ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Volume 16, No.1
January 2018

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
Semestral. - Continuação de : Revista de Estatística = ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

CREDITS

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: + 351 21 842 61 00
    Fax: + 351 21 845 40 84
    Web site: http://www.ine.pt
    Customer Support Service
    + 351 218 440 695
- **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*

INDEX

Statistical Properties and Sensitivity of a New Adaptive Sampling Method for Quality Control

Manuel do Carmo, Paulo Infante and *Jorge M. Mendes* 1

Finding the Optimal Threshold of a Parametric ROC Curve under a Continuous Diagnostic Measurement

Yi-Ting Hwang, Yu-Han Hung, Chun Chao Wang and *Harn-Jing Terng* 23

On the q-Generalized Extreme Value Distribution

Serge B. Provost, Abdus Saboor, Gauss M. Cordeiro and *Muhammad Mansoor* 45

Modelling Spatially Sampled Proportion Processes

Iosu Paradinas, Maria Grazia Pennino, Antonio López-Quílez, Marcial Marín, José María Bellido and *David Conesa* 71

Weibull Lindley distribution

A. Asgharzadeh, S. Nadarajah and *F. Sharafi* 87

Heuristic Tools for the Estimation of the Extremal Index: A Comparison of Methods

Marta Ferreira 115

Computationally Efficient Goodness-of-Fit Tests for the Error Distribution in Nonparametric Regression

G.I. Rivas-Martínez and *M.D. Jiménez-Gamero* 137

Abstracted/indexed in: *Current Index to Statistics, DOAJ, Google Scholar, Journal Citation Reports/Science Edition, Mathematical Reviews, Science Citation Index Expanded®, SCOPUS* and *Zentralblatt für Mathematic.*

REVSTAT – Statistical Journal
Volume 16, Number 1, January 2018, 1–22

STATISTICAL PROPERTIES AND SENSITIVITY OF A NEW ADAPTIVE SAMPLING METHOD FOR QUALITY CONTROL

Authors: Manuel do Carmo
– Universidade Europeia, Lisbon
and CIMA-UE, University of Évora, Portugal
manuel.carmo@universidadeeuropeia.pt

Paulo Infante
– ECT/DMAT and CIMA-UE, University of Évora,
Évora, Portugal
pinfante@uevora.pt

Jorge M. Mendes
– NOVA IMS, Universidade Nova de Lisboa,
Lisbon, Portugal
jmm@novaims.unl.pt

Received: November 2013 Revised: February 2016 Accepted: June 2016

Abstract:

- We present a new adaptive sampling method for statistical quality control. In this method, called LSI (***L**aplace* **s**ampling **i**ntervals), we use the probability distribution function of the *Laplace* standard distribution to obtain the sampling instants, depending on a k parameter that allows control of sampling costs. Several algebraic expressions concerning the statistical properties of the LSI method are presented. We compare the LSI method with fixed sampling intervals (FSI) and variable sampling intervals (VSI) methods using a *Shewhart* X-bar control chart and evaluate the sensitivity of these sampling methods when the lower sampling interval is truncated. The results obtained show that the new method is a viable alternative in various critical contexts and situations.

Key-Words:

- *adaptive sampling; sensitivity; simulation; Laplace sampling intervals.*

AMS Subject Classification:

- 62N10, 62P30.

1. INTRODUCTION

The success of a statistical quality control method is directly related to the type of control chart and especially to the sampling method used. The variability of the process is due to random causes (inherent to the process) or to the presence of assignable causes. The former cannot be economically identified and corrected, whereas the latter should be detected and eliminated. The choice of the control chart depends upon the characteristic being controlled. The quantitative characteristics are controlled using variable control charts ($\bar{X}$-charts, R-charts, or s-charts, for example) or special control charts (EWMA or CUSUM charts for continuous random variables, for example). For a long time the control charts used had fixed parameters (sampling intervals, sample sizes, and control limits). However, since final of the 1980s, new adaptive control charts have been developed for improved performance. In terms of their implementation, these charts can be classified in two broad categories. The first category encompasses control charts with adaptive parameters (sampling intervals, sample size, and control limits, depending on the sample information; see, for example, Reynolds *et al.* (1988), Daudin (1992), Prabhu *et al.* (1993), Costa (1994), Prabhu *et al.* (1994), Stoumbos & Reynolds (1997), Costa (1999), Rodrigues Dias (1999), Carot *et al.* (2002), Mahadik & Shirke (2009)). The second category encompasses control charts with predetermined parameters (parameters determined before the beginning of the process to be controlled; see for example, Banerjee & Rahim (1988), Rahim & Banerjee (1993), Lin & Chou (2005) and Rodrigues Dias & Infante (2008)).

Several measures have been developed to assess the statistical quality controls method's performance across time regarding to how quickly they detect assignable causes. The frequency of false alarms and the number of samples and analysed items are two examples. The ARL ("*average run length*") is perhaps the most widely used statistical measure for assessing the performance of a statistical control chart. The ARL is defined as the average number of samples that needs to be drawn before an out-of-control indication is given. If the control methods have constant and equal sampling intervals, then the time interval up to the detection of a change is directly proportional to the ARL. In the case of non-constant sampling intervals, the proportionality above fails and the ARL is not a measure of the efficiency of the control method. The $AATS$ ("*adjusted average time to signal*"), also known in the literature as "*steady-state performance*", is defined as the average interval of time from the instant at which a failure occurs in the system to the instant at which the control chart detects the failure. In the case of a *Shewhart* control chart with variable sampling intervals, $AATS = E(G) + E(D) \times (ARL - 1)$, where $E(D)$ is the average sampling interval and G represents the time interval between the instant at which the system fails and the instant at which the first sample, after the failure, is drawn. The $AATS$

is a measure that suits most practical situations. Morais (2002), Carmo (2004) and Rodrigues Dias & Carmo (2009) are important sources on the previously described approaches. In Morais & Pacheco (2001), stochastic order relations are established using the *RL* ("*run length*"), allowing comparison of different quality control methods without numerical computation of their performance.

In the following sections, we present a new sampling method called LSI ("*Laplace sampling intervals*"), an adaptive and continuous sampling method in which the sampling intervals are obtained on the basis of the probability density function of the *Laplace* standard distribution and depends on a scale parameter, k. The *AATS* will be used in section 3 to examine the statistical properties of this method and to compare its effectiveness with that of the FSI ("*fixed sampling intervals*") and VSI (*variable sampling intervals*) methods. In section 4, the sensitivity of this new method is compared to the sensitivity of the above mentioned methods. Finally, in section 5, conclusions are drawn and future work is proposed.

2. NEW SAMPLING METHOD: LSI (*LAPLACE* SAMPLING INTERVALS)

2.1. Methodology

Let X be a continuous quality variable so that when the system is in control state, X is a random variable with expected value $\mu = \mu_0$ and standard deviation $\sigma = \sigma_0$. If $x_1, x_2, ..., x_n$ are identical and independently distributed random variables with the same distribution of X, where n is the sample size, then $\bar{X}$ has the same expected value μ_0 and standard deviation $\sigma_0/\sqrt{n}$. As a consequence of the one assignable cause, corresponding to a failure of the system, the process state may change and then μ and σ may assume new values $\mu_1 = \mu_0 \pm \lambda\sigma_0$, and $\sigma_1 = \sigma_0$, with $\lambda > 0$. If t_i denotes a sampling instant of order i and $\overline{x}_i$ is the sample mean value of order i, according to the LSI method, the next sampling at the instant of order $i + 1$ is given by

$$t_{i+1} = t_i + k\,.\,l(u_i)\,, \qquad i = 0, 1, 2, ...\,, \tag{2.1}$$

where $u_i = \frac{\overline{x}_i - \mu_0}{\sigma_0}\sqrt{n}$, $t_0 = 0$, $\overline{x}_0 = \mu_0$, $l(u_i) = \frac{1}{2}\,e^{-|u_i|}$, n is the sample size, k is a convenient scale constant and $l(\cdot)$ is the density function of the standard *Laplace* variable. Therefore, according to (2.1), this sampling method considers consecutive sampling intervals $\delta_i = t_i - t_{i-1} = k\,.\,l(u_{i-1}) = k \times 0.5 \times e^{-|u_{i-1}|}$, $i = 1, 2, 3, ...$ These are values from independently and identically distributed continuous random variables D_i, $i = 1, 2, 3, ...$, with the same distribution of a generic variable D. When we obtain the value of k we have only sampling intervals, D_i, under control

$(E(D|\lambda=0)=1)$; when we obtain the values of the $AATS$ we have only sampling intervals out-of-control, subject to different shifts of the sample mean. Thus, D is a function of $\bar{X}$ and, consequently, of U, given by

$$D = \frac{k.e^{-\left|\frac{\bar{X}-\mu}{\sigma}\sqrt{n}\right|}}{2} = k.l(U)\ . \tag{2.2}$$

The constant k depends on several factors and, especially, on the costs associated with the production process (not imposing, so far, any limits on the control chart for means) and $U = \left(\bar{X}-\mu\right)\sqrt{n}/\sigma$. Using this adaptive and continuous method, the sampling frequency decreases (the sampling instants are spaced further apart in time) when the sample mean is marked close to the mean of the distribution. When the sample mean is marked close to control limits, the probability of a shift in the mean increases, and the sampling frequency increases (the sampling instants are less distant in time). Like the VSI sampling method, the LSI method is an adaptive method in which the time interval to the next sample depends on the information in the current sample. The disadvantage is that the sampling intervals function of the LSI chart is a continuous function of the chart statistic (and this implies an infinite number of possible sampling intervals). However, the sampling interval function is a very simple function of the chart statistic. It can be easy to implement in practice, particularly, in automatic monitoring. The NSI (*normal sampling intervals*) method, presented by Rodrigues Dias (1999) and studied in Infante (2004), showed limitations in practical applications. In this method the sampling instants are obtained using the density function of the standard Normal distribution; the smallest sampling interval is very small, which reduces the application of this sampling method. The idea emerged to study one analogous method in which the smallest sampling interval would be greater than in the NSI method which, therefore, allowed practical applications. Regarding the skewness and shape, the *Laplace* density function is similar to the Normal density function, as in the *Cauchy* density function, but having heavier tails. This fact addresses some of the difficulties seen in the practical application of the NSI method. In our preliminary work we simulated sampling intervals for three probability density functions ($pdf(x,\mu,\sigma)$): those of the Normal, *Cauchy*, and *Laplace* distributions. Considering "*3-sigma*" control limits and a time unit average sampling interval, in control, the following results were obtained:

a) Normal distribution: $pdf(0,0,1) = 0.399$, $k = 3.535$, smallest sampling interval = 0.016, largest sampling interval = 1.410.

b) *Cauchy* distribution: $pdf(0,$ not defined, not defined$) = 0.318$, $k = 4.778$, smallest sampling interval = 0.152, largest sampling interval = 1.521.

c) *Laplace* distribution: $pdf(0,0,1) = 0.500$, $k = 3.813$, smallest sampling interval = 0.095, largest sampling interval = 1.907.

Based on these results, we selected the *Laplace* distribution's probability density function. All the parameters are defined such that the sampling frequency

decreases close to the central region and the smallest sampling interval is more likely to apply in practice. In addition, the smallest and largest sampling intervals are approximately equal to the sampling pair most frequently used in the VSI method ($(d_1, d_2) = (0.1, 1.9)$). We are considering general sampling interval functions that are continuous functions of the chart statistic. Stoumbos *et al.* (2001) study what function would be the optimal function in some sense.

2.2. Statistical Properties

In the remainder of this paper we assume that X follows a normal distribution with expected value $\mu = \mu_0$ and standard deviation $\sigma = \sigma_0$. We will consider a *Shewhart* chart with LCL and UCL, respectively, lower and upper control limits, given by:

$$LCL = \mu_0 - L\frac{\sigma_0}{\sqrt{n}}, \qquad UCL = \mu_0 + L\frac{\sigma_0}{\sqrt{n}}, \tag{2.3}$$

where L is the coefficient of the control limits (in practice, typically around three units of standard deviation). As mentioned above, after shift, μ takes on the new value $\mu_1 = \mu_0 \pm \lambda\sigma_0$, where $\lambda > 0$ is the magnitude of the mean shift (in the present work, only mean shifts are considered). Therefore, if u_i denotes the standard sample mean, for values to $|u_i| > L$ the process is considered to be out-of-control, although this might be a false alarm.

Considering the assumptions in (2.2) and (2.3), and that $f^*(\overline{x})$ is the corresponding conditional density function of $\overline{x}$, given by

$$f^*(\overline{x}) = \frac{\sqrt{n}}{\beta\sigma\sqrt{2\pi}}\, e^{-\frac{n(\overline{x}-\mu)^2}{2\sigma^2}}, \qquad \overline{x} \in \,]LCL, UCL[\,, \tag{2.4}$$

then $f^*(\overline{x})\, d\overline{x}$ is the elementary probability of $\overline{x} \in \,]\overline{x}, \overline{x} + d\overline{x}[$ and the average sampling interval is given by

$$\begin{aligned} E\,(D|\lambda, n, L) &= \int_{LCL}^{UCL} k.l(\overline{x}).f^*(\overline{x})\, d\overline{x} \\ &= \int_{LCL}^{UCL} \frac{k.\sqrt{n}}{2\beta\sigma_0\sqrt{2\pi}}\, e^{-\left[\left|\frac{\overline{x}-\mu_0}{\sigma_0}\sqrt{n}\right| + \frac{n(\overline{x}-\mu_0-\lambda\sigma_0)^2}{2\sigma_0^2}\right]} d\overline{x}\,, \end{aligned} \tag{2.5}$$

where β is the probability of the sample mean lies between the control limits, and is given by

$$\beta = \Phi\big(L - \lambda\sqrt{n}\big) - \Phi\big(-L - \lambda\sqrt{n}\big)\,. \tag{2.6}$$

Considering $\Phi(u)$ as the distribution function of the standard normal random variable, the following expression for the average sampling interval arises

$$E\,(D|\lambda, n, L) = \frac{k\sqrt{e}}{2\beta}\left[e^{\lambda\sqrt{n}}.A(L, \lambda, n) + e^{-\lambda\sqrt{n}}.B(L, \lambda, n)\right], \tag{2.7}$$

where

$$
\begin{aligned}
A(L,\lambda,n) &= \Phi\big(-1-\lambda\sqrt{n}\big) - \Phi\big(-L-1-\lambda\sqrt{n}\big)\,, \\
B(L,\lambda,n) &= \Phi\big(L+1-\lambda\sqrt{n}\big) - \Phi\big(1-\lambda\sqrt{n}\big)\,.
\end{aligned} \tag{2.8}
$$

The expression (2.7) depends on the sample size, n, the coefficient of the control limits, L, the mean shifts, λ, and β (the probability of a Type II error if the sample mean is out of the control limits). Assuming that the values of n, L and λ are known, then E(D) is a linear function of k. When the process is in control, $\lambda = 0$, the average sampling interval is given by

$$
E\,(D|L) \;=\; \frac{k\sqrt{e}}{\beta}\left[\Phi(L+1) - \Phi(1)\right], \tag{2.9}
$$

where $\beta = 2\Phi(L) - 1$ and does not depend on the sample size, n. Therefore, if the average sampling interval is equal to a time unit (without loss of generality, the sampling period used in the FSI method), the constant k is given by

$$
k = \frac{\beta}{\sqrt{e}\left[\Phi(L+1) - \Phi(1)\right]}\,, \tag{2.10}
$$

which is equal to 3.8134, based on the usual "*3-sigma*" limits. This result was obtained by numerical integration using the R software.

The variance of the sampling intervals can be obtained by the equality $\mathrm{Var}(D) = E(D^2) - [E(D)]^2$. The expression for $E(D^2)$ is obtained using the same reasoning applied to derive (2.5), leading to

$$
\begin{aligned}
E\left(D^2|\lambda,nL\right) &= \int_{LCL}^{UCL} \big[k\,.\,l(\overline{x})\big]^2 . f^*(\overline{x})\,d\overline{x} \\
&= \frac{e^2k^2}{4\beta}\left[e^{2\lambda\sqrt{n}}\left[\Phi\big(-2-\lambda\sqrt{n}\big) - \Phi\big(-L-2-\lambda\sqrt{n}\big)\right]\right. \\
&\quad \left. +\; e^{-2\lambda\sqrt{n}}\left[\Phi\big(L+2-\lambda\sqrt{n}\big) - \Phi\big(2-\lambda\sqrt{n}\big)\right]\right],
\end{aligned} \tag{2.11}
$$

which allows us to obtain the desired variance.

As mentioned above, there are different measures that are commonly used to assess the effectiveness of control charts. In this study we use the *AATS* to compare the effectiveness of the LSI method with the effectiveness of the FSI and VSI methods. Let *RL* (*run length*) be a random variable denoting the number of samples to be drawn before a false alarm or a failure occurrence, regardless of the sampling method used. *RL* follows a geometric distribution with the parameter $1-\beta$, that is, with a mean and the variance, respectively, given by

$$
ARL(\lambda) = \frac{1}{1-\beta}\,, \tag{2.12}
$$

and

$$\mathrm{Var}[RL(\lambda)] = \frac{\beta}{(1-\beta)^2} \,. \tag{2.13}$$

In general, a process starts in control. Therefore, the time interval between occurrence of a failure and its detection is of particular importance. For example, in a production process in which the malfunction costs are high, the average total cost of a production cycle may increase. As the failure may occur in the interval between two samples, it is necessary to adjust the ATS (*average time to signal* — which is defined as the average interval of time between the beginning of the process and an out-of-control sign, eventually a false alarm, being given by the control chart). Thus, we consider G to be the time interval between the occurrence of a failure and the moment when the first sample is drawn after the mean shift. The $AATS$ (*adjusted average time to signal*) is given by

$$AATS = E(G) + (ARL - 1).E(D) \,, \tag{2.14}$$

where the expected value of G has to be determined. In the FSI method, the expected value of G is, approximately, half of the inspection period used. However, in this adaptive case, we do not have a constant sampling interval. The distribution of the variable G depends on when the shift of the mean occurs. Let us assume that the time when a shift occurs is uniformly distributed in each sampling interval. If a failure occurs in a sampling interval of length d, the average time until the next sample is drawn is $0.5 \times d$. Although the number of sampling intervals is infinite, we can assume that the probability of the shift occurring in a sampling interval of length d is proportional to the product of the length of the interval and the probability of selecting this interval, as long as the process is in control, as Reynolds *et al.* (1988) and Runger & Pignatiello (1991) assumed for the VSI method. Taking into account that the variable G is continuous, the expression for its expected value can be obtained using the same reasoning that Reynolds *et al.* (1988) used in the VSI case. Based on the assumptions stated above, we obtain the following expression for the expected value of G

$$E\,(G|L) = \frac{E\left(D^2|\lambda=0\right)}{2\,E\,(D|\lambda=0)} = \frac{k.e^{3/2}}{4}\,\frac{\Phi(L+2)-\Phi(2)}{\Phi(L+1)-\Phi(1)} \,, \tag{2.15}$$

which can be written as

$$E(G) = k.e^{3/2} \times C(L) \,, \tag{2.16}$$

with

$$C(L) = \frac{\Phi(L+2)-\Phi(2)}{4 \times \left[\Phi(L+1)-\Phi(1)\right]} \,. \tag{2.17}$$

Expression (2.17) depends only on the control limits and may be simplified to

$$E(G) = 0.036 \times k e^{3/2} \,. \tag{2.18}$$

This simplified expression will be useful in future algebraic treatments. This simplified version is originated by the data in Table 1, containing approximations for $C(L)$ for several values of L. It is clear that 0.036 is an excellent approximation of $C(L)$, particularly for values of $L \geq 2$. This approximation is not as good as one might expect for $L < 2$. However, this situation can be considered irrelevant in many applications, as it results in a high number of false alarms.

Table 1: Values of $C(L)$ for different multiples L of the standard deviation.

L	1	1.5	2	2.5	3	3.5	4	4.5	5
$C(L)$	0.0394	0.0369	0.0361	0.0359	0.0359	0.0358	0.0358	0.0358	0.0358

The values of the $AATS$ can be obtained as

$$AATS_{\text{LSI}} = 0.036 \times k \times e^{3/2} + \left(\frac{\beta}{1-\beta}\right) \times E(D)\,. \tag{2.19}$$

In this case, the distribution of the sampling interval D_i is the conditional distribution of the sample mean given that the process is out-of-control. $D_1, D_2, ...$ are independent of RL, and the variance of TS (*time to signal*) can be written as

$$\text{Var}(TS) = \text{Var}(G) + E(RL-1)\,\text{Var}(D) + \text{Var}(RL-1)\,[E(D)]^2\,, \tag{2.20}$$

for which we need the value of $\text{Var}(G)$. To get $\text{Var}(G)$, we begin by determining $E(G^2)$. According to Reynolds *et al.* (1988), for the VSI method, and Infante (2004), for the NSI method, the algebraic expression is given by

$$E(G^2) = \frac{E(D^3|\lambda=0)}{3E(D|\lambda=0)} = \frac{k^2 e^4}{12}\,\frac{\Phi(L+3)-\Phi(3)}{\Phi(L+1)-\Phi(1)}\,, \tag{2.21}$$

which depends only on L. Therefore, the variance of the variable G is given by

$$\begin{aligned} \text{Var}(G) &= E(G^2) - [E(G)]^2 \\ &= \frac{k^2 e^4}{12}\,\frac{\Phi(L+3)-\Phi(3)}{\Phi(L+1)-\Phi(1)} - \frac{k^2 e^3}{16}\left[\frac{\Phi(L+2)-\Phi(2)}{\Phi(L+1)-\Phi(1)}\right]^2. \end{aligned} \tag{2.22}$$

From (2.7), (2.11), (2.21) and (2.22) we obtain (2.20).

3. COMPARISONS BETWEEN THE LSI METHOD AND THE FSI AND VSI METHODS

As mentioned in the previous section, comparisons of the effectiveness of the LSI sampling method with the FSI and VSI methods is made using the $AATS$.

Thus, the two sampling methods in comparison are considered to be both in control, or in other words, the average sampling intervals are equal to one time unity ($d = 1$) and the control limits are "*3-sigma*" ($L = 3$). Comparisons are made for mean shifts, only. Because it is assumed that the characteristic X follows a normal distribution, the direction of the shift is of no importance at all. Under these assumptions the value of the parameter k in the LSI method is 3.8134.

3.1. Comparison between the LSI and FSI sampling methods

Assuming a fixed value for the sampling interval, d, the expected value of G (the random variable previously defined for the FSI sampling method) can be defined as half of the sampling interval, d. Infante & Rodrigues Dias (2002) and Carmo (2004), in independent studies, analysed this approximation for different lifetimes, and both concluded this approximation to be acceptable. Therefore, the $AATS$ of the fixed sampling method is given by

$$AATS_{FSI} = E(G) + (ARL - 1) \times d \cong \frac{d}{1-\beta} - \frac{d}{2}\,. \tag{3.1}$$

To compare the effectiveness of the two sampling methods, LSI and FSI, we assume that, in control, the average sampling interval of the LSI method is equal to the sampling interval of the FSI method (without loss of the generality, $d = 1$), obtaining a value of $k = 3.8134$ for the LSI method. Considering (2.19) and (3.1) the ratio

$$Q_{\mathrm{LSI/FSI}} = \frac{AATS_{\mathrm{FSI}} - AATS_{\mathrm{LSI}}}{AATS_{\mathrm{FSI}}} \times 100\% \tag{3.2}$$

represents a measure of the relative variation, in %, of the $AATS$ value when $AATS_{\mathrm{FSI}}$ is the reference. The results obtained for mean shifts with different sample sizes are illustrated in Figure 1.

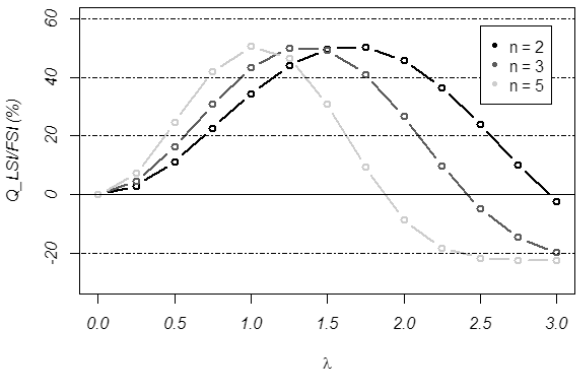


Figure 1: $Q_{LSI/FSI}(\%)$, as a function of λ and different values of n.

From Figure 1, the following conclusions arise:

i) The control chart for means with the LSI method detects small and moderate mean shifts more quickly than the control chart for means with the FSI method. This means that the LSI method is more sensitive to changes whose probability of detection is low. The reductions in the $AATS$ using the LSI method can be very large.

ii) For shifts in which the probability of the detection is high, the FSI method performs better than the LSI method. This is not surprising, because this is true for most known adaptive methods. In the situations described above, the average number of samples taken before detection of a failure is very small. Therefore, the value of $E(G)$ is of great importance. It is approximately equal to the time of system malfunction because only a single sample is required to detect the shift. For an average sampling interval, in control, equal to unity, $E(G) \cong 0.50$ in the FSI method and $E(G) \cong 0.61$ in the LSI method. However, the reduction obtained with FSI method, in terms of the $AATS$, is limited to a maximum of 22.5% (for $n = 5$), whereas the reduction obtained with LSI method has a maximum of 50.3%.

iii) For the different sample sizes considered, the $Q_{\mathrm{LSI/FSI}}$ values begin with an average rate of positive variation, reaching an absolute value maximum, and then reaching an average rate of negative variation. The average rate of positive variation increases more quickly as a function of λ when the sample size increases. The average rate of negative variation increases more quickly as a function of λ when the sample size decreases.

iv) In general, when the sample size increases, the values that maximize (λ) and the reductions obtained with the LSI method decrease. This makes sense because the probability of detection of the shift increases with the sample size.

3.2. Comparison between the LSI and VSI sampling methods

Looking for improvements in the performance of classical control charts, Reynolds *et al.* (1988) divided the region of continuation, $C =]-L, L[$, into two sub-regions, $C_1 =]-L, -w] \cup [w, L[$ and $C_2 =]-w, w[$, and used two sampling intervals, d_1 and d_2 , with $d_1 < d < d_2$. The VSI method allows us to anticipate the next sample (we use d_1 if the sample mean belongs to the C_1 region) or to delay it (using d_2 if the sample mean belongs to the C_2 region). Reynolds & Arnold (1989), Reynolds (1989), Runger & Pignatiello (1991), and Reynolds (1995), in different contexts, gave theoretical justifications for the use of two

sampling intervals. For two intervals the average sampling interval in the VSI method is given by:

$$E\,(D|\lambda,n) \;=\; \frac{d_1\times p_{11}+d_2\times p_{12}}{\beta}\;, \tag{3.3}$$

where β is given by (2.6), and

$$\begin{aligned} p_{11} \;&=\; \Phi\big(L-\lambda\sqrt{n}\big)-\Phi\big(w-\lambda\sqrt{n}\big)+\Phi\big(-w-\lambda\sqrt{n}\big)-\Phi\big(-L-\lambda\sqrt{n}\big)\,,\\ p_{12} \;&=\; \Phi\big(w-\lambda\sqrt{n}\big)-\Phi\big(-w-\lambda\sqrt{n}\big)\,, \end{aligned} \tag{3.4}$$

are the probabilities of a sample mean occurring in regions C_1 and C_2, respectively, when a mean shift occurs. W is given by

$$W \;=\; \Phi^{-1}\left[\frac{2\Phi(L)\times(d-d_1)+d_2-d}{2\,(d_2-d_1)}\right], \tag{3.5}$$

according to the expression presented by Runger & Pignatiello (1991), when the average sampling interval in the VSI method, in control, is equal to the sampling period, d, in the FSI method. According to Reynolds *et al.* (1988), the average time interval between the instant when a failure occurs and the instant when the first sample is drawn after the shift occurs is given by

$$E(G) \;=\; \frac{d_1^2\,p_{01}+d_2^2\,p_{02}}{2\,(d_1\,p_{01}+d_2\,p_{02})}\;. \tag{3.6}$$

The adjusted average time to signal, $AATS$, is given by

$$AATS_{\rm VSI} \;=\; \frac{d_1^2\,p_{01}+d_2^2\,p_{02}}{2\,(d_1\,p_{01}+d_2\,p_{02})}+\frac{d_1p_{11}+d_2\,p_{12}}{1-\beta}\;, \tag{3.7}$$

where

$$p_{01} \;=\; 2\,\big[\Phi(L)-\Phi(w)\big] \qquad \text{and} \qquad p_{02} \;=\; 2\,\Phi(w)-1 \tag{3.8}$$

are the probabilities of a sample mean belonging to the regions C_1 and C_2, respectively, when the process is in control.

To compare the effectiveness of the LSI and VSI methods, we assume that the average sampling intervals in both sampling methods are equal to the fixed sampling interval ($d=1$ and $k=3.8134$) in the expressions (2.19) and (3.7). Once again, the ratio

$$Q_{\rm LSI/VSI} \;=\; \frac{AATS_{\rm VSI}-AATS_{\rm LSI}}{AATS_{\rm VSI}}\times 100\%\;, \tag{3.9}$$

represents a measure in % of the relative variation of the $AATS$ value, with respect to the $AATS_{\rm VSI}$ reference. The results obtained for mean shifts with different sample sizes are presented in Table 2. The following conclusions are immediate:

Table 2: $Q_{\text{LSI/VSI}}(\%)$, as a function of λ, for different values of n and different sampling pairs in VSI.

n	(d_1, d_2)	λ											
		0.00	*0.25*	*0.50*	*0.75*	*1.00*	*1.25*	*1.50*	*1.75*	*2.00*	*2.25*	*2.50*	*3.00*
2	*(0.1, 1.9)*	0.1	−1.4	−5.9	−13.2	−21.6	−26.0	−19.8	−4.5	10.5	20.7	26.4	31.0
	(0.1, 1.5)	0.0	−0.9	−3.7	−8.6	−14.9	−20.5	−20.9	−14.0	−4.0	4.4	9.7	14.2
	$AATS_{\text{LSI}}$	*370.01*	*216.71*	*79.98*	*29.08*	*11.31*	*4.86*	*2.40*	*1.41*	*0.98*	*0.79*	*0.70*	*0.63*
3	*(0.1, 1.9)*	0.1	−2.2	−8.9	−19.0	−25.9	−18.2	1.1	17.0	25.5	29.5	31.2	32.2
	(0.1, 1.5)	0.0	−1.4	−5.7	−12.8	−20.1	−20.5	−10.6	1.2	8.8	12.7	14.4	15.4
	$AATS_{\text{LSI}}$	*370.01*	*175.53*	*50.46*	*15.24*	*5.27*	*2.23*	*1.22*	*0.86*	*0.71*	*0.66*	*0.63*	*0.61*
5	*(0.1, 1.9)*	0.1	−3.7	−14.6	−25.7	−15.4	9.3	23.9	29.6	31.5	32.1	32.3	32.3
	(0.1, 1.5)	0.0	−2.3	−9.5	−19.4	−19.4	−4.9	7.3	12.8	14.7	15.3	15.4	15.5
	$AATS_{\text{LSI}}$	*370.01*	*122.99*	*24.81*	*5.97*	*1.98*	*1.01*	*0.74*	*0.65*	*0.63*	*0.62*	*0.61*	*0.61*

i) In sample sizes more widely used in the literature, $n \geq 3$, the LSI method is quicker than the VSI method in detecting shifts of magnitude $\lambda > 1.5$, i.e., in situations whose probability of detection is high.

ii) The effectiveness of the LSI method increases when the sample size increases for moderate and large shifts in the mean. For small shifts in the mean, the effectiveness of the LSI method decreases as the sample size increases.

iii) If we consider $(d_1, d_2) = (0.1, 1.9)$ in VSI, the maximum reductions obtained with the LSI method are considerable (approximately 32%); in general, the performance of the LSI improves significantly when the sample size is larger; if the probability of occurrence of a shift is equal for all λ, using the LSI method could be a competitive advantage.

iv) If we consider $(d_1, d_2) = (0.1, 1.5)$ in VSI, the maximum reductions obtained with the LSI method are significantly smaller (approximately 16%) than those obtained with the other sampling pair in the VSI method; in general, the performance of the LSI improves as the sample size gets larger.

An example of application

In sections 3.1 and 3.2 we compared, in a critical way, the performances of the FSI and VSI methods with the performance of the LSI method in terms of the $AATS$. For a better perception in absolute terms of the LSI method, we present an example of application that allows checking the effectiveness in the detection of the shift.

Thereby, if there is a mean shift of magnitude $\lambda = 1.0$, and considering how unit of the time one hour, for the quality characteristic being monitored:

i) If we use the FSI sampling method, the first sample after the mean shift is drawn, on average, after 30 minutes, and we need 240 minutes, on average, to detect the shift.

ii) If we use the VSI sampling method:

a) with the sampling pair $(d_1, d_2) = (0.1, 1.9)$, the first sample after the mean shift is drawn, on average, after 54 minutes, and we need 103 minutes, on average, to detect the shift;

b) with the sampling pair $(d_1, d_2) = (0.1, 1.5)$, the first sample after the mean shift is drawn, on average, after 44 minutes, and we need 100 minutes, on average, to detect the shift.

iii) If we use the LSI sampling method: the first sample after the mean shift is drawn, on average, after 37 minutes, and we need 119 minutes, on average, to detect the shift.

In this case we can conclude that the use of the LSI method allows us to reduce the out-of-control period by 121 minutes, on average, compared to the FSI method, and increase the out-of-control by either 16 minutes or 19 minutes, compared to the VSI method, depending on whether we use the sampling pair $(d_1, d_2) = (0.1, 1.9)$ or the sampling pair $(d_1, d_2) = (0.1, 1.5)$.
On the other hand, if a shift of magnitude is of $\lambda = 1.5$, for the quality characteristic being monitored:

i) If we use the FSI sampling method, the first sample after the mean shift is drawn, on average, after 30 minutes, and we need 64 minutes, on average, to detect the shift.

ii) If we use the VSI sampling method:

a) with the sampling pair $(d_1, d_2) = (0.1, 1.9)$, the first sample after the mean shift is drawn, on average, after 54 minutes, and we need 58 minutes, on average, to detect the shift;

b) with the sampling pair $(d_1, d_2) = (0.1, 1.5)$, the first sample after the mean shift is drawn, on average, after 44 minutes, and we need 48 minutes, on average, to detect the shift.

iii) If we use the LSI sampling method: the first sample after the mean shift is drawn, on average, after 37 minutes, and we need 44 minutes, on average, to detect the shift.

We can conclude that the use of the LSI method allows us to reduce the out-of-control period by 20 minutes, on average, compared to the FSI method, and by either 14 minutes or 4 minutes, compared to the VSI method, depending on whether we use the sampling pair $(d_1, d_2) = (0.1, 1.9)$ or the sampling pair $(d_1, d_2) = (0.1, 1.5)$.
Thus, for this situation and others in which $\lambda > 1.5$, the use of the LSI method makes it possible to reduce the malfunction costs and makes the product more competitive by reducing its final price.

The influence of the sampling interval distribution on the standard deviation must be analysed as well. The values of the coefficient of variation of the TS for the different sampling methods (for the conditions previously described) are presented in Table 3. The results shown there allow us to conclude that for all methods and small mean shifts, the coefficients of variation are very close to 1.

Table 3: Values of the coefficient of variation of TS for the FSI, VSI and LSI methods.

CV		λ							
		0.00	*0.50*	*1.00*	*1.50*	*1.75*	*2.00*	*2.50*	*3.00*
FSI	$d = 1$	1.0000	1.0000	0.9998	0.9996	0.9989	0.9977	0.9798	0.9633
VSI	*(0.1, 1.9)*	1.0000	0.9968	0.8022	0.6143	0.6211	0.6261	0.6221	0.6112
	(0.1, 1.5)	1.0000	0.9981	0.8609	0.6071	0.6031	0.6087	0.6130	0.6133
LSI		0.9986	0.9818	0.7915	0.6357	0.6715	0.6943	0.7060	0.7067

For moderate mean shifts, the coefficient of variation for the LSI method is the smallest, although it is similar to the one of the VSI method when $d_2 = 1.9$. For $\lambda \geq 1.5$, the LSI method has a slightly larger coefficient of variation than the VSI method for all sampling pairs (due to greater dispersion in the sampling intervals underlying the Laplace distribution), but a smaller coefficient of variation than the FSI method.

4. SENSITIVITY ANALYSIS

To evaluate the consistency of the LSI method, a sensitivity analysis was performed. In this section the lower sampling interval is truncated, as it results in a situation similar to the VSI method. On the other hand, the concern in practical applications in certain industrial contexts in which one may be physically or administratively unable to take and analyse samples at very short time intervals justifies this type of study.

D is the random variable that represents the time interval between consecutive inspections, and d_1 is the smallest sampling interval possible. Hence, we have:

$$D \leq d_1 \iff \frac{k}{2}.e^{-|u|} \leq d_1 \iff u \geq -\ln\left(\frac{2\times d_1}{k}\right) \vee u \leq \ln\left(\frac{2\times d_1}{k}\right), \tag{4.1}$$

where $L^* = -\ln\left(\frac{2\times d_1}{k}\right)$ is a multiple of the standard deviation that can be interpreted as W in the VSI method. Let us consider D^* as the time interval between

consecutive samples when the sample mean is between $\mu_0 \pm L^* \sigma_0 \, n^{-0,5}$. The distribution of D^* is the conditional distribution of the mean, given that D^* falls between the control limits for the given mean shifts. The probability density function of D^* is given by

$$f^{**}(\bar{x}) \;=\; \frac{\sqrt{n}}{\beta^* \, \sigma_0 \sqrt{2\,\pi}} \; e^{-\frac{n(\bar{x}-\mu_0-\lambda\sigma_0)^2}{2\sigma_0^2}} \,, \tag{4.2}$$

with

$$\beta^* \;=\; \Phi\big(L^* - \lambda\sqrt{n}\big) - \Phi\big(-L^* - \lambda\sqrt{n}\big) \,. \tag{4.3}$$

Through reasoning similar to that which has been applied in the statistical properties of the LSI method, we have

$$E\big(D^* | L^*, \lambda, n\big) \;=\; \frac{\sqrt{e}\, k^*}{2\beta^*} \left[e^{\lambda\sqrt{n}} \times A(L^*, \lambda, n) + e^{-\lambda\sqrt{n}} \times B(L^*, \lambda, n) \right] , \tag{4.4}$$

where

$$\begin{aligned} A(L^*, \lambda, n) \;&=\; \Phi\big(-1 - \lambda\sqrt{n}\big) - \Phi\big(-L^* - 1 - \lambda\sqrt{n}\big) \,, \\ B(L^*, \lambda, n) \;&=\; \Phi\big(L^* + 1 - \lambda\sqrt{n}\big) - \Phi\big(1 - \lambda\sqrt{n}\big) \,, \end{aligned} \tag{4.5}$$

and k^* depends on the value of L^*. Thus, the probability of using the sampling interval d_1 is given by

$$\begin{aligned} p_1 \;&=\; P\big(D = d_1 | \lambda\big) \\ &=\; 1 - P\left(\mu_0 - L^* \frac{\sigma_0}{\sqrt{n}} \le \bar{X} \le \mu_0 + L^* \frac{\sigma_0}{\sqrt{n}} \;\middle|\; LCL \le \bar{X} \le UCL \right) \\ &=\; 1 - \frac{\beta^*}{\beta} \,. \end{aligned} \tag{4.6}$$

Based on the assumptions stated, the average sampling interval is given by

$$E(D) \;=\; \frac{\beta^*}{\beta} \times E(D^*) + d_1 \times \left(1 - \frac{\beta^*}{\beta} \right) . \tag{4.7}$$

Considering (4.7), "*3-sigma*" control limits and a unit average sampling interval, in control, the values of k^* and L^*, obtained by simulation are presented in Table 4 for the considered values of d_1.

Table 4: Values of k^* and L^* obtained by simulation for different values of d_1.

d_1	k^*	L^*
0.1	3.8134	2.9480
0.2	3.8099	2.2539
0.3	3.7942	1.8443
0.4	3.7591	1.5473
0.5	3.6976	1.3077

Examining the results in Table 4, we conclude that the value of k^* gets smaller as the value of d_1 increases, reducing the multiples of the standard deviation. This feature shows how the LSI method can be equated to the VSI method because when we increase the smaller sampling interval in the VSI method, the W value decreases.

To assess the impact of truncation of the lower sampling interval in terms of the $AATS$ values, we rewrite expression (2.14), adapted to the new conditions, as

$$AATS = E(G) + (ARL - 1)\,E(D) = E(G) + \frac{E(D^*)\,\beta^* + d_1\,(\beta - \beta^*)}{1 - \beta}\,, \tag{4.8}$$

where $E(G)$ value is obtained by simulation and is used in comparisons between the LSI method and the remaining methods. Intuitively, an increase in the value of d_1 leads to an increase in its probability of use. To prove that this intuition is correct, we perform a sensitivity study of the LSI method. We compare the $AATS$ values obtained using the LSI method in its original form with those obtained using the LSI method with truncation of the lower sampling interval.

The results are presented in Figure 2, using a measure of relative variation (sensitive to the lower sampling interval change) and the values of k^* and L^*, expressed in terms of % of the $AATS$ value (being $AATS_{\text{LSI}}$ the reference)

$$Q_{\text{LSI}^*/\text{LSI}} = \frac{AATS_{\text{LSI}^*} - AATS_{\text{LSI}}}{AATS_{\text{LSI}}} \times 100\%\,. \tag{4.9}$$

Analysing this figure, one concludes that the differences in the $AATS$ values increase as the probability of detecting mean shifts increases, reaching its maximum for shifts of magnitudes of $\lambda = 1.25$. From this point onward, the effectiveness of the methods tends to converge, becoming identical for large magnitudes of mean shifts. However, for $d_1 = 0.4$ and $d_1 = 0.5$, there are strong increases in the $AATS$ for some mean shifts. Even when d_1 is three times greater than the initial value, the maximum relative reduction in the $AATS$ using the non-truncated LSI method is only 12.7%.

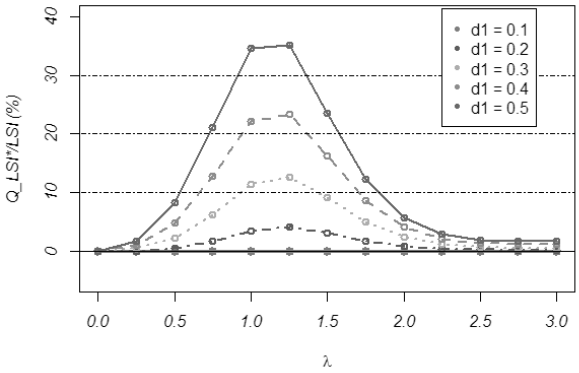


Figure 2: $Q_{\text{LSI}^*/\text{LSI}}(\%)$, as a function of λ for different values of d_1, with $n = 5$.

Comparison between the LSI* and FSI* sampling methods

For the conditions mentioned in the previous section, the $AATS$ values for the truncated LSI method and the FSI method for a sample size of five were compared. The assessment measures the effect that the change in the lower sampling interval can have on the performance of the LSI method, compared to what occurs in the FSI method. Thus, one considers the measure of performance presented in (3.2), taking into consideration the new values of the $AATS$ in LSI ($AATS_{\text{LSI}^*}$). From the results of $Q_{\text{LSI}^*/\text{FSI}}$, only for mean shifts, we can conclude the following: when the lower sampling interval is truncated, the LSI* method is more effective than the FSI* method for the same mean shifts; the increase in the sampling interval is not proportional to the reduction in effectiveness of the method; the FSI* method detects large mean shifts more quickly than does the LSI* method, maintaining the effectiveness presented previously.

Comparison between the LSI* and VSI* sampling methods

Using a similar methodology, the LSI* and VSI* methods were compared for the same and for new conditions. Considering a sample size of 5 units and the same number of false alarms, we truncate the lower sampling interval in both methods to the same values. To compare the effectiveness of the two methods, the performance measure defined in (3.9) is used, replacing $AATS_{\text{LSI}}$ with $AATS_{\text{LSI}^*}$ and $AATS_{\text{VSI}}$ with $AATS_{\text{VSI}^*}$. From the results presented in Table 5, for mean shifts, we can draw the following conclusions:

- **i**) In general, the performance of the LSI* method improves when the lower sampling interval gets larger for small shifts. In particular, when $\lambda = 1$ and $d_1 \geq 0.3$, LSI* is more effective than VSI*.
- **ii**) For moderate to large shifts, the performance of LSI* is better than VSI*, except when the lower sampling interval increases.

iii) When we use the VSI* method, the increases obtained are significantly greater than the reductions for the different sampling pairs.

Table 5: $Q_{\text{LSI}^*/\text{VSI}^*}$ (%), as a function of λ, for different values of d_1 equal to the smaller sampling interval in VSI.

d_1	(d_1, d_2)	λ											
		0.00	*0.25*	*0.50*	*0.75*	*1.00*	*1.25*	*1.50*	*1.75*	*2.00*	*2.25*	*2.50*	*3.00*
0.1	*(0.1, 1.9)*	0.1	−3.7	−14.6	−25.7	−15.4	9.3	23.9	29.6	31.5	32.1	32.2	32.3
	(0.1, 1.5)	0.0	−2.3	−9.5	−19.4	−19.4	−4.9	7.3	12.8	14.7	15.3	15.4	15.5
0.2	*(0.2, 1.9)*	0.1	−2.7	−9.7	−14.5	−4.9	12.0	22.3	26.5	28.0	28.4	28.6	28.6
	(0.2, 1.5)	0.0	−1.5	−5.9	−10.5	−8.3	0.7	7.6	10.8	11.8	12.2	12.3	12.3
0.3	*(0.3, 1.9)*	0.1	−1.8	−6.3	−8.0	−0.6	11.0	18.6	22.2	23.6	24.1	24.3	24.3
	(0.3, 1.5)	0.0	−1.0	−3.5	−5.6	−3.6	1.6	5.6	7.5	8.3	8.6	8.6	8.6
0.4	*(0.4, 1.9)*	0.0	−1.3	−4.0	−4.4	0.8	8.5	14.2	17.3	18.7	19.3	19.4	19.5
	(0.4, 1.5)	0.0	−0.6	−2.1	−3.0	−1.8	0.8	2.8	3.9	4.4	4.6	4.6	4.6
0.5	*(0.5, 1.9)*	0.0	−0.8	−2.5	−2.5	0.9	5.7	9.7	12.1	13.3	13.9	14.0	14.1
	(0.5, 1.5)	0.0	−0.4	−1.2	−1.8	−1.4	−0.6	0.0	0.2	0.3	0.3	0.3	0.3

For the purpose of illustration, consider the case of the lower sampling interval for the LSI* method being truncated to $d_1 = 0.2$. For the example given at the end of section 3.2 and for a mean shift of magnitude $\lambda = 1.5$, we conclude that the use of the LSI* method allows us to reduce the malfunction period by 18 minutes, on average, compared to the FSI* method, and by either 13 minutes or 4 minutes, compared to the VSI* method, depending on whether we use the sampling pair $(d_1, d_2) = (0.2, 1.9)$ or the sampling pair $(d_1, d_2) = (0.2, 1.5)$. These results demonstrate the good performance and sensitivity of the LSI* method.

Results concerning the robustness of the method have been obtained by Carmo *et al.* (2013) for a case in which the distribution of quality has a *t-Student* distribution and another case in which the distribution of quality is a mixture of two normal distributions with different standard deviations. In both cases the performance of the LSI method is better than the performance of FSI and VSI methods, and there are situations in which the LSI method detects mean shifts more quickly than the VSI method.

5. CONCLUSIONS

The LSI method detects small and moderate mean shifts in quality more quickly than the FSI method. For large mean shifts, FSI is more efficient. However, the gains achieved with the use of LSI are greater. In a production system in which the sampling costs are very significant (for example, in the production of a touchscreen display for the iPhone 5; 44 U.S. dollars per unit), and in which the quality changes are small or moderate, the use of LSI offers a competitive advantage in reducing sampling and malfunction costs.

When we use the sample pair $(d_1, d_2) = (0.1, 1.9)$ in VSI, the adaptive methods are subject to the same conditions. In other words, the smallest and largest sampling intervals in the LSI method are approximately the same. In LSI, the smallest interval is 0.095 and the largest interval is 1.907. For the sampling pairs considered, LSI detects moderate and large mean shifts more quickly.

We consider the LSI method to be not very sensitive because it has a similar performance to that of the non-truncated method for several mean shifts, particularly when the smallest sample interval is smaller than three times the original smallest interval.

For the reasons explained, and for simplicity, the use of the LSI method can offer a competitive advantage in automating tasks and using nano-scale measurement instruments.

Future research will involve a different approach to the calculation of $E(G)$, using different distributions for the lifetime of the system and assessing its impact. We will extend the study of the statistical properties and performance to the use of joint control charts ($\bar{X}$-chart and S-chart or $\bar{X}$-chart and R-chart) and special control charts (CUSUM and EWMA charts) to compare the LSI method with other adaptive methods (for example, VSS, VSSI, and VP).

Finally, it is our intention to conduct a study to determine the k value that minimizes a cost function by production cycle.

ACKNOWLEDGMENTS

The authors are grateful to the Editor and Referees for their careful reviews and helpful suggestions, which have improved the final manuscript considerably. This work is financed by national funds through FCT — Foundation for Science and Technology — under the project UID/MAT/04674/2013 (CIMA).

REFERENCES

[1] Banerjee, P.K. and Rahim, M.A. (1988). Economic design of $\bar{X}$ control charts under Weibull shock models, *Technometrics*, **30**(4), 407–414.

[2] Carmo, M. (2004). *Problemas de Amostragem em Controlo Estatístico de Qualidade*, Master Thesis, Universidade Nova de Lisboa, Lisboa.

[3] Carmo, M.; Infante, P. and Mendes, J.M. (2013). *Alguns resultados da robustez de um método de amostragem adaptativo em controlo de qualidade*. In "Estatística: Novos Desenvolvimentos e Inspirações" (M. Maia, P. Campos, and P.D. Silva, Eds.), SPE, Lisboa, 95–108.

[4] Carot, V.; Jabaloyes, J.M. and Carot, T. (2002). Combined double sampling and variable sampling interval $\bar{X}$ chart, *International Journal of Production Research*, **40**(9), 2175–2186.

[5] Costa, A.F.B. (1994). $\bar{X}$ chart with variable sample size, *Journal of Quality Technology*, **26**(3), 155–163.

[6] Costa, A.F.B. (1999). $\bar{X}$ chart with variable parameters, *Journal of Quality Technology*, **31**(4), 408–416.

[7] Daudin, J.J. (1992). Double sampling $\bar{X}$ chart, *Journal of Quality Technology*, **24**(2), 78–87.

[8] Infante, P. (2004). *Métodos de Amostragem em Controlo de Qualidade*, Doctoral Dissertation, Universidade de Évora, Évora.

[9] Infante, P. and Rodrigues Dias, J. (2002). Análise da importância da distribuição do tempo de vida no período de inspecção em controlo estatístico de qualidade, *Investigação Operacional*, **22**(2), 167–179.

[10] Lin, Y.C. and Chou, C.Y. (2005). Adaptive $\bar{X}$ control charts with sampling at fixed times, *Quality and Reliability Engineering International*, **21**(2), 163–175.

[11] Mahadik, S.B. and Shirke, D.T. (2009). A Special Variable Sample Size and Sampling Interval $\bar{X}$ Chart, *Communications in Statistics — Theory and Methods*, **38**, 1284–1299.

[12] Morais, M.C. (2002). *Ordenação Estocástica na Análise de Desempenho de Esquemas de Controlo de Qualidade*, Doctoral Dissertation, IST, Universidade Técnica de Lisboa, Lisboa.

[13] Morais, M.C. and Pacheco, A. (2001). *Ordenação Estocástica na Análise de Desempenho de Esquemas de Controlo de Qualidade*. In "A Estatística em Movimento" (M.M. Neves, J. Cadima, M.J. Martins and F. Rosado, Eds.), SPE, Lisboa, 247–260.

[14] Prabhu, S.S.; Montgomery, D.C. and Runger, G.C. (1994). A combined adaptive sample size and sampling interval $\bar{X}$ control scheme, *Journal of Quality Technology*, **26**(3), 164–176.

[15] Prabhu, S.S.; Runger, G.C. and Keats, J.B. (1993). An adaptive sample size $\bar{X}$ chart, *International Journal of Production Research*, **31**(2), 2895–2909.

[16] Rahim, M.A. and Banerjee, P.K. (1993). A generalized model for the economic design of $\bar{X}$ control charts for production systems with increasing failure rate and early replacement, *Naval Research Logistics*, **40**(6), 787–809.

[17] Reynolds, M.R. (1989). Optimal variable sampling interval control chart, *Sequential Analysis*, **8**(4), 361–379.

[18] Reynolds, M.R. (1995). Evaluation properties of variable sampling interval control charts, *Sequential Analysis*, **14**(1), 59–97.

[19] Reynolds, M.R.; Amin, R.W.; Arnold, J.C. and Nachlas, J.A. (1988). $\bar{X}$ charts with variables sampling intervals, *Technometrics*, **30**(2), 181–192.

[20] Reynolds, M.R. and Arnold, J.C. (1989). Optimal one-sided *Shewhart* control charts with variable sampling intervals, *Sequential Analysis*, **8**(1), 51–77.

[21] Rodrigues Dias, J. (1999). *A new method to obtain different sampling intervals in statistical quality control*, Universidade de Évora, Évora.

[22] Rodrigues Dias, J. and Carmo, M. (2009). *A new approach for comparing economic sampling methods in quality control in systems with different failure rates.* In "IRFt'2009-3rd International Conference on Integrity, Reliability Failure, Challenges and Opportunities" (S.A. Meguid and J.F. Silva Gomes, Eds.), F.E.U.P, Porto, 653–654.

[23] Rodrigues Dias, J. and Infante, P. (2008). Control charts with predetermined sampling intervals, *International Journal of Quality and Reliability Management*, **25**(4), 423–435.

[24] Runger, G.C. and Pignatiello, J.J. (1991). Adaptive sampling for process control, *Journal of Quality Technology*, **23**(2), 133–155.

[25] Stoumbos, Z.G.; Mittenthal, J. and Runger, G.C. (2001). Steady-state-optimal adaptive control charts based on variable sampling intervals, *Stochastic Analysis and Applications*, **19**(6), 1025–1057.

[26] Stoumbos, Z.G. and Reynolds, M.R. (1997). Control charts applying a sequential test at fixed sampling intervals, *Journal of Quality Technology*, **29**(1), 21–40.

REVSTAT – Statistical Journal
Volume 16, Number 1, January 2018, 23–43

FINDING THE OPTIMAL THRESHOLD OF A PARAMETRIC ROC CURVE UNDER A CONTINUOUS DIAGNOSTIC MEASUREMENT

Authors: Yi-Ting Hwang
– Department of Statistics, National Taipei University, Taipei, Taiwan
hwangyt@gm.ntpu.edu.tw

Yu-Han Hung
– Department of Statistics, National Taipei University, Taipei, Taiwan
lalamomok0914@hotmail.com

Chun Chao Wang
– Department of Statistics, National Taipei University, Taipei, Taiwan
ccw@gm.ntpu.edu.tw

Harn-Jing Terng
– Advpharma, Inc., New Taipei city, Taiwan
ternghj@advpharma.com.tw

Received: March 2015 Revised: July 2016 Accepted: July 2016

Abstract:

- The accuracy of a binary diagnostic test can easily be assessed by comparing the sensitivity and specificity with the status of respondents. When the result of a diagnostic test is continuous, the assessment of accuracy depends on a specified threshold. The receiver operating characteristic (ROC) curve, which includes all possible combinations of sensitivity and specificity, provides an appropriate measure for evaluating the overall accuracy of the diagnostic test. Nevertheless, in practice, a cutoff value is still required to make easier its clinical usage easier. The determination of a proper cutoff value depends on how important the practitioner views the specificity and sensitivity. Given particular values of specificity and sensitivity, this paper derives the optimal cutoff value under two parametric assumptions on the outcomes of the diagnostic test. Because the optimal cutoff value does not have a closed form, the numerical results are tabulated for some parameter settings to find the optimal cutoff value. Finally, real data are employed to illustrate the use of the proposed method.

Key-Words:

- *bilogistic model; binormal model; optimal threshold; sensitivity; specificity.*

AMS Subject Classification:

- 62C05.

1. INTRODUCTION

A diagnostic test that results in a continuous value is often evaluated using the receiver operating characteristic (ROC) curve. Let TP, FP, FN and TN denote the true positive decision, false positive decision, false negative decision and true negative decision, respectively. The following table provides 4 possible diagnostic test decisions:

True status	Test result	
	Positive	Negative
Case	TP	FN
Normal	FP	TN

Let $P[\text{TP}]$ be the probability that a true positive decision is made, and let $P[\text{TN}]$, $P[\text{FP}]$ and $P[\text{FN}]$ be defined similarly. The true positive rate (TPR) and the true negative rate (TNR) can be derived from $P[\text{TP}]$, $P[\text{TN}]$, $P[\text{FP}]$ and $P[\text{FN}]$ as

$$\text{TPR} = \frac{P[\text{TP}]}{P[D+]}, \tag{1.1}$$

$$\text{TNR} = \frac{P[\text{TN}]}{P[D-]}, \tag{1.2}$$

where $P[D+] = P[\text{TP}] + P[\text{FN}]$ denotes the prevalence of a disease and $P[D-] = P[\text{TN}] + P[\text{FP}] = 1 - P[D+]$.

