N+1}\right) F_1^{-1}\left(\frac{R^{(n)}_{i,t-2}}{N+1}\right) - \overline{m}_{\mathcal{D}e} \right\} \tag{4.8}$$

and

$$\overline{m}_{\mathcal{D}e} = \frac{1}{N(N-1)(N-2)} \sum\sum\sum_{1 \leqslant t_1 \neq t_2 \neq t_3 \leqslant N} sign\left(F_1^{-1}\left(\frac{t_1}{N+1}\right)\right) F_1^{-1}\left(\frac{t_2}{N+1}\right) F_1^{-1}\left(\frac{t_3}{N+1}\right),$$

where F_1 is the distribution function of the double-exponential and

$$F_1^{-1}(u) = \begin{cases} d\log(2u) & if \;\; 0 < u \leqslant \frac{1}{2} \\ -d\log(2-2u) & if \;\; \frac{1}{2} \leqslant u < 1. \end{cases}$$

5. ASYMPTOTIC RELATIVE EFFICIENCIES

In order to compare the performance of the parametric and non-parametric tests presented, we calculate the Asymptotic Relative Efficiencies (AREs) of rank based tests with respect to the Pseudo-Gaussian one. The results obtained are satisfactory. Hence, under $\mathbf{P}^{(n)}_{\theta^{(n)};g_1}$, for any g_1 and for different scores f_1, the asymptotic relative efficiencies of $\underset{\sim f_1}{Q}$ with respect to $\underline{Q}_{\mathcal{N}}$ are

$$ARE_{g_1}(\underset{\sim f_1}{Q}/\underline{Q}_{\mathcal{N}}) = \left(\frac{\lambda_{f_1,g_1}}{\lambda_{\mathcal{N}}}\right)^2 = \left(\frac{\mathcal{I}^2(f_1,g_1)\sigma^4(f_1,g_1)}{\sigma^2_{g_1}\sigma^4_{f_1}I(f_1)}\right)^2. \tag{5.1}$$

Table 1 gives the numerical values of (5.1) for $\underset{\sim f_1}{Q} = \underset{\sim vdW}{Q}, \underset{\sim W}{Q}, \underset{\sim La}{Q}, \underset{\sim t_5}{Q}, \underset{\sim s\mathcal{N}(2)}{Q}$ and $\underset{\sim st_5(2)}{Q}$ under densities g_1 that are normal, Logistic, Double exponential, Student-t_5, Skew-normal $s\mathcal{N}(2)$ and Skew-Student $st_5(2)$.

Note that for $f_1 = vdW$ these values are always greater than one, i.e., the van der Waerden test (vdW) always has an efficiency greater than or equal to one, the equality being realized only if the density underlying g is itself a Gaussian density ($\mathcal{N}$), which means that rank based tests are asymptotically more powerful than Gaussian tests (this result is proved in many cases, see for example, Chernoff and Savage (1958) [7] and Hallin (1993) [17] for ARMA models). Note also that each value is maximum in its corresponding column. Thus, at each of the densities, non-parametric tests perform better, compared to the Pseudo-Gaussian test.

Table 1: Asymptotic relative efficiencies of some rank tests compared to their Pseudo-Gaussian counterpart.

Scores f_1 \ Actual density g_1	$\mathcal{N}$	l	$\mathcal{D}e$	t_5	$s\mathcal{N}(2)$	$st_5(2)$
Van der Waerden	1.0000	1.1723	1.5244	1.3435	1.6328	1.7262
Wilcoxon	0.9347	1.2026	2.3421	1.5002	1.9782	1.7822
Laplace	0.4275	1.1337	4.0000	1.0349	1.5433	1.6889
Student-t_5	0.8160	1.1569	2.7812	1.5625	1.8922	1.9501
Skew-normal $s\mathcal{N}(2)$	0.9520	1.0989	1.5633	1.1490	2.2301	2.3301
Skew-Student $st_5(2)$	0.5179	0.9734	1.9331	1.2150	1.7325	3.0133

6. SIMULATION

To enhance the interpretation and validity of the theoretical results of the previous sections, we present a simulation experiment using R-programming. The purpose of this section is to evaluate the performance of the proposed tests, at asymptotic level $\alpha = 5\%$.

