ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Volume 14, No.3
June 2016

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
    (National network) : 808 201 808
    Other networks: + 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**
  - *150 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*
- **PRICE [VAT included]**
  - *€ 9,00*

INDEX

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 14, Number 3, June 2016, 229–244

PROPERTIES OF n-LAPLACE TRANSFORM RATIO ORDER AND $\mathcal{L}^{(n)}$-CLASS OF LIFE DISTRIBUTIONS

Authors: Jalil Jarrahiferiz
– Mathematics Department, Birjand Branch,
Islamic Azad University, Birjand, Iran
jarrahi@iaubir.ac.ir

GholamReza Mohtashami Borzadaran
– Statistics Department, School of Mathematical Sciences,
Ferdowsi University of Mashhad, Mashhad, Iran
grmohtashami@um.ac.ir

Abdolhamid Rezaei Roknabadi
– Statistics Department, School of Mathematical Sciences,
Ferdowsi University of Mashhad, Mashhad, Iran
rezaei@um.ac.ir

Received: December 2012 Revised: February 2014 Accepted: November 2014

Abstract:

- One notion of stochastic comparisons of non-negative random variables based on ratios of n^{th} derivative of Laplace transforms (n-Laplace transform order or shortly $\leq_{n\text{-Lt-r}}$ order) is introduced by Mulero *et al.* (2010). In addition, they studied some of its applications in frailty models. In this paper, we have focused on some further properties of this order. In particular, we have shown that $\leq_{n\text{-Lt-r}}$ order implies dual weak likelihood ratio order ($\leq_{\text{DWLR}}$ order). Moreover, $\leq_{n\text{-Lt-r}}$ order, under certain circumstances, implies likelihood ratio order ($\leq_{\text{lr}}$ order). Finally, the $\mathcal{L}^{(n)}$ ($\bar{\mathcal{L}}^{(n)}$)-class of life distribution is proposed and studied. This class reduces to $\mathcal{L}$ ($\bar{\mathcal{L}}$)-class if we take $n = 0$.

Key-Words:

- *likelihood ratio order; hazard rate order; shock models; dual weak likelihood ratio ordering; totally positive of order 2 (*TP_2*).*

AMS Subject Classification:

- 62E10, 60E05, 60E15.

1. INTRODUCTION, DEFINITIONS AND IMPLICATIONS

There are several stochastic orders that have been introduced in the literature based on Laplace transforms. For example, Laplace transform order ($\leq_{\mathrm{Lt}}$ order) compares two random variables via their Laplace transforms. Moreover, Laplace transform ratio order ($\leq_{\mathrm{Lt\text{-}r}}$ order) and reverse Laplace transform ratio order ($\leq_{\mathrm{r\text{-}Lt\text{-}r}}$ order) which are presented based on ratios of Laplace transforms, studied by Shaked and Wong (1997). Recently, Li *et al.* (2009) introduced differentiated Laplace transform order ($\leq_{\mathrm{d\text{-}Lt\text{-}r}}$ order) which is based on ratio of derivative of Laplace transforms, and then, Mulero *et al.* (2010) generalized differentiated Laplace transform order to n-Laplace transform order. In addition, one can see Rolski and Stoyan (1976), Alzaid *et al.* (1991) and Shaked and Shanthikumar (2007) for more details. The main purpose of this article is to study the n-Laplace transform ratio order. The $\mathcal{L}$ ($\bar{\mathcal{L}}$)-class of life distributions states that $\int_0^\infty e^{-st}\bar{F}(t)\,dt \geq (\leq) \int_0^\infty e^{-st}\bar{G}(t)\,dt$, where $\bar{G}(t) = e^{-t/\mu}$, $t \geq 0$ and $\mu = \int_0^\infty \bar{F}(t)\,dt$ which was introduced by Klefsjo (1983). He presented results concerning closure properties under some of this class reliability operations, under shock models and a certain cumulative damage model. Mitra *et al.* (1995), Sengupta (1995), Bhattacharjee and Sengupta (1996), Chaudhuri *et al.* (1996), Lin (1998), Lin and Hu (2000) and Klar (2002) have studied this topic.

Here, we give some preliminaries and definitions and study some new results that are used to present our main results. Various properties and its relationships to other stochastic orders, will be described in the next section.

Throughout the paper, we assume that X and Y are absolutely continuous and non-negative random variables and use the term *increasing* in place of *non-decreasing.*

For any absolutely continuous and non-negative random variable X with density function f and survival function $\bar{F}$, the Laplace transform of f is given by $L_X(s) = \int_0^\infty e^{-st}f(t)\,dt$, $s > 0$, and the Laplace transform of $\bar{F}$ is defined as

$$L_X^*(s) = \int_0^\infty e^{-st}\bar{F}(t)\,dt\,, \qquad s > 0\,.$$

It is easy to see that $L_X(s) = 1 - s\,L_X^*(s)$. For absolutely continuous and non-negative random variable Y with density function g and survival function $\bar{G}$, $L_Y(s)$ and $L_Y^*(s)$ can be defined similar to $L_X(s)$ and $L_X^*(s)$ respectively.

Think of X as representing the length of an interval. Let this interval be subject to a poissonian marking process with intensity s. Then the Laplace transform $L_X(s)$ is the probability that there are no marks in the interval.

$$\begin{aligned} P\{X \text{ has no marks}\} &= E\big(P\{X \text{ has no marks}\}\,|\,X\big) \\ &= E\big(P\{\text{the number of events in the interval } X \text{ is } 0\,|\,X\}\big) \\ &= E\big(e^{-sX}\big) \\ &= L_X(s)\,. \end{aligned}$$

Note that

$$\begin{aligned} P\{\text{there are } n \text{ events in the interval } X\,|\,X\} &= \frac{(sX)^n}{n!}\,e^{-sX}\,, \\ P\{\text{the number of events in the interval } X \text{ is } 0\,|\,X\} &= e^{-sX}\,, \end{aligned}$$

so,

$$E\big(P\{\text{there are } n \text{ events in the interval } X\,|\,X\}\big) = (-1)^n\,\frac{s^n}{n!}\,L_X^{(n)}(s)\,,$$

and

$$E\big(P\{\text{the number of events in the interval } X \text{ is } 0\,|\,X\}\big) = L_X(s)\,.$$

Example 1.1 (Thinning of a Renewal Stream). Assume that for a random point process the lengths of the time intervals between the points which are independent and equally distributed random variables with probability density f and Laplace transform $L_X(s)$. Such a point process is called a renewal stream. The process is subject to the following thinning operation. Each point is kept with probability $1-p$ and is removed with probability p and the removal of different points are independent. We will derive the Laplace transform $L_Y^{(n)}(s)$ for the time intervals in the new stream. Let Y be the length of the time interval from a point to the next in the thinned stream and let X be the length of the time interval from the same point to the next in the original stream of points. By conditioning with respect to whether the next point is kept or removed we get, if a catastrophe risk is added as described above,

$$\begin{aligned} L_Y^{(n)}(s) &= P(n \text{ catastrophe in } X) \\ &= (1-p)\,.\,P(n \text{ catastrophe in } X) \\ &\quad +\ p\,.\,P(n \text{ catastrophe in } X)\ P(n \text{ catastrophe in } Y)\,, \end{aligned}$$

we have used the fact that if the next point is removed the process starts from scratch again. Thus we have

$$L_Y^{(n)}(s) = (1-p)\,L_X^{(n)}(s) + p\,L_X^{(n)}(s)\,L_Y^{(n)}(s)\,,$$

which gives

$$L_Y^{(n)}(s) = \frac{(1-p)\,L_X^{(n)}(s)}{1-p\,L_X^{(n)}(s)}\,.$$

If, for instance, X has an exponential distribution with parameter λ, which is the case if the stream is a Poisson process,

$$L_X^{(n)}(s) \;=\; (-1)^n \int_0^{\infty} \lambda\, x^n\, e^{-x(s+\lambda)} \;=\; (-1)^n \frac{\lambda\, n!}{(s+\lambda)^{n+1}}$$

and

$$L_Y^{(n)}(s) \;=\; \frac{(-1)^n\,(1-p)\,\lambda\, n!}{(s+\lambda)^{n+1} - (-1)^n p\,\lambda\, n!}\,.$$

For instance, if $n = 0$ then

$$L_Y(s) \;=\; \frac{(1-p)\,\lambda}{s+\lambda(1-p)}\,,$$

thus the lengths of the time intervals in the new stream have exponential distributions with parameter $\lambda(1-p)$.

Thinning of streams of points appears in many applications in operations research, in technology and in biology. For instance, consider the stream of customers arriving at a supermarket, and make this stream thinner by considering only those customers which buy a certain item.

Example 1.2 (Waiting Time for the M/G/1 System). In this system, the customers arrive according to a Poisson process with parameter λ. There is one service station and we assume the queue discipline is “first come-first served”. Let X_n be the waiting time of customer number n which the density function of X_n is denoted by f_n. We now assume that the customers on arrival at the system are marked with probability $1-s$. If the waiting time X_n of customer number n is t, the conditional probability that m marked customer arrive during this time is

$$\sum_{k=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^k}{k!}\cdot\binom{k}{m}(1-s)^m\, s^{k-m}\,,$$

so,

$$\begin{aligned}
&P(m \text{ marked customer during } X_n) \;=\\
&\qquad = \left(\frac{1-s}{s}\right)^m \sum_{k=0}^{\infty}\frac{(\lambda s)^k}{k!}\binom{k}{m}\int_0^{\infty} e^{-\lambda t}\, t^k f_n(t)\,dt\\
&\qquad = \left(\frac{1-s}{s}\right)^m \sum_{k=0}^{\infty}\frac{(\lambda s)^k}{k!}\binom{k}{m}(-1)^k L_{X_n}^{(k)}(s)\,.
\end{aligned}$$

Recall that X is said to be smaller than Y in the Laplace transform order (denoted by $X \le_{\mathrm{Lt}} Y$), if $L_X(s) \ge L_Y(s)$, $\forall\, s > 0$. Shaked and Wong (1997) established and extensively investigated stochastic orderings based on ratios of

Laplace transform. They said that X is smaller than Y in the Laplace transform ratio order (and denoted by $X \leq_{\text{Lt-r}} Y$) if $\frac{L_X(s)}{L_Y(s)}\left(\frac{1-sL_X^*(s)}{1-sL_Y^*(s)}\right)$ is increasing in $s>0$. Also, X is smaller than Y in the reverse Laplace transform ratio order (denoted by $X \leq_{\text{r-Lt-r}} Y$) if $\frac{1-L_X(s)}{1-L_Y(s)}\left(\frac{L_X^*(s)}{L_Y^*(s)}\right)$ is increasing in $s>0$. It is evident that $X \leq_{\text{Lt-r}} (\leq_{\text{r-Lt-r}})\, Y$ implies $X \leq_{\text{Lt}} Y$.

Li *et al.* (2009) introduced a new stochastic order upon Laplace transform with applications. They said that X is smaller than Y in the differentiated Laplace transform ratio order (denoted by $X \leq_{\text{d-Lt-r}} Y$) if $\frac{L'_X(s)}{L'_Y(s)}$ is increasing in $s>0$. They demonstrated that

$$X \leq_{\text{d-Lt-r}} Y \implies X \leq_{\text{Lt-r}} (\leq_{\text{r-Lt-r}})\, Y\,.$$

For two random variables X and Y with densities f and g and survival functions $\bar{F}$ and $\bar{G}$ respectively, we say that X is smaller than Y in the likelihood ratio order ($X \leq_{\text{lr}} Y$) if $\frac{g(t)}{f(t)}$ is increasing in t and say that X is smaller than Y in the hazard rate order ($X \leq_{\text{hr}} Y$) if $\frac{\bar{G}(t)}{\bar{F}(t)}$ is increasing in t. For more details of other stochastic orders one can see Shaked and Shanthikumar (2007).

Indeed, their new order has been constructed using the first derivative of the Laplace transform of density functions rather than the own Laplace transform. In order to clarify and further to determine how does the comparison affect, Mulero *et al.* (2010) considered, in general, the n^{th} derivative of the Laplace transform. As a useful observation, for example, the order based on ratios of the Laplace transform as it increases or decreases, may be important to present much information about comparison of two random variables. Moreover, as shown in the continue, for a special shock model, it is highly motivated to be considered in the case of comparison of number of shocks according to $\leq_{\text{lr}}$ order. Thus, they introduced a new partial orderings as below:

Definition 1.1. We say that X is smaller than Y in n-Laplace transform ratio (denoted by $X \leq_{n\text{-Lt-r}} Y$) if

$$\frac{L_X^{(n)}(s)}{L_Y^{(n)}(s)} \quad \text{is increasing in } s>0\,, \tag{1.1}$$

in which $n \geq 0$ is an integer and $L_X^{(n)}(s)$ denotes n^{th} derivative of $L_X(s)$ and similarly for Y.

We can define $\leq_{n\text{-Lt}^*\text{-r}}$ order for n^{th} derivative of $L_X^*(s)$ in a same manner.

Example 1.3. As pointed out in Mulero (2010), when $X_i \sim Gamma(\alpha_i, \beta_i)$, $i=1,2$, then $X_1 \leq_{n\text{-Lt-r}} X_2$ holds if for every $n \geq 0$, $\beta_1 \geq \beta_2$ and $\alpha_2 \geq \alpha_1$. It can be seen in this case that if $\alpha_1=\alpha_2=1$ and $\beta_1 \geq \beta_2$ then $X_1 \leq_{n\text{-Lt}^*\text{-r}} X_2$.

2. MAIN RESULTS

In this section, we present some results for $\leq_{n\text{-Lt-r}}$ order and then we discuss $\leq_{n\text{-Lt-r}}$ order for shock models. The same results can be obtained for $\leq_{n\text{-Lt}^*\text{-r}}$.

2.1. Basic properties

First of all, stochastic orders which have connections to $\leq_{n\text{-Lt-r}}$ order have been described.

Theorem 2.1. *Let X_1 and X_2 be absolutely continuous and non-negative iid random variables with density functions $f_1(\cdot)$ and $f_2(\cdot)$ respectively, and n be a non-negative integer. Then for any n, we have:*

- (**a**) *If $X_1 \leq_{\mathrm{lr}} X_2$ then $X_1 \leq_{n\text{-Lt-r}} X_2$.*
- (**b**) *If f_1 and f_2 are both bounded on $[0,\infty)$ then for all n, $X_1 \leq_{n\text{-Lt-r}} X_2$ implies that $X_1 \leq_{\mathrm{lr}} X_2$.*
- (**c**) *If $X_1 \leq_{n+1\text{-Lt-r}} X_2$ then $X_1 \leq_{n\text{-Lt-r}} X_2$.*

Proof: As we know that a non-negative function $h(x,y)$ is said to be TP_2 (RR_2) if

$$\begin{vmatrix} h(x_1,y_1) & h(x_1,y_2) \\ h(x_2,y_1) & h(x_2,y_2) \end{vmatrix} \geq (\leq)\ 0$$

for every $x_1 \leq x_2$ and $y_1 \leq y_2$.

(**a**) It is easy to verify, that $t^n e^{-st}$ is RR_2 in $s>0$ and in $t>0$. So, by Karlin (1968, Lemma 1.1 on p. 99) it follows that

$$\int_0^\infty t^n e^{-st} f_j(t)\, dt\,,$$

is RR_2 in $j \in \{1,2\}$ and in $s>0$, that is,

$$\frac{\int_0^\infty t^n e^{-st} f_2(t)\, dt}{\int_0^\infty t^n e^{-st} f_1(t)\, dt}\,,$$

is decreasing in $s>0$. Hence we have the result.

(**b**) Let $X_1 \leq_{n\text{-Lt-r}} X_2$, so, by Widder (1946), we have

$$\lim_{n\to\infty} L_{X_1}^{(n)}(s)\Big|_{s=\frac{n+1}{t}} = f_1(t)\,,$$

and similarly we have for Y. So, $\frac{f_2(t)}{f_1(t)}$ is increasing in $t>0$.

(**c**) Since

$$\begin{aligned} -L^{(n)}_{X_i}(s) &= \int_s^\infty L^{(n+1)}_{X_i}(t)\,dt \\ &= \int_0^\infty \left[L^{(n+1)}_{X_i}(t)\, 1_{(s,\infty)}(t) \right] dt\,, \end{aligned}$$

and $L^{(n+1)}_{X_i}(s)$ is $\mathrm{TP}_2(i,t)$ and $1_{(s,\infty)}(t)$ is $\mathrm{TP}_2(t,s)$, so, by "basic composition theorem" in Karlin (1968), $L^{(n)}_{X_i}(s)$ is $\mathrm{TP}_2(i,s)$, and thus $\frac{L^{(n)}_{X_2}(s^{-1})}{L^{(n)}_{X_1}(s^{-1})}$ is increasing in $s>0$. $\square$

Note that the inverse of the above theorem necessarily does not establish, for this case, see the following example:

Example 2.1. Let $P(X=0) = P(X=1) = \frac{1}{2}$ and $P(Y=0) = \frac{2}{4}$, $P(Y=1) = \frac{1}{4}$, $P(Y=2) = \frac{1}{4}$. It is clear that $X \le_{\mathrm{lr}} Y$ is invalid, but $X \le_{n\text{-Lt-r}} Y$ and $X \le_{n\text{-Lt}^*\text{-r}} Y$ are true.

There is no relationship between $\le_{n\text{-Lt-r}}$ and $\le_{n\text{-Lt}^*\text{-r}}$ orderings which is mentioned by Shaked and Wong (1997).

Example 2.2. Let $P(X=1) = P(X=2) = P(X=3) = \frac{1}{3}$ and $P(Y=0) = P(Y=1) = \frac{1}{4}$ and $P(Y=2) = \frac{1}{2}$. Then, $X \le_{n\text{-Lt-r}} Y$ and $X \le_{n\text{-Lt}^*\text{-r}} Y$ are not hold.

The inverse of part (c) of Theorem 2.1 is not necessarily established. This review is the following example:

Example 2.3. Let $P(X=1) = \frac{3}{4}$, $P(X=2) = P(X=3) = \frac{1}{8}$, $P(Y=1) = \frac{1}{4}$ and $P(Y=2) = \frac{3}{4}$. Then,

(**i**) for $n=1$, Li *et al.* (2009) showed that $X \le_{1\text{-Lt-r}} Y$,

(**ii**) for $n=2$, we have, $\frac{d}{ds}\frac{L^{(n)}_X(s)}{L^{(n)}_Y(s)} = \frac{1}{2}\,\frac{68\,e^{-s}-18\,e^{-2s}-108\,e^{-3s}}{(1+12\,e^{-s})}$, which for $s=0.1$ is equal to -1.4, for $s=1$ is equal to 1.59 and for $s=10$ gives 0.0015, so $X \not\le_{2\text{-Lt-r}} Y$.

The next result can be easily built and thus is presented only with proof of part (c).

Theorem 2.2.

(**a**) *Let $\{X_j\}$ and $\{Y_j\}$ be two sequences of random variables such that $X_j \to X$ and $Y_j \to Y$ in distribution. If $X_j \leq_{n\text{-Lt-r}} Y_j$, $j = 1, 2, ...$, then $X \leq_{n\text{-Lt-r}} Y$.*

(**b**) *Let X and Y be non-negative random variables with moments μ_i and ν_i, $i = 1, 2, ...$, respectively, ($\mu_0 = \nu_0 = 1$). Then, $X \leq_{n\text{-Lt-r}} Y$ if and only if*

$$\frac{\sum_{i=0}^{\infty} \frac{(-s)^i}{i!}\, \mu_{n+i}}{\sum_{i=0}^{\infty} \frac{(-s)^i}{i!}\, \nu_{n+i}} \quad \text{is increasing in } \ s > 0\,.$$

(**c**) *Let X, Y and Θ be random variables such that $[X \,|\, \Theta = \theta] \leq_{n\text{-Lt-r}} [Y \,|\, \Theta = \theta']$ for all θ and θ' in the support of Θ. Then $X \leq_{n\text{-Lt-r}} Y$.*

Proof: We only present the proof of part (**c**). The proof of parts (**a**) and (**b**) is clear.

With similar arguments to Theorem 5.B.8 of Shaked and Shanthikumar (2007) we have

$$\frac{L_X^{(n)}(s)}{L_Y^{(n)}(s)} = \frac{E_\Theta\Big(L_{[X|\Theta]}^{(n)}(s)\Big)}{E_\Theta\Big(L_{[Y|\Theta]}^{(n)}(s)\Big)}\,.$$

On the other hand $\frac{d}{ds}\frac{L_{[X|\theta]}^{(n)}(s)}{L_{[Y|\theta']}^{(n)}(s)} \geq 0$, if and only if

$$L_{[X|\theta]}^{(n+1)}(s)\, L_{[Y|\theta']}^{(n)}(s) - L_{[X|\theta]}^{(n)}(s)\, L_{[Y|\theta']}^{(n+1)}(s) \geq 0\,,$$

for all θ and θ' in the support of Θ. Consequently,

$$E_\Omega\Big(L_{[X|\theta]}^{(n+1)}(s)\, L_{[Y|\theta']}^{(n)}(s) - L_{[X|\theta]}^{(n)}(s)\, L_{[Y|\theta']}^{(n+1)}(s)\Big) \geq 0\,,$$

where $\Omega = (\theta, \theta')$, and the proof is complete. $\square$

Theorem 2.3. *Let $f(t)$ and $g(t)$ be both bounded on $[0,\infty)$. If $X \leq_{n\text{-Lt-r}} Y$, then $X \leq_{\text{DWLR}} Y$.*

Proof: If $X \leq_{n\text{-Lt-r}} Y$, then $\frac{L_X^{(n)}(s)}{L_Y^{(n)}(s)} \geq \frac{E(X^n)}{E(Y^n)}$ so,

$$\lim_{n\to\infty} \frac{L_X^{(n)}(s)}{L_Y^{(n)}(s)}\bigg|_{s=\frac{n+1}{t}} \geq \lim_{n\to\infty} \frac{E(X^n)}{E(Y^n)} = c$$

where $0 < c \leq 1$. So by Widder (1946), if $f(t)$ and $g(t)$ are both bounded on $[0, \infty)$, then we have $f(t) \geq c g(t)$, from which we conclude $X \leq_{\text{DWLR}} Y$. (Note that if $c > 1$ then $\int f(t)\, dt \geq 1$.) $\square$

2.2. Shock models

A device is subjected to shocks arriving according to a Poisson process with parameter λ. Then the lifetime T_1 of the system is given by $T_1 = \sum_{j=1}^{N_1} X_j$, where N_1 denote the number of shocks survived by the system and X_j is the random interval time between the $j-1$ and j^{th} shocks. Suppose further that the device has probability $\bar{P}_k = P(N_1 > k)$ for all $k \in N$ of surviving the first k shocks, where $1 = \bar{P}_0 \geq \bar{P}_1 \geq \dots$. Also, let $p_{k+1} = \bar{P}_k - \bar{P}_{k+1}$, $k = 0, 1, 2, \dots$, then, the probability function of the device is given by

$$f(t_1) = \sum_{k=0}^{\infty} \frac{e^{-\lambda t_1} (\lambda t_1)^k}{k!} \lambda p_{k+1} .$$

The survival function of this device is given by

$$\bar{F}(t_1) = \sum_{k=0}^{\infty} \frac{e^{-\lambda t_1} (\lambda t_1)^k}{k!} \bar{P}_k .$$

Consider another device which is also subjected to shocks arriving according to a Poisson process with the same parameter λ. Then the lifetime T_2 of the system is given by $T_2 = \sum_{j=1}^{N_2} Y_j$, where N_2 denote the number of shocks survived by the system and Y_j is the random interval time between the $j-1$ and j^{th} shocks. The device has probability $\bar{Q}_k = P(N_2 > k)$ for all $k \in N$ of surviving the first k shocks, where $1 = \bar{Q}_0 \geq \bar{Q}_1 \geq \dots$. Also, $q_{k+1} = \bar{Q}_k - \bar{Q}_{k+1}$, $k = 0, 1, 2, \dots$, then, the probability function of the device is given by

$$g(t_2) = \sum_{k=0}^{\infty} \frac{e^{-\lambda t_2} (\lambda t_2)^k}{k!} \lambda q_{k+1} .$$

The corresponding survival function of this device is given by

$$\bar{G}(t_2) = \sum_{k=0}^{\infty} \frac{e^{-\lambda t_2} (\lambda t_2)^k}{k!} \bar{Q}_k .$$

Theorem 2.4. *Let N_1, N_2, T_1 and T_2 be random variables as above. If $N_1 \leq_{\text{lr}} N_2$ then $T_1 \leq_{n\text{-Lt-r}} T_2$.*

Proof: Let us denote $L_{T_i}^{(n)}(s) = L_i^{(n)}(s),\ i = 1, 2$. We have

$$L_i^{(n)}(s) = \sum_{k=0}^{\infty} (-1)^n \frac{(n+k)!}{k!} \frac{\lambda^{k+1}}{(\lambda + s)^{n+k+1}} p_{k+1} ,$$

in which $(-1)^n \frac{(n+k)!}{k!} \frac{\lambda^{k+1}}{(\lambda+s)^{n+k+1}}$ is $\text{RR}_2(s, k)$ and p_{k+1} is $\text{TP}_2(k, i)$, so, by Karlin (1968, Lemma 1.1 on p. 99) it follows that $L_i^{(n)}(s)$ is $\text{RR}_2(s, i)$, therefore $\frac{L_2^{(n)}(s)}{L_1^{(n)}(s)}$ is decreasing in $s > 0$, or equivalently, $T_1 \leq_{n\text{-Lt-r}} T_2$. $\square$

3. $\mathcal{L}^{(n)}$-CLASS

The $\mathcal{L}$ $(\bar{\mathcal{L}})$-class of life distributions for which $\int_0^\infty e^{-st}\bar{F}(t)\,dt \geq (\leq)$ $\int_0^\infty e^{-st}\bar{G}(t)\,dt$, where $\bar{G}(t)=e^{-t/\mu}$, $t\leq 0$ and $\mu=\int_0^\infty \bar{F}(t)\,dt$ has been introduced by Klefsjo (1983). He presented results concerning closure properties under some usual reliability operations and studied some shock models and a certain cumulative damage model. The $\mathcal{L}$-class is strictly larger than the well known HNBUE class (the harmonic new better than used in expectation class of life distributions) in which a life distribution F is said to be HNBUE if $\int_t^\infty \bar{F}(x)\,dx \leq \mu\exp(-\frac{t}{\mu})$ for all $t\geq 0$. The $\mathcal{L}$-class of life distributions has attracted a great deal of attention (for more details see Lin 1998).

3.1. Basic properties of $\mathcal{L}^{(n)}$-class

We will define the class $\mathcal{L}^{(n)}$ $(\bar{\mathcal{L}}^{(n)})$-class of life distributions based on n^{th} derivative of Laplace transform in the same manner of $\mathcal{L}$ $(\bar{\mathcal{L}})$-class.

Definition 3.1. Let X be a non-negative random variable with life distribution F, survival function $\bar{F}=1-F$ and finite mean $\mu=\int_0^\infty \bar{F}(t)\,dt$. We say that the life distributions F belongs to the $\mathcal{L}^{(n)}$-class if

$$\int_0^\infty t^n e^{-st}\bar{F}(t)\,dt \;\geq\; n!\left(\frac{\mu}{1+s\mu}\right)^{n+1}, \qquad \text{for} \quad s\geq 0\,. \tag{3.1}$$

If the reversed inequality holds we shall say that F belongs to the $\bar{\mathcal{L}}^{(n)}$-class.

Theorem 3.1.

- **(a)** *If $X\in\mathcal{L}^{(n)}$-class and $X\in\bar{\mathcal{L}}^{(n)}$-class, then X has exponential distribution with mean μ.*
- **(b)** $\mathcal{L}^{(n)}\subset\mathcal{L}^{(n-1)}$ *for all $n=1,2,\ldots$.*
- **(c)** $\mathcal{L}^{(n)}=\mathcal{L}_0^{(n)}\cup\mathcal{L}_+^{(n)}$ *, in which $\mathcal{L}_+^{(n)}$ denote the class of all distributions F having support $S_F\subset(0,\infty)$, mean $\mu<\infty$, and satisfying the relation (3.1). Also, denote by $\mathcal{L}_0^{(n)}=\{F_0\}$, where F_0 is the degenerate at 0.*
- **(d)** $F\in\mathcal{L}^{(n)}$ *if and only if* $\int_0^\infty t^n e^{-st}f(t)\,dt\leq n!\,\frac{\mu^n}{(1+s\mu)^{n+1}}$.

Proof: **(a)** By assumptions

$$\int_0^\infty t^n e^{-st}\big(\bar{F}(t)-e^{-t/\mu}\big)\,dt \;=\; 0\,. \tag{3.2}$$

Due to statistical completeness property of the exponential distribution, it follows that $\bar{F}(t) = e^{-t/\mu}$, $\forall\, t \geq 0$.

(**b**) By equation (3.1), the random variable X belongs to $\mathcal{L}^{(n)}$-class if

$$\int_0^\infty t^n e^{-st}\big(\bar{F}(t) - e^{-t/\mu}\big)\, dt \;\geq\; 0\,,$$

that gives

$$\int_x^\infty \int_0^\infty t^n e^{-st}\big(\bar{F}(t) - e^{-t/\mu}\big)\, dt\, ds \;\geq\; 0\,,$$

so

$$\int_0^\infty t^{n-1}\big(\bar{F}(t) - e^{-t/\mu}\big) \int_x^\infty t\, e^{-st}\, ds\, dt \;\geq\; 0\,,$$

therefore

$$\int_0^\infty t^{n-1} e^{-xt}\big(\bar{F}(t) - e^{-t/\mu}\big)\, dt \;\geq\; 0\,,$$

which means $X \in \mathcal{L}^{(n-1)}$-class.

(**c**) Using (3.1) we conclude that for all $s \geq 0$

$$n!\left(\frac{\mu}{1+s\,\mu}\right)^{n+1} \leq \big(1 - F(0)\big) \int_0^\infty t^n e^{-st}\, dt\,,$$

from which we get

$$\left(\frac{\mu}{1+s\,\mu}\right)^{n+1} \leq \frac{1}{s^{n+1}}\big(1 - F(0)\big)\,,$$

hence, $\mu^{n+1} \leq \big(1 - F(0)\big)\big(\mu + \frac{1}{s}\big)^{n+1}$. Letting $s \to \infty$ yields $\mu^{n+1}F(0) \leq 0$, and with similar discuss to Lin (1998) obtain the result.

(**d**) Using $L_X^{(n)}(s) = n\,{L_X^*}^{(n-1)}(s) - s\,{L_X^*}^{(n)}(s)$ for all $n = 1, 2, 3, \ldots$, the desired result follows. □

Theorem 3.2. *If X has distribution function F and Y has exponential distribution with mean μ, such that $E(X^n) = E(Y^n)$, then, $Y \leq_{n\text{-Lt}^*\text{-r}} X$ implies that $X \in \mathcal{L}^{(n)}$.*

Proof: Note that $Y \leq_{n\text{-Lt}^*\text{-r}} X$ so, $\frac{{L_Y^*}^{(n)}(s)}{{L_X^*}^{(n)}(s)}$ is increasing in $s \geq 0$. Hence,

$$\frac{\displaystyle\int_0^\infty t^n e^{-t(s+\frac{1}{\mu})}\, dt}{\displaystyle\int_0^\infty t^n e^{-st}\bar{F}(t)\, dt} \;\geq\; \lim_{s\to 0} \frac{\displaystyle\int_0^\infty t^n e^{-t(s+\frac{1}{\mu})}\, dt}{\displaystyle\int_0^\infty t^n e^{-st}\bar{F}(t)\, dt} = 1\,,$$

which means X is in $\mathcal{L}^{(n)}$. □

We are going to give another interesting characterization of the $\mathcal{L}_+^{(n)}$-class through the equilibrium transformation, which will be used to estimate the moments of $F \in \mathcal{L}_+^{(n)}$ and to characterize the exponential distribution. Let X be a non-negative random variable with distribution function F and finite mean $\mu > 0$. Then the equilibrium transformation F_e of F is defined by

$$F_e(x) = \frac{1}{\mu}\int_0^x \bar{F}(t)\,dt \qquad \text{for} \quad x \geq 0\,.$$

The distribution F_e is known by the names equilibrium distribution and let the random variable with distribution function F_e is denoted by X_e.

Theorem 3.3. *Let X be a positive random variable with distribution function F and finite mean $\mu > 0$. If $X \in \mathcal{L}_+^{(n)}$ and $E(X^n e^{-sX}) \geq E(X_e^n e^{-sX_e})$ then $X \in \mathcal{L}_+^{(n-1)}$.*

Proof: Note that

$$\begin{aligned}
E(X^n e^{-sX}) \;\geq\; E(X_e^n e^{-sX_e}) \;\Longleftrightarrow& \\
\Longleftrightarrow\;\; & \frac{1}{\mu}\int_0^\infty t^n e^{-st}\bar{F}(t)\,dt \;\leq\; \int_0^\infty t^n e^{-st} f(t)\,dt \\
\Longleftrightarrow\;\; & \frac{1}{\mu}\int_0^\infty t^n e^{-st}\bar{F}(t)\,dt \;\leq\; n\int_0^\infty t^{n-1} e^{-st}\bar{F}(t)\,dt \,-\, s\int_0^\infty t^n e^{-st}\bar{F}(t)\,dt \\
\Longleftrightarrow\;\; & \frac{1+s\mu}{\mu}\int_0^\infty t^n e^{-st}\bar{F}(t)\,dt \;\leq\; n\int_0^\infty t^{n-1} e^{-st}\bar{F}(t)\,dt\,,
\end{aligned}$$

since $X \in \mathcal{L}_+^{(n)}$ then $\int_0^\infty t^{n-1}e^{-st}\bar{F}(t) \geq (n-1)!\,(\frac{\mu}{1+s\mu})^n$ that means $X \in \mathcal{L}_+^{(n-1)}$. □

Block and Savits (1980) considered

$$a_n(s) = \frac{(-1)^n}{(n)!}\,L_X^{*\,(n)}(s)\,, \qquad n = 0,1,2,\ldots\,, \quad s > 0\,,$$

and set $\alpha_{n+1}(s) = s^{n+1}a_n(s)$ for $n = 0,1,2,...,$ $s > 0$. So, $X \in \mathcal{L}^{(n)}$ if and only if

$$\alpha_{n+1}(s) \;\geq\; \left(\frac{s\,\mu}{1+s\,\mu}\right)^{n+1}.$$

Block and Savits (1980) supposed that $\{N_s(t),\ t \geq 0\}$ be a Poisson process with rate $s > 0$. They showed, if X is a random variable with survival function $\bar{F}(u)$, then

$$\begin{aligned}
\alpha_{n+1}(s) \;&=\; s\int_0^\infty \frac{e^{-su}(s\,u)^n}{n!}\,\bar{F}(u)\,du \\
&=\; s\int_0^\infty P\{N_s(u) = n\}\,\bar{F}(u)\,du \\
&=\; s\int_0^\infty P\{N_s(u) > n\}\,dF(u) \\
&=\; P\{N_s(X) > n\}\,.
\end{aligned}$$

Furthermore, if $Y_1, Y_2, ...$ are the (exponential) arrival times for the process, then

$$\alpha_{n+1}(s) \;=\; P\left(\sum_{i=1}^{n+1} Y_i \le X\right) \;=\; \int_0^\infty G^{(n+1)}(u)\, dF(u)\,, \tag{3.3}$$

where $\bar{G}(u) = \exp(-s\,u)$, $u \ge 0$. Thus (3.3) shows that the $\{\alpha_n(s),\ n \ge 1\}$ are the discrete survival probabilities for a special case of the random threshold cumulative damage model of Esary *et al.* (1973).

3.2. $\mathcal{L}^{(n)}$ ($\bar{\mathcal{L}}^{(n)}$) class for discrete life distributions

Let ξ be a strictly positive integer valued random variable and denote $\bar{P}_k = P(\xi > k)$, $k = 0, 1, 2, ...$, the corresponding survival probabilities. Also, suppose that $1 = \bar{Q}_0 \ge \bar{Q}_1 \ge \bar{Q}_2 \ge ...$ denote the corresponding survival probabilities of a geometric distribution with mean

$$\mu \;=\; \sum_{k=0}^{\infty} \bar{Q}_k \;=\; \sum_{k=0}^{\infty} \bar{P}_k\,,$$

that is,

$$\bar{Q}_k = (1 - 1/\mu)^k\,, \qquad k = 0, 1, 2, ...$$

Since the discrete counterpart to Laplace transform is the probability generating function, we consider the following natural definition:

Definition 3.2. A discrete life distribution and its survival probabilities $\bar{P}_k,\ k = 0, 1, 2, ...$, with finite mean $\sum\limits_{k=0}^{\infty} \bar{P}_k = \mu$ are in $\mathcal{L}^{(n)}$ ($\bar{\mathcal{L}}^{(n)}$) class if

$$\sum_{k=n}^{\infty} \frac{k!}{(k-n)!}\, \bar{P}_k\, p^{k-n} \ge (\le)\ \frac{n!\, \mu(\mu-1)^n}{\big(p + (1-p)\,\mu\big)^{n+1}}\,, \qquad \text{for} \quad 0 \le p \le 1\,.$$

Example 3.1. Let $\bar{P}_0 = 1,\ \bar{P}_1 = \frac{1}{2},\ \bar{P}_2 = \frac{1}{4},\ \bar{P}_3 = \frac{1}{5}$ and $\bar{P}_k = 0$ for $k = 4, 5, 6, ...$. Then $\bar{P}_k$ is belong to $\mathcal{L}$-class but is not belong to $\mathcal{L}^{(1)}$.

Preservation of $\mathcal{L}^{(n)}$ and $\bar{\mathcal{L}}^{(n)}$ classes under mixture and convolutions are studied as follow:

Mixtures:

Let $\{F_\Theta\}$ be a family of life distribution, where Θ is random variable with distribution $H(\theta)$, the mixture F of F_Θ according to H is $F(t) = \int F_\Theta(t)\, dH(\theta)$. If each F_Θ is an exponential distribution and therefore DFR, then F is DFR that follows F not in $\mathcal{L}^{(n)}$.

Consider $\bar{G}(t) = \int \bar{G}_\Theta(t)\, dH(\theta)$, $t \geq 0$, where $\bar{G}_\Theta(t) = \exp(-t/\mu_\Theta)$ in which $\mu_\Theta = \int_0^\infty \bar{F}_\Theta(t)\, dt$ and $\mu = \int_0^\infty \bar{F}(t)\, dt$. If every F_Θ is in $\bar{\mathcal{L}}^{(n)}$ so,

$$\begin{aligned}\int_0^\infty t^n e^{-st}\big(\bar{G}(t) - \bar{F}(t)\big)\, dt \;&=\; \int_0^\infty t^n e^{-st} \int_\theta \big(\bar{G}_\Theta(t) - \bar{F}_\Theta(t)\big)\, dH(\theta)\, dt \\ &=\; \int_\theta \int_0^\infty t^n e^{-st}\big(\bar{G}_\Theta(t) - \bar{F}_\Theta(t)\big)\, dt\, dH(\theta) \;\geq\; 0\,.\end{aligned}$$

Convolutions:

Let θ_1 and θ_2 be two independent random variables with life distributions F_1 and F_2 with means μ_1 and μ_2, respectively, belonging to $\mathcal{L}^{(n)}$ class. If $\bar{G}_1(t) = \exp(-t/\mu_1)$, $t \geq 0$, and $\bar{G}_2(t) = \exp(-t/\mu_2)$, $t \geq 0$, then $\theta_1 + \theta_2$, that has life distribution $F_1 * F_2$, belongs to $\mathcal{L}^{(n)}$, too. With a similar argument to Klefsjo (1983), by using properties of the Laplace transform of convolutions we get

$$\int_0^\infty t^n e^{-st}\, \overline{F_1 * F_2}(t)\, dt \;\geq\; \int_0^\infty t^n e^{-st}\, \overline{G_1 * G_2}(t)\, dt\,,$$

due to the fact $G_1 * G_2$ is IFR, it follows that $\theta_1 + \theta_2$ is in $\mathcal{L}^{(n)}$ class. Note that since $G_1 * G_2$ is IFR then it also follows that $\bar{\mathcal{L}}^{(n)}$ class is not closed under convolutions.