A ROC curve is constructed from different values for the TPR and FPR. The determination of the TPR and FPR requires a cutoff value to classify the normal and diseased populations when the outcome is continuous. The ROC curve is then formed using TPRs and FPRs derived from all possible cutoff values. However, for practical use, the continuous outcome has to be dichotomized such that the investigator or practitioner can easily use it to discriminate the disease status. Nevertheless, the ROC curve does not provide direct information on how to determine such a cutoff value. It is thus important to find an optimal cutoff value (OCV) such that the probabilities of correct decisions are maximized.

Let S_D and S_N denote the outcome of the diagnostic measure for the disease group and the normal group, respectively, and let F_D and F_N denote the corresponding distribution functions. The ROC curve can be represented as

$$\text{ROC}(t) = \bar{F}_D(\bar{F}_N^{-1}(t)),$$

where $t \in (0, 1)$, $\bar{F}_D(t) = 1 - F_D(t)$ is the survival function of $F_D(t)$ and $\bar{F}_N(t)$ is

defined similarly. Because the FPR and TPR are functions of $\bar{F}_D$ and $\bar{F}_N$ as

$$\begin{aligned}\text{FPR}(c) &= P[S_N > c|N] = 1 - F_N(c) = \bar{F}_N(c),\\ \text{TPR}(c) &= P[S_D > c|D] = 1 - F_D(c) = \bar{F}_D(c),\end{aligned}$$

for a given cutoff $c \in (-\infty, \infty)$, the ROC curve can be represented in terms of the TPR and FPR.

To derive the OCV, an additional objective function is required. Three objectives have been discussed in the literature to find the OCV (Akobeng [1]; Kumar [5]). The first objective function is defined as the distance from the ROC curve to the point (0,1), that is,

$$C_1(c) = \sqrt{(1 - \text{TPR}(c))^2 + (\text{FPR}(c))^2} \tag{1.3}$$

and the OCV is the point at which $C_1(c)$ has the minimum. The second objective function proposed by Youden [9] is the vertical distance from the line of equality to the point on the ROC curve, which is

$$C_2(c) = \text{TPR}(c) + \text{TNR}(c) - 1, \tag{1.4}$$

and the OCV is the point that maximizes C_2. $C_2(c)$ is known as the Youden index. An alternative and equivalent representation of $C_2(c)$ is

$$|\text{TPR}(c) - (1 - \text{TNR}(c))|$$

expressed by Lee [6] and Krzanowski and Hand [4]. The third objective function is a weighted function of the probability of four diagnostic decisions, defined by Metz [8] as

$$C_3(c) = C_0 + C_{\text{TP}}P[\text{TP}] + C_{\text{TN}}P[\text{TN}] + C_{\text{FP}}P[\text{FP}] + C_{\text{FN}}P[\text{FN}], \tag{1.5}$$

where C_0 is the overhead cost, C_{TP} represents the average cost of the medical consequences of a true positive decision, and the remainder of the costs are defined similarly. Based on (1.1) and (1.2), expression (1.5) can be rewritten as

$$\begin{aligned}C_3(c) =& \{C_0 + C_{\text{FP}} \times P[D-] + C_{\text{FN}} \times P[D+]\}\\ &+ \{[C_{\text{FN}} - C_{\text{TP}}] \times P[D+]\} \times \text{TPR}(c)\\ &+ \{[C_{\text{TN}} - C_{\text{FP}}] \times P[D-]\} \times \text{TNR}(c)\end{aligned} \tag{1.6}$$

In particular, the first term on the right-hand side of (1.6) includes only the three costs and the prevalence, which do not depend on the decision of a diagnostic test. Because the determination of the OCV is not related to this term, it is neglected in the following discussion. Thus, in terms of (1.6), the best cutoff value is the one that minimizes C_3. The critical value occurs at

$$\frac{\partial \text{TPR(c)}}{\partial \text{TNR(c)}} = -\frac{(C_{\text{TN}} - C_{\text{FP}}) \times P[D-]}{(C_{\text{FN}} - C_{\text{TP}}) \times P[D+]},$$

which is the slope of a line of isoutility or the tangent line in the ROC space. Metz [8] concluded that the OCV on a ROC curve must be tangent to the highest line of isoutility that intersects with the ROC curve.

The OCV derived from the first and second objective functions is determined empirically (Kumar [5]). Under the binormal model and assuming that the slope of the tangent line to the ROC curve equals η, an explicit form for the OCV under $C_3(c)$ is derived and is referred to as P252 in Halperm *et al.* [3]. However, the third objective function uses not only the cost for each decision but also the prevalence of the disease. The latter can possibly be obtained empirically using the existing data, whereas the cost of the medical consequences is difficult to obtain. Thus, it is rarely used in the medical literature (Kumar [5]).

For a practitioner, sensitivity and specificity, which correspond to the TPR and TNR, are commonly used measures, and the importance of these two measures depends on the purpose of the diagnostic test. Thus, rather than the equal weight setting for the TPR and TNR as in (1.3) and (1.4), in this paper, we suggest using a more general objective function,

$$C(c) = \alpha \times \mathrm{TPR}(c) + \beta \times \mathrm{TNR}(c), \tag{1.7}$$

where $0 < \alpha, \beta < 1$ and $\alpha + \beta = 1$, to derive the OCV. The weight α can be regarded as the relative cost for an additional cost of classifying a TP compared to an additional cost of classifying a TN. Assuming the location and scale parametric assumption, the OCV can be then obtained under $C(c)$. In particular, when $\alpha = 0$, the objective function in (1.7) is the usual criterion for finding the OCV by minimizing the FPR or maximizing the specificity. Conversely, when $\beta = 0$, the objective function is the usual criterion for finding the OCV by maximizing the sensitivity. Section 2 describes the basic definition of the ROC curve and the derivation for the OCV. Section 3 presents the numerical results. Sections 4 and 5 provide a real application and discussions, respectively.

2. METHOD

Assume that F_D and F_N belong to a location and scale family. In other words, both distributions can be expressed by a standard form, say F, with different location and scale parameters. Let (μ_D, γ_D) and (μ_N, γ_N) denote the parameters for F_D and F_N, respectively. The FPR and TPR can be represented in terms of F as

$$\mathrm{TPR}(c) = P\Big[\frac{S_D - \mu_D}{\gamma_D} > \frac{c - \mu_D}{\gamma_D}\Big] = F\Big(\frac{\mu_D - c}{\gamma_D}\Big) \tag{2.1}$$

$$\mathrm{FPR}(c) = P\Big[\frac{S_N - \mu_N}{\gamma_N} > \frac{c - \mu_N}{\gamma_N}\Big] = F\Big(\frac{\mu_N - c}{\gamma_N}\Big). \tag{2.2}$$

Let t_p denote the critical value of F, i.e., $1 - F(t_p) = p$. Given $\mathrm{FPR}(c)$, the following relationship is obtained:

$$t_{\mathrm{FPR}} = F_N^{-1}(\mathrm{FPR}(c)) = -\frac{c - \mu_N}{\gamma_N},$$

and

$$c = \mu_N - \gamma_N \times t_{\mathrm{FPR}}. \tag{2.3}$$

Additionally, given $\mathrm{TPR}(c)$, we have

$$t_{\mathrm{TPR}} = F_D^{-1}(\mathrm{TPR}(c)) = -\frac{c - \mu_D}{\gamma_D},$$

and

$$c = \mu_D - \gamma_D \times t_{\mathrm{TPR}}. \tag{2.4}$$

Given FPR and TPR, (2.3) and (2.4) provide the relationship between two critical values as

$$t_{\mathrm{TPR}} = \frac{\mu_D - \mu_N}{\gamma_D} + \frac{\gamma_N}{\gamma_D} t_{\mathrm{FPR}} = a + b t_{\mathrm{FPR}}, \tag{2.5}$$

where $a = (\mu_D - \mu_N)/\gamma_D$ and $b = \gamma_N/\gamma_D$. From (2.5), a linear relationship exists between two critical values of F_D and F_N, where a is the intercept and b is the slope. Given $\mathrm{FPR}(c)$, the ROC curve can be represented as

$$\mathrm{ROC}(c) = P[S_D > c] = F\Big(\frac{\mu_D - c}{\gamma_D}\Big). \tag{2.6}$$

Substituting the value of c defined in (2.3) into (2.6) yields

$$\mathrm{ROC}(c) = P[S_D > c] = F\Big(\frac{\mu_D - \mu_N + \gamma_N \times t_{\mathrm{FPR}}}{\gamma_D}\Big) = F(a + b t_{\mathrm{FPR}}).$$

Under the location and scale family as defined in (2.1), (2.2) and (2.5), (1.7) becomes

$$C(c) = \alpha F\Big(a + b\Big(\frac{\mu_N - c}{\gamma_N}\Big)\Big) + \beta F\Big(\frac{c - \mu_N}{\gamma_N}\Big).$$

The OCV can then be determined by finding the critical value of $\frac{dC}{dc} = 0$, where

$$\frac{dC(c)}{dc} = \alpha f\Big(a + b\Big(\frac{\mu_N - c}{\gamma_N}\Big)\Big) \times \Big(-\frac{b}{\gamma_N}\Big) + \beta f\Big(\frac{c - \mu_N}{\gamma_N}\Big) \times \Big(\frac{1}{\gamma_N}\Big) \tag{2.7}$$

and $f(\cdot)$ is the density function of $F(\cdot)$. The following theorem discusses two location and scale families. The proof for Theorem 2.1 is provided in the Appendix, and the proof for Theorem 2.2 is similar.

Theorem 2.1. *Assume that $F(\cdot) = \Phi(\cdot)$ is a standard normal distribution function. To be consistent with the conventional notation, the scale parameters are denoted by σ_D and σ_N. Then,*

1. *When $b = 1$, we obtain*

$$OCV = \mu_N + \frac{a}{2}\sigma_N - \frac{\sigma_N}{a}\log\Big(\frac{1-\beta}{\beta}\Big). \tag{2.8}$$

2. *When $b \neq 1$, we obtain*

$$OCV = \frac{T \pm \sqrt{T^2 - 2(1-b^2)R/\sigma_N^2}}{(1-b^2)/\sigma_N^2}, \tag{2.9}$$

where

$$R = \frac{\mu_N^2 - (a\sigma_N + b\mu_N)^2}{2\sigma_N^2} + \log(\frac{\alpha b}{\beta}), \tag{2.10}$$

$$T = \frac{\mu_N - ab\sigma_N - b^2\mu_N}{\sigma_N^2}, \tag{2.11}$$

and R and T have to satisfy the condition $T^2 - 2(1-b^2)R/\sigma_N^2 > 0$.

Theorem 2.2. *Assume that $F(\cdot)$ is a standard logistic distribution function, i.e.,*

$$F(x) = [1 + \exp(-x)]^{-1}.$$

Then,

1. *When $b = 1$, we obtain a closed form for the OCV as*

$$OCV = -\sigma_D \log(q), \tag{2.12}$$

where

$$q = \frac{(\alpha-\beta) \pm \sqrt{\alpha\beta(\exp(a) + \exp(-a) - 2)}}{\Big[\beta\exp\Big(-\frac{\mu_N}{\gamma_N}\Big) - \alpha\exp\Big(-\frac{\mu_D}{\gamma_N}\Big)\Big]\exp\Big(\frac{\mu_D+\mu_N}{\gamma_N}\Big)}. \tag{2.13}$$

2. *When $b \neq 1$, the OCV is found numerically by solving the following nonlinear equation*

$$\begin{aligned}&\frac{\beta}{\gamma_N}\exp\Big(\frac{\mu_N}{\gamma_N}\Big)k^{-\frac{1}{\gamma_N}}\Big(\exp\Big(\frac{b\mu_N + a\gamma_N}{\gamma_N}\Big)k^{-\frac{1}{\gamma_D}} + 1\Big)^2 \\ &\quad = \frac{\alpha b}{\gamma_N}\exp\Big(\frac{(b\mu_N + a\gamma_N}{\sigma_N}\Big)k^{-\frac{1}{\gamma_D}}\Big(\exp\Big(\frac{\mu_N}{\gamma_N}\Big)k^{-\frac{1}{\gamma_N}} + 1\Big)^2,\end{aligned} \tag{2.14}$$

where $k = e^c$.

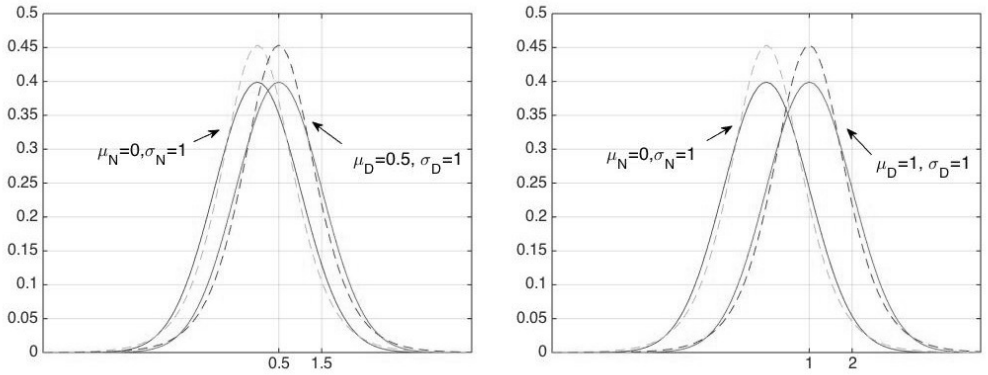
2.1. Relationship between the objective function and cutoff values

As c increases, the TPR decreases and the TNR increases. Because we assume that a case has a higher test value, the relative change in the TPR with respect to c is more rapid than that in the TNR. Furthermore, as expected, increasing μ_D means a smaller overlapping area in the densities for the normal and diseased populations and results in an increase in the TPR. When μ_D is fixed, the influence of σ_D on the TPR depends on c. When c is closer to μ_D, increasing σ_D reduces the TPR.

To understand how the parametric assumption influences the relationship between the objective function and the OCV, the basic features for the binormal and bilogistic models are discussed in the following. The common feature is that both distributions are symmetric about the location parameter. Nevertheless, the scale parameter in the normal distribution is the standard deviation, whereas the scale parameter in the logistic distribution is equal to the standard deviation times $\sqrt{3}/\pi$. Finally, the kurtosis of the normal distribution equals 3, whereas that of the logistic distribution equals 4.2.

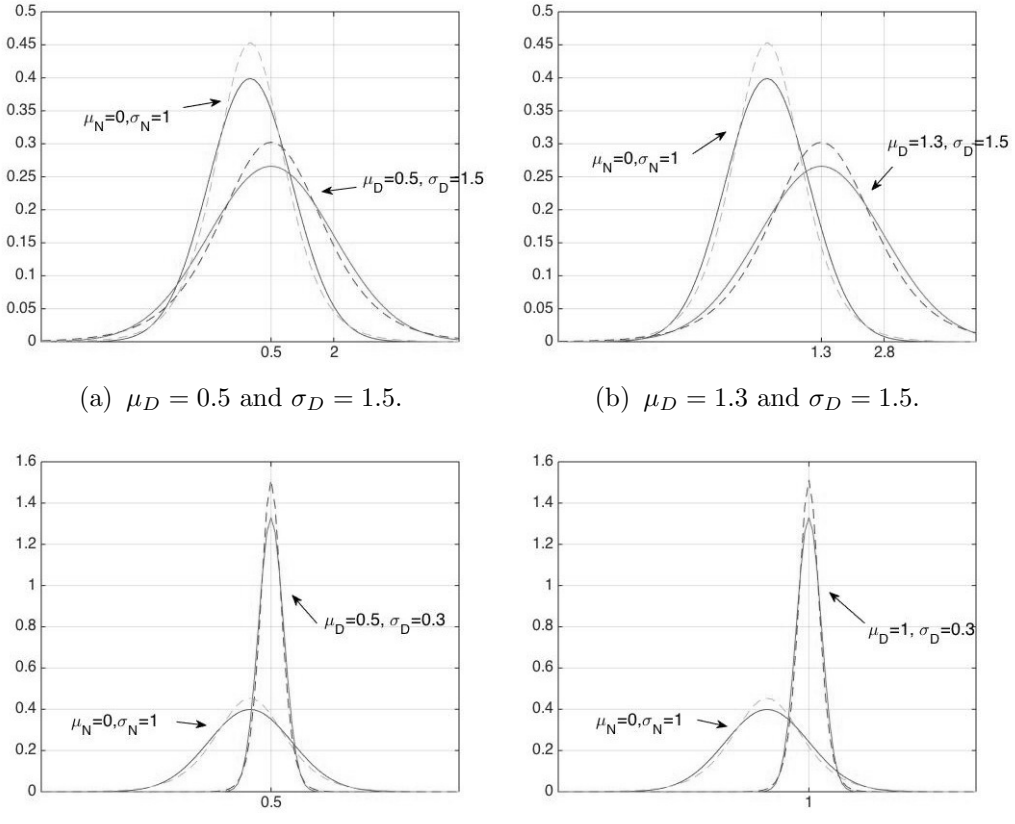
Assuming that $\mu_N = 0$ and $\sigma_N = 1$, Figures 1(a)–1(b) display the normal and logistic density functions for the normal and diseased populations when $b = 1$, and Figures 2(a)–2(d) display the situations when $b \neq 1$, where the solid line represents the normal distribution and the dashed line represents the logistic distribution and the left curve is for the control population and the right curve is for the diseased population. Under the same settings of μ_D and σ_D, the tail probability for the logistic distribution is slightly larger than that for the normal distribution. Furthermore, the mode of the logistic distribution is higher than that of the normal distribution because it has a larger kurtosis. These distinct features influence the TPR and TNR as shown in Table 1. Furthermore, due to a more concentrated feature for the logistic distribution, under the considered situation, the TNR of the logistic distribution is slightly larger than that of the normal distribution when c is closer to the μ_N, whereas for larger c, the TNR of the logistic distribution is slightly smaller. Thus, under the assumption that $\mu_N < \mu_D$, to have a higher TPR, the cutoff value for the logistic distribution is smaller than that for the normal distribution. In contrast, when investigating the TNR, the cutoff values for the logistic distribution might not be smaller.

The proposed objective function is a weighted function of the TPR and TNR. Figures 3(a)–3(b) show the relationship between the objective function C and the cutoff value c for various βs assuming that $\mu_N = 0$, $\sigma_N = 1$ and $\mu_D = 1$, $\sigma_D = 1$. For the binormal assumption, Figure 3(a) shows that when $\beta = 0.5$ and OCV=0.5, we obtain $C(\text{OCV}) = 0.6915$. When $\beta = 0.7$, that is, the specificity is more important than the sensitivity, we obtain OCV=1.3473 and $C(\text{OCV}) = 0.7470$.



(a) $\mu_D = 0.5$ and $\sigma_D = 1$. (b) $\mu_D = 1$ and $\sigma_D = 1$.

Figure 1: The probability density functions for normal distribution and logistic distributions for $\mu_N = 0$, $\sigma_N = 1$ and $b = 1$, where the solid line represents the normal curve and the dashed line represents the logistic curve.



(a) $\mu_D = 0.5$ and $\sigma_D = 1.5$. (b) $\mu_D = 1.3$ and $\sigma_D = 1.5$.

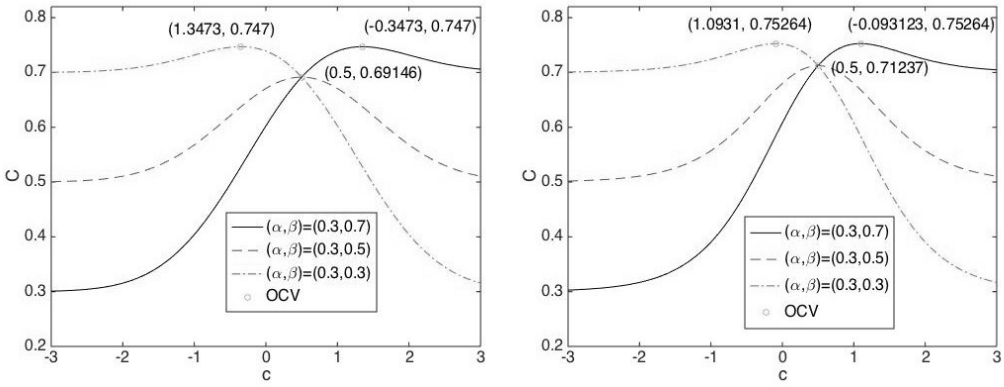
(c) $\mu_D = 0.5$ and $\sigma_D = 0.3$. (d) $\mu_D = 1$ and $\sigma_D = 0.3$.

Figure 2: The probability density functions for the normal distribution and logistic distribution for $\mu_N = 0$, $\sigma_N = 1$ and $b \neq 1$, where the solid line represents the normal curve and the dashed line represents the logistic curve.

Table 1: TPR and TNR under $c = 0.5, 1.5, 2$ for the binormal model and bilogistic model assuming $\mu_N = 0$ and $\sigma_N = 1$.

μ_D	σ_D	c	Normal distribution		Logistic distribution	
			TPR	TNR	TPR	TNR
0.5	1	0.5	0.5000	0.6915	0.5000	0.7124
		2	0.0668	0.9772	0.0618	0.9741
1	1	0.5	0.6915	0.6915	0.7124	0.7124
		2	0.1587	0.9772	0.1402	0.9741
0.5	1.5	0.5	0.5000	0.6915	0.5000	0.7124
		1.5	0.2525	0.9331	0.2298	0.9382
1	1.5	0.5	0.6306	0.6915	0.6467	0.7124
		1.5	0.3694	0.9332	0.3533	0.9382
0.5	0.3	0.5	0.5000	0.6915	0.5000	0.7124
		1.5	0.0004	0.9332	0.0024	0.9382
1	0.3	0.5	0.9522	0.6915	0.9536	0.7124
		1.5	0.0478	0.9332	0.0464	0.9382

Conversely, when $\beta = 0.3$, that is, the sensitivity is more important than the specificity, we obtain OCV=-0.3473 and $C(\text{OCV}) = 0.7470$. Figure 3(b) shows a similar pattern for when the bilogistic model is considered, but $C(\text{OCV})$ is slightly larger and the OCV is moving towards small values. This result arises from a larger kurtosis for the logistic distribution.



(a) Binormal model. (b) Bilogistic model.

Figure 3: Relationship between cutoff values and C under the binormal model and bilogistic model under various combinations of (α, β), where $\circ$ indicates the point at (OCV, C(OCV)).

2.2. Special cases

Depending on the purpose of the test, the investigator might be more interested in the specificity as long as the sensitivity reaches a specific limit, or vice versa. That is, an investigator might want to have a diagnostic test in which the sensitivity is at least larger than a pre-specified value L, where $0 < L < 1$. Then, the OCV is obtained by maximizing the specificity under the constraint that the sensitivity is larger than L, i.e., $\text{TPR} \geq L$. Likewise, the OCV can be obtained by maximizing the sensitivity under the constraint that the specificity is larger than L, i.e., $\text{TNR} \geq L$. The following derives the boundary for the TPR and TNR under the binormal and bilogistic models. The following proofs can be obtained in a straightforward manner.

Theorem 2.3. *Assume that $F(\cdot)$ is a standard normal distribution function and that $L > 0$ is a pre-specified constant. Then,*

1. *When $L \leq$ TPR, upper bounds of c and the TNR are*

$$c \leq \mu_D - \sigma_N \Phi^{-1}(L),$$
$$\text{TNR} \leq \Phi\Big(\frac{\mu_D - \mu_N - \sigma_N \Phi^{-1}(L)}{\sigma_N}\Big).$$

Thus, the OCV equals $\mu_D - \sigma_N \Phi^{-1}(L)$.

2. *When $L \leq$ TNR, a lower bound of c and an upper bound of the TNR are given as*

$$c \geq \mu_N - \sigma_N \Phi^{-1}(1 - L),$$
$$\text{TNR} \leq \Phi\Big(\frac{\mu_D - \mu_N + \sigma_N \Phi^{-1}(1 - L)}{\sigma_N}\Big).$$

Thus, the OCV equals $\mu_N - \sigma_N \Phi^{-1}(1 - L)$.

Theorem 2.4. *Assume that $F(\cdot)$ is a bilogistic distribution function and that $L > 0$ is a pre-specified constant. Then,*

1. *When $L \leq$ TPR, upper bounds of c and the TNR are*

$$c \leq \mu_D - \gamma_N \log(\frac{L}{1-L}),$$
$$\text{TNR} \leq \frac{1}{1 + exp\Big(\frac{\mu_N - \mu_D + \gamma_N \log\left(\frac{L}{1-L}\right)}{\gamma_N}\Big)}.$$

Thus, the OCV equals $\mu_D - \gamma_N \log(\frac{L}{1-L})$.

2. *When $L \leq \mathrm{TNR}$, a lower bound of c and an upper bound of the TNR are given as*

$$c \geq \mu_N - \gamma_N \log\Big(\frac{1-L}{L}\Big),$$
$$\mathrm{TNR} \leq \frac{\exp(\frac{\mu_D-\mu_N+\gamma_N\log(\frac{1-L}{L})}{\gamma_N})}{1+\exp\Big(\frac{\mu_D-\mu_N+\gamma_N\log(\frac{1-L}{L})}{\sigma_N}\Big)}.$$

Thus, the OCV equals $\mu_N - \gamma_N \log\Big(\frac{1-L}{L}\Big)$.

3. NUMERICAL RESULTS

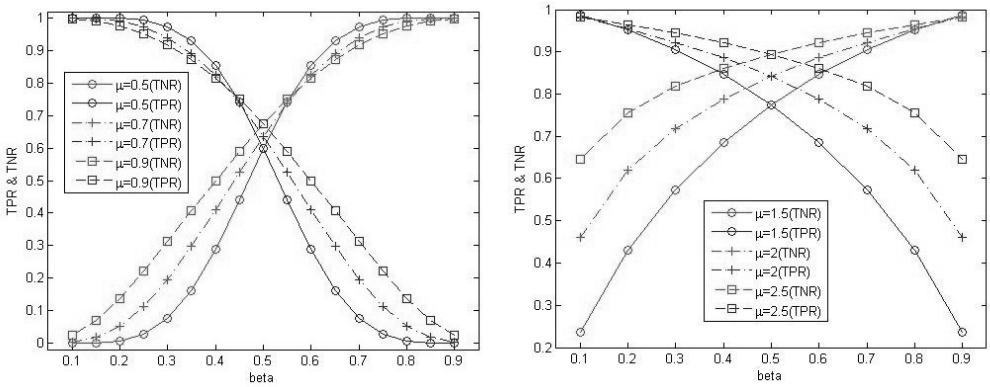
Based on the objective function defined in (1.6), Section 2 derives the OCV under the binormal and bilogistic models. When the binormal model is assumed, the OCV can be obtained explicitly, whereas under the bilogistic model, the OCV can be obtained explicitly only when b=1. The following discusses the OCV, TPR, and TNR under various settings for β and the location and scale parameters.

For simplicity, the standard normal distribution is assumed for the control population, i.e., $\mu_N = 0$ and $\sigma_N = 1$. Because the formula for determining the OCV varies with b, the following discussion considers $b = 1$ and $b \neq 1$ separately. For each scenario, the parameter setting is classified into two situations. The first scenario considers different values of μ_D given σ_D. The second scenario considers different values of σ_D given μ_D. Furthermore, the settings for μ_D and σ_D are discussed according to the effect size $\mathrm{ES} = \mu_D/\sigma_D$. Additionally, μ_D is assumed to be larger than μ_N. Moreover, because $\beta = 0$ and $\beta = 1$ correspond to special cases discussed in Section 2.2, the numerical results only consider $0.1 \leq \beta \leq 0.9$. Similar results for the bilogistic model are given in the Supplement.

3.1. Situation I when σ_D is fixed and μ_D is varied

The first situation discusses the numerical results when σ_D is fixed and ES is varied. For $\mathrm{ES} < 1$, μ_D equals 0.5, 0.7 and 0.9, whereas for $1 < \mathrm{ES}$, μ_D equals 1.5, 2 and 2.5. Figures 4(a)–4(b) display the relationship between TPR and TNR with respect to β when μ_D is varied and $\sigma_D = 1$. When β increases, the investigator is more interested in the TNR. As expected, the TNR increases while the TPR decreases. Increasing μ_D means that the difference in the testing result between two groups becomes more evident. Furthermore, for a fixed β

and σ_D, the OCV is a function of μ_D, as given in (2.8). Thus, as μ_D increases, the OCV increases, which corresponds to an increase in the TNR and a decrease in the TPR. Furthermore, due to a symmetric property, the OCV is located at TPR=TNR when $\beta = 0.5$. Table 2 presents the OCV, TPR and TPR for each scenario.



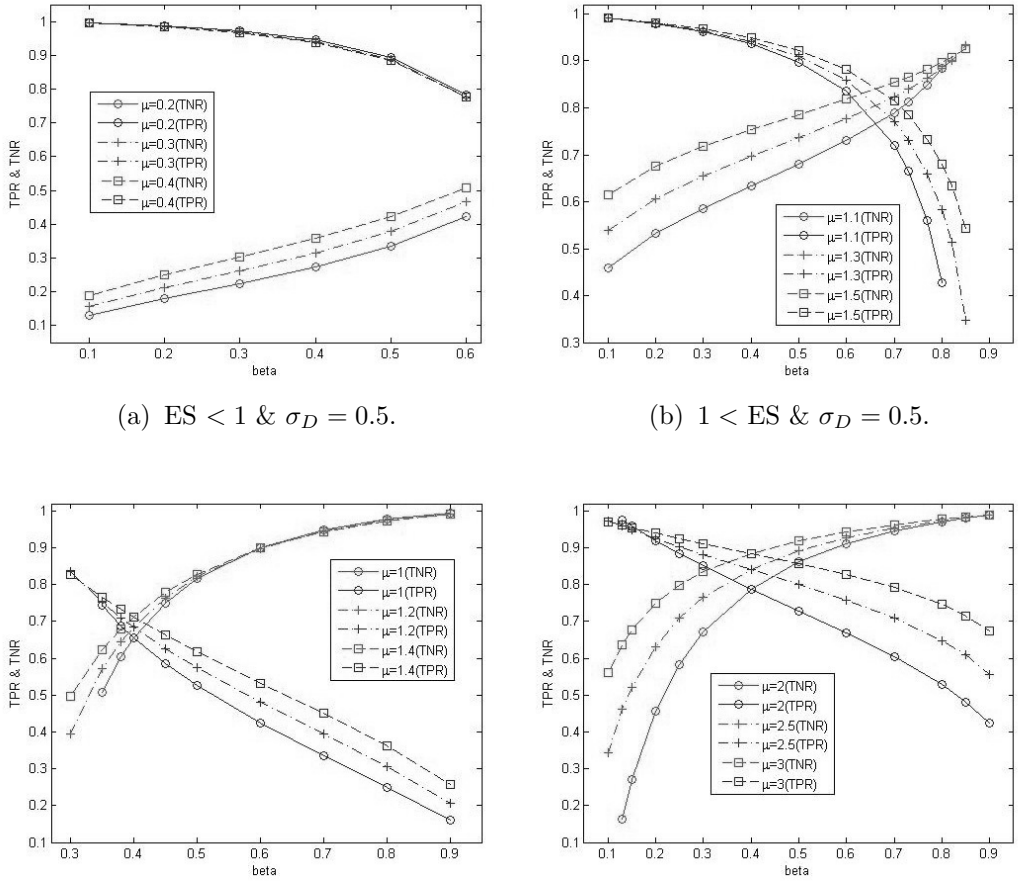
(a) ES < 1. (b) $1 <$ ES.

Figure 4: TNR and TPR at the OCV for various combinations of μ_D, β and ES under the binormal model and $b = 1$.

Table 2: Numerical results for TNR, TPR and OCV under the binormal model with various μ_Ds and $\sigma_D = 1$.

ES	μ_D	σ_D	Measures	β								
				0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.5	0.5	1	OCV	−4.1444	−2.5226	−1.4446	−0.5609	0.2500	1.0609	1.9446	3.0226	4.6444
			TPR	1.0000	0.9987	0.9741	0.8556	0.5987	0.2874	0.0743	0.0058	0.0000
			TNR	0.0000	0.0058	0.0743	0.2874	0.5987	0.8556	0.9741	0.9987	1.0000
0.7	0.7	1	OCV	−2.7889	−1.6304	−0.8604	−0.2292	0.3500	0.9292	1.5604	2.3304	3.4889
			TPR	0.9998	0.9901	0.9407	0.8236	0.6368	0.4093	0.1948	0.0515	0.0026
			TNR	0.0026	0.0515	0.1948	0.4093	0.6368	0.8236	0.9407	0.9901	0.9998
0.9	0.9	1	OCV	−1.9914	−1.0903	−0.4914	−0.0005	0.4500	0.9005	1.3914	1.9903	2.8914
			TPR	0.9981	0.9767	0.9180	0.8161	0.6736	0.4998	0.3116	0.1378	0.0232
			TNR	0.0232	0.1378	0.3116	0.4998	0.6736	0.8161	0.9180	0.9767	0.9981
1.5	1.5	1	OCV	−0.7148	−0.1742	0.1851	0.4797	0.7500	1.0203	1.3149	1.6742	2.2148
			TPR	0.9866	0.9530	0.9057	0.8462	0.7734	0.6843	0.5734	0.4309	0.2374
			TNR	0.2374	0.4309	0.5734	0.6843	0.7734	0.8462	0.9057	0.9530	0.9866
2	2	1	OCV	−0.0986	0.3069	0.5764	0.7973	1.0000	1.2027	1.4236	1.6931	2.0986
			TPR	0.9821	0.9548	0.9227	0.8855	0.8413	0.7874	0.7178	0.6205	0.4607
			TNR	0.4607	0.6205	0.7178	0.7874	0.8413	0.8855	0.9227	0.9548	0.9821
2.5	2.5	1	OCV	0.3711	0.6955	0.9111	1.0878	1.2500	1.4122	1.5889	1.8045	2.1289
			TPR	0.9834	0.9644	0.9440	0.9211	0.8944	0.8617	0.8189	0.7566	0.6447
			TNR	0.6447	0.7566	0.8189	0.8617	0.8944	0.9211	0.9440	0.9644	0.9834

Figures 5(a)–5(d) display the TPR and TNR at the OCV when β is varied and $\sigma_D \neq 1$. The pattern for the TPR and TNR with respect to β is no longer symmetric. Similar to $\sigma_D = 1$, as β increases, the TPR decreases and the TNR increases. However, the relationship between the TPR and TNR depends on σ_D, ES and β. When ES < 1 and $\sigma_D = 0.5$, the TPR is always larger than the TNR regardless of β. This is because $\sigma_D = 0.5$ means that the result obtained from the diseased group is more homogeneous, and the diagnostic test has a higher ability to detect a case even if ES < 1. However, when ES < 1 and $\sigma_D = 1.5$, the TPR is larger than the TNR only if $\beta < 0.4$. Furthermore, when ES > 1, the TPR is larger than the TNR only for some βs.



(a) ES < 1 & $\sigma_D = 0.5$. (b) $1 <$ ES & $\sigma_D = 0.5$.

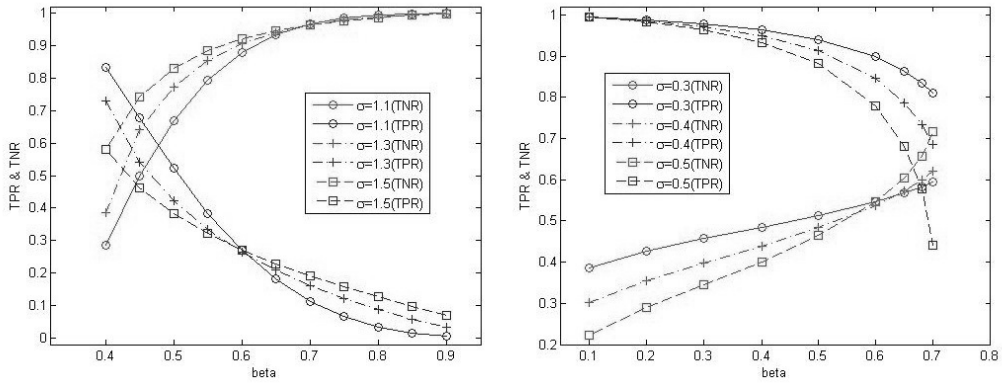
(c) ES < 1 & $\sigma_D = 1.5$. (d) $1 <$ ES & $\sigma_D = 1.5$.

Figure 5: TNR and TPR at the OCV when μ_D, σ_D, β and ES are varied, $b \neq 1$ and the binormal model are assumed.

3.2. Situation II when σ_D is varied and μ_D is fixed

Situation II provides numerical results for OCV, TPR and TNR when $\mu_D = 0.5$ and σ_D is varied. When $\mu_D = 0.5$, ES < 1 means that σ_D is larger than $\sigma_N = 1$, which means that it is easier to conclude a FN. Figure 6(a) shows the relationship between the TPR and TNR at the OCV with respect to β when σ_D is varied and ES < 1. The pattern of change for the TPR with respect to σ_D is related to β. When β increases, TPR expectedly decreases because β is the weight for the TNR. Nevertheless, when $0.5 < \beta$, the TPR becomes very small and slightly increases as σ_D increases. In addition, the TNR is large as long as $0.6 < \beta$, as listed in Figure 6(a).

When $\mu_D = 0.5$, $1 <$ ES means that σ_D is smaller than $\sigma_N = 1$, which indicates that it is easier to conclude a TP. Figure 6(b) displays the relationship between the TPR and TNR with respect to β when σ_D is varied and $1 <$ ES. Expectedly, as σ_D increases, the TPR decreases regardless of β. Unlike ES < 1, the relationship between the TNR and σ_D depends on β. When $\beta < 0.6$, the TNR decreases as σ_D increases, whereas when $0.6 < \beta$, the TNR increases as σ_D increases.



(a) ES < 1. (b) $1 <$ ES.

Figure 6: TNR and TPR at the OCV for various combinations of σ_D, β and ES under the binormal model and $\mu_D = 0.5$.

As β increases, the TNR is more important and results in a larger OCV. Table 3 demonstrates this trend. The impact of σ_D on the OCV is related to ES. When ES < 1, as σ_D increases, the OCV increases. Nevertheless, when ES > 1, the trend reverses.

Table 3: The relationship among OCV, TNR and TPR when the binormal model is assumed, $\mu_D = 0.5$ and σ_D is varied.

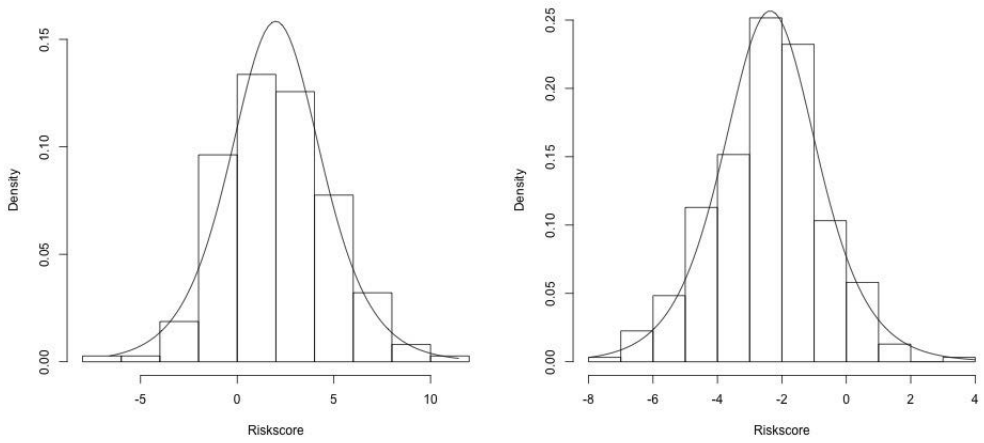
ES	μ_D	σ_D	Measures	β								
				0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.45	0.5	1.1	OCV	—	—	—	−0.5684	0.4400	1.1730	1.8288	2.5112	3.3878
			TPR	—	—	—	0.8343	0.5218	0.2703	0.1135	0.0337	0.0043
			TNR	—	—	—	0.2849	0.6700	0.8796	0.9663	0.9940	0.9996
0.38	0.5	1.3	OCV	—	—	—	−0.2929	0.7493	1.3147	1.7900	2.2693	2.8720
			TPR	—	—	—	0.7290	0.4239	0.2654	0.1605	0.0868	0.0340
			TNR	—	—	—	0.3848	0.7732	0.9057	0.9633	0.9884	0.9980
0.33	0.5	1.5	OCV	—	—	—	0.2000	0.9490	1.4109	1.8068	2.2097	2.7192
			TPR	—	—	—	0.5793	0.3824	0.2718	0.1918	0.1272	0.0695
			TNR	—	—	—	0.5793	0.8287	0.9209	0.9646	0.9864	0.9967
1.67	0.5	0.3	OCV	−0.2872	−0.1851	−0.1085	−0.0384	0.0344	0.1192	0.2368	—	—
			TPR	0.9957	0.9888	0.9787	0.9636	0.9397	0.8978	0.8098	—	—
			TNR	0.3870	0.4266	0.4568	0.4847	0.5137	0.5474	0.5936	—	—
1.25	0.5	0.4	OCV	−0.5196	−0.3711	−0.2583	−0.1532	−0.0417	0.0940	0.3072	—	—
			TPR	0.9946	0.9853	0.9710	0.9488	0.9122	0.8450	0.6851	—	—
			TNR	0.3017	0.3553	0.3981	0.4391	0.4833	0.5374	0.6207	—	—
1	0.5	0.5	OCV	−0.7609	−0.5570	−0.4001	−0.2518	−0.0904	0.1163	0.5753	—	—
			TPR	0.9942	0.9827	0.9641	0.9336	0.8812	0.7786	0.4401	—	—
			TNR	0.2234	0.2888	0.3445	0.4006	0.4640	0.5463	0.7175	—	—

— Numerical data are not available.

4. CASE STUDY

Early detection may improve the survival of patients with lung cancer. Chian *et al.* (2015) investigated peripheral blood mononuclear cell (PBMC)-derived gene expression signatures for their potential in the early detection of non-small cell lung cancer (NSCLC). PBMCs were obtained from 187 patients with NSCLC and from 310 non-cancer controls based on an age- and gender-matched case-control study. Controlling for gender, age and smoking status, 15 NSCLC-associated molecular markers were used to construct a risk score to distinguish subjects with lung cancer from controls. Detailed markers and the model construction are presented in Chian *et al.* (2016).

From the preventive perspective in health management, a higher sensitivity is preferred such that the disease can be detected earlier. Thus, β might be set to be smaller than 0.5. Nonetheless, cancer-specific clinicians often examine highly suspicious subjects. Thus, they may wish to have a higher specificity test. Figure 7 presents the histograms of the risk scores for the case and control groups for the PBMC data. The bilogistic model appears to be appropriate for these data. The maximum likelihood estimators of μ and γ are obtained for each group. The corresponding estimates of μ and γ for the case are 1.9911 and 1.5782 and those for the control are -2.3620 and 0.9739. Based on these estimates, the logistic density curves are plotted on top of the histogram in Figure 7.



(a) Case. (b) Control.

Figure 7: Histograms for risk scores for case and control groups for PBMC data, where the solid curve is the logistic density curve.

Under the bilogistic assumption, Table 4 lists the OCVs for β ranging from 0.1 to 0.9 for the risk score derived from the PBMC data. Figure 8 presents the corresponding TPR and TNR. For instance, when $\beta = 0.4$, the OCV equals -0.634. The test would expect to have equal chances at approximately 0.85 to identify a true positive or a true negative. Nevertheless, when $\beta = 0.6$, the test would have a higher chance to find a true negative.

Table 4: OCV for the PBMC data.

β	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
OCV	-2.864	-1.565	-1.027	-0.634	-0.291	0.044	0.409	0.861	1.581

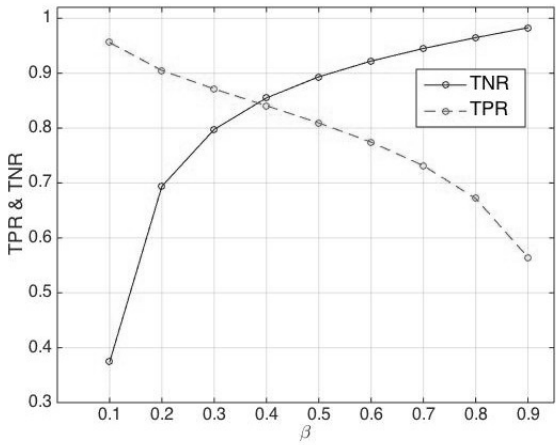


Figure 8: TPR and TNR under various βs for the PBMC data.

5. DISCUSSION AND CONCLUSION

The determination of the cutoff value is practically important. Because the ROC curve includes two important measures, TPR and TNR, to obtain the optimal operating point (OOP) or OCV, an additional objective function is required. One of two existing criteria can be regarded as the special case of the proposed criterion. The objective function C_3 requires information about the cost for the incorrect decision, which cannot be easily obtained. Furthermore, the OCV for this criterion is determined by setting the slope of the tangent line to the ROC curve to a pre-specified value (Halperm [3]). Because the slope is a function of the prevalence of the disease and costs, it is difficult to explain clinically (Kumar [5]).

The OCV is often obtained empirically (Kumar [5]). This paper derives the closed form for the OCV under the location and scale family. The binormal model is the most commonly used parametric assumption for the ROC curve. Under such an assumption, this paper provides exact formulas for the OCV. Furthermore, numerical results are presented under various scenarios. When $b = 1$, the TPR and TNR are related to the weight (β). In particular, increasing β means increasing the TNR. Nevertheless, when $b \neq 1$, regardless of β, the TNR might not be higher than 0.5. In particular, when the binomial model is violated, this paper provides another parametric choice, the bilogistic model. However, there is no closed form for the OCV. This paper provides a nonlinear equation for determining the OCV. In addition to discussing the OCV for the bilogistic model, the difference between these two parametric models is also addressed. The result of this paper can provide guidance for practitioners to choose the OCV.

Rather than choosing the OCV based on the sensitivity and specificity, Linnet *et al.* [7] suggested using the likelihood ratio

$$LR(c) = \frac{f(\frac{\mu_D - c}{\gamma_D})}{f(\frac{\mu_N - c}{\gamma_N})} \tag{5.1}$$

as an alternative for interpreting the test result. If (5.1) exceeds 1, then the relative frequency of the distribution of diseased individuals exceeds that of the normal individuals. In other words, given the index test result c, a respondent is more likely to have the disease. Their result can also be extended to the location and scale family.

APPENDIX: Proof of Theorem 2.1

Assume that F is the standard normal distribution function. To be consistent with the conventional notation, γ_D and γ_N are replaced by σ_D and σ_N, respectively. Therefore, (2.7) becomes

$$\text{(A.1)} \quad \frac{\partial C(c)}{\partial c} = \frac{-\alpha b}{\sqrt{2\pi}\sigma_N} \exp\Big(-\frac{[a + \frac{b(\mu_N - c)}{\sigma_N}]^2}{2}\Big) + \frac{\beta}{\sqrt{2\pi b}\sigma_N} \exp\Big(-\frac{b^2\big(c - \mu_N\big)^2}{2\sigma_N^2}\Big)$$

and set $\frac{\partial C(c)}{\partial c} = 0$ to obtain the OCV. An explicit formula for OCV can be determined and is dependent on b.

When $b = 1$, i.e., $\sigma_N^2 = \sigma_D^2$, the objective function and the corresponding derivative with respect to c are

$$\text{(A.2)} \qquad C = \alpha\Phi\Big(a + \frac{\mu_N - c}{\sigma_D}\Big) + \beta\Phi(\frac{c - \mu_N}{\sigma_D})$$

and

$$\text{(A.3)} \quad \frac{\partial C}{\partial c} = \frac{-\alpha}{\sqrt{2\pi}\sigma_D} \exp\left(-\frac{1}{2}\Big[a + \frac{\mu_N - c}{\sigma_D}\Big]^2\right) + \frac{\beta}{\sqrt{2\pi}\sigma_D} \exp\left(-\frac{1}{2}\Big(\frac{c - \mu_N}{\sigma_D}\Big)^2\right).$$

Let $\frac{\partial C}{\partial c} = 0$. We have

$$-\alpha b \exp\left(-\frac{[a\sigma_D + \mu_N - c]^2}{2\sigma_N^2}\right) + \beta \exp\left(-\frac{(c - \mu_N)^2}{2\sigma_D^2}\right) = 0,$$

which implies

$$\text{(A.4)} \qquad \log\Big(\frac{\alpha}{\beta}\Big) - \frac{[a\sigma_D + (\mu_N - c)]^2}{2\sigma_D^2} + \frac{(c - \mu_N)^2}{2\sigma_D^2} = 0.$$

After simplifying the preceding equation, we obtain

$$\frac{2(\mu_D - \mu_N)c + \mu_N^2 - \mu_D^2}{2\sigma_D^2} + \log(\frac{\alpha}{\beta}) = 0$$

and the OCV as given in (2.8).

When $b \neq 1$, the objective function and the corresponding derivative with respect to c are

$$C = \alpha\Phi\Big(a + b(\frac{\mu_N - c}{\sigma_N})\Big) + \beta\Phi\Big(\frac{c - \mu_N}{\sigma_N}\Big)$$

and

$$\frac{\partial C}{\partial c} = \frac{-\alpha b}{\sqrt{2\pi}\sigma_N} \exp\left(-\frac{1}{2}\left[a + b\left(\frac{\mu_N - c}{\sigma_N}\right)\right]^2\right) + \frac{\beta}{\sqrt{2\pi}\sigma_N} \exp\left(-\frac{1}{2}\left[\frac{c - \mu_N}{\sigma_N}\right]^2\right).$$

Let $\frac{\partial C}{\partial c} = 0$. We obtain

$$-\alpha b \exp\left(-\frac{[a\sigma_N + b(\mu_N - c)]^2}{2\sigma_N^2}\right) + \beta \exp\left[-\frac{(c-\mu_N)^2}{2\sigma_N^2}\right] = 0,$$

which implies

$$\log(\frac{\alpha b}{\beta}) - \frac{[a\sigma_N + b(\mu_N - c)]^2}{2\sigma_N^2} + \frac{(c-\mu_N)^2}{2\sigma_N^2} = 0. \tag{A.5}$$

Rearranging (A.5), we obtain

$$\frac{(1-b^2)}{2\sigma_N^2}c^2 - \frac{(\mu_N - ab\sigma_N - b^2\mu_N)}{\sigma_N^2}c + \frac{\mu_N^2 - (a\sigma_N + b\mu_N)^2}{2\sigma_N^2} + \log(\frac{\alpha b}{\beta}) = 0$$

and the OCV is equal to

$$c = \frac{T \pm \sqrt{T^2 - 2(1-b^2)R/\sigma_N^2}}{(1-b^2)/\sigma_N^2} \tag{A.6}$$

where R and T are defined in (2.10) and (2.11), respectively.

ACKNOWLEDGMENTS

This work has been supported by NSC102-2118-M-305-002-MY2 from the Ministry of Science and Technology, Taiwan. We also acknowledge the valuable suggestions from the referees.

REFERENCES

[1] Akobeng, A.K. (2007). Understanding diagnostic tests 3: receiver operating characteristic curves, *Acta Padiatrica*, **96**, 644–647.

[2] Chian, C.F.; Hwang, Y.T.; Terng, H.J. *et al.* (2016). Panels of tumor-derived RNA markers in peripheral blood of patients with non-small cell lung cancer: their dependence on age, gender and clinical stages. To appear in *Oncotarget*.

[3] Halperm, E.T.; Albert, M.; Krieger, A.M.; Metz, C.E. and Maidment, A.D. (1996). Comparison of receiver operating characteristic curves on the basis of optimal operating points, *Acad. Radiol.*, **3**, 245–253.

[4] Krzanowski, W.J. and Hand, D.J. (2009). *ROC Curves for Continuous Data*, CRC Press, New York.

[5] Kumar, R. and Indrayan, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers, *Indian Pediatrics*, **48**, 277–287.

[6] Lee, C.T. (2006). A solution for the most basic optimization problem associated with an ROC curve, *Statistical Methods for Medical Research*, **15**, 571–584.

[7] Linnet, K.; Bossuyt, P.M.M.; Moons, K.G.M. and Reitsma, J.B. (2012). Quantifying the accuracy of a diagnostic test or marker, *Clinical Chemistry*, **58**(9), 1292–1301.

[8] Metz, C.E. (1978). Basic principles of ROC analysis, *Seminars in Nuclear Medicine*, **8**(4), 283–298.

[9] Youden, W.J. (1950). Index for rating diagnostic tests, *Cancer*, **3**(1), 32–35.

REVSTAT – Statistical Journal
Volume 16, Number 1, January 2018, 45–70

ON THE q-GENERALIZED EXTREME VALUE DISTRIBUTION

Authors: Serge B. Provost
– Department of Statistical & Actuarial Sciences, The University of Western Ontario, London, Canada N6A5B7
provost@stats.uwo.ca

Abdus Saboor
– Department of Mathematics, Kohat University of Science & Technology, Kohat, Pakistan 26000
dr.abdussaboor@kust.edu.pk; saboorhangu@gmail.com

Gauss M. Cordeiro
– Departamento de Estatística, Universidade Federal de Pernambuco, 50740-540 Brazil
gausscordeiro@gmail.com

Muhammad Mansoor
– Department of Statistics, The Islamia University of Bahawalpur, Bahawalpur, Pakistan
mansoor.abbasi143@gmail.com

Received: December 2015 Revised: July 2016 Accepted: July 2016

Abstract:

- Asymmetrical models such as the Gumbel, logistic, Weibull and generalized extreme value distributions have been extensively utilized for modeling various random phenomena encountered for instance in the course of certain survival, financial or reliability studies. We hereby introduce q-analogues of the generalized extreme value and Gumbel distributions, the additional parameter q allowing for increased modeling flexibility. These extended models can yield several types of hazard rate functions, and their supports can be finite, infinite as well as bounded above or below. Closed form representations of some statistical functions of the proposed distributions are provided. It is also shown that they compare favorably to three related distributions in connection with the modeling of a certain hydrological data set. Finally, a simulation study confirms the suitability of the maximum likelihood method for estimating the model parameters.

Key-Words:

- *extreme value theory; generalized extreme value distribution; goodness-of-fit statistics; Gumbel distribution; moments; Monte Carlo simulations; q-analogues.*

AMS Subject Classification:

- 62E15, 62F10.

1. INTRODUCTION

Extreme value theory deals with the asymptotic behavior of extreme observations in a sample of realizations of a random variable. This theory can be applied to the prediction of the occurrence of rare events such as high flood levels, large jumps in the stock markets and sizeable insurance claims. It is based on the extremal types theorem which states that exactly three types of distributions, namely the Gumbel, Fréchet and Weibull models, referred to as types I, II and III extreme value distributions, can model the limiting distribution of properly normalized maxima (or minima) of sequences of independent and identically distributed random variables. As the *generalized extreme value* ($\mathcal{GEV}$) distribution, also called the Fisher–Tippett [10] distribution, encompasses all three types, it can be utilized as an approximation to model the maxima of long (finite) sequences of random variables. The $\mathcal{GEV}$ and Gumbel distributions are widely utilized in finance, actuarial science, hydrology, economics, material sciences, telecommunications, engineering, time series modelling, risk management, reliability analysis as well as several other fields of scientific investigation involving extreme events. For informative scholarly works on extreme value distributions and related results, the reader is referred to [5], [12], [15] and [7].

Being a limiting distribution, the $\mathcal{GEV}$ model may prove somewhat inadequate in practice, and generalizations thereof ought to provide greater flexibility for modeling purposes.The extended models being proposed in this paper, namely, the q-generalized extreme value and q-Gumbel distributions, are in fact q-analogues of the distributions of origin which are re-expressed in terms of an additional parameter denoted by q.

Mathai [17] developed a pathway model involving superstatistics, which arise in statistical mechanics in connection with the study of nonlinear and non-equilibrium systems. As explained for example in [8, 28], such systems exhibit spatio-temporal dynamics that are inhomogeneous and can be described by a "superposition of several statistics on different scales". The non-equilibrium steady-state macroscopic systems being considered are assumed to be made up of a large number of smaller cells that are temporarily in local equilibrium; moreover, each of these cells can take on a given value x of the variable of interest with probability density function $g(x)$ wherefrom one can determine the generalized Boltzmann factor, $B(\epsilon) = \int_0^\infty \mathrm{e}^{-\epsilon x} g(x)\,\mathrm{d}x$, ϵ denoting the energy of a microstate occurring within each cell. Such distributions are related to Tsallis statistics [27] which find applications in statistical mechanics, turbulence studies and Monte Carlo computational methods. Recently, several q-type superstatistical distributions such as the q-exponential, q-Weibull and q-logistic were developed in the context of statistical mechanics, information theory and reliability modelling, as discussed for instance in [30, 31, 20, 18, 14] and [21].