We simulated several $BLP(0,0,2,1)$ panel data described by

$$X_{i,t} = bX_{i,t-2}\varepsilon_{i,t-1} + \varepsilon_{i,t} \qquad i = 1,2,...,100, \quad t = 1,2,...,12, \tag{6.1}$$

where:

- $b = 0$ for null hypothesis, and $b = 0.05, 0.1, 0.15, 0.2$ for increasingly severe alternatives;
- The $(\varepsilon_{i,t})$'s are i.i.d. with a symmetric density - Gaussian $(\mathcal{N})$, logistic (l), double exponential $(\mathcal{D}e)$, Student with $\nu = 5$ degrees of freedom (t_5) — or with an asymmetric density — the skew-normal $s\mathcal{N}(\delta)$ and skew-Student $st_5(\delta)$ densities [4] (both with skewness parameter value $\delta = 2$).

We performed the simulations for $n = 100$ and $T = 12$. In each case we generated 2500 independent samples of size $N = n(T-2) = 1000$ from (6.1).

For each replication, we performed the following tests at asymptotic level $\alpha = 5\%$: the pseudo-Gaussian test based on $Q^{\dagger}_{\mathcal{N}}$, the van der Waerden test based on $Q^{\underset{\sim}{}}_{\sim vdW}$, the Wilcoxon test based on $Q^{\underset{\sim}{}}_{\sim W}$, the Laplace test based on $Q^{\underset{\sim}{}}_{\sim L}$, the rank tests based on Student with 5 degrees of freedom and data-driven skew-Student $st_{\hat{\nu}}(\hat{\delta})$ scores.

Rejection frequencies are reported in Table 2 and they amply confirm the excellent overall performances of our rank-based procedure with data-driven scores. It also appears from the skew-normal and skew-Student simulations that asymmetry significantly improves the superiority of rank tests over the pseudo-Gaussian one.

7. CONCLUSION

The problem of testing the null hypothesis of a randomness against *first-order superdiagonal panel model* $BLP(0,0,2,1)$ (in large n and small T) is considered for specified and unspecified error density. Optimal parametric and pseudo-Gaussian procedures are derived based on the Local Asymptotic Normality property. Moreover, the pseudo-Gaussian test appears to have quite poor performances under skewed and heavy-tailed distributions. Therefore a rank-based version of the test is considered. Particular cases such as van der Waerden, Wilcoxon, Laplace and data-driven scores are given. These tests exhibit remarkably high ARE values with respect to their pseudo-Gaussian counterpart. Simulations confirm the excellent overall performances of the proposed tests.

[4] See, for instance, Azzalini and Capitanio (2003) [2] for a definition of skew-normal and skew-Student densities.

Table 2: Rejection frequencies (out of 2500 replications), for $b = 0$ (null hypothesis) and various non-zero values of b (alternative hypotheses), with error density g_1 that is Gaussian ($\mathcal{N}$), logistic (l), double exponential ($\mathcal{D}e$), Student (t_5), skew-normal ($s\mathcal{N}(2)$) and skew-Student t5 ($st_5(2)$) of the pseudo-Gaussian and rank based (based on van der Waerden, Wilcoxon, Laplace, Student-t_5 and data-driven scores) procedures.