ACKNOWLEDGMENTS

The authors are very grateful to the referees for their valuable comments that improved the paper.

REFERENCES

[1] Alzaid, A.; Kim, J.S. and Proschan, F. (1991). Laplace ordering and its applications, *Journal of Applied Probability*, **28**, 116–130.

[2] Bhattacharjee, A. and Sengupta, D. (1996). On the coefficient of variation of the $\mathcal{L}$- and $\bar{\mathcal{L}}$-classes, *Statistics and Probability Letters*, **27**, 177–180.

[3] Block, H.W. and Savits, T.H. (1980). Laplace transforms for classes of life distributions, *Annals of Probability*, **8**, 465–474.

[4] Chaudhuri, G.; Deshpande, J. and Dharmadhikari, A. (1996). A lower bound on the $\mathcal{L}$-class of life distributions and its applications, *Calcutta Statist. Assoc. Bull.*, **46**, 269–274.

[5] Esary, J.D.; Marshall, A.W. and Proschan, F. (1973). Shock models and wear processes, *Annals of Probability*, **1**, 627–649.

[6] Karlin, S. (1968). *Total Positivity*, Stanford University Press, Stanford, CA.

[7] Klar, B. (2002). A note on the $\mathcal{L}$-class of life distributions, *journal of Applied probability*, **39**, 11–19.

[8] Klefsjo, B. (1983). A useful aging property based on the Laplace transform, *Journal of Applied Probability*, **20**, 615–620.

[9] Li, X.; Ling, X. and Li, P. (2009). A new stochastic order based upon Laplace transform with applications, *Journal of Statistical Planning and Inference*, **139**, 2624–2630.

[10] Lin, G. (1998). Characterizations of the $\mathcal{L}$-class of life distributions, *Statistics and Probability Letters*, **40**, 259–266.

[11] Lin, G. and Hu, C. (2000). A note on the $\mathcal{L}$-class of life distributions, *Sankhya*, **60**, 176–183.

[12] Mitra, M.; Basu, S. and Bhattacharjee, M. (1995). Characterizing the exponential law under Laplace order domination, *Calcutta Statist. Assoc. Bull.*, **45**, 171–178.

[13] Mulero, J.; Pellerey, F. and Rodríguez-Griñolo, R. (2010). Negative aging and stochastic comparisons of residual life times in multivariate frailty models, *Journal of Statistical Planning and Inference*, **140**, 1594–1600.

[14] Rolski, T. and Stoyan, D. (1976). On the comparison of waiting times in GI/G/1 queues, *Operations Research*, **24**, 197–199.

[15] Sengupta, D. (1995). Reliability bounds for the $\mathcal{L}$-class and Laplace order, *Journal of Applied Probability*, **32**, 832–835.

[16] Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*, Academic Press, New York.

[17] Shaked, M. and Wong, T. (1997). Stochastic orders based on ratios of Laplace transforms, *Journal of Applied Probability*, **34**, 404–419.

[18] Widder, D.V. (1946). *The Laplace Transform*, Princeton University Press.

REVSTAT – Statistical Journal
Volume 14, Number 3, June 2016, 245–271

A MIXTURE INTEGER-VALUED GARCH MODEL

Authors: Mamadou Lamine Diop
– LERSTAD, Université Gaston Berger, Saint Louis, Sénégal
diopml@yahoo.r

Aliou Diop
– LERSTAD, Université Gaston Berger, Saint Louis, Sénégal
aliou.diop@ugb.edu.sn

Abdou Kâ Diongue
– LERSTAD, Université Gaston Berger, Saint Louis, Sénégal
abdou.diongue@ugb.edu.sn

Received: July 2013 Revised: July 2014 Accepted: November 2014

Abstract:

- In this paper, we generalize the mixture integer-valued ARCH model (MINARCH) introduced by Zhu *et al.* (2010) (F. Zhu, Q. Li, D. Wang. A mixture integer-valued ARCH model, *J. Statist. Plann. Inference*, 140 (2010), 2025–2036.) to a mixture integer-valued GARCH (MINGARCH) for modeling time series of counts. This model includes the ability to take into account the moving average (MA) components of the series. We give the necessary and sufficient conditions for first and second order stationarity solutions. The estimation is done via the EM algorithm. The model selection problem is studied by using three information criterions. We also study the performance of the method via simulations and include a real data application.

Key-Words:

- *integer-valued; mixture models; GARCH; EM algorithm.*

AMS Subject Classification:

- 62F10, 62M10.

1. INTRODUCTION

Time series count data are widely observed in real-world applications (epidemiology, econometrics, insurance). Many different approaches have been proposed to model time series count data, which are able to describe different types of marginal distribution. Zeger (1988) discusses a model for regression analysis with a time series of counts by illustrating the technique with an analysis of trends in U.S. polio incidence, while Ferland *et al.* (2006) proposed an integer-valued autoregressive conditional heteroscedastic (INARCH) model to deal with integer-valued time series with overdispersion. Zhu (2011) proposed a negative binomial INGARCH (NBINGARCH) model that can deal with both overdispersion and potential extreme observations simultaneously. Zhu (2012) introduced a generalized Poisson INGARCH model, which can account for both overdispersion and underdispersion, among others.

In the literature, time series are often assumed to be driven by a unimodal innovation series. However, many time series may exhibit multimodality either in the marginal or the conditional distribution. For example, Martin (1992) proposed to model multimodal jump phenomena by a multipredictor autoregressive time series (MATS) model, while Wong and Li (2000) generalized the GMTD model to the full mixture autoregressive (MAR) model whose predictive distribution could also be multimodal. Muller and Sawitzki (1991) proposed and studied a method for analyzing the modality of a distribution.

Recently, Zhu *et al.* (2010) have used the idea of Saikkonen (2007) on the definition of a very general mixture model to generalize the INARCH model to the mixture (MINARCH) model, which has the advantages over the INARCH model because of its ability to handle multimodality and non-stationary components. But, they did not take into account the MA part of the model. Sometimes, as in the classical GARCH model, large number of lagged residuals must be included to specify the model correctly. As it is well known that computational problem may arise when the autoregressive polynomial in the conditional mean of the MINARCH model presents high order, we introduce in this paper the MINGARCH model which is a natural generalization of the MINARCH model.

The paper is organized as follows. In Section 2 we describe the MINGARCH model and the stationarity conditions. In Section 3, we discuss the estimation procedures by using an expectation-maximization (EM) algorithm introduced by Dempster *et al.* (1997) with a simulation study. We illustrate the usefulness of the model in Section 4 by an empirical example. A brief discussion and concluding remarks are given in Section 5.

2. THE MIXTURE INTEGER-VALUED GARCH MODEL

The MINGARCH(K; $p_1, ..., p_K$; $q_1, ..., q_K$) model is defined by:

$$
(2.1) \qquad \begin{cases} X_t \; = \sum_{k=1}^{K} \mathbb{1}(\eta_t = k)\, Y_{kt}\,, \\ Y_{kt}|\mathcal{F}_{t-1} : \mathcal{P}(\lambda_{kt})\,, \\ \lambda_{kt} \; = \alpha_{k0} + \sum_{i=1}^{p_k} \alpha_{ki}\, X_{t-i} + \sum_{j=1}^{q_k} \beta_{kj}\, \lambda_{k(t-j)}\,, \end{cases}
$$

where $\mathcal{P}(\lambda)$ is the Poisson distribution with parameter λ, $\alpha_{k0} > 0$, $\alpha_{ki} \geq 0$, $\beta_{kj} \geq 0$, $(i = 1, ..., p_k,\;\; j = 1, ..., q_k,\;\; k = 1, ..., K)$, $\mathbb{1}(\cdot)$ denotes the indicator function, p_k and q_k are respectively the MA and AR orders of λ_{kt}, $\mathcal{F}_{t-1}$ indicates the information given up to time $t-1$, η_t is a sequence of independent and identically distributed random variables with $\mathbb{P}(\eta_t = k) = \alpha_k$, $k = 1, ..., K$. It is assumed that X_{t-j} and η_t are independent for all t and $j > 0$, the variables Y_{kt} and η_t are conditionally independent given $\mathcal{F}_{t-1}$, $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_K$ for identifiability (see Titterington (1985)) and $\sum_{k=1}^{K} \alpha_k = 1$. If $\beta_{kj} = 0$, $k = 1, ..., K$, $j = 1, ..., q_k$, the model is denoted MINARCH(K; $p_1, ..., p_K$).

The MINGARCH model is able to handle the conditional overdispersion in integer-valued time series. In fact, the conditional mean and variance are given by

$$
\mathbb{E}\Big(X_t|\mathcal{F}_{t-1}\Big) \; = \; \sum_{k=1}^{K} \alpha_k\, \lambda_{kt}\,,
$$

and

$$
\mathrm{Var}\Big(X_t|\mathcal{F}_{t-1}\Big) \; = \; \mathbb{E}\Big(X_t|\mathcal{F}_{t-1}\Big) + \sum_{k=1}^{K} \alpha_k\, \lambda_{kt}^2 - \left(\sum_{k=1}^{K} \alpha_k\, \lambda_{kt}\right)^2.
$$

Using the Jensen's inequality, we can easily see that:

$$
\sum_{k=1}^{K} \alpha_k\, \lambda_{kt}^2 - \left(\sum_{k=1}^{K} \alpha_k\, \lambda_{kt}\right)^2 > \; 0\,.
$$

Hence the conditional variance is greater than the conditional mean. Furthermore

$$
\begin{aligned}
\mathrm{Var}\Big(X_t\Big) \; &= \; \mathbb{E}\Big(\mathrm{Var}\big(X_t|\mathcal{F}_{t-1}\big)\Big) + \mathrm{Var}\Big(\mathbb{E}\big(X_t|\mathcal{F}_{t-1}\big)\Big) \\
&= \; \mathbb{E}\left(\sum_{k=1}^{K} \alpha_k\, \lambda_{kt} + \sum_{k=1}^{K} \alpha_k\, \lambda_{kt}^2 - \left(\sum_{k=1}^{K} \alpha_k\, \lambda_{kt}\right)^2\right) + \mathrm{Var}\left(\sum_{k=1}^{K} \alpha_k\, \lambda_{kt}\right) \\
&= \; \mathbb{E}\Big(X_t\Big) + \sum_{k=1}^{K} \alpha_k\, \mathbb{E}\Big(\lambda_{kt}^2\Big) - \Big(\mathbb{E}\big(X_t\big)\Big)^2 \\
&\geq \; \mathbb{E}\Big(X_t\Big) + \mathbb{E}\left(\sum_{k=1}^{K} \alpha_k\, \lambda_{kt}^2 - \left(\sum_{k=1}^{K} \alpha_k\, \lambda_{kt}\right)^2\right).
\end{aligned}
$$

Then the variance is larger than the mean, which indicates that model (2.1) is also able to describe the time series count with overdispersion.

Let us now introduce the polynomials $D_k(B) = 1 - \beta_{k1} B - \beta_{k2} B^2 - \cdots - \beta_{kq} B^q$, $k = 1, ..., K$, where B is the backshift operator. In the following, we assume that

H_1 : For $k = 1, ..., K$, the roots of $D_k(z) = 0$ lie outside the unit circle,

H_2 : $\lambda_{kt} < \infty$ a.s. for any fixed t and k.

Let $p = \max(p_1, ..., p_K)$; $q = \max(q_1, ..., q_K)$; $\alpha_{ki} = 0$, for $i > p_k$; $\beta_{kj} = 0$, for $j > q_k$ and $L = \max(p, q)$.

First and second-order stationarity conditions for the MINGARCH model (2.1) are given in Theorem 2.1 and Theorem 2.2.

Theorem 2.1. *Assume that the conditions H_1 and H_2 hold. A necessary and sufficient condition for model (2.1) to be stationarity in the mean is that the roots of the equation:*

$$1 - \sum_{k=1}^{K} \alpha_k \left(\frac{\sum_{i=1}^{p_k} \alpha_{ki} Z^{-i}}{1 - \sum_{j=1}^{q_k} \beta_{kj} Z^{-j}} \right) = 0 \tag{2.2}$$

lie inside the unit circle.

Proof: Let $\mu_t = \mathbb{E}(X_t) = \sum_{k=1}^{K} \alpha_k \mathbb{E}(\lambda_{kt})$ for all $t \in \mathbb{Z}$. Since

$$\lambda_{kt} = \alpha_{k0} + \sum_{i=1}^{p_k} \alpha_{ki} X_{t-i} + \sum_{j=1}^{q_k} \beta_{kj} \lambda_{k(t-j)},$$

the recursion equation gives, for all $m > 1$,

$$\begin{aligned}
\lambda_{kt} = \alpha_{k_0} &+ \sum_{i=1}^{L} \alpha_{ki} X_{t-i} + \sum_{l=1}^{m} \sum_{j_1,...,j_l=1}^{L} \alpha_{k0} \beta_{kj_1} \cdots \beta_{kj_l} \\
&+ \sum_{l=1}^{m} \sum_{j_1,...,j_{l+1}=1}^{L} \alpha_{kj_{l+1}} \beta_{kj_1} \cdots \beta_{kj_l} X_{t-j_1-\cdots-j_l-j_{l+1}} \\
&+ \sum_{j_1,...,j_{m+1}=1}^{L} \beta_{kj_1} \cdots \beta_{kj_{m+1}} \lambda_{k(t-j_1-\cdots-j_{m+1})}.
\end{aligned}$$

Let $C_{k0} = \alpha_{k0} + \sum_{l=1}^{\infty} \sum_{j_1,...,j_l=1}^{L} \alpha_{k0} \beta_{kj_1} \cdots \beta_{kj_l}$. We define

$$\lambda'_{kt} = C_{k0} + \sum_{i=1}^{L} \alpha_{ki} X_{t-i} + \sum_{l=1}^{\infty} \sum_{j_1,...,j_{l+1}=1}^{L} \alpha_{kj_{l+1}} \beta_{kj_1} \cdots \beta_{kj_l} X_{t-j_1-j_2-\cdots-j_{l+1}}. \tag{2.3}$$

Since $\sum_{j=1}^{L}\beta_{kj}<1$ it is easy to see that $0\le\lambda'_{kt}<\infty$ a.s. for any fixed t and k.

We will show below that $\lambda_{kt}=\lambda'_{kt}$ almost surely as $m\to\infty$ for any fixed t and k. In what follows, C will denote any positive constant whose value is unimportant and may vary from line to line. Let t and k be fixed now. It follows that for any $m\ge 1$

$$\begin{aligned}\left|\lambda_{kt}-\lambda'_{kt}\right| \le & \sum_{l=m+1}^{\infty}\sum_{j_1,\dots,j_l=1}^{L}\alpha_{k0}\,\beta_{kj_1}\cdots\beta_{kj_l}\\ &+\sum_{l=m+1}^{\infty}\sum_{j_1,\dots,j_{l+1}=1}^{L}\alpha_{kj_{l+1}}\beta_{kj_1}\cdots\beta_{kj_l}\,X_{t-j_1-j_2-\cdots-j_{l+1}}\\ &+\sum_{j_1,\dots,j_{m+1}=1}^{L}\beta_{kj_1}\cdots\beta_{kj_{m+1}}\,\lambda_{k(t-j_1-\cdots-j_{m+1})}\,.\end{aligned}$$

Under H_2, we have

$$E\left\{\sum_{j_1,\dots,j_{l+1}=1}^{L}\alpha_{kj_{l+1}}\beta_{kj_1}\cdots\beta_{kj_l}\,X_{t-j_1-j_2-\cdots-j_{l+1}}\right\}\le C\left(\sum_{j=1}^{L}\beta_{kj}\right)^{l},$$

and

$$E\left\{\sum_{j_1,\dots,j_{m+1}=1}^{L}\beta_{kj_1}\cdots\beta_{kj_{m+1}}\,\lambda_{k(t-j_1-\cdots-j_{m+1})}\right\}\le C\left(\sum_{j=1}^{L}\beta_{kj}\right)^{m+1}.$$

The expectation of the right-hand side of the above is bounded by

$$\left(C_{k0}+C_1\left(1-\sum_{j=1}^{L}\beta_{kj}\right)^{-1}\right)\left(\sum_{j=1}^{L}\beta_{kj}\right)^{m+1}.$$

Let $A_m=\left\{|\lambda_{kt}-\lambda'_{kt}|>\frac{1}{m}\right\}$. Then

$$\mathbb{P}(A_m)\le m\left(C_{k0}+C_1\left(1-\sum_{j=1}^{L}\beta_{kj}\right)^{-1}\right)\left(\sum_{j=1}^{L}\beta_{kj}\right)^{m+1}.$$

Then, using Borel–Cantelli lemma and the fact that $A_m\subset A_{m+1}$, we can show that $\lambda_{kt}=\lambda'_{kt}$ a.s. Therefore,

$$\begin{aligned}\mu_t = &\sum_{k=1}^{K}\alpha_k\,C_{k0}+\sum_{i=1}^{L}\sum_{k=1}^{K}\alpha_k\,\alpha_{ki}\,\mu_{t-i}\\ &+\sum_{l=1}^{\infty}\sum_{j_1,\dots,j_{l+1}=1}^{L}\sum_{k=1}^{K}\alpha_k\,\alpha_{kj_{l+1}}\beta_{kj_1}\cdots\beta_{kj_l}\,\mu_{t-j_1-j_2-\cdots-j_{l+1}}\,.\end{aligned}\tag{2.4}$$

The necessary and sufficient condition for a non-homogeneous difference equation (2.4) to have a stable solution, which is finite and independent of t, is that all roots of the equation

$$1-\sum_{i=1}^{L}\sum_{k=1}^{K}\alpha_k\,\alpha_{ki}\,Z^{-i}-\sum_{l=1}^{\infty}\sum_{j_1,\ldots,j_{l+1}=1}^{L}\sum_{k=1}^{K}\alpha_k\,\alpha_{kj_{l+1}}\beta_{kj_1}\cdots\beta_{kj_l}\,Z^{-(j_1+j_2+\cdots+j_{l+1})}=0$$

lie inside the unit circle (see Goldberg (1958)). This equation is equivalent to

$$1-\sum_{k=1}^{K}\alpha_k\left(\sum_{i=1}^{p_k}\alpha_{ki}\,Z^{-i}\right)\sum_{l=0}^{\infty}\left(\sum_{j=1}^{q_k}\beta_{kj}\,Z^{-j}\right)^{l}=0\,.$$

Since $\sum_{j=1}^{q_k}\beta_{kj}<1,\ k=1,...,K$ and $\|Z\|<1$, the equation (2.2) follows. □

As an illustration, we consider in the following corollary the MINARCH(K; $p_1,...,p_K$).

Corollary 2.1. *A necessary and sufficient condition for the MINARCH(K; $p_1,...,p_K$) model to be first-order stationary is that the roots of the equation*

$$1-\sum_{i=1}^{p}\left(\sum_{k=1}^{K}\alpha_k\,\alpha_{ki}\right)Z^{-i}=0$$

lie inside the unit circle, where $p=\max(p_1,...,p_K)$.

Now, we consider the MINGARCH model with $p_k=q_k=1$ for all $k=1,...,K$. The following corollary gives a necessary and sufficient condition for the MINGARCH($K;1,...,1;1,...,1$) model to be stationary in the mean.

Corollary 2.2. *A necessary and sufficient condition for the MINGARCH($K;1,...,1;1,...,1$) model to be first-order stationarity is that the roots of the equation*

$$1+C_1Z^{-1}+C_2Z^{-2}+\cdots+C_KZ^{-K}=0$$

lie inside the unit circle where

$$C_1=-\sum_{k=1}^{K}(\delta_k+\alpha_k\gamma_k)$$

and

$$C_j=(-1)^j\left[\sum_{k_1>k_2>\ldots>k_j}^{K}\delta_{k_1}\delta_{k_2}\cdots\delta_{k_j}+\sum_{k=1}^{K}\alpha_k\gamma_k\left(\sum_{\substack{k_1>k_2>\ldots>k_{j-1}\\k_1\neq k,\,k_2\neq k,\,\ldots,\,k_{j-1}\neq k}}^{K}\delta_{k_1}\delta_{k_2}\ldots\delta_{k_{j-1}}\right)\right]$$

for $j=2,...,K$, with $\gamma_k=\alpha_{k1}$ and $\delta_k=\beta_{k1}$.

Proof: The equation (2.2) becomes

$$1-\sum_{k=1}^{K}\frac{\alpha_k\gamma_k Z^{-1}}{1-\delta_k Z^{-1}} = 0 .$$

Reducing to the same denominator, the preceding equation is equivalent to:

$$\prod_{k=1}^{K}\left(1-\delta_k Z^{-1}\right) - \sum_{k=1}^{K}\alpha_k\gamma_k Z^{-1}\prod_{\substack{k'=1\\k'\neq k}}^{K}\left(1-\delta_{k'} Z^{-1}\right) =$$
$$= 1+C_1Z^{-1}+C_2Z^{-2}+\cdots+C_K Z^{-K} = 0 . \qquad \square$$

From equation (2.4), we have

$$\mathbb{E}(X_t) = \mu = \sum_{k=1}^{K}\alpha_k C_{k0} + \mu\sum_{l=0}^{\infty}\sum_{j_1,\ldots,j_{l+1}=1}^{L}\sum_{k=1}^{K}\alpha_k\,\alpha_{kj_{l+1}}\beta_{kj_1}\cdots\beta_{kj_l} . \tag{2.5}$$

Hence

$$\mu = \frac{\sum_{k=1}^{K}\left(\frac{\alpha_k\,\alpha_{k0}}{1-\sum_{j=1}^{q_k}\beta_{kj}}\right)}{1-\sum_{k=1}^{K}\left(\frac{\sum_{i=1}^{p_k}\alpha_k\,\alpha_{ki}}{1-\sum_{j=1}^{q_k}\beta_{kj}}\right)} .$$

A necessary condition for first-order stationarity of model (2.1) is given in the following proposition.

Proposition 2.1. *Under conditions H_1 and H_2, a necessary condition for first-order stationarity of model (2.1) is*

$$\sum_{k=1}^{K}\left(\frac{\sum_{i=1}^{p_k}\alpha_k\,\alpha_{ki}}{1-\sum_{j=1}^{q_k}\beta_{kj}}\right) < 1 .$$

Remark 2.1.

1. As a special case, a necessary condition for the MINGARCH$(2;1,1;1,1)$ model to be stationary in the mean is:

$$\frac{\alpha_1\alpha_{11}}{1-\beta_{11}}+\frac{\alpha_2\,\alpha_{21}}{1-\beta_{21}} < 1 .$$

2. When the process (X_t) follows a MINARCH$(K;p_1,\ldots,p_K)$, the condition stated in Proposition 2.1, reduced $\sum_{k=1}^{K}\left(\sum_{i=1}^{p_k}\alpha_k\,\alpha_{ki}\right) < 1$ as in Zhu *et al.* (2010).

The second order stationarity condition for the MINGARCH model (2.1) in given the following theorem. Its proof is postponed in an Appendix.

Theorem 2.2. *Let $(X_t)_{t\in\mathbb{Z}}$ be a MINGARCH($K; p_1,...,p_K; q_1,...,q_K$) model. Assume that the conditions H_1 and H_2 hold. If the process $(X_t)_{t\in\mathbb{Z}}$ is first-order stationary then a necessary and sufficient condition for the process to be second-order stationary is that all roots of $1-c_1Z^{-1}-c_2Z^{-2}-\cdots-c_LZ^{-L}=0$ lie inside the unit circle, where*

$$c_u = \sum_{k=1}^{K}\alpha_k\left(\Delta_{k,u}-\sum_{v=1}^{L-1}\Lambda_{kv}\,b_{vu}\,\omega_{u0}\right), \quad u=1,...,L-1 \ \text{ and }\ c_L=\sum_{k=1}^{K}\alpha_k\,\Delta_{k,L}\,,$$

with

$$\begin{aligned}
\Delta_{k,i} &= \Delta^{(1)}_{k,i}+\Delta^{(2)}_{k,i}\,,\\
\Delta^{(1)}_{k,i} &= \sum_{l=0}^{\infty}\ \sum_{\substack{j_{l+2}=i\\ j_{l+2}=j_1+\cdots+j_{l+1}}}^{L}\alpha_{kj_{l+1}}\alpha_{kj_{l+2}}\,\beta_{kj_1}\cdots\beta_{kj_l}\,,\\
\Delta^{(2)}_{k,i} &= \sum_{\substack{l=0\\ l'=0}}^{\infty}\ \sum_{\substack{j_1+\cdots+j_{l+2}=i\\ j_1+\cdots+j_{l+2}=j'_1+\cdots+j'_{l'+1}}}^{L}\alpha_{kj_{l+2}}\,\beta_{kj_1}\cdots\beta_{kj_{l+1}}\,\alpha_{kj'_{l'+1}}\beta_{kj'_1}\cdots\beta_{kj'_{l'}}\,,\\
\Lambda_{kv} &= \Lambda^{(1)}_{kv}+\Lambda^{(2)}_{kv}\,,\\
\Lambda^{(1)}_{kv} &= \sum_{l=0}^{\infty}\ \sum_{|j_{l+2}-j_1-\cdots-j_{l+1}|=v}^{L}\alpha_{kj_{l+1}}\alpha_{kj_{l+2}}\,\beta_{kj_1}\cdots\beta_{kj_l}\,,\\
\Lambda^{(2)}_{kv} &= \sum_{\substack{l=0\\ l'=0}}^{\infty}\ \sum_{\left|j_1+\cdots+j_{l+2}-j'_1-\cdots-j'_{l'+1}\right|=v}^{L}\alpha_{kj_{l+2}}\,\beta_{kj_1}\cdots\beta_{kj_{l+1}}\,\alpha_{kj'_{l'+1}}\beta_{kj'_1}\cdots\beta_{kj'_{l'}}\,,
\end{aligned}$$

and $\Gamma=(\omega_{ij})_{i,j=1}^{L-1}$, $\Gamma^{-1}=(b_{ij})_{i,j=1}^{L-1}$, two matrices such that

$$\omega_{i0}=\sum_{l=0}^{\infty}\sum_{k=1}^{K}\alpha_k\,\delta_{i0kl}\,,\quad \omega_{iu}=\sum_{l=0}^{\infty}\sum_{k=1}^{K}\alpha_k\,\delta_{iukl}\ \text{ for } u\neq i\,,\quad \omega_{ii}=\sum_{l=0}^{\infty}\sum_{k=1}^{K}\alpha_k\,\delta_{iikl}-1\,,$$

$$\delta_{iukl} = \sum_{|i-j_1-\cdots-j_{l+1}|=u}\alpha_{kj_{l+1}}\,\beta_{kj_1}\cdots\beta_{kj_l}\,.$$

We remark that when (X_t) follows a MINARCH($K; p_1,...,p_K$), Theorem 2.2 reduces to Theorem 2 of Zhu *et al.* (2010), where $L=\max(p_1,...,p_K)$.

If the process (X_t) following a MINGARCH($K; p_1,...,p_K; q_1,...,q_K$) model is second-order stationary, then from (5.2), we have

$$\mathbb{E}(X_t^2) = \frac{c_0}{1-\sum_{u=1}^{L}c_u}\,,$$

where $c_0 > 0$ (see Appendix B). Hence, necessary second order stationary condition for a special case is given by the following proposition.

Proposition 2.2. *The second order stationary condition for a MINGARCH$(K; 1, ..., 1; 1, ..., 1)$ is $c_1 < 1$ where $c_1 = \sum_{k=1}^{K} \alpha_k \, \alpha_{k1}^2$.*

In the following theorem, we give a necessary and sufficient condition for the process (X_t) following a MINGARCH$(K; 1, ..., 1; 1, ..., 1)$ model to be m order stationary. The results for the general model MINGARCH$(K; p_1, ..., p_K; q_1, ..., q_K)$ are difficult to obtain and need further investigations.

Theorem 2.3. *The m-th moment of a MINGARCH$(K; 1, ..., 1; 1, ..., 1)$ model is finite if and only if*

$$\sum_{k=1}^{K} \alpha_k \left(\alpha_{k1} + \beta_{k1}\right)^m < 1 \,. \tag{2.6}$$

Proof: Since $Y_{kt}|\mathcal{F}_{t-1}$ is a Poisson random variable with mean $\lambda_{kt} = \alpha_{k0} + \alpha_{k1} X_{t-1} + \beta_{k1} \lambda_{k(t-1)}$ conditionally to time $t-1$, the m-th moment of X_t is given by

$$\mathbb{E}\left(X_t^m\right) \;=\; \sum_{k=1}^{K} \alpha_k \sum_{j=0}^{m} \left\{ {m \atop j} \right\} \mathbb{E}\big(\lambda_{kt}^j\big)$$

where $\left\{ {m \atop j} \right\}$ is the Stirling number of the second kind (see Gradshteyn and Ryzhik (2007), p. 1046) and

$$\lambda_{kt}^j \;=\; \sum_{n=0}^{j} \binom{j}{n} \alpha_{k0}^{j-n} \sum_{r=0}^{n} \binom{n}{r} \left(\alpha_{k1} X_{t-1}\right)^r \left(\beta_{k1} \lambda_{k(t-1)}\right)^{n-r}.$$

We mimic the proof of Proposition 6 in Ferland *et al.* (2006) by setting

$$\mathbf{\Lambda}_{k,t} \;=\; \left(\lambda_{kt}^m, ..., \lambda_{kt}^2, \lambda_{kt}\right)^T$$

and showing that for all k

$$\mathbb{E}\left(\mathbf{\Lambda}_{k,t}|\mathcal{F}_{t-2}\right) \;=\; \mathbf{d}_k + \mathbf{D}_k \mathbf{\Lambda}_{k,t-1}$$

where $\mathbf{d}_k$ and $\mathbf{D}_k$ are respectively a constant vector and an upper triangular matrix. The derivation of the required condition follows the great lines of the proof of Proposition 6 in Ferland *et al.* (2006). □

The result obtained in Theorem (2.6) is an extension of Proposition 6 in Ferland *et al.* (2006) for an INGARCH$(1, 1)$ process. When the β_{ki}'s equal to zero, the (necessary) condition (2.6) is a special case of the result obtained in Theorem 3 of Zhu *et al.* (2010).

3. PARAMETER ESTIMATION AND SIMULATION

3.1. Estimation procedure

In this section, we discuss the estimation of the parameters of a MINGARCH model by using the expectation-maximization (EM) algorithm (see Dempster *et al.* (1997)). Suppose that the observation $X = (X_1, ..., X_n)$ is generated from the MINGARCH model.

Let $Z = (Z_1, ..., Z_n)$ be the random variable where $Z_t = (Z_{1,t}, ..., Z_{K,t})^T$ is a vector whose components are defined by:

$$Z_{i,t} = \begin{cases} 1 & \text{if } X_t \text{ comes from the } i\text{-th component; } 1 \le i \le K\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

The vectors Z_t are not observed and its distribution is:

$$\mathbb{P}\big(Z_t = (1, 0, ..., 0)^T\big) = \alpha_1\,, \quad ..., \quad \mathbb{P}\big(Z_t = (0, 0, ..., 0, 1)^T\big) = \alpha_K\,.$$

Let $\alpha = (\alpha_1, ..., \alpha_{K-1})^T$, $\alpha_{(k)} = (\alpha_{k0}, \alpha_{k1}, ..., \alpha_{kp_k})^T$, $\beta_{(k)} = (\beta_{k1}, ..., \beta_{kq_k})^T$ $\theta_{(k)} = (\alpha^T_{(k)}, \beta^T_{(k)})$ and $\theta = (\alpha, \theta_{(1)}, ..., \theta_{(K)})^T \in \Theta$ (the parameter space).

Given Z_t, the distribution of the complete data (X_t, Z_t) is then given by

$$\prod_{k=1}^{K} \left(\alpha_k \frac{\lambda_{kt}^{X_t} \exp(-\lambda_{kt})}{X_t!}\right)^{Z_{kt}}.$$

Let l_t be the conditional log-likelihood function at time t. The log-likelihood is given by $l(\theta) = \sum_{t=1}^{n} l_t$.

$l(\theta)$ is the joint log-likelihood function of the first L random variables of the series and $l^*(\theta) = \sum_{t=L+1}^{n} l_t$ is called the conditional log-likelihood function. When the sample size n is large, the influence of $\sum_{t=1}^{L} l_t$ will be negligible. In this study, the parameters will be estimated by maximizing the conditional log-likelihood function l^* given by

$$l^*(\theta) = \sum_{t=L+1}^{n} \left\{ \sum_{k=1}^{K} Z_{kt} \log(\alpha_k) + X_t \sum_{k=1}^{K} Z_{kt} \log(\lambda_{kt}) - \sum_{k=1}^{K} Z_{kt}\lambda_{kt} - \log(X_t!) \right\}.$$

The first derivatives of the conditional log-likelihood with respect to θ are:

$$\text{(3.1)} \qquad \frac{\partial l*}{\partial \alpha_k} = \sum_{t=L+1}^{n} \left(\frac{Z_{kt}}{\alpha_k} - \frac{Z_{Kt}}{\alpha_K}\right), \qquad k = 1, ..., K-1\,,$$

$$(3.2) \qquad \frac{\partial l*}{\partial \alpha_{ki}} = \sum_{t=L+1}^{n} Z_{kt} \frac{X_t - \lambda_{kt}}{\lambda_{kt}} U(X_t, i)\,, \qquad k = 1, ..., K, \quad i = 0, ..., p_k\,,$$

$$(3.3) \qquad \frac{\partial l^*}{\partial \beta_{kj}} = \sum_{t=L+1}^{n} Z_{kt} \frac{X_t - \lambda_{kt}}{\lambda_{kt}} \lambda_{k,t-j}\,, \qquad k = 1, ..., K, \quad j = 1, ..., q_k\,,$$

where $U(X_t, i) = 1$ for $i = 0$ and $U(X_t, i) = X_{t-i}$ for $i > 0$.

Given that the process $\{Z_t\}$ is not observed, the data that we have do not allow the estimation of the parameter θ. An iterative procedure (EM) is proposed for estimating the parameters by maximizing the conditional log-likelihood function $l^*(\theta)$, which consists of two steps (E and M) that we describe in the following.

E-step:

Suppose that θ is known. The missing data $\mathbf{Z}$ are then replaced by their conditional expectations, conditional on the parameters and on the observed data X. In this case the conditional expectation of the k-th component of Z_t is just the conditional probability that the observation X_t comes from the k-th component of the mixture distribution conditional on θ and X. Let $\tau_{k,t}$ be the conditional expectation of Z_{kt}.
Then the E-step equation is given by:

$$\tau_{k,t} = \frac{\alpha_k {\lambda_{kt}}^{X_t} \exp(-\lambda_{kt})}{\sum_{i=1}^{K} \alpha_i {\lambda_{it}}^{X_t} \exp(-\lambda_{it})}$$

where $k = 1, 2, ..., K$ and $t = L+1, ..., n$. In practice, we take $Z_{kt} = \tau_{k,t}$ from the previous E-step of the EM procedure.

M-step:

The missing data Z are replaced by their conditional expectations on the parameters θ and on the observed data $X_1, ..., X_n$. The estimates of the parameters θ can then be obtained by maximizing the conditional log-likelihood function $l^*(\theta)$ by equating expressions (3.2)–(3.3) to 0.
The M-step equations become

$$\hat{\alpha}_k = \frac{1}{n-L} \sum_{t=L+1}^{n} \tau_{k,t}\,, \qquad k = 1, ..., K\,.$$

From the equation (3.2), we have:

$$\sum_{t=L+1}^{n} \frac{\tau_{t,k} X_t}{\hat{\lambda}_{kt}} U(X_t, i) = \sum_{t=L+1}^{n} \tau_{k,t}\, U(X_t, i)\,.$$

Then

$$\sum_{t=L+1}^{n}\left\{\frac{\tau_{k,t}\,X_t}{\sum_{j=0}^{p_k}\hat{\alpha}_{kj}\,U(X_t,j)+\sum_{j=1}^{q_k}\hat{\beta}_{kj}\hat{\lambda}_{k(t-j)}}\,U(X_t,i)\right\}=\sum_{t=L+1}^{n}\tau_{k,t}\,U(X_t,i)\,,$$

for $k=1,...,K, \quad i=0,...,p_k$.
Similarly equation (3.3) gives:

$$\sum_{t=L+1}^{n}\frac{\tau_{k,t}\,X_t}{\hat{\lambda}_{kt}}\,\hat{\lambda}_{k,t-j}\;=\sum_{t=L+1}^{n}\tau_{k,t}^{(s)}\,\hat{\lambda}_{k,t-j}^{(s)}\,.$$

Then

$$\sum_{t=L+1}^{n}\left\{\frac{\tau_{k,t}^{(s)}X_t}{\sum_{i=0}^{p_k}\hat{\alpha}_{ki}^{(s)}\,U(X_t,i)+\sum_{t=L+1}^{q_k}\hat{\beta}_{ki}^{(s)}\,\hat{\lambda}_{k,t-i}^{(s)}}\,\hat{\lambda}_{k,t-j}^{(s)}\right\}=\sum_{t=L+1}^{n}\tau_{k,t}^{(s)}\,\hat{\lambda}_{k,t-j}^{(s)}\,,$$

for $k=1,...,K,\; j=1,...,q_k$.

The estimate of θ is then obtained by iterating these two steps until convergence. The criterion used for checking convergence of the EM procedure is

$$\max\left\{\left|\frac{\theta_i^{(s+1)}-\theta_i^{(s)}}{\theta_i^{(s)}}\right|,\; s,i\geq 1\right\}\;\leq\;10^{-5}$$

where $\theta_i^{(s)}$ is the i-th component of θ obtained in the s-th iteration.

Among different strategies for choosing starting initial values for the EM algorithm (see Karlis and Xekalaki (2003), Melnykov and Melnykov (2012)), the random initialization method is employed in this paper (the initial values for $\theta_{(k)}$ are chosen randomly from a uniform distribution and the mixing proportions are generated from a Dirichlet distribution). The asymptotic properties are not treated in this paper but they have been studied by many authors. For example, Nityasuddhia and Böhning (2003) have studied the asymptotic properties of the EM algorithm estimate for normal mixture models. They show that the EM algorithm gives reasonable solutions of the score equations in an asymptotic unbiased sense. The performance of the EM algorithm is assessed by some simulation experiments.

3.2. Simulation studies

Monte Carlo experiment was conducted to investigate the performance of the EM estimation method. In all these simulation experiments, we use 100 independent realizations of the MINGARCH model defined in (2.1) with sizes $n = 100$, $n = 200$ and $n = 500$. The following two models were used in the experiment. The first, denoted Model (I), is a MINGARCH(2; 1,1; 1,1) model with parameter values

$$\begin{pmatrix} \alpha_1 & \alpha_{10} & \alpha_{11} & \beta_{11} \\ \alpha_2 & \alpha_{20} & \alpha_{21} & \beta_{21} \end{pmatrix} = \begin{pmatrix} 0.75 & 1.00 & 0.20 & 0.30 \\ 0.25 & 5.00 & 0.50 & 0.30 \end{pmatrix}.$$

The second, denoted Model (II), is a MINGARCH(3; 1,1,1; 1,1,1) model with parameter values

$$\begin{pmatrix} \alpha_1 & \alpha_{10} & \alpha_{11} & \beta_{11} \\ \alpha_2 & \alpha_{20} & \alpha_{21} & \beta_{21} \\ \alpha_3 & \alpha_{30} & \alpha_{31} & \beta_{31} \end{pmatrix} = \begin{pmatrix} 0.55 & 0.80 & 0.40 & 0.30 \\ 0.25 & 1.00 & 0.50 & 0.25 \\ 0.20 & 0.50 & 0.60 & 0.20 \end{pmatrix}.$$

The performances of the estimators are evaluated by the root mean square error (RMSE) and the mean absolute error (MAE).

Based on the results in Tables 1 and 2, we can see that as the sample size increases, the estimates seem to converge to the true parameter values.

Table 1: Results of the simulation study with model (I).