The cumulative distribution function (cdf) and probability density function (pdf) of the $\mathcal{GEV}$ distribution, including the Gumbel distribution as a limiting case wherein $\xi \to 0$, are respectively given by

$$F_1(x) = F_1(x;\mu,\sigma,\xi) = \begin{cases} \exp\left[-\left(1+\xi\left(\dfrac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right], & \xi \neq 0\,, \\ \exp\left[-\exp\left(-\left(\dfrac{x-\mu}{\sigma}\right)\right)\right], & \xi \to 0\,, \end{cases} \tag{1.1}$$

and

$$f_1(x) = f_1(x;\mu,\sigma,\xi) = \begin{cases} \dfrac{1}{\sigma}\left(1+\xi\left(\dfrac{x-\mu}{\sigma}\right)\right)^{(-1/\xi)-1} \\ \qquad \times \exp\left[-\left(1+\xi\left(\dfrac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right], & \xi \neq 0\,, \\ \dfrac{1}{\sigma}\exp\left[-\exp\left(-\left(\dfrac{x-\mu}{\sigma}\right)\right)\right] \\ \qquad \times \exp\left(-\left(\dfrac{x-\mu}{\sigma}\right)\right), & \xi \to 0\,, \end{cases}$$

where μ is a location parameter, σ is a positive scale parameter and ξ is the shape parameter. The support of the distribution is

$$x \in \begin{cases} (\mu-\sigma/\xi, \infty)\,, & \xi > 0\,, \\ (-\infty,\infty)\,, & \xi \to 0\,, \\ (-\infty, \mu-\sigma/\xi)\,, & \xi < 0\,. \end{cases} \tag{1.2}$$

On reparameterizing the $\mathcal{GEV}$ distribution by setting $m = \mu/\sigma$ and $s = \sigma^{-1}$ in (1.1) and (1.2), one has the following representations of the cdf and pdf:

$$F_2(x;s,m,\xi) = \begin{cases} \exp\Big[-\big(1+\xi(sx-m)\big)^{-1/\xi}\Big], & \xi \neq 0\,, \\ \exp\Big[-\exp\big(-(sx-m)\big)\Big], & \xi \to 0\,, \end{cases} \tag{1.3}$$

and

$$f_2(x;s,m,\xi) = \begin{cases} s\big(1+\xi(sx-m)\big)^{(-1/\xi)-1} \\ \qquad \times \exp\Big[-\big(1+\xi(sx-m)\big)^{-1/\xi}\Big], & \xi \neq 0\,, \\ s\exp\Big[-\exp\big(-(sx-m)\big)\Big] \\ \qquad \times \exp\big(-(sx-m)\big), & \xi \to 0\,. \end{cases}$$

The support then becomes

$$x \in \begin{cases} \left(\frac{m}{s} - \frac{1}{\xi s}, \infty\right), & \xi > 0\,, \\ (-\infty, \infty)\,, & \xi \to 0\,, \\ \left(-\infty, \frac{m}{s} - \frac{1}{\xi s}\right), & \xi < 0\,. \end{cases} \tag{1.4}$$

Paralleling the pathway approach advocated by Mathai [17], we now introduce the q-analogues of the $\mathcal{GEV}$ and Weibull distributions, namely, the *q-generalized extreme value* (q-$\mathcal{GEV}$) and q-Gumbel distributions. The cdf and pdf of the q-$\mathcal{GEV}$ and q-Gumbel (obtained by letting $\xi \to 0$ in the q-$\mathcal{GEV}$ model) distributions are respectively given by

(1.5)

$$F(x) = F(x; s, m, \xi, q) = \begin{cases} \left[1 + q\big(\xi(sx - m) + 1\big)^{-\frac{1}{\xi}}\right]^{-\frac{1}{q}}, & \xi \neq 0\,, \quad q \neq 0\,, \\ \left[1 + q\,\mathrm{e}^{-(sx-m)}\right]^{-\frac{1}{q}}, & \xi \to 0\,, \quad q \neq 0\,, \end{cases}$$

and

$$f(x) = f(x; s, m, \xi, q) = \begin{cases} s\big(1 + \xi(sx - m)\big)^{-\frac{1}{\xi}-1} \\ \quad \times \left[1 + q\big(\xi(sx - m) + 1\big)^{-\frac{1}{\xi}}\right]^{-\frac{1}{q}-1}, & \xi \neq 0\,, \quad q \neq 0\,, \\ s\,\mathrm{e}^{m-sx}\left[1 + q\,\mathrm{e}^{m-sx}\right]^{-\frac{1}{q}-1}, & \xi \to 0\,, \quad q \neq 0\,, \end{cases}$$

where the support of the distributions is as follows:

$$x \in \begin{cases} \left(\frac{m}{s} - \frac{1}{\xi s}, \infty\right), & q > 0\,, \quad \xi > 0\,, \\ \left(-\infty, \frac{m}{s} - \frac{1}{\xi s}\right), & q > 0\,, \quad \xi < 0\,, \\ \left(\frac{(-q)^{\xi} - 1}{\xi s} + \frac{m}{s}, \infty\right), & q < 0\,, \quad \xi > 0\,, \\ \left(\frac{(-q)^{\xi} - 1}{\xi s} + \frac{m}{s}, \frac{m}{s} - \frac{1}{\xi s}\right), & q < 0\,, \quad \xi < 0\,, \\ (-\infty, \infty)\,, & \xi \to 0\,, \quad q > 0\,, \\ \left(\frac{m + \ln(-q)}{s}, \infty\right), & \xi \to 0\,, \quad q < 0\,. \end{cases} \tag{1.6}$$

The intervals specifying the supports of these distributions are such that the terms being raised to non-integer powers remain positive for the respective domains of q and ξ.

The effects of the parameters q and ξ on the shape of the distributions are illustrated graphically in Figures 1 to 5. Plots of the hazard rates of X are displayed in Figures 6 and 7 for certain parameter values. These plots illustrate the impressive versatility of the proposed models.

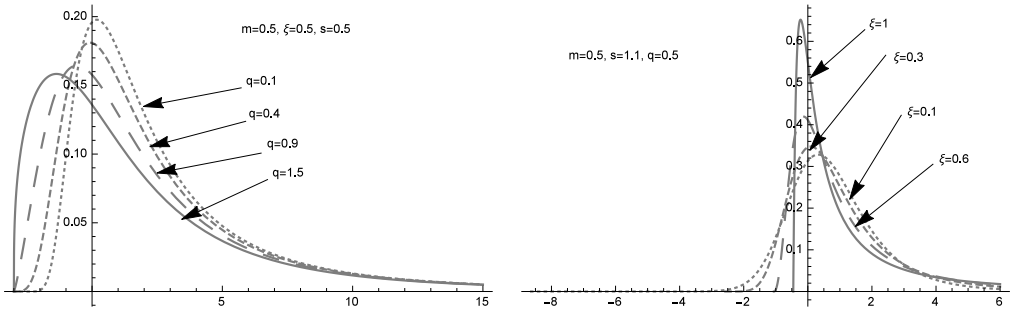


Figure 1: Plots of the q-$\mathcal{GEV}$ density function for certain parameter values $(q > 0,\ \xi > 0)$.

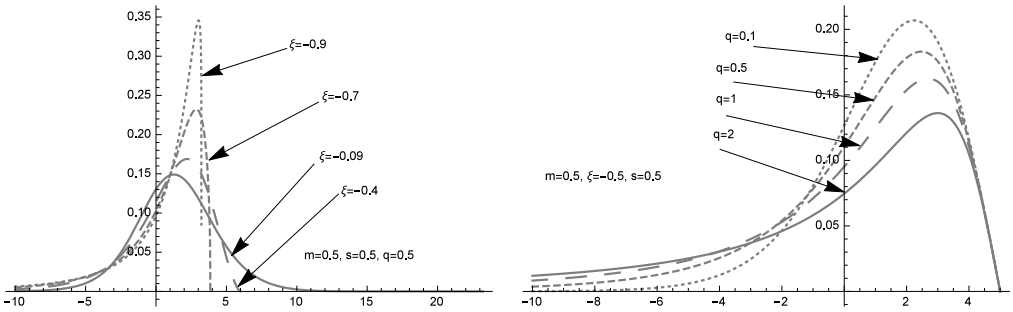


Figure 2: Plots of the q-$\mathcal{GEV}$ density function for certain parameter values $(q > 0,\ \xi < 0)$.

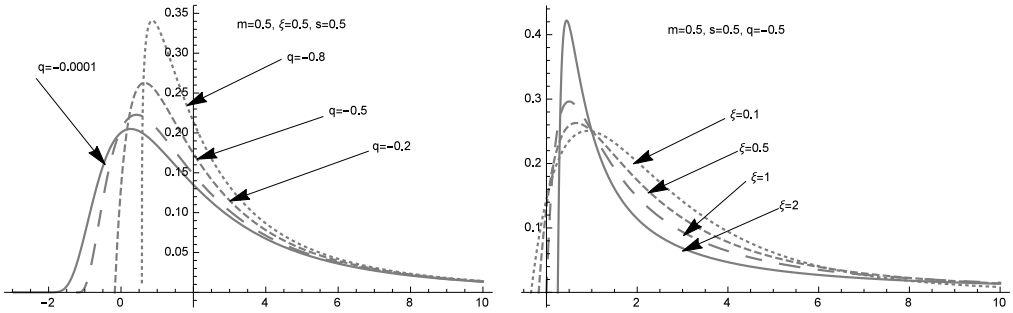


Figure 3: Plots of the q-$\mathcal{GEV}$ density function for certain parameter values $(q < 0,\ \xi > 0)$.

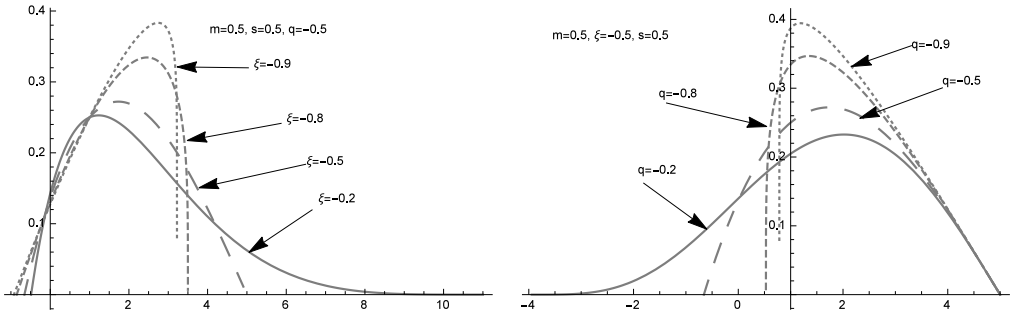


Figure 4: Plots of the q-$\mathcal{GEV}$ density function for certain parameter values ($q < 0,\ \xi < 0$).

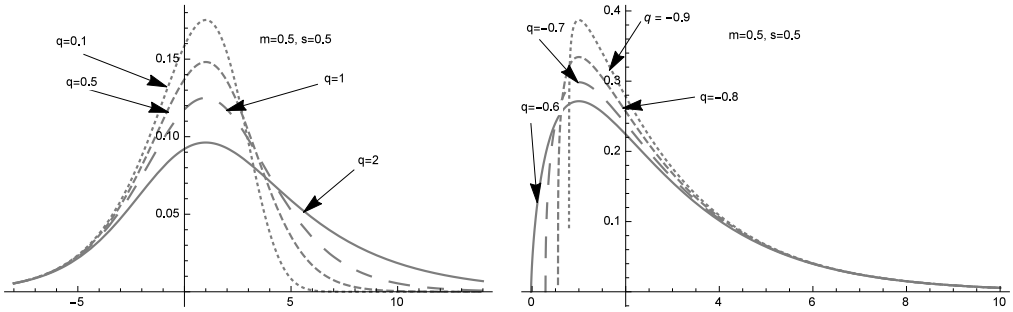


Figure 5: Plots of the q-Gumbel density function for certain parameter values. Right panel: $q > 0$; Left panel: $q < 0$.

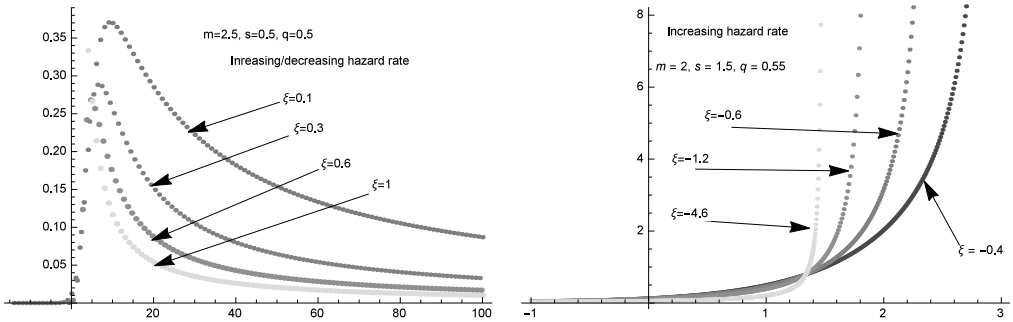


Figure 6: Plots of the q-$\mathcal{GEV}$ hazard rates for certain parameter values. Right panel: $\xi < 0$, $q > 0$; Left panel: $\xi > 0$, $q > 0$.

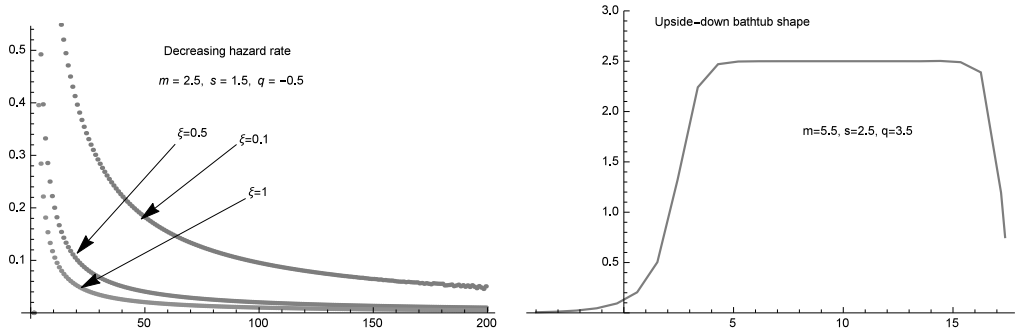


Figure 7: Plots of the q-$\mathcal{GEV}$ (left panel) and q-Gumbel (right panel) hazard rates for certain parameter values.

Remark 1.1. The $\mathcal{GEV}$ and Gumbel distributions are respectively obtained as limiting cases of the q-$\mathcal{GEV}$ and q-Gumbel distributions by letting q approach zero.

The paper is organized as follows. Section 2 contains computable representations of certain statistical functions of the q-$\mathcal{GEV}$ and q-Gumbel distributions. Section 3 explains how to determine the maximum likelihood estimators of the model parameters. In Section 4, the proposed distributions as well as three related models are fitted to an actual data set, and several statistics are utilized to assess goodness of fit. A Monte Carlo simulation study is carried out in Section 5 to verify the accuracy of the maximum likelihood estimates. Finally, some concluding remarks are included in the last section.

2. CERTAIN STATISTICAL FUNCTIONS

This section includes certain computable representations of the ordinary moments and the L-moments of the q-Gumbel (s, m, q) and q-$\mathcal{GEV}(s, m, \xi, q)$ random variables, which were obtained by making use of the symbolic computation package *Mathematica*. Closed form representations of their quantile functions as well as the moment-generating function of the q-Gumbel distribution are also provided. Whenever such closed form representations could be determined, the numerical results were found to agree to at least five decimals with those evaluated by numerical integration. Thus, numerical integration can arguably be employed to evaluate any required statistical function with great accuracy.The following identity can be particularly useful for evaluating the expected value of an integrable function of a continuous random variable denoted by $W(X)$:

$$E[W(X)] = \int_{-\infty}^{\infty} W(x)\, f(x)\, \mathrm{d}x = \int_{0}^{1} W(Q_X(p))\, \mathrm{d}p\,,$$

where $f(x)$ is the pdf of X and $Q_X(p)$ denotes the quantile function of X as defined in Section 2.1.

2.1. The quantile function

The quantile function is frequently utilized for determining confidence intervals or eliciting certain properties of a distribution. In order to obtain the quantile function of a random variable X, that is,

$$Q_X(p) = \inf\Big\{x \in \mathbb{R}\colon p \le F(x)\Big\}\,, \qquad p \in (0,1)\,,$$

one has to solve the equation $F(x) = p$ with respect to x for some fixed $p \in (0,1)$, where $F(x)$ denotes the cdf of X.

The following quantile functions of the q-$\mathcal{GEV}$ ($\xi \neq 0$) and q-Gumbel ($\xi \to 0$) can be readily obtained from their cdf's as specified by Equation (1.5):

$$(2.1)\; x_p \equiv Q_X(p) = F^{-1}(p) = \begin{cases} \dfrac{m}{s} + \dfrac{1}{s\xi}\left[\left(\dfrac{p^{-q}-1}{q}\right)^{-\xi} - 1\right], & \xi \neq 0\,, \\[2ex] \dfrac{m}{s} - \dfrac{1}{s}\ln\left(\dfrac{p^{-q}-1}{q}\right), & \xi \to 0\,. \end{cases}$$

2.2. Moments

Many key characteristics of a distribution can be inferred from its central moments. We first determine conditions under which the integer moments of the q-$\mathcal{GEV}$ distribution are finite. In light of the relationship given in the introduction of this section and the representation of quantile function of the q-$\mathcal{GEV}$ distribution specified by Equation (2.1), the k^{th} moment of this distribution can be evaluated as

$$\int_0^1 \left(\frac{1}{\xi}\right)^k \left(\left(\frac{p^{-q}-1}{q}\right)^{-\xi} - 1\right)^k \mathrm{d}p\,.$$

It is assumed without any loss of generality that $m = 0$ and $s = 1$. On applying the binomial expansion to $\big(\big(\frac{p^{-q}-1}{q}\big)^{-\xi} - 1\big)^k$, the k^{th} moment is expressible as a linear combination of the integrals

$$\int_0^1 \left(\frac{p^{-q}-1}{q}\right)^{j(-\xi)} \mathrm{d}p\,, \qquad j = 0, 1, ..., k\,.$$

Letting $\tau = \xi j$ and integrating, *Mathematica* provides the following condition for the existence of the integral when q is positive: $-\frac{1}{q} < \tau < 1$ or $\xi j < 1$ and $\xi j > -1/q$.
If q is negative, the condition for the existence of the k^{th} moment is $\tau < 1$, that is, $\xi j < 1$, $j = 1, ..., k$. Thus the conditions for the existence of the positive integer

moments of the q-$\mathcal{GEV}$ distribution are as follows: $\xi < \frac{1}{k}$ whenever $q > 0$ and $\xi > 0$; $\xi > -1/(kq)$ whenever $q > 0$ and $\xi < 0$; $\xi < 1/k$ whenever $q < 0$ and $\xi > 0$; no requirement being necessary when q and ξ are both negative.
Moreover, as in the case of the Gumbel distribution, the positive integer moments of the q-Gumbel distribution are finite whether q is positive or negative.
As determined by symbolic computations, the n^{th} ordinary moment of the q-$\mathcal{GEV}$ distribution can be expressed as follows:

$$\begin{aligned} E(X^n) &= \frac{(-1)^n}{\xi^n} - \frac{\sum_{i=0}^{n-1}(-1)^{i+1}\binom{n}{i}\Gamma\big(1-(n-i)\,\xi\big)\left(\frac{1}{q}\right)^{1-\xi(n-i)}}{\xi^n\,\Gamma\Big(1+\frac{1}{q}\Big)} \\ &\quad\times\ \Gamma\left((n-i)\,\xi+\frac{1}{q}\right), \qquad q>0\,, \\ &= \frac{1}{s^n}\left[\frac{(m\xi-1)^n}{\xi^n} - \Gamma\left(\frac{q-1}{q}\right)\sum_{i=0}^{n-1}\frac{c_i\,(m\xi-1)^i\,(-q)^{\xi(n-i)}}{\xi^{n-1}\,\Gamma\big(-(n-i)\,\xi-\frac{1}{q}+1\big)}\right. \\ &\quad\left.\times\,\big(I_{i\neq 0}-(1/\xi)\,I_{i=0}\big)\,\Gamma\big(I_{i=0}-(n-i)\,\xi\big)\right], \qquad q<0\,, \end{aligned}$$

where I denotes the indicator function and the c_i's are such that $c_i = 1$ if $i = 0$, $c_i = n!/(i!\,(n-i-1)!)$ if $1 \le i \le (n-1)/2$ and $c_i = n!/i!$ if $i > (n-1)/2$.

A necessary condition for the existence of the n^{th} moment of X is $\xi < 1/n$. The representation obtained for $q < 0$ also requires that $q\xi$ be greater than $-(1/n)$. As previously pointed out, numerical integration will provide accurate results when a closed form representation is unavailable.

It should be noted that, for instance, letting Y have a q-Gumbel distribution with pdf $f(y;1,0,q)$, is straightforward to determine the h^{th} moment of $X = (m+Y)/s$ — whose pdf is $f(x;s,m,q)$ — in terms of the first h moments of Y since

$$E(X^h) = \frac{1}{s^h}\sum_{j=0}^{h}\binom{h}{j}m^{h-j}\,E(Y^j)\,.$$

When q is positive, the h^{th} moment of the q-Gumbel distribution whose parameters m and s are respectively 0 and 1, is given by

$$\begin{aligned} E(X^h) &= h!\left[{}_{h+2}F_{h+1}\left(1,...,1,\frac{1}{q}+1;2,...,2;-q\right)\right. \\ &\quad\left.+\,(-1)^h\,q_{h+1}^{h-\frac{1}{q}}\,F_h\left(\frac{1}{q},...,\frac{1}{q};\frac{1}{q}+1,...,\frac{1}{q}+1;-\frac{1}{q}\right)\right], \end{aligned} \tag{2.2}$$

where the generalized hypergeometric function ${}_pF_q(a;b;z)$ admits the power series $\sum_{k=0}^{\infty}\frac{(a_1)_k\cdots(a_p)_k}{(b_1)_k\cdots(b_q)_k\,k!}\,z^k$.

The following closed form representation of the moment-generating function of the q-Gumbel distribution wherein $m = 0$, $s = 1$ was obtained assuming that $q < 0$:

$$M(t) \;=\; \frac{\Gamma\Big(\frac{q-1}{q}\Big)\,\Gamma(1-t)\,(-q)^t}{\Gamma\Big(-t-\frac{1}{q}+1\Big)}\,.$$

The h^{th} moment of this distribution when its parameter q is negative can then be obtained by differentiating $M(t)$. For instance when $q < 0$, the first and second moments of the q-Gumbel distribution are

$$E(X) \;=\; H_{-\frac{1}{q}} + \log(-q)$$

and

$$E(X^2) \;=\; \Big(H_{-\frac{1}{q}} + \log(-q)\Big)^2 - \psi^{(1)}\left(\frac{q-1}{q}\right) + \frac{\pi^2}{6}\,,$$

where H_δ denotes the Harmonic function $\int_0^1 \frac{1-x^\delta}{1-x}\,\mathrm{d}x$ and $\psi^{(1)}(\cdot)$ is the digamma function.

2.3. L-Moments

Unlike the conventional moments, the L-moments of a random variable whose mean is finite always exist, which explains their frequent use in extreme value theory. Since L-moments can be evaluated as linear combinations of probability weighted moments, which are defined for instance in [4], we first determine the latter.

The m^{th} order probability weighted moment of the q-Gumbel distribution is given by

$$\begin{aligned}
\beta_m \;&=\; \int_{-\infty}^{\infty} y\,F(y)^r\,\mathrm{d}F(y) \\
&=\; \frac{1}{s}\left[\mathrm{e}^m\,{}_3F_2\Big(1,1,\frac{k}{q}+\frac{1}{q}+1;2,2;-q\,\mathrm{e}^m\Big)\right. \\
&\quad -\frac{q\,(q+\mathrm{e}^{-m})^{\frac{k+1}{q}}\,\big(q\,(\mathrm{e}^m q+1)\big)^{-\frac{k+1}{q}}}{(k+1)^2} \\
&\quad \left.\times\;{}_2F_1\Big(\frac{k+1}{q},\frac{k+1}{q};\frac{k+q+1}{q};-\frac{\mathrm{e}^{-m}}{q}\Big)\right], \qquad q>0\,, \qquad (2.3)\\
&=\; \Re\left(\frac{\mathrm{e}^{\frac{i(k+1)\pi}{q}}(k+1)\,\pi\,\csc\Big(\frac{(k+1)\pi}{q}\Big)-\Big(\frac{1}{q}\Big)^{\frac{k-q+1}{q}}\;{}_2F_1\Big(\frac{k+1}{q},\frac{k+1}{q};\frac{k+q+1}{q};-\frac{1}{q}\Big)}{(k+1)^2}\right. \\
&\quad \left.+\;{}_3F_2\Big(1,1,\frac{k}{q}+\frac{1}{q}+1;2,2;-q\Big)\right), \qquad q<0\,,
\end{aligned}$$

where m is a nonnegative integer, $i=\sqrt{-1}$ and $\Re(s)$ denotes the real part of s. The first four L-moments of the q-Gumbel distribution are then obtained as follows: $\lambda_1=\beta_0,\ \lambda_2=2\beta_1-\beta_0,\ \lambda_3=6\beta_2-6\beta_1+\beta_0$ and $\lambda_4=20\beta_3-30\beta_2+12\beta_1-\beta_0$.

The L-moments of the q-$\mathcal{GEV}$ distribution, as well as other statistical functions of either of the newly introduced distributions, such as incomplete moments and mean deviations, can readily and accurately be evaluated by numerical integration. All the expressions included in this section were verified numerically for several values of the parameters, the code being available upon request.

3. MAXIMUM LIKELIHOOD ESTIMATION AND GOODNESS-OF-FIT STATISTICS

The parameters of the q-$\mathcal{GEV}$ and q-Gumbel distributions are estimated by making use of the maximum likelihood method. As well, several goodness-of-fit statistics to be utilized in Section 4 are defined in this section.

3.1. Maximum Likelihood Estimation

In order to estimate the parameters of the q-$\mathcal{GEV}$ and q-Gumbel distributions whose density functions are as specified in Equation (1.6), one has to maximize their respective log-likelihood functions with respect to the model parameters. Given the observations $x_i,\ i=1,...,n$, the log-likelihood functions of the q-$\mathcal{GEV}$ and q-Gumbel models are respectively given by

$$\begin{aligned}\ell(s,m,\xi,q) \;=\; & n\log(s)+\left(-\frac{1}{q}-1\right)\sum_{i=1}^{n}\log\Big(q\big(\xi\,(s\,x_i-m)+1\big)^{-1/\xi}+1\Big)\\ & +\;\left(-\frac{1}{\xi}-1\right)\sum_{i=1}^{n}\log\big(\xi\,(s\,x_i-m)+1\big)\,,\end{aligned}\tag{3.1}$$

whenever $\xi\neq 0$ and

$$\ell(s,m,q) \;=\; n\log(s)+\sum_{i=1}^{n}\log(s\,x_i-m)+\left(-\frac{1}{q}-1\right)\sum_{i=1}^{n}\log(1+q\,\mathrm{e}^{m-s x_i})$$

as $\xi\to 0$.

The associated log-likelihood system of equations are respectively

$$\begin{aligned}
\frac{\partial \ell(s,m,\xi,q)}{\partial s} &= \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n} -\frac{q x_i \big(\xi\,(s x_i - m)+1\big)^{-\frac{1}{\xi}-1}}{q\big(\xi\,(s x_i - m)+1\big)^{-\frac{1}{\xi}}+1} \\
&\quad + \left(-\frac{1}{\xi}-1\right)\sum_{i=1}^{n} \frac{\xi x_i}{\xi\,(s x_i - m)+1} + \frac{n}{s} \;=\; 0\,, \\
\frac{\partial \ell(s,m,\xi,q)}{\partial m} &= \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n} \frac{q\big(\xi\,(s x_i - m)+1\big)^{-\frac{1}{\xi}-1}}{q\big(\xi\,(s x_i - m)+1\big)^{-\frac{1}{\xi}}+1} \\
&\quad + \left(-\frac{1}{\xi}-1\right)\sum_{i=1}^{n} -\frac{\xi}{\xi\,(s x_i - m)+1} \;=\; 0\,, \\
\frac{\partial \ell(s,m,\xi,q)}{\partial \xi} &= \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n} \frac{q\big(\xi\,(s x_i - m)+1\big)^{-\frac{1}{\xi}}}{q\big(\xi\,(s x_i - m)+1\big)^{-\frac{1}{\xi}}+1} \\
&\quad \times \left(\frac{\log\big(\xi\,(s x_i - m)+1\big)}{\xi^2} - \frac{s x_i - m}{\xi\big(\xi\,(s x_i - m)+1\big)}\right) \\
&\quad + \frac{\sum_{i=1}^{n}\log\big(\xi\,(s x_i - m)+1\big)}{\xi^2} \\
&\quad + \left(-\frac{1}{\xi}-1\right)\sum_{i=1}^{n} \frac{s x_i - m}{\xi\,(s x_i - m)+1} \;=\; 0\,, \qquad (3.2) \\
\frac{\partial \ell(s,m,\xi,q)}{\partial q} &= \frac{\sum_{i=1}^{n}\log\Big(q\big(\xi\,(s x_i - m)+1\big)^{-\frac{1}{\xi}}+1\Big)}{q^2} \\
&\quad + \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n} \frac{\big(\xi\,(s x_i - m)+1\big)^{-\frac{1}{\xi}}}{q\big(\xi\,(s x_i - m)+1\big)^{-\frac{1}{\xi}}+1} \;=\; 0
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \ell(s,m,q)}{\partial s} &= \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n} -\frac{q x_i\, \mathrm{e}^{m-s x_i}}{q\,\mathrm{e}^{m-s x_i}+1} - \sum_{i=1}^{n} x_i + \frac{n}{s} \;=\; 0\,, \\
\frac{\partial \ell(s,m,q)}{\partial m} &= \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n} \frac{q\,\mathrm{e}^{m-s x_i}}{q\,\mathrm{e}^{m-s x_i}+1} + n\,, \qquad (3.3) \\
\frac{\partial \ell(s,m,q)}{\partial q} &= \frac{\sum_{i=1}^{n}\log(q\,\mathrm{e}^{m-s x_i}+1)}{q^2} + \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n} \frac{\mathrm{e}^{m-s x_i}}{q\,\mathrm{e}^{m-s x_i}+1}\,.
\end{aligned}$$

Solving the nonlinear systems specified by the sets of equations (3.2) and (3.3) respectively yields the maximum likelihood estimates ($\mathcal{MLE}$'s) of the parameters of the q-$\mathcal{GEV}$ and q-Gumbel distributions. Since these equations cannot

be solved analytically, iterative methods such as the Newton–Raphson technique are required. For both distributions, all the second order log-likelihood derivatives exist. In order to determine approximate confidence intervals for the parameters of the q-$\mathcal{GEV}$ and q-Gumbel distributions, one needs the 4×4 and 3×3 observed information matrices which are obtained by taking the opposite of the matrices of the second derivatives of the loglikelihood functions wherein the parameters are replaced by the $\mathcal{MLE}$'s, these matrices being denoted by $J(\upsilon_1) = \{J(\upsilon_1)_{rt}\}$ for $r, t = s, m, \xi, q$, where υ_1 denotes the vector of the parameters s, m, ξ, q, and $J(\upsilon_2) = \{J(\upsilon_2)_{rt}\}$ for $r, t = s, m, q$, where υ_2 is a vector whose components are s, m, q. Under standard regularity conditions, $(\upsilon_1 - \widehat{\upsilon}_1)$ asymptotically follows the multivariate normal distribution $\mathcal{N}_4(O, -J(\widehat{\upsilon}_1)^{-1})$ and the asymptotic distribution of $(\upsilon_2 - \widehat{\upsilon}_2)$ is $\mathcal{N}_3(O, -J(\widehat{\upsilon}_2)^{-1})$. These distributions can be utilized to construct approximate confidence intervals for the model parameters. Thus, denoting for example the total observed information matrix evaluated at $\widehat{\upsilon}_1$, that is, $-J(\widehat{\upsilon}_1)$, by $-\widehat{J}$, one would have the following approximate $100(1-\alpha)\%$ confidence intervals for the parameters of the q-$\mathcal{GEV}$ distribution

$$\widehat{s} \pm z_{\phi/2}\sqrt{(-\widehat{J}^{-1})_{ss}}\,, \qquad \widehat{m} \pm z_{\phi/2}\sqrt{(-\widehat{J}^{-1})_{mm}}\,,$$

$$\widehat{\xi} \pm z_{\phi/2}\sqrt{(-\widehat{J}^{-1})_{\xi\xi}}\,, \qquad \widehat{q} \pm z_{\phi/2}\sqrt{(-\widehat{J}^{-1})_{qq}}\,,$$

where $z_{\alpha/2}$ denotes the $100\,(1-\alpha/2)^{\text{th}}$ percentile of the standard normal distribution. The observed information matrices for the q-$\mathcal{GEV}$ and q-Gumbel models are provided in Appendices A and B.

One can determine the global maximum of the log-likelihood functions by setting certain initial values for the parameters in the maximizing routine being used. To that end, one could for instance make use of estimates of the parameters obtained for a sub-model such as those of the $\mathcal{GEV}$ distribution when assigning initial values to the parameters s, m, ξ of the q-$\mathcal{GEV}$ distribution. While Park and Sohn [23] obtained parameter estimates for the $\mathcal{GEV}$ distribution by making use of generalized weighted least squares and estimates of the three parameters are given in Chapter 30 of [4] in terms of probability weighted moments, Prescott and Walden [24] advocated the use of the maximum likelihood approach. It should be noted that, for both distributions under consideration, the $\mathcal{MLE}$'s do not appear to be particularly sensitive to the initial parameter values.

3.2. Goodness-of-fit statistics

In order to assess the relative adequacy of competing models, one has to rely on certain goodness-of-fit statistics. These may include the log-likelihood function evaluated at the $\mathcal{MLE}$'s denoted by $\widehat{\ell}$, Akaike's information criterion (AIC), the

corrected Akaike information criterion (CAIC), as well as the modified Anderson–Darling (A^*), the modified Cramér–von Mises (W^*) and the Kolmogrov–Smirnov (K–S) statistics. The smaller these statistics are, the better the fit. The AIC and AICC statistics are respectively given by

$$\text{AIC} = -2\ell(\widehat{\theta}) + 2p \qquad \text{and} \qquad \text{AICC} = \text{AIC} + \frac{2p\,(p+1)}{n-p-1}\,,$$

where $\ell(\widehat{\theta})$ denotes the log-likelihood function evaluated at the $\mathcal{MLE}$'s, p is the number of estimated parameters and n, the sample size.

The Anderson–Darling and Cramér–von Mises statistics can be evaluated by means of the following formulae:

$$A^* = \left(\frac{2.25}{n^2} + \frac{0.75}{n} + 1\right)\left[-n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\log\bigl(z_i\,(1-z_{n-i+1})\bigr)\right],$$

and

$$W^* = \left(\frac{0.5}{n} + 1\right)\left[\sum_{i=1}^{n}\left(z_i - \frac{2i-1}{2n}\right)^2 + \frac{1}{12n}\right],$$

where $z_i = \text{cdf}(y_i)$, the y_i's denoting the ordered observations.

As for the Kolmogrov–Smirnov statistic, it is defined by

$$\text{K–S} = \text{Max}\left[\frac{i}{n} - z_i,\, z_i - \frac{i-1}{n}\right].$$

As is explained in [2], unlike the asymptotic distributions of the AIC and AICC statistics, those of the A^* and W^* statistics have complicated forms requiring numerical techniques for determining specific percentiles.

4. APPLICATIONS

4.1. A hydrological data set

In this section, we fit five models to a rainfall precipitation data set which is freely available on the Korea Meteorological Administration (KMA) website `http://www.kma.go.kr` and represent the annual maximum daily rainfall amounts in millimeters in Seoul, Korea during the period 1961–2002. The selected models are the three-parameter $\mathcal{GEV}$, the Kumaraswamy generalized extreme value (Kum$\mathcal{GEV}$) [9], the exponentiated generalized Gumbel (EGGu) [3], and the newly introduced q-$\mathcal{GEV}$ and q-Gumbel distributions. Then, five statistics are employed

in order to assess goodness of fit. Table 1 displays certain descriptive statistics associated with the set of observations under consideration.

Table 1: Descriptive statistics for the Seoul rainfall data.

Mean	Median	SD	Kurtosis	Skewness	MD – mean	MD – median	Entropy
144.599	131.6	66.1781	3.80435	0.940673	48.7761	33.2	4.61435

MD := Mean deviation

The Kum$\mathcal{GEV}$ and EGGu density functions are respectively given by

$$f(x;a,b,\xi,\sigma,\mu) \;=\; \frac{1}{\sigma}\, a\, b u \exp(-a\, u)\left[1-\exp(-a\, u)\right]^{b-1},$$

where $u = \{1+\xi(x-\mu)/\sigma\}^{-1/\xi}$ with x such that $(1+\xi(x-\mu)/\sigma) > 0$; $a > 0$, $b > 0$, $\xi \in \mathbb{R}$, $\sigma > 0$ and $\mu \in \mathbb{R}$, and

$$f(x;\sigma,\mu,\alpha,\beta) \;=\; \frac{\alpha\beta}{\sigma}\, \mathrm{e}^{-\left(\frac{x-\mu}{\sigma}+\mathrm{e}^{\frac{\mu-x}{\sigma}}\right)} \left(1-\mathrm{e}^{-\mathrm{e}^{\frac{\mu-x}{\sigma}}}\right)^{\alpha-1} \left[1-\left(1-\mathrm{e}^{-\mathrm{e}^{\frac{\mu-x}{\sigma}}}\right)^{\alpha}\right]^{\beta-1},$$

where $x \in \mathbb{R}$, $\xi \in \mathbb{R}$, $\sigma > 0$, $\mu \in \mathbb{R}$, $\alpha > 0$ and $\beta > 0$.

The $\mathcal{MLE}$'s of the parameters are included in Table 2 for each of the fitted distributions. It can be seen from the values of the goodness–of–fit statistics appearing Table 3 that the two proposed distributions provide the most adequate models. The plots of the cdf's that are superimposed on the empirical cdf in the right panel of Figure 8 also suggest that they better fit the data. Additionally, asymptotic confidence intervals for the model parameters are included in Table 4.

Table 2: $\mathcal{MLE}$'s of the parameters (standard errors in parentheses) for the Seoul rainfall data.

Distribution	Estimates				
$\mathcal{GEV}(s,m,\xi)$	0.0212 (0.0015)	2.3781 (0.1666)	0.0028 (0.0570)		
Kum$\mathcal{GEV}(a,b,\xi,\sigma,\mu)$	18.289 (5.652)	15.412 (13.558)	21.175 (9.868)	1.1934 (0.440)	2.1339 (11.002)
EGGu$(\sigma,\mu,\alpha,\beta)$	85.686 (206.89)	−18.428 (509.13)	1.7687 (4.4618)	18.593 (201.49)	
q-$\mathcal{GEV}(s,m,\xi,q)$	0.0303 (0.0085)	4.1082 (1.6329)	0.1973 (0.0922)	1.1225 (1.266)	
q-Gumbel(s,m,q)	0.02045 (0.0026)	2.4323 (0.4135)	0.1129 (0.1746)		

Table 3: Goodness-of-fit statistics for the Seoul rainfall data.

Distribution	AIC	AICC	A^*	W^*	K–S	p-value (K–S)
$\mathcal{GEV}(s, m, \xi)$	1169.63	1169.87	0.9583	0.1325	0.0892	0.3725
Kum$\mathcal{GEV}(a, b, \xi, \sigma, \mu)$	1174.726	1175.33	0.8566	0.1505	0.0889	0.3767
EGGu$(\sigma, \mu, \alpha, \beta)$	1169.16	1169.56	0.6566	0.1099	0.0872	0.4007
q-$\mathcal{GEV}(s, m, \xi, q)$	1168.64	1169.04	0.4638	0.0678	0.0716	0.6535
q-Gumbel(s, m, q)	1166.94	1167.18	0.6279	0.1021	0.0862	0.4157

Table 4: Confidence intervals for the parameters of the q-Gumbel and q-$\mathcal{GEV}$ models (Seoul rainfall data).

CI (q-Gumbel)	s	m	q
95%	[0, 0.025546]	[2.4272, 2.4373]	[−0.229316, 0.4551]
99%	[0.01374, 0.027158]	[2.4255, 2.4390]	[−0.337568, 0.5633]

CI (q-$\mathcal{GEV}$)	s	m	ξ	q
95%	[0, 0.04096]	[0.9078, 7.3086]	[0.01698, 0.3776]	[−1.257536, 3.7075]
99%	[0, 0.02496]	[−0.1046, 8.3210]	[−0.040576, 0.43517]	[−2.042828, 4.4928]

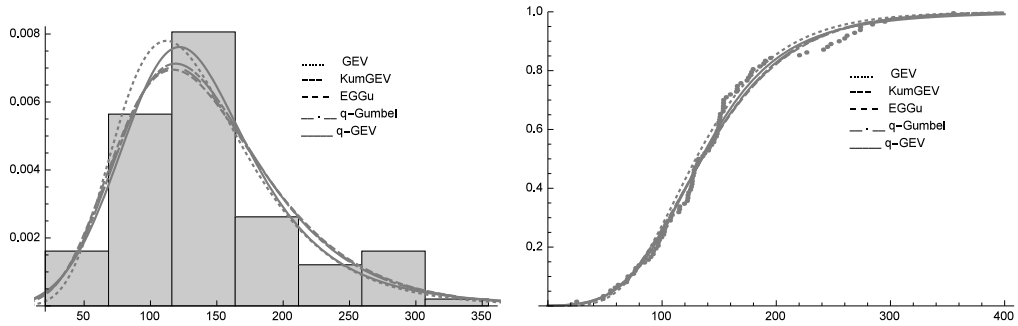


Figure 8: The $\mathcal{GEV}$, Kum$\mathcal{GEV}$, EGGu, q-Gumbel and q-$\mathcal{GEV}$ estimated pdf's superimposed on the histogram of the data (left panel); the estimated cdf's and empirical cdf (right panel).

4.2. Return level

A return period (sometimes referred to as recurrence interval) is an estimate of the likelihood of an event, such as a certain rainfall precipitation level or a given river discharge flow level. It is a statistical measure that is based on historical data, which proves especially useful in risk analysis as it represents the average

recurrence interval over an extended period of time. In fact, the return period is the inverse of the probability that the level will be exceeded in any one year — or, equivalently, the expected waiting time or mean number of years it will take for an exceeding level to occur. For example, a rainfall precipitation return level x_5 has a 20% (or one fifth) probability of being exceeded in any one year, which of course, does not mean that such a rainfall level will happen regularly every 5 years or only once in a five-year period, despite what the phrase "return period" might suggest.

Based on these considerations and assuming that the event components are independently distributed, the probability that an exceeding event will occur for the first time in t years is $p(1-p)^{t-1}$, $t = 1, 2, ...,$ which is the geometric probability mass function ([25]) whose mean is equal to $T = 1/p$, when the yearly exceedance probability $p = P(X \geq x_T)$ is assumed to remain constant throughout the future years of interest ([1] and [22]) . The probability of exceeding x_T can be estimated by the survival probability, $1 - F(x_T)$, the return period T then being equal to $1/P(X \geq x_T)$. Thus, for a given return period T, the corresponding return level can be obtained as follows:

$$x_T \;=\; F^{-1}(1 - 1/T)\,,$$

which yields

$$x_T \;=\; \frac{1}{s}\left\{m - \log\left(-\frac{1-(1-1/t)^{-q}}{q}\right)\right\}$$

for the q-Gumbel model and

$$x_T \;=\; -\frac{1}{\xi\, s}\left\{\left(-\frac{1-(1-1/t)^{-q}}{q}\right)^{-\xi}\left(-m\xi\left(-\frac{1-(1-1/t)^{-q}}{q}\right)^{\xi} - 1\right)\right\}$$

for the q-$\mathcal{GEV}$ model, where $x_T > 0$ and $T > 1$. When unknown, the parameters are replaced by their $\mathcal{MLE}$'s. The estimates of the return levels x_T obtained from the q-$\mathcal{GEV}$ distribution for the return periods, T=2, 5, 10, 20, 50, 100 years, which appear in Table 5, apply to the previously analyzed Seoul rainfall precipitation data.

Table 5: Return level estimates $\widehat{x}_T$ for given values of T (Seoul rainfall data).

T	$\widehat{x}_T$ (q-$\mathcal{GEV}$ model)
2	133.964
5	187.515
10	225.94
20	267.07
30	293.139
50	328.625
100	382.323

5. SIMULATION STUDY

The suitability of the maximum likelihood approach for estimating the parameters of the q-Gumbel and q-$\mathcal{GEV}$ distributions is assessed in this section. Samples of sizes 50, 100, 300 and 500 were generated from the quantile functions of these distributions by Monte Carlo simulations for several values of the parameters. The biases and mean squared errors (MSE's) of the resulting $\mathcal{MLE}$'s were determined for each combination of sample sizes and assumed parameter values on the basis of 5,000 replications.

The simulations results that were obtained for the q-Gumbel and q-$\mathcal{GEV}$ are respectively reported in Tables 6 and 7. As expected, the biases and MSE's generally decrease as the sample sizes increase. It should be noted that the $\mathcal{MLE}$'s remain fairly accurate even for moderately sized samples. Those results corroborate the appropriateness of the maximum likelihood methodology — as described in Section 3.1 — for estimating the parameters of the proposed models.

Table 6: Monte Carlo simulation results: biases and MSE's for the q-Gumbel model.

n	Actual values			Bias			MSE		
	q	s	m	$\widehat{q}$	$\widehat{s}$	$\widehat{m}$	$\widehat{q}$	$\widehat{s}$	$\widehat{m}$
50	0.5	1.0	0.0	−0.0015	0.0424	0.0109	0.2461	0.0749	0.1535
	1.5	2.0	1.0	0.1779	0.1933	0.2060	1.1077	0.5524	0.9587
	3.0	2.0	1.0	0.3521	0.1904	0.2552	2.3258	1.1566	1.7621
	−0.5	1.0	0.0	−0.1432	−0.0315	−0.0965	0.0863	0.0434	0.0698
	−1.5	2.0	1.0	0.0121	−0.0007	0.0114	0.0008	0.0013	0.0004
	−3.0	2.0	1.0	0.0109	−0.0036	−0.0001	0.0006	0.0001	0.0000
100	0.5	1.0	0.0	−0.0104	0.0160	−0.0008	0.0808	0.0240	0.0507
	1.5	2.0	1.0	0.0644	0.0775	0.0791	0.3300	0.1625	0.2677
	3.0	2.0	1.0	0.2278	0.1351	0.1708	1.5085	0.3438	0.7338
	−0.5	1.0	0.0	−0.0704	−0.0196	−0.0495	0.0282	0.0183	0.0258
	−1.5	2.0	1.0	0.0075	0.0053	0.0101	0.0002	0.0004	0.0002
	−3.0	2.0	1.0	0.0031	−0.001	0.0000	0.0001	0.0000	0.0000
300	0.5	1.0	0.0	−0.0020	0.0052	−0.0003	0.0243	0.0072	0.0148
	1.5	2.0	1.0	0.0192	0.0246	0.0246	0.0851	0.0411	0.0684
	3.0	2.0	1.0	0.0715	0.0404	0.0516	0.3001	0.0617	0.1298
	−0.5	1.0	0.0	−0.0275	−0.0099	−0.0201	0.0058	0.0052	0.0065
	−1.5	2.0	1.0	0.0039	0.0059	0.0070	0.0000	0.0001	0.0001
	−3.0	2.0	1.0	−0.0003	0.0001	0.0000	0.0000	0.0000	0.0000
500	0.5	1.0	0.0	−0.0013	0.0032	0.0002	0.0142	0.0041	0.0089
	1.5	2.0	1.0	0.0148	0.0175	0.0169	0.0483	0.0236	0.0384
	3.0	2.0	1.0	0.0421	0.0243	0.0322	0.1764	0.0355	0.0742
	−0.5	1.0	0.0	−0.0180	−0.0066	−0.0141	0.0030	0.0031	0.0036
	−1.5	2.0	1.0	0.0035	0.0057	0.0065	0.0000	0.0001	0.0001
	−3.0	2.0	1.0	−0.0002	0.0001	0.0000	0.0000	0.0000	0.0000

Table 7: Monte Carlo simulation results: biases and MSE's for the q-$\mathcal{GEV}$ model.

n	Actual values				Bias				MSE			
	q	s	m	ξ	$\widehat{q}$	$\widehat{s}$	$\widehat{m}$	$\widehat{\xi}$	$\widehat{q}$	$\widehat{s}$	$\widehat{m}$	$\widehat{\xi}$
50	0.5	1.0	0.0	0.5	0.9087	0.0420	0.3640	−0.0684	1.9675	0.0743	0.3694	0.0412
	1.5	1.0	1.0	0.5	0.5282	0.0210	0.1514	0.0031	0.7358	0.0522	0.2084	0.0096
	1.5	2.0	1.0	0.5	0.5474	0.0369	0.1576	0.0027	0.7963	0.2146	0.2457	0.0089
	1.5	2.0	1.0	1.5	0.1072	0.0004	0.0007	−0.0005	0.0271	0.0001	0.0000	0.0000
	−0.5	1.0	0.0	−0.5	−0.2784	−0.2644	−0.1517	−0.2271	0.1553	0.1932	0.0781	0.0966
	−1.5	2.0	1.0	−0.5	0.0026	0.0086	0.0115	−0.0164	0.0012	0.0012	0.0005	0.0027
	−1.5	2.0	1.0	−1.5	−0.0025	0.0023	0.0019	−0.0023	0.0000	0.0000	0.0000	0.0001
100	0.5	1.0	0.0	0.5	0.6083	0.0439	0.2602	−0.0343	0.9423	0.0289	0.1939	0.0115
	1.5	1.0	1.0	0.5	0.3917	0.0092	0.1279	−0.0088	0.4535	0.0223	0.1105	0.0056
	1.5	2.0	1.0	0.5	0.4033	0.0071	0.1314	−0.0084	0.4327	0.0902	0.1029	0.0053
	1.5	2.0	1.0	1.5	0.0827	−0.0002	0.0001	0.0000	0.0223	0.0000	0.0000	0.0000
	−0.5	1.0	0.0	−0.5	−0.1429	−0.1471	−0.0842	−0.1121	0.0514	0.0725	0.0291	0.0312
	−1.5	2.0	1.0	−0.5	−0.0003	0.0118	0.0096	−0.0076	0.0005	0.0004	0.0003	0.0008
	−1.5	2.0	1.0	−1.5	−0.0009	0.0009	0.0007	−0.0008	0.0000	0.0000	0.0000	0.0000
300	0.5	1.0	0.0	0.5	0.2501	0.0220	0.1144	−0.0132	0.2391	0.0087	0.0578	0.0026
	1.5	1.0	1.0	0.5	0.1988	0.0088	0.0822	−0.0118	0.1599	0.0066	0.0403	0.0024
	1.5	2.0	1.0	0.5	0.1974	0.0180	0.0801	−0.0117	0.1590	0.0252	0.0396	0.0023
	1.5	2.0	1.0	1.5	0.0352	0.0000	0.0000	0.0000	0.0133	0.0000	0.0000	0.0000
	−0.5	1.0	0.0	−0.5	−0.0491	−0.0539	−0.0320	−0.0363	0.0090	0.0163	0.0073	0.0050
	−1.5	2.0	1.0	−0.5	0.0005	0.0092	0.0069	−0.0019	0.0001	0.0002	0.0001	0.0001
	−1.5	2.0	1.0	−1.5	−0.0002	0.0002	0.0001	−0.0002	0.0000	0.0000	0.0000	0.0000
500	0.5	1.0	0.0	0.5	0.1581	0.0136	0.0734	0.0010	0.1418	0.0015	0.0347	0.0011
	1.5	1.0	1.0	0.5	0.1289	0.0051	0.0558	0.0005	0.0853	0.0015	0.0227	0.0005
	1.5	2.0	1.0	0.5	0.1330	0.0127	0.0566	0.0002	0.0873	0.0015	0.0227	0.0000
	1.5	2.0	1.0	1.5	0.0199	0.0000	0.0000	0.0000	0.0100	0.0000	0.0000	0.0000
	−0.5	1.0	0.0	−0.5	−0.0334	−0.0378	−0.0229	0.0025	0.0047	0.0025	0.0041	0.0025
	−1.5	2.0	1.0	−0.5	0.0011	0.0072	0.0057	0.0001	0.0000	0.0001	0.0001	0.0001
	−1.5	2.0	1.0	−1.5	−0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000

6. CONCLUDING REMARKS

The q-generalized extreme value and the q-Gumbel distributions introduced herein are truly versatile: they can be positively or negatively skewed; they can give rise to increasing, decreasing and upside-down bathtub shaped hazard rate functions, and their supports can be finite, bounded above or below, or infinite. The flexibility of these models was further confirmed by applying them to fit a certain data set consisting of annual maximum daily precipitations, and comparing them to three other models by means of several goodness-of-fit statistics. As well, the model parameters were successfully estimated by the method of maximum likelihood, the suitability of this approach having been supported by a simulation study. Moreover, we observed that numerical integration produces highly accurate results when evaluating various statistical functions of the q-analogues of the $\mathcal{GEV}$ and Gumbel random variables. In practice, the q-generalized extreme value

model ought to be more realistic and useful than its original counterpart, which is actually a limiting distribution, and the proposed extended models should lead to further advances in risk theory, biostatistics, hydrology, meteorology, survival analysis and engineering, among several other fields of research that have already benefited from the utilization of existing related models.