Underlying densities g_1	Test	b				
		0	0.05	0.1	0.15	0.2
Normal	Pseudo Gaussien	0.0520	0.2236	0.7224	0.9680	0.9996
	Van der Waerden	0.0512	0.2448	0.6844	0.9564	1.0000
	Wilcoxon	0.0508	0.2280	0.7400	0.9640	1.0000
	Laplace	0.0512	0.2160	0.6928	0.8840	0.9992
	Student-t_5	0.0496	0.2360	0.6560	0.9760	1.0000
	Data-Driven	0.0524	0.2800	0.7400	0.9760	1.0000
Logistic	Pseudo Gaussien	0.0464	0.2400	0.7144	0.9632	0.9992
	Van der Waerden	0.0488	0.2688	0.7204	0.9844	0.9996
	Wilcoxon	0.0520	0.3044	0.7880	0.9620	0.9980
	Laplace	0.0496	0.2960	0.7320	0.9840	0.9980
	Student-t_5	0.0560	0.2488	0.7640	0.9840	0.9996
	Data-Driven	0.0500	0.3240	0.8360	0.9920	1.0000
Double exponential	Pseudo Gaussien	0.0524	0.2236	0.6908	0.9544	0.9972
	Van der Waerden	0.0476	0.2324	0.7820	0.9956	0.9888
	Wilcoxon	0.0492	0.3720	0.8412	0.9884	0.9992
	Laplace	0.0520	0.4924	0.9080	0.9960	1.0000
	Student-t_5	0.0484	0.3920	0.8800	0.9920	1.0000
	Data-Driven	0.0480	0.3760	0.8760	0.9520	1.0000
Student-t_5	Pseudo Gaussien	0.0496	0.3248	0.8768	0.9932	0.9996
	Van der Waerden	0.0488	0.3044	0.8660	0.9924	1.0000
	Wilcoxon	0.0492	0.4964	0.9248	0.9732	0.9989
	Laplace	0.0488	0.4560	0.8840	0.9880	0.9996
	Student-t_5	0.0476	0.4640	0.9560	0.9960	1.0000
	Data-Driven	0.0540	0.4960	0.9720	1.0000	1.0000
Skew-normal $s\mathcal{N}(2)$	Pseudo Gaussien	0.0496	0.1264	0.4572	0.7900	0.9612
	Van der Waerden	0.0464	0.1328	0.4112	0.8084	0.9488
	Wilcoxon	0.0468	0.1440	0.4560	0.8240	0.9440
	Laplace	0.0492	0.2120	0.4824	0.7244	0.8680
	Student-t_5	0.0432	0.1760	0.4120	0.7360	0.9240
	Data-Driven	0.0460	0.2080	0.5360	0.8080	0.9400
Skew-Student $st_5(2)$	Pseudo Gaussien	0.0480	0.2240	0.6800	0.9392	0.9904
	Van der Waerden	0.0524	0.2368	0.7200	0.9240	0.9888
	Wilcoxon	0.0488	0.3120	0.7284	0.9688	0.9992
	Laplace	0.0540	0.3160	0.6800	0.9124	0.9640
	Student-t_5	0.0504	0.2840	0.7280	0.9440	0.9920
	Data-Driven	0.0484	0.3480	0.8360	0.9720	0.9960

ACKNOWLEDGMENTS

The authors wish to express their special thanks to the reviewers for their contribution to improve the quality of this paper.

REFERENCES

[1] Akharif, A. and Hallin, M. (2003). Efficient detection of random coefficients in autoregressive models, *Annals of Statistics*, **31**(2), 675–704.

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

[3] Benghabrit, Y. and Hallin, M. (1992). Optimal rank-based tests for against first-order superdiagonal bilinear dependence against diagonal bilinear dependence, *Journal of Statistical Planning and Inference*, **32**(1), 45–61.

[4] Bennala, N.; Hallin, M. and Paindaveine, D. (2012). Pseudo-Gaussian and rank-based optimal tests for random individual effects in large n small T panels, *Journal of Econometrics*, **170**, 50–67.

[5] Bouzaachane, K. (2006). *Modèles de séries chronologiques bilinéaires: estimations, algorithmes et applications*, PhD thesis, University of Mohammed V-Agdal, Morocco.

[6] Cassart, D.; Hallin, M. and Paindaveine, D. (2011). A class of optimal tests for symmetry based on local Edgeworth approximations, *Bernoulli*, **17**, 1063–1094.

[7] Chernoff, H. and Savage, I.R. (1958). Asymptotic normality and efficiency of certain nonparametric tests, *Annals of Mathematical Statistics*, **29**, 972–994.

[8] El Melhaoui, S. and Allal, J. (2006). Optimal rank-based detection of exponential component in autoregressive models, *Journal of Nonparametric Statistics*, **18**(7), 431–447.

[9] Fihri, M.; Mellouk, A. and Akharif, A. (2016). Rank and signed-rank tests for random coefficient regression model, *Journal of Statistics Applications and Probability*, **5**, 233–247.

[10] Ghosh, S. (1999). *Asymptotics, Nonparametrics, and Time Series*, CRC Press, 1st Edition.