Sample size	k		α_k	α_{k0}	α_{k1}	β_{k1}
100	1	True values	0.7500	1.0000	0.2000	0.3000
		Mean estimated	0.7410	1.1883	0.1833	0.2446
		RMSE	0.0523	0.5789	0.0623	0.2137
		MAE	0.0405	0.4726	0.0506	0.1801
	2	True values	0.2500	5.0000	0.5000	0.3000
		Mean estimated	0.2590	5.1660	0.4619	0.2901
		RMSE	0.0523	2.6410	0.2823	0.2588
		MAE	0.0405	2.2060	0.2103	0.2274
200	1	True values	0.7500	1.0000	0.2000	0.3000
		Mean estimated	0.7463	1.0093	0.1909	0.3054
		RMSE	0.0359	0.4429	0.0468	0.1773
		MAE	0.0291	0.3641	0.0381	0.1460
	2	True values	0.2500	5.0000	0.5000	0.3000
		Mean estimated	0.2537	5.2571	0.4612	0.2928
		RMSE	0.0359	2.2616	0.1728	0.2380
		MAE	0.0291	1.8728	0.1314	0.1976
500	1	True values	0.7500	1.0000	0.2000	0.3000
		Mean estimated	0.7510	1.0646	0.1959	0.2817
		RMSE	0.0259	0.2525	0.0272	0.1035
		MAE	0.0212	0.1867	0.0214	0.0783
	2	True values	0.2500	5.0000	0.5000	0.3000
		Mean estimated	0.2490	5.3064	0.5026	0.2688
		RMSE	0.0259	1.6316	0.0982	0.1702
		MAE	0.0212	1.3483	0.0774	0.1443

Table 2: Results of the simulation study with model (II).

Sample size	k		α_k	α_{k0}	α_{k1}	β_{k1}
100	1	True values	0.5500	0.8000	0.4000	0.3000
		Mean estimated	0.5435	0.7671	0.4429	0.2163
		RMSE	0.1063	0.4997	0.1898	0.2339
		MAE	0.0828	0.4054	0.1482	0.1977
	2	True values	0.2500	1.0000	0.5000	0.2500
		Mean estimated	0.2240	1.0888	0.5344	0.2532
		RMSE	0.0802	0.7182	0.3804	0.2563
		MAE	0.0607	0.5504	0.2420	0.2113
	3	True values	0.2000	0.5000	0.6000	0.2000
		Mean estimated	0.2323	0.9516	0.4475	0.2714
		RMSE	0.0600	0.7127	0.2413	0.2263
		MAE	0.0429	0.5490	0.1895	0.1850
200	1	True values	0.5500	0.8000	0.4000	0.3000
		Mean estimated	0.5286	0.7471	0.4113	0.2552
		RMSE	0.1117	0.4363	0.1563	0.1942
		MAE	0.0838	0.3566	0.1190	0.1545
	2	True values	0.2500	1.0000	0.5000	0.2500
		Mean estimated	0.2316	1.0570	0.5340	0.2433
		RMSE	0.0785	0.6025	0.2584	0.1928
		MAE	0.0602	0.4787	0.1751	0.1506
	3	True values	0.2000	0.5000	0.6000	0.2000
		Mean estimated	0.2397	0.8867	0.4450	0.3042
		RMSE	0.0652	0.6088	0.2306	0.2439
		MAE	0.0452	0.4959	0.1806	0.1825
500	1	True values	0.5500	0.8000	0.4000	0.3000
		Mean estimated	0.5556	0.7040	0.4248	0.2725
		RMSE	0.0825	0.3246	0.1171	0.1797
		MAE	0.0614	0.2595	0.0934	0.1407
	2	True values	0.2500	1.0000	0.5000	0.2500
		Mean estimated	0.2182	0.9508	0.5223	0.2656
		RMSE	0.0620	0.4723	0.2059	0.2132
		MAE	0.0487	0.3853	0.1569	0.1576
	3	True values	0.2000	0.5000	0.6000	0.2000
		Mean estimated	0.2261	0.8985	0.4690	0.2815
		RMSE	0.0536	0.5883	0.1963	0.1988
		MAE	0.0298	0.4780	0.1586	0.1506

The performance of the estimate improves when the sample size increases. But this performance varies depending on the parameters. Indeed the parameter estimate α_k seems to give good results for all sample sizes considered. For the parameter α_{k0}, the RMSE and the MAE are slightly higher.

4. REAL DATA EXAMPLE

In this section we investigate the time series representing a count of the calls monthly reported in the 22nd police car beat in Pittsburg, starting in January 1990 and ending in December 2001. The data are available online at the forecasting principles site (`http://www.forecastingprinciples.com`), in the section about crime data. The summary statistics are given in Table 3. Mean and variance are estimated as 6.3056 and 23.0249, respectively. Hence the data seem to be overdispersed. The histogram of the series in Figure 1 shows that the series seems to be bimodal. Using the bimodality index of Der and Everitt (2002), Zhu *et al.* (2010) show that the series is bimodal. Moreover, they found that the MINARCH model is more appropriate for this dataset than the INARCH model. The autocorrelation function in Figure 2 implies that the moving average polynomial order is at most equal to three (i.e. $0 \leq q \leq 3$) while when considering the partial autocorrelation function, we can choose p such that $1 \leq p \leq 3$. Thus, in the following, we consider a MINGARCH model (2.1) with $K = 1, 2, 3$.

Table 3: Summary statistics of the crime counts series.

Sample size	Minimum	Maximum	Median	Mean	Variance	Skewness	Kurtosis
144	0	30	5	6.3056	23.0249	1.9732	8.7530

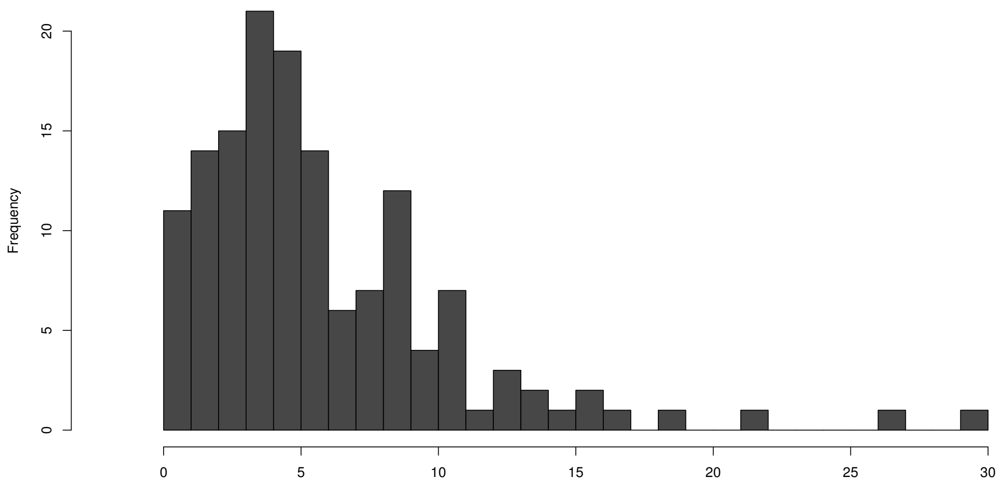

Figure 1: Histogram of the crime counts series.

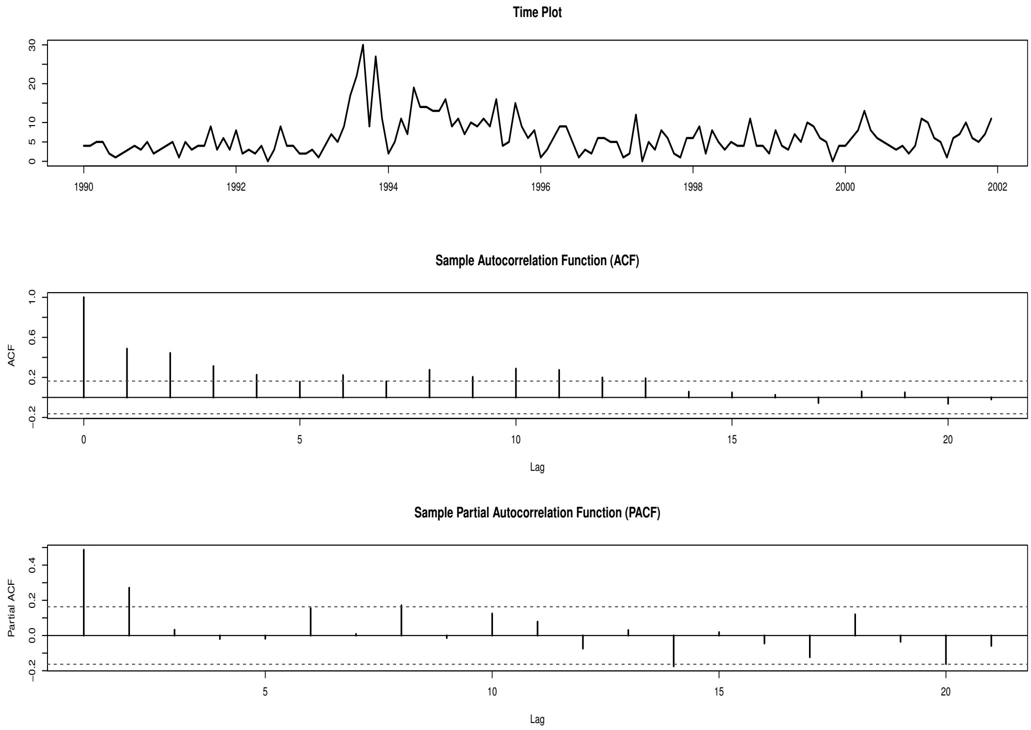

Figure 2: Crime counts series: the time plot, the sample autocorrelation and partial autocorrelation function.

The model selection criteria considered here are the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the mixture regression criterion (MRC) proposed by Naik *et al.* (2007). These two first criteria are both defined as minus twice the maximized log-likelihood plus a penalty term. The first choice is the maximum log-likelihood given by the EM estimation, it includes the information of the unobserved random variable $\mathbf{Z}$. The second choice is computed from the (conditional) probability density function of the MINGARCH model and is defined as

$$l' = \sum_{t=L+1}^{n} \log \left\{ \sum_{k=1}^{K} \alpha_k \frac{\lambda_{kt}^{X_t} \exp(-\lambda_{kt})}{X_t!} \right\}.$$

We use l' in this paper, it may have better performance in finite samples (see Wong and Li (2000)). The AIC and the BIC are given by:

$$\begin{aligned} AIC &= -2l' + 2\left(2K - 1 + \sum_{k=1}^{K} p_k + \sum_{k=1}^{K} q_k\right), \\ BIC &= -2l' + \log\Big(n - \max(p_{\max}, q_{\max})\Big)\left(2K - 1 + \sum_{k=1}^{K} p_k + \sum_{k=1}^{K} q_k\right). \end{aligned}$$

The MRC consists of three terms: the first measures the lack of fit, the second imposes a penalty for regression parameters, and the third is the clustering penalty function. For the MINGARCH model, the MRC is defined as

$$\mathrm{MRC} = \sum_{k=1}^{K} \widehat{n}_k \log(\widehat{\sigma}_k^2) + \sum_{k=1}^{K} \frac{\widehat{n}_k(\widehat{n}_k + \widehat{h}_k)}{\widehat{n}_k - \widehat{h}_k - 2} - 2\sum_{k=1}^{K} \widehat{n}_k \log(\widehat{\alpha}_k),$$

where $\widehat{n}_k = \mathrm{tr}(\widehat{W}_k)$, $\widehat{h}_k = \mathrm{tr}(\widehat{H}_k)$, $\widehat{\sigma}_k^2 = (U - V\theta_k^*)^T \widehat{W}_k^{1/2}(I - \widehat{H}_k)(U - V\theta_k^*)/\widehat{n}_k$ with
$\widehat{W}_k = \mathrm{diag}\left((\widehat{\tau}_{k,L+1}, ..., \widehat{\tau}_{kn})^T\right)$, $\widehat{V}_k = \widehat{W}_k^{1/2} V$, $\widehat{H}_k = \widehat{V}_k \left(\widehat{V}_k^T \widehat{V}_k\right)^{-1} \widehat{V}_k^T$, $k=1, ..., K$,
$V = (V_{L+1}, ..., V_n)^T$, $V_j = (1, X_{j-1}, ..., X_{j-p}, \lambda_{k_j(j-1)}, ..., \lambda_{k_j(j-q)})^T$,
$k_j \mid \tau_{k_j,j} = \max\{\tau_{1,j}, ..., \tau_{K,j}\}$, $j = L+1, ..., n$,
$\theta_k^* = \left({\alpha_{(k)}}^T, \mathbf{0}^T, {\beta_{(k)}}^T, \mathbf{0}^T\right)^T_{(p+q+1)\times 1}$, $U = (X_{L+1}, ..., X_n)^T$.

The problem of model selection for MINGARCH models requires two aspects. First, we must select the number of components K. Second, the model identification problem needs to be addressed (i.e. the **AR** polynomial order, p_k, and the **MA** polynomial order, q_k). In this paper we not discuss the selection problem for the number of components. We concentrate on the order selection of each component. The order of the components is chosen to be that minimizing the values of the three criterions. The results are given in Tables 4, 5 and 6.

Table 4: AIC, BIC and MRC values for the crime counts series, $K = 1$.

Order	AIC			BIC			MRC		
	$p=1$	$p=2$	$p=3$	$p=1$	$p=2$	$p=3$	$p=1$	$p=2$	$p=3$
$q=0$	832.52	813.55	811.50	838.44	**822.41**	823.29	558.37	546.64	**545.87**
$q=1$	815.11	813.62	813.31	824.01	825.44	852.80	550.56	548.57	547.16
$q=2$	813.04	815.06	813.39	824.87	829.84	831.09	548.57	550.50	549.37
$q=3$	**807.35**	809.35	811.82	846.83	827.04	832.46	546.31	548.31	547.19

Table 5: AIC, BIC and MRC values for the crime counts series, $K = 2$.

Order	AIC			BIC			MRC		
	$p=1$	$p=2$	$p=3$	$p=1$	$p=2$	$p=3$	$p=1$	$p=2$	$p=3$
$q=0$	767.11	760.23	757.05	781.93	<u>**780.92**</u>	783.59	607.93	606.19	596.61
$q=1$	760.55	758.91	755.24	781.29	785.52	787.68	557.14	548.20	550.90
$q=2$	756.33	760.51	757.54	782.94	793.02	795.88	<u>**537.59**</u>	540.95	548.81
$q=3$	**751.82**	755.78	759.26	838.69	794.12	803.49	538.54	541.92	546.76

Table 6: AIC, BIC and MRC values for the crime counts series, $K = 3$.

Order	AIC			BIC			MRC		
	$p=1$	$p=2$	$p=3$	$p=1$	$p=2$	$p=3$	$p=1$	$p=2$	$p=3$
$q=0$	766.75	760.89	759.21	**790.46**	793.41	800.49	647.45	660.88	720.25
$q=1$	559.52	759.75	758.96	792.11	801.13	809.09	575.82	577.42	631.02
$q=2$	756.75	766.67	756.16	798.13	816.91	824.13	**573.34**	642.36	649.52
$q=3$	**749.04**	757.72	763.83	799.17	804.11	945.47	573.38	573.62	938.79

For the AIC, the BIC and the MRC, the minimums are represented by the underlined values. Based on the results in these tables (4, 5 and 6), the BIC and the MRC retain the two-component mixture model respectively with $(p, q) = (2, 0)$ and $(p, q) = (1, 2)$, which confirm the bimodality observed in the histogram. In contrast, the AIC retains the three-component mixture model with $(p, q) = (1, 3)$, which confirms the phenomena often observed in a lot of applications, namely that the AIC overclusters and overfits the data (for instance, see Naik *et al.* (2007)). In practice, it is observed that the BIC criterion selects the model of dimension smaller than the AIC criterion, which is not surprising since the BIC penalizes more than the AIC (when $n > 7$). We notice also that the next smallest AIC, BIC and MRC values are obtained in the two-component model with respectively $(p, q) = (1, 3)$, $(p, q) = (1, 1)$ and $(p, q) = (1, 3)$. These results confirm the result of the histogram and lends substantial support to the two-component model with $p = 1$ and $q \neq 0$. The values of the AIC and MRC obtained in our model are better than those of the MINARCH model. The values of BIC suggest the MINARCH$(2; 2, 2)$ model, but the smallest value is near of the BIC value obtained with MINGARCH$(2; 1, 1; 2, 2)$ model (780.92 and 782.94). In addition, the AIC of the MINGARCH$(2; 1, 1; 2, 2)$ (selected by the MRC) is better than the one in the MINARCH$(2; 2, 2)$ model. Hence, our results indicate that the MINGARCH model should be preferred to the MINARCH for this dataset.

5. CONCLUDING REMARKS

In this paper, a new model which generalizes the MINARCH model is proposed. Conditions for stationarity of the model and estimation procedure based on EM algorithm are investigated. Moreover, we study the finite performance of the estimation method using Monte Carlo simulations. Finally, a real case study is proposed. In a forthcoming, we plan to study the ergodicity conditions of the model as well as the optimal choice of the parameter K. In addition, we plan to study necessary and sufficient conditions for the MINGARCH$(K; p_1, ..., p_K; q_1, ..., q_K)$ process to be m order stationary for $m > 2$.

APPENDIX A — Proof of Theorem 2.2

Let $\gamma_{it} = E(X_t X_{t-i})$ for $i = 0, 1, ..., L$,

$$\begin{aligned}
\gamma_{it} &= \sum_{k=1}^{K} \alpha_k \mathbb{E}(\lambda_{kt} X_{t-i}) \\
&= \sum_{k=1}^{K} \alpha_{k0}\alpha_k E(X_{t-i}) + \sum_{l=1}^{m}\sum_{k=1}^{K}\sum_{j_1,\ldots,j_l=1}^{L} \alpha_{k0}\alpha_k\beta_{kj_1}\cdots\beta_{kj_l} E(X_{t-i}) \\
&\quad + \sum_{l=1}^{m}\sum_{k=1}^{K}\sum_{j_1,\ldots,j_{l+1}=1}^{L} \alpha_k\alpha_{k_{j+1}}\beta_{kj_1}\cdots\beta_{kj_l} E(X_{t-j_1-\cdots-j_{l+1}} X_{t-i}) \\
&\quad + \sum_{k=1}^{K}\sum_{j_1,\ldots,j_{m+1}=1}^{L} \alpha_k\beta_{kj_1}\cdots\beta_{kj_{m+1}} E(\lambda_{k(t-j_1-\cdots-j_{m+1})} X_{t-i})\,.
\end{aligned}$$

Using the same arguments as in the proof of Theorem 2.1, we can show that almost surely

$$\begin{aligned}
\gamma_{it} &= \sum_{k=1}^{K} \alpha_{k0}\alpha_k E(X_{t-i}) + \sum_{l=1}^{\infty}\sum_{k=1}^{K}\sum_{j_1,\ldots,j_l=1}^{L} \alpha_{k0}\alpha_k\beta_{kj_1}\cdots\beta_{kj_l} E(X_{t-i}) \\
&\quad + \sum_{k=1}^{K}\sum_{j=1}^{L} \alpha_{kj}\alpha_k E(X_{t-j} X_{t-i}) \\
&\quad + \sum_{l=1}^{\infty}\sum_{k=1}^{K}\sum_{j_1,\ldots,j_{l+1}=1}^{L} \alpha_k\alpha_{kj_{l+1}}\beta_{kj_1}\cdots\beta_{kj_l} E(X_{t-j_1-\cdots-j_{l+1}} X_{t-i}) \\
&= I + II + III + IV
\end{aligned}$$

with

$$\begin{aligned}
III &= \sum_{k=1}^{K}\sum_{j=1}^{L} \alpha_{kj}\alpha_k E(X_{t-j} X_{t-i}) \\
&= \sum_{k=1}^{K} \alpha_{ki}\alpha_k\gamma_{0,t-i} + \sum_{k=1}^{K}\sum_{j=1,i\neq j}^{L} \alpha_{kj}\alpha_k\gamma_{|j-i|,t} \\
&= \sum_{k=1}^{K} \alpha_{ki}\alpha_k\gamma_{0,t-i} \\
&\quad + \sum_{k=1}^{K} \alpha_k \Big(\sum_{|j-i|=1} \alpha_{ki}\gamma_{1,t} + \cdots + \sum_{|j-i|=i} \alpha_{kj}\gamma_{i,t} + \cdots + \sum_{|j-i|=L-1} \alpha_{kj}\gamma_{L-1,t} \Big) \\
&= \sum_{k=1}^{K} \alpha_k\delta_{i0k0}\gamma_{0,t-i} + \sum_{k=1}^{K}\sum_{u=1}^{L-1} \alpha_k\delta_{iuk0}\gamma_{u,t}
\end{aligned}$$

and

$$
\begin{aligned}
IV &= \sum_{l=1}^{\infty}\sum_{k=1}^{K}\sum_{j_1,\dots,j_{l+1}=1}^{L} \alpha_k \alpha_{kj_{l+1}} \beta_{kj_1}\cdots\beta_{kj_l}\gamma_{|i-j_1-\cdots-j_{l+1}|,t} \\
&= \sum_{l=1}^{\infty}\sum_{k=1}^{K}\sum_{j_1+\cdots+j_{l+1}=i}^{L} \alpha_k \alpha_{kj_{l+1}} \beta_{kj_1}\cdots\beta_{kj_l}\gamma_{0,t-i} \\
&\quad + \sum_{l=1}^{\infty}\sum_{k=1}^{K}\sum_{j_1+\cdots+j_{l+1}\neq i}^{L} \alpha_k \alpha_{kj_{l+1}} \beta_{kj_1}\cdots\beta_{kj_l}\gamma_{|i-j_1-\cdots-j_{l+1}|,t} \\
&= \sum_{l=1}^{\infty}\sum_{k=1}^{K}\alpha_k\delta_{i0kl}\gamma_{0,t-i} + \sum_{l=1}^{\infty}\sum_{k=1}^{K}\sum_{u=1}^{L-1}\alpha_k\delta_{iukl}\gamma_{u,t}
\end{aligned}
$$

where

$$
\delta_{iukl} = \sum_{|i-j_1-\cdots-j_{l+1}|=u} \alpha_{kj_{l+1}}\beta_{kj_1}\cdots\beta_{kj_l}.
$$

Then

$$
III + IV = \sum_{l=0}^{\infty}\sum_{k=1}^{K}\alpha_k\delta_{i0kl}\gamma_{0,t-i} + \sum_{l=0}^{\infty}\sum_{k=1}^{K}\sum_{u=1}^{L-1}\alpha_k\delta_{iukl}\gamma_{u,t}
$$

where the first term of this summation ($l = 0$) is III.

Moreover, using the same notation, we get

$$
\begin{aligned}
I + II &= \left(\sum_{k=1}^{K}\alpha_{k0}\alpha_k + \sum_{l=1}^{\infty}\sum_{k=1}^{K}\sum_{j_1,\dots,j_l=1}^{L}\alpha_{k0}\alpha_k\beta_{kj_1}\cdots\beta_{kj_l}\right)\mu \\
&= \left(\sum_{l=0}^{\infty}\sum_{k=1}^{K}\sum_{j_1,\dots,j_l=1}^{L}\alpha_{k0}\alpha_k\beta_{kj_1}\cdots\beta_{kj_l}\right)\mu =: K_1
\end{aligned}
$$

Finally, for $i = 1, \dots, L$

$$
K_1 + \omega_{i0}\gamma_{0,t-i} + \sum_{u=1}^{L-1}\omega_{iu}\gamma_{u,t} = 0
$$

where

$$
\omega_{i0} = \sum_{l=0}^{\infty}\sum_{k=1}^{K}\alpha_k\delta_{i0kl}, \quad \omega_{iu} = \sum_{l=0}^{\infty}\sum_{k=1}^{K}\alpha_k\delta_{iukl}
$$

$$
\text{for} \quad u \neq i \ \text{ and } \ \omega_{ii} = \sum_{l=0}^{\infty}\sum_{k=1}^{K}\alpha_k\delta_{iikl} - 1.
$$

Let $\Gamma = (\omega_{ij})_{i,j=1}^{L-1}$ and $\Gamma^{-1} = (b_{ij})_{i,j=1}^{L-1}$. The invertibility of the matrix Γ is checked in Appendix B.

Then

$$\Gamma(\gamma_{1,t},...,\gamma_{L-1,t})^T = -\big(K_1+\omega_{10}\gamma_{0,t-1},...,K_1+\omega_{(L-1)0}\gamma_{0,t-(L-1)}\big)$$

which is equivalent to

$$(\gamma_{1,t},...,\gamma_{L-1,t})^T = -\Gamma^{-1}\big(K_1+\omega_{10}\gamma_{0,t-1},...,K_1+\omega_{(L-1)0}\gamma_{0,t-(L-1)}\big).$$

We can show that

$$\gamma_{i,t} = -K_1\sum_{u=1}^{L-1} b_{iu} - \sum_{u=1}^{L-1} b_{iu}\omega_{u0}\gamma_{0,t-u}\,.$$

The second moment is given by:

$$\gamma_{0,t} \;=\; \mathbb{E}(X_t) + \sum_{k=1}^{K}\alpha_k\mathbb{E}(\lambda_{kt}^2)\,.$$

For $k=1,...,K$, we have

$$\begin{aligned}\lambda_{kt}^2 \;&=\; \Big(\alpha_{k0}+\sum_{i=1}^{L}\alpha_{ki}X_{t-i}+\sum_{j=1}^{L}\beta_{kj}\lambda_{k(t-j)}\Big)\lambda_{kt}\\ &=\; \alpha_{k0}\lambda_{kt}+\sum_{i=1}^{L}\alpha_{ki}X_{t-i}\lambda_{kt}+\sum_{j=1}^{L}\beta_{kj}\lambda_{k(t-j)}\lambda_{kt}\,.\end{aligned}$$

The hypothesis H_1 implies that the process $\{\lambda_{kt},\ t\in\mathbb{Z}\}$ is first-order stationary. Hence

$$\mathbb{E}(\lambda_{kt}) = \frac{\alpha_{k0}+\sum_{i=1}^{L}\alpha_{ki}\mu}{1-\sum_{j=1}^{L}\beta_{kj}} \qquad \text{for}\ \ k=1,...,K.$$

We have

$$\mathbb{E}\Big(\sum_{i=1}^{L}\alpha_{ki}X_{t-i}\lambda_{kt}\Big) \;=$$

$$\begin{aligned}&= \mathbb{E}\Big(C_{k0}\sum_{i=1}^{L}\alpha_{ki}X_{t-i}+\sum_{i=1}^{L}\alpha_{ki}X_{t-i}\sum_{l=0}^{\infty}\sum_{j_1,\dots,j_{l+1}=1}^{L}\alpha_{kj_{l+1}}\beta_{kj_1}\cdots\beta_{kj_l}X_{t-j_1-j_2-\cdots-j_{l+1}}\Big)\\ &= \mathbb{E}\Big(C_{k0}\sum_{i=1}^{L}\alpha_{ki}X_{t-i}+\sum_{l=0}^{\infty}\sum_{j_1,\dots,j_{l+2}=1}^{L}\alpha_{kj_{l+1}}\alpha_{kj_{l+2}}\beta_{kj_1}\cdots\beta_{kj_l}X_{t-j_1-j_2-\cdots-j_{l+1}}X_{t-j_{l+2}}\Big)\\ &= C_{k0}\mu\sum_{i=1}^{L}\alpha_{ki}+\sum_{i=1}^{L}\Delta_{k,i}^{(1)}\gamma_{0,t-i}+\sum_{v=1}^{L-1}\Lambda_{kv}^{(1)}\gamma_{v,t}\end{aligned}$$

where

$$\begin{aligned}\Delta_{k,i}^{(1)} \;&=\; \sum_{l=0}^{\infty}\sum_{\substack{j_{l+2}=i\\ j_{l+2}=j_1+\cdots+j_{l+1}}}^{L}\alpha_{kj_{l+1}}\alpha_{kj_{l+2}}\beta_{kj_1}\cdots\beta_{kj_l}\,,\\ \Lambda_{kv}^{(1)} \;&=\; \sum_{l=0}^{\infty}\sum_{|j_{l+2}-j_1-\cdots-j_{l+1}|=v}^{L}\alpha_{kj_{l+1}}\alpha_{kj_{l+2}}\beta_{kj_1}\cdots\beta_{kj_l}\,.\end{aligned}$$

Moreover

$$\begin{aligned}\sum_{j=1}^{L}\beta_{kj}\lambda_{k(t-j)} &= \sum_{j=1}^{L}\beta_{kj}\left\{C_{k0}+\sum_{l=0}^{\infty}\sum_{j_1,\ldots,j_{l+1}=1}^{L}\alpha_{kj_{l+1}}\beta_{kj_1}\cdots\beta_{kj_l}X_{t-j-j_1-j_2-\cdots-j_{l+1}}\right\}\\ &= C_{k0}\sum_{j=1}^{L}\beta_{kj}+\sum_{l=0}^{\infty}\sum_{j_1,\ldots,j_{l+2}=1}^{L}\alpha_{kj_{l+2}}\beta_{kj_1}\cdots\beta_{kj_{l+1}}X_{t-j_1-j_2-\cdots-j_{l+2}}\,.\end{aligned}$$

Hence

$$\sum_{j=1}^{L}\beta_{kj}\lambda_{k(t-j)}\lambda_{kt} =$$

$$\begin{aligned}&= \left\{C_{k0}\sum_{j=1}^{L}\beta_{kj}+\sum_{l=0}^{\infty}\sum_{j_1,\ldots,j_{l+2}=1}^{L}\alpha_{kj_{l+2}}\beta_{kj_1}\cdots\beta_{kj_{l+1}}X_{t-j_1-j_2-\cdots-j_{l+2}}\right\}\\ &\quad\times\left\{C_{k0}+\sum_{l=0}^{\infty}\sum_{j_1,\ldots,j_{l+1}=1}^{L}\alpha_{kj_{l+1}}\beta_{kj_1}\cdots\beta_{kj_l}X_{t-j_1-j_2-\cdots-j_{l+1}}\right\}\\ &= C_{k0}^2\sum_{j=1}^{L}\beta_{kj}+C_{k0}\sum_{j=1}^{L}\beta_{kj}\sum_{l=0}^{\infty}\sum_{j_1,\ldots,j_{l+1}=1}^{L}\alpha_{kj_{l+1}}\beta_{kj_1}\cdots\beta_{kj_l}X_{t-j_1-j_2-\cdots-j_{l+1}}\\ &+ C_{k0}\sum_{l=0}^{\infty}\sum_{j_1,\ldots,j_{l+2}=1}^{L}\alpha_{kj_{l+2}}\beta_{kj_1}\cdots\beta_{kj_{l+1}}X_{t-j_1-j_2-\cdots-j_{l+2}}\\ &+\sum_{\substack{l=0\\ l'=0}}^{\infty}\sum_{\substack{j_1,\ldots,j_{l+2}=1\\ j'_1,\ldots,j'_{l'+1}=1}}^{L}\alpha_{kj_{l+2}}\beta_{kj_1}\cdots\beta_{kj_{l+1}}\alpha_{kj'_{l'+1}}\beta_{kj'_1}\cdots\beta_{kj'_{l'}}X_{t-j_1-j_2-\cdots-j_{l+2}}X_{t-j'_1-j'_2-\cdots-j'_{l'+1}}.\end{aligned}$$

The term $\mathbb{E}\Big(\sum_{j=1}^{L}\beta_{kj}\lambda_{k(t-j)}\lambda_{kt}\Big)$ is given by

$$\begin{aligned}\mathbb{E}\Big(\sum_{j=1}^{L}\beta_{kj}\lambda_{k(t-j)}\lambda_{kt}\Big) &= C_{k0}^2\sum_{j=1}^{L}\beta_{kj}+2C_{k0}\mu\sum_{l=0}^{\infty}\sum_{j_1,\ldots,j_{l+2}=1}^{L}\alpha_{kj_{l+2}}\beta_{kj_1}\cdots\beta_{kj_{l+1}}\\ &\quad+\sum_{i=1}^{L}\Delta_{k,i}^{(2)}\gamma_{0,t-i}+\sum_{v=1}^{L-1}\Lambda_{kv}^{(2)}\gamma_{v,t}\end{aligned}$$

where

$$\begin{aligned}\Delta_{k,i}^{(2)} &= \sum_{\substack{l=0\\ l'=0}}^{\infty}\sum_{\substack{j_1+\cdots+j_{l+2}=i\\ j_1+\cdots+j_{l+2}=j'_1+\cdots+j'_{l'+1}}}^{L}\alpha_{kj_{l+2}}\beta_{kj_1}\cdots\beta_{kj_{l+1}}\alpha_{kj'_{l'+1}}\beta_{kj'_1}\cdots\beta_{kj'_{l'}}\\ \Lambda_{kv}^{(2)} &= \sum_{\substack{l=0\\ l'=0}}^{\infty}\sum_{\left|j_1+\cdots+j_{l+2}-j'_1-\cdots-j'_{l'+1}\right|=v}^{L}\alpha_{kj_{l+2}}\beta_{kj_1}\cdots\beta_{kj_{l+1}}\alpha_{kj'_{l'+1}}\beta_{kj'_1}\cdots\beta_{kj'_{l'}}.\end{aligned}$$

Let $\Delta_{k,i} = \Delta_{k,i}^{(1)} + \Delta_{k,i}^{(2)}$ and $\Lambda_{kv} = \Lambda_{kv}^{(1)} + \Lambda_{kv}^{(2)}$.

For $k = 1, ..., K$, the expectation of λ_{kt}^2 is given by

$$\mathbb{E}(\lambda_{k,t}^2) = C_k + \sum_{i=1}^{L} \Delta_{k,i}\gamma_{0,t-i} + \sum_{v=1}^{L-1} \Lambda_{kv}\gamma_{v,t}$$

with

$$C_k = \alpha_{k0}\mathbb{E}(\lambda_{kt}) + C_{k0}\sum_{i=1}^{L}\alpha_{ki}\mu + C_{k0}^2\sum_{j=1}^{L}\beta_{kj} + 2C_{k0}\sum_{l=0}^{\infty}\sum_{j_1,...,j_{l+2}=1}^{L}\alpha_{kj_{l+2}}\beta_{kj_1}\cdots\beta_{kj_{l+1}}\mu \,.$$

Then

$$\begin{aligned}
\gamma_{0,t} &= \mu + \sum_{k=1}^{K}\alpha_k\left(C_k + \sum_{i=1}^{L}\Delta_{k,i}\gamma_{0,t-i} + \sum_{v=1}^{L-1}\Lambda_{kv}\gamma_{v,t}\right) \\
&= \mu + \sum_{k=1}^{K}\alpha_k\left[C_k + \sum_{u=1}^{L}\Delta_{k,u}\gamma_{0,t-u}\right. \\
&\quad \left. + \sum_{v=1}^{L-1}\Lambda_{kv}\left(-K_1\sum_{u=1}^{L-1}b_{vu} - \sum_{u=1}^{L-1}b_{vu}\omega_{u0}\gamma_{0,t-u}\right)\right] \\
&= c_0 + \sum_{k=1}^{K}\alpha_k\left[\sum_{u=1}^{L}\Delta_{k,u}\gamma_{0,t-u} - \sum_{u=1}^{L-1}\left(\sum_{v=1}^{L-1}\Lambda_{kv}b_{vu}\omega_{u0}\right)\gamma_{0,t-u}\right]
\end{aligned}$$

where

$$c_0 = \mu + \sum_{k=1}^{K}\alpha_k C_k - K_1\sum_{k=1}^{K}\alpha_k\sum_{v=1}^{L-1}\Lambda_{kv}\sum_{u=1}^{L-1}b_{vu}.$$

Hence

$$\gamma_{0,t} = c_0 + \sum_{k=1}^{K}\alpha_k\left[\sum_{u=1}^{L-1}\left(\Delta_{k,u} - \sum_{v=1}^{L-1}\Lambda_{kv}b_{vu}\omega_{u0}\right)\gamma_{0,t-u} + \Delta_{k,L}\gamma_{0,t-L}\right]. \tag{5.1}$$

Let

$$c_u = \sum_{k=1}^{K}\alpha_k\left(\Delta_{k,u} - \sum_{v=1}^{L-1}\Lambda_{kv}b_{vu}\omega_{u0}\right), \quad u = 1, ..., L-1 \quad \text{and} \quad c_L = \sum_{k=1}^{K}\alpha_k\Delta_{k,L}\,.$$

Then the equation (5.1) is equivalent to:

$$\gamma_{0,t} = c_0 + \sum_{u=1}^{L}c_u\gamma_{0,t-u}\,. \tag{5.2}$$

The necessary and sufficient condition for a non-homogeneous difference equation (5.2) to have a stable solution, which is finite and independent of t, is that all roots of the equation: $1 - c_1Z^{-1} - c_2Z^{-2} - \cdots - c_LZ^{-L} = 0$ lie inside the unit circle.

APPENDIX B — Invertibility of Γ and positivity of c_0

In the following lines, we establish the invertibility of Γ and check the positivity of c_0. The same ideas were already used in the paper by Gonçalves *et al.* (2013).

Invertibility of Γ:

We show that the matrix $\Gamma = (\omega_{ij})_{i,j=1}^{L-1}$ is strictly diagonally dominant by rows. For $i = 1, ..., L-1$,

$$\begin{aligned}
|\omega_{ii}| - \sum_{\substack{u=1\\u\neq i}}^{L-1} |\omega_{iu}| &= 1 - \sum_{l=0}^{\infty}\sum_{k=1}^{K} \alpha_k \delta_{iikl} - \sum_{\substack{u=1\\u\neq i}}^{L-1}\sum_{l=0}^{\infty}\sum_{k=1}^{K} \alpha_k \delta_{iukl} \\
&= 1 - \sum_{u=1}^{L-1}\sum_{l=0}^{\infty}\sum_{k=1}^{K} \alpha_k \delta_{iukl}\,.
\end{aligned}$$

We have

$$\begin{aligned}
\sum_{u=1}^{L-1}\sum_{l=0}^{\infty}\sum_{k=1}^{K} \alpha_k \delta_{iukl} &= \sum_{u=1}^{L-1}\sum_{l=0}^{\infty}\sum_{k=1}^{K} \alpha_k \sum_{|i-j_1-\cdots-j_{l+1}|=u} \alpha_{kj_{l+1}} \beta_{kj_1} \cdots \beta_{kj_l} \\
&\leq \sum_{j_1,\ldots,j_{l+1}=1}^{L} \sum_{l=0}^{\infty}\sum_{k=1}^{K} \alpha_k \alpha_{kj_{l+1}} \beta_{kj_1} \cdots \beta_{kj_l}\,.
\end{aligned}$$

Based on the necessary condition for first-order stationarity in equation (2.5), we have

$$\sum_{j_1,\ldots,j_{l+1}=1}^{L} \sum_{l=0}^{\infty}\sum_{k=1}^{K} \alpha_k \alpha_{kj_{l+1}} \beta_{kj_1} \cdots \beta_{kj_l} < 1\,.$$

Hence, $|\omega_{ii}| - \sum_{\substack{u=1\\u\neq i}}^{L-1} |\omega_{iu}| > 0$. Then Γ is strictly diagonally dominant by rows. Hence, the matrix Γ is invertible by using the Levy–Desplanques Theorem (see Horn and Jonhson (2013), pp. 352, 392).

Positivity of c_0:

$$c_0 = \mu + \sum_{k=1}^{K} \alpha_k C_k - K_1 \sum_{k=1}^{K} \alpha_k \sum_{v=1}^{L-1} \Lambda_{kv} \sum_{u=1}^{L-1} b_{vu}\,.$$

To prove the positivity of c_0, it suffices to show that $b_{vu} \leq 0, v = 1, ..., L-1, u = 1, ..., L-1$. Indeed, it is easily seen that $-\Gamma$ is strictly diagonally dominant by rows. In addition, $-\omega_{ij} < 0$ for $i \neq j$ and $-\omega_{ii} > 0$ for $i = 1, ..., L-1$. Then $-\Gamma$ is a nonsingular M-matrix (see Quarteroni *et al.* (2000), p. 30, Property 1.20). This implies that $-\Gamma$ is inverse-positive that is $(-\Gamma)^{-1} \geq 0$. Hence, $\Gamma^{-1} \leq 0$, therefore $b_{vu} \leq 0$ for $v = 1, ..., L-1,\ u = 1, ..., L-1$.

ACKNOWLEDGMENTS

We wish to thank the referee as well as the associate editor for helpful comments and suggestions, which have permitted to improve the presentation of this paper.