APPENDIX A

The 4×4 total observed information matrix associated with the q-$\mathcal{GEV}$ distribution is given by $-J(\upsilon_1)$ wherein the parameters are replaced by their $\mathcal{MLE}$'s where

$$J(\upsilon_1) = \begin{pmatrix} J(\upsilon_1)_{ss} & J(\upsilon_1)_{sm} & J(\upsilon_1)_{s\xi} & J(\upsilon_1)_{sq} \\ J(\upsilon_1)_{ms} & J(\upsilon_1)_{mm} & J(\upsilon_1)_{m\xi} & J(\upsilon_1)_{mq} \\ J(\upsilon_1)_{\xi s} & J(\upsilon_1)_{\xi m} & J(\upsilon_1)_{\xi\xi} & J(\upsilon_1)_{\xi q} \\ J(\upsilon_1)_{qs} & J(\upsilon_1)_{qm} & J(\upsilon_1)_{q\xi} & J(\upsilon_1)_{qq} \end{pmatrix},$$

with

$$\begin{aligned} J(\upsilon_1)_{ss} = & \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n}\left(-\frac{q^2x_i^2\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{2}{\xi}-2}}{\left(q\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}}+1\right)^2}\right. \\ & \left.-\frac{\left(-\frac{1}{\xi}-1\right)\xi q x_i^2\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}-2}}{q\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}}+1}\right) \\ & +\left(-\frac{1}{\xi}-1\right)\sum_{i=1}^{n}-\frac{\xi^2x_i^2}{\left(\xi\left(sx_i-m\right)+1\right)^2}-\frac{n}{s^2}, \end{aligned}$$

$$\begin{aligned} J(\upsilon_1)_{sm} = & \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n}\left[-\frac{q^2x_i\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{2}{\xi}-2}}{\left(q\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}}+1\right)^2}\right. \\ & \left.+\frac{\left(-\frac{1}{\xi}-1\right)\xi q x_i\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}-2}}{q\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}}+1}\right] \\ & +\left(-\frac{1}{\xi}-1\right)\sum_{i=1}^{n}\frac{\xi^2x_i}{\left(\xi\left(sx_i-m\right)+1\right)^2}, \end{aligned}$$

$$\begin{aligned} J(\upsilon_1)_{s\xi} = & \sum_{i=1}^{n}\frac{\xi x_i}{\left(\xi\left(sx_i-m\right)+1\right)\xi^2} \\ & +\left(-\frac{1}{\xi}-1\right)\sum_{i=1}^{n}\left(\frac{x_i}{\xi\left(sx_i-m\right)+1}-\frac{\xi x_i\left(sx_i-m\right)}{\left(\xi\left(sx_i-m\right)+1\right)^2}\right) \end{aligned}$$

$$
\begin{aligned}
&+\left(-\frac{1}{q}-1\right)\sum_{i=1}^{n}\left[\frac{q^2x_i\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{2}{\xi}-1}}{\left(q\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}}+1\right)^2}\right.\\
&\times\left(\frac{\log\left(\xi\left(sx_i-m\right)+1\right)}{\xi^2}-\frac{sx_i-m}{\xi\left(\xi\left(sx_i-m\right)+1\right)}\right)\\
&\left.-\frac{qx_i\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}-1}\left(\frac{\log\left(\xi\left(sx_i-m\right)+1\right)}{\xi^2}+\frac{\left(-\frac{1}{\xi}-1\right)\left(sx_i-m\right)}{\xi\left(sx_i-m\right)+1}\right)}{q\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}}+1}\right],
\end{aligned}
$$

$$
\begin{aligned}
J(\upsilon_1)_{sq} \;=\; &\frac{\sum_{i=1}^{n}-\frac{qx_i\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}-1}}{q\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}}+1}}{q^2}\\
&+\left(-\frac{1}{q}-1\right)\sum_{i=1}^{n}\left(\frac{qx_i\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{2}{\xi}-1}}{\left(q\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}}+1\right)^2}\right.\\
&\left.-\frac{x_i\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}-1}}{q\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}}+1}\right),
\end{aligned}
$$

$$
\begin{aligned}
J(\upsilon_1)_{mm} \;=\; &\left(-\frac{1}{q}-1\right)\sum_{i=1}^{n}\left(-\frac{q^2\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{2}{\xi}-2}}{\left(q\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}}+1\right)^2}\right.\\
&\left.-\frac{\left(-\frac{1}{\xi}-1\right)\xi q\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}-2}}{q\left(\xi\left(sx_i-m\right)+1\right)^{-\frac{1}{\xi}}+1}\right)\\
&+\;\left(-\frac{1}{\xi}-1\right)\sum_{i=1}^{n}-\frac{\xi^2}{\left(\xi\left(sx_i-m\right)+1\right)^2},
\end{aligned}
$$

$$
\begin{aligned}
J(\upsilon_1)_{m\xi} \;=\; &\sum_{i=1}^{n}-2\log(x_i)\,x_i^{\xi}+\sum_{i=1}^{n}\left(-\frac{\mathrm{e}^{sx_i+mx_i^{\xi}}(-1+q)\log(x_i)\,x_i^{\xi}}{\left(-\mathrm{e}^{sx_i+mx_i^{\xi}}(-1+q)+2q\right)}\right.\\
&-\frac{\mathrm{e}^{2sx_i+2mx_i^{\xi}}(-1+q)^2m\log(x_i)\,x_i^{2\xi}}{\left(-\mathrm{e}^{sx_i+mx_i^{\xi}}(-1+q)+2q\right)^2}\\
&\left.-\frac{\mathrm{e}^{sx_i+mx_i^{\xi}}(-1+q)m\log(x_i)\,x_i^{2\xi}}{\left(-\mathrm{e}^{sx_i+mx_i^{\xi}}(-1+q)+2q\right)}\right)\\
&+\sum_{i=1}^{n}\left(\frac{x_i^{\xi}}{sx_i+\xi mx_i^{\xi}}+\frac{\xi\log(x_i)\,x_i^{\xi}}{sx_i+\xi mx_i^{\xi}}-\frac{\xi x_i^{\xi}\left(mx_i^{\xi}+\xi m\log(x_i)\,x_i^{\xi}\right)}{\left(sx_i+\xi mx_i^{\xi}\right)^2}\right),
\end{aligned}
$$

$$J(\upsilon_1)_{mq} = \sum_{i=1}^{n}\left(\frac{\mathrm{e}^{sx_i+mx_i^{\xi}}\left(2-\mathrm{e}^{sx_i+mx_i^{\xi}}\right)(-1+q)\,x_i^{\xi}}{\left(-\mathrm{e}^{sx_i+mx_i^{\xi}}(-1+q)+2q\right)^2}-\frac{\mathrm{e}^{sx_i+mx_i^{\xi}}\,x_i^{\xi}}{-\mathrm{e}^{sx_i+mx_i^{\xi}}(-1+q)+2q}\right),$$

$$\begin{aligned}J(\upsilon_1)_{\xi\xi} = \sum_{i=1}^{n}\Bigg(&-\frac{(q-1)\,m^2x_i^{2\xi}\log^2(x_i)\,\mathrm{e}^{mx_i^{\xi}+sx_i}}{2q-(q-1)\,\mathrm{e}^{mx_i^{\xi}+sx_i}}\\ &-\frac{(q-1)^2m^2x_i^{2\xi}\log^2(x_i)\,\mathrm{e}^{2mx_i^{\xi}+2sx_i}}{\left(2q-(q-1)\,\mathrm{e}^{mx_i^{\xi}+sx_i}\right)^2}-\frac{(q-1)\,mx_i^{\xi}\log^2(x_i)\,\mathrm{e}^{mx_i^{\xi}+sx_i}}{2q-(q-1)\,\mathrm{e}^{mx_i^{\xi}+sx_i}}\Bigg)\\ &+\sum_{i=1}^{n}\left(\frac{\xi mx_i^{\xi}\log^2(x_i)+2mx_i^{\xi}\log(x_i)}{\xi mx_i^{\xi}+sx_i}-\frac{\left(mx_i^{\xi}+\xi mx_i^{\xi}\log(x_i)\right)^2}{\left(\xi mx_i^{\xi}+sx_i\right)^2}\right)\\ &+\sum_{i=1}^{n}-2mx_i^{\xi}\log^2(x_i)\,,\end{aligned}$$

$$\begin{aligned}J(\upsilon_1)_{\xi q} = \sum_{i=1}^{n}\Bigg(&\frac{(q-1)\,mx_i^{\xi}\log(x_i)\,\mathrm{e}^{mx_i^{\xi}+sx_i}\left(2-\mathrm{e}^{mx_i^{\xi}+sx_i}\right)}{\left(2q-(q-1)\,\mathrm{e}^{mx_i^{\xi}+sx_i}\right)^2}\\ &-\frac{mx_i^{\xi}\log(x_i)\,\mathrm{e}^{mx_i^{\xi}+sx_i}}{2q-(q-1)\,\mathrm{e}^{mx_i^{\xi}+sx_i}}\Bigg)\end{aligned}$$

and

$$J(\upsilon_1)_{qq}(s,m,\xi,q) = \sum_{i=1}^{n}-\frac{\left(2-\mathrm{e}^{sx_i+mx_i^{\xi}}\right)^2}{\left(-\mathrm{e}^{sx_i+mx_i^{\xi}}(-1+q)+2q\right)^2}\,.$$

APPENDIX B

The 3×3 total observed information matrix associated with the q-Gumbel distribution is given by $-J(\upsilon_2)$ wherein the parameters are replaced by their $\mathcal{MLE}$'s where

$$J(\upsilon_2) = \begin{pmatrix} J(\upsilon_2)_{ss} & J(\upsilon_2)_{sm} & J(\upsilon_2)_{sq}\\ J(\upsilon_2)_{ms} & J(\upsilon_2)_{mm} & J(\upsilon_2)_{mq}\\ J(\upsilon_2)_{qs} & J(\upsilon_2)_{qm} & J(\upsilon_2)_{qq}\end{pmatrix},$$

with

$$J(\upsilon_2)_{ss} = \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n}\left(\frac{qx_i^2\,\mathrm{e}^{m-sx_i}}{q\mathrm{e}^{m-sx_i}+1}-\frac{q^2x_i^2\,\mathrm{e}^{2m-2sx_i}}{(q\mathrm{e}^{m-sx_i}+1)^2}\right)-\frac{n}{s^2}\,,$$

$$J(\upsilon_2)_{sm} = \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n}\left(\frac{q^2x_i\,\mathrm{e}^{2m-2sx_i}}{(q\mathrm{e}^{m-sx_i}+1)^2}-\frac{qx_i\,\mathrm{e}^{m-sx_i}}{q\mathrm{e}^{m-sx_i}+1}\right),$$

$$J(\upsilon_2)_{sq} = \frac{\sum_{i=1}^{n} -\frac{q x_i \mathrm{e}^{m-sx_i}}{q\mathrm{e}^{m-sx_i}+1}}{q^2} + \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n}\left(\frac{q x_i \mathrm{e}^{2m-2sx_i}}{(q\mathrm{e}^{m-sx_i}+1)^2} - \frac{x_i \mathrm{e}^{m-sx_i}}{q\mathrm{e}^{m-sx_i}+1}\right),$$

$$J(\upsilon_2)_{mm} = \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n}\left(\frac{q\mathrm{e}^{m-sx_i}}{q\mathrm{e}^{m-sx_i}+1} - \frac{q^2\mathrm{e}^{2m-2sx_i}}{(q\mathrm{e}^{m-sx_i}+1)^2}\right),$$

$$J(\upsilon_2)_{mq} = \frac{\sum_{i=1}^{n} \frac{q\mathrm{e}^{m-sx_i}}{q\mathrm{e}^{m-sx_i}+1}}{q^2} + \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n}\left(\frac{\mathrm{e}^{m-sx_i}}{q\mathrm{e}^{m-sx_i}+1} - \frac{q\mathrm{e}^{2m-2sx_i}}{(q\mathrm{e}^{m-sx_i}+1)^2}\right)$$

and

$$J(\upsilon_2)_{qq} = -\frac{2\sum_{i=1}^{n}\log(q\mathrm{e}^{m-sx_i}+1)}{q^3} + \frac{2\sum_{i=1}^{n}\frac{\mathrm{e}^{m-sx_i}}{q\mathrm{e}^{m-sx_i}+1}}{q^2} + \left(-\frac{1}{q}-1\right)\sum_{i=1}^{n} -\frac{\mathrm{e}^{2m-2sx_i}}{(q\mathrm{e}^{m-sx_i}+1)^2}.$$

ACKNOWLEDGMENTS

The financial support of the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged by the Serge B. Provost. The research programs of Abdus Saboor and Gauss Cordeiro are respectively supported by the Higher Education Commission of Pakistan under NRPU project No. 3104 and CNPq, Brazil. We are very grateful to two reviewers for their constructive and insightful comments and suggestions.

REFERENCES

[1] Ahammed, F.; Hewa, G.A. and Argue, J.R. (2014). Variability of annual daily maximum rainfall of Dhaka, Bangladesh, *Atmospheric Research*, **137**, 176–182.

[2] Anderson, T.W. and Darling, D.A. (1952). Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes, *Annals of Mathematical Statistics*, **23**, 193–212.

[3] Andrade, A.; Rodrigues, H.; Bourguignon, M. and Cordeiro, G. (2015). The exponentiated generalized Gumbel distribution, *Revista Colombiana de Estadística*, **38**, 123–143.

[4] Balakrishnan, N. (1995). *Recent Advances in Life-Testing and Reliability*, CRC Press, Boca Raton.

[5] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, "Springer Series in Statistics", Springer, Heidelberg.

[6] Conniffe, D. (2007). The generalised extreme value distribution as utility function, *The Economic and Social Review*, **38**, 275–288.

[7] Cordeiro, G.M.; Nadarajah, S. and Ortega, E.M.M. (2012). The Kumaraswamy Gumbel distribution, *Statistical Methods and Applications*, **21**, 139–168.

[8] Ebeling, W. and Sokolov, M. (2005). *Statistical Thermodynamics and Stochastic Theory of Nonequilibrium Systems*, World Scientific, Singapore.

[9] Eljabri, S.S. (2013). *New Statistical Models for Extreme Values*, Doctoral Thesis, Manchester, UK: The University of Manchester.

[10] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample, *Proc. Cambridge Philosophical Society*, **24**, 180–190.

[11] Gradshteyn, I.S. and Ryzhik, I.M. (2007). *Table of Integrals, Series, and Products*, Edited by Alan Jeffrey and Daniel Zwillinger, 7th edn., Academic Press, New York.

[12] Gumbel, E.J. (1958). *Statistics of Extremes*, Columbia University Press, New York.

[13] Hoskig, J.R.M.; Wallis, J.R. and Wood, E.E. (1985). Estimation of the generalized extreme value distribution by the moment of probability-weighted moments, *Technometrics*, **27**, 339–349.

[14] Jose, K.K. and Naik, S.R. (2009). On the q-Weibull distribution and its applications, *Communications in Statistics: Theory and Methods*, **38**, 912–926.

[15] Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions, Theory and Applications*, Imperial College Press, London.

[16] Markose, S. and Alentorn, A. (2011). The generalized extreme value distribution, implied tail index and option pricing, *The Journal of Derivatives*, **18**, 35–60.

[17] Mathai, A.M. (2005). A pathway to matrix-variate gamma and normal densities, *Linear Algebra and Its Applications*, **396**, 317–328.

[18] Mathai, A.M. and Haubold, H.J. (2007). Pathway model, superstatistics, Tsallis statistics and a generalized measure of entropy, *Physica A*, **375**, 110–122.

[19] Mathai, A.M. and Haubold, H. (2011). A pathway from Bayesian statistical analysis to superstatistics, *Applied Mathematics and Computation*, **218**, 799–804.

[20] Mathai, A.M. and Provost, S.B. (2006). On q-logistic and related models, *IEEE Transactions on Reliability*, **55**, 237–344.

[21] Mathai, A.N. and Provost, S.B. (2011). The q-extended inverse Gaussian distribution, *The Journal of Probability and Statistical Science*, **9**, 1–20.

[22] Nadarajah, S. and Choi, D. (2007). Maximum daily rainfall in South Korea, *J. Earth Syst. Sci.*, **116**, 311–320.

[23] Park, H.W. and Sohn, H. (2006). Parameter estimation of the generalized extreme value distribution for structural health monitoring, *Probab. Eng. Math.*, **21**, 366–376.

[24] Prescott, P. and Waldon, A.T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme value distribution, *Biometrika*, **67**, 723–724.

[25] Salas, J.D.; Govindaraju, R.; Anderson, M.; Arabi, M.; Frances, F.; Suarez, W.; Lavado, W. and Green, T.R. (2014). *Introduction to Hydrology.* In "Handbook of Environmental Engineering, Volume 15" (L.K. Wang and C.T. Yang, Eds.), Modern Water Resources Engineering, New York, NY: Humana Press-Springer Science, 1–126.

[26] Sankarasubramanian, A. and Srinivasan, K. (1996). Evaluation of sampling properties of general extreme value (GEV) distribution-L-moments vs conventional moments, *Proceedings of North American Water and Environment Congress & Destructive Water*, 152–158.

[27] Tsallis, C. (2000). *Nonextensive statistical mechanics and its applications.* In "Lecture Notes in Physics" (S. Abe and Y. Okamoto, Eds.), Springer-Verlag, Berlin.

[28] Uffink, J. (2007). *Compendium of the foundations of classical statistical physics.* In "Philosophy of Physics, Part B" (J. Butterfield and J. Earman, Eds.), Elsevier, Amsterdam, 923–1074.

[29] Wang, X. and Dey, D.K. (2010). Generalized extreme value regression for binary response data, an application to B2B electronic payments system adoption, *The Annals of Applied Statistics*, **4**, 2000–2023.

[30] Wilk, G. and Wlodarczyk, Z. (2000). Interpretation of the nonextensivity parameter q in some applications of Tsallis statistics and Lévy distributions, *Phys. Rev. Lett.*, **84**, 2770–2773.

[31] Wilk, G. and Wlodarczyk, Z. (2001). Non-exponential decays and nonextensivity, *Physica A*, **290**, 55–58.

REVSTAT – Statistical Journal
Volume 16, Number 1, January 2018, 71–86

MODELLING SPATIALLY SAMPLED PROPORTION PROCESSES

Authors: Iosu Paradinas
– Departament d'Estadística i Investigació Operativa, Universitat de València, C/ Dr. Moliner 50, Burjassot, 46100, Valencia, Spain and Asociación Ipar Perspective, C/ Karabiondo 17, 48600, Sopela, Bizkaia
paradinas.iosu@gmail.com

Maria Grazia Pennino
– Instituto Español de Oceanografía, Centro Oceanográfico de Murcia, C/ Varadero 1, San Pedro del Pinatar, 30740, Murcia, Spain
grazia.pennino@mu.ieo.es

Antonio López-Quílez
– Departament d'Estadística i Investigació Operativa, Universitat de València, C/ Dr. Moliner 50, Burjassot, 46100, Valencia, Spain
antonio.lopez@uv.es

Marcial Marín
– Departament d'Estadística i Investigació Operativa, Universitat de València, C/ Dr. Moliner 50, Burjassot, 46100, Valencia, Spain
pitroig72@gmail.com

José María Bellido
– Instituto Español de Oceanografía, Centro Oceanográfico de Murcia, C/ Varadero 1, San Pedro del Pinatar, 30740, Murcia, Spain
josem.bellido@mu.ieo.es

David Conesa
– Departament d'Estadística i Investigació Operativa, Universitat de València, C/ Dr. Moliner 50, Burjassot, 46100, Valencia, Spain
David.V.Conesa@uv.es

Received: April 2016 Revised: July 2016 Accepted: July 2016

Abstract:

- Many ecological processes are measured as proportions and are spatially sampled. In all these cases the standard procedure has long been the transformation of proportional data with the arcsine square root or logit transformation, without considering the spatial correlation in any way. This paper presents a robust regression model to analyse this kind of data using a beta regression and including a spatially correlated term within the Bayesian framework. As a practical example, we apply the proposed approach to a spatio-temporally sampled fishery discard dataset.

Key-Words:

- *modelling proportions; beta regression; spatial modelling; Bayesian hierarchical modelling.*

1. INTRODUCTION

Many ecological processes are spatially sampled and measured as proportions; one example is, sea-grass coverage in a area. The traditional approach in ecology has been to, first transform proportional data to approximate normality, and then analyse them using Gaussian linear models, such as analysis of variance or linear regression.

A very common transformation is the arcsine square root transformation. This transformation can be useful to stabilise variances and normalise the data but there are several reasons why it should be avoided. Firstly, model parameters cannot be easily interpreted in terms of the original response [Warton and Hui, 2011, Ferrari and Cribari-Neto, 2004]. Secondly, the efficacy of the arcsine transformation in normalising proportional data is heavily dependent on the sample size, and does not perform well at extreme ends of the distribution [Warton and Hui, 2011, Wilson and Hardy, 2002]. Thirdly, measures of proportions typically display asymmetry, and hence inference based on the normality assumption can be misleading [Ferrari and Cribari-Neto, 2004].

An alternative that is becoming more prevalent in ecological analyses is the logistic regression, an analytical method designed to deal with binomial proportional data [Steel *et al.*, 1997, Wilson and Hardy, 2002, Warton and Hui, 2011], i.e. proportions measured as *x out of n*. The logistic regression provides a more biologically and ecologically interpretative analysis and is not sensitive to sample size. Nonetheless, such binomial data is prone to overdispersion, resulting in an incorrect quantification of the uncertainty when applying the proposed binomial generalised linear model (GLM). In these cases, the inclusion of a random intercept term using generalised linear mixed models (GLMMs) may improve the assessment of uncertainty [Wilson and Hardy, 2002].

When data are non-binomial, that is, observations do not follow the *x out of n* pattern, the logistic regression is no longer applicable. As an alternative approach, Warton and Hui [2011] suggested the logit transformation of the data, which overcomes the problems of interpretability and range of the confidence/credible intervals using the arcsine square root transformation. However, any transformation of the data (y_t) implies that regression parameters are only interpretable in terms of the transformed mean of y_t and not the mean of the original data.

The beta distribution is a well known distribution that satisfies the characteristics of proportions, bounded to the [0, 1] interval with asymmetric shapes. It has long been used in a wide range of applications involving proportions and probabilities [Gupta and Nadarajah, 2004]. However, only recently has it been

applied to linear regression modelling [Ferrari and Cribari-Neto, 2004, Smithson and Verkuilen, 2006, Liu and Kong, 2015] and time-series analysis [Da-Silva and Migon, 2016], allowing bounded estimates and intervals with model parameters that are directly interpretable in terms of the mean of the response.

Aside from the likelihood function, it is well known that changes in ecological processes in time and space are driven by a set of factors and interactions. Understanding these drivers is very often the ultimate goal among scientists seeking to manage natural resources effectively. However, the immeasurable complexity of ecological spatial processes often means that the spatial variability of the data exceed the variability explained by the explanatory variables. This phenomenon usually results in spatially autocorrelated model residuals that can yield incorrect results and a restricted predictive capacity of the models [Fortin and Dale, 2009, Legendre *et al.*, 2002].

A good solution to improve model fit and prediction is to introduce spatial terms in our models. Spatial terms are based on the principle that close observations have more in common than distant observations [Tobler, 1970]. Consequently, by applying a distance-based function, these terms are capable of improving fine scale predictions and identifying hidden spatial hot and/or cold spots that may be important for management purposes. In addition, from a management perspective it is crucial to address the uncertainty associated with our predictions and estimates. In this respect, the Bayesian hierarchical approach is able to accommodate complex systems and obtain a proper uncertainty assessment by relying on quite straightforward probability rules [Clark, 2005].

The reminder of this article goes as follows. First, we summarise the characteristics of the hierarchical spatial beta regression. Then, we introduce the principles of the Integrated Nested Laplace Approximation (INLA from now on) using the Stochastic Partial Differential Equations (SPDE) approach (`http://www.r-inla.org`) [Rue *et al.*, 2009] as an effective way to deal with spatially sampled proportional data. As an example, we apply this approach to a fishery discards database to identify discard proportion high-density areas in the Western Mediterranean Sea. Finally, we end up with some conclusions.

2. HIERARCHICAL SPATIAL BETA REGRESSION

Traditionally the beta distribution is denoted by two scaling parameters $Be(a, b)$. In order to apply regression it is necessary to reparametrize its density distribution in terms of its mean $\mu = \frac{a}{a+b}$ and a dispersion $\phi = a + b$, so that:

$$\pi(y) = \frac{\Gamma(\phi)}{\gamma(\mu\phi)\gamma(\phi(1-\mu))}\, y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \qquad 0 < y < 1\,, \tag{2.1}$$

where Γ is the gamma function, $E(y) = \mu$ and $Var(y) = \frac{\mu(1-\mu)}{1+\phi}$. Note that here, as opposed to the Gaussian distribution, the variance depends on the mean, which translates into maximum variance at the centre of the distribution and minimum at the edges, to support the truncated nature of the beta distribution.

It is also important to note that the probability density (2.1) does not provide a satisfactory description of the data at both ends of the distribution, zero and one. An ad hoc solution may be to add a small error value to the observations to satisfy this criterion [Warton and Hui, 2011]; otherwise zero and one inflated models are required [Liu and Kong, 2015].

Following the $Be(\mu, \phi)$ reparametrisation, a given set of observations $y_1, ..., y_n$, that represent proportions, can be related to a set of covariates and functions using a similar approach to the generalised linear model:

$$Logit(\mu_i) = \eta_i \tag{2.2}$$
$$\eta_i = \alpha + \sum_{j=1}^{n_\beta} \beta_j z_{ji} + \sum_{k=1}^{n_k} f_k(u_{ki}) + v_i$$

where η_i enters the likelihood through a logit link, α is the intercept of the model, β_j are the fixed effects of the model, $f_k()$ denote any smooth effects (including spatial dependence effects) and v_i are unstructured error terms (random variables).

At the time of writing, a handful of R packages allow beta regression: `betareg` [Grün *et al.*, 2011], `mgcv` [Wood, 2011] and `gamlss` [Stasinopoulos and Rigby, 2007] in the frequentist field and `Bayesianbetareg` [Marin *et al.*, 2014], `zoib` Liu and Kong [2015] and `R-INLA` (the implementation of INLA in R [Martins *et al.*, 2013]) in the Bayesian counterpart. `zoib` allows zero/one inflated beta regression but only `R-INLA` allows a wide range of flexible hierarchical models to be fitted at a user-friendly and computationally efficient environment, as we will show in the following Section.

Indeed, Bayesian hierarchical methods are becoming very popular in many fields due to the complexity of the relationships involved in natural systems [Clark, 2005]. Modelling these relationships often requires specifying sub-models inside the additive predictor that allow a suspected hidden or latent effect to be inferred that characterise these relationships.

A good example may be the use of spatial latent fields that apply distance-based functions to model the spatial dependence of the data. In these cases, the main intensity of the process is driven by a set of covariates $X\beta$, also called large-scale variation, to which a spatial term is added based on a correlation function $f_w()$ that describe the unobserved small-scale variation. Consequently, we end up with a spatial correlation model, which depends on its own hyperparameters, as

part of a broader model that characterises the intensity of the process; in other words, we have a hierarchical model with a spatial latent variable.

A popular point-referenced spatial model, the geostatistical model, has the characteristic that the spatial covariance function $f_w()$ is continuous over the range of the spatial effect. Based on this function, it is customary to assume a Gaussian latent field $W \sim N(0, Q(\kappa\tau))$ with covariance matrix Q that depends on two hyperparameters, in the case of `R-INLA`, κ and τ. These hyperparameters determine the range and the variance of the spatial latent field. When we include this in the additive predictor of a beta distributed process Y, we obtain a hierarchical model with at least three stages:

- First stage: $Y|\beta, W \sim Be(\boldsymbol{X}\boldsymbol{\beta} + W, \rho)$
  where Y are conditionally independent given W.
- Second stage: $W|\kappa, \tau \sim N(0, Q(\kappa\tau))$
  where W is a Gaussian latent spatial model.
- Third stage: priors on $(\beta, \rho, \kappa, \tau)$.

A common problem with this kind of hierarchical model is that there is no closed expression for the marginal posterior distributions of the parameters and hyperparameters, so numerical approximations are needed. The typical approach to approximate these posteriors is to use MCMC simulation methods. Unfortunately, MCMC can get very computationally inefficient when applied to complex models such as spatial models.

3. THE INLA APPROACH FOR GEOSTATISTICAL MODELS

Performing inference and prediction under a geostatistical Gaussian field W entail the so-called "big n problem" [Banerjee *et al.*, 2003]. This problem is related to the dense covariance matrix Q, which traduces into very high MCMC computational costs. In this vein, the stochastic partial differential equations (SPDE) approach in `R-INLA` allows reducing the required number of computations from $O(n^3)$ [Stein *et al.*, 2004] to $O(n^{3/2})$ [Cameletti *et al.*, 2013] in the two dimensional spatial domain. In what follows, we first present the INLA method followed by the SPDE approach.

The INLA algorithm, proposed by Rue *et al.* [2009], is a numerical approximation method to perform Bayesian inference. The most remarkable feature of INLA, as opposed to MCMC, is that it allows the posterior distributions of latent Gaussian models to be accurately approximated through Laplace approximations [Laplace, 1986, Tierney and Kadane, 1986], even for complex models without becoming computationally prohibitive. INLA exploits the fact that latent Gaussian

models admit conditional independence properties [Rue and Held, 2005], which allows expressing them as computationally efficient Gaussian Markov random fields (GMRFs) with a sparse precision matrix [Rue and Held, 2005].

The estimation of the latent components, collected in a set of parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{W}\}$ and hyperparameters $\boldsymbol{\Omega} = \{\rho, \kappa, \tau\}$ in `R-INLA`, is computed in three steps. First, the posterior marginal distribution of the hyperparameters is approximated by using the Laplace integration method

$$p(\boldsymbol{\Omega}|Y) \approx \left.\frac{p(Y|\boldsymbol{\theta}, \boldsymbol{\Omega})p(\boldsymbol{\theta}|\boldsymbol{\Omega})p(\boldsymbol{\Omega})}{\widetilde{p}(\boldsymbol{\theta}|\boldsymbol{\Omega}, Y)}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*(\boldsymbol{\Omega})} = \widetilde{p}(\boldsymbol{\Omega}|Y), \tag{3.1}$$

where $\widetilde{p}(\boldsymbol{\theta}|\boldsymbol{\Omega}, Y)$ is the Gaussian approximation, given by the Laplace method, of $p(\boldsymbol{\theta}|\boldsymbol{\Omega}, Y)$ and $\boldsymbol{\theta}^*(\boldsymbol{\Omega})$ is the mode for a given $\boldsymbol{\Omega}$.

Then, `R-INLA` approximates $p(\theta_i|\boldsymbol{\Omega}, Y)$ by using again the Laplace integration method

$$p(\theta_i|\boldsymbol{\Omega}, Y) \approx \left.\frac{p(\boldsymbol{\theta}|\boldsymbol{\Omega}, Y)}{\widetilde{p}(\boldsymbol{\theta}_{-i}|\theta_i, \boldsymbol{\Omega}, Y)}\right|_{\boldsymbol{\theta}_{-i}=\boldsymbol{\theta}^*_{-i}(\theta_i,\Omega)} = \widetilde{p}(\theta_i|\boldsymbol{\Omega}, Y), \tag{3.2}$$

where $\widetilde{p}(\boldsymbol{\theta}_{-i}|\theta_i, \boldsymbol{\Omega}, Y)$ is the Laplace Gaussian approximation to $p(\boldsymbol{\theta}_{-i}|\theta_i, \boldsymbol{\Omega}, Y)$ and $\boldsymbol{\theta}^*_{-i}(\theta_i, \boldsymbol{\Omega})$ is its mode. This strategy can be very computationally expensive since $\widetilde{p}(\boldsymbol{\theta}_{-i}|\theta_i, \boldsymbol{\Omega}, Y)$ has to be recomputed for each value of $\boldsymbol{\theta}$ and $\boldsymbol{\Omega}$. See section 3.2 in Rue *et al.* [2009] for a more detailed text on the different approximation approaches available in `R-INLA`.

Finally, `R-INLA` approximates the marginal posterior distributions based on the previous two steps

$$p(\theta_i|Y) \approx \int \widetilde{p}(\theta_i|\boldsymbol{\Omega}, Y)\widetilde{p}(\boldsymbol{\Omega}|Y)d\boldsymbol{\Omega}, \tag{3.3}$$

where the integral can be numerically solved through a finite weighted sum applied in certain integration points and then interpolating in between. For a more detailed text on the selection of integration points see section 3.1(c) in Rue *et al.* [2009].

As mentioned above, INLA exploits the good computational properties of GMRFs to perform fast Bayesian inference. Nevertheless, continuous GFs (like the ones involved in geostatistical models) are continuously indexed, thus, in principle, not applicable in INLA. In this regard, Lindgren *et al.* [2011] provided a clever approximation of a GF with Matérn covariance function (3.4) to a GMRF using a fractional stochastic partial differential equation.

Lindgren *et al.* [2011]'s approximation of a GF requires that its covariance function is of the Matérn family. Following Lindgren *et al.* [2011]'s notation, the Matérn covariance function for an stationary and isotropic GF is

$$\mathcal{C}(d) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa||s_i - s_j||)^\nu K_\nu(\kappa||s_i - s_j||), \tag{3.4}$$

where κ is a scaling parameter that determines the effective range r of the spatial effect.

The approximation by Lindgren *et al.* [2011] fall on the fact that a GF $z(\boldsymbol{s})$ with Matérn covariance function is a solution to the linear fractional SPDE

$$(\kappa^2 - \Delta)^{\alpha/2} z(\tau \boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^d, \alpha = \nu + d/2, \kappa > 0, \nu > 0, \tag{3.5}$$

where Δ is the Laplacian, d is the dimension of the GF $z(\boldsymbol{s})$, ν is the smoothness parameter of the Matérn function and $\mathcal{W}$ is the Gaussian spatial white noise process.

Finally, the solution to the SPDE can be approximated using the Finite Element Method [Zienkiewicz *et al.*, 1977] through a deterministic basis function representation defined on a triangulation of the domain $\mathcal{D}$ (see Figure 1 for the triangulation used in the case study of the following Section). The triangulation, so-called *mesh*, of the study area is based on Delaunay triangulations [Delaunay, 1934], which, as opposed to a regular grid, allows a flexible partition of the region into triangles that can satisfy different types of constraints to better accommodate different characteristics of the study area.

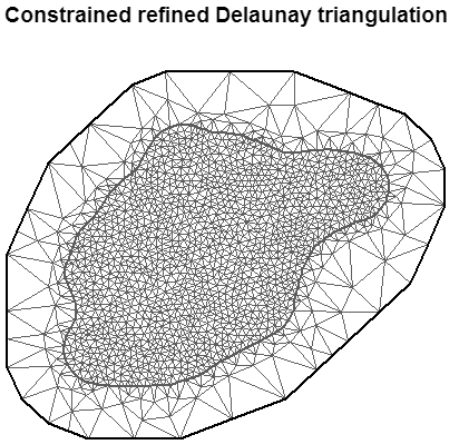


Figure 1: Triangulation of the study area. The outer ring of sparse triangles allows us avoid having a border effect inside the study area.

4. APPLICATION TO TRAWL DISCARD PROPORTIONS

The modelling approach proposed to tackle spatially sampled proportions was applied to a trawl fishery discard database in the Spanish Mediterranean. Fishery discards, i.e. the part of the catch that is thrown back to the sea dead,

constitute an unnecessary biomass loss from the marine systems. A repeatedly proposed discard mitigation measure is the spatial management of fishery resources [Kelleher, 2005, Bellido *et al.*, 2011, Pennino *et al.*, 2014]. In this regard, spatial beta regression is specially important to the fishery discards framework since it allows the spatial assessment of discard proportions, which allows assessing the economic benefit of a fishing operation against its ecological impact due to the discard portion of the catch.

4.1. Data

Trawl discard data were collected according to European Comission [2009] regulation, which establishes a métier-based discard sampling programme. Specifically this study was based on bottom trawl data for the south-eastern part of the Spanish Mediterranean Sea (Figure 2) [see Pennino *et al.*, 2014, for a more detailed description of the métiers].

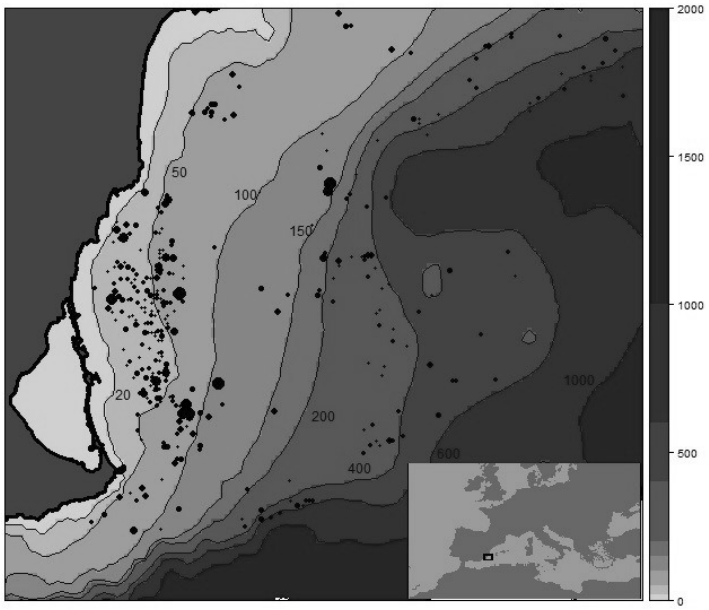


Figure 2: Map of the study area with bathymetric contours in meters. Black dots represent the centroids of the 391 sampled hauls and size plotted according to the observed discard proportion.

The database, provided by the *Instituto Español de Oceanografía* (IEO, Spanish Oceanographic Institute), contains a total of 391 hauls collected between 2009 and 2012, including catch and discard data disaggregated by species. The characteristics of each fishing operation (date, geolocation and depth) were also extracted directly from this database.

A discard proportion response variable of regulated species was created as the fraction of discarded biomass of the total catch. Unlike total discards, discard proportions represent benefit versus loss, and are therefore a better indicator to assess whether or not discards are disproportionate to the catch.

4.2. Modelling trawl discard proportions

The analysis of trawl discard proportions included the total catch of each fishing haul, the mean bathymetry of the haul, a geostatistical term and a vessel effect as predictors (Table 1). Therefore, assuming that the discard proportion Y_i at location i follows a beta distribution, the final model can be expressed as:

$$\begin{aligned} Y_i &\sim \text{Be}(\mu_i, \phi_i), \quad i = 1, ..., n \\ \text{logit}(\mu_i) &= \beta_c c_i + d_i + W_i \\ \beta_c &\sim N(0, 0.001) \\ \Delta^2 d_j &= d_j - 2d_{j+1} + d_{j+2} \sim N(0, \rho_d), \quad j = 1, ..., m \\ \log\rho_D &\sim LogGamma(0.5, 0.00005) \\ \boldsymbol{W} &\sim N(0, \mathbf{Q}(\kappa, \tau)) \\ 2\log\kappa &\sim N(\mu_\kappa, \rho_\kappa) \\ \log\tau &\sim N(\mu_\tau, \rho_\tau) \end{aligned} \tag{4.1}$$

where the mean of discard proportions enters the model through the logit link, i indexes the location of each haul and j indexes different depths (d_j, representing the different values of bathimetry starting at $d_1 = 40$ metres till $d_{m=30} = 720$ metres). In the last two rows μ stands for the mean of the normal distributions while ρ denotes its corresponding precision.

Table 1: List of covariates included in the analysis and the effect assigned to them.

Variable	Description	Unit	Effect
Total catch	Total catch of the haul	Kilograms	Linear
Location	Geolocation	UTM	Geostatistical
Depth	Mean depth of the haul	Meters	Non-linear effect
Vessel	Sampled vessel ID	—	Random noise effect

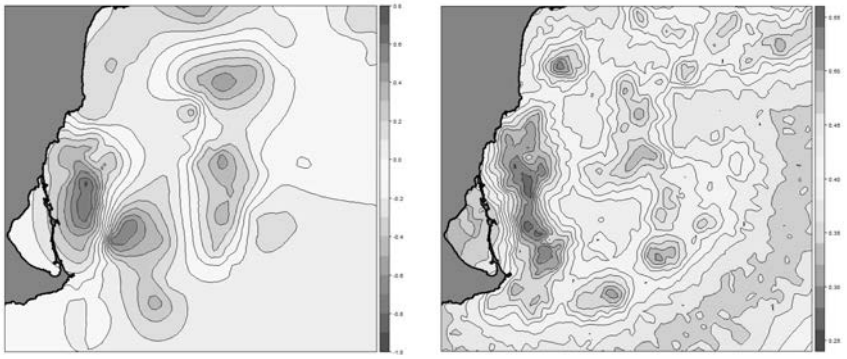
Based on the work by Rochet and Trenkel [2005], who found that discard proportions are not fully proportional to the catch, the total catch of each haul $C = (c_1, ..., c_n)$ was introduced as a linear effect with vague normal prior distributions as implemented by default in `R-INLA`. The exploratory analysis revealed non-linear relationships between depth and the discard proportion, so a second order random walk (RW2) latent model was applied based on constant depth

increments d_j. These RW2 models, which perform as Bayesian smoothing splines [Fahrmeir and Lang, 2001], can be expressed as a computationally efficient GMRF [Rue and Held, 2005], and are therefore applicable in INLA. The smoothing of the bathymetric effect was selected visually by subsequently changing its prior distribution while models were scaled to have a generalized variance equal to one [Sørbye and Rue, 2014].

The two dimensional geostatistical latent model $\boldsymbol{W}$, introduced to identify fine-scale hot-spots, depends on two hyperparameters κ and τ that define the variance and the range of the spatial effect. Specifically, and with the smoothing parameter of the Matérn (3.4) fixed ($\nu = 1$), the range of the spatial terms is approximately $\sqrt{8}/\kappa$ and the variance $1/(4\pi\kappa^2\tau^2)$. The priors for κ and τ are specified over the logτ and 2logκ. Default `R-INLA` prior distributions were used, where μ_κ is specified so that the range of the field is 20% of the longest distance in the field and μ_τ is chosen so that the mean variance of the field is one. The rest of the prior distributions in use are described in (4.1).

4.3. Results

Figure 3 shows the posterior mean and the standard deviation of the spatial component, which represents the intrinsic spatial variability of the data without the rest of the independent variables. This effect highlights (in blue), high discard proportion areas or hot-spots. Similarly, two cold-spots were found (in red), one in the coastal shallow waters in front of the lagoon and another in the mid-northern part of the 150–300 meter strata. These hot-spots characterise the areas where more discards are expected as compared to other areas with similar environmental conditions. As a consequence, a marine spatial planning framework could consider these areas for protection so that discarded/wasted biomass is minimised.



(a) Mean (b) Standard deviation

Figure 3: Posterior predictive mean and standard deviation maps of the spatial component of discard proportions.

As expected, the total catch of the haul had a positive effect on the expected discard proportions (posterior mean = 0.038; 95% CI = [0.0027, 0.0049]), i.e. the discard proportion increases with total catch increments. The bathymetric effect showed a negative relationship of discard proportions to depth, suggesting that the highest discard proportions are located in shallow waters and decrease with depth (Figure 4).

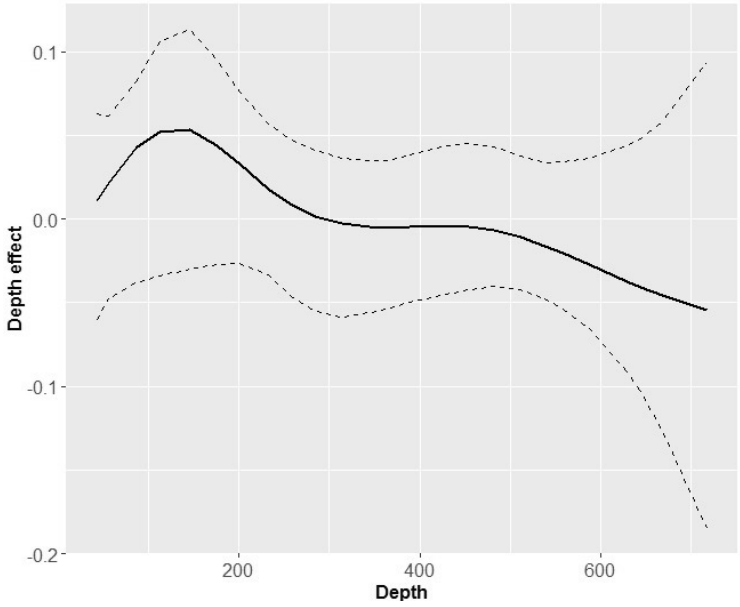


Figure 4: Marginal effect of the bathymetry in the linear predictor. The continuous line represents the mean effect and dashed lines their 95% credible intervals.

Finally, no vessel effects was identified in the study area suggesting that discard proportions are reasonably homogeneous across vessels.

5. CONCLUSIONS

In this paper, we use a Bayesian hierarchical spatial beta model to analyse spatially sampled proportion data. To this end, we use a simple reparametrisation of the beta distribution to apply regression on the mean of the process. The Bayesian approach allows a straightforward quantification of uncertainty, which is important for decision making, while the hierarchical structure allows a more natural model specification, especially when including complex latent models such as geostatistical terms.

Beta regression overcomes all the drawbacks of the traditional data transformations [Warton and Hui, 2011, Ferrari and Cribari-Neto, 2004]. First, it allows a direct interpretation of model parameters in terms of the original data; second, the analysis is not sensitive to the sample size; and lastly, posterior distributions are expected to concentrate well within the bounded range of proportions. It is only when observations on the extremes of the distribution are present, i.e. 0 and 1, that the beta distribution does not provide a satisfactory description of the data. A possible solution to this problem is to add some small value to the proportion, which introduces minimal bias while still satisfying the above criteria [Warton and Hui, 2011]; otherwise, zero and/or one inflated models may be required [Ospina and Ferrari, 2012], now available in the `zoib` package [Liu and Kong, 2015] for R.

The incorporation of spatial random effects in beta regression models can be very useful in a wide range of disciplines. For example mapping plant coverage in ecology; mapping budget allocation in econometrics; mapping the percentage of retirees in sociology, mapping sex-ratios in species, etc. Furthermore, combining the Bayesian spatial hierarchical modelling approach [Banerjee *et al.*, 2003] and the temporal extension of Da-Silva and Migon [2016], the beta regression framework can be extended to the spatio-temporal domain. Consequently, it is possible to tackle problems such as the evolution of plant epidemics [Stein *et al.*, 1994], the spatio-temporal evolution of temperature [Hengl *et al.*, 2012] or the understanding of the spatial dynamism of species over time, as in Paradinas *et al.* [2015]. It must be taken into account that the computational burden of these models can be even more demanding than in the purely spatial domain, making `R-INLA` and its SPDE module two almost necessary tools to deal with them.

The Bayesian analysis of fisheries distribution is a very important field of research in marine ecology [Muñoz *et al.*, 2013, Quiroz *et al.*, 2015]. The case study presented here applies spatial beta regression to identify fishery discard hot-spots based on discard proportions, which, as opposed to total discard units, assess the biomass benefit against the amount of wasted biomass that constitute discards. Our results have identified at least one high discard proportion hot-spot in the study area. Under a marine spatial planning framework that seeks to minimise the ecological impact of the fishing activity, the characterisation of hot-spots could be specially useful for policy makers, as it would allow them to protect those hot-spots as areas of special interest.

To conclude, we would like to mention that the geostatistical beta regression approach proposed here to analyse proportions is not only applicable to non-binomial proportional data but also to binomial proportional data, i.e. proportions measured as *x out of n*. In fact, applying beta regression in these cases may be an easier and more natural approach to avoid the usual problem of overdispersion in logistic regression than that proposed in Wilson and Hardy [2002] using GLMMs.

ACKNOWLEDGMENTS

The authors very much appreciate the suggestions and comments made by two anonymous reviewers on the original manuscript that greatly improved the final version. D. C. and A. L. Q. would also like to thank the Ministerio de Educación y Ciencia (Spain) for financial support (jointly financed by the European Regional Development Fund) via Research Grants MTM2010-19528 and MTM2013-42323-P. The authors have no conflict of interests to declare.

REFERENCES

[1] Banerjee, S.; Gelfand, A.E. and Carlin, B.P. (2003). *Hierarchical Modeling and Analysis for Spatial Data*, Crc Press, 2003.

[2] Bellido, J.M.; Santos, M.B.; Pennino, M.G.; Valeiras, X. and Pierce, G.J. (2011). Fishery discards and bycatch: solutions for an ecosystem approach to fisheries management?, *Hydrobiologia*, **670**(1), 317–333.

[3] Cameletti, M.; Lindgren, F.; Simpson, D. and Rue, H. (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach, *AStA Advances in Statistical Analysis*, **97**(2), 109–131.

[4] Clark, J.S. (2005). Why environmental scientists are becoming Bayesians, *Ecology Letters*, **8**(1), 2–14.

[5] Da-Silva, C.Q. and Migon, H.S. (2016). Hierarchical dynamic beta model, *Revstat Statistical Journal*, **14**(1), 49–73.

[6] Delaunay, B. (1934). Sur la sphere vide, *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk*, **7**(793-800), 1–2.

[7] European Comission (2009). Comission Decision 2010/93/EU, adopting a multiannual Community programme for the collection, management and use of data in the fisheries sector for the period 2011–2013. Technical report, European Comission.

[8] Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors, *Applied Statistics*, **50**, 201–220.

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

[10] Fortin, M.J. and Dale, M.R.T. (2009). Spatial autocorrelation in ecological studies: a legacy of solutions and myths, *Geographical Analysis*, **41**(4), 392–397.

[11] Grün, B.; Kosmidis, I. and Zeileis, A. (2011). Extended beta regression in R: Shaken, Stirred, Mixed, and Partitioned. Technical report, Working Papers in Economics and Statistics.

[12] Gupta, A.K. and Nadarajah, S. (2004). *Handbook of Beta Distribution and Its Applications*, CRC Press.

[13] Hengl, T.; Heuvelink, G.B.M.; Tadić, M.P. and Pebesma, E.J. (2012). Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images, *Theoretical and Applied Climatology*, **107**(1-2), 265–277.

[14] Kelleher, K. (2005). *Discards in the world's marine fisheries: an update*, Food & Agriculture Org., Fisheries Technical Paper Number 470.

[15] Laplace, P.S. (1986). Memoir on the probability of the causes of events, *Statistical Science*, **1**(3), 364–378.

[16] Legendre, P.; Dale, M.R.T.; Fortin, M.J.; Gurevitch, J.; Hohn, M. and Myers, D. (2002). The consequences of spatial structure for the design and analysis of ecological field surveys, *Ecography*, **25**(5), 601–615.

[17] Lindgren, F.; Rue, H. and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(4), 423–498.

[18] Liu, F. and Kong, Y. (2015). *zoib: an R package for Bayesian Inference for Beta Regression and Zero/One Inflated Beta Regression.*

[19] Marin, M.; Rojas, J.; Jaimes, D.; Rojas, H.A.G.; Corrales, M.; Zarate, M.F.; Duplat, R.; Villarraga, F. and Cepeda-Cuervo, E. (2014). *Bayesianbetareg: Bayesian Beta regression: joint mean and precision modeling*, URL: `http://CRAN.R-project.org/package=Bayesianbetareg`. R package version 1.2.

[20] Martins, T.G.; Simpson, D.; Lindgren, F. and Rue, H. (2013). Bayesian computing with INLA: new features, *Computational Statistics & Data Analysis*, **67**, 68–83.

[21] Muñoz, F.; Pennino, M.G.; Conesa, D.; López-Quílez, A. and Bellido, J.M. (2013). Estimation and prediction of the spatial occurrence of fish species using Bayesian latent Gaussian models, *Stochastic Environmental Research and Risk Assessment*, **27**, 1171–1180.

[22] Ospina, R. and Ferrari, S.L.P. (2012). A general class of zero-or-one inflated beta regression models, *Computational Statistics & Data Analysis*, **56**(6), 1609–1623.

[23] Paradinas, I.; Conesa, D.; Pennino, M.G.; Muñoz, F.; Fernández, A.M.; López-Quílez, A. and Bellido, J.M. (2015). Bayesian spatio-temporal approach to identifying fish nurseries by validating persistence areas, *Marine Ecology Progress Series*, **528**, 245–255.

[24] Pennino, M.G.; Muñoz, F.; Conesa, D.; López-Quílez, A. and Bellido, J.M. (2014). Bayesian spatio-temporal discard model in a demersal trawl fishery, *Journal of Sea Research*, **90**, 44–53.

[25] Quiroz, Z.C.; Prates, M.O. and Rue, H. (2015). A Bayesian approach to estimate the biomass of anchovies off the coast of Perú, *Biometrics*, **71**(1), 208–217.

[26] Rochet, M.J. and Trenkel, V.M. (2005). Factors for the variability of discards: assumptions and field evidence, *Canadian Journal of Fisheries and Aquatic Sciences*, **62**(1), 224–235.

[27] Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, CRC Press.

[28] Rue, H.; Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**(2), 319–392.

[29] Smithson, M. and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables, *Psychological Methods*, **11**(1), 54–71.

[30] Sørbye, S.H. and Rue, H. (2014). Scaling intrinsic Gaussian Markov random field priors in spatial modelling, *Spatial Statistics*, **8**, 39–51.

[31] Stasinopoulos, D.M. and Rigby, R.A. (2007). Generalized additive models for location scale and shape (GAMLSS) in R, *Journal of Statistical Software*, **23**(7), 1–46.

[32] Steel, R.G.D.; Torrie, J.H. and Dickey, D.A. (1997). *Principles and Procedures of Statistics: A Biological Approach*, McGraw-Hill.

[33] Stein, A.; Kocks, C.G.; Zadoks, J.C.; Frinking, H.D.; Ruissen, M.A. and Myers, D.E. (1994). A geostatistical analysis of the spatio-temporal development of downy mildew epidemics in cabbage, *Phytopathology*, **84**(10), 1227–1238.

[34] Stein, M.L.; Chi, Z. and Welty, L.J. (2004). Approximating likelihoods for large spatial data sets, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **66**(2), 275–296.

[35] Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association*, **81**(393), 82–86.

[36] Tobler, W.R. (1970). A computer movie simulating urban growth in the Detroit region, *Economic Geography*, **46**, 234–240.

[37] Warton, D.I. and Hui, F.K.C. (2011). The arcsine is asinine: the analysis of proportions in ecology, *Ecology*, **92**(1), 3–10.

[38] Wilson, K. and Hardy, I.C.W. (2002). *Statistical Analysis of Sex Ratios: An Introduction*, Cambridge University Press.

[39] S.N.Wood (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(1), 3–36.

[40] Zienkiewicz, O.C.; Taylor, R.L.; Zienkiewicz, O.C. and Taylor, R.L. (1977). *The Finite Element Method*, volume 3, McGraw-Hill London.

REVSTAT – Statistical Journal
Volume 16, Number 1, January 2018, 87–113

WEIBULL LINDLEY DISTRIBUTION

Authors: A. Asgharzadeh
– Department of Statistics, University of Mazandaran, Babolsar, Iran

S. Nadarajah
– School of Mathematics, University of Manchester, Manchester M13 9PL, UK
mbbsssn2@manchester.ac.uk

F. Sharafi
– Department of Statistics, University of Mazandaran, Babolsar, Iran

Received: January 2016 Revised: July 2016 Accepted: July 2016

Abstract:

- A new distribution is introduced based on compounding Lindley and Weibull distributions. This distribution contains Lindley and Weibull distributions as special cases. Several properties of the distribution are derived including the hazard rate function, moments, moment generating function, and Lorenz curve. An estimation procedure by the method of maximum likelihood and a simulation study to assess its performance are given. Three real data applications are presented to show that the new distribution fits better than known generalizations of the Lindley distribution.

Key-Words:

- *Lindley distribution; maximum likelihood estimation; Weibull distribution.*

AMS Subject Classification:

- 62E99.

1. INTRODUCTION

The Lindley distribution was first proposed by Lindley [20] in the context of fiducial and Bayesian inference. In recent years, this distribution has been studied and generalized by several authors, see Ghitany *et al.* [15], Zalerzadeh and Dolati [29], Ghitany *et al.* [14], Bakouch *et al.* [4], Barreto-Souza and Bakouch [5] and Ghitany *et al.* [13].

In this paper, we introduce a new generalization of the Lindley distribution referred to as the Weibull Lindley (WL) distribution by compounding Lindley and Weibull distributions. The compounding approach gives new distributions that extend well-known families of distributions and at the same time offer more flexibility for modeling lifetime data. The flexibility of such compound distributions comes in terms of one or more hazard rate shapes, that may be decreasing or increasing or bathtub shaped or upside down bathtub shaped or unimodal.

Many recent distributions have been introduced by using a compounding approach. For example, Adamidis and Loukas [1] proposed a distribution by taking the minimum of N independent and identical exponential random variables, where N is a geometric random variable. But this distribution allows for only decreasing hazard rates. Kus [18] proposed a distribution by taking the minimum of N independent and identical exponential random variables, where N is a Poisson random variable. But this distribution also allows for only decreasing hazard rates. Barreto-Souza *et al.* [6] proposed a distribution by taking the minimum of N independent and identical Weibull random variables, where N is a geometric random variable. But this distribution does not allow for bathtub shaped hazard rates, the most realistic hazard rates. Morais and Barreto-Souza [22] proposed a distribution by taking the minimum of N independent and identical Weibull random variables, where N is a power series random variable. But this distribution also does not allow for bathtub shaped hazard rates. Asgharzadeh *et al.* [3] proposed a distribution by taking the minimum of N independent and identical Pareto type II random variables, where N is a Poisson random variable. But this distribution allows for only decreasing hazard rates. Silva *et al.* [27] proposed a distribution by taking the minimum of N independent and identical extended Weibull random variables, where N is a power series random variable. This distribution does allow for bathtub shaped hazard rates, but that is expected since the extended Weibull distribution contains as particular cases many generalizations of the Weibull distribution. Bourguignon *et al.* [7] proposed a distribution by taking the minimum of N independent and identical Birnbaum–Saunders random variables, where N is a power series random variable. But this distribution does not allow for bathtub shaped hazard rates.

The WL distribution introduced here is obtained by compounding just two random variables (Lindley and Weibull random variables). Besides the WL dis-

tribution has just three parameters, less than several of the distributions cited above.

Let Y denote a Lindley random variable with parameter $\lambda > 0$ and survival function $\overline{G}(y) = \frac{1+\lambda+\lambda y}{1+\lambda} e^{-\lambda y}$, $y > 0$. Let Z denote a Weibull random variable with parameters $\alpha > 0$ and $\beta > 0$, and survival function $\overline{Q}(z) = e^{-(\beta z)^{\alpha}}$, $z > 0$. Assume Y and Z are independent random variables. We define $X = \min(Y, Z)$ as a WL random variable and write $X \sim WL(\alpha, \beta, \lambda)$. The survival function of X is

$$\overline{F}(x) = \overline{G}(x)\overline{Q}(x).$$

The cumulative distribution function (cdf) of X can be written as

$$F(x) = 1 - \frac{1+\lambda+\lambda x}{1+\lambda} e^{-\lambda x - (\beta x)^{\alpha}} \tag{1.1}$$

for $x > 0$, $\alpha > 0$, $\beta \geq 0$ and $\lambda \geq 0$. The probability density function (pdf) of X is

$$f(x) = \frac{1}{1+\lambda} \left[\alpha\lambda(\beta x)^{\alpha} + \alpha\beta(1+\lambda)(\beta x)^{\alpha-1} + \lambda^2(1+x)\right] e^{-\lambda x - (\beta x)^{\alpha}} \tag{1.2}$$

for $x > 0$, $\alpha > 0$, $\beta \geq 0$ and $\lambda \geq 0$.

Some special cases of the WL distribution are: the Weibull distribution with parameters α and β for $\lambda = 0$; the Rayleigh distribution with parameter α for $\lambda = 0$ and $\beta = 2$; the exponential distribution with parameter β for $\lambda = 0$ and $\alpha = 1$; the Lindley distribution with parameter λ for $\beta = 0$.

The WL distribution can be used very effectively for analyzing lifetime data. Some possible motivations for the WL distribution are:

- The WL distribution accommodates different hazard rate shapes, that may be decreasing or increasing or bathtub shaped, see Figure 2. Bathtub shaped hazard rates are very important in practice. None of the known generalizations of the Lindley distribution accommodate a bathtub shaped hazard rate function.
- The WL distribution has closed form expressions for survival and hazard rate functions, which is not the case for some generalizations of the Lindley distribution. Hence, the likelihood function for the WL distribution takes explicit forms for ordinary type-II censored data and progressively type-II censored data. Hence, the WL distribution could be a suitable model to analyse ordinary type-II censored data and progressively type-II censored data.
- The Lindley and Weibull distributions are special cases of the WL distribution.

- Suppose a system is composed of two independent components in series; let Y and Z denote their lifetimes; suppose Y is a Lindley random variable and Z is a Weibull random variable; then the lifetime of the system is a WL random variable.
- Suppose a system is composed of n independent components in series; let $Y, Z_1, Z_2, ..., Z_{n-1}$ denote their lifetimes; suppose Y is a Lindley random variable and $Z_1, Z_2, ..., Z_{n-1}$ are identical Weibull random variables; then the lifetime of the system is also a WL random variable.
- The pdf of the WL distribution can be bimodal, see Figure 1. This is not the case for the Weibull distribution or any generalization of the Lindley distribution. So, any bimodal data set (see Figure 8 for example) cannot be adequately modeled by any of the known generalizations of the Lindley distribution.
- Additive hazard rates arise in many practical situations, for example, event-history analysis (Yamaguchi [28]), modeling of excess mortalities (Gail and Benichou [12], page 391), modeling of breast cancer data (Cadarso-Suarez *et al.* [8]), modelling of hazard rate influenced by periodic fluctuations of temperature (Nair *et al.* [24], page 268), and "biologic" and "statistical" interactions in epidemiology (Andersen and Skrondal [2]). Hence, it is import to have distributions based on additive hazard rates. The WEL distribution is the first generalization of the Lindley distribution based on additive hazard rates.