[11] Grahn, T. (1995). A conditional least squares approach to bilinear time series, *Journal of Time Series Analysis*, **16**, 509–529.

[12] Granger, C.W.J. and Andersen, A.P. (1978). *An Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprecht, Gettingen.

[13] Guegan, D. (1981). Étude d'un modèle non linéaire, le modèle superdiagonal d'ordre 1, *C.R. Acad. Sci. Paris, Série 1, Math.*, **293**, 95–99.

[14] Guegan, D. (1988). *Modèles bilinéaires et polynomiaux de sèries chronologiques: étude probabiliste et analyse statistique*, Thèse d'État, Universitè Joseph Fourier, Grenoble.

[15] Guegan, D. and Pham, D.T. (1992). Power of the score test against bilinear time series models, *Statistica Sinica*, **2**, 157–170.

[16] Hájek, J.; Šidák, Z. and Sen, P.K. (1999). *Theory of Rank Tests*, Academic Press, New York – London, 2nd edition.

[17] Hallin, M. (1993). On the Pitman non-admissibility of correlogram-based methods, *Journal of Time Series Analysis*, **16**, 607–612.

[18] Hallin, M.; Ingenbleek, J.F. and Puri, M.L. (1985). Linear serial rank tests for randomness against ARMA alternatives, *The Annals of Statistics*, **13**, 1156–1181.

[19] Hallin, M. and Mèlard, G. (1988). Rank-Based Tests for Randomness against First-Order Serial Dependence, *Journal of the American Statistical Association*, **83**, 1117–1128.

[20] Hallin, M. and Werker, B.J.M. (1999). *Optimal testing for semiparametric AR models: from Gaussian Lagarange multipliers to autoregression rank scores and adaptive tests.* In "Asymptotics, Nonparametrics and Time Series" (S. Ghosh, Ed.), Dekker, New York, 295–350.

[21] Hallin, M. and Werker, B.J.M. (2003). Semi-parametric efficiency, distribution-freeness and invariance, *Bernoulli*, **9**(1), 137–165.

[22] Hwang, S.Y. and Basawa, I.V. (1993). Asymptotic optimal inference for a class of nonlinear time series models, *Stochastic Processes and their Applications*, **46**, 91–113.

[23] Le Cam, L.M. (1986). *Asymptotic Methods in Statistical Decision Theory*, Springer Verlag, New York.

[24] Le Cam, L.M. and Yang, G.L. (2000). *Asymptotics in Statistics: Some Basic Concepts*, Springer Verlag, New York, 2nd edition.

[25] Liu, J. and Brockwell, P.J. (1988). On the general bilinear time series model, *Journal of Applied Probability*, **25**, 553–564.

[26] Maravall, A. (1983). An application of nonlinear time series forecasting, *Journal of Business and Economic Statistics*, **1**, 66–74.

[27] Phan, T.D. and Tran, L.T. (1980). Quelques résultats sur les modèles bilinéaires de séries chronologiques, *C.R. Acad. Sci. Paris*, **290**, 335–338.

[28] Phan, T.D. and Tran, L.T. (1981). On the first-order bilinear time series model, *Journal of Applied Probability*, **18**, 617–627.

[29] Saikkonen, P. and Luukkonen, R. (1991). Power properties of a time series linearity test against some simple bilinear alternatives, *Statistica Sinica*, **1**, 453–464.

[30] Sesay, S.A.O. and Subba Rao, T. (1991). Difference equations for higher order moments and cumulant for the bilinear time series model $\mathrm{BL}(p, 0, p, 1)$, *Journal of Time Series Analysis*, **10**, 159–169.

[31] Subba Rao, T. (1981). On the theory of bilinear time series models, *Journal of the Royal Statistical Society, Series B*, **43**, 244–255.

[32] Subba Rao, T. and Gabr, M.M. (1984). *An Introduclion to Bispectral Analysis and Bilinear Time Series Models*, Lecture Notes in Statistics No. 24, Springer, Berlin.

[33] Swensen, A.R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend, *Journal of Multivariate Analysis*, **16**, 54–70.

[34] Tan, L. and Wang, L. (2015). The LASSO method for bilinear time series models, *Communications in Statistics – Simulation and Computation*, **45**, 1–11.