REFERENCES

[1] Dempster, A.P.; Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Statist. Soc.*, **Series B 39**, 1–38.

[2] Der, G. and Everitt, B.S. (2002). *A Handbook of Statistical Analysis using SAS, second ed.*, Chapman and Hall/CRC, London.

[3] Ferland, R.; Latour, A. and Oraichi, D. (2006). Integer-valued GARCH process, *J. Time Ser. Anal.*, **27**, 923–942.

[4] Goldberg, S. (1958). *Introduction to Difference Equation*, Wiley, New York.

[5] Gonçalves, E.; Mendes-Lopes, N. and Silva, P. (2013). Infinitely divisible distributions in integer valued GARCH models, *Preprint Number 13-38*, Universidade de Coimbra.

[6] Gradshteyn, I.S. and Ryzhik, I.M. (2007). *Table of Integrals, Series, and Products*, seventh ed., Academic Press, San Diego.

[7] Horn, R.A. and Jonhson, C.R. (2013). *Matrix Analysis*, Cambridge University Press, New York, 2nd ed.

[8] Karlis, D. and Xekalaki, E. (2003). Choosing initial values for the EM algorithm for finite mixtures, *Comput. Statist. and Data Anal.*, **41**, 577–590.

[9] Martin, V.L. (1992). Threshold time series models as multimodal distribution jump processes, *J. Time Ser. Anal.*, **13**, 79–94.

[10] Melnykov, V. and Melnykov, I. (2012). Initializing the EM algorithm in Gaussian mixture models with an unknown number of components, *Comput. Statist. and Data Anal.*, **56**, 1381–1395.

[11] Muller, D.W. and Sawitzki, G. (1991). Excess Mass Estimates and Tests for Multimodality, *J. Amer. Statist. Assoc.*, **86**, 738–746.

[12] Naik, P.A.; Shi, P. and Tsai, C.L. (2007). Extending the Akaike information criterion to mixture regression models, *J. Amer. Statist. Assoc.*, **102**, 244–254.

[13] Nityasuddhia, D. and Böhning, D. (2003). Asymptotic properties of the EM algorithm estimate for normal mixture models with component specific variances, *Comput. Statist. and Data Anal.*, **41**, 591–601.

[14] Quarteroni, A.; Sacco, R. and Saleri, F. (2000). *Numerical Mathematics*, Texts in Applied Mathematics 37, Springer-Verlag, New York.

[15] SAIKKONEN, P. (2007). Stability of mixtures of vector autoregressions with autoregressive conditional heteroskedasticity, *Stat. Sinica*, **17**, 221–239.

[16] TITTERINGTON, D.M.; SMITH, A.F.M. and MARKOV, U.E. (1985). Statistical Analysis of Finite Mixture Distributions, Wiley, New York.

[17] WONG, C.S. and LI, W.K. (2000). On a mixture autoregressive model, *J. Roy. Statist. Soc.*, **Series B 62**, 95–115.

[18] ZEGER, S.L. (1988). A regression model for time series of counts, *Biometrika*, **75**, 621–629.

[19] ZHU, F. (2011). A negative binomial integer-valued GARCH model, *J. Time Ser. Anal.*, **32**, 54–67.

[20] ZHU, F. (2012). Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models, *J. Math. Anal. Appl.*, **389**, 58–71.

[21] ZHU, F.; LI, Q. and WANG, D. (2010). A mixture integer-valued ARCH model, *J. Statist. Plann. Inference*, **140**, 2025–2036.

REVSTAT – Statistical Journal
Volume 14, Number 3, June 2016, 273–296

MEAN-OF-ORDER-p LOCATION-INVARIANT EXTREME VALUE INDEX ESTIMATION

Authors: M. Ivette Gomes
– CEAUL and DEIO, FCUL, Universidade de Lisboa, Portugal
ivette.gomes@fc.ul.pt

Lígia Henriques-Rodrigues
– Universidade de São Paulo, IME, São Paulo, Brasil, and CEAUL, Portugal
ligiahr@ime.usp.br

B.G. Manjunath
– CEAUL, Universidade de Lisboa, Portugal
bgmanjunath@gmail.com

Received: September 2014 Revised: January 2015 Accepted: January 2015

Abstract:

- A simple generalisation of the classical Hill estimator of a positive extreme value index (EVI) has been recently introduced in the literature. Indeed, the Hill estimator can be regarded as the logarithm of the mean of order $p = 0$ of a certain set of statistics. Instead of such a geometric mean, we can more generally consider the mean of order p (MOP) of those statistics, with p real, and even an optimal MOP (OMOP) class of EVI-estimators. These estimators are scale invariant but not location invariant. With PORT standing for peaks over random threshold, new classes of PORT-MOP and PORT-OMOP EVI-estimators are now introduced. These classes are dependent on an extra tuning parameter q, $0 \leq q < 1$, and they are both location and scale invariant, a property also played by the EVI. The asymptotic normal behaviour of those PORT classes is derived. These EVI-estimators are further studied for finite samples, through a Monte-Carlo simulation study. An adequate choice of the *tuning* parameters under play is put forward, and some concluding remarks are provided.

Key-Words:

- *bootstrap and/or heuristic threshold selection; heavy tails; location/scale invariant semi-parametric estimation; Monte-Carlo simulation; optimal levels; statistics of extremes.*

AMS Subject Classification:

- 62G32, 65C05.

1. INTRODUCTION

Given a sample of size n of *independent, identically distributed* random variables (RVs), $\underline{\mathbf{X}}_n := (X_1, ..., X_n)$, with a common *cumulative distribution function* (CDF) F, let us denote by $X_{1:n} \leq \cdots \leq X_{n:n}$ the associated ascending order statistics. As usual in a framework of *extreme value theory* (EVT), let us further assume that there exist sequences of real constants $\{a_n > 0\}$ and $\{b_n \in \mathbb{R}\}$ such that the maximum, linearly normalised, i.e. $(X_{n:n} - b_n)/a_n$, converges in distribution to a non-degenerate RV. Then, the limit distribution is necessarily of the type of the general *extreme value* (EV) CDF, given by

$$\text{(1.1)} \qquad \mathrm{EV}_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}),\ 1+\xi x > 0, & \text{if } \xi \neq 0, \\ \exp(-\exp(-x)),\ x \in \mathbb{R}, & \text{if } \xi = 0. \end{cases}$$

The CDF F is said to belong to the max-domain of attraction of EV_ξ, and we consider the common notation $F \in \mathcal{D}_\mathcal{M}(\mathrm{EV}_\xi)$. The parameter ξ is the *extreme value index* (EVI), the primary parameter of extreme events.

The EVI measures the heaviness of the survival function or right tail-function

$$\text{(1.2)} \qquad \overline{F}(x) := 1 - F(x),$$

and the heavier the right tail, the larger ξ is. Let us further use the notation $\mathcal{R}_a$ for the class of regularly varying functions at infinity, with an index of regular variation equal to $a \in \mathbb{R}$, i.e. positive measurable functions $g(\cdot)$ such that for all $x > 0$, $g(tx)/g(t) \to x^a$, as $t \to \infty$ (see Bingham *et al.*, 1987, among others, for details on the theory of regular variation). In this paper we work with Pareto-type underlying models, i.e. with a positive EVI, a quite common assumption in many areas of application, like bibliometrics, biostatistics, computer science, insurance, finance, social sciences and telecommunications, among others. The right-tail function $\overline{F}$, in (1.2), belongs then to $\mathcal{R}_{-1/\xi}$. Indeed, and more generally,

$$\text{(1.3)} \qquad \overline{F} \in \mathcal{D}_\mathcal{M}(\mathrm{EV}_{\xi>0}) =: \mathcal{D}_\mathcal{M}{}^+ \qquad \Longleftrightarrow \qquad \overline{F} \in \mathcal{R}_{-1/\xi},$$

a result due to Gnedenko (1943).

For the class of Pareto-type models in (1.3), the most well-known EVI-estimators are the Hill (H) estimators (Hill, 1975), which are the averages of the log-excesses,

$$V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}, \quad 1 \leq i \leq k < n.$$

We can thus define the H-class of EVI-estimators as:

$$\text{(1.4)} \qquad \mathrm{H}(k) := \mathrm{H}(k; \underline{\mathbf{X}}_n) := \frac{1}{k}\sum_{i=1}^{k} V_{ik}, \quad 1 \leq k < n.$$

We can further write

$$\mathrm{H}(k) = \sum_{i=1}^{k} \ln\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{1/k} = \ln\left(\prod_{i=1}^{k} \frac{X_{n-i+1:n}}{X_{n-k:n}}\right)^{1/k}, \quad 1 \le k < n.$$

The Hill estimator is thus the logarithm of the *geometric mean* (or *mean of order* 0) of

$$U_{ik} := X_{n-i+1:n}/X_{n-k:n}, \quad 1 \le i \le k < n.$$

Brilhante *et al.* (2013) considered as basic statistics, the *mean of order p* (MOP) of U_{ik}, $1 \le i \le k$, for $p \ge 0$. More generally, Gomes and Caeiro (2014) considered those same statistics for any $p \in \mathbb{R}$, i.e. the class of statistics

$$\mathrm{M}_p(k) = \begin{cases} \left(\frac{1}{k}\sum\limits_{i=1}^{k} U_{ik}^p\right)^{1/p}, & \text{if } p \ne 0, \\[2ex] \left(\prod\limits_{i=1}^{k} U_{ik}\right)^{1/k}, & \text{if } p = 0, \end{cases}$$

and the following associated class of MOP EVI-estimators:

$$\mathrm{H}_p(k) = \mathrm{H}_p(k; \underline{\mathbf{X}}_n) \equiv \mathrm{MOP}(k) := \begin{cases} \Big(1 - \mathrm{M}_p^{-p}(k)\Big)/p, & \text{if } p < 1/\xi, \\[2ex] \ln \mathrm{M}_0(k) = \mathrm{H}(k), & \text{if } p = 0, \end{cases} \tag{1.5}$$

with $\mathrm{H}_0(k) \equiv \mathrm{H}(k)$, given in (1.4). This class of MOP EVI-estimators depends now on this *tuning* parameter $p \in \mathbb{R}$, it is highly flexible, but, as often desired, it is not location-invariant, depending strongly on possible shifts in the model underlying the data. To make the EVI-estimators $\mathrm{H}_p(k)$, in (1.5), location-invariant, it is thus sensible to use the *peaks over a random threshold* (PORT) technique now applied to the MOP EVI-estimation. The PORT methodology, introduced in Araújo Santos *et al.* (2006) and further studied in Gomes *et al.* (2008a), is based on a *sample of excesses* over a random threshold $X_{n_q:n}$, $n_q := \lfloor nq \rfloor + 1$, where $\lfloor x \rfloor$ denotes the integer part of x, i.e. it is based on the sample of size $n^{(q)} = n - n_q$, defined by

$$\underline{\mathbf{X}}_n^{(q)} := \big(X_{n:n} - X_{n_q:n}, ..., X_{n_q+1:n} - X_{n_q:n}\big). \tag{1.6}$$

After the introduction, in Section 2, of a few technical details in the field of EVT and a brief reference to the most simple *minimum-variance reduced-bias* (MVRB) EVI-estimators, the *corrected-Hill* (CH) EVI-estimators introduced and studied in Caeiro *et al.* (2005), we refer a class of *optimal* MOP (OMOP) EVI-estimators recently studied in Brilhante *et al.* (2014). We further introduce the new classes of PORT-MOP and PORT-OMOP EVI-estimators. Section 3 is essentially dedicated to consistency and asymptotic normal behaviour of these new classes of EVI-estimators, with a brief reference to the known asymptotic behaviour of the CH and MOP EVI-estimators. Section 4 is dedicated to the finite sample properties of the new classes of estimators, comparatively to the behaviour of the aforementioned MVRB and even PORT-MVRB EVI-estimators, done through a small-scale simulation study. In Section 5, we refer possible methods for the adaptive choice of the tuning parameters (k, p, q), either based on the bootstrap or on heuristic methodologies, and provide some concluding remarks.

2. PRELIMINARY RESULTS IN THE AREA OF EVT

In the area of EVT and whenever working with large values, i.e. with the right tail of the model F underlying the available sample, the model F is usually said to be *heavy-tailed* whenever (1.3) holds. Moreover, with the notation $F^{\leftarrow}(t) := \inf\{x : F(x) \geq t\}$ for the generalised inverse function of F, the condition $F \in \mathcal{D}_{\mathcal{M}}^{+}$ is equivalent to say that the tail quantile function $U(t) := F^{\leftarrow}(1 - 1/t)$ is of regular variation with index ξ (de Haan, 1984). We thus assume the validity of any of the following first-order conditions:

$$(2.1) \qquad F \in \mathcal{D}_{\mathcal{M}}^{+} \quad \Longleftrightarrow \quad \overline{F} \in \mathcal{R}_{-1/\xi} \quad \Longleftrightarrow \quad U \in \mathcal{R}_{\xi}.$$

The *second-order parameter* ρ (≤ 0) rules the rate of convergence in the first-order condition, in (2.1), and can be defined as the non-positive parameter appearing in the limiting relation

$$(2.2) \qquad \lim_{t\to\infty} \frac{\ln U(tx) - \ln U(t) - \xi \ln x}{A(t)} = \psi_\rho(x) := \begin{cases} \frac{x^\rho - 1}{\rho}, & \text{if } \rho < 0, \\ \ln x, & \text{if } \rho = 0, \end{cases}$$

which is assumed to hold for every $x > 0$, and where $|A|$ must then be of regular variation with index ρ (Geluk and de Haan, 1987). For related details on the topic, see Beirlant *et al.* (2004) and de Haan and Ferreira (2006).

Whenever dealing with bias reduction in the field of extremes, it is usual to consider a slightly more restrict class than $\mathcal{D}_{\mathcal{M}}^{+}$, the class of models

$$(2.3) \qquad U(t) = C\, t^\xi \bigl\{1 + A(t)/\rho + o(t^\rho)\bigr\}, \qquad A(t) = \xi\beta t^\rho,$$

as $t \to \infty$, where $C > 0$, $\xi > 0$, $\rho < 0$ and $\beta \neq 0$ (Hall and Welsh, 1985). This means that the slowly varying function $L(t)$ in $U(t) = t^\xi L(t)$ is assumed to behave asymptotically as a constant. To assume (2.3) is equivalent to choose $A(t) = \xi\beta t^\rho$, $\rho < 0$, in the more general second-order condition in (2.2). Models like the log-Gamma $(\rho = 0)$ are thus excluded from this class. The standard Pareto $(\rho = -\infty)$ is also excluded. But most heavy-tailed models used in applications, like the EV_ξ, in (1.1), the Fréchet, $F(x) = \exp(-x^{-1/\xi})$, $x \geq 0$, both for $\xi > 0$, and the well-known Student's t CDFs, among others, belong to Hall–Welsh class.

2.1. The CH class of EVI-estimators

Due to its simplicity and just as mentioned above, the most popular EVI-estimators, consistent only for non-negative values of ξ, are Hill estimators in (1.4). We further consider the simplest class of CH EVI-estimators, the one introduced in Caeiro *et al.* (2005),

$$\mathrm{CH}(k) = \mathrm{CH}(k; \underline{\mathbf{X}}_n) := \mathrm{H}(k)\left(1 - \frac{\hat{\beta}(n/k)^{\hat{\rho}}}{1-\hat{\rho}}\right). \tag{2.4}$$

The estimators in (2.4) can be second-order MVRB EVI-estimators, for adequate levels k and an adequate external estimation of the vector of second-order parameters, (β, ρ), in (2.3), algorithmically given in Gomes and Pestana (2007), among others, i.e. the use of $\mathrm{CH}(k)$, and an adequate estimation of (β, ρ), enables us to eliminate the dominant component of the bias of the Hill estimator, $\mathrm{H}(k)$, keeping its asymptotic variance. Like that, and theoretically, $\mathrm{CH}(k)$ outperforms $\mathrm{H}(k)$ for all k.

We again suggest the use of the class of β-estimators in Gomes and Martins (2002) and the simplest class of ρ-estimators in Fraga Alves *et al.* (2003). In the simulations, we have considered only models with $|\rho| \le 1$. Indeed, this is the case where alternatives to the H-class of EVI-estimators are welcome due to the high bias of H EVI-estimators for moderate up to large values of k, including the optimal k in the sense of minimal *root mean square error* (RMSE). In such cases, we suggest the use of the *tuning* parameter $\tau = 0$ in the simplest class of ρ-estimators in Fraga Alves *et al.* (2003), given by

$$\hat{\rho}_\tau(k) \equiv \hat{\rho}_\tau(k; \underline{\mathbf{X}}_n) := \min\left(0,\ \frac{3(R_n^{(\tau)}(k; \underline{\mathbf{X}}_n) - 1)}{R_n^{(\tau)}(k; \underline{\mathbf{X}}_n) - 3}\right), \tag{2.5}$$

and dependent on the statistics

$$R_n^{(\tau)}(k; \underline{\mathbf{X}}_n) := \frac{\left(M_n^{(1)}(k; \underline{\mathbf{X}}_n)\right)^\tau - \left(M_n^{(2)}(k; \underline{\mathbf{X}}_n)/2\right)^{\tau/2}}{\left(M_n^{(2)}(k; \underline{\mathbf{X}}_n)/2\right)^{\tau/2} - \left(M_n^{(3)}(k; \underline{\mathbf{X}}_n)/6\right)^{\tau/3}}, \quad \tau \in \mathbb{R},$$

with the usual notation $a^{b\tau} = b \ln a$ if $\tau = 0$, and where

$$M_n^{(j)}(k; \underline{\mathbf{X}}_n) := \frac{1}{k}\sum_{i=1}^{k} \{\ln X_{n-i+1:n} - \ln X_{n-k:n}\}^j, \quad j = 1, 2, 3.$$

As already suggested in previous papers, we have here decided for the computation of $\hat{\rho}_\tau(k)$ at $k = k_1$, given by $k_1 = \lfloor n^{1-\epsilon} \rfloor$, $\epsilon = 0.001$, the threshold used in Caeiro *et al.* (2005) and Gomes and Pestana (2007).

For the estimation of the scale second-order parameter β, in (2.3), and again on the basis of a sample $\underline{\mathbf{X}}_n$, we consider

$$\hat{\beta}_{\hat{\rho}}(k) \equiv \hat{\beta}_{\hat{\rho}}(k; \underline{\mathbf{X}}_n) := \left(\frac{k}{n}\right)^{\hat{\rho}} \frac{d_{\hat{\rho}}(k)\ D_0(k) - D_{\hat{\rho}}(k)}{d_{\hat{\rho}}(k)\ D_{\hat{\rho}}(k) - D_{2\hat{\rho}}(k)}, \tag{2.6}$$

dependent on the estimator $\hat{\rho} = \hat{\rho}_0(k_1; \underline{\mathbf{X}}_n)$, with $\hat{\rho}_\tau(k)$ defined in (2.5), and where, for any $\alpha \leq 0$,

$$d_\alpha(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\alpha} \quad \text{and}$$

$$D_\alpha(k) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^{-\alpha} U_i, \quad U_i := i\left(\ln\frac{X_{n-i+1:n}}{X_{n-i:n}}\right),$$

with U_i, $1 \leq i \leq k$, the *scaled log-spacings* associated with $\underline{\mathbf{X}}_n$. Details on the distributional behaviour of the estimator in (2.6) can be found in Gomes and Martins (2002) and more recently in Gomes *et al.* (2008b) and Caeiro *et al.* (2009). Interesting alternative classes of estimators of the 'shape' and 'scale' second-order parameters have recently been introduced. References to those classes can be found in recent overviews on reduced-bias estimation (Chapter 6 of Reiss and Thomas, 2007; Beirlant *et al.*, 2012; Gomes and Guillou, 2014).

2.2. The OMOP class of EVI-estimators

Working in the class of models in (2.3) for technical simplicity, Brilhante *et al.* (2014) noticed that there is an optimal value $p \equiv p_{\text{M}} = \varphi_\rho/\xi$, with

$$\varphi_\rho = 1 - \rho/2 - \sqrt{\rho^2 - 4\rho + 2}\Big/2 \in \left(0, 1 - \sqrt{2}/2\right), \tag{2.7}$$

which maximises the asymptotic efficiency of the class of estimators in (1.5). They then considered the MOP EVI-estimator associated with the optimal $p \equiv p_{\text{M}}$ estimated through $\hat{p}_{\text{M}}$, based on any initial consistent estimator of ξ and ρ, i.e. an *optimal* MOP (OMOP) class of EVI-estimators. Here, we estimate the optimal k-value for the H EVI-estimation, $k_{0|0} := \arg\min_k \text{RMSE}\big(\text{H}_0(k)\big)$, computing, as given in Hall (1982),

$$\hat{k}_{0|0} \equiv \hat{k}_{0|\text{H}_0} = \left((1-\hat{\rho})n^{-\hat{\rho}}/\big(\hat{\beta}\ \sqrt{-2\hat{\rho}}\big)\right)^{2/(1-2\hat{\rho})},$$

the associated observed value of the EVI-estimator $\text{H}_{00} := \text{H}(\hat{k}_{0|0})$, and, with φ_ρ given in (2.7), the OMOP EVI-estimators

$$\text{H}^*(k) \equiv \text{H}^*(k; \underline{\mathbf{X}}_n) := \text{H}_{\hat{p}_{\text{M}}}(k; \underline{\mathbf{X}}_n), \quad 1 \leq k < n, \quad \hat{p}_{\text{M}} = \varphi_{\hat{\rho}}/\text{H}_{00}. \tag{2.8}$$

Neither the H nor the CH nor the MOP EVI-estimators are invariant for changes in location, but they can easily be made location-invariant with the technique introduced in Araújo Santos *et al.* (2006), briefly discribed in the following Section.

2.3. The PORT methodology

The EVI-estimators in (1.4), (1.5), (2.4) and (2.8) are scale-invariant, but not location-invariant, as often desired, due to the fact that the EVI itself enjoys such a property, i.e. it is location and scale invariant. Indeed, note that a general first-order condition to have $F \in \mathcal{D}_{\mathcal{M}}(\mathrm{EV}_\xi)$, given in de Haan (1984), can be written as

$$F \in \mathcal{D}_{\mathcal{M}}(\mathrm{EV}_\xi) \iff \lim_{t\to\infty} \frac{U(tx)-U(t)}{a(t)} = \psi_\xi(x), \tag{2.9}$$

for an adequate function $a(\cdot)$, with an absolute value necessarily in $\mathcal{R}_\xi$, and where $\psi_\rho(\cdot)$ is the Box–Cox function, already defined in (2.2). If a shift s is induced in data associated with the RV X, i.e. if we consider $Y = X+s$, the relationship between the tail quantile functions of Y and X is given by $U_Y(t) = s + U_X(t)$. Consequently, $U_Y(tx) - U_Y(t) = U_X(tx) - U_X(t)$ and from (2.9), the EVI, ξ, is the same for X and $Y = X+s$, for any shift $s \in \mathbb{R}$.

Just as mentioned above, the class of PORT-Hill estimators is based on the *sample of excesses* in (1.6). In this article, we shall work with PORT-MOP and PORT-OMOP EVI-estimators, generally denoted E. They have the same functional form of the associated EVI-estimators in (1.5) and (2.8) but with the original sample $\underline{\mathbf{X}}_n$ replaced everywhere by the sample of excesses $\underline{\mathbf{X}}_n^{(q)}$, in (1.6). Consequently, they are given by the functional equations,

$$\mathrm{E}^{(q)}(k) := \mathrm{E}\big(k; \underline{\mathbf{X}}_n^{(q)}\big), \qquad \text{with } \mathrm{E} \equiv \mathrm{H}_p \text{ and } \mathrm{E} \equiv \mathrm{H}^*. \tag{2.10}$$

These estimators are now invariant for both changes of location and scale, and depend on the extra *tuning parameter* q, which only influences the asymptotic bias, making them highly flexible and even able to compare favourably with the MVRB EVI-estimators in (2.4), for a large variety of underlying models in the domain of attraction for maxima of the EV_ξ CDF, in (1.1). In the simulation procedure, we further include the PORT-MVRB EVI-estimators,

$$\mathrm{CH}^{(q)}(k) = \mathrm{CH}\big(k; \underline{\mathbf{X}}_n^{(q)}\big), \tag{2.11}$$

studied by simulation in Gomes *et al.* (2011a, 2013), with $\underline{\mathbf{X}}_n^{(q)}$ and $\mathrm{CH}\big(k; \underline{\mathbf{X}}_n\big)$ respectively given in (1.6) and (2.4).

3. ASYMPTOTIC BEHAVIOUR OF EVI-ESTIMATORS

Consistency of the Hill EVI-estimators, $\mathrm{H} \equiv \mathrm{H}_0$, written both in (1.4) and (1.5), is achieved in the whole $\mathcal{D}_{\mathcal{M}}^+$ whenever we work with intermediate values of k, i.e.

$$k = k_n \to \infty, \ 1 \le k < n, \quad \text{and} \quad k_n = o(n), \ \text{as } n \to \infty. \tag{3.1}$$

3.1. Asymptotic normal behaviour of MOP and OMOP EVI-estimators

Let us consider the notation $\mathcal{N}(\mu, \sigma^2)$ for a normal RV with mean value μ and variance σ^2. Under the aforementioned second-order framework, in (2.2), and as a generalization of the results in de Haan and Peng (1998), Brilhante *et al.* (2013) derived, for the MOP EVI-estimators in (1.5) and $0 \le p \le 1/(2\xi)$, the asymptotic distributional representation,

$$\sqrt{k}\Big(\mathrm{H}_p(k) - \xi\Big) \ \stackrel{d}{=} \ \mathcal{N}\left(0, \tfrac{\xi^2(1-p\xi)^2}{1-2p\xi}\right) + \frac{(1-p\xi)\sqrt{k}A(n/k)}{1-\rho-p\xi}\big(1 + o_p(1)\big),$$

more generally valid for $p \in \mathbb{R}$ (Gomes and Caeiro, 2014). For the OMOP EVI-estimators, in (2.8), Brilhante *et al.* (2014) got the obvious validity of a similar asymptotic distributional representation, but with $p\xi$ replaced by φ_ρ, in (2.7), i.e.

$$\sqrt{k}\Big(\mathrm{H}^*(k) - \xi\Big) \ \stackrel{d}{=} \ \mathcal{N}\left(0, \tfrac{\xi^2(1-\varphi_\rho)^2}{1-2\varphi_\rho}\right) + \frac{(1-\varphi_\rho)\ \sqrt{k}A(n/k)}{1-\varphi_\rho-\rho}\big(1 + o_p(1)\big).$$

The asymptotic variance increases when p moves away from $p = 0$, but the bias decreases and, at optimal levels in the sense of minimal RMSE, the OMOP EVI-estimators outperform the H EVI-estimators.

Under the same conditions as before, but with $\mathrm{CH}(k)$ given in (2.4) and assuming that (2.3) holds, an adequate estimation of the second-order parameters, (β, ρ), enables to guarantee that $\sqrt{k}\big(\mathrm{CH}(k) - \xi\big)$ can be asymptotically normal with variance also equal to ξ^2 but with a null mean value. Indeed, from the results in Caeiro *et al.* (2005), we know that it is possible to get

$$\sqrt{k}\Big(\mathrm{CH}(k) - \xi\Big) \stackrel{d}{=} \mathcal{N}\left(0, \xi^2\right) + o_p\Big(\sqrt{k}A(n/k)\Big).$$

On the basis of the results in the aforementioned papers, and generally denoting by $\mathrm{E}(k)$ any of the EVI-estimators in (1.5) and (2.8), we can state the following theorem.

Theorem 3.1. (de Haan and Peng, 1998; Caeiro *et al.*, 2005; Brilhante *et al.*, 2013, 2014)**.** *Under the validity of the first-order condition, in* (2.1)*, and for intermediate sequences* $k = k_n$*, i.e. if* (3.1) *holds, the classes of* EVI*-estimators* $\mathrm{H}_p(k)$*, in* (1.5)*, for* $p < 1/\xi$*, and the* EVI*-estimators in* (2.4) *and* (2.8) *are consistent for the estimation of* ξ*. If we assume the validity of the second-order condition in* (2.2) *and additionally assume that we are working with values of* k *such that* $\lambda_{_A} := \lim_{n\to\infty} \sqrt{k}\, A(n/k)$ *is finite, we can then guarantee that for* $p < 1/(2\xi)$ *whenever dealing with* $\mathrm{H}_p(k)$*,*

$$\sqrt{k}\,(\mathrm{E}(k) - \xi) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \mathcal{N}\left(\lambda_{_A} b_\bullet, \sigma_\bullet^2\right),$$

where

$$b_{_{\mathrm{H}_p}} = \frac{1-p\xi}{1-\rho-p\xi}, \quad b_{_{\mathrm{H}^*}} = \frac{1-\varphi_\rho}{1-\rho-\varphi_\rho},$$
$$\sigma^2_{_{\mathrm{H}_p}} = \frac{\xi^2(1-p\xi)^2}{1-2p\xi}, \quad \sigma^2_{_{\mathrm{H}^*}} = \frac{\xi^2(1-\varphi_\rho)^2}{1-2\varphi_\rho}.$$

If we further assume to be working in Hall–Welsh *class of models in* (2.3)*, and estimate* β *and* ρ *consistently through* $\hat\beta$ *and* $\hat\rho$*, with* $\hat\rho - \rho = o_p(1/\ln n)$*, we get the aforementioned normal behaviour also for* $\mathrm{E} = \mathrm{CH}$*, in* (2.4)*, but now with* $b_{_{\mathrm{CH}}} = 0$ *and* $\sigma^2_{_{\mathrm{CH}}} = \sigma^2_{_{\mathrm{H}}} = \xi^2$*.*

Remark 3.1. Note again that $\sigma^2_{_{\mathrm{H}}} < \sigma^2_{_{\mathrm{H}_p}}$ for all $\xi > 0$ and $0 \neq p < 1/\xi$. The other way round, $b_{_{\mathrm{H}}} \geq b_{_{\mathrm{H}_p}}$ for all ξ. And as can be seen in Brilhante *et al.* (2013; 2014), at the optimal p, $\mathrm{H}_p(k)$ can asymptotically outperform $\mathrm{H}(k)$ at optimal levels in the sense of minimal RMSE, in the whole (ξ,ρ)-plane. As far as we know, such a property is so far achieved only by this class of EVI-estimators. See also Paulauskas and Vaiciulis (2013).

3.2. Asymptotic behaviour of PORT-MOP EVI-estimators

Note first that if there is a possible shift s in the model, i.e. if the CDF $F(x) \equiv F_s(x) = F(x;s)$ depends on (x,s) through the difference $x - s$, the parameter ξ does not change, as mentioned above in Section 2.3, but the parameter ρ, as well as the A-function, in (2.2), depend on such a shift s, i.e. $\rho = \rho_s$, $A = A_s$, and

$$(A_s(t),\rho_s) := \begin{cases} \big(-\xi s/U_0(t), -\xi\big), & \text{if } \xi+\rho_0 < 0 \ \wedge \ s \neq 0, \\ \big(A_0(t) - \xi s/U_0(t), \rho_0\big), & \text{if } \xi+\rho_0 = 0 \ \wedge \ s \neq 0, \\ \big(A_0(t),\rho_0\big), & \text{otherwise.} \end{cases}$$

Further details on the influence of such a shift in $\big(\beta, \rho, A(\cdot)\big)$ and on the estimation of 'shape' and 'scale' second-order parameters can be found in Henriques-Rodrigues *et al.* (2014, 2015).

To study the asymptotic properties of the PORT-MOP (and PORT-OMOP) EVI-estimators for $p \neq 0$, it is convenient to study first the behaviour of the statistics,

$$W_p(k;q) := \frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{n-i+1:n}-X_{n_q:n}}{X_{n-k:n}-X_{n_q:n}}\right)^p, \qquad p\neq 0, \tag{3.2}$$

for $X = X_0 \frown F_0$. Indeed,

$$\mathrm{H}_p\Big(k;\underline{\mathbf{X}}_n^{(q)}\Big) = \frac{1-W_p^{-1}(k;q)}{p} \quad \text{if} \quad p\neq 0. \tag{3.3}$$

Remark 3.2. Note that with

$$Q_r(k;q) = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{i}{k}\right)^r \frac{X_{n-i+1:n}-X_{n_q:n}}{X_{n-k:n}-X_{n_q:n}},$$

the statistics studied in Caeiro *et al.* (2014), we get, with $W_p(k;q)$ given in (3.2), $W_1(k;q) = Q_0(k;q)$.

Remark 3.3. It is also worth noting that, as already detected in Fraga Alves *et al.* (2009), for invariant versions of the mixed moment, and in Caeiro *et al.* (2014), for invariant versions of the Pareto probability weighted moment EVI-estimators, due to the fact that

$$X_{\lfloor nq\rfloor+1:n} - U_0(1/(1-q)) = O_p\Big(1/\sqrt{n}\Big),$$

$X_{n_q:n}$ can be replaced by the q-quantile

$$\chi_q := U_0(1/(1-q)). \tag{3.4}$$

The asymptotic behaviour of the statistics $W_p(k;q)$, in (3.2), comes then straightforwardly from the behaviour of the non-shifted statistics, as stated in the following proposition.

Theorem 3.2. *Under the second order framework in* (2.2)*, and for intermediate k, i.e. whenever* (3.1) *holds, we can guarantee, under general broad conditions, the asymptotic normality of* $W_p(k;q)$*, in* (3.2)*. Indeed, we can write, for* $p\xi < 1/2$,

$$\begin{aligned} W_p(k;q) \;\stackrel{d}{=}\; & \frac{1}{1-p\xi} + \frac{\sigma_p(\xi)\mathcal{N}(0,1)}{\sqrt{k}} + \frac{pA_0(n/k)(1+o_p(1))}{(1-p\xi)(1-p\xi-\rho_0)} \\ & + \frac{p\xi\chi_q(1+o_p(1))}{(1-p\xi)(1-(p-1)\xi)U_0(n/k)}, \end{aligned} \tag{3.5}$$

where

$$\sigma_p^2(\xi) := \frac{(p\xi)^2}{(1-p\xi)^2(1-2p\xi)}. \tag{3.6}$$

Proof: It is well-known that $U_0(X_{i:n}) \stackrel{d}{=} Y_{i:n}$, where Y is a standard unit Pareto RV, with CDF $F_Y(y) = 1 - 1/y,\ y > 1$. Moreover, $Y_{n-i+1:n}/Y_{n-k:n} \stackrel{d}{=} Y_{k-i+1:k},\ 1 \le i \le k$. Under the second order framework in (2.2), and thinking on the fact that we are now working with $s = 0$ due to the location invariance property of the statistics in (3.2), we can write

$$\begin{aligned}\frac{X_{n-i+1:n}}{X_{n-k:n}} &\stackrel{d}{=} \frac{U_0\big(\frac{Y_{n-i+1:n}}{Y_{n-k:n}}Y_{n-k:n}\big)}{U_0(Y_{n-k:n})}\\ &\stackrel{d}{=} Y_{k-i+1:k}^{\xi}\left(1 + \frac{Y_{k-i+1:k}^{\rho}-1}{\rho}A_0(Y_{n-k:n})(1+o_p(1))\right).\end{aligned}$$

Next, with the notation $\chi_q = U_0(1/(1-q))$, already introduced in (3.4),

$$\begin{aligned}\frac{X_{n-i+1:n}-\chi_q}{X_{n-k:n}-\chi_q} &= \frac{X_{n-i+1:n}}{X_{n-k:n}}\left(\frac{1-\chi_q/X_{n-i+1:n}}{1-\chi_q/X_{n-k:n}}\right)\\ &= \frac{X_{n-i+1:n}}{X_{n-k:n}}\left(1 + \frac{\chi_q}{X_{n-k:n}}\left(1 - \frac{X_{n-k:n}}{X_{n-i+1:n}}\right)(1+o_p(1))\right).\end{aligned}$$

Consequently,

$$\begin{aligned}W_p(k;q) &:= \frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{n-i+1:n}-X_{n_q:n}}{X_{n-k:n}-X_{n_q:n}}\right)^p\\ &= \frac{1}{k}\sum_{i=1}^{k}\left(\frac{X_{n-i+1:n}}{X_{n-k:n}}\left(1+\frac{\chi_q}{X_{n-k:n}}\left(1-\frac{X_{n-k:n}}{X_{n-i+1:n}}\right)(1+o_p(1))\right)\right)^p,\end{aligned}$$

and we can write

$$\begin{aligned}W_p(k;q) \stackrel{d}{=} &\ \frac{1}{k}\sum_{i=1}^{k}Y_{i:k}^{p\xi} + \frac{p\xi\chi_q}{U_0(n/k)}\frac{1}{k}\sum_{i=1}^{k}Y_{i:k}^{p\xi}\ \frac{Y_{i:k}^{-\xi}-1}{-\xi}(1+o_p(1))\\ &+ \frac{p}{k}\sum_{i=1}^{k}Y_{i:k}^{p\xi}\ \frac{Y_{i:k}^{\rho}-1}{\rho}A_0(n/k)(1+o_p(1)).\end{aligned}$$

Since, for $p\xi < 1$

$$\frac{1}{k}\sum_{i=1}^{k}Y_{i:k}^{p\xi} \xrightarrow{\mathbb{P}} \frac{1}{1-p\xi}$$

and if we further assume that $\rho < 0$,

$$\frac{1}{k}\sum_{i=1}^{k}Y_{i:k}^{p\xi}\left(\frac{Y_{i:k}^{\rho}-1}{\rho}\right) \xrightarrow{\mathbb{P}} \frac{1}{(1-p\xi)(1-p\xi-\rho)},$$

equation (3.5) follows. Moreover, $\sigma_p^2(\xi)$, given in (3.6), is merely the variance of $\sum_{i=1}^{k} Y_{i:k}^{p\xi}/k = \sum_{i=1}^{k} Y_i^{p\xi}/k$. $\square$

We next state the main theoretical result in this article, related to the shift invariant versions of the EVI-estimators in (1.5) and (2.8), i.e. the shift-invariant EVI-estimators, generally denoted $\mathrm{E}^{(q)}(k)$ in (2.10). Again, the asymptotic variance is kept at the same level of the unshifted EVI-estimators, but the dominant component of bias changes only in a few cases.