The rest of this paper is organized as follows: various mathematical properties of the WL distribution are derived in Sections 2 to 4; estimation and simulation procedures for the WL distribution are derived in Section 5; three real data applications are illustrated in Section 6.

Some of the mathematical properties derived in Sections 2 to 4 involve infinite series: namely, (3.1), (3.2) and (4.1). Extensive computations not reported here showed that the relative errors between (3.1), (3.2) and (4.1) and their versions with the infinite series in each truncated at twenty did not exceed 10^{-20}. This shows that (3.1), (3.2) and (4.1) can be computed for most practical uses with their infinite sums truncated at twenty. The computations were performed using Maple. Maple took only a fraction of a second to compute the truncated versions of (3.1), (3.2) and (4.1). The computational times for the truncated versions were significantly smaller than those for the untruncated versions and those based on numerical integration.

Throughout this paper, we report conclusions on various properties of the WL distribution: the last four paragraphs of Section 2.1 reporting conclusions on the shape of the pdf of the WL distribution; the last paragraph of Section 2.2 reporting conclusions on the shape of the hazard rate function of the WL dis-

tribution; Section 2.3 reporting conclusions on the shape of the quartiles of the WL distribution; the last paragraph of Section 3 reporting conclusions on the mean, variance, skewness and kurtosis of the WL distribution; the last paragraph of Section 4 reporting conclusions on the Lorenz curve of the WL distribution. These conclusions are the result of extensive graphical analyses based on a wide range of parameter values (although the graphics presented here are based on a few choices of parameter values). However, we have no analytical proofs for these conclusions.

2. SHAPES

Here, we study the shapes of the pdf, (1.2), the corresponding hazard rate function and the corresponding quartiles. Shape properties are important because they allow the practitioner to see if the distribution can be fitted to a given data set (this can be seen, for example, by comparing the shape of the histogram of the data with possible shapes of the pdf). Shape properties of the hazard rate function are useful to see if the distribution can model increasing failure rates, decreasing failure rates or bathtub shaped failure rates. Shape properties of the hazard rate function has implications, for example, to the design of safe systems in a wide variety of applications. Quartiles are fundamental for estimation (for example, quartile estimators) and simulation.

2.1. Shape of probability density function

We can see from (1.2) that

$$\lim_{x\to 0} f(x) = \begin{cases} \infty, & \alpha < 1, \\ \dfrac{\beta(1+\lambda)+\lambda^2}{1+\lambda}, & \alpha = 1, \\ \dfrac{\lambda^2}{1+\lambda}, & \alpha > 1 \end{cases}$$

and

$$f(x) \sim \begin{cases} \dfrac{\lambda^2}{1+\lambda} x e^{-\lambda x-(\beta x)^\alpha}, & \alpha < 1, \\ \dfrac{(\beta+\lambda)\lambda}{1+\lambda} x e^{-\lambda x-\beta x}, & \alpha = 1, \\ \dfrac{\alpha\beta^\alpha\lambda}{1+\lambda} e^{-\lambda x-(\beta x)^\alpha}, & \alpha > 1 \end{cases}$$

as $x \to \infty$.

Note also that $f(x)$ can be written as

$$f(x) = g(x)\overline{Q}(x) + q(x)\overline{G}(x),$$

where

$$g(x) = \frac{\lambda^2(1+x)}{1+\lambda}e^{-\lambda x}$$

and

$$q(x) = \alpha\beta^{\alpha}x^{\alpha-1}e^{-(\beta x)^{\alpha}}.$$

So, the first derivative of $f(x)$ is

$$f'(x) = g'(x)\overline{Q}(x) + q'(x)\overline{G}(x) - 2g(x)q(x).$$

Therefore, $f(x)$ is decreasing if $g'(x) < 0$ and $q'(x) < 0$. This is possible if $\lambda \geq 1$ and $\alpha \leq 1$.

The first derivative of $f(x)$ is

$$\begin{aligned} f'(x) \;=\; & \frac{e^{-\lambda x-(\beta x)^{\alpha}}}{1+\lambda}\Bigg[(1+\lambda+\lambda x)\left[\alpha\beta^2(\alpha-1)(\beta x)^{\alpha-2} - (\alpha\beta)^2(\beta x)^{2\alpha-2}\right] \\ & -2\alpha\beta\lambda^2(\beta x)^{\alpha-1} + \lambda^2(1-\lambda-\lambda x)\Bigg]. \end{aligned}$$

So, the modes of $f(x)$ at say $x = x_0$ are the roots of

$$\begin{aligned} (2.1) \qquad (1+\lambda+\lambda x)\left[\alpha\beta^2(\alpha-1)(\beta x)^{\alpha-2} - (\alpha\beta)^2(\beta x)^{2\alpha-2}\right] = \\ = \; 2\alpha\beta\lambda^2(\beta x)^{\alpha-1} - \lambda^2(1-\lambda-\lambda x). \end{aligned}$$

The roots of (2.1) are difficult to find in general. However, if $\beta = 0$ then $x_0 = \frac{1-\lambda}{\lambda}$, the mode of the Lindley distribution, for $0 < \lambda < 1$.

We now study (2.1) graphically. Figure 1 shows possible shapes of the pdf of the WL distribution for selected (α, β, λ).

The left plot in Figure 1 shows bimodal shapes of the pdf with a maximum followed by a minimum. The x coordinates of the (local minimum, local maximum) are $(0.395, 0.933)$ for $\lambda = 1.2$, $(0.451, 0.921)$ for $\lambda = 1.4$, $(0.500, 0.906)$ for $\lambda = 1.6$ and $(0.547, 0.889)$ for $\lambda = 1.8$. The location of the minimum moves more to the right with increasing values of λ. The location of the maximum moves more to the left with increasing values of λ.

The right plot in Figure 1 shows unimodal shapes of the pdf. The x coordinates of the mode are 0.370 for $\lambda = 1.2$, 0.298 for $\lambda = 1.4$, 0.227 for $\lambda = 1.6$ and 0.157 for $\lambda = 1.8$. The location of the mode moves more to the left with increasing values of λ.

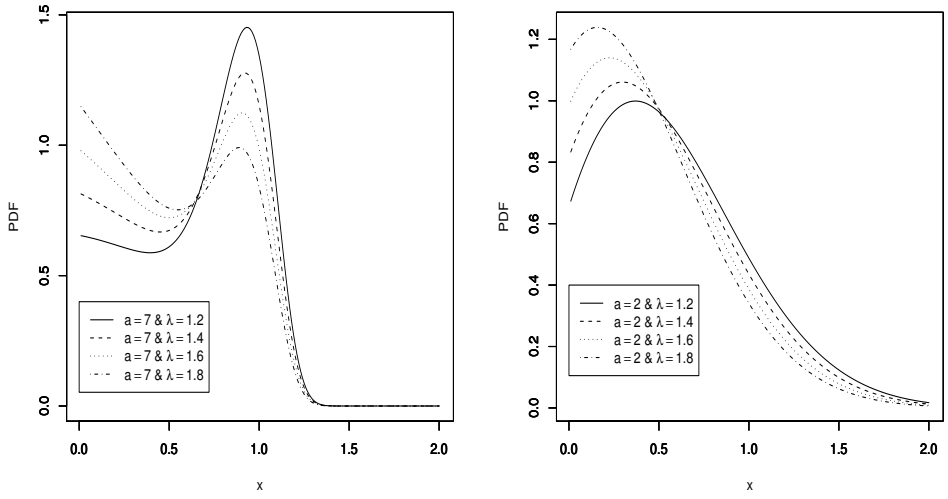


Figure 1: Pdfs of the WL distribution for $\beta = 1$.

Monotonically decreasing shapes and monotonically decreasing shapes containing an inflexion point are also possible for the pdf.

In each plot, the upper tails of the pdf become lighter with increasing values of λ. The lower tails of the pdf become heavier with increasing values of λ.

2.2. Shape of hazard rate function

Using (1.1) and (1.2), the hazard rate function of the WL distribution can be obtained as

$$h(x) = \frac{\lambda^2(1+x)}{1+\lambda+\lambda x} + \alpha\beta(\beta x)^{\alpha-1}. \tag{2.2}$$

It is obvious that the hazard rate functions of Lindley and Weibull distributions are contained as particular cases for $\beta = 0$ and $\lambda = 0$, respectively. Also, (2.2) can be expressed as

$$h_X(x) = h_Y(x) + h_Z(x),$$

i.e., the hazard rate function of the WL distribution is the sum of the hazard rate functions of Lindley and Weibull distributions. As a result, the hazard rate function of the WL distribution can exhibit monotonically increasing, monotonically decreasing and bathtub shapes, see Figure 2.

We can see from (2.2) that

$$\lim_{x\to 0} h(x) = \begin{cases} \infty, & \alpha < 1, \\ \dfrac{\beta(1+\lambda)+\lambda^2}{1+\lambda}, & \alpha = 1, \\ \dfrac{\lambda^2}{1+\lambda}, & \alpha > 1 \end{cases}$$

and

$$\lim_{x\to\infty} h(x) = \begin{cases} \lambda, & \alpha < 1, \\ \lambda+\beta, & \alpha = 1, \\ \infty, & \alpha > 1. \end{cases}$$

Bathtub shapes of the the hazard rate function appear possible when α is close enough to 0 or α is close enough to 1, see Figure 2. Monotonically decreasing shapes are possible for all values of α in between (i.e., in between α being close enough to 0 and α being close enough to 1). Monotonically increasing shapes are possible for all other values of α.

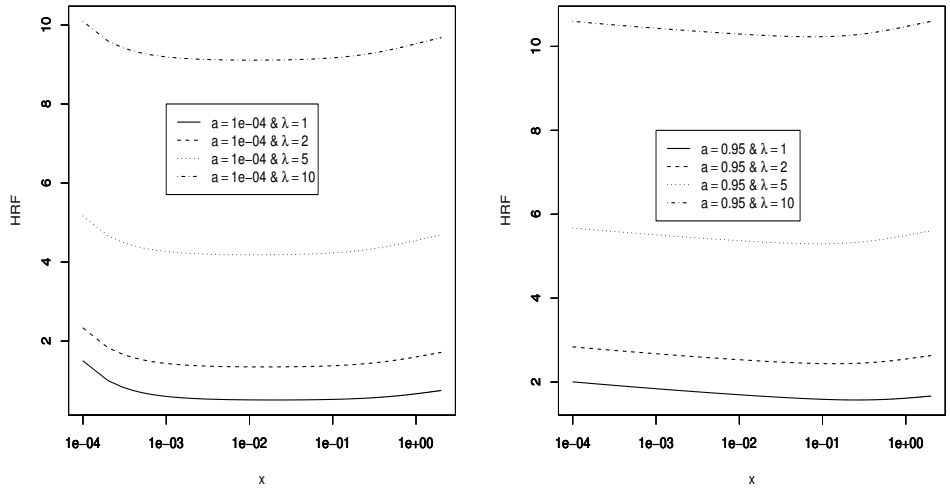


Figure 2: Hazard rate function of the WL distribution for $\beta = 1$.

2.3. Shape of quartiles

The p-th quartile say x_p of a WL random variable defined by $F(x_p) = p$ is the root of

$$x_p = \frac{1+\lambda}{\lambda}\left[(1-p)e^{\lambda x_p + (\beta x_p)^\alpha} - 1\right]$$

for $0 < p < 1$. Numerical investigations showed that x_p are monotonic decreasing functions of λ and monotonic increasing functions of α except for high quartiles.

3. MOMENT GENERATING FUNCTION AND MOMENTS

The moment generating function is fundamental for computing moments, factorial moments and cumulants of a random variable. It can also be used for estimation (for example, estimation methods based on empirical moment generating functions).

Moments are fundamental for any distribution. For instance, the first four moments can be used to describe any data fairly well. Moments are also useful for estimation (for example, the method of moments).

Several interesting characteristics of a distribution can be studied by moments and moment generating function. Let $X \sim WL(\alpha, \beta, \lambda)$. Then the moment generating function of X can be expressed as

$$\begin{aligned} M_X(t) &= E\left(e^{tX}\right) \\ &= \sum_{i=0}^{\infty} \frac{(-1)^i \beta^{i\alpha}}{i!(\lambda - t)^{i\alpha}} \left\{ \frac{\lambda^2 \Gamma(i\alpha + 1)}{(1+\lambda)(\lambda - t)} + \frac{\lambda^2 \Gamma(i\alpha + 2)}{(1+\lambda)(\lambda - t)^2} \right. \\ &\qquad \left. + \frac{\alpha \beta^{\alpha} \Gamma\left(\alpha(i+1)\right)}{(\lambda - t)^{\alpha}} + \frac{\alpha \lambda \beta^{\alpha} \Gamma\left(\alpha(i+1) + 1\right)}{(1+\lambda)(\lambda - t)^{\alpha+1}} \right\}, \end{aligned} \tag{3.1}$$

where $\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt$ denotes the gamma function. The r-th raw moment of X can be expressed as

$$\begin{aligned} \mu'_r = E\left(X^r\right) &= \frac{\lambda^2}{1+\lambda} \sum_{i=0}^{\infty} \frac{(-1)^i \beta^{i\alpha}}{i! \lambda^{i\alpha + r + 1}} \left[\Gamma(i\alpha + r + 1) + \frac{\Gamma(i\alpha + r + 2)}{\lambda} \right] \\ &\quad + \alpha \beta^{\alpha} \sum_{i=0}^{\infty} \frac{(-1)^i \beta^{i\alpha}}{i! \lambda^{i\alpha + \alpha + r}} \left[\Gamma(i\alpha + \alpha + r) + \frac{\Gamma(i\alpha + \alpha + r + 1)}{1+\lambda} \right]. \end{aligned} \tag{3.2}$$

The central moments, coefficient of variation, skewness and kurtosis of X can be readily obtained using the raw moments of X. Numerical investigations of the behavior of the mean, variance, skewness and kurtosis versus α and λ showed the following: i) mean is a monotonic decreasing function of λ; ii) mean is either a monotonic increasing function of α or initially decreases before increasing with respect to α; iii) variance is either a monotonic decreasing function of λ or initially increases before decreasing with respect to λ; iv) variance is either a monotonic decreasing function of α or a monotonic increasing function of α; v) skewness is either a monotonic decreasing function of λ or initially decreases before increasing with respect to λ; vi) skewness is either a monotonic decreasing function of α or initially increases before decreasing with respect to α; vii) kurtosis is either a monotonic decreasing function of λ, a monotonic increasing function of λ, initially decreases before increasing with respect to λ, or initially increases, then decreases

before increasing with respect to λ; viii) kurtosis is either a monotonic decreasing function of α, initially increases before decreasing with respect to α, initially decreases before increasing with respect to α, or initially increases, then decreases before increasing with respect to α.

4. LORENZ CURVE

The Lorenz curve for a positive random variable X is defined as the graph of the ratio

$$L\left(F(x)\right) = \frac{E(X \mid X \leq x)F(x)}{E(x)} = \frac{\int_0^x xf(x)dx}{\int_0^{+\infty} xf(x)dx}$$

against $F(x)$ with the property $L(p) \leq p$, $L(0) = 0$ and $L(1) = 1$. If X represents annual income, $L(p)$ is the proportion of total income that accrues to individuals having the $100p$ percent lowest incomes. If all individuals earn the same income then $L(p) = p$ for all p. The area between the line $L(p) = p$ and the Lorenz curve may be regarded as a measure of inequality of income, or more generally, of the variability of X.

The Lorenz curve has also received applications in areas other than income modeling: hierarchy theory for digraphs (Egghe [9]); depression and cognition (Maldonado *et al.* [21]); disease risk to optimize health benefits under cost constraints (Gail [11]); seasonal variation of environmental radon gas (Groves-Kirkby *et al.* [16]); statistical nonuniformity of sediment transport rate (Radice [26]).

For the WL distribution, we have

$$\begin{aligned}\int_0^x xf(x)dx \;=\; & \frac{1}{1+\lambda}\sum_{i=0}^{\infty}\frac{(-1)^i\beta^{i\alpha}}{i!\lambda^{i\alpha}}\gamma(i\alpha+2,\lambda x)\\ & +\frac{1}{1+\lambda}\sum_{i=0}^{\infty}\frac{(-1)^i\beta^{i\alpha}}{i!\lambda^{i\alpha+1}}\gamma(i\alpha+3,\lambda x)\\ & +\alpha\beta\sum_{i=0}^{\infty}\frac{(-1)^i\beta^{i\alpha+\alpha-1}}{i!\lambda^{i\alpha+\alpha+1}}\gamma(i\alpha+\alpha+1,\lambda x)\\ & +\frac{\alpha}{1+\lambda}\sum_{i=0}^{\infty}\frac{(-1)^i\beta^{i\alpha+\alpha}}{i!\lambda^{i\alpha+\alpha+1}}\gamma(i\alpha+\alpha+2,\lambda x),\end{aligned}$$

where $\gamma(a,x) = \int_0^x t^{a-1}e^{-t}dt$ denotes the incomplete gamma function. So, the

Lorenz curve for the WL distribution is

$$
\begin{aligned}
L(p) \;=\; & \frac{1}{\mu}\Bigg[\sum_{i=0}^{\infty}\frac{(-1)^i\beta^{i\alpha}}{i!\lambda^{i\alpha}}\gamma(i\alpha+2,\lambda x)+\frac{1}{1+\lambda}\sum_{i=0}^{\infty}\frac{(-1)^i\beta^{i\alpha}}{i!\lambda^{i\alpha+1}}\gamma(i\alpha+3,\lambda x) \\
& +\alpha\beta\sum_{i=0}^{\infty}\frac{(-1)^i\beta^{i\alpha+\alpha-1}}{i!\lambda^{i\alpha+\alpha+1}}\gamma(i\alpha+\alpha+1,\lambda x) \\
& +\frac{\alpha}{1+\lambda}\sum_{i=0}^{\infty}\frac{(-1)^i\beta^{i\alpha+\alpha}}{i!\lambda^{i\alpha+\alpha+1}}\gamma(i\alpha+\alpha+2,\lambda x)\Bigg]. \qquad (4.1)
\end{aligned}
$$

Possible shapes of (4.1) versus α and λ are shown in Figure 3. When $\alpha = 0.5$, the curves bend further towards the diagonal line as λ increases. When $\alpha = 1$, the curves bend further away from the diagonal line as λ increases. For each fixed λ, the curves bend further towards the diagonal line as α increases.

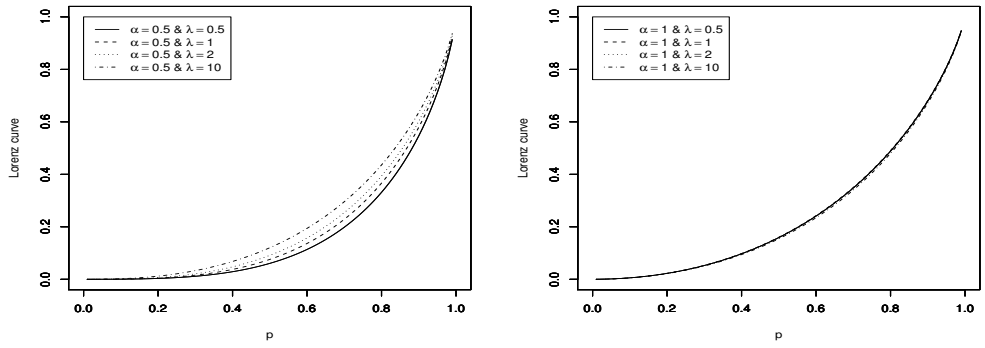


Figure 3: Lorenz curve of the WL distribution for $\beta = 1$.

5. ESTIMATION AND SIMULATION

Maximum likelihood estimation of the three parameters of the WL distribution is considered in Section 5.1. An assessment of the performance of the maximum likelihood estimators is performed in Section 5.2. Maximum likelihood estimation of the three parameters in the presence of censored data is considered in Section 5.3. A scheme for simulating from the WL distribution is given in Section 5.4.

5.1. Maximum likelihood estimation

Suppose $x_1, ..., x_n$ is a random sample from the WL distribution. The log-likelihood function is

$$\ell(\alpha,\beta,\lambda) = \sum_{i=0}^{n} \log\left[\lambda^2(1+x_i) + \alpha\lambda(\beta x_i)^{\alpha} + \alpha\beta(1+\lambda)(\beta x_i)^{\alpha-1}\right] - \lambda\sum_{i=0}^{n} x_i$$

$$- \sum_{i=0}^{n}(\beta x_i)^{\alpha} - n\log(1+\lambda). \tag{5.1}$$

The maximum likelihood estimators of (α, β, λ) can be obtained by solving the likelihood equations

$$\frac{\partial\ell(\alpha,\beta,\lambda)}{\partial\alpha} = \sum_{i=0}^{n}\frac{1}{W(x_j)}\Big\{[\beta(1+\lambda) + \alpha\beta(1+\lambda)\log(\beta x_i)](\beta x)^{\alpha-1}$$

$$+ [\lambda + \alpha\lambda\log(\beta x_i)](\beta x_i)^{\alpha}\Big\} - \lambda\sum_{i=0}^{n}(\beta x)^{\alpha}\log(\beta x_i) = 0, \tag{5.2}$$

$$\frac{\partial\ell(\alpha,\beta,\lambda)}{\partial\beta} = \sum_{i=0}^{n}\frac{1}{\beta W(x_i)}\left[\alpha\beta(1+x_i)(\beta x_i)^{\alpha-1} + \alpha^2\lambda(\beta x_i)^{\alpha}\right]$$

$$- \frac{\alpha}{\beta}\sum_{i=0}^{n}(\beta x_i)^{\alpha} = 0 \tag{5.3}$$

and

$$\frac{\partial\ell(\alpha,\beta,\lambda)}{\partial\lambda} = \sum_{i=0}^{n}\frac{1}{W(x_i)}\left[\alpha(\beta x_i)^{\alpha} + \alpha\beta(\beta x_i)^{\alpha-1} + 2\lambda(1+x_i)\right]$$

$$- \sum_{i=0}^{n} x_i - \frac{n}{1+\lambda} = 0, \tag{5.4}$$

where $W(x) = \lambda^2(1+x) + \alpha\lambda(\beta x)^{\alpha} + \alpha\beta(1+\lambda)(\beta x)^{\alpha-1}$. Alternatively, the MLEs can be obtained by maximizing (5.1) numerically. We shall use the latter approach in Sections 5.2 and 6. The maximization was performed by using the nlm function in R (R Development Core Team [25]). In Sections 5.2 and 6, the function nlm was executed with the initial values taken to be:

(**i**) The true parameter values (applicable for Section 5.2 only);

(**ii**) $\alpha = 0.01, 0.02, ..., 10$, $\beta = 0.01, 0.02, ..., 10$ and $\lambda = 0.01, 0.02, ..., 10$;

(**iii**) The moments estimates, i.e., the solutions $E(X) = (1/n)\sum_{i=1}^{n} x_i$, $E\left(X^2\right) = (1/n)\sum_{i=1}^{n} x_i^2$ and $E\left(X^3\right) = (1/n)\sum_{i=1}^{n} x_i^3$, where $E(X)$,

$E\left(X^2\right)$ and $E\left(X^3\right)$ are given by (3.2). These equations do not give explicit solutions. They were solved numerically using a quasi-Newton algorithm. Numerical investigations showed that this involved roughly the same amount of time as solving of $\frac{\partial \ell(\alpha,\beta,\lambda)}{\partial \alpha} = 0$, $\frac{\partial \ell(\alpha,\beta,\lambda)}{\partial \beta} = 0$ and $\frac{\partial \ell(\alpha,\beta,\lambda)}{\partial \lambda} = 0$ using a quasi-Newton algorithm.

In the cases of i) and iii), the function nlm converged all the time and the MLEs were unique. In the case of ii), the MLEs were unique whenever the function nlm converged. In the case of ii), the function nlm did not converge about five percent of the time.

For interval estimation of (α, β, λ), we consider the observed Fisher information matrix given by

$$I_F(\alpha,\beta,\lambda) = -\begin{pmatrix} I_{\alpha\alpha} & I_{\alpha\beta} & I_{\alpha\lambda} \\ I_{\beta\alpha} & I_{\beta\beta} & I_{\beta\lambda} \\ I_{\lambda\alpha} & I_{\lambda\beta} & I_{\lambda\lambda} \end{pmatrix},$$

where $I_{\phi_1\phi_2} = \partial^2\ell/\partial\phi_1\partial\phi_2$.

Under certain regularity conditions (see, for example, Ferguson [10]) and Lehmann and Casella [19], pages 461-463) and for large n, the distribution of $\sqrt{n}\left(\widehat{\alpha}-\alpha, \widehat{\beta}-\beta, \widehat{\lambda}-\lambda\right)$ can be approximated by a trivariate normal distribution with zero means and variance-covariance matrix given by the inverse of the observed information matrix evaluated at the maximum likelihood estimates. This approximation can be used to construct approximate confidence intervals, confidence regions, and testing hypotheses for the parameters. For example, an asymptotic confidence interval for α with level $1-\gamma$ is $\left(\widehat{\alpha} \mp z_{1-\gamma/2}\sqrt{I^{\widehat{\alpha},\widehat{\alpha}}}\right)$, where $I^{\widehat{\alpha},\widehat{\alpha}}$ is the $(1,1)$-th element of the inverse of $I_F\left(\widehat{\alpha},\widehat{\beta},\widehat{\lambda}\right)$ and $z_{1-\gamma/2}$ is the $(1-\gamma/2)$-th quartile of the standard normal distribution.

5.2. Simulation study

Here, we assess the performance of the maximum likelihood estimators given by (5.2)–(5.4) with respect to sample size n. The assessment was based on a simulation study:

1. Generate ten thousand samples of size n from (1.2). The inversion method was used to generate samples, i.e., variates of the WL distribution were generated using

$$U = \frac{1+\lambda}{\lambda}\left[(1-p)e^{\lambda X+(\beta X)^\alpha} - 1\right],$$

where $U \sim U(0,1)$ is a uniform variate on the unit interval.

2. Compute the maximum likelihood estimates for the ten thousand samples, say $\left(\widehat{\alpha}_i, \widehat{\beta}_i, \widehat{\lambda}_i\right)$ for $i = 1, 2, ..., 10000$.

3. Compute the biases and mean squared errors given by

$$\text{bias}_h(n) = \frac{1}{10000}\sum_{i=1}^{10000}\left(\widehat{h}_i - h\right), \quad \text{MSE}_h(n) = \frac{1}{10000}\sum_{i=1}^{10000}\left(\widehat{h}_i - h\right)^2$$

for $h = \alpha, \beta, \lambda$.

We repeated these steps for $n = 10, 11, ..., 100$ with $\alpha = 1$, $\beta = 1$ and $\lambda = 1$, so computing $\text{bias}_h(n)$ and $\text{MSE}_h(n)$ for $h = \alpha, \beta, \lambda$ and $n = 10, 11, ..., 100$.

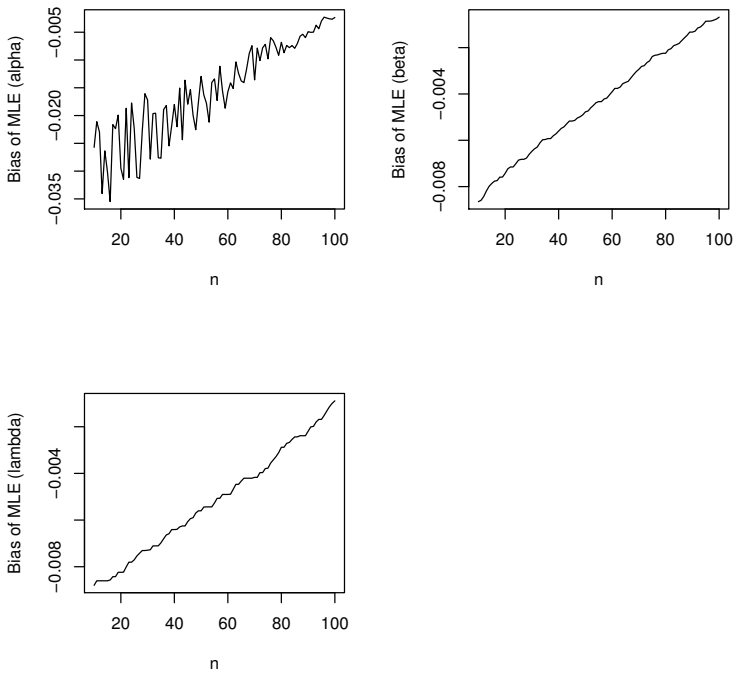


Figure 4: Biases of $\left(\widehat{\alpha}, \widehat{\beta}, \widehat{\lambda}\right)$ versus n.

Figures 4 and 5 show how the three biases and the three mean squared errors vary with respect to n. The following observations can be made: the biases for each parameter are negative; the biases appear largest for the parameter, α; the biases appear smallest for the parameters, β and λ; the biases for each parameter increase to zero as $n \to \infty$; the mean squared errors for each parameter decrease to zero as $n \to \infty$; the mean squared errors appear largest for the parameter, α; the mean squared errors appear smallest for the parameter, λ. These observations are for only one choice for (α, β, λ), namely that $(\alpha, \beta, \lambda) = (1, 1, 1)$. But the results were similar for a wide range of other values of (α, β, λ). In particular,

the biases for each parameter always increased to zero as $n \to \infty$ and the mean squared errors for each parameter always decreased to zero as $n \to \infty$.

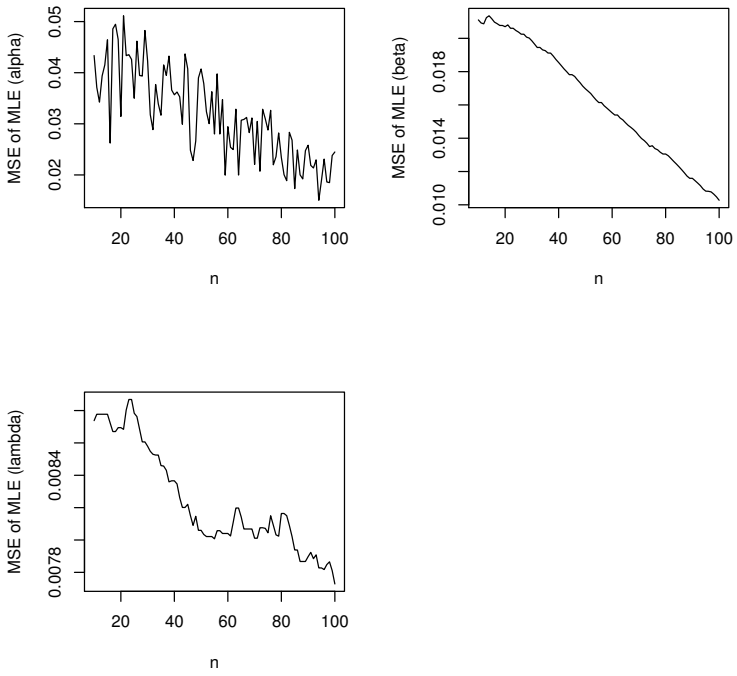


Figure 5: Mean squared errors of $\left(\widehat{\alpha}, \widehat{\beta}, \widehat{\lambda}\right)$ versus n.

Section 6 presents three real data applications. The sample size for the first data set is eight hundred and seventy seven. The sample size for the second data set is twenty six. The sample size for the third data set is two hundred and ninety five. We shall see later in Section 6 that the WL distribution provides good fits to the three data sets. Based on this fact, the biases for $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{\lambda}$ can be expected to be less than 0.025, 0.007 and 0.0075, respectively, for all of the data sets. The mean squared errors for $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{\lambda}$ can be expected to be less than 0.04, 0.02 and 0.0088, respectively, for all of the data sets. Hence, the point estimates given in Section 6 for all data sets can be considered accurate enough.

5.3. Censored maximum likelihood estimation

Often with lifetime data, we encounter censored data. There are different forms of censoring: type I censoring, type II censoring, etc. Here, we consider the general case of multi-censored data: there are n subjects of which

- n_0 are known to have failed at the times $x_1, ..., x_{n_0}$;
- n_1 are known to have failed in the interval $[s_{j-1}, s_j]$, $j = 1, ..., n_1$;
- n_2 survived to a time r_j, $j = 1, ..., n_2$ but not observed any longer.

Note that $n = n_0 + n_1 + n_2$ and that type I censoring and type II censoring are contained as particular cases of multi-censoring. The log-likelihood function of the model parameters for this multi-censoring data is:

$$\begin{aligned}\ell(\alpha, \beta, \lambda) \;=\; & \sum_{i=0}^{n_0} \log\left[\lambda^2 (1 + x_i) + \alpha\lambda (\beta x_i)^{\alpha} + \alpha\beta(1 + \lambda) (\beta x_i)^{\alpha-1}\right] - \lambda \sum_{i=0}^{n_0} x_i \\ & - \sum_{i=0}^{n_0} (\beta x_i)^{\alpha} - n_0 \log(1 + \lambda) \\ & + \sum_{i=1}^{n_1} \log\left\{F(s_i) - F(s_{i-1})\right\} \\ & + \sum_{i=1}^{n_2} \log\left\{1 - F(r_i)\right\}, \end{aligned} \tag{5.5}$$

where $F(\cdot)$ is given by (1.1). The MLEs can be obtained by maximizing (5.5) numerically. The maximization can be performed by using the nlm function in R.

5.4. Generating data

Section 5.2 gave an inversion method for simulating from the WL distribution. Here, we present an alternative method for simulation.

We know that a WL random variable is the minimum of independent Weibull and Lindley random variables. So, to generate a random sample from the WL distribution, the following algorithm can also be used:

1. First generate a random sample $v_1, ..., v_n$ from Weibull(α, β);

2. Independently, generate a random sample $w_1, ..., w_n$ from Lindley(λ);

3. Set $x_i = \min(v_i, w_i)$ for $i = 1, ..., n$.

Then $x_1, x_2, ..., x_n$ will be a random sample from WL(α, β, λ).

6. APPLICATIONS

In this section, we fit the WL distribution to three real data sets. We compare the fits of the WL distribution to the fits of some related distributions: the extended Lindley (EL) distribution due to Bakouch *et al.* [4] with the pdf

$$f(x) = \frac{\lambda(1+\lambda+\lambda x)^{\alpha-1}}{(1+\lambda)^{\alpha}}\left[\beta(1+\lambda+\lambda x)(\lambda x)^{\beta-1}-\alpha\right]e^{-(\lambda x)^{\beta}}$$

for $x>0$, $\alpha\in(-\infty,0)\cup\{0,1\}$, $\beta>0$ and $\lambda>0$; the weighted Lindley (WEL) distribution due to Ghitany *et al.* [14] with the pdf

$$f(x) = \frac{\theta^{c+1}}{(\theta+c)\Gamma(c)}x^{c-1}(1+x)e^{-\theta x}$$

for $x>0$, $c>0$ and $\theta>0$; the exponential Poisson Lindley (EPL) distribution due to Barreto-Souza and Bakouch [5] with the pdf

$$f(x) = \frac{\beta\theta^2(1+\theta)^2e^{-\beta x}\left(3+\theta-e^{-\beta x}\right)}{\left(1+3\theta+\theta^2\right)\left(1+\theta-e^{-\beta x}\right)^3}$$

for $x>0$, $\theta>0$ and $\beta>0$; the Lindley distribution with the pdf

$$f(x) = \frac{\lambda^2}{\lambda+1}(1+x)e^{-\lambda x}$$

for $x>0$ and $\lambda>0$; the generalized Lindley (GL1) distribution due to Zalerzadeh and Dolati [29] with the pdf

$$f(x) = \frac{\theta^2(\theta x)^{\alpha-1}(\alpha+\gamma x)e^{-\theta x}}{(\gamma+\theta)\Gamma(\alpha+1)}$$

for $x>0$, $\alpha>0$, $\theta>0$ and $\gamma>0$; the power Lindley (PL) distribution due to Ghitany *et al.* [13] with the pdf

$$f(x) = \frac{\alpha\beta^2}{\beta+1}(1+x^{\alpha})x^{\alpha-1}e^{-\beta x^{\alpha}}$$

for $x>0$, $\alpha>0$ and $\beta>0$; and, the generalized Lindley (GL2) distribution due to Nadarajah *et al.* [23] with the pdf

$$f(x) = \frac{\alpha\lambda^2}{1+\lambda}(1+x)\left[1-\frac{1+\lambda+\lambda x}{1+\lambda}e^{-\lambda x}\right]^{\alpha-1}e^{-\lambda x}$$

for $x>0$, $\alpha>0$ and $\lambda>0$.

6.1. Data set 1

The first data are times to reinfection of STD for eight hundred and seventy seven patients. The data were taken from Section 1.12 of Klein and Moeschberger [17]. We fitted the eight distributions to the data. Table 1 gives the parameter estimates, standard errors obtained by inverting the observed information matrix, log-likelihood values, values of Akaike information criterion (AIC), values of Bayesian information criterion (BIC), p values based on the Kolmogorov–Smirnov (KS) statistic, p values based on the Anderson–Darling (AD) statistic, and p values based on the Cramér–von Mises (CVM) statistic. The fitted pdfs of the three best fitting distributions as well as the empirical histogram are shown in Figure 6. The corresponding probability plots are shown in Figure 7.

Table 1: Parameter estimates, standard errors, log-likelihoods, AICs, BICs and goodness-of-fit measures for data set 1.

Distribution	Parameter estimates (s.e)	$-\log L$	AIC	BIC	KS	AD	CVM
EL	$\widehat{\lambda} = 8.806\times10^{-1}(1.302\times10^{-2})$, $\widehat{\alpha} = -9.804\times10^{-1}(3.034\times10^{-2})$, $\widehat{\beta} = 9.935\times10^{-7}(8.098\times10^{-3})$	9203.4	18412.9	18427.2	4.080×10^{-4}	2.700×10^{-5}	1.979×10^{-4}
WEL	$\widehat{\theta} = 2.878\times10^{-3}(1.076\times10^{-4})$, $\widehat{c} = 9.359\times10^{-2}(1.324\times10^{-2})$	6082.4	12168.7	12178.3	9.474×10^{-3}	2.879×10^{-4}	6.021×10^{-2}
EPL	$\widehat{\theta} = 2.326(7.568\times10^{-1})$, $\widehat{\beta} = 2.190\times10^{-3}(1.854\times10^{-4})$	6055.1	12114.1	12123.7	8.151×10^{-2}	1.395×10^{-2}	1.174×10^{-1}
Lindley	$\widehat{\lambda} = 5.397\times10^{-3}(1.313\times10^{-4})$	6413.0	12828.1	12832.9	1.996×10^{-3}	1.762×10^{-4}	5.579×10^{-4}
GL1	$\widehat{\theta} = 7.872\times10^{-2}(4.339\times10^{-3})$, $\widehat{\alpha} = 1.453\times10^{-6}(4.164\times10^{-2})$, $\widehat{\gamma} = 7.595\times10^{-1}(6.225\times10^{-2})$	27827.1	55660.1	55674.4	1.864×10^{-5}	2.149×10^{-5}	1.697×10^{-4}
PL	$\widehat{\alpha} = 5.696\times10^{-1}(1.364\times10^{-2})$, $\widehat{\beta} = 7.671\times10^{-2}(6.420\times10^{-3})$	6056.3	12116.7	12126.2	6.872×10^{-2}	1.232×10^{-3}	7.611×10^{-2}
GL2	$\widehat{\lambda} = 2.980\times10^{-3}(1.391\times10^{-4})$, $\widehat{\alpha} = 3.660\times10^{-1}(1.509\times10^{-2})$	6031.8	12067.5	12077.1	1.695×10^{-1}	7.568×10^{-2}	1.480×10^{-1}
WL	$\widehat{\lambda} = 2.331\times10^{-3}(2.714\times10^{-4})$, $\widehat{\alpha} = 6.435\times10^{-1}(3.870\times10^{-2})$, $\widehat{\beta} = 1.740\times10^{-3}(2.792\times10^{-4})$	6022.9	12051.7	12066.0	3.131×10^{-1}	8.243×10^{-2}	2.735×10^{-1}

We can see that the WL distribution gives the smallest AIC value, the smallest BIC value, the largest p value based on the KS statistic, the largest p value based on the AD statistic, and the largest p value based on the CVM statistic. The second smallest AIC, BIC values and the second largest p values are given by the GL2 distribution. The third smallest AIC, BIC values and the third largest p values are given by the EPL distribution. The fourth smallest AIC, BIC values and the fourth largest p values are given by the PL distribution. The fifth smallest AIC, BIC values and the fifth largest p values are given by the WEL distribution. The sixth smallest AIC, BIC values and the sixth largest p values are given by the Lindley distribution. The seventh smallest AIC, BIC values and the seventh largest p values are given by the EL distribution. The largest AIC, BIC values and the smallest p values are given by the GL1 distribution.

Hence, the WL distribution provides the best fit based on the AIC values, BIC values, p values based on the KS statistic, p values based on the AD statistic, and p values based on the CVM statistic. The density and probability plots also show that the WL distribution provides the best fit.

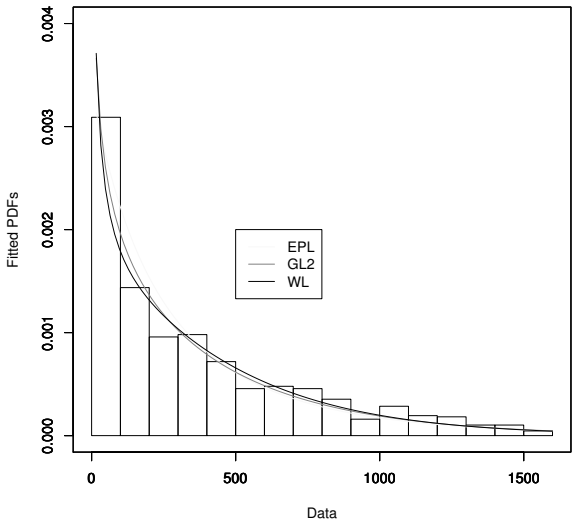


Figure 6: Fitted pdfs of the three best fitting distributions for data set 1.

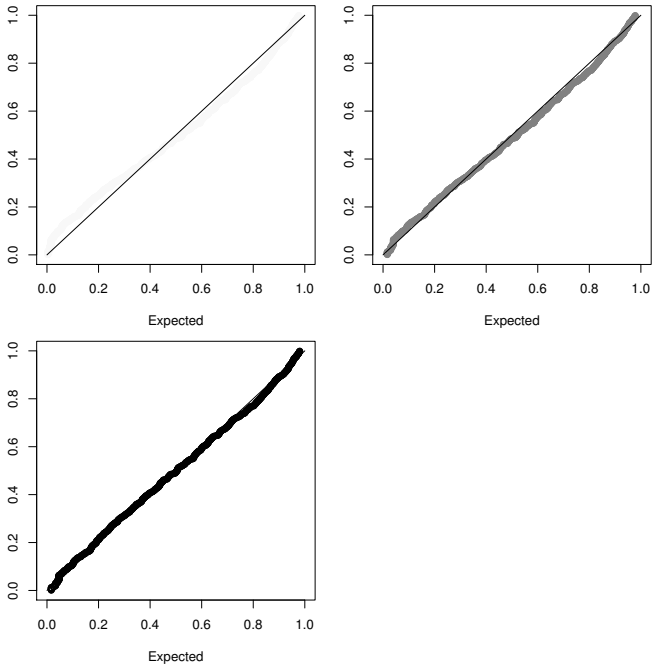


Figure 7: PP plots of the three best fitting distributions for data set 1 (yellow for the EPL distribution, red for the GL2 distribution and black for the WL distribution).

6.2. Data set 2

The second data are times to death of twenty six psychiatric patients. The data were taken from Section 1.15 of Klein and Moeschberger [17]. The eight distributions were fitted to this data. The parameter estimates, standard errors and the various measures are given in Table 2. The corresponding density and probability plots are shown in Figures 8 and 9, respectively.

Table 2: Parameter estimates, standard errors, log-likelihoods, AICs, BICs and goodness-of-fit measures for data set 2.

Distribution	Parameter estimates (s.e)	$-\log L$	AIC	BIC	KS	AD	CVM
EL	$\widehat{\lambda} = 7.510 \times 10^{-1}(9.604 \times 10^{-2})$, $\widehat{\alpha} = -8.534 \times 10^{-1}(2.414 \times 10^{-1})$, $\widehat{\beta} = 2.050 \times 10^{-6}(1.197 \times 10^{-1})$	164.9	335.8	339.5	1.163×10^{-3}	3.780×10^{-5}	9.083×10^{-5}
WEL	$\widehat{\theta} = 7.727 \times 10^{-2}(2.090 \times 10^{-2})$, $\widehat{c} = 1.107(4.659 \times 10^{-1})$	107.7	219.3	221.8	4.691×10^{-1}	1.392×10^{-3}	3.064×10^{-2}
EPL	$\widehat{\theta} = 1.267 \times 10^{4}(4.662 \times 10^{5})$, $\widehat{\beta} = 3.784 \times 10^{-2}(7.442 \times 10^{-3})$	111.1	226.3	228.8	1.840×10^{-2}	4.964×10^{-5}	1.024×10^{-4}
Lindley	$\widehat{\lambda} = 7.311 \times 10^{-2}(1.016 \times 10^{-2})$	107.7	217.4	218.6	7.924×10^{-1}	1.126×10^{-2}	4.895×10^{-2}
GL1	$\widehat{\theta} = 8.282 \times 10^{-2}(2.152 \times 10^{-2})$, $\widehat{\alpha} = 1.420(5.569 \times 10^{-1})$, $\widehat{\gamma} = 2.742 \times 10^{-1}(3.199 \times 10^{-1})$	107.1	220.3	224.0	2.920×10^{-2}	1.163×10^{-4}	7.787×10^{-4}
PL	$\widehat{\alpha} = 1.225(2.069 \times 10^{-1})$, $\widehat{\beta} = 3.452 \times 10^{-2}(2.470 \times 10^{-2})$	106.9	217.8	220.3	6.957×10^{-1}	8.737×10^{-3}	3.715×10^{-2}
GL2	$\widehat{\lambda} = 7.547 \times 10^{-2}(1.398 \times 10^{-2})$, $\widehat{\alpha} = 1.069(2.874 \times 10^{-1})$	107.7	219.3	221.8	1.081×10^{-1}	1.582×10^{-4}	4.384×10^{-3}
WL	$\widehat{\lambda} = 4.340 \times 10^{-2}(1.051 \times 10^{-2})$, $\widehat{\alpha} = 9.901(2.822)$, $\widehat{\beta} = 2.832 \times 10^{-2}(1.014 \times 10^{-3})$	93.4	192.8	196.5	8.998×10^{-1}	8.372×10^{-1}	2.849×10^{-1}

We can see again that the WL distribution gives the smallest AIC value, the smallest BIC value, the largest p value based on the KS statistic, the largest p value based on the AD statistic, and the largest p value based on the CVM statistic. The second smallest AIC, BIC values and the second largest p values are given by the Lindley distribution. The third smallest AIC, BIC values and the third largest p values are given by the PL distribution. The fourth smallest AIC, BIC values and the fourth largest p values are given by the WEL distribution. The fifth smallest AIC, BIC values and the fifth largest p values are given by the GL2 distribution. The sixth smallest AIC, BIC values and the sixth largest p values are given by the GL1 distribution. The seventh smallest AIC, BIC values and the seventh largest p values are given by the EPL distribution. The largest AIC, BIC values and the smallest p values are given by the EL distribution.

Hence, the WL distribution again provides the best fit based on the AIC values, BIC values, p values based on the KS statistic, p values based on the AD

statistic, and p values based on the CVM statistic. The density and probability plots again show that the WL distribution provides the best fit.

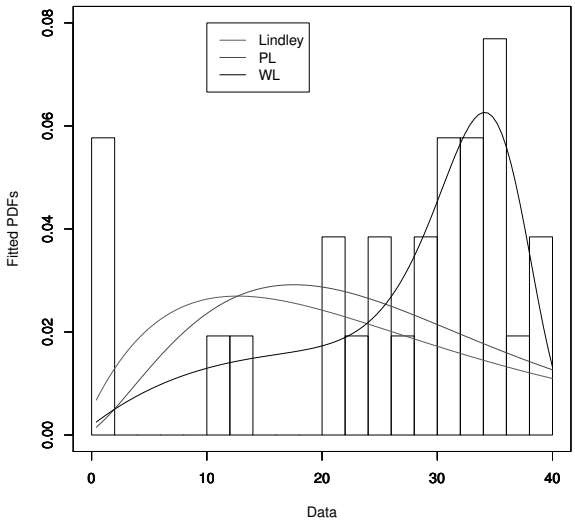


Figure 8: Fitted pdfs of the three best fitting distributions for data set 2.

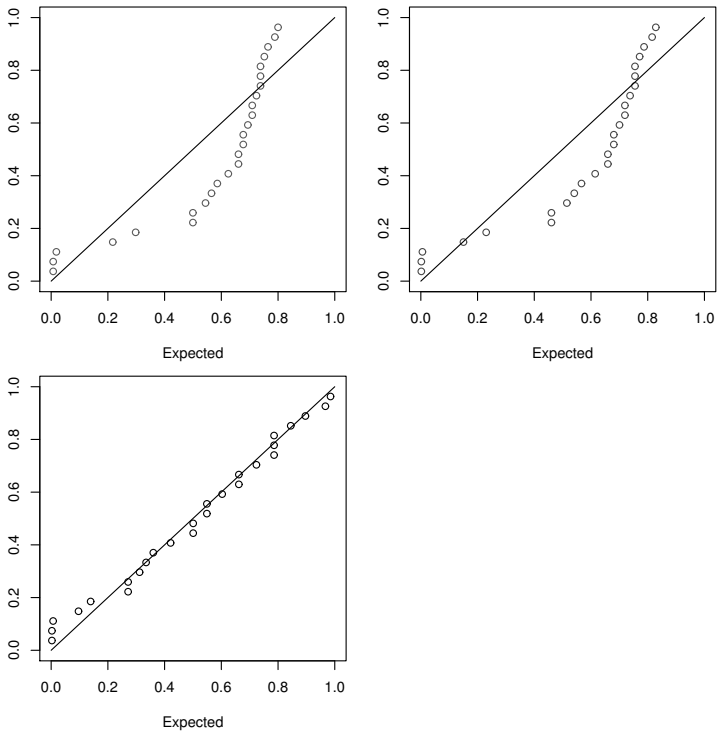


Figure 9: PP plots of the three best fitting distributions for data set 2 (brown for the Lindley distribution, blue for the PL distribution and black for the WL distribution).

6.3. Data set 3

The third data are times to infection for AIDS for two hundred and ninety five patients. The data were taken from Section 1.19 of Klein and Moeschberger [17]. The eight distributions were fitted to this data. The parameter estimates, standard errors and the various measures are given in Table 3. The corresponding density and probability plots are shown in Figures 10 and 11, respectively.

Table 3: Parameter estimates, standard errors, log-likelihoods, AICs, BICs and goodness-of-fit measures for data set 3.

Distribution	Parameter estimates (s.e)	$-\log L$	AIC	BIC	KS	AD	CVM
EL	$\widehat{\lambda} = 2.066\times10^{-1}(5.340\times10^{-3})$, $\widehat{\alpha} = -1.425\times10^{-1}(8.097\times10^{-2})$, $\widehat{\beta} = 3.503(2.776\times10^{-1})$	537.5	1080.9	1092.0	7.775×10^{-1}	9.719×10^{-1}	1.866×10^{-1}
WEL	$\widehat{\theta} = 1.396(1.130\times10^{-1})$, $\widehat{c} = 5.025(4.472\times10^{-1})$	563.3	1130.6	1138.0	2.072×10^{-1}	1.269×10^{-1}	9.133×10^{-3}
EPL	$\widehat{\theta} = 1.145\times10^{4}(1.091\times10^{5})$, $\widehat{\beta} = 2.403\times10^{-1}(1.402\times10^{-2})$	713.2	1430.4	1437.8	6.394×10^{-2}	2.022×10^{-2}	7.102×10^{-4}
Lindley	$\widehat{\lambda} = 4.106\times10^{-1}(1.731\times10^{-2})$	659.7	1321.4	1325.1	1.113×10^{-1}	8.193×10^{-2}	1.443×10^{-3}
GL1	$\widehat{\theta} = 1.402(1.157\times10^{-1})$, $\widehat{\alpha} = 5.099(5.533\times10^{-1})$, $\widehat{\gamma} = 3.872(4.373)$	563.3	1132.5	1143.6	1.744×10^{-1}	9.239×10^{-2}	3.008×10^{-3}
PL	$\widehat{\alpha} = 2.099(8.683\times10^{-2})$, $\widehat{\beta} = 8.357\times10^{-2}(1.176\times10^{-2})$	544.7	1093.5	1100.9	5.437×10^{-1}	1.664×10^{-1}	3.248×10^{-2}
GL2	$\widehat{\lambda} = 7.544\times10^{-1}(3.321\times10^{-2})$, $\widehat{\alpha} = 4.536(4.812\times10^{-1})$	571.4	1146.9	1154.2	1.136×10^{-1}	9.136×10^{-2}	1.616×10^{-3}
WL	$\widehat{\lambda} = 1.595\times10^{-1}(3.235\times10^{-2})$, $\widehat{\alpha} = 4.036(4.329\times10^{-1})$, $\widehat{\beta} = 1.949\times10^{-1}(6.412\times10^{-3})$	535.7	1077.4	1088.4	8.059×10^{-1}	9.908×10^{-1}	8.666×10^{-1}

We can see yet again that the WL distribution gives the smallest AIC value, the smallest BIC value, the largest p value based on the KS statistic, the largest p value based on the AD statistic, and the largest p value based on the CVM statistic. The second smallest AIC, BIC values and the second largest p values are given by the EL distribution. The third smallest AIC, BIC values and the third largest p values are given by the PL distribution. The fourth smallest AIC, BIC values and the fourth largest p values are given by the WEL distribution. The fifth smallest AIC, BIC values and the fifth largest p values are given by the GL1 distribution. The sixth smallest AIC, BIC values and the sixth largest p values are given by the GL2 distribution. The seventh smallest AIC, BIC values and the seventh largest p values are given by the Lindley distribution. The largest AIC, BIC values and the smallest p values are given by the EPL distribution.

Hence, the WL distribution yet again provides the best fit based on the AIC values, BIC values, p values based on the KS statistic, p values based on the AD

statistic, and p values based on the CVM statistic. The density and probability plots yet again show that the WL distribution provides the best fit.

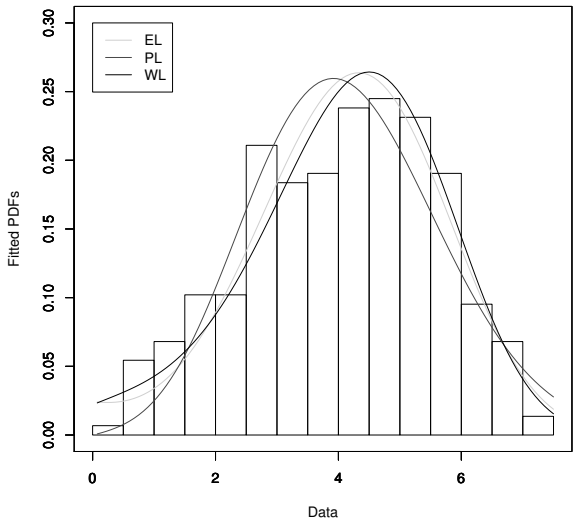


Figure 10: Fitted pdfs of the three best fitting distributions for data set 3.

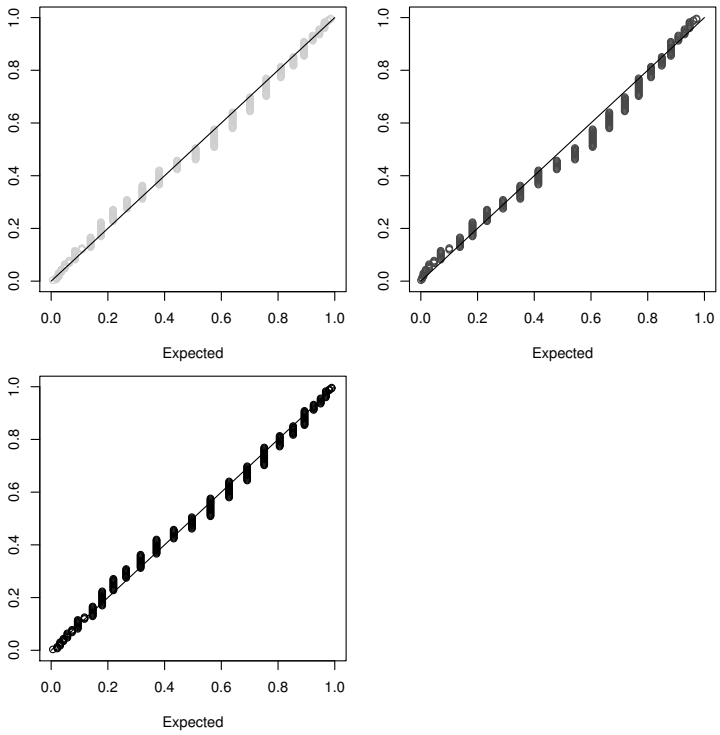


Figure 11: PP plots of the three best fitting distributions for data set 3 (pink for the EL distribution, blue for the PL distribution and black for the WL distribution).