[35] Van der Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge.

[36] Weiss, A.A. (1986). ARCH and bilinear time series models: Comparison and combination, *Journal of Business and Economic Statistics*, **4**, 59–70.

REVSTAT – STATISTICAL JOURNAL

Background

Statistics Portugal (INE, I.P.), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of a scientific statistical journal called Revista de Estatística. The original language used in this publication was Portuguese and the idea behind it was to publish it, three times a year, containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided that the publication should also include papers in English. This step was taken to achieve a broader dissemination, and to encourage foreign contributors to submit their work for publication.

At the time, the Editorial Board was mainly comprised of Portuguese university professors. It is now comprised of international university faculties and this has been the first step aimed at changing the character of Revista de Estatística from a national to an international scientific journal.

We have also initiated a policy of publishing special volumes that may be thematic highlighting areas of interest or associated with scientific events in Statistics. For example, in 2001, a special issue of Revista de Estatística was published containing three volumes of extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

In 2003, the name of the Journal has been changed to REVSTAT - STATISTICAL JOURNAL, now fully published in English, with a prestigious international editorial board, aiming to become a reference scientific journal that promotes the dissemination of relevant research results in Statistics.

The editorial policy of REVSTAT Statistical Journal is mainly placed on the originality and importance of the research.

All articles consistent with REVSTAT aims and scope will undergo scientific evaluation by at least two reviewers, one from the Editorial Board and another external.

The only working language allowed is English.

Abstract and Indexing Services

The REVSTAT is covered by the following abstracting/indexing services:

- Current Index to Statistics
- Google Scholar
- Mathematical Reviews® (MathSciNet®)
- Science Citation Index Expanded
- Zentralblatt für Mathematic
- Scimago Journal & Country Rank
- Scopus

Instructions to Authors

Articles must be written in English and will be submitted according to the following guidelines:

The corresponding author sends the manuscript in PDF format to the Executive Editor (revstat@ine.pt) with the Subject "New Submission to REVSTAT"; a MS#REVSTAT reference will be assigned later.

Optionally, in a mail cover letter, authors are welcome to suggest one of the Editors or Associate Editors, whose opinion may be considered suitable to be taken into account.

The submitted manuscript should be original and not have been previously published nor about to be published elsewhere in any form or language, avoiding concerns about self-plagiarism'.

Content published in this journal is peer-reviewed (Single Blind).

All research articles will be refereed by at least two researchers, including one from the Editorial Board unless the submitted manuscript is judged unsuitable for REVSTAT or does not contain substantial methodological novelty, in which case is desk rejected.

Manuscripts should be typed only in black, in double-spacing, with a left margin of at least 3 cm, with numbered lines, and with less than 25 pages. Figures (minimum of 300dpi) will be reproduced online in colours, if produced this way; however, authors should take into account that the printed version is always in black and grey tones.

The first page should include the name, ORCID iD (optional), Institution, country, and mail-address of the author(s) and a summary of fewer than one hundred words, followed by a maximum of six keywords and the AMS 2000 subject classification.

Authors are encouraged to submit articles using LaTeX, in the REVSTAT style, which is available at the LaTeX2e MACROS webpage.

References about the format and other useful information on the submission are available in the LaTeX2e Templates page.

Acknowledgments of people, grants or funds should be placed in a short section before the References title page. Note that religious beliefs, ethnic background, citizenship and political orientations of the author(s) are not allowed in the text.

Supplementary files (in REVSTAT style) may be published online along with an article, containing data, programming code, extra figures, or extra proofs, etc; however, REVSTAT is not responsible for any supporting information supplied by the author(s).

Any contact with REVSTAT must always contain the assigned REVSTAT reference number.

Accepted papers

Authors of accepted papers are requested to provide the LaTex files to the Secretary of the REVSTAT revstat@ine.pt. The authors should also mention if figure files were included, and submit electronic figures separately in .gif, .jpg, .png or .pdf format. Figures must be a minimum of 300dpi.

Copyright and reprints

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, Statistics Portugal, in order to ensure the widest possible dissemination of information, namely through the Statistics Portugal website (http://www.ine.pt).

After assigning copyright, authors may use their own material in other publications provided that REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.

Editorial Board