Theorem 3.3. *Under the second order framework in* (2.2)*, with* $p\xi < 1/2$*, and for intermediate* k*, i.e. if* (3.1) *holds, the asymptotic bias of the* PORT-MOP *and* PORT-OMOP EVI-*estimators, in* (2.10)*, is going to be ruled by*

$$B(t) = \begin{cases} \xi\chi_q/U_0(t), & \textit{if } \xi+\rho_0<0 \wedge \chi_q\neq 0, \\[2ex] A_0(t)+\xi\chi_q/U_0(t), & \textit{if } \xi+\rho_0=0 \wedge \chi_q\neq 0, \\[2ex] A_0(t), & \textit{otherwise}, \end{cases}$$

with χ_q *defined in* (3.4)*. If we assume that* $\sqrt{k}\,A_0(n/k) \to \lambda_{_A}$ *and/or* $\sqrt{k}/U_0(n/k) \to \lambda_{_U}$*, finite, as* $n\to\infty$*, and with* E *denoting either* H_p *or* H^**, as given in* (2.10)*,*

$$\sqrt{k}\left(\mathrm{E}^{(q)}(k)-\xi\right) \underset{n\to\infty}{\overset{d}{\longrightarrow}} \mathcal{N}\left(b_{\mathrm{E}|q}, \sigma^2_{\mathrm{E}}\right),$$

where

$$b_{\mathrm{E}|q} = \begin{cases} \frac{\xi(1-p\xi)\chi_q}{1-(p-1)\xi}\lambda_{_U}, & \textit{if } \xi+\rho_0<0 \wedge \chi_q\neq 0, \\[2ex] \frac{1-p\xi}{1-(p-1)\xi}\,\lambda_{_A} + \frac{\xi(1-p\xi)\chi_q}{1-(p-1)\xi}\lambda_{_U}, & \textit{if } \xi+\rho_0=0 \wedge \chi_q\neq 0, \\[2ex] \frac{1-p\xi}{1-p\xi-\rho_0}\,\lambda_{_A}, & \textit{otherwise}. \end{cases}$$

Proof: For $p\neq 0$, (3.3) and the use of Taylor's expansion $(1+x)^{-1} = 1-x+o(x)$, as $x\to 0$, enables us to get

$$\begin{aligned} \mathrm{H}_p^{(q)}(k) \;&\overset{d}{=}\; \xi + \frac{\sigma_p(\xi)(1-p\xi)^2\mathcal{N}(0,1)(1+o_p(1))}{|p|\sqrt{k}} \\ &\qquad + \frac{(1-p\xi)A_0(n/k)(1+o_p(1))}{(1-p\xi-\rho_0)} + \frac{\xi(1-p\xi)\chi_q(1+o_p(1))}{(1-(p-1)\xi)U_0(n/k)}. \end{aligned}$$

Consequently, the result in the theorem follows. □

4. FINITE SAMPLE PROPERTIES OF THE EVI-ESTIMATORS

We have implemented multi-sample Monte-Carlo simulation experiments of size 5000×20, i.e. 20 independent replicates with 5000 runs each, for the classes of MOP and PORT-MOP EVI-estimators associated with $p = p_\ell = 2\ell/(5\xi)$, $\ell = 0,\ 1,\ 2$, and also for the OMOP and PORT-OMOP EVI-estimators. The values $q = 0$ and $q = 0.25$ were considered. We further proceeded to the comparison with the MVRB and the PORT MVRB EVI-estimators, for the same values of q as mentioned above. Sample sizes from $n = 100$ until $n = 5000$ were simulated from a set of underlying models that include the ones shown here as an illustration, the EV model, with CDF $F(x) = \mathrm{EV}_\xi(x)$, with $\mathrm{EV}_\xi(x)$ given in (1.1), $\xi = 0.1, 0.25$, and the *Student-*t_ν, with $\nu = 4, 2$ degrees-of-freedom ($\xi = 1/\nu = 0.25, 0.5$). For details on multi-sample simulation, see Gomes and Oliveira (2001), among others. For the EV parents, results are presented essentially for $q = 0$, the value of q associated with the best performance of the PORT methodology for these models. For Student parents we consider $q = 0.25$. This is due to the fact that for the Student model the left endpoint is infinite and we cannot thus consider $q = 0$ (see Araújo Santos *et al.*, 2006, and Gomes *et al.*, 2008a, for further details related to the topic).

Remark 4.1. Note that, as already stated in the aforementioned articles dealing with a PORT framework, if there are only positive observed values in the sample, we gain nothing with the use of the PORT methodology. The other way round, if there are negative elements in the sample, as happens with EV and Student models and, in practice, with log-returns in financial data, among other types of data, the gain is quite high, as we shall see in the following. This is the main reason for the choice of the aforementioned parents.

4.1. Mean values and mean square error patterns as k-functionals

For each value of n and for each of the above-mentioned models, we have first simulated the mean value (E) and the RMSE of the estimators under consideration, as functions of the number of top order statistics k involved in the estimation. Apart from the MOP, H_p, in (1.5), $p = 0$ $\big(\mathrm{H}_0 \equiv \mathrm{H}\big)$ and $p = p_\ell = 2\ell/(5\xi)$, $\ell = 1$ (for which asymptotic normality holds), and $\ell = 2$ (where only consistency was proved), the OMOP (H^*), in (2.8), and the MVRB (CH) EVI-estimators, in (2.4), we have also included their PORT versions, respectively given in (2.10) and (2.11), for the above mentioned values of q.

The results are illustrated in Figure 1, for an EV_ξ underlying parent, with $\xi = 0.25$ and $q = 0$. In this case, and for all k, there is a clear reduction in

RMSE, as well as in bias, with the obtention of estimates closer to the target value ξ, particularly when we consider H_{p_2} and the associated PORT-version. However, at optimal levels, even the PORT-H^* and PORT-H_{p_1} versions beat the MVRB EVI-estimators. Indeed, the PORT-H_{p_1} can even beat the PORT-MVRB EVI-estimators, as happens in this illustration.

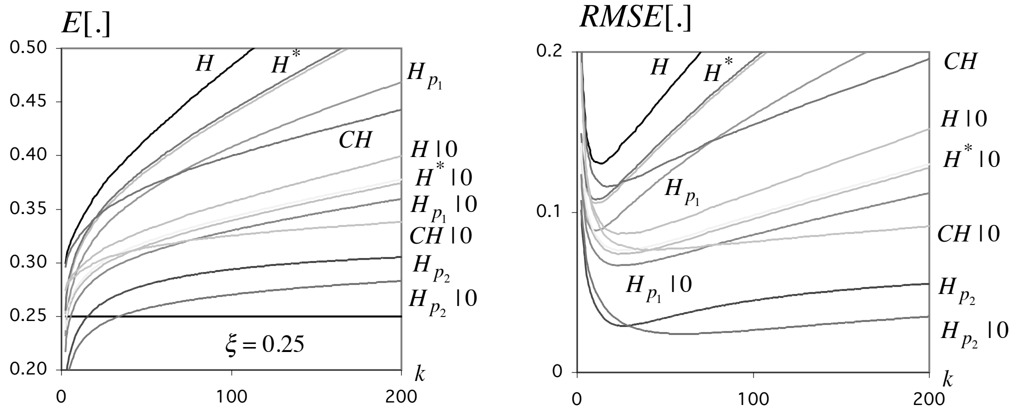

Figure 1: Mean values (*left*) and root mean square errors (*right*) of H, H^* (OMOP), CH, and H_p, $p = p_\ell = 2\ell/(5\xi)$, $\ell = 1, 2$ (MOP), together with their PORT versions, associated with $q = 0$ and generally denoted $\bullet|0$, for $EV_{0.25}$ underlying parents and sample size $n = 1000$.

Similar patterns have been obtained for all other simulated models, with the PORT-MVRB outperforming the PORT-MOP only in a few cases and for large sample sizes n.

4.2. Mean values and relative efficiency indicators at optimal levels

Table 1 is also related to the EV_ξ model, with $\xi = 0.25$. We there present, for different sample sizes n, the simulated mean values at optimal levels (levels where RMSEs are minima as functions of k) of the EVI-estimators under consideration in this study. Information on standard errors, computed on the basis of the 20 replicates with 5000 runs each, are available from the authors, upon request. Among the estimators considered, and distinguishing 3 regions, a first one with (H, CH, H^*, H_{p_1}), a second one with the associated PORT versions, (H|0, CH|0, $H^*|0$, $H_{p_1}|0$), and a third one with (H_{p_2}, $H_{p_2}|0$), for which an asymptotic normal behaviour is not available, the one providing the smallest squared bias is underlined and written in **bold** whenever there is an out-performance of the behaviour achieved in the previous region.

Table 1: Simulated mean values of the semi-parametric EVI-estimators under consideration, at their simulated optimal levels for underlying $EV_{0.25}$ parents.

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$
H	0.4202	0.3915	0.3646	0.3482	0.3348	0.3212
CH	0.3816	0.3716	0.3533	0.3416	0.3295	0.3174
H^*	0.3398	0.3351	0.3303	0.3226	0.3167	0.3082
H_{p_1}	**0.3059**	**0.3077**	**0.3034**	**0.3013**	**0.2998**	**0.2940**
H|0	0.3663	0.3464	0.3261	0.3154	0.3053	0.2957
CH|0	0.3510	0.3369	0.3210	0.3114	0.3033	0.2945
$H^*\|0$	0.3292	0.3208	0.3106	0.3046	0.2980	0.2904
$H_{p_1}\|0$	**0.3052**	**0.3001**	**0.2963**	**0.2928**	**0.2895**	**0.2848**
H_{p_2}	0.2723	0.2698	0.2669	0.2651	0.2638	0.2620
$H_{p_2}\|0$	**0.2669**	**0.2650**	**0.2625**	**0.2614**	**0.2603**	**0.2590**

We have further computed the Hill estimator, given in (1.5) when $p = 0$, at the simulated value of $k_{0|0} = \arg\min_k \text{RMSE}\big(\text{H}_0(k)\big)$, the simulated optimal k in the sense of minimum RMSE, not relevant in practice, but providing an indication of the best possible performance of Hill's estimator. Such an estimator is denoted by $\widetilde{\text{H}}_{00}$. For any of the estimators under study, generally denoted $\text{E}(k)$, we have also computed E_0, the estimator $\text{E}(k)$ computed at the simulated value of $k_{0|E} := \arg\min_k \text{RMSE}\big(\text{E}(k)\big)$. The simulated indicators are

$$\text{REFF}_{\text{E}|0} := \frac{\text{RMSE}\left(\widetilde{\text{H}}_{00}\right)}{\text{RMSE}\left(\text{E}_0\right)}. \tag{4.1}$$

Remark 4.2. Note that, as usual, an indicator higher than one means a better performance than the Hill estimator. Consequently, the higher these indicators are, the better the associated EVI-estimators perform, comparatively to $\widetilde{\text{H}}_{00}$.

Table 2: Simulated values of RMSE_0 (first row) and of $\text{REFF}_{\bullet|0}$ indicators, for underlying $EV_{0.25}$ parents.

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$
RMSE_0	0.246	0.200	0.157	0.133	0.113	0.092
CH	1.3256	1.2374	1.1711	1.1304	1.1008	1.0716
H^*	1.4391	1.3384	1.2491	1.2021	1.1653	1.1333
H_{p_1}	**1.9307**	**1.7443**	**1.5646**	**1.4633**	**1.3785**	**1.2999**
H|0	1.4875	1.4991	1.5169	1.5309	1.5405	1.5542
CH|0	1.9212	1.8505	1.7790	1.7366	1.6958	1.6633
$H^*\|0$	1.8966	1.8156	1.7511	1.7217	1.6995	1.6868
$H_{p_1}\|0$	**2.3988**	**2.2171**	**2.0478**	**1.9564**	**1.8828**	**1.8230**
H_{p_2}	6.4033	5.6755	4.9396	4.4849	4.0943	3.6784
$H_{p_2}\|0$	**7.5643**	**6.7594**	**5.9369**	**5.4315**	**4.9769**	**4.4991**

Again as an illustration of the results obtained, we present Table 2. In the first row, we provide RMSE_0, the RMSE of $\widetilde{\mathrm{H}}_{00}$, so that we can easily recover the RMSE of all other estimators. The following rows provide the REFF-indicators for the different EVI-estimators under study. A similar mark (underlined and **bold**) is used for the highest REFF indicator, again considering the aforementioned three regions.

For a better visualization of the results presented in Table 1 and Table 2, we further present Figure 2. Due to the high REFF-indicators of H_{p_2} and associated PORT estimators, we present them in a different scale, at the top of Figure 2, *right*, the one related to the REFF-indicators.

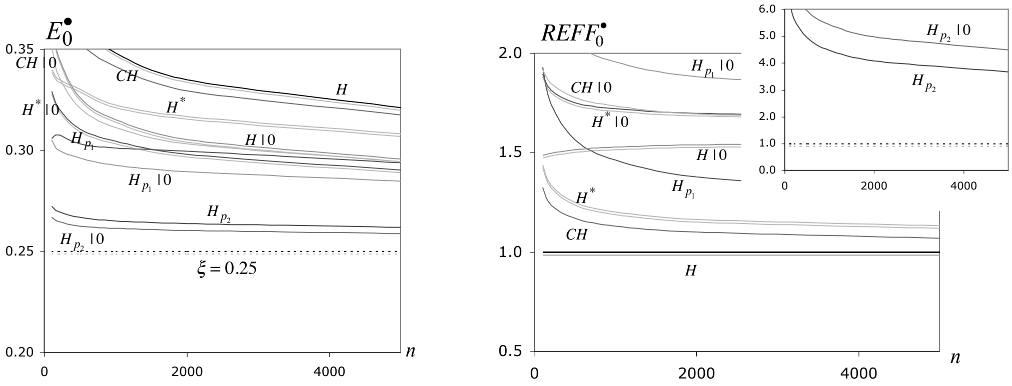

Figure 2: Mean values (*left*) and REFF-indicators (*right*) at optimal levels of the different estimators under study, for an underlying $\mathrm{EV}_{0.25}$ parent and sample sizes $n = 100(100)500$ and $500(500)5000$.

Tables 3–4, 5–6 and 7–8 are similar to Tables 1–2, respectively for $\mathrm{EV}_{0.1}$, Student-t_4 and Student-t_2 underlying parents.

Table 3: Simulated mean values of the semi-parametric EVI-estimators under consideration, at their simulated optimal levels for underlying $\mathrm{EV}_{0.1}$ parents.

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$
H	0.2918	0.2644	0.2403	0.2225	0.2089	0.1952
CH	0.2714	0.2544	0.2341	0.2214	0.2076	0.1946
H^*	0.1895	0.1745	0.1605	0.1516	0.1442	0.1464
H_{p_1}	**0.1601**	**0.1496**	**0.1396**	**0.1330**	**0.1274**	**0.1315**
H|0	0.2404	0.2191	0.2009	0.1895	0.1801	0.1688
CH|0	0.2346	0.2176	0.1989	0.1887	0.1793	0.1689
$\mathrm{H}^*\|0$	0.1611	0.1499	0.1435	0.1441	0.1458	0.14440
$\mathrm{H}_{p_1}\|0$	**0.1400**	**0.1317**	**0.1278**	**0.1290**	**0.1271**	**0.1291**
H_{p_2}	0.1159	0.1149	0.1133	0.1127	0.1114	0.1105
$\mathrm{H}_{p_2}\|0$	**0.1131**	**0.1124**	**0.1110**	**0.1104**	**0.1098**	**0.1090**

Table 4: Simulated values of RMSE_0 (first row) and of $\mathrm{REFF}_{\bullet|0}$ indicators, for underlying $\mathrm{EV}_{0.1}$ parents.

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$
RMSE_0	0.2524	0.2109	0.1732	0.1511	0.1329	0.1136
CH	1.1778	1.1141	1.0684	1.0450	1.0293	1.0186
H^*	2.0954	1.9436	1.7846	1.6708	1.5618	1.4483
H_{p_1}	**3.0221**	**2.7527**	**2.4758**	**2.2837**	**2.1044**	**1.9174**
H|0	1.4292	1.4185	1.4153	1.4093	1.4006	1.3967
CH|0	1.5680	1.5140	1.4760	1.4509	1.4290	1.4134
$\mathrm{H}^*\|0$	2.5865	2.3621	2.1291	1.9935	1.8775	1.7709
$\mathrm{H}_{p_1}\|0$	**3.5906**	**3.2188**	**2.8408**	**2.6229**	**2.4277**	**2.2369**
H_{p_2}	12.1731	10.5862	9.1739	8.3307	7.6068	6.8415
$\mathrm{H}_{p_2}\|0$	**13.3178**	**11.6827**	**10.1972**	**9.2846**	**8.5188**	**7.6951**

Table 5: Simulated mean values of the semi-parametric EVI-estimators under consideration, at their simulated optimal levels for underlying Student-t_4 parents ($\xi = 0.25$).

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$
H	0.3607	0.3392	0.3167	0.3055	0.2959	0.2862
CH	0.3109	0.3104	0.3005	0.2939	0.2879	0.2805
H^*	0.3236	0.3135	0.3028	0.2959	0.2891	0.2818
H_{p_1}	**0.2964**	**0.2914**	**0.2881**	**0.2844**	**0.2810**	**0.2765**
H|0.25	0.3078	0.2935	0.2806	0.2728	0.2672	0.2613
CH|0.25	0.2869	0.2783	**0.2686**	**0.2641**	**0.2599**	**0.2561**
$\mathrm{H}^*\|0.25$	0.2923	0.2861	0.2764	0.2699	0.2658	0.2607
$\mathrm{H}_{p_1}\|0.25$	**0.2797**	**0.2762**	0.2709	0.2671	0.2640	0.2599
H_{p_2}	0.2662	0.2646	0.2616	0.2604	0.2589	0.2575
$\mathrm{H}_{p_2}\|0.25$	**0.2613**	**0.2591**	**0.2570**	**0.2558**	**0.2550**	**0.2539**

Table 6: Simulated values of RMSE_0 (first row) and of $\mathrm{REFF}_{\bullet|0}$ indicators, for underlying Student-t_4 parents.

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$
RMSE_0	0.1830	0.1431	0.1059	0.0854	0.0696	0.0535
CH	1.4349	1.3982	1.3615	1.3223	**1.2834**	**1.2358**
H^*	1.2984	1.2280	1.1625	1.1297	1.1046	1.0822
H_{p_1}	**1.7501**	**1.5845**	**1.4200**	**1.3285**	1.2554	1.1819
H|0.25	1.6242	1.6823	1.7745	1.8702	1.9850	2.1777
CH|0.25	2.4005	**2.5115**	**2.7219**	**2.8846**	**3.1153**	**3.5054**
$\mathrm{H}^*\|0.25$	1.9459	1.9360	1.9712	2.0386	2.1329	2.3108
$\mathrm{H}_{p_1}\|0.25$	**2.4223**	2.3048	2.2245	2.2166	2.2410	2.3346
H_{p_2}	5.3556	4.7308	4.0399	3.5993	3.2243	2.7827
$\mathrm{H}_{p_2}\|0.25$	**6.6674**	**6.0186**	**5.2884**	**4.8145**	**4.3920**	**3.8883**

Table 7: Simulated mean values of the semi-parametric EVI-estimators under consideration, at their simulated optimal levels for underlying Student-t_2 parents ($\xi = 0.5$).

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$
H	0.6015	0.5769	0.5560	0.5439	0.5355	0.5257
CH	**0.4644**	**0.5059**	**0.5117**	**0.5073**	**0.5041**	**0.5019**
H^*	0.5823	0.5671	0.5510	0.5404	0.5324	0.5233
H_{p_1}	0.5553	0.5486	0.5393	0.5325	0.5261	0.5182
H|0.25	0.5203	0.5139	0.5063	0.5037	0.5020	0.5009
CH|0.25	**0.4885**	**0.4940**	**0.4974**	**0.4988**	**0.4995**	**0.4997**
$\mathrm{H}^*\|0.25$	0.5194	0.5142	0.5070	0.5035	0.5018	0.5009
$\mathrm{H}_{p_1}\|0.25$	0.5186	0.5130	0.5078	0.5048	0.5023	0.5011
H_{p_2}	0.5206	0.5168	0.5137	0.5111	0.5086	0.5053
$\mathrm{H}_{p_2}\|0.25$	0.5120	0.5096	0.5072	0.5051	0.5036	0.5018

Table 8: Simulated values of RMSE_0 (first row) and of $\mathrm{REFF}_{\bullet|0}$ indicators, for underlying Student-t_2 parents.

	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$
RMSE_0	0.2028	0.1528	0.1078	0.0835	0.0652	0.0470
CH	0.9803	**1.4180**	**1.7059**	**1.9437**	**2.2267**	**2.6414**
H^*	1.1363	1.1047	1.0811	1.0695	1.0666	1.0644
H_{p_1}	**1.4333**	1.3224	1.2344	1.1957	1.1841	1.1844
H|0.25	1.8476	1.9699	2.2126	2.4120	2.6709	3.0481
CH|0.25	**2.4870**	**2.6495**	**2.9310**	**3.1988**	**3.5307**	**4.0413**
$\mathrm{H}^*\|0.25$	1.9814	2.0820	2.3071	2.5030	2.7652	3.1490
$\mathrm{H}_{p_1}\|0.25$	2.2140	2.2306	2.3726	2.5269	2.7644	3.1234
H_{p_2}	3.7572	3.2811	2.7464	2.4304	2.2496	2.1766
$\mathrm{H}_{p_2}\|0.25$	**4.5942**	**4.1347**	**3.6354**	**3.3598**	3.2719	3.3502

Remark 4.3. As intuitively expected, $\mathrm{H}_{p|\bullet}$ are decreasing in p, approaching the true value of ξ, or all simulated models.

Remark 4.4. For adequate values of q and p, the PORT-MOP EVI-estimators are able to outperform the MVRB and even the PORT-MVRB, in some cases.

5. AN ADAPTIVE CHOICE OF (k, p, q) AND CONCLUDING REMARKS

Apart from heuristic choices based on sample path stability, similar to the ones in Neves *et al.* (2015), we suggest the use of the double-bootstrap methodology, briefly described in the following Section.

5.1. Bootstrap adaptive PORT-MOP EVI-estimation

A reasonably sophisticated and time-consuming algorithm, that has proved to work properly in many situations, is the double-bootstrap algorithm. The basic framework for such algorithm is related to the fact that for any class of EVI-estimators, generally denoted $\mathrm{E}(k)$,

$$k_{0|\mathrm{E}}(n) = \arg\min_k \mathrm{RMSE}\big(\mathrm{E}(k)\big) = k_{A|\mathrm{E}}(n)\big(1+o(1)\big), \tag{5.1}$$

with $k_{A|\mathrm{E}}(n) := \arg\min_k \mathrm{ARMSE}\big(\mathrm{E}(k)\big)$ and ARMSE standing for *asymptotic root mean square error*. The bootstrap methodology can then enable us to consistently estimate the optimal sample fraction, $k_{0|\mathrm{E}}(n)/n$, with $k_{0|\mathrm{E}}(n)$ given in (5.1), on the basis of a consistent estimator of $k_{A|\mathrm{E}}(n)$, in a way similar to the one used in Draisma *et al.* (1999), Danielson *et al.* (2001) and Gomes and Oliveira (2001), for the classical adaptive Hill EVI-estimation, performed through $\mathrm{H}(k) \equiv \mathrm{H}_0(k)$, in (1.4), in Brilhante *et al.* (2013), for the MOP EVI-estimation through $\mathrm{H}_p(k)$, in (1.5), in Gomes *et al.* (2011b, 2012), for second-order reduced-bias estimation, and in Gomes *et al.* (2015) for the CH and PORT-CH EVI-estimation.

The bootstrap methodology is applied to sub-samples of size $m_1 = o(n)$ and $m_2 = m_1^2/n$, is practically independent on m_1 for an adequate PORT EVI-estimation and it is essentially based on the relationship between the optimal sample fraction of the EVI-estimator under consideration, and the one of the auxiliary statistics

$$T_{k,n} \equiv T(k|\mathrm{E}) := \mathrm{E}([k/2]) - \mathrm{E}(k), \quad k = 2, ..., n-1,$$

which converge in probability to the known value zero, for any intermediate k, and have an asymptotic behaviour strongly related with the asymptotic behaviour of $\mathrm{E}(k)$. For details, see Gomes *et al.* (2015), where an algorithm for the optimal choice of (k, q) is provided for the PORT-MVRB EVI-estimators, in (2.11). Indeed, for the adaptive choice of (k, p, q) based on minimal bootstrap RMSE, an algorithm of the type of the one in Gomes *et al.* (2015) can be conceived with the inclusion of the MOP and PORT-MOP together with the Hill, the PORT-Hill, the MVRB and the PORT-MVRB. This is however a topic out of the scope of this article.

5.2. Overall comments

A few concluding remarks:

- For both mean values and RMSEs at optimal levels, and for all simulated models, if we restrict ourselves to the region of values of p where we can guarantee asymptotic normality, i.e. $p < 1/(2\xi)$, the best results were obtained for the value of p closer to $1/(2\xi)$, i.e. $p = 2/(5\xi)$. The OMOP is not at all competitive with the MOP, regarding both bias and MSE.
- For the simulated models, the MOP can clearly beat the MVRB, being beaten by the MVRB only for Student-t_2 parents. A similar comment applies to the behaviour of the PORT-MOP comparatively to the PORT-MVRB EVI-estimators.
- The improvement achieved with the use of the PORT-MOP EVI-estimation can be highly significant, as illustrated. Indeed, the PORT-MOP can, for an adequate (p, q) beat the MVRB EVI-estimators for all k, being often able to beat the optimal PORT-MVRB. This is surely due to the small increase in the variance and the high reduction of bias of the PORT-MOP comparatively with the PORT-MVRB, a topic not yet investigated, due to the deep involvement of a third-order framework.

ACKNOWLEDGMENTS

Research partially supported by National Funds through FCT – Fundação para a Ciência e a Tecnologia, project PEst-OE/MAT/UI0006/2014.

REFERENCES

[1] Araújo Santos, P.; Fraga Alves, M.I. and Gomes, M.I. (2006). Peaks over random threshold methodology for tail index and quantile estimation, *Revstat*, **4**(3), 227–247.

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

[3] Beirlant, J.; Caeiro, F. and Gomes, M.I. (2012). An overview and open research topics in the field of statistics of univariate extremes, *Revstat*, **10**(1), 1–31.

[4] Bingham, N.H.; Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, Cambridge University Press.

[5] Brilhante, F.; Gomes, M.I. and Pestana, D. (2013). A simple generalization of the Hill estimator, *Computational Statistics and Data Analysis*, **57**(1), 518–535.

[6] Brilhante, M.F.; Gomes, M.I. and Pestana, D. (2014). *The mean-of-order p extreme value index estimator revisited.* In "New Advances in Statistical Modeling and Application" (A. Pacheco, R. Santos, M.R. Oliveira and C.D. Paulino, Eds.), Studies in Theoretical and Applied Statistics, Selected Papers of the Statistical Societies, Springer-Verlag, Berlin and Heidelberg, 163–175.

[7] Caeiro, F.; Gomes, M.I. and Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator, *Revstat*, **3**(2), 113–136.

[8] Caeiro, F.; Gomes, M.I. and Henriques-Rodrigues, L. (2009). Reduced-bias tail index estimators under a third order framework, *Communications in Statistics—Theory & Methods*, **38**(7), 1019–1040.

[9] Caeiro, F.; Gomes, M.I. and Henriques-Rodrigues, L. (2014). A location invariant probability weighted moment EVI estimator, *International J. Computer Mathematics*, DOI: 10.1080/00207160.2014.975217

[10] Danielsson, J.; Haan, L. de; Peng, L. and de Vries, C.G. (2001). Using a bootstrap method to choose the sample fraction in the tail index estimation, *J. Multivariate Analysis*, **76**, 226–248.

[11] Draisma, G.; Haan, L. de; Peng, L. and Pereira, T.T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index, *Extremes*, **2**(4), 367–404.

[12] Fraga Alves, M.I.; Gomes, M.I. and Haan, L. de (2003). A new class of semi-parametric estimators of the second order parameter, *Portugaliae Mathematica*, **60**(2), 193–213.

[13] Fraga Alves, M.I.; Gomes, M.I.; Haan, L. de and Neves, C. (2009). The mixed moment estimator and location invariant alternatives, *Extremes*, **12**, 149–185.

[14] Geluk, J. and Haan, L. de (1987). *Regular Variation, Extensions and Tauberian Theorems*, CWI Tract 40, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands.

[15] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Mathematics*, **44**(6), 423–453.

[16] Gomes, M.I. and Caeiro, F. (2014). *Efficiency of partially reduced-bias mean-of-order-p versus minimum-variance reduced-bias extreme value index estimation.* In "Proceedings of COMPSTAT 2014" (M. Gilli, G. Gonzalez-Rodriguez and A. Nieto-Reyes, Eds.), The International Statistical Institute/International Association for Statistical Computing, 289–298.

[17] Gomes, M.I. and Guillou, A. (2014). Extreme value theory and statistics of univariate extremes: A review, *International Statistical Review*, DOI: 10.1111/insr.12058.

[18] Gomes, M.I. and Martins, M.J. (2002). "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter, *Extremes*, **5**(1), 5–31.

[19] Gomes, M.I. and Oliveira, O. (2001). The bootstrap methodology in Statistics of Extremes: choice of the optimal sample fraction, *Extremes*, **4**(4), 331–358.

[20] Gomes, M.I. and Pestana, D. (2007). A sturdy reduced-bias extreme quantile (VaR) estimator, *J. American Statistical Association*, **102**(477), 280–292.

[21] Gomes, M.I.; Fraga Alves, M.I. and Araújo Santos, P. (2008a). PORT Hill and moment estimators for heavy-tailed models, *Communications in Statistics—Simulation & Computation*, **37**, 1281–1306.

[22] Gomes, M.I.; Haan, L. de and Henriques-Rodrigues, L. (2008b). Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses, *J. Royal Statistical Society B*, **70**(1), 31–52.

[23] Gomes, M.I.; Henriques-Rodrigues, L. and Miranda, C. (2011a). Reduced-bias location-invariant extreme value index estimation: a simulation study, *Communications in Statistics—Simulation & Computation*, **40**(3), 424–447.

[24] Gomes, M.I.; Mendonça, S. and Pestana, D. (2011b). Adaptive reduced-bias tail index and VaR estimation via the bootstrap methodology, *Communications in Statistics—Theory & Methods*, **40**(16), 2946–2968.

[25] Gomes, M.I.; Figueiredo, F. and Neves, M.M. (2012). Adaptive estimation of heavy right tails: the bootstrap methodology in action, *Extremes*, **15**, 463–489.

[26] 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, *J. Statistical Computation and Simulation*, **83**(6), 1129–1144.

[27] Gomes, M.I.; Caeiro, F.; Henriques-Rodrigues, L. and Manjunath, B.G. (2015). *Bootstrap methods in statistics of extremes.* In "Handbook of Extreme Value Theory and Its Applications to Finance and Insurance" (F. Longin, Ed.), Handbook Series in Financial Engineering and Econometrics (Ruey Tsay Adv. Ed.), John Wiley and Sons, in press.

[28] Haan, L. de (1984). *Slow variation and characterization of domains of attraction*, In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), 31–48, D. Reidel, Dordrecht, Holland.

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

[30] Haan, L. de and Peng, L. (1998). Comparison of extreme value index estimators, *Statistica Neerlandica*, **52**, 60–70.

[31] Hall, P. (1982). On some simple estimates of an exponent of regular variation, *J. Royal Statistical Society B*, **44**, 37–42.

[32] Hall, P. and Welsh, A.H. (1985). Adaptative estimates of parameters of regular variation, *Annals of Statistics*, **13**, 331–341.

[33] Henriques-Rodrigues, L.; Gomes, M.I.; Fraga Alves, M.I. and Neves, C. (2014). PORT estimation of a shape second-order parameter, *Revstat*, **13**(3), 299–328.

[34] Henriques-Rodrigues, L.; Gomes, M.I. and Manjunath, B.G. (2015). Estimation of a scale second-order parameter related to the PORT methodology, *J. Statistical Theory and Practice*, **9**(3), 571–599.

[35] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics*, **3**, 1163–1174.

[36] Neves, M.M.; Gomes, M.I.; Figueiredo, F. and Prata-Gomes, D. (2015). Modeling extreme events: sample fraction adaptive choice in parameter estimation, *J. Statistical Theory and Practice*, **9**(1), 184–199.

[37] Paulauskas, V. and Vaiciulis, M. (2013). On the improvement of Hill and some others estimators, *Lithuanian Mathematical J.*, **53**, 336–355.

[38] Reiss, R.-D. and Thomas, M. (2001; 2007). *Statistical Analysis of Extreme Values, with Application to Insurance, Finance, Hydrology and Other Fields*, 2nd edition; 3rd edition, Birkhäuser Verlag.

REVSTAT – Statistical Journal
Volume 14, Number 3, June 2016, 297–309

SEQUENTIAL ESTIMATION OF A COMMON LOCATION PARAMETER OF TWO POPULATIONS

Authors: Agnieszka Stępień-Baran
– Institute of Applied Mathematics and Mechanics, University of Warsaw, Banacha 2, 02-097 Warszawa, Poland
A.Stepien-Baran@mimuw.edu.pl

Received: June 2014 Revised: January 2015 Accepted: January 2015

Abstract:

- The problem of sequentially estimating a common location parameter of two independent populations from the same distribution with an unknown location parameter and known but different scale parameters is considered in the case when the observations become available at random times. Certain classes of sequential estimation procedures are derived under a location invariant loss function and with the observation cost determined by convex functions of the stopping time and the number of observations up to that time.

Key-Words:

- *location parameter; location invariant loss function; minimum risk equivariant estimator; optimal stopping time; risk function.*

AMS Subject Classification:

- 62L12, 62L15, 62F10.

1. INTRODUCTION

The paper concerns the invariant problem of sequentially estimating a common location parameter of two independent populations from the same distribution with an unknown location parameter and known but different scale parameters, in the special case when the observations arrive at random times. For example, in studying the effectiveness of experimental safety devices of mobile constructions relevant data may become available only as a result of accidents. Medical data (such as data on drug abuse or an asymptomatic disease) can sometimes only be obtained when patients voluntarily seek help or are somehow otherwise identified and examined, at random times. Other examples are data resulting from an undersea survey of containerized radioactive waste, from archeological discoveries, from market research or from planning the assortment of production (when the orders come forward at random times).

The estimation problem of a common location of two independent populations has been extensively discussed in the literature. Rao and Reddy (1988) studied the estimation of the unknown common location parameter of two symmetric distributions with different scale parameters. They derived asymptotic distributions and the asymptotic relative efficiencies of proposed estimators: the mean, the median, the average of the mean and the median and the Hodges–Lehmann estimator. Baklizi (2004) considered estimation of the common location parameter of several exponentials. It is found that the proposed estimators are effective in taking advantage of the available prior information. Farsipour and Asgharzadeh (2002) investigated the problem of estimating the common mean of two normal distributions. They derived a class of risk unbiased estimators which linearly combines the means of the two samples from both distributions. Mitra and Sinha (2007) studied some aspects of the problem of estimation of a common mean of two normal populations from an asymptotic point of view. They also considered the Bayes estimate of the common mean under Jeffrey's prior. Chang *et al.* (2012) considered the problem of estimating the common mean of two normal distributions with unknown ordered variances. They gave a broad class of estimators which includes the estimators proposed by Nair (1982) and Elfessi *et al.* (1992) and showed that the estimators stochastically dominate the estimators which do not take into account the order restriction on variances, including the one given by Graybill and Deal (1959). Then they proposed a broad class of individual estimators of two ordered means when unknown variances are ordered.

The problem of estimating with delayed observations was investigated by Starr *et al.* (1976), who considered the case of Bayes estimation of a mean of normally distributed observations with known variance. Some of their results were generalized by Magiera (1996). He dealt with estimation of the mean value parameter of the exponential family of distributions. Jokiel-Rokita and Stępień (2009) studied the model with delayed observations for estimating a location parameter.

We consider the following model. Let the samples $(X_1, ..., X_{n_1})$ and $(Y_1, ..., Y_{n_2})$ be independent and have a joint distribution P_θ with a Lebesgue p.d.f.

$$f\left(\frac{x_1-\theta}{\sigma_1}, ..., \frac{x_{n_1}-\theta}{\sigma_1}\right),$$

and

$$f\left(\frac{y_1-\theta}{\sigma_2}, ..., \frac{y_{n_2}-\theta}{\sigma_2}\right),$$

respectively, where f is known, σ_1, $\sigma_2 > 0$ are known and different scale parameters, and $\theta \in \mathbb{R}$ is an unknown location parameter.

The set of observations is bounded, i.e., the statistician can receive at most $N = n_1 + n_2$ observations. It is assumed that X_i is observed at time t_i, $i = 1, ..., n_1$, where $t_1, ..., t_{n_1}$ are the values of the order statistics of positive i.i.d. random variables $U_1, ..., U_{n_1}$ which are obtained before the conducted observations $X_1, ..., X_{n_1}$ and independent of $X_1, ..., X_{n_1}$. Similarly: Y_i is observed at time s_i, $i = 1, ..., n_2$, where $s_1, ..., s_{n_2}$ are the values of the order statistics of positive i.i.d. random variables $V_1, ..., V_{n_2}$ which are obtained before the conducted observations $Y_1, ..., Y_{n_2}$ and independent of $Y_1, ..., Y_{n_2}$. Furthermore, it is assumed that the samples $(U_1, ..., U_{n_1})$ and $(V_1, ..., V_{n_2})$ are independent.

Let

$$k_1(t) = \sum_{i=1}^{n_1} \mathbf{1}_{[0,t]}(U_i) \tag{1.1}$$

and

$$k_2(t) = \sum_{i=1}^{n_2} \mathbf{1}_{[0,t]}(V_i) \tag{1.2}$$

denote the number of observations which have been made by time $t \geq 0$ for the sample $(X_1, ..., X_{n_1})$ and $(Y_1, ..., Y_{n_2})$, respectively, and let $\mathcal{F}_{1,t} = \sigma\big\{k_1(r),\ r \leq t,\ X_1, ..., X_{k_1(t)}\big\}$ and $\mathcal{F}_{2,t} = \sigma\left\{k_2(r),\ r \leq t,\ Y_1, ..., Y_{k_2(t)}\right\}$ be the informations which is available at time t.

The problem is to estimate the parameter θ. If observation is stopped at time t, the loss incurred is defined by

$$\mathcal{L}_t(\theta, d) := \mathcal{L}(\theta, d) + c_A k_1(t) + c_B k_2(t) + c_1(t) + c_2(t), \tag{1.3}$$

where $\mathcal{L}(\theta, d)$ denotes the loss associated with estimation, when θ is the true value of the parameter and d is the chosen estimate. The functions $c_1(t)$ and $c_2(t)$ represents the cost of observing the processes up to time t ($k_1(t)$ and $k_2(t)$, respectively). It is supposed to be a differentiable and increasing convex functions such that $c_1(0) = 0$ and $c_2(0) = 0$. The constants $c_A \geq 0$ and $c_B \geq 0$ are the cost of taking one observation X_i and Y_i, respectively.

The family $\{P_\theta : \theta \in \mathbb{R}\}$ is invariant under the location transformations $x \mapsto x + \alpha$ ($y \mapsto y + \alpha$) with $\alpha \in \mathbb{R}$. Consequently, the decision problem is invariant under location transformations if and only if $\mathcal{L}(\theta, a) = \mathcal{L}(\theta + \alpha, a + \alpha)$ for all $\alpha \in \mathbb{R}$, which is equivalent to

$$\mathcal{L}(\theta, a) = \delta(a - \theta) \tag{1.4}$$

for a Borel function $\delta(\cdot)$ on $\mathbb{R}$. An estimator d of the parameter θ is *location equivariant* if and only if

$$d(X_1 + \alpha, ..., X_{n_1} + \alpha, Y_1 + \alpha, ..., Y_{n_2} + \alpha) = d(X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2}) + \alpha.$$

Suppose that we agree to take at least one observation. If we observe the process for $t \geq t_1$ units of time, then the conditional expected loss, given $k_1(t)$ and $k_2(t)$, associated with an equivariant estimator $d\left(\mathbf{X}_{k_1(t)}, \mathbf{Y}_{k_2(t)}\right)$ based on the random size samples $\mathbf{X}_{k_1(t)} = \left(X_1, ..., X_{k_1(t)}\right)$ and $\mathbf{Y}_{k_2(t)} = \left(Y_1, ..., Y_{k_2(t)}\right)$ is of the form

$$\begin{aligned} \mathcal{R}_t\left(\theta, d\left(\mathbf{X}_{k_1(t)}, \mathbf{Y}_{k_2(t)}\right)\right) :&= E_\theta\left[\mathcal{L}_t\left(\theta, d\left(\mathbf{X}_{k_1(t)}, \mathbf{Y}_{k_2(t)}\right)\right) \Big| k_1(t), k_2(t)\right] \\ &= h_1(k_1(t)) + h_2(k_2(t)) + c_1(t) + c_2(t), \end{aligned} \tag{1.5}$$

where E_θ means the expectation with respect to the conditional distribution given θ. The functions h_1 and h_2 depend only on the loss function δ.

The form of the risk function $\mathcal{R}_t(\theta, d)$, given by (1.5), follows from the fact that the risk of any equivariant estimator of the parameter θ in the invariant problem of estimation is independent of θ (see e.g. Lehmann and Casella 1998, Theorem 3.1.4). Hence, if an equivariant estimator exists which minimizes the constant risk, it is called *the minimum risk equivariant (MRE) estimator.*

In Section 2 we present the method of finding a stopping time which minimizes the expected risk associated with a MRE estimator of the parameter θ over all stopping times. We consider a situation when the common distributions of the random variables $U_1, ..., U_{n_1}$ and $V_1, ..., V_{n_2}$, respectively, which can be interpreted as the lifetimes of n_1 and n_2 objects are known exactly. In Section 3 we apply the results of Section 2 to estimate a common location parameter of two normal distributions under the squared error loss and a LINEX loss function. Additionally, in Section 4 some illustrative simulations are given.