7. CONCLUSIONS

We have introduced a three-parameter generalization of the Lindley distribution referred to as the Weibull Lindley distribution. We have provided at least seven possible motivations for this new distribution. We have studied shapes of probability density and hazard rate functions, moments, moment generating function, Lorenz curve, maximum likelihood estimators in the presence of complete data and maximum likelihood estimators in the presence of censored data. We have assessed the finite sample performance of the maximum likelihood estimators by simulation. We have provided three real data applications.

We have seen that the probability density function can be bimodal, unimodal, monotonically decreasing or monotonically decreasing with an inflexion point. The hazard rate function can be monotonically increasing, monotonically decreasing or bathtub shaped. The maximum likelihood estimators appear to be regular for sample sizes larger than twenty. The data applications show that the Weibull Lindley distribution provides better fits than all known generalizations of the Lindley distribution for at least three data sets.

ACKNOWLEDGMENTS

The authors would like to thank the Editor, the Associate Editor and the referee for careful reading and comments which greatly improved the paper.

REFERENCES

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

[2] Andersen, P.K. and Skrondal, A. (2015). A competing risks approach to "biologic" interaction, *Lifetime Data Analysis*, **21**, 300–314.

[3] Asgharzadeh, A.; Bakouch, H.S. and Esmaeili, L. (2013). Pareto Poisson–Lindley distribution with applications, *Journal of Applied Statistics*, **40**, 1717–1734.

[4] Bakouch, H.S.; Al-Zahrani, B.M.; Al-Shomrani, A.A.; Marchi, V.A.A. and Louzada, F. (2012). An extended Lindley distribution, *Journal of the Korean Statistical Society*, **41**, 75–85.

[5] Barreto-Souza, W. and Bakouch, H.S. (2013). A new lifetime model with decreasing failure rate, *Statistics*, **47**, 465–476.

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

[7] Bourguignon, M.; Silva, R.B. and Cordeiro, G.M. (2014). A new class of fatigue life distributions, *Journal of Statistical Computation and Simulation*, **84**, 2619–2635.

[8] Cadarso-Suarez, C.; Meira-Machado, L.; Kneib, T. and Gude, F. (2010). Flexible hazard ratio curves for continuous predictors in multi-state models: An application to breast cancer data, *Statistical Modelling*, **10**, 291–314.

[9] Egghe, L. (2002). Development of hierarchy theory for digraphs using concentration theory based on a new type of Lorenz curve, *Mathematical and Computer Modelling*, **36**, 587–602.

[10] Ferguson, T.S. (1996). *A Course in Large Sample Theory*, Chapman and Hall, London.

[11] Gail, M.H. (2009). Applying the Lorenz curve to disease risk to optimize health benefits under cost constraints, *Statistics and Its Interface*, **2**, 117–121.

[12] Gail, M.H. and Benichou, J. (2000). *Encyclopedia of Epidemiologic Methods*, John Wiley and Sons.

[13] Ghitany, M.E.; Al-Mutairi, D.K.; Balakrishnan, N. and Al-Enezi, L.J. (2013). Power Lindley distribution and associated inference, *Computational Statistics and Data Analysis*, **64**, 20–33.

[14] Ghitany, M.E.; Alqallaf, F.; Al-Mutairi, D.K. and Husain, H.A. (2011). A two-parameter weighted Lindley distribution and its applications to survival data, *Mathematics and Computers in Simulation*, **81**, 1190–1201.

[15] Ghitany, M.E.; Atieh, B. and Nadarajah, S. (2008). Lindley distribution and its application, *Mathematics and Computers in Simulation*, **78**, 493–506.

[16] Groves-Kirkby, C.J.; Denman, A.R. and Phillips, P.S. (2009). Lorenz curve and Gini coefficient: Novel tools for analysing seasonal variation of environmental radon gas, *Journal of Environmental Management*, **90**, 2480–2487.

[17] Klein, J.P. and Moeschberger, M.L. (1997). *Survival Analysis Techniques for Censored and Truncated Data*, Springer Verlag, New York.

[18] Kus, C. (2007). A new lifetime distribution, *Computational Statistics and Data Analysis*, **51**, 4497–4509.

[19] Lehmann, L.E. and Casella, G. (1998). *Theory of Point Estimation*, second edition, Springer Verlag, New York.

[20] Lindley, D.V. (1958). Fiducial distributions and Bayes' theorem, *Journal of the Royal Statistical Society*, A, **20**, 102–107.

[21] Maldonado, A.; Perez-Ocon, R. and Herrera, A. (2007). Depression and cognition: New insights from the Lorenz curve and the Gini index, *International Journal of Clinical and Health Psychology*, **7**, 21–39.

[22] Morais, A.L. and Barreto-Souza, W. (2011). A compound class of Weibull and power series distributions, *Computational Statistics and Data Analysis*, **55**, 1410–1425.

[23] Nadarajah, S.; Bakouch, H.S. and Tahmasbi, R. (2011). A generalized Lindley distribution, *Sankhyā*, B, **73**, 331–359.

[24] Nair, U.N.; Sankaran, P.G. and Balakrishnan, N. (2010). *Quantile-Based Reliability Analysis*, Birkhauser, Boston.

[25] R Development Core Team (2014). *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

[26] Radice, A. (2009). Use of the Lorenz curve to quantify statistical nonuniformity of sediment transport rate, *Journal of Hydraulic Engineering*, **135**, 320–326.

[27] Silva, R.B.; Bourguignon, M.; Dias, C.R.B. and Cordeiro, G.M. (2013). The compound class of extended Weibull power series distributions, *Computational Statistics and Data Analysis*, **58**, 352–367.

[28] Yamaguchi, K. (1993). Modeling time-varying effects of covariates in event-history analysis using statistics from the saturated hazard rate model, *Sociological Methodology*, **23**, 279–317.

[29] Zakerzadeh, H. and Dolati, A. (2009). Generalized Lindley distribution, *Journal of Mathematical Extension*, **3**, 13–25.

REVSTAT – Statistical Journal
Volume 16, Number 1, January 2018, 115–136

HEURISTIC TOOLS FOR THE ESTIMATION OF THE EXTREMAL INDEX: A COMPARISON OF METHODS

Author: Marta Ferreira
– Center of Mathematics of Minho University and CEMAT, CEAUL — University of Lisbon, Portugal
msferreira@math.uminho.pt

Received: July 2015 Revised: July 2016 Accepted: July 2016

Abstract:

- Clustering of exceedances of a critical level is a phenomenon that concerns risk managers in many areas. The extremal index θ measures the propensity of the large observations in a dataset to cluster. Thus the estimation of θ is an important issue recurrently addressed in literature. Besides a declustering parameter, inference also depends on a threshold. This choice is actually a crucial topic and is transversal to many other extremal parameters. In this paper we analyze a threshold-free heuristic procedure. We also make comparisons with other heuristic procedures already developed within the extremal index estimation. Our study is based on simulation. We illustrate with an application to environmental data.

Key-Words:

- *extreme value theory; extremal index estimation; heuristic methods.*

AMS Subject Classification:

- 62G32, 60G70.

1. INTRODUCTION

In many environmental applications, extreme events are the main aspects of practical concern. Financial time series are increasingly being analyzed to assess the risk from extreme events. A description of extreme events is usually based on observations that exceed a high threshold. Serial dependence leads to large values occurring close in time and thus forming clusters. Clustering of extremes does not take place in an independent and identically distributed (i.i.d.) setting.

Consider $\{X_n\}_{n\geq 1}$ a stationary sequence with common distribution function (df) F and $\{Y_n\}_{n\geq 1}$ an i.i.d. sequence with the same parent df F. We say $\{X_n\}_{n\geq 1}$ has extremal index θ $(0 < \theta \leq 1)$ if, for all $\tau > 0$, there is a sequence of levels $u_n \equiv u_n(\tau)$, $n \geq 1$, such that

$$P\big(\max(Y_1,\ldots,Y_n) \leq u_n\big) = F^n(u_n) \underset{n\to\infty}{\longrightarrow} e^{-\tau}$$

$$\text{and} \quad P\big(\max(X_1,\ldots,X_n) \leq u_n\big) \underset{n\to\infty}{\longrightarrow} e^{-\theta\tau} \tag{1.1}$$

(Leadbetter *et al.* [20] 1983). The sequence $u_n \equiv u_n(\tau)$, $n \geq 1$, satisfying $F^n(u_n) \to \exp(-\tau)$ or, equivalently, $n(1-F(u_n)) \to \tau$, as $n \to \infty$, is usually denoted as normalized levels.

There are several characterizations of the extremal index bringing out different estimators. Many of these estimators can be stated as functions of a number k of upper order statistics. Analogous to the semiparametric estimation of various tail measures (e.g., the tail index and tail dependence coefficients in a multivariate framework), there is a proverbial tradeoff between bias and variance. The first increases with k (large bias for a large amount of top order statistics used in estimates) and the second increases as k gets smaller (large variability as fewer top order statistics are considered). A typical path is plotted in Figure 1. After the great variability in the beginning, there is a stable sample path, as function of k, around the true value and then the bias starts to stand out and dominate.

Thus k needs to be chosen from the stability zone that mediates the variance domain and the bias domain. There are several methods developed in literature towards this choice of k concerning the estimation of tail measures. A survey within the tail index estimation can be seen in Beirlant *et al.* ([2] 2004). More recently, a general procedure was introduced in Gomes *et al.* ([14] 2013) for the tail index estimation, which was latter adopted in Neves *et al.* ([23] 2015) to estimate the extremal index. This consists of a pure heuristic procedure to find the "plateau" region of the estimates path from which we may infer the true value of the parameter. The methodology in Frahm *et al.* ([12] 2005), developed within the estimation of the tail dependence coefficient of random pairs, also seeks a stable region but after a smoothing of the sample path; see Frahm *et al.* ([12]

2005) and Ferreira and Silva ([9] 2014). In Ferreira ([5] 2014) and Ferreira ([6] 2015a) it was also adapted to the tail index estimation.

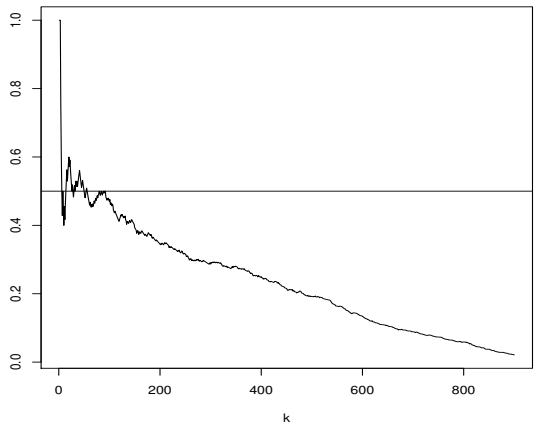


Figure 1: Runs estimates sample path of a moving maximum process, $X_i = \max(a_0 Z_i, a_1 Z_{i-1}, a_2 Z_{i-2})$, $i \geq 1$, where $\{Z_i\}_{i \geq -1}$ is an i.i.d. unit Fréchet sequence, $a_0 = 1/3$, $a_1 = 1/6$ and $a_2 = 1/2$, with run length $r = 2$. The horizontal line corresponds to the true value.

Here we are going to apply the methodology of Frahm *et al.* ([12] 2005) to several estimators of the extremal index. For comparison, we also analyze the performance of the procedure in Neves *et al.* ([23] 2015) applied to those estimators. As these are threshold-free methods, we also compare with the blocks and sliding estimation threshold-free procedure presented in Robert *et al.* ([25] 2009). The description of the methods is addressed in Section 2. The comparison of the procedures is assessed through simulation in Section 3 and an illustration with real data is stated in Section 4. A small discussion is presented in Section 5.

2. ESTIMATION METHODS

The extremal index can be interpreted in different ways, leading to different estimators. In O'Brien ([24] 1974) it is proved that

$$P\big(\max(X_2, ..., X_{r_n} \leq u_n | X_1 > u_n)\big) \underset{n \to \infty}{\longrightarrow} \theta\,, \tag{2.1}$$

where r_n is such that $r_n \to \infty$ and $r_n = o(n)$. Under a mild mixing condition, Hsing *et al.* ([17], 1988) stated that

$$E\big[\textstyle\sum_{i=1}^{r_n} \mathbf{1}_{\{X_j > u_n\}} | \sum_{i=1}^{r_n} \mathbf{1}_{\{X_j > u_n\}} \geq 1\big] \to \theta^{-1}, \tag{2.2}$$

with $\mathbf{1}(\cdot)$ denoting the indicator function, i.e., the limiting mean number of exceedances of u_n in an interval of length r_n corresponds to the arithmetic inverse of the extremal index, given that there are exceedances.

Also under a slight mixing condition, Ferro and Segers ([11] 2003) show that

$$P\big(\overline{F}(u_n)\,T(u_n) > t\big) \underset{n\to\infty}{\longrightarrow} \theta e^{-\theta t}, \qquad t > 0, \tag{2.3}$$

where $T(u_n) = \min\{n \geq 1\colon X_{n+1} > u_n | X_1 > u_n\}$, i.e., the process of inter-exceedance times normalized by exceedances of u_n follows a mixture of a point mass and an exponential distribution $Exp(\theta^{-1})$.

Relations (2.1)–(2.3) yield the most common approaches to estimate θ, respectively, the runs, the blocks and the intervals method.

The blocks and the runs estimators are based on their own clusters identification procedure and both correspond to the ratio between the number of clusters and the number of exceedances of a high threshold u_n (Hsing [15] 1991; Weissman and Novak [29] 1998; Nandagopalan [22] 1990; Hsing [16] 1993). The intervals estimator is based on an inter-exceedance times method (Ferro and Segers [11] 2003).

More precisely, the runs estimator is expressed as

$$\widehat{\theta}^{\mathtt{R}} \;=\; (N_n(u_n))^{-1} \sum_{i=1}^{n-r} \mathbf{1}_{\{X_i > u_n\}}\, \mathbf{1}_{\{X_{i+1} \leq u_n\}} \cdots \mathbf{1}_{\{X_{i+r} \leq u_n\}}, \tag{2.4}$$

where $N_n(u_n)$ is the number of exceedances of u_n. Independent clusters are identified as runs of observations above u_n, separated by r_n consecutive values under u_n.

By considering $b_n = [n/r_n]$ blocks of length r_n ($[\cdot]$ means the integer part), the simple blocks estimator corresponds to

$$\widehat{\theta}^{\mathtt{B}} = \frac{C_n(u_n)}{N_n(u_n)} \tag{2.5}$$

where $C_n(u)$ is the number of clusters, i.e, in this context it corresponds to the number of blocks in which at least one exceedance of u_n occurs. The variant

$$\widehat{\theta}^{\mathtt{BL}} = \frac{\log(1 - C_n(u_n)/k_n)}{r_n \log(1 - N_n(u_n)/n)} \tag{2.6}$$

has been proposed in Smith and Weissman ([27] 1994) as having a better asymptotic behavior of second order.

After some considerations and based on the result in (2.3), the intervals estimator is stated as

$$\widehat{\theta}^{\mathtt{I}} = \begin{cases} 1 \wedge \frac{2\left(\sum_{i=1}^{N-1} T_i\right)^2}{(N-1)\sum_{i=1}^{N-1} T_i^2} & , \text{if } \max\{T_i : 1 \le i \le N-1\} \le 2 \\ 1 \wedge \frac{2\left(\sum_{i=1}^{N-1} (T_i-1)\right)^2}{(N-1)\sum_{i=1}^{N-1} (T_i-1)(T_i-2)} & , \text{if } \max\{T_i : 1 \le i \le N-1\} > 2, \end{cases} \tag{2.7}$$

where T_i denotes the i-th inter-exceedance time, $i = 1, \ldots, N-1$ and $N \equiv N_n(u_n)$.

The analysis of convenient local dependence conditions may eliminate the need for a cluster identification scheme, such as the local dependence condition $\mathrm{D}^{(m)}(u_n)$ of Chernick *et al.* ([4] 1991), with m some positive integer. Consider notation $M_{i,j} = \max\{X_{i+1}, \ldots, X_j\}$, for $i < j$, $M_{i,j} = -\infty$ if $i \ge j$ and $M_{0,j} = M_j$. Under condition $\mathrm{D}(u_n)$ of Leadbetter ([19] 1974) which holds whenever $\alpha_{n,\,l_n} \to 0$, as $n \to \infty$, for $l_n = o(n)$, where

$$\begin{aligned} \alpha_{n,\,l} \;=\; \sup\Big\{ \Big| & P\big(M_{i_1,i_1+p} \le u_n, M_{j_1,j_1+q} \le u_n\big) \\ & - \; P(M_{i_1,i_1+p} \le u_n)\, P(M_{j_1,j_1+q} \le u_n) \Big| : \\ & 1 \le i_1 < i_1 + p + l \le j_1 < j_1 + q \le n \Big\}, \end{aligned}$$

we say that $\mathrm{D}^{(m)}(u_n)$ is satisfied by $\{X_n\}_{n\ge1}$ if, for some $\{b_n\}_{n\ge1}$ such that, as $n \to \infty$,

$$b_n \to \infty\,, \qquad b_n \alpha_{n,\,l_n} \to 0\,, \qquad b_n l_n / n \to 0,$$

we have

$$n P\big(X_1 > u_n, M_{1,m} \le u_n < M_{m,r_n}\big) \to 0\,, \qquad n \to \infty,$$

with $\{r_n = [n/b_n]\}_{n\ge1}$. The stronger conditions

$$n \sum_{j=m+1}^{r_n} P\big(X_1 > u_n, M_{1,m} \le u_n < X_j\big) \to 0\,, \qquad n \to \infty,$$

also stated in Chernick *et al.* ([4] 1991), lead to $\mathrm{D}'(u_n)$ if $m = 1$ and $\mathrm{D}''(u_n)$ if $m = 2$, considered in Leadbetter *et al.* ([20] 1983) and Leadbetter and Nandagopalan ([21] 1989), respectively. Condition $\mathrm{D}'(u_n)$ inhibits clustering of exceedances and thus resembles an i.i.d. behavior and brings out $\theta = 1$, whilst $\mathrm{D}''(u_n)$ allows clustering but inhibits the occurrence of two or more upcrossings. Moreover, if condition $\mathrm{D}^{(m_0)}(u_n)$ holds then $\mathrm{D}^{(m)}(u_n)$ also holds for all $m \ge m_0$.

Ferreira and Ferreira ([10] 2015) stated a new estimator that works under $\mathrm{D}^{(m)}(u_n)$. More precisely, if $\{X_n\}_{n\ge1}$ satisfies condition $\mathrm{D}^{(m)}(u_n)$, we can estimate θ by

$$\widehat{\theta}^{\mathtt{FF}} = \frac{U_n^Z(u_n)}{N_n(u_n)} \tag{2.8}$$

where $U_n^Z(u_n)$ is the number of upcrossings of u_n within $\{Z_1, \ldots, Z_{[n/(m-1)]}\}$ with $Z_n = M_{(n-1)(m-1),n(m-1)}$, $n \geq 1$. Other estimators developed in the same context were also considered in that work whose overall performance did not surpass $\widehat{\theta}^{\mathtt{FF}}$. Estimation approaches working only for series that satisfy condition $\mathrm{D}^{(2)}(u_n)$ can also be seen in Süveges ([28] 2007), Ferreira and Ferreira ([8] 2012) and Ferreira ([7] 2015b).

Chernick *et al.* ([4] 1991) also show that, under $\mathrm{D}^{(m)}(u_n)$, the extremal index exists and can be computed by the limit

$$\theta = \lim_{n\to\infty} P\big(M_{1,m} \leq u_n | X_1 > u_n\big). \tag{2.9}$$

Observe that the runs estimator in (2.4) corresponds to the empirical counterpart of (2.9) by considering $r = m$. Diagnostic tools to analyze condition $\mathrm{D}^{(m)}(u_n)$ may be seen in Süveges ([28] 2007) and Ferreira and Ferreira ([10] 2015).

Observe also that taking $r = 2$ in (2.4) corresponds to the Nandagopalan's runs estimator derived in Nandagopalan ([22] 1990) under $\mathrm{D}''(u_n)$.

The disjoint blocks and the sliding blocks estimators presented in Robert *et al.* ([25] 2009) are derived from the extremal index definition in (1.1).

Consider, for r positive integer,

$$F_r(u) := P(M_r \leq u), \quad \tau_r(u) := r(1 - F(u)) \quad \text{and} \quad \theta_r(u) := -\frac{\log F_r(u)}{\tau_r(u)}\,.$$

We have $\theta = \lim_{r\to\infty} \theta_r(u_r)$ for normalized levels $u_r = u_r(\tau)$ according to definition in (1.1). The estimation of the block maxima df F_r through $b = [n/r]$ disjoint blocks or $n - r + 1$ sliding blocks, that is

$$\widehat{F}_{n,r}^{\mathtt{DJ}}(u) := \frac{1}{b}\sum_{i=1}^{b} \mathbb{1}_{\{M_{(i-1)r,ir} \leq u\}} \quad \text{and} \quad \widehat{F}_{n,r}^{\mathtt{SL}}(u) := \frac{1}{n-r+1}\sum_{i=1}^{n-r+1} \mathbb{1}_{\{M_{i-1,i-1+r} \leq u\}}$$

originates the estimators, respectively,

$$\widehat{\theta}^{\mathtt{DJ}} = -\frac{\log \widehat{F}_{n,r}^{\mathtt{DJ}}(u_n)}{\widehat{\tau}_{n,r}(u_n)} \quad \text{and} \quad \widehat{\theta}^{\mathtt{SL}} = -\frac{\log \widehat{F}_{n,r}^{\mathtt{SL}}(u_n)}{\widehat{\tau}_{n,r}(u_n)}\,, \tag{2.10}$$

with

$$\widehat{\tau}_{n,r}(u_n) = \frac{rN_n(u_n)}{n}\,.$$

In order to achieve consistency in the estimators above, τ must be actually taken as an intermediate sequence τ_n, $n \geq 1$, that is,

$$\tau_n \to \infty \quad \text{and} \quad \tau_n/n \to 0.$$

Gomes *et al.* ([13] 2008) and Neves *et al.* ([23] 2015) considered the levels u_n in the interval between the $k+1$ and the k-th upper order statistics,

$[X_{n-k:n}, X_{n-k+1:n})$, for the Nandagopalan's runs estimators. The advantage is to move the framework to a similar context of the semiparametric estimation of other important tail measures existing in the literature which allows new estimation methods of the extremal index by adapting the existing ones. In this paper we will consider those levels $u_n \in [X_{n-k:n}, X_{n-k+1:n})$ in the estimators (2.4)–(2.8) and (2.10) and denote, respectively,

$$\widehat{\theta}_k^{\mathtt{R}},\ \widehat{\theta}_k^{\mathtt{B}},\ \widehat{\theta}_k^{\mathtt{BL}},\ \widehat{\theta}_k^{\mathtt{I}},\ \widehat{\theta}_k^{\mathtt{FF}},\ \widehat{\theta}_k^{\mathtt{DJ}} \text{ and } \widehat{\theta}_k^{\mathtt{SL}}. \tag{2.11}$$

Observe that we are replacing τ by k. Indeed, we consider that $k \equiv k_n$, $n \geq 1$, is replacing τ_n and thus it is also an intermediate sequence on behalf of consistency. Estimators in (2.11) are functions of k, the number of order statistics higher than the chosen level, where an increasing/decreasing k increases the bias/variance (see Figure 1). Thus the choice of k is central in the estimation, not only of the extremal index, but also of many other tail measures, making this topic largely addressed in literature (see, e.g., Beirlant *et al.* [2] 2004).

The "plateau-finding" algorithm of Frahm *et al.* ([12] 2005), applied to the estimation of the tail dependence coefficient of random pairs and here adopted to estimate the extremal index, is based on a smoothing of the estimator's sample path by a simple box kernel with integer bandwidth $d > 0$. The resulting trajectory thus corresponds to the moving average of $2d + 1$ successive points of the initial one and will be used in the rest of the procedure that consists on the application of a plateau definition and respective finding criterium. In the following we detail the method which we denote Algorithm 1.

Algorithm 1:

For a sample $(X_1, ..., X_n)$, consider bandwidth $d = [wn] \in \mathbb{N}$ and compute the means of $2d + 1$ successive points of $\widehat{\theta}_k$, $1 \leq k < n$, with smoothing degree $w = 0.005$ (thus each moving average is about 1% of the data, as suggested in Frahm *et al.* [12] 2005). In the resulting smoothed values, $\overline{\widehat{\theta}}_1, \ldots, \overline{\widehat{\theta}}_{n-2d}$, define the plateaus $p_k = (\overline{\widehat{\theta}}_k, \ldots, \overline{\widehat{\theta}}_{k+m-1})$, $k = 1, \ldots, n - 2d - m + 1$, with length $m = [\sqrt{n - 2d}]$. The algorithm stops at the first plateau satisfying

$$\sum_{i=k+1}^{k+m-1} \left| \overline{\widehat{\theta}}_i - \overline{\widehat{\theta}}_k \right| \leq 2s,$$

where s is the empirical standard deviation of $\overline{\widehat{\theta}}_1, \ldots, \overline{\widehat{\theta}}_{n-2d}$. Estimate θ as the mean of the values of the chosen plane region (consider the estimate zero if no stable region fulfills the stopping condition).

For comparison, we also consider another heuristic procedure introduced in Gomes *et al.* ([14] 2013), also seeking the plane region that presumably includes

the "optimal" sample fraction k to be estimated. The algorithm is described below and denoted Algorithm 2.

Algorithms 2 and 3:

For a sample $(X_1, ..., X_n)$, obtain the minimum value j_0, such that the rounded values to j decimal places of $\widehat{\theta}_k$, $1 \leq k < n$, denoted $\widehat{\theta}_k(j)$ are not all equal. Identify the set of values of k associated to equal consecutive values of $\widehat{\theta}_k(j_0)$. Consider the set with largest range $\ell := k_{\texttt{max}} - k_{\texttt{min}}$. Take all the estimates $\widehat{\theta}_k(j_0 + 2)$ with $k_{\texttt{max}} \leq k \leq k_{\texttt{min}}$, i.e., the estimates with two more decimal points and obtain the mode. Denote $\mathcal{K}$ the set of k-values associated with this mode. Consider $\widehat{\theta}_{\widehat{k}}$, where $\widehat{k}$ is the maximum of $\mathcal{K}$.

We also consider the variant $\widehat{\theta}_{\widetilde{k}}$ by taking $\widetilde{k} = \ell$ as mentioned in Neves *et al.* ([23] 2015). This will be denoted Algorithm 3.

Observe that the described methodologies are all threshold-free. Robert *et al.* ([25] 2009) also presented a threshold-free procedure based on blocks and sliding estimators defined in (2.10). It is described downwards and will be called Algorithm 4:

Algorithm 4:

For a sample $(X_1, ..., X_n)$, choose a block size r, take $b = [n/r]$, $\tau = 1$ and $u = X_{n-[b\tau]+1:n}$. Consider $N_{a,b}(u) := \sum_{a<i\leq b} \mathbf{1}_{\{X_i>u\}}$, $\overline{N}_{n,r}(u) := (1/(n-r+1)) \times \sum_{i=0}^{n-r} N_{i,i+r}(u)$, $\widehat{\sigma}^2_{n,r}(u) := \sum_{i=0}^{n-r}(N_{i,i+r}(u) - \overline{N}_{n,r}(u))^2$ and $\widehat{c}^2_{n,r}(u) := \frac{\widehat{\theta}}{\widehat{\tau}_{n,r}(u)} \times \widehat{\sigma}^2_{n,r}(u) - 1$. Calculate $\widehat{\theta} = \widehat{\theta}^{\Gamma}_{[b\tau]}$, $\Gamma \in \{\texttt{SL}, \texttt{DJ}\}$ and $\widehat{c} = \widehat{c}_{n,r}(u)$. Obtain $\widehat{\mu}$ through

$$\mu = \begin{cases} \mu_{\texttt{SL}} := \theta\alpha^{-2}(e^{\alpha} - 1 - \alpha) + \alpha^{-1}\theta c^2 \\ \mu_{\texttt{DJ}} := \theta(2\alpha)^{-1}(e^{\alpha} - 1) + \alpha^{-1}\theta c^2, \end{cases}$$

replacing θ, c and α by, respectively, $\widehat{\theta}$, $\widehat{c}$ and $\widehat{\theta}\tau$. Obtain the bias-corrected $\widehat{\theta} - \widehat{\mu}/b$ and estimate the variance by evaluating $v = 2(\theta^2/\alpha^3)(e^{\alpha} - 1 - \alpha - \alpha^2/2) + \theta^2 c^2/\alpha$ at $\theta = \widehat{\theta}$, $c = \widehat{c}$ and $\alpha = \widehat{\theta}\tau$. Take $\widehat{\alpha}$ as the value that minimizes v when $\theta = \widehat{\theta}$ and $c = \widehat{c}$. Now repeat the procedure for the founded optimal value $\tau = \widehat{\alpha}/\widehat{\theta}$.

In the sequel, we use the abbreviations A1, A2, A3 and A4, respectively, to refer the algorithms above.

3. SIMULATION STUDY

We are going to analyze through simulation the performance of the estimators in (2.11) within the methodologies A1–A4 described above. This study is based on the following models:

- Max-autoregressive process (MAR), $X_i = \alpha X_{i-1} \vee \epsilon_i$, where $0 < \alpha < 1$ and $\{\epsilon_i\}_{i\geq 1}$ is an i.i.d. sequence of r.v.'s with d.f. $F_\epsilon(x) = \exp(-(1-\alpha)/x)$, $x > 0$. This process has $\theta = 1 - \alpha$. We consider $\alpha = 1/2$ and hence $\theta = 1/2$.
- Moving maxima process (MM), $X_i = \bigvee_{j=0,\ldots,m} \alpha_j \epsilon_{i-j}$, with $\sum_{j=0}^{m} \alpha_j = 1$ and $\alpha_j \geq 0$, $\{\epsilon_i\}_{i\geq 1}$ is an i.i.d. sequence of unit Féchet distributed r.v.'s. This process has $\theta = \bigvee_{j=0,\ldots,m} \alpha_j$. We consider $m = 3$, $\alpha_0 = 1/3$, $\alpha_1 = 1/6$, $\alpha_2 = 1/2$ leading to $\theta = 1/2$.
- Autoregressive Gaussian process (AR), $X_i = \alpha X_{i-1} + \epsilon_i$, where $\{\epsilon_i\}_{i\geq 1}$ is an i.i.d. sequence of $\mathrm{N}(0, 1-\alpha^2)$ distributed r.v.'s. This process satisfies condition $\mathrm{D}'(u_n)$ and thus $\theta = 1$ (Leadbetter *et al.* [20] 1983).
- A first order autoregressive process, with Cauchy marginals (ARCauchy) of Chernick ([3] 1978), $X_i = sX_{i-1} + \epsilon_i$, with $|s| < 1$. The extremal index is given by $1 - s^2$. We take $s = -3/5$ and thus $\theta = 0.64$.
- A negatively correlated uniform autoregressive process (ARUnif) of Chernick *et al.* ([4], 1991), $X_i = -(1/s)X_{i-1} + \epsilon_i$, where $\{\epsilon_i\}_{i\geq 1}$ is an i.i.d. sequence such that $P(\epsilon_1 = j/s) = 1/s$ for $j = 1, \ldots, s$. We have $\theta = 1 - 1/s^2$. Here we consider $s = 2$ and thus $\theta = 3/4$.
- Bivariate extreme value Markov process with standard Gumbel marginals and logistic dependence function, i.e.,

$$P\big(X_i \leq x, X_{i+1} \leq y\big) \;=\; \exp\big(-(x^{1/\alpha} + y^{1/\alpha})^{\alpha}\big).$$

  We consider the dependence parameter $\alpha = 0.5$ which gives $\theta = 0.328$ (Smith [26] 1992), and denote the process MCBEV.
- A GARCH(1,1) process, $X_i = \sigma_i \epsilon_i$, with $\sigma_i^2 = \alpha + \lambda X_{i-1}^2 + \beta \sigma_{i-1}^2$, $\alpha, \lambda, \beta > 0$, where $\{\epsilon_i\}_{i\geq 1}$ is an i.i.d. sequence of standard Gaussian r.v.'s. We consider $\alpha = 10^{-6}$, $\lambda = 1/4$ and $\beta = 7/10$ resulting in $\theta = 0.447$ (see details in Laurini and Tawn, [18] 2012).

We consider samples of sizes $n = 100$, 1000, 5000 and generate 100 independent replications of each and for each model. We compare the estimation procedures by computing the absolute mean bias and the root mean square error (rmse).

Remark 3.1. Observe that the methods being compared avoid threshold selection but need a cluster identification parameter, whether be it a block size or a run length. Recall that the dependence condition $\mathrm{D}^{(m)}(u_n)$ of Chernick *et al.* ([4], 1991) is a diagnostic tool for cluster identification within the runs estimator $\widehat{\theta}_k^{\mathtt{R}}$ and estimator $\widehat{\theta}_k^{\mathtt{FF}}$ defined in (2.8). More precisely, we take the run length r equal to m in the first (see discussion concerning (2.9)) and cycles of size $m-1$ in the second as stated in (2.8). The MAR process satisfies condition $\mathrm{D}^{(2)}(u_n)$, whilst the processes MM, ARCauchy and ARUnif satisfy condition $\mathrm{D}^{(3)}(u_n)$. See Ferreira and Ferreira ([10] 2015) and references therein for more details. In this latter reference, we validated conditions $\mathrm{D}^{(4)}(u_n)$ and $\mathrm{D}^{(5)}(u_n)$ for the processes MCBEV and GARCH, respectively. In what concerns the remaining estimators which are based on blocks schemes, the respective cluster parameters were chosen according to an overall good performance found on further simulations.

The results are presented in Tables 1–6 (the bold numbers correspond to the smallest estimates obtained in each model). A high bias is observed in the AR model and also in models ARCauchy, ARUnif and GARCH concerning the runs, the intervals and the FF estimator, under algorithms A2 and A3. The lowest values of rmse rely frequently on blocks $\widehat{\theta}_k^{\mathtt{B}}$ and $\widehat{\theta}_k^{\mathtt{BL}}$ estimators under algorithms A1, A2 and A3, followed by estimators $\widehat{\theta}_k^{\mathtt{R}}$ and $\widehat{\theta}_k^{\mathtt{FF}}$ within algorithm A1. In the AR process, the results differ from the others where the estimators $\widehat{\theta}_k^{\mathtt{DJ}}$ and $\widehat{\theta}_k^{\mathtt{SL}}$ tend to behave better over the four algorithms. Observe that in this case we have the boundary value $\theta = 1$ (as in i.i.d. sequences) where inference is usually problematic (see Ancona-Navarrete and Tawn [1] 2000). The intervals estimator, $\widehat{\theta}_k^{\mathtt{I}}$, is parameter-free under the methods in study and may be considered within MAR, AR, MM and MCBEV models. The worst performances concern mainly estimators $\widehat{\theta}_k^{\mathtt{FF}}$ and $\widehat{\theta}_k^{\mathtt{R}}$ for algorithm A2, where the method is returning a too high k, corresponding to estimates with very large bias. In Gomes *et al.* ([13] 2008) it was presented a reduced-bias version of Nandagopalan's estimator based on the Generalized Jackknife (GJ) methodology, which is given by

$$\widehat{\theta}_k^{\mathtt{NGJ}} = 5\widehat{\theta}^{\mathtt{R}}_{[k/2]+1} - 2\left(\widehat{\theta}^{\mathtt{R}}_{[k/4]+1} + \widehat{\theta}_k^{\mathtt{R}}\right). \tag{3.1}$$

Notice that Nandagopalan's estimator corresponds to the runs estimator whenever we take the run length 2, which in turn requires $\mathrm{D}''(u_n)$. In our examples, only models MAR and AR satisfy this condition. We have also applied the estimator (3.1) to all models within algorithms A2 and A3. Indeed, except in the GARCH case, the rmse of $\widehat{\theta}_k^{\mathtt{NGJ}}$ decreases to about the half of the rmse of the runs estimator, mostly for larger sample sizes ($n \geq 1000$) and with algorithm A2. In the case of algorithm A3, the rmse of $\widehat{\theta}_k^{\mathtt{NGJ}}$ is smaller than the runs estimator only within the largest sample size ($n = 5000$) of models MAR and ARUnif.

Table 1: Root mean squared errors obtained for simulated samples of size $n = 100$. For estimators $\widehat{\theta}_k^{\mathtt{BL}}$ and $\widehat{\theta}_k^{\mathtt{B}}$ we considered blocks of length 3 except in MCBEV and GARCH models where we used blocks of length 4 and 5, respectively. For estimators $\widehat{\theta}_k^{\mathtt{DJ}}$ and $\widehat{\theta}_k^{\mathtt{SL}}$ we considered blocks of length 5. For estimator $\widehat{\theta}_k^{\mathtt{R}}$ ($\widehat{\theta}_k^{\mathtt{FF}}$) we considered runs (cycles) of length 2 in MAR and AR, of length 3 in MM, ARCauchy and ARUnif, of length 4 in MCBEV and length 5 in GARCH. See Remark 3.1.

A1	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.1338	0.3603	0.1215	0.1534	0.1528	0.1249	0.1782
$\widehat{\theta}_k^{\mathtt{I}}$	0.2754	0.2899	0.2271	0.3549	0.2487	0.3071	0.5170
$\widehat{\theta}_k^{\mathtt{FF}}$	0.1469	0.5336	0.1488	0.1831	0.1474	0.1352	0.2018
$\widehat{\theta}_k^{\mathtt{BL}}$	0.1467	0.4548	0.1780	0.2869	0.1820	0.1544	0.1626
$\widehat{\theta}_k^{\mathtt{B}}$	0.1233	0.4826	0.1594	0.1948	0.1459	0.1380	0.1539
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.2971	0.5041	0.2803	0.3296	0.3406	0.2909	0.5316
$\widehat{\theta}_k^{\mathtt{SL}}$	0.2583	**0.2327**	0.2474	0.2907	0.3419	0.3052	0.5185

A2	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.3183	0.7928	0.4133	0.6283	0.7399	0.2663	0.4366
$\widehat{\theta}_k^{\mathtt{NGJ}}$	0.3003	0.5448	0.5071	0.3925	0.2500	0.3404	0.5569
$\widehat{\theta}_k^{\mathtt{I}}$	0.2621	0.5387	0.3032	0.2500	0.2500	0.2133	0.5511
$\widehat{\theta}_k^{\mathtt{FF}}$	0.4353	0.9410	0.4627	0.6400	0.7500	0.4447	0.2910
$\widehat{\theta}_k^{\mathtt{BL}}$	0.0996	0.5217	0.0948	0.2176	**0.1079**	0.0734	0.1793
$\widehat{\theta}_k^{\mathtt{B}}$	0.1216	0.6186	0.1162	0.2118	0.3018	**0.0495**	0.2108
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.2646	0.4831	0.2146	0.2801	0.3010	0.2405	0.3230
$\widehat{\theta}_k^{\mathtt{SL}}$	0.3035	0.4431	0.2110	0.3158	0.4289	0.2457	0.4511

A3	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.1199	0.4192	0.1746	0.3026	0.2923	0.1158	0.3151
$\widehat{\theta}_k^{\mathtt{NGJ}}$	0.6235	0.5285	0.5323	0.3673	0.2505	0.6538	0.6584
$\widehat{\theta}_k^{\mathtt{I}}$	0.2770	0.2846	0.2527	0.2500	0.2500	0.3095	0.4959
$\widehat{\theta}_k^{\mathtt{FF}}$	0.1801	0.6620	0.2295	0.4697	0.4486	0.3133	0.1766
$\widehat{\theta}_k^{\mathtt{BL}}$	0.1663	0.4521	0.1890	0.2587	0.2051	0.1517	0.1580
$\widehat{\theta}_k^{\mathtt{B}}$	**0.0967**	0.5245	**0.0775**	**0.1105**	0.1179	0.0570	**0.1439**
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.4813	0.2996	0.4017	0.4394	0.3769	0.4782	0.5875
$\widehat{\theta}_k^{\mathtt{SL}}$	0.4802	0.4092	0.4041	0.4609	0.4946	0.5056	0.6435

A4	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.3001	0.2423	0.2748	0.2888	0.3457	0.3496	0.4886
$\widehat{\theta}_k^{\mathtt{SL}}$	0.2482	0.2445	0.2302	0.2644	0.3279	0.3562	0.4811

Table 2: Absolute bias obtained for simulated samples of size $n = 100$. For estimators $\widehat{\theta}_k^{\mathtt{BL}}$ and $\widehat{\theta}_k^{\mathtt{B}}$ we considered blocks of length 3 except in MCBEV and GARCH models where we used blocks of length 4 and 5, respectively. For estimators $\widehat{\theta}_k^{\mathtt{DJ}}$ and $\widehat{\theta}_k^{\mathtt{SL}}$ we considered blocks of length 5. For estimator $\widehat{\theta}_k^{\mathtt{R}}$ ($\widehat{\theta}_k^{\mathtt{FF}}$) we considered runs (cycles) of length 2 in MAR and AR, of length 3 in MM, ARCauchy and ARUnif, of length 4 in MCBEV and length 5 in GARCH. See Remark 3.1.

A1	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.0152	0.3280	0.0833	0.0618	0.0569	0.0076	**0.0303**
$\widehat{\theta}_k^{\mathtt{I}}$	0.1577	0.2040	0.1234	0.3545	0.2487	0.1875	0.5006
$\widehat{\theta}_k^{\mathtt{FF}}$	0.0780	0.5172	0.1243	0.1276	**0.0003**	0.1070	0.1841
$\widehat{\theta}_k^{\mathtt{BL}}$	0.0063	0.4319	0.0709	0.0307	0.0321	0.0439	0.0771
$\widehat{\theta}_k^{\mathtt{B}}$	0.0120	0.4657	0.0432	0.0350	0.0756	0.0265	0.1071
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0454	0.3721	0.1328	0.2646	0.2934	0.2091	0.4143
$\widehat{\theta}_k^{\mathtt{SL}}$	0.2050	0.1729	0.2186	0.2519	0.2993	0.2797	0.4928

A2	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.2971	0.7778	0.4045	0.6283	0.7399	0.2586	0.4366
$\widehat{\theta}_k^{\mathtt{NGJ}}$	0.0057	0.4671	0.3341	0.3714	0.2500	0.0889	0.5222
$\widehat{\theta}_k^{\mathtt{I}}$	**0.0031**	0.4575	0.0879	0.2500	0.2500	0.0017	0.5505
$\widehat{\theta}_k^{\mathtt{FF}}$	0.4294	0.9393	0.4599	0.6400	0.7500	0.4445	0.2861
$\widehat{\theta}_k^{\mathtt{BL}}$	0.0371	0.5127	**0.0113**	0.1278	0.0389	0.0097	0.1673
$\widehat{\theta}_k^{\mathtt{B}}$	0.1135	0.6172	0.1051	0.1628	0.2906	0.0311	0.2047
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0107	0.3768	0.0187	0.0198	0.1110	0.0554	0.1470
$\widehat{\theta}_k^{\mathtt{SL}}$	0.1030	0.2913	0.0594	0.0748	0.1952	0.1127	0.2979

A3	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.0488	0.4056	0.1598	0.2875	0.2800	0.0937	0.3110
$\widehat{\theta}_k^{\mathtt{NGJ}}$	0.0057	0.4671	0.3341	0.3714	0.2500	0.0889	0.5222
$\widehat{\theta}_k^{\mathtt{I}}$	0.1607	0.2045	0.1672	0.2500	0.2500	0.2044	0.4792
$\widehat{\theta}_k^{\mathtt{FF}}$	0.1593	0.6558	0.2182	0.4652	0.4444	0.3097	0.1711
$\widehat{\theta}_k^{\mathtt{BL}}$	0.0275	0.4163	0.0766	0.0566	0.0702	0.0565	0.0561
$\widehat{\theta}_k^{\mathtt{B}}$	0.0547	0.5184	0.0120	**0.0179**	0.1023	**0.0025**	0.1280
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.3553	**0.0206**	0.2558	0.3441	0.3146	0.3330	0.4882
$\widehat{\theta}_k^{\mathtt{SL}}$	0.2852	0.0424	0.2375	0.2385	0.2952	0.3101	0.4835

A4	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.2433	0.1865	0.2096	0.2462	0.3059	0.3099	0.4571
$\widehat{\theta}_k^{\mathtt{SL}}$	0.2110	0.2031	0.1921	0.2238	0.2821	0.3219	0.4544

Table 3: Root mean squared errors obtained for simulated samples of size $n = 1000$. For estimators $\widehat{\theta}_k^{\mathtt{BL}}$ and $\widehat{\theta}_k^{\mathtt{B}}$ we considered blocks of length 3 except in MCBEV and GARCH models where we used blocks of length 4 and 5, respectively. For estimators $\widehat{\theta}_k^{\mathtt{DJ}}$ and $\widehat{\theta}_k^{\mathtt{SL}}$ we considered blocks of length 20. For estimator $\widehat{\theta}_k^{\mathtt{R}}$ ($\widehat{\theta}_k^{\mathtt{FF}}$) we considered runs (cycles) of length 2 in MAR and AR, of length 3 in MM, ARCauchy and ARUnif, of length 4 in MCBEV and length 5 in GARCH. See Remark 3.1.

A1	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.0666	0.2877	0.0645	0.0874	0.0583	0.0508	0.1059
$\widehat{\theta}_k^{\mathtt{I}}$	0.0795	0.3243	0.0658	0.2312	0.2498	0.0670	0.2217
$\widehat{\theta}_k^{\mathtt{FF}}$	0.0853	0.4148	0.0744	0.1014	0.0566	0.0809	0.1295
$\widehat{\theta}_k^{\mathtt{BL}}$	0.0467	0.4381	0.0767	0.0882	0.0944	0.0836	0.0845
$\widehat{\theta}_k^{\mathtt{B}}$	0.0532	0.4499	0.0680	0.0624	0.1022	0.0675	0.0993
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.2161	0.4323	0.2225	0.1859	0.3059	0.1351	0.2101
$\widehat{\theta}_k^{\mathtt{SL}}$	0.1342	0.3379	0.1101	0.1303	0.1563	0.1057	0.1877

A2	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.1915	0.7503	0.4357	0.6388	0.7489	0.2656	0.4457
$\widehat{\theta}_k^{\mathtt{NGJ}}$	0.3003	0.5448	0.5071	0.3925	0.2500	0.3404	0.5569
$\widehat{\theta}_k^{\mathtt{I}}$	0.1227	0.5738	0.0688	0.3600	0.2500	0.0565	0.5202
$\widehat{\theta}_k^{\mathtt{FF}}$	0.4435	0.9582	0.4794	0.6397	0.7500	0.2768	0.4458
$\widehat{\theta}_k^{\mathtt{BL}}$	**0.0283**	0.4765	0.0647	0.2051	0.0631	0.0680	**0.0615**
$\widehat{\theta}_k^{\mathtt{B}}$	0.0926	0.5931	0.0803	0.1206	0.2387	**0.0188**	0.1398
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0932	0.3744	0.3845	0.1216	0.2352	0.1007	0.2247
$\widehat{\theta}_k^{\mathtt{SL}}$	0.1211	0.3330	0.1085	0.1636	0.2534	0.1402	0.2763

A3	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.0950	0.4033	0.1573	0.3373	0.3289	0.1097	0.3671
$\widehat{\theta}_k^{\mathtt{NGJ}}$	0.1910	0.2989	0.3553	0.3774	0.2954	0.2495	0.4270
$\widehat{\theta}_k^{\mathtt{I}}$	0.0781	0.3848	0.0640	0.3600	0.4290	0.0573	0.3668
$\widehat{\theta}_k^{\mathtt{FF}}$	0.1726	0.6209	0.1978	0.4741	0.4548	0.1633	0.2951
$\widehat{\theta}_k^{\mathtt{BL}}$	0.0321	0.4584	0.0641	0.0948	**0.0562**	0.0728	0.0657
$\widehat{\theta}_k^{\mathtt{B}}$	0.0839	0.5430	**0.0385**	**0.0320**	0.0868	0.0298	0.1542
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.1130	0.2348	0.1095	0.1169	0.1568	0.0986	0.1913
$\widehat{\theta}_k^{\mathtt{SL}}$	0.0992	0.2507	0.0937	0.1143	0.1497	0.1013	0.1851

A4	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.1203	0.2496	0.1350	0.1409	0.1757	0.1336	0.2330
$\widehat{\theta}_k^{\mathtt{SL}}$	0.1054	**0.2313**	0.0931	0.1300	0.1753	0.1279	0.2279

Table 4: Absolute bias obtained for simulated samples of size $n = 1000$. For estimators $\widehat{\theta}_k^{\mathtt{BL}}$ and $\widehat{\theta}_k^{\mathtt{B}}$ we considered blocks of length 3 except in MCBEV and GARCH models where we used blocks of length 4 and 5, respectively. For estimators $\widehat{\theta}_k^{\mathtt{DJ}}$ and $\widehat{\theta}_k^{\mathtt{SL}}$ we considered blocks of length 20. For estimator $\widehat{\theta}_k^{\mathtt{R}}$ ($\widehat{\theta}_k^{\mathtt{FF}}$) we considered runs (cycles) of length 2 in MAR and AR, of length 3 in MM, ARCauchy and ARUnif, of length 4 in MCBEV and length 5 in GARCH. See Remark 3.1.

A1	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.0301	0.2748	0.0481	0.0607	**0.0128**	0.0134	0.0608
$\widehat{\theta}_k^{\mathtt{I}}$	**0.0007**	0.2901	0.0016	0.1815	0.2498	0.0107	0.1523
$\widehat{\theta}_k^{\mathtt{FF}}$	0.0604	0.4010	0.0593	0.0780	0.0074	0.0707	0.1184
$\widehat{\theta}_k^{\mathtt{BL}}$	0.0032	0.4350	0.0595	0.0400	0.0625	0.0683	0.0703
$\widehat{\theta}_k^{\mathtt{B}}$	0.0146	0.4461	0.0420	0.0174	0.0881	0.0494	0.0886
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0368	0.3275	0.0347	0.0332	0.0237	0.0260	0.0934
$\widehat{\theta}_k^{\mathtt{SL}}$	0.0417	0.2780	0.0459	0.0562	0.1221	0.0800	0.1571

A2	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.1768	0.7390	0.4300	0.6388	0.7489	0.2584	0.4457
$\widehat{\theta}_k^{\mathtt{NGJ}}$	0.0437	0.3963	0.2290	0.6301	0.2500	0.1622	0.6187
$\widehat{\theta}_k^{\mathtt{I}}$	0.1075	0.5675	0.0368	0.3600	0.2500	0.0334	0.5136
$\widehat{\theta}_k^{\mathtt{FF}}$	0.4389	0.9578	0.4783	0.6397	0.7500	0.2700	0.4458
$\widehat{\theta}_k^{\mathtt{BL}}$	0.0029	0.4747	0.0533	0.1550	0.0156	0.0629	**0.0570**
$\widehat{\theta}_k^{\mathtt{B}}$	0.0894	0.5922	0.0650	0.0608	0.2316	0.0041	0.1370
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0283	0.3423	0.3701	0.0453	0.1888	0.0847	0.1956
$\widehat{\theta}_k^{\mathtt{SL}}$	0.0872	0.3160	0.0839	0.1281	0.2122	0.1176	0.2455

A3	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.0853	0.4012	0.1554	0.3363	0.3278	0.1070	0.3667
$\widehat{\theta}_k^{\mathtt{NGJ}}$	0.0031	0.2511	0.1745	0.3805	0.2954	0.0456	0.3930
$\widehat{\theta}_k^{\mathtt{I}}$	0.0231	0.3732	**0.0015**	0.3373	0.2490	**0.0106**	0.3145
$\widehat{\theta}_k^{\mathtt{FF}}$	0.1703	0.6203	0.1965	0.4735	0.4545	0.1620	0.2945
$\widehat{\theta}_k^{\mathtt{BL}}$	0.0006	0.4565	0.0550	0.0677	0.0254	0.0660	0.0597
$\widehat{\theta}_k^{\mathtt{B}}$	0.0775	0.5422	0.0314	**0.0017**	0.0855	0.0124	0.1522
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0597	0.2013	0.0679	0.0519	0.0959	0.0792	0.1490
$\widehat{\theta}_k^{\mathtt{SL}}$	0.0647	0.2340	0.0673	0.0700	0.1147	0.0848	0.1533

A4	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0658	0.2172	0.0699	0.0860	0.1134	0.1104	0.2088
$\widehat{\theta}_k^{\mathtt{SL}}$	0.0668	**0.1988**	0.0460	0.0833	0.1302	0.1106	0.2057

Table 5: Root mean squared errors obtained for simulated samples of size $n = 5000$. For estimators $\widehat{\theta}_k^{\mathtt{BL}}$ and $\widehat{\theta}_k^{\mathtt{B}}$ we considered blocks of length 3 except in MCBEV and GARCH models where we used blocks of length 4 and 5, respectively. For estimators $\widehat{\theta}_k^{\mathtt{DJ}}$ and $\widehat{\theta}_k^{\mathtt{SL}}$ we considered blocks of length 20. For estimator $\widehat{\theta}_k^{\mathtt{R}}$ ($\widehat{\theta}_k^{\mathtt{FF}}$) we considered runs (cycles) of length 2 in MAR and AR, of length 3 in MM, ARCauchy and ARUnif, of length 4 in MCBEV and length 5 in GARCH. See Remark 3.1.

A1	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.0434	0.2504	0.0395	0.0518	0.0314	0.0409	0.1257
$\widehat{\theta}_k^{\mathtt{I}}$	0.0460	0.2904	0.0327	0.0799	0.2499	0.0387	0.1004
$\widehat{\theta}_k^{\mathtt{FF}}$	0.0541	0.4309	0.0423	0.0558	**0.0303**	0.0608	0.1072
$\widehat{\theta}_k^{\mathtt{BL}}$	0.0231	0.4234	0.0554	0.0585	0.0803	0.0715	0.0683
$\widehat{\theta}_k^{\mathtt{B}}$	0.0290	0.4272	0.0572	0.0457	0.0865	0.0656	0.0823
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0587	0.2439	0.0755	0.0895	0.1030	0.0789	0.1415
$\widehat{\theta}_k^{\mathtt{SL}}$	0.0592	0.2317	0.0557	0.0796	0.0954	0.0781	0.1407

A2	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.1500	0.7498	0.4382	0.6396	0.7498	0.2648	0.4464
$\widehat{\theta}_k^{\mathtt{NGJ}}$	0.0700	0.3918	0.2273	0.4596	0.2500	0.1532	0.6150
$\widehat{\theta}_k^{\mathtt{I}}$	0.0940	0.5357	0.0446	0.3600	0.2500	0.0370	0.5351
$\widehat{\theta}_k^{\mathtt{FF}}$	0.4416	0.9505	0.4915	0.6398	0.7500	0.2749	0.4456
$\widehat{\theta}_k^{\mathtt{BL}}$	**0.0113**	0.4786	0.0606	0.2282	0.0510	0.0688	**0.0547**
$\widehat{\theta}_k^{\mathtt{B}}$	0.0838	0.5925	0.0502	0.0505	0.1818	0.0189	0.1308
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0862	0.3085	0.0942	0.1364	0.2416	0.1171	0.2121
$\widehat{\theta}_k^{\mathtt{SL}}$	0.1056	0.2843	0.1056	0.1595	0.2502	0.1350	0.2456

A3	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.0970	0.2576	0.1530	0.3369	0.3316	0.1063	0.3688
$\widehat{\theta}_k^{\mathtt{NGJ}}$	0.0700	0.3918	0.2273	0.4596	0.2500	0.1532	0.6150
$\widehat{\theta}_k^{\mathtt{I}}$	0.0611	0.4150	**0.0300**	0.3600	0.2499	0.0378	0.3648
$\widehat{\theta}_k^{\mathtt{FF}}$	0.1663	0.6147	0.1941	0.4709	0.4544	0.1587	0.2940
$\widehat{\theta}_k^{\mathtt{BL}}$	0.0190	0.4818	0.0606	0.1082	0.0253	0.0728	0.0777
$\widehat{\theta}_k^{\mathtt{B}}$	0.0787	0.5461	0.0366	**0.0308**	0.0858	**0.0164**	0.1553
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0716	0.2528	0.0777	0.0992	0.1202	0.1035	0.1692
$\widehat{\theta}_k^{\mathtt{SL}}$	0.0842	0.2695	0.0908	0.1062	0.1312	0.1152	0.1769

A4	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0664	**0.2149**	0.0561	0.0989	0.1147	0.1259	0.2109
$\widehat{\theta}_k^{\mathtt{SL}}$	0.0644	0.2155	0.0523	0.0916	0.1066	0.1224	0.2100

Table 6: Absolute bias obtained for simulated samples of size $n = 5000$. For estimators $\widehat{\theta}_k^{\mathtt{BL}}$ and $\widehat{\theta}_k^{\mathtt{B}}$ we considered blocks of length 3 except in MCBEV and GARCH models where we used blocks of length 4 and 5, respectively. For estimators $\widehat{\theta}_k^{\mathtt{DJ}}$ and $\widehat{\theta}_k^{\mathtt{SL}}$ we considered blocks of length 20. For estimator $\widehat{\theta}_k^{\mathtt{R}}$ ($\widehat{\theta}_k^{\mathtt{FF}}$) we considered runs (cycles) of length 2 in MAR and AR, of length 3 in MM, ARCauchy and ARUnif, of length 4 in MCBEV and length 5 in GARCH. See Remark 3.1.