2. THE OPTIMAL STOPPING TIME

Suppose that in the estimation problem of the parameter θ with the loss function $\mathcal{L}(\theta, d)$ there exists an MRE estimator, denoted by d^*. We look for a stopping time τ^* which minimizes the expected risk

$$E\left[\mathcal{R}_\tau\left(\theta, d^*\left(\mathbf{X}_{k_1(\tau)}, \mathbf{Y}_{k_2(\tau)}\right)\right)\right] = E[h_1(k_1(\tau)) + h_2(k_2(\tau)) + c_1(\tau) + c_2(\tau)] \tag{2.1}$$

over all stopping times $\tau \geq t_1$, $\tau \in \mathcal{T}$, where $\mathcal{T}$ denotes the class of $(\mathcal{F}_{1,t}, \mathcal{F}_{2,t})$-measurable functions. Such a stopping time will be called *an optimal stopping time*. Then we construct *an optimal sequential estimation procedure* of the form $\left(\tau^*, d^*\left(\mathbf{X}_{k_1(\tau^*)}, \mathbf{Y}_{k_2(\tau^*)}\right)\right)$.

Let the random variables $U_1, ..., U_{n_1}$ be independent and have a common known distribution function G_1. Suppose that $G_1(0) = 0$, $G_1(t) > 0$ for $t > 0$, G_1 is absolutely continuous with density g_1, and g_1 is the right hand derivative of G_1 on $(0, \infty)$. Denote the class of such G_1 by $\mathcal{G}_1$. Let $\zeta_1 = \sup\{t : G_1(t) < 1\}$, and $\rho_1(t) = g_1(t)[1 - G_1(t)]^{-1}$, $0 \leq t < \zeta_1$, denote the failure rate. Under the above assumptions the process $k_1(t)$, given by (1.1), is a nonstationary Markov chain with respect to $\mathcal{F}_{1,t}$, $0 \leq t \leq \zeta_1$ (see Starr *et al.* (1976)). The random variables $V_1, ..., V_{n_2}$ satisfy the analogous assumptions. Namely, let the random variables $V_1, ..., V_{n_2}$ be independent and have a common known distribution function G_2. Suppose that $G_2(0) = 0$, $G_2(t) > 0$ for $t > 0$, G_2 is absolutely continuous with density g_2, and g_2 is the right hand derivative of G_2 on $(0, \infty)$. Denote the class of such G_2 by $\mathcal{G}_2$. Let $\zeta_2 = \sup\{t : G_2(t) < 1\}$, and $\rho_2(t) = g_2(t)[1 - G_2(t)]^{-1}$, $0 \leq t < \zeta_2$, denote the failure rate. Under the above assumptions the process $k_2(t)$, given by (1.2), is a nonstationary Markov chain with respect to $\mathcal{F}_{2,t}$, $0 \leq t \leq \zeta_2$.

The infinitesimal operator $\mathcal{A}_{1,t}$ of the processes $k_1(t)$ at $\widetilde{h}_1$ is defined by

$$\mathcal{A}_{1,t}\widetilde{h}_1(k) := \lim_{s \to 0^+} s^{-1} E\left[\widetilde{h}_1(k_1(t+s)) - \widetilde{h}_1(k_1(t)) | k_1(t) = k\right]. \tag{2.2}$$

The domain $D_{\mathcal{A}_{1,t}}$ of $\mathcal{A}_{1,t}$ is the set of all bounded Borel measurable functions $\widetilde{h}_1$ on the set $\{0, 1, ..., n_1\}$ for which the limit in (2.2) exists boundedly pointwise for every $k \in \{0, 1, ..., n_1\}$. The infinitesimal operator $\mathcal{A}_{2,t}$ of the processes $k_2(t)$ is defined analogously.

To determine an optimal stopping time we use the following lemma which provides the form of the infinitesimal operator $\mathcal{A}_{1,t}$ of the process $k_1(t)$, given by (1.1).

Lemma 2.1. *Let $\widetilde{h}_1$ be a given real-valued function on the set $\{0, 1, ..., n_1\}$. The infinitesimal operator $\mathcal{A}_{1,t}$ of the process $k_1(t)$, given by (1.1), at $\widetilde{h}_1$ is of the form*

$$\mathcal{A}_{1,t}\widetilde{h}_1(k) = (n_1 - k)\left[\widetilde{h}_1(k+1) - \widetilde{h}_1(k)\right]\rho_1(t).$$

Proof: Fix $k \in \{0, 1, ..., n_1\}$. It is clear that

$$E\left[\widetilde{h}_1(k_1(t+s)) - \widetilde{h}_1(k_1(t))|k_1(t)=k\right] =$$

$$= \sum_{i=k+1}^{n_1} \left[\widetilde{h}_1(i) - \widetilde{h}_1(k)\right] P\left(k_1(t+s)=i|k_1(t)=k\right)$$

$$= \left[\widetilde{h}_1(k+1) - \widetilde{h}_1(k)\right] P\left(k_1(t+s)=k+1|k_1(t)=k\right)$$

$$+ \sum_{i=k+2}^{n_1} \left[\widetilde{h}_1(i) - \widetilde{h}_1(k)\right] P\left(k_1(t+s)=i|k_1(t)=k\right)$$

$$= \left[\widetilde{h}_1(k+1) - \widetilde{h}_1(k)\right] (n_1-k)\frac{G_1(t+s)-G_1(t)}{1-G_1(t)} \left[\frac{1-G_1(t+s)}{1-G_1(t)}\right]^{n_1-k-1}$$

$$+ \sum_{i=k+2}^{n_1} \left[\widetilde{h}_1(i) - \widetilde{h}_1(k)\right] P\left(k_1(t+s)=i|k_1(t)=k\right)$$

$$\leq \left[\widetilde{h}_1(k+1) - \widetilde{h}_1(k)\right] (n_1-k)\frac{G_1(t+s)-G_1(t)}{1-G_1(t)} \left[\frac{1-G_1(t+s)}{1-G_1(t)}\right]^{n_1-k-1}$$

$$+ 2\sup_{i\leq n_1} \left|\widetilde{h}_1(i)\right| P\left(k_1(t+s) \geq k+2|k_1(t)=k\right)$$

$$= \left[\widetilde{h}_1(k+1) - \widetilde{h}_1(k)\right] (n_1-k)\frac{G_1(t+s)-G_1(t)}{1-G_1(t)} \left[\frac{1-G_1(t+s)}{1-G_1(t)}\right]^{n_1-k-1}$$

$$+ 2\sup_{i\leq n_1} \left|\widetilde{h}_1(i)\right| \left\{1 - \left[\frac{1-G_1(t+s)}{1-G_1(t)}\right]^{n_1-k} \left[1-(n_1-k)\frac{G_1(t+s)-G_1(t)}{[1-G_1(t+s)]}\right]\right\}.$$

Now it is easy to see that

$$\lim_{s\to 0^+} \frac{E\left[\widetilde{h}_1(k_1(t+s)) - \widetilde{h}_1(k_1(t))|k_1(t)=k\right]}{s} =$$

$$= (n_1-k)\left[\widetilde{h}_1(k+1) - \widetilde{h}_1(k)\right]\rho_1(t)$$

and the lemma is proved. □

The infinitesimal operator $\mathcal{A}_{2,t}$ of the processes $k_2(t)$ is calculated analogously and we have

$$\mathcal{A}_{2,t}\widetilde{h}_2(k) = (n_2-k)\left[\widetilde{h}_2(k+1) - \widetilde{h}_2(k)\right]\rho_2(t).$$

Let $\widetilde{h}_1(k) = h_1(k)$ for $k = 1, ..., n_1$ and $\widetilde{h}_1(0) = 0$, and $\widetilde{h}_2(k) = h_2(k)$ for $k = 1, ..., n_2$ and $\widetilde{h}_2(0) = 0$. The following theorem determines the optimal stopping time τ^* for a large class of possible h_1 and h_2.

Theorem 2.1. *Suppose that $G_1 \in \mathcal{G}_1$ has non-increasing failure rate ρ_1, $G_2 \in \mathcal{G}_2$ has non-increasing failure rate ρ_2, and the functions $h_1(k)$ and $h_2(k)$ in* *formula (1.5) are such that* $h_1(k)-h_1(k+1)$ *is non-increasing for* $k \in \{1,...,n_1-1\}$ *and* $h_2(k)-h_2(k+1)$ *is non-increasing for* $k \in \{1,...,n_2-1\}$*. Then the stopping time*

$$\begin{aligned}\tau^* &= \inf\Big\{t \ge t_1:\ \mathcal{A}_{1,t}\widetilde{h}_1(k_1(t)) + \mathcal{A}_{2,t}\widetilde{h}_2(k_2(t)) + c_1'(t) + c_2'(t) \ge 0\Big\}\\ &= \inf\Big\{t \ge t_1:\ (n_1-k_1(t))[h_1(k_1(t)) - h_1(k_1(t)+1)]\rho_1(t)\\ &\qquad + (n_2-k_2(t))[h_2(k_2(t)) - h_2(k_2(t)+1)]\rho_2(t) \le c_1'(t)+c_2'(t)\Big\} \qquad (2.3)\end{aligned}$$

minimizes the expected risk given by (2.1) over all stopping times $\tau \ge t_1$, $\tau \in \mathcal{T}$.

Proof: The proof follows Starr *et al.* (1976), Theorem 2.1. Using Dynkin's formula, we have

$$\begin{aligned}&E[\widetilde{h}_1(k_1(\tau)) + \widetilde{h}_2(k_2(\tau)) + c_1(\tau) + c_2(\tau)] = \\ &\qquad = E\left\{\int_0^\tau [\mathcal{A}_{1,t}\widetilde{h}_1(k_1(t)) + \mathcal{A}_{2,t}\widetilde{h}_2(k_2(t)) + c_1'(t) + c_2'(t)]dt\right\}\end{aligned}$$

for all stopping times τ. In particular for $\tau \ge t_1$ we have

$$\begin{aligned}&E[h_1(k_1(\tau^*)) + h_2(k_2(\tau^*)) + c_1(\tau^*) + c_2(\tau^*)] - \\ &\qquad - E[h_1(k_1(\tau)) + h_2(k_2(\tau)) + c_1(\tau) + c_2(\tau)] = \\ &= E[\widetilde{h}_1(k_1(\tau^*)) + \widetilde{h}_2(k_2(\tau^*)) + c_1(\tau^*) + c_2(\tau^*)] \\ &\quad - E[\widetilde{h}_1(k_1(\tau)) + \widetilde{h}_2(k_2(\tau)) + c_1(\tau) + c_2(\tau)] \\ &= E\left\{\int_\tau^{\tau^*} [\mathcal{A}_{1,t}\widetilde{h}_1(k_1(t)) + \mathcal{A}_{2,t}\widetilde{h}_2(k_2(t)) + c_1'(t) + c_2'(t)]dt\mathbf{1}(\tau<\tau^*)\right\} \\ &\quad - E\left\{\int_{\tau^*}^{\tau} [\mathcal{A}_{1,t}\widetilde{h}_1(k_1(t)) + \mathcal{A}_{2,t}\widetilde{h}_2(k_2(t)) + c_1'(t) + c_2'(t)]dt\right\}\mathbf{1}(\tau>\tau^*). \qquad (2.4)\end{aligned}$$

Taking into account the assumptions concerning the function $h_1(k)$, $h_2(k)$, $c_1(t)$, $c_2(t)$, $\rho_1(t)$ and $\rho_2(t)$ we have that (2.4) is less or equal to zero. Thus, the stopping time τ^* is optimal. □

3. SPECIAL CASE

In this section we use the solutions of Section 2 to estimate a common location parameter of two normal distributions under the squared error loss

$$\mathcal{L}(\theta, d) = (d-\theta)^2 \qquad (3.1)$$

and under a LINEX loss function

$$\mathcal{L}(\theta, d) = \exp[a(d-\theta)] - a(d-\theta) - 1, \qquad (3.2)$$

where $a \neq 0$. Taking MRE estimators as optimal estimators of a location parameter of two normal distributions, we construct optimal sequential estimation procedures under the aforementioned loss functions in the model with observations which are available at random times.

Let X_i, $i = 1, ..., n_1$, be independent random variables from the normal distribution $\mathcal{N}(\theta, \sigma_1^2)$ and Y_i, $i = 1, ..., n_2$, be independent random variables from the normal distribution $\mathcal{N}(\theta, \sigma_2^2)$, where $\theta \in \mathbb{R}$ is an unknown location parameter and σ_1, $\sigma_2 > 0$ are known. We assume that the samples $(X_1, ..., X_{n_1})$ and $(Y_1, ..., Y_{n_2})$ are independent and $\sigma_1 \neq \sigma_2$.

Let

$$\overline{X}_{k_1(t)} = \frac{1}{k_1(t)} \sum_{i=1}^{k_1(t)} X_i, \quad \overline{Y}_{k_2(t)} = \frac{1}{k_2(t)} \sum_{i=1}^{k_2(t)} Y_i$$

denote the sample means based on the random size sample $\mathbf{X}_{k_1(t)} = \big(X_1, ..., X_{k_1(t)}\big)$ and $\mathbf{Y}_{k_2(t)} = \big(Y_1, ..., Y_{k_2(t)}\big)$, respectively, where $k_1(t)$ is given by (1.1) and $k_2(t)$ is given by (1.2).

The following theorem provides the MRE estimator of the parameter θ and the corresponding risk function under the loss function given by (3.1) and (3.2), respectively.

Theorem 3.1. *For any stopping time t*

(**a**) *If the loss function is given by (3.1), then the MRE estimator of the parameter θ is*

$$d_S^* \left(\mathbf{X}_{k_1(t)}, \mathbf{Y}_{k_2(t)}\right) = \omega \overline{X}_{k_1(t)} + (1-\omega)\overline{Y}_{k_2(t)}$$

with $\omega \in (0,1)$, and the risk function of the estimator d_S^* has the form

$$\mathcal{R}_t\left(\theta, d_S^*\right) = \frac{\omega^2 \sigma_1^2}{2k_1(t)} + \frac{(1-\omega)^2 \sigma_2^2}{2k_2(t)} + c_A k_1(t) + c_B k_2(t) + c_1(t) + c_2(t).$$

(**b**) *If the loss function is given by (3.2), then the MRE estimator of the parameter θ is*

$$\begin{aligned} d_L^*\big(\mathbf{X}_{k_1(t)}, \mathbf{Y}_{k_2(t)}\big) = \ & \omega \left(\overline{X}_{k_1(t)} - \frac{a\sigma_1^2}{2k_1(t)}\right) + (1-\omega) \left(\overline{Y}_{k_2(t)} - \frac{a\sigma_2^2}{2k_2(t)}\right) \\ & + \omega(1-\omega)\ a\left(\frac{\sigma_1^2}{2k_1(t)} + \frac{\sigma_2^2}{2k_2(t)}\right) \end{aligned}$$

with $\omega \in (0,1)$, and the risk function of the estimator d_L^* has the form

$$\mathcal{R}_t(\theta, d_L^*) = \frac{\omega^2 a^2 \sigma_1^2}{2k_1(t)} + \frac{(1-\omega)^2 a^2 \sigma_2^2}{2k_2(t)} + c_A k_1(t) + c_B k_2(t) + c_1(t) + c_2(t).$$

Proof: The forms of the MRE estimators d_S^* and d_L^* are obtained from the general formula for the MRE estimators of the location parameter under the loss function (1.4) (see e.g. Shao (2003), Theorem 4.5). The formulas for the risk functions $\mathcal{R}_t(\theta, d_S^*)$ and $\mathcal{R}_t(\theta, d_L^*)$ follow from straightforward calculations. □

On the basis of Theorems 2.1 and 3.1 we construct optimal sequential estimation procedures of the form $\left(\tau^*, d^*\left(\mathbf{X}_{k_1(\tau^*)}, \mathbf{Y}_{k_2(\tau^*)}\right)\right)$, where τ^* is defined by (2.3), and d^* is the corresponding sequential MRE estimator of θ based on the random size samples $\mathbf{X}_{k_1(\tau^*)}$ and $\mathbf{Y}_{k_2(\tau^*)}$.

The next theorem determines the optimal sequential estimation procedure under the loss function $\mathcal{L}(\theta, d)$ given by (3.1) and (3.2), respectively.

Theorem 3.2. *Suppose that $G_1 \in \mathcal{G}_1$ has non-increasing failure rate ρ_1 and $G_2 \in \mathcal{G}_2$ has non-increasing failure rate ρ_2.*

(**a**) *Under the loss function $\mathcal{L}_t(\theta, d)$ given by (1.3) with $\mathcal{L}(\theta, d)$ of the form (3.1), the sequential estimation procedure $\left(\tau_S^*, d_S^*\left(\mathbf{X}_{k_1(\tau_S^*)}, \mathbf{Y}_{k_2(\tau_S^*)}\right)\right)$, where*

$$\tau_S^* = \inf\left\{t \geq t_1 :\ (n_1 - k_1(t))\left[\frac{\omega^2\sigma_1^2}{2k_1(t)} - \frac{\omega^2\sigma_1^2}{2(k_1(t)+1)} - c_A\right]\rho_1(t)\right.$$
$$\left. + (n_2 - k_2(t))\left[\frac{(1-\omega)^2\sigma_2^2}{2k_2(t)} - \frac{(1-\omega)^2\sigma_2^2}{2(k_2(t)+1)} - c_B\right]\rho_2(t) \leq c_1'(t) + c_2'(t)\right\}$$

and

$$d_S^*\left(\mathbf{X}_{k_1(\tau_S^*)}, \mathbf{Y}_{k_2(\tau_S^*)}\right) = \omega\overline{X}_{k_1(\tau_S^*)} + (1-\omega)\overline{Y}_{k_2(\tau_S^*)}$$

is optimal.

(**b**) *Under the loss function $\mathcal{L}_t(\theta, d)$ given by (1.3) with $\mathcal{L}(\theta, d)$ of the form (3.2), the sequential estimation procedure $\left(\tau_L^*, d_L^*\left(\mathbf{X}_{k_1(\tau_L^*)}, \mathbf{Y}_{k_2(\tau_L^*)}\right)\right)$, where*

$$\tau_L^* = \inf\left\{t \geq t_1 :\ (n_1 - k_1(t))\left[\frac{\omega^2 a^2\sigma_1^2}{2k_1(t)} - \frac{\omega^2 a^2\sigma_1^2}{2(k_1(t)+1)} - c_A\right]\rho_1(t)\right.$$
$$\left. + (n_2 - k_2(t))\left[\frac{(1-\omega)^2 a^2\sigma_2^2}{2k_2(t)} - \frac{(1-\omega)^2 a^2\sigma_2^2}{2(k_2(t)+1)} - c_B\right]\rho_2(t) \leq c_1'(t) + c_2'(t)\right\}$$

and

$$\begin{aligned} d_L^*\left(\mathbf{X}_{k_1(\tau_L^*)}, \mathbf{Y}_{k_2(\tau_L^*)}\right) = {} & \omega\left(\overline{X}_{k_1(\tau_L^*)} - \frac{a\sigma_1^2}{2k_1(\tau_L^*)}\right) \\ & + (1-\omega)\left(\overline{Y}_{k_2(\tau_L^*)} - \frac{a\sigma_2^2}{2k_2(\tau_L^*)}\right) \\ & + \omega(1-\omega)\, a\left(\frac{\sigma_1^2}{2k_1(\tau_L^*)} + \frac{\sigma_2^2}{2k_2(\tau_L^*)}\right) \end{aligned}$$

is optimal.

Proof: We have to show that the assumptions of Theorem 2.1 are satisfied, i.e., the functions $h_1(k) - h_1(k+1)$ and $h_2(k) - h_2(k+1)$ are non-increasing on the set $\{1, ..., n_1-1\}$ and $\{1, ..., n_2-1\}$, respectively. Hence, we need to verify the condition $2h_1(k+1) - h_1(k) - h_1(k+2) \le 0$ and $2h_2(k+1) - h_2(k) - h_2(k+2) \le 0$, which are equivalent to $h_1(k+1) \le (h_1(k) + h_1(k+2))/2$ and $h_2(k+1) \le (h_2(k) + h_2(k+2))/2$. This can be reduced to the verification that h_1 and h_2 are convex on the interval $[1, n_1-1]$ and $[1, n_2-1]$, respectively. It is easy to see that

(**a**)
$$h_1''(k) = \frac{\omega^2\sigma_1^2}{k^3}, \quad h_2''(k) = \frac{(1-\omega)^2\sigma_2^2}{k^3}$$

and $h_1''(k) > 0$, $h_2''(k) > 0$ for $k \ge 1$;

(**b**)
$$h_1''(k) = \frac{\omega^2 a^2\sigma_1^2}{k^3}, \quad h_2''(k) = \frac{(1-\omega)^2 a^2\sigma_2^2}{k^3}$$

and $h_1''(k) > 0$, $h_2''(k) > 0$ for $k \ge 1$. □

4. SIMULATION RESULTS

In this section we present some results of the numerical study. The first table contains the results of the simulation study for $X_1, ..., X_{n_1} \sim \mathcal{N}(0,1)$, $n_1 = 30$ and $Y_1, ..., Y_{n_2} \sim \mathcal{N}(0,25)$, $n_2 = 50$: the means of τ_S^*, d_S^*, τ_L^* and d_L^* for $a = 2$, over the 1000 replications, when $\omega = 0.25$, $\rho_1(t) = 1\,(U_i \sim \mathcal{E}(1))$, $c_1(t) = t^2$ and $\rho_2(t) = (2\cdot\sqrt{3t})^{-1}, (V_i \sim \mathcal{W}e(1/2,3))$, $c_2(t) = e^t - 1$.

c_A	c_B	Mean(τ_S^*)	Mean(d_S^*)	Mean(τ_L^*)	Mean(d_L^*)
0.005	0.000001	0.2388	−0.0094	0.4705	−0.8777
0.000001	0.005	0.2339	−0.0099	0.4629	−0.8949
0.005	0.005	0.2281	0.0123	0.4537	−0.8943
0.000001	0.000001	0.2455	−0.0406	0.4687	−0.9203

The second table contains the results of the simulation study for $X_1, ..., X_{n_1} \sim \mathcal{N}(0,1)$, $n_1 = 30$ and $Y_1, ..., Y_{n_2} \sim \mathcal{N}(0,25)$, $n_2 = 50$: the means of τ_S^*, d_S^*, τ_L^* and d_L^* for $a = 2$, over the 1000 replications, when $\omega = 0.5$, $\rho_1(t) = (2\cdot\sqrt{3t})^{-1}\,(U_i \sim \mathcal{W}e(1/2,3))$, $c_1(t) = e^t - 1$ and $\rho_2(t) = 1\,(V_i \sim \mathcal{E}(1))$, $c_2(t) = t^2$.

c_A	c_B	Mean(τ_S^*)	Mean(d_S^*)	Mean(τ_L^*)	Mean(d_L^*)
0.005	0.000001	0.3634	0.0036	0.6073	−0.4517
0.000001	0.005	0.3657	−0.0088	0.6144	−0.4831
0.005	0.005	0.3610	−0.0097	0.6102	−0.4881
0.000001	0.000001	0.3665	0.0134	0.6165	−0.4453

The third table contains the results of the simulation study for $X_1,...,X_{n_1} \sim \mathcal{N}(0,1)$, $n_1 = 30$ and $Y_1,...,Y_{n_2} \sim \mathcal{N}(0,25)$, $n_2 = 50$: the means of τ_S^*, d_S^*, τ_L^* and d_L^* for $a = 2$, over the 1000 replications, when $\omega = 0.75$, $\rho_1(t) = (2 \cdot \sqrt{3t})^{-1}$ ($U_i \sim \mathcal{W}e(1/2,3)$), $c_1(t) = t^2$ and $\rho_2(t) = 1$ ($V_i \sim \mathcal{E}(1)$), $c_2(t) = e^t - 1$.

c_A	c_B	Mean(τ_S^*)	Mean(d_S^*)	Mean(τ_L^*)	Mean(d_L^*)
0.005	0.000001	0.2975	−0.0107	0.5129	−0.1816
0.000001	0.005	0.2956	0.0284	0.5076	−0.1504
0.005	0.005	0.2908	0.0014	0.5010	−0.1417
0.000001	0.000001	0.3021	0.0106	0.5110	−0.1710

Simulation results above are consistent with expectations. The both procedures are working properly. In case of Linex loss function, decision function is biased, however it is MRE estimator because ω is fixed. It could be applicable especially in a case when one sample is more preferable than second one.

REFERENCES

[1] Rao, M.S. and Reddy, M.K. (1988). Estimation of a common location parameter of two distributions, *Aligarh J. Statist.*, **8**, 6–13.

[2] Baklizi, A. (2004). Shrinkage estimation of the common location parameter of several exponentials, *Comm. Statist. Simulation Comput.*, **33**(2), 321–339.

[3] Farsipour, N.S. and Asgharzadeh, A. (2002). On the admissibility of estimators of the common mean of two normal populations under symmetric and asymmetric loss functions, *South African Statist. J.*, **36**(1), 39–54.

[4] Mitra, P.K. and Sinha, B.K. (2007). On some aspects of estimation of a common mean of two independent normal populations, *J. Statist. Plann. Inference*, **137**(1), 184–193.

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

[6] Nair, K.A. (1982). An estimator of the common mean of two normal populations, *J. Statist. Plann. Inference*, **6**(2), 119–122.

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

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

[9] Starr, N.; Wardrop, R. and Woodroofe, M. (1976). Estimating a mean from delayed observations, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **35**(2), 103–113.

[10] Magiera, R. (1996). On a class of sequential estimation problems for one-parameter exponential families, *Sankhyā Ser. A*, **58**(1), 160–170.

[11] Jokiel-Rokita, A. and Stępień, A. (2009). Sequential estimation of a location parameter from delayed observations, *Statist. Papers*, **50**(2), 363–372.

[12] Lehmann, E.L. and Casella, G. (1998). *Theory Of Point Estimation*, Springer-Verlag, New York.

[13] Shao, J. (2003). *Mathematical Statistics*, Springer-Verlag, New York.

REVSTAT – Statistical Journal
Volume 14, Number 3, June 2016, 311–325

AN RKHS FRAMEWORK FOR SPARSE FUNCTIONAL VARYING COEFFICIENT MODEL

Authors: Behdad Mostafaiy
– Department of Statistics, Faculty of Mathematical Science, Shahid Beheshti University, Iran
behdad.mostafaiy@gmail.com

Mohammad Reza Faridrohani
– Department of Statistics, Faculty of Mathematical Science, Shahid Beheshti University, Iran
m_faridrohani@sbu.ac.ir

S. Mohammad E. Hosseininasab
– Department of Statistics, Faculty of Mathematical Science, Shahid Beheshti University, Iran
m_hosseininasab@sbu.ac.ir

Received: July 2014 Revised: November 2014 Accepted: February 2015

Abstract:

- We study functional varying coefficient model in which both the response and the predictor are functions of a common variable such as time. We demonstrate the estimation of the slope function for the case of sparse and noise-contaminated longitudinal data. So far, a few methods have been introduced based on varying coefficient model. To estimate the slope function, we consider a regularization method using a reproducing kernel Hilbert space framework. Despite the generality of the regularization method, the procedure is easy to implement. Our numerical results show that the introduced procedure performs well in some senses.

Key-Words:

- *functional varying coefficient model; regularization; reproducing kernel Hilbert space; sparsity.*

AMS Subject Classification:

- 62G05, 62P10.

1. INTRODUCTION

Due to rapid development of science and technology, it is possible to collect data that are naturally functions. This type of data , that are referred as functional data, has many applications in various fields of science, including, for example, environmental science, chemometrics, engineering, biomedical studies, public health, and econometrics. Functional data analysis deals with situations in which the individual observed data are infinite-dimensional, such as curves. See Ramsay and Silverman (2002, 2005) and Ferraty and Vieu (2006) for comprehensive discussions on methods and applications for functional data.

Functional linear model is one of the most useful methods to explore the relationship between two sets of observations. There are various types of functional linear model that have been widely studied in the literature. In this paper, we consider a functional linear model where one observe a random sample $\{(X_i, Y_i) : i = 1, 2, ..., n\}$ corresponds to functional varying coefficient model, i.e.,

$$Y(t) = \alpha(t) + \beta(t)X(t) + Z(t), \tag{1.1}$$

where α and β are smoothed functions, and $Z(t)$ is a noise term with zero mean and finite variance. Without loss of generality, we assume that $E[X(t)] = E[Y(t)] = 0$, then the functional linear model (1.1) becomes,

$$Y(t) = \beta(t)X(t) + Z(t). \tag{1.2}$$

Relation (1.2) models Y via X pointwisely, and allows β to vary with time. Fan and Zhang (2008) have provided a review of statistical methods proposed for various varying coefficient models according to three approaches. These approaches are based on polynomial spline, smoothing splines and local polynomial smoothing. See also Wu *et al.* (1998), Huang *et al.* (2002, 2004), Hoover *et al.* (1998), Chiang *et al.* (2001), Wu and Chiang (2000), and Kauermann and Tutz (1999). Fan and Zhang (1999), Wang and Xia (2009), and Lin and Ying (2001) applied another approaches for varying coefficient models. Most of these papers did not examine sparse and irregular designs and face some problems in implementing these designs.

In many experiments though, for example most longitudinal studies, the functional trajectories of the involved smooth random processes are not directly observable. In these cases, the observed data are noisy, sparse and irregularly spaced measurements of these trajectories.

Following the notation in Yao *et al.* (2005a), let U_{ij} and V_{ij} the jth observations of the random trajectories $X_i(\cdot)$ and $Y_i(\cdot)$ at a random time points T_{ij}, respectively, where T_{ij} are independently drawn from a distribution on compact domain $T \subset \mathbb{R}$. Assume that U_{ij} and V_{ij} are contaminated with measurement errors ε_{ij} and ϵ_{ij}, respectively. These errors are assumed to be i.i.d. with mean zero and finite variance σ_X^2 for ε_{ij} and σ_Y^2 for ϵ_{ij}. Therefore, the models may be represented in the following forms:

$$(1.3)\qquad \begin{aligned} U_{ij} &= X_i(T_{ij}) + \varepsilon_{ij}, & j &= 1, ..., m; & i &= 1, ..., n\,, \\ V_{ij} &= Y_i(T_{ij}) + \epsilon_{ij}, & j &= 1, ..., m; & i &= 1, ..., n\,. \end{aligned}$$

Functional data analysis of model (1.3) has been extended by Yao et. al (2005a, 2005b). See also Li and Hsing (2010), and Yang et. al (2011). Şentürk and Müller (2010), and Şentürk and Nguyen (2011) have considered functional varying coefficient in model (1.1). The model given in Şentürk and Müller (2010) is a model with one covariate process that incorporates a history index. Their estimation approach is based on least square estimation. Şentürk and Nguyen (2011) have studied a model with error-prone time-dependent variables and time-invariant covariates. They used covariance representation techniques to estimate the slope function. More references that studied varying coefficient models for model (1.3) include Şentürk and Müller (2008), Noh and Park (2010), Chiou *et al.* (2012), and Şentürk *et al.* (2013).

In this paper, we assume that the slope function β belongs to a reproducing kernel Hilbert space (RKHS) $\mathcal{H}$, and investigate the regularization method for estimating β. By simulation, we show that our estimation method perform well as sampling frequency and sample size increase. We do our simulation study in two different settings. One is when locations are same and equidistant for all curves, that is, $T_{1j} = T_{2j} = \cdots = T_{nj} = \frac{2j}{2m+1}$ for all $j = 1, 2, ..., m$. Another setting is when T_{ij} are independently sampled from T. These settings are referred as *common design* and *stochastic design*, respectively (see Cai and Yuan (2011)).

The paper is organized as follows. In section 2, our estimation procedure is introduced. The numerical results are given in Section 3. Section 4 collects the obtained results and discusses possible extensions of our work.

2. ESTIMATION PROCEDURE

In this section, we introduce a regularization method for estimating the slope function β using a reproducing kernel Hilbert space (RKHS) framework. First, we review some basic facts of RKHS. A Hilbert space $\mathcal{H}$ of functions on a set T with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is called an RKHS if there exists a bivariate function $K(\cdot, \cdot)$ on $T \times T$ such that for every $t \in T$ and $f \in \mathcal{H}$,

(i) $K(\cdot, t) \in \mathcal{H}$,

(ii) $f(t) = \langle\, f,\, K(\cdot, t)\,\rangle_{\mathcal{H}}$.

Relation (ii) is termed the reproducing property of K, and K is called reproducing kernel of $\mathcal{H}$. Every reproducing kernel determine unique RKHS. In addition, an RKHS has unique reproducing kernel. For any $s_1, ..., s_{m'}, {s'}_1, ..., {s'}_{n'} \in T$ and $a_1, ..., a_{m'}, b_1, ..., b_{n'} \in \mathbb{R}$, we have

$$\langle \sum_{i=1}^{m'} a_i K(\cdot, s_i), \sum_{j=1}^{n'} b_j K(\cdot, {s'}_j) \rangle_{\mathcal{H}} = \sum_{i=1}^{m'} \sum_{j=1}^{n'} a_i b_j K(s_i, {s'}_j) \tag{2.1}$$

More details on RKHS can be found in Aronszajn (1950), Berlinet and Thomas-Agnan (2004) and Wahba (1990).

Now, we investigate the method of regularization to estimate β. We assume that $\beta \in \mathcal{H}(K)$, where $\mathcal{H}(K)$ is an RKHS with reproducing kernel K. We estimate β via

$$\hat{\beta}_\lambda = \underset{\beta \in \mathcal{H}(K)}{\arg\min} \left\{ \ell_{mn}(\beta) + \lambda \|\beta\|^2_{\mathcal{H}(K)} \right\} \tag{2.2}$$

where

$$\ell_{mn}(\beta) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} (V_{ij} - U_{ij}\beta(T_{ij}))^2$$

and $\lambda > 0$ is tuning parameter that control tradeoff between fidelity to the data measured by ℓ_{mn} and smoothness of the solution measured by RKHS norm.

Remark 1. We can define minimization problem (2.2) in more general sense. For example, one may replace $\|\beta\|^2_{\mathcal{H}(K)}$ by $J(\beta)$ and then define

$$\hat{\beta}_\lambda = \underset{\beta \in \mathcal{H}(K)}{\arg\min} \{\ell_{mn}(\beta) + \lambda J(\beta)\}$$

where the penalty functional J is a squared semi-norm on $\mathcal{H}(K)$ such that the null space

$$\mathcal{H}_0(K) = \{g \in \mathcal{H}(K) : J(g) = 0\}$$

be a finite dimensional linear subspace of $\mathcal{H}(K)$.

The representer theorem gives the solution of regularization problem (2.2) in a finite dimensional subspace, although it is taken over an infinite dimensional subspace (see Wahba 1990).

Theorem 1. *Consider minimization problem* (2.2), *then there exist constants* a_{ij}, $i = 1, ..., n$, $j = 1, ..., m$, *such that*

$$\hat{\beta}_\lambda(t) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} K(t, T_{ij}).$$

The proof of this Theorem is similar to that of Theorem 1.3.1 in Wahba (1990) and so we omit it.

In order to see how calculate this estimate, let $T := [0,1]$ and $\mathcal{H} := \mathcal{W}_2^r$ where $\mathcal{W}_2^r$ is the rth order Sobolev–Hilbert space:

$$\mathcal{W}_2^r = \Big\{ g : [0,1] \to \mathbb{R} \mid \ g, g^{(1)}, ..., g^{(r-1)} \text{are absolutely continuous} \\ \text{and } \ g^{(r)} \in L^2([0,1]) \Big\}.$$

Sobolev spaces have many applications in nonparametric function estimation. The smoothness of a function that belongs to some Sobolev spaces is guaranteed by existing its derivatives in some orders. To further study about Sobolev spaces see, for example, Adams (1975). There are various norms that we can equip to $\mathcal{W}_2^r$ so that $\mathcal{W}_2^r$ be an RKHS (see Berlinet and Thomas-Agnan (2004)). If we endow $\mathcal{W}_2^r$ with squared norm $\|g\|^2_{\mathcal{W}_2^r} = \sum_{k=0}^{r-1} \left(\int g^{(k)}\right)^2 + \int [g^{(r)}]^2$, then it is an RKHS with reproducing kernel

$$K_r(s,t) = \frac{1}{(r!)^2} B_r(s) B_r(t) + \frac{(-1)^{r-1}}{(2r)!} B_{2r}(|s-t|),$$

where $B_r(.)$ is the rth Bernoulli polynomial. By Theorem 1, it suffices to consider β of the following form:

$$\beta(t) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} K_r(t, T_{ij})$$

for some $\mathbf{a} = [a_{11}, ..., a_{1m}, a_{21}, ..., a_{nm}]' \in \mathbb{R}^{nm}$. Using equation (2.1) yields

$$\begin{aligned} \|\beta\|^2_{\mathcal{W}_2^r} &= \sum_{i_1=1}^{n} \sum_{j_1=1}^{m} \sum_{i_2=1}^{n} \sum_{j_2=1}^{m} a_{i_2 j_2} a_{i_1 j_1} K_r(T_{i_1 j_1}, T_{i_2 j_2}) \\ &= \mathbf{a}' \mathbf{P} \mathbf{a} \end{aligned}$$

where

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \mathbf{P}_{13} & \cdots & \mathbf{P}_{1n} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \mathbf{P}_{23} & \cdots & \mathbf{P}_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{P}_{n1} & \mathbf{P}_{n2} & \mathbf{P}_{n3} & \cdots & \mathbf{P}_{nn} \end{pmatrix}$$

and

$$\mathbf{P}_{i_1 i_2} = [K_r(T_{i_1 j_1}, T_{i_2 j_2})]_{1 \le j_1, j_2 \le m}, \quad 1 \le i_1, i_2 \le n$$

Define $\mathbf{U} = [U_{11}, ..., U_{1m}, U_{21}, ..., U_{nm}]'$ and $\mathbf{V} = [V_{11}, ..., V_{1m}, V_{21}, ..., V_{nm}]'$ then

$$\ell_{mn}(\beta) + \lambda \|\beta\|^2_{\mathcal{W}_2^r} = \frac{1}{nm} \|\mathbf{V} - \mathbf{U} \circ (\mathbf{Pa})\|^2_{\ell_2} + \lambda \mathbf{a}' \mathbf{P} \mathbf{a}, \tag{2.3}$$

where $\mathbf{A} \circ \mathbf{B}$ is the Hadamard product of two matrices $\mathbf{A}$ and $\mathbf{B}$. So finding minimizer of left hand side of (2.3) over $\mathcal{W}_2^r$ is equivalent to finding a vector

$\mathbf{a} \in \mathbb{R}^{nm}$ which minimizes right hand side of (2.3). Let $\mathbf{Q}$ be an $nm \times nm$ matrix such that

$$(\text{the } i\text{th column of } \mathbf{Q}) = (\text{the } i\text{th column of } \mathbf{P}) \circ (\mathbf{U} \circ \mathbf{U})), \qquad i = 1, 2, ..., nm$$

It can be seen that the minimizer of (2.3) is

$$\mathbf{a} = (\mathbf{Q} + nm\lambda \mathbf{I})^{-1} (\mathbf{U} \circ \mathbf{V})$$

3. SIMULATION STUDY

In our simulation study, we carried out a set of simulation studies to emphasize the practical implementation of our methodology. Let true slope function be

$$\beta(t) = \sum_{k=1}^{50} \zeta_k \phi_k(t), \quad t \in [0, 1],$$

where $\zeta_1 = 0.3$, $\phi_1(t) = 1$, $\zeta_k = 4(-1)^{k+1}k^{-2}$ and $\phi_k(t) = \cos((k-1)\pi t)$ for $k \geq 1$. It is clear that this function belongs to the second order Sobolev space ($r = 2$). Random functions X_i's are generated independently as follows:

$$X(t) = \sqrt{2}\sin(\pi t)\xi_1 + \sqrt{2}\cos(\pi t)\xi_2, \quad t \in [0, 1],$$

where ξ_1 and ξ_2 are independent random variables with $\xi_i \sim N(0, i)$, $i = 1, 2$. The response trajectories are generated according to model (1.2) with

$$Z(t) = \sqrt{2}\sin(\pi t)Z_1 + \sqrt{2}\cos(\pi t)Z_2, \quad t \in [0, 1],$$

where Z_1 and Z_2 are i.i.d. random variables from $N(0, 0.1)$. Design points are selected based on common or random design. Noisy observations of each curve obtain according to model (1.3) in each curve.

The fifty curves from $X(t)$ and $Y(t)$ were given in the top panels of Figure 1, the left panel for $X(t)$ and the right panel for $Y(t)$. The lower panels of Figure 1 shows the observed data for $m = 5$ random design points based on stochastic design, the left panel for U and the right panel for V.

We use integrated squared error, $\|\hat{\beta}_\lambda - \beta\|_{L_2}^2 = \int_0^1 \left(\hat{\beta}_\lambda(t) - \beta(t)\right)^2 dt$, to assess goodness of fit of the model. The integrated squared error, $\|\hat{\beta}_\lambda - \beta\|_{L_2}^2$, as a function of smoothing parameter λ is shown in Figure 2 for both designs, the right panel for stochastic design, and the left panel for common design. The best choice for smoothing parameter is the value of λ that minimizes $\|\hat{\beta}_\lambda - \beta\|_{L_2}^2$.

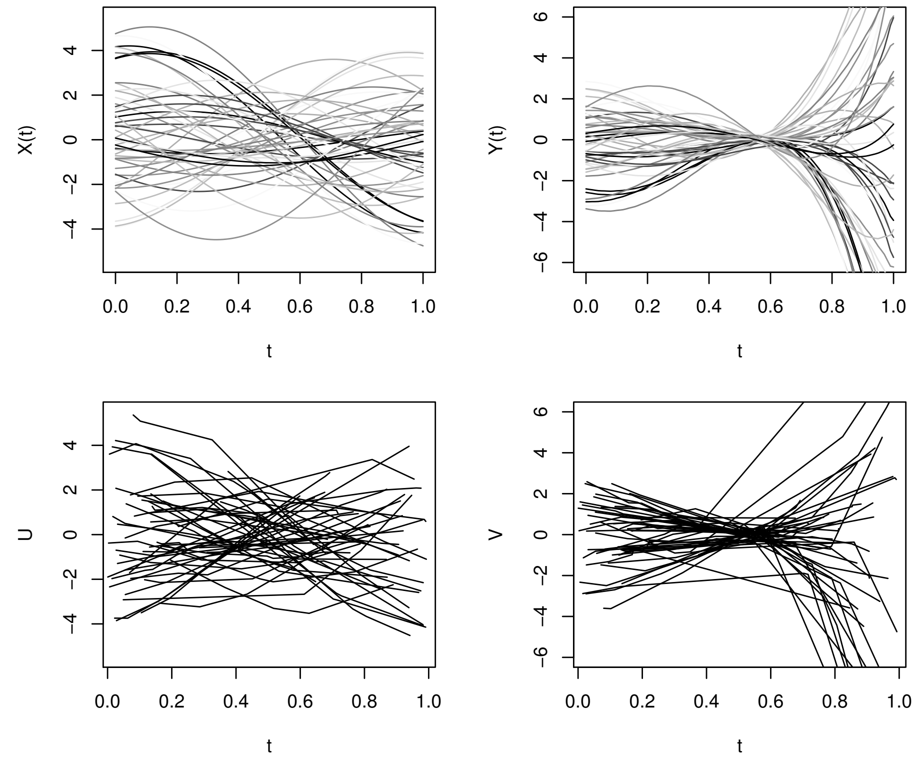

Figure 1: The top panels give 50 simulated curves, the left panel for X and the right panel for Y. Noisy observations at 5 random location based on stochastic design are shown in the lower panels, the left panel for U and the right panel for V.

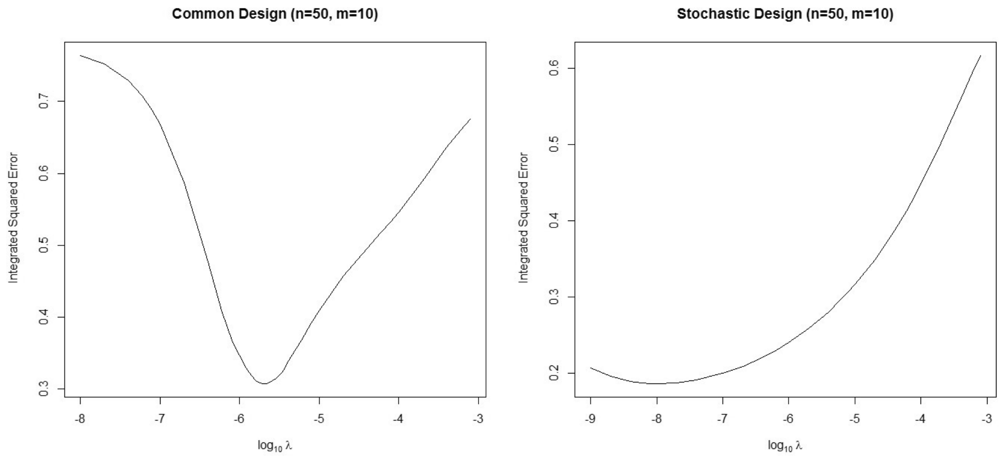

Figure 2: Sensitivity of integrated squared error, $\|\hat{\beta}_\lambda - \beta\|^2_{L_2}$, with respect to smoothing parameter λ for both designs. The right panel for stochastic design, and the left panel for common design.

We calculated $\|\hat{\beta}_\lambda - \beta\|^2_{L_2}$ for different combinations of $n \in \{25,\ 50,\ 100,\ 200\}$ and $m \in \{3,\ 5,\ 10,\ 20\}$. Table 1 presents the obtained value of the smoothing parameter for each simulated data set. As we see in Table 1, the smoothing parameter for common design is much greater than the smoothing parameter for stochastic design. This is because, in the stochastic design we observe random functions X and Y in many different points over whole of n samples, while in the common design, we observe random functions X and Y in only m equidistant points. In addition, the results of a Monte Carlo approximations of $\|\hat{\beta}_\lambda - \beta\|^2_{L_2}$ for common and stochastic design are reported in Tables 2 and 3 respectively. It can be seen from both Tables that the averaged integrated squared error and variance of estimated slope function decrease as either m or n increases. On the other hand, the values of $\|\hat{\beta}_\lambda - \beta\|^2_{L_2}$ for stochastic design is smaller than that is for common design. These results imply that stochastic design has better performance than common design.

Table 1: The value of smoothing parameter for common and stochastic design.

Type of design	n	m = 3	5	10	15	20
common design	25	8×10^{-5}	2.5×10^{-5}	3.5×10^{-6}	5×10^{-7}	10^{-7}
stochastic design	25	5×10^{-7}	10^{-7}	5×10^{-8}	2×10^{-8}	10^{-8}
common design	50	7.5×10^{-5}	2×10^{-5}	3×10^{-6}	5×10^{-7}	9×10^{-8}
stochastic design	50	10^{-7}	5×10^{-8}	1.5×10^{-8}	4×10^{-9}	2×10^{-9}
common design	100	7×10^{-5}	1.5×10^{-5}	2.5×10^{-6}	4.5×10^{-7}	8.5×10^{-8}
stochastic design	100	2.5×10^{-8}	7.5×10^{-9}	2.5×10^{-9}	10^{-9}	5.5×10^{-10}
common design	200	6.5×10^{-5}	10^{-5}	2×10^{-6}	4×10^{-7}	8×10^{-8}
stochastic design	200	7.5×10^{-9}	2.5×10^{-9}	6.5×10^{-10}	4.5×10^{-10}	3.5×10^{-10}

Table 2: Averaged integrated squared error $\|\hat{\beta}_\lambda - \beta\|^2_{L_2}$ and variance of $\hat{\beta}_\lambda$ (in the parentheses) for common design.

n	m = 3	5	10	15	20
25	0.7683 (0.0091)	0.4967 (0.0085)	0.3167 (0.0074)	0.2498 (0.0064)	0.2136 (0.0051)
50	0.7628 (0.0050)	0.4882 (0.0041)	0.3112 (0.0033)	0.2458 (0.0031)	0.2108 (0.0027)
100	0.7596 (0.0024)	0.4854 (0.0019)	0.3085 (0.0015)	0.2436 (0.0013)	0.2098 (0.0011)
200	0.7538 (0.0012)	0.4827 (0.0009)	0.3029 (0.0007)	0.2400 (0.0006)	0.2054 (0.0004)

Table 3: Averaged integrated squared error $\|\hat{\beta}_\lambda - \beta\|^2_{L_2}$ and variance of $\hat{\beta}_\lambda$ (in the parentheses) for stochastic design.

n	m = 3	5	10	15	20
25	0.3435 (0.1335)	0.2866 (0.1111)	0.2324 (0.0673)	0.1944 (0.0406)	0.1902 (0.0354)
50	0.2660 (0.0902)	0.2316 (0.0667)	0.1858 (0.0413)	0.1794 (0.0313)	0.1676 (0.0305)
100	0.2111 (0.0569)	0.1812 (0.0392)	0.1609 (0.0258)	0.1528 (0.0214)	0.1487 (0.0172)
200	0.1776 (0.0366)	0.1577 (0.0245)	0.1437 (0.0165)	0.1365 (0.0111)	0.1312 (0.0084)

4. APPLICATION

The human immune deficiency virus (HIV) attacks immune cells called CD4+ and leads to AIDS. CD4+ cells are a specific kind of white blood cell and are a necessary part of the immune system. They lead the attack against infections. The CD4+ cell count measures the number of CD4+ cells in a sample of blood. CD4+ cell counts are reported as the number of cells in a cubic millimetre of blood. A normal CD4+ cell count is around 1100 cells per cubic millimetre of blood. The CD4+ cell counts can vary time to time. When someone is infected with HIV the number of CD4+ cells they have goes down. So an infected person's CD4+ cell number can be used to monitor disease progression.

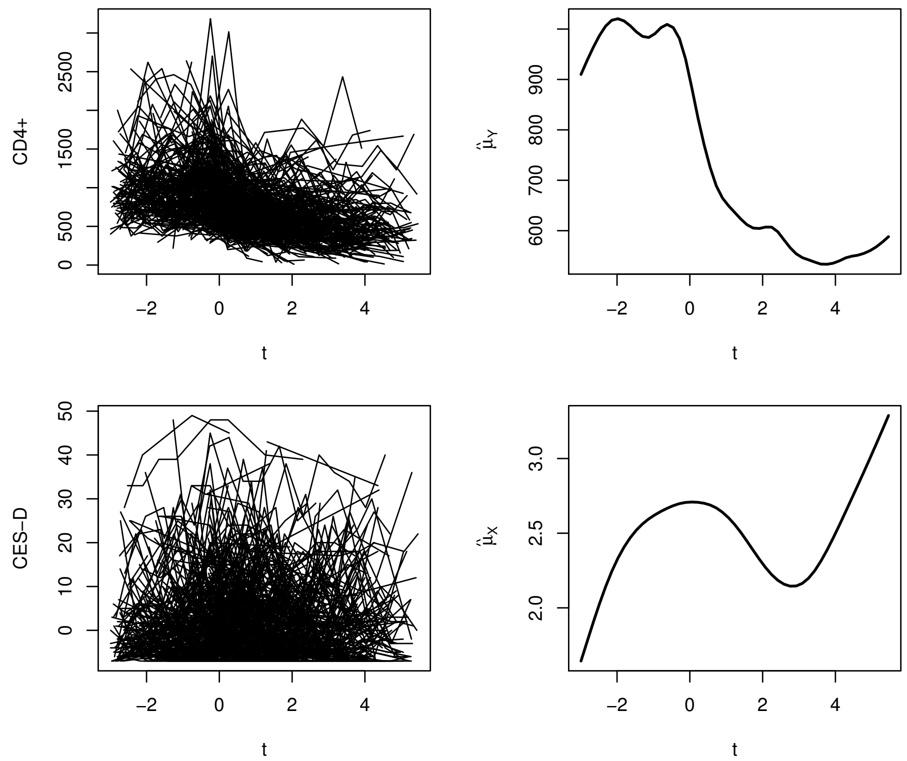

Figure 3: The top panel provides observed individual trajectories and the smooth estimate of the mean function for CD4+ cell counts. The bottom panel includes observed individual trajectories and the smooth estimate of the mean function for CES-D scores.

The CES-D scale is a short self-report scale designed to measure depressive symptomatology during the past week. A higher score indicates greater depressive symptoms. It is interesting to explore whether there is an association between depressive symptoms and CD4+ cell counts over time. The data, reported by Kaslow *et al.* (1987), recorded CD4+ cell counts, CES-D scores and other variables over time for a total of 369 infected men enrolled in the Multicenter AIDS Cohort Study. The measurements were scheduled at each half-yearly visit. But because of missing appointments among other factors, the actual measurement times are random, irregular and sparse. For both CD4+ and CES-D the number of observations ranged from 1 to 12, with a median of 6 measurements per subject, yielding a total of 2376 records.

In this dataset, both the CD4+ cell counts and CES-D scores are considered as functions of time since seroconversion (time when HIV becomes detectable). We model the response process CD4+ cell counts and the predictor process CES-D scores via functional varying coefficient model (1.1). Individual trajectories of CD4+ cell counts and CES-D scores are shown in Figure 3, along with the smooth estimated mean functions of CD4+ cell counts and CES-D scores. The estimated mean function of CD4+ cell counts shows a drastic decreasing from seroconversion to around 2 years after seroconversion. Also the estimated mean function of CES-D scores is decreasing in this period.

We used 5-fold cross-validation to choose the smoothing parameter λ. The procedure is as follows. Divide the data into 5 roughly equal parts at random. For each part, fit the model with parameter λ to the remaining 4 parts. Let $\hat{\beta}_\lambda^{(-k)}$ be the estimated slope function by dropping the kth part, $k = 1, 2, ..., 5$. The cross-validation error is given by

$$CV(\lambda) = \frac{1}{5}\sum_{k=1}^{5}\sum_{i\,\in\,k\text{th part}} \frac{1}{m_i}\sum_{j=1}^{m_i}\Big(V_{ij} - \hat{\mu}_Y(T_{ij}) - [U_{ij} - \hat{\mu}_X(T_{ij})]\hat{\beta}_\lambda^{(-k)}(T_{ij})\Big)^2,$$

where m_i is number of measurements for ith subject and, $\hat{\mu}_X(t)$ and $\hat{\mu}_Y(t)$ are the smooth estimates of mean function for CD4+ cell counts and CES-D scores respectively. To estimate the mean function under sparse and irregular designs, we refer the readers to Yao *et al.* (2005a), Li and Hsing (2010), and Cai and Yuan (2011). Now calculate $CV(\lambda)$ for different values of λ and choose the optimal value of λ as the minimal of $CV(\lambda)$. Here we obtained $\lambda = 100$.

The estimated slope function $\hat{\beta}$ and intercept function $\hat{\alpha}$ are displayed in Figure 4, where we used $\hat{\alpha}(t) = \hat{\mu}_Y(t) - \hat{\mu}_X(t)\hat{\beta}(t)$ to estimate the intercept function. Since the value of CES-D and the estimated slope function are small with respect to the value CD4+ cell counts, the shapes of $\hat{\mu}_Y(t)$ and $\hat{\alpha}(t)$ are obtained almost similar. In Figure 5, we provided difference $\hat{\mu}_Y(t) - \hat{\alpha}(t)$. By comparing Figures 4 and 5 we see that there is a minor association between CD4+ cell counts and CES-D scores in earlier and later times. In addition the association in other times is negligible.

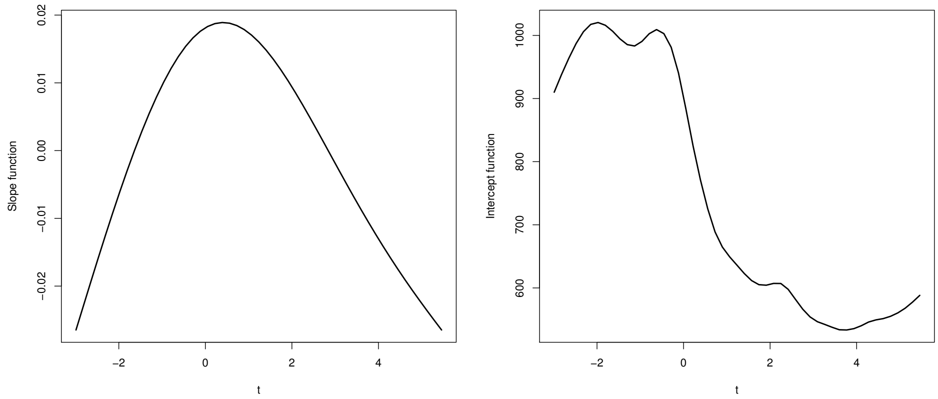

Figure 4: The left panel shows the estimated slope function and the right panel displays the estimated intercept function.

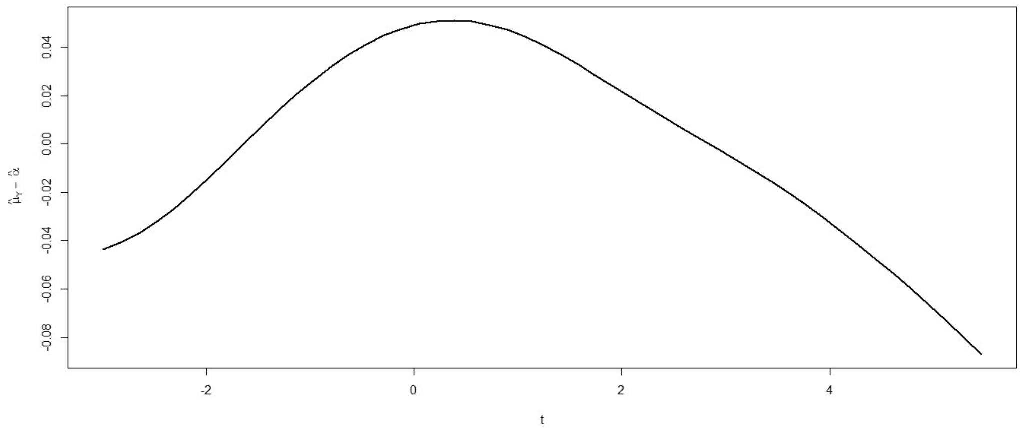

Figure 5: The difference $\hat{\mu}_Y(t) - \hat{\alpha}(t)$.

5. CONCLUSIONS AND EXTENSIONS

We have presented a regularization method to estimate the slope function in functional varying coefficient model using an RKHS approach. Our procedure is easy to implement in the numerical scheme and do not need resorting some numerical techniques to compute the slope function. As we saw in the simulation study, increasing either m or n leads to improved estimates, in the sense of integrated squared error and variance. In this paper, we have assumed that all sampling points on each curve are same. We note that this assumption is not necessary and we may have different sampling points on each curves. Let m_i be the sampling frequency on ith curve. It is suffice to define

$$\hat{\beta}_\lambda = \arg\min_{\beta \in \mathcal{H}} \left\{ \ell_{mn}(\beta) + \lambda \|\beta\|^2_{\mathcal{H}} \right\}$$

where

$$\ell_{mn}(\beta) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} \left(V_{ij} - U_{ij}\beta(T_{ij})\right)^2$$

and then use the given procedure with some mild modifications.

Obtaining rates of convergence and studying optimality of the estimators, in some sense, are interesting problems in nonparametric function estimation. Şentürk and Müller (2010) have given rate of convergence for functional varying coefficient model with sparse and noise-contaminated data in the supremum of absolute error sense but they have not studied optimality of their estimators. Another interesting problem is estimating derivatives of $\beta(t)$ in this model. These ideas will be explored in future works.

ACKNOWLEDGMENTS

The authors are grateful to the referee for valuable comments and suggestions that improved the paper.

REFERENCES

[1] Adams, R.A. (1975). *Sobolev Spaces*, Academic Press, New York.

[2] Aronszajn, N. (1950). Theory of reproducing kernel, *Transactions of the American Mathematical Society*, **68**, 337–404.

[3] Berlinet, A. and Thomas-Agnan, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer-Academic Publishers, Dordrecht.

[4] Cai, T. and Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition, *The Annals of Statistics*, **39**, 2330–2355.

[5] Chiang, C.; Rice J.A. and Wu, C.O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables, *Journal of the American Statistical Association*, **96**, 605–617.

[6] Chiou, J.M.; Ma, Y. and Tsai, C.L. (2012). Functional random effect time-varying coefficient model for longitudinal data, *Stat.*, **1**, 75–89.

[7] Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models, *The Annals of Statistics*, **27**, 1491–1518.

[8] Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models, *Statistics and Its Interface*, **1**, 179–195.

[9] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*, Springer, New York.

[10] Hoover, D.R.; Rice, J.A.; Wu, C.O. and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data, *Biometrika*, **85**, 809–822.

[11] Huang, J.Z.; Wu, C.O. and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements, *Biometrika*, **89**, 111–128.

[12] Huang, J.Z.; Wu, C.O. and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data, *Statistica Sinica*, **14**, 763–788.

[13] Kaslow, R.; Ostrow, D.; Detels, R.; Phair, J.; Polk, B. and Rinaldo, C. (1987). The multicenter AIDS cohort study: Rationale, organization and selected characteristics of the participants, *American Journal of Epidemiology*, **129**, 310–318.

[14] Kauermann, G. and Tutz, G. (1999). On model diagnostics using varying coefficient models, *Biometrika*, **86**, 119–128.

[15] Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data, *The Annals of Statistics*, **38**, 3321–3351.

[16] Lin, D.Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis for longitudinal data, *Journal American Statistical Association*, **96**, 103–126.

[17] Noh, H.S. and Park, B.U. (2010). Sparse varying coefficient models for longitudinal data, *Statistica Sinica*, **20**, 1183–1202.

[18] Ramsay, J.O. and Silverman, B.W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*, Springer, New York.

[19] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis, 2nd Edition*, Springer, New York.

[20] Sentürk, D. and Müller, H.G. (2008). Generalized varying coefficient models for longitudinal data, *Biometrika*, **95**, 653–666.

[21] Sentürk, D. and Müller, H.G. (2010). Functional varying coefficient models for longitudinal data, *Journal of the American Statistical Association*, **105**, 1256–1264.

[22] Sentürk, D. and Nguyen, D.V. (2011). Varying coefficient models for sparse noise-contaminated longitudinal data, *Statistica Sinica*, **21**, 1831–1856.

[23] Sentürk, D.; Dalrymple, L.S.; Mohammed, S.M.; Kaysen, G.A. and Nguyen, D.V. (2013). Modeling time-varying effects with generalized and unsynchronized longitudinal data, *Statistics in Medicine*, **32**, 2971–2987.

[24] Wahba, G. (1990). *Spline models for observational data*, SIAM, Philadelphia.

[25] WANG, H. and XIA, C. (2009). Shrinkage estimation of the varying coefficient model, *Journal of the American Statistical Association*, **104**, 747–757.

[26] WU, C.O.; CHIANG, C.T. and HOOVER, D.R. (1998). Asymptotic confidence regions for kernel smoothing of a varying coefficient models with longitudinal data, *Journal of the American Statistical Association*, **93**, 1388–1402.

[27] WU, C.O. and CHIANG, C.T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable, *Statistica Sinica*, **10**, 433–456.

[28] YANG, W.; MÜLLER, H.G. and STADTMÜLLER, U. (2011). Functional singular component analysis, *Journal of the Royal Statistical Society Series B Statistical Methodology*, **73**, 303–324.

[29] YAO, F.; MÜLLER, H.G. and WANG, J. (2005a). Functional data analysis for sparse longitudinal data, *Journal of the American Statistical Association*, **100**, 577–590.

[30] YAO, F.; MÜLLER, H.G. and WANG, J.L. (2005b). Functional linear regression analysis for longitudinal data, *The Annals of Statistics*, **33**, 2873–2903.

REVSTAT – Statistical Journal
Volume 14, Number 3, June 2016, 327–348

GAMMA KERNEL ESTIMATION OF THE DENSITY DERIVATIVE ON THE POSITIVE SEMI-AXIS BY DEPENDENT DATA

Author: L.A. Markovich
– Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia, and
Institute for Information Transmission Problems, Moscow, Bolshoy Karetny per. 19, build. 1, Moscow 127051, Russia
kimo1@mail.ru

Received: December 2014 Revised: January 2015 Accepted: February 2015

Abstract:

- We estimate the derivative of a probability density function defined on $[0, \infty)$. For this purpose, we choose the class of kernel estimators with asymmetric gamma kernel functions. The use of gamma kernels is fruitful due to the fact that they are non-negative, change their shape depending on the position on the semi-axis and possess good boundary properties for a wide class of densities. We find an optimal bandwidth of the kernel as a minimum of the mean integrated squared error by dependent data with strong mixing. This bandwidth differs from that proposed for the gamma kernel density estimation. To this end, we derive the covariance of derivatives of the density and deduce its upper bound. Finally, the obtained results are applied to the case of a first-order autoregressive process with strong mixing. The accuracy of the estimates is checked by a simulation study. The comparison of the proposed estimates based on independent and dependent data is provided.

Key-Words:

- *density derivative; dependent data; gamma kernel; nonparametric estimation.*

AMS Subject Classification:

- 60G35, 60A05.

1. INTRODUCTION

Kernel density estimation is a non-parametric method to estimate a probability density function (pdf) $f(x)$. It was originally studied in [20], [22] for symmetric kernels and univariate independent identically distributed (i.i.d.) data. When the support of the underlying pdf is unbounded, this approach performs well. If the pdf has a support on $[0,\infty)$, the use of classical estimation methods with symmetric kernels yield a large bias on the zero boundary and leads to a bad quality of the estimates [30]. This is due to the fact that symmetric kernel estimators assign nonzero weight at the interval $(-\infty, 0]$. There are several methods to reduce the boundary bias effect, for example, the data reflection [25], boundary kernels [19], the hybrid method [14], the local linear estimator [18], [17] among others. Another approach is to use asymmetric kernels. In case of univariate nonnegative i.i.d. random variables (r.v.s), the pdf estimators with gamma kernels were proposed in [8]. In [5] the gamma-kernel estimator was developed for univariate dependent data. The gamma kernel is nonnegative and it changes its shape depending on the position on the semi-axis. Estimators constructed with gamma kernels have no boundary bias if $f''(0) = 0$ holds, i.e. when the underlying density $f(x)$ has a shoulder at $x = 0$ (see formula (4.3) in [31]). This shoulder property is fulfilled particularly for a wide exponential class of pdfs which satisfy important integral condition

$$\int_0^\infty x^{-1/2} f(x)dx < \infty \tag{1.1}$$

assumed in [8]. In [31] the half normal and standard exponential pdfs are considered as examples such that the boundary kernel $K_c(t)$ (p. 553 in [31]) gives the better estimate than the gamma-kernel estimator considered in [8]. At the same time, the exponential distribution does not satisfy both the shoulder condition and the condition (1.1). The half normal density satisfies the shoulder condition, but it does not satisfy (1.1). Since (1.1) is not valid for the latter pdfs, such comparison is not appropriate.

Alternative asymmetrical kernel estimators like inverse Gaussian and reciprocal inverse Gaussian estimators were studied in [24]. The comparison of these asymmetric kernels with the gamma kernel is given in [6].

Along with the density estimation it is often necessary to estimate the derivative of a pdf. Derivative estimation is important in the exploration of structures in curves, comparison of regression curves, analysis of human growth data, mean shift clustering or hypothesis testing. The estimation of the density derivative is required to estimate the logarithmic derivative of the density function. The latter has a practical importance in finance, actuary mathematics, climatology and signal processing. However, the problem of the density derivative estimation has received less attention. It is due to a significant increasing complexity of calculations, especially for the multivariate case. The boundary bias problem for the multivariate pdf becomes more solid [4]. The pioneering papers devoted to univariate symmetrical kernel density derivative estimation are [7], [26].

The paper does not focus on the boundary performance but on finding of the optimal bandwidth that is appropriate for the pdf derivative estimation in case of dependent data satisfying a strong mixing condition. In [30] an optimal mean integrated squared error (MISE) of the kernel estimate of the first derivative of order $n^{-\frac{4}{7}}$ was indicated. This corresponds to the optimal bandwidth of order $n^{-\frac{1}{7}}$ for symmetrical kernels. The estimation of the univariate density derivative using a gamma kernel estimator by independent data was proposed in [11], [12]. This allows us to achieve the optimal MISE of the same order $n^{-4/7}$ with a bandwidth of order $n^{-\frac{2}{7}}$.

1.1. Contributions of this paper

It is shown that in the case of dependent data, assuming strong mixing, we can estimate the derivative of the pdf using the same technique that has been applied for independent data in [11]. Lemma 2.1, Section 2.1 contains the upper bound of the covariance. The mathematical technic applied for the derivative estimation is similar to one applied for the pdf. However, formulas became much more complicated, particularly because one has to deal with the special Digamma function that includes the bandwidth b. Thus, one has to pick out the order by b from complicated expressions containing logarithms and the special function. In Section 2.2 we find the optimal bandwidth $b \sim n^{-2/7}$ which is different from the optimal bandwidth $b_2^* \sim n^{-2/5}$ proposed for the pdf estimation (see [8], p. 476).

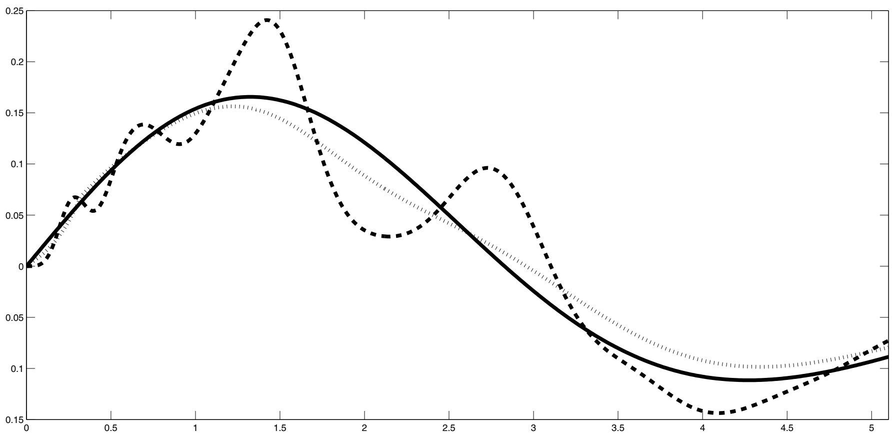

Figure 1: Nonparametric gamma-kernel estimation of Maxwell density derivative function for sample size n=2000. The pdf derivative (solid line), the estimate with b (dotted gray line), the estimate with b_2^* (dashed line).

In Fig. 1 it is shown that the use of b_2^* to estimate the pdf derivative leads to a bad quality (for simplicity the i.i.d. data were taken). We prove that the optimal $MISE$ of the pdf derivative has the same rate of convergence to the true pdf derivative as for the independent case, namely $O(n^{-4/7})$. We show in Section 2.3 that for the strong mixing autoregressive process of the first order (AR(1)) all results are valid without additional conditions. In Section 3 a simulation study for i.i.d. and dependent samples is performed. The flexibility of the gamma kernel allows us to fit accurately the multi-modal pdf derivatives.

1.2. Practical motivation

In practice it is often necessary to deal with sequences of observations that are derived from stationary processes satisfying the strong mixing condition. As an example of such processes one can take autoregressive processes like in Section 2.3. Along with the evaluation of the density function and its derivative by dependent samples, the estimation of the logarithmic derivative of the density is an actual problem. The logarithmic pdf derivative is the ratio of the derivative of the pdf to the pdf itself. The pdf derivative estimation is necessary for an optimal filtering in the signal processing and control of nonlinear processes where only the exponential pdf class is used, [10]. Moreover, the pdf derivative gives information about the slope of the pdf curve, its local extremes, significant features in data and it is useful in regression analysis [9]. The pdf derivative also plays a key role in clustering via mode seeking [23].

1.3. Theoretical background

Let $\{X_i; i = 1, 2, ...\}$ be a strongly stationary sequence with an unknown probability density function $f(x)$, which is defined on $x \in [0, \infty)$. We assume that the sequence $\{X_i\}$ is α-mixing with coefficient

$$\alpha(i) = \sup_k \sup_{\substack{A \in \mathcal{F}_1^k(X) \\ B \in \mathcal{F}_{k+i}^{\infty}(X)}} |P(A \cap B) - P(A)P(B)|.$$

Here, $\mathcal{F}_i^k(X)$ is the σ-field of events generated by $\{X_j, i \le j \le k\}$ and $\alpha(i) \to 0$ as $i \to \infty$. For these sequences we will use a notation $\{X_j\}_{j \ge 1} \in \mathcal{S}(\alpha)$. Let $f_i(x, y)$ be a joint density of X_1 and X_{1+i}, $i = 1, 2, ...$

Our objective is to estimate the derivative $f'(x)$ by a known sequence of observations $\{X_i\}$. We use the non-symmetric gamma kernel estimator that was defined in [8] by the formula

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{\rho_b(x),b}(X_i). \tag{1.2}$$

Here

$$K_{\rho_b(x),b}(t) = \frac{t^{\rho_b(x)-1}\exp(-t/b)}{b^{\rho_b(x)}\Gamma(\rho_b(x))} \tag{1.3}$$

is the kernel function, b is a smoothing parameter (bandwidth) such that $b \to 0$ as $n \to \infty$, $\Gamma(\cdot)$ is a standard gamma function and

$$\rho_b(x) \;=\; \begin{cases} \rho_1(x) = x/b, & \text{if} \quad x \geq 2b, \\ \rho_2(x) = (x/(2b))^2 + 1, & \text{if} \quad x \in [0, 2b). \end{cases} \tag{1.4}$$

The use of gamma kernels is due to the fact that they are nonnegative, change their shape depending on the position on the semi-axis and possess better boundary bias than symmetrical kernels. The boundary bias becomes larger for multivariate densities. Hence, to overcome this problem the gamma kernels were applied in [4]. Earlier the gamma kernels were only used for the density estimation of identically distributed sequences in [4], [8] and for stationary sequences in [5].

To our best knowledge, the gamma kernels have been applied to the density derivative estimation at first time in [11]. In this paper the derivative $f'(x)$ was estimated under the assumption that $\{X_1, X_2, ..., X_n\}$ are i.i.d. random variables as derivative of (1.2). This implies that

$$\hat{f}'_n(x) \;=\; \frac{1}{n}\sum_{i=1}^{n} K'_{\rho_b(x),b}(X_i) \tag{1.5}$$

holds, where

$$K'_{\rho_b(x),b}(t) \;=\; \begin{cases} K'_{\rho_1(x),b}(t) = \frac{1}{b}K_{\rho_1(x),b}(t)L_1(t), & \text{if} \quad x \geq 2b, \\ K'_{\rho_2(x),b}(t) = \frac{x}{2b^2}K_{\rho_2(x),b}(t)L_2(t), & \text{if} \quad x \in [0, 2b), \end{cases} \tag{1.6}$$

is the derivative of $K_{\rho(x),b}(t)$,

$$\begin{aligned} L_1(t) \;&=\; L_1(t,x) = \ln t - \ln b - \Psi(\rho_1(x)), \\ L_2(t) \;&=\; L_2(t,x) = \ln t - \ln b - \Psi(\rho_2(x)), \end{aligned} \tag{1.7}$$

Here $\Psi(x)$ denotes the Digamma function (the logarithmic derivative of the gamma function). The unknown smoothing parameter b was obtained as the minimum of the mean integrated squared error ($MISE$) which, as known, is equal to

$$MISE(\hat{f}'_n(x)) \;=\; \mathsf{E}\int_0^{\infty}(f'(x) - \hat{f}'_n(x))^2 dx.$$

Remark 1.1. The latter integral can be splitted into two integrals $\int_0^{2b}$ and $\int_{2b}^{\infty}$. In the case when $x \geq 2b$ the integral $\int_0^{2b}$ tends to zero when $b \to 0$. Hence, we omit the consideration of this integral in contrast to [31]. The first integral has the same order by b as the second one, thus it cannot affect on the selection of the optimal bandwidth.

The following theorem has been proved.

Theorem 1.1 ([11]). *If $b \to 0$ and $nb^{3/2} \to \infty$ as $n \to \infty$, the integrals*

$$\int_0^{\infty} P(x)dx, \quad \int_0^{\infty} x^{-3/2} f(x)dx$$

are finite and $\int_0^{\infty} P(x)dx \neq 0$, then the leading term of a MISE expansion of the density derivative estimate $\hat{f}'(x)$ is equal to

$$\begin{aligned} MISE(\hat{f}_n'(x)) &= \frac{b^2}{16}\int_0^{\infty} P(x)dx \\ &+ \int_0^{\infty} \frac{n^{-1}b^{-3/2}x^{-3/2}}{4\sqrt{\pi}}\left(f(x) + b\left(\frac{f(x)}{2x} - \frac{f'(x)}{2}\right)\right)dx \\ &+ o(b^2 + n^{-1}(b^{-3/2})), \end{aligned} \tag{1.8}$$

where

$$P(x) = \left(\frac{f(x)}{3x^2} + f''(x)\right)^2.$$

Taking the derivative of (1.8) in b leads to equation

$$\begin{aligned} \frac{b}{8}\int_0^{\infty}\left(\frac{f(x)}{3x^2} + f''(x)\right)^2 dx - \frac{3n^{-1}b^{-\frac{5}{2}}}{8\sqrt{\pi}}\int_0^{\infty} x^{-\frac{3}{2}} f(x)dx \\ + \frac{n^{-1}b^{-\frac{3}{2}}}{16\sqrt{\pi}}\int_0^{\infty} x^{-\frac{3}{2}}\left(\frac{f(x)}{x} - f'(x)\right)dx = 0. \end{aligned} \tag{1.9}$$

Neglecting the term with $b^{-3/2}$ as compared to the term $b^{-5/2}$, the equation becomes simpler and its solution is equal to the optimal global bandwidth

$$b_0 = \left(\frac{3\int_0^{\infty} x^{-3/2} f(x)dx}{\sqrt{\pi}\int_0^{\infty}\left(\frac{f(x)}{3x^2} + f''(x)\right)^2 dx}\right)^{2/7} n^{-2/7}. \tag{1.10}$$

The substitution of b_0 into (1.8) yields an optimal $MISE$ with the rate of convergence $O(n^{-\frac{4}{7}})$. The unknown density and its second derivative in (1.10) were estimated by the rule of thumb method [12].

In [30], p. 49, it was indicated an optimal $MISE$ of the first derivative kernel estimate $n^{-\frac{4}{7}}$ with the bandwidth of order $n^{-\frac{1}{7}}$ for symmetrical kernels. Nevertheless, our procedure achieves the same order $n^{-4/7}$ with a bandwidth of order $n^{-\frac{2}{7}}$. Moreover, our advantage concerns the reduction of the bias of the density derivative at the zero boundary by means of asymmetric kernels. Gamma kernels allow us to avoid boundary transformations which is especially important for multivariate cases.

Further results presented in Section 2.2 will be based on Theorem 1.1.

2. MAIN RESULTS

2.1. Estimation of the density derivative by dependent data

Here, we estimate the density derivative by means of the kernel estimator (1.5) by dependent data. Thus, its mean squared error is determined as

$$MSE(\widehat{f'}_n(x)) = (Bias(\widehat{f'}_n(x)))^2 + var(\widehat{f'}_n(x)), \tag{2.1}$$

where, due to the stationarity of the process X_i, the variance is given by

$$\begin{aligned}
\mathrm{var}(\widehat{f'}_n(x)) &= \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} K'_b(X_i)\right) = \frac{1}{n^2}\mathrm{var}\left(\sum_{i=1}^{n} K'_b(X_i)\right) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathrm{var}(K'_b(X_i)) + 2\sum_{1\le i<j\le n}\mathrm{cov}(K'_b(X_i), K'_b(X_j))\right) \\
&= \frac{1}{n}\mathrm{var}(K'_b(X_i)) + \frac{2}{n^2}\sum_{1\le i<j\le n}\mathrm{cov}(K'_b(X_i), K'_b(X_j)) \\
&= \frac{1}{n}\mathrm{var}(K'_b(X_i)) + \frac{2}{n}\sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)\mathrm{cov}(K'_b(X_1), K'_b(X_{1+i})) \\
&= V(x) + C(x).
\end{aligned}$$

For simplicity we use here and further the notation $K'_{\rho_b(x),b}(t) = K'_b(t)$ in (1.5).