A1	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.0180	0.2431	0.0293	0.0273	0.0035	**0.0015**	0.1093
$\widehat{\theta}_k^{\mathtt{I}}$	0.0065	0.2784	**0.0024**	0.0522	0.2499	0.0118	**0.0482**
$\widehat{\theta}_k^{\mathtt{FF}}$	0.0360	0.4132	0.0321	0.0343	**0.0015**	0.0500	0.1023
$\widehat{\theta}_k^{\mathtt{BL}}$	0.0044	0.4223	0.0050	0.0363	0.0660	0.0669	0.0638
$\widehat{\theta}_k^{\mathtt{B}}$	0.0015	0.4258	0.0462	0.0238	0.0774	0.0586	0.0779
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0372	0.2241	0.0383	0.0640	0.0752	0.0613	0.1229
$\widehat{\theta}_k^{\mathtt{SL}}$	0.0379	0.2251	0.0469	0.0570	0.0711	0.0648	0.1202

A2	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.1394	0.7498	0.4374	0.6396	0.7498	0.2612	0.4464
$\widehat{\theta}_k^{\mathtt{NGJ}}$	0.0475	0.3896	0.2176	0.4050	0.2500	0.1428	0.6102
$\widehat{\theta}_k^{\mathtt{I}}$	0.0852	0.5310	0.0280	0.3600	0.2500	0.0267	0.5330
$\widehat{\theta}_k^{\mathtt{FF}}$	0.4406	0.9501	0.4908	0.6398	0.7500	0.2688	0.4456
$\widehat{\theta}_k^{\mathtt{BL}}$	**0.0009**	0.4783	0.0580	0.2202	0.0349	0.0680	0.0532
$\widehat{\theta}_k^{\mathtt{B}}$	0.0814	0.5919	0.0308	**0.0040**	0.1602	0.0048	0.1284
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0747	0.3002	0.0856	0.1167	0.2156	0.1073	0.1987
$\widehat{\theta}_k^{\mathtt{SL}}$	0.0959	0.2793	0.0976	0.1393	0.2163	0.1240	0.2304

A3	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{R}}$	0.0947	0.2573	0.1526	0.3367	0.3314	0.1055	0.3687
$\widehat{\theta}_k^{\mathtt{NGJ}}$	0.0107	0.2890	0.1773	0.3896	0.2990	0.0859	0.4441
$\widehat{\theta}_k^{\mathtt{I}}$	0.0546	0.4124	0.0164	0.3600	0.2499	0.0210	0.3389
$\widehat{\theta}_k^{\mathtt{FF}}$	0.1657	0.6146	0.1939	0.4708	0.4544	0.1584	0.2939
$\widehat{\theta}_k^{\mathtt{BL}}$	0.0011	0.4813	0.0580	0.1006	0.0152	0.0701	0.0765
$\widehat{\theta}_k^{\mathtt{B}}$	0.0777	0.5460	0.0346	0.0064	0.0856	0.0114	0.1545
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0609	0.2470	0.0692	0.0857	0.1075	0.0990	0.1592
$\widehat{\theta}_k^{\mathtt{SL}}$	0.0758	0.2660	0.0854	0.0988	0.1217	0.1121	0.1688

A4	MAR	AR	MM	ARCauchy	ARUnif	MCBEV	GARCH
$\widehat{\theta}_k^{\mathtt{DJ}}$	0.0562	**0.2079**	0.0451	0.0835	0.0853	0.1212	0.2049
$\widehat{\theta}_k^{\mathtt{SL}}$	0.0570	0.2097	0.0449	0.0812	0.0778	0.1190	0.2065

4. APPLICATION TO REAL DATA

We consider the daily maximum temperatures (in degrees Celsius) at Uccle (Belgium), from 1901 to 1999, on the warmest month of July (thus stationarity is assumed), consisting in $n = 3051$ observations. The data is available at "`http://lstat.kuleuven.be/Wiley/Data/ecad00045TX.txt`" and is plotted in Figure 2. The extremal index of this series was analyzed in Beirlant *et al.* ([2] 2004), where the respective estimates, obtained through parametric modeling, ranged between 0.49 and 0.56.

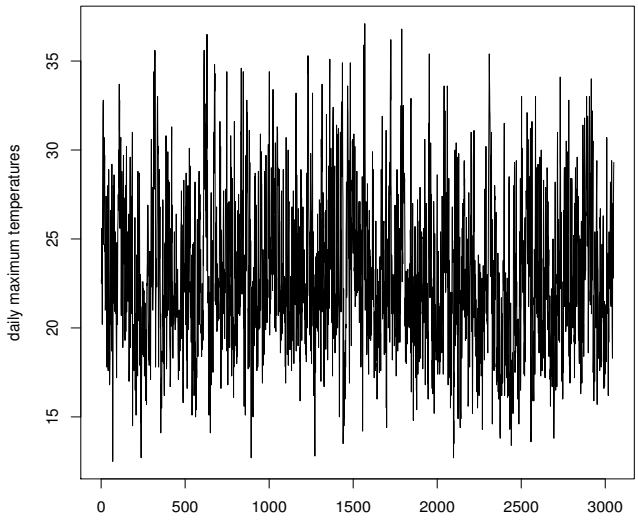


Figure 2: July daily maximum temperatures (in degrees Celsius) at Uccle, over the years 1901–1999.

We start by checking if we can validate some condition $\mathrm{D}^{(k)}(u_n)$. To this end, we use the empirical methodology of Ferreira and Ferreira ([10], 2015) by calculating the proportion of anti-$\mathrm{D}^{(m)}(u_n)$ events among the exceedances for several pairs of normalized levels u_n and block sizes r_n:

$$p(u_n, r_n) = \frac{\sum_{j=1}^{n-r_n+1} \mathbb{1}_{\left\{X_j > u_n, X_{j+1} \le u_n, \ldots, X_{j+m-1} \le u_n, M_{j+m-1, r_n+j-1} > u_n\right\}}}{\sum_{j=1}^{n} \mathbb{1}_{\{X_j > u_n\}}}.$$

More precisely, for each fixed $\tau > 0$, we take u_n as the empirical $(1 - \tau/n)$-th quantile for increasing sample sizes n and choose the sequence $\{b_n = [n/r_n]\}_n$ growing at a slower rate than n, e.g., $b_n = [(\log n)^a]$, for some $a > 0$. If $\mathrm{D}^{(m)}(u_n)$ holds with b_n, the points $(n, p(u_n, r_n))$ approach zero as $n \to \infty$. Based on the

suggested declustering parameter $r = 4$ in Beirlant *et al.* ([2] 2004), we have analyzed the proportions of anti-$\mathrm{D}^{(4)}(u_n)$, plotted in Figure 3 (right panel) for $\tau = 15$ (full line) and $\tau = 20$ (dashed line), with $k_n = [(\log n)^{2.5}]$. Observe that the values are small and almost indistinguishable from the proportions of anti-$\mathrm{D}^{(3)}(u_n)$ (left panel). We have also taken $k_n = [(\log n)^3]$ which led to null proportions in both cases. Therefore, we assume the validity of the $\mathrm{D}^{(3)}(u_n)$ local condition and consider run length 3 for the runs estimator and cycles of length 2 for the FF estimator in (2.8); see Remark 3.1. We also take block-length 3 in the blocks estimators. The disjoint and slides methods were implemented with block-length 15.

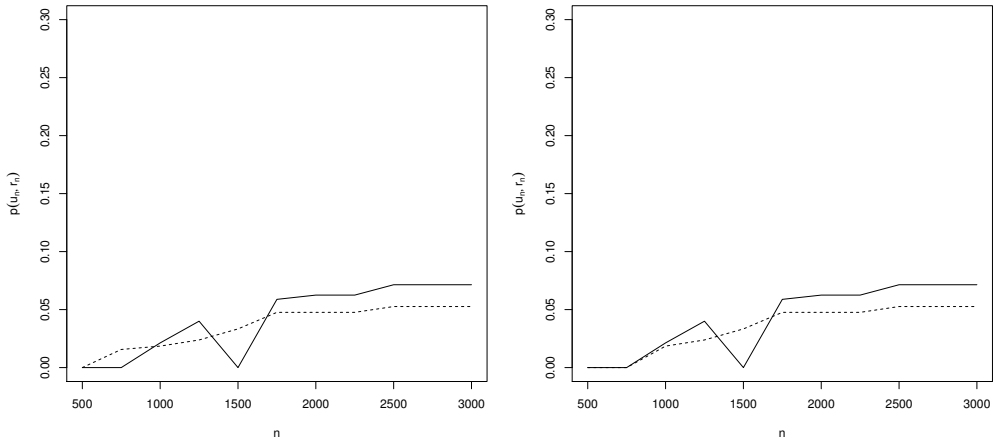


Figure 3: Observed proportions of anti-$\mathrm{D}^{(3)}(u_n)$ (left) and anti-$\mathrm{D}^{(4)}(u_n)$ conditions for Uccle data, for $\tau = 15$ (full line) and $\tau = 20$ (dashed line), with $k_n = [(\log n)^{2.5}]$.

The sample paths of the considered estimators in (2.11) and (3.1) are in Figure 4. Under algorithm A4, we obtained the estimate 0.51 for both disjoint and slide estimators. We have also applied the bias-reduced GJ Nandagopalan's runs estimator in (3.1) from which the values 0.41 and 0.57 were derived under A2 and A3, respectively. The remaining estimates are summarized in Table 7. The results are mostly in agreement with the simulation study.

Table 7: Extremal index estimates for Uccle data.

	$\widehat{\theta}_k^{\mathtt{R}}$	$\widehat{\theta}_k^{\mathtt{I}}$	$\widehat{\theta}_k^{\mathtt{FF}}$	$\widehat{\theta}_k^{\mathtt{B}}$	$\widehat{\theta}_k^{\mathtt{BL}}$	$\widehat{\theta}_k^{\mathtt{DJ}}$	$\widehat{\theta}_k^{\mathtt{SL}}$
A1	0.49	0.47	0.46	0.50	0.51	0.53	0.57
A2	0.10	0.33	0.05	0.39	0.50	0.52	0.53
A3	0.32	0.30	0.28	0.42	0.50	0.49	0.53

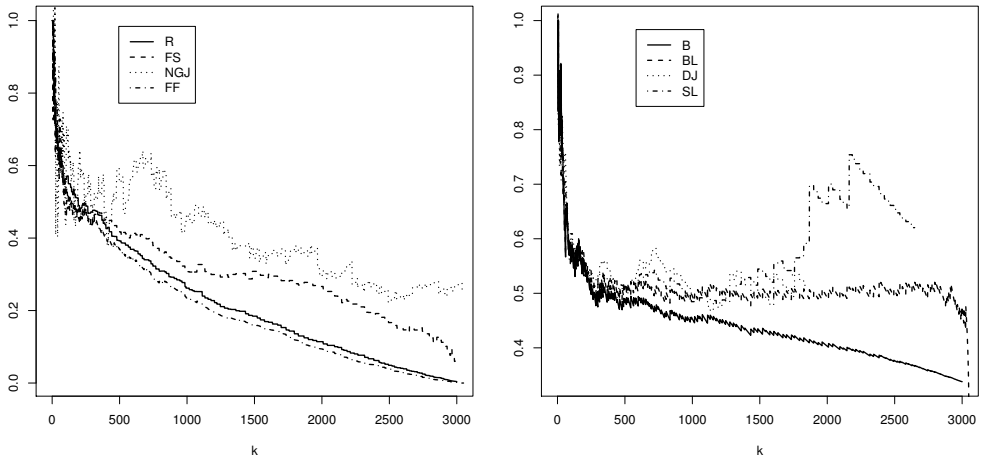


Figure 4: Sample paths of estimators in (2.11) and estimator (3.1) for Uccle data.

5. DISCUSSION

We have analyzed several estimators of the extremal index under different methodologies. The procedure based in Frahm *et al.* ([12] 2005) revealed an overall satisfactory performance. The best results were mostly observed within the blocks estimators, $\widehat{\theta}_k^{\mathtt{B}}$ and $\widehat{\theta}_k^{\mathtt{BL}}$, under the methodology of Neves *et al.* ([23] 2015). The large biases observed in the AR process makes inference within weak dependence, i.e., $\theta = 1$, an open topic to explore in this framework. Other methods to analyze the local dependence D-conditions are also welcome. The bias-reduced GJ Nandagopalan's estimator is sensitive to the restricted condition D$''$ and a generalization of the method to the broader runs estimator may be more advantageous. These points will be addressed in a future work.

ACKNOWLEDGMENTS

This research was financed by Portuguese Funds through FCT — Fundação para a Ciência e a Tecnologia, through the projects UID/MAT/00013/2013 and UID/MAT/00006/2013 and by the Research Center CEMAT through the Project UID/Multi/04621/2013.
The author also acknowledges the valuable suggestions from the referees.

REFERENCES

[1] Ancona-Navarrete, M.A. and Tawn, J.A. (2000). A comparison of methods for estimating the extremal index, *Extremes*, **3**, 5–38.

[2] Beirlant, J.; Goegebeur, Y.; Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Application*, John Wiley.

[3] Chernick, M.R. (1978). *Mixing Conditions and Limit Theorems for Maxima of Some Stationary Sequences*, PhD dissertation, Stanford University.

[4] Chernick, M.R.; Hsing, T. and McCormick, W.P. (1991). Calculating the extremal index for a class of stationary sequences, *Advaces in Applied Probability*, **23**, 835–850.

[5] Ferreira, M. (2014). *A heuristic procedure to estimate the tail index*. In "Proceedings of the 14th International Conference in Computational Science and Its Applications — ICCSA 2014" (B. Apduhan, A.M. Rocha, S. Misra, D. Taniar, O. Gervasi and B. Murgante, Eds.), IEEE-Computer Society 4264a241, 241–245.

[6] Ferreira, M. (2015a). Estimating the tail index: Another algorithmic method, *ProbStat Forum*, **8**(5), 45–53.

[7] Ferreira, M. (2015b). Estimating the extremal index through the tail dependence concept, *Discussiones Mathematicae Probability and Statistics*, **35**, 61–74.

[8] Ferreira, M. and Ferreira, H. (2012). On extremal dependence: some contributions, *TEST*, **21**(3), 566–583.

[9] Ferreira, M. and Silva, S. (2014). An Analysis of a Heuristic Procedure to Evaluate Tail (in)dependence, *Journal of Probability and Statistics*, Vol. **2014**, Article ID 913621, 15 pages.

[10] Ferreira, H. and Ferreira, M. (2015). Estimating the extremal index through local dependence. (Submitted.)

[11] Ferro, C.A. and Segers, J. (2003). Inference for clusters of extremes, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**, 545–556.

[12] Frahm, G.; Junker, M. and Schmidt, R. (2005). Estimating the tail-dependence coefficient: properties and pitfalls, *Insurance: Mathematics and Economics*, **37**(1), 80–100.

[13] Gomes, M.I.; Hall, A. and Miranda, C. (2008). Subsampling techniques and the jackknife methodology in the estimation of the extremal index, *Journal of Statistical Computation and Simulation*, **52**, 2022–2041.

[14] Gomes, M.I.; Henriques-Rodrigues, L.; Fraga Alves, M.I. and Manjunath, B.G. (2013). Adaptive PORT-MVRB estimation: an empirical comparison of two heuristic algorithms, *Journal of Statistical Computation and Simulation*, **83**(6), 1129–1144.

[15] Hsing, T. (1991). Estimating the parameters of rare events, *Stochastic Processes and their Applications*, **37**, 117–139.

[16] Hsing, T. (1993). Extremal index estimation for a weakly dependent stationary sequence, *Annals of Statistics*, **21**, 2043–2071.

[17] Hsing, T.; Hüsler, J. and Leadbetter, M.R. (1988). On the exceedance point process for a stationary sequence, *Probability Theory and Related Fields*, **78**, 97–112.

[18] Laurini, F. and Tawn, J. (2012). The extremal index for GARCH(1,1) processes, *Extremes*, **15**, 511–529.

[19] Leadbetter, M.R. (2012). On extremes values in stationary sequences, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **15**, 511–529.

[20] Leadbetter, M.R.; Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.

[21] Leadbetter, M.R. and Nandagopalan, S. (1989). *On exceedance point process for stationary sequences under mild oscillation restrictions*. In "Extreme value theory: proceedings, Oberwolfach 1987" (J. Hüsler and D. Reiss, Eds.), Springer, New York, 69–80.

[22] Nandagopalan, S. (1990). *Multivariate Extremes and Estimation of the Extremal Index*, Ph.D. Thesis, University of North Carolina, Chapel Hill.

[23] Neves, M.; Gomes, M.I.; Figueiredo, F. and Prata Gomes, D. (2012). Modeling Extreme Events: Sample Fraction Adaptive Choice in Parameter Estimation, *Journal of Statistical Theory and Practice*, **9**(1), 184–199.

[24] O'Brien, G.L. (1974). The maximum term of uniformly mixing stationary processes, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **30**, 57–63.

[25] Robert, C.Y.; Segers, J. and Ferro, C.A. (2009). A sliding blocks estimator for the extremal index, *Electronic Journal of Statistics*, **3**, 993–1020.

[26] Smith, R.L. (1992). The extremal index for a Markov chain, *Journal of Applied Probability*, **29**, 37–45.

[27] Smith, R.L. and Weissman, I. (1994). Estimating the extremal index, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **56**, 515–528.

[28] Süveges, M. (2007). Likelihood estimation of the extremal index, *Extremes*, **10**, 41–55.

[29] Weissman, I. and Novak, S.Y. (1998). On blocks and runs estimators of the extremal index, *Journal of Statistical Planning and Inference*, **66**(2), 281–288.

REVSTAT – Statistical Journal
Volume 16, Number 1, January 2018, 137–166

COMPUTATIONALLY EFFICIENT GOODNESS-OF-FIT TESTS FOR THE ERROR DISTRIBUTION IN NONPARAMETRIC REGRESSION

Authors: G.I. Rivas-Martínez
– Laboratorio de Sistemas de Potencia y Control, Universidad Nacional de Asunción, Paraguay
gusyri@hotmail.com

M.D. Jiménez-Gamero
– Dpto. de Estadística e Investigación Operativa, Universidad de Sevilla, España
dolores@us.es

Received: April 2016 Revised: August 2016 Accepted: August 2016

Abstract:

- Several procedures have been proposed for testing goodness-of-fit to the error distribution in nonparametric regression models. The null distribution of the associated test statistics is usually approximated by means of a parametric bootstrap which, under certain conditions, provides a consistent estimator. This paper considers a goodness-of-fit test whose test statistic is an L_2 norm of the difference between the empirical characteristic function of the residuals and a parametric estimate of the characteristic function in the null hypothesis. It is proposed to approximate the null distribution through a weighted bootstrap which also produces a consistent estimator of the null distribution but, from a computational point of view, is more efficient than the parametric bootstrap.

Key-Words:

- *goodness-of-fit; empirical characteristic function; regression residuals; weighted bootstrap; consistency.*

AMS Subject Classification:

- 62G08, 62G09, 62G10.

1. INTRODUCTION

Let (X, Y) be a bivariate random vector satisfying the general nonparametric regression model

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1.1}$$

where $m(x) = E(Y \mid X = x)$ is the regression function, $\sigma^2(x) = \mathrm{Var}(Y \mid X = x)$ is the conditional variance function and ε is the regression error, which is assumed to be independent of X. Note that, by construction, $E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = 1$. The covariate X is continuous with density function f_X. The regression function, the variance function, the error distribution and that of the covariate are unknown and no parametric models are assumed for them.

Because the knowledge of the error distribution will improve the statistical analysis of model (1.1), several authors have proposed tests for such distribution, that is, tests of the null hypothesis

$$H_0 \colon F \in \mathcal{F},$$

versus the alternative

$$H_1 \colon F \notin \mathcal{F},$$

where F stands for the cumulative distribution function (CDF) of ε and $\mathcal{F}$ is a parametric family,

$$\mathcal{F} = \left\{ F(\cdot; \theta),\ \ \theta \in \Theta \right\}, \qquad \Theta \subseteq \mathbb{R}^p.$$

Examples are the tests in Neumeyer *et al.* [17] and Heuchenne and Van Keilegom [6], which are based on comparing the empirical CDF of the residuals to a parametric estimator of the CDF under the null hypothesis. Since the equality of the CFDs can be also interpreted in terms of the associated characteristic functions (CFs), Hušková and Meintanis [11] have proposed a test for H_0 that is based on comparing the empirical CF of the residuals to a parametric estimator of the CF under the null hypothesis. As commented in Jiménez-Gamero [13], it is interesting to observe that the last paper requires weaker conditions for the validity of the procedures than the ones based on the CDF. Nevertheless, in all cases the limit distribution of the proposed test statistics is unknown, even under the null distribution, because it depends on the unknown value of the parameter θ. To overcome this difficulty, these papers propose to use a parametric bootstrap (PB) for approximating the null distribution of the test statistic. Although very easy to implement, the PB can become very computationally expensive as the sample size and/or the number of unknown parameters increase.

This paper studies another method for estimating the null distribution of the test statistic $T_{n,w}(\hat{\theta})$ in [11]. Specifically, a weighted bootstrap (WB) approximation in the sense of Burke [2] is considered (see also Zhu [23]). This method

has been previously suggested in Kojadinovic and Yan [15], to approximate the null distribution of goodness-of-fit (GOF) tests based on the empirical CDF, and in Jiménez-Gamero and Kim [14], to approximate the null distribution of GOF tests based on the empirical CF (ECF), among others. Both papers assume observable independent and identically distributed (IID) data. They show that the properties of the WB are quite similar to those of the PB (it provides a consistent estimator of the null distribution and the resulting test is able to detect any alternative) but, from a computational point of view, it is more efficient. In view of the good properties of the WB in these and other papers, it is also expected to work satisfactorily for estimating the null distribution of the test statistic considered in this paper. The purpose of the current study is to investigate, both theoretically and empirically, the use of the WB for approximating the null distribution of $T_{n,w}(\hat{\theta})$. A main difference between the setting in this paper and the one in [14, 15] is that in our case the errors are not observable. So we replace the errors by the residuals, but the residuals are not independent.

The paper is organized as follows. Section 2 describes the test statistic and explains some problems with the WB approximation. Section 3 gives a solution to the problems described in the previous section and proves the consistency of the proposed WB approximation. It also shows that the resulting test is consistent, in the sense of being able to detect any alternative. The application of the proposed WB approximation requires the estimation of certain functions appearing in the linear expansion of the parameter estimators. The estimation of such functions is dealt with in Section 4. Section 5 reports the results of some simulation experiments designed to study the finite sample performance of the proposed approximation and to compare it to the PB. From this numerical study it is concluded that both approximations behave quite closely but, from a computational point of view, the WB outperforms the PB. Section 6 concludes and outlines possible extensions of the results presented in this paper. All proofs and technical details are deferred to the last section.

The following notation will be used along the paper: all vectors are column vectors; for any vector a, a_k denotes its k-th coordinate and $\|a\|$ its Euclidean norm; the superscript ^{T} denotes transpose; E_θ and P_θ denote expectation and probability, respectively, assuming that the data has CDF $F(\cdot;\theta)$; P_* denotes the conditional probability law, given the data; all limits in this paper are taken when $n \to \infty$; $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution; $\xrightarrow{P}$ denotes convergence in probability; $\xrightarrow{a.s.}$ denotes the almost sure convergence; for any complex number $z = a + \mathrm{i}b$, $|z|$ is its modulus; an unspecified integral denotes integration over the whole real line $\mathbb{R}$; for a given non-negative real-valued function w we denote $\|\cdot\|_w$ to the norm and $\langle\cdot,\cdot\rangle_w$ to the scalar product in the Hilbert space $L^2(w) = \{g : \mathbb{R} \to \mathbb{C}, \int |g(t)|^2 w(t)dt < \infty\}$; if F is a CDF, then $L^2(F) = \{g : \mathbb{R} \to \mathbb{C}, \int |g(t)|^2 dF(t) < \infty\}$; for any real function $f(t;\theta)$ differentiable at $t \in \mathbb{R}$ and at $\theta = (\theta_1, \theta_2, ..., \theta_p)^T \in \mathbb{R}^p$ the following notations will be

used:

$$f'(t;\theta) = \frac{\partial}{\partial t} f(t;\theta)\,, \qquad f_{(r)}(t;\theta) = \frac{\partial}{\partial \theta_r} f(t;\theta), \quad 1 \le r \le p,$$

$$\nabla f(t;\theta) = \Big(f_{(1)}(t;\theta), f_{(2)}(t;\theta), ..., f_{(p)}(t;\theta) \Big)^T.$$

2. THE TEST STATISTIC

Let $(X_1, Y_1), ..., (X_n, Y_n)$ be IID from model (1.1), that is, $Y_j = m(X_j) + \sigma(X_j)\varepsilon_j$, $1 \le j \le n$. Since the hypothesis H_0 is on the common error distribution, $\varepsilon_1, ..., \varepsilon_n$, and the errors are not observable, the inference must be based on the residuals,

$$\hat{\varepsilon}_j = \frac{Y_j - \hat{m}(X_j)}{\hat{\sigma}(X_j)}, \quad 1 \le j \le n,$$

where $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$ are estimators of $m(\cdot)$ and $\sigma(\cdot)$, respectively. Several choices are possible for $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$. Here, as in [11], we use the following kernel estimators for the density function f_X of X, the regression function $m(\cdot)$ and the variance function $\sigma^2(\cdot)$,

$$\begin{aligned}
\hat{f}_X(x) &= \frac{1}{n} \sum_{j=1}^{n} K_{h_n}(X_j - x)\,, \\
\hat{m}(x) &= \frac{1}{n\hat{f}_X(x)} \sum_{j=1}^{n} K_{h_n}(X_j - x)\, Y_j, \\
\hat{\sigma}^2(x) &= \frac{1}{n\hat{f}_X(x)} \sum_{j=1}^{n} K_{h_n}(X_j - x)\, \{Y_j - \hat{m}(x)\}^2\,,
\end{aligned}$$

where $K_{h_n}(\cdot) = \frac{1}{h_n} K(\frac{\cdot}{h_n})$, $K(\cdot)$ is a kernel and h_n is the bandwidth, satisfying certain conditions that will be specified later.

Hušková and Meintanis [11] proposed the following test for testing H_0,

$$\Psi = \begin{cases} 1, \text{ if } \quad T_{n,w}(\hat{\theta}) \ge t_{n,\omega,\alpha}, \\ 0, \text{ otherwise,} \end{cases}$$

where $t_{n,\omega,\alpha}$ is the $1-\alpha$ percentile of the null distribution of $T_{n,\omega}(\hat{\theta})$,

$$T_{n,\omega}(\hat{\theta}) = n \int |c_n(t) - c(t,\hat{\theta})|^2 \omega(t) dt = n \|c_n(t) - c(t,\hat{\theta})\|_w^2, \tag{2.1}$$

$c_n(t)$ is the ECF of the residuals,

$$c_n(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(\mathrm{i}t\hat{\varepsilon}_j) = \frac{1}{n} \sum_{j=1}^{n} \cos(t\hat{\varepsilon}_j) + \mathrm{i}\frac{1}{n} \sum_{j=1}^{n} \sin(t\hat{\varepsilon}_j),$$

$c(t;\theta)$ is the CF associated to $F(\varepsilon;\theta)$, that is, $c(t;\theta) = E_\theta\{\exp(\mathrm{i}t\varepsilon)\} = R(t;\theta) + \mathrm{i}I(t;\theta)$, $\omega(t)$ is a nonnegative function such that $\int \omega(t)dt < \infty$, which may depend on θ, and $\hat{\theta}$ is a consistent estimator of θ satisfying the following assumption.

(**A.1**) Under H_0, $\sqrt{n}(\hat{\theta} - \theta_0) = \dfrac{1}{\sqrt{n}} \displaystyle\sum_{j=1}^{n} \psi(\varepsilon_j;\theta_0) + o_p(1)$, where θ_0 is the true parameter value, $E_{\theta_0}\{\psi(\varepsilon_j;\theta_0)\} = 0$ and $E_{\theta_0}\{\|\psi(\varepsilon_j;\theta_0)\|^2\} < \infty$.

Assumption (A.1) implies that, when the null hypothesis is true and θ_0 denotes the true parameter value, $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically normally distributed. This assumption is satisfied by commonly used estimators such as maximum likelihood estimators and method of moment estimators when $\varepsilon_1, ..., \varepsilon_n$ are observable and, in such a case, the expression of the function ψ is well-known (see, for example, [1, Ch. 5]). In our setting, the errors are not observable and the expression of the function ψ differs from the observable case. This topic will be discussed in detail in Section 4.

Theorem 1 in [11] states that if $\hat{\theta}$ satisfies (A.1), H_0 is true and θ_0 is the true parameter value, under certain additional conditions (assumptions (A.2)–(A.7) in Section 7),

$$T_{n,\omega}(\hat{\theta}) \xrightarrow{\mathcal{L}} \|Z(t;\theta_0)\|^2_\omega, \tag{2.2}$$

where $\{Z(t;\theta_0),\, t \in \mathbb{R}\}$ is a centered Gaussian process on $L_2(\omega)$ with covariance structure of the form $Cov_{\theta_0}\{Z_1(\varepsilon;t,\theta_0,\psi), Z_1(\varepsilon;s,\theta_0,\psi)\}$,

$$\begin{aligned} Z_1(\varepsilon;t,\theta,\psi) \,=\, & \cos(t\varepsilon) + \sin(t\varepsilon) - R(t;\theta) - I(t;\theta) - t\varepsilon\{R(t;\theta) - I(t;\theta)\} \\ & -t\tfrac{\varepsilon^2-1}{2}\{R'(t;\theta) + I'(t;\theta)\} - \psi^T(\varepsilon;\theta)\{\nabla R(t;\theta) + \nabla I(t;\theta)\}. \end{aligned} \tag{2.3}$$

Clearly, the asymptotic null distribution of $T_{n,\omega}(\hat{\theta})$ is unknown. It depends on the hypothetical the error distribution, on the chosen estimator and the true unknown value of the parameter.

In order to try to approximate the null distribution of $T_{n,\omega}(\hat{\theta})$ we first observe that it resembles a degree-2 V-statistic, because

$$T_{n,\omega}(\hat{\theta}) = \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \rho(\hat{\varepsilon}_j, \hat{\varepsilon}_k; \hat{\theta}),$$

with $\rho(\varepsilon, z;\theta) = u(\varepsilon - z) - u_0(\varepsilon;\theta) - u_0(z;\theta) + u_{00}(\theta)$, $u_0(\varepsilon;\theta) = \int u(\varepsilon - z)dF(z;\theta)$, $u_{00}(\theta) = \int u(\varepsilon - z)dF(\varepsilon;\theta)dF(z;\theta)$, and $u(t) = \int \cos(t\varepsilon)\omega(\varepsilon)d\varepsilon$.

Dehling and Mikosch [4] (see also Hušková and Janssen [10]) showed that if $\varepsilon_1, ..., \varepsilon_n$ are IID, $\xi_1, ..., \xi_n$ are IID with $E(\xi_1) = 0$ and $\mathrm{Var}(\xi_1) = 1$, independent of $\varepsilon_1, ..., \varepsilon_n$ and $V_n = \frac{1}{n^2}\sum_{1\le j,k\le n} g(\varepsilon_j, \varepsilon_k)$ is a degenerate degree-2 V-statistic,

then the conditional distribution, given $\varepsilon_1, ..., \varepsilon_n$, of

$$\frac{1}{n} \sum_{1 \leq j,k \leq n} g(\varepsilon_j, \varepsilon_k) \xi_j \xi_k$$

consistently estimates that of nV_n. In the light of this result, since $\hat{\varepsilon}_j$ and $\hat{\theta}$ are approximations to ε_j and θ, respectively, one may be tempted to estimate the null distribution of $T_{n,\omega}(\hat{\theta})$ by means of the conditional distribution, given $(X_1, Y_1), ..., (X_n, Y_n)$, of

$$W^* = \frac{1}{n} \sum_{1 \leq j,k \leq n} \rho(\hat{\varepsilon}_j, \hat{\varepsilon}_k; \hat{\theta}) \xi_j \xi_k. \tag{2.4}$$

We will see that this approach is wrong. The next result gives the limit distribution of W^*. The required assumptions are listed in Section 7.

Theorem 2.1. *Suppose that $\|\hat{\theta} - \theta_1\| = o_p(1)$, for some $\theta_1 \in \Theta$, that assumptions (A.2)–(A.6) hold, that the first partial derivatives $R_{(r)}(t;\theta)$, $I_{(r)}(t;\theta)$, $1 \leq r \leq p$, exist and are continuous functions $\forall \theta \in U(\theta_1) \subseteq \Theta$, an open neighborhood of θ_1, and they are bounded by functions in $L_2(\omega)$, $\forall \theta \in U(\theta_1)$, then*

$$\sup_x |P_* \{W^* \leq x\} - P\{W_0 \leq x\}| \xrightarrow{P} 0,$$

where $W_0 = \|Z_0(t;\theta_1)\|^2_\omega$, $\{Z_0(t;\theta_1),\, t \in \mathbb{R}\}$ is a centered Gaussian process on $L_2(\omega)$ with covariance structure of the form $Cov\{Z_0(\varepsilon;t,\theta_1), Z_0(\varepsilon;s,\theta_1)\}$, $Z_0(\varepsilon;t,\theta) = \cos(t\varepsilon) + \sin(t\varepsilon) - R(t;\theta) - I(t;\theta)$.

From the result in Theorem 2.1 and (2.2), it is clear that the conditional distribution of W^* does not provide a consistent estimator of the null distribution of $T_{n,\omega}(\hat{\theta})$ because replacing $m(\cdot)$, $\sigma(\cdot)$ and θ by $\hat{m}(\cdot)$, $\hat{\sigma}(\cdot)$ and $\hat{\theta}$, respectively, has an impact on the asymptotic null distribution of the test statistic that is not captured by the conditional distribution of W^*. The next Section shows how to deal with this problem.

Before ending this section we do some comments on the behaviour of $\hat{\theta}$ under the alternative. Theorem 2.1 assumes that $\hat{\theta}$ has a limit (in probability), θ_1. In practice, to estimate θ one proceeds as if H_0 were true. For example, θ is usually estimated by its quasi maximum likelihood estimator, which maximizes the likelihood under the null hypothesis (with the errors replaced by the residuals). If H_0 is true, under certain assumptions, the resulting estimator converges to the true parameter value (see Section 4); if H_0 is not true, then proceeding as in White [22] for observable data, it can shown that, under certain conditions, the estimator also converges to a well-defined limit. Similar comments could be done for other estimators.

3. CONSISTENCY OF THE WB APPROXIMATION

If assumptions (A.1)–(A.7) hold and H_0 is true, from the proof of Theorem 1 in [11], it follows that

$$T_{n,\omega}(\hat{\theta}) = T_{1,n,\omega}(\theta_0) + o_p(1), \tag{3.1}$$

where

$$T_{1,n,\omega}(\theta) = \|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} Z_1(\varepsilon_j; t, \theta, \psi)\|^2_\omega,$$

with $Z_1(\varepsilon; t, \theta, \psi)$ as defined in (2.3). Now, from (3.1) and applying the results in [4], we get that the conditional distribution, given $(X_1, Y_1), ..., (X_n, Y_n)$, of

$$T^*_{1,n,\omega}(\theta_0) = \|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} Z_1(\varepsilon_j; t, \theta_0, \psi)\xi_j\|^2_\omega,$$

provides a consistent estimator of the distribution of $T_{n,\omega}(\hat{\theta})$, when H_0 is true. From a practical point of view, this result is useless because $Z_1(\varepsilon_j; t, \theta_0, \psi)$ depends on the non-observable error ε_j, on the unknown value of θ_0 and on the function $\psi(\varepsilon_j; \theta_0)$, whose explicit expression is usually unknown. Suppose that $\|\hat{\theta} - \theta_1\| = o_p(1)$, for some $\theta_1 \in \Theta$, θ_1 being the true parameter value if H_0 is true. To overcome these difficulties we replace ε_j by $\hat{\varepsilon}_j$, θ_0 by $\hat{\theta}$ and $\psi(\varepsilon_j; \theta_0)$ by $\psi_n(\hat{\varepsilon}_j; \hat{\theta})$, where $\psi_n(\cdot; \hat{\theta})$ is a function of the data which approximates ψ in such a way that

$$\begin{aligned} &\frac{1}{n}\sum_{j=1}^{n} \|\psi_n(\hat{\varepsilon}_j; \hat{\theta}) - \psi_1(\varepsilon_j; \theta_1)\|^2 \xrightarrow{P} 0, \\ &\text{with } E\{\|\psi_1(\varepsilon; \theta_1)\|^2\} < \infty \text{ and } \psi_1(\varepsilon; \theta_1) = \psi(\varepsilon; \theta_1) \text{ if } H_0 \text{ is true.} \end{aligned} \tag{3.2}$$

The choice of ψ_n will depend on ψ, that is, on the estimator of θ considered. Section 4 studies some proposals for ψ_n satisfying (3.2) for two common choices for $\hat{\theta}$: the maximum likelihood estimator and the method of moments estimator, both based on the residuals. So, the null distribution of $T_{n,\omega}(\hat{\theta})$ is now estimated by means of the conditional distribution, given $(X_1, Y_1), ..., (X_n, Y_n)$, of

$$T^*_{2,n,\omega}(\hat{\theta}) = \|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} Z_1(\hat{\varepsilon}_j; t, \hat{\theta}, \psi_n)\xi_j\|^2_\omega.$$

The next theorem gives the limit of the conditional distribution of $T^*_{2,n,\omega}(\hat{\theta})$, given (X_1, Y_1), ...(X_n, Y_n).

Theorem 3.1. *Suppose that $\|\hat{\theta} - \theta_1\| = o_p(1)$, for some $\theta_1 \in \Theta$, θ_1 being the true parameter value if H_0 is true, and that assumptions (A.1)–(A.7) and* (3.2) *hold, then*

$$\sup_x \left| P_* \left\{ T^*_{2,n,\omega}(\hat{\theta}) \le x \right\} - P\{T_2 \le x\} \right| \xrightarrow{P} 0,$$

where $T_2 = \|Z_2(t;\theta_1)\|^2_\omega$, $\{Z_2(t;\theta_1),\, t \in \mathbb{R}\}$ *is a centered Gaussian process on* $L_2(\omega)$ *with covariance structure of the form* $Cov\{Z_1(\varepsilon;t,\theta_1,\psi_1), Z_1(\varepsilon;s,\theta_1,\psi_1)\}$.

The result in Theorem 3.1 is valid whether the null hypothesis H_0 is true or not. An immediate consequence of this fact and (2.2) is the following.

Corollary 3.1. *If* H_0 *is true and the assumptions in Theorem 3.1 hold, then*

$$\sup_x \left| P_* \left\{ T^*_{2,n,\omega}(\hat{\theta}) \leq x \right\} - P_{\theta_1} \left\{ T_{n,\omega}(\hat{\theta}) \leq x \right\} \right| \xrightarrow{P} 0.$$

Let $\alpha \in (0,1)$ and

$$\Psi_* = \begin{cases} 1, \text{ if } \quad T_{n,\omega}(\hat{\theta}) \geq t^*_{2,n,\omega,\alpha}, \\ 0, \text{ otherwise,} \end{cases}$$

where $t^*_{2,n,\omega,\alpha}$ is the $1-\alpha$ percentile of the conditional distribution of $T^*_{2,n,\omega}(\hat{\theta})$, or equivalently, $\Psi_* = 1$ if $p^* \leq \alpha$, where $p^* = P_* \left\{ T^*_{2,n,\omega}(\hat{\theta}) \geq T_{n,\omega}(\hat{\theta})_{obs} \right\}$ and $T_{n,\omega}(\hat{\theta})_{obs}$ is the observed value of the test statistic. The result in Corollary 3.1 states that Ψ_* is asymptotically correct, in the sense that its type I error is asymptotically equal to the nominal value α.

Corollary 3.2. *Suppose that* H_0 *is not true and let* $c(t)$ *denote the true CF of the errors. If the assumptions in Theorem 3.1 hold and* ω *is such that*

$$\kappa = \|c(t) - c(t;\theta_1)\|^2_\omega > 0, \tag{3.3}$$

then $P(\Psi_* = 1) \to 1$.

Corollary 3.2 shows that, if ω is such that (3.3) holds, then the test Ψ_* is consistent in the sense of being able to asymptotically detect any (fixed) alternative. Since two distinct characteristic functions can be equal in a finite interval (Feller [5, p.506]), a general way to ensure (3.3) is to take ω positive for almost all (with respect to the Lebesgue measure) points in $\mathbb{R}$.

Remark 3.1. If model (1.1) is homoscedastic, that is, if $\sigma(x) = \sigma$, $\forall x$, for some unknown $\sigma > 0$, we can use the residuals $\tilde{\varepsilon}_j = Y_j - \hat{m}(X_j)$, $1 \leq j \leq n$, and consider σ as a parameter of the family $\mathcal{F}$. In this framework, the result in Theorem 3.1 (with weaker assumptions) keeps on being true with the following simpler expression for $Z_1(\varepsilon;t,\theta,\psi)$,

$$\begin{aligned} Z_1(\varepsilon;t,\theta,\psi) \;=\; & \cos(t\varepsilon) - R(t;\theta) + \sin(t\varepsilon) - I(t;\theta) - t\varepsilon R(t;\theta) + t\varepsilon I(t;\theta) \\ & -\psi^T(\varepsilon;\theta)\{\nabla R(t;\theta) + \nabla I(t;\theta)\}. \end{aligned}$$

Remark 3.2. If the null hypothesis is simple, then the result in Theorem 3.1 (with weaker assumptions) is also true with the following simpler expression for $Z_1(\varepsilon;t,\theta,\psi) = Z_1(\varepsilon;t)$,

$$\begin{aligned} Z_1(\varepsilon;t) \;=\; & \cos(t\varepsilon) - R(t) + \sin(t\varepsilon) - I(t) - t\varepsilon R(t) + t\varepsilon I(t) \\ & - t\frac{\varepsilon^2-1}{2}\{R'(t) + I'(t)\}, \end{aligned}$$

where $R(t)$ and $I(t)$ denote the real and the imaginary parts of the CF of the law in the null hypothesis.

Remark 3.3. If model (1.1) is homoscedastic and the null hypothesis is simple, which implies that $\sigma(x) = \sigma$, $\forall x$, for some known $\sigma > 0$, as observed in Remark 3.1, we can use the residuals $\tilde{\varepsilon} = Y_j - \hat{m}(X_j)$, $1 \le j \le n$. In this setting, the result in Theorem 3.1 (with weaker assumptions) is also true with the following simpler expression for $Z_1(\varepsilon;t,\theta,\psi) = Z_1(\varepsilon;t)$,

$$Z_1(\varepsilon;t) \;=\; \cos(t\varepsilon) - R(t) + \sin(t\varepsilon) - I(t) - t\varepsilon R(t) + t\varepsilon I(t),$$

where $R(t)$ and $I(t)$ denote the real and the imaginary parts of the CF of the law in the null hypothesis.

Remark 3.4. When the null hypothesis is simple, the asymptotic null distribution of the test statistic $T_{n,\omega}(\hat{\theta})$ does not depend on unknown parameters. So, in this case the asymptotic null distribution could be used to approximate the null distribution. The simulations carried out (reported in Section 5) reveal that, for small to moderate sample sizes, the WB provides a better fit.

Remark 3.5. Theorem 3 in [11] shows that the PB null distribution estimator of $T_{n,\omega}(\hat{\theta})$ satisfies a result which is similar to that stated in Corollary 3.1 for the WB estimator. Nevertheless, although the tests Ψ_* and the one obtained by approximating $t_{n,\omega,\alpha}$ through its PB estimator, are both of them consistent against all fixed alternatives, their powers will be different for finite sample sizes.

So far we have assumed that the weight function does not depend on θ, but in some cases it does. Such dependence is motivated by the recommendations in Epps and Pulley [8], who suggest to choose $\omega(t)$ giving high weight where the ECF is a relatively precise estimator of the population CF. It entails taking $\omega(t) = \nu\{|c(t;\hat{\theta})|\}$, for some ν, a nonnegative increasing function. For example, if $\int |c(t;\theta)|^2 dt < \infty$, one could choose $\omega(t) = |c(t;\hat{\theta})|^2 / \int |c(x;\hat{\theta})|^2 dx$, which is the choice for ω in Epps and Pulley [8] (see also Epps [7]). In addition, as observed in Jiménez-Gamero *et al.* [12], such choice for $\omega(t)$ may have some computational

advantages when the density (under the null hypothesis) of $\varepsilon_1 - \varepsilon_2$, $\varepsilon_1 - \varepsilon_2 + \varepsilon_3$ and $\varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4$ is known since from expression (14) in [12], the test statistic (2.1) can be expressed as

$$\frac{1}{f_{\varepsilon_1-\varepsilon_2}(0;\hat{\theta})}\left\{\frac{1}{n}\sum_{j,k=1}^{n} f_{\varepsilon_1-\varepsilon_2}(\hat{\varepsilon}_j - \hat{\varepsilon}_k;\hat{\theta}) - 2\sum_{j=1}^{n} f_{\varepsilon_1-\varepsilon_2+\varepsilon_3}(\hat{\varepsilon}_j;\hat{\theta}) + n f_{\varepsilon_1-\varepsilon_2+\varepsilon_3-\varepsilon_4}(0;\hat{\theta})\right\},$$

where $f_U(x;\theta)$ is the density function of U.

If the weight function ω depends on θ, $\omega(t) = \omega(t;\theta)$, then the test statistic (2.1) becomes

$$T_{n,\hat{\omega}}(\hat{\theta}) = n\int |c_n(t) - c(t;\hat{\theta})|^2 \omega(t;\hat{\theta})dt = n\|c_n(t) - c(t;\hat{\theta})\|^2_{\hat{\omega}},$$

where the subindex $\hat{\omega}$ means that the weight function depends on $\hat{\theta}$, that is, $\omega(t) = \omega(t;\hat{\theta})$. To deal with this case we will assume that the weight function is smooth as a function of θ, as expressed in the next assumption.

(**A.8**) $|\omega(t;\theta_1) - \omega(t;\theta)| \le \omega_0(t;\theta_1)\|\theta - \theta_1\|$, $\forall\,\theta$ in an open neighborhood of θ_1, with $\omega_0(t;\theta_1)$ satisfying $\int \omega_0(t;\theta_1)dt < \infty$.

If assumption (A.8) holds, assumptions (A.2), (A.7) hold with $\omega(t) = \omega_0(t;\theta)$ and H_0 is true, then

$$T_{n,\hat{\omega}}(\hat{\theta}) = T^1_{n,\omega}(\hat{\theta}) + o_p(1),$$

with $T^1_{n,\omega}(\hat{\theta}) = n\int |c_n(t) - c(t,\hat{\theta})|^2\omega(t;\theta_1)dt$.

Let $T^*_{3,n,\omega}(\hat{\theta}) = \|\frac{1}{\sqrt{n}}\ \sum_{j=1}^n Z_1(\hat{\varepsilon}_j;t,\hat{\theta},\psi_n)\xi_j\|^2_{\hat{\omega}}$ and

$$\Psi_{1*} = \begin{cases} 1, \text{ if } \quad T_{n,\hat{\omega}}(\hat{\theta}) \ge t^*_{3,n,\omega,\alpha}, \\ 0, \text{ otherwise,} \end{cases}$$

where $t^*_{3,n,\omega,\alpha}$ is the $1-\alpha$ percentile of the conditional distribution of $T^*_{3,n,\omega}(\hat{\theta})$. Now, proceeding as in the case where ω does not depend on the parameter θ, we state the following result.

Theorem 3.2. *Suppose that $\|\hat{\theta} - \theta_1\| = o_p(1)$, for some $\theta_1 \in \Theta$, θ_1 being the true parameter value if H_0 is true, that assumptions (A.1)–(A.8) and (3.2) hold, where both (A.2) and (A.7) hold with $\omega(t) = \omega_0(t;\theta_1)$ and $\omega(t) = \omega(t;\theta_1)$.*

(**a**) *If H_0 is true, then*

$$\sup_x \left| P_*\left\{T^*_{3,n,\omega}(\hat{\theta}) \le x\right\} - P_{\theta_1}\left\{T_{n,\hat{\omega}}(\hat{\theta}) \le x\right\}\right| \xrightarrow{P} 0.$$

(**b**) *If H_0 is not true and (3.3) holds with $\omega(t) = \omega(t;\theta_1)$, then $P(\Psi_{1*} = 1) \to 1$.*

The observation in Remark 3.1 also applies in this case.

Remark 3.6. The results stated up to now keep on being true if instead of using the raw multipliers, $\xi_1, ..., \xi_n$, we use the centered multipliers, $\xi_1 - \bar{\xi}, ..., \xi_n - \bar{\xi}$, as suggested in [2, 15], where $\bar{\xi} = \frac{1}{n}\sum_{j=1}^{n} \xi_j$.

Remark 3.7. In practice, to calculate the WB approximation to the null distribution of $T_{n,\omega}(\hat{\theta})$ (analogously for $T_{n,\hat{\omega}}(\hat{\theta})$) we proceed as follows:

1. Calculate the residuals $\hat{\varepsilon}_1, ..., \hat{\varepsilon}_n$ (or $\tilde{\varepsilon}_1, ..., \tilde{\varepsilon}_n$, if the model is homoscedastic).
2. Calculate $\hat{\theta}$ and the observed value of the test statistic $T_{n,\omega}(\hat{\theta})_{obs}$.
3. Calculate $m_{jk} = \langle Z_1(\hat{\varepsilon}_j; t, \hat{\theta}, \psi_n),\ Z_1(\hat{\varepsilon}_k; t, \hat{\theta}, \psi_n)\rangle_\omega$, $1 \le j \le k \le n$, and take $m_{jk} = m_{kj}$.
4. For some large integer B, repeat the following steps for every $b \in \{1, ..., B\}$:
   (**a**) Generate n IID variables $\xi_1, ..., \xi_n$ with mean 0 and variance 1.
   (**b**) Calculate $T^{*b}_{2,n,w}(\hat{\theta}) = \frac{1}{n}\sum_{j,k} \xi_j \xi_k m_{jk}$ (or $T^{*b}_{2,n,\omega}(\hat{\theta}) = \frac{1}{n}\sum_{j,k}(\xi_j - \bar{\xi}) \cdot (\xi_k - \bar{\xi}) m_{jk}$, as noted in Remark 3.6).
5. Approximate the p-value by $\hat{p} = \frac{1}{B}\sum_{b=1}^{B} I\{T^{*b}_{2,n,\omega}(\hat{\theta}) > T_{n,\omega}(\hat{\theta})_{obs}\}$.

4. PARAMETER ESTIMATORS

The maximum likelihood estimator (MLE) satisfies Assumption (A.1) for observable random variables. In our case, the errors are not observable. It seems reasonable to replace the errors by the residuals in the likelihood and then maximize in θ the resulting function. Specifically, assume that the CDF $F(x;\theta)$ has a Radon–Nikodym derivative $f(x;\theta)$ with respect to some σ-finite measure over $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the class of Borel sets of $\mathbb{R}$. To estimate θ we treat the residuals as it they were the true errors and consider

$$\hat{\theta}_{ML} = \arg\max_{\theta\in\Theta} \sum_{j=1}^{n} \log f(\hat{\varepsilon}_j; \theta).$$

Theorem 3.1 in Heuchenne and Van Keilegom [6] shows that (under certain conditions) $\hat{\theta}_{ML}$ satisfies (A.1) with $\psi(\varepsilon;\theta) = \psi_{ML}(\varepsilon;\theta)$ given by

$$\psi_{ML}(\varepsilon;\theta) = \rho(\varepsilon;\theta) + \varepsilon\rho_1(\theta) + \frac{\varepsilon^2 - 1}{2}\rho_2(\theta), \tag{4.1}$$

where $\rho_1(\theta) = E_\theta\{\rho'(\varepsilon;\theta)\}$, $\rho_2(\theta) = E_\theta\{\varepsilon\rho'(\varepsilon;\theta)\}$, $\rho(\varepsilon;\theta) = -A(\theta)^{-1}\nabla \log f(\varepsilon;\theta)$, $A(\theta) = (A_{rs}(\theta))$ and

$$A_{rs}(\theta) = E_\theta\left(\frac{\partial}{\partial\theta_r}\log f(\varepsilon;\theta)\frac{\partial}{\partial\theta_s}\log f(\varepsilon;\theta)\right), \quad 1 \le s, r \le p.$$

In view of (4.1), a natural choice for $\psi_n(\varepsilon;\theta)$ is $\psi_n(\varepsilon;\theta) = \psi_{n,ML}(\varepsilon;\theta)$ with

$$\psi_{n,ML}(\varepsilon;\theta) = \rho_n(\varepsilon;\theta) + \varepsilon\hat{\rho}_1(\theta) + \frac{\varepsilon^2-1}{2}\hat{\rho}_2(\theta),$$

where

$$\begin{aligned}
\rho_n(\varepsilon;\theta) &= -\hat{A}_n(\theta)^{-1}\nabla \log f(\varepsilon;\theta),\\
\hat{\rho}_1(\theta) &= \frac{1}{n}\sum_{j=1}^{n}\rho_n'(\hat{\varepsilon}_j;\theta),\\
\hat{\rho}_2(\theta) &= \frac{1}{n}\sum_{j=1}^{n}\hat{\varepsilon}_j\rho_n'(\hat{\varepsilon}_j;\theta),\\
\rho_n'(\varepsilon;\theta) &= -\hat{A}_n(\theta)^{-1}\frac{\partial}{\partial\varepsilon}\nabla \log f(\varepsilon;\theta),\\
\hat{A}_n(\theta) &= (\hat{A}_{n,rs}(\theta)),\\
\hat{A}_{n,rs}(\theta) &= \frac{1}{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta_r}\log f(\hat{\varepsilon}_j;\theta)\frac{\partial}{\partial\theta_s}\log f(\hat{\varepsilon}_j;\theta), \quad 1 \le s, r \le p.
\end{aligned}$$

The next theorem shows that $\psi_{n,ML}(\varepsilon;\theta)$ satisfies (3.2). Let $A_F(\theta) = (A_{F,rs}(\theta))$, with $A_{F,rs}(\theta) = E\left(\frac{\partial}{\partial\theta_r}\log f(\varepsilon;\theta)\,\frac{\partial}{\partial\theta_s}\log f(\varepsilon;\theta)\right)$, $1 \le s, r \le p$, $\rho_{1,F}(\theta) = E\{\rho_F'(\varepsilon;\theta)\}$, $\rho_{2,F}(\theta) = E\{\varepsilon\rho_F'(\varepsilon;\theta)\}$ and $\rho_F(\varepsilon;\theta) = -A_F(\theta)^{-1}\nabla \log f(\varepsilon;\theta)$.

Theorem 4.1. *Suppose that $\|\hat{\theta} - \theta_1\| = o_p(1)$, for some $\theta_1 \in \Theta$, θ_1 being the true parameter value if H_0 is true, and that assumptions (A.3)–(A.6), (A.9) hold, then $\psi_{n,ML}(\varepsilon;\theta)$ satisfies*

$$\frac{1}{n}\sum_{j=1}^{n}\|\psi_{n,ML}(\hat{\varepsilon}_j;\hat{\theta}) - \psi_1(\varepsilon_j;\theta_1)\|^2 \xrightarrow{P} 0,$$

with $\psi_1(\varepsilon;\theta) = \rho_F(\varepsilon;\theta) + \varepsilon\rho_{1,F}(\theta) + \frac{\varepsilon^2-1}{2}\rho_{2,F}(\theta)$.

Clearly, $\psi_1(\varepsilon_j;\theta)$ in Theorem 4.1 satisfies $\psi_1(\varepsilon_j;\theta_1) = \psi_{ML}(\varepsilon;\theta_1)$ when H_0 is true.

Remark 4.1. If model (1.1) is homoscedastic then the expressions for $\psi_{ML}(\varepsilon;\theta)$ and $\psi_{n,ML}(\varepsilon;\theta)$ simplify to $\psi_{ML}(\varepsilon;\theta) = \rho(\varepsilon;\theta) + \varepsilon\rho_1(\theta)$ and $\psi_{n,ML}(\varepsilon;\theta) = \rho_n(\varepsilon;\theta) + \varepsilon\hat{\rho}_1(\theta)$, respectively.

Another estimator that is commonly used is the method of moment estimator (MME). Although these estimators are not usually optimal, they are frequently used because their calculation is less time consuming than that of MLEs. MMEs satisfy Assumption (A.1) for observable random variables. As noticed before, in our setting the errors are not observable. Next, we study if (A.1) still holds when the errors are replaced by the residuals. Assume that, under the null hypothesis, $\theta_0 = g(\mu_0)$, for some known function $g = (g_1, ..., g_p)^T$, $g_r : \mathbb{R}^{k-1} \to \mathbb{R}$, $1 \leq r \leq p$, $\mu_0 = (\mu_{0,2}, ..., \mu_{0,k})^T$ and $\mu_{0,s} = E_{\theta_0}(\varepsilon^s)$, $\forall s$. The first moment has not been included because, by construction, it is known and equal to 0. In heteroscedastic models the second order moment is also known (thus in this case $\mu_0 = (\mu_{0,3}, ..., \mu_{0,k})^T$), but it is not in homoscedastic models (thus in this case $\mu_0 = (\mu_{0,2}, ..., \mu_{0,k})^T$). Nevertheless, we will work with $\mu_0 = (\mu_{0,2}, ..., \mu_{0,k})^T$, by implicitly understanding that in heteroscedastic models $g(\mu_{0,2}, ..., \mu_{0,k}) = g(\mu_{0,3}, ..., \mu_{0,k})$. Let $\hat{\theta}_{MM} = g(\hat{\mu})$, with $\hat{\mu} = (\hat{\mu}_2, ..., \hat{\mu}_k)^T$, $\hat{\mu}_s = \frac{1}{n}\sum_{j=1}^n \hat{\varepsilon}_j^s$, $\forall s$. The next theorem states that, under certain conditions, assumption (A.1) holds for $\hat{\theta}_{MM}$. Let $\nabla g_r(x) = \left(\frac{\partial}{\partial x_2} g_r(x), ..., \frac{\partial}{\partial x_k} g_r(x)\right)^T$, $1 \leq r \leq p$, and let $\nabla g(x)$ be the $p \times (k-1)$-matrix with rows $\nabla g_1(x)^T, ..., \nabla g_p(x)^T$, for any $x = (x_2, ..., x_k)^T \in \mathbb{R}^{k-1}$.