Thus, (2.1) can be written as

$$MSE(\widehat{f'}(x)) = B(x)^2 + V(x) + C(x), \tag{2.2}$$

where

$$B(x) = Bias(\widehat{f'}_n(x)).$$

The bias of the estimate does not change, but the variance contains a covariance. The next lemma is devoted to its finding.

Lemma 2.1. *Let*

1. $\{X_j\}_{j\geq 1} \in \mathcal{S}(\alpha)$ *and* $\int\limits_1^{\infty} \alpha(\tau)^{\upsilon} d\tau < \infty, \quad 0 < \upsilon < 1$ *hold,*

2. $f(x)$ *be a twice continuously differentiable function,*

3. $b \to 0$ *and* $nb^{-(\upsilon+1)/2} \to \infty$ *as* $n \to \infty$.

Then the covariance $C(x)$ *is bounded by*

$$\begin{aligned}
|C(x)| &= \left| \frac{2}{n} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) cov(K'_{\rho_b(x),b}(X_1), K'_{\rho_b(x),b}(X_{1+i})) \right| \\
&\leq \left(2^{-\frac{\upsilon+3}{2}} \pi^{\frac{1-\upsilon}{2}} x^{-\frac{\upsilon+5}{2}} \frac{b^{-\frac{\upsilon+1}{2}}}{n} \left(b^2 C_2(\upsilon, x) + bC_1(\upsilon, x) + C_3(\upsilon, x) \right)^{1-\upsilon} \right. \\
&\quad \left. + o(b^2) \right) \int\limits_1^{\infty} \alpha(\tau)^{\upsilon} d\tau,
\end{aligned} \tag{2.3}$$

where $K'_{\rho_b(x)}$ *is defined by* (1.6) *and* $C_1(\upsilon, x)$, $C_2(\upsilon, x)$ *and* $C_1(\upsilon, x)$ *are given by* (4.8).

A similar lemma was proved in [10] for symmetrical kernels and not strictly positive x.

2.2. Mean integrated squared error of $\hat{f'}_n(x)$

Using the upper bound (2.3) we can obtain the upper bound of the MISE and find the expression of the optimal bandwidth b as the minimum of the latter.

Theorem 2.1. *If the conditions of Theorem 1.1 and Lemma 2.1 hold, then the MISE expansion for the estimate* $\hat{f'}_n(x)$ *of the density derivative is equal to*

$$\begin{aligned}
MISE(f'(x)) \leq & \int\limits_0^{\infty} \frac{n^{-1} b^{-\frac{3}{2}} x^{-\frac{3}{2}}}{4\sqrt{\pi}} \left(f(x) + \frac{b}{2} \left(\frac{f(x)}{x} - f'(x) \right) \right) dx \\
& + \int\limits_0^{\infty} \left(2^{-\frac{\upsilon+3}{2}} \pi^{\frac{1-\upsilon}{2}} x^{-\frac{\upsilon+5}{2}} \frac{b^{-\frac{\upsilon+1}{2}}}{n} C_3(\upsilon, x)^{1-\upsilon} \right) \int\limits_1^{\infty} \alpha(\tau)^{\upsilon} d\tau dx \\
& + \frac{b^2}{16} \int\limits_0^{\infty} P(x) dx + o(b^2 + n^{-1}(b^{-\frac{3}{2}})).
\end{aligned} \tag{2.4}$$

and the optimal bandwidth is $b_{opt} = o(n^{-2/7})$ *and the* $MISE_{opt} = O(n^{-4/7})$.

Remark 2.1. It is evident from the formula (2.4) that the term responsible for the covariance has the order $\frac{b^{-\frac{\upsilon+1}{2}}}{n}$, $0 < \upsilon < 1$. Thus, it does not influence the order of MISE irrespective of the mixing coefficient $\alpha(\tau)$.

The proof is given in Appendix 4.

2.3. Example of a strong mixing process

We use the first-order autoregressive process as an example of a process that satisfies Theorem 1.1. X_i determines a first-order autoregressive (AR(1)) process with the innovation r.v. ϵ_0 and the autoregressive parameter $\rho \in (-1, 1)$ if

$$X_i = \rho X_{i-1} + \epsilon_i, \quad i = ... - 1, 0, 1, ..., \tag{2.5}$$

holds and ϵ_i is a sequence of i.i.d. r.v.s. Let AR(1) process (2.5) be strong mixing with mixing numbers $\alpha(\tau)$, $\tau = 1, 2, ...$

$$\alpha(\tau) \leq \widetilde{\alpha}(\tau) \equiv \begin{cases} 2(C+1)\mathsf{E}|X_i|^\nu |\rho^\nu|^\tau, & \text{if} \quad \tau \geq \tau_0, \\ 1, & \text{if} \quad 1 \leq \tau < \tau_0, \end{cases} \tag{2.6}$$

where $\nu = \min\{p, q, 1\}$ and $p > 0, q > 0, C > 0, \tau_0 > 0$ hold. In [2] it was proved that with some conditions AR(1) is a strongly mixing process.

In Appendix 4 we prove the following lemma.

Lemma 2.2. *Under the conditions* (2.6) *the AR(1) process* (2.5) *satisfies Lemma 2.1 and Theorem 2.1.*

3. SIMULATION RESULTS

To investigate the performance of the gamma-kernel estimator we select the following positive defined pdfs: the Maxwell ($\sigma = 2$), the Weibull ($a = 1, b = 4$) and the Gamma ($\alpha = 2.43, \beta = 1$) pdf,

$$\begin{aligned} f_M(x) &= \frac{\sqrt{2}x^2 \exp(-x^2/2\sigma^2)}{\sigma^3\sqrt{\pi}}, \\ f_W(x) &= sx^{s-1}\exp(-x^s), \\ f_G(x) &= \frac{x^{\alpha-1}\exp(-x/\beta)}{\beta^\alpha\Gamma(\alpha)}. \end{aligned}$$

Their derivatives

$$
\begin{aligned}
f'_M(x) &= -\frac{\sqrt{2}x\exp(-x^2/2\sigma^2)(x^2-2\sigma^2)}{\sigma^5\sqrt{\pi}}, \\
f'_W(x) &= -sx^{s-2}\exp(-x^s)(sx^s-s+1), \\
f'_G(x) &= \frac{x^{\alpha-2}\exp(-x/\beta)(\beta+x-\alpha\beta)}{\beta^{\alpha+1}\Gamma(\alpha)}
\end{aligned}
\tag{3.1}
$$

are to be estimated. The Weibull and the Gamma pdfs are frequently used in a wide range of applications in engineering, signal processing, medical research, quality control, actuarial science and climatology among others. For example, most total insurance claim distributions are shaped like gamma pdfs [13]. The gamma distribution is also used to model rainfalls [1]. Gamma class pdfs, like Erlang and χ^2 pdfs are widely used in modeling insurance portfolios [15].

We generate Maxwell, Weibull and Gamma i.i.d. samples with sample sizes $n \in \{100, 500, 1000, 2000\}$ using standard Matlab generators. To get the dependent data we generate Markov chains with the same stationary distributions using the Metropolis–Hastings algorithm [16]. Due to the existence of the probability of rejecting a move from the previous point to the next one, the variance of such Markov sequence $\{X_t\}$ is corrupted by the function of the latter rejecting probability (see [27], Theorem 3.1). The Metropolis–Hastings Markov chains [16] are geometrically ergodic for the underlying light-tailed distributions. Hence, they satisfy the strong mixing condition [21].

The gamma kernel estimates (1.2) with the optimal bandwidth (1.10) for the derivatives (3.1) can be seen in Figures 2–4. The optimal bandwidth (1.10) is counted for every replication of the simulation using the rule of thumb method, where as a reference density we take the gamma pdf.

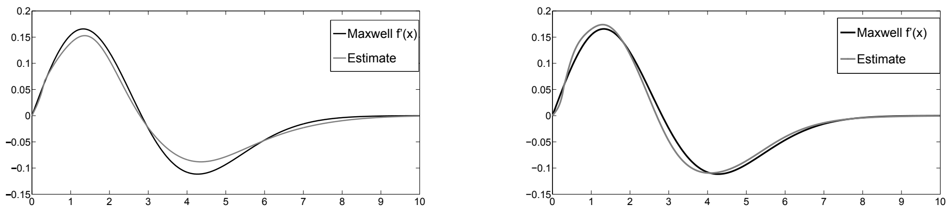

Figure 2: Estimates of the Maxwell pdf derivative by i.i.d. data (left) and by dependent data (right): the $f'_M(x)$ (black line), gamma kernel estimate from the rule of thumb (grey line) for the sample size $n = 2000$.

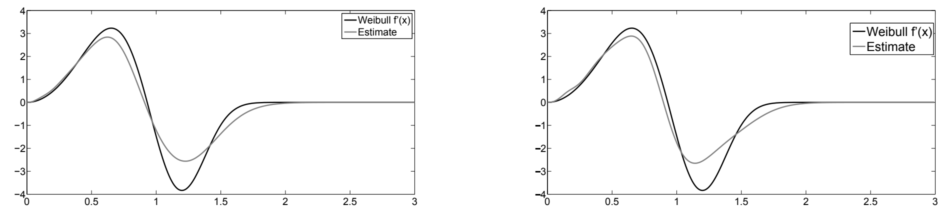

Figure 3: Estimates of the Weibull pdf derivative by i.i.d. data (left) and by dependent data (right): the $f'_W(x)$ (black line), gamma kernel estimate from the rule of thumb (grey line) for the sample size $n = 2000$.

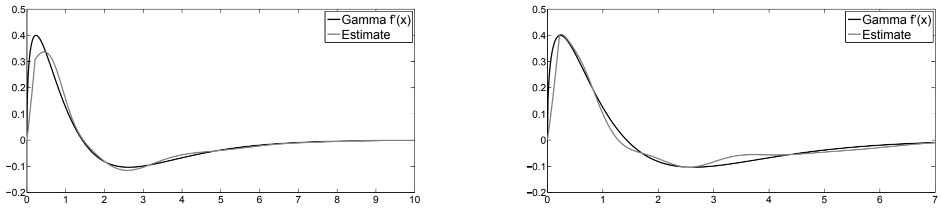

Figure 4: Estimates of the Gamma pdf derivative by i.i.d. data (left) and by dependent data (right): the $f'_G(x)$ (black line), gamma kernel estimate from the rule of thumb (grey line) for the sample size $n = 2000$.

The estimation error of the pdf derivative is calculated by the following formula

$$m = \int_0^\infty (f'(x) - \hat{f}'(x))^2 dx,$$

where $f'(x)$ is a true derivative and $\hat{f}'(x)$ is its estimate. Values of m's averaged over 500 simulated samples and the standard deviations for the underlying distributions are given in Table 1 for i.i.d. r.v.s and in Table 2 for dependent data.

Table 1: Mean errors m and standard deviations for i.i.d. r.v.s.

Distribution	n			
	100	500	1000	2000
Gamma	0.032792 (0.011967)	0.015208 (0.0044094)	0.010675 (0.0027815)	0.0074668 (0.0016452)
Weibull	2.0056 (0.52931)	1.1987 (0.25172)	0.9157 (0.18333)	0.69155 (0.12178)
Maxwell	0.0077597 (0.0033915)	0.0035692 (0.0015351)	0.0028675 (0.00099263)	0.0020923 (0.00068739)

Table 2: Mean errors m and standard deviations for strong mixed r.v.s.

Distribution	n			
	100	500	1000	2000
Gamma	0.039226 (0.015824)	0.018124 (0.006055)	0.01252 (0.0038485)	0.0086675 (0.0023361)
Weibull	2.2052 (1.1585)	1.3009 (0.5957)	0.97509 (0.41041)	0.75382 (0.28755)
Maxwell	0.0077694 (0.006793)	0.0039277 (0.0028336)	0.002878 (0.0020021)	0.0027313 (0.0016573)

As expected, the mean error and the standard deviation decrease when the sample size rises, and this holds both for i.i.d. and the dependent case. The performance of the gamma kernel changes when dependence is introduced, but the results in both tables are close. The mean errors are very close due to the fact the bandwidth parameter is selected to minimize this error. However, the standard deviations for the dependent data are higher than for the i.i.d. r.v.s. For example, for the sample size of 500 the mean errors and the standard deviations for the Maxwell pdf for the i.i.d. r.v.s are 0.0035692 (0.0015351) and for dependent r.v.s 0.0039277 (0.0028336). They differ due to the contribution of the Metropolis–Hastings rejecting probability. This difference is less pronounced for larger sample sizes.

The Metropolis–Hastings algorithm gives opportunity to generate AR processes with known pdfs. As a consequence we know their derivatives and can find mean errors and standard deviations of the gamma-kernel density derivatives estimates for the dependent data. In the case when we consider the noise distribution $\{\epsilon\}$ of the AR model (2.5) and the autoregressive parameter ρ that influences on the dependence rate (2.6), we cannot indicate in general the true pdf of the process. Hence, we consider the histogram based on 200000 observations as a true pdf. As the noise distribution $\{\epsilon\}$ let us take the Gamma distribution $(\alpha = 1.5, \beta = 1)$ and the Maxwell distribution $(\sigma = 1)$. In [5] it was proved that, as in the i.i.d. case, the gamma-kernel estimator of the pdf achieves the same optimal rate of convergence in terms of the mean integrated squared error as for strongly mixed r.v.s. For the various parameters $\rho \in \{0.1, 0.2, 0.3, 0.4\}$ the gamma estimates for the densities of the AR models are given in Figures 5–6.

Since the gamma-kernel estimators perform good for the various dependence rates it is also true for the gamma-kernel pdf derivative estimators, but the bandwidth parameter must be selected differently.

Hence, this findings confirms the fact that the covariance term (2.3) of the pdf derivative is negligible in comparison with its variance and implies that one can use the same optimal bandwidth (1.10), both for independent and strongly mixed dependent data.

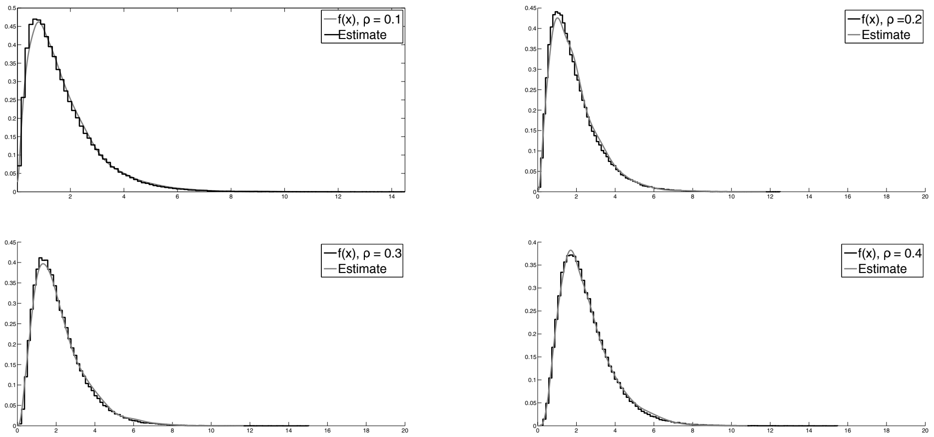

Figure 5: Gamma-kernel estimates of the pdf of the AR model with the Gamma noise and $\rho \in \{0.1, 0.2, 0.3, 0.4\}$ for the sample size $n = 2000$.

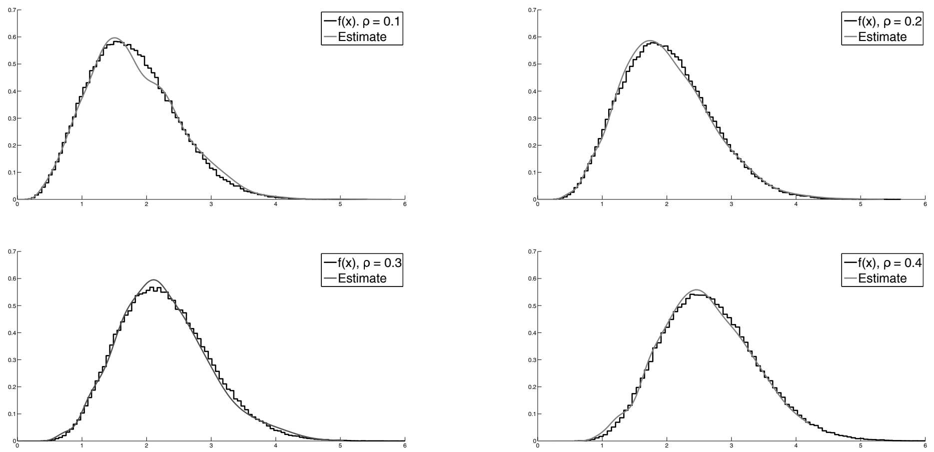

Figure 6: Gamma-kernel estimates of the pdf of the AR model with the Maxwell noise and $\rho \in \{0.1, 0.2, 0.3, 0.4\}$ for the sample size $n = 2000$.

4. APPENDIX

Proof of Lemma 2.1: Taking an integral from (2.2) we get

$$MISE(\widehat{f}'(x)) = \int_0^\infty (B(x)^2 + V(x) + C(x))dx, \tag{4.1}$$

where

$$C(x) = \frac{2}{n}\sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)\operatorname{cov}(K_b'(X_1), K_b'(X_{1+i})). \tag{4.2}$$

To evaluate the covariance we shall apply Davydov's inequality

$$|\operatorname{cov}(K_b'(X_1), K_b'(X_{1+i}))| \le 2\pi\alpha(i)^{1/r} \parallel K_b'(X_1) \parallel_q \parallel K_b'(X_{1+i}) \parallel_p, \tag{4.3}$$

where $p^{-1} + q^{-1} + r^{-1} = 1$, $1 \le p, q, r \le \infty$, [3].

The latter norm for the case $x \ge 2b$ is determined by

$$\begin{aligned} \parallel K_b'(X_1) \parallel_q &= \left(\int \left(\frac{1}{b}K(y)L_1(y)\right)^q f(y)dy\right)^{1/q} \\ &= \frac{1}{b}\left(\mathsf{E}\left(K(\xi_1)^{q-1}L_1(\xi_1)^q f(\xi_1)\right)\right)^{1/q}, \end{aligned} \tag{4.4}$$

where $L_1(t)$ is introduced in (1.7). The kernel $K(\xi_1)$ was used in (4.4) as a density function and ξ_1 is a $Gamma(\rho_1(x), b)$ random variable.

In the case $x \in [0, 2b)$, similarly we have

$$\begin{aligned} \parallel K_b'(X_1) \parallel_q &= \left(\int \left(\frac{x}{2b^2}K(y)L_2(y)\right)^q f(y)dy\right)^{1/q} \\ &= \frac{x}{2b^2}\left(\mathsf{E}\left(K(\xi_2)^{q-1}L_2(\xi_2)^q f(\xi_2)\right)\right)^{1/q}, \end{aligned} \tag{4.5}$$

where $L_2(t)$ is determined by (1.7), and ξ_2 is a $Gamma(\rho_2(x), b)$ random variable. Expressions (4.4) and (4.5) are constructed similarly, thus to a certain point, we will not make differences between them.

By the standard theory of the gamma distribution it is known that $\mu = \mathsf{E}(\xi) = \rho_b(x)b$ and the variance is given by $var(\xi) = \rho_b(x)b^2$. For simplicity, we further use the notation ρ instead of $\rho_b(x)$ defined in (1.4).

The Taylor expansion of both mathematical expectations in (4.4), (4.5) in the neighborhood of μ is represented by

$$\begin{aligned} \mathsf{E}\left(K(\xi)^{q-1}L(\xi)^q f(\xi)\right) &= K(\mu)^{q-1}L(\mu)^q f(\mu) + (K(\xi)^{q-1}L(\xi)^q f(\xi))'|_{\xi=\mu}\mathsf{E}(\xi-\mu) \\ &+ \left(K(\xi)^{q-1}L(\xi)^q f(\xi)\right)''|_{\xi=\mu}\frac{\mathsf{E}(\xi-\mu)^2}{2} + o\left(\mathsf{E}(\xi-\mu)^2\right). \end{aligned}$$

In the case when $x \geq 2b$, $\mu = \rho b = x$, $var(\xi) = \rho b^2 = xb$, we get

$$\mathsf{E}\left(K(\xi)^{q-1}L(\xi)^q f(\xi)\right) =$$

$$= \frac{K(x)^{q-1}}{b}\Bigg(qL(x)^{q+1}f'(x) - L(x)^q f(x)L'(x)$$
$$- L(x)^{q+1}f'(x) + bL(x)^q f''(x) + q^2L(x)^q f(x)L'(x) + bq^2L(x)^{q-2}(L'(x))^2 f(x)$$
$$+ 2bqL(x)^{q-1}L'(x)f'(x) + bqL(x)^{q-1}f(x)L''(x) - bqL(x)^{q-2}(L'(x))^2\Bigg)$$
$$+ \frac{K(x)^{q-1}L(x)(q-1)}{b^2}\Bigg((q-1)f(x)L(x)^{q+1} + bL(x)^q f'(x)$$
$$+ bqL(x)^{q-1}f(x)L'(x)\Bigg) + o\left(b^2\right).$$

Using Stirling's formula

$$\Gamma(z) = \sqrt{\frac{2\pi}{z}}\left(\frac{z}{e}\right)^z\left(1 + O\left(\frac{1}{z}\right)\right),$$

we can rewrite the kernel function as

$$K(t) = \frac{t^{\rho-1}\exp(-t/b)}{b^\rho\Gamma(\rho)} = \frac{t^{\rho-1}\exp(-t/b)\exp(\rho)}{b^\rho\sqrt{2\pi}\rho^{\rho-\frac{1}{2}}(1+O(1/\rho))}.$$

Taking $\rho = \rho_1(x)$ according to (1.4), $t = x$, it holds

$$K(\rho_1(x)b) = \frac{1}{\sqrt{2\pi}}\frac{x^{x/b-1}\exp((x-x)/b)}{b^{\frac{x}{b}}\frac{x}{b}^{\frac{x}{b}-\frac{1}{2}}(1+O(b/x))} = \frac{x^{-\frac{1}{2}}b^{-\frac{1}{2}}}{\sqrt{2\pi}(1+O(b/x))}.$$

Hence, its upper bound is given by

$$K(x) \leq \frac{1}{\sqrt{2\pi xb}}. \tag{4.6}$$

Next, using the property of the Digamma function $\Psi(x) = \ln(x) - \frac{1}{2x} - \frac{1}{12x^2} + \frac{1}{120x^4} + O(1/x^6)$, the first equation in (1.7) can de rewritten as

$$L_1(\rho_1 b) \; = \; \ln(\rho_1 b) - \ln(b) - \Psi(\rho_1) = \frac{b}{2x} + \frac{b^2}{12x^2} + o(b^2). \tag{4.7}$$

Then substituting (4.6) in (4.4) and using the expressions (4.6) and (4.7), we deduce

$$\parallel K_b'(X_1) \parallel_q \leq \pi^{\frac{1-q}{2q}}(2x)^{\frac{1-q}{2q}-1}b^{\frac{1-q}{2q}}\left(b^2C_2(q,x) + bC_1(q,x) + C_3(q,x)\right)^{1/q} + o(b^2),$$

where we used the notations

$$
\begin{aligned}
(4.8) \qquad C_1(q,x) &= -f(x)\frac{2q^3-9q^2+4q-33}{24x} - f'(x)\frac{q+1}{2} + f''(x)\frac{x}{2}, \\
C_2(q,x) &= f(x)\frac{2q+54x-q^2x+21q^3x+q^4x+93qx}{144x^3} \\
&\quad - f'(x)\frac{(q+1)^2}{12x} + f''(x)\frac{q+1}{12}, \\
C_3(q,x) &= -f(x)\frac{(q+1)(q-2)}{2}.
\end{aligned}
$$

The same steps can be done for $\| K_b'(X_{1+i}) \|_p$ from (4.3). Then, if $p=q$ holds, one can represent Davydov's inequality (4.3) as

$$
\begin{aligned}
(4.9) \qquad & |cov(K_b'(X_1), K_b'(X_{1+i}))| \le \\
&\le 2\pi\alpha(i)^{\frac{1}{r}}\pi^{\frac{1-q}{q}}(2x)^{\frac{1-q}{q}-2}b^{\frac{1-q}{q}}\left(b^2C_2(q,x)+bC_1(q,x)+C_3(q,x)\right)^{2/q} + o(b^2).
\end{aligned}
$$

Using (4.9) and taking $p=q=2+\delta$, $r=\frac{2+\delta}{\delta}$ it can be deduced that the covariance (4.2) is given by

$$
\begin{aligned}
|C(x)| &= \left|\frac{2}{n}\sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)cov(K_b'(X_1),K_b'(X_{1+i}))\right| \\
&\le \left|\left(2^{-\frac{2\delta+3}{\delta+2}}\pi^{\frac{1}{\delta+2}}x^{-\frac{3\delta+5}{\delta+2}}\frac{b^{-\frac{\delta+1}{\delta+2}}}{n}\left(b^2C_2(\delta,x)+bC_1(\delta,x)+C_3(\delta,x)\right)^{\frac{2}{2+\delta}}\right)\right. \\
&\quad \left.\cdot\sum_{i=1}^{n-1}\left(1-\frac{i}{n}\right)\alpha(i)^{\frac{\delta}{2+\delta}}\right| + o(b^2).
\end{aligned}
$$

Then we can estimate the covariance by the previous expressions

$$
\begin{aligned}
|C(x)| &\le S(b,x,\delta,n)\sum_{\tau=2}^{n}\left(1-\frac{\tau-1}{n}\right)\alpha(\tau-1)^{\frac{\delta}{2+\delta}} + o(b^2) \\
&\le S(b,x,\delta,n)\sum_{\tau=2}^{\infty}\alpha(\tau-1)^{\frac{\delta}{2+\delta}} + o(b^2) \le S(b,x,\delta,n)\int_1^{\infty}\alpha(\tau)^{\frac{\delta}{2+\delta}}d\tau + o(b^2),
\end{aligned}
$$

where we used the following notation

$$
S(b,x,\delta,n) = 2^{-\frac{2\delta+3}{\delta+2}}\pi^{\frac{1}{\delta+2}}x^{-\frac{3\delta+5}{\delta+2}}\frac{b^{-\frac{\delta+1}{\delta+2}}}{n}\left(b^2C_2(\delta,x)+bC_1(\delta,x)+C_3(\delta,x)\right)^{\frac{2}{2+\delta}}.
$$

Let us denote $\frac{\delta}{2+\delta}=\upsilon$, $0<\upsilon<1$. Then, in this notations, we get the estimate of the covariance

$$
|C(x)| \le \left(2^{-\frac{\upsilon+3}{2}}\pi^{\frac{1-\upsilon}{2}}x^{-\frac{\upsilon+5}{2}}\frac{b^{-\frac{\upsilon+1}{2}}}{n}\left(bC_1(\upsilon,x)+C_3(\upsilon,x)\right)^{1-\upsilon} + o(b^2)\right)\int_1^{\infty}\alpha(\tau)^{\upsilon}d\tau.
$$

By $0 < \upsilon < 1$ then it follows

$$|C(x)| \sim \frac{1}{n} b^{-\frac{\upsilon+1}{2}}. \qquad \square$$

Remark 4.1. The main contribution to MISE (4.1) is provided by the part corresponding to $x \geq 2b$, so we will not do similar calculations here and further for $x \in [0, 2b)$ as $b \to 0$.

Proof of Theorem 2.1: Regarding the dependent case it is known that the MISE contains the bias, the variance and the covariance. By (1.8) it follows that the integrated sum of the squared bias and variance is the following expression

$$\begin{aligned}
\int_0^\infty (B(x)^2 + V(x))dx &= \frac{b^2}{16}\int_0^\infty P(x)dx \\
&+ \int_0^\infty \frac{n^{-1}b^{-\frac{3}{2}}x^{-\frac{3}{2}}}{4\sqrt{\pi}}\left(f(x) + \frac{b}{2}\left(\frac{f(x)}{x} - f'(x)\right)\right)dx \qquad (4.10)\\
&+ o(b^2 + n^{-1}b^{-\frac{3}{2}}).
\end{aligned}$$

This corresponds to the independent case.

By integration of (2.3) we get the upper bound of the integrated covariance

$$\int_0^\infty C(x)dx \leq \int_0^\infty\left(2^{-\frac{\upsilon+3}{2}}\pi^{\frac{1-\upsilon}{2}}x^{-\frac{\upsilon+5}{2}}\frac{b^{-\frac{\upsilon+1}{2}}}{n}C_3(\upsilon,x)^{1-\upsilon} + o(b^2)\right)\int_1^\infty \alpha(\tau)^\upsilon d\tau dx. \qquad (4.11)$$

Combining (4.10) and (4.11), one can write

$$\begin{aligned}
MISE(f'(x)) \leq & \int_0^\infty \frac{n^{-1}b^{-3/2}x^{-3/2}}{4\sqrt{\pi}}\left(f(x) + \frac{b}{2}\left(\frac{f(x)}{x} - f'(x)\right)\right)dx \\
&+ \int_0^\infty 2^{-\frac{\upsilon+3}{2}}\pi^{\frac{1-\upsilon}{2}}x^{-\frac{\upsilon+5}{2}}\frac{b^{-\frac{\upsilon+1}{2}}}{n}C_3(\upsilon,x)^{1-\upsilon}dx\int_1^\infty \alpha(\tau)^\upsilon d\tau \\
&+ \frac{b^2}{16}\int_0^\infty P(x)dx + o(b^2 + n^{-1}b^{-\frac{5}{2}}).
\end{aligned}$$

The derivative of this expression in b leads to

$$\begin{aligned}
&\frac{b}{8}\int_0^\infty P(x)dx - \frac{3n^{-1}b^{-\frac{5}{2}}}{8\sqrt{\pi}}\int_0^\infty x^{-\frac{3}{2}}f(x)dx + \frac{n^{-1}b^{-\frac{3}{2}}}{16\sqrt{\pi}}\int_0^\infty x^{-\frac{3}{2}}\left(\frac{f(x)}{x} - f'(x)\right)dx \\
&- \int_0^\infty \frac{\upsilon+1}{2}2^{-\frac{\upsilon+3}{2}}\pi^{\frac{1-\upsilon}{2}}x^{-\frac{\upsilon+5}{2}}\frac{b^{-\frac{\upsilon+3}{2}}}{n}C_3(\upsilon,x)^{1-\upsilon}dx\int_1^\infty \alpha(\tau)^\upsilon d\tau = 0. \qquad (4.12)
\end{aligned}$$

Since $0 < \upsilon < 1$ holds as in Lemma 2.1, the third term in (4.12) by b has the worst rate

$$c_1 b^{-\frac{\upsilon+3}{2}} = O\left(b^{-\frac{3}{2}}\right),$$

where c_1 is a constant.

Neglecting terms with $b^{-3/2}$ and $b^{-\frac{\upsilon+3}{2}}$ in comparison to the term containing $b^{-5/2}$, we simplify the equation

$$\frac{b^{7/2}}{8}\int\limits_0^\infty P(x)dx - \frac{3n^{-1}}{8\sqrt{\pi}}\int\limits_0^\infty x^{-\frac{3}{2}} f(x)dx + o(b^{7/2}) = 0.$$

The optimal $b = o(n^{-2/7})$ is the same as in (1.10). Let us insert such b in (2.4)

$$\begin{aligned} MISE_{opt}(\hat{f}'(x)) &= \int_0^\infty \frac{P(x)n^{-\frac{4}{7}}}{16} T^{\frac{4}{7}} dx + \int\limits_0^\infty \frac{n^{-4/7}T^{-3/7}x^{-3/2}}{4\sqrt{\pi}} f(x)dx \\ &+ \int\limits_0^\infty \frac{n^{-6/7}T^{-1/7}x^{-3/2}}{8\sqrt{\pi}} \left(\frac{f(x)}{x} - f'(x)\right) dx \\ &+ \int\limits_0^\infty \Bigg(2^{-\frac{\upsilon+3}{2}} \pi^{\frac{1-\upsilon}{2}} x^{-\frac{\upsilon+5}{2}} \frac{T^{-\frac{\upsilon+1}{7}}}{n^{\frac{6-\upsilon}{7}}} C_3(\upsilon, x)^{1-\upsilon} dx \int\limits_1^\infty \alpha(\tau)^\upsilon d\tau, \end{aligned} \tag{4.13}$$

where

$$T = \frac{3\int_0^\infty x^{-3/2} f(x)dx}{\sqrt{\pi}\int_0^\infty \left(\frac{f(x)}{3x^2} + f''(x)\right)^2 dx}.$$

The last term in (4.13) has the rate $o(n^{\frac{\upsilon-6}{7}})$. By $0 < \upsilon < 1$ we get that the optimal rate of convergence of MISE is given by $MISE_{opt}(\hat{f}'(x)) = O(n^{-4/7})$. □

Proof of Lemma 2.2: We have to prove that $\alpha(\tau)$ defined by (2.6) satisfies the conditions of Lemma 2.1. Conditions 2 and 3 of Lemma 2.1 only refer to the density distribution. Thus, we remain to check only the first condition of Lemma 2.1.

To this end, using (2.6) we get

$$\begin{aligned} \int\limits_1^\infty \alpha(\tau)^\upsilon d\tau &\le \int\limits_1^{\tau_0} d\tau + \int\limits_{\tau_0}^\infty \left(2(C+1)\mathsf{E}|X_i|^\nu |\rho^\nu|^\tau\right)^\upsilon d\tau \\ &= \tau_0 - 1 + \left(2(C+1)\mathsf{E}|X_i|^\nu\right)^\upsilon \int\limits_{\tau_0}^\infty \left(|\rho^\nu|^\tau\right)^\upsilon d\tau. \end{aligned} \tag{4.14}$$

The integral in (4.14) can be taken in general as

$$\int_{\tau_0}^{\infty} (|\rho^{\nu}|^{\tau})^{\upsilon} \, d\tau = \frac{|\rho^{\nu}|^{\tau\upsilon}}{\upsilon \ln(|\rho^{\nu}|)}\Big|_{\tau_0}^{\infty}$$

Thus, to satisfy the first condition of Lemma 2.1, it must be

$$|\rho^{\nu}|^{\tau\upsilon}\Big|_{\tau=\infty} < \infty. \tag{4.15}$$

Since $\rho \in (-1, 1)$ holds, it follows $|\rho| \in [0, 1)$. For $\rho = 0$ (4.15) is satisfied. For $|\rho| \in (0, 1)$ one can rewrite (4.15) as

$$\left(\frac{1}{\xi}\right)^{\nu\tau\upsilon}\Big|_{\tau=\infty} < \infty, \quad \xi > 1,$$

which is valid as $\nu\upsilon > 0$. The latter is true since $0 < \upsilon < 1$ and $\nu = \min\{p, q, 1\} > 0$. Thus, the strong mixing AR(1) process (2.5) satisfies Lemma 2.1. Hence, it satisfies the conditions of Theorem 2.1. □

ACKNOWLEDGMENTS

I am grateful to my supervisor DrSci Alexander Dobrovidov for an interesting topic. The work was partly supported by the Russian Foundation for Basic Research grant 16-08-01285 A and the Russian Science Foundation grant 14-50-00150.

REFERENCES

[1] Aksoy, H. (2000). Use of Gamma distribution in hydrological analysis, *Turk. J. Engin. Environ. Sci.*, **24**, 419–428.

[2] Andrews, D.W.K. (1983). *First order autoregressive processes and strong mixing*, Yale University, New Haven, Connecticut.

[3] Bosq, S. (1996). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*, Springer, New York.

[4] Bouezmarnia, T. and Rombouts, J.V.K. (2007). Nonparametric density estimation for multivariate bounded data, *Journal of Statistical Planning and Inference*, **140**(1), 139–152.

[5] Bouezmarnia, T. and Rombouts, J.V.K. (2010). Nonparametric density estimation for positive times series, *Computational Statistics and Data Analysis*, **54**(2), 245–261.

[6] Bouezmarnia, T. and Scaillet, O. (2003). Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data, *Econometric Theory*, **21**, 390–412.

[7] Bhattacharya, P.K. (1967). Estimation of a probability density function and its derivatives, *The Indian Journal of Statistics*, **A 29**, 373–382.

[8] Song Xi Chen (2000). Probability density function estimation using gamma kernels, *Annals of the Institute of Statistical Mathematics*, **54**, 471–480.

[9] De Brabanter, K. and De Brabanter, J. and De Moor, B. (2011). *Nonparametric derivative estimation*, "Proc. of the 23rd Benelux Conference on Artificial Intelligence (BNAIC)", Gent, Belgium, 75–81.

[10] Dobrovidov, A.V. and Koshkin, G.M. and Vasiliev, V.A. (2012). *Non-Parametric State Space Models*, Kendrick press, USA.

[11] Dobrovidov, A.V. and Markovich, L.A. (2013). *Nonparametric gamma kernel estimators of density derivatives on positive semi-axis*, "Proc. of IFAC MIM 2013": Petersburg, Russia, June 19–21, 944–949.

[12] Dobrovidov, A.V. and Markovich, L.A. (2013). *Data-driven bandwidth choice for gamma kernel estimates of density derivatives on the positive semi-axis*, "Proc. of IFAC International Workshop on Adaptation and Learning in Control and Signal Processing Caen", France, 500–505.

[13] Furman, E. (2008). On a multivariate gamma distribution, *Statist. Probab. Lett.*, **78**, 2353–2360.

[14] Hall, P. and Wehrly, T.E. (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimators, *J. Amer. Statist. Assoc.*, **86**, 665–672.

[15] Hürlimann, W. (2001). Analytical evaluation of economic risk capital for portfolios of gamma risks, *ASTIN Bulletin*, **31**, 107–122.

[16] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**(1), 97–109.

[17] Jones, M.C. (1993). Simple boundary correction for density estimation kernel, *Statistics and Computing*, **3**, 135–146.

[18] Lejeune, M. and Sarda, P. (1992). Smooth estimators of distribution and density functions, *Computational Statistics and Data Analysis*, **14**, 457–471.

[19] Müller, H.G. (1991). Smooth optimum kernel estimators near endpoints, *Biometrika*, **78**(3), 521–530.

[20] Parzen, E. (1962). On estimation of a probability density function and mode, *The Annals of Mathematical Statistics*, **33**(3), 1065.

[21] Roberts, G.O.; Rosenthal, J.S.; Segers, J. and Sousa, B. (2007). Extremal indices, geometric ergodicity of Markov chains, and MCMC, *Extremes*, **9**(3-4), 213–229.

[22] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *The Annals of Mathematical Statistics*, **27**(3), 832.

[23] Sasaki, H.; Hyvärinen, A. and Sugiyama, M. (2014). *Clustering via mode seeking by direct estimation of the gradient of a log-density.* In "Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2014)", to appear.

[24] Scaillet, O. (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels, *Journal of Nonparametric Statistics*, **16**, 217–226.

[25] Schuster, E.F. (1985) Incorporating support constraints into nonparametric estimators of densities, *Commun. Statist. Theory Methods*, **14**, 1123–1136.

[26] Schuster, E.F. (1969) Estimation of a probability function and its derivatives, *Ann. Math. Statist.*, **40**, 1187–1195.

[27] Sköld, M. and Roberts, G.O. (2003). Density estimates from the Metropolis–Hastings Algorithm, *Scand. J. Stat.*, **30**, 699–718.

[28] Tsypkin, Ya.Z. (1985). Optimality in adaptive control systems. Uncertainty and Control, *Springer, Lecture Notes in Control and Information Sciences Berlin, Heidelberg*, **70**, 153–214.

[29] Turlach, B.A. (1993). Bandwidth Selection in Kernel Density Estimation: A Review, *CORE and Institut de Statistique*.

[30] Wand, M.P. and Jones, M.C. (1995). Kernel Smoothing, *Chapman and Hall*, London.

[31] Zhang, S. (2010). A note on the performance of the gamma kernel estimators at the boundary, *Statis. Probab. Lett.*, **80**, 548–557.

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 February, one in April, one in June and the other in October.

— On average, five articles will be published per issue

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.

Copyright

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, Statistics Portugal, in order to ensure the widest possible dissemination of information, namely through the Statistics Portugal website (http://www.ine.pt).

After assigning copyright, authors may use their own material in other publications provided that REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.