Theorem 4.2. *Suppose that assumptions (A.3)–(A.6) hold, that g is continuously differentiable at μ_0, that $\mu_{0,2k} < \infty$ and that H_0 is true, then*

$$\sqrt{n}(\hat{\theta}_{MM} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \psi_{MM}(\varepsilon_j; \mu_0) + o_p(1),$$

where $\psi_{MM}(\varepsilon; \mu_0) = \nabla g(\mu_0) v$, $v = (v_2, ..., v_k)^T$, $v_s = \varepsilon^s - \mu_{0,s} - \mu_{0,s-1}\varepsilon - \mu_{0,s}\frac{\varepsilon^2 - 1}{2}$, $2 \leq s \leq k$.

In the light of the result in Theorem 4.2, to approximate $\psi_{MM}(\varepsilon; \mu)$ we could replace the population moments by their empirical counterparts based on the residuals. The next theorem shows that this approximation for $\psi_{MM}(\varepsilon; \theta)$ satisfies (3.2). Let $\mu_{F,s} = E(\varepsilon^s)$ and $\mu_F = (\mu_{F,2}, ..., \mu_{F,k})^T$.

Theorem 4.3. *Suppose that assumptions (A.3)–(A.6), (A.10) hold and that $\mu_{F,2k} < \infty$, then*

$$\frac{1}{n} \sum_{j=1}^n \|\psi_{MM}(\hat{\varepsilon}_j; \hat{\mu}) - \psi_{MM}(\varepsilon_j; \mu_F)\|^2 \xrightarrow{P} 0.$$

Clearly, $\psi_{MM}(\varepsilon_j; \mu_F) = \psi_{MM}(\varepsilon_j; \mu_0)$ when H_0 is true.

Remark 4.2. If model (1.1) is homoscedastic then the expressions for $\psi_{MM}(\varepsilon; \mu)$ simplifies to $\psi_{MM}(\varepsilon; \mu_0) = \nabla g(\mu_0) v$, $v = (v_2, ..., v_k)^T$, $v_s = \varepsilon^s - \mu_{0,s} - \mu_{0,s-1}\varepsilon$, $2 \leq s \leq k$.

5. FINITE SAMPLE PERFORMANCE

With the aim of studying the finite sample performance of the proposed procedure, two simulation experiments were carried out: first, a homoscedastic regression model was considered, and then a heteroscedastic regression model. The main goal of these experiments is to compare the approximations provided by the asymptotic null distribution (when the null hypothesis is simple), the PB (as described in [11]) and the WB proposed in this paper, in three senses: closeness of the approximation under the null, the power for fixed alternatives of the resulting test and the consumed time (for the PB and the WB). This section reports and summarizes the numerical results obtained. All computations were performed using programs written in the R language [20].

In both models the hypotheses $H_0 : \varepsilon \sim N(0, \theta)$, that corresponds to testing that the error distribution is normal with CF $\exp(-0.5\theta t^2)$, and $H_0 : \varepsilon \sim \mathcal{L}(0, \theta)$, that corresponds to testing that the error distribution is Laplace with CF $\frac{1}{1+\theta t^2}$, were studied. As in Hušková and Meintanis [11], and following the recommendations in Epps and Pulley [8], the weight functions considered were: $\omega(t; \theta) = \exp(-\lambda\theta t^2)$, when testing normality, and $\omega(t; \theta) = (1 + \theta t^2)^4 \exp(-\lambda t^2)$, when testing for the Laplace distribution. For the homoscedastic model two cases were considered: θ known and θ unknown. In this second case, the parameter was estimated by a MME. Specifically, $\hat{\theta} = \frac{1}{n}\sum_{j=1}^{n} \hat{\varepsilon}_j^2$, for testing normality, and $\hat{\theta} = \frac{1}{2n}\sum_{j=1}^{n} \hat{\varepsilon}_j^2$, for the Laplace distribution. To estimate the regression function and the conditional variance, the Epanechnikov kernel $K(u) = 0.75 \times (1 - u^2)$ was employed.

As for the choice of the bandwidth, in a recent review about GOF problems in nonparametric regression, González-Manteiga and Crujeiras [9] say that the bandwidth selection for tests based on smoothing is a "really tough problem" and "it is far from being solved" (see also the discussions of Sperlich [21] and de Uña-Álvarez [3] to the mentioned article). Because of this reason, to choose h, we proceeded as in the simulation study in Pardo-Fernández *et al.* [18]: we took $h = c \times n^a$, where c and a are real constants and n is the sample size; to determine c, a and λ some preliminary simulations were performed with the purpose of finding values giving type I error close to the nominal. For all tried combinations of $c \in (1,\ 1.8)$, $a \in (-0.50,\ -0.25)$ and $\lambda \in (0.03,\ 0.54)$ good results were obtained for the WB. Here we only report the results for $c = 1.2$, $a = -0.375$ and $\lambda = 0.04$.

The error distribution were generated from: the normal distribution (denoted as N in the tables), the Laplace distribution (denoted as LP), the logistic distribution (denoted as LG), the Gumbel distribution (denoted as G), the beta distribution with parameters $a = 1$ and $b = 0.5$ (denoted as β), the chi-squared

distribution with 3 degrees of freedom (denoted as χ^2_3) and the Student t distribution with 5 degrees of freedom (denoted as t_5). All aforementioned distributions were conveniently centered and scaled to have mean 0 and variance 1.

To approximate the p-value, 1000 replications were generated for both the PB and the WB. For the WB, the raw multipliers and the centered multipliers were considered, denoted by WB1 and WB2 in the tables, respectively. The multipliers were generated from a univariate standard normal distribution. As for the asymptotic distribution (when the null hypothesis is simple, denoted as A in the tables), it is rather difficult to calculate because it coincides with that of $\sum_{j\geq 1}\lambda_j\chi^2_{1,j}$, where $\chi^2_{1,1},\chi^2_{1,2},...$ are independent chi-squared variables with one degree of freedom, the set $\{\lambda_j,\, j\geq 1\}$ are the non-null eigenvalues of the integral equation $\int C(t,s)G_j(t)dt=\lambda_j G_j(s)$, with corresponding eigenfunctions $\{G_j(\cdot),\, j\geq 1\}$, $C(t,s)$ is the covariance kernel of $Z_1(\varepsilon;t)$ (see Remarks 3.2 and 3.3 for the expression of $Z_1(\varepsilon;t)$), and determining the eigenvalues of an integral equation is tricky. Because of this reason, we approximated it by generating 10,000 samples of size 1000 obeying H_0 and calculated the test statistic at each sample, obtaining 10,000 values. The empirical CDF of these 10,000 values was taken as an approximation to the asymptotic null distribution.

1000 samples with size $n=25$ were generated from each distribution and the fractions of p-values less than or equal to 0.05 and 0.1 were calculated. The experiment was repeated for $n=50$, 100.

5.1. Homoscedastic model

The reported results correspond to the model

$$Y_j = X_j + X_j^2 + \varepsilon_j, \quad 1\leq j\leq n,$$

where X_j follows the uniform $(0,1)$ distribution. We first considered that θ is known. Since the model is homoscedastic and the null hypothesis is simple, the simplifications in Remark 3.3 can be applied. Table 1 displays the results obtained for the type I error and the power for testing normality and Table 2 for testing GOF to the Laplace distribution. Looking at these tables it can be concluded that, in terms of type I error, both the PB and the WB behave very close to the nominal levels, while the asymptotic approximation is a bit conservative, specially for testing GOF for the Laplace distribution. As for the power, the test based on the WB approximation seems to be a bit more powerful than one based on the PB. In most cases (all but alternatives β and χ^2_3 in Table 2) the WB approximation is also more powerful than one based on the asymptotic approximation.

Tables 3 and 4 show the results when θ is assumed to be unknown. In this case, the simplifications in Remark 3.1 can be applied. Looking at these tables it

can be concluded that, in terms of the type I error, as before, both the PB and the WB behave very close to the nominal levels. As for the power, for $n = 25, 50$ in some cases the WB is more powerful than the PB, but in others cases the opposite is observed; for $n = 100$ the test based on the WB approximation seems to be a bit more powerful than one based on the PB.

Table 1: (Homoscedastic model, simple null hypothesis) Percentage of rejections for the normality null hypothesis at the significance levels 5% (upper entry) and 10% (lower entry).

	$n = 25$				$n = 50$				$n = 100$			
	A	PB	WB1	WB2	A	PB	WB1	WB2	A	PB	WB1	WB2
N	3.60	6.10	4.10	6.40	4.00	5.00	4.12	4.84	5.20	4.74	4.12	4.74
	8.20	11.50	10.20	12.30	9.00	10.04	9.24	10.48	10.20	9.64	9.34	10.40
LP	25.50	36.10	57.80	64.80	45.40	56.30	86.60	88.30	76.60	77.70	98.90	99.00
	35.90	48.70	70.60	74.70	57.40	68.50	90.40	91.00	83.60	83.20	99.60	99.70
LG	10.30	57.60	56.40	63.10	12.70	88.10	87.30	89.30	17.80	99.90	100.00	100.00
	18.10	70.40	72.00	76.00	20.60	93.20	94.30	95.10	27.80	99.90	100.00	100.00
G	18.40	33.50	45.80	52.00	36.70	61.80	87.30	89.30	71.70	90.70	100.00	100.00
	30.60	46.30	62.80	67.70	49.80	74.40	94.30	95.10	81.80	96.70	100.00	100.00
β	54.10	37.50	76.20	83.10	87.50	61.20	98.40	99.00	99.70	85.30	100.00	100.00
	65.20	49.40	87.60	89.60	92.70	69.90	99.60	98.80	99.90	90.70	100.00	100.00
χ^2_3	48.60	44.20	76.50	82.50	84.60	73.40	98.40	98.60	99.90	94.50	100.00	100.00
	61.30	57.30	87.80	89.60	92.70	83.10	99.10	99.30	99.90	97.00	100.00	100.00
t_5	15.50	44.50	49.10	55.00	24.50	74.00	87.30	89.20	39.30	97.50	99.90	99.90
	25.00	59.50	63.00	67.90	35.40	84.70	93.70	94.90	51.10	99.50	100.00	100.00

Table 2: (Homoscedastic model, simple null hypothesis) Percentage of rejections for the Laplace null hypothesis at the significance levels 5% (upper entry) and 10% (lower entry).

	$n = 25$				$n = 50$				$n = 100$			
	A	PB	WB1	WB2	A	PB	WB1	WB2	A	PB	WB1	WB2
N	3.70	22.60	17.20	19.10	4.20	42.60	38.10	39.30	8.30	69.70	68.60	69.20
	8.40	30.90	25.00	27.20	9.10	51.80	48.10	50.20	14.60	77.60	78.10	78.20
LP	2.70	4.70	3.60	4.20	3.80	4.80	3.80	3.80	3.90	5.50	4.40	4.50
	7.30	9.40	7.70	8.90	8.20	10.60	8.00	9.20	8.90	9.20	9.00	9.10
LG	4.20	25.60	18.90	20.60	4.70	40.60	36.90	37.50	5.90	69.90	70.00	70.70
	7.50	35.00	28.50	31.20	9.30	48.80	46.60	47.50	11.90	78.30	78.30	79.00
G	6.00	23.30	17.70	18.80	11.60	41.60	36.60	38.20	27.10	67.10	68.20	68.90
	10.90	31.70	25.90	28.20	20.70	50.60	47.50	48.60	40.40	77.10	77.10	77.80
β	35.50	12.80	13.40	15.30	78.60	19.30	30.90	32.60	99.30	36.20	66.00	66.60
	48.80	19.20	21.80	24.30	86.20	27.70	43.20	44.50	99.60	46.60	74.00	75.60
χ^2_3	17.50	20.00	16.00	17.80	44.40	34.00	32.60	33.80	92.30	61.50	65.60	66.50
	27.20	27.60	24.00	25.70	59.10	44.00	43.70	44.60	96.90	72.00	76.20	76.70
t_5	3.20	21.50	16.20	18.10	5.40	39.30	35.00	36.70	8.70	71.60	70.80	71.40
	8.00	31.10	24.60	27.10	10.00	49.60	45.90	48.00	14.10	79.80	80.10	80.70

Table 3: (Homoscedastic model, composite null hypothesis) Percentage of rejections for the normality null hypothesis at significance levels 5% (upper entry) and 10% (lower entry).

	$n = 25$			$n = 50$			$n = 100$		
	PB	WB1	WB2	PB	WB1	WB2	PB	WB1	WB2
N	6.50	5.60	7.30	5.20	4.80	5.60	5.40	5.10	5.20
	10.70	10.90	14.50	10.00	9.90	11.10	9.20	9.20	9.60
LP	29.90	15.30	21.50	33.60	40.40	43.50	38.30	80.50	81.60
	40.50	26.20	30.60	44.10	56.30	58.70	54.70	90.40	91.00
LG	30.30	44.10	50.50	47.80	86.60	89.00	94.90	99.90	99.90
	40.30	60.60	65.80	63.90	93.80	94.60	98.50	99.99	99.99
G	29.10	18.50	21.50	35.70	42.50	43.50	51.80	80.50	83.60
	43.50	29.20	30.60	51.30	58.30	59.70	66.10	90.40	95.90
β	18.00	16.40	20.40	23.30	39.40	42.70	67.30	80.80	82.10
	25.40	27.10	30.50	32.80	55.30	56.60	72.10	89.70	91.60
χ^2_3	37.30	51.40	53.80	58.90	77.30	80.70	83.10	89.90	91.30
	48.50	63.20	64.20	67.80	85.40	87.20	91.50	97.70	98.80
t_5	40.40	14.50	21.50	52.90	38.90	42.40	76.80	82.30	83.10
	58.70	28.70	31.40	69.20	53.70	56.00	88.20	89.50	90.30

Table 4: (Homoscedastic model, composite null hypothesis) Percentage of rejections for the Laplace null hypothesis at significance levels 5% (upper entry) and 10% (lower entry).

	$n = 25$			$n = 50$			$n = 100$		
	PB	WB1	WB2	PB	WB1	WB2	PB	WB1	WB2
N	53.20	56.90	58.80	62.80	64.40	66.20	69.30	71.40	77.20
	66.30	68.20	71.10	74.50	75.40	76.60	80.60	80.90	81.20
LP	4.30	3.80	4.50	4.60	4.60	4.40	5.00	4.70	4.90
	9.20	8.30	9.20	10.30	9.30	10.40	9.50	9.80	9.50
LG	52.40	48.20	50.50	60.40	58.50	60.20	74.60	77.50	78.50
	65.30	62.00	65.70	72.10	71.70	73.90	90.80	93.20	93.70
G	52.20	47.20	50.30	50.40	51.10	58.70	63.80	65.50	66.90
	64.30	60.70	64.60	62.20	61.50	73.20	80.40	82.30	83.10
β	50.50	57.00	62.90	55.80	60.60	65.60	76.40	83.50	87.70
	63.60	71.50	76.50	72.30	74.20	77.30	87.70	95.60	98.80
χ^2_3	37.50	67.30	70.60	41.40	78.50	80.10	43.50	88.00	88.40
	51.50	79.60	82.30	54.10	91.30	93.20	59.60	97.30	98.30
t_5	33.30	42.20	44.60	38.10	44.80	44.80	44.10	51.20	52.00
	46.40	52.80	56.70	52.30	56.40	58.90	60.90	65.00	65.80

5.2. Heteroscedastic model

The reported results correspond to the model

$$Y_j = X_j + X_j^2 + (X_j + 0.5)\varepsilon_j, \quad 1 \leq j \leq n,$$

where X_j follows the uniform $(0, 1)$ distribution. Since the model is heteroscedastic and the null hypothesis is simple, the simplifications in Remark 3.2 can be applied. Table 5 displays the results obtained for the type I error and the power for testing normality and Table 6 for testing GOF to the Laplace distribution. Similar conclusions to those given for Tables 1 and 2 can be also expressed in this case.

Table 5: (Heteroscedastic model) Percentage of rejections for the normality null hypothesis at the significance levels 5% (upper entry) and 10% (lower entry).

	$n = 25$				$n = 50$				$n = 100$			
	A	PB	WB1	WB2	A	PB	WB1	WB2	A	PB	WB1	WB2
N	4.50	6.00	5.40	6.50	4.90	5.36	4.82	5.92	4.90	5.32	5.08	5.74
	10.30	10.80	10.20	12.50	10.50	10.70	10.12	11.74	9.40	10.30	10.64	11.24
LP	16.40	43.00	60.00	64.30	34.20	60.00	87.10	88.50	65.00	75.60	99.50	99.50
	23.40	54.00	70.40	73.70	44.50	71.40	92.10	92.80	73.80	82.30	99.80	99.80
LG	7.40	57.60	56.40	63.10	8.50	91.90	94.70	95.20	12.60	99.80	100.00	100.00
	12.10	70.40	72.00	76.00	15.00	95.90	97.40	98.10	20.20	99.90	100.00	100.00
G	19.40	39.10	56.40	63.10	36.90	68.10	94.70	95.20	67.20	93.60	100.00	100.00
	29.90	55.00	72.00	76.00	49.90	80.30	97.40	98.10	76.10	97.60	100.00	100.00
β	43.00	16.10	57.60	63.30	86.20	77.00	99.80	99.80	99.90	95.20	100.00	100.00
	56.20	26.10	70.00	74.70	92.30	86.20	100.00	100.00	100.00	97.60	100.00	100.00
χ^2_3	50.90	41.60	85.50	89.10	83.00	71.30	99.70	99.70	99.20	95.70	100.00	100.00
	61.80	54.80	92.50	93.90	91.00	83.10	99.90	99.90	99.70	98.80	100.00	100.00
t_5	9.20	51.00	59.10	65.40	15.90	80.20	92.90	94.30	27.90	99.00	100.00	100.00
	16.20	65.70	71.60	76.50	23.40	89.30	97.70	98.00	36.80	99.90	100.00	100.00

Table 6: (Heteroscedastic model) Percentage of rejections for the Laplace null hypothesis at the significance levels 5% (upper entry) and 10% (lower entry).

	$n = 25$				$n = 50$				$n = 100$			
	A	PB	WB1	WB2	A	PB	WB1	WB2	A	PB	WB1	WB2
N	2.00	31.80	25.00	27.10	2.60	55.30	51.20	52.50	2.80	86.10	85.70	86.20
	4.90	40.30	34.70	37.10	7.40	64.80	61.80	62.90	7.80	90.50	91.20	91.40
LP	2.10	4.60	3.70	4.60	3.00	5.70	4.00	4.40	3.60	4.40	4.00	4.40
	6.80	10.00	8.00	9.60	7.30	11.50	9.20	10.20	7.80	9.10	8.40	9.00
LG	2.10	33.80	27.10	29.30	2.30	54.80	50.80	52.30	3.10	85.00	84.40	84.70
	6.30	43.80	37.60	40.20	6.80	64.40	61.60	62.50	7.00	89.30	89.60	89.90
G	2.10	31.30	23.50	25.50	2.80	53.90	50.20	51.50	3.00	85.30	85.10	85.60
	6.70	41.10	34.40	37.10	6.80	65.10	62.70	63.70	7.50	91.10	91.10	91.50
β	3.00	19.20	18.40	21.00	6.00	33.50	43.20	45.90	27.60	56.70	81.20	81.50
	8.00	27.40	29.10	31.70	14.60	43.70	55.30	56.80	39.60	68.50	87.30	87.80
χ^2_3	2.70	22.30	18.60	20.80	3.40	43.10	42.90	44.50	5.60	78.40	81.30	81.90
	7.10	30.80	27.30	30.10	7.60	54.50	54.10	56.60	12.70	84.10	87.30	87.70
t_5	2.90	30.60	22.80	24.50	3.90	56.80	53.20	53.90	4.60	84.30	83.90	84.30
	6.30	41.50	33.70	38.00	6.50	66.70	64.30	65.20	9.40	90.20	90.40	90.70

5.3. Time consumed

Table 7 compares the PB and the WB (with raw and centered multipliers) in terms of the required CPU time. This table shows the CPU time consumed in seconds to get a p-value for testing GOF for the normal and the Laplace distributions in the homoscedastic (for both single and composite null hypothesis) and the heteroscedastic models with sample sizes $n = 25$, 50, 100, 200. Looking at this table it becomes evident that the WB is more efficient than the PB, in terms of the required computing time, specially for larger sample sizes. The difference in time when using the raw and the centered multipliers is rather small.

Table 7: CPU time consumed for the calculation of one p-value in seconds for testing normality and Laplace distribution for the homoscedastic model and composite null hypothesis (upper entry), the heteroscedastic model (middle entry) and the homoscedastic model and single null hypothesis (lower entry).

n	Normal distribution			Laplace distribution		
	PB/WB1	WB1	WB2	PB/WB1	WB1	WB2
25	2.72	0.71	0.74	3.49	1.00	1.01
	7.45	0.33	0.35	7.17	0.54	0.60
	4.42	0.31	0.34	5.34	0.50	0.55
50	5.61	0.71	0.70	7.51	1.08	1.09
	30.88	0.17	0.22	38.15	0.26	0.25
	15.63	0.19	0.19	23.68	0.28	0.25
100	12.15	0.84	0.86	23.40	1.11	1.12
	52.80	0.25	0.27	74.33	0.42	0.45
	30.64	0.25	0.26	64.56	0.37	0.39
200	27.56	1.25	1.27	76.37	1.54	1.58
	66.19	0.59	0.62	127.80	0.83	0.83
	41.14	0.56	0.58	117.51	0.78	0.76

The gain in computational efficiency of the WB over the PB stems from the fact that one does not have to re-estimate the parameters at each iteration, which slows down the process considerably. Note that in the WB the parameter θ, the regression function $m(.)$ and the conditional variance function $\sigma(\cdot)$ are estimated only one time. For the WB approximation, once the set $\{m_{jk},\ 1 \leq j \leq k \leq n\}$ is computed, the WB replicates $T^{*1}_{2,n,\omega}(\hat{\theta}), ..., T^{*B}_{2,n,\omega}(\hat{\theta})$ can be calculated very rapidly.

6. CONCLUSIONS

This paper proposes a WB approximation for the null distribution of a test statistic for testing GOF to the error distribution in nonparametric models. It provides a consistent estimator. The WB and the PB share this property. Nevertheless, from a computational point of view, the WB approximation is more efficient, in the sense of requiring less computation time. The numerical examples support these attributes. In addition, in cases were the asymptotic null distribution does not depend on unknown quantities, the simulations carried out declare that, for small to moderate sample sizes, the WB provides a better fit than the asymptotic distribution.

To derive the results in this paper we considered certain estimators for the regression function and the conditional variance function. In addition, we assumed that the covariate was univariate. The results could be extended by considering other estimators (such as other local polynomial estimators) as well as covariates with higher dimension. The null distribution of other test statistics (for example, those based on the empirical CDF) could be similarly approximated.

7. APPENDIX

7.1. Assumptions

(**A.2**) The weight function ω satisfies

$$\omega(t) = \omega(-t), \quad \forall t, \tag{7.1}$$

$\omega(t) \geq 0$, $\forall t$, and $\int t^4 \omega(t) dt < \infty$.

There is no restriction in assuming that the weight function $\omega(t)$ satisfies (7.1) because otherwise by defining $\omega_1(t) = 0.5\{\omega(t) + \omega(-t)\}$, which satisfies (7.1), we have that $T_{n,\omega}(\hat{\theta}) = T_{n,\omega_1}(\hat{\theta})$.

(**A.3**) $\varepsilon_1, ..., \varepsilon_n$ are IID with $E(\varepsilon_j^4) < \infty$ and $\varepsilon_1, ..., \varepsilon_n$ and $X_1, ..., X_n$ are independent.

Recall that by construction we have that $E(\varepsilon_j) = 0$ and $\mathrm{Var}(\varepsilon_j) = 1$.

(**A.4**) (**i**) X has a compact support S.
(**ii**) f_X, m and σ are twice continuously differentiable on S.
(**iii**) $\inf_{x \in S} f_X(x) > 0$ and $\inf_{x \in S} \sigma(x) > 0$.

(**A.5**) $nh_n^4 \to 0$, $nh_n^2/\ln n \to \infty$.

(**A.6**) K is a twice continuously differentiable symmetric pdf with compact support.

Assumptions (A.4)–(A.6) are mainly needed to guarantee the uniform consistency of the kernel estimators $\hat{f}_X(\cdot)$, $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$ for $f_X(\cdot)$, $m(\cdot)$ and $\sigma(\cdot)$, respectively.

(**A.7**) The first partial derivatives $R'(t;\theta)$, $I'(t;\theta)$, $R_{(r)}(t;\theta)$, $I_{(r)}(t;\theta)$, $1 \le r \le p$, exist and are continuous functions $\forall t \in \mathbb{R}$, $\forall \theta$ in an open neighborhood of θ_1. In addition, $R'(t;\theta)$, $I'(t;\theta)$, $R_{(r)}(t;\theta)$, $I_{(r)}(t;\theta)$, $tR'(t;\theta)$, $tI'(t;\theta)$, $tR_{(r)}(t;\theta)$, $tI_{(r)}(t;\theta)$, $1 \le r \le p$, are bounded by functions in $L_2(\omega)$, $\forall \theta$ in an open neighborhood of θ_1.

The following assumption will be used for the maximum likelihood estimator of the parameter.

(**A.9**) The following functions exist $\forall \theta$ in an open neighborhood of θ_1:

$$u_r(x;\theta) = \tfrac{\partial}{\partial \theta_r} \log f(x;\theta)\,,$$

$$u_{1,r}(x;\theta) = \tfrac{\partial^2}{\partial x \partial \theta_r} \log f(x;\theta)\,, \quad u_{0,r,s}(x;\theta) = \tfrac{\partial^2}{\partial \theta_r \partial \theta_s} \log f(x;\theta)\,,$$

$$u_{2,r}(x;\theta) = \tfrac{\partial^3}{\partial x^2 \partial \theta_r} \log f(x;\theta)\,, \quad u_{1,r,s}(x;\theta) = \tfrac{\partial^3}{\partial x \partial \theta_r \partial \theta_s} \log f(x;\theta)\,,$$

and satisfy

$$|u_{1,r}(a_1 + a_2 x;\theta)| \le b_{1,r}(x), \quad \text{with } xb_{1,r}(x),\, b_{1,r}(x) \in L_2(F)\,,$$
$$|u_{0,r,s}(a_1 + a_2 x;\theta)| \le b_{0,r,s}(x) \in L_2(F)\,,$$
$$|u_{2,r}(a_1 + a_2 x;\theta)| \le b_{2,r}(x) \in L_2(F)\,,$$
$$|u_{1,r,s}(a_1 + a_2 x;\theta)| \le b_{1,r,s}(x) \in L_2(F)\,,$$

$\forall a_1, a_2, \theta$ such that $|a_1|, |a_2 - 1|, |\theta - \theta_1| \le \delta$, for some small δ, $1 \le r, s \le p$. In addition, the following expectations exist:

$$E\big\{u_r(\varepsilon;\theta_1)\, u_s(\varepsilon;\theta_1)\big\}\,,$$
$$E\big\{\varepsilon u_{1,r}(\varepsilon;\theta_1)\big\}\,,$$

$1 \le r, s \le p$.

The following assumption will be used for the method of moment estimator of the parameter, which assumes that under the null hypothesis, $\theta_0 = g(\mu_0)$, for some known function $g = (g_1, ..., g_p)^T$, $g_r : \mathbb{R}^{k-1} \to \mathbb{R}$, $1 \le r \le p$:

(**A.10**) g_r is twice continuously differentiable at a neighborhood of μ_F, $1 \le r \le p$.

7.2. Proofs

We now sketch the proofs of the results stated in the previous sections, as well as some preliminary results. Along this section M denotes a generic positive constant taking many different values.

Lemma 7.1. *Suppose that assumptions (A.3)–(A.6) hold, then*

(a) $\frac{1}{n}\sum_{j=1}^{n}(\varepsilon_j - \hat{\varepsilon}_j)^2 = o_p(1)$.
(b) $\frac{1}{n}\sum_{j=1}^{n}(\hat{\varepsilon}_j^2 - \varepsilon_j^2)^2 = o_p(1)$.
(c) $\frac{1}{n}\sum_{j=1}^{n}(\hat{\varepsilon}_j^2 - 1)^2 = O_p(1)$.
(d) $\frac{1}{n}\sum_{j=1}^{n}\hat{\varepsilon}_j^2 = O_p(1)$.

Proof: First, observe that under the considered assumptions (see, for example, Masry [16])

$$\sup_{x\in S}|\hat{m}(x) - m(x)| \;=\; o_p(n^{-1/4}), \tag{7.2}$$

$$\sup_{x\in S}|\hat{\sigma}(x) - \sigma(x)| \;=\; o_p(n^{-1/4}). \tag{7.3}$$

The difference between the residuals and the errors can be written as follows

$$\hat{\varepsilon}_j - \varepsilon_j = \varepsilon_j\left(\frac{\sigma(X_j) - \hat{\sigma}(X_j)}{\hat{\sigma}(X_j)}\right) + \left(\frac{m(X_j) - \hat{m}(X_j)}{\hat{\sigma}(X_j)}\right). \tag{7.4}$$

The results in (a)–(d) follow from (7.2)–(7.4). □

Lemma 7.2. *If $\|\hat{\theta} - \theta_1\| = o_p(1)$ and (A.7) holds, then*

(a) $\|t\{R'(t;\hat{\theta}) - R'(t;\theta_1)\}\|_\omega^2 = o_p(1)$,
$\|t\{I'(t;\hat{\theta}) - I'(t;\theta_1)\}\|_\omega^2 = o_p(1)$.
(b) $\int\|\nabla R(t;\hat{\theta}) - \nabla R(t;\theta_1)\|^2\omega(t)dt = o_p(1)$,
$\int\|\nabla I(t;\hat{\theta}) - \nabla I(t;\theta_1)\|^2\omega(t)dt = o_p(1)$.
(c) $\|R(t;\hat{\theta}) - R(t;\theta_1)\|_\omega^2 = o_p(1)$,
$\|I(t;\hat{\theta}) - I(t;\theta_1)\|_\omega^2 = o_p(1)$.
(d) $\|t\{R(t;\hat{\theta}) - R(t;\theta_1)\}\|_\omega^2 = o_p(1)$,
$\|t\{I(t;\hat{\theta}) - I(t;\theta_1)\}\|_\omega^2 = o_p(1)$.

Proof: (a) From (A.7) $tR'(t;\theta) \in L_2(\omega)$, $\forall\theta$ in a neighborhood of θ_1. Since $\hat{\theta} \xrightarrow{P} \theta_1$, the integral $\int\{R'(t;\hat{\theta}) - R'(t;\theta_1)\}^2t^2\omega(t)dt$ is finite with probability tending to 1. Thus, $\forall\epsilon > 0, \exists M = M(\epsilon) > 0$ such that

$$\int_{\mathbb{R}\setminus[-M,M]}\{R'(t;\hat{\theta}) - R'(t;\theta_1)\}^2t^2\omega(t)dt < \epsilon, \tag{7.5}$$

with probability tending to 1. $tR'(t;\theta)$ is a uniformly continuous function in $[-M,M]\times B_\delta(\theta_1)=C$, where $B_\delta(\theta_1)=\{\theta:\|\theta-\theta_1\|\leq\delta\}$. Thus, $\forall\epsilon>0$, $\exists\ \rho=\rho(\epsilon)>0$ such that $\forall(t_a,\theta_a),(t_b,\theta_b)\in C$ satisfying $\|(t_a,\theta_a)-(t_b,\theta_b)\|<\rho$, we have $|t_1R'(t_a;\theta_a)-t_2R'(t_b;\theta_b)|<\epsilon/\iota$, with $\iota=\int\omega(t)dt$. As a consequence

$$\int_{-M}^{M}\{R'(t;\hat\theta)-R'(t;\theta_1)\}^2t^2\omega(t)dt<\epsilon, \tag{7.6}$$

with probability tending to 1. As ϵ is arbitrary, the result in (a) for the real part follows from (7.5) and (7.6). The proof for the imaginary part is parallel.

(b) The proof of this part is quite similar to that of part (a).

Parts (c) and (d) can be proven by applying the mean value theorem. □

Proof of Theorem 2.1: W^* can be expressed as $W^*=W_1+W_2+2W_3$, where $W_3^2\leq W_1W_2$, $W_1=\|\frac{1}{\sqrt{n}}\sum_{j=1}^n Z_0(\varepsilon_j;t,\theta_1)\xi_j\|_\omega^2$, $W_2=\|\frac{1}{\sqrt{n}}\sum_{j=1}^n\{Z_0(\hat\varepsilon_j;t,\hat\theta)-Z_0(\varepsilon_j;t,\theta_1)\}\xi_j\|_\omega^2$. From the results in [4],

$$\sup_x|P_*\{W_1\leq x\}-P\{W_0\leq x\}|\xrightarrow{a.s.}0.$$

Thus, to show the result it suffices to see that $W_2=o_{p_*}(1)$ in probability. With this aim, observe that W_2 can be expressed as $W_2=\sum_{j=1}^4S_j+\sum_{j\neq k}S_{jk}$, with $S_{jk}^2\leq S_jS_k$, $1\leq j,k\leq4$. In the proof of Theorem 3.1 it is given the expression of S_j and it is also proven that $S_j=o_{p_*}(1)$ in probability, $1\leq j\leq4$. This proves the result. □

Proof of Theorem 3.1: $T_{2,n,\omega}^*(\hat\theta)$ can be expressed as $T_{2,n,\omega}^*(\hat\theta)=D_1+D_2+2D_3$, where $D_1=\|\frac{1}{\sqrt{n}}\sum_{j=1}^n Z_2(\varepsilon_j;t,\theta_1)\xi_j\|_\omega^2$, $D_2=\|\frac{1}{\sqrt{n}}\sum_{j=1}^n\{Z_2(\hat\varepsilon_j;t,\hat\theta)-Z_2(\varepsilon_j;t,\theta_1)\}\xi_j\|_\omega^2$, $D_3^2\leq D_1D_2$. From the results in [4],

$$\sup_x|P_*\{D_1\leq x\}-P\{T_2\leq x\}|\xrightarrow{a.s.}0.$$

Thus, to show the result it suffices to see that $D_2=o_{p_*}(1)$ in probability. With this aim, observe that D_2 can be expressed as

$$D_2=\sum_{j=1}^{10}S_j+\sum_{k<j}S_{jk},$$

with $S_{jk}^2\leq S_jS_k$, $1\leq j,k\leq10$,

$$S_1=\|\tfrac{1}{\sqrt{n}}\textstyle\sum_{j=1}^n\{\cos(t\varepsilon_j)-\cos(t\hat\varepsilon_j)\}\xi_j\|_\omega^2,$$

$$S_2=\|\tfrac{1}{\sqrt{n}}\textstyle\sum_{j=1}^n\{\sin(t\varepsilon_j)-\sin(t\hat\varepsilon_j)\}\xi_j\|_\omega^2,$$

$$S_3=\|\tfrac{1}{\sqrt{n}}\{R(t;\hat\theta)-R(t;\theta_1)\}\left(\textstyle\sum_{j=1}^n\xi_j\right)\|_\omega^2,$$

$$S_4 = \|\tfrac{1}{\sqrt{n}}\{I(t;\hat{\theta}) - I(t;\theta_1)\}\left(\textstyle\sum_{j=1}^n \xi_j\right)\|_\omega^2,$$
$$S_5 = \|\tfrac{t}{\sqrt{n}}\textstyle\sum_{j=1}^n\{\hat{\varepsilon}_j R(t;\hat{\theta}) - \varepsilon_j R(t;\theta_1)\}\xi_j\|_\omega^2,$$
$$S_6 = \|\tfrac{t}{\sqrt{n}}\textstyle\sum_{j=1}^n\{\hat{\varepsilon}_j I(t;\hat{\theta}) - \varepsilon_j I(t;\theta_1)\}\xi_j\|_\omega^2,$$
$$S_7 = \|\tfrac{t}{2\sqrt{n}}\textstyle\sum_{j=1}^n\{(\hat{\varepsilon}_j^2 - 1)R'(t;\hat{\theta}) - (\varepsilon_j^2 - 1)R'(t;\theta_1)\}\xi_j\|_\omega^2,$$
$$S_8 = \|\tfrac{t}{2\sqrt{n}}\textstyle\sum_{j=1}^n\{(\hat{\varepsilon}_j^2 - 1)I'(t;\hat{\theta}) - (\varepsilon_j^2 - 1)I'(t;\theta_1)\}\xi_j\|_\omega^2,$$
$$S_9 = \|\tfrac{1}{\sqrt{n}}\textstyle\sum_{j=1}^n\{\psi_n^T(\hat{\varepsilon}_j;\hat{\theta})\nabla R(t;\hat{\theta}) - \psi_1^T(\varepsilon_j;\theta)\nabla R(t;\theta_1)\}\xi_j\|_\omega^2,$$
$$S_{10} = \|\tfrac{1}{\sqrt{n}}\textstyle\sum_{j=1}^n\{\psi_n^T(\hat{\varepsilon}_j;\hat{\theta})\nabla I(t;\hat{\theta}) - \psi_1^T(\varepsilon_j;\theta)\nabla I(t;\theta_1)\}\xi_j\|_\omega^2.$$

We will show that $S_j = o_{p_*}(1)$ in probability, $1 \le j \le 10$. By the mean value theorem,

$$S_1 = \frac{1}{n}\sum_{j,k=1}^n \xi_j\xi_k(\varepsilon_j - \hat{\varepsilon}_j)(\varepsilon_k - \hat{\varepsilon}_k)\int t^2 \sin(t\,\widetilde{\varepsilon}_j)\sin(t\,\widetilde{\varepsilon}_k)\omega(t)dt,$$

where $\widetilde{\varepsilon}_j = \alpha_j\varepsilon_j + (1-\alpha_j)\hat{\varepsilon}_j$, for some $\alpha_j \in (0,1)$. Then, from Lemma 7.1 (a),

$$E_*(S_1) \le \frac{1}{n}\sum_{j=1}^n(\varepsilon_j - \hat{\varepsilon}_j)^2\int t^2\omega(t)dt = o_p(1),$$

which implies $S_1 = o_{p_*}(1)$ in probability. Analogously, $S_2 = o_{p_*}(1)$ in probability.

Since $S_3 = \left(\frac{1}{\sqrt{n}}\sum_{j=1}^n \xi_j\right)^2 \|R(t;\hat{\theta}) - R(t;\theta_1)\|_\omega^2$, the central limit theorem and Lemma 7.2 (c) imply that $S_3 = o_{p_*}(1)$ in probability. Analogously, $S_4 = o_{p_*}(1)$ in probability.

Observe that $S_5 = S_{51} + S_{52} + 2S_{53}$, with $S_{53}^2 \le S_{51}S_{52}$,

$$S_{51} = \tfrac{1}{n}\textstyle\sum_{j,k=1}^n(\hat{\varepsilon}_j - \varepsilon_j)(\hat{\varepsilon}_k - \varepsilon_k)\xi_j\xi_k\|tR(t;\hat{\theta})\|_\omega^2,$$
$$S_{52} = \tfrac{1}{n}\textstyle\sum_{j,k=1}^n \varepsilon_j\varepsilon_k\xi_j\xi_k\|t\{R(t;\hat{\theta}) - R(t;\theta_1)\}\|_\omega^2.$$

From Lemma 7.1 (a) and Assumption (A.2), it follows that $E_*(S_{51}) = o_p(1)$ and thus $S_{51} = o_{p_*}(1)$, in probability. From Lemma 7.2 (d), it follows that $E_*(S_{52}) = o_p(1)$ and thus $S_{52} = o_{p_*}(1)$, in probability. Therefore, $S_5 = o_{p_*}(1)$, in probability. Analogously, $S_6 = o_{p_*}(1)$, in probability.

Observe that $S_7 = S_{71} + S_{72} + 2S_{73}$, with $S_{73}^2 \le S_{71}S_{72}$,

$$S_{71} = \tfrac{1}{4}\tfrac{1}{n}\textstyle\sum_{j,k=1}^n(\hat{\varepsilon}_j^2 - 1)(\hat{\varepsilon}_k^2 - 1)\xi_j\xi_k\|t\{R'(t;\hat{\theta}) - R'(t;\theta_1)\}\|_\omega^2,$$
$$S_{72} = \tfrac{1}{4}\tfrac{1}{n}\textstyle\sum_{j,k=1}^n(\hat{\varepsilon}_j^2 - \varepsilon_j^2)(\hat{\varepsilon}_k^2 - \varepsilon_k^2)\xi_j\xi_k\|tR'(t;\theta_1)\|_\omega^2.$$

From Lemma 7.1 (c) and Lemma 7.2 (a), it follows that $E_*(S_{71}) = o_p(1)$ and thus $S_{71} = o_{p_*}(1)$, in probability.

From Lemma 7.1 (b) and (A.7), it follows that $E_*(S_{72}) = o_p(1)$ and thus $S_{72} = o_{p_*}(1)$, in probability. Therefore, $S_7 = o_{p_*}(1)$, in probability. Analogously, $S_8 = o_{p_*}(1)$, in probability.

Observe that $S_9 = S_{91} + S_{92} + 2S_{93}$, with $S_{93}^2 \leq S_{91}S_{92}$,

$$S_{91} = \| \tfrac{1}{\sqrt{n}} \textstyle\sum_{j=1}^n \{\psi_n(\hat{\varepsilon}_j;\hat{\theta}) - \psi_1(\varepsilon_j;\theta_1)\}^T \nabla R(t;\hat{\theta})\xi_j\|_\omega^2,$$
$$S_{92} = \| \tfrac{1}{\sqrt{n}} \textstyle\sum_{j=1}^n \psi_1(\varepsilon_j;\theta_1)^T \{\nabla R(t;\hat{\theta}) - \nabla R(t;\theta_1)\}\xi_j\|_\omega^2.$$

From (3.2) and (A.7), it follows that $E_*(S_{91}) = o_p(1)$ and thus $S_{91} = o_{p_*}(1)$, in probability. From (A.1) and Lemma 7.2 (b), it follows that $E_*(S_{92}) = o_p(1)$ and thus $S_{92} = o_{p_*}(1)$, in probability. Therefore, $S_9 = o_{p_*}(1)$, in probability. Analogously, $S_{10} = o_{p_*}(1)$, in probability. This completes the proof. □

Proof of Corollary 3.2: From Theorem 3.1 it follows that $T_{2,n,\omega}^*(\hat{\theta}) = O_{p_*}(1)$ in probability. From Theorem 2 in [11], $\frac{T_{n,\omega}(\theta)}{n} \xrightarrow{P} \kappa > 0$. These two facts imply the result. □

Lemma 7.3. *Suppose that $\|\hat{\theta} - \theta_1\| = o_p(1)$, for some $\theta_1 \in \Theta$, and that assumptions (A.3)–(A.6), (A.9) hold, then*

(**a**) $\frac{1}{n}\sum_{j=1}^n \|\nabla \log f(\hat{\varepsilon}_j;\hat{\theta}) - \nabla \log f(\varepsilon_j;\theta_1)\|^2 = o_p(1)$.

(**b**) $\hat{A}_{n,rs}(\hat{\theta}) = A_{F,rs}(\theta_1) + o_p(1)$, $1 \leq r, s \leq p$.

(**c**) $\hat{\rho}_1(\hat{\theta}) = \rho_{1,F}(\theta_1) + o_p(1)$.

(**d**) $\hat{\rho}_2(\hat{\theta}) = \rho_{2,F}(\theta_1) + o_p(1)$.

Proof: (a) From the mean value theorem and (A.9),

$$\begin{aligned}
&\tfrac{1}{n}\textstyle\sum_{j=1}^n \left\{ \tfrac{\partial}{\partial\theta_r} \log f(\hat{\varepsilon}_j;\hat{\theta}) - \tfrac{\partial}{\partial\theta_r} \log f(\varepsilon_j;\theta_1) \right\}^2 \\
&\quad = \tfrac{1}{n}\textstyle\sum_{j=1}^n \left\{ \tfrac{\partial^2}{\partial\varepsilon\partial\theta_r} \log f(\tilde{\varepsilon}_j;\tilde{\theta})(\hat{\varepsilon}_j - \varepsilon_j) + \sum_{s=1}^p \tfrac{\partial^2}{\partial\theta_r\partial\theta_s} \log f(\tilde{\varepsilon}_j;\tilde{\theta})(\hat{\theta}_s - \theta_{1s}) \right\}^2 \\
&\quad \leq S_{r,1} + S_{r,2} + 2S_{r,3},
\end{aligned}$$

with $S_{r,3}^2 \leq S_{r,1}S_{r,2}$, $\tilde{\varepsilon}_j = (1-\alpha_j)\hat{\varepsilon}_j + \alpha_j\varepsilon_j$, for some $\alpha_j \in (0,1)$, $1 \leq j \leq n$, $\tilde{\theta} = (1-\alpha)\hat{\theta} + \alpha\theta_1$, for some $\alpha \in (0,1)$,

$$S_{r,1} \;=\; \|\hat{\theta} - \theta_1\|^2 \frac{1}{n}\sum_{j=1}^n \sum_{s=1}^p b_{0,r,s}^2(\varepsilon_j)$$

and

$$S_{r,2} \;=\; \frac{1}{n}\sum_{j=1}^n b_{1,r}^2(\varepsilon_j)(\hat{\varepsilon}_j - \varepsilon_j) = o_p(1).$$

From (A.9), (7.2)–(7.4), it follows that $S_{r,1} = o_p(1)$, $S_{r,2} = o_p(1)$, $1 \le r \le p$. This proves (a).

The proof of parts (b)–(d) follows similar steps to that of part (a). □

Proof of Theorem 4.1: Observe that $\frac{1}{n}\sum_{j=1}^{n} \|\psi_{1n}(\hat{\varepsilon}_j;\hat{\theta}) - \psi(\varepsilon_j;\theta_1)\|^2 \le D_1 + D_2 + D_3 + D_4$, with $D_4^2 \le \sum_{j\neq k} D_j D_k$,

$$\begin{aligned} D_1 &= \frac{1}{n}\sum_{j=1}^{n} \|\hat{A}_n(\hat{\theta})^{-1}\nabla \log f(\hat{\varepsilon}_j;\hat{\theta}) - A_F(\theta_1)^{-1}\nabla \log f(\varepsilon_j;\theta_1)\|^2, \\ D_2 &= \frac{1}{n}\sum_{j=1}^{n} \|\hat{\varepsilon}_j\hat{\rho}_1(\hat{\theta}) - \varepsilon_j\rho_{F,1}(\theta_1)\|^2, \\ D_3 &= \frac{1}{n}\sum_{j=1}^{n} \|\frac{\hat{\varepsilon}_j^2 - 1}{2}\hat{\rho}_2(\hat{\theta}) - \frac{\varepsilon_j^2 - 1}{2}\rho_{F,2}(\theta_1)\|^2. \end{aligned}$$

By using the results in Lemmas 7.1 and 7.3 one obtain $D_j = o_p(1)$, $1 \le j \le 3$, and hence the result. □

Proof of Theorem 4.2: From (7.2)–(7.4),

$$\begin{aligned} \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \hat{\varepsilon}_j^s &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \varepsilon_j^s + \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \varepsilon_j^{s-1}\frac{m(X_j) - \hat{m}(X_j)}{\hat{\sigma}(X_j)} \\ &\quad + \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \varepsilon_j^s\frac{\sigma(X_j) - \hat{\sigma}(X_j)}{\hat{\sigma}(X_j)} + o_p(1). \end{aligned} \tag{7.7}$$

Taking into account the following facts

$$\begin{aligned} &(\mathbf{m.1}) \quad \sup_{x\in S}\left|\frac{\hat{m}(x)-m(x)}{\hat{\sigma}(x)} - \frac{\hat{m}(x)-m(x)}{\sigma(x)}\right| = o_p(n^{-1/2}), \\ &(\mathbf{m.2}) \quad \sup_{x\in S}\left|\hat{m}(x) - m(x) - \frac{1}{nf_X(x)}\sum_{k=1}^{n_v} K_{h_n}(x - X_k)\sigma(X_k)\varepsilon_k\right| \\ &\qquad = o_p(n^{-1/2}), \end{aligned}$$

it follows that

$$\begin{aligned} \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \varepsilon_j^{s-1}\frac{m(X_j) - \hat{m}(X_j)}{\hat{\sigma}(X_j)} &= \\ &= \frac{-1}{n\sqrt{n}}\sum_{j,k=1}^{n} \varepsilon_j^{s-1}\varepsilon_k\frac{\sigma(X_k)}{f_X(X_j)\sigma(X_j)}K_{h_n}(X_j - X_k) + o_p(1). \end{aligned}$$

Now, by using projections, we get (see, for example, the proof of Theorem 2 in [18] for a similar development)

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n} \varepsilon_j^{s-1}\frac{m(X_j) - \hat{m}(X_j)}{\hat{\sigma}(X_j)} = -\mu_{F,s-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} \varepsilon_j + o_p(1). \tag{7.8}$$

Next we deal with the third term in the right-hand side of (7.7). Taking into account the following facts

$$\begin{aligned}
&(\mathbf{s.1}) \quad \sup_{x\in S}\left|\tfrac{\hat\sigma(x)-\sigma(x)}{\hat\sigma(x)}-\tfrac{\hat\sigma(x)-\sigma(x)}{\sigma(x)}\right| = o_p(n^{-1/2}),\\
&(\mathbf{s.2}) \quad \sup_{x\in S}\left|\hat\sigma(x)-\sigma(x)-\tfrac{\hat\sigma^2(x)-\sigma^2(x)}{2\sigma(x)}\right| = o_p(n^{-1/2}),\\
&(\mathbf{s.3}) \quad \sup_{x\in S}\Big|\hat\sigma^2(x)-\sigma^2(x)-\tfrac{1}{nf_X(x)}\textstyle\sum_{j=1}^n K_{h_n}(X_j-x)\\
&\qquad \cdot\Big[\{Y_j-m(x)\}^2-\sigma^2(x)\Big]\Big| = o_p(n^{-1/2}),
\end{aligned}$$

it follows that

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n \varepsilon_j^s\frac{\sigma(X_j)-\hat\sigma(X_j)}{\hat\sigma(X_j)} =$$

$$= \frac{1}{2n\sqrt{n}}\sum_{j,k=1}^n \varepsilon_j^s\frac{1}{f_X(X_j)\sigma^2(X_j)}K_{h_n}(X_j-X_k)\left[\sigma^2(X_j)-\{Y_k-m(X_j)\}^2\right]+o_p(1).$$

Now, by using projections, we get (see, for example, the proof of Lemma 11 in [19] for a similar development)

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n \varepsilon_j^s\frac{\sigma(X_j)-\hat\sigma(X_j)}{\hat\sigma(X_j)} = -\frac{\mu_{F,s}}{2}\frac{1}{\sqrt{n}}\sum_{j=1}^n(\varepsilon_j^2-1)+o_p(1). \tag{7.9}$$

The result follows from (7.7)–(7.9). □

Proof of Theorem 4.3: Notice that

$$\hat\mu_s-\mu_{F,s} = \frac{1}{n}\sum_{j=1}^n(\hat\varepsilon_j^s-\varepsilon_j^s)+\frac{1}{n}\sum_{j=1}^n(\varepsilon_j^s-\mu_{F,s}).$$

From (7.2)–(7.4), the first term in the right-hand side of the above equality is $o_p(1)$; from the SLLN, the second term in the right-hand side of the above equality is $o(1)$ a.s. Therefore $\hat\mu_s-\mu_{F,s}=o_p(1)$, $2\le s\le k$. The result follows from this fact and (A.10). □

ACKNOWLEDGMENTS

The authors thank the anonymous referee and the AE for their helpful suggestions and constructive comments. G.I. Rivas-Martínez acknowledges financial support from Fundación Carolina, Universidad Nacional de Asunción and Universidad de Sevilla. M.D. Jiménez-Gamero acknowledges financial support from grant MTM2014-55966-P of the Spanish Ministry of Economy and Competitiveness, and grant MTM2017-89422-P of the Spanish Ministry of Economy, Industry and Competitiveness, ERDF support included.

REFERENCES

[1] Bickel, P.J. and Doksum, K.A. (2001). *Mathematical Statistics*, Prentice Hall, New Jersey.

[2] Burke, M.D. (2000). Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap, *Statistics & Probability Letters*, **46**, 13–20.

[3] de Uña-Álvarez, J. (2013). Comments on: An updated review of Goodness-of-Fit tests for regression models, *Test*, **22**, 414–418.

[4] Dehling, H. and Mikosch, T. (1994). Random quadratic forms and the bootstrap for U-statistics, *Journal of Multivariate Analysis*, **51**, 392–413.

[5] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. 2, Wiley, New York.

[6] Heuchenne, C. and Van Keilegom, I. (2010). Goodness of fit tests for the error distribution in nonparametric regression, *Computational Statistics & Data Analysis*, **54**, 1942–1951.

[7] Epps, T.W. (2005). Tests for location-scale families based on the empirical characteristic function, *Metrika*, **62**, 99–114.

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

[9] González-Manteiga, W. and Crujeiras, R. (2013). An updated review of Goodness-of-Fit tests for regression models, *Test*, **22**, 361–411.

[10] Hušková, M. and Janssen, P. (1993). Consistency of the generalized bootstrap for degenerate U-statistics, *The Annals of Statistics*, **21**, 1811–1823.

[11] Hušková, M. and Meintanis, S.G. (2010). Test for the error distribution in nonparametric possiby heteroscedatic regression models, *Test*, **19**, 92–112.

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

[13] Jiménez-Gamero, M.D. (2013). Comments on: An updated review of Goodness-of-Fit tests for regression models, *Test*, **22**, 412–413.

[14] Jiménez-Gamero, M.D. and Kim, H-M. (2015). Fast goodness-of-fit test based on the characteristic function, *Computational Statistics & Data Analysis*, **89**, 172–191.

[15] Kojadinovic, I. and Yan, J. (2012). Goodness-of-fit testing based on a weighted bootstrap: A fast sample alternative to the parametric bootstrap, *The Canadian Journal of Statistics*, **40**, 480–500.

[16] Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates, *Journal of Time Series Analysis*, **17**, 571–600.

[17] Neumeyer, N.; Dette, H. and Nagel, E-R. (2006). Bootstrap test for the error distribution in linear and nonparametric regression models, *Australian & New Zealand Journal of Statistics*, **48**, 129–156.

[18] Pardo-Fernández, J.C.; Jiménez-Gamero, M.D. and El Ghouch, A. (2015). A nonparametric ANOVA-type test for regression curves based on characteristic functions, *Scandinavian Journal of Statistics*, **42**, 197–213.

[19] Pardo-Fernández, J.C.; Jiménez-Gamero, M.D. and El Ghouch, A. (2015). Tests for the equality of conditional variance functions in nonparametric regression, *Electronic Journal of Statistics*, **9**, 1826–1851.

[20] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, URL: `http://www.R-project.org/`, Vienna, Austria.

[21] Sperlich, S. (2013). Comments on: An updated review of Goodness-of-Fit tests for regression models, *Test*, **22**, 419–427.

[22] White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica*, **50**, 1–25.

[23] Zhu, LX. (2005). *Lecture Notes in Statistics. Nonparametric Monte Carlo test and their applications*, Springer, Berlin.

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 the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal.

In 2001, the *Revista de Estatística* published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT - STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

— The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

— All research articles will be refereed by at least two persons, one from the Editorial Board and another external.

— The only working language allowed will be English. — Four volumes are scheduled for publication, one in January, one in April, one in July and the other in October.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract and Indexing Services

The REVSTAT is covered by the following abstracting/indexing services:

- Current Index to Statistics
- Google Scholar
- Mathematical Reviews
- Science Citation Index Expanded
- Zentralblatt für Mathematic

Instructions to Authors, special-issue editors and publishers

The articles should be written in English and may be submitted in two different ways:

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt) and to one of the two Editors or Associate Editors, whose opinion the author wants to be taken into account, together to the following e-mail address: revstat@fc.ul.pt

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt), together with the corresponding PDF or PostScript file to the following e-mail address: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Manuscripts (text, tables and figures) should be typed only in black on one side, in double-spacing, with a left margin of at least 3 cm and with less than 30 pages. The first page should include the name, institution and address of the author(s) and a summary of less than one hundred words, followed by a maximum of six key words and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style. This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to PC Windows System (Zip format), Macintosh, Linux and Solaris Systems (StuffIt format), and Mackintosh System (BinHex Format), are available in the REVSTAT link of the Statistics Portugal Website: http://www.ine.pt/revstat/inicio.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: revstat@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Editorial Board